.

Statistical Inference via Convex Optimization

Anatoli Juditsky, Arkadi Nemirovski

Spring 2020

Source: A. Juditsky, A. Nemirovski *Statistical Inference via Convex Optimization*, Princeton University Press, April 2020 `https://www.isye.gatech.edu/~nemirovs/StatOptNoSolutions.pdf`

Preface

♣ **Fact:** Many inference procedures in Statistics reduce to optimization

♠ **Example: MLE – Maximum Likelihood Estimation**

Problem: *Given a parametric family* $\{p_\theta(\cdot) : \theta \in \Theta\}$ *of probability densities on* $\mathbb{R}^d$ *and a random observation* ω *drawn from some density* $p_{\theta_\star}(\cdot)$ *from the family, estimate the parameter* $\theta_\star$.

Maximum Likelihood Estimate: Given ω, maximize $p_\theta(\omega)$ over $\theta \in \Theta$ and use the maximizer $\widehat{\theta} = \widehat{\theta}(\omega)$ as an estimate of $\theta_\star$.

Note: In MLE, optimization is used for number crunching only and has nothing to do with motivation and performance analysis of MLE.

Fact: *Most of traditional applications of Optimization in Statistics are of "number crunching" nature. While often vitally important, "number crunching" applications are beyond our scope.*

♣ **What is in our scope,** are *inference routines motivated and justified by Optimization Theory* – Convex Analysis, Optimality Conditions, Duality...
As a matter of fact, *our "working horse" will be Convex Optimization.* This choice is motivated by

• *nice geometry* of *convex* sets, functions, and optimization problems

• *computational tractability* of convex optimization implying *computational efficiency* of statistical inferences stemming from Convex Optimization.

Major topics to be covered:

- Sparsity-Oriented Signal Processing
- Hypothesis Testing
- Signal Recovery from Indirect Observations in Linear and Generalized Linear Models

.

SPARSITY-ORIENTED SIGNAL PROCESSING

- Signal Recovery from Indirect Observations
- Sparse ℓ_1 Recovery: Motivation
- Validating ℓ_1 Recovery
    - s-Goodness and Nullspace Property
    - Quantifying Nullspace Property
    - Regular and Penalized ℓ_1 Recoveries
    - Restricted Isometry Property
    - Tractability Issues

Sparsity-oriented Signal Processing: Problem's Setting

♠ **Basic Signal Processing problem** is to recover *unknown* signal $x_* \in \mathbb{R}^n$ from its observation

$$y = A(x_*) + \xi$$

• $x \mapsto A(x) : \mathbb{R}^n \to \mathbb{R}^m$: *known* "signal-to-observation" transformation

• ξ: observation noise.

♣ In many applications, the signal-to-observation transformation is just *linear*:

$$A(x) = Ax \text{ for some known } m \times n \text{ matrix } A.$$

♠ Assume from now on that $A(\cdot)$ is linear

$\Rightarrow$ the recovery problem is just *to solve a system of linear equations*

$$Ax = b := Ax_*$$

given $m \times n$ matrix A and a noisy observation y of the "true" right hand side b.

♣ **Problem of interest:** *to solve a linear system*

$$Ax = b := Ax_*$$

given $m \times n$ matrix A and a noisy observation y of the "true" right hand side b.

♠ As of now, there are two typical settings of the problem:

• $m \geq n$ (typically, $m \gg n$) — we have (much) more observations than unknowns. This is the classical case studied in numerical Linear Algebra (where noise is non-random) and Statistics (where noise is random).

Unless A is "pathological," the only difficulty here is the presence of noise. The challenge is to reproduce well the true signal while suppressing as much as possible the influence of noise.

• $m < n$ (and even $m \ll n$) – we have (much) less observations than unknowns. Till early 2000's, this case was thought of as completely meaningless. Indeed, as Linear Algebra says, *an under-determined* (with more unknowns than equations) *system of linear equations either has no solutions at all, or has infinitely many solutions which can be arbitrarily far away from each other.*

$\Rightarrow$ *When $m < n$, the true signal cannot be recovered from observations even in the noiseless case!*

♠ **Remedy:** Add some information on the true signal.

♣ **Problem of interest:** *to solve a linear system*

$$Ax = b := Ax_*$$

given $m \times n$ matrix A and a noisy observation y of the "true" right hand side b in the case of $m \ll n$

♠ **Sparsity-oriented remedy [a.k.a.** *Compressed Sensing***]:** *Reduce the problem to the one where the signal is sparse – has $s \ll n$ nonzero entries, and utilize sparsity in your recovery routine.*

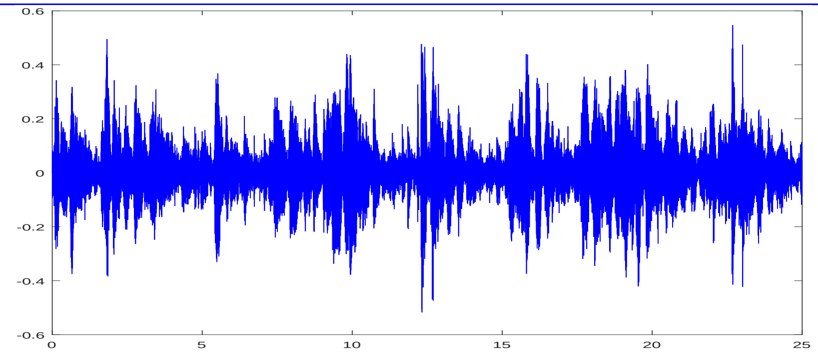
♠ **Fact:** *Many real-life signals x when presented by their coefficients in properly selected basis ("dictionary") B:*

$$x = Bu$$

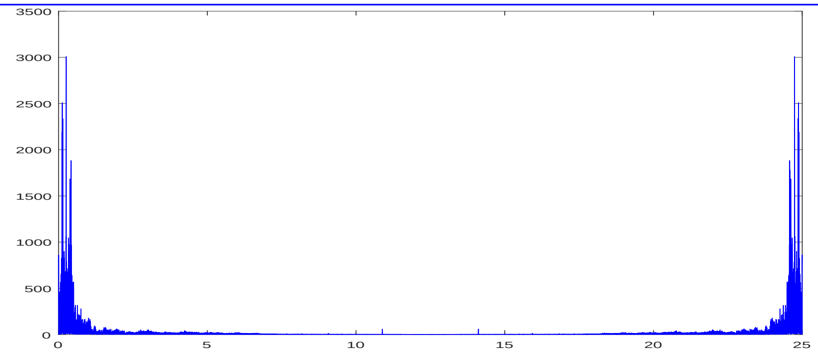
- columns of B: vectors of basis B
- u: coefficients of x in basis B

become sparse (or nearly so): u has just $s \ll n$ nonzero entries (or can be well approximated by vector with $s \ll n$ nonzero entries).

Illustration: 25 sec fragment of audio signal "Mail must go through" (dimension 1,058,400) and its Discrete Fourier Transform:



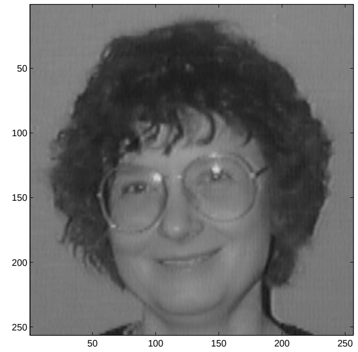
How mail goes through in time domain



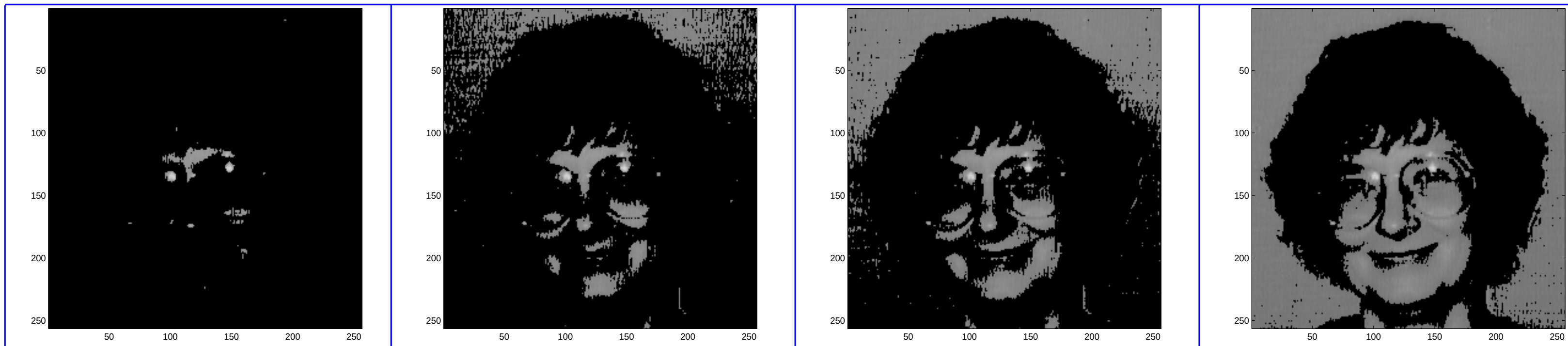
How mail goes through in frequency domain

% of leading Fourier coefficients kept	energy
100%	100%
25%	99.8%
15%	99.6%
5%	98.2%
1%	79.0%

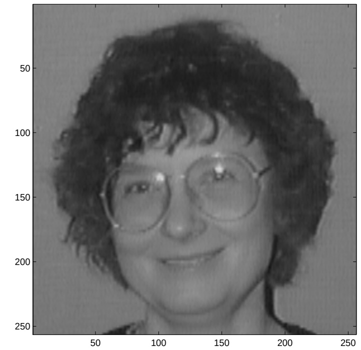
Illustration: The 256×256 image



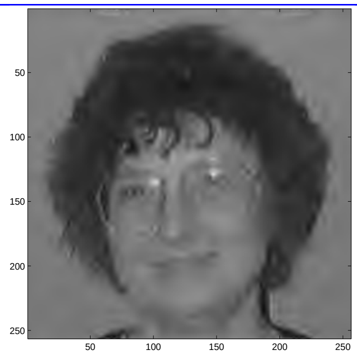
can be thought of as $256^2 = 65536$-dimensional vector (write down the intensities of pixels column by column). "As is," this vector is not sparse and cannot be approximated well by highly sparse vectors. This is what happens when we keep several leading (i.e., largest in magnitude) entries and zero out all other entries:



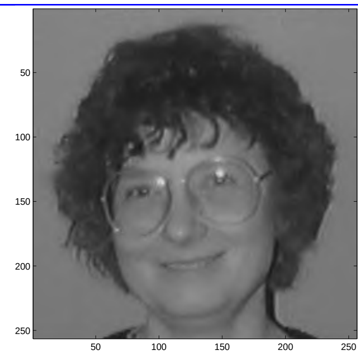
1% of leading entries kept	10% of leading entries kept	25% of leading entries kept	50% of leading entries kept



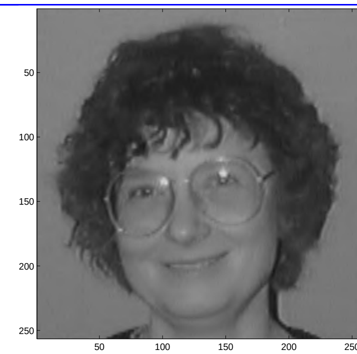
However, the image (same as other “non-pathological” images) is nearly sparse when represented in *wavelet* basis:



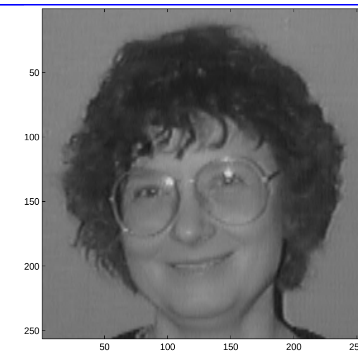
1% of leading wavelet coeff. (99.70% of energy) kept



5% of leading wavelet coeff. (99.93% of energy) kept

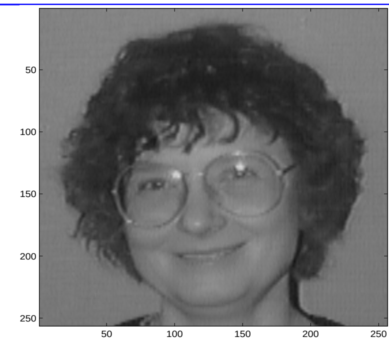


10% of leading wavelet coeff. (99.96% of energy) kept

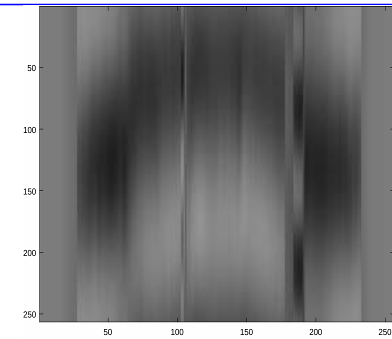


25% of leading wavelet coeff. (99.99% of energy) kept

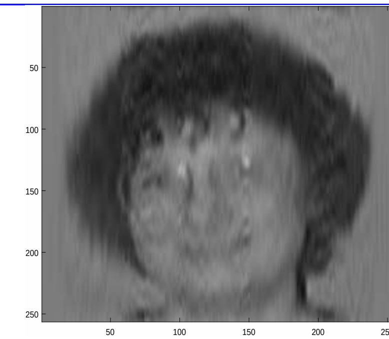
♠ Similar, albeit less intense, phenomenon takes place when representing typical images in frequency domain:



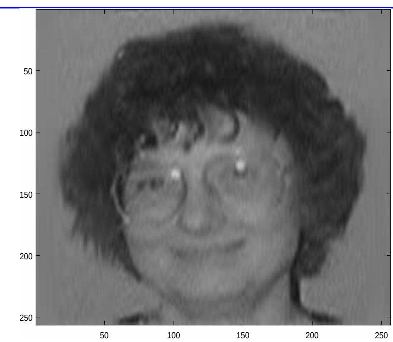
True image (100% of energy) kept



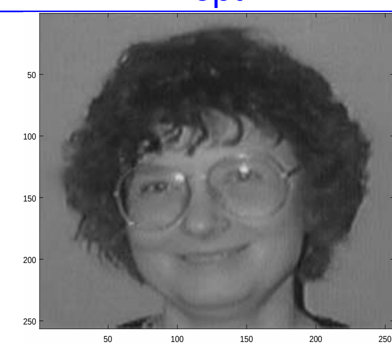
1% of leading Fourier coeff. (96.41% of energy) kept



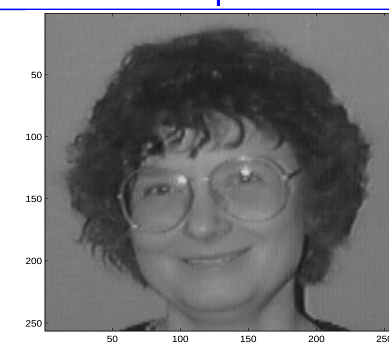
5% of leading Fourier coeff. (99.46% of energy) kept



10% of leading Fourier coeff. (99.76% of energy) kept



15% of leading Fourier coeff. (99.95% of energy) kept



25% of leading Fourier coeff. (99.99% of energy) kept

♠ When recovering a signal x_* admitting a sparse (or nearly so) representation Bu_* in a *known* basis B from observations

$$y = Ax_* + \xi,$$

the situation reduces to the one when the signal to be recovered is just sparse.

Indeed, we can first recover *sparse* u_* from observations

$$y = Ax_* + \xi = [AB]u_* + \xi.$$

After an estimate $\widehat{u}$ of u_* is built, we can estimate x_* by $B\widehat{u}$.

$\Rightarrow$ In fact, sparse recovery is about how to recover a *sparse* n-dimensional signal x from $m \ll n$ observations

$$y = Ax_* + \xi.$$

(?) How to recover a *sparse* (or nearly so) n-dimensional signal x from $m \ll n$ observations

$$y = Ax_* + \xi \text{ ?}$$

♠ To get an idea, consider the case when x_* is exactly sparse – has $s \ll n$ nonzero entries – and there is no observation noise:

$$y = Ax_*$$

• *If we knew the positions* $i_1, ..., i_s$ *of the nonzero entries in* x_*, we could recover x_* by solving the system *with just* s *unknowns:*

$$y = \left[A_{i_1}, ..., A_{i_s}\right] \cdot \left[x_{i_1}; ...; x_{i_s}\right]. \quad (!)$$

When $s \leq m$ (which, with $s \ll n$, still allows for $m \ll n$), we would get *over-determined* system of linear equations on the nonzero entries in x. Assuming A "non-pathologic," so that every $s \leq m$ columns of A are linearly independent, (!) has a unique solution which can be easily found.

But: *We never know in advance where the nonzeros in* x *are located!*

(?) How to recover a *sparse* n-dimensional signal x_* from $m \ll n$ observations

$$y = Ax_* \,?$$

♠ A straightforward way to account for the fact that we *never know where the nonzeros in x_* stand*, is to look for *the sparsest* solution to the system $y = Ax$. This amounts to solving the optimization problem

$$\min_x \mathrm{nnz}(x) \text{ s.t. } y = Ax \qquad (!)$$

- nnz(x): # of nonzero entries in x.

• It is easily seen that *if x_* is s-sparse and every $2s$ columns in A are linearly independent* (which is so when $2s \leq m$, unless A is pathological), *then x_* is the unique optimal solution to (!)*, and thus our procedure recovers x_* *exactly*.

But: nnz(z) is a bad (nonconvex and discontinuous) function, so that (!) is a disastrously complicated combinatorial problem. Seemingly, the only "theoretically solid" way to solve (!) is to use brute force search where we test one by one all collections of potential locations of nonzero entries in a solution. Brute force is completely unrealistic: to recover s-sparse signal, it would require looking through *at least*

$$N = \binom{n}{s-1} = \frac{n!}{(s-1)!(n-s+1)!}$$

candidate solutions.

• with $s = 17, n = 128$, N is as large as $1.49 \cdot 10^{21}$

• with $s = 49, n = 1024$, N is as large as $3.94 \cdot 10^{84}$

(?) How to recover a *sparse* n-dimensional signal x_* from $m \ll n$ observations

$$y = Ax_* \text{ ?}$$

• Solving problem

$$\min_x \text{nnz}(x) \text{ s.t. } y = Ax \qquad (!)$$

would yield the desired recovery, but (!) is heavily computationally intractable...

♠ **Partial remedy:** Replace the difficult to minimize objective $\text{nnz}(\theta)$ with an "easy-to-minimize" objective, specifically, with $\|\theta\|_1 = \sum_i |\theta_i|$, thus arriving at ℓ_1*-recovery*

$$\widehat{x} = \text{argmin}_x \left\{\sum_i |x_i| : Ax = y := Ax_*\right\} \quad (!!)$$

♠ **Observation:** (!!) is just an LO program!

Indeed,

• the constraints in (!!) are linear equalities.

• $|x_i| = \max[x_i, -x_i]$, so that the terms in the objective can be "linearized."

♠ The LO reformulation of (!!) is

$$\min_{x,z} \left\{\sum_j z_j : Ax = y, z_j \geq x_j, z_j \geq -x_j \, \forall j \leq n\right\}.$$

• *In the noiseless case*, ℓ_1 recovery is given by

$$\widehat{x} = \mathrm{argmin}_x \left\{\sum_i |x_i| : Ax = y := Ax_*\right\}$$

♠ When the observation y is noisy:

$$y = Ax_* + \xi$$

the constraint $Ax = y$ on a candidate recovery should be relaxed.

• *When we know an upper bound* δ on some norm $\|\xi\|$ of the noise ξ, a natural version of ℓ_1 recovery is

$$\widehat{x} \in \mathrm{Argmin}_x \left\{\sum_i |x_i| : \|Ax - y\| \le \delta\right\} \qquad (*)$$

Note: When $\|\xi\| = \|\xi\|_\infty := \max_i |\xi_i|$ ("uniform norm"), $(*)$ reduces to the LO program

$$\min_{x,z}\left\{\sum_j z_j : \begin{array}{l} -z_j \le x_j \le z_j,\ 1 \le j \le n \\ y_i - \delta \le [Ax]_i \le y_i + \delta, 1 \le i \le m \end{array}\right\}$$

• *When the noise* ξ *is random with zero mean*, there are reasons to define ℓ_1 recovery by *Dantzig Selector:*

$$\widehat{x} \in \mathrm{Argmin}_x \left\{\sum_i |x_i| : \|Q(Ax - y)\|_\infty \le \delta\right\}$$

with $M \times m$ *contrast matrix* Q and $\delta > 0$ chosen according to noise's structure and intensity. This again is reducible to LO program, specifically,

$$\min_{x,z}\left\{\sum_j z_j : \begin{array}{l} -z_j \le x_j \le z_j,\ 1 \le j \le n \\ -\delta \le [QAx - Qy]_i \le \delta, 1 \le i \le M \end{array}\right\}$$

• **Note:** In Dantzig Selector proper, $Q = A^T$.

(?) How to recover a *sparse* (or nearly so) n-dimensional signal x_* from $m \ll n$ observations

$$y = Ax_* + \xi \text{ ?}$$

(!) Use ℓ_1 minimization

$$\widehat{x} \in \mathrm{Argmin}_x \left\{\sum_i |x_i| : \|Ax - y\| \leq \delta\right\}$$

♣ Compressed Sensing theory shows that *under appropriate assumptions on A, in a meaningful range of sizes m, n and sparsities s, ℓ_1-minimization recovers the unknown signal x_**

— *exactly,* when x_* is s-sparse and there is no observation noise,

— *within inaccuracy* $\leq C(A)[\delta_n + \delta_s]$ in the general case

- δ_n: magnitude of noise
- δ_s: deviation of x_* from its best s-sparse approximation

♠ **Bad news:** "Appropriate assumptions on A" are *difficult to verify*

Partial remedy: there are conservative *verifiable* sufficient conditions for "appropriate assumptions."

♠ **Good news:** *For A drawn at random from natural distributions, "appropriate assumptions" are satisfied with overwhelming probability.*

• E.g., when entries in $m \times n$ matrix A are, independently of each other, sampled from Gaussian distribution, the resulting matrix, *with probability approaching 1 as m, n grow*, ensures the validity of ℓ_1 recovery of sparse signals with as many as

$$s = O(1)\frac{m}{\ln(n/m)}$$

nonzero entries.

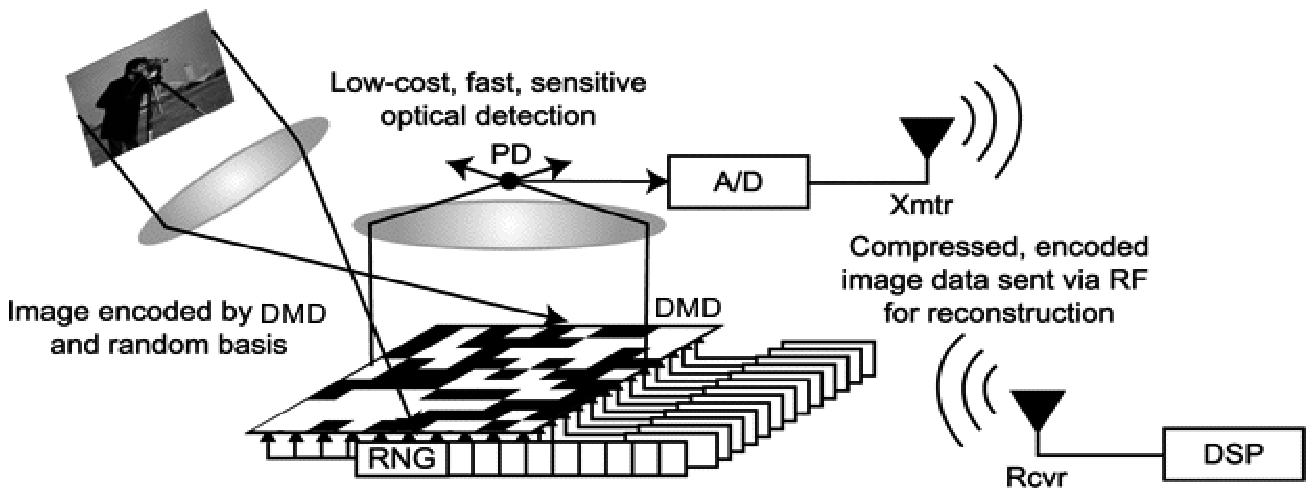
♠ **More good news:** In many applications (Imaging, Radars, Magnetic Resonance Tomography,...), signal acquisition via randomly generated matrices A makes perfect sense and results in significant acceleration of the acquisition process; see

David Donoho, Gauss Prize Lecture *"Compressed sensing – from blackboard to bedside"* (ICM2018), `https://www.youtube.com/watch?v=mr-oT5gMboM`

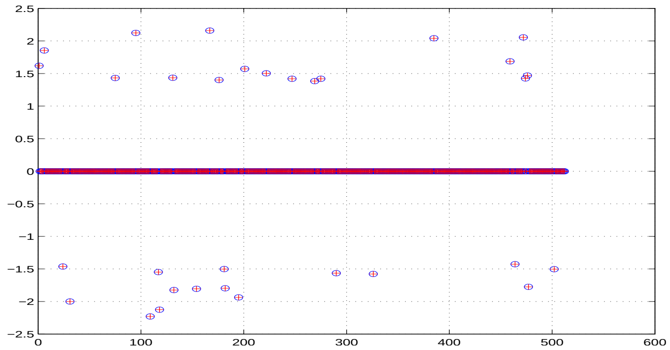
In these applications, signals of interest are sparse in properly selected bases

$\Rightarrow$ *With accelerated acquisition, no information is lost!*

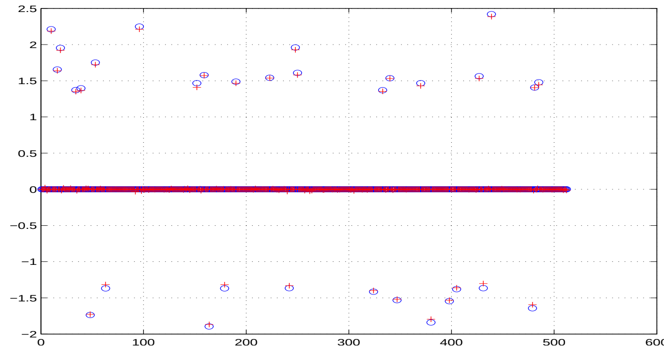
♠ Example: Single-Pixel Camera:



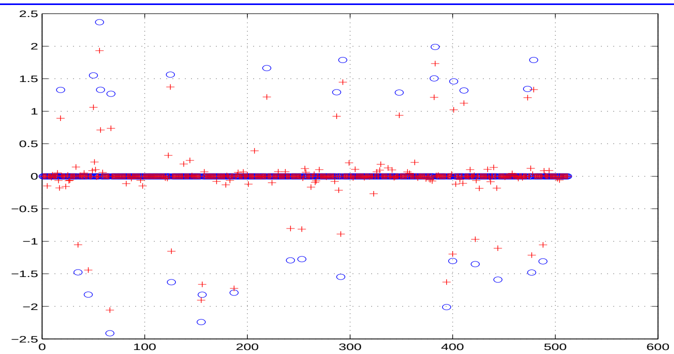
How it works:
Sparse recovery via Dantzig Selector



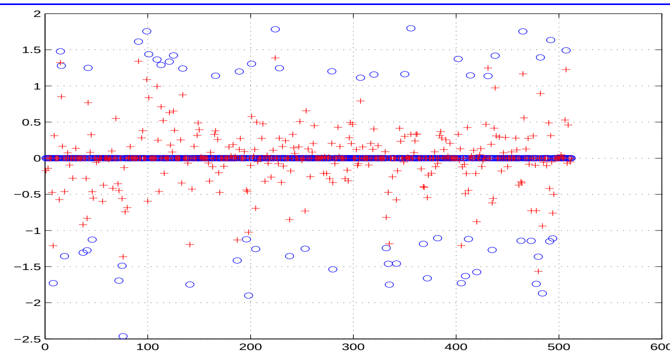
32 nonzero entries
No noise
Recovery error: $7.8 \cdot 10^{-9}$



32 nonzero entries
Noise's StD 0.01
Recovery error: 0.064



32 nonzero entries
Noise's StD 0.10
Recovery error: 0.66



64 nonzero entries
No noise
Recovery error: 1.47

o: signal $+$: recovery
256×512 Gaussian sensing matrix A

Validity of sparse signal recovery via ℓ_1 minimization

♠ **Notational convention:** From now on, for a vector $x \in \mathbb{R}^n$

• $I_x = \{j : x_j \neq 0\}$ is the *support* of x.

• for a subset I of the index set $\{1, ..., n\}$, x_I is the vector obtained from x by zeroing out entries with indexes *not* in I, and I^o is the complement of I:

$$I^o = \{i \in \{1, ..., n\} : i \notin I\}.$$

• for $s \leq n$, x^s is the vector obtained from x by zeroing our all but the s largest in magnitude entries.

x^s is the best s-sparse approximation of x in any one of the ℓ_p norms, $1 \leq p \leq \infty$.

• for $s \leq n$ and $p \in [1, \infty]$, we set

$$\|x\|_{s,p} = \|x^s\|_p.$$

Validity of ℓ_1 minimization in the noiseless case

♣ The minimal requirement on sensing matrix A which makes ℓ_1-minimization valid is to guarantee the correct recovery of *exactly* s-sparse signals in the *noiseless* case, and we start with investigating this property.

♠ **s-Goodness:** *An $m \times n$ sensing matrix A is called s-good, if whenever the true signal x underlying noiseless observations is s-sparse, this signal will be recovered exactly by ℓ_1-minimization.*

Equivalently: A is s-good, if

$$\begin{array}{l} \mathrm{nnz}(x_*) \leq s \\ \Rightarrow x_* \text{ is the unique optimal solution to} \\ \qquad \min_x\{\|x\|_1 : Ax = Ax_*\} \end{array}$$

♠ **Necessary and sufficient condition** for s-goodness is Nullspace Property:

$$\boxed{\begin{array}{c} \textit{For every } 0 \neq z \in \mathrm{Ker} A := \{z : Az = 0\} \textit{ it holds} \\ \|z\|_{s,1} < \frac{1}{2}\|z\|_1. \end{array}}$$

• Nullspace Property can be *derived* from LO Optimality Conditions, same as can be verified directly.

• **s-goodness $\Rightarrow$ Nullspace Property**:

Nullspace Property does *not* take place

$\Rightarrow \exists 0 \neq z \in \mathrm{Ker} A : \|z^s\|_1 \geq \frac{1}{2}\|z\|_1$

$\Rightarrow Az^s = A[z^s - z], \|z^s\|_1 \geq \|z^s - z\|_1$

$\Rightarrow$ s-sparse signal $x_* = z^s$ is not the unique optimal solution to $\min_x\{\|x\|_1 : Ax = Ax_*\}$ – contradiction

• **Nullspace Property $\Rightarrow$ s-goodness**: Let Nullspace Property take place and x_* be s-sparse, and let u be an optimal solution to $\min_x\{\|x\|_1 : Ax = Ax_*\}$.

Denoting by I the support of x_*, for $z = u - x_*$ we have $z \in \mathrm{Ker} A$ and

$$\begin{array}{rl}
 & z_I = u_I - [x_*]_I = u_I - x_* \ \& \ z_{I^o} = u_{I^o} \\
\Rightarrow & \|z_I\|_1 \geq \|x_*\|_1 - \|u_I\|_1 \ \& \ \|z_{I^o}\|_1 = \|u_{I^o}\|_1 \\
\Rightarrow & \|z_I\|_1 - \|z_{I^o}\|_1 \geq \|x_*\|_1 - \|u_I\|_1 - \|u_{I^o}\|_1 \\
 & = \|x_*\|_1 - \|u\|_1 \geq 0 \\
\Rightarrow & \|z_I\|_1 - \|z_{I^o}\|_1 \geq 0 \\
\Rightarrow & \|z\|_{s,1} \geq \|z_I\|_1 \geq \frac{1}{2}[\|z_I\|_1 + \|z_{I^o}\|_1] = \frac{1}{2}\|z\|_1 \\
\Rightarrow & z = 0
\end{array}$$

♣ **Questions to be addressed:**

♠ What happens when A is s-good, but ℓ_1 recovery is “imperfect,” e.g.

- x is not exactly s-sparse, and/or
- there is observation noise

♠ How to verify, given A and s, that A is s-good

Quantifying Nullspace Property and Imperfect ℓ_1 Recovery

♣ *In order to address the above questions, we need to "quantify" Nullspace Property.*

♠ Nullspace Property states that

$$\{z \in \mathrm{Ker}A \ \& \ \|z\|_1 = 1\} \Rightarrow \|z\|_{s,1} < 1/2\},$$

or, which is the same,

$$\exists \kappa < 1/2 : \|z\|_{s,1} \leq \kappa \|z\|_1 \ \forall z \in \mathrm{Ker}A \qquad (!)$$

♠ **Equivalent form** of *necessary and sufficient* condition (!) for s-goodness of $m \times n$ sensing matrix A reads:

$A \in \mathbb{R}^{m\times n}$ is s-good if and only if for some constant $\kappa < 1/2$ and some (and then any) norm $\|\cdot\|$ on $\mathbb{R}^m$ one has

$$\exists C < \infty : \|x\|_{s,1} \leq C\|Ax\| + \kappa \|x\|_1 \ \forall x \in \mathbb{R}^n \qquad (!!)$$

Indeed, (!!) clearly implies (!). Assume (!), and let $\bar{x}$ be $\|\cdot\|_1$-closest to x element of $\mathrm{Ker}A$, so that $\|x-\bar{x}\|_1 \leq c\|Ax\|$ with c independent of x. We have

$$\begin{aligned}\|x\|_{s,1} &\leq \|\bar{x}\|_{s,1} + \|x-\bar{x}\|_1 \leq \kappa\|\bar{x}\|_1 + \|x-\bar{x}\|_1 \\ &\leq \kappa\|x\|_1 + [1+\kappa]\|x-\bar{x}\|_1 \leq [1+\kappa]c\|Ax\| + \kappa\|x\|_1\end{aligned}$$

$$\exists C : \|x\|_{s,1} \leq C\|Ax\| + \kappa\|x\|_1 \ \forall x \in \mathbb{R}^n \qquad (!!)$$

♠ It makes sense to rewrite the latter condition in a more flexible form linking

- $m \times n$ sensing matrix A,
- sparsity level s,
- $m \times N$ *contrast matrix* H,
- *norm* $\|\cdot\|$ on $\mathbb{R}^N$,
- *condition's parameter* $q \in [1, \infty]$, and
- *parameter* $\kappa \in (0, 1/2)$

Condition $\mathbf{Q}_q(s, \kappa)$:

$$\boxed{\|x\|_{s,q} := \|x^s\|_q \leq s^{\frac{1}{q}}\|H^T Ax\| + \kappa s^{\frac{1}{q}-1}\|x\|_1 \ \forall x \in \mathbb{R}^n}$$

♠ We treat condition $\mathbf{Q}_q(s, \kappa)$ *as a condition on contrast matrix* H *and norm* $\|\cdot\|$.

♠ **Note:** A is s-good if and only if the Nullspace Property holds, or, which is the same, *if and only if the condition* $\mathbf{Q}_1(s, \kappa)$ *with some* $\kappa < 1/2$ *is satisfiable* (e.g., with $N = n$, $H = CA^T$ with properly selected C, and $\|\cdot\| = \|\cdot\|_\infty$).

Condition $\mathbf{Q}_q(s,\kappa)$:

$$\boxed{\|x\|_{s,q} := \|x^s\|_q \leq s^{\frac{1}{q}}\|H^TAx\| + \kappa s^{\frac{1}{q}-1}\|x\|_1 \; \forall x \in \mathbb{R}^n}$$

♠ **Immediate observations:**

• *The larger is q, the stronger is $\mathbf{Q}_q(s,\kappa)$: If $H, \|\cdot\|$ satisfy $\mathbf{Q}_q(s,\kappa)$ and $p \in [1,q]$, then $H, \|\cdot\|$ satisfy $\mathbf{Q}_p(s,\kappa)$.*

Indeed, if $H, \|\cdot\|$ satisfy $\mathbf{Q}_q(s,\kappa)$ and $1 \leq p \leq q$, then

$$\begin{aligned} \|x\|_{s,p} &\leq \|x\|_{s,q} s^{\frac{1}{p}-\frac{1}{q}} \leq s^{\frac{1}{p}-\frac{1}{q}}\left[s^{\frac{1}{q}}\|H^TAx\| + \kappa s^{\frac{1}{q}-1}\|x\|_1\right] \\ &= s^{\frac{1}{p}}\|H^TAx\| + \kappa s^{\frac{1}{p}-1}\|x\|_1. \end{aligned}$$

• *Satisfiability of the weakest condition $\mathbf{Q}_1(s,\kappa)$ for some $\kappa < 1/2$ is necessary and sufficient for s-goodness of A.*

♠ **Fact:** Conditions $\mathbf{Q}_q(s,\kappa)$ underly instructive bounds on recovery error for imperfect ℓ_1 recovery.

Example A: Regular ℓ_1-Recovery

♣ **Regular ℓ_1 recovery** of signal x from observations

$$y = Ax + \eta$$

is given by

$$\widehat{x}_{\mathrm{reg}}(y) \in \mathop{\mathrm{Argmin}}_u \left\{ \|u\|_1 : \|H^T(Au - y)\| \le \rho \right\}$$

where H, $\|\cdot\|$, $\rho \ge 0$ are construction's parameters.

♠ **Theorem.** *Let s be a positive integer, $q \in [1, \infty]$ and $\kappa \in (0, 1/2)$. Assume that $H, \|\cdot\|$ satisfy $\mathbf{Q}_q(s, \kappa)$, and let*

$$\Xi_\rho = \{\eta : \|H^T \eta\| \le \rho\}.$$

Then for all $x \in \mathbb{R}^n$ and $\eta \in \Xi_\rho$ one has

$$\|\widehat{x}_{\mathrm{reg}}(Ax + \eta) - x\|_p \le \frac{4(2s)^{\frac{1}{p}}}{1 - 2\kappa} \left[\rho + \frac{\|x - x^s\|_1}{2s} \right], \; 1 \le p \le q.$$

Note: Regular ℓ_1 recovery requires a priori information on noise needed to select ρ with "meaningful" Ξ_ρ and does *not* require a priori information on sparsity s.

$$\boxed{\begin{array}{l}\forall \eta \in \Xi_\rho = \{\eta : \|H^T\eta\| \leq \rho\}\ \forall x : \\ \|\widehat{x}_{\rm reg}(Ax+\eta) - x\|_p \leq \frac{4(2s)^{\frac{1}{p}}}{1-2\kappa}\left[\rho + \frac{\|x-x^s\|_1}{2s}\right],\ 1 \leq p \leq q.\end{array}}$$

♠ **Comments:**

A. ρ stems from observation errors:

- $\eta \equiv 0 \Rightarrow$ we can set $\rho = 0$, resulting in zero recovering error for *exactly* s*-sparse* signals
- η is "uncertain but bounded" : $\eta \in \mathcal{U}$ for some known and bounded $\mathcal{U}$

$\Rightarrow$ we can set $\rho = \max_{u \in \mathcal{U}} \|H^T u\|$

- $\eta \sim \mathcal{N}(0, \sigma^2 I_m) \Rightarrow$ given tolerance β and setting

$$\rho = \sigma\sqrt{2\ln(N/\beta)}\max_i \|\mathrm{Col}_i[H]\|_2$$

we get

$$\mathrm{Prob}\{\eta : \|H^T\eta\|_\infty \leq \rho\} \geq 1 - \beta$$

When $\|\cdot\| = \|\cdot\|_\infty$, this allows to build explicitly "confidence domains" for regular ℓ_1 recovery.

B. Pay attention to the factor s^{-1} at the "near-sparsity" term $\|x - x^s\|_1$.

C. Adjusting H and $\|\cdot\|$, we can, to some extent, account for the nature of observation errors.

Example B: Penalized ℓ_1 Recovery

Penalized ℓ_1 recovery of signal x from observations

$$y = Ax + \eta$$

is given by

$$\widehat{x}_{\text{pen}}(y) \in \operatorname*{Argmin}_u \left\{ \|u\|_1 + \lambda \|H^T(Au - y)\| \right\}$$

where H, $\|\cdot\|$, $\lambda > 0$ are construction's parameters.

♠ **Theorem.** *Given A, positive integer s, and $q \in [1,\infty]$, assume that $H, \|\cdot\|$ satisfy $\mathbf{Q}_q(s,\kappa)$ with $\kappa < 1/2$, and let $\lambda \geq 2s$. Then for all $\eta \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$, for $1 \leq p \leq q$ it holds*

$$\|\widehat{x}_{\text{pen}}(Ax+\eta) - x\|_p \leq \frac{4\lambda^{\frac{1}{p}}\left[\frac{1}{2} + \frac{\lambda}{4s}\right]}{1-2\kappa}\left[\|H^T\eta\| + \frac{\|x - x^s\|_1}{2s}\right].$$

In particular, with $\lambda = 2s$, for $1 \leq p \leq q$ it holds

$$\|\widehat{x}_{\text{pen}}(Ax+\eta) - x\|_p \leq \frac{4(2s)^{\frac{1}{p}}}{1-2\kappa}\left[\|H^T\eta\| + \frac{\|x - x^s\|_1}{2s}\right].$$

Note: Penalized ℓ_1 recovery requires a priori knowledge of sparsity level s and does *not* require any information on noise.

Note: *When $\lambda = 2s$, for all x it holds*

$$\begin{array}{l} \forall (\rho \geq 0, \eta \in \Xi_\rho := \{\eta : \|H^T\eta\| \leq \rho\}) : \\ \|\widehat{x}_{\text{pen}}(Ax+\eta) - x\|_p \leq \frac{4(2s)^{\frac{1}{p}}}{1-2\kappa}\left[\rho + \frac{\|x-x^s\|_1}{2s}\right],\ 1 \leq p \leq q. \end{array}$$

$$\boxed{\begin{array}{c} H, \|\cdot\| \text{ satisfy } \mathbf{Q}_q(s,\kappa) \\ y = Ax + \eta, \eta \sim \mathcal{N}(0, \sigma^2 I_N) \\ x \in \mathbb{R}^n \text{ is } s\text{-sparse} \end{array}}$$

$$\Downarrow$$

$$\text{Prob}\left\{\|\widehat{x}_{\text{reg}}(Ax+\eta) - x\|_p \leq C(H,\kappa,\ln(1/\epsilon))\sigma s^{\frac{1}{p}}\right\} \geq 1-\epsilon$$

$$\text{Prob}\left\{\|\widehat{x}_{\text{pen}}(Ax+\eta) - x\|_p \leq C(H,\kappa,\ln(1/\epsilon))\sigma s^{\frac{1}{p}}\right\} \geq 1-\epsilon$$

$$1 \leq p \leq q$$

Note: Given *direct observations* $y = x + \eta$ of *s-dimensional* signal x with $\eta \sim \mathcal{N}(0, \sigma^2 I_s)$, the expected $\|\cdot\|_p$-norm of recovery error in optimal recovery is $O(1)\sigma s^{\frac{1}{p}}$.

How it works:
Regular vs. Penalized ℓ_1 Recovery

Problem: Given noisy observations of $m = n/2$ of randomly selected entries in time series $z = (z_1, ..., z_n)$ with nearly s-sparse Discrete Cosine Transform (DCT), we want to recover the time series.

Model: Treating as the signal x underlying observations the DCT of z and assuming for the sake of definiteness the observation noise to be white Gaussian, our observation becomes

$$y = Ax + \sigma\xi, \qquad [\xi \sim \mathcal{N}(0, I_m)]$$

where A is the $m \times n$ submatrix of the matrix F of Inverse DCT with rows indexed by the observed entries in z. Applying ℓ_1 minimization, we convert y into an estimate $\widehat{x}$ of x, and take $F\widehat{x}$ as the estimate of time series z.

Experiment: • $m = 256$, $n = 2m = 512$;

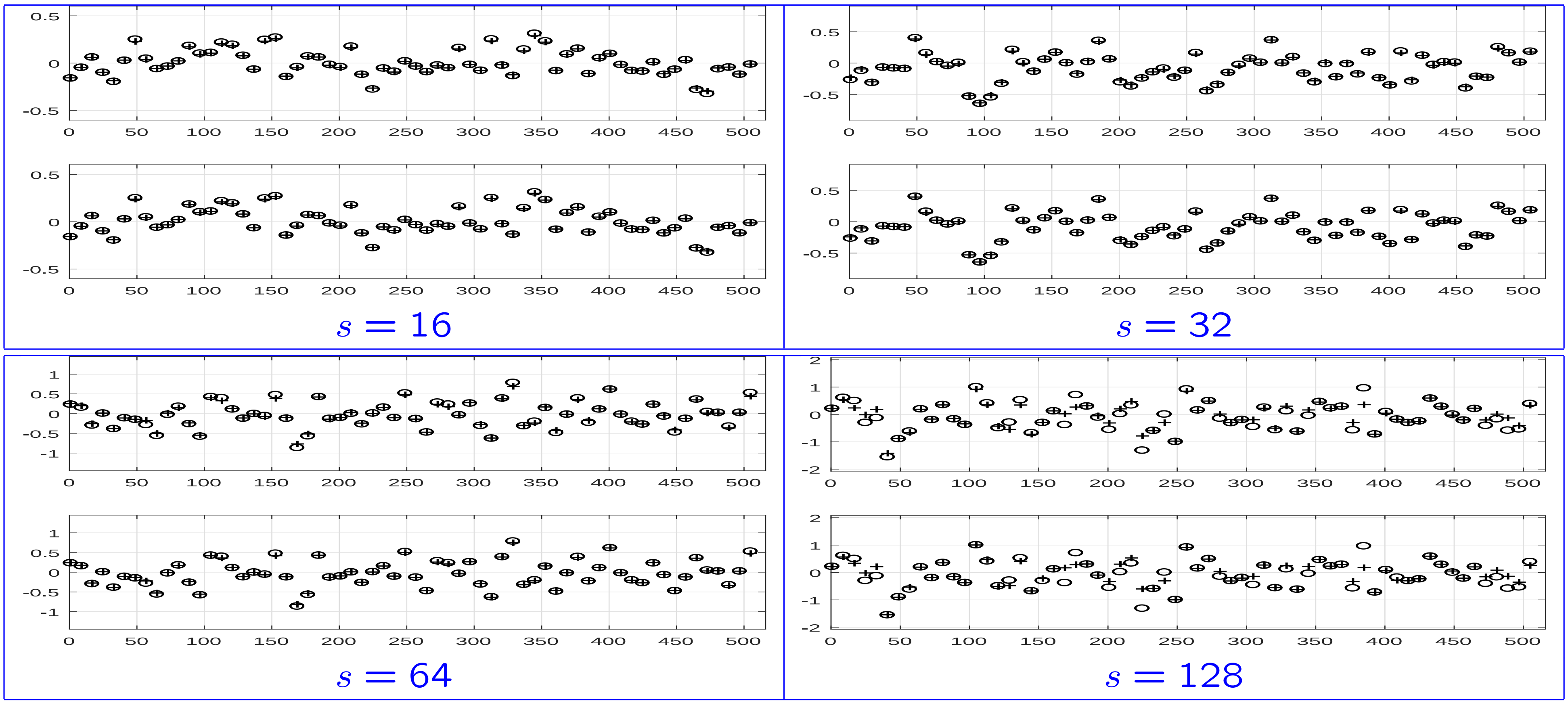
• $\sigma = 0.01$;

• near s-sparsity: $\|x - x^s\|_1 \leq 1$;

• contrast pair is $(H = \sqrt{n/m}A, \|\cdot\|_\infty)$;

• parameter ρ of regular recovery ensures $\mathrm{Prob}_{\zeta\sim\mathcal{N}(0,\sigma^2)}\{|\zeta| > \rho\} = 0.01/n$;

• in penalized recovery, $\lambda = 2s$.



	$s=16$	$s=32$	$s=64$	$s=128$
$\|z-\widehat{z}\|_2$	0.2417	0.3871	0.8178	4.8256
$\|z-\widehat{z}\|_\infty$	0.0343	0.0514	0.1744	0.8272

recovery errors, regular recovery

	$s=16$	$s=32$	$s=64$	$s=128$
$\|z-\widehat{z}\|_2$	0.1399	0.2385	0.4216	5.3431
$\|z-\widehat{z}\|_\infty$	0.0177	0.0362	0.1023	0.9141

recovery errors, penalized recovery

Top plots: regular ℓ_1 recovery, bottom plots: penalized ℓ_1 recovery

o: true signal +: recovery

[to make plots readable, every 8-th entry in time series is displayed]

Note: the actual level of s-goodness of A is at most 24!

How to Verify Validity Conditions for ℓ_1-Recovery ?

♣ **Bad news:** *Given A and s, the Nullspace Property is difficult to verify. Similarly, when $q < \infty$ and $\kappa < 1/2$, it is difficult to verify whether the condition $\mathbf{Q}_q(s,\kappa)$ is satisfied by given $H, \|\cdot\|$, same as it is difficult to verify whether the condition is satisfiable at all.*

♠ **Relatively good news:** There are natural ensembles of *random* sensing matrices for which properly selected $H, \|\cdot\|$ *with overwhelming probability* satisfy $\mathbf{Q}_2(s,\kappa)$ and thus are s-good.

♣ **Definition.** *An $m \times n$ sensing matrix A satisfies Restricted Isometry Property* RIP(δ, k), *if*

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \ \forall (x : \mathrm{nnz}(x) \leq k).$$

♠ **Theorem** *Let $m \times n$ sensing matrix A satisfy* RIP$(\delta, 2s)$ *for some $\delta < 1/3$ and positive integer s. Then*

- *The pair $\left(H = \frac{s^{-1/2}}{\sqrt{1-\delta}} I_m, \|\cdot\|_2\right)$ satisfies the condition $\mathbf{Q}_2\left(s, \frac{\delta}{1-\delta}\right)$;*
- *The pair $(H = \frac{1}{1-\delta} A, \|\cdot\|_\infty)$ satisfies the condition $\mathbf{Q}_2\left(s, \frac{\delta}{1-\delta}\right)$.*

♠ **Theorem** *Given* $\delta \in (0, \frac{1}{5}]$*, with properly selected positive* $c = c(\delta)$, $d = d(\delta)$, $f = f(\delta)$ *for all* $m \leq n$ *and all positive integers* k *such that*

$$k \leq \frac{m}{c \ln(n/m) + d}$$

the probability for a random $m \times n$ *matrix* A *with independent* $\mathcal{N}(0, \frac{1}{m})$ *entries to satisfy* $\mathrm{RIP}(\delta, k)$ *is at least*

$$1 - \exp\{-fm\}.$$

Similar result holds true for *Rademacher matrices* – those with i.i.d. entries taking values $\pm 1/\sqrt{m}$ with probabilities 0.5.
Note: k can be "nearly" as large as m !

Sketch of the proof

♠ Let A be Gaussian random $m \times n$ matrix from Theorem, $I \subset \{1, ..., n\}$ be fixed k-element index set, and $A_I = [A_{ij} : i \le m, j \in I]$. Let us fix $\alpha \in (0, 0.1]$.

Fact: *For fixed $u \in \mathbb{R}^k$ with $\|u\|_2 = 1$ one has*

$$\mathrm{Prob}\{A : \|A_I u\|_2^2 \notin [1-\alpha, 1+\alpha]\} \le 2\mathrm{e}^{-\frac{m}{5}\alpha^2}.$$

[observe that $A_I u \sim \mathcal{N}(0, \frac{1}{m} I_m)$ and use standard bounds on the tails of χ^2-distribution]

$\Rightarrow$ *Let Γ be α-net on the unit sphere S_k in $\mathbb{R}^k$. Then*

$$\mathrm{Prob}\{A : \exists u \in S_k : \|A_I u\|_2^2 \notin [1-4\alpha, 1+4\alpha]\} \le \pi := 2|\Gamma|\mathrm{e}^{-\frac{m}{5}\alpha^2}$$

[By Fact, $\mathrm{Prob}\underbrace{\{A : \|A_I u\|_2^2 \in [1-\alpha, 1+\alpha]\ \forall u \in \Gamma\}}_{\mathcal{E}} \ge 1 - \pi$. Since the quadratic form $f(u) := u^T A_I^T A_I u$ is Lipschitz continuous on S_k with constant $2M := 2\max_{u \in S_k} \underbrace{\|A_I u\|_2^2}_{f(u)}$, we have

$$A \in \mathcal{E} \Rightarrow \left\{ \begin{array}{l} \min_{u \in S_k} f(u) \ge \min_{u \in \Gamma} f(u) - 2\alpha M \ge 1 - \alpha - 2\alpha M \\ M = \max_{u \in S_k} f(u) \le \max_{u \in \Gamma} f(u) + 2\alpha M \le 1 + \alpha + 2\alpha M \end{array} \right.,$$

and the conclusion follows.]

$\Rightarrow \forall (I, |I| = k)$:

$$\mathrm{Prob}\{A : (1-4\alpha) I_k \preceq A_I^T A_I \preceq (1+4\alpha) I_k\} \ge 1 - 2\underbrace{[1 + 2/\alpha]^k}_{\mathcal{F}} \mathrm{e}^{-\frac{m}{5}\alpha^2}$$

[Comparing volumes, the cardinality of a minimal α-net on S_k is $\le \mathcal{F}$]

$\Rightarrow \mathrm{Prob}\{A : A$ is *not* $\mathrm{RIP}(4\alpha, k)\} \le \binom{n}{k} [1 + 2/\alpha]^k \mathrm{e}^{-\frac{m}{5}\alpha^2}$

$\Rightarrow$ Theorem.

♠ **Bad news:** No (series of) explicitly computable (even by a randomized computation) $\mathrm{RIP}(0.1, k)$ "low" ($2m \leq n$) $m \times n$ matrices with "large" k (namely, $k \gg \sqrt{m}$) are known.

♡ The natural idea – "generate at random a low $m \times n$ matrix and check whether it satisfies $\mathrm{RIP}(0.1, k)$" with "large" k; if yes, output the matrix" – fails: while typical random matrices do possess $\mathrm{RIP}(0.1, k)$ with "large" k, we do *not* know how to verify this property in a computationally efficient fashion.

♡ Designing/checking RIP matrices is similar to other situations where we do know that a typical randomly selected object possesses some property, but we neither can point out an individual object with this property, nor can check efficiently whether a given object possesses it. Some examples:

• **Complexity of Boolean functions** [Shannon, 1949]: *For a Boolean function f of n Boolean variables, the minimal number of AND, OR, NOT switches in a circuit computing the function is upper-bounded by $O(1)\frac{2^n}{n}$, and as n grows, this bound becomes sharp with overwhelming probability.*

<u>However:</u> No individual functions with nonlinear "Boolean complexity" are known...

• **Lindenstrauss-Johnson Theorem** *For a Gaussian "low" $m \times n$ matrix A, the image $\{Ax : x \in B_n\}$ of the unit n-dimensional box $B_n = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$ under the mapping $x \mapsto Ax$ with overwhelming, as $n \to \infty$, probability is in-between two similar ellipsoids with the ratio of linear sizes not exceeding* $1 + O(1)(n/m)^2$.

<u>However:</u> No individual matrices A with AB_n reasonably close to an ellipsoid are known...

Note: For every $\epsilon \in (0,1)$ and every n, one can *explicitly* point out a polytope P given by $O(1)n\ln(1/\epsilon)$ linear inequalities on $O(1)n\ln(1/\epsilon)$ variables such that the projection of P onto the plane of the first n variables is in-between $\{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ and $\{x \in \mathbb{R}^n : \|x\|_2 \leq 1 + \epsilon\}$. However, this "fast polyhedral approximation" of Euclidean ball deals with polytopes P quite different from boxes...

♠ We have seen that RIP-matrices A yield easy-to-satisfy condition $\mathbf{Q}_2(s,\kappa)$. Unfortunately, RIP is difficult to verify...

♠ **Good news:** *Condition* $\mathbf{Q}_\infty(s,\kappa)$ *is fully computationally tractable.*

♠ **Theorem** *Let A be an $m\times n$ sensing matrix, s be a sparsity level, and $\kappa\geq 0$. Whenever $\bar{H},\|\cdot\|$ satisfy $\mathbf{Q}_\infty(s,\kappa)$, there exists an $m\times n$ matrix H such that*

$$\|\mathrm{Col}_j[I_n-H^TA]\|_\infty\leq s^{-1}\kappa,\ 1\leq j\leq n.$$

As a result, $H,\|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty(s,\kappa)$. Besides this,

$$\|H^T\eta\|_\infty\leq\|\bar{H}^T\eta\|\ \forall\eta\in\mathbb{R}^m.$$

In addition, $m\times n$ contrast matrix H such that $H,\|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty(s,\kappa)$ with as small κ as possible can be found as follows: we consider n LP programs

$$\mathrm{Opt}_i=\min_{\nu,h}\left\{\nu:\|A^Th-e^i\|_\infty\leq\nu\right\},\qquad (\#_i)$$

where e^i is i-th basic orth in $\mathbb{R}^n$, find optimal solutions Opt_i,h_i to these problems, and make h_i, $i=1,...,n$, the columns of H; the corresponding value of κ is

$$\kappa_*=s\max_i\mathrm{Opt}_i.$$

Finally, there exists a transparent alternative description of the quantities Opt_i (and thus – of κ_*):

$$\mathrm{Opt}_i=\max_x\left\{x_i:\|x\|_1\leq 1,Ax=0\right\}.$$

Let A be an $m\times n$ sensing matrix, s be a sparsity level, and $\kappa\geq 0$. Whenever $\bar{H}$, $\|\cdot\|$ satisfy $\mathbf{Q}_\infty(s,\kappa)$, there exists an $m\times n$ matrix H such that

$$\|\mathrm{Col}_j[I_n-H^TA]\|_\infty\leq s^{-1}\kappa,\ 1\leq j\leq n,$$

As a result, H, $\|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty(s,\kappa)$. Besides this,

$$\|H^T\eta\|_\infty\leq\|\bar{H}^T\eta\|\ \forall\eta\in\mathbb{R}^m.$$

Proof uses **Basic fact of Convex Geometry:** A norm $\|\cdot\|$ on $\mathbb{R}^N$ induces the *conjugate* norm

$$\|f\|_*=\max_{h:\|h\|\leq 1}f^Th.$$

One always has $|f^Th|\leq\|f\|_*\|h\|$ & $\|h\|=\max\limits_{f:\|f\|_*\leq 1}f^Th$

Now,

$$\begin{array}{l}
i\leq n\\
\Rightarrow x_i\leq\|x\|_{s,\infty}\leq\|\bar{H}^TAx\|+s^{-1}\kappa\|x\|_1\ \forall x\ \text{[by } Q_\infty(s,\kappa)]\\
\Rightarrow \max\limits_x\left\{x_i-\|\bar{H}^TAx\|:\|x\|_1\leq 1\right\}\leq s^{-1}\kappa\\
\Leftrightarrow \max\limits_{x:\|x\|_1\leq 1}\ \min\limits_{f\in\mathbb{R}^N,\|f\|_*\leq 1}\left[[e^i]^Tx-f^T\bar{H}^TAx\right]\leq s^{-1}\kappa\ \text{[since } \|\bar{H}^TAx\|=\max_{f:\|f\|_*\leq 1}f^T\bar{H}^TAx]\\
\Leftrightarrow \min\limits_{f:\|f\|_*\leq 1}\underbrace{\max\limits_{x:\|x\|_1\leq 1}\left[[e^i-A^T\bar{H}f]^Tx\right]}_{=\|e^i-A^T\bar{H}f\|_\infty}\leq s^{-1}\kappa\\
\Leftrightarrow \forall i\leq n\,\exists f_i\in\mathbb{R}^N:\|e^i-A^T\bar{H}f_i\|_\infty\leq s^{-1}\kappa\ \&\ \|f_i\|_*\leq 1.
\end{array}$$

$$\forall i \leq n \exists f_i \in \mathbb{R}^N : \|e^i - A^T \bar{H} f_i\|_\infty \leq \kappa \;\&\; \|f_i\|_* \leq 1.$$

Let $h_i = \bar{H} f_i$ and $H = [h_1, ..., h_n]$. Then

$$\begin{array}{l}
[I_n - H^T A]_{ij} = [I_n - A^T H]_{ji} = [e^i - A^T h_i]_j = [e^i - A^T \bar{H} f_i]_j \\
\Rightarrow \max_{i,j} |[I_n - H^T A]_{ij}| \leq \max\limits_i \max\limits_j |[e^i - A^T \bar{H} f_i]_j| \\
\qquad\qquad\qquad\qquad\qquad\quad \leq \max\limits_i \|e^i - A^T \bar{H} f_i\|_\infty \leq s^{-1}\kappa \\
\Rightarrow \|\mathrm{Col}_i[I_n - H^T A]\|_\infty \leq s^{-1}\kappa \,\forall i
\end{array}$$

Further,

$$\begin{array}{l}
\|\mathrm{Col}_i[I_n - H^T A]\|_\infty \leq s^{-1}\kappa \,\forall i \\
\Rightarrow \|[I_n - H^T A]x\|_\infty \leq s^{-1}\kappa \|x\|_1 \,\forall x \in \mathbb{R}^n \\
\Rightarrow \|x\|_\infty - \|H^T A x\|_\infty \leq s^{-1}\kappa \|x\|_1 \,\forall x \in \mathbb{R}^n \\
\Rightarrow H, \|\cdot\|_\infty \text{ satisfy } \mathbf{Q}_\infty(s, \kappa)
\end{array}$$

In addition,

$$\begin{array}{l}
\|H^T \eta\|_\infty = \max_i |h_i^T \eta| = \max_i |f_i^T \bar{H} \eta| \leq \max_i \|f_i\|_* \|\bar{H}^T \eta\| \\
\leq \|\bar{H}^T \eta\| \;\forall \eta.
\end{array}$$

... In addition, $m \times n$ contrast matrix H such that $H, \|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty(s,\kappa)$ with as small κ as possible can be found as follows: we consider n LP programs

$$\mathrm{Opt}_i = \min_{\nu,h}\left\{\nu : \|A^T h - e^i\|_\infty \le \nu\right\}, \qquad (\#_i)$$

where e^i is i-th basic orth in $\mathbb{R}^n$, find optimal solutions Opt_i, h_i to these problems, and make h_i, $i = 1, ..., n$, the columns of H; the corresponding value of κ is $\kappa_* = s\max_i \mathrm{Opt}_i$.

Proof: By the above reasoning, if $H, \|\cdot\|$ satisfy $\mathbf{Q}_\infty(s,\kappa)$, then $\forall (i \le n) \exists h_i : \|e^i - A^T h_i\|_\infty \le s^{-1}\kappa$, and if h_i, $i \le n$, satisfy $\|e^i - A^T h_i\|_\infty \le s^{-1}\kappa$ for some κ, then $H := [h_1, ..., h_n], \|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty(s,\kappa)$.

... Finally, there exists a transparent alternative description of the quantities Opt_i (and thus – of κ_*);

$$\mathrm{Opt}_i = \max_x \left\{x_i : \|x\|_1 \le 1, Ax = 0\right\}.$$

Proof:

$$\begin{array}{l}
\mathrm{Opt}_i = \min\left\{t : -t \le e^i_j - [A^T h]_j \le t, \forall j\right\}\\
= \max_{\lambda,\mu}\left\{[\lambda - \mu]_i : \begin{array}{c} A^T[\lambda-\mu] = 0\\ \sum_i \lambda_i + \sum_i \mu_i = 1\\ \lambda \ge 0, \mu \ge 0\end{array}\right\} \text{[LP duality]}\\
= \max_x \left\{x_i : A^T x = 0, \|x\|_1 \le 1\right\}
\end{array}$$

Illustration

♠ k**-th Hadamard matrix** $\mathcal{H}_k$ is $n_k \times n_k$ matrix, $n_k = 2^k$, with entries ± 1 given by the recurrence

$$\mathcal{H}_0 = [1]; \mathcal{H}_{k+1} = \begin{bmatrix} \mathcal{H}_k & \mathcal{H}_k \\ \mathcal{H}_k & -\mathcal{H}_k \end{bmatrix}$$

Note: *$\mathcal{H}_k$ is symmetric and is proportional to orthogonal matrix:* $\mathcal{H}_k^T \mathcal{H}_k = n_k I_{n_k}$

$\Rightarrow$ When $k > 0$, the only eigenvalues of $\mathcal{H}_k$ are $\sqrt{n_k}$ and $-\sqrt{n_k}$ with multiplicities $n_k/2$ each.

• Let $k > 1$, $m_k = n_k/2 = 2^{k-1}$, and let $a_1, ..., a_{m_k}$ be an orthonormal system of eigenvectors of $\mathcal{H}_k$ with eigenvalue $\sqrt{n_k}$. Let A_k be the $m_k \times n_k$ matrix with the rows $a_1^T, ..., a_{m_k}^T$.

Fact: *Let $s < \frac{1}{2}\sqrt{n_k} = 2^{k/2-1}$. Then the matrix A_k is s-good. Moreover, there exists (and can be efficiently computed) contrast matrix H_k such that $(H_k, \|\cdot\|_\infty)$ satisfies the condition $\mathbf{Q}_\infty(s, \kappa_s = s/\sqrt{n_k})$, and $\|\mathrm{Col}_i[H_k]\|_2 \leq \sqrt{2 + 2/\sqrt{n_k}}$ for all j.*

♠ **Verifiable Sufficient condition for satisfiability of $\mathbf{Q}_q(s,\kappa)$:** *Let $m\times n$ matrix H satisfy the condition*

$$\|\mathrm{Col}_j[I_n - H^TA]\|_{s,q} \le s^{\frac{1}{q}-1}\kappa,\ 1\le j\le n \tag{!}$$

Then $H, \|\cdot\|_\infty$ satisfy $\mathbf{Q}_q(s,\kappa)$.

Proof:

$$\begin{array}{l}
(!) \Rightarrow \|[I_n - H^TA]x\|_{s,q} \le s^{\frac{1}{q}-1}\kappa\|x\|_1\ \forall x\\
\Rightarrow \|x\|_{s,q} - \|H^TAx\|_{s,q} \le s^{\frac{1}{q}-1}\kappa\|x\|_1\ \forall x\\
\Rightarrow \|x\|_{s,q} \le \|H^TAx\|_{s,q} + s^{\frac{1}{q}-1}\kappa\|x\|_1\\
\Rightarrow \|x\|_{s,q} \le s^{\frac{1}{q}}\|H^TAx\|_\infty + s^{\frac{1}{q}-1}\kappa\|x\|_1\ \forall x
\end{array}$$

Note: (!) is an explicit system of convex constraints on H

$\Rightarrow$ *The sufficient condition* (!) *for $H, \|\cdot\|_\infty$ to satisfy $\mathbf{Q}_q(s,\kappa)$ is computationally tractable.*

Note: *When $q=\infty$, feasibility of* (!) *is necessary and sufficient for satisfiability of $\mathbf{Q}_\infty(s,\kappa)$: $(H\in\mathbb{R}^{m\times n}, \|\cdot\|_\infty)$ satisfies $\mathbf{Q}_\infty(s,\kappa)$ if and only if*

$$\|\mathrm{Col}_j[I_n - H^TA]\|_\infty \le s^{-1}\kappa\ \forall j.$$

♠ Let $m\times n$ matrix H satisfy the condition

$$\|\mathrm{Col}_j[I_n - H^TA]\|_{s,q} \le s^{\frac{1}{q}-1}\kappa,\ 1\le j\le n \qquad (!)$$

Then $H,\|\cdot\|$ satisfy $\mathbf{Q}_q(s,\kappa)$.

The above statement, whatever simple, has an instructive origin. Consider the following problem:

(?) *Given a convex function* $\phi(x):\mathbb{R}^n\to\mathbb{R}$ *and a convex set*

$$X=\{x\in \mathrm{Conv}\{f_1,...,f_N\}: Ax=0\}$$
$$[A\in\mathbb{R}^{m\times n}]$$

we want to compute/upper-bound efficiently the quantity

$$\phi_* = \max_{x\in X}\phi(x).$$

Example: Verifying the Nullspace Property of matrix A reduces to checking whether the quantity

$$\phi_* := \max_{x\in X}\left[\phi(x) := \|x\|_{s,1}\right],$$
$$X=\{x\in \mathrm{Conv}\{\pm e_1,\pm e_2,...,\pm e_n\}: Ax=0\}$$
$$[e_i: \text{basic orths}]$$

is or is not $< 1/2$.

$$\phi_* = \max_{x\in X} \phi(x),\ \ X = \{x \in \mathrm{Conv}\{f_1, ..., f_N\} : Ax = 0\}$$

• ϕ_* is the maximum of a convex function over a bounded polyhedral set and as such is in general NP-hard to compute. However, we can point out a simple scheme for efficient upper-bounding ϕ_*:

$$\begin{array}{lcl}
\forall H \in \mathbb{R}^{m\times n}: & & \\
\phi_* & = & \max_x\{\phi(x) : x \in \mathrm{Conv}\{f_1, ..., f_N\}, Ax = 0\} \\
 & = & \max_x\{\phi([I - H^TA]x) : x \in \mathrm{Conv}\{f_1, ..., f_N\}, Ax = 0\} \\
 & \leq & \max_x\{\phi([I - H^TA]x) : x \in \mathrm{Conv}\{f_1, ..., f_N\}\} \\
 & = & \max_{j\leq N} \phi([I - H^TA]f_j),
\end{array}$$

$$\Rightarrow \boxed{\phi_* \leq \overline{\phi} := \min_H \left[\max_{j\leq N}\phi([I - H^TA]f_j)\right]}$$

and $\overline{\phi}$ is efficiently computable (as the optimal value in a convex problem).

• **Note:** As applied to

$$\phi(x) = \|x\|_{s,1},\ \ X = \{x \in \mathrm{Conv}\{\pm e_1, ..., \pm e_n\} : Ax = 0\},$$

the above bounding scheme results in the verifiable sufficient condition

$$\exists(\kappa < 1/2, H) : \|\mathrm{Col}_j[I - H^TA]\|_{s,1} \leq \kappa,\ 1 \leq j \leq n$$

for s-goodness of A. This hint leads to the verifiable sufficient conditions for $\mathbf{Q}_q(s, \kappa)$.

♠ **Bad news:** *When $m \times n$ sensing matrix A is "essentially non-square", namely, $n \geq 2m$, the above verifiable sufficient conditions for the validity of $\mathbf{Q}_q(s,\kappa)$ can be satisfiable only in the range*

$$s \leq \sqrt{2m} \qquad (!)$$

which is much less than the range

$$s \leq O(1)\frac{m}{\ln(n/m)}$$

where random Gaussian/Rademacher $m \times n$ sensing matrices satisfy $\mathrm{RIP}(\frac{1}{4}, 2s)$ *with overwhelming probability, thus implying satisfiability of $\mathbf{Q}_2(s, \frac{1}{3})$.*

Note:

A. No series of individual essentially non-square $m \times n$ sensing matrices A with $m, n \to \infty$ which are provably s-good for $s \geq O(1)\sqrt{m}$ are known

B. For $k = 1, 2, ...$ one can easily point out individual $2^{k-1} \times 2^k$ sensing matrices for which condition $\mathbf{Q}_\infty(s, \frac{1}{3})$ is satisfiable whenever $s \leq \frac{\sqrt{2m}}{3}$.

C. *Whenever A satisfies* $\mathrm{RIP}(\delta, 2k)$ *and $s \leq \frac{1-\delta}{3\delta}\sqrt{k}$, the pair $(H = \frac{\sqrt{k}}{1-\delta}A, \|\cdot\|_\infty)$ satisfies $\mathbf{Q}_\infty(s, \frac{1}{3})$*

D. For properly selected $C > 0$ and every m, n, one can point out individual $m \times n$ sensing matrix which is $C\sqrt{m}$-good.

♣ **Mutual Incoherence.** Let A be $m\times n$ sensing matrix without zero columns. Mutual Incoherence of A is the quantity

$$\mu(A) = \max_{i\neq j} \frac{|\mathrm{Col}_i^T[A]\mathrm{Col}_j[A]|}{\mathrm{Col}_i^T[A]\mathrm{Col}_i[A]}$$

Observation: *The $m \times n$ matrix H with columns $\frac{\mathrm{Col}_j[A]}{\mathrm{Col}_j^T[A]\mathrm{Col}_j[A]}$, $j = 1, ..., n$, satisfies*

$$\forall j : \|\mathrm{Col}_j[I_n - H^T A]\|_\infty \leq \frac{\mu(A)}{1+\mu(A)}$$

$\Rightarrow H, \|\cdot\|_\infty$ satisfy $\mathbf{Q}_\infty\left(s, \frac{s\mu(A)}{1+\mu(A)}\right)$ for every s. In particular, A is s-good, provided that

$$\frac{2\mu(A)}{1+\mu(A)} < \frac{1}{s}.$$

.

HYPOTHESIS TESTING, I

- *Preliminaries*
  - *Tests & Risks*
  - *Repeated Observations*
  - *2-Point Lower Risk Bound*
- *Pairwise Tests via Euclidean Separation*
- *From Pairwise to Multiple Hypothesis Testing*

♣ **Hypothesis Testing Problem:** Given

• *observation space* Ω where our observations take values,

• L *families* $\mathcal{P}_1$, $\mathcal{P}_2$,...,$\mathcal{P}_L$ of probability distributions on Ω, and

• *an observation* ω – a realization of random variable with *unknown* probability distribution P *known to belong to one of the families* $\mathcal{P}_\ell$: $P \in \bigcup\limits_{\ell=1}^{L} \mathcal{P}_\ell$,

we want to decide to which one of the families $\mathcal{P}_\ell$ the distribution P belongs.

Equivalent wording: *Given the outlined data, we want to decide on L hypotheses $H_1, ..., H_L$, with ℓ-th hypothesis H_ℓ stating that $P \in \mathcal{P}_\ell$.*

♣ **A test** is a function $\mathcal{T}(\cdot)$ on Ω. The value $\mathcal{T}(\omega)$ of this function at a point $\omega \in \Omega$ is a subset of the set $\{1, ..., L\}$.

• relation $\ell \in \mathcal{T}(\omega)$ is interpreted as "given observation ω, the test accepts the hypothesis H_ℓ"

• relation $\ell \notin \mathcal{T}(\omega)$ is interpreted as "given observation ω, the test rejects the hypothesis H_ℓ"

♠ $\mathcal{T}$ is called *simple*, if $\mathcal{T}(\omega)$ is a singleton for every $\omega \in \Omega$.

♣ **For a simple test** $\mathcal{T}$, its *risks* are defined as follows:

♠ ℓ-th partial risk of $\mathcal{T}$ is the (worst-case) probability to reject ℓ-th hypothesis when it is true:

$$\mathrm{Risk}_\ell(\mathcal{T}|H_1,...,H_L) = \sup_{P\in\mathcal{P}_\ell} \mathrm{Prob}_{\omega\sim P}\{\ell \not\in \mathcal{T}(\omega)\}$$

♠ total risk of $\mathcal{T}$ is the sum of all partial risks:

$$\mathrm{Risk}_{\mathrm{tot}}(\mathcal{T}|H_1,...,H_L) = \sum_{1\le\ell\le L} \mathrm{Risk}_\ell(\mathcal{T}|H_1,...,H_L).$$

♠ risk of $\mathcal{T}$ is the maximum of all partial risks:

$$\mathrm{Risk}(\mathcal{T}|H_1,...,H_L) = \max_{1\le\ell\le L} \mathrm{Risk}_\ell(\mathcal{T}|H_1,...,H_L).$$

♣ **Note:** What was called test is in fact a *deterministic* test.

A *randomized* test is a *deterministic* function $\mathcal{T}(\omega,\eta)$ of observation ω and *independent of ω random variable $\eta \sim P_\eta$* with once for ever fixed distribution (say, $P_\eta = \text{Uniform}[0,1]$). The values $\mathcal{T}(\omega,\eta)$ of $\mathcal{T}$ are subsets of $\{1,...,L\}$ (singletons for a simple test).

- Given observation ω, we "flip a coin" (draw a realization of η), accept hypotheses H_ℓ, $\ell \in \mathcal{T}(\omega,\eta)$, and reject all other hypotheses.
- Partial risks of randomized test are

$$\text{Risk}_\ell(\mathcal{T}|H_1,...,H_L) = \sup_{P\in\mathcal{P}_\ell} \text{Prob}_{(\omega,\eta)\sim P\times P_\eta}\{\ell \notin \mathcal{T}(\omega,\eta)\}.$$

Exactly as above, these risks give rise to the *total risk* and *risk* of $\mathcal{T}$.

♣ **Testing from repeated observations.** There are situations where an inference can be based on several observations $\omega_1, ..., \omega_K$ rather than on a single observation. Our related setup is as follows:

♠ We are given L families $\mathcal{P}_\ell$, $\ell = 1, ..., L$, of probability distributions on observation space Ω and a collection

$$\omega^K = (\omega_1, ..., \omega_K)$$

and want to make conclusions on how the distribution of ω^K "is positioned" w.r.t. the families $\mathcal{P}_\ell$, $1 \le \ell \le L$. Specifically, we are interested in three situations of type:

♠ **A. Stationary K-repeated observations:** $\omega_1, ..., \omega_K$ are *independently of each other* drawn from a distribution P. Our goal is to decide, given ω^K, on the hypotheses $P \in \mathcal{P}_\ell$, $\ell = 1, ..., L$.

Equivalently: Families $\mathcal{P}_\ell$ give rise to the families

$$\mathcal{P}_\ell^{\odot,K} = \{P^K = \underbrace{P \times ... \times P}_{K} : P \in \mathcal{P}_\ell\}$$

of probability distributions on $\Omega^K = \underbrace{\Omega \times ... \times \Omega}_{K}$ – *direct powers* of families $\mathcal{P}_\ell$.

Given observation $\omega^K \in \Omega^K$, we want to decide on the hypotheses

$$H_\ell^{\odot,K} : \omega^K \sim P^K \in \mathcal{P}_\ell^{\odot,K}, \; 1 \le \ell \le L.$$

♠ **B. Semi-stationary K-repeated observations:** *"The nature" selects somehow a sequence $P_1, ..., P_K$ of distributions on Ω, and then draws, independently across k, observations ω_k from these distributions:*

$$\omega_k \sim P_k \text{ are independent across } k \leq K$$

Our goal is to decide, given $\omega^K = (\omega_1, ..., \omega_K)$, on the hypotheses $\{P_k \in \mathcal{P}_\ell, 1 \leq k \leq K\}$, $\ell = 1, ..., L$.
Equivalently: Families $\mathcal{P}_\ell$ give rise to the families

$$\mathcal{P}_\ell^{\oplus,K} = \bigoplus_{k=1}^{K} \mathcal{P}_\ell := \{P^K = P_1 \times ... \times P_K : P_k \in \mathcal{P}_\ell,\ 1 \leq k \leq K\}$$

of probability distributions on $\Omega^K = \underbrace{\Omega \times ... \times \Omega}_{K}$ – *direct products* of families $\mathcal{P}_\ell$.

Given observation $\omega^K \in \Omega^K$, we want to decide on the hypotheses

$$H_\ell^{\oplus,K} : \omega^K \sim P^K \in \mathcal{P}_\ell^{\oplus,K},\ 1 \leq \ell \leq L.$$

♠ **C. Quasi-stationary K-repeated observations:** We observe random sequence $\omega^K = (\omega_1, ..., \omega_K)$ generated as follows:

> *There exists a random sequence $\zeta_1, ..., \zeta_K$ of driving factors such that for $1 \leq k \leq K$*
> - *ω_k is a deterministic function of $\zeta^k = (\zeta_1, ..., \zeta_k)$*
> - *conditional, ζ^{k-1} given, distribution of ω_k always belongs to $\mathcal{P}_\ell$.*

Our goal is to decide, given ω^K, on the underlying ℓ.

Equivalently: Families $\mathcal{P}_\ell$ of probability distributions on Ω, $1 \leq \ell \leq L$, give rise to the *quasi-direct products* $\mathcal{P}_\ell^{\otimes,K} = \bigotimes_{k=1}^K \mathcal{P}_\ell$ of families $\mathcal{P}_\ell$. The family $\bigotimes_{k=1}^K \mathcal{P}_\ell$ is comprised of all probability distributions on $\Omega^K = \underbrace{\Omega \times ... \times \Omega}_{K}$ which can be obtained from $\mathcal{P}_\ell$ via the above "driving factors" mechanism.

Given observation $\omega^K \in \Omega^K$, we want to decide on the hypotheses

$$H_\ell^{\otimes,K} : \omega^K \sim P^K \in \mathcal{P}_\ell^{\otimes,K},\ 1 \leq \ell \leq L.$$

♣ **Important fact: 2-point lower risk bound.** Consider *simple pairwise test* deciding on two simple hypotheses on the distribution P of observation $\omega \in \Omega$:

$$H_1 : P = P_1,\; H_2 : P = P_2.$$

Let P_1, P_2 have densities p_1, p_2 w.r.t. some reference measure Π on Ω. Then *the total risk of every test $\mathcal{T}$ deciding on H_1, H_2 admits lower bound as follows:*

$$\mathrm{Risk}_{\mathrm{tot}}(\mathcal{T}|H_1,H_2) \geq \int_\Omega \min[p_1(\omega),p_2(\omega)]\Pi(d\omega).$$

As a result,

$$\mathrm{Risk}(\mathcal{T}|H_1,H_2) \geq \frac{1}{2}\int_\Omega \min[p_1(\omega),p_2(\omega)]\Pi(d\omega). \qquad (*)$$

Note: *The bound does not depend on the choice of Π (for example, we can always take $\Pi = P_1 + P_2$).*

$$\mathrm{Risk}(\mathcal{T}|H_1,H_2) \geq \frac{1}{2}\int\limits_\Omega \min[p_1(\omega),p_2(\omega)]\Pi(d\omega). \qquad (?)$$

Proof (for deterministic test). Simple test deciding on H_1, H_2 must accept H_1 and reject H_2 on some subset Ω_1 of Ω and must reject H_1 and accept H_2 on the complement $\Omega_2 = \Omega\backslash\Omega_1$ of this set. We have

$$\begin{array}{ll}
\mathrm{Risk}_1(\mathcal{T}|H_1,H_2) & = \int\limits_{\Omega_2} p_1(\omega)\Pi(d\omega) \geq \int\limits_{\Omega_2} \min[p_1(\omega),p_2(\omega)]\Pi(d\omega)\\
\mathrm{Risk}_2(\mathcal{T}|H_1,H_2) & = \int\limits_{\Omega_1} p_2(\omega)\Pi(d\omega) \geq \int\limits_{\Omega_1} \min[p_1(\omega),p_2(\omega)]\Pi(d\omega)\\
\Rightarrow \mathrm{Risk}_{\mathrm{tot}}(\mathcal{T}|H_1,H_2) & \geq \int\limits_{\Omega_2} \min[p_1(\omega),p_2(\omega)]\Pi(d\omega) + \int\limits_{\Omega_1} \min[p_1(\omega),p_2(\omega)]\Pi(d\omega)\\
& = \int\limits_{\Omega} \min[p_1(\omega),p_2(\omega)]\Pi(d\omega) \qquad \square
\end{array}$$

♠ **Corollary.** *Consider L hypotheses $H_\ell : P \in \mathcal{P}_\ell$, $\ell = 1, 2, ..., L$, on the distribution P of observation $\omega \in \Omega$, let $\ell \neq \ell'$ and let $P_\ell \in \mathcal{P}_\ell$, $P_{\ell'} \in \mathcal{P}_{\ell'}$. The risk of any simple test $\mathcal{T}$ deciding on $H_1, ..., H_L$ can be lower-bounded as*

$$\mathrm{Risk}(\mathcal{T}|H_1, ..., H_L) \geq \frac{1}{2} \int\limits_\Omega \min\left[P_\ell(d\omega), P_{\ell'}(d\omega)\right],$$

where, by convention, the integral in the right hand side is

$$\int\limits_\Omega \min[p_\ell(\omega), p_{\ell'}(\omega)]\Pi(d\omega),$$

with p_ℓ, $p_{\ell'}$ being the densities of P_ℓ, $P_{\ell'}$ w.r.t. $\Pi = P_\ell + P_{\ell'}$.

Indeed, risk of $\mathcal{T}$ cannot be less than the risk of the naturally induced by $\mathcal{T}$ simple test deciding on two simple hypotheses $P = P_\ell$, $P = P_{\ell'}$, specifically, the simple test which, given observation ω accepts the hypothesis $P = P_1$ whenever $\ell \in \mathcal{T}(\omega)$ and accepts the hypothesis $P = P_{\ell'}$ otherwise.

Pairwise Hypothesis Testing via Euclidean Separation

♣ **Situation:** Let $\Omega = \mathbb{R}^d$, and let our observation be

$$\omega = x + \xi \qquad (*)$$

where the deterministic vector x is the signal of interest, and ξ is random observation noise with probability density $p(\cdot)$ of the form

$$p(u) = f(\|u\|_2)$$

where $f(\cdot)$ is a strictly monotonically decreasing function on the nonnegative ray.

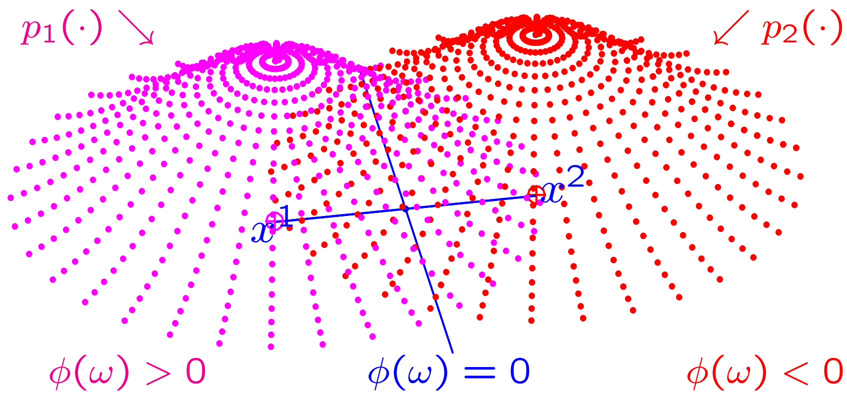
Simple example: *standard (zero mean, unit covariance)* Gaussian noise: $p(u) = (2\pi)^{-d/2} \mathrm{e}^{-u^T u/2}$.

Our goal is to decide on two simple hypotheses on the signal underlying observation, the first stating that $x = x^1$, and the second stating that $x = x^2$, where x^1, x^2 are two given points.

Equivalent wording: We are given two probability distributions, P_1 and P_2, on $\mathbb{R}^d$, with densities $p_1(u) = p(u - x^1)$ and $p_2(u) = p(u - x^2)$, and want to decide on two simple hypotheses $H_1 : P = P_1$, $H_2 : P = P_2$ on the distribution P of our observation.

♠ Assuming $x^1 \neq x^2$, let $2\delta = \|x^1 - x^2\|_2$, $e = \frac{x^1 - x^2}{\|x^1 - x^2\|_2}$,

$$\Pi = \{\omega : \|\omega - x^1\|_2 = \|\omega - x^2\|_2\} = \{\omega : \phi(\omega) = 0\},\ \phi(\omega) = e^T\omega - \underbrace{\tfrac{1}{2}e^T[x^1 + x^2]}_{c}$$



Consider test $\mathcal{T}$ which, given observation $\omega = x + \xi$, accepts the hypothesis $H_1 : P = P_1$ (i.e., $x = x^1$) when $\phi(\omega) \geq 0$, and accepts the hypothesis $H_2 : P = P_2$ (i.e., $x = x^2$) otherwise. We have

$$\begin{array}{rcl}\mathrm{Risk}_1(\mathcal{T}|H_1, H_2) & = & \int\limits_{\omega:\phi(\omega)<0} p_1(\omega)d\omega = \int\limits_{u:e^Tu\geq\delta} f(\|u\|_2)du \\ & = & \int\limits_{\omega:\phi(\omega)\geq 0} p_2(\omega)d\omega = \mathrm{Risk}_2(\mathcal{T}|H_1, H_2)\end{array}$$

Since $p(u)$ is strictly decreasing function of $\|u\|_2$, we have $\min[p_1(u), p_2(u)] = \left\{\begin{array}{ll} p_1(u), & \phi(u)\geq 0 \\ p_2(u), & \phi(u)\leq 0\end{array}\right.$, whence

$$\begin{array}{l}\mathrm{Risk}_1(\mathcal{T}|H_1, H_2) + \mathrm{Risk}_2(\mathcal{T}|H_1, H_2) \\ \quad = \int\limits_{\omega:\phi(\omega)<0} p_1(\omega)d\omega + \int\limits_{\omega:\phi(\omega)\geq 0} p_2(\omega)d\omega) = \int\limits_{\mathbb{R}^d} \min[p_1(u), p_2(u)]du\end{array}$$

$\Rightarrow$ *Test $\mathcal{T}$ is the minimum risk simple test deciding on H_1, H_2.*

♣ **Extension:** Given observation $\omega = x + \xi$ with observation noise ξ possessing probability density

$$p(u) = f(\|u\|_2),$$

where $f(\cdot)$ is a strictly decreasing function on the nonnegative ray, we want do decide on two composite hypotheses H_1, H_2:

$$H_1 : x \in X_1, \quad H_2 : x \in X_2,$$

where X_1, X_2 are nonempty *nonintersecting, closed and convex* sets, and one of the sets is bounded.

♠ **Elementary fact:** *With X_1, X_2 as above, consider the convex minimization problem*

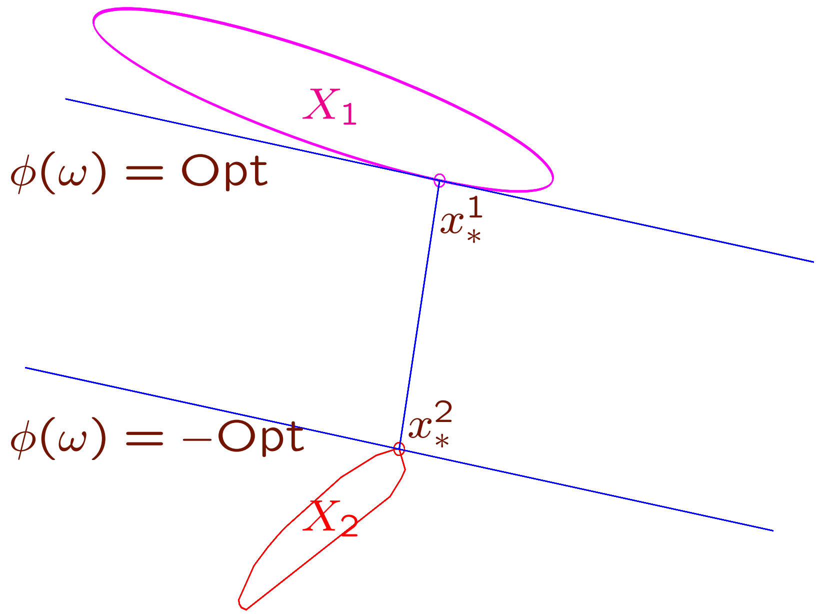
$$\mathrm{Opt} = \min_{x^1 \in X_1, x^2 \in X_2} \tfrac{1}{2}\|x^1 - x^2\|_2.$$

The problem is solvable. Let (x_*^1, x_*^2) be an optimal solution, and let

$$\phi(\omega) = e^T\omega - c, \; e = \frac{x_*^1 - x_*^2}{\|x_*^1 - x_*^2\|_2}, \, c = \frac{1}{2}e^T[x_*^1 + x_*^2]$$

Then the stripe $\{\omega : -\mathrm{Opt} \le \phi(\omega) \le \mathrm{Opt}\}$ separates X_1 and X_2:

$$\phi(x^1) \ge \phi(x_*^1) = \mathrm{Opt}\, \forall x^1 \in X_1,$$
$$\phi(x^2) \ge \phi(x_*^2) = -\mathrm{Opt}\, \forall x^2 \in X_2$$



♠ We have associated with two non-intersecting closed convex X_1, X_2, one of the sets being bounded,

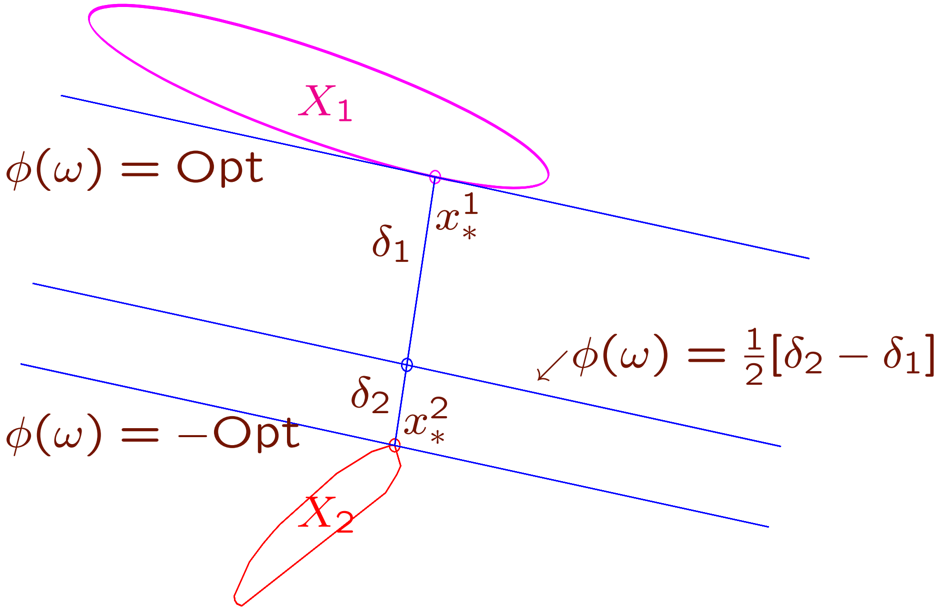
— convex optimization problem

$$\text{Opt} = \min_{x^1 \in X_1, x^2 \in X_2} \frac{1}{2}\|x^1 - x^2\|_2$$

— linear function

$$\phi(\omega) = e^T\omega - \frac{1}{2}e^T[x^1_* + x^2_*],\ e = \frac{1}{2\text{Opt}}[x^1_* - x^2_*]$$

where $[x^1_*, x^2_*]$ is an optimal solution to the above problem. While this solution not necessarily is uniquely defined by X_1, X_2, $\phi(\cdot)$ *is uniquely defined by* X_1, X_2.



♠ Given $\delta_1\geq 0, \delta_2\geq 0$ with $\delta_1+\delta_2=2\mathrm{Opt}$, $\phi(\cdot)$ specifies simple *Euclidean Separation Test $\mathcal{T}$ induced by $X_1, X_2, \delta_1, \delta_2$*:

$$\mathcal{T}(\omega)=\left\{\begin{array}{ll}\{1\}, & \phi(\omega)\geq\frac{1}{2}[\delta_2-\delta_1]\\ \{2\}, & \text{otherwise}\end{array}\right.$$

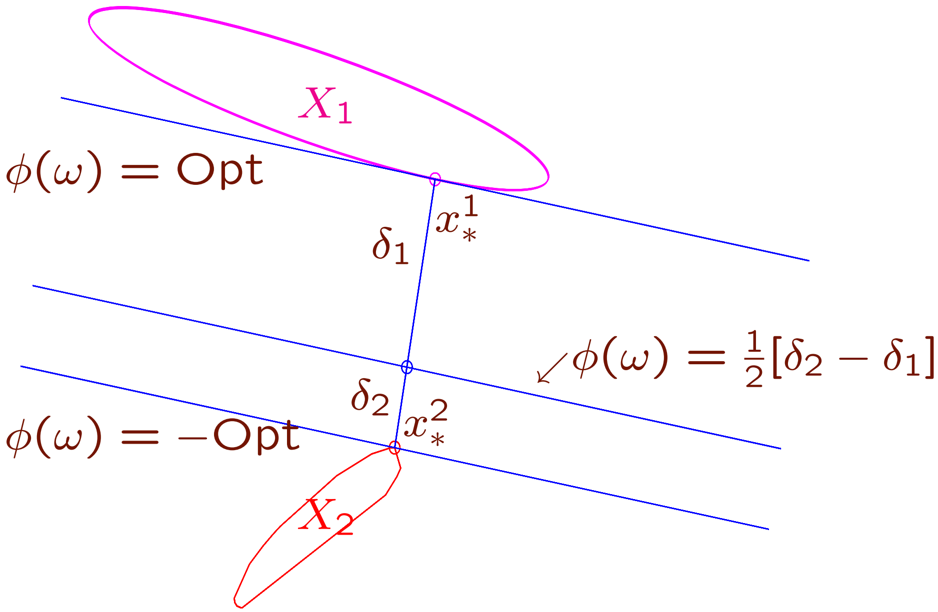
♠ **Fact:** *Let $\xi\sim p(\cdot)$, where $p(u)=f(\|u\|_2)$ with strictly decreasing $f(t), t\geq 0$. Given observation $\omega=x+\xi$ the Euclidean Separation Test $\mathcal{T}$ decides on the hypotheses $H_1: x\in X_1$, $H_2: x\in X_2$ with risks satisfying*

$$\mathrm{Risk}_1(\mathcal{T}|H_1,H_2)\leq\int_{\delta_1}^{\infty}\gamma(s)ds,\ \ \mathrm{Risk}_2(\mathcal{T}|H_1,H_2)\leq\int_{\delta_2}^{\infty}\gamma(s)ds$$

where $\gamma(\cdot)$ is the univariate marginal density of ξ, that is, probability density of the scalar random variable $h^T\xi$, where $\|h\|_2=1$.

♡ *In addition, when $\delta_1=\delta_2=\mathrm{Opt}$, $\mathcal{T}$ is the minimum risk test deciding on H_1, H_2, and*

$$\mathrm{Risk}(\mathcal{T}|H_1,H_2)=\int_{\mathrm{Opt}}^{\infty}\gamma(s)ds.$$



♣ **Extension:** Under the premise of Fact: the observation is $\omega = x + \xi$ with $\xi \sim p(\cdot) = f(\|\cdot\|_2)$, where

- $f : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly decreasing, and
- the hypotheses to be decided upon are $H_1 : x \in X_1$, $H_2 : x \in X_2$ with closed convex nonintersecting and nonempty X_1, X_2, one of the sets being bounded,

the risk bounds $\mathrm{Risk}_\ell(\mathcal{T}|H_1, H_2) \leq \int\limits_{\delta_\ell}^{\infty} \gamma(s)ds,\ \ell = 1, 2$ for the Euclidean Separation Test stem from the following observation:

Under the circumstances, for every half-space $E = \{u \in \mathbb{R}^d : e^T u \geq \delta\}$, where $\|e\|_2 = 1$ and $\delta \geq 0$, one has

$$\mathrm{Prob}_{\xi\sim p(\cdot)}\{\xi \in E\} \leq \int\limits_{\delta}^{\infty} \gamma(s)ds.$$

♣ Given an even probability density $\gamma(\cdot)$ on the axis such that $\int\limits_{\delta}^{\infty} \gamma(s)ds < \frac{1}{2}$ whenever $\delta > 0$, let us associate with it the family $\mathcal{P}_\gamma^d$ of all probability distributions P on $\mathbb{R}^d$ such that

A: distribution P possesses even density, and

B: whenever $e \in \mathbb{R}^d$, $\|e\|_2 = 1$, and $\delta \geq 0$, we have

$$\mathrm{Prob}_{\xi\sim P}\{\xi : e^T\xi \geq \delta\} \leq \Gamma(\delta) := \int\limits_{\delta}^{\infty} \gamma(s)ds$$

By the same reasons as in Fact, we have the following

♠ **Proposition.** *Whenever the distribution P of noise ξ in observation $\omega = x + \xi$ belongs to $\mathcal{P}_\gamma^d$ and X_1, X_2 are non-intersecting closed convex sets, one of the sets being bounded, the risks of the Euclidean Separation Test $\mathcal{T}$ induced by X_1, X_2 and δ_1, δ_2 can be upper-bounded as*

$$\mathrm{Risk}_\ell(\mathcal{T}|H_1, H_2) \leq \Gamma(\delta_\ell) := \int_{\delta_\ell}^{\infty} \gamma(s)ds, \ \ell = 1, 2.$$

♠ **Example: Gaussian mixtures.** Let η be an d-dimensional Gaussian random vector with zero mean and covariance matrix Θ (notation: $\eta \sim \mathcal{N}(0,\Theta)$). Let, further, Z be *independent of* η positive random variable. *Gaussian mixture* is the probability distribution of the random vector $\xi = \sqrt{Z}\eta$. Examples of Gaussian mixtures are:

- Gaussian distribution $\mathcal{N}(0,\Theta)$ (take Z identically equal to 1),
- multidimensional Student's t-distribution with ν degrees of freedom (ν/Z has χ^2-distribution with ν degrees of freedom)

♠ **Immediate Observations:**

•*Let Z be a random variable taking values in $[0,1]$, let $\eta \sim \mathcal{N}(0,\Theta)$ with $\Theta \preceq I_d$ (i.e., the matrix $I_d - \Theta$ is positive semidefinite) be independent of Z, and let*

$$\gamma_{\mathcal{G}}(s) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-s^2/2}$$

be the standard (zero mean, unit variance) Gaussian density on the axis. Then the distribution of the Gaussian mixture $\xi = \sqrt{Z}\eta$ belongs to the family $\mathcal{P}^d_{\gamma_{\mathcal{G}}}$.

•*With γ given by the distribution P_Z of Z according to*

$$\gamma_Z(s) = \int_{z>0} \frac{1}{\sqrt{2\pi z}}\mathrm{e}^{-\frac{s^2}{2z}} P_Z(dz),$$

the distribution of random variable $\sqrt{Z}\eta$, with $\eta \sim \mathcal{N}(0,\Theta)$, $\Theta \preceq I_d$, independent of Z, belongs to the family $\mathcal{P}^d_{\gamma_Z}$.

From Euclidean Separation to Majority Test

♣ Let $\gamma(\cdot)$, $\mathcal{P}^d_\gamma$, X_1, X_2 be as in Proposition, and assume we have access to semi-stationary K-repeated observations

$$\omega^K = \{\omega_k = x_k + \xi_k : 1 \leq k \leq K\}$$

where

- $\{x_k : 1 \leq k \leq K\}$ is a deterministic sequence of signals,
- $\xi_k \sim P_k, 1 \leq k \leq K$, are independent across k noises, and
- $\{P_k, 1 \leq k \leq K\}$ is a deterministic sequence of distributions from $\mathcal{P}^d_\gamma$.

Given ω^k, we want to decide on the hypotheses

$$H^K_1 : x_k \in X_1, 1 \leq k \leq K \text{ and } H^K_2 : x_k \in X_2, 1 \leq k \leq K.$$

Equivalently: The sets X_ℓ, $\ell = 1, 2$, give rise to families $\mathcal{P}_\ell$ of probability distributions on $\Omega = \mathbb{R}^d$; $\mathcal{P}_\ell$ is comprised of distributions P of random vectors of the form $x + \xi$, with deterministic $x \in X_\ell$ and with the distribution of noise ξ belonging to $\mathcal{P}^d_\gamma$. The families $\mathcal{P}_\ell$, in turn, give rise to hypotheses

$$H^K_\ell = H^{\oplus,K}_\ell : P^K \in \mathcal{P}^{\oplus,K}_\ell, \ \ell = 1, 2,$$

on the distribution P^K of K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$. Given ω^K, we want to decide on the hypotheses H^K_1, H^K_2.

$$\boxed{\begin{array}{c}\omega^K=\{\omega_k=x_k+\xi_k:1\leq k\leq K\}\\ H^K_\ell: x_k\in X_\ell,\,1\leq k\leq K,\,\xi_k\sim P_k\in\mathcal{P}^d_\gamma: \text{ independent across } k\end{array}}$$

♠ Let us use the *majority test* $\mathcal{T}^{\rm maj}_K$ defined as follows:

• we build the Euclidean separator of X_1, X_2, thus arriving at the affine function

$$\phi(\omega)=e^T\omega-c \qquad [\|e\|_2=1]$$

such that the stripe

$$\{\omega:-\mathrm{Opt}\leq\phi(\omega)\leq\mathrm{Opt}\}$$

with

$$\mathrm{Opt}=\min_{x^1\in X_1,x^2\in X_2}\tfrac{1}{2}\|x^1-x^2\|_2,$$

separates X_1, X_2;

• given $(\omega_1,...,\omega_K)$, we compute reals $v_k=\phi(\omega_k)$, $1\leq k\leq K$, and accept H^K_1 when the number of nonnegative v_k's is at least $K/2$, otherwise we accept H^K_2.

♠ **Risk analysis.** Assume that H_1^K takes place, so that $\{x_k\}$ form some deterministic sequence of points from X_1, and ξ_k are drawn, independently across k, from some distributions $P_k \in \mathcal{P}_\gamma^d$. With $\{x_k\}$ and $\{P_k\}$ fixed, v_k are independent across k, and probability for v_k to be negative is, by our previous results, $\leq \epsilon_\star := \Gamma(\mathrm{Opt}) := \int_{\mathrm{Opt}}^\infty \gamma(s)ds$, where

$$\mathrm{Opt} = \min_{x^1\in X_1, x^2\in X_2} \tfrac{1}{2}\|x^1 - x^2\|_2.$$

Consequently, the probability to reject H_1^K under the circumstances is

$$\leq \epsilon_K := \sum_{K/2\leq k\leq K} \binom{K}{k} \epsilon_\star^k (1-\epsilon_\star)^{K-k}.$$

By "symmetric" reasoning, the probability to reject H_2^K when the hypothesis is true is $\leq \epsilon_K$ as well. We arrive at

♠ **Proposition.** *The risk of $\mathcal{T}_K^{\mathrm{maj}}$ can be upper-bounded as*

$$\mathrm{Risk}(\mathcal{T}_K^{\mathrm{maj}}|H_1^K, H_2^K) \leq \sum_{K/2\leq k\leq K} \binom{K}{k} \epsilon_\star^k (1-\epsilon_\star)^{K-k}$$

$$\left[\epsilon_\star = \int_{\mathrm{Opt}}^\infty \gamma(s)ds, \mathrm{Opt} = \min_{x^1\in X_1, x^2\in X_2} \tfrac{1}{2}\|x^1 - x^2\|_2\right]$$

Fact: *Conclusion remains true in the case of quasi-stationary observations.*

$$\text{Risk}(\mathcal{T}_K^{\text{maj}}|H_1^K,H_2^K)\leq\sum_{K/2\leq k\leq K}\binom{K}{k}\epsilon_\star^k(1-\epsilon_\star)^{K-k}$$
$$\left[\epsilon_\star=\int_{\text{Opt}}^{\infty}\gamma(s)ds,\ \text{Opt}=\min_{x^1\in X_1,x^2\in X_2}\tfrac{1}{2}\|x^1-x^2\|_2\right]$$

♣ **Quiz:** We have used "evident" observation as follows:

Let w_1,... w_K be independent random variables taking values 0 and 1, and let the probability for w_i to take value 1 be some $p_i\in[0,1]$. Then for every fixed M the probability of the event *"at least M of $w_1,...,w_K$ are equal to 1"* as a function of $p_1,...,p_K$ is nondecreasing in every one of p_i's. (In our context, w_i were the signs of v_i).

Why this observation is true?

From Pairwise to Multiple Hypotheses Testing

♣ **Situation:** We are given L families of probability distributions $\mathcal{P}_\ell$, $1 \leq \ell \leq L$, on observation space Ω, and observe a realization of random variable $\omega \sim P$ taking values in Ω. Given ω, we want to decide on the L hypotheses

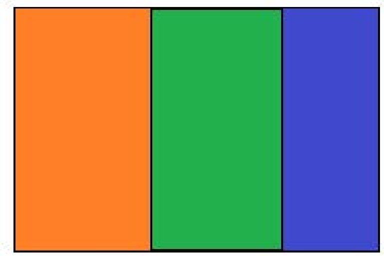
$$H_\ell : P \in \mathcal{P}_\ell,\ 1 \leq \ell \leq L.$$

Our ideal goal would be to find a low-risk simple test deciding on the hypotheses.

However: It may happen that the " ideal goal" is not achievable, for example, when some pairs of families $\mathcal{P}_\ell$ have nonempty intersections. When $\mathcal{P}_\ell \cap \mathcal{P}_{\ell'} \neq \emptyset$ for some $\ell \neq \ell'$, there is no way to decide on the hypotheses with risk $< 1/2$.

But: *Impossibility to decide reliably on all L hypotheses "individually" does not mean that no meaningful inferences can be done.*

♠ **Example:** Consider the 3 colored rectangles on the plane:



and 3 hypotheses, with H_ℓ, $1 \leq \ell \leq 3$, stating that our observation is $\omega = x + \xi$ with deterministic "signal" x belonging to ℓ-th rectangle and $\xi \sim \mathcal{N}(0, \sigma^2 I_2)$.
♡ Whatever small σ be, no test can decide on the 3 hypotheses with risk $< 1/2$; e.g., there is no way to decide reliably on H_1 vs. H_2. However, *we may hope that when σ is small, an observation allows us to discard reliably some of the hypotheses. For example, if H_1 is true, we hopefully can discard H_3.*

♠ *When handling multiple hypotheses which cannot be reliably decided upon "as they are," it makes sense to speak about testing the hypotheses "up to closeness."*

$$\boxed{\omega \sim P,\ \ H_\ell : P \in \mathcal{P}_\ell,\ 1 \leq \ell \leq L}$$

♣ **Closeness relation** *$\mathcal{C}$ on L hypotheses $H_1, ..., H_L$ is defined as some set of pairs (ℓ, ℓ') with $1 \leq \ell, \ell' \leq L$; we interpret the relation $(\ell, \ell') \in \mathcal{C}$ as the fact that the hypotheses H_ℓ and H'_ℓ are close to each other.*
We always assume that

- *$\mathcal{C}$ contains all "diagonal pairs" (ℓ, ℓ), $1 \leq \ell \leq L$ ("every hypothesis is close to itself")*
- *$(\ell, \ell') \in \mathcal{C}$ if and only if $(\ell', \ell) \in \mathcal{C}$* ("closeness is symmetric relation")

Note: By symmetry of $\mathcal{C}$, the relation $(\ell, \ell') \in \mathcal{T}$ is in fact a property of *un*ordered pair $\{\ell, \ell'\}$.

♠ **"Up to closeness" risks.** Let $\mathcal{T}$ be a test deciding on $H_1,...,H_L$; given observation ω, $\mathcal{T}$ accepts all hypotheses H_ℓ with indexes $\ell\in\mathcal{T}(\omega)$ and rejects all other hypotheses.

We say that ℓ-th partial $\mathcal{C}$-risk of test $\mathcal{T}$ is $\leq\epsilon$, if *whenever* H_ℓ *is true:* $\omega\sim P\in\mathcal{P}_\ell$, *the* P*-probability of the event*

$\mathcal{T}$ *accepts* H_ℓ: $\ell\in\mathcal{T}(\omega)$ and *all hypotheses* $H_{\ell'}$ *accepted by* $\mathcal{T}$ *are* $\mathcal{C}$*-close to* H_ℓ: $(\ell,\ell')\in\mathcal{C}\;\forall\ell'\in\mathcal{T}(\omega)$

is at least $1-\epsilon$.

♠ **ℓ-th partial $\mathcal{C}$-risk of $\mathcal{T}$** is the smallest ϵ with the outlined property:

$$\begin{array}{l}\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)\\ \quad=\sup\limits_{P\in\mathcal{P}_\ell}\mathrm{Prob}_{\omega\sim P}\left\{[\ell\not\in\mathcal{T}(\omega)]\text{ or }[\exists\ell'\in\mathcal{T}(\omega):(\ell,\ell')\not\in\mathcal{C}]\right\}\end{array}$$

♠ **$\mathcal{C}$-risk of $\mathcal{T}$** is the largest of the partial $\mathcal{C}$-risks of the test:

$$\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)=\max_{1\leq\ell\leq L}\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L).$$

$$\boxed{\begin{array}{c}\omega \sim P,\ \ H_\ell : P \in \mathcal{P}_\ell,\ 1 \leq \ell \leq L\\ \mathcal{C}:\ \text{closeness relation}\end{array}}$$

♣ **Multiple Hypothesis Testing via Pairwise Tests.** Assume that for every *un*ordered pair $\{\ell, \ell'\}$ with $(\ell, \ell') \not\in \mathcal{C}$ we are given a *simple* test $\mathcal{T}_{\{\ell,\ell'\}}$ deciding on H_ℓ vs. $H_{\ell'}$ via observation ω.

Our goal is to "assemble" the tests $\mathcal{T}_{\{\ell,\ell'\}}$, $(\ell, \ell') \not\in \mathcal{C}$, into a test $\mathcal{T}$ deciding on $H_1, ..., H_L$ up to closeness $\mathcal{C}$.

♠ **The construction:**

• For $(\ell,\ell')\not\in\mathcal{C}$, so that $\ell\neq\ell'$, we define function $T_{\ell\ell'}(\omega)$ as follows:

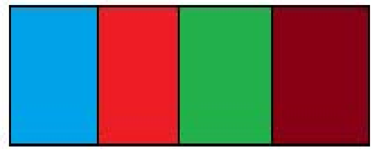
$$T_{\ell\ell'}(\omega)=\left\{\begin{array}{rl}1, & \mathcal{T}_{\{\ell,\ell'\}}(\omega)=\{\ell\}\\ -1, & \mathcal{T}_{\{\ell,\ell'\}}(\omega)=\{\ell'\}\end{array}\right. .$$

Note: $\mathcal{T}_{\{\ell,\ell'\}}$ is a simple test $\Rightarrow$ $T_{\ell\ell'}(\cdot)$ is well defined and takes values ±1.

♡ For $(\ell,\ell')\in\mathcal{C}$, we set $T_{\ell\ell'}(\cdot)\equiv 0$.

Note: By construction, we have $T_{\ell\ell'}(\omega)\equiv -T_{\ell'\ell}(\omega)$, $1\leq\ell,\ell'\leq L$.

• The test $\mathcal{T}$ is as follows: *given observation ω, we build the $L\times L$ matrix* $T(\omega)=[T_{\ell\ell'}(\omega)]$ *and accept exactly those of the hypotheses H_ℓ for which ℓ-th row in $T(\omega)$ is nonnegative*, that is, all tests $\mathcal{T}_{\{\ell,\ell'\}}$ with $(\ell,\ell')\not\in\mathcal{C}$ accept H_ℓ, observation being ω.



Example: • $L = 4$

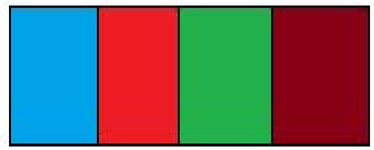
• $\mathcal{C} = \{(1,1),(2,2),(3,3),(4,4),\{1,2\},\{2,3\},\{3,4\}\}$

Given tests $\mathcal{T}_{\{1,3\}}$, $\mathcal{T}_{\{1,4\}}$, $\mathcal{T}_{\{2,4\}}$ and observation ω

♠ When $\mathcal{T}_{\{1,3\}}$ accepts H_1, $\mathcal{T}_{\{1,4\}}$ accepts H_1, $\mathcal{T}_{\{2,4\}}$ accepts H_4, we get

$$T(\omega) = \left[\begin{array}{c|c|c|c} 0 & 0 & +1 & +1 \\ \hline 0 & 0 & 0 & -1 \\ \hline -1 & 0 & 0 & 0 \\ \hline -1 & +1 & 0 & 0 \end{array}\right]$$

$\Rightarrow$ Aggregated test $\mathcal{T}$ accepts H_1



♠ When $\mathcal{T}_{\{1,3\}}$ accepts H_1, $\mathcal{T}_{\{1,4\}}$ accepts H_1, $\mathcal{T}_{\{2,4\}}$ accepts H_2, we get

$$T(\omega) = \left[\begin{array}{c|c|c|c} 0 & 0 & +1 & +1 \\ \hline 0 & 0 & 0 & +1 \\ \hline -1 & 0 & 0 & 0 \\ \hline -1 & -1 & 0 & 0 \end{array}\right]$$

$\Rightarrow$ Aggregated test $\mathcal{T}$ accepts H_1 and H_2

♠ **Observation:** *When $\mathcal{T}$ accepts some hypothesis H_ℓ, all hypotheses accepted by $\mathcal{T}$ are $\mathcal{C}$-close to H_ℓ.*

Indeed, if ℓ-th row in $T(\omega)$ is nonnegative and ℓ' is *not* $\mathcal{C}$-close to ℓ, we have $T_{\ell\ell'}(\omega) \geq 0$ and $T_{\ell\ell'}(\omega) \in \{-1, 1\}$

$\Rightarrow T_{\ell\ell'}(\omega) = 1$

$\Rightarrow T_{\ell'\ell}(\omega) = -T_{\ell\ell'}(\omega) = -1$

$\Rightarrow \ell'$-th row in $T(\omega)$ is *not* nonnegative

$\Rightarrow \ell'$ is *not* accepted.

♠ **Risk analysis.** For $(\ell,\ell')\not\in\mathcal{C}$, let

$$\begin{array}{rcl}\epsilon_{\ell\ell'} &=& \mathrm{Risk}_1(\mathcal{T}_{\{\ell,\ell'\}}|H_\ell,H_{\ell'}) = \sup\limits_{P\in\mathcal{P}_\ell}\mathrm{Prob}_{\omega\sim P}\{\ell\not\in\mathcal{T}_{\{\ell,\ell'\}}(\omega)\}\\ &=& \sup\limits_{P\in\mathcal{P}_\ell}\mathrm{Prob}_{\omega\sim P}\{T_{\ell\ell'}(\omega)=-1\} = \sup\limits_{P\in\mathcal{P}_\ell}\mathrm{Prob}_{\omega\sim P}\{T_{\ell'\ell}(\omega)=1\}\\ &=& \sup\limits_{P\in\mathcal{P}_\ell}\mathrm{Prob}_{\omega\sim P}\{\ell'\in\mathcal{T}_{\{\ell,\ell'\}}(\omega)\}\\ &=& \mathrm{Risk}_2(\mathcal{T}_{\{\ell,\ell'\}}|H_{\ell'},H_\ell).\end{array}$$

♠ **Proposition.** *One has*

$$\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)\leq\epsilon_\ell:=\sum_{\ell':(\ell,\ell')\not\in\mathcal{C}}\epsilon_{\ell\ell'}.$$

Indeed, let us fix ℓ, and let H_ℓ be true. Let $P\in\mathcal{P}_\ell$ be the distribution of observation ω, and let $I=\{\ell'\leq L:(\ell,\ell')\not\in\mathcal{C}\}$. For $\ell'\in I$, let $E_{\ell'}$ be the event $\{\omega:T_{\ell\ell'}(\omega)=-1\}$. We have $\mathrm{Prob}_{\omega\sim P}(E_{\ell'})\leq\epsilon_{\ell\ell'}$ (by definition of $\epsilon_{\ell\ell'}$) $\Rightarrow$ $\mathrm{Prob}_{\omega\sim P}\big(\underbrace{\cup_{\ell'\in I}E_{\ell'}}_{E}\big)\leq\epsilon_\ell$.

When the event E does *not* take place, we have $T_{\ell\ell'}(\omega)=1$ for all $\ell'\in I$
$\Rightarrow T_{\ell\ell'}(\omega)\geq 0$ for all ℓ', $1\leq\ell'\leq L$
$\Rightarrow \ell\in\mathcal{T}(\omega)$
$\Rightarrow$ (by Observation) $\{\ell\in\mathcal{T}(\omega)\}\ \&\ \{(\ell,\ell')\in\mathcal{C}\ \forall\ell'\in\mathcal{T}(\omega)\}$.
By definition of partial $\mathcal{C}$-risk, we get

$$\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)\leq\mathrm{Prob}_{\omega\sim P}(E)\leq\epsilon_\ell. \qquad\square$$

Testing Multiple Hypotheses via Euclidean Separation

♣ **Situation:** We are given L nonempty, closed and bounded convex sets $X_\ell \subset \mathbb{R}^d$, $1 \le \ell \le L$, and a family $\mathcal{P}^d_\gamma$ of noise distributions, a closeness $\mathcal{C}$, and semi-stationary K-repeated observation

$$\omega^K = \{\omega_k = x_k + \xi_k, 1 \le k \le K\},$$

so that

- $\{x_k, 1 \le k \le K\}$, is a deterministic sequence of signals,
- $\xi_k \sim P_k$, $1 \le k \le K$, are independent across k noises, and
- $\{P_k, 1 \le k \le K\}$, is a deterministic sequence of distributions from $\mathcal{P}^d_\gamma$.

Given ω^K, we want to decide up to closeness $\mathcal{C}$ on L hypotheses

$$H_\ell : \{x_k \in X_\ell, 1 \le k \le K\}.$$

Given ω^K, we want to decide up to closeness $\mathcal{C}$ on L hypotheses
$H_\ell : \{x_k \in X_\ell, 1 \leq k \leq K\}$.

Equivalently: The sets $X_\ell \subset \mathbb{R}^d$ along with $\mathcal{P}_\gamma^d$ specify L families of distributions $\mathcal{P}_\ell$, $1 \leq \ell \leq L$; specifically, $\mathcal{P}_\ell$ is comprised of probability distributions of random variables $x + \xi$, where deterministic x belongs to X_ℓ, and the distribution of random noise ξ belongs to $\mathcal{P}_\gamma^d$. Given ω^K, we want to decide, up to closeness $\mathcal{C}$, on L hypotheses

$$H_\ell : P^K \in \mathcal{P}_\ell^{\oplus,K},\ 1 \leq \ell \leq L$$

on the distribution P^K of observation ω^K.

♠ **Standing Assumption:** *Whenever ℓ, ℓ' are not $\mathcal{C}$-close: $(\ell, \ell') \notin \mathcal{C}$, the sets X_ℓ, $X_{\ell'}$ do not intersect.*

♠ **Strategy:** We intend to assemble pairwise Euclidean separation tests.

♠ **Building blocks.** For $(\ell,\ell')\not\in\mathcal{C}$, we solve convex optimization problems

$$\mathrm{Opt}_{\ell\ell'}=\min_{u\in X_\ell,v\in X_{\ell'}}\tfrac{1}{2}\|u-v\|_2. \qquad (P_{\ell\ell'})$$

Note: By Standing Assumption, $\mathrm{Opt}_{\ell\ell'}>0$. Optimal solution (u_*,v_*) to $(P_{\ell\ell'})$ defines affine functions

$$\begin{array}{c}\phi_{\ell\ell'}(\omega)=e_{\ell\ell'}^T\omega-c_{\ell\ell'}\\ e_{\ell\ell'}=\frac{u_*-v_*}{\|u_*-v_*\|_2},\ c_{\ell\ell'}=\frac{1}{2}e_{\ell\ell'}^T[u_*+v_*]\end{array}$$

Note: *We have* $\phi_{\ell\ell'}(\cdot)\equiv-\phi_{\ell'\ell}(\cdot)$ *for all* $(\ell,\ell')\not\in\mathcal{C}$.

♡ As we know, whenever $\delta_{\ell\ell'}\geq 0,\delta_{\ell'\ell}\geq 0$ satisfy

$$2\mathrm{Opt}_{\ell\ell'}=\delta_{\ell\ell'}+\delta_{\ell'\ell}$$

it holds

$$\begin{array}{rl}\forall(u\in X_\ell,P\in\mathcal{P}^d_\gamma): & \mathrm{Prob}_{\xi\sim P}\{\phi(u+\xi)<\frac{1}{2}[\delta_{\ell'\ell}-\delta_{\ell\ell'}]\}\\ & \leq\Gamma(\delta_{\ell\ell'}):=\int\limits_{\delta_{\ell\ell'}}^{\infty}\gamma(s)ds\\ \forall(v\in X_{\ell'},P\in\mathcal{P}^d_\gamma): & \mathrm{Prob}_{\xi\sim P}\{\phi(u+\xi)\geq\frac{1}{2}[\delta_{\ell'\ell}-\delta_{\ell\ell'}]\}\\ & \leq\Gamma(\delta_{\ell'\ell}):=\int\limits_{\delta_{\ell'\ell}}^{\infty}\gamma(s)ds\end{array}$$

$\ell, \ell' : (\ell, \ell') \not\in \mathcal{C}$		
$\Rightarrow$	$\mathrm{Opt}_{\ell\ell'} = \min\limits_{u\in X_\ell, v\in X_{\ell'}} \frac{1}{2}\|u-v\|_2 > 0 = \mathrm{Opt}_{\ell'\ell}$	
$\Rightarrow$	$u_*, v_*, \phi_{\ell\ell'}(\omega) = e_{\ell\ell'}^T\omega - c_{\ell\ell'} \equiv -\phi_{\ell'\ell}(\omega)$ $\left[e_{\ell\ell'} = \frac{u_*-v_*}{\|u_*-v_*\|_2}, c_{\ell\ell'} = \frac{1}{2}e_{\ell\ell'}^T[u_*+v_*]\right]$	
	$\delta_{\ell\ell'} \geq 0, \delta_{\ell'\ell} \geq 0, 2\mathrm{Opt}_{\ell\ell'} = \delta_{\ell\ell'} + \delta_{\ell'\ell}$	$(*)$
$\Rightarrow$	$\forall (u \in X_\ell, P \in \mathcal{P}_\gamma^d): \quad \mathrm{Prob}_{\xi\sim P}\left\{\phi(u+\xi) < \frac{1}{2}[\delta_{\ell'\ell} - \delta_{\ell\ell'}]\right\} \leq \Gamma(\delta_{\ell\ell'}) := \int_{\delta_{\ell\ell'}}^{\infty} \gamma(s)ds$ $\forall (v \in X_{\ell'}, P \in \mathcal{P}_\gamma^d): \quad \mathrm{Prob}_{\xi\sim P}\left\{\phi(v+\xi) \geq \frac{1}{2}[\delta_{\ell'\ell} - \delta_{\ell\ell'}]\right\} \leq \Gamma(\delta_{\ell'\ell}) := \int_{\delta_{\ell'\ell}}^{\infty} \gamma(s)ds$	$(!)$

♠ **Assembling building blocks, case of $K = 1$.**

• For ℓ, ℓ' with $(\ell, \ell') \not\in \mathcal{C}$ we select $\delta_{\ell\ell'}$ satisfying $(*)$, thus arriving at pairwise simple tests

$$\mathcal{T}_{\{\ell,\ell'\}}(\omega) = \left\{ \begin{array}{ll} \{\ell\}, & \phi_{\ell\ell'}(\omega) \geq \frac{1}{2}[\delta_{\ell'\ell} - \delta_{\ell\ell'}] \\ \{\ell'\}, & \phi_{\ell\ell'}(\omega) < \frac{1}{2}[\delta_{\ell'\ell} - \delta_{\ell\ell'}] \end{array} \right.$$

• Further, we use out general construction to assemble pairwise tests $\{\mathcal{T}_{\{\ell,\ell'\}} : (\ell, \ell') \not\in \mathcal{C}\}$ into single-observation test $\mathcal{T}$ deciding on $H_1, ..., H_L$

Note: *By* $(!)$*, the associated with tests* $\mathcal{T}_{\{\ell,\ell'\}}$ *quantities* $\epsilon_{\ell\ell'}$ *satisfy the relations* $\epsilon_{\ell\ell'} \leq \Gamma(\delta_{\ell\ell'}) := \int_{\delta_{\ell\ell'}}^{\infty} \gamma(s)ds$*, whence* $\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1, ..., H_L) \leq \sum\limits_{\ell':(\ell,\ell')\not\in\mathcal{C}} \Gamma(\delta_{\ell\ell'})$.

	$\ell,\ell':(\ell,\ell')\not\in\mathcal{C}\Rightarrow \mathrm{Opt}_{\ell\ell'}=\min\limits_{u\in X_\ell,v\in X_{\ell'}}\frac{1}{2}\|u-v\|_2$	
$\Rightarrow$	$\delta_{\ell\ell'}\geq 0,\delta_{\ell'\ell}\geq 0, 2\mathrm{Opt}_{\ell\ell'}=\delta_{\ell\ell'}+\delta_{\ell'\ell}$	
$\Rightarrow$	$\mathcal{T}:\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}\|H_1,...,H_L)\leq \sum\limits_{\ell':(\ell,\ell')\not\in\mathcal{C}}\Gamma(\delta_{\ell\ell'})$	$\left[\Gamma(\delta)=\int\limits_\delta^\infty\gamma(s)ds\right]$

♠ Single-observation case $K=1$: optimizing the construction over the "free parameters" $\delta_{\ell\ell'}$, $(\ell,\ell')\not\in\mathcal{C}$, of the construction.

♡ A natural model here is as follows: given nonnegative *weight matrix* W and nonnegative vectors α, β, we want to minimize "scale factor" t under the constraint

$$W\cdot[\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)]_{\ell=1}^L\leq\alpha+t\beta$$

This problem can be safely approximated by the optimization problem

$$\min_{\{\delta_{\ell\ell'}\},t}\left\{t:\begin{array}{l}W\cdot\left[\sum_{\ell':(\ell,\ell')\not\in\mathcal{C}}\Gamma(\delta_{\ell\ell'})\right]_{\ell=1}^L\leq\alpha+t\beta\\ \delta_{\ell\ell'}\geq 0,\delta_{\ell\ell'}+\delta_{\ell'\ell}=2\mathrm{Opt}_{\ell\ell'},\ (\ell,\ell')\not\in\mathcal{C}\end{array}\right\}\qquad(\#)$$

Note: *Assuming* $\gamma(\cdot)$ *nonincreasing on* $\mathbb{R}_+$ (as is the case, e.g., for Gaussian mixtures), *function* $\Gamma(\delta)=\int\limits_\delta^\infty\gamma(s)ds$ *is convex on* $\mathbb{R}_+$

$\Rightarrow$ $(\#)$ *is an explicit Convex Programming problem!*

$\ell, \ell' : (\ell, \ell') \not\in \mathcal{C}$
$\Rightarrow$ $\mathrm{Opt}_{\ell\ell'} = \min_{u\in X_\ell, v\in X_{\ell'}} \frac{1}{2}\|u-v\|_2 > 0 = \mathrm{Opt}_{\ell'\ell}$
$\Rightarrow$ $u_*, v_*, \phi_{\ell\ell'}(\omega) = e_{\ell\ell'}^T\omega - c_{\ell\ell'} \equiv -\phi_{\ell'\ell}(\omega)$ $\left[e_{\ell\ell'} = \frac{u_*-v_*}{\|u_*-v_*\|_2}, c_{\ell\ell'} = \frac{1}{2}e_{\ell\ell'}^T[u_*+v_*]\right]$
$\Rightarrow$ $\forall(u\in X_\ell, P\in\mathcal{P}_\gamma^d): \quad \mathrm{Prob}_{\xi\sim P}\{\phi(u+\xi)<0\} \leq \Gamma(\mathrm{Opt}_{\ell\ell'})$ $\forall(v\in X_{\ell'}, P\in\mathcal{P}_\gamma^d): \quad \mathrm{Prob}_{\xi\sim P}\{\phi(v+\xi)\geq 0\} \leq \Gamma(\mathrm{Opt}_{\ell\ell'})$ $\left[\Gamma(\delta) := \int_\delta^\infty \gamma(s)ds\right]$ (!)

♠ **Case of K-repeated observations, $K > 1$.** In the case of semi-stationary K-repeated observations $\omega^k = (\omega_1, ..., \omega_K)$, we act as follows:

- For $(\ell, \ell') \not\in \mathcal{C}$, we build majority tests

$$\mathcal{T}_{\{\ell,\ell'\}}(\omega^K) = \left\{\begin{array}{ll}\{\ell\}, & \mathrm{Card}\{k\leq K : \phi_{\ell\ell'}(\omega_k)\geq 0\} \geq K/2\\ \{\ell'\}, & \text{otherwise}\end{array}\right.$$

- Further, we use out general construction to assemble simple tests

$$\{\mathcal{T}_{\{\ell,\ell'\}} : (\ell,\ell') \not\in \mathcal{C}\}$$

into test $\mathcal{T}_K$ deciding on $H_1^K, ..., H_L^K$ via observation ω^K

Note: *By our results on majority tests, the associated with tests $\mathcal{T}_{\{\ell,\ell'\}}$ quantities $\epsilon_{\ell\ell'}$ satisfy the relations*

$$\epsilon_{\ell\ell'} \leq \sum_{K/2\leq k\leq K} \binom{K}{k} [\Gamma(\mathrm{Opt}_{\ell\ell'})]^k [1-\Gamma(\mathrm{Opt}_{\ell\ell'})]^{K-k}$$

whence

$$\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}_K|H_1,...,H_L) \leq \sum_{\ell':(\ell,\ell')\not\in\mathcal{C}} \sum_{K/2\leq k\leq K} \binom{K}{k} [\Gamma(\mathrm{Opt}_{\ell\ell'})]^k [1-\Gamma(\mathrm{Opt}_{\ell\ell'})]^{K-k}.$$

Note: By Standing Assumption, $\mathrm{Opt}_{\ell\ell'} > 0$ when $(\ell,\ell') \not\in \mathcal{C} \Rightarrow \Gamma(\mathrm{Opt}_{\ell\ell'}) < 1/2$
$\Rightarrow$ *Risks* $\mathrm{Risk}_\ell^{\mathcal{C}}(\mathcal{T}_K|H_1,...,H_L)$ *go to 0 exponentially fast as* $K\to\infty$.

.

HYPOTHESIS TESTING, II

- Detector-Based Tests
  - Detectors & Detector-Based Pairwise Tests
  - Testing “up to Closeness”
  - Simple Observation Schemes
    - Minimum Risk Detectors
    - Near-Optimal Tests
    - Sequential Hypothesis Testing
    - Measurement Design
- Recovering linear forms in Simple o.s.

Detectors & Detector-Based Pairwise Tests

♣ **Situation:** *Given two families $\mathcal{P}_1$, $\mathcal{P}_2$ of probability distributions on a given observation space Ω and an observation $\omega \sim P$ with P known to belong to $\mathcal{P}_1 \cup \mathcal{P}_2$, we want to decide whether $P \in \mathcal{P}_1$ (hypothesis H_1) or $P \in \mathcal{P}_2$ (hypothesis H_2).*

♣ **Detectors.** A *detector* is a function $\phi : \Omega \to \mathbb{R}$. *Risks* of a detector ϕ w.r.t. $\mathcal{P}_1, \mathcal{P}_2$ are defined as

$$\mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] = \sup_{P\in\mathcal{P}_1} \int_\Omega e^{-\phi(\omega)} P(d\omega), \ \mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2] = \sup_{P\in\mathcal{P}_2} \int_\Omega e^{\phi(\omega)} P(d\omega)$$
$$\mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] = \mathrm{Risk}_2[-\phi|\mathcal{P}_2,\mathcal{P}_1]$$

♠ **Simple test** $\mathcal{T}_\phi$ associated with detector ϕ, given observation ω,

- accepts H_1 when $\phi(\omega) \geq 0$,
- accepts H_2 when $\phi(\omega) < 0$.

♣ **Immediate observation:**

$$\boxed{\begin{array}{rcl} \mathrm{Risk}_1(\mathcal{T}_\phi|H_1,H_2) & \leq & \mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] \\ \mathrm{Risk}_2(\mathcal{T}_\phi|H_1,H_2) & \leq & \mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2] \end{array}}$$

Reason: $\mathrm{Prob}_{\omega\sim P}\{\omega : \psi(\omega) \geq 0\} \leq \int e^{\psi(\omega)} P(d\omega)$.

Elementary Calculus of Detectors

$$\boxed{\mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] = \sup_{P\in\mathcal{P}_1} \int_\Omega \mathrm{e}^{-\phi(\omega)} P(d\omega),\ \mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2] = \sup_{P\in\mathcal{P}_2} \int_\Omega \mathrm{e}^{\phi(\omega)} P(d\omega)}$$

♣ Detectors admit simple "calculus:"

♣ **Renormalization:** $\phi(\cdot) \Rightarrow \phi_a(\cdot) = \phi(\cdot) - a$

$$\Rightarrow \left\{ \begin{array}{rcl} \mathrm{Risk}_1[\phi_a|\mathcal{P}_1,\mathcal{P}_2] &=& \mathrm{e}^{a}\mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] \\ \mathrm{Risk}_2[\phi_a|\mathcal{P}_1,\mathcal{P}_2] &=& \mathrm{e}^{-a}\mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2] \end{array} \right.$$

$\Rightarrow$ *What matters, is the product*

$$[\mathrm{Risk}[\phi|\mathcal{P}_1,\mathcal{P}_2]]^2 := \mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2]\mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2]$$

of partial risks of a detector. Shifting the detector by constant, we can distribute this product between factors as we want, e.g., always can make the detector balanced:

$$\mathrm{Risk}[\phi|\mathcal{P}_1,\mathcal{P}_2] = \mathrm{Risk}_1[\phi|\mathcal{P}_1,\mathcal{P}_2] = \mathrm{Risk}_2[\phi|\mathcal{P}_1,\mathcal{P}_2].$$

♣ **Detectors are well-suited for passing to multiple observations.** For $1 \le k \le K$, let

- $\mathcal{P}_{1,k}, \mathcal{P}_{2,k}$ be families of probability distributions on observation spaces Ω_k,
- ϕ_k be detectors on Ω_k.

♡ Families $\{\mathcal{P}_{1,k}, \mathcal{P}_{2,k}\}_{k=1}^K$ give rise to families of product distributions on $\Omega^K = \Omega_1 \times ... \times \Omega_K$:

$$\begin{array}{l}\mathcal{P}_1^K = \{P^K = P_1 \times ... \times P_K : P_k \in \mathcal{P}_{1,k},\ 1 \le k \le K\},\\ \mathcal{P}_2^K = \{P^K = P_1 \times ... \times P_K : P_k \in \mathcal{P}_{2,k},\ 1 \le k \le K\},\end{array}$$

and detectors $\phi_1, .., \phi_K$ give rise to detector ϕ^K on Ω^K:

$$\phi^K(\underbrace{\omega_1, ..., \omega_K}_{\omega^K}) = \sum_{k=1}^K \phi_k(\omega_k).$$

♠ **Observation:** *For $\chi = 1, 2$, we have*

$$\text{Risk}_\chi[\phi^K|\mathcal{P}_1^K, \mathcal{P}_2^K] = \prod_{k=1}^K \text{Risk}_\chi[\phi_k|\mathcal{P}_{1,k}, \mathcal{P}_{2,k}]. \qquad (!)$$

$$\boxed{\phi^K(\underbrace{\omega_1,...,\omega_K}_{\omega^K}) = \sum_{k=1}^K \phi_k(\omega_k).}$$

♡ In the sequel, we refer to families $\mathcal{P}_\chi^K$ as to *direct products* of families of distributions $\mathcal{P}_{\chi,k}$ over $1 \leq k \leq K$:

$$\mathcal{P}_\chi^K = \mathcal{P}_\chi^{\oplus,K} = \bigoplus_{k=1}^K \mathcal{P}_{\chi,k} := \{P^K = P_1 \times ... \times P_K : P_k \in \mathcal{P}_{\chi.k}, 1 \leq k \leq K\}.$$

We can define also *quasi-direct products*

$$\mathcal{P}_\chi^{\otimes,K} = \bigotimes_{k=1}^K \mathcal{P}_{\chi,k}$$

of the families $\mathcal{P}_{\chi,k}$ over $1 \leq k \leq K$. By definition, $\mathcal{P}_\chi^{\otimes,K}$ is comprised of al distributions P^K of random sequences $\omega^K = (\omega_,...,\omega_K)$, $\omega_k \in \Omega_k$, which can be generated as follows: in the nature there exists a random sequence $\zeta^K = (\zeta_1,...,\zeta_K)$ of “driving factors” such that for every $k \leq K$, ω_k is a deterministic function of $\zeta^k = (\zeta_1,...,\zeta_k)$, and the conditional, ζ^{k-1} being fixed, distribution of ω_k always belongs to $\mathcal{P}_{\chi,k}$.

♠ *It is immediately seen that for $\chi = 1, 2$ it holds*

$$\mathrm{Risk}_\chi[\phi^K|\mathcal{P}_1^{\otimes,K}, \mathcal{P}_2^{\otimes,K}] = \prod_{k=1}^K \mathrm{Risk}_\chi[\phi_k|\mathcal{P}_{1,k}, \mathcal{P}_{2,k}].$$

♣ **From pairwise detectors to detectors for unions.** Assume that we are given an observation space Ω along with

- R families $\mathcal{R}_r$, $r = 1, ..., R$ of "red" probability distributions on Ω,
- B families $\mathcal{B}_b$, $b = 1, ..., B$ of "brown" probability distributions on Ω,
- detectors $\phi_{rb}(\cdot)$, $1 \leq r \leq R$, $1 \leq b \leq B$.

Let us aggregate the red and the brown families as follows

$$\mathcal{R} = \bigcup_{r=1}^{R} \mathcal{R}_r,\ \mathcal{B} = \bigcup_{b=1}^{B} \mathcal{B}_b$$

and assemble detectors ϕ_{rb} into a single detector

$$\phi(\omega) = \max_{r\leq R}\min_{b\leq B}\phi_{rb}(\omega).$$

♠ **Observation:** *We have*

$$\begin{array}{rcl}\mathrm{Risk}_1[\phi|\mathcal{R},\mathcal{B}] & \leq & \max_{r\leq R}\sum_{b\leq B}\mathrm{Risk}_1[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b],\\ \mathrm{Risk}_2[\phi|\mathcal{R},\mathcal{B}] & \leq & \max_{b\leq B}\sum_{r\leq R}\mathrm{Risk}_2[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b].\end{array}$$

♠ **Observation:** *We have*

$$\begin{array}{rcl}\mathrm{Risk}_1[\phi|\mathcal{R},\mathcal{B}] & \leq & \max_{r\leq R}\sum_{b\leq B}\mathrm{Risk}_1[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b],\\ \mathrm{Risk}_2[\phi|\mathcal{R},\mathcal{B}] & \leq & \max_{b\leq B}\sum_{r\leq R}\mathrm{Risk}_2[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b].\end{array}$$

Indeed,

$$\begin{array}{rl} P\in\mathcal{R}_{r_*}\Rightarrow & \int e^{-[\max_r \min_b \phi_{rb}(\omega)]}P(d\omega)=\int e^{\min_r \max_b[-\phi_{rb}(\omega)]}P(d\omega)\\ & \leq \int e^{\max_b[-\phi_{r_*b}(\omega)]}P(d\omega)\leq \sum_b\int e^{-\phi_{r_*b}(\omega)}P(d\omega)\leq \sum_b \mathrm{Risk}_1[\phi_{r_*b}|\mathcal{R}_{r_*},\mathcal{B}_b]\\ \Rightarrow & \mathrm{Risk}_1[\phi|\mathcal{R},\mathcal{B}]\leq \max_{r\leq R}\sum_{b\leq B}\mathrm{Risk}_1[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b];\\ P\in\mathcal{B}_{b_*}\Rightarrow & \int e^{\max_r \min_b \phi_{rb}(\omega)}P(d\omega)\leq\int e^{\max_r \phi_{rb_*}(\omega)}P(d\omega)\\ & \leq\sum_r\int e^{\phi_{rb_*}(\omega)}P(d\omega)\leq\sum_r \mathrm{Risk}_2[\phi_{rb_*}|\mathcal{R}_r,\mathcal{B}_{b_*}]\\ \Rightarrow & \mathrm{Risk}_2[\phi|\mathcal{R},\mathcal{B}]\leq \max_{b\leq B}\sum_{r\leq R}\mathrm{Risk}_2[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b].\end{array}$$

♠ **Refinement:** W.l.o.g. we can assume that the detectors ϕ_{rb} are balanced:

$$\epsilon_{rb} := \mathrm{Risk}[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b] = \mathrm{Risk}_1[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b] = \mathrm{Risk}_2[\phi_{rb}|\mathcal{R}_r,\mathcal{B}_b].$$

Consider matrices

$$E = \left[\begin{array}{ccc} \epsilon_{1,1} & \cdots & \epsilon_{1,B} \\ \vdots & \cdots & \vdots \\ \epsilon_{R,1} & \cdots & \epsilon_{R,B} \end{array}\right],\ F = \left[\begin{array}{c|c} & E \\ \hline E^T & \end{array}\right]$$

♡ The maximal eigenvalue θ of F is the spectral norm $\|E\|_{2,2}$ of E, and the leading eigenvector $[g; f]$ of F can be selected to be positive (*Perron-Frobenius Theorem*).

Note: $\theta g = Ef$ & $\theta f = E^T g$

♡ Let us pass from the detectors ϕ_{rb} to shifted detectors $\psi_{rb} = \phi_{rb} - \ln(f_b/g_r)$ and assemble the shifted detectors into the detector

$$\psi(\omega) = \max_{r\le R}\min_{b\le B}\psi_{rb}(\omega)$$

By previous observation

$$\begin{array}{l} \mathrm{Risk}_1(\psi|\mathcal{R},\mathcal{B}) \le \max_r\sum_b \mathrm{Risk}_1[\psi_{rb}|\mathcal{R}_r,\mathcal{B}_b] = \max_r\sum_b \epsilon_{rb}(f_b/g_r) \\ \quad = \max_r[(Ef)_r/g_r] = \theta = \|E\|_{2,2} \\ \mathrm{Risk}_2(\psi|\mathcal{R},\mathcal{B}) \le \max_b\sum_r \mathrm{Risk}_2[\psi_{rb}|\mathcal{R}_r,\mathcal{B}_b] = \max_b\sum_r \epsilon_{rb}(g_r/f_b) \\ \quad = \max_b[(E^Tg)_b/f_b] = \theta = \|E\|_{2,2} \end{array}$$

$\Rightarrow$ *Partial risks of detector ψ on aggregated families $\mathcal{R}$, $\mathcal{B}$ are $\le \|E\|_{2,2}$.*

Detector-Based Tests "Up to Closeness"

♠ **Situation:** We are given

• L families of probability distributions $\mathcal{P}_\ell$, $\ell = 1, ..., L$, on observation space Ω, giving rise to L hypotheses H_ℓ, on the distribution P of random observation $\omega \in \Omega$:

$$H_\ell : P \in \mathcal{P}_\ell, \ 1 \leq \ell \leq L;$$

• closeness relation $\mathcal{C}$;

• system of balanced detectors

$$\left\{\phi_{\ell\ell'} : \ell < \ell', (\ell, \ell') \notin \mathcal{C}\right\}$$

along with upper bounds $\epsilon_{\ell\ell'}$ on detectors' risks:

$$\forall(\ell, \ell' : \ell < \ell', (\ell, \ell') \notin \mathcal{C}) : \left\{ \begin{array}{l} \int_\Omega \mathrm{e}^{-\phi_{\ell\ell'}(\omega)} P(d\omega) \leq \epsilon_{\ell\ell'} \ \forall P \in \mathcal{P}_\ell \\ \int_\Omega \mathrm{e}^{\phi_{\ell\ell'}(\omega)} P(d\omega) \leq \epsilon_{\ell\ell'} \ \forall P \in \mathcal{P}_{\ell'} \end{array} \right.$$

• Our goal is to build single-observation test deciding on hypotheses $H_1, ..., H_L$ up to closeness $\mathcal{C}$.

♠ **Construction:** Let us set

$$\phi_{\ell\ell'}(\omega) = \left\{ \begin{array}{ll} -\phi_{\ell'\ell}(\omega), & \ell > \ell', (\ell,\ell') \not\in \mathcal{C} \\ 0, & (\ell,\ell') \not\in \mathcal{C} \end{array} \right., \ \epsilon_{\ell\ell'} = \left\{ \begin{array}{ll} \epsilon_{\ell'\ell}, & \ell > \ell', (\ell,\ell') \not\in \mathcal{C} \\ 1, & (\ell,\ell') \not\in \mathcal{C} \end{array} \right.,$$

thus ensuring that

$$\begin{array}{c} \phi_{\ell\ell'}(\cdot) \equiv -\phi_{\ell'\ell}(\cdot),\ \epsilon_{\ell\ell'} = \epsilon_{\ell'\ell},\ 1 \leq \ell, \ell' \leq L \\ \int_\Omega \mathrm{e}^{-\phi_{\ell\ell'}(\omega)} P(d\omega) \leq \epsilon_{\ell\ell'}\, \forall (P \in \mathcal{P}_\ell,\ 1 \leq \ell, \ell' \leq L) \end{array}$$

• Given shifts $a_{\ell\ell'} = -a_{\ell'\ell}$, we specify test $\mathcal{T}$ as follows: *Given observation ω, $\mathcal{T}$ accepts all hypotheses H_ℓ such that*

$$\phi_{\ell\ell'}(\omega) > a_{\ell\ell'}\ \forall (\ell' : (\ell,\ell') \not\in \mathcal{C})$$

and rejects all other hypotheses.

♠ **Proposition.** *The $\mathcal{C}$-risk of $\mathcal{T}$ can be upper-bounded as*

$$\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L) \leq \max_{\ell \leq L} \sum_{\ell' : (\ell,\ell') \not\in \mathcal{C}} \epsilon_{\ell\ell'} \mathrm{e}^{a_{\ell\ell'}}$$

♠ **Optimal shifts:** Consider the symmetric nonnegative matrix

$$E = [\epsilon_{\ell\ell'}\chi_{\ell\ell'}]_{\ell,\ell'=1}^{L},\ \chi_{\ell\ell'} = \left\{\begin{array}{ll} 1, & (\ell,\ell')\not\in\mathcal{C} \\ 0, & (\ell,\ell')\in\mathcal{C}\end{array}\right.,$$

and let $\theta = \|E\|_{2,2}$ be the spectral norm of E, or, which is the same under the circumstances, the largest eigenvalue of E. By Perron-Frobenius Theorem, for every $\theta' > \theta$ there exists a positive vector f such that

$$Ef \leq \theta' f;$$

the same holds true when $\theta' = \theta$, provided the leading eigenvector of E (which always can selected to be nonnegative) is positive.

Fact: *With $\alpha_{\ell\ell'} = \ln(f_{\ell'}/f_\ell)$, the risk bound from Proposition reads*

$$\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L) \leq \theta'.$$

Thus, assembling the detectors $\phi_{\ell\ell'}$ appropriately, one can get a test with $\mathcal{C}$-risk arbitrarily close to $\|E\|_{2,2}$.

♠ **Utilizing repeated observations.** Assuming K-repeated observations allowed, we can apply the above construction to

- K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$ in the role of ω,
- quasi-direct powers $\mathcal{P}_\ell^{\otimes,K} = \mathcal{P}_\ell \otimes ... \otimes \mathcal{P}_\ell$ of families $\mathcal{P}_\ell$ in the role of these families, and respective hypotheses $H_\ell^{\otimes,K}$ in the role of hypotheses H_ℓ,
- detectors $\phi_{\ell\ell'}^{(K)}(\omega^K) = \sum_{k=1}^K \phi_{\ell\ell'}(\omega_k)$ in the role of detectors $\phi_{\ell\ell'}$, which allows to replace $\epsilon_{\ell\ell'}$ with $\epsilon_{\ell\ell'}^K$.

As a result, we get K-observation test $\mathcal{T}^K$ such that

$$\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}^K | H_1^{\otimes,K}, ..., H_L^{\otimes,K}) \leq \theta'_K$$

where θ'_K can be made arbitrarily close (under favorable circumstances, even equal) to the quantity

$$\theta_K = \left\| \left[\epsilon_{\ell\ell'}^K \chi_{\ell\ell'} \right]_{\ell\ell'=1}^K \right\|_{2,2}, \ \chi_{\ell\ell'} = \begin{cases} 1, & (\ell,\ell') \notin \mathcal{C} \\ 0, & (\ell,\ell') \in \mathcal{C} \end{cases}$$

In particular, in the case when $\epsilon_{\ell\ell'} < 1$ whenever $(\ell,\ell') \notin \mathcal{C}$, we can ensure that the $\mathcal{C}$-risk of $\mathcal{T}^K$ converges to 0 exponentially fast as $K \to \infty$.

♣ **"Universality" of detector-based tests.** *Let $\mathcal{P}_\chi$, $\chi = 1, 2$, be two families of probability distributions on observation space Ω, and let H_χ, $\chi = 1, 2$, be associated hypotheses on the distribution of an observation.*

Assume that there exists a simple deterministic or randomized test $\mathcal{T}$ deciding on H_1, H_2 with risk $\leq \epsilon \in (0, 1/2)$. Then there exists a detector ϕ with

$$\mathrm{Risk}[\phi|\mathcal{P}_1, \mathcal{P}_2] \leq \epsilon_+ := 2\sqrt{\epsilon[1-\epsilon]} < 1.$$

Indeed, let $\mathcal{T}$ be deterministic, let $\Omega_\chi = \{\omega \in \Omega : \mathcal{T}(\omega) = \{\chi\}\}$, $\chi = 1, 2$, and let

$$\phi(\omega) = \begin{cases} \frac{1}{2}\ln\left([1-\epsilon]/\epsilon\right), & \omega \in \Omega_1 \\ \frac{1}{2}\ln\left(\epsilon/[1-\epsilon]\right), & \omega \in \Omega_2 \end{cases}$$

Then

$$\boxed{\begin{array}{l} P \in \mathcal{P}_1, \epsilon' = \int_{\Omega_2} P(d\omega)\ [\leq \epsilon] \Rightarrow \\ \int \mathrm{e}^{-\phi(\omega)} P(d\omega) = \sqrt{\epsilon/[1-\epsilon]}(1-\epsilon') + \sqrt{[1-\epsilon]/\epsilon}\,\epsilon' \\ = \sqrt{\epsilon/[1-\epsilon]} + \underbrace{\left[\sqrt{[1-\epsilon]/\epsilon} - \sqrt{\epsilon/[1-\epsilon]}\right]}_{\geq 0} \underbrace{\epsilon'}_{\leq \epsilon} \\ \leq \sqrt{\epsilon/[1-\epsilon]} + \left[\sqrt{[1-\epsilon]/\epsilon} - \sqrt{\epsilon/[1-\epsilon]}\right]\epsilon = 2\sqrt{\epsilon[1-\epsilon]} \\ \hline P \in \mathcal{P}_2, \epsilon' = \int_{\Omega_1} P(d\omega)\ [\leq \epsilon] \Rightarrow \\ \int \mathrm{e}^{\phi(\omega)} P(d\omega) = \sqrt{\epsilon/[1-\epsilon]}(1-\epsilon') + \sqrt{[1-\epsilon]/\epsilon}\,\epsilon' \leq 2\sqrt{\epsilon[1-\epsilon]} \end{array}}$$

$$\Rightarrow \mathrm{Risk}_\chi[\phi|\mathcal{P}_1, \mathcal{P}_2] \leq 2\sqrt{\epsilon[1-\epsilon]}.$$

Now let $\mathcal{T}$ be randomized. Setting $\mathcal{P}_\chi^+ = \{P\times\mathrm{Uniform}[0,1] : P\in\mathcal{P}_\chi\}$, $\chi=1,2$, $\Omega^+ = \Omega\times[0,1]$, by above there exists a bounded detector $\phi_+ : \Omega^+ \to \mathbb{R}$ such that

$$\begin{array}{l}\forall (P\in\mathcal{P}_1) : \int_\Omega \left[\int_0^1 \mathrm{e}^{-\phi_+(\omega,s)}ds\right] P(d\omega)ds \leq \epsilon_+ = 2\sqrt{\epsilon[1-\epsilon]},\\ \forall (P\in\mathcal{P}_2) : \int_\Omega \left[\int_0^1 \mathrm{e}^{\phi_+(\omega,s)}ds\right] P(d\omega) \leq \epsilon_+,\end{array}$$

whence, setting $\phi(\omega) = \int_0^1 \phi(\omega,s)ds$ and applying Jensen's Inequality,

$$\begin{array}{l}\forall (P\in\mathcal{P}_1) : \int_\Omega \mathrm{e}^{-\phi(\omega)} P(d\omega) \leq \epsilon_+,\\ \forall (P\in\mathcal{P}_2) : \int_\Omega \mathrm{e}^{\phi(\omega)} P(d\omega) \leq \epsilon_+\end{array} \qquad \square$$

♠ Risk $2\sqrt{\epsilon[1-\epsilon]}$ of the detector-based test induced by simple test $\mathcal{T}$ is "much worse" than the risk ϵ of $\mathcal{T}$.

However: *When repeated observations are allowed, we can compensate for risk deterioration* $\epsilon \mapsto 2\sqrt{\epsilon[1-\epsilon]}$ *by passing in the detector-based test from a single observation to a moderate number of them.*

$$\inf_{\phi}\left\{\text{Risk}[\phi|\mathcal{P}_1,\mathcal{P}_2]=\min\left\{\epsilon:\begin{array}{rcl}\int_\Omega \mathrm{e}^{-\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_1)\\ \int_\Omega \mathrm{e}^{\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_2)\end{array}\right\}\right\} \qquad (!)$$

Note:

- The optimization problem specifying risk has constraints *convex* in (ϕ,ϵ)
- When passing from families $\mathcal{P}_\chi$, $\chi=1,2$, to their convex hulls, the risk of a detector remains intact.

♣ **Bottom line:** *It would be nice to be able to solve* $(!)$, *thus arriving at the lowest risk detector-based tests.*

But: (!) is an optimization problem with *infinite-dimensional* decision "vector" and *infinitely many* constraints.

$\Rightarrow$ $(!)$ *in general is intractable.*

Simple observation schemes: A series of special cases where (!) is efficiently solvable via Convex Optimization.

Preliminaries from Convex Programming: Saddle Points

♣ Let $X \subset \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$ be nonempty sets, and let $F(x,\lambda)$ be a real-valued function on $X \times \Lambda$. This function gives rise to two optimization problems

$$\begin{array}{ll}\mathrm{Opt}(P) = \inf\limits_{x\in X} \overbrace{\sup\limits_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} & (P)\\ \mathrm{Opt}(D) = \sup\limits_{\lambda\in\Lambda} \underbrace{\inf\limits_{x\in X} F(x,\lambda)}_{\underline{\mathbf{F}}(\lambda)} & (D)\end{array}$$

$$
\begin{array}{rcll}
\mathrm{Opt}(P) &=& \inf\limits_{x\in X} \overbrace{\sup\limits_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} & (P)\\
\mathrm{Opt}(D) &=& \sup\limits_{\lambda\in\Lambda} \underbrace{\inf\limits_{x\in X} F(x,\lambda)}_{\underline{\mathbf{F}}(\lambda)} & (D)
\end{array}
$$

Game interpretation: Player I chooses $x \in X$, player II chooses $\lambda \in \Lambda$. With choices of the players x, λ, player I pays to player II the sum of $F(x, \lambda)$. What should the players do to optimize their wealth?

♢If Player I chooses x first, and Player II knows this choice when choosing λ, II will maximize his profit, and the loss of I will be $\overline{F}(x)$. To minimize his loss, I should solve (P), thus ensuring himself loss $\mathrm{Opt}(P)$ or less.

♢If Player II chooses λ first, and Player I knows this choice when choosing x, I will minimize his loss, and the profit of II will be $\underline{\mathbf{F}}(\lambda)$. To maximize his profit, II should solve (D), thus ensuring himself profit $\mathrm{Opt}(D)$ or more.

$$\mathrm{Opt}(P) = \inf_{x\in X} \overbrace{\sup_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} \quad (P)$$
$$\mathrm{Opt}(D) = \sup_{\lambda\in\Lambda} \underbrace{\inf_{x\in X} F(x,\lambda)}_{\underline{\mathbf{F}}(\lambda)} \quad (D)$$

Observation: For Player I, second situation seems better, so that it is natural to guess that his anticipated loss in this situation is $\leq$ his anticipated loss in the first situation:

$$\mathrm{Opt}(D) \equiv \sup_{\lambda\in\Lambda} \inf_{x\in X} F(x,\lambda) \leq \inf_{x\in X} \sup_{\lambda\in\Lambda} F(x,\lambda) \equiv \mathrm{Opt}(P).$$

This indeed is true: assuming $\mathrm{Opt}(P) < \infty$ (otherwise the inequality is evident),

$$\begin{array}{l}
\forall(\epsilon > 0): \quad \exists x_\epsilon \in X: \sup_{\lambda\in\Lambda} F(x_\epsilon,\lambda) \leq \mathrm{Opt}(P) + \epsilon \\
\Rightarrow \forall \lambda \in \Lambda: \underline{\mathbf{F}}(\lambda) = \inf_{x\in X} F(x,\lambda) \leq F(x_\epsilon,\lambda) \leq \mathrm{Opt}(P) + \epsilon \\
\Rightarrow \mathrm{Opt}(D) \equiv \sup_{\lambda\in\Lambda} \underline{\mathbf{F}}(\lambda) \leq \mathrm{Opt}(P) + \epsilon \\
\Rightarrow \mathrm{Opt}(D) \leq \mathrm{Opt}(P).
\end{array}$$

$$\mathrm{Opt}(P) = \inf_{x\in X} \overbrace{\sup_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} \quad (P)$$
$$\mathrm{Opt}(D) = \sup_{\lambda\in\Lambda} \underbrace{\inf_{x\in X} F(x,\lambda)}_{\underline{\mathbf{F}}(\lambda)} \quad (D)$$

♣ What should the players do when making their choices simultaneously?

A "good case" when we can answer this question – F has a *saddle point*.

Definition: *We call a point* $(x_*, \lambda_*) \in X \times \Lambda$ *a saddle point of* F*, if*

$$F(x, \lambda_*) \geq F(x_*, \lambda_*) \geq F(x_*, \lambda) \;\forall (x \in X, \lambda \in \Lambda).$$

In game terms, a saddle point is an *equilibrium* – no one of the players can improve his wealth, provided the adversary keeps his choice unchanged.

Proposition [Existence and Structure of saddle points]: *F has a saddle point if and only if both (P) and (D) are solvable with equal optimal values. In this case, the saddle points of F are exactly the pairs (x_*, λ_*), where x_* is an optimal solution to (P), and λ_* is an optimal solution to (D).*

$$\begin{array}{ll}\mathrm{Opt}(P)=\inf\limits_{x\in X}\overbrace{\sup\limits_{\lambda\in\Lambda}F(x,\lambda)}^{\overline{F}(x)} & (P)\\ \mathrm{Opt}(D)=\sup\limits_{\lambda\in\Lambda}\underbrace{\inf\limits_{x\in X}F(x,\lambda)}_{\underline{\mathbf{F}}(\lambda)} & (D)\end{array}$$

Proof, $\Rightarrow$: Assume that (x_*,λ_*) is a saddle point of F, and let us prove that x_* solves (P), λ_* solves (D), and $\mathrm{Opt}(P)=\mathrm{Opt}(D)$.
Indeed, we have

$$F(x,\lambda_*)\geq F(x_*,\lambda_*)\geq F(x_*,\lambda)\ \forall(x\in X,\lambda\in\Lambda)$$

whence

$$\begin{array}{l}\mathrm{Opt}(P)\leq\overline{F}(x_*)=\sup\limits_{\lambda\in\Lambda}F(x_*,\lambda)=F(x_*,\lambda_*)\\ \mathrm{Opt}(D)\geq\underline{\mathbf{F}}(\lambda_*)=\inf\limits_{x\in X}F(x,\lambda_*)=F(x_*,\lambda_*)\end{array}$$

Since $\mathrm{Opt}(P)\geq\mathrm{Opt}(D)$, we see that all inequalities in the chain

$$\mathrm{Opt}(P)\leq\overline{F}(x_*)=F(x_*,\lambda_*)=\underline{\mathbf{F}}(\lambda_*)\leq\mathrm{Opt}(D)$$

are equalities. Thus, x_* solves (P), λ_* solves (D) and $\mathrm{Opt}(P)=\mathrm{Opt}(D)$.

$$\mathrm{Opt}(P) = \inf_{x\in X} \overbrace{\sup_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} \quad (P)$$
$$\mathrm{Opt}(D) = \sup_{\lambda\in\Lambda} \underbrace{\inf_{x\in X} F(x,\lambda)}_{\underline{F}(\lambda)} \quad (D)$$

Proof, $\Leftarrow$. Assume that (P), (D) have optimal solutions x_*, λ_* and $\mathrm{Opt}(P) = \mathrm{Opt}(D)$, and let us prove that (x_*, λ_*) is a saddle point. We have

$$\begin{array}{l} \mathrm{Opt}(P) = \overline{F}(x_*) = \sup\limits_{\lambda\in\Lambda} F(x_*,\lambda) \geq F(x_*,\lambda_*) \\ \mathrm{Opt}(D) = \underline{\mathbf{F}}(\lambda_*) = \inf\limits_{x\in X} F(x,\lambda_*) \leq F(x_*,\lambda_*) \end{array} \qquad (*)$$

Since $\mathrm{Opt}(P) = \mathrm{Opt}(D)$, all inequalities in $(*)$ are equalities, so that

$$\sup_{\lambda\in\Lambda} F(x_*,\lambda) = F(x_*,\lambda_*) = \inf_{x\in X} F(x,\lambda_*).$$

Existence of Saddle Points

♣ **Theorem [Sion-Kakutani]** *Let $X \subset \mathbb{R}^n$, $\Lambda \subset \mathbb{R}^m$ be nonempty convex closed sets and $F(x,\lambda) : X \times \Lambda \to \mathbb{R}$ be a continuous function which is convex in $x \in X$ and concave in $\lambda \in \Lambda$. Assume that Λ is compact.*

(i) *"MinMax equals MaxMin:" One has*

$$\mathrm{SadVal} := \inf_{x \in X} \sup_{\lambda \in \Lambda} F(x,\lambda) = \sup_{\lambda \in \Lambda} \min_{x \in X} F(x,\lambda)$$

Note: SadVal *is either real, or $-\infty$.*

(ii) *Assume that there exists $\bar{\lambda} \in \Lambda$ such that for every $a \in \mathbb{R}$ the set*

$$X_a : \{x \in X : F(x,\bar{\lambda}) \le a\}$$

is bounded (e.g., since X is bounded).
Then SadVal *is real, and F possesses a saddle point on $X \times \Lambda$.*

Proof of Sion-Kakutani Theorem

MinMax Lemma [von Neumann] *Let X ba a nonempty convex compact set and $f_1, ..., f_N$ be continuous convex functions on X. then the quantity*

$$\mathrm{Opt} = \min_{x \in X} \max[f_1(x), f_2(x), ..., f_N(x)]$$

is the minimum over X of certain convex combination of f_i:

$$\exists \mu^* \in \mathbb{R}^N_+, \sum_i \mu_i^* = 1 : \mathrm{Opt} = \min_x \sum_{i=1}^N \mu_i^* f_i(x).$$

Note: for every collection of nonnegative weights μ_i summing up to one we have $\sum_i \mu_i f_i(x) \le \max_i f_i(x)$ and therefore

$$\min_X \sum_i \mu_i f_i(x) \le \mathrm{Opt}.$$

Proof of MinMax Lemma: Assuming w.l.o.g. that $\mathrm{Opt} = 0$ (replace f_i with $f_i - \mathrm{Opt}$!), consider two convex sets in $\mathbb{R}^N$:

$$S = \{0\},\ T = \{y \in \mathbb{R}^N : \exists x \in X : y \geq f(x) := [f_1(x); ...; f_N(x)]\}.$$

From convexity of X and f_i's it follows that T is convex. Besides this, T clearly possesses a nonempty interior.
We claim that $S = \{0\} \not\in \mathrm{int}\, T$. Indeed, assuming the opposite, T contains a negative vector, whence, by definition of T, $f_i(\bar{x}) < 0$ for some $\bar{x} \in X$ and all i, so that $\min_X \max_i f_i(x) < 0$, while we are in the case $\mathrm{Opt} = 0$.
By Separation Theorem, the fact that $S = \{0\} \not\in \mathrm{int}\, T \neq \emptyset$ implies that S and T can be separated: there exists $\lambda = [\lambda_1; ...; \lambda_N] \neq 0$ such that

$$0 = \max_{s \in S} \lambda^T s \leq \inf_{y \in Y} \sum_i \lambda_i y_i. \qquad (*)$$

since T contains all positive vectors with large enough entries, $(*)$ implies that $\lambda \geq 0$, and since $f(x) \in T$ for all $x \in X$, $(*)$ says that

$$\mathrm{Opt} = 0 \leq \sum_i \lambda_i f_i(x)\ \forall x \in X \qquad (!)$$

Since $0 \neq \lambda \geq 0$, the weights $\mu_i^* = \lambda_i / \sum_j \lambda_j$ are well defined, nonnegative, sum up to 1, and by $(!)$ we have

$$\mathrm{Opt} = 0 \leq \sum_i \mu_i^* f_i(x)\ \forall x \in X \qquad \square$$

Proof of Sion-Kakutani Theorem: We should prove that problems

$$\begin{array}{rcl}\mathrm{Opt}(P) & = & \inf\limits_{x\in X} \overbrace{\sup\limits_{\lambda\in\Lambda} F(x,\lambda)}^{\overline{F}(x)} \quad (P)\\ \mathrm{Opt}(D) & = & \sup\limits_{\lambda\in\Lambda} \underbrace{\inf\limits_{x\in X} F(x,\lambda)}_{\underline{F}(\lambda)} \quad (D)\end{array}$$

are solvable with equal optimal values.

1^0. Since X is compact and $F(x,\lambda)$ is continuous on $X\times\Lambda$, the function $\underline{F}(\lambda)$ is continuous on Λ. Besides this, the sets

$$\Lambda^a = \{\lambda\in\Lambda : \underline{F}(\lambda)\geq a\}$$

are contained in the sets

$$\Lambda_a = \{\lambda\in\Lambda : F(\bar{x},\lambda)\geq a\}$$

and therefore are bounded. Finally, Λ is closed, so that the *continuous* function $\underline{F}(\cdot)$ with *bounded* level sets Λ^a attains it maximum on a *closed* set Λ. Thus, (D) is solvable.

2^0. Consider the sets

$$X(\lambda) = \{x \in X : F(x,\lambda) \leq \mathrm{Opt}(D)\}.$$

These are closed convex subsets of a compact set X. Let us prove that every finite collection of these sets has a nonempty intersection. Indeed, assume that $X(\lambda^1) \cap ... \cap X(\lambda^N) = \emptyset$, so that

$$\max_{j=1,...,N} F(x,\lambda^j) > \mathrm{Opt}(D)\ \forall x \in X$$

$$\Rightarrow \min_{x\in X} \max_j F(x,\lambda^j) > \mathrm{Opt}(D)$$

by compactness of X and continuity of F.
By MinMax Lemma, there exist weights $\mu_j \geq 0, \sum_j \mu_j = 1$, such that

$$\min_{x\in X} \underbrace{\sum_j \mu_j F(x,\lambda^j)}_{\substack{\leq F(x,\sum_j \mu_j \lambda_j) \\ \text{since } F \text{ is concave in } \lambda}} > \mathrm{Opt}(D),$$

$$\Rightarrow \underline{F}(\sum_j \mu_j \lambda_j) := \min_{x\in X} F(x,\sum_j \mu_j \lambda_j) \geq \min_{x\in X} \sum_j \mu_j F(x,\lambda_j) > \mathrm{Opt}(D),$$

which is impossible.

3^0. Since every finite collection of closed convex subsets $X(\lambda)$ of the compact set X has a nonempty intersection, all these sets have a nonempty intersection:

$$\exists x_* \in X : F(x_*, \lambda) \leq \mathtt{Opt}(D)\ \forall \lambda.$$

Due to $\mathtt{Opt}(P) \geq \mathtt{Opt}(D)$, this is possible iff x_* is optimal for (P) and $\mathtt{Opt}(P) = \mathtt{Opt}(D)$.

Simple Observation Schemes

♣ **Simple Observation Scheme** is a collection

$$\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F}),$$

where

• (Ω, Π) is a (complete separable metric) *observation space* Ω with (σ-finite σ-additive) *reference measure* Π,

$$\mathrm{supp}\,\Pi = \Omega;$$

• $\{p_\mu(\cdot) : \mu \in \mathcal{M}\}$ is a parametric family of probability densities, taken w.r.t. Π, on Ω, and

• $\mathcal{M}$ is a relatively open *convex* set in some $\mathbb{R}^n$

• $p_\mu(\omega)$: *positive* and continuous in $\mu \in \mathcal{M}, \omega \in \Omega$

• $\mathcal{F}$ is a *finite-dimensional* space of continuous functions on Ω containing constants and such that

$$\ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}\ \forall \mu, \nu \in \mathcal{M}$$

• *For $\phi \in \mathcal{F}$, the function*

$$\mu \mapsto \ln\left(\int_\Omega \mathrm{e}^{\phi(\omega)} p_\mu(\omega) P(d\omega)\right)$$

is finite and concave in $\mu \in \mathcal{M}$.

♠ Example 1: Gaussian o.s.

- $(\Omega, \Pi) = (\mathbb{R}^d, \text{mes}_d)$ is $\mathbb{R}^d$ with Lebesgue measure,
- $\{p_\mu(\cdot) = \mathcal{N}(\mu, I_d) : \mu \in \mathbb{R}^d\}$,
- $\mathcal{F} = \{\text{affine functions on } \Omega\} \Rightarrow \left\{ \begin{array}{l} \ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}, \\ \ln\left(\int\limits_\Omega e^{a^T\omega+b} p_\mu(\omega)\Pi(d\omega)\right) = a^T\mu + b + \frac{a^Ta}{2} : \text{ is concave in } \mu. \end{array} \right.$
- Gaussian o.s. is the standard observation model in Signal Processing.

♠ Example 2: Poisson o.s.

- (Ω, Π), is the nonnegative part $\mathbf{Z}^d_+$ of integer lattice in $\mathbb{R}^d$ equipped with counting measure,
- $\{p_\mu(\omega) = \prod\limits_{i=1}^d \frac{\mu_i^{\omega_i} e^{-\mu_i}}{\omega_i!} : \mu \in \mathcal{M} := \mathbb{R}^d_{++}\}$ is the family of distributions of random vectors with independent across i Poisson entries $\omega_i \sim \text{Poisson}(\mu_i)$,
- $\mathcal{F} = \{\text{affine functions on } \Omega\} \Rightarrow \left\{ \begin{array}{l} \ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}, \\ \ln\left(\int\limits_\Omega e^{a^T\omega+b} p_\mu(\omega)\Pi(d\omega)\right) = b + \sum_i (e^{a_i} - 1)\mu_i \text{ is concave in } \mu. \end{array} \right.$

Poisson o.s. arises in *Poisson Imaging*, including

- *Positron Emission Tomography*,
- *Large binocular Telescope*,
- *Nanoscale Fluorescent Microscopy*.

♠ **Example 3: Discrete o.s.**

• (Ω, Π) is finite set $\{1, ..., d\}$ with counting measure,

• $\{p_\mu(\omega) = \mu_\omega,\ \mu \in \mathcal{M} = \{\mu > 0 : \sum_{\omega=1}^d \mu_\omega = 1\}\}$ is the set of non-vanishing probability distributions on Ω,

• $\mathcal{F} = \{\text{all functions on } \Omega\} \Rightarrow \left\{\begin{array}{l} \ln(p_\mu(\cdot)/p_\nu(\cdot)) \in \mathcal{F}, \\ \ln\left(\int\limits_\Omega e^{\phi(\omega)} p_\mu(\omega)\Pi(d\omega)\right) = \ln\left(\sum_{\omega\in\Omega} e^{\phi(\omega)}\mu_\omega\right) \text{ is concave in } \mu. \end{array}\right.$

♠ **Example 4: Direct product of simple o.s.'s.**

Simple o.s.'s

$$\mathcal{O}_k = \Big((\Omega_k, \Pi_k), \{p_{\mu_k,k}(\cdot) : \mu_k \in \mathcal{M}_k\}, \mathcal{F}_k\Big), 1 \le k \le K$$

give rise to their *direct product* $\bigotimes_{k=1}^K \mathcal{O}_k$ defined as the o.s.

$$\Big((\Omega^K, \Pi^K), \{p_{\mu^K}(\cdot) : \mu^K \in \mathcal{M}^K\}, \mathcal{F}^K\Big),$$

where

- $\Omega^K = \Omega_1 \times, ... \times \Omega_K$, $\Pi^K = \Pi_1 \times ... \times \Pi_K$
- $\mathcal{M}^K = \mathcal{M}_1 \times ... \times \mathcal{M}_K$, $p_{(\mu_1,...,\mu_K)}(\omega_1, ..., \omega_K) = \prod\limits_{k=1}^K p_{\mu_k,k}(\omega_k)$
- $\mathcal{F}^K = \{\phi(\underbrace{\omega_1, ..., \omega_K}_{\omega^K}) = \sum_{k=1}^K \phi_k(\omega_k) : \phi_k \in \mathcal{F}_k,\ 1 \le k \le K\}$

♡ **Fact:** *Direct product of simple o.s.'s is a simple o.s.*

♠ **Example 5: Power of a simple o.s.**

When all K o.s.'s in direct product $\mathcal{O}^K = \bigotimes_{k=1}^K \mathcal{O}_k$ are identical to each other:

$$\mathcal{O}_k = \mathcal{O} := ((\Omega, \Pi), \{p_\mu(\cdot) : \mu \in \mathcal{M}\}, \mathcal{F}),\ 1 \le k \le K$$

we can "restrict $\mathcal{O}^K$ to its diagonal," arriving at *K-th power $\mathcal{O}^{(K)}$ of $\mathcal{O}$*:

$$\mathcal{O}^{(K)} = \left((\Omega^K, \Pi^K), \{p_\mu^{(K)}(\cdot) : \mu \in \mathcal{M}\}, \mathcal{F}^{(K)}\right),$$

$$p_\mu^{(K)}(\omega_1, ..., \omega_K) = \prod_{k=1}^K p_\mu(\omega_k),\ \mathcal{F}^{(K)} = \{\phi^{(K)}(\omega^K) = \sum_{k=1}^K \phi(\omega_k) : \phi \in \mathcal{F}\}$$

♡ **Fact:** *Power of a simple o.s. is a simple o.s.*

$$\boxed{\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2) = \min_{\phi(\cdot),\epsilon}\left\{\epsilon : \begin{array}{lcl}\int_\Omega e^{-\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_1)\\ \int_\Omega e^{\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_2)\end{array}\right\}} \qquad (!)$$

♣ **Main Result.** *Let* $\mathcal{O} = ((\Omega,\Pi),\{p_\mu(\cdot):\mu\in\mathcal{M}\},\mathcal{F})$ *be a simple o.s., and let* M_1, M_2 *be two nonempty compact convex subsets of* $\mathcal{M}$. *These subsets give rise to two families of probability distributions* $\mathcal{P}_1$, $\mathcal{P}_2$ *on* Ω *and two hypotheses on the distribution* P *of random observation* $\omega\in\Omega$*:*

$$\mathcal{P}_\chi = \{P: \text{ the density of } P \text{ is } p_\mu \text{ with } \mu\in\mathcal{M}_\chi\},\ H_\chi : P\in\mathcal{P}_\chi,\ \chi=1,2.$$

Consider the function

$$\Phi(\phi;\mu,\nu) = \tfrac{1}{2}\left[\ln\left(\int_\Omega e^{-\phi(\omega)}p_\mu(\omega)\Pi(d\omega)\right) + \ln\left(\int_\Omega e^{\phi(\omega)}p_\nu(\omega)\Pi(d\omega)\right)\right] : \mathcal{F}\times[M_1\times M_2]\to\mathbb{R}.$$

Then

A. $\Phi(\phi;\mu,\nu)$ *is continuous on its domain, convex in* $\phi\in\mathcal{F}$, *concave in* (μ,ν) *on* $\mathcal{M}_1\times\mathcal{M}_2$ *and possesses saddle point (*min *in* ϕ, max *in* (μ,ν)*):*

$$\begin{array}{l}\exists(\phi_*\in\mathcal{F},(\mu^*,\nu^*)\in M_1\times M_2):\\ \Phi(\phi;\mu^*,\nu^*)\geq\Phi(\phi_*;\mu^*,\nu^*)\geq\Phi(\phi_*;\mu,\nu)\ \forall(\phi\in\mathcal{F},(\mu,\nu)\in M_1\times M_2)\end{array}$$

$$\boxed{\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)=\min_{\phi(\cdot),\epsilon}\left\{\epsilon:\begin{array}{lcl}\int_\Omega \mathrm{e}^{-\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_1)\\ \int_\Omega \mathrm{e}^{\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_2)\end{array}\right\}} \qquad (!)$$

$$\boxed{\begin{array}{r}\Phi(\phi;\mu,\nu)=\frac{1}{2}\left[\ln\left(\int_\Omega \mathrm{e}^{-\phi(\omega)}p_\mu(\omega)\Pi(d\omega)\right)+\ln\left(\int_\Omega \mathrm{e}^{\phi(\omega)}p_\nu(\omega)\Pi(d\omega)\right)\right]:\\ \mathcal{F}\times[M_1\times M_2]\to\mathbb{R}.\end{array}}$$

B. *The component ϕ_* of a saddle point $(\phi_*,(\mu^*,\nu^*))$ of Φ is an optimal solution to (!), and*

$$\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)=\exp\{\Phi(\phi_*;\mu^*,\nu^*)\}.$$

$$\boxed{\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)=\min_{\phi(\cdot),\epsilon}\left\{\epsilon:\begin{array}{lcl}\int_\Omega e^{-\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_1)\\ \int_\Omega e^{\phi(\omega)}P(d\omega) & \leq & \epsilon\,\forall(P\in\mathcal{P}_2)\end{array}\right\}} \qquad (!)$$

$$\boxed{\begin{array}{r}\Phi(\phi;\mu,\nu)=\frac{1}{2}\left[\ln\left(\int_\Omega e^{-\phi(\omega)}p_\mu(\omega)\Pi(d\omega)\right)+\ln\left(\int_\Omega e^{\phi(\omega)}p_\nu(\omega)\Pi(d\omega)\right)\right]:\\ \mathcal{F}\times[M_1\times M_2]\to\mathbb{R}.\end{array}}$$

C. *A saddle point $(\phi_*,(\mu^*,\nu^*))$ can be found as follows. We solve the optimization problem*

$$\mathrm{SadVal}=\max_{\mu\in M_1,\nu\in M_2}\ln\left(\int_\Omega\sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right);$$

which is a solvable convex optimization problem, and take an optimal solution to the problem as (μ^*,ν^*). We then set

$$\phi_*(\omega)=\tfrac{1}{2}\ln\left(p_{\mu^*}(\omega)/p_{\nu^*}(\omega)\right),$$

thus getting an optimal detector $\phi_*\in\mathcal{F}$. For this detector and the associated simple test $\mathcal{T}_{\phi_*}$,

$$\begin{array}{c}\mathrm{Risk}(\mathcal{T}_{\phi_*}|H_1,H_2)\leq\mathrm{Risk}[\phi_*|\mathcal{P}_1,\mathcal{P}_2]=\mathrm{Risk}_1[\phi_*|\mathcal{P}_1,\mathcal{P}_2]=\mathrm{Risk}_2[\phi_*|\mathcal{P}_1,\mathcal{P}_2]\\ =\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)=e^{\mathrm{SadVal}}=\int_\Omega\sqrt{p_{\mu^*}(\omega)p_{\nu^*}(\omega)}\Pi(d\omega).\end{array}$$

Informal explanation of Main Result

A. Question: Assume that we are given two distributions, one with density $p(\omega) > 0$, and another with density $q(\omega) > 0$, What is the smallest risk detector for the "families" $\mathcal{P}_1 = \{p\}$ and $\mathcal{P}_2 = \{q\}$?
Answer: We want to solve the problem

$$\min_{\phi(\cdot)} \max \left[\int_\Omega \exp\{-\phi(\omega)\} p(\omega) \Pi(d\omega), \int_\Omega \exp\{\phi(\omega)\} q(\omega) \Pi(d\omega) \right].$$

As we remember, what matters is the product of partial risks; shifting $\phi(\cdot)$ by constant, we can redistribute the product between the factors as we want.
$\Rightarrow$ *All we need is to solve the problem*

$$\min_{\phi(\cdot)} \frac{1}{2} \left[\ln \left(\int_\Omega \exp\{-\phi(\omega)\} p(\omega) \Pi(d\omega) \right) + \ln \left(\int_\Omega \exp\{\phi(\omega)\} q(\omega) \Pi(d\omega) \right) \right]$$

The (balanced) optimal solution is just $\phi_*(\omega) = \frac{1}{2} \ln (p(\omega)/q(\omega))$, and its risk on the pair $\{p\}$, $\{q\}$ is $\int_\Omega \sqrt{p(\omega)q(\omega)} \Pi(d\omega)$. The simplest way to see it is represent a candidate solution in the form of $\phi_*(\omega) + \delta(\omega)$ and to note that in terms of $\delta(\cdot)$ the objective to be minimized becomes

$$\Phi[\delta] = \frac{1}{2} \left[\ln \left(\int_\Omega \exp\{-\delta(\omega)\} \sqrt{p(\omega)q(\omega)} \Pi(d\omega) \right) + \ln \left(\int_\Omega \exp\{\delta(\omega)\} \sqrt{p(\omega)q(\omega)} \Pi(d\omega) \right) \right]$$

We see that $\Phi[\delta]$ *is convex and even functional of* $\delta(\cdot)$, *and thus it attains its minimum when* $\delta(\cdot) = 0$.
Note: *We lose nothing when assuming that we select the best detector from some linear space $\mathcal{F}$ of functions on Ω rather than from the space of all functions on Ω; all that matters is for $\mathcal{F}$ to contain $\ln(p(\cdot)/q(\cdot))$.*

B. Now let us try to find the minimum risk detector for "massive" families of probability densities $\mathcal{P}_1 = \{p_\mu(\cdot) : \mu \in M_1\}$, $\mathcal{P}_2 = \{p_\mu(\cdot) : \mu \in M_2\}$, where $\{p_\mu(\cdot), \mu \in \mathcal{M}\}$ is a parametric family of positive probability densities, and M_1 and M_2 are given subsets of $\mathcal{M}$.
By the same "redistributing partial risks" argument all we need is to solve the optimization problem

$$\mathrm{Opt} = \min_{\phi(\cdot)} \frac{1}{2}\left[\max_{\mu\in M_1} \ln\left(\int_\Omega \exp\{-\phi(\omega)\} p_\mu(\omega)\Pi(d\omega)\right) + \max_{\nu\in M_2} \ln\left(\int_\Omega \exp\{\phi(\omega)\} p_\nu(\omega)\Pi(d\omega)\right)\right]$$

• Let us look at all pairs $p_\mu(\cdot)$, $p_\nu(\cdot)$ with $\mu \in M_1$ and $\nu \in M_2$ and at the optimal for these pairs detectors $\phi_{\mu\nu}(\omega) = \frac{1}{2}\ln(p_\mu(\omega)/p_\nu(\omega))$ and their risks $\int_\Omega \sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)$. These risks clearly lower-bound Opt.
$\Rightarrow$ The quantity

$$\underline{\mathrm{Opt}} = \max_{\mu\in M_1, \nu\in M_2} \ln\left(\int_\Omega \sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right) \qquad (!)$$

lower-bounds Opt.
• We now can make an *educated guess* that $\underline{\mathrm{Opt}}$ *is equal to* Opt, *and the optimal detector for the "worst" pair* $\mu \in M_1$, $\nu \in M_2$ *– one which is an optimal solution to* (!) *– is an optimal solution to the problem of interest.*
♣ Simplicity of the observation scheme in question and compactness and convexity of M_1 and M_2 turn out to be the conditions which make our educated guess true, and make the problem of computing the optimal detector convex and thus computationally tractable!

Implementation

♠ **Gaussian o.s.** $\mathcal{P}_\chi = \{\mathcal{N}(\mu, I_d) : \mu \in M_\chi\}$, $\chi = 1, 2$:

- Problem $\max_{\mu\in\mathcal{M}_1,\nu\in\mathcal{M}_2} \ln\left(\int \sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right)$ reads

$$\max_{\mu\in M_1,\nu\in M_2}\left[-\frac{1}{8}\|\mu-\nu\|_2^2\right]$$

- The optimal balanced detector and its risk are given by

$$\boxed{\begin{array}{rcl} \phi_*(\omega) & = & \frac{1}{2}[\mu^*-\nu^*]\omega - c, \\ & & (\mu^*,\nu^*) \in \underset{\mu\in M_1,\nu\in M_2}{\mathrm{Argmin}} \|\mu-\nu\|_2^2 \\ & & c = \frac{1}{4}[\mu^*-\nu^*]^T[\mu^*+\nu^*] \\ \varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2) & = & \exp\left\{-\frac{\|\mu^*-\nu^*\|_2^2}{8}\right\} \end{array}}$$

Note: We are in the “signal plus noise” model of observations with noise $\sim \mathcal{N}(0, I_d)$. The test $\mathcal{T}_{\phi_*}$ is nothing but the pairwise Euclidean separation test associated with $X_\chi = M_\chi$, $\chi = 1, 2$.

♠ **Poisson o.s.** $\mathcal{P}_\chi = \{\bigotimes_{i=1}^{d} \mathrm{Poisson}(\mu_i) : \mu = [\mu_1; ...; \mu_d] \in M_\chi\}$, $\chi = 1, 2$:

- Problem $\max_{\mu\in\mathcal{M}_1,\nu\in\mathcal{M}_2} \ln\left(\int \sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right)$ reads

$$\max_{\mu\in M_1,\nu\in M_2} \underbrace{\left[-\frac{1}{2}\sum_{i=1}^{d}(\sqrt{\mu_i}-\sqrt{\nu_i})^2\right]}_{\sum_i[\sqrt{\mu_i\nu_i}-\frac{1}{2}\mu_i-\frac{1}{2}\nu_i]}$$

- The optimal balanced detector and its risk are given by

$$\boxed{\begin{array}{rcl} \phi_*(\omega) & = & \frac{1}{2}\sum_{i=1}^{d}[\ln(\mu_i^*/\nu_i^*)\omega_i + \nu_i^* - \mu_i^*], \\ & & (\mu^*,\nu^*) \in \mathop{\mathrm{Argmax}}\limits_{\mu\in M_1,\nu\in M_2} \sum_i[\sqrt{\mu_i\nu_i} - \frac{1}{2}\mu_i - \frac{1}{2}\nu_i] \\ \varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2) & = & \exp\left\{-\frac{1}{2}\sum_i\left(\sqrt{\mu_i^*}-\sqrt{\nu_i^*}\right)^2\right\} \end{array}}$$

♠ Discrete o.s.

$$\mathcal{P}_\chi = \{\mu \in M_\chi\}, M_\chi \subset \mathbf{\Delta}_d^o = \{\mu \in \mathbb{R}^d_+ : \sum_\omega \mu_\omega = 1, \mu > 0\},\\ \chi = 1,2$$

- Problem $\max_{\mu\in\mathcal{M}_1,\nu\in\mathcal{M}_2} \ln\left(\int \sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right)$ reads

$$\max_{\mu\in M_1,\nu\in M_2} \sum_\omega \sqrt{\mu_\omega \nu_\omega}$$

- The optimal balanced detector and its risk are given by

$$\boxed{\begin{array}{rcl} \phi_*(\omega) & = & \frac{1}{2}\ln(\mu^*_\omega/\nu^*_\omega), \omega \in \Omega = \{1,...,d\} \\ & & (\mu^*,\nu^*) \in \underset{\mu\in M_1,\nu\in M_2}{\text{Argmin}} \sum_\omega \sqrt{\mu_\omega\nu_\omega} \\ \varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2) & = & \sum_\omega \sqrt{\mu^*_\omega \nu^*_\omega} \end{array}}$$

♠ **Direct product of simple o.s.'s.** Let
$$\mathcal{O}_k = \big((\Omega_k, \Pi_k), \{p_{\mu_k,k}(\cdot) : \mu_k \in \mathcal{M}_k\}, \mathcal{F}_k\big),\ 1 \le k \le K,$$
be simple o.s.'s, and $M_{\chi,k} \subset \mathcal{M}_k$, $\chi = 1, 2$, be nonempty convex compact sets. Consider the simple o.s.

$$\big((\Omega^K, \Pi^K), \{p_{\mu^K} : \mu^K \in \mathcal{M}^K\}, \mathcal{F}^K\big) = \bigotimes_{k=1}^{K} \mathcal{O}_k$$

along with two compact convex sets

$$M_\chi = M_{\chi,1} \times \ldots \times M_{\chi,K},\ \chi = 1, 2.$$

♡ **Question:** *What is the problem*

$$\max_{\mu^K \in M_1, \nu^K \in M_2} \ln\left(\int_{\Omega^K} \sqrt{p_{\mu^K}(\omega^K) p_{\nu^K}(\omega^K)}\,\Pi^K(d\omega^K)\right)$$

responsible for the smallest risk detector for the families of distributions $\mathcal{P}_1^{(K)}$, $\mathcal{P}_2^{(K)}$ *associated in* $\mathcal{O}^K$ *with the sets* M_1, M_2 *?*

♡ **Answer:** *This is the separable problem*

$$\max_{\{\mu_k \in M_{1,k}, \nu_k \in M_{2,k}\}_{k=1}^K} \sum_{k=1}^K \ln\left(\int_{\Omega_k} \sqrt{p_{\mu_k,k}(\omega_k) p_{\nu_k,k}(\omega_k)}\,\Pi_k(d\omega_k)\right)$$

$\Rightarrow$ *Minimum risk balanced detector for $\mathcal{P}_1^{(K)}$, $\mathcal{P}_2^{(K)}$ can be chosen as*

$$\begin{array}{c}\phi_*^K(\omega_1,...,\omega_K)=\sum_{k=1}^K\phi_{*,k}(\omega_k),\\ \phi_{*,k}(\omega_k)=\frac{1}{2}\ln\left(p_{\mu_k^*,k}(\omega)/p_{\nu_k^*,k}(\omega)\right)\\ \left[(\mu_k^*,\nu_k^*)\in\underset{\mu_k\in M_{1,k},\nu_k\in M_{2,k}}{\mathrm{Argmax}}\ln\left(\int_{\Omega_k}\sqrt{p_{\mu_k,k}(\omega_k)p_{\nu_k,k}(\omega_k)}\Pi_k(d\omega_k)\right)\right]\end{array}$$

and

$$\varepsilon_\star(\mathcal{P}_1^{(K)},\mathcal{P}_2^{(K)})=\prod_{k=1}^K\varepsilon_\star(\mathcal{P}_{1,k},\mathcal{P}_{2,k}),$$

where $\mathcal{P}_{\chi_k}$ are the families of distributions associated in $\mathcal{O}_k$ with $M_{\chi,k}$, $\chi=1,2$.

♠ **Remark:** The families of distributions $\mathcal{P}_\chi^{(K)}$ are direct products of the families $\mathcal{P}_{\chi,k}$ over $k=1,...K$. From Detector Calculus, extending $\mathcal{P}_\chi^{(K)}$ to families $\mathcal{P}_\chi^{\otimes,K}$ of quasi-direct products of families $\mathcal{P}_{\chi,k}$, $k=1,...,K$, we still have

$$\mathrm{Risk}[\phi_*^K|\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}] \leq \underbrace{\prod_{k=1}^K \varepsilon_\star(\mathcal{P}_{1,k},\mathcal{P}_{2,k})}_{=:\epsilon_K},$$

whence also $\epsilon_K = \varepsilon_\star(\mathcal{P}_1^K,\mathcal{P}_2^K) \leq \varepsilon_\star(\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}) \leq \epsilon_K$

$$\Rightarrow \varepsilon_\star(\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}) = \prod_{k=1}^K \varepsilon_\star(\mathcal{P}_{1,k},\mathcal{P}_{2,k}).$$

♠ **Power of a simple o.s.** Let

$$\mathcal{O} = ((\Omega,\Pi),\{p_\mu(\cdot):\mu\in\mathcal{M}\},\mathcal{F})$$

be a simple o.s., and $M_\chi\subset\mathcal{M}$, $\chi=1,2$, be nonempty convex compact sets. Consider the K-th power of $\mathcal{O}$, that is, the simple o.s.

$$\mathcal{O}^{(K)} = \left((\Omega^K,\Pi^K),\{p_\mu^{(K)}(\omega_1,...,\omega_K)=\prod_{k=1}^K p_\mu(\omega_k):\mu\in\mathcal{M}\},\mathcal{F}^{(K)}\right).$$

♡ **Question:** *What is the problem*

$$\max_{\mu\in M_1,\nu\in M_2} \ln\left(\int_{\Omega^K}\sqrt{p_\mu^{(K)}(\omega^K)p_\nu^{(K)}(\omega^K)}\Pi^K(d\omega^K)\right)$$

responsible for the smallest risk detector for the families of distributions $\mathcal{P}_\chi^K$ associated in $\mathcal{O}^{(K)}$ with the sets M_χ, $\chi=1,2$?

♡ **Answer:** *This is the separable problem*

$$\max_{\mu\in M_1,\nu\in M_2}\underbrace{\sum\nolimits_{k=1}^K \ln\left(\int_\Omega\sqrt{p_\mu(\omega_k)p_\nu(\omega_k)}\Pi(d\omega_k)\right)}_{K\ln\left(\int_\Omega\sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right)}$$

⇒ *Minimum risk balanced detector for $\mathcal{P}_1^K$, $\mathcal{P}_2^K$ can be chosen as*

$$\phi_*^{(K)}(\omega_1,...,\omega_K)=\sum\nolimits_{k=1}^K\phi_*(\omega_k) \text{ with } \phi_*(\omega_k)=\tfrac{1}{2}\ln\left(p_{\mu^*}(\omega)/p_{\nu^*}(\omega)\right)$$

$$\left[(\mu^*,\nu^*)\in\mathop{\mathrm{Argmax}}_{\mu\in M_1,\nu\in M_2}\ln\left(\int_\Omega\sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right)\right]$$

and

$$\varepsilon_\star(\mathcal{P}_1^K,\mathcal{P}_2^K)=[\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)]^K,$$

where $\mathcal{P}_\chi$ are the families of distributions associated in $\mathcal{O}$ with M_χ, $\chi=1,2$.

♠ **Remark:** The families of distributions $\mathcal{P}_\chi^K$ are direct powers $\mathcal{P}_\chi^{\oplus,K}$ of the families $\mathcal{P}_\chi$. From Detector Calculus, extending $\mathcal{P}_\chi^K$ to families $\mathcal{P}_\chi^{\otimes,K}$ of quasi-direct powers of families $\mathcal{P}_\chi$, we still have

$$\mathrm{Risk}[\phi_*^{(K)}|\mathcal{P}_1^{\oplus,K},\mathcal{P}_2^{\oplus,K}] \leq \underbrace{[\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)]^K}_{=:\epsilon_K},$$

whence also $\epsilon_K = \varepsilon_\star(\mathcal{P}_1^K,\mathcal{P}_2^K) \leq \varepsilon_\star(\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}) \leq \epsilon_K$

$$\Rightarrow \varepsilon_\star(\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}) = [\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)]^K.$$

Near-Optimality of Minimum Risk Detector-Based Tests in Simple Observation Schemes

♣ **Proposition A.** *Let*

$$\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$$

be a simple o.s., and $M_\chi \subset \mathcal{M}$, $\chi = 1, 2$, *be nonempty convex compact sets, giving rise to families of distributions*

$$\mathcal{P}_\chi = \{P : P \text{ has density } p_\mu(\cdot) \text{ w.r.t. } \Pi \text{ with } \mu \in M_\chi\},\ \chi = 1, 2,$$

hypotheses

$$H_\chi : P \in \mathcal{P}_\chi,\ \chi = 1, 2,$$

on the distribution of a random observation $\omega \in \Omega$, *and minimum risk detector* ϕ_* *for* $\mathcal{P}_1$, $\mathcal{P}_2$.

Assume that in the nature there exists a simple single-observation test, deterministic or randomized, $\mathcal{T}$ *with*

$$\mathrm{Risk}(\mathcal{T}|H_1, H_2) \leq \epsilon < 1/2.$$

Then the risk of the simple test $\mathcal{T}_{\phi_*}$ *accepting* H_1 *when* $\phi_*(\omega) \geq 0$ *and accepting* H_2 *otherwise "is comparable" to* ϵ*:*

$$\mathrm{Risk}(\mathcal{T}_{\phi_*}|H_1, H_2) \leq \epsilon_+ := 2\sqrt{\epsilon(1-\epsilon)} < 1.$$

Proof. From what we called "universality" of detector-based tests, there exists a detector ϕ with $\mathrm{Risk}[\phi|\mathcal{P}_1,\mathcal{P}_2] \leq \epsilon_+$, and $\mathrm{Risk}[\phi_*|\mathcal{P}_1,\mathcal{P}_2]$ can be only less than $\mathrm{Risk}[\phi|\mathcal{P}_1,\mathcal{P}_2]$. $\square$

♣ **Proposition B.** *Let* $\mathcal{O} = ((\Omega,\Pi),\{p_\mu : \mu\in\mathcal{M}\},\mathcal{F})$ *be a simple o.s., and* $M_\chi\subset\mathcal{M}$, $\chi=1,2$, *be nonempty convex compact sets, giving rise to families of distributions*

$$\mathcal{P}_\chi = \{P : P \text{ has density } p_\mu(\cdot) \text{ w.r.t. } \Pi \text{ with } \mu\in M_\chi\},\ \chi=1,2$$

their direct powers

$$\mathcal{P}_\chi^{\odot,K} = \{P\times\ldots\times P : P\in\mathcal{P}_\chi\},\ \chi=1,2,\ K=1,2,\ldots$$

hypotheses $H_\chi^K : P\in\mathcal{P}_\chi^{\odot,K}$, $\chi=1,2$, $K=1,2,\ldots$ *on the distribution* P *of random* K*-repeated observation* $\omega^K=(\omega_1,\ldots\omega_K)\in\Omega^K$, *and minimum risk detector* ϕ_* *for* $\mathcal{P}_1,\mathcal{P}_2$.

Assume that in the nature there positive integer K_* *and a simple* K_**-observation test, deterministic or randomized,* $\mathcal{T}_{K_*}$ *capable to decide on the hypotheses* $H_\chi^{K_*}$, $\chi=1,2$, *with risk* $\leq\epsilon<1/2$. *Then the test* $\mathcal{T}_{\phi_*,K}$ *deciding on* H_χ^K, $\chi=1,2$, *by accepting* H_1^K *whenever* $\phi^{(K)}(\omega^K):=\sum_{k=1}^K\phi_*(\omega_k)\geq 0$ *and accepting* H_2^K *otherwise, satisfies*

$$\mathrm{Risk}(\mathcal{T}_{\phi_*,}|H_1^K,H_2^K)\leq\epsilon\ \forall K\geq\widehat{K}_* = \frac{2}{1-\frac{\ln(4(1-\epsilon))}{\ln(1/\epsilon)}}K_*.$$

Moreover, this risk bound remains true when the hypotheses H_χ^K *are extended to* $H_\chi^{\otimes,K}$ *stating that the distribution* P *of* ω^K *belongs to the quasi-direct* K*-th power of* $\mathcal{P}_\chi$, $\chi=1,2$. Note that $\widehat{K}_*/K_*\to 2$ as $\epsilon\to+0$.

Proof. As we know, K_*-th power $\mathcal{O}^{(K_*)}$ of $\mathcal{O}$ is simple o.s. along with $\mathcal{O}$, and $\phi_*^{(K_*)}$ is the minimum risk detector for the families $\mathcal{P}_\chi^{\odot,K_*}$, $\chi = 1,2$, the risk of this detector being $[\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)]^{K_*}$. By Proposition A as applied to $\mathcal{O}^{(K_*)}$ in the role of $\mathcal{O}$, we have

$$[\varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2)]^{K_*} \le 2\sqrt{\epsilon(1-\epsilon)} \Rightarrow \varepsilon_\star(\mathcal{P}_1,\mathcal{P}_2) \le [2\sqrt{\epsilon(1-\epsilon)}]^{1/K_*} < 1.$$

By Detector Calculus, it follows that for $K = 1,2,...$ it holds

$$\mathrm{Risk}[\phi_*^{(K)}|\mathcal{P}_1^{\otimes,K},\mathcal{P}_2^{\otimes,K}] \le [2\sqrt{\epsilon(1-\epsilon)}]^{K/K_*}$$

and the right hand side is $\le \epsilon$ whenever $K \ge \widehat{K}_*$. $\square$

Near-Optimality of Detector-Based Up to Closeness Testing in Simple Observation Schemes

♣ **Situation:** We are given a simple o.s.

$$\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$$

and a collection of nonempty convex compact subsets M_ℓ, $1 \leq \ell \leq L$ giving rise to

• Families $\mathcal{P}_\ell = \{P : P$ admits density p_μ, $\mu \in M_\ell$ w.r.t. $\Pi\}$, $\ell = 1, ..., L$, along with quasi-direct powers $\mathcal{P}_\ell^{\otimes,K} = \bigotimes_{k=1}^K \mathcal{P}_\ell$ of $\mathcal{P}_\ell$ and hypotheses $H_\ell^{\otimes,K} : P \in \mathcal{P}_\ell^{\otimes,K}$ on the distribution P of K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$,

• minimum-risk balanced single-observation detectors $\phi_{\ell\ell'}(\omega)$ for $\mathcal{P}_\ell$, $\mathcal{P}_{\ell'}$ along with their risks $\varepsilon_\star(\mathcal{P}_\ell, \mathcal{P}_{\ell'})$, $1 \leq \ell < \ell' \leq L$, and K-repeated versions

$$\phi_{\ell\ell'}^K(\omega^K) = \sum_{k=1}^K \phi_{\ell\ell'}(\omega_k)$$

of $\phi_{\ell\ell'}$ such that

$$\text{Risk}[\pi_{\ell\ell'}^{(K)} | H_\ell^{\otimes,K}, H_{\ell'}^{\otimes,K}] \leq [\varepsilon_\star(\mathcal{P}_\ell, \mathcal{P}_{\ell'})]^K .$$

♠ Assume that in addition to the above data, we are given a closeness relation $\mathcal{C}$ on $\{1, ..., L\}$. Applying Calculus of Detectors, for every positive integer K, setting

$$\theta_K = \left\| \left[\varepsilon_\star^K(\mathcal{P}_\ell, \mathcal{P}_{\ell'}) \cdot \left\{ \begin{array}{ll} 1, & (\ell,\ell') \notin \mathcal{C} \\ 0, & (\ell,\ell') \in \mathcal{C} \end{array} \right. \right]_{\ell,\ell'=1}^{L} \right\|_{2,2}$$

we can assemble the outlined data, in a computationally efficient fashion, into a K-observation test $\mathcal{T}^K$ deciding on $H_\ell^{\otimes,K}$, $1 \le \ell \le L$, with $\mathcal{C}$-risk upper-bounded as follows:

$$\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}^K | H_1^{\otimes,K}, ..., H_L^{\otimes,K}) \le \varkappa \theta_K$$

($\varkappa > 1$ can be selected to be as close to 1 as we want).

♠ **Proposition.** *In the just described situation, assume that for some $\epsilon < 1/2$ and K_* in the nature there exists a K_*-observation test $\mathcal{T}$, deterministic or randomized, deciding on the hypotheses*

$$H_\ell^{\odot,K_*}: \omega^{K_*} = (\omega_1, ..., \omega_{K_*}) \text{ is an i.i.d. sample drawn from a } P \in \mathcal{P}_\ell,$$

$\ell = 1, ..., L$, *with $\mathcal{C}$-risk $\leq \epsilon$. Then the test $\mathcal{T}^K$ with*

$$K \geq 2 \underbrace{\left[\frac{1 + \ln(\varkappa L)/\ln(1/\epsilon)}{1 - \ln(4(1-\epsilon))/\ln(1/\epsilon)}\right]}_{\to 1 \text{ as } \epsilon \to +0} K_*$$

decides on $H_\ell^{\otimes,K}$, $\ell = 1, ..., L$, with $\mathcal{C}$-risk $\leq \epsilon$ as well.

Proof. • Let us fix ℓ, ℓ' such that $(\ell, \ell') \notin \mathcal{C}$, and let us convert $\mathcal{T}$ into a simple K_*-observation test $\widetilde{\mathcal{T}}$ deciding on $H_\ell^{\odot,K_*}$, $H_{\ell'}^{\odot,K_*}$ as follows: whenever $\ell \in \mathcal{T}(\omega^{K_*})$, $\widetilde{\mathcal{T}}$ accepts $H_\ell^{\odot,K_*}$ and rejects $H_{\ell'}^{\odot,K_*}$, otherwise the test accepts $H_{\ell'}^{\odot,K_*}$ and rejects $H_\ell^{\odot,K_*}$. It is immediately seen that

$$\text{Risk}(\widetilde{\mathcal{T}}|H_\ell^{\odot,K_*}, H_{\ell'}^{\odot,K_*}) \le \epsilon.$$

Indeed, let $P^{K_*} = P \times ... \times P$ be the distribution of ω^{K_*}. Whenever P^{K_*} obeys $H_\ell^{\odot,K_*}$, $\mathcal{T}$ must accept the hypothesis with P^{K_*}-probability $\ge 1 - \epsilon$, whence

$$\text{Risk}_1(\widetilde{\mathcal{T}}|H_\ell^{\odot,K_*}, H_{\ell'}^{\odot,K_*}) \le \epsilon.$$

If P^{K_*} obeys $H_{\ell'}^{\odot,K_*}$, the P^{K_*}-probability of the event "$\mathcal{T}$ accepts $H_{\ell'}^{\odot,K_*}$ and rejects $H_\ell^{\odot,K_*}$" is $\le \epsilon$, since $H_{\ell'}^{\odot,K_*}$, $H_\ell^{\odot,K_*}$ are not $\mathcal{C}$-close to each other
$\Rightarrow$ P^{K_*}-probability to reject $H_\ell^{\odot,K_*}$ is at least $1 - \epsilon$
$\Rightarrow$ $\text{Risk}_2(\widetilde{\mathcal{T}}|H_\ell^{\odot,K_*}, H_{\ell'}^{\odot,K_*}) \le \epsilon$.

$\boxed{H_\ell^{\odot,K_*}, H_{\ell'}^{\odot,K_*} \text{ can be decided upon by a simple test with risk } \leq \epsilon}$

- $H_{\ell'}^{\odot,K_*}, H_\ell^{\odot,K_*}$ can be decided upon with risk $\leq \epsilon < 1/2$

$\Rightarrow$ $\varepsilon_\star(\mathcal{P}_\ell^{\odot,K_*}, \mathcal{P}_{\ell'}^{\odot,K_*}) \leq 2\sqrt{\epsilon(1-\epsilon)} < 1$ (Calculus of Detectors)

$\Rightarrow$ $\varepsilon_\star(\mathcal{P}_\ell, \mathcal{P}_{\ell'}) \leq \left[2\sqrt{\epsilon(1-\epsilon)}\right]^{1/K_*} < 1$ (since $\mathcal{O}$ is a simple o.s.)

$\Rightarrow$ $\theta_K \leq \left[2\sqrt{\epsilon(1-\epsilon)}\right]^{K/K_*} L$

$\Rightarrow$ $\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}^K | H_1^{\otimes,K}, ..., H_L^{\otimes,K}) \leq \varkappa\theta_K \leq \epsilon$ when

$$K/K^* \geq 2\frac{1+\ln(\varkappa L)/\ln(1/\epsilon)}{1-\ln(4(1-\epsilon))/\ln(1/\epsilon)}.$$

How it works: Illustration I
Predicting Outcome of Elections via Opinion Polls

Situation: L candidates are running for office, with just one to be elected, and every voter has already decided whom to vote for in the forthcoming elections. We want to predict elections' outcome via Opinion poll where K randomly selected voters reveal their choices. How large should be K in order to predict the winner with a given confidence?

Model: Assume that K voters to be interviewed are drawn from the population uniformly and independently of each other. Denoting by μ_ℓ the fraction of voters intending to vote for candidate $\#\ell$ in the entire population, we get a probability distribution μ on the L-element set of candidates.

Note: *Outcomes of K interviews form K-element i.i.d. sample ω^K drawn from μ.*

♠ Given small "winning margin" $\delta > 0$ and assuming that the distribution μ of voters' preferences is *not* a "δ-tie" – the difference between the largest and the second largest entries in μ is *at least* δ — predicting the winner can be modeled as deciding on L convex hypotheses

$$H_\ell : \mu \in \mathcal{P}_\ell := \{p \in \mathbf{\Delta}_L : p_\ell \geq \delta + \max_{j \neq \ell} p_j\},\ \ell = 1, ..., L$$
$$[\mathbf{\Delta}_L = \{p \in \mathbb{R}^L_+ : \sum_\ell p_\ell = 1\}]$$

in Discrete o.s. via stationary K-repeated observation.

$$H_\ell : \mu \in \mathcal{P}_\ell := \{p \in \mathbf{\Delta}_L : p_\ell \geq \delta + \max_{j \neq \ell} p_j\},\ \ell = 1, ..., L$$

♠ Our machinery applies as follows:

• We solve $L(L-1)/2$ convex optimization problems

$$\epsilon_{ij} = \max_{\mu,\nu} \left\{ \sum_i \sqrt{\mu_i \nu_i} : \mu \in \mathcal{P}_i, \nu \in \mathcal{P}_j \right\},\ 1 \leq i < j \leq L.$$

with optimal solutions μ^{ij}, ν^{ij} giving rise to detectors

$$\phi_{ij}(\omega) = \frac{1}{2} \ln \left(\mu^{ij}_\omega / \nu^{ij}_\omega \right),\ \omega \in \Omega = \{1, 2, ..., L\}$$

We set also

$$\epsilon_{ji} = \epsilon_{ij},\ \phi_{ji}(\cdot) = -\phi_{ij}(\cdot),\ 1 \leq i < j \leq m,\ \epsilon_{ii} = 0,\ \phi_{ii}(\cdot) \equiv 0, i \leq m$$

• We build the symmetric matrix $E = \left[\epsilon^K_{ij} \right]_{i,j \leq L}$ The Perron-Frobenius eigenvector f of E gives rise to the detectors

$$\phi^{(K)}_{ij}(\omega^K) = \sum_{k=1}^K \phi_{ij}(\omega_k) + \ln(f_i/f_j)$$

and the test which accepts H_ℓ if and only if $p^{(K)}_{\ell j}(\omega^K) > 0$ for all $j \neq \ell$.

The risk of this test does not exceed the spectral norm of E.

♠ *Given δ and upper bound ϵ on the risk of predicting elections' outcome, we can specify the smallest size K of Opinion poll resulting in prediction of required quality.*

♣ Results, confidence level $1 - \epsilon = 0.95$:

	winning margin δ			
	10%	5%	2.5%	1%
$L = 2$	$166 \vee 597$	$664 \vee 2,394$	$2,657 \vee 9,584$	$16,607 \vee 59,912$
$L = 4$	$166 \vee 815$	$664 \vee 3,272$	$2,657 \vee 13,098$	$16,607 \vee 81,882$
$L = 8$	$166 \vee 984$	$664 \vee 3949$	$2,657 \vee 15,809$	$16,694 \vee 98,811$

- upper bounds on poll sizes are given by our machinery
- lower bounds on poll sizes stem from lower bounding of pairwise risks

♠ USA Presidential Elections-2016:

State	Actual margin	Poll size, lower bound	Poll size, upper bound
Georgia	5.1%	638	2,301
Wisconsin	0.77%	28,008	101,043
Pennsylvania	0.72%	32,030	115,555
Michigan	0.23%	313,864	1,132,333

Note: the *total* number of Michigan voters participated in Presidential Elections-2016 was 4,799,284

Variation: Comparative Drug Study

♣ **Situation:** We want to carry out a clinical study aimed at comparing the effects of two drugs, A and B. The effect of a drug on a particular patient is categorical with μ mutually exclusive values, say, ternary: "positive effect," "no effect," or "negative effect."

The study is organized as follows: in a single trial we

— draw trial's subject at random, from the uniform distribution on the pool of animals (or people) participating in the study

— flip a coin, with probability α for heads and β for tails, to decide which drug, A or B, to administer.

After the subject is administered the drug, we record the effect.

Model: Let us associate with k-th member of the pool 2μ-dimensional vector x^k as follows:

— the first μ entries encode the effect on the member of drug A: when it is $\iota\in\{1,...,\mu\}$, we write 1 in position ι and 0 in other positions of the first half of x^k

— the last μ entries encode the effect of drug B: when it is ι, we write 1 in position $\mu+\iota$ and 0 in the remaining positions of the second half of x^k.

Example: With ternary effect,

— $x=[1;0;0;0;0;1]$ encodes *"positive effect of drug A, negative effect of drug B"*

— $x=[0;1;0;1;0;0]$ encodes *"no effect of drug A, positive effect of drug B"*

— $x=[1;0;1;0;0;1]$ is illegitimate

Let x be the average of the vectors $\{x^k\}_k$ taken over the pool of all candidates.

Note: x *encodes the probabilities* $p_{U\iota}$ *of possible outcomes "administered drug* $U\in\{A,B\}$*, observed effect* $\iota\in\{1,...,\mu\}$*" of a single trial:*

$$p_{A\iota}=\alpha x_\iota,\;\; p_{B\iota}=\beta x_{\mu+\iota}$$

$\Rightarrow$ *The distribution* p *of outcomes of a single trial is linearly parameterized by the (unknown in advance) vector* x *known to belong to the convex set*

$$\Delta^\mu=\{x\in\mathbb{R}^{2\mu}_+:\sum_{\iota=1}^{\mu}x_\iota=\sum_{\iota=1}^{\mu}x_{\mu+\iota}=1\}$$

... the distribution p of outcomes of a single trial is linearly parameterized by the (unknown in advance) vector x known to belong to the convex compact set

$$\Delta^\mu = \{x \in \mathbb{R}^{2\mu}_+ : \sum_{\iota=1}^{\mu} x_\iota = \sum_{\iota=1}^{\mu} x_{\mu+\iota} = 1\}$$

$\Rightarrow$ Various questions about relative performance of the drugs, like

Which of the drugs have more chances to have positive effect?

reduce to testing convex hypotheses in Discrete o.s.

Example 1: Assume that the effect is ternary:

$$\iota = 1 \Rightarrow \text{positive effect; } \iota = 2 \Rightarrow \text{no effect; } \iota = 3 \Rightarrow \text{negative effect}$$

and we want to decide via K experiments on the hypotheses

- the chances for A to have positive effect are at least by margin $\delta > 0$ larger than those for B
- the chances for A to have positive effect are smaller than those for B

Equivalently: *Given stationary K-repeated observation ω^K with components ω_k taking values $(U, \iota) \in \{A, B\} \times \{1, 2, 3\}$, and the distribution p affinely parameterized by $x \in \Delta^3$, decide on the hypotheses*

$$H_A : p \in \mathcal{P}(X_A), \; H_B : p \in \mathcal{P}(X_B)$$

where

$$\mathcal{P}(X) = \{p(x) : x \in X\}$$

and

$$p(x)_{U\iota} = \begin{cases} \alpha x_\iota, & U = A \\ \beta x_{\mu+\iota}, & U = B \end{cases}, \; X_A = \{x \in \Delta^3 : x_1 \geq x_4 + \delta\}, \; X_B = \{x \in \Delta^3 : x_1 \leq x_4\}.$$

Numerical results: With ternary effect, the number K of observations needed to decide 0.95-reliably on the hypotheses

- *The chances to get an outcome from $\mathcal{I}$ with drug A are at least by margin δ larger than the chances to get an outcome from $\mathcal{J}$ with drug B*
- *The chances to get an outcome from $\mathcal{I}$ with drug A are smaller than the chances to get an outcome from $\mathcal{J}$ with drug B*

are independent of *proper and nonempty* subsets $\mathcal{I}$, $\mathcal{J}$ of the set

{"positive effect," "no effect," "negative effect"}

of outcomes of a single trial and is as follows:

δ	0.50	0.25	0.15	0.10	0.05
K	87	375	591	2,388	9,578

$\alpha = 0.5, \beta = 0.5$

δ	0.50	0.25	0.15	0.10	0.05
K	117	501	788	3,185	12,771

$\alpha = 0.75, \beta = 0.25$

Note: When trials using different drugs require different amounts of resources (money, time, clinical facilities, etc.), one could use easy-to-compute dependency of K on $\alpha = 1 - \beta$ to optimize our study under constraints on how reliable and how "costly" it should be.

δ					
0.50	Cost$(B) = 1$ 0.50/87/87	Cost$(B) = 2$ 0.58/127/91	Cost$(B) = 3$ 0.63/163/96	Cost$(B) = 4$ 0.66/197/96	Cost$(B) = 5$ 0.68/229/104
	Cost$(B) = 6$ 0.72/260/104	Cost$(B) = 7$ 0.72/291/104	Cost$(B) = 8$ 0.72/322/117	Cost$(B) = 9$ 0.72/351/117	Cost$(B) = 10$ 0.76/380/117
0.25	Cost$(B) = 1$ 0.50/375/375	Cost$(B) = 2$ 0.59/547/391	Cost$(B) = 3$ 0.63/700/412	Cost$(B) = 4$ 0.67/845/412	Cost$(B) = 5$ 0.68/983/447
	Cost$(B) = 6$ 0.72/1118/447	Cost$(B) = 7$ 0.72/1252/447	Cost$(B) = 8$ 0.73/1378/501	Cost$(B) = 9$ 0.75/1503/501	Cost$(B) = 10$ 0.77/1628/501
0.13	Cost$(B) = 1$ 0.50/1525/1525	Cost$(B) = 2$ 0.50/2225/1589	Cost$(B) = 3$ 0.64/2849/1676	Cost$(B) = 4$ 0.67/3436/1676	Cost$(B) = 5$ 0.69/3995/1816
	Cost$(B) = 6$ 0.71/4540/1816	Cost$(B) = 7$ 0.73/5085/1816	Cost$(B) = 8$ 0.74/5596/2035	Cost$(B) = 9$ 0.75/6105/2035	Cost$(B) = 10$ 0.76/6614/2035
0.06	Cost$(B) = 1$ 0.50/6127/6127	Cost$(B) = 2$ 0.59/8935/6382	Cost$(B) = 3$ 0.63/11446/6733	Cost$(B) = 4$ 0.67/13803/6733	Cost$(B) = 5$ 0.69/16047/7294
	Cost$(B) = 6$ 0.71/18235/7294	Cost$(B) = 7$ 0.72/20423/7294	Cost$(B) = 8$ 0.74/22468/8170	Cost$(B) = 9$ 0.75/24510/8170	Cost$(B) = 10$ 0.76/26553/8170

Optimized Study

X/XX/XXX in cells: X: α; XX: cost of study; XXX: K
Cost(A)=1

How it Works: Illustration II
Selecting the Best in a Family of Estimates

♣ **Problem:**

• We are given a simple o.s. $\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$ and have access to stationary K-repeated observations

$$\omega_k \sim p_{A(x_*)}(\cdot),\ k = 1, ..., K,$$

of unknown signal x_* known to belong to a given convex compact set $X \subset \mathbb{R}^n$.
[$x \mapsto A(x)$: affine mapping such that $A(X) \subset \mathcal{M}$].

• We are given M candidate estimates $x_i \in \mathbb{R}^n$, $1 \leq i \leq M$, of x_*, a norm $\|\cdot\|$ on $\mathbb{R}^n$, and a reliability tolerance $\epsilon \in (0, 1)$

• **Ideal Goal:** *Use observations $\omega_1, ..., \omega_K$ to identify $(1-\epsilon)$-reliably the $\|\cdot\|$-closest to x_* point among $x_1, ..., x_M$.*

• **Actual Goal:** *Given $\alpha \geq 1$, $\beta \geq 0$ and a grid $\Gamma = \{r_0 > r_1 > ... > r_N > 0\}$, use observations $\omega_1, ..., \omega_K$ to identify $(1-\epsilon)$-reliably a point $x_{i(\omega^K)}$ such that*

$$\|x_* - x_{i(\omega^K)}\| \leq \alpha\rho(x_*) + \beta$$
$$\left[\begin{array}{c} \rho(x) := \min\{r : r \in \Gamma, r \geq \min_i \|x - x_i\|\} \\ \rho(x) \text{ is grid approximation of } \min_i \|x - x_i\| \end{array}\right]$$

Note: We select r_0 large enough to ensure that $X \subset \cup_i \{x : \|x - x_i\| \leq r_0\}$, r_N to be small enough, and Γ to be dense enough. For example, we can set $\Gamma = \{10^{10}[0.9]^{-s}, 0 \leq s \leq 438\}$, resulting in $r_N < 10^{-10}$. In our application this 439-point grid approximation of $\mathbb{R}_+$ for all practical purposes is as good as $\mathbb{R}_+$ itself.

♠ **Proposed solution:** Use testing hypotheses up to closeness.

Recall the recipe for deciding via i.i.d. observations $\omega_k \sim P$ on L *convex* hypotheses $H_\ell : P \in \mathcal{P}_\ell$ in *simple o.s.* up to closeness $\mathcal{C}$:

A. For $\ell < \ell'$ such that ℓ, ℓ' are not $\mathcal{C}$-close to each other, compute the optimal single-observation detector $\phi_{\ell\ell'}$ for $\mathcal{P}_\ell, \mathcal{P}'_\ell$ and its risk $\epsilon_{\ell\ell'}$. Set $\epsilon_{\ell'\ell} = \epsilon_{\ell\ell'}$ and $\phi_{\ell'\ell}(\cdot) = -\phi_{\ell\ell'}(\cdot)$.
For ℓ, ℓ' $\mathcal{C}$-close to each other, set $\epsilon_{\ell\ell'} = 0$.

B. If some of $\epsilon_{\ell\ell'}$ are equal to 1, terminate – our machinery does not work. Otherwise look at symmetric $L \times L$ matrices $E_K = [\epsilon^K_{\ell\ell'}]_{\ell,\ell'}$ and find the smallest K such that

$$\|E_K\|_{2,2} \leq \epsilon \qquad [\epsilon : \text{ desired } \mathcal{C}\text{-risk of would-be test}]$$

With the resulting K, the detectors $\phi_{\ell\ell'}$ can be assembled in K-observation test $\mathcal{T}^K$ deciding on $H_1, ..., H_L$ up to closeness $\mathcal{C}$ with risk $\leq \epsilon$.

Test $\mathcal{T}^K$ works as follows:

- find Perron-Frobenius eigenvector f of E_K.
- Given ω^K, for ℓ, ℓ' not $\mathcal{C}$-close to each other, compute the quantities $\phi^K_{\ell\ell'} = \sum_{k=1}^K \phi_{\ell\ell'}(\omega_k) + \ln(f_\ell / f_{\ell'})$
- accept all hypotheses H_ℓ, if any, such that $\phi^K_{\ell\ell'} > 0$ for all ℓ' not $\mathcal{C}$-close to ℓ.

Goal: *Given* $\alpha \geq 1$, $\beta \geq 0$ *and a grid* $\Gamma = \{r_0 > r_1 > ... > r_N > 0\}$, *use observations* $\omega_1, ..., \omega_K$ *to identify* $(1-\epsilon)$*-reliably a point* $x_{i(\omega^K)}$ *such that*

$$\|x_* - x_{i(\omega^K)}\| \leq \alpha\rho(x_*) + \beta$$
$$\left[\ \rho(x) := \min\{r : r \in \Gamma, r \geq \min_i \|x - x_i\|\}\ \right]$$

Construction:

• We look at $M(N+1)$ hypotheses

$$H_{ij} : \omega_k \sim p_{A(x)}(\cdot) \text{ for some } x \in X_{ij} := \{x \in X : \|x - x_i\| \leq r_j\}.$$

and discard those which are empty: $X_{ij} = \emptyset$. We end up with a list of $L \leq M(N+1)$ hypotheses $\{H_{ij} : ij \in \mathcal{I}\}$.

• We define closeness $\mathcal{C} = \mathcal{C}_{\alpha,\beta}$: ij $\mathcal{C}$*-close to* $i'j'$ *iff*

$$\|x_i - x_{i'}\| \leq \bar{\alpha}(r_j + r_{j'}) + \beta \qquad \left[\bar{\alpha} = \tfrac{\alpha-1}{2}\right]$$

• We apply the above recipe to build K-observation test $\mathcal{T}^K$ deciding on H_{ij}, $ij \in \mathcal{I}$, up to closeness $\mathcal{C}$. If the recipe fails to work, reject (α, β). Otherwise, given ω^K, we apply $\mathcal{T}^K$.

— *If* $\mathcal{T}^K(\omega^K) \neq \emptyset$, the test accepts some hypotheses H_{ij}. We select among them the one, $H_{i_*j_*}$, with the largest j, and claim that x_{i_*} is the desired point: $\|x_* - x_{i_*}\| \leq \alpha\rho(x_*) + \beta$.

— *If* $\mathcal{T}^K(\omega^K) = \emptyset$, we can do whatever we want, e.g., return x_1 as the closest to x_* point among x_i.

♣ **Fact:** *In the situation in question, whenever* (α,β) *is not rejected, the resulting inference* $\omega^K \mapsto i_* = i_*(\omega^K)$ *meets the design specifications:*

$$\left(x_* \in X, \omega_k \sim p_{A(x_*)}(\cdot) \text{ independent across } k \leq K\right)$$
$$\Rightarrow \mathrm{Prob}\{\|x_* - x_{i_*(\omega^K)}\| \leq \alpha\rho(x_*) + \beta\} \geq 1 - \epsilon$$

Indeed, let $i_\maltese$ be the index of the closest to x_* point among x_i:

$$\|x_* - x_{i_\maltese}\| \leq \rho(x_*) = r_{j_\maltese}.$$

Then $H_{i_\maltese j_\maltese}$ is true, and since the $\mathcal{C}$-risk of $\mathcal{T}^K$ is $\leq \epsilon$, the stemming from x_* probability of the event

"$\mathcal{T}^K$ accepts $H_{i_\maltese j_\maltese}$, and every other hypothesis accepted by $\mathcal{T}^K$ is $\mathcal{C}$-close to $H_{i_\maltese j_\maltese}$"

is $\geq 1-\epsilon$. When this event takes place, $j_\star \geq j_\maltese$, whence $r_{j_*} \leq r_{j_\maltese} = \rho(x_*)$, and $H_{i_* j_*}$ is $\mathcal{C}$-close to $H_{i_\maltese j_\maltese}$, whence

$$\|x_{i_*} - x_{i_\maltese}\| \leq \frac{\alpha-1}{2}[r_{j_*} + r_{j_\maltese}] + \beta \leq (\alpha-1)\rho(x_*) + \beta$$

$$\Rightarrow \|x_{i_*} - x_*\| \leq \|x_{i_*} - x_{i_\maltese}\| + \|x_{i_\maltese} - x_*\| \leq (\alpha-1)\rho(x_*) + \beta + \rho(x_*) = \alpha\rho(x_*) + \beta \qquad \square$$

♣ **Fact:** *In the situation in question, assume that for some* $\epsilon \in (0, 1/2)$, $a, b \geq 0$ *and positive integer* K_* *in the nature there exists an inference* $\omega^{K_*} \to i_*(\omega^{K_*})$ *such that*

$$\begin{array}{l}(x_* \in X, \omega_k \sim p_{A(x_*)} \text{ independent across } k) \\ \qquad \Rightarrow \text{Prob}\{\|x_* - x_{i_*(\omega^{K_*})}\| \leq a\rho(x_*) + b\} \geq 1 - \epsilon.\end{array}$$

Then the pair $(\alpha = 2a + 3, \beta = 2b)$ *is not rejected by the above construction, and the number of observations* K *required by it to infer from* ω^K *index* $\widehat{i}(\omega^K)$ *such that*

$$\begin{array}{l}(x_* \in X, \omega_k \sim p_{A(x_*)} \text{ independent across } k) \\ \qquad \Rightarrow \text{Prob}\{\|x_* - x_{\widehat{i}(\omega^K)}\| \leq \alpha\rho(x_*) + \beta\} \geq 1 - \epsilon\end{array}$$

is comparable to K_*: $K \leq \text{Ceil}\left(\frac{2\ln(M(N+1))/\ln(1/\epsilon)}{1-\ln(4(1-\epsilon))/\ln(1/\epsilon)} K_*\right)$.

Indeed, let $H_{ij}: \|x_* - x_i\| \leq r_j$ and $H_{i'j'}: \|x_* - x_{i'}\| \leq r_{j'}$ be not $\mathcal{C}_{\alpha,\beta}$-close to each other.
<u>Claim:</u> *H_{ij} and $H_{i'j'}$ can be decided upon via K_* observations with risk $\leq \epsilon$.*
Here is K_*-observation test $\mathcal{T}$ deciding on H_{ij} vs. $H_{i'j'}$ with risk $\leq \epsilon$:

> *Given ω^{K_*}, apply the inference $\omega^{K_*} \to i_*(\omega^{K_*})$ and check whether $\|x_i - x_{i_*(\omega^{K_*})}\| \leq (a+1)r_j + b$. If it is the case, accept H_{ij}, otherwise accept $H_{i'j'}$.*

Let us prove that the risk of $\mathcal{T}$ is $\leq \epsilon$. Indeed, let the event $\mathcal{E}: \|x_* - x_{i_*(\omega^{K_*})}\| \leq a\rho(x_*) + b$ take place (it happens with probability $\geq 1-\epsilon$). Then
— if H_{ij} is true, we have $\|x_* - x_i\| \leq r_j$, whence $\rho(x_*) \leq r_j$ and thus $\|x_* - x_{i_*(\omega^{K_*})}\| \leq ar_j + b$. Red relations imply that $\|x_i - x_{i_*(\omega^{K_*})}\| \leq (a+1)r_j + b$, thus $\mathcal{T}$ accepts H_{ij}. Thus, *when $\mathcal{E}$ takes place and H_{ij} is true, $\mathcal{T}$ accepts H_{ij}.*
— if $H_{i'j'}$ is true, we, same as above, have $\|x_{i'} - x_{i_*(\omega^{K_*})}\| \leq (a+1)r_{j'} + b$. Assuming that $\mathcal{T}$ rejects $H_{i'j'}$ we have also $\|x_i - x_{i_*(\omega^{K_*})}\| \leq (a+1)r_j + b$, implying that $\|x_i - x_{i'}\| \leq (a+1)[r_j + r_{j'}] + 2b$, which is not the case since H_{ij} and $H_{i'j'}$ are not $\mathcal{C}_{2a+3,2b}$-close to each other.
<u>Bottom line:</u> *When $\mathcal{E}$ takes place, $\mathcal{T}$ makes no errors, so that the risk of $\mathcal{T}$ is $\leq \epsilon$.* $\square$

$\Rightarrow$ *Whenever H_{ij}, $H_{i'j'}$ are not $\mathcal{C}_{\alpha,\beta}$-close to each other, we have $\epsilon_{ij,i'j'} \leq [2\sqrt{\epsilon(1-\epsilon)}]^{1/K_*} < 1$*

$\Rightarrow$ *$\mathcal{T}^K$ with announced K is well defined and has $\mathcal{C}_{\alpha,\beta}$-risk $\leq \epsilon$.* $\square$

♣ **Numerical illustration:** Given noisy observation

$$\omega = Ax + \sigma\xi,\ \xi \sim \mathcal{N}(0, I_n)$$

of the "discretized primitive" Ax of a signal $x = [x^1; ...; x^n] \in \mathbb{R}^n$:

$$[Ax]_j = \frac{1}{n}\sum_{s=1}^{j} x^s,\ 1 \le j \le n,$$

for $i = 1, ..., \kappa$ we have built Least Squares polynomial, of order $i-1$, approximations x_i of x:

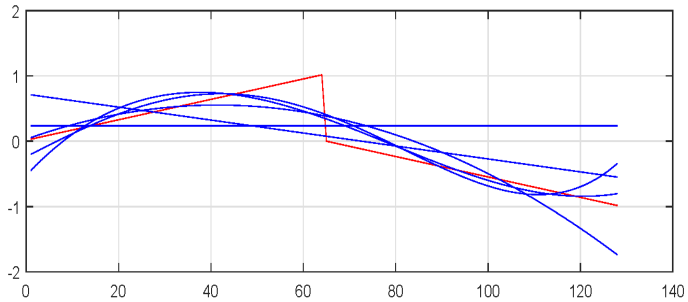
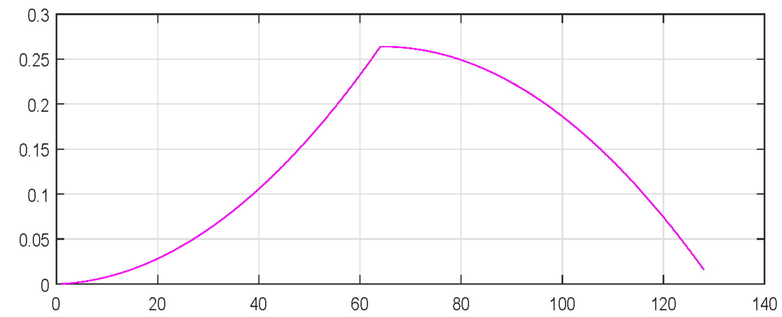
$$x_i = \mathrm{argmin}_{x\in\mathcal{X}_i} \|Ax - \omega\|_2^2$$
$$\left[\mathcal{X}_i = \{x = [x^1; ...; x^n] : x^s \text{ is polynomial, of degree} \le i-1, \text{ in } s\}\right]$$

and now want to use K additional observations to identify the nearly closest to x_*, in the norm

$$\|u\| = \frac{1}{n}\sum_{i=1}^{n} |u_i|$$

on $\mathbb{R}^n$, among the points x_i, $1 \le i \le \kappa$.

♠ **Experiment** [$\epsilon = 0.01$, $n = 128$, $\sigma = 0.01$, $\kappa = 5$, $\alpha = 3$, $\beta = 0.05$]

Left: x_* and x_i. Right: the primitive of x_*

i	1	2	3	4	5
$\|x - x_i\|$	0.5348	0.33947	0.23342	0.16313	0.16885

distances from x_* to x_i

- Computation yielded $K = 3$. But
  - — with $K = 3$, in sample of 1000 simulations, *not a single case of wrong identification of the exactly closest to x_* point was observed*, i.e., we always got $\|x - x_{i(\omega^3)}\| = \rho(x_*)$, in spite of the theoretical guarantee as poor as $\|x_* - x_{i(\omega^3)}\| \leq 3\rho(x_*) + 0.05$
  - — the same was true when $K = 3$ was replaced with $K = 1$;
  - — replacing $K = 3$ with $K = 1$ *and increasing σ from 0.01 to 0.05*, the procedure started to make imperfect conclusions. However, the exactly closest to x_* point x_4 was identified correctly in as many as 961 of 1000 simulations, and the empirical mean $\mathbf{E}\{\|x_* - x_{i(\omega^1)}\| - \rho(x_*)\}$ was as small as 0.0024.

How it Works: Illustration III
Recovering Linear-Fractional Function of a Signal

♣ **Problem:** An unknown signal x known to belong to a given convex compact set $X \subset \mathbb{R}^n$ is observed according to

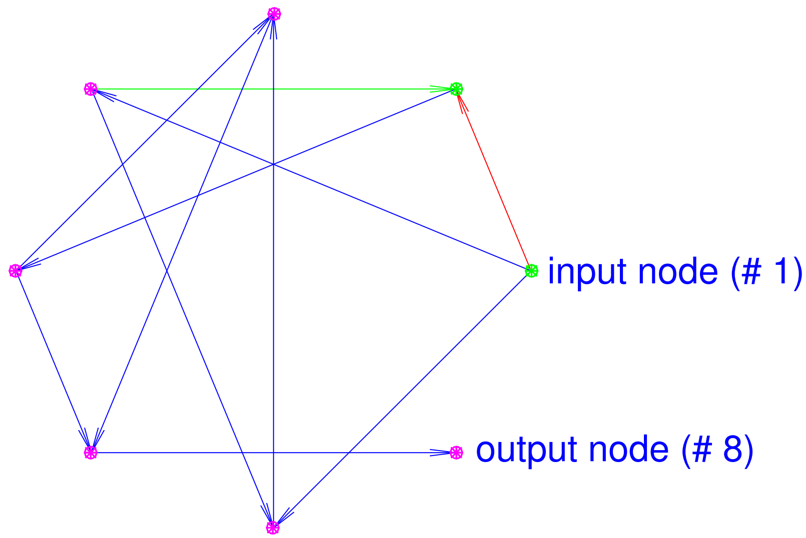
$$\omega = Ax + \sigma\xi, \ \xi \sim \mathcal{N}(0, I_d)$$

Our goal is to recover the value at x of a linear-fractional functional $F(z) = f^T z / e^T z$, with $e^T z > 0$, $z \in X$.

♠ **Illustration:** We are given noisy measurements of voltages V_i at *some* nodes i and currents I_{ij} in *some* arcs (i, j) of an electric circuit, and want to recover the resistance of a particular arc $(\widehat{i}, \widehat{j})$:

$$r_{\widehat{ij}} = \frac{V_{\widehat{j}} - V_{\widehat{i}}}{I_{\widehat{ij}}}$$

Circuit with 8 nodes and 11 arcs



$$x = [\text{voltages at nodes; currents in arcs}]$$

$$Ax = [\text{observable voltages; observable currents}]$$

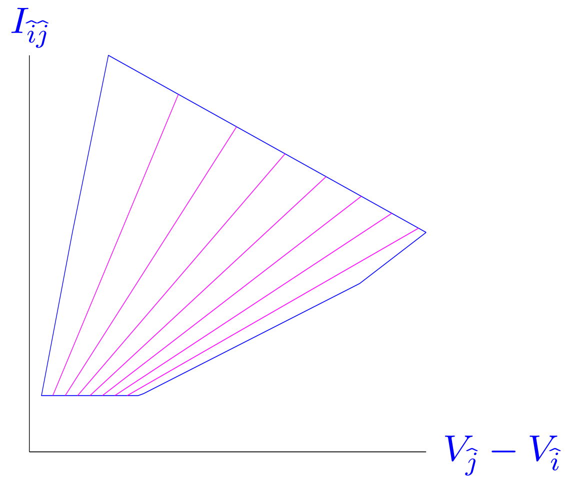
- Currents are measured in blue arcs only
- Voltages are measured in magenta nodes only
- We want to recover resistance of red arc
- $X:$ $\left\{\begin{array}{l}\textit{conservation of current, except for nodes \#\#1,8}\\ \textit{zero voltage at node \#1, nonnegative currents}\\ \textit{current in red arc at least 1, total of currents at most 33}\\ \textit{Ohm Law, resistances of arcs between 1 and 10}\end{array}\right.$

♠ **Strategy:** Given L,

• split the range $\Delta = [\min_{x\in X} F(x), \max_{x\in X} F(x)]$ into L consecutive bins Δ_ℓ of length $\delta_L = \mathrm{length}(\Delta)/L$,

• define the convex compact sets

$$X_\ell = \{x \in X : F(x) \in \Delta_\ell\},\ \ M_\ell = \{Ax : x \in X_\ell\},\ 1 \le \ell \le L$$



2D projections of X and $X_1, ..., X_8$

• decide on L hypotheses $H_\ell : P = \mathcal{N}(\mu, \sigma^2 I), \mu \in M_\ell$ on the distribution P of observation $\omega = Ax + \sigma\xi$ *up to closeness* $\mathcal{C}$ *"H_ℓ is close to $H_{\ell'}$ iff* $|\ell - \ell'| \le 1$*"*

• estimate $F(x)$ by the center of masses of all accepted bins.

♠ **Fact:** *For the resulting test $\mathcal{T}$, with probability $\ge 1 - \mathrm{Risk}^{\mathcal{C}}(\mathcal{T}|H_1, ..., H_L)$ the estimation error does not exceed δ_L.*

♠ **Implementation and results:** Given target risk ϵ and L, we selected the largest σ for which $\mathrm{Risk}^{\mathcal{C}}(\mathcal{T}|H_1,...,H_L)$ is $\leq \epsilon$.

• This is what we get in our Illustration for $\epsilon = 0.01$: $\Delta = [1, 10]$

L	8	16	32
δ_L	$9/8 \approx 1.13$	$9/16 \approx 0.56$	$9/32 \approx 0.28$
σ	0.024	0.010	0.005
$\sigma_{\rm opt}/\sigma \leq$	1.31	1.31	1.33
σ	0.031	0.013	0.006
$\sigma_{\rm opt}/\sigma \leq$	1.01	1.06	1.08

• $\sigma_{\rm opt}$ – the largest σ for which "in the nature" there exists a test deciding on $H_1,...,H_L$ with $\mathcal{C}$-risk ≤ 0.01

• **Red data:** Risks $\epsilon_{\ell\ell'}$ of pairwise tests are bounded via risks of optimal detectors, $\mathcal{C}$-risk of $\mathcal{T}$ is bounded by

$$\left\|\left[\epsilon_{\ell\ell'} \cdot \chi_{(\ell,\ell')\notin\mathcal{C}}\right]_{\ell,\ell'=1}^{L}\right\|_{2,2};$$

• **Brown data:** Risks $\epsilon_{\ell\ell'}$ of pairwise tests are bounded via error function, $\mathcal{C}$-risk of $\mathcal{T}$ is bounded by

$$\max_{\ell} \sum_{\ell':(\ell,\ell')\notin\mathcal{C}} \epsilon_{\ell\ell'}.$$

Illustration III Revisited
Recovering N-Convex Functionals

♣ **Fact:** The construction used to recover linear-fractional function can be extended to recovering N*-convex functionals*.

♠ **Definition:** *Let $X \subset \mathbb{R}^n$ be a convex compact set, $F : X \to \mathbb{R}$ be a continuous function, and N be a positive integer. We say that F is N-convex, if for every real a the sets*

$$X^{a,\geq} = \{x \in X : F(x) \geq a\}, \; X^{a,\leq} = \{x \in X : F(x) \leq a\}$$

can be represented as the unions of at most N convex compact sets.

Examples: A. *Fractional-linear function $F(x) = \frac{e(x)}{d(x)}$ with positive on X denominator is 1-convex:*

$$\{x \in X : F(x) \underset{\leq}{\overset{\geq}{}} a\} = \{x \in X : e(x) - ad(x) \underset{\leq}{\overset{\geq}{}} 0\}$$

B. *If* F_χ *is* N_χ*-convex on* X, $\chi = 1,2$, *then* $\max[F_1,F_2]$ *and* $\min[F_1,F_2]$ *are* $\max[N_1+N_2, N_1N_2]$*-convex on* X*:*

$$\Rightarrow \left\{\begin{array}{l} \left\{\begin{array}{l} X_\chi^{a,\geq} := \{x\in X : F_\chi(x)\geq a\} = \bigcup_{\nu=1}^{N_\chi} U^a_{\nu,\chi} \\ X_\chi^{a,\leq} := \{x\in X : F_\chi(x)\leq a\} = \bigcup_{\nu=1}^{N_\chi} V^a_{\nu,\chi} \end{array}\right. , \chi = 1,2 \text{ [}U,V\text{: convex]} \\ \{x\in X : \max[F_1(x),F_2(x)]\geq a\} = \left[\bigcup\limits_{\mu\leq N_1} U^a_{\mu,1}\right]\bigcup\left[\bigcup\limits_{\nu\leq N_2} U^a_{\nu,2}\right] \\ \{x\in X : \max[F_1(x),F_2(x)]\leq a\} = \bigcup\limits_{\mu\leq N_1,\nu\leq N_2}\left[V^a_{\mu,1}\bigcap V^a_{\nu,2}\right] \end{array}\right.$$

C. Conditional quantile. Let a probabilistic vector $0 < p \in \mathbb{R}^n$ represent probability distribution on finite subset $S = \{s_1 < s_2 < ... < s_n\}$ of the real axis.

Regularized α-quantile $q_\alpha[p]$ is defined as follows:

— we pass from p to the probability distribution P in $\Delta = [s_1, s_n]$ by assigning probability mass p_1 to s_1 and uniformly spreading the probability masses p_i, $i > 1$, over the segments $[s_{i-1}, s_i]$

— $q_\alpha[p]$ is the usual α-quantile of P:

$$q_\alpha[p] = \min\{s \in \Delta : \mathrm{Prob}_{\xi\sim P}\{\xi \leq s\} \geq \alpha\}$$

Fact: *Let $X = \{x(t,s) : t \in T, s \in S\}$ be a convex compact set comprised of nonvanishing probability distributions on 2D grid $T \times S$, let $t \in T$, and let $\left\{x_{|t}(s) = \frac{x(t,s)}{\sum_{s' \in S} x(t,s')}, s \in S\right\}$ be the conditional, given t, probability distribution on S induced by $x \in X$. Then*

$$f_{\alpha,t}(x) = q_\alpha[x_{|t}(\cdot)]$$

is 1-convex function of $x \in X$.

♠ **Problem of interest:** *Given*

- *convex compact set* $X \subset \mathbb{R}^n$,
- N*-convex functional* $F : X \to \mathbb{R}$,
- *a collection* X_j, $\ell = 1, ..., J$, *of convex compact subsets of* X,
- *stationary* K*-repeated observations* $\omega_1, ..., \omega_K$ *stemming, via simple o.s.,*

  from unknown signal $x \in \bigcup_{j=1}^{J} X_j$,

we want to recover $F(x)$.

Strategy: Given L, we

- Split the range $\Delta = [\min_{x \in X} F(x), \max_{x \in X} F(x)]$ into L consecutive bins Δ_ℓ of length $\delta_L = \text{length}(\Delta)/L$,
- Observe that by N-convexity of F every one of the sets

$$\{x \in \bigcup_{j=1}^{J} X_j : F(x) \in \Delta_\ell\}$$

is the union of at most $N^2 J$ convex compact sets Y_s^ℓ

- Associate with the nonempty among the sets Y_s^ℓ the hypotheses "observation stems from a signal from Y_s^ℓ"
- Define closeness $\mathcal{C}$ on the resulting collection of hypotheses $H_1, ..., H_{\mathcal{L}}$, $\mathcal{L} \leq N^2 JL$, by claiming H_μ and H_ν $\mathcal{C}$-close iff both hypotheses stem from the same or from two consecutive bins Δ_ℓ
- Use our machinery for testing multiple convex hypotheses in simple o.s. to build a test $\mathcal{T}_K$ deciding on $H_1, ..., H_{\mathcal{L}}$ up to closeness $\mathcal{C}$ via K-repeated observation.

• Apply the test $\mathcal{T}_K$ to observations $\omega_1, ..., \omega_K$ and take as the estimate of $F(x)$ the center of masses of all bins associated with the hypotheses accepted by the test.

♠ Same as in the above fractional-linear example, it is immediately seen that

• *The probability for the recovery error to be* $> \delta_L$ *is upper-bounded by the* $\mathcal{C}$*-risk of* $\mathcal{T}_K$.

In addition, *with our estimate, the number of observations K required to ensure recovery error $\leq \delta_L$ with a given reliability $1-\epsilon$, $\epsilon \ll 1$, is within logarithmic in N, J, L factor off the "ideal" number of observations needed to achieve, with reliability $1-\epsilon$, recovery error $\delta_L/2$.*

Sequential Hypothesis Testing

♣ **Motivating example: Opinion polls.** Recall the elections' story:

• Population-wide elections with L candidates are to be held.

• Preferences of a voter are represented by L-dimensional basic orth with 1 in position ℓ meaning voting for candidate $\#\ell$.

Equivalently: Preference ω of a voter is a vertex in the L-dimensional probabilistic simplex

$$\mathbf{\Delta}_L = \{p \in \mathbb{R}^L : p \geq 0, \sum_\ell p_\ell = 1\}.$$

• The average $\mu = [\mu_1; ...; \mu_L]$ of preferences of all voters "encodes" election's outcome: μ_ℓ is the fraction of voters supporting ℓ-th candidate, and the winner corresponds to the largest entry in μ (assumed to be uniquely defined).

Note: *μ is a probabilistic vector: $\mu \in \mathbf{\Delta}_L$. We think of μ as of a probability distribution on the L-element set $\Omega = \mathrm{Ext}(\mathbf{\Delta}_L)$ of vertices of $\mathbf{\Delta}_L$.*

• **Our goal** is to design *opinion poll* – to select K voters at random from the uniform distribution on the voters' population and to observe their preferences, in order to predict, with reliability $1 - \epsilon$, election's outcome.

♠ **Poll's model** is drawing stationary K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$, $\omega_k \in \Omega$, from distribution μ.

♠ We assume that the elections never end with "near tie," that is, the fraction of votes for the winner is at least by a known margin δ larger than the fraction of votes for every no-winner, and introduce L hypotheses on the distribution μ from which $\omega_1, ..., \omega_K$ are drawn:

$$H_\ell : \mu \in \mathcal{P}_\ell = \{\mu \in \mathbf{\Delta}_L : \mu_\ell \geq \mu_{\ell'} + \delta, \forall \ell' \neq \ell\}, \ell = 1, ..., L$$

Our goal is to specify K in a way which allows to decide on $H_1, ..., H_L$ via stationary K-repeated observations with risk $\leq \epsilon$.

♠ We are in the case of Discrete o.s., and can use our machinery to build a near-optimal K-observation test deciding on $H_1, ..., H_L$ up to trivial closeness $\mathcal{C}$ "H_ℓ is close to $H_{\ell'}$ iff $\ell = \ell'$" and then select the smallest K for which the $\mathcal{C}$-risk of this test is $\leq \epsilon$.

♠ **Illustration** $L = 2$**:** In this case Ω is two-point set of basic orths in $\mathbb{R}^2$, the minimum risk single-observation detector is

$$\phi_*(\omega) = \frac{1}{2}\ln\left(\frac{1+\delta}{1-\delta}\right)[\omega^1 - \omega^2] : \Omega \to \mathbb{R}$$

and $\mathrm{Risk}[\phi_*|\mathcal{P}_1, \mathcal{P}_2] = 1 - \delta^2$

$$\Rightarrow K = \mathrm{Ceil}\left(\frac{\ln(1/\epsilon)}{\ln(1/(1-\delta^2))}\right) \asymp \frac{1}{\delta^2}\ln(1/\epsilon).$$

δ	0.3162	0.1000	0.0316	0.0100
$\underline{K} \vee K$	$16 \vee 57$	$166 \vee 597$	$1,660 \vee 5,989$	$16,607 \vee 59,912$

$\underline{K}$: *lower* bound on optimal poll size

Poll sizes, $\epsilon = 0.05$

δ	0.3162	0.1000	0.0316	0.0100
$\underline{K} \vee K$	$16 \vee 57$	$166 \vee 597$	$1,660 \vee 5,989$	$16,607 \vee 59,912$

$\underline{K}$: *lower* bound on optimal poll size

Bad news: *Required size of opinion poll grows rapidly as "winning margin" decreases.*

♠ **Question:** *Can we do better?*

♠ **Partial remedy:** Let us pass to *sequential tests*, where we *attempt* to make conclusion *before* all K respondents required by the worst-case-oriented analysis are interviewed.

Hope: If elections are about to be "landslide" (i.e., in unknown to us actual distribution μ_* of voters' preferences the winner beats all other candidates by margin $\delta_* \gg \delta$), the winner hopefully can be identified after a relatively small number of interviews.

♣ **Strategy.** We select a number S of *attempts* and associate with attempt s number $K(s)$ of observations, $K(1) < ... < K(S)$.
s-th attempt to make inference is made when $K(s)$ observations are collected. When it happens, we apply to the collected so far observation $\omega^{K(s)} = (\omega_1, ..., \omega_{K(s)})$ a test $\mathcal{T}_s$ which, depending on $\omega^{K(s)}$,

– either accepts exactly one of the hypotheses $H_1, ..., H_L$, in which case we terminate,

– or claims that information collected so far does not allow to make an inference, in which case we pass to collecting more observations (when $s < S$) or terminate (when $s = S$).

♠ **Specifications:** We want the overall procedure to be

• *conclusive:* an inference should be made in one of the S attempts (thus, when attempt S is reached, making inference becomes a *must*);

• *reliable:* whenever the true distribution μ_* underlying observations obeys one of our L hypotheses, the μ_*-probability for this hypothesis to be eventually accepted should be $\geq 1 - \epsilon$, where $\epsilon \in (0, 1)$ is a given in advance risk bound.

♠ An implementation:

• We select somehow the number of attempts S and set $\delta_s = \delta^{s/S}$ so that $\delta_1 > \delta_2 > ... > \delta_S = \delta$. Besides this, we split risk bound ϵ into S parts ϵ_s: $\epsilon_s > 0$, $s \leq S$ & $\sum_{s=1}^{S} \epsilon_s = \epsilon$;

• For $s < S$, we define $2L$ hypotheses

$$\begin{array}{rcl} H^s_{2\ell-1} & = & H_\ell : \mu \in \mathcal{P}^s_{2\ell-1} = \mathcal{P}_\ell := \{\mu \in \mathbf{\Delta}_L : \mu_\ell \geq \delta_S + \max_{\ell' \neq \ell} \mu_{\ell'}\} \\ & & \text{“weak hypothesis”} \\ H^s_{2\ell} & = & \{\mu \in \mathcal{P}^s_{2\ell} := \{\mu \in \mathbf{\Delta}_L : \mu_\ell \geq \delta_s + \max_{\ell' \neq \ell} \mu_{\ell'}\} \subset \mathcal{P}_\ell \\ & & \text{“strong hypothesis”} \end{array}$$

$1 \leq \ell \leq L$, and assign $H^s_{2\ell-1}$ and $H^s_{2\ell}$ with color ℓ, $1 \leq \ell \leq L$.

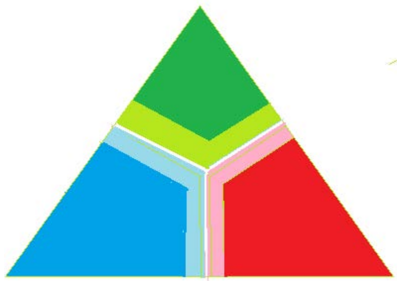
• For $s = S$ we introduce L hypotheses $H^S_\ell = H_\ell$, $\ell = 1, ..., L$, with H^S_ℓ assigned color ℓ.

• For $s < S$, we introduce closeness relation $\mathcal{C}_s$ on the collection of hypotheses $H^s_1, ..., H^s_{2L}$ as follows:

• the only hypotheses close to a strong hypothesis $H^s_{2\ell}$ are the hypotheses $H^s_{2\ell}$ and $H^s_{2\ell-1}$ of the same color;

• the only hypotheses close to a weak hypothesis $H^s_{2\ell-1}$ are all weak hypotheses and the strong hypothesis $H_{2\ell}$ of the same color as $H_{2\ell-1}$.

• For $s = S$, the $\mathcal{C}_s$-closeness is trivial: $H^S_\ell \equiv H_\ell$ is $\mathcal{C}_S$-close to $H^S_{\ell'} \equiv H_{\ell'}$ if and only if $\ell = \ell'$.



3-candidate hypotheses in probabilistic simplex Δ_3

[weak green]	M_1	dark green + light green: candidate A wins with margin $\geq \delta_S$
[strong green]	M_1^s	dark green: candidate A wins with margin $\geq \delta_s > \delta_S$
[weak red]	M_2	dark red + pink: candidate B wins with margin $\geq \delta_S$
[strong red]	M_2^s	dark red: candidate B wins with margin $\geq \delta_s > \delta_S$
[weak blue]	M_3	dark blue + light blue: candidate C wins with margin $\geq \delta_S$
[strong blue]	M_3^s	dark blue: candidate C wins with margin $\geq \delta_s > \delta_S$

- $H^s_{2\ell-1}$: $\mu \in M_\ell$ [weak hypothesis]
  weak hypothesis $H^s_{2\ell-1}$ is $\mathcal{C}_s$-close to itself, to all other weak hypotheses and to strong hypothesis $H^s_{2\ell}$ of the same color as $H^s_{2\ell-1}$
- $H^s_{2\ell}$: $\mu \in M^s_\ell$ [strong hypothesis]
  strong hypothesis $H^s_{2\ell}$ is $\mathcal{S}$-close only to itself and to weak hypothesis $H^s_{2\ell-1}$ of the same color as $H^s_{2\ell-1}$

• **Note:** We are in the case of stationary repeated observations in Discrete o.s., the hypotheses H_j^s are of the form *"i.i.d. observations $\omega_1, \omega_2, ...$ are drawn from distribution $\mu \in M_j^s$ with nonempty closed convex sets $M_j^s \subset \Delta_L$,"* *and sets M_j^s, $M_{j'}^s$ with $(j,j') \not\in \mathcal{C}_s$ do not intersect*

$\Rightarrow$ the risks of the minimum-risk pairwise detectors for $\mathcal{P}_j^s$, $\mathcal{P}_{j'}^s$, $(j,j') \not\in \mathcal{C}_s$, are < 1

$\Rightarrow$ we can efficiently find out the smallest $K = K(s)$ for which our machinery produces a test $\mathcal{T} = \mathcal{T}_s$ deciding, via stationary $K(s)$-repeated observations, on the hypotheses $\{H_j^s\}_j$ with $\mathcal{C}_s$-risk $\leq \epsilon_s$.

• It is easily seen that $K(1) < K(2) < ... < K(S-1)$. In addition, discarding all attempts $s < S$ with $K(s) < K(S)$ and renumbering the remaining attempts, we may assume w.l.o.g. that $K(1) < K(2) < ... < K(S)$.

♠ **Our inference routine** works as follows: we observe ω_k, $k = 1, 2, ..., K(S)$ (i.e., carry interviews with one by one randomly selected voters), and perform s-th attempt to make conclusion when $K(s)$ observations are acquired ($K(s)$ interviews are completed).

At s-th attempt, we apply the test $\mathcal{T}_s$ to observation $\omega^{K(s)}$. If the test does accept some of the hypotheses H_j^s *and all accepted hypotheses have the same color ℓ*, we accept ℓ-th of our original hypotheses $H_1, ..., H_L$ (i.e., predict that ℓ-th candidate will be the winner) and terminate, otherwise we proceed to next observations (i.e., next interviews) (when $s < S$) or claim the winner to be, say, the first candidate and terminate (when $s = S$).

♠ **Facts:**

- *The risk of the outlined sequential hypothesis testing procedure is $\leq \epsilon$: whenever the distribution μ_* underlying observations obeys hypothesis H_ℓ for some $\ell \leq L$, the μ_*-probability of the event "H_ℓ is the only accepted hypothesis" is at least $1-\epsilon$.*

- *The worst-case volume of observations $K(S)$ is within logarithmic factor from the minimal number of observations allowing to decide on the hypotheses $H_1, ..., H_L$ with risk $\leq \epsilon$.*

- *Whenever the distribution μ_* underlying observations obeys strong hypothesis $H^s_{2\ell}$ for some ℓ and s ("distribution μ_* of voters' preferences corresponds to winning margin at least δ_s"), the conclusion, with μ_*-probability $\geq 1-\epsilon$, will be made in course of the first s attempts (i.e., in course of the first $K(s)$ interviews).*

Informally: In landslide elections, the winner will be predicted reliably after a small number of interviews.

How it Works: 2-Candidate Elections

♠ **Setup:**

- # of candidates $L = 2$
- # $\delta_s = 10^{-s/4}$
- range of # of attempts S: $1 \le S \le 8$

♠ **Numerical Results:**

S	1	2	4	5	6	8
$\delta = \delta_S$	0.5623	0.3162	0.1000	0.0562	0.0316	0.0100
K	25	88	287	917	9206	92098
$K(S)$	25	152	1594	5056	16005	160118

Volume K of non-sequential test, number of attempts S and worst-case volume $K(S)$ of sequential test as functions of winning margin $\delta = \delta_S$. Risk ϵ is set to 0.01.

Note: Worst-case volume of sequential test is essentially worse than the volume of non-sequential test.
But: *When drawing the true distribution μ_* of voters' preferences at random from the uniform distribution on the set of μ's with winning margin ≥ 0.01, the typical size of observations used by Sequential test with $S = 8$ prior to termination is $\ll K(S)$*:

Empirical Volume of Sequential test

median	mean	60%	65%	75%	80%	85%	90%	95%	100%
177	9182	177	397	617	1223	1829	8766	87911	160118

Column "X%": empirical X%-quantile of test's volume. Data over 1,000 experiments.
Empirical risk: 0.01

Measurement Design

♣ **Observation:** In our Hypothesis Testing setup, observation scheme is our "environment" and is completely out of our control. However, there are situations where the observation scheme is under our *partial* control.

♠ **Example: Opinion Poll revisited.** In our original Opinion Poll problem, a particular voter was represented by basic orth $\omega = [0; ...; 0; 1; 0; ...; 0] \in \mathbb{R}^L$, with entry 1 in position ℓ meaning that the voter prefers candidate ℓ to all other candidates. Our goal was to predict the winner by observing preferences of respondents selected at random from uniform distribution on voters' population.

However: Imagine we can split voters in I non-intersecting groups (say, according to age, education, gender, income, occupation,...) in such a way that we have certain a priori knowledge of the distribution of preferences within the groups. In this situation, our poll can be organized as follows:

- We assign the groups with nonnegative weights q_i summing up to 1
- To organize an interview, we first select at random one of the groups, with probability q_i to select group i, and then select a respondent from i-th group at random, from uniform distribution on the group.

- We assign the groups with nonnegative weights q_i summing up to 1
- To organize an interview, we first select at random one of the groups, with probability q_i to select group i, and then select a respondent from i-th group at random, from uniform distribution on the group.

Note: When q_i is equal to the fraction θ_i of group i in the entire population, the above policy reduces to the initial one. It can make sense, however, to use q_i different from θ_i, with $q_i \ll \theta_i$ if a priori information about preferences of voters from i-th group is rich, and $q_i \gg \theta_i$ if this information is poor. Hopefully, this will allow us to make more reliable predictions with the same total number of interviews.

♠ **The model** of outlined situation is as follows:

- We characterize distribution of preferences within group i by vector $\mu^i \in \mathbf{\Delta}_L$. for $1 \leq \ell \leq L$, ℓ-th entry in μ^i is the fraction of voters *in group* i voting for candidate ℓ;
  Note: The population-wide distribution of voters' preferences is $\mu = \sum_{i=1}^I \theta_i \mu^i$.
- A priori information on distribution of preferences of voters from group i is modeled as the inclusion $\mu^i \in M^i$, for some known subset $M^i \subset \mathbf{\Delta}_L$ *which we assume to be nonempty convex compact set.*
- Output of particular interview is pair (i, j), where $i \in \{1, ..., I\}$ is selected at random according to probability distribution q, and j is the candidate preferred by respondent selected from group i at random, according to uniform distribution on the group.

⇒ *Our observation* (outcome of an interview) *becomes*

$$\omega := (i, \ell) \in \Omega = \{1, ..., I\} \times \{1, ..., L\}, \text{ Prob}\{\omega = (i, j)\} = p(i, j) := q_i \mu^i_j.$$

The hypotheses to be decided upon are

$$H_\ell[q] : p \in \mathcal{P}_\ell[q] := \left\{ \{p_{ij} = q_i \mu^i_j\}_{\substack{1 \leq i \leq I,\\ 1 \leq j \leq L}} : \begin{array}{l} \mu^i \in M^i \, \forall i, \\ \left[\sum_i \theta_i \mu^i\right]_\ell \geq \delta + \left[\sum_i \theta_i \mu^i\right]_{\ell'} \quad \forall (\ell' \neq \ell) \end{array} \right\}$$

$H_\ell[q]$, $\ell = 1, ..., L$, states that the "signal" $\vec{\mu} = [\mu^1; ...; \mu^I]$ underlying distribution p of observations ω induces population-wide distribution $\sum_i \theta_i \mu^i$ of votes resulting in electing candidate ℓ with winning margin $\geq \delta$.

$$H_\ell[q] : p \in \mathcal{P}_\ell[q] := \left\{ \{p_{ij} = q_i\mu^i_j\}_{\substack{1\le i\le I,\\ 1\le j\le L}} : \begin{array}{c} \mu^i \in M^i\,\forall i, \\ \left[\sum_i \theta_i \mu^i\right]_\ell \ge \delta + \left[\sum_i \theta_i \mu^i\right]_{\ell'} \quad \forall(\ell' \ne \ell) \end{array} \right\}$$

♠ **Note:** Hypotheses $H_\ell[q]$ are of the form

$$H_\ell[q] = \{p = A[q]\vec{\mu} : \vec{\mu} := [\mu^1; ...; \mu^L] \in \mathcal{M}^\ell\},$$
$$[A[q]\vec{\mu}]_{ij} = q_i\mu^i_j,$$

where $\mathcal{M}^\ell$, $\ell = 1, ..., L$, are nonempty nonintersecting convex compact subsets in $\underbrace{\mathbf{\Delta}_L \times ... \times \mathbf{\Delta}_L}_{I}$

Note: Opinion Poll with K interviews corresponds to stationary K-repeated observation in Discrete o.s. with (IL)-element observation space Ω

$\Rightarrow$ *Given K, we can use our machinery to design a near-optimal detector-based test $\mathcal{T}_K$ deciding via stationary K-repeated observation* (i.e., via the outcomes of K interviews) *on hypotheses $H_\ell[q]$, $\ell = 1, ..., L$ up to trivial closeness "$H_\ell[q]$ is close to $H_{\ell'}[q]$ iff $\ell = \ell'$." This test will predict the winner with reliability $1 - \mathrm{Risk}(\mathcal{T}_K|H_1[q], ..., H_L[q])$.*

$$H_\ell[q] = \{p = A[q]\vec{\mu} : \vec{\mu} := [\mu^1; ...; \mu^L] \in \mathcal{M}^\ell\},$$
$$[A[q]\vec{\mu}]_{ij} = q_i \mu^i_j,$$

♠ By our theory, setting $\chi_{\ell\ell'} = \left\{\begin{array}{ll} 0, & \ell = \ell' \\ 1, & \ell \neq \ell' \end{array}\right.$, we have

$$\mathrm{Risk}(\mathcal{T}_K | H_1[q], ..., H_L[q]) \leq \epsilon_K[q] := \left\| \left[\epsilon^K_{\ell\ell'}[q]\chi_{\ell\ell'}\right]_{\ell,\ell'=1}^L \right\|_{2,2},$$
$$\begin{array}{rcl} \epsilon_{\ell\ell'}[q] & = & \max\limits_{\vec{\mu}\in\mathcal{M}^\ell, \vec{\nu}\in\mathcal{M}^{\ell'}} \sum_{i,j} \sqrt{[A[q]\vec{\mu}]_{ij}[A[q]\vec{\nu}]_{ij}} \\ & = & \max\limits_{\vec{\mu}\in\mathcal{M}^\ell, \vec{\nu}\in\mathcal{M}^{\ell'}} \underbrace{\sum_{i=1}^I q_i \left[\sum_{j=1}^L \sqrt{\mu^i_j \nu^i_j}\right]}_{\Phi(q;\vec{\mu},\vec{\nu})} \end{array}$$

Note: $\Phi(q; \vec{\mu}, \vec{\nu})$ is linear in q.

♣ *Let us carry out Measurement Design – optimization of $\epsilon_K[q]$ in q.*

♠ **Main observation:** $\epsilon_K[q] = \Gamma(\Psi(q))$, where

• $\Gamma(Q) = \|[(Q_{\ell\ell'})^K \chi_{\ell\ell'}]_{\ell,\ell'=1}^L\|_{2,2}$ is efficiently computable *convex* and *entrywise nondecreasing* function on the space of nonnegative $L \times L$ matrices

• $\Psi(q)$ is matrix-valued function with efficiently computable *convex in q and non-negative* entries

$$\Psi_{\ell\ell'}(q) = \max_{\vec{\mu}\in\mathcal{M}^\ell, \vec{\nu}\in\mathcal{M}^{\ell'}} \Phi(q; \vec{\mu}, \vec{\nu})$$

$\Rightarrow$ *Optimal selection of q_i's reduces to solving explicit convex problem*

$$\boxed{\min_q \left\{\Gamma(\Psi(q)) : q = [q_1; ...; q_I] \geq 0, \sum_{i=1}^I q_i = 1\right\}}$$

How it Works: Measurement Design in Election Polls

♠ Setup:

- Opinion Poll problem with L candidates and winning margin $\delta = 0.05$
- Reliability tolerance $\epsilon = 0.01$
- A priori information on voters' preferences in groups:

$$M^i = \{\mu^i \in \mathbf{\Delta}_L : p^i_\ell - u_i \le \mu^i_\ell \le p^i_\ell + u_i, \ell \le L\}$$

- p^i: radomly selected probabilistic vector • u_i: uncertainty level

♠ Sample of results:

L	I	Group sizes θ Uncertainty levels u	K_{ini}	q_{opt}	K_{opt}
2	2	$\theta = [0.50; 0.50]$ $u = [0.03; 1.00]$	1212	[0.44; 0.56]	1194
2	2	[0.50; 0.50] [0.02; 1.00]	2699	[0.00; 1.00]	1948
3	3	[0.33; 0.33; 0.33] [0.02; 0.03; 1.00]	3177	[0.00; 0.46; 0.54]	2726
5	4	[0.25; 0.25; 0.25; 0.25] [0.02; 0.02; 0.03; 1.00]	2556	[0.00; 0.13; 0.32; 0.55]	2086
5	4	[0.25; 0.25; 0.25; 0.25] [1.00; 1.00; 1.00; 1.00]	4788	[0.25; 0.25; 0.25; 0.25]	4788

Effect of measurement design. K_{ini} and K_{opt} are the poll sizes required for 0.99-reliable prediction of the winner when $q_i = \theta_i$ and $q = q_{\text{opt}}$, respectively.
Note: Uncertainty$= 1.00 \Leftrightarrow$ No a priori information

♣ In numerous situations, we do have partial control of observation scheme and thus can look for optimal Measurement Design.

However: the situations where optimal Measurement Design can be found efficiently, like in design of Election Polls, are rare.

Additional examples of these rare situations are *Poisson o.s. and Gaussian o.s. with time control.*

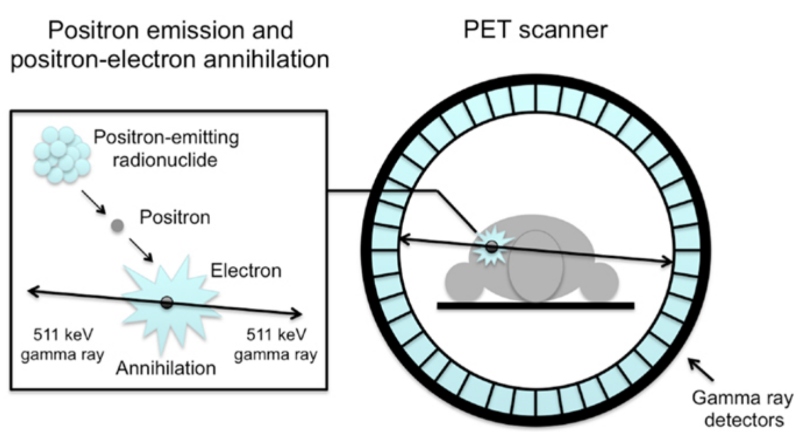
♠ **Poisson o.s. with time control.** Typical models where Poisson o.s. arises are as follows:

• "in the nature" there exists a "signal" x known to belong to some convex compact set $\subset \mathbb{R}^n$

For example, in Positron Emission Tomography, x is (discretized) density of radioactive tracer administered to patient

• We observe random vector $\omega \in \mathbb{R}^m$ with independent entries $\omega_i \sim \text{Poisson}(a_i^T x)$, and want to make inferences on x.

For example, in PET, tracer disintegrates, and every disintegration act results in pair of gamma-quants flying in opposite directions along a randomly oriented line passing through disintegration point. This line is registered when two detector cells are (nearly) simultaneously hit:



The data acquired in PET study are the numbers ω_i of lines registered in *bins* (pairs of detector cells) $i = 1, ..., m$ over a time horizon T, and

$$\boxed{\begin{array}{c}\omega_i \sim \mathrm{Poisson}(T \sum_{j=1}^n p_{ij} x_j) \\ \left[\begin{array}{c} p_{ij}\text{: probability for line emanated from voxel } j = 1, ..., n \\ \text{to cross pair } i = 1, ..., m \text{ of detector cells}\end{array}\right]\end{array} \Rightarrow \ A = T\left[p_{ij}\right]_{i \le m, j \le n}}$$

$$\boxed{\omega = \{\omega_i \sim \text{Poisson}([Ax]_i)\}_{i \leq m}}$$

- In some situations, the sensing matrix A can be partially controlled:

$$A = A[q] := \text{Diag}\{q\} A_*$$

- A_*: given $m \times n$ matrix; • $q \in \mathcal{Q}$: vector of control parameters.

For example, in a full body PET scan the position of the patient w.r.t. the scanner is updated several times to cover the entire body.

The data acquired in position ι form subvector ω^ι in the entire observation $\omega = [\omega^1; ...; \omega^I]$:

$$\omega_i^\iota \sim \text{Poisson}([t_\iota A^\iota x]_i, \; 1 \leq i \leq \bar{m} = m/I$$
$$\left[\; A^\iota : \text{ given matrices; } t_\iota : \text{ duration of study in position } \iota \; \right]$$

implying that $\omega = \text{Diag}\{q\} A_*$ with properly selected A_* and q of the form

$$q = \left[\underbrace{t_1; ...; t_1}_{\bar{m}}; ...; \underbrace{t_I; ...; t_I}_{\bar{m}} \right]$$

H_ℓ^q : $\omega_i \sim \text{Poisson}([A[q]x]_i)$ are independent across $i \leq m$ and $x \in X_\ell$
$$A = A[q] := \text{Diag}\{q\}A_*$$
• A_*: given $m \times n$ matrix; • $q \in \mathcal{Q}$: control parameters.

• Let our goal be to decide, up to a given closeness $\mathcal{C}$, on L hypotheses on the distribution of Poisson observation ω:

$$H_\ell^q : \omega \sim \text{Poisson}([A[q]x]_1) \times \ldots \times \text{Poisson}([A[q]x]_m) \ \& \ x \in X_\ell$$

X_ℓ: given convex compact sets, $1 \leq \ell \leq L$.

♠ By our theory, the (upper bound on the) $\mathcal{C}$-risk of near-optimal test deciding on H_ℓ^q, $\ell = 1, \ldots, L$, is $\epsilon(q) = \left\|[\exp\{\text{Opt}_{\ell\ell'}(q)\}\chi_{\ell\ell'}]_{\ell,\ell'=1}^L\right\|_{2,2}$ where

$$\chi_{\ell\ell'} = \begin{cases} 0, & (\ell,\ell') \in \mathcal{C} \\ 1, & (\ell,\ell') \notin \mathcal{C} \end{cases}, \text{Opt}_{\ell\ell'}(q) = \max_{u \in X_\ell, v \in X_{\ell'}} -\tfrac{1}{2}\sum_{i=1}^m \left(\sqrt{[A[q]u]_i} - \sqrt{[A[q]v]_i}\right)^2$$

• As in Opinion Polls, $\epsilon(q) = \Gamma(\Psi(q))$, where

• $\Gamma(Q) = \| [\exp\{Q_{\ell\ell'}\}\chi_{\ell\ell'}]_{\ell,\ell'=1}^L \|_{2,2}$ is a convex entrywise nondecreasing function of $Q \in \mathbb{R}_+^{L \times L}$

• $[\Psi(q)]_{\ell\ell'} = \exp\left\{\max_{u \in X_\ell, v \in X_{\ell'}} \sum_{i=1}^m q_i \left(\sqrt{[A_*u]_i[A_*v]_i} - \tfrac{1}{2}[A_*u]_i - \tfrac{1}{2}[A_*v]_i\right)\right\}$ is efficiently computable and convex in q

$\Rightarrow$ *Assuming the set $\mathcal{Q} \subset \mathbb{R}_+^m$ of allowed controls q convex, optimizing $\epsilon(q)$ over $q \in \mathcal{Q}$ is an explicitly given convex optimization problem.*

♣ An efficiently solvable Measurement Design problem in Gaussian o.s.

$$\omega = A[q]x + \xi,\ \xi \sim \mathcal{N}(0, I_m)$$

[• $A[q]$ partially controlled sensing matrix; • $q \in \mathcal{Q}$: control parameters.]

is the one where

$$A[q] = \mathrm{Diag}\{\sqrt{q_1}, ..., \sqrt{q_m}\}A_* \ \&\ \mathcal{Q} \subset \mathbb{R}^m_+ \text{ is a convex compact set}$$

In this case, minimizing $\mathcal{Q}$-risk of test deciding up to closeness $\mathcal{C}$ on L hypotheses

$$H_\ell^q : \omega \sim \mathcal{N}(A[q]x, I_m),\ x \in X_\ell,\ 1 \leq \ell \leq L$$

associated with nonempty convex compact sets X_ℓ reduces to solving convex problem

$$\min_{q \in \mathcal{Q}} \Gamma(\Psi(q))$$

where

$$\Gamma(Q) = \|\ [\exp\{Q_{\ell\ell'}/8\}\chi_{\ell\ell'}]_{\ell,\ell' \leq L}\ \|_{2,2}$$

is convex entrywise nondecreasing function of $L \times L$ matrix Q, and

$$\begin{array}{rcl}[\Psi(q)]_{\ell\ell'} & = & \max\limits_{u \in X_\ell, v \in X_{\ell'}} \left[-\|A[q](u-v)\|_2^2\right] \\ & = & -\min\limits_{u \in X_\ell, v \in X_{\ell'}} (u-v)^T A_*^T \mathrm{Diag}\{q\} A_*(u-v)\end{array}$$

is efficiently computable convex function of $q \in \mathcal{Q}$.

♠ **Illustration.** In some applications, "the physics" beyond Gaussian o.s. $\omega = Ax + \xi$ is as follows. There are m sensors measuring analogous vector-valued continuous time signal $x(t)$ (nearly constant on the observation horizon). The output of sensor $\#i$ is

$$\omega_i = \frac{1}{|\Delta_i|}\int_{\Delta_i}[a_{i,*}^T x(t) + B_i(t)]dt$$

- Δ_i : continuous time interval on which sensor #i is on
- $B_i(t)$: "Brownian motion:" $\frac{1}{|\Delta|}\int_{\Delta} B_i(t)dt \sim \mathcal{N}(0, |\Delta|^{-1})$, $\int_{\Delta} B_i(t)dt$, $\int_{\Delta'} B_i(t)dt$ are independent when $\Delta \cap \Delta' = \emptyset$
- Brownian motions $B_i(t)$ are independent across i

• When all sensors work in parallel for unit time, we arrive at the standard Gaussian o.s. $\omega = A_* x + \xi$, $\xi \sim \mathcal{N}(0, I_m)$.
• When sensors work on consecutive segments $\Delta_1, ..., \Delta_m$ of durations $q_i = |\Delta_i|$, we arrive at

$$\omega_i = a_{i,*}^T x + q_i^{-1/2}\xi_i,\ \xi_i \sim \mathcal{N}(0, 1) \text{ are independent across } i$$

Rescaling observations: $\omega_i \mapsto \sqrt{q_i}\omega_i$, we arrive at partially controlled o.s.

$$\omega = \mathrm{Diag}\{\sqrt{q_1}, ..., \sqrt{q_m}\}A_* x + \xi,\ \xi \sim \mathcal{N}(0, I_m)$$

A natural selection of $\mathcal{Q}$ is, e.g., $\mathcal{Q} = \{q \geq 0 : \sum_i q_i = m\}$ – setting the overall "time budget" to the same value as in the case of sensors working for unit time in paralel.

Recovering Linear Functionals in Simple o.s.

♣ **Situation:** Given are:

- Simple o.s. $\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$
- Convex compact set $X \subset \mathbb{R}^n$ and *affine* mapping $x \mapsto A(x) : X \to \mathcal{M}$
- Linear function $g^T x$ on $\mathbb{R}^n$

Given observation

$$\omega \sim p_{A(x)}$$

stemming from *unknown* signal x known to belong to X, we want to recover $g^T x$.

♣ Given reliability tolerance $\epsilon \in (0, 1)$, we quantify performance of a candidate estimate $\widehat{g}(\cdot) : \Omega \to \mathbb{R}$ by its ϵ*-risk*

$$\mathrm{Risk}_\epsilon[\widehat{g}|X] = \min\left\{\rho : \mathrm{Prob}_{\omega \sim p_{A(x)}}\left\{|\widehat{g}(\omega) - g^T x| > \rho\right\} \leq \epsilon \,\forall x \in X\right\}.$$

♣ We intend to build, *in a computationally efficient manner*, a *provably near-optimal* in terms of its ϵ-risk estimate of the form

$$\widehat{g}(\omega) = \phi(\omega) + \varkappa$$

with $\phi \in \mathcal{F}$.

♣ **Construction:** Let us set

$$\Phi(\phi;\mu) = \ln\left(\mathbf{E}_{\omega\sim p_\mu}\{\exp\{\phi(\omega)\}\}\right)$$

Recall that Φ is continuous real-valued convex-concave function on $\mathcal{F}\times\mathcal{M}$.

Main observation: *Let $\psi\in\mathcal{F}$ and $\alpha>0$. Then for $x,y\in X$ one has*

$$\begin{array}{ll}
\ln\left(\mathrm{Prob}_{\omega\sim p_{A(x)}}\left\{\psi(\omega)>g^Tx+\rho\right\}\right)\leq\Phi(\psi/\alpha;A(x))-\frac{\rho+g^Tx}{\alpha} & (a)\\
\ln\left(\mathrm{Prob}_{\omega\sim p_{A(y)}}\left\{\psi(\omega)<g^Ty-\rho\right\}\right)\leq\Phi(-\psi/\alpha;A(y))-\frac{\rho-g^Ty}{\alpha} & (b)
\end{array}$$

As a result, for every $\psi\in\mathcal{F}$ and $\alpha>0$, setting

$$\begin{array}{rcl}
\Psi_+(\alpha,\psi) & = & \max_{x\in X}\left[\alpha\Phi(\psi/\alpha;A(x))-g^Tx+\alpha\ln(2/\epsilon)\right],\\
\Psi_-(\alpha,\psi) & = & \max_{y\in X}\left[\alpha\Phi(-\psi/\alpha;A(y))+g^Ty+\alpha\ln(2/\epsilon)\right],\\
\varkappa & = & \frac{1}{2}\left[\Psi_-(\alpha,\psi)-\Psi_+(\alpha,\psi)\right],
\end{array}$$

for the estimate $\phi(\omega)=\psi(\omega)+\varkappa$ we have

$$\mathrm{Risk}_\epsilon[\phi(\cdot)|X]\leq\frac{1}{2}\left[\Psi_+(\alpha,\psi)+\Psi_-(\alpha,\psi)\right]$$

$$\Phi(\phi;\mu)=\ln\left(\mathbf{E}_{\omega\sim p_\mu}\{\exp\{\phi(\omega)\}\}\right)$$

Claim: For every $\psi\in\mathcal{F}, \alpha>0$ and all $x,y\in X$ one has

$$\begin{array}{ll}\ln\left(\mathrm{Prob}_{\omega\sim p_{A(x)}}\left\{\psi(\omega)>g^Tx+\rho\right\}\right)\leq\Phi(\psi/\alpha;A(x))-\frac{\rho+g^Tx}{\alpha} & (a)\\ \ln\left(\mathrm{Prob}_{\omega\sim p_{A(y)}}\left\{\psi(\omega)<g^Ty-\rho\right\}\right)\leq\Phi(-\psi/\alpha;A(y))-\frac{\rho-g^Ty}{\alpha} & (b)\end{array}$$

Indeed,

$$\exp\{\Phi(\psi/\alpha;A(x))\}=\mathbf{E}_{\omega\sim p_{A(x)}}\{\exp\{\psi(\omega)/\alpha\}\}=\mathbf{E}_{\omega\sim p_{A(x)}}\left\{\exp\{\tfrac{\psi(\omega)-g^Tx-\rho}{\alpha}\}\right\}\exp\{\tfrac{g^Tx+\rho}{\alpha}\}$$
$$\geq\mathrm{Prob}_{\omega\sim p_{A(x)}}\left\{\psi(\omega)>g^Tx+\rho\right\}\exp\{\tfrac{g^Tx+\rho}{\alpha}\}\Rightarrow(a);$$
$$\exp\{\Phi(-\psi/\alpha;A(y))\}=\mathbf{E}_{\omega\sim p_{A(y)}}\{\exp\{-\psi(\omega)/\alpha\}\}=\mathbf{E}_{\omega\sim p_{A(y)}}\left\{\exp\{\tfrac{-\psi(\omega)+g^Ty-\rho}{\alpha}\}\right\}\exp\{\tfrac{-g^Ty+\rho}{\alpha}\}$$
$$\geq\mathrm{Prob}_{\omega\sim p_{A(y)}}\left\{\psi(\omega)<g^Ty-\rho\right\}\exp\{\tfrac{-g^Ty+\rho}{\alpha}\}\Rightarrow(b).$$

$$\ln\left(\mathrm{Prob}_{\omega\sim p_{A(x)}}\left\{\psi(\omega) > g^Tx+\rho\right\}\right) \leq \Phi(\psi/\alpha; A(x)) - \frac{\rho+g^Tx}{\alpha} \quad (a)$$
$$\ln\left(\mathrm{Prob}_{\omega\sim p_{A(y)}}\left\{\psi(\omega) < g^Ty-\rho\right\}\right) \leq \Phi(-\psi/\alpha; A(y)) - \frac{\rho-g^Ty}{\alpha} \quad (b)$$

Claim: For every $\psi\in\mathcal{F}$ and $\alpha>0$, setting

$$\begin{array}{rcl}\Psi_+(\alpha,\psi) &=& \max_{x\in X}\left[\alpha\Phi(\psi/\alpha; A(x)) - g^Tx + \alpha\ln(2/\epsilon)\right],\\ \Psi_-(\alpha,\psi) &=& \max_{y\in X}\left[\alpha\Phi(-\psi/\alpha; A(y)) + g^Ty + \alpha\ln(2/\epsilon)\right],\\ \varkappa &=& \frac{1}{2}\left[\Psi_-(\alpha,\psi) - \Psi_+(\alpha,\psi)\right]\end{array}$$

we have

$$\mathrm{Risk}_\epsilon[\psi(\cdot)+\kappa|X] \leq \frac{1}{2}\left[\Psi_+(\alpha,\psi)+\Psi_-(\alpha,\psi)\right] \qquad (*)$$

Indeed, given $\psi\in\mathcal{F}$, $\alpha>0$, $z\in X$, let $\Psi_\pm = \Psi_\pm(\alpha,\psi)$, $\Psi=\frac{1}{2}\left[\Psi_+ + \Psi_-\right]$. We have

$$\begin{array}{l}\mathrm{Prob}_{\omega\sim p_{A(z)}}\left\{\psi(\omega)+\kappa > g^Tz+\Psi\right\} = \mathrm{Prob}_{\omega\sim p_{A(z)}}\left\{\psi(\omega) > g^Tz+\Psi_+\right\}\\ \leq \exp\{\Phi(\psi/\alpha; A(z)) - (\Psi_+ + g^Tz)/\alpha\} \text{ [by } (a)]\\ \leq \exp\{\Phi(\psi/\alpha; A(z)) - (\alpha\Phi(\psi/\alpha; A(z)) - g^Tz + \alpha\ln(2/\epsilon) + g^Tz)/\alpha\} = \epsilon/2\end{array}$$

and

$$\begin{array}{l}\mathrm{Prob}_{\omega\sim p_{A(z)}}\left\{\psi(\omega)+\kappa < g^Tz-\Psi\right\} = \mathrm{Prob}_{\omega\sim p_{A(z)}}\left\{\psi(\omega) < g^Tz-\Psi_-\right\}\\ \leq \exp\{\Phi(-\psi/\alpha; A(z)) - (\Psi_- - g^Tz)/\alpha\} \text{ [by } (b)]\\ \leq \exp\{\Phi(-\psi/\alpha; A(z)) - (\alpha\Phi(-\psi/\alpha; A(z)) + g^Tz + \alpha\ln(2/\epsilon) - g^Tz)/\alpha\} = \epsilon/2\end{array}$$

and $(*)$ follows.

♣ **Result:** We have justified the first claim in the following

Theorem [Ju&N'09] *In the situation in question, consider convex* (due to convexity-concavity of Φ) *optimization problem*

$$\mathrm{Opt} = \inf_{\alpha>0,\psi\in\mathcal{F}} \left\{ \Psi(\alpha,\omega) := \frac{1}{2}\left[\Psi_+(\alpha,\psi) + \Psi_-(\alpha,\psi)\right] \right\}.$$

A feasible solution α,ψ to this problem gives rise to estimate $\phi(\omega) = \psi(\omega) + \varkappa$ such that

$$\mathrm{Risk}_\epsilon[\phi|X] \leq \Psi(\alpha,\omega).$$

and the right hand side in this bound can be made arbitrarily close to Opt.

In addition, when $\epsilon < 1/2$, Opt *is within moderate factor of the minimax optimal ϵ-risk*

$$\mathrm{RiskOpt}_\epsilon[X] = \inf_{\widehat{g}(\cdot)} \mathrm{Risk}_\epsilon[\widehat{g}|X],$$

specifically,

$$\mathrm{Opt} \leq \frac{2\ln(2/\epsilon)}{\ln\left(\frac{1}{4\epsilon(1-\epsilon)}\right)} \mathrm{RiskOpt}_\epsilon[X].$$

Note: The "Gaussian o.s." version of this result is due to D. Donoho (1994).

Note: The above scheme is applicable to every simple o.s., in particular, to K-th degree of simple o.s. $\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$, that is, to the case where instead of estimation via *single* observation ω we speak about estimating via stationary K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$ with $\omega_1, ..., \omega_K$ supplied by $\mathcal{O}$.

In terms of $\mathcal{O}$, our Main Observation reads:

Let $\psi \in \mathcal{F}$, $\alpha > 0$, and $\psi^K(\omega^K) = \sum_{k=1}^K \psi(\omega_k)$. Then for $x, y \in X$ one has

$$\begin{array}{lr} \ln\left(\mathrm{Prob}_{\omega^K \sim p^K_{A(x)}}\left\{\psi^K(\omega^K) > g^T x + \rho\right\}\right) \leq K\Phi(\psi/\alpha; A(x)) - \frac{\rho + g^T x}{\alpha} & (a) \\ \ln\left(\mathrm{Prob}_{\omega^K \sim p^K_{A(y)}}\left\{\psi^K(\omega^K) < g^T y - \rho\right\}\right) \leq K\Phi(-\psi/\alpha; A(y)) - \frac{\rho - g^T y}{\alpha} & (b) \end{array}$$

As a result, for every $\psi \in \mathcal{F}$ and $\alpha > 0$, setting

$$\begin{array}{rcl} \Psi_+(\alpha, \psi) & = & \max_{x \in X}\left[K\alpha\Phi(\psi/\alpha; A(x)) - g^T x + \alpha \ln(2/\epsilon)\right], \\ \Psi_-(\alpha, \psi) & = & \max_{y \in X}\left[K\alpha\Phi(-\psi/\alpha; A(y)) + g^T y + \alpha \ln(2/\epsilon)\right], \\ \varkappa & = & \frac{1}{2}\left[\Psi_-(\alpha, \psi) - \Psi_+(\alpha, \psi)\right], \end{array}$$

for the estimate $\phi(\omega^K) = \sum_{k=1}^K \psi(\omega_K) + \varkappa$ we have

$$\mathrm{Risk}_\epsilon[\phi(\cdot)|X] \leq \frac{1}{2}\left[\Psi_+(\alpha, \psi) + \Psi_-(\alpha, \psi)\right]$$

Example: Gaussian o.s. Here $\mathcal{F}=\{\phi(\omega)=\psi_0+\psi^T\omega\}$; on a close inspection, we lose nothing when setting $\psi_0=0$.

$\Rightarrow \Phi(\psi,\mu)=\ln\left(c_n\int e^{\psi^T\omega-(\omega-\mu)^T(\omega-\mu)/2}d\omega\right)=\{\psi^T\mu+\frac{1}{2}\psi^T\psi\}$

$$\Rightarrow \begin{cases} \Psi_+(\alpha,\psi) & = & \max\limits_{x\in X}\left[\psi^TA(x)-g^Tx\right]+\left[K\frac{\psi^T\psi}{2\alpha}+\alpha\ln(2/\epsilon)\right] \\ \Psi_-(\alpha,\psi) & = & \max\limits_{x\in X}\left[g^Ty-\psi^TA(y)\right]+\left[K\frac{\psi^T\psi}{2\alpha}+\alpha\ln(2/\epsilon)\right] \end{cases}$$

$\Rightarrow$ *The optimization problem* $\min\limits_{\alpha>0,\psi}\frac{1}{2}\left[\Psi_+(\alpha,\psi)+\Psi_-(\alpha,\psi)\right]$ *responsible for good estimates admits analytical elimination of* α *and results in the optimization problem*

$$\min_{\psi}\left\{\frac{1}{2}\max_{x\in X}\left[\psi^TA(x)-g^Tx\right]+\frac{1}{2}\max_{y\in X}\left[g^Ty-\psi^TA(y)\right]+\sqrt{2K\ln(2/\epsilon)}\|\phi\|_2\right\}$$

in ψ*-variable only.*

Numerical Illustration

Covering story: At the North-bound part of a highway leaving Atlanta there at $n+1$ crossings where cars traveling North enter/exit the highway.

- Arrivals of cars traveling North and entering the highway at crossing # j, $j=0,1,...,n-1$, form Poisson process with (unknown) parameter $x_j \leq 1$; the arrival processes are mutually independent
- A car on a highway traveling North and approaching a crossing exits the highway at this crossing with given probability p
- *For $i=1,...,n$, we observe the total number ω_i of cars traveling North and exiting the highway at crossing # i on time horizon $[0,T]$ and want to recover x_j for a particular value of j.*

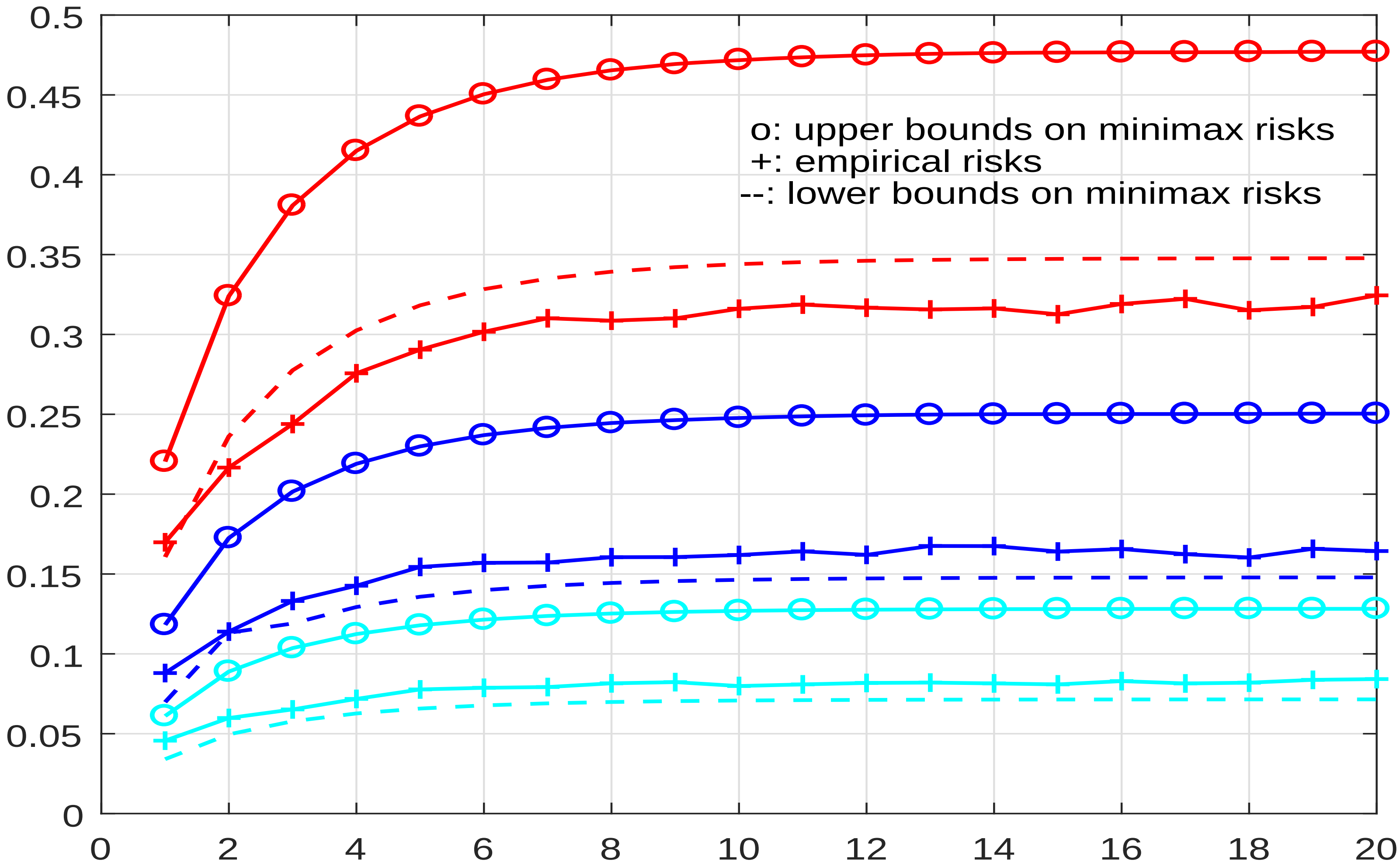
Model: Observation $\omega = [\omega_1; ...; \omega_n]$ is collection of independent of each other Poisson random variables; the vector of their Poisson parameters is TAx, with

$$A = \left[\begin{array}{c|c|c|c|c} p & & & & \\ \hline p(1-p) & p & & & \\ \hline p(1-p)^2 & p(1-p) & p & & \\ \hline \vdots & \vdots & \vdots & \ddots & \\ \hline p(1-p)^{n-1} & p(1-p)^{n-2} & p(1-p)^{n-3} & ... & p \end{array}\right]$$

$\Rightarrow$ *Our problem is to recover linear form of signal*

$$x \in X = \{x \in \mathbb{R}^n : 0 \leq x_j \leq 1, 0 \leq j < n\}$$

observed via Poisson o.s.



• $n = 20$ • $p = 1/3$ • $\epsilon = 0.01$

red: $T = 500$ blue: $T = 2000$ cyan: $T = 8000$

Risks of recovering x_j vs. j

Note: empirical risks are at most by 5% worse than lower bounds on minimax optimal 0.01-risks

Extension: Recovering Linear Form on Union of Convex Sets

♣ **Situation:** Given are:

- Simple o.s. $\mathcal{O} = ((\Omega, \Pi), \{p_\mu : \mu \in \mathcal{M}\}, \mathcal{F})$
- Convex compact sets $X_i \subset \mathbb{R}^n$, $i \leq I$, and *affine* mappings $x \mapsto A_i(x) : X_i \to \mathcal{M}$
- Linear function $g^T x$ on $\mathbb{R}^n$

Given stationary K-repeated observation $\omega^K = (\omega_1, ..., \omega_K)$, with

$$\omega_k \sim p_{A_i(x)},\ 1 \leq k \leq K,$$

stemming from *unknown* signal x known to belong to X_i with some *unknown* $i \leq I$, we want to recover $g^T x$.

♠ **Construction:**

A. Given "reliability tolerance" $0 < \epsilon < 1$, for $1 \le i, j \le I$, let

$$\begin{array}{rcl}\Phi_{ij}(\alpha,\phi;x,y) & = & \frac{1}{2}K\alpha\left[\Phi_{\mathcal{O}}(\phi/\alpha;A_i(x))+\Phi_{\mathcal{O}}(-\phi/\alpha;A_j(y))\right]+\frac{1}{2}g^T[y-x]+\alpha\ln(2I/\epsilon):\\ & & \{\alpha>0,\phi\in\mathcal{F}\}\times[X_i\times X_j]\to\mathbb{R},\\ \Psi_{ij}(\alpha,\phi) & = & \max\limits_{x\in X_i,y\in X_j}\Phi_{ij}(\alpha,\phi;x,y)=\frac{1}{2}\left[\Psi_{i,+}(\alpha,\phi)+\Psi_{j,-}(\alpha,\phi)\right]:\{\alpha>0\}\times\mathcal{F}\to\mathbb{R},\end{array}$$

where

$$\begin{array}{rcl}\Psi_{\ell,+}(\alpha,\psi) & = & \max\limits_{x\in X_\ell}\left[K\alpha\Phi_{\mathcal{O}}(\psi/\alpha;A_\ell(x))-g^Tx+\alpha\ln(2I/\epsilon)\right]:\{\alpha>0,\psi\in\mathcal{F}\}\to\mathbb{R},\\ \Psi_{\ell,-}(\alpha,\psi) & = & \max\limits_{x\in X_\ell}\left[K\alpha\Phi_{\mathcal{O}}(-\psi/\alpha;A_\ell(x))+g^Tx+\alpha\ln(2I/\epsilon)\right]:\{\alpha>0,\psi\in\mathcal{F}\}\to\mathbb{R}\end{array}$$

and $\Phi_{\mathcal{O}}(\phi;\mu)=\ln\left(\int_\Omega e^{\phi(\omega)}p_\mu(\omega)\Pi(d\omega)\right)$

Comment: It is easy to verify that *whenever* $\alpha_{ij}>0,\phi_{ij}\in\mathcal{F}$, *setting*

$$\begin{array}{rcl}\rho_{ij} & = & \Psi_{ij}(\alpha_{ij},\phi_{ij})=\frac{1}{2}\left[\Psi_{i,+}(\alpha_{ij},\phi_{ij})+\Psi_{j,-}(\alpha_{ij},\phi_{ij})\right]\\ \varkappa_{ij} & = & \frac{1}{2}\left[\Psi_{j,-}(\alpha_{ij},\phi_{ij})-\Psi_{i,+}(\alpha_{ij},\phi_{ij})\right]\\ g_{ij}(\omega^K) & = & \sum_{k=1}^K\phi_{ij}(\omega_k)+\varkappa_{ij}\end{array}$$

we ensure that

$$\begin{array}{l}x\in X_i,\omega^K\sim p^K_{A_i(x)}\Rightarrow \mathrm{Prob}\{g_{ij}(\omega^K)>g^Tx+\rho_{ij}\}\le\frac{\epsilon}{2I}\\ y\in X_j,\omega^K\sim p^K_{A_j(x)}\Rightarrow \mathrm{Prob}\{g_{ij}(\omega^K)<g^Ty-\rho_{ij}\}\le\frac{\epsilon}{2I}\end{array}$$

B. For $1 \leq i,j \leq I$, we find feasible near-optimal solutions α_{ij},ϕ_{ij} to (convex by their origin) optimization problems

$$\mathrm{Opt}_{ij} = \min_{\alpha>0,\phi\in\mathcal{F}} \Psi_{ij}(\alpha,\phi),$$

and set

$$\begin{array}{c}\rho_{ij} = \Psi_{ij}(\alpha_{ij},\phi_{ij}),\ \varkappa_{ij} = \frac{1}{2}\left[\Psi_{j,-}(\alpha_{ij},\phi_{ij}) - \Psi_{i,+}(\alpha_{ij},\phi_{ij})\right]\\ g_{ij}(\omega^K) = \sum_{k=1}^K \phi_{ij}(\omega_k) + \varkappa_{ij}\end{array}$$

Given observation ω^K, we set

$$G = [g_{ij}(\omega^K)]_{\substack{i\leq I,\\ j\leq I}},\ r_i = \max_j g_{ij}(\omega^K),\ c_j = \min_i g_{ij}(\omega^K)$$

and take the quantity

$$\widehat{g}(\omega^K) = \frac{1}{2}\left[\min_i \rho_i + \max_j c_j\right]$$

as the estimate of $g^T x$.

♠ **Proposition:** *ϵ-risk of the estimate $\widehat{g}$ does not exceed $\rho = \max_{i,j}\rho_{ij}$, i.e., whenever $\ell \leq I$ and $x \in X_\ell$, the $p^K_{A_\ell(x)}$-probability of the event $|g^T x - \widehat{g}(\omega^K)| > \rho$ is $\leq \epsilon$. Note that ρ can be made arbitrarily close to $\mathrm{Opt}(K) = \max_{i,j}\mathrm{Opt}_{ij}$.*

Sketch of the proof: Let $\omega^K \sim p^K_{A_\ell(x)}$. From comment to **A** it follows that the $p^K_{A_\ell(x)}$-probability of the event

$$\forall i,j: g_{\ell j} \leq g^T x + \rho_{\ell j} \;\&\; g_{i\ell} \leq g^T x - \rho_{i\ell} \qquad [g_{ij} = g_{ij}(\omega^K)]$$

is at least $1-\epsilon$.

When this event takes place, we have

- all entries in ℓ-th row of $G = [g_{ij}]$ by magenta inequalities are $\leq g^T x + \rho$,
- all entries in ℓ-to column of G, by red inequalities, are $\geq g^T x - \rho$
- $r_i = \max\limits_j g_{ij}$, $c_j = \min\limits_i g_{ij}$ (by definition of r_i and c_j)

$\Rightarrow f^T x - \rho \leq \min\limits_i g_{i\ell} \leq \min\limits_i r_i \leq r_\ell \leq g^T x + \rho \Rightarrow f^T x \in [\min\limits_i r_i - \rho, \min\limits_i r_i + \rho]$

and similarly $f^T x \in [\max\limits_j c_j - \rho, \max\limits_j c_j + \rho]$ $\square$

Near-Optimality: *Let* $\epsilon \in (0, 1/2)$ *and* K_* *be a positive integer, and let* $\mathrm{Risk}^*_\epsilon(K_*)$ *be the minimax optimal* ϵ*-risk, the number of observations being* K_* (that is, the infimum, over all Borel K_*-observation estimates, of ϵ-risks of the estimates) *Then for every integer* K *satisfying*

$$K > \frac{2\ln(2I/\epsilon)}{\ln([4\epsilon(1-\epsilon)]^{-1})} K_*$$

one has

$$\mathrm{Opt}(K) \leq \mathrm{Risk}^*_\epsilon(K_*).$$

In addition, assuming that every i, j *there exists* $\bar{x}_{ij} \in X_i \cap X_j$ *such that* $A_i(\bar{x}_{ij}) = A_j(\bar{x}_{ij})$ *one has*

$$K \geq K_* \Rightarrow \mathrm{Opt}(K) \leq \frac{2\ln(2I/\epsilon)}{\ln([4\epsilon(1-\epsilon)]^{-1})} \mathrm{Risk}^*_\epsilon(K_*).$$

Sketch of the proof [first claim only]: Since $\mathrm{Opt}(K)=\max\limits_{i,j}\mathrm{Opt}_{ij}(K)$, all we need to verify is that when

$$K>\frac{2\ln(2I/\epsilon)}{\ln([4\epsilon(1-\epsilon)]^{-1})}K_* \qquad (*)$$

we have $\mathrm{Opt}_{ij}(K)\leq \mathrm{Risk}^*_\epsilon(K_*)$ for every i,j.

- Recall that $\mathrm{Opt}_{ij}(K)=\inf\limits_{\alpha>0,\phi\in\mathcal{F}}\left[\Psi_{ij}(\alpha,\phi):=\max\limits_{x\in X_i,y\in X_j}\Phi_{ij}(\alpha,\phi;x,y)\right]$ and by its origin, Φ_{ij} is convex in α,ϕ and concave in x,y, whence

$$\begin{array}{rcl}\mathrm{Opt}_{ij}(K) & = & \max\limits_{x\in X_i,y\in X_j}\inf\limits_{\alpha>0,\phi\in\mathcal{F}}\Phi_{ij}(\alpha,\phi;x,y)\\ & \underbrace{=}_{(!)} & \max\limits_{x,y}\left\{\frac{1}{2}g^T[y-x]: x\in X_i, y\in X_j, \left[\int\sqrt{p_{A_i(x)}(\omega)p_{A_j(y)}(\omega)}\Pi(d\omega)\right]^K\geq\frac{\epsilon}{2I}\right\}.\end{array}$$ with

(!) given by straightforward computation,

Assuming, on the contrary to what should be proved, that $\mathrm{Opt}_{ij}(K)>\mathrm{Risk}^*_\epsilon(K_*)$, we can find $\bar{x}\in X_i$, $\bar{y}\in X_j$ such that with $\mu=A_i(\bar{x})$, $\nu=A_j(\bar{x})$ it holds

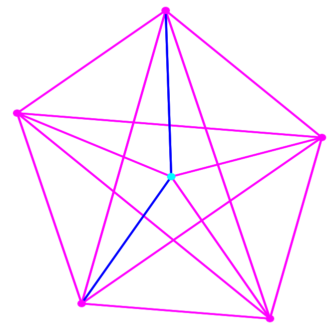
$$\frac{1}{2}g^T[\bar{y}-\bar{x}]>\mathrm{Risk}^*_\epsilon(K_*)\ \&\ \left[\int\sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right]^K\geq\frac{\epsilon}{2I} \qquad (!)$$

By first relation in (!), two simple hypotheses stating that the distributions of ω^{K_*} is $p_\mu^{K_*}$, resp., $p_\nu^{K_*}$ can be decided upon with risk $\leq\epsilon$, whence by elementary results about Hellinger affinity,

$$2\sqrt{\epsilon(1-\epsilon)}\geq\int\sqrt{p_\mu^{K_*}(\omega^K)p_\nu^{K_*}(\omega^K)}\Pi^K(d\omega^K)=\left[\int\sqrt{p_\mu(\omega)p_\nu(\omega)}\Pi(d\omega)\right]^{K_*}.$$

This combines with $(*)$ to imply the inequality *opposite* to (!), which is a desired contradiction. $\square$

Toy Illustration: Recovering Origin-Destination Traffics

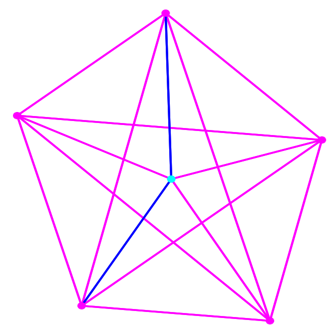


♠ **Covering story:** Nodes in the network represent five villages (magenta dots) and crossing with no population (cyan dot), and arcs represent road segments.

- There are two states of the road net:
  — normal: some normal traveling times in all segments,
  — abnormal: normal traveling times in magenta segments and much larger than normal traveling times in blue segments.
- There are $L = 7$ origin-destination pairs, ℓ-th with its own traffic x_ℓ. The travelers know normal and abnormal traveling times of the arcs and the state of the network and select the fastest routs between their origins and destinations. As a result, the total traffic in arc γ is $\sum_\ell A^\chi_{\gamma\ell} x_\ell$ where $\chi \in \{\mathrm{normal}, \mathrm{abnormal}\}$ is the state of the network.
- We do *not* know network's state and traffics in origin-destination pairs. All we know are

— the number L of origin-destination pairs and an upper bound T on the total traffic $\sum_\ell x_\ell$
— the sensing matrices $A^\chi = [A^\chi_{\gamma\ell}]_{\gamma\in\Gamma,\ell\leq L}$, where $\chi \in \{\mathrm{normal}, \mathrm{abnormal}\}$, and Γ is the set of $M = 29$ arcs where we measure traffic.

- *Given noisy measurements of traffics in the arcs of Γ: $y_\gamma = [A^\chi x]_\gamma + \sigma\xi_\gamma$, with independent across γ noises $\xi_\gamma \sim \mathcal{N}(0,1)$ and known σ, we want to recover origin-destination traffics x_ℓ, $\ell \leq L$.*



♠ **Model:** The unknown signal x lives in $X = \{x \in \mathbb{R}^L_+ : \sum_\ell x_\ell \leq T\}$. We set $X^1 = X^2 = X$ and

$$A_1(x) = A^{\text{normal}}x,\ A_2(x) = A^{\text{abnormal}}x.$$

⇒ *The problem of recovering x_ℓ for a particular ℓ is covered by the Gaussian case of our setup, and we can use the above machinery to recover x_ℓ's one by one.*

	$\|\cdot\|_\infty$ recovery errors			computed upper
σ	mean	median	maximal	bound on 0.01 risk
2^{-3}	0.478	0.480	0.994	0.665
2^{-5}	0.119	0.112	0.224	0.166
2^{-7}	0.030	0.028	0.066	0.042
2^{-9}	0.008	0.007	0.017	0.011
2^{-11}	0.002	0.001	0.005	0.003

Numerical results over 100 simulations

- Pay attention to clear "numerical consistency."

Note: For every σ, our estimate is a “nonlinear aggregation” of 4 estimates which are *affine* in observations. In the reported instance, this estimate is consistent.
In contrast: In the same instance, even in the noiseless case, the worst-case recovery error for *every affine* estimate of x_2 is ≥ 0.25.
Explanation: We are observing in Gaussian noise either Ax, or Bx, with unknown x belonging to the known signal set $X = \{x \in \mathbb{R}^7_+ : \sum_\ell x_\ell \leq T\}$. *We do know A and B,* but do *not* know from which one of the matrices A, B the observation comes. In this situation, the *ultimate obstacle* for high-accuracy recovering $g^T x$ in the low-noise case is

— for our estimate – the fact that $g^T x - g^T y$ is not identically zero on the intersection of $X \times X$ and the linear subspace $\mathcal{L} = \{[x; y] : Ax = By\}$ of pairs (x, y) of "non-distinguishable signals." *In the reported instance, this obstacle is absent – the only common point of $\mathcal{L}$ and $X \times X$ is the origin.*

— for an affine estimate – the fact that the vector $[g; -g]$ is not orthogonal to $\mathcal{L}$. *In the reported instance this obstacle is present – the vector $[e_2; -e_2]$ is far from being orthogonal to $\mathcal{L}$.*

Another Illustration

♣ **SetUp:** Given $J = 100$ points $x_j \in \mathbb{R}^{20}$ and stationary K-repeated observation

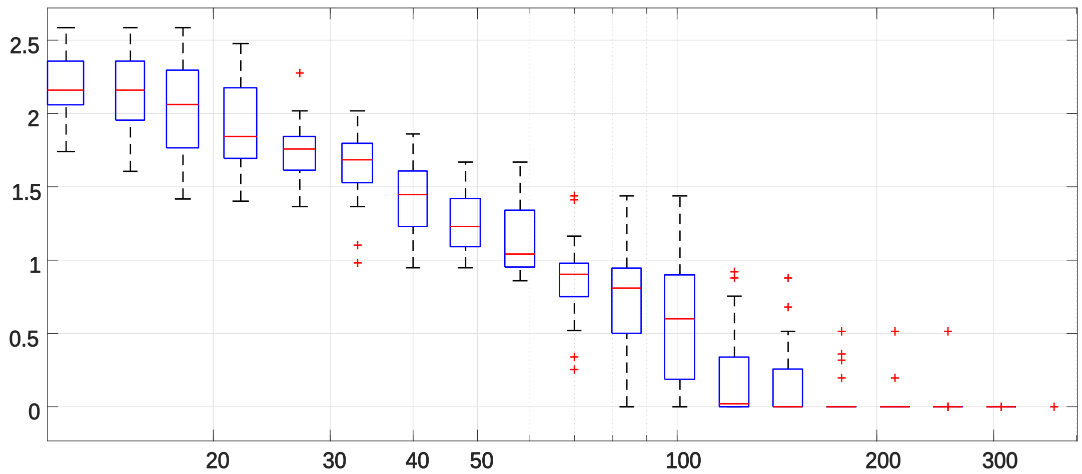
$$\omega^K = (\omega_1, .., \omega_K),\ \omega_k \sim \mathcal{N}(Ax, I_{20})$$

of one of the points (we do not know which one!), we want to recover the first entry of the point.

- A: randomly generated matrix
- $\epsilon = 0.01$.

Note: we are in the situation where $X_i = \{x_i\}$ are singletons.

♠ **Results:**



Recovery error vs. K, data over 20 randomly generated collections $\{x_i\}_{i=1}^{100}$

HYPOTHESIS TESTING, III

- *Beyond simple observation schemes*

♣ **Goal:** *to extend our detector-based hypothesis testing machinery beyond the scope of simple o.s.'s*

♠ **Starting point:** "Executive Summary" of what happened with simple o.s.'s.

0. Basic problem of interest: *Given two families $\mathcal{P}_1$ and $\mathcal{P}_2$ of probability distributions on observation space Ω and an observation $\omega \sim P \in \mathcal{P}_1 \cup \mathcal{P}_2$, we want to decide on the hypothesis $H_1 : P \in \mathcal{P}_1$ vs. the alternative $P \in \mathcal{P}_2$.*

1. Basic tool: A family $\mathcal{F}$ of *candidate detectors* $\phi(\cdot) : \Omega \to \mathbb{R}$. Associated tests $\mathcal{T}_\phi$ were of the form

$$\phi(\omega) \left\{ \begin{array}{lcl} > 0 & \Rightarrow & \text{accept } H_1, \text{ reject } H_2 \\ \leq 0 & \Rightarrow & \text{accept } H_2, \text{ reject } H_1 \end{array} \right. ,$$

and we upper-bounded the risk of $\mathcal{T}_\phi$ by the *risk of detector* ϕ

$$\text{Risk}[\phi|\mathcal{P}_1, \mathcal{P}_2] = \max\left[\sup_{P \in \mathcal{P}_1} \mathbf{E}_{\omega\sim P}\{\exp\{-\phi(\omega)\}\}, \sup_{P\in\mathcal{P}_2} \mathbf{E}_{\omega\sim P}\{\exp\{\phi(\omega)\}\}\right].$$

Basic tool: A family $\mathcal{F}$ of *candidate detectors* $\phi(\cdot) : \Omega \to \mathbb{R}$...

2. *In simple o.s.'s we dealt with the families $\mathcal{F}$ of candidate detectors were in fact comprised of affine functions of ω.*

Indeed, this was the case with Gaussian and Poisson o.s.'s, but seemingly was *not* the case with Discrete o.s. – there $\Omega = \{1, ..., d\}$ and $\mathcal{F}$ was comprised of *whatever* functions of $\omega \in \Omega$.

However: When encoding the points $1, 2, ..., d \in \Omega$ with the standard basic orths $e_1, ..., e_d$ in $\mathbb{R}^d$ — when identifying Ω with the set of vertices of d-dimensional probabilistic simplex — *every function on Ω becomes affine function of $\omega \in \Omega$!*

Note: When the families $\mathcal{F}$ associated with simple o.s.'s in question are comprised of affine functions of $\omega \in \Omega$, so are the families associated with direct products/direct powers of these simple o.s.'s!

3. *The key element of our setup was convex-concave function* $\Phi(h;\mu):\mathbb{R}^d\times\mathcal{M}\to\mathbb{R}$. *Our families* $\mathcal{P}_1=\{P_\mu:\mu\in M_1\}$, $\mathcal{P}_2=\{P_\mu:\mu\in M_2\}$ *of a parametric family of distributions* $\{P_\mu:\mu\in\mathcal{M}\}$ *on* Ω, *and* Φ *was linked to this family by the relation*

$$\ln\left(\mathbf{E}_{\omega\sim P_\mu}\left\{e^{h^T\omega}\right\}\right)=\Phi(h;\mu). \qquad (!)$$

We dealt with the situation when M_1, M_2 were convex compact subsets of $\mathcal{M}$, and (!) allowed us to pose the problem of finding minimum risk affine detector $\phi(\omega)=h^T\omega+\kappa$ as the convex-concave saddle point problem

$$\mathrm{SadVal}=\min_h\max_{\mu\in M_1,\nu\in M_2}\frac{1}{2}\left[\Phi(-h;\mu)+\Phi(h;\nu)\right], \qquad (*)$$

and the risk of affine detector stemming from the h-component of a saddle point was $\exp\{\mathrm{SadVal}\}$.

- An additional reasoning demonstrated that *in the case of simple o.s., this construction yields minimum risk detectors.*

♠ **In the forthcoming extension**, we

• Still stick to detector-based tests and detectors *affine* in ω

• Relax the assumption that $\mathcal{P}_1 = \{P_\nu : \nu \in M_1\}$, $\mathcal{P}_2 = \{P_\nu : \nu \in M_2\}$ for convex compact sets M_1, M_2 and parametric family $\mathcal{P} = \{P_\nu : \nu \in \mathcal{M}\}$ such that

$$\ln\left(\mathbf{E}_{\omega\sim P_\nu}\left\{e^{h^T\omega}\right\}\right) = \Phi(h;\nu). \tag{!}$$

for a known to us convex-concave function $\Phi(h;\nu)$.

Instead, we assume that

• we are given a convex-concave function $\Phi(h;\nu) : \mathbb{R}^d \times \mathcal{M} \to \mathbb{R}$

• $\mathcal{P}_1$ and $\mathcal{P}_2$ are sub-families of a family $\mathcal{P}$ of distributions on $\mathbb{R}^d$, and *every $P \in \mathcal{P}$ can be assigned* (perhaps in many ways!) *a value of parameter $\nu \in \mathcal{M}$ in such a way that*

$$\forall h : \ln\left(\mathbf{E}_{\omega\sim P}\left\{e^{h^T\omega}\right\}\right) \le \Phi(h;\nu). \tag{!!}$$

• $\mathcal{P}_\chi$, $\chi = 1, 2$, can be associated with convex compact sets $\mathcal{M}_\chi$ in such a way that

$$\ln\left(\mathbf{E}_{\omega\sim P}\left\{e^{h^T\omega}\right\}\right) \le \begin{cases} \Phi(h;\nu)\ \forall h \text{ and some } \nu \in M_1, & P \in \mathcal{P}_1 \\ \Phi(h;\nu)\ \forall h \text{ and some } \nu \in M_2, & P \in \mathcal{P}_2 \end{cases}$$

We assume that

• we are given a convex-concave function $\Phi(h;\nu):\mathbb{R}^d\times\mathcal{M}\to\mathbb{R}$

• $\mathcal{P}_1$ and $\mathcal{P}_2$ are sub-families of a family $\mathcal{P}$ of distributions on $\mathbb{R}^d$, and *every $P\in\mathcal{P}$ can be assigned* (perhaps in many ways!) *a value of parameter $\nu\in\mathcal{M}$ in such a way that*

$$\forall h:\ \ln\left(\mathbf{E}_{\omega\sim P}\left\{e^{h^T\omega}\right\}\right)\leq\Phi(h;\nu). \tag{!!}$$

• $\mathcal{P}_\chi$, $\chi=1,2$, can be associated with convex compact sets $\mathcal{M}_\chi$ in such a way that

$$\ln\left(\mathbf{E}_{\omega\sim P}\left\{e^{h^T\omega}\right\}\right)\leq\left\{\begin{array}{ll}\Phi(h;\nu)\ \forall h \text{ and some } \nu\in M_1, & P\in\mathcal{P}_1\\ \Phi(h;\nu)\ \forall h \text{ and some } \nu\in M_2, & P\in\mathcal{P}_2\end{array}\right.$$

♠ With this extension, the convex-concave saddle point problem

$$\mathrm{SadVal}=\min_h\max_{\mu\in M_1,\nu\in M_2}\frac{1}{2}\left[\Phi(-h;\mu)+\Phi(h;\nu)\right], \tag{$*$}$$

still supplies "presumably good" affine detector with risk $\leq\exp\{\mathrm{SadVal}\}$.

Bad news: the resulting tests not necessarily are near-optimal

Good news: Our new setup covers situations going far beyond simple o.s.'s, e.g., the case of *sub-Gaussian* distributions, where the "parameter" $\mu=(u,\Theta)\in\mathbb{R}^d\times\mathbf{S}^d_+$ of a distribution P satisfies

$$\ln\left(\mathbf{E}_{\omega\sim P}\{e^{h^T\omega}\}\right)\leq h^Tu+\frac{1}{2}h^T\Theta h\ \forall h.$$

Setup

♣ Given an observation space $\Omega = \mathbb{R}^d$, consider a triple $\mathcal{H}, \mathcal{M}, \Phi$, where

- $\mathcal{H}$ is a nonempty closed convex set in Ω symmetric w.r.t. the origin,
- $\mathcal{M}$ is a compact convex set in some $\mathbb{R}^n$,
- $\Phi(h;\mu) : \mathcal{H} \times \mathcal{M} \to \mathbb{R}$ is a continuous function *convex in* $h \in \mathcal{H}$ and *concave in* $\mu \in \mathcal{M}$.

♣ $\mathcal{H}, \mathcal{M}, \Phi$ specify a family $\mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ of probability distributions on Ω. A probability distribution P belongs to the family iff there exists $\mu \in \mathcal{M}$ such that

$$\ln\left(\int_\Omega e^{h^T\omega} P(d\omega)\right) \leq \Phi(h;\mu) \ \forall h \in \mathcal{H} \qquad (*)$$

We refer to μ ensuring $(*)$ as to *parameter* of distribution P.

- **Warning:** A distribution P may have many different parameters!

♡ We refer to triple $\mathcal{H}, \mathcal{M}, \Phi$ satisfying the above requirements as to *regular data*, and to $\mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ – as to the *simple family of distributions* induced by these data.

♠ **Example 1: Gaussian and sub-Gaussian distributions.** When

• $\mathcal{M} = \{(u, \Theta)\} \subset \mathbb{R}^d \times \mathrm{int}\,\mathbf{S}^d_+$ is a convex compact set such that $\Theta \succ 0$ for all $(u, \Theta) \in \mathcal{M}$,

• $\mathcal{H} = \mathbb{R}^d$,

• $\Phi(h; u, \Theta) = h^T u + \frac{1}{2} h^T \Theta h$,

$\mathcal{S} = \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ contains all probability distributions P which are *sub-Gaussian with parameters* (u, Θ), meaning that

$$\ln\left(\int_\Omega e^{h^T \omega} P(d\omega)\right) \leq h^T u + \frac{1}{2} h^T \Theta h \;\; \forall h, \qquad (1)$$

and, in addition, the "parameter" (u, Θ) belongs to $\mathcal{M}$.

Note: Whenever P is sub-Gaussian with parameters (u, Θ), u is the expectation of P.

Note: $\mathcal{N}(u, \Theta) \in \mathcal{S}$ whenever $(u, \Theta) \in \mathcal{M}$; for $P = \mathcal{N}(u, \Theta)$, (1) is an identity.

♠ **Example 2: Poisson distributions.** When

- $\mathcal{M} \subset \mathbb{R}^d_+$ is a convex compact set,
- $\mathcal{H} = \mathbb{R}^d$,
- $\Phi(h; \mu) = \sum_{i=1}^d \mu_i(\mathrm{e}^{h_i} - 1)$,

$\mathcal{S} = \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ contains distributions of all d-dimensional random vectors ω_i with independent across i entries $\omega_i \sim \mathrm{Poisson}(\mu_i)$ such that $\mu = [\mu_1; ...; \mu_d] \in \mathcal{M}$.

♠ **Example 3: Discrete distributions.** When

- $\mathcal{M} = \{\mu \in \mathbb{R}^d : \mu \geq 0, \sum_j \mu_j = 1\}$ is the probabilistic simplex in $\mathbb{R}^d$,
- $\mathcal{H} = \mathbb{R}^d$,
- $\Phi(h; \mu) = \ln\left(\sum_{i=1}^d \mu_i \mathrm{e}^{h_i}\right)$,

$\mathcal{S} = \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ contains all discrete distributions supported on the vertices of the probabilistic simplex.

♠ **Example 4: Distributions with bounded support.** Let $X \subset \mathbb{R}^d$ be a nonempty convex compact set with support function $\phi_X(\cdot)$:

$$\phi_X(y) = \max_{x \in X} y^T x : \mathbb{R}^d \to \mathbb{R}^d.$$

When $\mathcal{M} = X$, $\mathcal{H} = \mathbb{R}^d$ and

$$\Phi(h;\mu) = h^T \mu + \frac{1}{8}[\phi_X(h) + \phi_X(-h)]^2, \tag{2}$$

$\mathcal{S} = \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$ contains all probability distributions supported on X, and for such a distribution P, $\mu = \int_X \omega P(d\omega)$ is a parameter of P.

• **Note:** When G, $0 \in G$, is a convex compact set, the conclusion in Example 4 remains valid when function (2) is replaced with the smaller function

$$\Phi(h;\mu) = \min_{g \in G} \left[\mu^T (h-g) + \frac{1}{8}[\phi_X(h-g) + \phi_X(g-h)]^2 + \phi_X(g) \right].$$

♣ **Fact:** *Simple families of probability distributions admit "calculus:"*

♠ [summation] For $1 \leq \ell \leq L$, let λ_ℓ be reals, and let $\mathcal{H}_\ell, \mathcal{M}_\ell, \Phi_\ell$ be regular data with common observation space: $\mathcal{H}_\ell \subset \Omega = \mathbb{R}^d$. Setting

$$\mathcal{H} = \{h \in \mathbb{R}^d : \lambda_\ell h \in \mathcal{H}_\ell, 1 \leq \ell \leq L\}, \mathcal{M} = \mathcal{M}_1 \times ... \times \mathcal{M}_L,$$
$$\Phi(h; \mu_1, ..., \mu_L) = \sum_{\ell=1}^L \Phi_\ell(\lambda_\ell h; \mu_\ell),$$

we get regular data with the following property:

Whenever random vectors $\xi_\ell \sim P_\ell \in \mathcal{S}[\mathcal{H}_\ell, \mathcal{M}_\ell, \Phi_\ell]$, $1 \leq \ell \leq L$, are independent across ℓ, the distribution P of the random vector $\xi = \sum_{\ell=1}^L \lambda_\ell \xi_\ell$ belongs to $\mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$. Denoting by μ_ℓ parameters of P_ℓ, $\mu = [\mu_1; ...; \mu_L]$ can be taken as parameter of P.

♠ [direct product] For $1 \leq \ell \leq L$, let $\mathcal{H}_\ell, \mathcal{M}_\ell, \Phi_\ell$ be regular data with observation spaces $\Omega_\ell = \mathbb{R}^{d_\ell}$. Setting

$$\mathcal{H} = \mathcal{H}_1 \times ... \times \mathcal{H}_L \subset \Omega = \mathbb{R}^{d_1+...+d_L}. \mathcal{M} = \mathcal{M}_1 \times ... \times \mathcal{M}_L,$$
$$\Phi(h_1, ..., h_L; \mu_1, ..., \mu_L) = \sum_{\ell=1}^L \Phi_\ell(h_\ell; \mu_\ell),$$

we get regular data with the following property:

Whenever $P_\ell \in \mathcal{S}[\mathcal{H}_\ell, \mathcal{M}_\ell, \Phi_\ell]$, $1 \leq \ell \leq L$, the direct product distribution $P = P_1 \times ... \times P_L$ belongs to $\mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$. Denoting by μ_ℓ parameters of P_ℓ, $\mu = [\mu_1; ...; \mu_L]$ can be taken as parameter of P.

♠ [marginal distribution] Let $\mathcal{H}, \mathcal{M}, \Phi$ be regular data with observation space $\mathbb{R}^d$, and let $\omega \mapsto A\omega + a : \mathbb{R}^d \mapsto \Omega = \mathbb{R}^\delta$. Setting

$$\bar{\mathcal{H}} = \{h \in \mathbb{R}^\delta : A^T h \in \mathcal{H}\}, \ \bar{\Phi}(h;\mu) = h^T a + \Phi(A^T h;\mu),$$

we get regular data $\bar{\mathcal{H}}, \mathcal{M}, \bar{\Phi}$ with the following property:

Whenever $\xi \sim P \in \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi]$, the distribution $\bar{P}$ of the random variable $\omega = A\xi + a$ belongs to the simple family $\mathcal{S}[\bar{\mathcal{H}}, \mathcal{M}, \bar{\Phi}]$, and parameter of P is a parameter of $\bar{P}$ as well.

♣ **Main observation:** *When deciding on simple families of distributions, affine tests and their risks can be efficiently computed via Convex Programming:*

♡ **Theorem.** *Let $\mathcal{H}_\chi, \mathcal{M}_\chi, \Phi_\chi$, $\chi = 1, 2$, be two collections of regular data with compact $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{H}_1 = \mathcal{H}_2 =: \mathcal{H}$, and let*

$$\Psi(h) = \max_{\mu_1 \in \mathcal{M}_1, \mu_2 \in \mathcal{M}_2} \underbrace{\frac{1}{2}\left[\Phi_1(-h;\mu_1) + \Phi_2(h,\mu_2)\right]}_{\Phi(h;\mu_1,\mu_2)} : \mathcal{H} \to \mathbb{R}$$

Then Ψ is efficiently computable convex function, and for every $h \in \mathcal{H}$, setting

$$\phi(\omega) = h^T\omega + \underbrace{\frac{1}{2}\left[\max_{\mu_1 \in \mathcal{M}_1} \Phi_1(-h;\mu_1) - \max_{\mu_2 \in \mathcal{M}_2} \Phi_2(h;\mu_2)\right]}_{\varkappa},$$

one has

$$\mathrm{Risk}[\phi|\mathcal{P}_1, \mathcal{P}_2] \leq \exp\{\Psi(h)\} \qquad [\mathcal{P}_\chi = \mathcal{S}[\mathcal{H}, \mathcal{M}_\chi, \Phi_\chi]]$$

In particular, if convex-concave function $\Phi(h;\mu_1,\mu_2)$ possesses a saddle point $h_*, (\mu_1^*, \mu_2^*)$ on $\mathcal{H} \times (\mathcal{M}_1 \times \mathcal{M}_2)$, the affine detector

$$\phi_*(\omega) = h_*^T\omega + \tfrac{1}{2}\left[\Phi_1(-h;\mu_1^*) - \Phi_2(h^*;\mu_2^*)\right]$$

admits risk bound

$$\mathrm{Risk}[\phi_*|\mathcal{P}_1, \mathcal{P}_2] \leq \exp\{\Phi(h^*;\mu_1^*,\mu_2^*)\}$$

Indeed, let $h \in \mathcal{H}$. Selecting $\mu_1^* \in \mathop{\mathrm{Argmax}}\limits_{\mu_1 \in \mathcal{M}_1} \Phi_1(-h; \mu_1)$, $\mu_2^* \in \mathop{\mathrm{Argmax}}\limits_{\mu_2 \in \mathcal{M}_2} \Phi_2(h; \mu_2)$, we have

$$P \in \mathcal{P}_1 := \mathcal{S}[\mathcal{H}, \mathcal{M}_1, \Phi_1] \Rightarrow \exists \mu_1 \in \mathcal{M}_1 : \mathbf{E}_{\omega \sim P}\left\{ \mathrm{e}^{-h^T\omega} \right\} \leq \mathrm{e}^{\Phi_1(-h;\mu_1)}$$

$$\Rightarrow \mathbf{E}_{\omega \sim P}\left\{ \mathrm{e}^{-\phi(\omega)} \right\} \leq \mathrm{e}^{\Phi_1(-h;\mu_1^*)-\kappa} = \mathrm{e}^{\Psi(h)} \Rightarrow \mathrm{Risk}_1[\phi|\mathcal{P}_1, \mathcal{P}_2] \leq \mathrm{e}^{\Psi(h)}.$$

Similarly,

$$P \in \mathcal{P}_2 := \mathcal{S}[\mathcal{H}, \mathcal{M}_2, \Phi_2] \Rightarrow \exists \mu_2 \in \mathcal{M}_2 : \mathbf{E}_{\omega \sim P}\left\{ \mathrm{e}^{h^T\omega} \right\} \leq \mathrm{e}^{\Phi_2(h;\mu_2)}$$

$$\Rightarrow \mathbf{E}_{\omega \sim P}\left\{ \mathrm{e}^{\phi(\omega)} \right\} \leq \mathrm{e}^{\Phi_2(h;\mu_2^*)+\kappa} = \mathrm{e}^{\Psi(h)} \Rightarrow \mathrm{Risk}_2[\phi|\mathcal{P}_1, \mathcal{P}_2] \leq \mathrm{e}^{\Psi(h)}.$$

♠ **Numerical Illustration.** Given observation

$$\omega = Ax + \sigma A \text{Diag}\left\{\sqrt{x_1}, ..., \sqrt{x_n}\right\} \xi \qquad [\xi \sim \mathcal{N}(0, I_n)]$$

of an unknown signal x known to belong to a given convex compact set $M \subset \mathbb{R}^n_{++}$, we want to decide on two hypotheses $H_\chi : x \in X_\chi$, $\chi = 1, 2$, with risk 0.01.

X_χ: convex compact subsets of X.

Novelty: *Noise intensity depends on the signal!*

• Introducing regular data $\mathcal{H}_\chi = \mathbb{R}^n$, $\mathcal{M}_\chi = X_\chi$,

$$\Phi_\chi(h, \mu) = h^T A\mu + \frac{\sigma^2}{2} h^T [A\text{Diag}\{\mu\}A^T]h \qquad [\chi = 1, 2]$$

distribution of observations under H_χ belongs to $\mathcal{S}[\mathcal{H}, \mathcal{M}_\chi, \Phi_\chi]$.

• An affine detector for families $\mathcal{P}_\chi$ of distributions obeying H_χ, $\chi = 1, 2$, is given by the saddle point of the function

$$\Phi(h; \mu_1, \mu_2) := \frac{1}{2}\left[h^T[\mu_2 - \mu_1] + \frac{\sigma^2}{2} h^T A \mathrm{Diag}\{\mu_1 + \mu_2\} A^T h\right]$$

♡ **Data:** $n = 16$, $\sigma = 0.1$, target risk 0.01,

- $A = U\mathrm{Diag}\{0.01^{(i-1)/15}, i \leq 16\}V$ with random orthogonal U, V,
- $X_1 = \left\{x \in \mathbb{R}^{16} : \begin{array}{l} 0.001 \leq x_1 \leq \delta \\ 0.001 \leq x_i \leq 1,\ i \geq 2 \end{array}\right\}$
- $X_2 = \left\{x \in \mathbb{R}^{16} : \begin{array}{l} 2\delta \leq x_1 \leq 1 \\ 0.001 \leq x_i \leq 1,\ i \geq 2 \end{array}\right\}$

♡ **Results:**

$\delta = 0.1 \Rightarrow \mathrm{Risk}[\phi_*|\mathcal{P}_1, \mathcal{P}_2] = 0.4346 \Rightarrow$ 6-repeated observation

$\delta = 0.01 \Rightarrow \mathrm{Risk}[\phi_*|\mathcal{P}_1, \mathcal{P}_2] = 0.9201 \Rightarrow$ 56-repeated observation

• Safe "Gaussian o.s. approximation" of the above observation scheme requires 37-repeated observations to handle $\delta = 0.1$ and 3685-repeated observation to handle $\delta = 0.01$).

♣ **Sub-Gaussian case.** For $\chi = 1, 2$, let $U_\chi \subset \Omega = \mathbb{R}^d$ and $\mathcal{V}_\chi \subset \operatorname{int} \mathbf{S}^d_+$ be convex compact sets. Setting

$$\mathcal{M}_\chi = U_\chi \times \mathcal{V}_\chi,\ \ \Phi(h; u, \Theta) = h^T u + \frac{1}{2} h^T \Theta h : \mathcal{H} \times \mathcal{M}_\chi \to \mathbb{R},$$

the regular data $\mathcal{H} = \mathbb{R}^d, \mathcal{M}_\chi, \Phi$ specify the families

$$\mathcal{P}_\chi = \mathcal{S}[\mathbb{R}^d, U_\chi \times \mathcal{V}_\chi, \Phi]$$

of sub-Gaussian distributions with parameters from $U_\chi \times \mathcal{V}_\chi$.

♠ Saddle point problem responsible for design of affine detector for $\mathcal{P}_1, \mathcal{P}_2$ reads

$$\mathrm{SadVal} = \min_{h \in \mathbb{R}^d} \max_{\substack{u_1 \in U_1, u_2 \in U_2 \\ \Theta_1 \in \mathcal{V}_1, \Theta_2 \in \mathcal{V}_2}} \frac{1}{2} \left[h^T (u_2 - u_1) + \frac{1}{2} h^T [\Theta_1 + \Theta_2] h \right]$$

• Saddle point $(h_*; (u_1^*, u_2^*, \Theta_1^*, \Theta_2^*))$ does exist and satisfies

$$\begin{array}{c} h_* = [\Theta_1^* + \Theta_2^*]^{-1} [u_1^* - u_2^*], \\ \mathrm{SadVal} = -\frac{1}{4} [u_1^* - u_2^*][\Theta_1^* + \Theta_2^*]^{-1} [u_1^* - u_2^*] = -\frac{1}{4} h_*^T [u_1^* - u_2^*] \end{array}$$

• The associated affine detector and its risk are

$$\begin{array}{l} \phi_*(\omega) = h_*^T \left[\omega - \frac{1}{2} [u_1^* + u_2^*] \right] = [u_1^* - u_2^*]^T [\Theta_1^* + \Theta_2^*]^{-1} \left[\omega - \frac{1}{2} [u_1^* + u_2^*] \right] \\ \mathrm{Risk}(\phi_* | \mathcal{P}_1, \mathcal{P}_2) \\ \qquad \le \exp\{\mathrm{SadVal}\} = \exp\{-\frac{1}{4} [u_1^* - u_2^*][\Theta_1^* + \Theta_2^*]^{-1} [u_1^* - u_2^*]\} \end{array}$$

$\heartsuit$ **Note:** In the *symmetric case* $\mathcal{V}_1 = \mathcal{V}_2$ $(h_*; (u_1^*, u_2^*, \Theta_1^*, \Theta_2^*))$ can be selected to have $\Theta_1^* = \Theta_2^* =: \Theta_*$. In this case, *the affine detector we end up with is the minimum risk detector for* $\mathcal{P}_1, \mathcal{P}_2$.

What is "affine?" Quadratic Lifting

♣ We have developed a technique for building "presumably good" *affine* detectors for simple families of distributions.

But: Given observation $\zeta \sim P$, we can subject it to *nonlinear* transformation $\zeta \mapsto \omega = \psi(\zeta)$, e.g., to *quadratic lifting*

$$\zeta \mapsto \omega = (\zeta, \zeta\zeta^T)$$

and treat as our observation ω rather than the "true" observation ζ.

Note: *Affine in ω detectors are nonlinear in ζ.*

Example: Detectors affine in the quadratic lifting $\omega = (\zeta, \zeta\zeta^T)$ of ζ are exactly the *quadratic* functions of ζ.

♠ We can try to apply our machinery for building affine detectors to nonlinear transformations of true observations, thus arriving at nonlinear detectors.

- **Bottleneck:** To apply the outlined strategy to a pair $\mathcal{P}_1, \mathcal{P}_2$ of families of distributions of interest, we need to cover the families $\mathcal{P}_\chi^+$ of distributions of $\omega = \psi(\zeta)$ induced by distributions $P \in \mathcal{P}_\chi$ of ζ, $\chi = 1, 2$, by simple families of distributions.
- **What is ahead:** Simple "coverings" of quadratic lifts of (sub)Gaussian distributions.

♣ **Situation:** Given are:

- a compact nonempty set $U \subset \mathbb{R}^n$
- an affine mapping $u \mapsto \mathcal{A}(u) = A[u; 1] : \mathbb{R}^n \to \mathbb{R}^d$
- a convex compact set $\mathcal{V} \subset \mathrm{int}\, \mathbf{S}^d_+$.

• The above data specify families of probability distributions of random observations

$$\omega = (\zeta, \zeta\zeta^T),\ \zeta = \mathcal{A}(u) + \xi \in \mathbb{R}^d, \qquad (*)$$

specifically,

— the family $\mathcal{G}$ of all distributions of ω induced by deterministic $u \in U$ and *Gaussian* noise $\xi \sim \mathcal{N}(0, \Theta \in \mathcal{V})$

— the family $\mathcal{SG}$ of all distributions of ω induced by deterministic $u \in U$ and *sub-Gaussian*, with parameters $(0, \Theta \in \mathcal{V})$ noise ξ

♡ **Goal:** To cover $\mathcal{G}$ ($\mathcal{SG}$) by a simple family of distributions.

Gaussian case

♣ **Proposition.** *Given the above data* $U, \mathcal{A}(u) = A[u;1], \mathcal{V}$, *let us select*

- $\gamma \in (0,1)$
- *a computationally tractable convex compact set*

$$\mathcal{Z} \subset \mathcal{Z}^+ = \{Z \in \mathbf{S}^{n+1} : Z \succeq 0, Z_{n+1,n+1} = 1\}$$

such that $[u;1][u;1]^T \in \mathcal{Z}\ \forall u \in U$

- *A matrix* $\Theta_* \in \mathbf{S}^d$ *and* $\delta \in [0,2]$ *such that*

$$\forall(\Theta \in \mathcal{V}) : \Theta \preceq \Theta_* \ \&\ \|\Theta^{1/2}\Theta_*^{-1/2} - I_d\| \leq \delta \qquad [\|\cdot\| \text{ is the spectral norm}]$$

Let us set

$$B = \left[\begin{array}{c} A \\ 0,\ldots,0,1 \end{array}\right] \in \mathbb{R}^{(d+1)\times(n+1)},\ \mathcal{M} = \mathcal{V} \times \mathcal{Z},\ \mathcal{H} = \{(h,H) \in \mathbb{R}^d \times \mathbf{S}^d : -\gamma\Theta_*^{-1} \preceq H \preceq \gamma\Theta_*^{-1}\}$$

$$\Phi_{\mathcal{A},\mathcal{Z}}(h,H;\Theta,Z) = -\frac{1}{2}\ln\mathrm{Det}(I - \Theta_*^{1/2}H\Theta_*^{1/2}) + \frac{1}{2}\mathrm{Tr}([\Theta - \Theta_*]H) + \frac{\delta(2+\delta)\|\Theta_*^{1/2}H\Theta_*^{1/2}\|_F^2}{2(1-\|\Theta_*^{1/2}H\Theta_*^{1/2}\|)}$$

$$[\|\cdot\|_F - \text{Frobenius norm}]$$

$$+\frac{1}{2}\mathrm{Tr}\left(ZB^T\left[\left[\begin{array}{c|c} H & h \\ \hline h^T & \end{array}\right] + [H,h]^T\left[\Theta_*^{-1} - H\right]^{-1}[H,h]\right]B\right) : \mathcal{H} \times \mathcal{M} \to \mathbb{R}$$

Then $\mathcal{H}, \mathcal{M}, \Phi_{\mathcal{A},\mathcal{Z}}$ *is efficiently computable regular data, and* $\mathcal{G} \subset \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi_{\mathcal{A},\mathcal{Z}}]$.

Sub-Gaussian case

♣ **Proposition.** *Given the above data* $U, \mathcal{A}(u) = A[u;1], \mathcal{V}$, *let us select*

- $\gamma, \gamma^+ \in (0,1)$ *with* $\gamma < \gamma^+$
- *a computationally tractable convex compact set*

$$\mathcal{Z} \subset \mathcal{Z}^+ = \{Z \in \mathbf{S}^{n+1} : Z \succeq 0, Z_{n+1,n+1} = 1\}$$

such that $[u;1][u;1]^T \in \mathcal{Z}\ \forall u \in U$

- *A matrix* $\Theta_* \in \mathbf{S}^d$ *and* $\delta \in [0,2]$ *such that*

$$\forall(\Theta \in \mathcal{V}) : \Theta \preceq \Theta_* \ \&\ \|\Theta^{1/2}\Theta_*^{-1/2} - I_d\| \leq \delta$$

Let us set

$$B = \left[\begin{array}{c} A \\ 0,...,0,1 \end{array}\right] \in \mathbb{R}^{(d+1)\times(n+1)},\ \mathcal{H} = \{(h,H) \in \mathbb{R}^d \times \mathbf{S}^d : -\gamma\Theta_*^{-1} \preceq H \preceq \gamma\Theta_*^{-1}\}$$
$$\mathcal{H}^+ = \{(h,H,G) \in \mathbb{R}^d \times \mathbf{S}^d \times \mathbf{S}^d : -\gamma^+\Theta_*^{-1} \preceq H \preceq G \preceq \gamma^+\Theta_*^{-1}, 0 \preceq G\},\ \mathcal{M} = \mathcal{Z}$$
$$\Phi_{\mathcal{A},\mathcal{Z}}(h,H;Z) = \min_{G:(h,H,G)\in\mathcal{H}^+} \Big\{ -\tfrac{1}{2}\ln\mathrm{Det}(I - \Theta_*^{1/2} G \Theta_*^{1/2})$$
$$+\tfrac{1}{2}\mathrm{Tr}\left(Z B^T \left[\left[\begin{array}{c|c} H & h \\ \hline h^T & \end{array} \right] + [H,h]^T \left[\Theta_*^{-1} - G\right]^{-1} [H,h] \right] B \right) \Big\} : \mathcal{H} \times \mathcal{M} \to \mathbb{R}$$

Then $\mathcal{H}, \mathcal{M}, \Phi_{\mathcal{A},\mathcal{Z}}$ *is efficiently computable regular data, and* $\mathcal{SG} \subset \mathcal{S}[\mathcal{H}, \mathcal{M}, \Phi_{\mathcal{A},\mathcal{Z}}]$.

♠ **How to specify $\mathcal{Z}$.** To apply the above construction, one should specify a computationally tractable convex compact set

$$\mathcal{Z} \subset \mathcal{Z}^+ = \{Z \in \mathbf{S}^{n+1} : Z \succeq 0, Z_{n+1,n+1} = 1\}$$

the smaller the better, such that $u \in U \to [u; 1][u; 1]^T \in \mathcal{Z}$

• The ideal selection is

$$\mathcal{Z} = \mathcal{Z}[U] = \mathrm{Conv}\{[u; 1][u; 1]^T : u \in U\}$$

However: $\mathcal{Z}[U]$ usually is computationally intractable.
Important exception:

$$Q \succ 0, U = \{u : u^T Q u \leq 1\} \Rightarrow \mathcal{Z}[U] = \{Z \in \mathcal{Z}^+ : \sum_{i,j=1}^{n} Z_{ij} Q_{ij} \leq 1\}$$

♡ **"Simple" case:** When U is given by quadratic inequalities:

$$U = \{u \in \mathbb{R}^n : [u; 1]^T Q_s [u; 1] \leq q_s,\ 1 \leq s \leq S\}$$

we can set

$$\mathcal{Z} = \{Z \in \mathbf{S}^{n+1} : Z \succeq 0, Z_{n+1,n+1} = 1, \mathrm{Tr}(Q_s Z) \leq q_s,\ 1 \leq s \leq S\}. \tag{$*$}$$

• **Warning:** $(*)$ can yield very conservative outer approximation of $\mathcal{Z}[U]$. This conservatism with luck can be reduced by passing from the original description of U to an equivalent one, with emphasis on eliminating/updating linear constraints. For example,

• a constraint of the form $|a^T u - c| \leq r$ should be replaced with $(a^T u - c)^2 \leq r^2$

Note: every linear constraint in the description of U can be written as $\alpha - a^T u \geq 0$ and augmented by redundant constraint $a^T u \geq \beta$, with appropriately selected β. The resulting pair of constraints is equivalent to $|a^T u - c| \leq r$ with $c = \frac{1}{2}[\alpha + \beta]$ and $r = \frac{1}{2}[\alpha - \beta]$.

• It could make sense to write the linear constraints in the description of U in the form $\alpha - a^T u \geq 0$ and add to these constraints their pairwise products.

Quadratic Lifting – Does it Pay?

♣ **Situation:** Let for $\chi = 2, 1$ be given

- convex compact sets $U_\chi \subset \mathbb{R}^{n_\chi}$
- affine mappings $u_\chi \mapsto \mathcal{A}_\chi(u_\chi) : \mathbb{R}^{n_\chi} \to \mathbb{R}^d$
- convex compact sets $\mathcal{V}_\chi \subset \operatorname{int} \mathbf{S}^d_+$.

These data define families $\mathcal{G}_\chi$ of Gaussian distributions:

$$\mathcal{G}_\chi = \{\mathcal{N}(\mathcal{A}_\chi(u_\chi), \Theta_\chi) : u_\chi \in U_\chi, \Theta_\chi \in \mathcal{V}_\chi\}$$

♠ Our machinery offers two types of detectors for $\mathcal{G}_1$, $\mathcal{G}_2$:
♠ Affine detector ϕ_{aff} yielded by the solution to the saddle point problem

$$\mathrm{SadVal}_{\text{aff}} = \min_{h \in \mathbb{R}^d} \max_{\substack{u_1 \in U_1, u_2 \in U_2 \\ \Theta_1 \in \mathcal{V}_1, \Theta_2 \in \mathcal{V}_2}} \frac{1}{2}\left[h^T[\mathcal{A}_2(u_2) - \mathcal{A}_1(u_1)] + \frac{1}{2}h^T[\Theta_1 + \Theta_2]h\right]$$

with $\mathrm{Risk}(\phi_{\text{aff}}|\mathcal{G}_1, \mathcal{G}_2) \leq \exp\{\mathrm{SadVal}_{\text{aff}}\}$
♠ Quadratic detector ϕ_{lift} yielded by the solution to the saddle point problem

$$\mathrm{SadVal}_{\text{lift}} = \min_{(h,H) \in \mathcal{H}} \max_{\substack{\Theta_1 \in \mathcal{V}_1 \\ \Theta_2 \in \mathcal{V}_2}} \frac{1}{2}\left[\Phi_{\mathcal{A}_1,\mathcal{Z}_1}(-h, -H; \Theta_1) + \Phi_{\mathcal{A}_2,\mathcal{Z}_2}(h, H; \Theta_2)\right]$$

with $\mathrm{Risk}(\phi_{\text{lift}}|\mathcal{G}_1, \mathcal{G}_2) \leq \exp\{\mathrm{SadVal}_{\text{lift}}\}$

♠ **Fact:** *Assume that the sets $\mathcal{V}_\chi$ contain $\succeq$-largest elements. Then with proper selection of the "design parameters" $\mathcal{Z}_\chi, \Theta_*^{(\chi)}$ participating in the construction of $\Phi_{\mathcal{A}_\chi,\mathcal{Z}_\chi}$, $\chi = 1,2$, passing from affine to quadratic detectors helps:*

$$\mathrm{SadVal}_{\mathrm{lift}} \leq \mathrm{SadVal}_{\mathrm{aff}}$$

♡ **Numerical illustration:**

• $U_1 = U_1^\rho = \{u \in \mathbb{R}^{12} : u_i \geq \rho, 1 \leq i \leq 12\}, U_2 = U_2^\rho = -U_1^\rho$, $A_1 = A_2 \in \mathbb{R}^{8\times 13}$;

• $\mathcal{V}_\chi = \{\Theta_*^{(\chi)} = \sigma_\chi^2 I_8\}$

ρ	σ_1	σ_2	unrestricted H and h	$H = 0$	$h = 0$
0.5	2	2	0.31	0.31	1.00
0.5	1	4	0.24	0.39	0.62
0.01	1	4	0.41	1.00	0.41

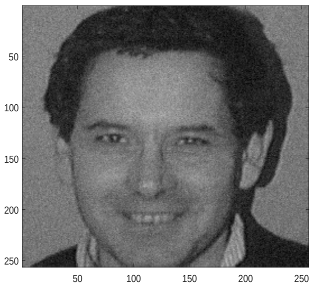
Risk of quadratic detector $\phi(\zeta) = h^T\zeta + \frac{1}{2}\zeta^T H\zeta + \varkappa$

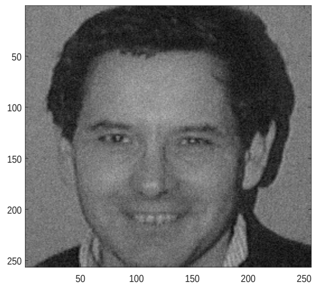
♣ We see that • when deciding on families of Gaussian distributions with common covariance matrix and expectations varying in associated with the families convex sets, passing from affine to quadratic detectors does not help.

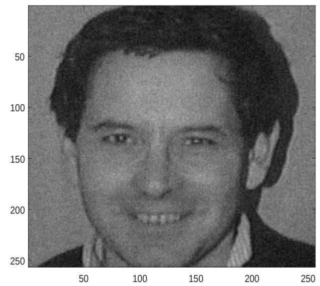
• in general, both affine and purely quadratic components in a quadratic detector are useful.

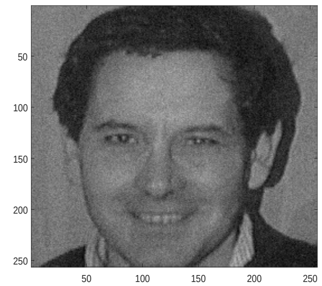
• when deciding on families of Gaussian distributions in the case where distributions from different families can have close expectations, affine detectors are useless, while the quadratic ones are not.

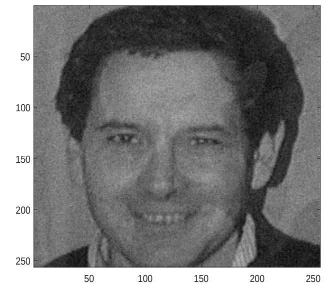
Illustration: Simple Change Point Detection

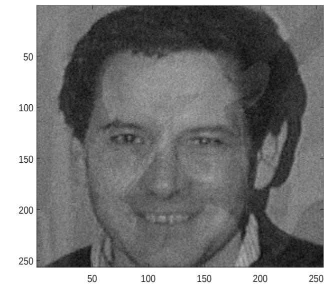

1

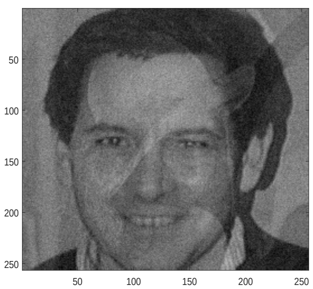

2

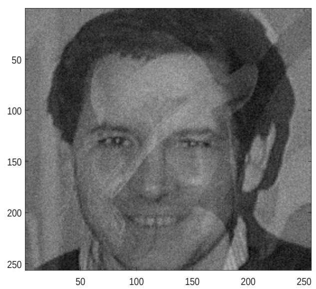

3

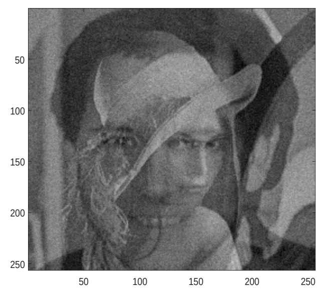

4

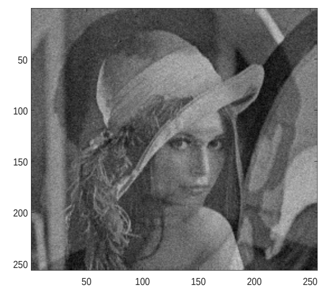

5

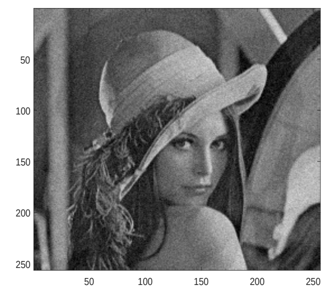

6

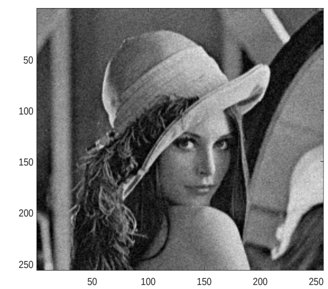

7


8


15


20


28


34

Frames from a noisy “movie”

When the picture starts to change?

♣ **Model:** *We observe one by one vectors ("vectorized" 2D images)*

$$\omega_t = x_t + \xi_t,$$

- x_t: deterministic image
- $\xi_t \sim \mathcal{N}(0, \sigma^2 I_d)$: independent across t observation noises.

Note: We know a range $[\underline{\sigma}, \overline{\sigma}]$ of σ, but perhaps do not know σ exactly.

- We know that $x_1 = x_2$ and want to check whether $x_1 = ... = x_K$ ("no change") or there is a change.

♠ **Goal:** *Given an upper bound* $\epsilon > 0$ *on the probability of false alarm, we want to design a sequential change detection routine capable to detect change, if any.*

♠ **Approach:**

- Pass from observations ω_t, $1 \leq t \leq K$, to observations

$$\zeta_t = \omega_t - \omega_1 = \underbrace{x_t - x_1}_{y_t} + \underbrace{\xi_t - \xi_1}_{\eta_t},\ 2 \leq t \leq K$$

- Test hypothesis $H_0 : y_2 = ... = y_K = 0$ vs. alternative

$$\bigcup_{k=2}^{K} H_k^\rho,\ H_k^\rho : y_2 = ... = y_{k-1} = 0, \|y_k\|_2 \geq \rho$$

via our machinery for testing

magenta hypothesis H_0

vs.

brown hypotheses $H_2^\rho, , ..., H_K^\rho$

via quadratic liftings $\zeta_t\zeta_t^T$ of observations ζ_t up to closeness

$\mathcal{C}$: *all brown hypotheses are close to each other and are not close to the magenta hypothesis*

- We intend to find the smallest ρ for which the $\mathcal{C}$-risk of the resulting inference is $\leq \epsilon$, and utilize this inference in change point detection.

How It Works

♠ **Setup:** $\dim y = 256^2 = 65536$, $\overline{\sigma} = 10$, $\overline{\sigma}^2/\underline{\sigma}^2 = 2$, $K = 9$, $\epsilon = 0.01$

♠ **Inference:** At time $t = 2, ..., K$, compute

$$\phi_*(\zeta_t) = -2.7138\frac{\|\zeta_t\|_2^2}{10^5} + 366.9548.$$

$\phi_*(\zeta_t) < 0 \Rightarrow$ *conclude that the change took place and terminate*

$\phi_*(\zeta_t) \geq 0 \Rightarrow$ *conclude that there was no change so far and proceed to the next image, if any*

♠ **Note:**

- *When magenta hypothesis H_0 holds true, the probability not to claim change on time horizon $2, ..., K$ is at least 0.99.*
- *When a brown hypothesis H_k^ρ holds true, the change at time $\leq K$ is detected with probability at least 0.99*, provided $\rho \geq \rho_* = 2716.6$ (average per pixel energy in y_k at least by 12% larger than $\overline{\sigma}^2$)
- *No test can 0.99-reliably decide via $\zeta_1, ..., \zeta_k$ on H_k^ρ vs. H_0* when $\rho/\rho_* < 0.965$.
- *In the movie, the change takes place at time 3 and is detected at time 4.*

.

ESTIMATING SIGNALS IN GAUSSIAN O.S. AND BEYOND

- *Problem of interest*
- *Developing tools*
  - *Conic Programming*
  - *Conic Duality*
- *Optimizing linear estimates*
  - *Ellitopic case*
  - *Spectratopic case*
- *Near-optimality of linear estimates*
- *Beyond linearity: polyhedral estimates*

♣ **Situation:** "In the nature" there exists a signal x known to belong to a given convex compact set $\mathcal{X} \subset \mathbb{R}^n$. We observe corrupted by noise affine image of the signal ("indirect observations"):

$$\omega = Ax + \xi \in \mathbb{R}^m$$

- A: given $m \times n$ sensing matrix
- ξ: $\mathcal{N}(0, \sigma^2 I)$ observation noise

♠ **Goal:** To recover the image Bx of x under a given linear mapping

- B: given $\nu \times n$ matrix.

♠ **Risk** of a candidate estimate $\widehat{x}(\cdot) : \Omega \to \mathbb{R}^\nu$ is defined as

$$\mathrm{Risk}[\widehat{x}|\mathcal{X}] = \sup_{x \in \mathcal{X}} \sqrt{\mathbf{E}_\xi \left\{ \|Bx - \widehat{x}(Ax + \xi)\|_2^2 \right\}}$$

$\Rightarrow$ Risk^2 is the worst-case, over $x \in \mathcal{X}$, expected $\|\cdot\|_2^2$ recovery error.

♠ With this worst-case quantification of risk, the "golden standard" is the *minimax risk*

$$\mathrm{RiskOpt}[\mathcal{X}] = \inf_{\widehat{x}} \mathrm{Risk}[\widehat{x}|\mathcal{X}],$$

inf being taken over *all* estimates – all (measurable) functions $\widehat{x}(\cdot) : \mathbb{R}^m \to \mathbb{R}^\nu$. .

♣ **Agenda:** Under appropriate assumptions on $\mathcal{X}$, we shall show that

A. *One can build, in a computationally efficient fashion, (nearly) the best, in terms of risk, estimate in the family of linear estimates*

$$\widehat{x}(\omega) = \widehat{x}_H(\omega) = H^T\omega \qquad [H \in \mathbb{R}^{m\times\nu}]$$

B. *The resulting linear estimate is nearly optimal among all estimates, linear and nonlinear alike.*

C. *Under appropriate assumptions on a norm $\|\cdot\|$ and a family $\mathcal{P}$ of distributions of observation noise, the results of* ***A***, ***B*** *can be extended to the situation where*
— the recovery error is measured in norm $\|\cdot\|$,
— distribution P of observation noise is known to belong to $\mathcal{P}$,
— the $\|\cdot\|_2$-risk

$$\mathrm{Risk}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \sqrt{\mathbf{E}_\xi\left\{\|Bx - \widehat{x}(Ax+\sigma\xi)\|_2^2\right\}}$$

is replaced with $(\|\cdot\|,\mathcal{P})$-risk

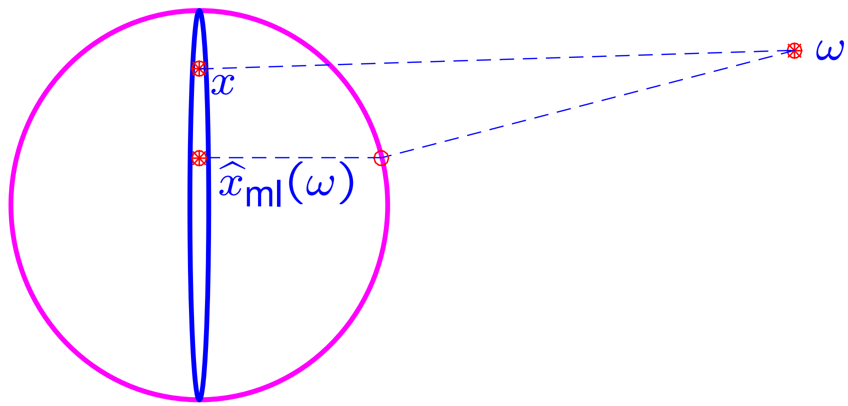
$$\mathrm{Risk}_{\|\cdot\|,\mathcal{P}}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}}\sup_{P\in\mathcal{P}} \mathbf{E}_{\xi\sim P}\left\{\|Bx - \widehat{x}(Ax+\xi)\|\right\}$$

Why linear estimates?

♠ As it was announced, *a "nearly optimal" linear estimate can be built in a computationally efficient fashion.*

♠ **In contrast,**

- *Exactly minimax optimal* estimate is *unknown* even in the simplest case when the observation is $\omega = x + \xi$ with $\xi \sim \mathcal{N}(0, \sigma^2)$ and $x \in \mathcal{X} = [-1, 1]$
- The "magic wand" of Statistics – the Maximum Likelihood estimate — is known to be optimal in the "noise goes to 0" asymptotics and can be disastrously bad before this asymptotics starts.



blue: $\mathcal{X}$ magenta: $A\mathcal{X}$

- $\mathcal{X} = \{x \in \mathbb{R}^n : x_n^2 + \epsilon^{-2} \sum_{i=1}^{n-1} x_i^2 \leq 1\}$
- $A = \text{Diag}\{1/\epsilon, ..., 1/\epsilon, 1\}$, $\eta \sim \mathcal{N}(0, \sigma^2 I_n)$, $B = I_n$

$\Rightarrow$ MLE: $\widehat{x}_{\rm ml}(\omega) = A^{-1} \cdot \text{argmin}_{\|u\|_2 \leq 1} \|\omega - u\|_2$

When $\sigma \ll 1$, $\sigma^2 n \geq O(1)$, and $\epsilon \leq O(\sigma)$, the risk of MLE is $O(1)$, while the risk of the linear estimate $\widehat{x}(\omega) = \omega_n$ is $O(\sigma) \ll O(1)$.

Note: As $\sigma \to 0$, the ML estimate regains optimality, but this happens the later the larger is n.

Developing Tools, Optimization
"Structure-Revealing" Representation of Convex Problem: Conic Programming

♣ When passing from a Linear Programming program

$$\min_x \left\{c^T x : Ax - b \geq 0\right\}$$

to a convex one, the traditional wisdom is to replace linear inequality constraints

$$a_i^T x - b_i \geq 0$$

with nonlinear ones:

$$g_i(x) \geq 0 \qquad [g_i \text{ are concave}]$$

♠ There exists, however, another way to introduce nonlinearity, namely, to replace the coordinate-wise *vector* inequality

$$y \geq z \Leftrightarrow y - z \in \mathbb{R}^m_+ = \{u \in \mathbb{R}^m : u_i \geq 0\,\forall i\} \qquad [y, z \in \mathbb{R}^m]$$

with another *vector* inequality

$$y \geq_{\mathbf{K}} z \Leftrightarrow y - z \in \mathbf{K} \qquad [y, z \in \mathbb{R}^m]$$

where $\mathbf{K}$ is a *regular cone* (i.e., closed, pointed and convex cone with a nonempty interior) in $\mathbb{R}^m$.

$$y \geq_{\mathbf{K}} z \Leftrightarrow y - z \in \mathbf{K} \qquad [y, z \in \mathbb{R}^m]$$

$\mathbf{K}$: closed, pointed and convex cone in $\mathbb{R}^m$ with a nonempty interior.

Requirements on $\mathbf{K}$ ensure that $\geq_{\mathbf{K}}$ obeys the usual rules for inequalities:

- $\geq_{\mathbf{K}}$ is a *partial order*:

$$\begin{array}{ll} x \geq_{\mathbf{K}} x \,\forall x & \text{[reflexivity]} \\ (x \geq_{\mathbf{K}} y \,\&\, y \geq_{\mathbf{K}} x) \Rightarrow x = y & \text{[antisymmetry]} \\ (x \geq_{\mathbf{K}} y, y \geq_{\mathbf{K}} z) \Rightarrow x \geq_{\mathbf{K}} z & \text{[transitivity]} \end{array}$$

- $\geq_{\mathbf{K}}$ *is compatible with linear operations:* the validity of $\geq_{\mathbf{K}}$ inequality is preserved when we multiply both sides by the same nonnegative real and add to it another valid $\geq_{\mathbf{K}}$-inequality;

- *in a sequence of $\geq_{\mathbf{K}}$-inequalities, one can pass to limits:*

$$\{a_i \geq_{\mathbf{K}} b_i,\ i = 1, 2, \ldots \,\&\, a_i \to a \,\&\, b_i \to b\} \Rightarrow a \geq_{\mathbf{K}} b$$

- *one can define the strict version $>_{\mathbf{K}}$ of $\geq_{\mathbf{K}}$:*

$$a >_{\mathbf{K}} b \Leftrightarrow a - b \in \operatorname{int} \mathbf{K}.$$

  Arithmetics of $>_{\mathbf{K}}$ and $\geq_{\mathbf{K}}$ inequalities is completely similar to the arithmetics of the usual coordinate-wise $\geq$ and $>$.

♣ LP problem:

$$\min_x \left\{c^T x : Ax - b \geq 0\right\} \Leftrightarrow \min_x \left\{c^T x : Ax - b \in \mathbb{R}^m_+\right\}$$

♣ General Conic problem:

$$\min_x \left\{c^T x : Ax - b \geq_{\mathbf{K}} 0\right\} \Leftrightarrow \min_x \left\{c^T x : Ax - b \in \mathbf{K}\right\}$$

- (A, b) – *data* of conic problem
- $\mathbf{K}$ - structure of conic problem

♠ Note: Every convex problem admits equivalent conic reformulation

♠ Note: With conic formulation, convexity is "built in"; with the standard MP formulation convexity should be kept in mind as an additional property.

♣ (??) A general convex cone has no more structure than a general convex function. Why conic reformulation is "structure-revealing"?

♣ (!!) As a matter of fact, just 3 types of cones allow to represent an extremely wide spectrum ("essentially all") of convex problems!

$$\boxed{\min_x \left\{c^T x : Ax - b \geq_{\mathbf{K}} 0\right\} \Leftrightarrow \min_x \left\{c^T x : Ax - b \in \mathbf{K}\right\}}$$

♠ Three Magic Families of cones:

- *$\mathcal{LP}$: Nonnegative orthants $\mathbb{R}^m_+$* – direct products of m nonnegative rays $\mathbb{R}_+ = \{s \in \mathbb{R} : s \geq 0\}$ giving rise to Linear Programming programs
$$\min_s \left\{c^T x : a_\ell^T x - b_\ell \geq 0, 1 \leq \ell \leq q\right\}.$$

- *$\mathcal{CQP}$: Direct products of Lorentz cones*
$\mathbf{L}^p_+ = \{u \in \mathbb{R}^p : u_p \geq \left(\sum_{i=1}^{p-1} u_i^2\right)^{1/2}\}$ giving rise to Conic Quadratic programs
$$\min_x \left\{c^T x : \|A_\ell x - b_\ell\|_2 \leq c_\ell^T x - d_\ell, 1 \leq \ell \leq q\right\}.$$

- *$\mathcal{SDP}$: Direct products of Semidefinite cones*
$\mathbf{S}^p_+ = \{M \in \mathbf{S}^p : M \succeq 0\}$ giving rise to Semidefinite programs
$$\min_x \left\{c^T x : \underbrace{\lambda_{\min}(\mathcal{A}^\ell(x)) \geq 0}_{\Leftrightarrow \mathcal{A}^\ell(x) \succeq 0}, 1 \leq \ell \leq q\right\}.$$
where $\mathbf{S}^p$ is the space of $p \times p$ real symmetric matrices, $\mathcal{A}_\ell(x) \in \mathbf{S}^p$ are affine in x and $\lambda_{\min}(S)$ is the minimal eigenvalue of $S \in \mathbf{S}^p$.

What can be reduced to $\mathcal{LP}/\mathcal{CQP}/\mathcal{SDP}$?
Calculus of Conic programs

♣ Let $\mathcal{K}$ be a family of regular cones closed w.r.t. taking direct products.

♠ **Definition:** • A *$\mathcal{K}$-representation of a set $X \subset \mathbb{R}^n$* is a representation

$$X = \{x \in \mathbb{R}^n : \exists u \in \mathbb{R}^m : Ax + Bu - b \in \mathbf{K}\} \qquad (*)$$

where $\mathbf{K} \in \mathcal{K}$.

• X is called *$\mathcal{K}$-representable*, if X admits a $\mathcal{K}$-r.

♡ **Note:** *Minimizing a linear objective $c^T x$ over a $\mathcal{K}$-representable set X reduces to a conic program on a cone from $\mathcal{K}$.*

Indeed, given $(*)$, problem $\min\limits_{x \in X} c^T x$ is equivalent to

$$\mathrm{Opt} = \min_{x,u}\left\{c^T x : Ax + Bu - b \in \mathbf{K}\right\}$$

♠ **Definition:** • *A $\mathcal{K}$-representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a $\mathcal{K}$-representation of the epigraph of f:*

$$\begin{array}{rcl} \mathrm{Epi}\{f\} & := & \{(x,t) : t \geq f(x)\} \\ & = & \{x, t : \exists v : Px + pt + Qv - q \in \mathbf{K}\}, \quad \mathbf{K} \in \mathcal{K} \end{array}$$

• f is called *$\mathcal{K}$-representable*, if f admits a $\mathcal{K}$-r.

♡ **Note:**

- *A level set of a $\mathcal{K}$-r. function is $\mathcal{K}$-r.:*

$$\begin{array}{rcl}\text{Epi}\{f\} & := & \{(x,t) : t \geq f(x)\} \\ & = & \{x,t : \exists v : Px + pt + Qu - q \in \mathbf{K}\} \\ \Rightarrow \{x : f(x) \leq c\} & = & \{x : \exists v : Px + Qu - [q - cp] \in \mathbf{K}\}\end{array}$$

- *Minimization of a $\mathcal{K}$-r. function f over a $\mathcal{K}$-r. set X reduces to a conic program on a cone from $\mathcal{K}$:*

$$\left.\begin{array}{rcl} x \in X & \Leftrightarrow & \exists u : Ax + Bu - b \in \mathbf{K}_X \\ t \geq f(x) & \Leftrightarrow & \exists v : Px + pt + Qv - q \in \mathbf{K}_f \end{array}\right\} \Rightarrow$$

$$\boxed{\begin{array}{c} \min_{x \in X} f(x) \\ \Updownarrow \\ \min\limits_{t,x,u,v} \left\{t : [Ax + Bu - b; Px + pt + Qv - q] \in \underbrace{\mathbf{K}_X \times \mathbf{K}_f}_{\in \mathcal{K}}\right\} \end{array}}$$

♣ Investigating "expressive abilities" of generic Magic conic problems reduces to answering the question

What are $\mathcal{LP}/\mathcal{CQP}/\mathcal{SDP}$-r. functions/sets?

♠ **"Built-in" restriction is Convexity:** *A $\mathcal{K}$-representable set/function must be convex.*

♠ **Good news:** *Convexity, essentially, is the only restriction: for all practical purposes, all convex sets/functions arising in applications are $\mathcal{SDP}$-r. Quite rich families of convex functions/sets are $\mathcal{LP}/\mathcal{CQP}$-r.*

♡ **Note:** Nonnegative orthants are direct products of (1-dimensional) Lorentz cones, and Lorentz cones are intersections of semidefinite cones and properly selected linear subspaces $\Rightarrow \mathcal{LP} \subset \mathcal{CQP} \subset \mathcal{SDP}$.

♣ *Let $\mathcal{K}$ be a family of regular cones closed w.r.t. taking direct products and passing from a cone $\mathbf{K}$ to its dual cone*

$$\mathbf{K}_* = \{\lambda : \langle \lambda, \xi \rangle \geq 0 \ \forall \xi \in \mathbf{K}\}$$

Note: $\mathbf{K}_*$ is regular cone provided $\mathbf{K}$ is so, and

$$(\mathbf{K}_*)_* = \mathbf{K}$$

♠ **Fact:** *$\mathcal{K}$-representable sets/functions admit fully algorithmic calculus: all basic convexity-preserving operations with functions/sets, as applied to $\mathcal{K}$-r. operands, produce $\mathcal{K}$-r. results, and the resulting $\mathcal{K}$-r.'s are readily given by $\mathcal{K}$-r.'s of the operands. "Calculus rules" are independent of what $\mathcal{K}$ is.*

$\Rightarrow$ *Starting with "raw materials"* (characteristic for $\mathcal{K}$ elementary $\mathcal{K}$-r. sets/functions) *and applying calculus rules, we can recognize $\mathcal{K}$-representability and get explicit $\mathcal{K}$-r.'s of sets/functions of interest.*

♣ **Basics of "calculus of $\mathcal{K}$-representability":**

♠ **[Sets:]** If $X_1, ..., X_k$ are $\mathcal{K}$-r. sets, so are their

- *intersections,*
- *direct products,*
- *images under affine mappings,*
- *inverse images under affine mappings.*

♠ **[Functions:]** If $f_1, ..., f_k$ are $\mathcal{K}$-r. functions, so are their

- *linear combinations with nonnegative coefficients,*
- *superpositions with affine mappings.*

Moreover, *if $F, f_1, ..., f_k$ are $\mathcal{K}$-r. functions, so is the superposition $F(f_1(x), ..., f_k(x))$ provided that F is monotonically nondecreasing in its arguments.*

♠ More advanced convexity-preserving operations preserve $\mathcal{K}$-representability under (pretty mild!) regularity conditions. This includes

• **for sets:** taking *conic hulls* and *convex hulls of (finite) unions* and passing from a set to its *recessive cone*, or *polar*, or *support function*

• **for functions:** *partial minimization*, *projective transformation*, and *taking Fenchel dual*.

♠ **Note:** *Calculus rules are simple and algorithmic*

$\Rightarrow$ *Calculus can be run on a compiler* [used in `cvx`].

Illustration

$$\min c^T x + d^T y$$

$$y \geq 0,\ Ax + By \leq b$$

$$2y_1^{-\frac{7}{2}} y_2^{-3} y_3^{\frac{-1}{5}} + 3y_2^{-\frac{3}{2}} y_4^{-\frac{2}{3}} \leq e^T x + 4y_1^{\frac{1}{5}} y_2^{\frac{2}{5}} y_3^{\frac{2}{5}} + 5y_3^{\frac{1}{3}} y_4^{\frac{2}{5}}$$

$$\left[\begin{array}{c|c|c|c} x_1 - x_2 & x_3 + x_2 & & \\ \hline x_3 + x_2 & x_2 - x_4 & x_5 - 6 & \\ \hline & x_5 - 6 & x_6 + x_7 & -x_8 \\ \hline & & -x_8 & x_5 \end{array}\right] \succeq 0$$

$$\text{Det}\left(\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_2 & x_6 & x_7 & x_8 & x_9 \\ x_3 & x_7 & x_{10} & x_{11} & x_{12} \\ x_4 & x_8 & x_{11} & x_{13} & x_{14} \\ x_5 & x_9 & x_{12} & x_{14} & x_{15} \end{bmatrix}\right) \geq 1$$

$$\text{Sum of 2 largest singular values of } \begin{bmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \\ x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} \end{bmatrix} \text{ is } \leq 6$$

$$1 - \sum_{i=1}^{6} [x_i - x_{i+1}] s^i \leq 0,\ \tfrac{3}{2} \leq s \leq 6$$

$$\sum_{i=1}^{4} x_{2i} \cos(i\phi) - \sum_{i=1}^{4} x_i \sin(i\phi) \leq 1,\ \tfrac{\pi}{3} \leq \phi \leq \tfrac{\pi}{2}$$

- the blue part of the problem is in $\mathcal{LP}$
- the blue-magenta part of the problem is in $\mathcal{CQP}$ and can be approximated, *in a polynomial time fashion*, by $\mathcal{LP}$
- the entire problem is in $\mathcal{SDP}$

and the reductions to $\mathcal{LP}/\mathcal{CQP}/\mathcal{SDP}$ are “fully algorithmic.”

Conic Duality

♣ Conic Programming admits nice Duality Theory completely similar to LP Duality.

Primal problem:

$$\min_x \left\{ c^T x : \left\{ \begin{array}{rcl} Ax - b & \geq_{\mathbf{K}} & 0 \\ Rx & = & r \end{array} \right\} \right.$$

$\Leftrightarrow$ [passing to primal slack $\xi = Ax - b$]

$$\boxed{\min_\xi \left\{ e^T \xi : \xi \in [\mathcal{L} - b] \cap \mathbf{K} \right\} \qquad (\mathcal{P})}$$

$$\left[\begin{array}{c} e : A^T e + R^T f = c \text{ for some } f \\ \mathcal{L} = \{Au : Ru = 0\} \end{array} \right]$$

Dual problem:

$$\max_{y,z} \left\{ b^T y : A^T y + R^T z = c,\ y \geq_{\mathbf{K}_*} 0 \right\}$$

$$\Leftrightarrow \quad \max_y \left\{ b^T y : y \in \mathbf{K}_*, \exists z : A^T y + R^T z = c \right\}$$

$$\boxed{\max_y \left\{ b^T y : y \in [\mathcal{L}^\perp + e] \bigcap \mathbf{K}_* \right] \qquad (\mathcal{D})}$$

[$\mathbf{K}_*$: cone dual to $\mathbf{K}$]

Note:

- the dual problem is conic along with primal
- the duality is completely symmetric

Note: Cones from Magic Families are self-dual, so that the dual of a Linear/Conic Quadratic/Semidefinite program is of exactly the same type.

Derivation of the Dual Problem

♣ **Primal problem:**

$$\mathrm{Opt}(P) = \min_x \left\{ c^T x : \begin{array}{l} A_i x - b_i \in \mathbf{K}^i, i \leq m \\ Rx = r \end{array} \right\} \quad (P)$$

♠ **Goal:** find a systematic way to bound $\mathrm{Opt}(P)$ *from below*.

♠ **Simple observation:** *When $y_i \in \mathbf{K}^i_*$, the scalar inequality $y_i^T A_i x \geq y_i^T b_i$ is a consequence of the constraint $A_i x - b_i \in \mathbf{K}^i$. If z is a vector of the same dimension as r, the scalar inequality $z^T R x \geq z^T r$ is a consequence of the constraint $Rx = r$.*

$\Rightarrow$ *Whenever $y_i \in \mathbf{K}^i_*$ for all i and z is a vector of the same dimension as r, the scalar linear inequality*

$$[\sum_i A_i^T y_i + R^T z]^T x \geq \sum_i b_i^T y_i + r^T z$$

is a consequence of the constraints in (P)

$\Rightarrow$ *Whenever $y_i \in \mathbf{K}^i_*$ for all i and z is a vector of the same dimension as r such that*

$$\sum_i A_i^T y_i + R^T z = c,$$

the quantity $\sum_i b_i^T y_i + r^T z$ is a lower bound on $\mathrm{Opt}(P)$.

• The *Dual problem*

$$\mathrm{Opt}(D) = \max_{y_i, z} \left\{ \sum_i b_i^T y_i + r^T z : \begin{array}{l} y_i \in \mathbf{K}^i_*, i \leq m \\ \sum_i A_i^T y_i + R^T z = c \end{array} \right\} \quad (D)$$

is just the problem of maximizing this lower bound on $\mathrm{Opt}(P)$.

♣ **Definition:** A conic problem

$$\min_x \left\{ c^T x : \begin{array}{l} A_i x - b_i \in \mathbf{K}^i,\ i \le m \\ Ax \le b \\ Rx = r \end{array} \right\} \qquad (C)$$

is called *strictly feasible*, if there exists a *feasible* solution $\bar{x}$ where all conic and $\le$ constraints are satisfied *strictly*:

$$A_i\bar{x} - b_i \in \mathrm{int}\,\mathbf{K}^i\ \forall i\ \&\ A\bar{x} < b,$$

and is called *essentially strictly feasible*, if there exists a *feasible* solution $\bar{x}$ where all *non-polyhedral* constraints are satisfied strictly:

$$A_i\bar{x} - b_i \in \mathrm{int}\,\mathbf{K}^i\ \forall i.$$

♣ **Conic Programming Duality Theorem.** *Consider a conic problem*

$$\mathrm{Opt}(P) = \min_x \left\{ c^T x : \begin{array}{l} A_i x - b_i \in \mathbf{K}^i, i \le m \\ Rx = r \end{array} \right\} \quad (P)$$

along with its dual

$$\mathrm{Opt}(D) = \max_{y_i, z} \left\{ \sum_i b_i^T y_i + r^T z : \begin{array}{l} y_i \in \mathbf{K}^i_*, i \le m \\ \sum_i A_i^T y_i + R^T z = c \end{array} \right\} \quad (D)$$

Then:

♠ [Symmetry] *Duality is symmetric: the dual problem is conic, and its dual is (equivalent to) the primal problem;*

♠ [Weak duality] *One has* $\mathrm{Opt}(D) \le \mathrm{Opt}(P)$*;*

♠ [Strong duality] *Let one of the problems be essentially strictly feasible and bounded. Then the other problem is solvable, and*

$$\mathrm{Opt}(D) = \mathrm{Opt}(P).$$

In particular, if both problems are essentially strictly feasible, both are solvable with equal optimal values.

$$\min_x \left\{ c^T x : \begin{array}{l} A_i x - b_i \in \mathbf{K}^i, i \le m \\ Rx = r \end{array} \right\} \qquad (P)$$

$$\Updownarrow$$

$$\max_{y_i, z} \left\{ \sum_i b_i^T y_i + r^T z : \begin{array}{l} y_i \in \mathbf{K}^i_*, i \le m \\ \sum_i A_i^T y_i + R^T z = c \end{array} \right\} \qquad (D)$$

Conic Programming Optimality Conditions:

Let both (P) and (D) be essentially strictly feasible. Then a pair $(x, [\{y_i\}, z])$ of primal and dual feasible solutions is comprised of optimal solutions to the respective problems if and only if

- [Zero Duality Gap]

$$\mathrm{DualityGap}(x, [\{y_i\}, z]) := c^T x - [\sum_i b_i^T y_i + r^T z] = 0$$

Indeed, $\mathrm{DualityGap}(x, [\{y_i\}, z]) = \underbrace{[c^T x - \mathrm{Opt}(P)]}_{\ge 0} + \underbrace{[\mathrm{Opt}(D) - [\sum_i b_i^T y_i + r^T z]]}_{\ge 0}$

and if and only if

- [Complementary Slackness]

$$[A_i x - b_i]^T y_i = 0, \; i \le m$$

Indeed, $\sum_i \underbrace{[A_i x - b_i]^T y_i}_{\ge 0} = [\sum_i A_i^T y_i] x - \sum_i b_i^T y_i = \underbrace{[c - R^T z]^T x}_{= c^T x - r^T z} - \sum_i b_i^T y_i = c^T x - [\sum_i b_i^T y_i + r^T z]$

$$= \mathrm{DualityGap}(x, [\{y_i\}, z])$$

♣ Conic Duality, same as the LP one, is

- *fully algorithmic:* to write down the dual, given the primal, is a purely mechanical process
- *fully symmetric:* the dual problem "remembers" the primal one

♡ Cf. Lagrange Duality:

$$\min_x \{f(x) : g_i(x) \leq 0,\, i = 1, ..., m\} \quad (P)$$
$$\Downarrow$$
$$\max_{y \geq 0} \underline{L}(y) \quad (D)$$
$$\left[\underline{L}(y) = \min_x \left\{ f(x) + \sum_i y_i g_i(x) \right\}\right]$$

- Dual "exists in the nature", but is given implicitly; its objective, typically, is not available in a closed form
- Duality is asymmetric: given $\underline{L}(\cdot)$, we, typically, cannot recover f and g_i...

Developing Tools, Optimization
Schur Complement Lemma

♣ **Lemma:** *Symmetric block matrix* $\left[\begin{array}{c|c} P & S^T \\ \hline S & R \end{array}\right]$ *with* $R \succ 0$ *is positive semidefinite if and only if the matrix* $P - S^T R^{-1} S$ *is so.*

Proof: $\left[\begin{array}{c|c} P & S^T \\ \hline S & R \end{array}\right] \succeq 0$ iff

$$
\begin{array}{rcl}
0 & \leq & \min\limits_{u,v}[u^T P u + 2u^T S^T v + v^T R v] \\
 & = & \min\limits_{u}\left[\underbrace{\min\limits_{v}[u^T P u + 2u^T S^T v + v^T R v]}_{\text{achieved when } v = -R^{-1}Su}\right] \\
 & = & \min\limits_{u} u^T \left[P - S^T R^{-1} S\right] u.
\end{array}
$$

Optimizing Linear Estimates

♣ **Situation:** "In the nature" there exists a signal x known to belong to a given convex compact set $\mathcal{X} \subset \mathbb{R}^n$. We observe corrupted by noise affine image of the signal:

$$\omega = Ax + \sigma\xi \in \Omega = \mathbb{R}^m$$

- A: given $m \times n$ sensing matrix • ξ: random noise

♠ **Goal:** To recover the image Bx of x

- B: given $\nu \times n$ matrix.

♠ **Risk** of a candidate estimate $\widehat{x}(\cdot) : \Omega \to \mathbb{R}^\nu$ is

$$\mathrm{Risk}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \sqrt{\mathbf{E}_\xi\left\{\|Bx - \widehat{x}(Ax+\sigma\xi)\|_2^2\right\}}$$

♣ **Assumption on noise:** *ξ is zero mean with unit covariance matrix.*

$\Rightarrow$ *The risk of a linear estimate $\widehat{x}_H(\omega) = H^T\omega$ (H: contrast matrix) is given by*

$$\begin{array}{rcl}\mathrm{Risk}^2[\widehat{x}_H|\mathcal{X}] & = & \max\limits_{x\in\mathcal{X}} \mathbf{E}_\xi\left\{\|[B-H^TA]x - \sigma H^T\xi\|_2^2\right\}\\ & = & \max\limits_{x\in\mathcal{X}} \left\{\|[B-H^TA]x\|_2^2 + \sigma^2\mathbf{E}_\xi\{\mathrm{Tr}(H^T\xi\xi^TH)\}\right\}\\ & = & \sigma^2\mathrm{Tr}(H^TH) + \underbrace{\max\limits_{x\in\mathcal{X}}\mathrm{Tr}([B-H^TA]xx^T[B^T-A^TH])}_{\Psi(H)}.\end{array}$$

$$\boxed{\mathrm{Risk}^2[\widehat{x}_H|\mathcal{X}] = \sigma^2\mathrm{Tr}(H^TH) + \Psi(H),\ \Psi(H) = \max_{x\in\mathcal{X}} \mathrm{Tr}([B - H^TA]xx^T[B^T - A^TH]).}$$

♡ **Note:** Ψ is convex $\Rightarrow$ *building the minimum risk linear estimate reduces to solving convex minimization problem*

$$\mathrm{Opt} = \min_H \left[\Psi(H) + \sigma^2\mathrm{Tr}(H^TH)\right]. \qquad (*)$$

But: Convex function Ψ is given implicitly and can be difficult to compute, making $(*)$ difficult as well.

$$
\begin{array}{c}
\mathrm{Opt} = \min_H \left[\sigma^2 \mathrm{Tr}(H^T H) + \Psi(H)\right] \\
\Psi(H) = \max\limits_{x\in\mathcal{X}} \mathrm{Tr}([B - H^T A] x x^T [B^T - A^T H])
\end{array} \qquad (*)
$$

♡ **Fact:** Basically, the only cases when $(*)$ is known to be easy are those when

- $\mathcal{X}$ is given as a convex hull of finite set of moderate cardinality
- $\mathcal{X}$ is an ellipsoid.

$\mathcal{X}$ is a box $\Rightarrow$ computing Ψ is NP-hard...

♠ When Ψ is difficult to compute, we can to replace Ψ in the design problem $(*)$ with an efficiently computable convex upper bound $\Psi^+(H)$.

We are about to consider a family of sets $\mathcal{X}$ – *ellitopes* – for which reasonably tight bounds Ψ^+ of desired type are available.

♣ **A basic ellitope** is a set $\mathcal{Y} \subset \mathbb{R}^N$ given as

$$\mathcal{Y} = \{y \in \mathbb{R}^N : \exists t \in \mathcal{T} : y^T S_k y \leq t_k,\, k \leq K\}$$

where

- $S_k \succeq 0$ are positive semidefinite matrices with $\sum_k S_k \succ 0$
- $\mathcal{T}$ is a convex compact subset of K-dimensional nonnegative orthant $\mathbb{R}^K_+$ such that
    - $\mathcal{T}$ contains some positive vectors
    - $\mathcal{T}$ is *monotone*: if $0 \leq t' \leq t$ and $t \in \mathcal{T}$, then $t' \in \mathcal{T}$ as well.

♠ **An ellitope** $\mathcal{X}$ is linear image of a basic ellitope:

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists y \in \mathbb{R}^N, t \in \mathcal{T} : x = Fy,\, y^T S_k y \leq t_k,\, k \leq K\}$$

- F is a given $n \times N$ matrix,

♠ **Note:** *Every ellitope is a symmetric w.r.t. the origin convex compact set.*

Examples of basic ellitopes:

A. Ellipsoid centered at the origin

$(K = 1, \mathcal{T} = [0; 1])$

B. (Bounded) intersection of K ellispoids/elliptic cylinders centered at the origin

$(\mathcal{T} = \{t \in \mathbb{R}^K : 0 \le t_k \le 1,\ k \le K\})$

C. Box $\{x \in \mathbb{R}^n : -1 \le x_i \le 1\}$

$(\mathcal{T} = \{t \in \mathbb{R}^n : 0 \le t_k \le 1,\ k \le K = n\}, x^T S_k x = x_k^2)$

D. ℓ_p-ball $\mathcal{X} = \{x \in \mathbb{R}^n : \|x\|_p \le 1\}$ with $p \ge 2$

$(\mathcal{T} = \{t \in \mathbb{R}^n_+ : \|t\|_{p/2} \le 1\}, x^T S_k x = x_k^2,\ k \le K = n)$

♠ *Ellitopes admit fully algorithmic calculus:* if $\mathcal{X}_i$, $1 \le i \le I$, are ellitopes, so are their

- intersection $\bigcap_i \mathcal{X}_i$
- direct product $\mathcal{X}_1 \times \ldots \times \mathcal{X}_I$
- arithmetic sum $\mathcal{X}_1 + \ldots + \mathcal{X}_I$
- linear images $\{Ax : x \in \mathcal{X}_i\}$
- inverse linear images $\{y : Ay \in \mathcal{X}_i\}$ under linear embedding A

♣ **Observation:** *Let*

$$\mathcal{X} = \{x : \exists (t \in \mathcal{T}, y) : x = Fy,\ y^T S_k y \le t_k\,, k \le K\} \qquad (*)$$

be an ellitope. Given a quadratic form $x^T W x$, $W \in \mathbf{S}^n$, *we have*

$$\max_{x\in\mathcal{X}} x^T W x \le \min_{\lambda} \left\{ \phi_{\mathcal{T}}(\lambda) : \lambda \ge 0, \sum_{k=1}^{K} \lambda_k S_k \succeq F^T W F \right\}$$

$$\phi_{\mathcal{T}}(\lambda) = \max_{t\in\mathcal{T}} t^T\lambda : \ \textit{support function of } \mathcal{T}$$

Indeed, we have

$$\begin{array}{l}\lambda \ge 0, F^T W F \preceq \sum_k \lambda_k S_k, x \in \mathcal{X} \Rightarrow \exists (t \in \mathcal{T}, y) : y^T S_k y \le t_k\ \forall k \le K, x = Fy \\ \Rightarrow \exists (t \in \mathcal{T}, y) : x^T W x = y^T F^T W F y \le \sum_k \lambda_k y^T S_k y \le \sum_k \lambda_k t_k \le \phi_{\mathcal{T}}(\lambda) \\ \Rightarrow x^T W x \le \phi_{\mathcal{T}}(\lambda).\end{array}$$

$$\mathcal{X} = \{x : \exists (t \in \mathcal{T}, y) : x = Fy,\ y^T S_k y \leq t_k\,, k \leq K\} \qquad (*)$$

♠ **Corollary:** *Let $\mathcal{X}$ be the ellitope $(*)$. Then the function*

$$\begin{array}{rcl}\Psi(H) & = & \max\limits_{x\in\mathcal{X}} \mathrm{Tr}((B - H^T A)xx^T(B^T - A^T H)) \\ & = & \max\limits_{x\in\mathcal{X}} x^T[(B^T - A^T H)(B - H^T A)]x\end{array}$$

can be upper-bounded as

$$\begin{array}{rcl}\Psi(H) \leq \overline{\Psi}(H) & := & \min\limits_{\lambda} \left\{\phi_{\mathcal{T}}(\lambda) : \lambda \geq 0, F^T[B^T - A^T H][B - H^T A]F \preceq \sum\limits_k \lambda_k S_k\right\} \\ & & \text{[Schur Complement Lemma]} \\ & = & \min\limits_{\lambda} \left\{\phi_{\mathcal{T}}(\lambda) : \lambda \geq 0, \left[\begin{array}{c|c} \sum_k \lambda_k S_k & F^T[B^T - A^T H] \\ \hline [B - H^T A]F & I_\nu \end{array}\right] \succeq 0\right\}\end{array}$$

The function $\overline{\Psi}(H)$ is real-valued and convex, and is efficiently computable whenever $\phi_{\mathcal{T}}$ is so, that is, whenever $\mathcal{T}$ is computationally tractable.

♠ **Bottom line:** *Given matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{\nu\times n}$ and an ellitope*

$$\mathcal{X} = \{x : \exists (t \in \mathcal{T}, y) : x = Fy,\ y^T S_k y \le t_k,\ k \le K\} \qquad (*)$$

contained in $\mathbb{R}^n$, consider the convex optimization problem

$$\mathrm{Opt} = \min_{H,\lambda} \left\{ \phi_{\mathcal{T}}(\lambda) + \sigma^2 \mathrm{Tr}(H^T H) : \begin{array}{c} \lambda \ge 0, \\ \left[\begin{array}{c|c} \sum_k \lambda_k S_k & F^T[B^T - A^T H] \\ \hline [B - H^T A]F & I_\nu \end{array}\right] \succeq 0 \end{array} \right\}.$$
$$\left[\phi_{\mathcal{T}}(\lambda) = \max_{t\in\mathcal{T}} \lambda^T t\right]$$

Assuming the noise ξ in observation $\omega = Ax + \sigma\xi$ zero mean with unit covariance matrix, the risk of the linear estimate $\widehat{x}_{H_*}(\cdot)$ induced by the optimal solution H_* to the problem (this solution clearly exists provided $\sigma > 0$) satisfies the risk bound

$$\mathrm{Risk}[\widehat{x}_{H_*}|\mathcal{X}] \le \sqrt{\mathrm{Opt}}.$$

♠ **Note:** We shall see eventually that *in the case of $\xi \sim \mathcal{N}(0, I_m)$, Opt is "nearly" the same as the ideal minimax risk*

$$\mathrm{RiskOpt} = \inf_{\widehat{x}(\cdot)} \mathrm{Risk}[\widehat{x}|\mathcal{X}],$$

where $\inf$ is taken w.r.t. all, not necessarily linear, estimates $\widehat{x}(\cdot)$.

How It Works: Inverse Heat Equation

♣ **Situation:** Square plate is heated at time 0 and is rest to cool; the temperature at the plate's boundary is all the time is kept 0.
Given given noisy measurements, taken along m points, of plate's temperature at time t_1, we want to recover distribution of temperature at a given time t_0, $0 < t_0 < t_1$.
♠ **The model:** The temperature field $u(t;p,q)$ evolves according to *Heat Equation*

$$\frac{\partial}{\partial t}u(t;p,q) = \left[\frac{\partial^2}{\partial p^2} + \frac{\partial^2}{\partial q^2}\right] u(t;p,q),\ t \geq 0, (p,q) \in S$$

• t: time • $S = \{(p,q), -1 \leq p,q \leq 1\}$: the plate

with boundary conditions $u(t;p,q)\Big|_{(p,q)\in\partial S} \equiv 0$.

♡ It is convenient to represent $u(t;p,q)$ by its expansion

$$u(t;p,q) = \sum_{k,\ell} x_{k\ell}(t)\phi_k(p)\phi_\ell(q), \qquad (*)$$

$$\phi_k(s) = \begin{cases} \cos(\omega_{2i-1}s), \omega_{2i-1} = (i-1/2)\pi & k = 2i-1 \\ \sin(\omega_{2i}s), \omega_{2i} = i\pi & k = 2i \end{cases}$$

Note: $\phi_k(s)$ are harmonic oscillations vanishing at $s = \pm 1$.

$$u(t;p,q)=\sum_{k,\ell}x_{k\ell}(t)\phi_k(p)\phi_\ell(q), \qquad (*)$$

$$\phi_k(s)=\begin{cases}\cos(\omega_{2i-1}s),\omega_{2i-1}=(i-1/2)\pi & k=2i-1\\ \sin(\omega_{2i}s),\omega_{2i}=i\pi & k=2i\end{cases}$$

Note:

- $\{\phi_{k\ell}(p,q)=\phi_k(p)\phi_\ell(q)\}_{k,\ell}$ form an orthonormal basis in $L_2(S)$
- $\phi_{k\ell}(\cdot)$ meet the boundary conditions

$$\phi_{k\ell}(p,q)\Big|_{(p,q)\in\partial S}=0$$

- in terms of the coefficients $x_{k\ell}(t)$, the Heat Equation becomes

$$\frac{d}{dt}x_{k\ell}(t)=-[\omega_k^2+\omega_\ell^2]x_{k\ell}(t)\Rightarrow x_{k\ell}(t)=\mathrm{e}^{-[\omega_k^2+\omega_\ell^2]t}x_{k\ell}(0).$$

♡ We select integer discretization parameter N and

- restrict $(*)$ to terms with $1\leq k,\ell\leq 2N-1$
- discretize the spatial variable (p,q) to reside in the grid

$$G_N=\{P_{ij}=(p_i,p_j)=(\frac{i}{N}-1,\frac{j}{N}-1),\ 1\leq i,j\leq 2N-1\}$$

Note: Restricting functions $\phi_{k\ell}(\cdot,\cdot)$, $1\leq k,\ell\leq 2N-1$ on grid G_N, we get orthogonal basis in $\mathbb{R}^{(2N-1)\times(2N-1)}$.

♠ We arrive at the model as follows:

- The signal x underlying observation is

$$x=\{x_{k\ell}:=x_{k\ell}(t_0),1\leq k,\ell\leq 2N-1\}\in\mathbb{R}^{(2N-1)\times(2N-1)}$$

- The observation is

$$\omega=A(x)+\sigma\xi\in\mathbb{R}^m,\ \xi\sim\mathcal{N}(0,I_m)$$
$$[A(x)]_\nu=\sum_{k,\ell=1}^{2N-1}x_{k\ell}\mathrm{e}^{-[\omega_k^2+\omega_\ell^2][t_1-t_0]}\phi_k(p_{i(\nu)})\phi_\ell(p_{j(\nu)})x_{k\ell}$$

- $(p_{i(\nu)},p_{j(\nu)})\in S$, $1\leq\nu\leq m$: measurement points

- We want to recover the restriction $B(x)$ of $u(t_0;p,q)$ to some grid, say, square grid

$$G_K=\{(r_i=\frac{i}{K}-1,r_j=\frac{j}{K}-1),\ 1\leq i,j\leq 2K-1\}\subset S,$$

resulting in

$$[B(x)]_{ij}=\sum_{k,\ell=1}^{2N-1}\phi_k(r_i)\phi_\ell(r_j)x_{k\ell}$$

• We assume that the initial distribution of temperatures $[u(0;p_i,p_j)]_{i,j=1}^{2N-1}$ satisfies $\|u\|_2 \leq R$, for some given R, implying that x resides in the ellitope, namely, the ellipsoid

$$\mathcal{X} = \left\{ \{x_{k\ell}\} \in \mathbb{R}^{(2N-1)\times(2N-1)} : \sum_{k,\ell} \left[\mathrm{e}^{[\omega_k^2+\omega_\ell^2]t_0} x_{k\ell} \right]^2 \leq R^2 \right\}$$

$$u(t;p_i,p_j) = \sum_{k,\ell} \mathrm{e}^{-[\omega_k^2+\omega_\ell^2][t-t_0]}\phi_k(p_i)\phi_\ell(p_j)x_{k\ell}$$
$$[A(x)]_\nu = \sum_{k,\ell=1}^{2N-1} x_{k\ell}\mathrm{e}^{-[\omega_k^2+\omega_\ell^2][t_1-t_0]}\phi_k(p_{i(\nu)})\phi_\ell(p_{j(\nu)})x_{k\ell}$$

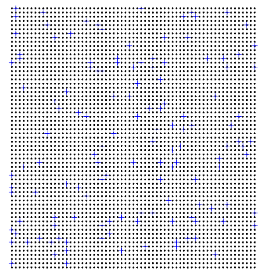
♣ **Bad news:** *Contributions of high frequency* (with large $\omega_k^2+\omega_\ell^2$) signal components $x_{k\ell}$ to $A(x)$ decrease exponentially fast with high decay rate as t_1-t_0 grows
$\Rightarrow$ *High frequency components $x_{k\ell}$ are impossible to recover from observations at time t_1, unless t_1 is very small.*

$$\mathcal{X} = \left\{\{x_{k\ell}\} : \sum_{k,\ell}\left[\mathrm{e}^{[\omega_k^2+\omega_\ell^2]t_0}x_{k\ell}\right]^2 \le R^2\right\}$$
$$[B(x)]_{ij} = \sum_{k,\ell=1}^{2N-1}\phi_k(r_i)\phi_\ell(r_j)x_{k\ell}$$

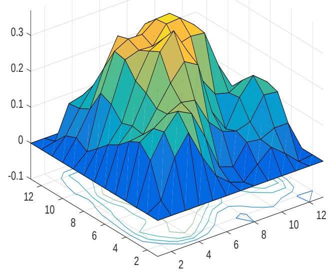
♣ **Good news:** *High frequency components $x_{k\ell}$ of $x\in\mathcal{X}$ are very small, provided t_0 is not too small*
$\Rightarrow$ *There is no necessity to recover well high frequency components of signal from observations!*

♣ **Numerical results** $N = 32$, $m = 125$, $K = 6$, $t_0 = 0.01$, $t_1 = 0.02$, $\sigma = 0.001$, $R = 15$

♡ *Minimax risk of optimal linear estimate: 0.1707*

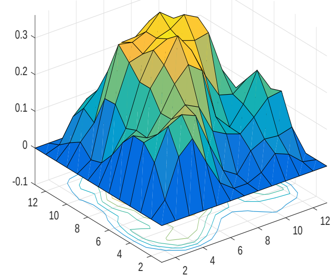


63×63 grid G_{63} and $m = 125$ measurement points



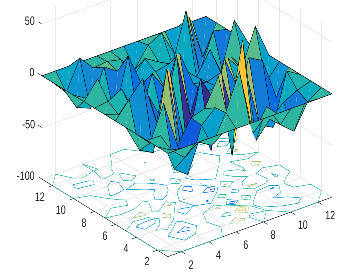
b

$$\|b\|_2 = 2.13 \qquad \|b\|_\infty = 0.43$$



$\widehat{b}$

$$\|\widehat{b} - b\|_2 = 0.15 \qquad \|\widehat{b} - b\|_\infty = 0.03$$



$\widetilde{b}$

$$\|\widetilde{b} - b\|_2 = 7.99 \qquad \|\widetilde{b} - b\|_\infty = 1.80$$

Sample results

- left: $b = B(x)$
- center: sample optimal linear recovery $\widehat{b} = H_*^T \omega$ of $b = B(x)$
- right: naive recovery $\widetilde{b} = B(\widetilde{x})$, $\widetilde{x}$: Least Squares solution to $A(x) = \omega$

How It Works: Denoising & Deblurring Images

- A grayscale image can be thought of as 2D $m \times n$ array $[x_{ij}]_{\substack{0\leq i<m,\\ 0\leq j<n}}$ with entries (pixels' intensities) in $[0,255]$
- Taking picture can be modeled as observing noisy convolution

$$\omega_{ij} = \underbrace{\sum_{\substack{0\leq p<\mu,\\ 0\leq q<\nu}} \kappa_{pq} x_{i-p,j-q}}_{\kappa\star x} + \xi_{ij},\ 0 \leq i < m+\mu-1, 0 \leq j < n+\nu-1 \qquad (*)$$

$$\left[\xi_{ij} \sim \mathcal{N}(0,\sigma^2) \text{ independent across } i,j\right]$$

of the image and a given *blurring kernel* $[\kappa_{pq}]_{\substack{0\leq p<\mu,\\ 0\leq q<\nu}}$.

Note: In $(*)$, $x_{ij} = 0$ outside of the actual range $\{0 \leq i < m, 0 \leq j < n\}$ of i,j.

Note: "Centering" image – subtracting from x_{ij} entries in x the midpoint S of the range $[0,255]$ of pixels' intensities and updating ω_{ij} accordingly, the images become 2D arrays from the box

$$\mathcal{X}_\infty = \{x \in \mathbb{R}^{m\times n} : |x_{ij}| \leq S\},$$

and the recovery problem falls into our framework.

$$x \mapsto \kappa \star x + \xi \;?? \Rightarrow ??\; \widehat{x} \approx x$$

Bad news: Linear dimensions mn of typical images are in the range of 10^5–10^6, making straightforward design of linear estimates $\omega \mapsto \widehat{x} = H^T\omega$ intractable—linear dimensions of contrast matrices should be in the range of 10^{10}–10^{12}.

Good news: Extending x and κ to $M := [m+\mu] \times N := [n+\nu]$ arrays x^+, κ^+ by adding zero entries to x, κ, and passing to 2D Discrete Fourier Transforms $\chi = \mathcal{F}x^+$, $\theta = \mathcal{F}\kappa^+$ of these arrays, observation scheme becomes extremely simple:

$$\zeta := \mathcal{F}\omega = \theta \bullet \chi + \sqrt{MN}\sigma\eta$$

[$\bullet$: entrywise product; η : (complex-valued) white Gaussian noise with unit covariance matrix]

Note: DFT multiplies $\|\cdot\|_2$ by $\sqrt{MN}$ $\Rightarrow$ *when the recovery error is measured in* $\|\cdot\|_2$, *recovering* x *from* ω *is equivalent to recovering* χ *from* ζ

Besides this, *when a priori information on* x *translates into simple constraints on* χ, *like*

$$\sum_{r,s} \beta_{rs}|\chi_{rs}|^2 \leq \beta \text{ and/or } |\chi_{rs}| \leq \gamma_{rs}\ \forall r,s,\ 0 \leq r < M, 0 \leq s < N \qquad (!)$$

frequency representations χ *of signals of interest become points of a simple (complexified) ellitope, and sensing matrix* A *becomes (complex-valued) diagonal*

$\Rightarrow$ *Working in frequency domain, we lose nothing when looking for linear estimates with diagonal (complex-valued) contrast matrices.*

Moreover, *when the number of constraints* $(!)$ *is small, designing the best linear estimate with diagonal contrast matrix reduces to solving a low-dimensional convex problem and takes few seconds even when* MN *is in the range of millions.*

... in frequency domain recovery problem becomes $\zeta = \theta \bullet \chi + \sqrt{MN}\sigma\eta$?? $\Rightarrow$?? $\widehat{\chi} \approx \chi$, and easy-to-utilize a priori information on χ are constraints of the form

$$\sum_{r,s} \beta_{rs}|\chi_{rs}|^2 \leq \beta \text{ and/or } |\chi_{rs}| \leq \gamma_{rs}\ \forall r,s,\ 0 \leq r < M, 0 \leq s < N \qquad (!)$$

Note: Our "built in" box constraint $\|x\|_\infty \leq L$ does *not* translate into a simple constraint on χ; the best simple (conservative!) frequency domain version of this constraint is the bound

$$\|\chi\|_2 \leq \sqrt{MN} \cdot \sqrt{mn} L \qquad (E)$$

on the $\|\cdot\|_2$-norm of χ.

Similarly, the standard in Image Reconstruction bounds

$$\mathrm{TV}(x) := \sum_{i,j} |x_{i+1,j} - x_{i,j}| + \sum_{i,j} |x_{i,j+1} - x_{i,j}| \leq U$$

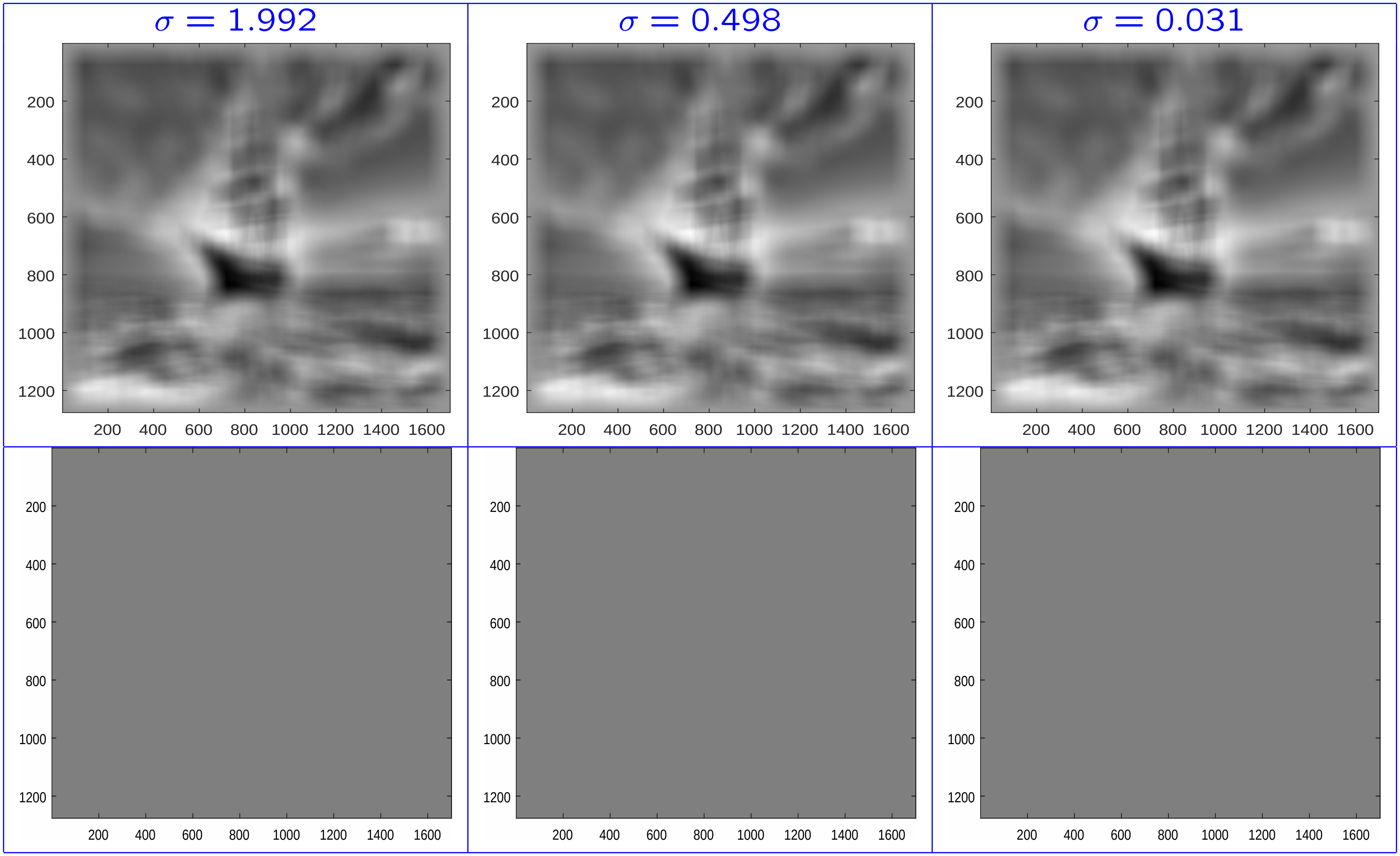
on *Total Variation* of x do *not* translate into simple constraints on χ.

• **However:** we can impose on χ, in addition to (E), *empirical* upper bounds on $\|\chi\|_\infty$ and $\|\chi\|_1$ by inspecting a "representative library" of images.

• **Warning:** When the blur is present (i.e., κ is not a δ-function), the recovery problem can easily become ill-posed, since convolution can "kill" come frequencies (formally: some of the entries in θ can be very small).

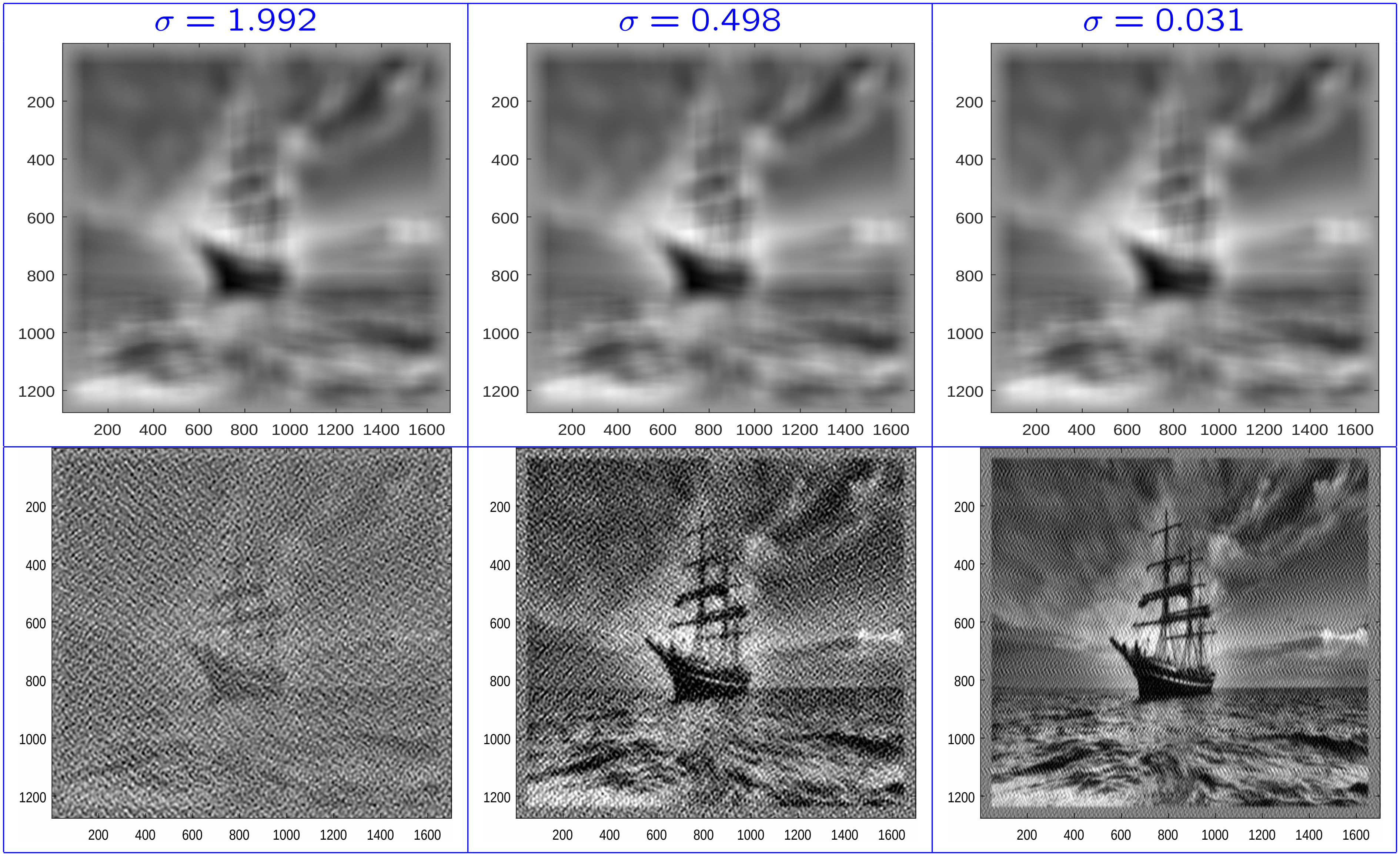
Blurred noisy observations (top) and recoveries (bottom) of 1200×1600 image, ill-posed case

[with bound on signal's energy]

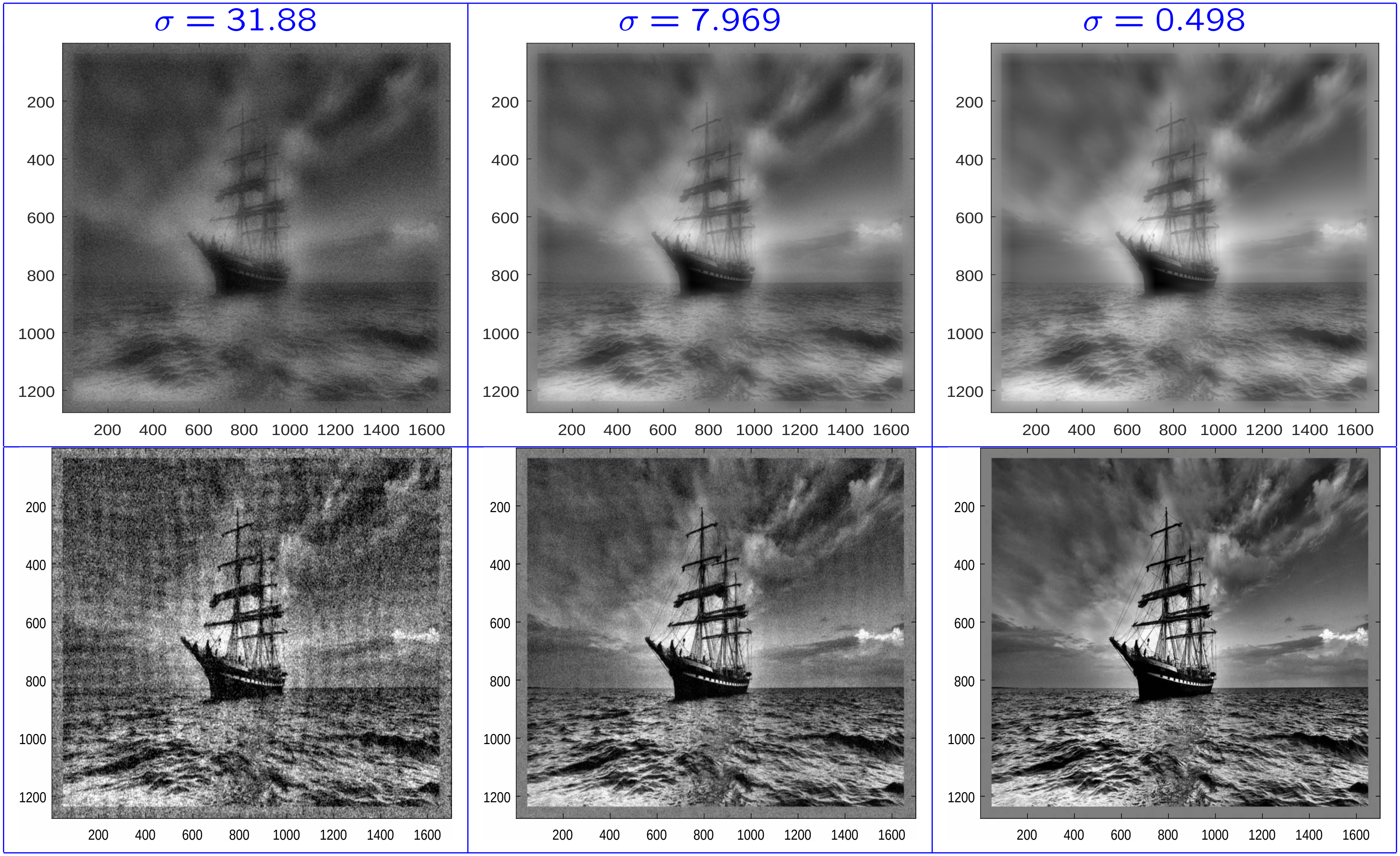


Blurred noisy observations (top) and recoveries (bottom) of 1200×1600 image, ill-posed case

[with rudimentary form of Total Variation constraints]



Blurred noisy observations (top) and recoveries (bottom) of 1200×1600 image, well-posed case



Byproduct on Semidefinite Relaxation

♠ **Theorem** *Let C be a symmetric $n \times n$ matrix and $\mathcal{X}$ be an ellitope:*

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists (t \in \mathcal{T}, y) : x = Fy, y^T S_k y \le t_k \; \forall k \le K\}.$$

Then the efficiently computable quantity

$$\mathrm{Opt} = \min_\lambda \left\{ \phi_{\mathcal{T}}(\lambda) : \lambda \ge 0, F^T C F \preceq \sum_k \lambda_k S_k \right\}$$
$$\left[\phi_{\mathcal{T}}(\lambda) = \max_{t \in \mathcal{T}} \lambda^T t\right]$$

is a tight upper bound on

$$\mathrm{Opt}_* = \max_{x \in \mathcal{X}} x^T C x :$$

namely,

$$\mathrm{Opt}_* \le \mathrm{Opt} \le 3 \ln(\sqrt{3} K) \mathrm{Opt}_*.$$

Note: Opt_* is difficult to compute within 4% accuracy when $\mathcal{X}$ is as simple as the unit box in $\mathbb{R}^n$.

♣ Let $\mathcal{X}$ be given by quadratic inequalities:

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \le t_k, k \le K\} \ne \emptyset$$
$[\mathcal{T}$: nonempty convex compact set$]$

We have

$$\mathrm{Opt}_* := \max_{x\in\mathcal{X}} x^T C x \le \mathrm{Opt} := \min_\lambda\{\phi_{\mathcal{T}}(\lambda) : \lambda \ge 0,\ C \preceq \sum_k \lambda_k S_k\} \le \Theta \cdot \mathrm{Opt}_*$$

What can be said about tightness factor Θ ?

Facts:

A. *Assuming $K = 1$ and Slater condition: $\bar{x}^T S_1 \bar{x} < t$ for some $\bar{x}$ and* some $t \in \mathcal{T}$, one can set $\Theta = 1$.

[famous $\mathcal{S}$**-Lemma**]

B. *Assuming that $x^T S_k x = x_k^2$, $k \le K = \dim x$, $\mathcal{T} = [0;1]^K$, and C is Laplacian of a graph* (i.e., off-diagonal entries in C are nonpositive and all row sums are zero), *one can set* $\Theta = 1.1382...$

[MAXCUT Theorem of Goemans and Williamson, 1995]

Note: Laplacian of a graph always is $\succeq 0$

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \leq t_k, k \leq K\} \neq \emptyset$$

[$\mathcal{T}$: nonempty convex compact set]

$$\Rightarrow \mathrm{Opt}_* := \max_{x \in \mathcal{X}} x^T C x \leq \mathrm{Opt} := \min_\lambda \{\phi_{\mathcal{T}}(\lambda) : \lambda \geq 0,\, C \preceq \sum_k \lambda_k S_k\} \leq \Theta \cdot \mathrm{Opt}_*$$

C. *Assuming that $C \succeq 0$ and all matrices S_k are diagonal, one can set* $\Theta = \frac{\pi}{2} = 1.5708...$

[$\frac{\pi}{2}$ Theorem, Nesterov, 1998]

D. *Assuming $\mathcal{X}$ is an ellitope* (i.e., $S_k \succeq 0, \sum_k S_k \succ 0$ and $\mathcal{T}$ contains a positive vector), *one can set* $\Theta = 3\ln(\sqrt{3}K)$

Note: In the case of $\mathbf{D}$, Θ indeed can be as large as $O(\ln(K))$

♠ A byproduct of Theorem is the following useful fact:

Theorem [upper-bounding of operator norms] *Let $\|\cdot\|_x$ be a norm on $\mathbb{R}^N$ such that the unit ball $\mathcal{X}$ of the norm is an ellitope:*

$$\mathcal{X} \;:=\; \{x : \|x\|_x \leq 1\} = \{x : \exists (t \in \mathcal{T}, y) : x = Py, y^T S_k y \leq t_k, k \leq K\}$$

For example, $\|\cdot\|_x = \|\cdot\|_p$ with $2 \leq p \leq \infty$

Let, further, $\|\cdot\|$ be a norm on $\mathbb{R}^M$ such that the unit ball $\mathcal{B}_*$ of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$ is an ellitope:

$$\mathcal{B}_* \;:=\; \{u \in \mathbb{R}^m : u^T v \leq 1\, \forall (v, \|v\| \leq 1)\} = \{u : \exists (r \in \mathcal{R}, z) : u = Qz, z^T R_\ell z \leq r_\ell, \ell \leq L\}$$

For example, $\|\cdot\| = \|\cdot\|_r$ with $1 \leq r \leq 2$.

Then the efficiently computable quantity

$$\mathrm{Opt}(C) \;=\; \min_{\lambda,\mu}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) :\; \lambda \geq 0, \mu \geq 0 \left[\begin{array}{c|c} \sum_\ell \mu_\ell R_\ell & \frac{1}{2} Q^T C P \\ \hline \frac{1}{2} P^T C^T Q & \sum_k \lambda_k S_k \end{array}\right] \succeq 0 \right\}$$
$$\left[C \in \mathbb{R}^{M\times N}\right]$$

is a convex in C upper bound on the operator norm

$$\|C\|_{\|\cdot\|_x \to \|\cdot\|} = \max_x \{\|Cx\| : \|x\|_x \leq 1\}$$

of the mapping $x \mapsto Cx$, and this bound is reasonably tight:

$$\|C\|_{\|\cdot\|_x \to \|\cdot\|} \leq \mathrm{Opt}(C) \leq 3\ln(\sqrt{3}(K+L))\|C\|_{\|\cdot\|_x \to \|\cdot\|}.$$

Indeed, the operator norm in question is the maximum of a quadratic form over an ellitope:

$$\|z\| = \max_u \left\{u^T z : u \in \mathcal{B}_*\right\}$$

$\Rightarrow$

$$\|C\|_{\|\cdot\|_x \to \|\cdot\|} = \max\left\{u^T C x : x \in \mathcal{X}, u \in \mathcal{B}_*\right\}$$

$\Rightarrow$

$$\begin{array}{rcl}\|C\|_{\|\cdot\|_x \to \|\cdot\|} & = & \frac{1}{2} \max\limits_{x \in \mathcal{X}, u \in \mathcal{B}_*} [x; u]^T \left[\begin{array}{c|c} & C \\ \hline C^T & \end{array}\right] [x; u] \\ & = & \frac{1}{2} \max\limits_{[y;z] \in \mathcal{W}} [y; z]^T \left[\begin{array}{c|c} & Q^T C P \\ \hline P^T C^T Q & \end{array}\right] [y; z]\end{array}$$

where $\mathcal{W}$ is the basic ellitope given by

$$\mathcal{W} = \left\{[y; z] : \exists [t; r] \in \mathcal{T} \times \mathcal{R} : \begin{array}{l} y^T S_k y \leq t_k, k \leq K \\ z^T R_\ell z \leq r_\ell, \ell \leq L \end{array}\right\}.$$

What is inside

♠ In the above results on tightness of semidefinite relaxation, we speak about tightness of the Semidefinite Relaxation upper bound on the maximum of a quadratic form over an ellitope:

$$\mathrm{Opt}_* = \max_{x,t}\left\{x^TCx : x^TS_kx \leq t_k, k \leq K, t \in \mathcal{T}\right\} \qquad (*)$$

♠ **Fact:** Semidefinite relaxation admits an alternative description as follows:
Let us associate with $(*)$ another optimization problem where instead of deterministic candidate solutions (x,t) we are looking for random solutions (ξ,τ) satisfying the constraints at average:

$$\mathrm{Opt}^+ = \max_{\xi,\tau}\left\{\mathbf{E}\{\xi^TC\xi\} : \begin{array}{l}\mathbf{E}\{\xi^TS_k\xi\} \leq \mathbf{E}\{\tau_k\}\\ \mathbf{E}\{\tau\} \in \mathcal{T}\end{array}\right\} \qquad (!)$$

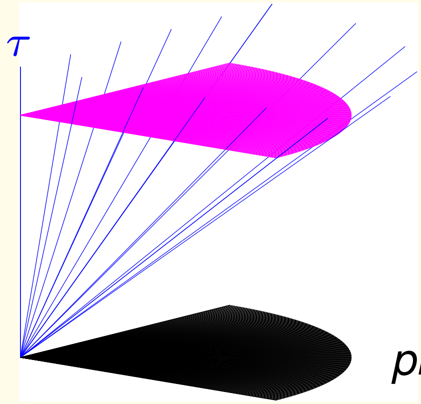
• **Immediate observation:** *Property of a random solution (ξ,τ) to be feasible for $(!)$ depends solely on the matrix $Q = \mathbf{E}\{\xi\xi^T\}$ and the vector $t = \mathbf{E}\{\tau\}$, so that*

$$\mathrm{Opt}^+ \;=\; \max_{Q,t}\left\{\mathrm{Tr}(CQ) : \begin{array}{l}\mathrm{Tr}(S_kQ) \leq t_k\\ Q \succeq 0, t \in \mathcal{T}\end{array}\right\} \qquad (\#)$$

$$\text{Opt}^+ = \max_{Q,t} \left\{ \text{Tr}(CQ) : \begin{array}{l} \text{Tr}(S_k Q) \le t_k \\ Q \succeq 0, t \in \mathcal{T} \end{array} \right\} \quad (\#)$$

Note: (#) is not a conic problem, the obstacle being the constraint $t \in \mathcal{T}$. We can easily make this constraint conic.

• Let $\mathcal{T}^+ = \{[t; 1] \in \mathbb{R}^{K+1} : t \in \mathcal{T}\}$, and let $\mathbf{T} \in \mathbb{R}^{K+1}$ be the set of nonnegative multiples of vectors from $\mathcal{T}^+$:



Sets $\mathcal{T}$, $\mathcal{T}^+$ and cone $\mathbf{T}$

- $\mathbf{T}$ is a regular cone (since $\mathcal{T}$ is a convex compact set with a nonempty interior)
- $\mathcal{T} = \{t : [t; 1] \in \mathbf{T}\}$
- The cone $\mathbf{T}_*$ dual to $\mathbf{T}$ is $\{[y; s] \in \mathbb{R}^{K+1} : s \ge \phi_{\mathcal{T}}(-y)\}$

Indeed, $\{[y; s] \in \mathbf{T}_*\} \Leftrightarrow \{y^T t + s\tau \ge 0 \,\forall [t; \tau] \in \mathbf{T}\}$

$\Leftrightarrow \{y^T t + s \ge 0 \,\forall t : [t; 1] \in \mathbf{T}\} \Leftrightarrow s \ge -y^T t \,\forall t \in \mathcal{T}\}$

$\Leftrightarrow s \ge \max_{t \in \mathcal{T}} [-y]^T t$

$\Leftrightarrow \{s \ge \phi_{\mathcal{T}}(-y)\}$

$$
\begin{array}{rcl}
\mathrm{Opt}_* & = & \max\limits_{x,t} \left\{ x^T C x : \exists (t \in \mathcal{T}) : x^T S_k x \le t_k \right\} \quad (*) \\
\mathrm{Opt}^+ & = & \max\limits_{Q,t} \left\{ \mathrm{Tr}(CQ) : \begin{array}{l} \mathrm{Tr}(S_k Q) \le t_k \\ Q \succeq 0, [t;1] \in \mathbf{T} \end{array} \right\} \quad (\#) \\
\end{array}
$$

$$
[\mathbf{T}_* = \{[y;s] : s \ge \phi_{\mathcal{T}}(-y)\}]
$$

♠ **Note:** $(\#)$ is strictly feasible and bounded, and *the problem*

$$
\mathrm{Opt} = \min_{\lambda} \left\{ \phi_{\mathcal{T}}(\lambda) : \lambda \ge 0, C \preceq \sum_k \lambda_k S_k \right\}
$$

specifying Semidefinite relaxation upper bound on Opt *is is nothing but the conic dual to* $(\#) \Rightarrow \mathrm{Opt}^+ = \mathrm{Opt}$.

• $(\#)$ suggests the following recipe for quantifying the conservatism of the upper bound Opt on Opt_*:

— *Find an optimal solution* Q_*, t_* *to* $(\#)$ *and treat* $Q_* \succeq 0$ *as the covariance matrix of random vector* ξ (many options!)

— *Random solutions* (ξ, t_*) *satisfy* $(*)$ *"at average." Try to "correct" them to get feasible solutions to* $(*)$ *and look how "costly" this correction is in terms of the objective.*

$$\text{Opt}^+ = \max_{Q,t} \left\{ \text{Tr}(CQ) : \begin{array}{l} \text{Tr}(S_k Q) \leq t_k \\ Q \succeq 0, [t;1] \in \mathbf{T} \end{array} \right\} \qquad (\#)$$

For example, in Goemans-Williamson MAXCUT and in Nesterov's $\pi/2$ Theorems, where $x^T Cx$ is maximized over the unit box

$$\mathcal{X} = \{\|x\|_\infty \leq 1\} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} := [0,1]^n : x_k^2 \leq t_k, k \leq n\},$$

that is, $\mathbf{T} = \{[t;\tau] : 0 \leq t_k \leq \tau,\, k \leq n\}$, $(\#)$ reads

$$\text{Opt}^+ = \max_{Q,t} \left\{ \text{Tr}(CQ) : \begin{array}{l} \text{Tr}(Q_{kk}) \leq t_k,\, k \leq K = n \\ Q \succeq 0, [t;1] \in \mathbf{T} = \{t : 0 \leq t_k \leq 1,\, k \leq n\} \end{array} \right\} \quad (\#)$$

one selects $\xi \sim \mathcal{N}(0, Q_*)$ and "corrects" ξ according to $\xi \mapsto \text{sign}[\xi]$.

$$
\begin{array}{rcll}
\mathrm{Opt}_* & = & \max\limits_{x,t}\left\{x^TCx : \exists(t\in\mathcal{T}) : x^TS_kx \le t_k\right\} & (*)\\
\mathrm{Opt}^+ & = & \max\limits_{Q,t}\left\{\mathrm{Tr}(CQ) : \begin{array}{l}\mathrm{Tr}(S_kQ)\le t_k\\ Q\succeq 0, [t;1]\in\mathbf{T}\end{array}\right\} & (\#)
\end{array}
$$

♠ This is how the above recipe works in the general ellitopic case:

A. Let (Q_*, t^*) be an optimal solution to $(\#)$. Set

$$\bar{C} := Q_*^{1/2}CQ_*^{1/2} = UDU^T$$

(U is orthogonal, D is diagonal).

B. Let $\xi = Q_*^{1/2}U\zeta$ with Rademacher random ζ (ζ_i take values ± 1 with probability $1/2$ and are independent across i), so that

$$\mathbf{E}\{\xi\xi^T\} = \mathbf{E}\{Q_*^{1/2}U\zeta\zeta^TU^TQ_*^{1/2}\} = Q_*^{1/2}U\underbrace{\mathbf{E}\{\zeta\zeta^T\}}_{I}U^TQ_*^{1/2} = Q_*.$$

$$\bar{C} := Q_*^{1/2} C Q_*^{1/2} = UDU^T, \quad \xi = Q_*^{1/2} U\zeta$$

Note that

$$\begin{array}{rcl}
\xi^T C\xi & = & \zeta^T U^T [Q_*^{1/2} C Q_*^{1/2}] U\zeta = \zeta^T D\zeta \\
& \equiv & \mathrm{Tr}(D) = \mathrm{Tr}(Q_*^{1/2} C Q_*^{1/2}) \equiv \mathrm{Tr}(CQ_*) \\
& = & \mathrm{Opt}, \\
\mathbf{E}\{\xi^T S_k \xi\} & = & \mathbf{E}\{\zeta^T U^T Q_*^{1/2} S_k Q_*^{1/2} U\zeta\} \\
& = & \mathrm{Tr}(U^T Q_*^{1/2} S_k Q_*^{1/2} U) \\
& = & \mathrm{Tr}(Q_*^{1/2} S_k Q_*^{1/2}) = \mathrm{Tr}(S_k Q_*) \\
& \leq & t_k^*, \; k \leq K
\end{array}$$

$$\begin{array}{rcll} \xi^T C\xi & \equiv & \mathrm{Opt} & (a)\\ \mathbf{E}\{\xi^T S_k\xi\} & \leq & t_k^*, k\leq K & (b)\end{array}$$

C. Since $S_k \succeq 0$ and ξ is "light-tail" (it comes from Rademacher random vector), simple bounds on probabilities of large deviations combine with (b) to imply that

$$\begin{array}{l}\forall(\gamma\geq 0, k\leq K):\\ \mathrm{Prob}\{\xi : \xi^T S_k\xi > \gamma t_k^*\} \leq O(1)\exp\{-O(1)\gamma\}\end{array}$$

$\Rightarrow$ with $\gamma_* = O(1)\ln(K+1)$, there exists a realization $\widehat{\xi}$ of ξ such that $\widehat{\xi}^T S_k\widehat{\xi} \leq \gamma_* t_k^*$, $k\leq K$

$\Rightarrow$ $(\xi^* = \widehat{\xi}/\sqrt{\gamma_*}, t^*)$ is feasible for

$$\mathrm{Opt}_* \;=\; \max_{x,t}\left\{x^T Cx : \exists(t\in\mathcal{T}) : x^T S_k x\leq t_k\right\} \quad (*)$$

$\Rightarrow$ $\mathrm{Opt}_* \geq \widehat{\xi}^T C\widehat{\xi}/\gamma_* = \mathrm{Opt}/\gamma_*$ (look at (a)!)

♠ "Simple bounds on probabilities of large deviations" stem from the following

Mini-Lemma: *Let P be positive semidefinite $N \times N$ matrix with trace ≤ 1 and ζ be N-dimensional Rademacher random vector. Then*

$$\mathbf{E}\left\{\exp\left\{\zeta^T P\zeta/3\right\}\right\} \leq \sqrt{3}.$$

♠ **Mini-Lemma $\Rightarrow$ bounds:** We have

$$\xi^T S_k \xi = \zeta^T \underbrace{U^T Q_*^{1/2} S_k Q_*^{1/2} U}_{t_k^* P_k} \zeta$$

and $\mathrm{Tr}(P_k) = \mathrm{Tr}(Q_*^{1/2} S_k Q_*^{1/2})/t_k^* = \mathrm{Tr}(S_k Q_*)/t_k^* \leq 1$

$\Rightarrow$ [Mini-Lemma] $\mathbf{E}\left\{\exp\left\{\zeta^T P_k\zeta/3\right\}\right\} \leq \sqrt{3}$

$\Rightarrow \mathrm{Prob}\{\zeta^T P_k \zeta > 3\rho\} \leq \sqrt{3}\mathrm{e}^{-\rho}$

$\Rightarrow \mathrm{Prob}\{\xi^T S_k \xi > \gamma t_k^*\} = \mathrm{Prob}\{\zeta^T P_k \zeta > \gamma\} \leq \sqrt{3}\mathrm{e}^{-\gamma/3}$.

Proof of Mini-Lemma: Let $P = \sum_i \sigma_i f_i f_i^T$ be the eigenvalue decomposition of P, so that $f_i^T f_i = 1$, $\sigma_i \geq 0$, and $\sum_i \sigma_i \leq 1$. The function

$$f(\sigma_1, ..., \sigma_N) = \mathbf{E}\left\{\mathrm{e}^{\frac{1}{3}\sum_i \sigma_i \zeta^T f_i f_i^T \zeta}\right\}$$

is convex on the simplex $\{\sigma \geq 0, \sum_i \sigma_i \leq 1\}$ and thus attains it maximum over the simplex at a vertex, implying that for some $h = f_i$, $h^T h = 1$, it holds

$$\mathbf{E}\{\mathrm{e}^{\frac{1}{3}\zeta^T P \zeta}\} \leq \mathbf{E}\{\mathrm{e}^{\frac{1}{3}(h^T\zeta)^2}\}.$$

Let $\xi \sim \mathcal{N}(0,1)$ be independent of ζ. We have

$$\begin{array}{l}
\mathbf{E}_\zeta\left\{\exp\{\frac{1}{3}(h^T\zeta)^2\}\right\} = \mathbf{E}_\zeta\left\{\mathbf{E}_\xi\left\{\exp\{[\sqrt{2/3}h^T\zeta]\xi\}\right\}\right\} \\
= \mathbf{E}_\xi\left\{\mathbf{E}_\zeta\left\{\exp\{[\sqrt{2/3}h^T\zeta]\xi\}\right\}\right\} \\
= \mathbf{E}_\xi\left\{\prod_{s=1}^N \mathbf{E}_\zeta\left\{\exp\{\sqrt{2/3}\xi h_s \zeta_s\}\right\}\right\} \\
= \mathbf{E}_\xi\left\{\prod_{s=1}^N \cosh(\sqrt{2/3}\xi h_s)\right\} \leq \mathbf{E}_\xi\left\{\prod_{s=1}^N \exp\{\xi^2 h_s^2/3\}\right\} \\
= \mathbf{E}_\xi\left\{\exp\{\xi^2/3\}\right\} = \sqrt{3}
\end{array}$$

$\square$

Extensions

♣ So far, we have considered a problem of recovering Bx from observation

$$\omega = Ax + \xi \in \mathbb{R}^m$$

where

• x is unknown signal known to belong to a given basic ellitope

$$\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \leq t_k, k \leq K\}$$

Note: Assuming signal set $\mathcal{X}$ *basic* ellitope rather than ellitope is w.l.o.g.: when $\mathcal{X} = F\mathcal{Y}$ with basic ellitope $\mathcal{Y}$, we lose nothing when assuming that the signal is y rather than $x = Fy$ and replacing A, B with AF, BF.

• $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{\nu \times n}$ are given matrices

• $\xi \sim \mathcal{N}(0, \sigma^2 I_m)$ is observation noise

• (squared) risk of a candidate estimate is the worst-case, over $x \in \mathcal{X}$, expected squared $\|\cdot\|_2$-norm of recovery error:

$$\mathrm{Risk}^2[\widehat{x}|\mathcal{X}] = \sup_{x \in \mathcal{X}} \mathbf{E}\left\{\|Bx - \widehat{x}(Ax + \xi)\|_2^2\right\}.$$

♠ We are about to extend our results to the situation where

• Noise ξ not necessary is zero mean Gaussian; we allow the distribution P of noise to be unknown in advance and to depend on signal x.

♡ **Assumption:** *We are given a convex compact set* $\Pi \subset \mathrm{int}\,\mathbf{S}^m_+$ *such that the variance matrix of* P *admits an upper bound from* Π*:*

$$P \in \mathcal{P}[\Pi] := \left\{P : \exists Q \in \Pi : \mathrm{Var}[P] := \mathbf{E}_{\xi\sim P}\{\xi\xi^T\} \preceq Q\right\}$$

• We measure recovering error in a given norm $\|\cdot\|$, not necessarily the Euclidean one, and define risk of a candidate estimate $\widehat{x}(\cdot)$ as

$$\mathrm{Risk}_{\|\cdot\|,\Pi}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \sup_{P\in\mathcal{P}[\Pi]} \mathbf{E}_{\xi\sim P}\left\{\|Bx - \widehat{x}(Ax+\xi)\|\right\}$$

♡ **Assumption:** *The unit ball* $\mathcal{B}_*$ *of the norm conjugate to* $\|\cdot\|$ *is an ellitope:*

$$\|u\| = \max_{h\in\mathcal{B}_*} h^T u,$$
$$\mathcal{B}_* = \{h : \exists (y \in \mathbb{R}^M, r \in \mathcal{R}) : h = Fy,\ y^T R_\ell y \preceq r_\ell\ \forall \ell \leq L\}$$

$$\boxed{\begin{array}{c}\mathcal{X}=\{x\in\mathbb{R}^n: \exists t\in\mathcal{T}: x^TS_kx\leq t_k\ \forall k\leq K\}\\ \omega=Ax+\xi\ ??\Rightarrow??\ \widehat{x}_H(\omega)=H^T\omega\approx Bx\end{array}}$$

Building Presumably Good Linear Estimate

♣ We have

$$\begin{array}{rcl}\mathrm{Risk}_{\|\cdot\|,\Pi}[\widehat{x}_H|\mathcal{X}] &=& \sup\limits_{x\in\mathcal{X}}\sup\limits_{P\in\mathcal{P}[\Pi]}\mathbf{E}_{\xi\sim P}\left\{\|Bx-H^T[Ax+\xi]\|\right\}\\ &\leq& \sup\limits_{x\in\mathcal{X}}\sup\limits_{P\in\mathcal{P}[\Pi]}\mathbf{E}_{\xi\sim P}\left\{\|[B-H^TAx]\|+\|H^T\xi\|\right\}\\ &\leq& \Phi(H)+\Psi_\Pi[H],\\ && \Phi(H)=\max\limits_{x\in\mathcal{X}}\|[B-H^TA]x\|,\\ && \Psi_\Pi(H)=\sup\limits_{P\in\mathcal{P}[\Pi]}\mathbf{E}_{\xi\sim P}\left\{\|H^T\xi\|\right\}\end{array}$$

Next,

$$\mathcal{B}_* = \{u = My : y \in \mathcal{Y}\},$$
$$\mathcal{Y} = \{y : \exists r \in \mathcal{R} : y^T R_\ell y \le r_\ell \ \forall \ell \le L\}$$

whence

$$\begin{array}{rcl}
\Phi(H) & := & \max\limits_{x\in\mathcal{X}} \|[B - H^T A]x\| = \max\limits_{[u;x]\in\mathcal{B}_*\times\mathcal{X}} [u;x]^T \left[\begin{array}{c|c} & \frac{1}{2}[B - H^T A] \\ \hline \frac{1}{2}[B^T - A^T H] & \end{array}\right][u;x] \\
& = & \max\limits_{[y;x]\in\mathcal{Y}\times\mathcal{X}} [y;x]^T \left[\begin{array}{c|c} & \frac{1}{2}F^T[B - H^T A] \\ \hline \frac{1}{2}[B^T - A^T H]Fy & \end{array}\right][y;x] \\
& & \text{[semidefinite relaxation; note that } \mathcal{Y}\times\mathcal{X} \text{ is an ellitope]} \\
& \le & \overline{\Phi}(H) := \min\limits_{\lambda,\mu} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) : \begin{array}{l} \lambda \ge 0, \mu \ge 0, \\ \left[\begin{array}{c|c} \sum_\ell \mu_\ell R_\ell & \frac{1}{2}F^T[H^T A - B] \\ \hline \frac{1}{2}[A^T H - B^T]F & \sum_k \lambda_k S_k \end{array}\right] \succeq 0 \end{array} \right\} \\
& & \left[\phi_{\mathcal{T}}(\lambda) = \max\limits_{t\in\mathcal{T}} \lambda^T t, \phi_{\mathcal{R}}(\mu) = \max\limits_{r\in\mathcal{R}} \mu^T r\right]
\end{array}$$

$$
\begin{array}{c}
\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \le t_k, k \le K\}\\
\mathcal{B}_* = \{u = My : y \in \mathcal{Y}\}, \mathcal{Y} = \{y : \exists r \in \mathcal{R} : y^T R_\ell y \le r_\ell \ \forall \ell \le L\}\\
\omega = Ax + \xi \ \Rightarrow \ \widehat{x}_H(\omega) = H^T\omega \approx Bx\\
\Downarrow\\
\mathrm{Risk}_{\|\cdot\|,\Pi}[\widehat{x}_H|\mathcal{X}] \le \overline{\Phi}(H) + \Psi_\Pi(H),
\end{array}
$$

$$
\begin{array}{rcl}
\overline{\Phi}(H) &=& \min\limits_{\lambda,\mu}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) : \begin{array}{c}\lambda \ge 0, \mu \ge 0,\\ \left[\begin{array}{c|c}\sum_\ell \mu_\ell R_\ell & \frac{1}{2}M^T[H^TA - B]\\ \hline \frac{1}{2}[A^TH - B^T]M & \sum_k \lambda_k S_k\end{array}\right] \succeq 0\end{array}\right\}\\
\Psi_\Pi(H) &=& \sup\limits_{P\in\mathcal{P}[\Pi]} \mathbf{E}_{\xi\sim P}\left\{\|H^T\xi\|\right\}
\end{array}
$$

♣ **Lemma:** *One has*

$$
\Psi_\Pi(H) \le \overline{\Psi}_\Pi(H) := \min_{\Theta,\varkappa}\left\{\Gamma_\Pi(\Theta) + \phi_{\mathcal{R}}(\varkappa) : \begin{array}{c}\varkappa \ge 0\\ \left[\begin{array}{c|c}\sum_\ell \varkappa_\ell R_\ell & \frac{1}{2}M^TH^T\\ \hline \frac{1}{2}HM & \Theta\end{array}\right] \succeq 0\end{array}\right\}
$$

$$
\Gamma_\Pi(\Theta) = \max_{Q\in\Pi} \mathrm{Tr}(Q\Theta).
$$

Lemma:

$$\boxed{\begin{array}{c}\|z\| = \max_y \left\{z^T My : \exists r \in \mathcal{R} : y^T R_\ell y \le r_\ell, \ell \le L\right\}\\ \Gamma_\Pi(\Theta) = \max\limits_{Q\in\Pi} \mathrm{Tr}(Q\Theta).\end{array}}$$

$$\Downarrow$$

$$\boxed{\Psi_\Pi(H) \le \overline{\Psi}_\Pi(H) := \min_{\Theta,\varkappa}\left\{\Gamma_\Pi(\Theta) + \phi_{\mathcal{R}}(\varkappa) : \begin{array}{c}\varkappa \ge 0\\ \left[\begin{array}{c|c}\sum_\ell \varkappa_\ell R_\ell & \frac{1}{2}M^T H^T\\ \hline \frac{1}{2}HM & \Theta\end{array}\right] \succeq 0\end{array}\right\}}$$

Indeed, let $(\varkappa, \Theta)$ be feasible for the problem specifying $\overline{\Psi}_\Pi$, and let $\mathrm{Var}[P] \preceq Q \in \Pi$. We have

$$\begin{array}{rcl}\|H^T\xi\| & = & \max\limits_{u\in\mathcal{B}_*}[-u^T H^T \xi] = \max\limits_{y\in\mathcal{Y}}[-y^T M^T H^T \xi] \le \max\limits_{y\in\mathcal{Y}}\left[y^T[\sum_\ell \varkappa_\ell R_\ell]y + \xi^T\Theta\xi\right]\\ & = & \max\limits_{r\in\mathcal{R},y}\left\{y^T[\sum_\ell \varkappa_\ell R_\ell]y + \xi^T\Theta\xi : y^T R_\ell y \le r_\ell, \ell \le L\right\} \le \max\limits_{r\in\mathcal{R}}\left\{\sum_\ell \varkappa_\ell r_\ell + \xi^T\Theta\xi\right\}\\ & \le & \phi_{\mathcal{R}}(\varkappa) + \xi^T\Theta\xi = \phi_{\mathcal{R}}(\varkappa) + \mathrm{Tr}(\Theta[\xi\xi^T]).\end{array}$$

Taking expectation in ξ, we get

$$\mathbf{E}_{\xi\sim P}\left\{\|H^T\xi\|\right\} \le \phi_{\mathcal{R}}(\varkappa) + \mathrm{Tr}(\Theta\mathrm{Var}[P]) \le \phi_{\mathcal{R}}(\varkappa) + \Gamma_\Pi(\Theta).$$

and the conclusion of Lemma follows. $\square$

Illustration: When $\|\cdot\| = \|\cdot\|_p$, $p \in [1,2]$, Lemma implies that whenever $\mathrm{Var}[P] \preceq Q$, one has

$$\mathbf{E}_{\xi\sim P}\left\{\|H^T\xi\|_p\right\} \leq \left\|\left[\|\mathrm{Col}_1[Q^{1/2}H]\|_2; ...; \|\mathrm{Col}_k[Q^{1/2}H]\|_2\right]\right\|_p$$

♠ **Summary:** *Consider convex optimization problem*

$$\mathrm{Opt} = \min_{H,\lambda,\mu,\varkappa,\Theta} \Bigg\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) + \phi_{\mathcal{R}}(\varkappa) + \Gamma_{\Pi}(\Theta) : \lambda \geq 0, \mu \geq 0, \varkappa \geq 0,$$

$$\left[\begin{array}{c|c} \sum_\ell \mu_\ell R_\ell & \frac{1}{2}M^T[H^TA - B] \\ \hline \frac{1}{2}[A^TH - B^T]M & \sum_k \lambda_k S_k \end{array}\right] \succeq 0, \left[\begin{array}{c|c} \sum_\ell \varkappa_\ell R_\ell & \frac{1}{2}M^TH^T \\ \hline \frac{1}{2}HM & \Theta \end{array}\right] \succeq 0 \Bigg\}$$

$$\left[\Gamma_{\Pi}(\Theta) = \max_{Q\in\Pi} \mathrm{Tr}(\Theta Q)\right]$$

The problem is solvable, and the H-component H_* of its optimal solution yields linear estimate

$$\widehat{x}_{H_*}(\omega) = H_*^T \omega$$

such that

$$\mathrm{Risk}_{\|\cdot\|,\Pi}[\widehat{x}_{H_*}|\mathcal{X}] \leq \mathrm{Opt}.$$

Fact: *In the case of zero mean Gaussian observation noise, the estimate $\widehat{x}_{H_*}$ is near-optimal:*

♠ **Theorem:** *We have*

$$\mathrm{Risk}_{\|\cdot\|,\Pi}[\widehat{x}_{H_*}|\mathcal{X}] \le \mathrm{Opt} \le O(1)\sqrt{\ln(2K)\ln(2L)}\,\mathrm{RiskOpt}_{\|\cdot\|,\Pi}[\mathcal{X}],$$

where

- $O(1)$ *is a positive absolute constant,*
- K *and* L *are "sizes" of the ellitopes*

$$\begin{array}{rcl}\mathcal{X} &=& \{x : \exists t \in \mathcal{T} : x^T S_k x \le t_k, k \le K\},\\ \mathcal{B}_* &=& M\mathcal{Y},\ \mathcal{Y} = \{y : \exists r \in \mathcal{R} : y^T R_\ell y \le r_\ell, \ell \le L\},\end{array}$$

- $\mathrm{RiskOpt}_{\|\cdot\|,\Pi}[\mathcal{X}] = \inf\limits_{\widehat{x}(\cdot)} \sup\limits_{Q\in\Pi} \sup\limits_{x\in\mathcal{X}} \mathbf{E}_{\xi\sim\mathcal{N}(0,Q)}\{\|x - \widehat{x}(Ax+\xi)\|\}$ *is the minimax optimal risk taken w.r.t. Gaussian zero mean observation noises with covariance matrices from* Π.

Variation: Recovery of partially stochastic signals

♣ So far, we have considered the problem of recovering the image Bx of unknown *deterministic* signal x known to belong to a given signal set $\mathcal{X}$ from noisy observations

$$\omega = Ax + \xi$$

of linear image of the signal.

In some applications, it makes sense to consider similar problem when the signal has a *stochastic* component.

Example: Kalman's Filter. Consider linear dynamical system

$$\begin{array}{rcl} y_1 &=& \zeta_0 \\ y_{t+1} &=& P_t y_t + u_t + \zeta_t, \ , \ t = 1, 2, ..., T \\ \omega_t &=& C_t y_t + \xi_t \end{array}$$

- $y_t \in \mathbb{R}^n$: states
- u_t: controls
- $\omega_t \in \mathbb{R}^m$: observations
- ζ_t: random "process noise"
- ξ_t: random observation noise
- P_t, C_t: known matrices.

What we want is to recover from observations $\omega_1, ..., \omega_T$ linear image

$$z := R[y_1; ...; y_{T+1}]$$

of the state trajectory, e.g., y_T ("filtering") or y_{T+1} ("forecast").

Note: In the classical Kalman Filter,

— $\zeta_0, ..., \zeta_T$ are independent of each other zero mean Gaussian

— $\xi_1, ..., \xi_T$ are independent of each other and of ζ_t's zero mean Gaussian

— $u_1, ..., u_T$ are deterministic and known (reduces to the case when $u_t \equiv 0$)

$$y_1 = \zeta_0,\; y_{t+1} = P_t y_t + u_t + \zeta_t,\; \omega_t = C_t y_t + \xi_t$$
$$\boxed{(\omega_1, ..., \omega_T) \;??\Rightarrow??\; z = R[y_1; ...; y_{T+1}]}$$

• When modeling the situation as an estimation problem, we can use state equation to express the states y_t as known linear functions of controls u_t and process noises ζ_t, thus arriving at the model

$$\omega = A[u;\zeta] + \xi \;??\Rightarrow?? z = B[u;\zeta]$$
$$[\omega = [\omega_1; ...; \omega_T], u = [u_1; ...; u_T],\; \zeta = [\zeta_0; ...; \zeta_T],\; \xi = [\xi_1; ...; \xi_T]]$$

• When quantifying the performance of a candidate estimate $\widehat{x}(\omega)$, it makes sense to look at risk of the form

$$\mathrm{Risk}[\widehat{x}] = \sup_u \mathbf{E}_{\xi,\zeta}\left\{\|B[u;\zeta] - \widehat{x}(A[u;\zeta] + \xi)\|\right\}.$$

Situation: We observe noisy linear image

$$\omega = A[u;\zeta] + \xi = A_d u + A_s\zeta + \xi$$

of "signal" $x = [u;\zeta]$ with deterministic component u and stochastic component ζ.
We assume that

- u is "uncertain-but-bounded" – is known to belong to a given set $\mathcal{U}$
- ζ and ξ have partially known distributions, specifically, for given $Q_\zeta \succ 0, Q_\xi \succ 0$ it holds

$$\mathrm{Var}[\xi] = \mathbf{E}\{\xi\xi^T\} \preceq Q_\xi,\ \ \mathrm{Var}[\zeta] = \mathbf{E}\{\zeta\zeta^T\} \preceq Q_\zeta$$

Given matrix $B = [B_d, B_s]$ and a norm $\|\cdot\|$ on the image space of B, we want to recover $B[u;\zeta] = B_d u + B_s\zeta$, quantifying the recovery error in $\|\cdot\|$. The performance of a candidate estimate $\widehat{x}(\cdot)$ is quantified by

$$\mathrm{Risk}[\widehat{x}] = \sup_{u\in\mathcal{U}}\sup_{P\in\mathcal{P}} \mathbf{E}_{[\xi;\zeta]\sim P}\{\|B[u;\zeta] - \widehat{x}(A[u;\zeta]+\xi)\|\}$$
$$\left[\mathcal{P} : \text{probability distributions such that } \mathbf{E}_{[\xi;\zeta]\sim P}\left\{\xi\xi^T\right\} \preceq Q_\xi,\ \mathbf{E}_{[\xi;\zeta]\sim P}\left\{\zeta\zeta^T\right\} \preceq Q_\zeta\right]$$

Goal: To build "presumably good" *linear* estimate $\widehat{x}_H(\omega) = H^T\omega$.

$$\omega = A_d u + A_s\zeta + \xi \ \&\ u \in \mathcal{U} \ \&\ \mathrm{Var}[\xi] \preceq Q_\xi \ \&\ \mathrm{Var}[\zeta] \preceq Q_\zeta$$
$$??\Downarrow??$$
$$\widehat{x}_H(\omega) := H^T\omega \approx B_d u + B_s\zeta$$

Assumption: *$\mathcal{U}$ is a basic ellitope, and the unit ball of the norm $\|\cdot\|_*$ dual to $\|\cdot\|$ is an ellitope:*

$$\begin{array}{c}\mathcal{U} = \{u : \exists t \in \mathcal{T} : u^T S_k u \le t_k,\ k \le K\}\\ \{v : \|v\|_* \le 1\} = \{v : \exists r \in \mathcal{R}, w : v = Mw, w^T R_\ell w \le r_\ell,\ \ell \le L\}\end{array}$$

• For a candidate linear estimate $\widehat{x}_H(\omega) = H^T\omega$, $u \in \mathcal{U}$, and a distribution P of $[\xi;\zeta]$ satisfying the bounds on the matrices of second moments of ξ and ζ we have

$$\begin{array}{l}\mathbf{E}_{[\xi,\zeta]\sim P}\left\{\|B_d u + B_s\zeta - H^T[A_d u + A_s\zeta + \xi]\|\right\}\\ \qquad \le \|B_d - H^T A_d]u\| + \mathbf{E}_{[\xi;\zeta]\sim P}\left\{\|H^T\xi\|\right\} + \mathbf{E}_{[\xi;\zeta]\sim P}\left\{\|[B_s - H^T A_s]\zeta\|\right\}\end{array}$$

As we know,

$$u \in \mathcal{U} \Rightarrow \|[B_d - H^T A_d]u\| \le \min_{\lambda\ge 0,\nu\ge 0}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\nu) : \left[\begin{array}{c|c}\sum_\ell \nu_\ell R_\ell & \frac{1}{2}M^T[B_d - H^T A_d^T]\\ \hline \frac{1}{2}[B_d^T - A_d^T H] & \sum_k \lambda_k S_k\end{array}\right] \succeq 0\right\}$$

$$\mathrm{Var}[\xi] \preceq Q_\xi \Rightarrow \mathbf{E}_\xi\left\{\|H^T\xi\|\right\} \le \min_{\mu\ge 0,G}\left\{\mathrm{Tr}(GQ_\xi) + \phi_{\mathcal{R}}(\mu) : \left[\begin{array}{c|c}G & \frac{1}{2}HM\\ \hline \frac{1}{2}M^T H^T & \sum_\ell \mu_\ell R_\ell\end{array}\right] \succeq 0\right\}$$

$$\mathrm{Var}[\zeta] \preceq Q_\zeta \Rightarrow \mathbf{E}_\zeta\left\{\|[B_s - H^T A_s]\zeta\|\right\} \le \min_{\mu\ge 0,G}\left\{\mathrm{Tr}(GQ_\xi) + \phi_{\mathcal{R}}(\mu) : \left[\begin{array}{c|c}G & \frac{1}{2}[B_s^T - A_s^T H]M\\ \hline \frac{1}{2}M^T[B_s - H^T A_s] & \sum_\ell \mu_\ell R_\ell\end{array}\right] \succeq 0\right\}$$

$$\omega = A_d u + A_s\zeta + \xi \;\&\; u \in \{u : \exists t \in \mathcal{T} : u^T S_k u \le t_k,\ k \le K\} \;\&\; \mathrm{Var}[\xi] \preceq Q_\xi \;\&\; \mathrm{Var}[\zeta] \preceq Q_\zeta$$
$$??\Downarrow??$$
$$\widehat{x}_H(\omega) := H^T\omega \approx B_d u + B_s\zeta$$
$$\{v : \|v\|_* \le 1\} = \{v : \exists r \in \mathcal{R}, w : v = Mw, w^T R_\ell w \le r_\ell,\ \ell \le L\}$$

Bottom line: *In the situation at hand, consider the convex optimization problem*

$$\mathrm{Opt} = \min_{\substack{H,\lambda,\nu,\\ \mu,\mu',G,G'}} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\nu) + \phi_{\mathcal{R}}(\mu) + \phi_{\mathcal{R}}(\mu') + \mathrm{Tr}(Q_\xi G) + \mathrm{Tr}(Q_\zeta G') : \begin{array}{l} \lambda \ge 0, \nu \ge 0, \mu \ge 0, \mu' \ge 0, \left[\begin{array}{c|c} \sum_\ell \nu_\ell R_\ell & \frac{1}{2}M^T[B_d - H^T A_d^T] \\ \hline \frac{1}{2}[B_d^T - A_d^T H] & \sum_k \lambda_k S_k \end{array}\right] \succeq 0 \\ \left[\begin{array}{c|c} G & \frac{1}{2}HM \\ \hline \frac{1}{2}M^T H^T & \sum_\ell \mu_\ell R_\ell \end{array}\right] \succeq 0, \left[\begin{array}{c|c} G' & \frac{1}{2}[B_s^T - A_s^T H]M \\ \hline \frac{1}{2}M^T[B_s - H^T A_s] & \sum_\ell \mu'_\ell R_\ell \end{array}\right] \succeq 0 \end{array} \right\}$$

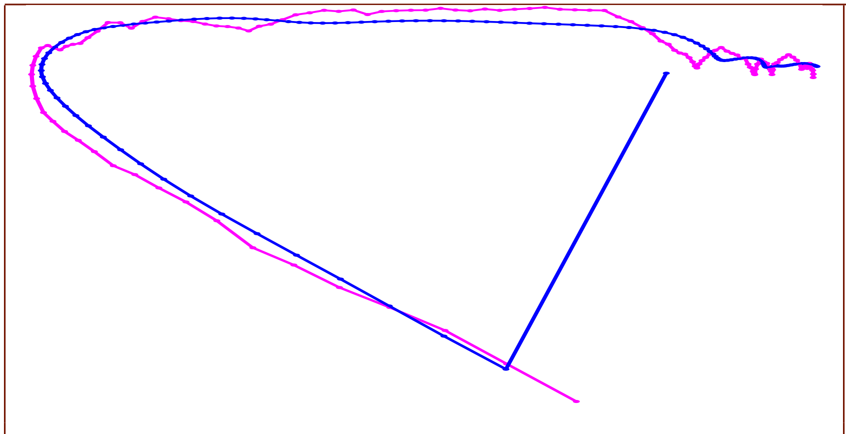
The problem is efficiently solvable, and the H-component H_* of its optimal solution gives rise to linear estimate $\widehat{x}_{H_*}(\omega) = H_*^T\omega$ such that

$$\mathrm{Risk}[\widehat{x}_{H_*}] \le \mathrm{Opt}.$$

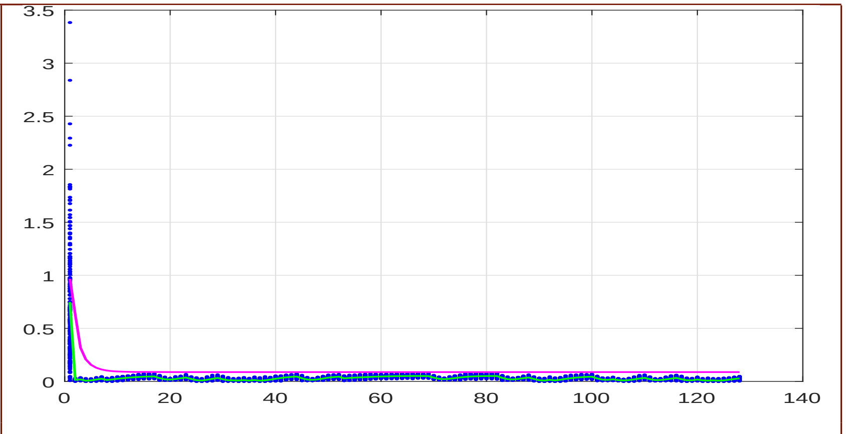
How it works

- **System:** Discretized pendulum $\ddot{x} = -\dot{x} - \kappa x$:

$$\begin{array}{rcl}\begin{bmatrix} x_{t+1} \\ v_{t+1} \end{bmatrix} & = & \begin{bmatrix} 0.9990 & 0.0951 \\ -0.0190 & 0.9039 \end{bmatrix}\begin{bmatrix} x_t \\ v_t \end{bmatrix} + (u_t + \zeta_t)\begin{bmatrix} 0.0048 \\ 0.0951 \end{bmatrix}, \ 1 \le t \le 128 \\ \omega_t & = & x_t + \xi_t \\ & & \left[\begin{bmatrix} x_1 \\ v_1 \end{bmatrix} \sim \mathcal{N}(0, I),\ \zeta_t \sim \mathcal{N}(0, 0.05^2),\ \xi_t \sim \mathcal{N}(0, 0.05^2),\ |u_t| \le 0.1\right]\end{array}$$



sample trajectory and forecasts



errors/mean errors/error bounds

Recovery under uncertain-but-bounded noise

♣ So far, we have considered recovering Bx, $x \in \mathcal{X}$, from observation

$$\omega = Ax + \xi$$

affected by *random* noise ξ. We are about to consider the case when ξ is "uncertain-but-bounded:" all we know is that

$$\xi \in \mathcal{H}$$

with a given convex and compact set $\mathcal{H}$.

♠ In the case in question, natural definition of risk of a candidate estimate $\widehat{x}(\cdot)$ is

$$\mathrm{Risk}_{\mathcal{H},\|\cdot\|}[\widehat{x}(\cdot)|\mathcal{X}] = \sup_{x\in\mathcal{X},\xi\in\mathcal{H}} \|Bx - \widehat{x}(Ax+\xi)\|.$$

♠ **Observation:** *Signal recovery under uncertain-but-bounded noise reduces to the situation where there is no observation noise at all.*

Indeed, let us treat as the signal the pair $z = [x;\xi] \in Z := \mathcal{X} \times \mathcal{H}$ and replace A with $\bar{A} = [A, I]$ and B with $\bar{B} = [B, 0]$, so that

$$\omega = \bar{A}[x;\xi] \qquad \& \qquad Bx = \bar{B}[x;\xi],$$

thus reducing signal recovery to recovering $\bar{B}z$, $z \in Z$, from noiseless observation $\bar{A}z$.

♣ Let us focus on the problem of recovering the image $Bx \in \mathbb{R}^\nu$ of unknown signal $x \in \mathbb{R}^n$ known to belong to signal set $\mathcal{X} \subset \mathbb{R}^n$ via observation

$$\omega = Ax \in \mathbb{R}^m.$$

Given norm $\|\cdot\|$ on $\mathbb{R}^\nu$, we quantify the performance of an estimate $\widehat{x}(\cdot) : \mathbb{R}^m \to \mathbb{R}^\nu$ by its $\|\cdot\|$-risk

$$\mathrm{Risk}_{\|\cdot\|}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \|Bx - \widehat{x}(Ax)\|.$$

♠ **Observation:** *Assuming that $\mathcal{X}$ is computationally tractable convex compact set and $\|\cdot\|$ is computationally tractable, it is easy to build an efficiently computable optimal within factor 2 nonlinear estimate:*

Given ω, let us solve the convex feasibility problem

$$\text{Find } y \in \mathcal{Y}[\omega] := \{y \in \mathcal{X} : Ay = \omega\}.$$

and take, as $\widehat{x}(\omega)$, the vector By, where y is (any) solution to the feasibility problem.

Note: When ω stems from a signal $x \in \mathcal{X}$, the set $\mathcal{Y}[\omega]$ contains x

$\Rightarrow \widehat{x}(\cdot)$ *is well defined*

$$x \in \mathcal{X}, \omega = Ax \;\Rightarrow\; \widehat{x}(\omega) = By$$
$$[y \in \mathcal{Y}[\omega] = \{y \in \mathcal{X} : Ay = \omega\}]$$

♠ **Performance analysis:** Let

$$\begin{array}{rcl} \mathfrak{R} & = & \max\limits_{y,z} \left\{ \frac{1}{2}\|B[y-z]\| : y, z \in \mathcal{X}, A[y-z] = 0 \right\} \\ & = & \frac{1}{2}\|B[y_* - z_*]\| \quad [y_*, z_* \in \mathcal{X}, A[y_* - z_*] = 0] \end{array}$$

Claim A: *For every estimate* $\widetilde{x}(\cdot)$ *it holds* $\mathrm{Risk}_{\|\cdot\|}[\widetilde{x}|\mathcal{X}] \geq \mathfrak{R}$.

Indeed, the observation $\omega = Ay_* = Az_*$ stems from both y_* and z_*, whence the $\|\cdot\|$-risk of every estimate is at least $\frac{1}{2}\|y_* - z_*\| = \mathfrak{R}$.

Claim B: *One has* $\mathrm{Risk}_{\|\cdot\|}[\widehat{x}|\mathcal{X}] \leq 2\mathfrak{R}$.

Indeed, let $\omega = Ax$ with $x \in \mathcal{X}$, and let $\widehat{x}(\omega) = B\widehat{y}$ with $\widehat{y} \in \mathcal{Y}[\omega]$. Then both $x, \widehat{y}$ belong to $\mathcal{Y}[\omega]$

$$\Rightarrow \frac{1}{2}\|B[x - \widehat{y}]\| \leq \mathfrak{R}.$$

♣ We have built optimal, within factor 2, estimate. How to upper-bound its $\|\cdot\|$-risk?
♠ **Observation:** Let $\mathcal{X}$ and the unit ball $\mathcal{B}_*$ of the norm $\|\cdot\|_*$ *conjugate* to $\|\cdot\|$ be ellitopes:

$$\begin{array}{rcl}\mathcal{X} & = & \left\{x = Py : y \in \mathcal{Y} := \{y : \exists t \in \mathcal{T} : y^T S_k y \leq t_k,\, k \leq K\}\right\}\\ \mathcal{B}_* & = & \left\{u = Qv : v \in \mathcal{V} := \{v : \exists r \in \mathcal{R} : v^T R_\ell v \leq r_\ell,\, \ell \leq L\}\right\}\end{array}$$

Then the $\|\cdot\|$-risk of the optimal, within factor 2, efficiently computable nonlinear estimate $\widehat{x}(\cdot)$ cam be tightly lower- and upper-bounded as follows.

• Assuming $\mathrm{Ker} A \cap \mathcal{X} \neq \{0\}$ (otherwise the risk is zero), the set $\mathcal{X}_A = \{x \in \mathcal{X} : Ax = 0\}$ is an ellitope:

$$\mathcal{X}_A = \left\{x = Fw, w \in \mathcal{W} := \{w : \exists t \in \mathcal{T} : w^T T_k w \leq t_k, k \leq K\}\right\}$$

Indeed, setting $E = \{y : APy = 0\}$, the set

$$\mathcal{Y}_A = \{y \in E : \exists t \in \mathcal{T} : y^T S_k y \leq t_k, k \leq K\}$$

is a basic ellitope in some $\mathbb{R}^{N'}$ $\Rightarrow$ $\mathcal{X}_A = \{Py : y \in \mathcal{Y}_A\}$ is an ellitope.

$$\begin{array}{rcl}\Rightarrow \mathfrak{R} & := & \max\limits_{x',x'' \in \mathcal{X}} \left\{\frac{1}{2}\|B[x'-x'']\| : A[x'-x''] = 0\right\} = \max\limits_{x \in \mathcal{X}_A} \|Bx\| = \max\limits_{w \in \mathcal{W}} \|BFw\|\\ & = & \|BF\|_{\|\cdot\|_w \to \|\cdot\|} \ [\|\cdot\|_w\text{: norm with the unit ball } \mathcal{W}]\end{array}$$

$\Rightarrow \mathfrak{R} \leq \mathrm{Opt} \leq 3\ln(\sqrt{3}[K+L])\mathfrak{R}$, with Opt given by

$$\mathrm{Opt} = \min_{\lambda,\mu}\left\{\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) : \begin{array}{c}\lambda \geq 0, \mu \geq 0\\ \left[\begin{array}{c|c}\sum_\ell \mu_\ell R_\ell & \frac{1}{2}Q^T BF\\ \hline \frac{1}{2}F^T B^T Q & \sum_k \lambda_k T_k\end{array}\right] \succeq 0\end{array}\right\}.$$

$\Rightarrow$ The optimal $\|\cdot\|$-risk is $\geq \mathfrak{R} \geq \frac{\mathrm{Opt}}{3\ln(\sqrt{3}[K+L])}$, and $\mathrm{Risk}_{\|\cdot\|}[\widehat{x}|\mathcal{X}] \leq 2\mathfrak{R} \leq 2\mathrm{Opt}$.

♠ In fact, under mild assumptions a *linear* estimate is near-optimal:

Theorem. *Consider the problem of recovering Bx in $\|\cdot\|$, $x \in \mathcal{X}$, via observation $\omega = Ax$. Let the signal set $\mathcal{X}$ and the unit ball $\mathcal{B}_*$ of the norm conjugate to $\|\cdot\|$ be ellitopes:*

$$\begin{array}{rcl}\mathcal{X} &=& \{x = Py : y \in \mathcal{Y} := \{y : \exists t \in \mathcal{T} : y^T S_k y \leq t_k,\, k \leq K\}\} \\ \mathcal{B}_* &=& \{u = Qz : z \in \mathcal{Z} = \{\exists r \in \mathcal{R} : z^T R_\ell z \leq r_\ell,\, \ell \leq L\}\}\end{array}$$

Then the linear estimate $\widehat{x}(\omega) = H_*^T \omega$ yielded by the H-component of optimal solution to the efficiently solvable optimization problem

$$\mathrm{Opt} = \min_{\lambda,\mu,H} \left\{ \phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) : \lambda \geq 0, \mu \geq 0 \left[\begin{array}{c|c} \sum_\ell \mu_\ell R_\ell & \frac{1}{2}[B - H^T A]P \\ \hline \frac{1}{2}P^T[B^T - A^T H] & \sum_k \lambda_k S_k \end{array}\right] \succeq 0 \right\}$$

is near-optimal:

$$\mathrm{Risk}_{\|\cdot\|}[\widehat{x}_{H_*}|\mathcal{X}] \leq \mathrm{Opt} \leq O(1)\ln(K+L)\mathrm{Risk}^*_{\|\cdot\|}[\mathcal{X}],$$

where

$$\mathrm{Risk}^*_{\|\cdot\|}[\mathcal{X}] = \inf_{\widehat{x}(\cdot)} \mathrm{Risk}_{\|\cdot\|}[\widehat{x}|\mathcal{X}],$$

inf *being taken over all estimates, linear and nonlinear alike, is the minimax optimal risk.*

From Ellitopes to Spectratopes

♠ **Fact:** *All our results extend from ellitopes – sets of the form*

$$\left\{y\in\mathbb{R}^N:\exists t\in\mathcal{T},z:y=Pz,z^TS_kz\leq t_k,k\leq K\right\}\qquad (E)$$
$$\left[\begin{array}{c}S_k\succeq 0,\sum_k S_k\succ 0\\ \mathcal{T}\subset\mathbb{R}^K_+:\text{ monotone convex compact, }\mathcal{T}\bigcap\mathrm{int}\,\mathbb{R}^K_+\neq\emptyset\end{array}\right]$$

which played the roles of signal sets, ranges of bounded noise, and unit balls of the norms conjugate to the norms $\|\cdot\|$ in which the recovering error is measured, to a wider family – spectratopes

basic spectratope: $\mathcal{Y}=\{y\in\mathbb{R}^N:\exists t\in\mathcal{T},S_k^2[y]\preceq t_kI_{d_k},k\leq K\}$

spectratope: $\mathcal{Z}=\{z=Py,y\in\mathcal{Y}\}$ $\qquad (S)$

$$\left[\begin{array}{l}S_k[y]=\sum_j y_jS^{kj},S^{kj}\in\mathbf{S}^{d_k}:\text{ linear mapings with values in }\mathbf{S}^{d_k}\\ y\neq 0\Rightarrow\sum_k S_k^2[y]\neq 0\text{ [equivalent to }\mathcal{Y}\text{ being bounded]}\\ \mathcal{T}\text{ as in }(E)\end{array}\right]$$

Note:

- Every ellitope is a spectratope.

It suffices to verify that basic ellitope $\mathcal{X} = \{x : \exists t \in \mathcal{T} : x^T S_k x \leq t_k, k \leq K\}$ is a basic spectratope.

Indeed, representing $S_k = \sum_{i=1}^{r_k} f_{ki} f_{ki}^T$, we have

$$\begin{array}{rcl}\mathcal{X} & = & \left\{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \leq t_k, k \leq K\right\} \\ & = & \left\{x \in \mathbb{R}^n : \exists \{t_{ki} \geq 0, 1 \leq k \leq K, 1 \leq i \leq r_i\} : [\sum_i t_{1i}; ...; \sum_i t_{Ki}] \in \mathcal{T} : [f_{ki}^T x]^2 \preceq t_{ki} I_1 \forall (k \leq K, i \leq r_k)\right\}\end{array}$$

- Denoting by $\|\cdot\|_{2,2}$ the spectral norm, matrix box

$$\mathcal{X} = \{x \in \mathbb{R}^{p\times q} : \|x\|_{2,2} \leq 1\} = \{x \in \mathbb{R}^{p\times q} : \left[\begin{array}{c|c} & x \\ \hline x^T & \end{array}\right]^2 \preceq I_{p+q}\}$$

and its symmetric version

$$\mathcal{X} = \{x \in \mathbf{S}^n : -I_n \preceq x \preceq I_n\} = \{x \in \mathbf{S}^n : x^2 \preceq I_n\}$$

are spectratopes $\Rightarrow$ access to matrix boxes as signal sets and nuclear norm as the recovery norm

- Spectratopes admit the same fully algorithmic calculus as ellitopes basic spectratope: $\mathcal{Y} = \{y \in \mathbb{R}^N : \exists t \in \mathcal{T}, S_k^2[y] \preceq t_k I_{d_k}, k \leq K\}$

spectratope: $\mathcal{Z} = \{z = Py, y \in \mathcal{Y}\}$ (S)

$\left[\begin{array}{l} S_k[y] = \sum_j y_j S^{kj}, S^{kj} \in \mathbf{S}^{d_k}: \text{ linear mapings with values in } \mathbf{S}^{d_k} \\ y \neq 0 \Rightarrow \sum_k S_k^2[y] \neq 0 \text{ [equivalent to } \mathcal{Y} \text{ being bounded]} \\ \mathcal{T} \in \mathbb{R}_+^K \text{ monotone convex compact set intersecting } \mathrm{int}\, \mathbb{R}_+^K \end{array}\right]$

♠ **Modifications** of the results when passing from ellitopes to spectratopes are as follows:

A. *The "size" K of an ellitope (E)* (logs of these sizes participate in our tightness factors) *in the case of spectratope (S) becomes $D = \sum_k d_k$*

B. Semidefinite relaxation bound for the quantity

$$\begin{array}{rcl}\mathrm{Opt}_* & = & \max\limits_{y}\left\{y^TBy : \exists t\in\mathcal{T}, z : y = Pz, S_k^2[z]\preceq t_kI_{d_k}, k\leq K\right\}\\ & = & \max\limits_{z,t}\left\{z^T\widehat{B}z : t\in\mathcal{T}, S_k^2[z]\preceq t_kI_{d_k}, k\leq K\right\}, \ \widehat{B}=P^TBP\end{array}$$

is as follows. We associate with $S_k[z]=\sum_j z_jS^{kj}$, $S^{kj}\in\mathbf{S}^{d_k}$, two linear mappings:

$$\begin{array}{ll}Q\mapsto\mathcal{S}_k[Q]:\mathbf{S}^{\dim z}\to\mathbf{S}^{d_k}: & \mathcal{S}_k[Q]=\sum_{i,j}\frac{1}{2}Q_{ij}[S^{ki}S^{kj}+S^{kj}S^{ki}]=\sum_{i,j}Q_{ij}S^{ki}S^{kj}\\ \Lambda\mapsto\mathcal{S}_k^*[\Lambda]:\mathbf{S}^{d_k}\to\mathbf{S}^{\dim z}: & \left[\mathcal{S}_k^*[\Lambda]\right]_{ij}=\frac{1}{2}\mathrm{Tr}(\Lambda[S^{ki}S^{kj}+S^{kj}S^{ki}])=\mathrm{Tr}(\Lambda S^{ki}S^{kj})\end{array}$$

Note: • $S_k^2[z]=\mathcal{S}_k[zz^T]$

• *the mappings $\mathcal{S}_k$ and $\mathcal{S}_k^*$ are conjugates of each other w.r.t. to the Frobenius inner product:*

$$\mathrm{Tr}(\mathcal{S}_k[Q]\Lambda)=\mathrm{Tr}(Q\mathcal{S}_k^*[\Lambda])\ \forall(Q\in\mathbf{S}^{\dim z},\Lambda\in\mathbf{S}^{d_k})$$

Selecting $\Lambda_k\succeq 0$, $k\leq K$, such that $\sum_k\mathcal{S}_k^*[\Lambda_K]\succeq\widehat{B}$, for

$$z\in\mathcal{Z}=\{z:\exists t\in\mathcal{T}:S_k^2[z]\preceq t_kI_{d_k},k\leq K\}$$

we have

$$\begin{array}{l}\exists t\in\mathcal{T}:S_k^2[z]\preceq t_kI_{d_k}\forall k\Rightarrow z^T\widehat{B}z\leq z^T\left[\sum_k\mathcal{S}_k^*[\Lambda_k]\right]z=\sum_k z^T\mathcal{S}_k^*[\Lambda_k]z=\sum_k\mathrm{Tr}(\mathcal{S}_k^*[\Lambda_k][zz^T])\\ =\sum_k\mathrm{Tr}(\Lambda_k\mathcal{S}_k[zz^T])=\sum_k\mathrm{Tr}(\Lambda_kS_k^2[z])\leq\sum_k t_k\mathrm{Tr}(\Lambda_k)\leq\phi_{\mathcal{T}}(\lambda[\Lambda]),\\ \qquad\phi_{\mathcal{T}}(\lambda)=\max\limits_{t\in\mathcal{T}}t^T\lambda,\ \lambda[\Lambda]=[\mathrm{Tr}(\Lambda_1);...;\mathrm{Tr}(\Lambda_K)]\end{array}$$

$$\Rightarrow\boxed{\mathrm{Opt}_*\leq\mathrm{Opt}:=\min_{\Lambda=\{\Lambda_k,k\leq K\}}\left\{\phi_{\mathcal{T}}(\lambda[\Lambda]):\Lambda_k\succeq 0,k\leq K,\widehat{B}\preceq\sum_k\mathcal{S}_k^*[\Lambda_k]\right\}}$$

♠ **Theorem.** *Semidefinite relaxation bound*

$$\mathrm{Opt} := \min_{\Lambda=\{\Lambda_k, k\leq K\}} \left\{\phi_{\mathcal{T}}(\lambda[\Lambda]) : \Lambda_k \succeq 0, k \leq K, \widehat{B} \preceq \sum_k \mathcal{S}_k^*[\Lambda_k]\right\}$$

on the quantity

$$\begin{array}{rcl}\mathrm{Opt}_* &=& \max_y \left\{y^T B y : \exists t \in \mathcal{T}, z : y = Pz, S_k^2[z] \preceq t_k I_{d_k}, k \leq K\right\} \\ &=& \max_{z,t} \left\{z^T \widehat{B} z : t \in \mathcal{T}, S_k^2[z] \preceq t_k I_{d_k}, k \leq K\right\}\end{array}$$

is tight:

$$\mathrm{Opt}_* \leq \mathrm{Opt} \leq 2\ln(2\sum_k d_k)\mathrm{Opt}_*.$$

Note: Proof follows the one for the ellitopic case.

But: The role of elementary Mini-Lemma in the spectratopic case is played by the following fundamental matrix concentration result:

Noncommutative Khintchine Inequality [Lust-Picard 1986, Pisier 1998, Buchholz 2001] *Let* $A_i \in \mathbf{S}^d$, $1 \leq i \leq N$, *be deterministic matrices such that*

$$\sum_i A_i^2 \preceq I_d,$$

and let ζ *be* N*-dimensional* $\mathcal{N}(0, I_N)$ *or Rademacher random vector. Then for all* $s \geq 0$ *it holds*

$$\mathrm{Prob}\left\{\|\sum_i \zeta_i A_i\|_{2,2} \geq s\right\} \leq 2d\exp\{-s^2/2\}.$$

C. Assuming that the signal set $\mathcal{X}$ and the unit ball $\mathcal{B}_*$ of the norm conjugate to $\|\cdot\|$ spectratopes:

$$\begin{array}{rcl}\mathcal{X} &=& \{x\in\mathbb{R}^n: \exists t\in\mathcal{T}: S_k^2[x]\preceq t_k I_{d_k}, k\leq K\}\\ \mathcal{B}_* &=& M\mathcal{Y},\ \mathcal{Y}=\{y\in\mathbb{R}^N: \exists r\in\mathcal{R}: R_\ell^2[y]\preceq r_\ell I_{f_\ell}, \ell\leq L\}\end{array}$$

and that the distribution of noise in observation $\omega = Ax+\xi$ belongs to $\mathcal{P}[\Pi]$, the problem responsible for building presumably good linear estimate of Bx via ω reads

$$\mathrm{Opt}=\min_{H,\Lambda,\Upsilon,\Upsilon',\Theta}\left\{\begin{array}{l}\phi_{\mathcal{T}}(\lambda[\Lambda])+\phi_{\mathcal{R}}(\lambda[\Upsilon])+\phi_{\mathcal{R}}(\lambda[\Upsilon'])+\Gamma_\Pi(\Theta):\\ \Lambda=\{\Lambda_k\succeq 0\}_{k\leq K}, \Upsilon=\{\Upsilon_\ell\succeq 0\}_{\ell\leq L}, \Upsilon'=\{\Upsilon'_\ell\succeq 0\}_{\ell\leq L}\\ \left[\begin{array}{c|c}\sum_\ell \mathcal{R}_\ell^*[\Upsilon_\ell] & \frac{1}{2}M^T[H^TA-B]\\ \hline \frac{1}{2}[A^TH-B^T]M & \sum_k \mathcal{S}_k^*[\Lambda_k]\end{array}\right]\succeq 0\\ \left[\begin{array}{c|c}\sum_\ell \mathcal{R}_\ell^*[\Upsilon'_\ell] & \frac{1}{2}M^TH^T\\ \hline \frac{1}{2}HM & \Theta\end{array}\right]\succeq 0\end{array}\right\}$$

$$\left[\begin{array}{c}\Gamma_\Pi(\Theta)=\max\limits_{Q\in\Pi}\mathrm{Tr}(\Theta Q), \phi_G(h)=\max\limits_{g\in G} g^Th\\ S_k[x]=\sum_i xS_k[x]=\sum_i x_iS^{ki}\Rightarrow \mathcal{S}_k^*[\Lambda_k]=\left[\mathrm{Tr}(S^{kp}\Lambda_kS^{kq})\right]_{p,q}\leq n\\ R_\ell[y]=\sum_i y_iR^{ki}\Rightarrow \mathcal{R}_\ell^*[\Upsilon_\ell]=\left[\mathrm{Tr}(R^{\ell p}\Upsilon_\ell R^{\ell q})\right]_{p,q\leq N}\\ \lambda[\{U_1,...,U_s]\}=[\mathrm{Tr}(U_1);...;\mathrm{Tr}(U_s)]\end{array}\right]$$

The risk $\mathrm{Risk}_{\|\cdot\|,\mathcal{P}[\Pi]}[\widehat{x}_{H_*}|\mathcal{X}]$ of the linear estimate $\widehat{x}_{H_*}(\omega)=H_*^T\omega$ yielded by the H-component of optimal solution to the problem does not exceed Opt.

$$\begin{array}{rcl}\mathcal{X} &=& \{x\in\mathbb{R}^n : \exists t\in\mathcal{T} : S_k^2[x]\preceq t_k I_{d_k}, k\leq K\}\\ \mathcal{B}_* &=& M\mathcal{Y},\ \mathcal{Y}=\{y\in\mathbb{R}^N : \exists r\in\mathcal{R} : R_\ell^2[y]\preceq r_\ell I_{f_\ell}, \ell\leq L\}\end{array}$$

D. Near-optimality statement reads as follows:

The $\|\cdot\|$-risk of the just defined presumably good linear estimate $\widehat{x}_{H_*}$ is within moderate factor of minimax optimal Gaussian risk:

$$\mathrm{Risk}_{\|\cdot\|,\mathcal{P}[\Pi]}[\widehat{x}_{H_*}|\mathcal{X}] \leq \mathrm{Opt} \leq O(1)\sqrt{\ln(2D)\ln(2F)}\,\mathrm{RiskOpt}_{\|\cdot\|,\mathcal{P}[\Pi]}[\mathcal{X}]$$

where

$$D=\sum_k d_k,\ F=\sum_\ell f_\ell$$

are the spectratopic sizes of $\mathcal{X}$ and $\mathcal{B}_*$, and

$$\mathrm{RiskOpt}_{\|\cdot\|,\mathcal{P}[\Pi]}[\mathcal{X}] = \inf_{\widehat{x}}\sup_{Q\in\Pi}\max_{x\in\mathcal{X}}\mathbf{E}_{\xi\sim\mathcal{N}(0,Q)}\{\|Bx-\widehat{x}(Ax+\xi)\|\}$$

is the Gaussian minimax optimal risk, i.e., the minimax risk associated with zero mean Gaussian noises with covariance matrices from Π.

Proof of Near-Optimality: Executive Sketch

Preliminaries, 1. We shall use the following important

Anderson's Lemma. *Let $f:\mathbb{R}^N\to\mathbb{R}$ be an even nonnegative summable function such that the sets $\{u: f(u)\geq t\}$ are convex for all $t\geq 0$, and let X be a symmetric w.r.t. the origin closed convex subset of $\mathbb{R}^n$. Then the function*

$$\int_{X+\tau e} f(u)du \qquad [e\in\mathbb{R}^N]$$

is nonincreasing in $\tau\geq 0$. As a result, if $W\in\mathbf{S}^N_+$, $\|\cdot\|$ is a norm on $\mathbb{R}^\nu$ and Y is an $\nu\times N$ matrix, one has

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,W)}\{\|Y\eta+e\|\geq r\}\geq \mathrm{Prob}\{\|Y\eta\|\geq r\}\ \forall(e\in\mathbb{R}^N, r\geq 0).$$

Preliminaries, 2. By simple saddle point argument, the optimal value Opt in the problem specifying the presumably good linear estimate is *as if* the distribution of noise were zero mean with appropriately selected covariance matrix $Q_* \in \Pi$.

From now on we restrict the observation noise to be $\mathcal{N}(0, Q_*)$.

Preliminaries, 3. The crucial role in the proof is played by the following

Main Lemma. *Let the unit ball $\mathcal{B}_*$ of the norm conjugate to norm $\|\cdot\|$ on $\mathbb{R}^\nu$ be a spectratope:*

$$\mathcal{B}_* = M\mathcal{Y}, \mathcal{Y} = \{y \in \mathbb{R}^N : \exists r \in \mathcal{R} : R_\ell^2[y] \preceq r_\ell I_{f_\ell}, \ell \leq L\},$$

let Y be an $S \times \nu$ matrix, and $\eta \sim \mathcal{N}(0, \Sigma)$ with $0 \prec \Sigma \in \mathbf{S}^S$. Then the upper bound

$$\Psi_\Sigma(Y) = \min_{\Upsilon, \Theta}\left\{\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(\Sigma\Theta) : \begin{array}{c}\Upsilon = \{\Upsilon_\ell \succeq 0\}_{\ell \leq L} \\ \left[\begin{array}{c|c}\sum_\ell \mathcal{R}_\ell^*[\Upsilon_\ell] & \frac{1}{2}M^TY^T \\ \hline \frac{1}{2}YM & \Theta\end{array}\right] \succeq 0\end{array}\right\}$$

on the quantity $\mathbf{E}_\eta\left\{\|Y^T\eta\|\right\}$ is tight:

$$\mathbf{E}_{\eta\sim\mathcal{N}(0,\Sigma)}\left\{\|Y^T\eta\|\right\} \leq \Psi_\Sigma(Y) \leq O(1)\sqrt{\ln(2F)}\mathbf{E}_{\eta\sim\mathcal{N}(0,\Sigma)}\left\{\|Y^T\eta\|\right\}, F = \sum_\ell f_\ell$$

Besides this, for every $\delta \in (0,1)$ it holds

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,\Sigma)}\left\{\|Y^T\eta\| > \frac{\delta}{\sqrt{\ln(2F/\delta)}}\Psi_\Sigma(Y)\right\} \geq 1 - O(1)\delta.$$

Proof of Main Lemma heavily utilizes Conic Duality.

$$\boxed{\omega = Ax + \xi : x \in \mathcal{X}, \xi \sim \mathcal{N}(0, Q_*), Q_* \succ 0}$$

Step 1: All we need is to upper-bound Opt in terms of the minimax optimal risk

$$\mathrm{RiskOpt}_{\|\cdot\|}[\mathcal{X}] = \inf_{\widehat{x}} \sup_{x \in \mathcal{X}} \mathbf{E}_{\xi \sim \mathcal{N}(0,Q_*)} \left\{\|Bx - \widehat{x}(Ax + \xi)\|\right\}.$$

Technically it is easier to upper-bound Opt in terms of the minimax ϵ-risk Risk_ϵ:

$$\mathrm{Risk}_\epsilon \quad := \quad \inf_{\widehat{x}} \min \left\{\rho : \mathrm{Prob}_{\xi \sim \mathcal{N}(0,Q_*)} \left\{\|Bx - \widehat{x}(Ax + \xi)\| > \rho\right\} \leq \epsilon \, \forall x \in \mathcal{X}\right\}$$

In the proof we use once for ever fixed ϵ, namely, $\epsilon = \frac{1}{8}$.

Note: $\mathrm{Risk}_{\frac{1}{8}} \leq 8\mathrm{Risk}_{\|\cdot\|}[\mathcal{X}] \Rightarrow$ upper-bounding Opt in terms of $\mathrm{Risk}_{\frac{1}{8}}$ automatically implies upper-bounding of Opt in terms of $\mathrm{Risk}_{\|\cdot\|}$.

$$\boxed{\omega = Ax + \xi : x \in \mathcal{X}, \xi \sim \mathcal{N}(0, Q_*), Q_* \succ 0}$$

Step 2: Let $W \in \mathbf{S}^n_+$. Consider the *Bayesian* version of our estimation problem, where the observation is

$$\omega = A\eta + \xi$$

$$\xi \sim \mathcal{N}(0, Q_*), \eta \sim \mathcal{N}(0, W) \text{ *are independent of each other*}$$

Fact [well known]: Since $[\omega; \eta]$ is zero mean Gaussian, *the conditional, given* ω*, expectation* $\mathbf{E}_{|\omega}\{B\eta\}$ *of* $B\eta$ *is a linear function* $\bar{H}^T\omega$ *of* ω.

Given this fact, Anderson's Lemma, and Main Lemma, we, with moderate effort, arrive at the following

♠ **Intermediate Conclusion:** *Given* $W \succ 0$ *and setting*

$$\begin{array}{rcl}
\overline{\Psi}(H) & = & \min\limits_{\Upsilon,\Theta}\left\{\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(Q_*\Theta) : \begin{array}{l}\Upsilon = \{\Upsilon_\ell \succeq 0\}_{\ell \le L} \\ \left[\begin{array}{c|c}\sum_\ell \mathcal{R}^*_\ell[\Upsilon_\ell] & \frac{1}{2}M^TH^T \\ \hline \frac{1}{2}HM & \Theta\end{array}\right] \succeq 0\end{array}\right\} \\
\overline{\Phi}(W,H) & = & \min\limits_{\Upsilon,\Theta}\left\{\phi_{\mathcal{R}}(\lambda[\Upsilon]) + \mathrm{Tr}(W\Theta) : \begin{array}{l}\Upsilon = \{\Upsilon_\ell \succeq 0\}_{\ell \le L} \\ \left[\begin{array}{c|c}\sum_\ell \mathcal{R}^*_\ell[\Upsilon_\ell] & \frac{1}{2}M^T[B - H^TA] \\ \hline \frac{1}{2}[B^T - A^TH]M & \Theta\end{array}\right] \succeq 0\end{array}\right\}
\end{array}$$

for an appropriate absolute constant $O(1) > 0$ *and every estimate* $\widehat{x}(\cdot)$ *we have*

$$\mathrm{Prob}_{[\xi;\eta] \sim \mathcal{N}(0,Q_*) \times \mathcal{N}(0,W)}\left\{\|B\eta - \widehat{x}(A\eta + \xi)\| \ge \frac{O(1)}{\sqrt{\ln(2F)}} \inf_H \left[\overline{\Phi}(W,H) + \overline{\Psi}(H)\right]\right\} \ge \frac{1}{4}$$

where $F = \sum_\ell f_\ell$ *is the spectratopic size of* $\mathcal{B}_*$.

$$\boxed{\mathcal{X} = \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : S_k^2[x] \preceq t_k I_{d_k}, k \leq K\}}$$

For appropriate positive absolute constant $O(1)$, for every $W \in \mathbf{S}^n_+$ and every estimate $\widehat{x}(\cdot)$ one has

$$\mathrm{Prob}_{[\xi;\eta]\sim\mathcal{N}(0,Q_*)\times\mathcal{N}(0,W)}\left\{\|B\eta - \widehat{x}(A\eta+\xi)\| \geq \frac{O(1)}{\sqrt{\ln(2F)}} \inf_H \left[\overline{\Phi}(W,H) + \overline{\Psi}(H)\right]\right\} \geq \tfrac{1}{4}. \qquad (!)$$

Concluding steps: Consider the parametric family of convex sets

$$\mathcal{W}[\varkappa] = \{W \in \mathbf{S}^n_+ : \exists t \in \mathcal{T} : \mathcal{S}_k[W] \leq \varkappa t_k I_{d_k}, k \leq K\} \qquad \left[\mathcal{S}_k[W] = \textstyle\sum_{i,j} W_{ij} S^{ki} S^{kj}\right]$$

where $\varkappa \in (0,1]$, and the parametric family of convex-concave saddle point problems

$$\mathrm{Opt}(\varkappa) = \sup_{W\in\mathcal{W}[\varkappa]} \inf_H \left[\overline{\Phi}(W,H) + \overline{\Psi}(H)\right]. \qquad (*_\varkappa)$$

Note: When $W \in \mathcal{W}[\varkappa]$ and $\eta \sim \mathcal{N}(0,W)$, the vector $\eta/\sqrt{\varkappa}$ "belongs to $\mathcal{X}$ at average:"

$$\exists t \in \mathcal{T} : \forall k \leq K : \varkappa t_k I_{d_k} \succeq \mathcal{S}_k[W] = \textstyle\sum_{i,j} \mathbf{E}_{\eta\sim\mathcal{N}(0,W)}\{\eta_i\eta_j\} S^{ki}S^{kj} = \mathbf{E}_{\eta\sim\mathcal{N}(0,W)}\{\sum_{i,j}\eta_i\eta_j S^{ki}S^{kj}\} = \mathbf{E}_{\eta\sim\mathcal{N}(0,W)}\{S_k^2[\eta]\}.$$

- It is not difficult to verify that *for every* $\varkappa \in (0,1]$*:*

a. *The convex-concave saddle point problem* $(*_\varkappa)$ *has a solution* $(W[\varkappa], H[\varkappa])$

b. $\mathrm{Opt}(\varkappa) \geq \sqrt{\varkappa}\,\mathrm{Opt}(1)$

c. $\mathrm{Opt}(1) = \mathrm{Opt}$ (miracle stemming from Conic Duality)

d. *As* $\varkappa \searrow 0$, $\mathrm{Prob}_{\eta\sim\mathcal{N}(0,W[\varkappa])}\{\eta \notin \mathcal{X}\}$ *rapidly goes to 0:*

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,W[\varkappa])}\{\eta \notin \mathcal{X}\} \leq 2\exp\{-\tfrac{1}{2\varkappa}\} \textstyle\sum_k d_k$$

(stems from Noncommutative Khintchine Inequality)

- By **b**, **c** and (!), for every estimate $\widehat{x}$ and every $\varkappa \in (0,1]$ we have

$$\mathrm{Prob}_{[\xi;\eta]\sim\mathcal{N}(0,Q_*)\times\mathcal{N}(0,W[\varkappa])}\left\{\|B\eta - \widehat{x}(A\eta+\xi)\| \geq \frac{O(1)}{\sqrt{\ln(2F)}}\sqrt{\varkappa}\,\mathrm{Opt}\right\} > \tfrac{1}{4}.$$

For every estimate $\widehat{x}$ and every $\varkappa \in (0,1]$ we have

$$\mathrm{Prob}_{[\xi;\eta]\sim\mathcal{N}(0,Q_*)\times\mathcal{N}(0,W[\varkappa])}\left\{\|B\eta-\widehat{x}(A\eta+\xi)\| \geq \frac{O(1)}{\sqrt{\ln(2F)}}\sqrt{\varkappa}\mathrm{Opt}\right\} > \frac{1}{4}, \qquad (!)$$

and as $\varkappa \searrow 0$, $\mathrm{Prob}_{\eta\sim\mathcal{N}(0,W[\varkappa])}\{\eta \notin \mathcal{X}\}$ rapidly goes to 0:

$$\mathrm{Prob}_{\eta\sim\mathcal{N}(0,W[\varkappa])}\{\eta \notin \mathcal{X}\} \leq 2\exp\{-\frac{1}{2\varkappa}\}D, \qquad (!!)$$

where $D = \sum_k d_k$ is the spectratopic size of $\mathcal{X}$.

These facts easily combine to yield the target upper bound

$$\mathrm{Opt} \leq O(1)\sqrt{\ln(2D)\ln(2F)}\mathrm{Risk}_{\frac{1}{8}}$$

on Opt in terms of $\mathrm{Risk}_{\frac{1}{8}}$.

Indeed, with $\varkappa = O(1)/\ln(2D)$ probability for $\eta \sim \mathcal{N}(0,W[\varkappa])$ to be outside of $\mathcal{X}$ is $< 1/8$ by (!!)
$\Rightarrow$ invoking (!),

$$\mathrm{Prob}_{[\xi;\eta]\sim\mathcal{N}(0,Q_*)\times\mathcal{N}(0,W[\varkappa])}\left\{\|B\eta-\widehat{x}(A\eta+\xi)\| \geq \underbrace{\frac{O(1)}{\sqrt{\ln(2F)}}\sqrt{\varkappa}\mathrm{Opt}}_{\mathfrak{R}} \;\&\; \eta \in \mathcal{X}\right\} > \frac{1}{4}-\frac{1}{8}=\frac{1}{8}$$

$\Rightarrow \mathrm{Risk}_{\frac{1}{8}} \geq \mathfrak{R}$. $\square$

Beyond linearity: Polyhedral estimates

♣ As before, our problem of interest is: *given noisy observation*

$$\omega = Ax + \xi \in \mathbb{R}^m,\ \xi \sim P_x,$$

of unknown signal x *known to belong to a given convex compact signal set* $\mathcal{X} \subset \mathbb{R}^n$*, we want to recover* $Bx \in \mathbb{R}^\nu$ *in a given norm* $\|\cdot\|$*.*

We have seen that under reasonable assumptions on problem's data, efficiently computable via convex Programming *linear in* ω estimates are near-optimal.

However: *There are meaningful situations which go beyond the scope of "reasonable assumptions," moreover, situations where linear estimation is provably far from being near-optimal.*

Example: Let $\mathcal{X} = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ be the unit ℓ_1-ball, observations be direct:

$$\omega = x + \sigma\eta,\ \eta \sim \mathcal{N}(0, I_n),$$

and we want to recover $Bx \equiv x$ in Euclidean norm. For a linear estimate $H^T\omega$, worst-case expected squared recovery error is

$$\max_{x \in \mathcal{X}} \mathbf{E}_{\eta \sim \mathcal{N}(0,I_n)} \left\{\|H^T(x + \sigma\eta) - x\|_2^2\right\} = \max_i \|\mathrm{Row}_i[I - H]\|_2^2 + \sigma^2 \mathrm{Tr}(H^T H)$$

Its minimum over $n \times n$ matrices H is achieved at the scalar matrix $H = hI_n$ with $h = \frac{1}{\sigma^2 n + 1}$ and equals

$$\mathrm{Risk}_{\mathrm{lin}}^2 = \frac{\sigma^2 n}{\sigma^2 n + 1}.$$

When $\sigma^2 n \geq 1$, this squared risk is at least $1/2$.

• Now consider the estimate as follows: *given ω, we estimate x by the optimal solution $\widehat{x}(\omega)$ to the convex optimization problem*

$$\mathrm{Opt}(\omega) = \min\left\{\|\omega - y\|_\infty : y \in \mathcal{X}\right\}.$$

Observe that when $\omega = x + \sigma\eta$ with $x \in X$, setting $\widehat{x} = \widehat{x}(\omega)$ we have

$$\begin{array}{rl}
 & \mathrm{Opt}(\omega) \leq \|\omega - x\|_\infty = \sigma\|\eta\|_\infty \\
\Rightarrow & \|x - \widehat{x}\|_\infty \leq \|x - \omega\|_\infty + \underbrace{\|\omega - \widehat{x}\|_\infty}_{\mathrm{Opt}(\omega) \leq \|x-\omega\|_\infty} \leq 2\sigma\|\eta\|_\infty \\
\Rightarrow & \|x - \widehat{x}\|_2^2 \leq \|x - \widehat{x}\|_\infty \underbrace{\|x - \widehat{x}\|_1}_{\leq 2} \leq 4\sigma\|\eta\|_\infty \\
\Rightarrow & \mathrm{Risk}^2[\widehat{x}] := \max_{x\in\mathcal{X}} \mathbf{E}\left\{\|x - \widehat{x}(Ax + \sigma\eta)\|_2^2\right\} \leq 4\sigma \mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\|\eta\|_\infty\right\}
\end{array}$$

It is easily seen that $\mathbf{E}_{\eta\sim\mathcal{N}(0,I_n)}\left\{\|\eta\|_\infty\right\} \leq 2\sqrt{\ln(2n)}$, whence

$$\mathrm{Risk}^2[\widehat{x}] \leq 8\sigma\sqrt{\ln(2n)}\ \&\ \mathrm{Risk}^2_{\mathrm{lin}} = \frac{\sigma^2 n}{\sigma^2 n + 1}.$$

$\Rightarrow$ *When σ is small and $\sigma^2 n$ is of order of 1, an appropriate nonlinear estimate significantly outperforms the best linear one – for the former, squared risk is nearly $O(\sigma)$, and for the latter it is $O(1)$.*

♠ **What is ahead:** *nonlinear polyhedral estimates with the "scope of near-optimality" strictly wider than the one for linear estimates.*

Polyhedral Estimate: Motivation

♣ To motivate Polyhedral Estimate, let us start with the problem where

$$\omega = Ax_* + \sigma\xi,\ \xi \sim \mathcal{N}(0, I_m)$$

with unknown x_* known to belong to a convex compact signal set $\mathcal{X} \subset \mathbb{R}^n$, and we want to recover Bx_* in norm $\|\cdot\|$. *Let us once for ever fix reliability tolerance $\epsilon \ll 1$.*

♠ The simplest inference we can make from observation is:

Let us select somehow *in advance* N vectors $h_i \in \mathbb{R}^m$. Then with confidence $1-\epsilon$ x_* belongs to the "confidence box"

$$\mathcal{B} := \{|h_i^T[\omega - Ax]| \le \rho_i, i \le N\} \qquad \left[\rho_i = \sigma\sqrt{2\ln(2N/\epsilon)}\|h_i\|_2\right]$$

Indeed, with $\delta_i := h_i^T[\omega - Ax_*] = \sigma h_i^T\xi$ one has $\mathrm{Prob}\{|\delta_i| \le \rho_i\ \forall i\} \ge 1 - 2\sum_i \exp\{-\frac{\rho_i^2}{2\sigma^2}\} \ge 1-\epsilon$.

Acting *as if* $\mathcal{B}$ were summarising all information on x_* contained in ω, we could *select a point $\widetilde{x} \in \mathcal{X} \cap \mathcal{B}$, take it as estimate of x_*, and recover Bx_* by $B\widetilde{x}$.*

Note: Assuming $x_* \in \mathcal{B}$, all we know with our "as if" is that $x_* \in \mathcal{B}$, $\widetilde{x} \in \mathcal{B}$ and $x_* \in \mathcal{X}$, $\widetilde{x} \in \mathcal{X}$, or, which is the same,

$$\Delta := \frac{1}{2}[x_* - \widetilde{x}] \in \mathcal{X}_\mathrm{s} := \frac{1}{2}[\mathcal{X} - \mathcal{X}]\ \&\ |h_i^T A\Delta| \le \rho_i, i \le N,$$

⇒ all we can say about the recovery error is that *with probability $\ge 1-\epsilon$, it holds*

$$\|Bx_* - B\widetilde{x}\| = 2\|B\Delta\| \le \mathfrak{R} := \max_z\{2\|Bz\| : z \in \mathcal{X}_\mathrm{s}, |h_i^T Az| \le \rho_i, 1 \le i \le N\}.$$

♠ *Choosing in advance* $h_i \in \mathbb{R}^m$, $i \leq N$, *and given* $\omega = Ax_* + \sigma\xi$, *take, as estimate of* Bx_*, *vector* $B\widetilde{x}$ *with* $\widetilde{x} \in \mathcal{X} \cap \mathcal{B}$, *where the "confidence box"* $\mathcal{B}$ *is given by*

$$\mathcal{B} = \{x : |h_i^T[\omega - Ax]| \leq \rho_i := \sigma\sqrt{2\ln(2N/\epsilon)}\|h_i\|_2,\, i \leq N\},$$

thus ensuring that

$$\|Bx_* - B\widetilde{x}\| \leq \mathfrak{R} := \max_z\{2\|Bz\| : z \in \mathcal{X}_\mathsf{s},\, |h_i^T Az| \leq \rho_i, 1 \leq i \leq N\}$$

with confidence $1 - \epsilon$.

Small modification: with probability $1 - \epsilon$ the set $\mathcal{B} \cap \mathcal{X}$ contains x_* and thus is nonempty; however, it can be empty with positive probability.

$\Rightarrow$ *It is better to replace the rule for selecting* $\widetilde{x}$ *with*

$$\widetilde{x} \in \operatorname*{Argmin}_x \left\{\max_i |h_i^T[\omega - Ax]|/\rho_i : x \in \mathcal{X}\right\}$$

which is always well defined and results in $\widetilde{x} \in \mathcal{B} \cap \mathcal{X}$ *provided* $x_* \in \mathcal{B}$ *and thus preserves the risk bound*

$$\|Bx_* - B\widetilde{x}\| \leq \mathfrak{R} \text{ with confidence } 1 - \epsilon$$

Illustration: When $\mathcal{X}=\{x\in\mathbb{R}^n:\|x\|_1\leq 1\}$ and $A=B=I_n$, selecting $N=n$ and taking as h_i the standard basic orths, we arrive at the recovery

$$\omega\mapsto \mathop{\mathrm{Argmin}}_{x\in\mathcal{X}}\|x-\omega\|_\infty$$

and

$$\mathfrak{R}=\max_z\left\{2\|z\|_2:\underbrace{\|z\|_1\leq 1}_{z\in\mathcal{X}_{\mathrm{s}}}\ \&\ \|z\|_\infty\leq\sigma\sqrt{2\ln(2n/\epsilon)}\right\}\leq 2\sqrt{\sigma\sqrt{2\ln(2n/\epsilon)}}$$

where the concluding inequality is due to $\|z\|_2^2\leq\|z\|_1\|z\|_\infty$.

• To say that $h^T\omega$ estimates h^TAx_* within accuracy 0.1 is the same as to say that $10h^T\omega$ estimates $10h^TAx_*$ within accuracy 1. *It is technically convenient to scale h_i to make $\rho_i = 1$, that is, to ensure that*

$$\|h_i\|_2 \leq [\sigma\sqrt{2\ln(2N/\epsilon)}]^{-1}.$$

With this convention, setting $H = [h_1, ..., h_N]$, our recovering routine becomes

$$\omega \mapsto \widetilde{x} \in \mathop{\mathrm{Argmin}}_{x\in\mathcal{X}} \|H^T[\omega - Ax]\|_\infty \mapsto \widehat{x} = B\widetilde{x}$$

and the formula for $\mathfrak{R}$ becomes

$$\mathfrak{R} = \max_z \{2\|Bz\|_2 : z \in \mathcal{X}_{\mathsf{s}} \;\&\; \|H^TAz\|_\infty \leq 1\}$$

$$\mathcal{X} \subset \mathbb{R}^n \;\&\; \omega = Ax + \xi \in \mathbb{R}^m,\ \xi \sim P_x \text{ with } x \in \mathcal{X} \quad ?? \Rightarrow ?? \quad \widehat{x}(\omega) \approx Bx \in \mathbb{R}^\nu$$

Polyhedral Estimate: Construction. Generic polyhedral estimate stems from the above motivation and is as follows:
The estimate is specified by $m \times N$ *contrast matrix* H *and is given by*

$$\omega \mapsto \bar{x}(\omega) \in \underset{y \in \mathcal{X}}{\text{Argmin}} \, \|H^T[\omega - Ay]\|_\infty \mapsto \widehat{x}_H(\omega) = B\bar{x}(\omega)$$

Risk Analysis. In what follows, it is convenient to quantify the performance of a candidate estimate $\widehat{x}(\cdot)$ by its ϵ-*risk* rather the worst-case, over $x \in \mathcal{X}$, expected error. Specifically, given reliability tolerance $\epsilon \in (0,1)$, we define $(\epsilon, \|\cdot\|)$-risk of a candidate estimate $\widehat{x}(\cdot) : \mathbb{R}^m \to \mathbb{R}^\nu$ as the worst case, over $x \in \mathcal{X}$, width of "$\|\cdot\|$ – $(1-\epsilon)$-confidence interval:"

$$\text{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}(\cdot)|\mathcal{X}] = \min\left\{\rho : \text{Prob}_{\xi \sim P_x}\{\|\widehat{x}(Ax+\xi) - Bx\| > \rho\} \leq \epsilon \,\forall x \in \mathcal{X}\right\}$$

$\mathcal{X} \subset \mathbb{R}^n$ & $\omega = Ax + \xi \in \mathbb{R}^m$, $\xi \sim P_x$ with $x \in \mathcal{X}$?? $\Rightarrow$?? $\widehat{x}(\omega) \approx Bx \in \mathbb{R}^\nu$

Immediate observation: *Given reliability tolerance $\epsilon \in (0,1)$, assume that contrast matrix H satisfies*

$$\mathrm{Prob}_{\xi \sim P_x}\{\|H^T\xi\|_\infty \leq 1\} \geq 1-\epsilon \;\forall x \in \mathcal{X} \qquad (!)$$

Let $\mathcal{X}_\mathrm{s} = \frac{1}{2}[\mathcal{X} - \mathcal{X}] = \left\{\frac{1}{2}[x - x'] : x, x' \in \mathcal{X}\right\}$ *and*

$$\mathfrak{R} = \max_z \left\{2\|Bz\| : z \in \mathcal{X}_\mathrm{s}, \|H^T Az\|_\infty \leq 1\right\}$$

For the polyhedral estimate $\widehat{x}_H$ associated with the contrast matrix H we have

$$\mathrm{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}_H|\mathcal{X}] \leq \mathfrak{R}.$$

Indeed, let us fix $x \in \mathcal{X}$, and let $\mathcal{E} = \{\xi : \|H^T\xi\|_\infty \leq 1\}$, so that $P_x\{\mathcal{E}\} \geq 1-\epsilon$. For $\xi \in \mathcal{E}$, setting $\widehat{x} = \widehat{x}(Ax+\xi)$, we have $\widehat{x} = B\bar{x}$ with $\bar{x} \in \mathop{\mathrm{Argmin}}\limits_{y \in \mathcal{X}} F(y) := \|H^T[Ax + \xi - Ay]\|_\infty$

We have $x \in \mathcal{X}$ and $F(x) \leq \|H^T\xi\|_\infty \leq 1$ since $\xi \in \mathcal{E}$

$\Rightarrow$ $\bar{x} \in \mathcal{X}$ and $F(\bar{x}) \leq 1$

$\Rightarrow$ $2 \geq F(x) + F(\bar{x}) = \|H^T\xi\|_\infty + \|H^T A[x - \bar{x}] + H^T\xi\|_\infty \geq \|H^T A[x - \bar{x}]\|_\infty$

$\Rightarrow$ for $z = \frac{1}{2}[x - \bar{x}] \in \mathcal{X}_\mathrm{s}$ it holds $\|H^T Az\|_\infty \leq 1$ $\Rightarrow$ $\|Bx - \widehat{x}\| = \|Bx - B\bar{x}\| = 2\|z\| \leq \mathfrak{R}$.

$\Rightarrow$ when $x \in \mathcal{X}$ and $\xi \in \mathcal{E}$ (which happens with P_x-probability at least $1-\epsilon$) it holds

$$\|x - \widehat{x}(Ax + \xi)\| \leq \mathfrak{R}.$$

$$\mathrm{Prob}_{\xi\sim P_x}\{\|H^T\xi\|_\infty\leq 1\}\geq 1-\epsilon\ \forall x\in\mathcal{X}\qquad (!)$$

$$\Rightarrow\qquad \widehat{x}_H(\omega)=B\bar{x}(\omega),\ \bar{x}(\omega)\in\mathop{\mathrm{Argmin}}_{x\in\mathcal{X}}\|H^T[\omega-Ax]\|_\infty:$$

$$\mathrm{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}_H|\mathcal{X}]\leq\mathfrak{R}:=\max_z\left\{2\|Bz\|:z\in\mathcal{X}_{\mathrm{s}},\|H^TAz\|_\infty\leq 1\right\}.\quad (*)$$

Questions to be addressed:

A. How to define a set $\mathcal{H}_\epsilon$, the wider the better, of contrast matrices H satisfying $(!)$

B. How to upper-bound $\mathfrak{R}$ efficiently

Note: Optimization problem in $(*)$ is a difficult problem of *maximizing convex* function over a convex set.

C. How to optimize, to the largest extent possible, $\mathfrak{R}$ over $H\in\mathcal{H}_\epsilon$

A. How to define a set $\mathcal{H}_\epsilon$ of contrast matrices H satisfying

$$\mathrm{Prob}_{\xi\sim P_x}\{\|H^T\xi\|_\infty \leq 1\} \geq 1-\epsilon\ ?$$

Answering Question A. In the sequel, we restrict ourselves with 3 observation schemes:

A.I. Sub-Gaussian case: *For every $x\in\mathcal{X}$, the distribution P_x of observation noise is sub-Gaussian with parameters $(0,\sigma^2 I_m)$:*

$$\mathbf{E}_{\xi\sim P_x}\{\exp\{h^T\xi\}\} \leq \frac{\sigma^2}{2}h^T h\ \forall h.$$

Given positive integer N and setting

$$\begin{array}{c}\pi_G(h) = \vartheta_G\|h\|_2 \ \text{ where } \ \vartheta_G = \sigma\sqrt{2\ln(2N/\epsilon)},\\ \mathcal{H}_\epsilon = \mathcal{H}_\epsilon^G = \{H\in\mathbb{R}^{m\times N} : \pi_G(\mathrm{Col}_j[H]) \leq 1,\ 1\leq j\leq N\}\end{array}$$

we ensure that for every $H\in\mathcal{H}_\epsilon$ and every $(0,\sigma^2 I_m)$-sub-Gaussian ξ it holds

$$\mathrm{Prob}\{\|H^T\xi\|_\infty \leq 1\} \geq 1-\epsilon.$$

Note: $\pi_G(h)$ decreases as $O(\sigma)$ as $\sigma\to+0$

A.II. Discrete case: *$\mathcal{X}$ is a convex compact subset of the probabilistic simplex $\mathbf{\Delta}_n = \{x \in \mathbb{R}^n : x \geq 0, \sum_i x_i = 1\}$, A is column-stochastic matrix, and observation ω stemming from signal $x \in \mathcal{X}$ is*

$$\omega = \frac{1}{K}\sum_{k=1}^{K} \zeta_k$$

with independent across $k \leq K$ random vectors ζ_k, each taking values e_i with probabilities $[Ax]_i$, $i = 1, ..., m$, e_i being the basic orths in $\mathbb{R}^m$.
Setting

$$\begin{array}{c}\pi_D(h) = 2\sqrt{\vartheta_D \max_{x\in\mathcal{X}} \sum_i [Ax]_i h_i^2 + \frac{16}{9}\vartheta_D^2 \|h\|_\infty^2} \text{ with } \vartheta_D = \frac{\ln(2N/\epsilon)}{K},\\ \mathcal{H}_\epsilon = \mathcal{H}_\epsilon^D := \{H \in \mathbb{R}^{m\times N} : \pi_D(\mathrm{Col}_j[H]) \leq 1, j \leq m\},\end{array}$$

we ensure that for every $H \in \mathcal{H}_\epsilon$ and every $x \in \mathcal{X}$, for the *zero mean i.i.d. random noise* $\xi_x = \omega - Ax$, with the above ω, it holds

$$\mathrm{Prob}\{\|H^T \xi_x\|_\infty \leq 1\} \geq 1 - \epsilon.$$

Note: $\pi_D(h)$ decreases as $O(1/\sqrt{K})$ as K grows

Note: The crucial role in the justification of the above bounds on probabilities of large deviations of histograms from true distributions is played by the fundamental

Bernstein Inequality: *Let $X_1, ..., X_N$ be independent zero mean random variables with variations $\sigma_1^2, ..., \sigma_N^2$ such that $|X_i| \leq M < \infty$ for all i and some M. Then for every $t \geq 0$ one has*

$$\mathrm{Prob}\left\{\sum_{i=1}^{N} X_i \geq t\right\} \leq \exp\left\{-\frac{t^2}{2\left[\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{3}Mt\right]}\right\}.$$

A.III. Poisson case: *$\mathcal{X}$ is a convex compact subset of the nonnegative orthant $\mathbb{R}^n_+$, A is entrywise nonnegative, and the observation ω stemming from $x \in \mathcal{X}$ is random vector with independent across i entries $\omega_i \sim$ Poisson$([Ax]_i)$.*

In the Poisson case we set

$$\pi_P(h) = 2\sqrt{\vartheta_P \max_{x\in\mathcal{X}} \sum_i [Ax]_i h_i^2 + \tfrac{4}{9}\vartheta_P^2 \|h\|_\infty^2} \text{ with } \vartheta_P = \ln(2N/\epsilon),$$
$$\mathcal{H}_\epsilon = \mathcal{H}_\epsilon^P := \{H \in \mathbb{R}^{m\times N} : \pi_P(\mathrm{Col}_j[H]) \le 1, 1 \le j \le N\}.$$

thus ensuring that for every $H \in \mathcal{H}_\epsilon$ and every $x \in \mathcal{X}$, for the *zero mean* random noise $\xi_x = \omega - Ax$, with the above ω, it holds

$$\mathrm{Prob}\{\|H^T\xi_x\|_\infty \le 1\} \ge 1 - \epsilon.$$

Note: In all 3 cases, the set $\mathcal{H}_\epsilon$ of "legitimate" in our context $m \times N$ contrast matrices is of the form

$$\mathcal{H}_\epsilon = \{H \in \mathbb{R}^{m\times N} : \pi(\mathrm{Col}_j(H)) \le 1, j \le M\}$$

where $\pi(\cdot)$ is norm of the form

$$\pi(h) = \sqrt{\alpha \max_{y\in Y} \sum_i y_i h_i^2 + \beta \|h\|_\infty^2\}} \qquad [Y \subset \mathbb{R}^m_+\text{: convex compact set}]$$

with $\alpha > 0$, $\beta \ge 0$ logarithmically depending on N/ϵ.

$$\text{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}_H|\mathcal{X}] \le \mathfrak{R}[H] = \max_z \left\{2\|Bz\| : z \in \mathcal{X}_s, \|H^T Az\|_\infty \le 1\right\}$$

B. How to upper-bound $\mathfrak{R}[H]$? **C.** How to optimize $\mathfrak{R}[H]$ over H ?

Answering Questions B, C, Version I

♠ The reference case for what follows is the one of $\|\cdot\| = \|\cdot\|_\infty$. In this case $\mathfrak{R}[H]$ is easy to compute by solving ν convex optimization problems

$$\begin{array}{rcl}\varsigma_\ell[H] & = & \max_z \left\{[Bz]_\ell : z \in \mathcal{X}_s, \|H^T Az\|_\infty \le 1\right\} \\ & = & \max_z \left\{|[Bz]_\ell| : z \in \mathcal{X}_s, \|H^T Az\|_\infty \le 1\right\},\end{array}$$

$\ell = 1, ..., \nu$, and taking the maximum of their optimal values as $\frac{1}{2}\mathfrak{R}[H]$.

♠ Assume that we restrict H to be an $m \times N$ matrix with a given $N \ge \nu$ satisfying, for a given norm $\pi(\cdot)$, the constraints

$$\pi(\text{Col}_j[H]) \le 1,\ 1 \le j \le N. \qquad (*)$$

It turns out that *under constraints $(*)$ on H, it is easy to minimize simultaneously all $\varsigma_\ell[H]$, $\ell \le \nu$, over H.*

Note: In the observation schemes we are considering, the design restriction $H \in \mathcal{H}_\epsilon$ on a candidate contrast matrix H indeed is given by constraints $(*)$ with appropriate norm π !

$$\varsigma_\ell[H] = \max_z \left\{[Bz]_\ell : z \in \mathcal{X}_s, \|H^T Az\|_\infty \le 1\right\}, \ell = 1, ..., \nu$$
$$\left[H \in \mathbb{R}^{m\times N} : \pi(\mathrm{Col}_j[H]) \le 1, 1 \le j \le N \,\&\, N \ge \nu\right]$$

Optimizing $\varsigma_\ell[H]$ over H

♠ Given a vector $b \in \mathbb{R}^n$ and a norm $\pi(\cdot)$ on $\mathbb{R}^m$, consider convex-concave saddle point problem

$$\mathrm{Opt}[b] = \inf_{g\in\mathbb{R}^m} \max_{x\in\mathcal{X}_s} \left\{\phi(g, x) := [b - A^T g]^T x + \pi(g)\right\} \qquad (SP)$$

Claim: *(SP) has a saddle point. This saddle point induces vector $\bar h = \bar h[b] \in \mathbb{R}^m$ with $\pi(\bar h) = 1$ such that $\max_x \left\{|b^T x| : x \in \mathcal{X}_s, |\bar h^T Ax| \le 1\right\} \le \mathrm{Opt}[b]$. In addition, for any matrix $G = [g^1, ..., g^M] \in \mathbb{R}^{m\times M}$ with $\pi(g^j) \le 1$, $1 \le j \le M$, one has*

$$\begin{array}{rcl}\max_x \left\{|b^T x| : x \in \mathcal{X}_s, \|G^T Ax\|_\infty \le 1\right\} &=& \max_x \left\{b^T x : x \in \mathcal{X}_s, \|G^T Ax\|_\infty \le 1\right\}\\ &\ge& \mathrm{Opt}[b].\end{array}$$

Corollary: *Let $\overline{H}$ be the $m \times \nu$ matrix with the columns $\bar h_\ell = \bar h[B_\ell]$, where B_ℓ^T is ℓ-th row of B, $1 \le \ell \le \nu$. Then $\pi(\mathrm{Col}_j[\overline{H}]) \le 1$, $j \le \nu$, and $\overline{H}$ minimizes simultaneously all quantities $\varsigma_\ell[H]$, $\ell \le \nu$, over $m \times N$ contrast matrices H satisfying $\pi(\mathrm{Col}_j[H]) \le 1$, $1 \le j \le N$. The resulting value of ς_ℓ is $\mathrm{Opt}[B_\ell]$, $\ell \le \nu$.*

Building $\bar{h}$: The convex-concave saddle point problem

$$\mathrm{Opt}[b] = \inf_{g\in\mathbb{R}^m} \max_{x\in\mathcal{X}_s} \left\{\phi(g,x) := [b - A^T g]^T x + \pi(g)\right\} \qquad (SP)$$

induces primal and dual problems

$$\begin{array}{rcll}
\mathrm{Opt}(P) & = & \inf_{g\in\mathbb{R}^m} \left[\overline{\phi}(g) := \max_{x\in\mathcal{X}_s} \phi(g,x)\right] & (P)\\
& = & \inf_{g\in\mathbb{R}^m} \left[\pi(g) + \max_{x\in\mathcal{X}_s} [b - A^T g]^T x\right], & \\
\mathrm{Opt}(D) & = & \max_{x\in\mathcal{X}_s} \left[\underline{\phi}(g) := \inf_{g\in\mathbb{R}^m} \phi(g,x)\right] & (D)\\
& = & \max_{x\in\mathcal{X}_s} \left[\inf_{g\in\mathbb{R}^m} \left[b^T x - [Ax]^T g + \pi(g)\right]\right] & \\
& = & \max_x \left[b^T x : \ x \in \mathcal{X}_s, \ \theta(Ax) \le 1\right] &
\end{array}$$

where $\theta(\cdot)$ is the norm conjugate to $\pi(\cdot)$ (we have used the evident fact that $\inf_{g\in\mathbb{R}^m}[f^T g + \pi(g)]$ is either $-\infty$ or 0 depending on whether $\theta(f) > 1$ or $\theta(f) \le 1$). Since $\mathcal{X}_s$ is compact, we have $\mathrm{Opt}(P) = \mathrm{Opt}(D) = \mathrm{Opt}[b]$ by the Sion-Kakutani theorem. Besides this, (D) is solvable (this is evident) and (P) is solvable as well, since $\overline{\phi}(g)$ is continuous due to the compactness of $\mathcal{X}_s$, and $\overline{\phi}(g) \ge \pi(g)$, so that $\overline{\phi}(\cdot)$ has bounded level sets. Let $\bar{g}$ be an optimal solution to (P). We select $\bar{h} = \bar{h}[b] \in \mathbb{R}^m$ in such a way that

$$\bar{g} = \pi(\bar{g})\bar{h} \ \& \ \pi(\bar{h}) = 1.$$

♠ The construction just outlined basically resolves the question of how to build the "legitimate" contrast matrix leading to the best, in terms of its risk bound, polyhedral estimate, *provided that the recovery norm is* $\|\cdot\|_\infty$.

♠ In fact, this construction has other consequences. Let us make the following assumptions:

A.1. *The recovery norm is* $\|\cdot\| = \|\cdot\|_r$ *with some* $r \in [1,\infty]$

A.2. *We have at our disposal a sequence* $\gamma = \{\gamma_i > 0,\, 1 \leq i \leq \nu\}$ *and* $\rho \in [1,\infty]$ *such that the image of* $\mathcal{X}_\mathsf{s}$ *under the mapping* $x \mapsto Bx$ *is contained in the "scaled* $\|\cdot\|_\rho$*-ball"*

$$\mathcal{Y} = \{y \in \mathbb{R}^\nu : \|\mathrm{Diag}\{\gamma\} y\|_\rho \leq 1\}.$$

Observation: *Let B_ℓ^T be ℓ-th row in B, $1 \leq \ell \leq \nu$. Under assumptions **A.1-2**, let $\epsilon \in (0,1)$ and a positive real $N \geq \nu$ be given, and let $\pi(\cdot)$ be a norm on $\mathbb{R}^m$ such that*

$$\forall (h:\ \pi(h) \leq 1, x \in \mathcal{X}) : \mathrm{Prob}\{|h^T \xi_x| \leq 1\} \geq 1 - \epsilon/N.$$

Let, next, an $m \times N$ matrix H and positive reals ς_ℓ, $1 \leq \ell \leq \nu$, satisfy the relations

$$\begin{array}{ll} (a) & \pi(\mathrm{Col}_j[H]) \leq 1,\ 1 \leq j \leq N; \\ (b) & \max_x \left\{ B_\ell^T x :\ x \in \mathcal{X}_\mathsf{s},\ \|H^T A x\|_\infty \leq 1 \right\} \leq \varsigma_\ell,\ 1 \leq \ell \leq \nu. \end{array}$$

Then the quantity

$$\mathfrak{R}[H] = \max_z \left\{ 2\|Bz\| : z \in \mathcal{X}_\mathsf{s}, \|H^T A z\|_\infty \leq 1 \right\}$$

can be upper-bounded as follows:

$$\mathfrak{R}[H] \leq \Psi(\varsigma) := 2 \max_v \left\{ \|[v_1/\gamma_1; ...; v_\nu/\gamma_\nu]\|_r :\ \|v\|_\rho \leq 1,\ 0 \leq v_\ell \leq \gamma_\ell \varsigma_\ell,\ 1 \leq \ell \leq \nu \right\}.$$

implying that

$$\mathrm{Risk}_{\epsilon, \|\cdot\|}[\widehat{w}_H | \mathcal{X}] \leq \Psi(\varsigma).$$

Function Ψ is nondecreasing on the nonnegative orthant and is easy to compute.

Note: We know how to make all ς_ℓ as small as possible under the restriction

$$\pi(\mathrm{Col}_j[H]) \leq 1,\ 1 \leq j \leq N;$$

we should select as H the $m \times N$ matrix with the columns $\bar{h}[B_\ell]$, $1 \leq \ell \leq \nu$, and, say, zero columns with indexes $> \nu$, resulting in

$$\varsigma_\ell = \mathrm{Opt}[B_\ell] := \inf_{g \in \mathbb{R}^m} \max_{x \in \mathcal{X}_{\mathrm{s}}} \left\{ \phi(g,x) := B_\ell^T x - g^T A x + \pi(g) \right\}$$

where B_ℓ^T is ℓ-th row of B.

Note: There is no reason to use $N > \nu$; $N = \nu$ already results in the best legitimate contrast.

Note: An attractive feature of the contrast design we have just developed is that it is *completely independent* of the entities participating in Assumptions **A.1-2** – these entities affect theoretical risk bounds of the resulting polyhedral estimate, *but not the estimate itself*.

Near-optimality. Unfortunately, for the proposed polyhedral estimate no really general results on near-optimality are known.

However: *There are important special cases* where near-optimality can be justified, most notably,

Simple diagonal case (one of the typical cases considered in the traditional Nonparametric Statistics), where

• $\mathcal{X} = \{x \in \mathbb{R}^n : \|Dx\|_\rho \leq 1\}$, where $D = \mathrm{Diag}\{\ell^\delta, \ell = 1, 2, ..., n\}$,
• $\|\cdot\| = \|\cdot\|_r$ with $1 \leq \rho \leq r < \infty$,
• $m = \nu = n$, $A = \mathrm{Diag}\{\ell^{-\alpha}, \ell = 1, ..., n\}$, $B = \mathrm{Diag}\{\ell^{-\beta}, \ell = 1, ..., n\}$,

with

$$\beta \geq \alpha \geq 0,\ \delta \geq 0 \ \&\ (\beta - \alpha) r < 1$$

• We are in Sub-Gaussian case: ξ_x is $(0, \sigma^2 I_n)$-sub-Gaussian, $x \in \mathcal{X}$.

Assuming that σ, ϵ, n are in the range $0 < \sqrt{\ln(2n/\epsilon)}\sigma \leq 1$ and n is large enough:

$$n \geq c\vartheta_G^{-\frac{1}{\alpha+\delta+1/\rho}} \qquad [\vartheta_G = \sigma\sqrt{2\ln(2n/\epsilon)}]$$

(here and what follows c and C depend solely on $\alpha, \beta, \delta, r, \rho$) our design results in

$$H = [\sigma\varkappa]^{-1} I_n \text{ with } \varkappa = \sqrt{2\ln(2n/\epsilon)}$$
$$\mathrm{Risk}_{\epsilon,\|\cdot\|_r}[\widehat{x}_H|\mathcal{X}] \leq C\,[\sigma\varkappa]^\varphi,\ \varphi = \frac{\beta + \delta + 1/\rho - 1/r}{\alpha + \delta + 1/\rho},$$

while the minimax optimal $(\epsilon, \|\cdot\|_r)$-risk is $\geq c\sigma^\varphi$.

$\Rightarrow$ *the risk of our polyhedral estimate is within logarithmic in n/ϵ factor of the minimax optimal risk.*

Not so good news: The above near-optimality result is obtained by the traditional for classical Non-Parametric Statistics *analytical closed form* risk analysis, this is where heavy structural restrictions on $\mathcal{X}$, A, and B come from.

Paying debts for Version I: Proofs

♠ **Observation to be verified:** *Let B_ℓ^T be ℓ-th row in B, $1\leq\ell\leq\nu$. Under assumptions*

$$\textbf{A.1:}\quad \|\cdot\|=\|\cdot\|_r \qquad \textbf{A.2:}\quad B\mathcal{X}_{\mathsf{s}}\subset\mathcal{Y}=\{y:\|\mathrm{Diag}\{\gamma\}y\|_\rho\leq 1\},$$

let $\epsilon\in(0,1)$ and a positive real $N\geq\nu$ be given, and let $\pi(\cdot)$ be a norm on $\mathbb{R}^m$ such that

$$\forall(h:\ \pi(h)\leq 1, x\in\mathcal{X}):\mathrm{Prob}\{|h^T\xi_x|\leq 1\}\geq 1-\epsilon/N.$$

Let, next, an $m\times N$ matrix H and positive reals ς_ℓ, $1\leq\ell\leq\nu$, satisfy the relations

$$\begin{array}{ll}(a) & \pi(\mathrm{Col}_j[H])\leq 1,\ 1\leq j\leq N;\\ (b) & \max_x\left\{B_\ell^Tx:\ x\in\mathcal{X}_{\mathsf{s}},\ \|H^TAx\|_\infty\leq 1\right\}\leq\varsigma_\ell,\ 1\leq\ell\leq\nu.\end{array}$$

Then the quantity

$$\mathfrak{R}[H]=\max_z\left\{2\|Bz\|:z\in\mathcal{X}_{\mathsf{s}},\|H^TAz\|_\infty\leq 1\right\}$$

can be upper-bounded as follows:

$$\mathfrak{R}[H]\leq\Psi(\varsigma):=2\max_v\left\{\|[v_1/\gamma_1;...;v_\nu/\gamma_\nu]\|_r:\ \|v\|_\rho\leq 1,\ 0\leq v_\ell\leq\gamma_\ell\varsigma_\ell,\ 1\leq\ell\leq\nu\right\}.$$

implying that

$$\mathrm{Risk}_{\epsilon,\|\cdot\|}[\widehat{w}_H|\mathcal{X}]\leq\Psi(\varsigma).$$

Function Ψ is nondecreasing on the nonnegative orthant and is easy to compute.

Proof. Let $\bar{z} \in \mathcal{X}_{\mathsf{s}}$ and $\|H^T A\bar{z}\|_\infty \leq 1$. Setting $y = B\bar{z}$, we have $y \in \mathcal{Y}$ due to $\bar{z} \in \mathcal{X}_{\mathsf{s}}$ and **A.2**. Thus, $\|\mathrm{Diag}\{\gamma\}y\|_p \leq 1$. Besides this, by (b) relations $\bar{z} \in \mathcal{X}_{\mathsf{s}}$ and $\|H^T A\bar{z}\|_\infty \leq 1$ combine with the symmetry of $\mathcal{X}_{\mathsf{s}}$ to imply that $|y_\ell| = |B_\ell^T \bar{z}| \leq \varsigma_\ell$, $\ell \leq \nu$. Taking into account that $\|\cdot\| = \|\cdot\|_r$ by **A.1**, we see that

$$\begin{array}{rcl}\mathfrak{R}[H] & = & \max_z \left\{2\|Bz\|_r : z \in \mathcal{X}_{\mathsf{s}}, \|H^T Az\|_\infty \leq 1\right\} \\ & \leq & 2\max_y \{\|y\|_r : |y_\ell| \leq \varsigma_\ell, \ell \leq \nu \ \&\ \|\mathrm{Diag}\{\gamma\}y\|_\rho \leq 1\} \\ & = & 2\max_v \{\|[v_1/\gamma_1; ...; v_\nu/\gamma_\nu]\|_r : \|v\|_\rho \leq 1, 0 \leq v_\ell \leq \gamma_\ell\varsigma_\ell, \ell \leq \nu\} = \Psi(\varsigma),\end{array}$$

as claimed. It is evident that Ψ is nondecreasing on the nonnegative orthant, and it is easy to verify that Ψ is efficiently computable. $\square$

♠ **Claim to be verified:** *Given a vector* $b \in \mathbb{R}^n$ *and a norm* $\pi(\cdot)$ *on* $\mathbb{R}^m$, *consider convex-concave saddle point problem*

$$\mathrm{Opt}[b] = \inf_{g\in\mathbb{R}^m} \left[\pi(g) + \max_{x\in\mathcal{X}_{\mathsf{S}}} [b - A^T g]^T x\right] \qquad (SP)$$

(SP) *has a saddle point. The* g*-component* $\bar{g}$ *of a saddle point induces vector* $\bar{h} = \bar{h}[b]$ *given by*

$$\bar{g} = \pi(\bar{g})\bar{h} \ \& \ \pi(\bar{h}) = 1$$

such that

$$\max_x \left\{|b^T x| : x \in \mathcal{X}_{\mathsf{S}}, |\bar{h}^T A x| \leq 1\right\} \leq \mathrm{Opt}[b].$$

In addition, for any matrix $G = [g^1, ..., g^M] \in \mathbb{R}^{m\times M}$ *with* $\pi(g^j) \leq 1$, $1 \leq j \leq M$, *one has*

$$\begin{array}{rcl} \max_x \left\{|b^T x| : x \in \mathcal{X}_{\mathsf{S}}, \|G^T A x\|_\infty \leq 1\right\} & = & \max_x \left\{b^T x : x \in \mathcal{X}_{\mathsf{S}}, \|G^T A x\|_\infty \leq 1\right\} \\ & \geq & \mathrm{Opt}[b]. \end{array}$$

Proof, Step 1: Building $\bar{h}$. The induced by the convex-concave saddle point problem

$$\mathrm{Opt}[b] = \inf_{g\in\mathbb{R}^m} \max_{x\in\mathcal{X}_S} \left\{\phi(g,x) := [b - A^T g]^T x + \pi(g)\right\} \tag{SP}$$

primal and dual problems are

$$\begin{array}{rcll}
\mathrm{Opt}(P) & = & \inf_{g\in\mathbb{R}^m}\left[\overline{\phi}(g) := \max_{x\in\mathcal{X}_S}\phi(g,x)\right] & (P)\\
& = & \inf_{g\in\mathbb{R}^m}\left[\pi(g) + \max_{x\in\mathcal{X}_S}[b - A^T g]^T x\right], & \\
\mathrm{Opt}(D) & = & \max_{x\in\mathcal{X}_S}\left[\underline{\phi}(g) := \inf_{g\in\mathbb{R}^m}\phi(g,x)\right] & (D)\\
& = & \max_{x\in\mathcal{X}_S}\left[\inf_{g\in\mathbb{R}^m}\left[b^T x - [Ax]^T g + \pi(g)\right]\right] & \\
& = & \max_x\left[b^T x:\ x\in\mathcal{X}_S,\ \theta(Ax)\le 1\right] &
\end{array}$$

where $\theta(\cdot)$ is the norm conjugate to $\pi(\cdot)$ (we have used the evident fact that $\inf_{g\in\mathbb{R}^m}[f^T g + \pi(g)]$ is either $-\infty$ or 0 depending on whether $\theta(f) > 1$ or $\theta(f)\le 1$). Since $\mathcal{X}_S$ is compact, we have $\mathrm{Opt}(P) = \mathrm{Opt}(D) = \mathrm{Opt}[b]$ by the Sion-Kakutani theorem. Besides this, (D) is solvable (this is evident) and (P) is solvable as well, since $\overline{\phi}(g)$ is continuous due to the compactness of $\mathcal{X}_S$, and $\overline{\phi}(g)\ge\pi(g)$, so that $\overline{\phi}(\cdot)$ has bounded level sets. Let $\bar{g}$ be an optimal solution to (P). $\bar{h}$ is the vector given by

$$\bar{g} = \pi(\bar{g})\bar{h}\ \&\ \pi(\bar{h}) = 1.$$

$$\begin{array}{rcll}
\mathrm{Opt}[b] & = & \inf_{g\in\mathbb{R}^m}\left[\overline{\phi}(g) := \max_{x\in\mathcal{X}_{\mathsf{S}}}\phi(g,x) := [b-A^Tg]^Tx+\pi(g)\right] & (P)\\
& = & \inf_{g\in\mathbb{R}^m}\left[\pi(g)+\max_{x\in\mathcal{X}_{\mathsf{S}}}[b-A^Tg]^Tx\right], & \\
& = & \max_{x\in\mathcal{X}_{\mathsf{S}}}\left[\underline{\phi}(g) := \inf_{g\in\mathbb{R}^m}\phi(g,x)\right] & \\
& = & \max_{x\in\mathcal{X}_{\mathsf{S}}}\left[\inf_{g\in\mathbb{R}^m}\left[b^Tx-[Ax]^Tg+\pi(g)\right]\right] & \\
& = & \max_x\left[b^Tx:\ x\in\mathcal{X}_{\mathsf{S}},\ \theta(Ax)\le 1\right] & (D)
\end{array}$$

where $\theta(\cdot)$ is the norm conjugate to $\pi(\cdot)$.

Proof, Step 2. To justify Claim we are proving, it remains to verify the following

Fact: *When $\bar{g}=\pi(\bar{g})\bar{h}$, $\pi(\bar{h})=1$, is an optimal solution (which does exist) to (P), one has*

$$\max_x\left\{|b^Tx| : x\in\mathcal{X}_{\mathsf{S}}, |\bar{h}^TAx|\le 1\right\}\le \mathrm{Opt}[b], \tag{1}$$

and for any matrix $G=[g^1,...,g^M]\in\mathbb{R}^{m\times M}$ with $\pi(g^j)\le 1$, $1\le j\le M$, one has

$$\max_x\left\{|b^Tx| : x\in\mathcal{X}_{\mathsf{S}}, \|G^TAx\|_\infty\le 1\right\} = \max_x\left\{b^Tx : x\in\mathcal{X}_{\mathsf{S}}, \|G^TAx\|_\infty\le 1\right\}\ge \mathrm{Opt}[b]. \tag{2}$$

Justifying Fact: Let x be a feasible solution to the optimization problem in (1). Replacing, if necessary, x with $-x$, we can assume that $|b^Tx|=b^Tx$. We now have

$$\begin{array}{l}
|b^Tx| = b^Tx = [\bar{g}^TAx-\pi(\bar{g})]+\underbrace{[b-A^T\bar{g}]^Tx+\pi(\bar{g})}_{\le\overline{\phi}(\bar{g})=\mathrm{Opt}[b]}\le \mathrm{Opt}[b]+[\pi(\bar{g})\bar{h}^TAx-\pi(\bar{g})]\\
\le \mathrm{Opt}[b]+\pi(\bar{g})\underbrace{|\bar{h}^TAx|}_{\le 1}-\pi(\bar{g})\le \mathrm{Opt}[b],
\end{array}$$

as claimed in (1). The equality in (2) is due to the symmetry of $\mathcal{X}_{\mathsf{S}}$ w.r.t. the origin. To verify the inequality in (2), let $\bar{x}$ be an optimal solution to (D), so that $\bar{x}\in\mathcal{X}_{\mathsf{S}}$ and $\theta(A\bar{x})\le 1$, implying, due to the fact that the columns of G are of $\pi(\cdot)$-norm ≤ 1, that $\bar{x}$ is a feasible solution to the optimization problem in (2). As a result, the second quantity in (2) is at least $b^T\bar{x}=\mathrm{Opt}[b]$, and (2) follows. $\square$

Answering Questions B, C, Version II

♣ Our second approach to **B**, **C** resembles what we did when building linear estimates – it is based on a kind of semidefinite relaxation

♠ **Definition.** *Given a nonempty convex compact set $\mathcal{Y} \in \mathbb{R}^N$, we say that $\mathbf{Y}$ is compatible with $\mathcal{Y}$, if $\mathbf{Y} = \{(V,\tau)\}$ is a closed convex cone contained in $\mathbf{S}^N_+ \times \mathbb{R}_+$ and such that*

— $\forall (V,\tau) \in \mathbf{Y} : \max_{y\in\mathcal{Y}} y^T V y \le \tau$

— *relations $(V,\tau) \in \mathbf{Y}$ and $\tau' \ge \tau$ imply that $(V,\tau') \in \mathbf{Y}$*

— *$\mathbf{Y}$ contains a pair (V,τ) with $V \succ 0$.*

• We say that a cone $\mathbf{Y}$ compatible with $\mathcal{Y}$ is *sharp*, if the only pair $(V,0) \in \mathbf{Y}$ is with $V = 0$.

Example: When $\mathcal{Y} = \{y \in \mathbb{R}^n : \|y\|_2 \le 1\}$, the cone

$$\mathbf{Y} = \{(V,\tau) : V \in \mathbf{S}^n_+, V \preceq \tau I_n\},$$

is the largest cone compatible with $\mathcal{Y}$, and this cone is sharp.

Fact: *When* $\text{Lin}(\mathcal{Y}) = \mathbb{R}^N$*, every cone compatible with* $\mathcal{Y}$ *is sharp*

Fact: *When* $\mathbf{Y}$ *is compatible with a shift of* $\mathcal{Y}$*,* $\mathbf{Y}$ *is compatible with* $\mathcal{Y}_{\text{s}} = \frac{1}{2}[\mathcal{Y} - \mathcal{Y}]$

Indeed, $\mathcal{Y}_s$ remains intact when shifting $\mathcal{Y}$, so that we can assume that $\mathbf{Y}$ is compatible with $\mathcal{Y}$. When $(V, \tau) \in \mathbf{Y}$ and $y, y' \in \mathcal{Y}$, we have $\frac{1}{4}(y - y')^T V(y - y') + \frac{1}{4}\underbrace{(y + y')^T V(y + y')}_{\geq 0} = \frac{1}{2}[y^T Vy + (y')^T Vy'] \leq \tau \Rightarrow \frac{1}{4}(y - y')V(y - y') \leq \tau,\ y, y' \in Y$. $\square$

♠ The role of compatibility in our context stems from the following

Observation: *Assume that we have at our disposal cones* $\mathbf{X}$ *and* $\mathbf{V}$ *compatible, respectively, with* $\mathcal{X}_s$ *and with the unit ball* $\mathcal{B}_* = \{v \in \mathbb{R}^\nu : \|u\|_* \leq 1\}$ *of the norm* $\|\cdot\|_*$ *conjugate to the norm* $\|\cdot\|$ *in which we measure the recovery error. Given contrast matrix* $H = [h_1, h_2, ..., h_N]$ *satisfying*

$$\mathrm{Prob}_{\xi\sim P_x}\{\|H^T\xi\|_\infty \leq 1\} \geq 1-\epsilon \;\forall x \in \mathcal{X} \qquad (!)$$

let

$$\mathrm{Opt}(H) = \min_{\lambda,(U,\mu),(V,\tau)} \left\{ 4\sum_j \lambda_j + 4\mu + \tau : \begin{array}{c} \lambda \in \mathbb{R}^N_+, (U,\mu) \in \mathbf{X}, (V,\tau) \in \mathbf{V} \\ \left[\begin{array}{c|c} V & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T H \mathrm{Diag}\{\lambda\} H^T A + U \end{array}\right] \succeq 0 \end{array} \right\}$$

$\mathrm{Opt}(H)$ *is an efficiently computable upper bound on the quantity*

$$\mathfrak{R} = \max_z \left\{ 2\|Bz\| : z \in \mathcal{X}_s, \|H^T A z\|_\infty \leq 1 \right\} \qquad (\#)$$

and thus. due to $(!)$*– upper bound on the* $(\epsilon, \|\cdot\|)$*-risk of the polyhedral estimate* $\widehat{x}_H(\cdot)$ *on* $\mathcal{X}$*.*

When $\mathbf{X}$ *and* $\mathbf{V}$ *are sharp, the optimization problem specifying* $\mathrm{Opt}(H)$ *is solvable.*

Situation: $\mathbf{X}$ is compatible with $\mathcal{X}_{\mathsf{s}}$, $\mathbf{V}$ is compatible with $\mathcal{B}_*$, $H=[h_1,...,h_N]$,

$$\mathrm{Opt}(H)=\min_{\lambda,(U,\mu),(V,\tau)}\left\{4\sum_j\lambda_j+4\mu+\tau:\begin{array}{c}\lambda\in\mathbb{R}^N_+,(U,\mu)\in\mathbf{X},(V,\tau)\in\mathbf{V}\\ \left[\begin{array}{c|c}V&\frac{1}{2}B\\ \hline \frac{1}{2}B^T&A^TH\mathrm{Diag}\{\lambda\}H^TA+U\end{array}\right]\succeq 0\end{array}\right\}\qquad(*)$$

$$\mathfrak{R}=\max_z\left\{2\|Bz\|:z\in\mathcal{X}_{\mathsf{s}},\|H^TAz\|_\infty\leq 1\right\}\qquad(\#)$$

Claim: $\mathfrak{R}\leq\mathrm{Opt}(H)$

Immediate reason: *When $\lambda\geq 0$, the bunch of two-sided linear inequalities $\|H^TAz\|_\infty\leq 1$ in $(\#)$ implies*, by taking weighted sum of their squares, *that $z^TA^TH\mathrm{Diag}\{\lambda\}H^TAz\leq\sum_j\lambda_j$ on the feasible set of $(\#)$.* The rest is readily given by the semidefinite constraint in $(*)$.

Formal proof: Let $\lambda,(U,\mu),(V,\tau)$ be a feasible solution to $(*)$ and z be a feasible solution to $(\#)$. Setting $w=2z$, we have $w\in 2\mathcal{X}_{\mathsf{s}}$ and $\|H^TAw\|_\infty\leq 2$. Let $u\in\mathcal{B}_*$. By the semidefinite constraint in $(*)$ we have

$$u^TBw\leq u^TVu+w^TA^TH\mathrm{Diag}(\lambda)H^TAw+w^TUw=u^TVu+\sum_j\lambda_j\underbrace{(h_j^TAw)^2}_{\leq 4}+w^TUw$$

$$\leq\tau+4\sum_j\lambda_j+4\mu.$$

Taking supremum over $u\in\mathcal{B}_*$, we get $2\|Bz\|\leq\tau+4\sum_j\lambda_j+4\mu$ for every feasible solution z to $(\#)$ $\Rightarrow$ $\mathfrak{R}\leq\tau+4\sum_j\lambda_j+4\mu$. Since $\lambda,(U,\mu),(V,\tau)$ is an arbitrary feasible solution to $(*)$, we get $\mathfrak{R}\leq\mathrm{Opt}(H)$. $\square$

$$
\boxed{\begin{array}{c}
H\in\mathbb{R}^{m\times N}: \forall x\in\mathcal{X}: \mathrm{Prob}_{\xi\sim P_x}\{\|H^T\xi\|_\infty\leq 1\}\geq 1-\epsilon \qquad (!)\\
\mathbf{X} \text{ is compatible with } \mathcal{X}_{\mathrm{S}},\ \mathbf{V} \text{ is compatible with } \mathcal{B}_*\\
\hline
\Rightarrow \mathrm{Opt}(H)=\min\limits_{\lambda,(U,\mu),(V,\tau)}\left\{4\sum_j\lambda_j+4\mu+\tau: \begin{array}{c}\lambda\in\mathbb{R}^N_+,(U,\mu)\in\mathbf{X},(V,\tau)\in\mathbf{V}\\ \left[\begin{array}{c|c}V & \frac{1}{2}B\\ \hline \frac{1}{2}B^T & A^THD\mathrm{iag}\{\lambda\}H^TA+U\end{array}\right]\succeq 0\end{array}\right\} \qquad (*)\\
\hline
\Rightarrow \widehat{x}_H(\omega)=B\,\mathop{\mathrm{Argmin}}\limits_{x\in\mathcal{X}}\|H^T[Ax-\omega]\|_\infty
\end{array}}
$$

$$\Downarrow$$

$$\boxed{\mathrm{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}_H|\mathcal{X}]\leq\mathrm{Opt}(H)}$$

♣ What is ahead:

In sub-Gaussian/Discrete/Poisson o.s., to enforce $(!)$ we impose on the columns h_j of H the restriction $\pi(h_j)\leq 1$, with adjusted to N, ϵ, and the o.s. norm π, thus defining the set

$$\mathcal{H}=\{H=[h_1,...,h_N]\in\mathbb{R}^{m\times N}:\pi(h_j)\leq 1, j\leq N\}$$

of “legitimate” contrasts. What matters are not the contrasts $H\in\mathcal{H}$ *per se*, but the conic set

$$\mathbf{H}_*=\{(G,\mu):\exists\lambda\geq 0, h_1,...,h_N: G=\sum_j\lambda_jh_jh_j^T, \pi(h_j)\leq 1\,\forall j, \sum_j\lambda_j\leq\mu\}$$

of pairs $(A^THD\mathrm{iag}\{\lambda\}H^TA,\sum_j\lambda_j)$ we can get from $H\in\mathcal{H}$ and $\lambda\geq 0$ and thus can use in $(*)$.

♠ **Questions** to be addressed:

I. *How to build a tight inner approximation of (usually difficult to handle) set $\mathbf{H}_*$ by something appropriate for optimizing $\mathrm{Opt}(H)$ over H* (which now becomes optimization over (G,μ)) ?

II. *How to build cones $\mathbf{X}$, $\mathbf{V}$, the larger the better, compatible with $\mathcal{X}_{\mathrm{S}}$, $\mathcal{B}_*$?*

Question: Given a norm π on $\mathbb{R}^m$ and positive integer N, how to build a tight inner approximation of the conic set

$$\mathbf{H}_* = \{(G,\mu) : \exists \lambda \geq 0, h_1, ..., h_N : G = \sum_j \lambda_j h_j h_j^T,\ \pi(h_j) \leq 1\ \forall j, \sum_j \lambda_j \leq \mu\}$$

by something appropriate for subsequent optimization over this something?

Fact: *The norm $\pi(\cdot)$ associated with sub-Gaussian/Discrete/Poisson case is of special form:*

$$\pi^2(h) = \theta([h]^2),\ \theta(u) = \max_{z\in\mathcal{Z}} z^T u,\ \ [[h_1; ...; h_m]]^2 = [h_1^2; h_2^2; ...; h_m^2], \qquad (!)$$

where $\mathcal{Z}$ is a convex compact subset of $\mathbb{R}^m_+$ with a nonempty interior.

Assumption: *From now on we assume that $\pi(\cdot)$ is given by $(!)$, and that $N \geq m$.*

Observation: *When the columns h_j of an $m \times N$ matrix H satisfy $\pi(h_j) \leq 1$, and $\lambda \geq 0$, μ satisfy $\sum_j \lambda_j \leq \mu$, we have*

$$\theta\left(\mathrm{Dg}\{\textstyle\sum_j \lambda_j h_j h_j^T\}\right) \leq \mu \qquad (*)$$

where $\mathrm{Dg}\{G\} \in \mathbb{R}^m$ is the diagonal of a matrix $G \in \mathbf{S}^m$.

Indeed, $\theta(\cdot)$ clearly is convex and homogeneous of degree 1, whence under the premise of Observation one has

$$\theta(\mathrm{Dg}\{\textstyle\sum_i \lambda_j h_j h_j^T\}) = \theta(\sum_{j=1}^N \lambda_j [h_j]^2) \leq \sum_j \lambda_j \theta([h_j]^2) \leq \left[\sum_j \lambda_j\right]\left[\max_j \pi^2(h_j)\right] \leq \mu$$

Observation: *Given norm $\pi(\cdot)$ such that*

$$\pi^2(h) = \theta([h]^2),\ \theta(u) = \max_{z\in\mathcal{Z}\subset\mathbb{R}^m_+} z^T u \qquad (*)$$

and setting

$$\mathbf{H}_* = \{(G,\mu) : \exists \lambda \geq 0, h_1,...,h_N : G = \sum_j \lambda_j h_j h_j^T,\ \pi(h_j) \leq 1\, \forall j, \sum_j \lambda_j \leq \mu\}$$

we have

$$(G,\mu) \in \mathbf{H}_* \Rightarrow G \succeq 0\ \&\ \theta(\mathrm{Dg}\{G\}) \leq \mu$$

Fact: *Observation can be "nearly inverted:" one has*

$$\mathbf{H} := \{(G,\mu) : G \succeq 0, \varkappa\theta(\mathrm{Dg}\{G\}) \leq \mu\} \subset \mathbf{H}_* \subset \{(G,\mu) : G \succeq 0, \theta(\mathrm{Dg}\{G\}) \leq \mu\},$$

where
— $\varkappa = 1$ when π is proportional to $\|\cdot\|_2$, and
— $\varkappa = 4\ln(4m^2)$ for a general norm π of the form $(*)$.
Thus, $\mathbf{H}$ is a reasonably tight computationally tractable (provided $\mathcal{Z}$ is so) inner approximation of $\mathbf{H}_*$.

Illustration I:

$$\pi(z) = \|z\|_2 \Rightarrow \pi^2(z) = \|[z]^2\|_1 \Rightarrow \theta(u) = \sum_i \max[u_i, 0] = \max_{z \in [0,1]^m} z^T u.$$

Here the claim reads

If $G \in \mathbf{S}^m_+$, *then we can find a representation* $G = \sum_j \lambda_j h_j h_j^T$ *with* $\pi(h_j) \equiv \|h_j\|_2 \leq 1$ *and* $\lambda_j \geq 0$ *such that* $\sum_j \lambda_j \leq \theta(\mathrm{Dg}(G)) \equiv \mathrm{Tr}(G)$.

This indeed is true and λ_j, h_j are given by eigenvalue decomposition of G.

Illustration II:

$$\pi(z) = \|z\|_\infty \Rightarrow \pi^2(z) = \|[z]^2\|_\infty \Rightarrow \theta(u) = \max[\max_i u_i, 0] = \max_{z\geq 0, \sum_i z_i \leq 1} z^T u.$$

Here the claim reads

If $G \in \mathbf{S}^m_+$, then we can find a representation $G = \sum_j \lambda_j h_j h_j^T$ with $\pi(h_j) \equiv \|h_j\|_\infty \leq 1$ and $\lambda_j \geq 0$ such that $\sum_j \lambda_j \leq \varkappa \max_i G_{ii}$, where $\varkappa = 4\ln(4m^2)$.

The construction is as follows. Assume w.l.o.g. that $\max_i G_{ii} = 1$.

• Set $G = FF^T$, so that **(a):** *the rows in F are of Euclidean norm ≤ 1*

• Let U be once for ever fixed orthogonal $m \times m$ matrix such that **(b):** $|U_{ij}| \leq \sqrt{2/m}$ (such a matrix does exist)

• With Rademacher random χ, we have $G = H_\chi H_\chi^T$, $H_\chi := F\mathrm{Diag}\{\chi\}U$. From **(a-b)** it is easily seen that *the probability for H_χ to have magnitudes of all entries $\leq \alpha = \sqrt{\varkappa/m}$ is at least* $1/2$

Indeed, ij-th entry in H_χ is $\sum_k F_{ik}\chi_k U_{kj}$, and the typical value of the *square* of this entry is

$$\mathbf{E}_\chi\left\{[\textstyle\sum_k F_{ik}\chi_k U_{kj}]^2\right\} = \sum_k F_{ik}^2 \underbrace{U_{kj}^2}_{\leq 2/m} \leq \frac{2}{m}\sum_k F_{ik}^2 \leq \frac{2}{m}.$$

$\Rightarrow$ We can rapidly find, in a randomized fashion, $\bar{H}$ such that $\bar{H}\bar{H}^T = G$ and the magnitudes of entries in $\bar{H}$ do not exceed α

$\Rightarrow$ Denoting by h_j the columns of $\bar{H}/\alpha$ and setting $\lambda_j = \alpha^2$, $j \leq m$, we have

$$\|h_j\|_\infty \leq 1 \ \&\ G = \sum_j \lambda_j h_j h_j^T \ \&\ \sum_j \lambda_j = m\alpha^2 = \varkappa = \varkappa \max_i G_{ii},$$

as required.

Claim: Relations

$$\mathbf{H}_* = \{(G,\mu) : \exists \lambda \geq 0, h_1,...,h_N : G = \sum_j \lambda_j h_j h_j^T,\ \pi(h_j) \leq 1\,\forall j, \sum_j \lambda_j \leq \mu\}$$
$$\pi^2(u) = \theta(u) := \max_{z\in\mathcal{Z}\subset\mathbb{R}^m_+} z^T u$$

imply that

$$\mathbf{H} := \{(G,\mu) : G \succeq 0, \varkappa\theta(\mathrm{Dg}\{G\}) \leq \mu\} \subset \mathbf{H}_* \subset \{(G,\mu) : G \succeq 0, \theta(\mathrm{Dg}\{G\}) \leq \mu\} \qquad (*)$$

Proof. The right inclusion in $(*)$ has been proved. Let us prove the left inclusion. By homogeneity it suffices to prove that when $G \succeq 0$ satisfies $\theta(\mathrm{Dg}\{G\}) \leq 1$, we can represent G as $G = \sum_j \lambda_j h_j h_j^T$ with $\lambda \geq 0$ satisfying

$$\sum_j \lambda_j \leq \varkappa.$$

Case of $\pi(\cdot) = \alpha\|\cdot\|_2$: Here $\mathcal{Z} = \{[\alpha^2;...;\alpha^2]\}$, $\theta(u) = \alpha^2 \sum_j u_j$, and on the close inspection we should prove that when $G \succeq 0$ and $\mathrm{Tr}(G) \leq 1$, we have $G = \sum_j \lambda_j h_j h_j^T$, with $\lambda \geq 0$, $\sum_j \lambda_j = 1$, and $\|h_j\|_2 \leq 1$ for all j – the fact readily given by eigenvalue decomposition of G.
General case: Since $G \succeq 0$, we have $G = Q^2$ with some $Q \in \mathbf{S}^m$. Setting $\sigma_i = G_{ii}$, we have

$$1 \geq \theta(\sigma)\ \&\ \sum_j Q_{ij}^2 = \sigma_i$$

$$G, Q \in \mathbf{S}^m \ \& \ G = Q^2 \ \& \ \sum_j Q_{ij}^2 = \sigma_i \text{ with } \theta(\sigma) \leq 1$$

• Let U be $m \times m$ orthonormal matrix with magnitudes of entries not exceeding $\gamma = \sqrt{2/m}$ (matrices of this type do exist). For a random Rademacher vector χ, setting $Q_\chi = Q\mathrm{Diag}\{\chi\}U$, we get

$$Q_\chi Q_\chi^T \equiv G.$$

On the other hand, $[Q_\chi]_{ij} = \sum_{\ell=1}^m Q_{i\ell}\chi_\ell U_{\ell j}$, whence

$$\mathbf{E}_\chi\left\{[Q_\chi]_{ij}^2\right\} = \sum_{\ell=1}^m Q_{i\ell}^2 U_{\ell j}^2 \leq (2/m)\sum_{\ell=1}^m Q_{i\ell}^2 = 2\sigma_i/m.$$

It is easily seen that when $\gamma \geq 1$, we have for every i, j:

$$\mathrm{Prob}\left\{[Q_\chi]_{ij}^2 > 2\gamma\sigma_i/m\right\} \leq 2\exp\{-\gamma/2\}.$$

$\Rightarrow$ *Setting $\gamma = 2\ln(4m^2) = \varkappa/2$, the probability for χ to ensure $[Q_\chi]_{ij}^2 \leq 2\gamma\sigma_i/m$ for all i, j is at least 1/2*

$\Rightarrow \exists Q_{\bar{\chi}} = [q_1, ..., q_m]$: $G = Q_{\bar{\chi}} Q_{\bar{\chi}}^T = \sum_j q_j q_j^T$ and $[q_j]^2 \leq \frac{2\gamma}{m}\sigma \Rightarrow \pi^2(q_j) \leq \frac{2\gamma}{m}\theta(\sigma) \leq \frac{2\gamma}{m}$

$\Rightarrow G = \sum_j \lambda_j h_j h_j^T$ with $h_j = \sqrt{\frac{m}{2\gamma}} q_j$ and $\lambda_j = \frac{2\gamma}{m}$

$\Rightarrow G = \sum_j \lambda_j h_j h_j^T$ with $\pi(h_j) \leq 1$ and $\lambda \geq 0$, $\sum_j \lambda_j = 2\gamma = \varkappa$. $\square$

Compatibility of closed convex cone $\mathbf{Y} = \{(U,\tau)\} \subset \mathbf{S}^N_+ \times \mathbb{R}_+$ with convex compact set $\mathcal{Y} \subset \mathbb{R}^N$ means that

• $y^T U y \leq \tau \ \forall (y \in \mathcal{Y}, (U,\tau) \in \mathbf{Y})$

• $\exists (\bar{U}, \bar{\tau}) \in \mathbf{Y} : \bar{U} \succ 0$

• $(U,\tau) \in \mathbf{Y}, \tau' \geq \tau \Rightarrow (U,\tau') \in \mathbf{Y}$.

How to build cone $\mathbf{U}$, the wider the better, compatible with a given convex compact set $\mathcal{Y}$?

♠ We know two sources of cones compatible with $\mathcal{Y}$:

— cones coming from semidefinite relaxation on ellitopes/spectratopes

— cones coming from *absolute norms*.

Compatibility via ellitopes/spectratopes

Fact: *Let $\mathcal{Y}$ be a convex compact subset of an ellitope:*

$$\mathcal{Y} \subset \mathcal{Z} = \{z \in \mathbb{R}^N : \exists (t \in \mathcal{T}, x) : z = Px, x^T S_k x \le t_k, k \le K\}$$
$$[S_k \succeq 0, \sum_k S_k \succ 0]$$

Then the cone

$$\mathbf{Y} = \{(U, \tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists \lambda \ge 0 : P^T U P \preceq \sum_k \lambda_k S_k, \phi_{\mathcal{T}}(\lambda) := \max_{t \in \mathcal{T}} t^T \lambda \le \tau\}$$

is compatible with $\mathcal{Y}$.
When $\mathcal{Y}$ is a subset of spectratope:

$$\mathcal{Y} \subset \mathcal{Z} = \{z \in \mathbb{R}^N : \exists (t \in \mathcal{T}, x) : z = Px,\ S_k^2[x] \preceq t_k I_{d_k}, k \le K\},$$
$$[S_k[x] = \sum_{j=1}^n x_j S^{kj}, S^{kj} \in \mathbf{S}^{d_k}]$$

the cone

$$\mathbf{Y} = \{(U, \tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists \{\Lambda_k \succeq 0\} : P^T U P \preceq \sum_k \mathcal{S}_k^*[\Lambda_k], \phi_{\mathcal{T}}(\lambda[\Lambda]) \le \tau\}$$
$$\left[\left[\mathcal{S}_k^*[\Lambda_k]\right]_{ij} = \mathrm{Tr}(S^{ki} \Lambda_k S^{kj}),\ [\lambda[\Lambda]]_k = \mathrm{Tr}(\Lambda_k)\right]$$

is compatible with $\mathcal{Y}$.

This is readily given by what we know on semidefinite relaxation on ellitopes/spectratopes.

Compatibility via absolute norms

♠ **Preliminaries: absolute norms.** A norm $\|\cdot\|$ on $\mathbb{R}^N$ is called *absolute*, if it depends solely on the magnitudes of entries of a vector:

$$\|z\| = \|\mathrm{abs}[z]\|,\ \mathrm{abs}[[z_1;...;z_N]] = [|z_1|;...,|z_N|].$$

Examples: The ℓ_s norms $\|\cdot\|_s$, are absolute; similarly, the *block ℓ_s-norm*

$$\|[z^1;...;z^K]\| = \|[\|z^1\|_{s_1};\|z^2\|_{s_2};...;\|z^K\|_{s_K}]\|_s \qquad [s,s_1,...,s_K \in [1,\infty]]$$

is absolute.

Facts:

• *An absolute norm $\|\cdot\|$ is monotone in the magnitudes of entries: if $\mathrm{abs}[z] \leq \mathrm{abs}[z']$, then $\|z\| \leq \|z'\|$.*

• *The norm $\|y\|_* = \max\limits_{x:\|x\|\leq 1} y^T x$ conjugate to an absolute norm $\|\cdot\|$ is absolute as well.*

Observation: *An absolute norm* $p(\cdot)$ *on* $\mathbb{R}^N$ *can be "lifted" to an absolute norm* $p^+(\cdot)$ *on* $\mathbf{S}^N$ *by setting*

$$p^+(X) = p\Big([p(\mathrm{Col}_1[X]); p(\mathrm{Col}_2[X]); ...; p(\mathrm{Col}_N[X])]\Big), \ X \in \mathbf{S}^N.$$

p^+ *indeed is an absolute norm, and*

$$p^+(xx^T) = p^2(x) \ \forall x \in \mathbb{R}^N.$$

Example: When $p(\cdot)$ is ℓ_π-norm on $\mathbb{R}^N$, $p^+(\cdot)$ is ℓ_π norm on $\mathbf{S}^N$.

♣ We say that an absolute norm $r(\cdot)$ *fits* an absolute norm $p(\cdot)$ on $\mathbb{R}^N$,

$$p(x) \leq 1 \Rightarrow r([x]^2) \leq 1.$$

Example: When $p(\cdot) = \|\cdot\|_s$, $s \in [1,\infty]$, the norm

$$r(\cdot) = \begin{cases} \|\cdot\|_1, & 1 \leq s \leq 2 \\ \|\cdot\|_{s/2}, & s \geq 2 \end{cases}$$

fits $p(\cdot)$.

Fact: *Let p be an absolute norm on $\mathbb{R}^N$, let absolute norm $r(\cdot)$ fit $p(\cdot)$, and let $\mathcal{Y} \subset B_p := \{x \in \mathbb{R}^N : p(x) \leq 1\}$. Then the set*

$$\mathbf{Y} = \left\{ (U,\tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists (W \in \mathbf{S}^n, w \in \mathbb{R}^N_+) : \begin{array}{l} U \preceq W + \mathrm{Diag}\{w\} \\ \|W\|_{p^{+*}} + r_*(w) \leq \tau \end{array} \right\}$$

where p^{+*} is the norm on $\mathbf{S}^N$ conjugate to p^+, and $r_*(\cdot)$ is the norm on $\mathbb{R}^N$ conjugate to $r(\cdot)$, is compatible with $\mathcal{Y}$.
Besides this, $p^{+*}(\cdot) \leq q^+(\cdot)$, where $q(\cdot)$ is the norm conjugate to $p(\cdot)$.

Fact: *Let p be an absolute norm on $\mathbb{R}^N$, let absolute norm $r(\cdot)$ fit $p(\cdot)$, and let*

$$\mathcal{Y} \subset B_p := \{x \in \mathbb{R}^N : p(x) \leq 1\}.$$

Then the set

$$\mathbf{Y} = \left\{(U,\tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists (W \in \mathbf{S}^n, w \in \mathbb{R}^N_+) : \begin{array}{c} U \preceq W + \mathrm{Diag}\{w\} \\ \|W\|_{p^{+*}} + r_*(w) \leq \tau \end{array}\right\}$$

where p^{+*} is the norm on $\mathbf{S}^N$ conjugate to p^+, and $r_*(\cdot)$ is the norm on $\mathbb{R}^N$ conjugate to $r(\cdot)$, is compatible with $\mathcal{Y}$. Besides this, $p^{+*}(\cdot) \leq q^+(\cdot)$, where $q(\cdot)$ is the norm conjugate to $p(\cdot)$.

Indeed, let $(U,\tau) \in \mathbf{Y}$, so that $U \preceq W + \mathrm{Diag}\{w\}$ with $w \geq 0$ and $\|W\|_{p^{+*}} + r_*(w) \leq \tau$. For $y \in \mathcal{Y}$ we have $p(y) \leq 1$ due to $\mathcal{Y} \subset B_p$, whence

$$\begin{array}{c} y^TUy = \mathrm{Tr}(U[yy^T]) \leq \mathrm{Tr}(W[yy^T]) + \mathrm{Tr}(\mathrm{Diag}\{w\}yy^T) \leq p^{+*}(W)p^+(yy^T) + w^T[y]^2 \\ \leq p^{+*}(W)\underbrace{p^2(y)}_{\leq 1} + r_*(w)\underbrace{r([y]^2)}_{\leq 1} \leq p^{+*}(W) + r_*(w) \leq \tau \end{array}$$

$\Rightarrow \max_{y\in\mathcal{Y}} y^TUy \leq \tau$.

Besides this, when $U, V \in \mathbf{S}^n$, denoting U_j and V_j the columns of U and V, we have

$$\begin{array}{c} \mathrm{Tr}(UV) = \sum_j U_j^TV_j \leq \sum_j p(U_j)q(V_j) \leq [p(U_1);...;p(U_N)]^T[q(V_1);...;q(V_N)] \\ \leq p([p(U_1);...;p(U_N)])q([q(V_1);...;q(V_N)]) = p^+(U)q^+(V) \end{array}$$

$\Rightarrow \|V\|_{p^{+*}} = \max\limits_{U:p^+(U)\leq 1} \mathrm{Tr}(UV) \leq q^+(V)$. $\square$

Example: Let $p(\cdot) = \|\cdot\|_s$ with $s \in [1, \infty]$. In this case
— we can take $r(x) = \|x\|_{\bar{s}}$, $\bar{s} = \max[s/2, 1]$, resulting in

$$r_*(w) = \|w\|_{\bar{s}_*},\ \bar{s}_* = \frac{\bar{s}}{\bar{s}-1} = \begin{cases} +\infty, & 1 \le s \le 2 \\ \frac{s}{s-2}, & s > 2 \end{cases}$$

— $p^{+*}(\cdot)$ is $\|\cdot\|_{s_*}$ on $\mathbf{S}^N$, $s_* = \frac{s}{s-1}$
and we conclude that the cone

$$\mathbf{Y}_s = \left\{(U, \tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists (w \ge 0, W) : U \preceq W + \mathrm{Diag}\{w\}, \|W\|_{s_*} + \|w\|_{\bar{s}_*} \le \tau\right\}$$

is compatible with any subset of the unit ℓ_s ball.
Note: It is easily seen that when $s \in [2, \infty]$, the expression for $\mathbf{Y}$ provably simplifies to

$$\mathbf{Y}_s = \left\{(U, \tau) \in \mathbf{S}^N_+ \times \mathbb{R}_+ : \exists (w \ge 0) : U \preceq \mathrm{Diag}\{w\}, \|w\|_{\frac{s}{s-2}} \le \tau\right\}$$

In the case in question $\mathbf{Y}_s$ is an ellitope, and $\mathbf{Y}_s$ happens to be exactly the cone compatible with this ellitope, as given by our "ellitopic" construction.
Note: In our context, the larger is a cone compatible with the set $\mathcal{Y}$ in question (for us, this is either $\mathcal{X}_{\mathrm{S}}$, or $\mathcal{B}_*$), the better. The "ideal" choice would be

$$\mathbf{Y} = \mathbf{Y}_*[\mathcal{Y}] = \{(U, \tau) : U \succeq 0, \tau \ge \max_{y \in \mathcal{Y}} y^T U y\}.$$

This ideal cone is typically intractable computationally, this is why we have developed techniques for building tractable approximations of this cone from inside.
However: *When $\mathcal{Y} = \{y \in \mathbb{R}^N : \|y\|_2 \le 1\}$, the cone*

$$\mathbf{Y}_2 = \{(U, \tau) : 0 \preceq U \preceq \tau I_N\}$$

is exactly the same as the "ideal" cone $\mathbf{Y}_*[\mathcal{Y}]$.

Ellitopic case, Signal-Independent White Gaussian Noise

♠ Assume that
- the o.s. is Gaussian: $\omega = Ax + \sigma\xi$, $\xi \sim \mathcal{N}(0, I_m)$
- the signal set $\mathcal{X}$ and the unit ball $\mathcal{B}_*$ of the norm conjugate to the one used to measure the recovery error are ellitopes:

$$\begin{array}{rcl}\mathcal{X} & = & \{x \in \mathbb{R}^n : \exists t \in \mathcal{T} : x^T S_k x \le t_k,\, k \le K\} \\ \mathcal{B}_* & = & \{u \in \mathbb{R}^m : \exists (r \in \mathcal{R}, z) : u = Mz, z^T R_\ell z \le r_\ell,\, \ell \le L\}\end{array}$$

In this case, our compatibility-based recipe for building presumably good polyhedral estimate combines with the machinery for building cones compatible with ellitopes to result in the polyhedral estimate $\widehat{x}_H$ yielded by the optimal solution to the convex optimization problem

$$\mathrm{Opt} = \min_{\Theta,U,\lambda,\mu} \left\{ 2\left[\phi_{\mathcal{T}}(\lambda) + \phi_{\mathcal{R}}(\mu) + \varkappa^2\sigma^2 \mathrm{Tr}(\Theta)\right] : \begin{array}{c} \Theta \succeq 0,\, U \succeq 0,\, \lambda \ge 0,\, \mu \ge 0, \\ \left[\begin{array}{c|c} U & \frac{1}{2}B \\ \hline \frac{1}{2}B^T & A^T\Theta A + \sum_k \lambda_k S_k \end{array}\right] \succeq 0, \\ M^T U M \preceq \sum_\ell \mu_\ell R_\ell \end{array} \right\}$$

$$\varkappa = \sqrt{2\ln(2m/\epsilon)},\ \phi_{\mathcal{Z}}(\nu) = \max_{z\in\mathcal{Z}} \nu^T z.$$

The $m \times m$ contrast matrix H is given by the Θ-component Θ_* of an optimal solution to the problem: the columns h_j of H are the eigenvectors of Θ_* normalized to satisfy $\|h_j\|_2 = (\varkappa\sigma)^{-1}$, and

$$\mathrm{Risk}_{\epsilon,\|\cdot\|}[\widehat{x}_H|\mathcal{X}] \le \mathrm{Opt}.$$

Proposition: *Assume that $\epsilon \le 1/8$. Then the resulting estimate is near-optimal:*

$$\mathrm{Opt} \le O(1)\varkappa\sqrt{\ln(2K)\ln(2L)}\mathrm{RiskOpt}_{\frac{1}{8}} \le O(1)\varkappa\sqrt{\ln(2K)\ln(2L)}\mathrm{RiskOpt}_{\epsilon},$$

where $\mathrm{RiskOpt}_\epsilon$ is the infimum, over all possible estimates, of $(\epsilon, \|\cdot\|)$-risks of the estimates on $\mathcal{X}$.

Note: Similar result holds true in the case when $\mathcal{X}$ and $\mathcal{B}_*$ are spectratopes.

How It Works

♠ **Setup:**

- Unknown signal x is restriction of function $h(t)$ of continuous time on the n-element equidistant grid on $[0, 4]$, with the magnitude of h known to be ≤ 1
- We want to recover the result of "numerical double-integration" of h – the vector Bx with

$$B_{ij} = \begin{cases} \frac{16}{n^2}[i - j + 1] & , i \geq j \\ 0 & , i > j \end{cases}$$

- We observe in Gaussian noise $\mathcal{N}(0, \sigma^2 I_m)$ the restriction of x onto m randomly selected points of the grid; this selection specifies A.
- The recovery error is measured in $\|\cdot\|_2$.

♠ We are in the case when the signal set $\mathcal{X}$ is the unit box:

$$\mathcal{X} = \{x \in \mathbb{R}^n : x_i^2 \leq 1, \ 1 \leq i \leq n\}$$

Note that our $\mathcal{X}$ and $\mathcal{B}_*$ are ellitopes, so that we can build efficiently

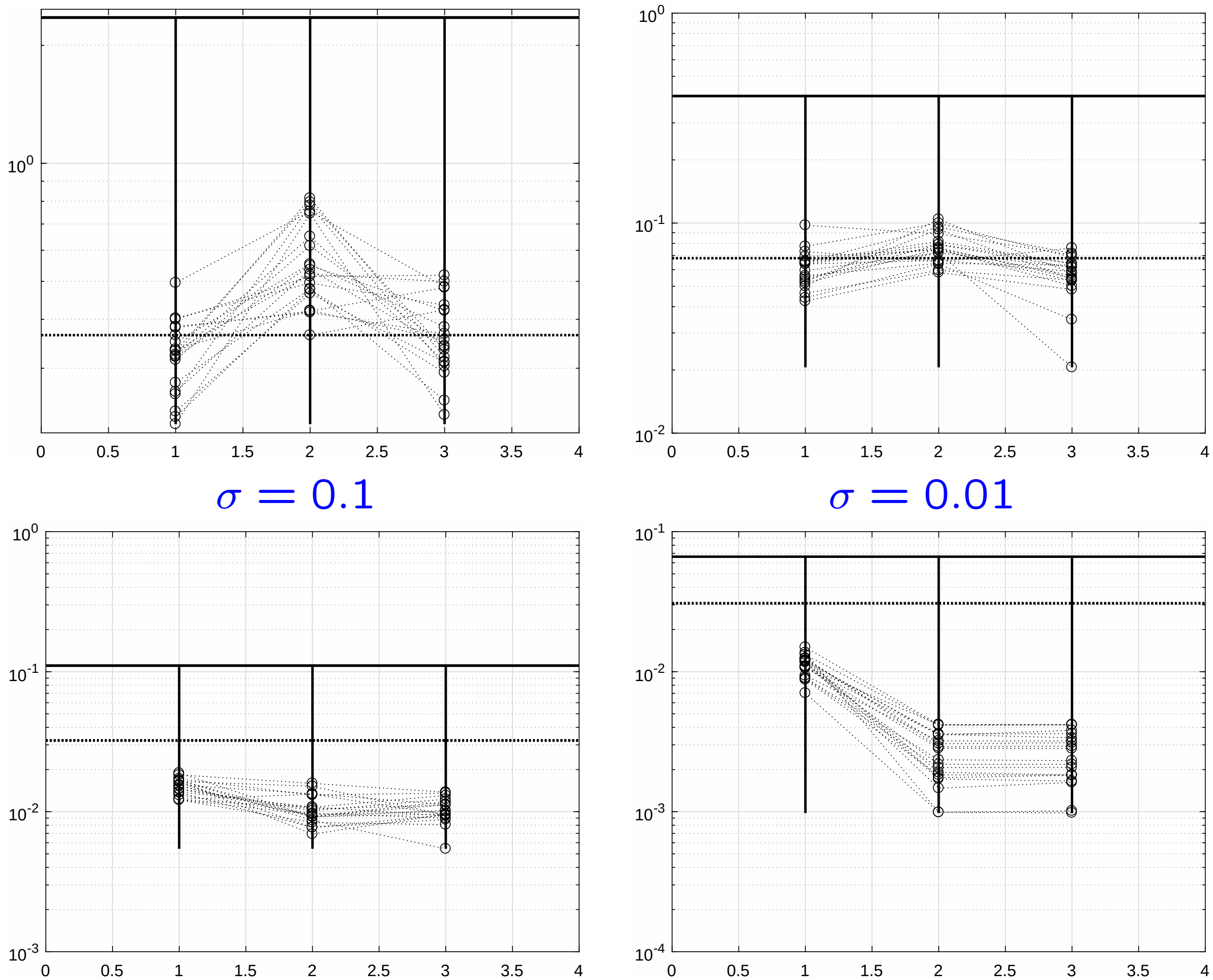
— the provably near-optimal linear estimate *Lin*,

— the polyhedral estimate *PolyI*,

— the provably near-optimal polyhedral estimate *PolyII*,

with *PolyI*, *PolyII* yielded by the first, resp. the second of our techniques for designing polyhedral estimates.

♠ In the experiments to be reported, $n = 64$, $m = 32$, and $\epsilon = 0.1$.



$\sigma = 0.1$ $\sigma = 0.01$

$\sigma = 0.001$ $\sigma = 0.0001$

Recovery errors for *Lin* (left column), *PolyI* (right column), and *PolyII* (middle column)

Horizontal lines: solid — upper bound on $\mathrm{Risk}_{0.1,\|\cdot\|_2}$ of *PolyII* dotted — upper bound on $\mathrm{Risk}_{\|\cdot\|_2}$ of *Lin*.

Data over 20 simulations per each value of σ

How It Works (continued)
Denoising and Deblurring Images

• Grayscale $m \times n$ image is $m \times n$ array with entries in the range $[0, 255]$. Subtracting from the entries $R = 127.5$, we represent the image by matrix $x \in \mathbb{R}^{m\times n}$ with entries in the range $[-R, R]$.
• Let us look how Polyhedral Estimate works when recovering images $x \in \mathbb{R}^{m\times n}$ from their blurred noisy observation

$$\omega = \kappa \star x + \xi$$

with $p \times q$ kernel κ and White Gaussian observation noise: entries of ξ are $\sim \mathcal{N}(0, \sigma^2)$ and independent of each other.
• Same as with linear estimates, we pass to frequency domain, where the observation becomes

$$\zeta = \theta \bullet \chi + \eta$$

[χ: DFT of x^+; θ: DFT of κ^+; η: complex-valued white Gaussian noise; $\bullet$: entrywise product
x^+, κ^+: $[m + p - 1] \times [n + q - 1]$ arrays obtained from x, κ by adding zero rows and columns]

and a priori information on χ reduces to a small number of (empirically identified) simple constraints of the form

$$0 \le |\chi_{rs}| \le \gamma_{rs} \,\forall r, s \;\&\; \sum_{r,s} \alpha^{(k)}_{rs} |\chi_{rs}| \le \alpha^{(k)} mn, \;\; \sqrt{\sum_{r,s} \beta^{(k)}_{rs} |\chi_{rs}|^2} \le \beta^{(k)} \sqrt{mn}, \; k \le K$$

By both theoretical and computational reasons, we use the simplest possible – proportional to the unit – contrast matrix, resulting in extremely simple (nothing more than Bisection!) recovery routine

$$\widehat{\chi} = \operatorname*{argmin}_{\chi} \left\{ \max_{r,s} |\zeta_{rs} - \theta_{rs}\chi_{rs}| : \sum_{r,s} \alpha^{(k)}_{rs} |\chi_{rs}| \le \alpha^{(k)} mn, \; \sqrt{\sum_{r,s} \beta^{(k)}_{rs} |\chi_{rs}|^2} \le \beta^{(k)} \sqrt{mn}, \; k \le K, \;\; |\chi_{rs}| \le \gamma_{rs} \,\forall r, s \right\}$$

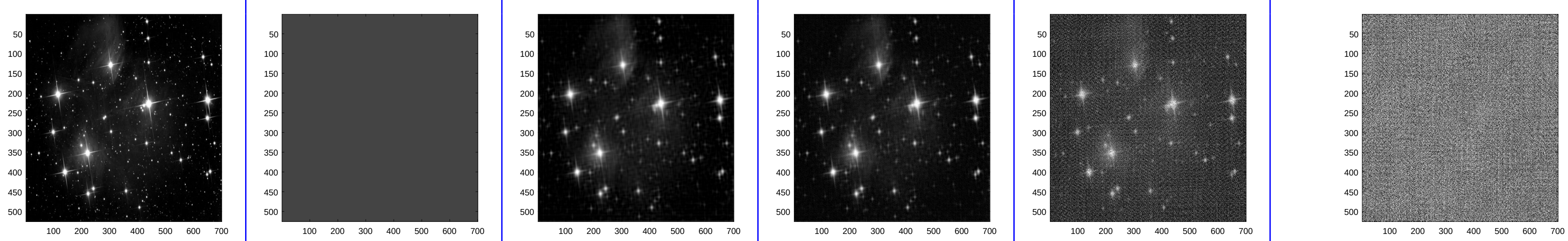
$$\zeta = \theta \bullet \chi + \eta$$

♠ In our implementation, constraints

$$\sum_{r,s} \alpha_{rs}^{(k)} |\chi_{rs}| \leq \alpha^{(k)} mn, \ \sqrt{\sum_{r,s} \beta_{rs}^{(k)} |\chi_{rs}|^2} \leq \beta^{(k)} \sqrt{mn}, \ k \leq K, \ 0 \leq |\chi_{rs}| \leq \gamma_{rs} \forall r, s \qquad (*)$$

express upper bounds on the ℓ_1, ℓ_2 and ℓ_∞ norms of the *Fourier transform* of an image x and its first order finite difference derivatives. These bounds come from analysing a small library of “real life” images.

Note: When the blur operator is ill-conditioned (some entries in θ are nearly zeros, *which is the case in all experiments to follow*), the recovery is sensitive (but not too sensitive) to the bounds in $(*)$. This is what happens when the right hand sides in $(*)$, as given by the library, are multiplied by a common factor γ:



True image | $\gamma = 0.5$ | $\gamma = 1$ | $\gamma = 10$ | $\gamma = 100$ | $\gamma = 1000$

Conditioning of blur: Card$\{i : |\theta_i| \leq 10^{-4} \max_i |\theta_i| = 10^{-4}\} = 4364$ (1.1% of the total of $mn = 367500$ entries in θ)

♠ A real life option (<u>not</u> used in the experiments to follow) is to tune γ manually.

♠ **Alternative** to the recovery routine

$$\widehat{\chi} = \underset{\chi}{\mathrm{argmin}} \left\{ \max_{r,s} |\zeta_{rs} - \theta_{rs}\chi_{rs}| : \sum_{r,s} \alpha^{(k)}_{rs}|\chi_{rs}| \leq \alpha^{(k)}mn,\ \sqrt{\sum_{r,s} \beta^{(k)}_{rs}|\chi_{rs}|^2} \leq \beta^{(k)}\sqrt{mn},\ k \leq K,\ |\chi_{rs}| \leq \gamma_{rs}\,\forall r,s \right\} \tag{A}$$

is what in Compressed Sensing was called **Regular recovery**:

$$\widehat{\chi} = \underset{\chi}{\mathrm{argmin}} \left\{ \|\chi\|_1 := \sum\nolimits_{r,s} |\chi_{rs}| : |\zeta_{rs} - \theta_{rs}\chi_{rs}| \leq \rho \right\} = \left[\widehat{\chi}_{rs} = \left\{ \begin{array}{ll} 0, & |\zeta_{rs}| \leq \rho \\ \frac{[1-\rho/|\zeta_{rs}|]\zeta_{rs}}{\theta_{rs}}, & |\zeta_{rs}| > \rho \end{array} \right. \right]_{r,s} \tag{B}$$

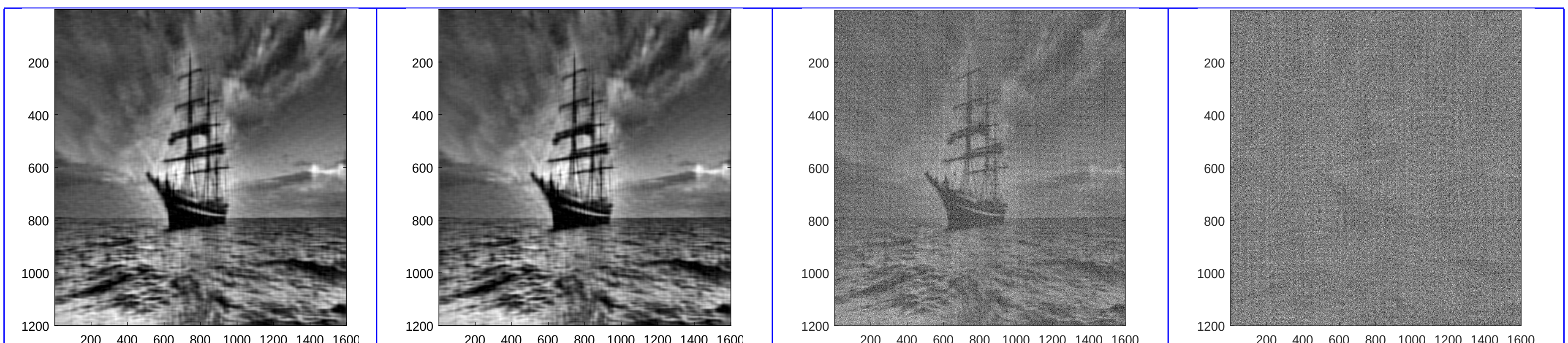
- ρ: $\|\cdot\|_\infty$-norm of the DFT η of observation noise is $\leq \rho$ with probability close to 1.

Note: $\|\cdot\|_1$-minimization is irrelevant here: the constraint imposes individual lower bounds on magnitudes of χ_{rs}, making irrelevant which *absolute* norm of χ is minimized under this constraint.

♠ **Note:** (A) *does not require* knowledge of noise's intensity σ, but *does require* knowledge of "empirical constants" in right hand sides of the constraints. In contrast, (B) *does not require* knowledge of "empirical constants," but *does require* knowledge of σ to specify ρ.

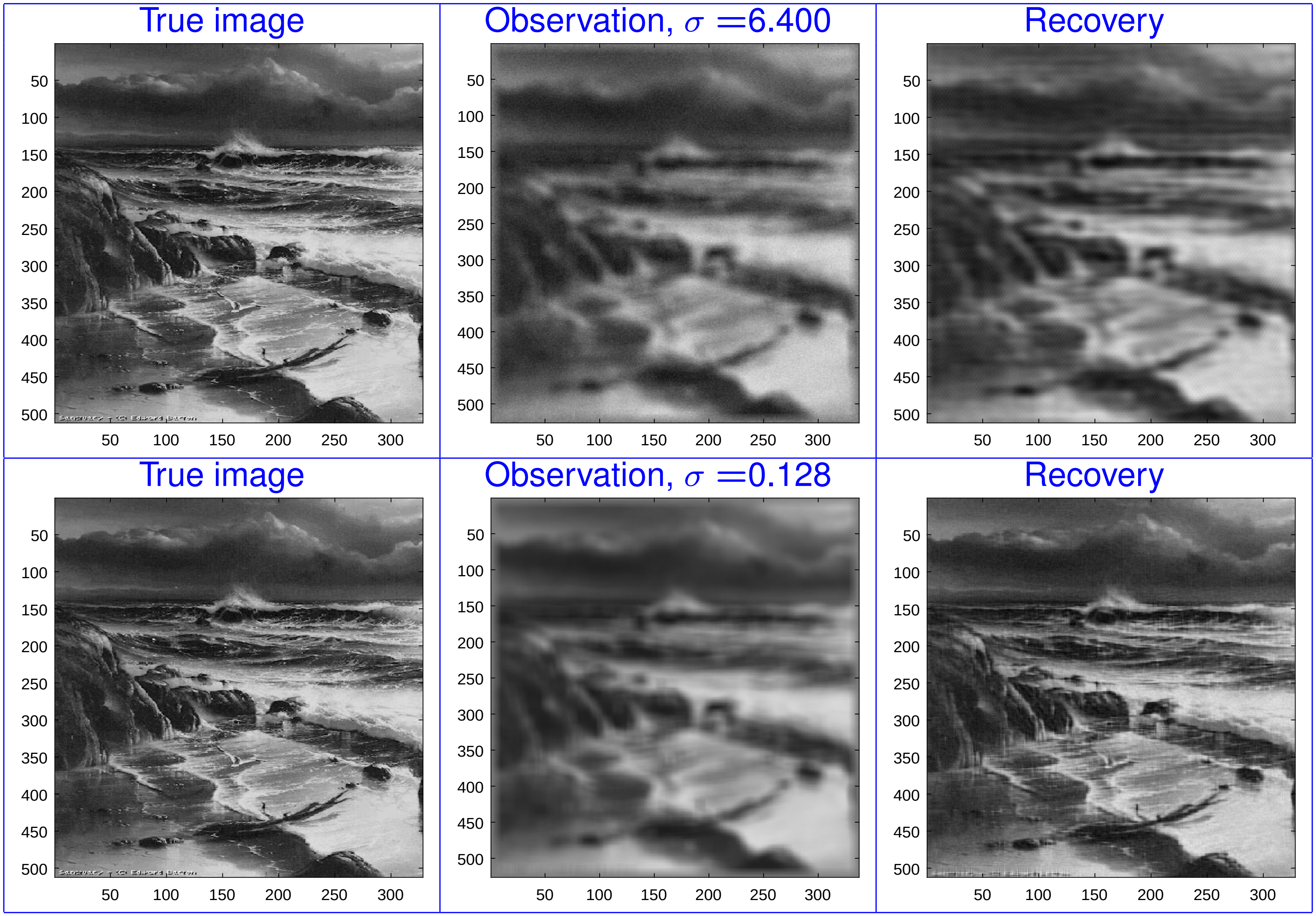
♠ In our experiments, with "properly selected" empirical constants and σ known, both recoveries were of the same quality.

Note: *Underestimating the actual noise intensity by factor like 2-3 "kills" (B):*

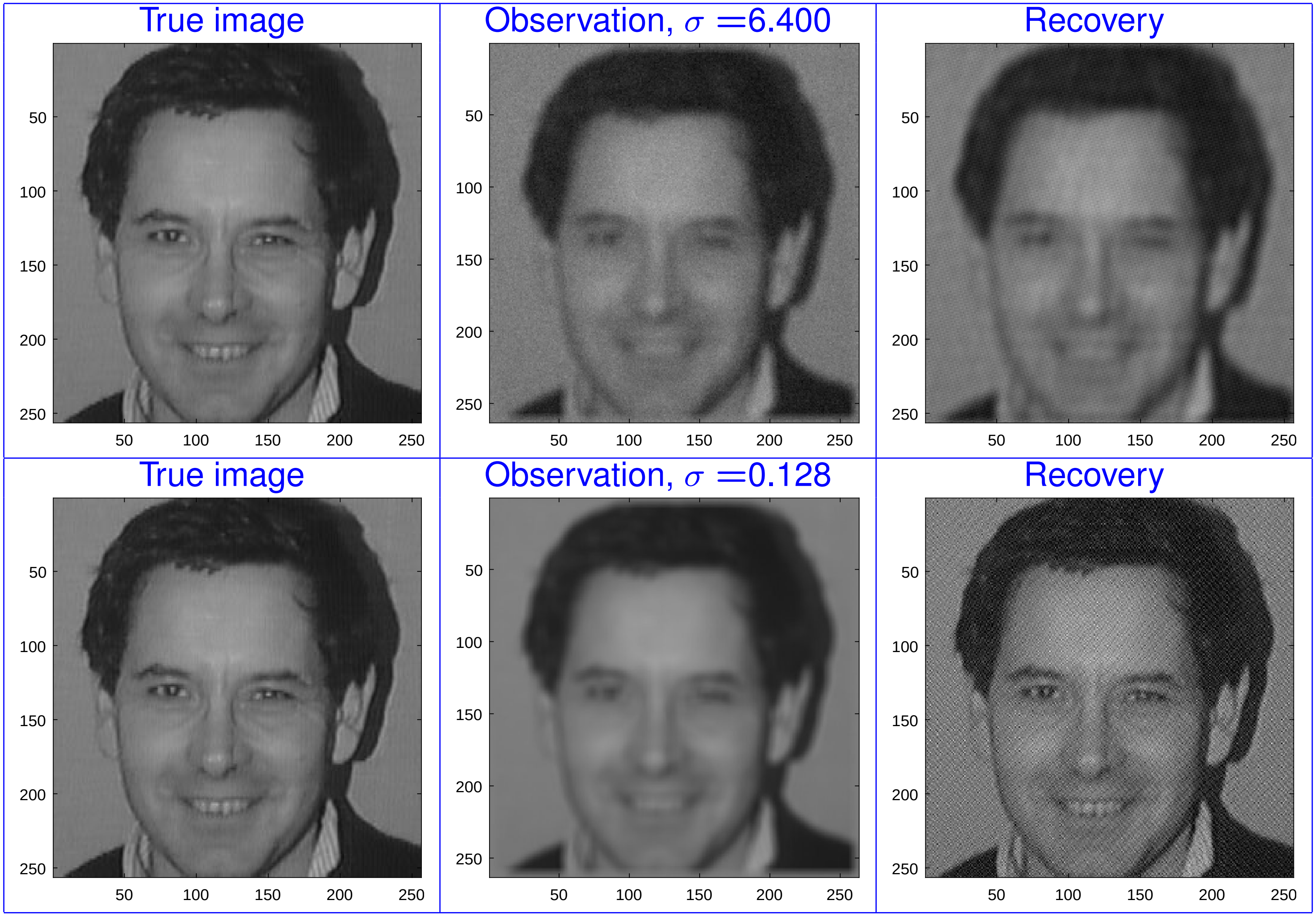


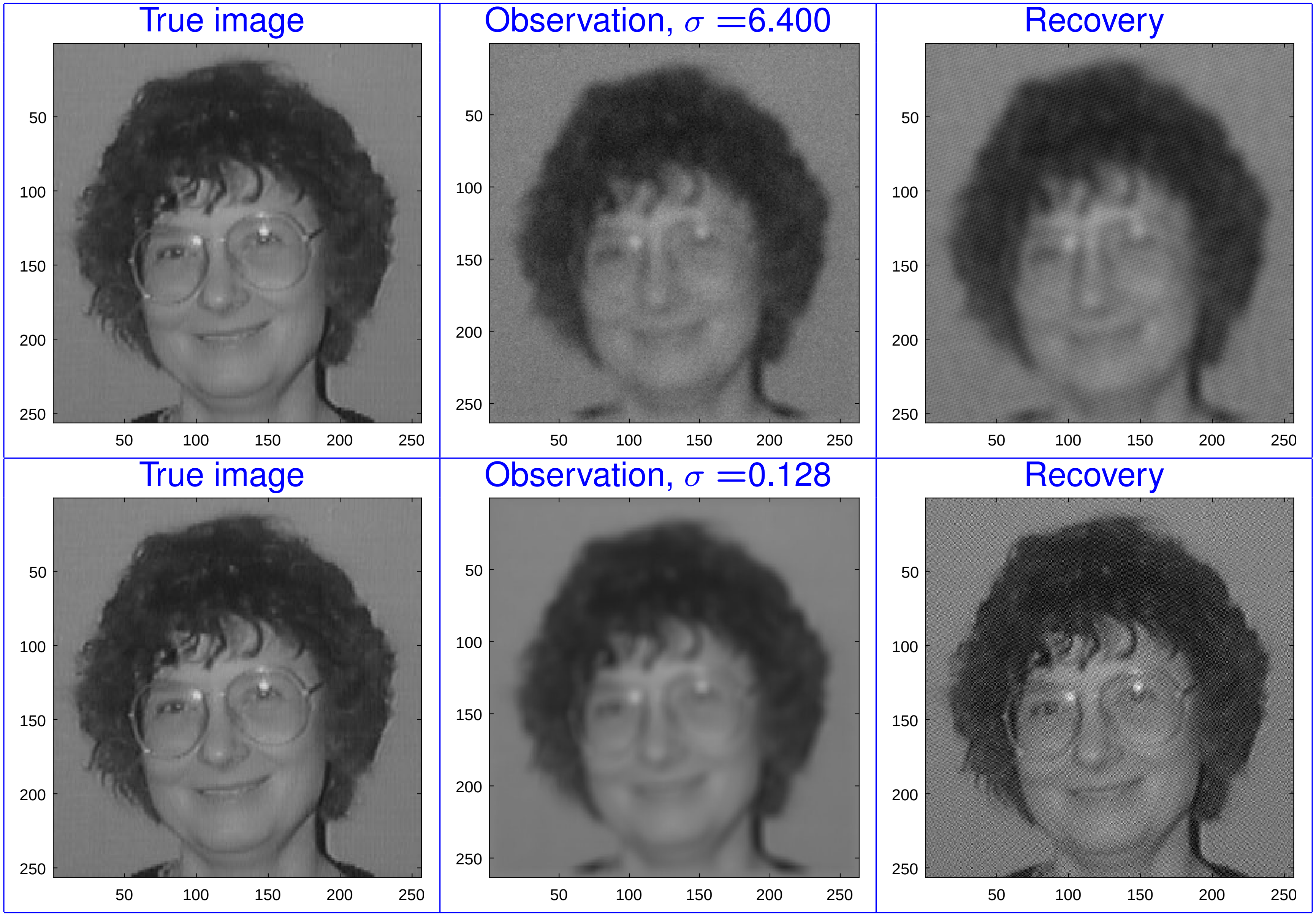
(A) | (B), ρ specified by $\sigma = 0.64$, true σ: 0.64 | (B), ρ specified by $\sigma = 0.32$, true σ: 0.64 | (B), ρ specified by $\sigma = 0.22$, true σ: 0.64

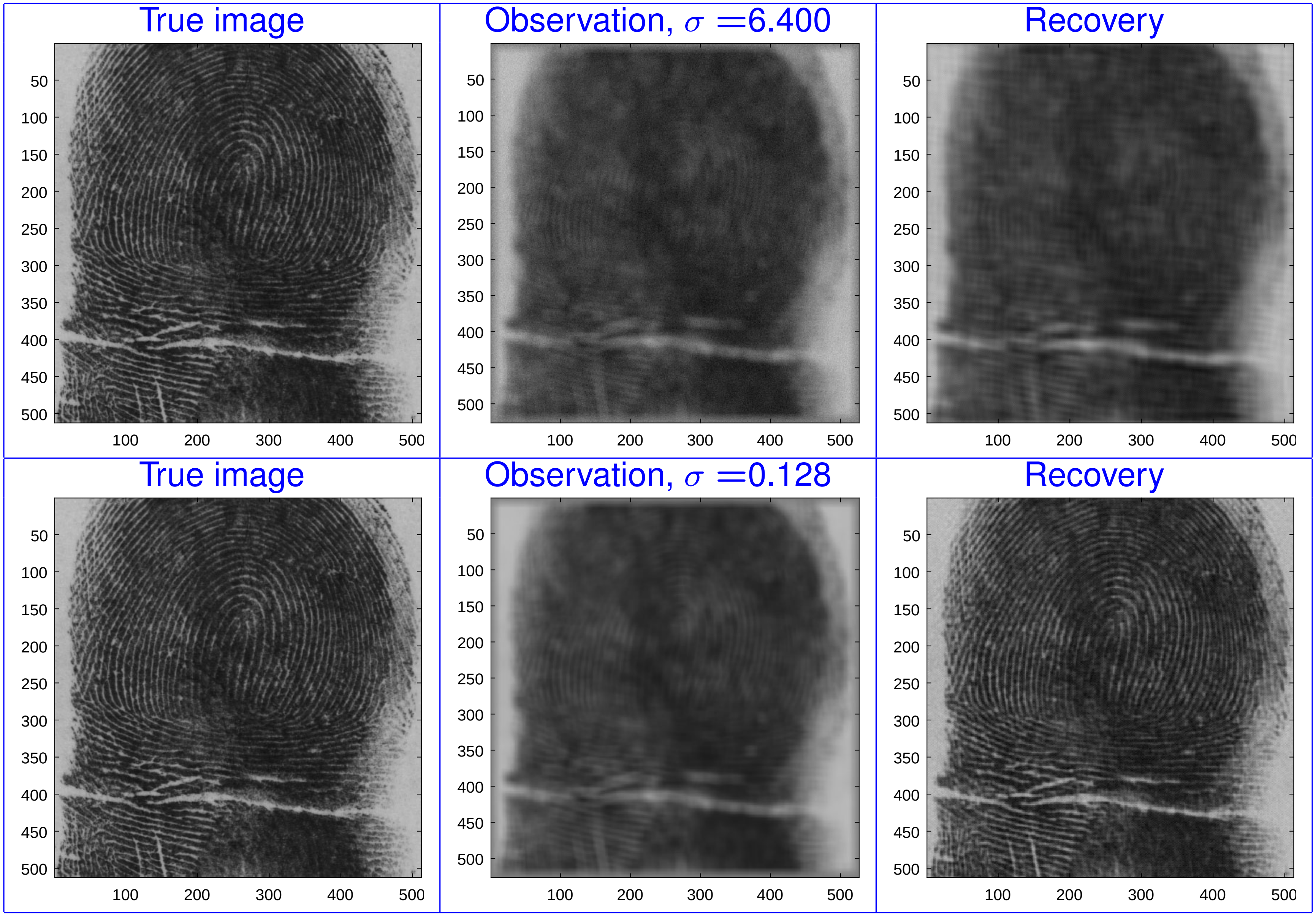
Recovery (A), Illustrations

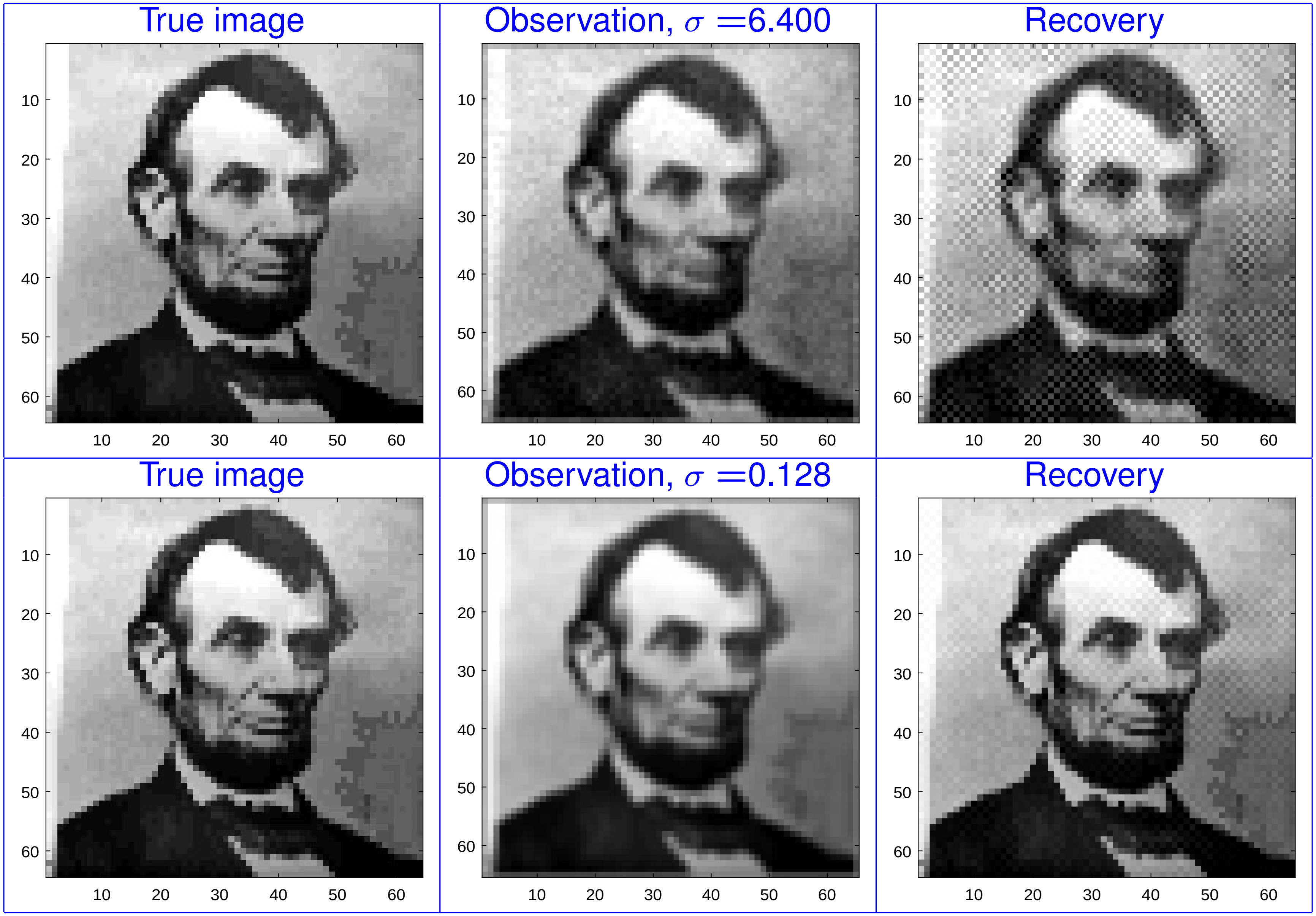


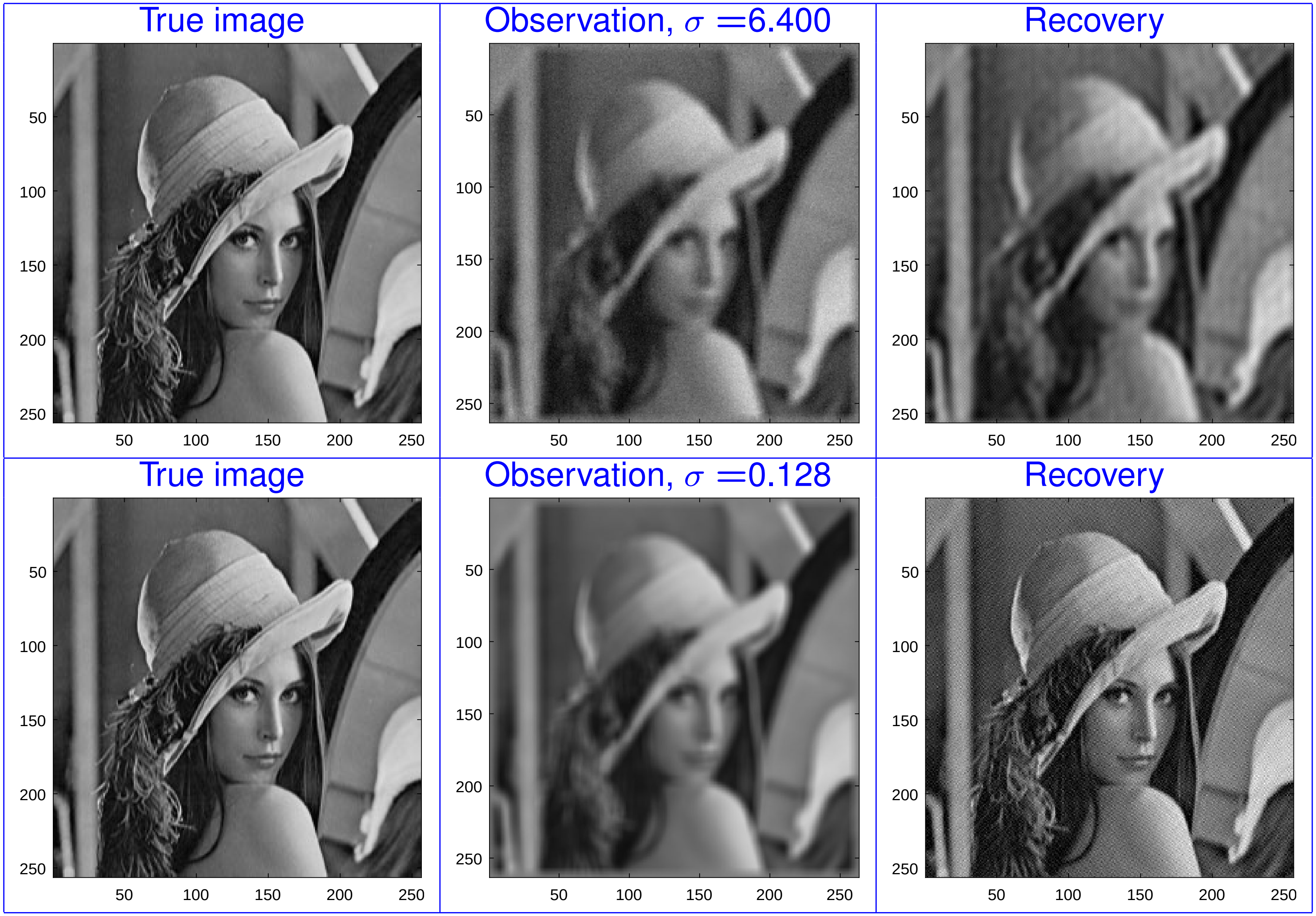

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250

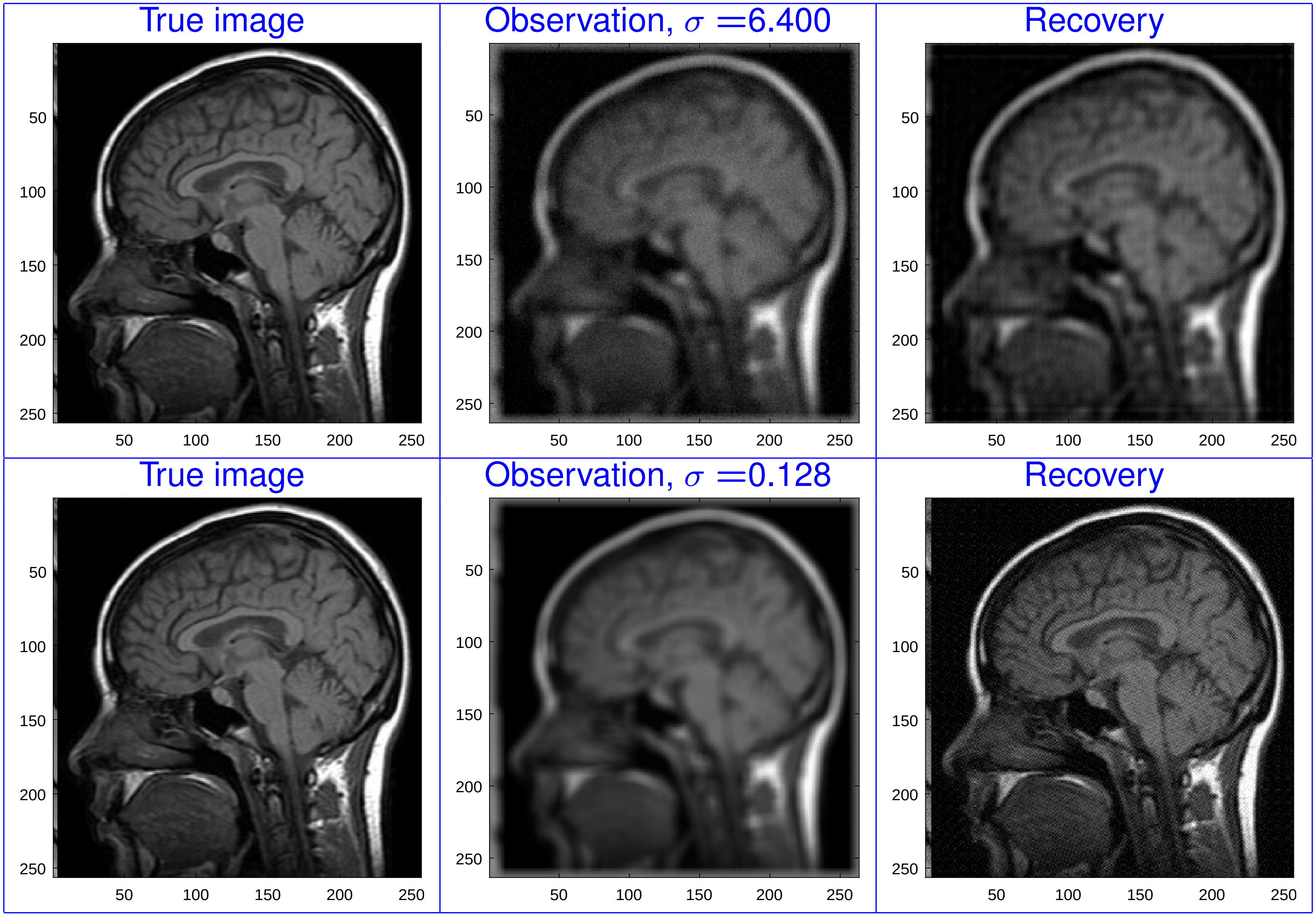

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250

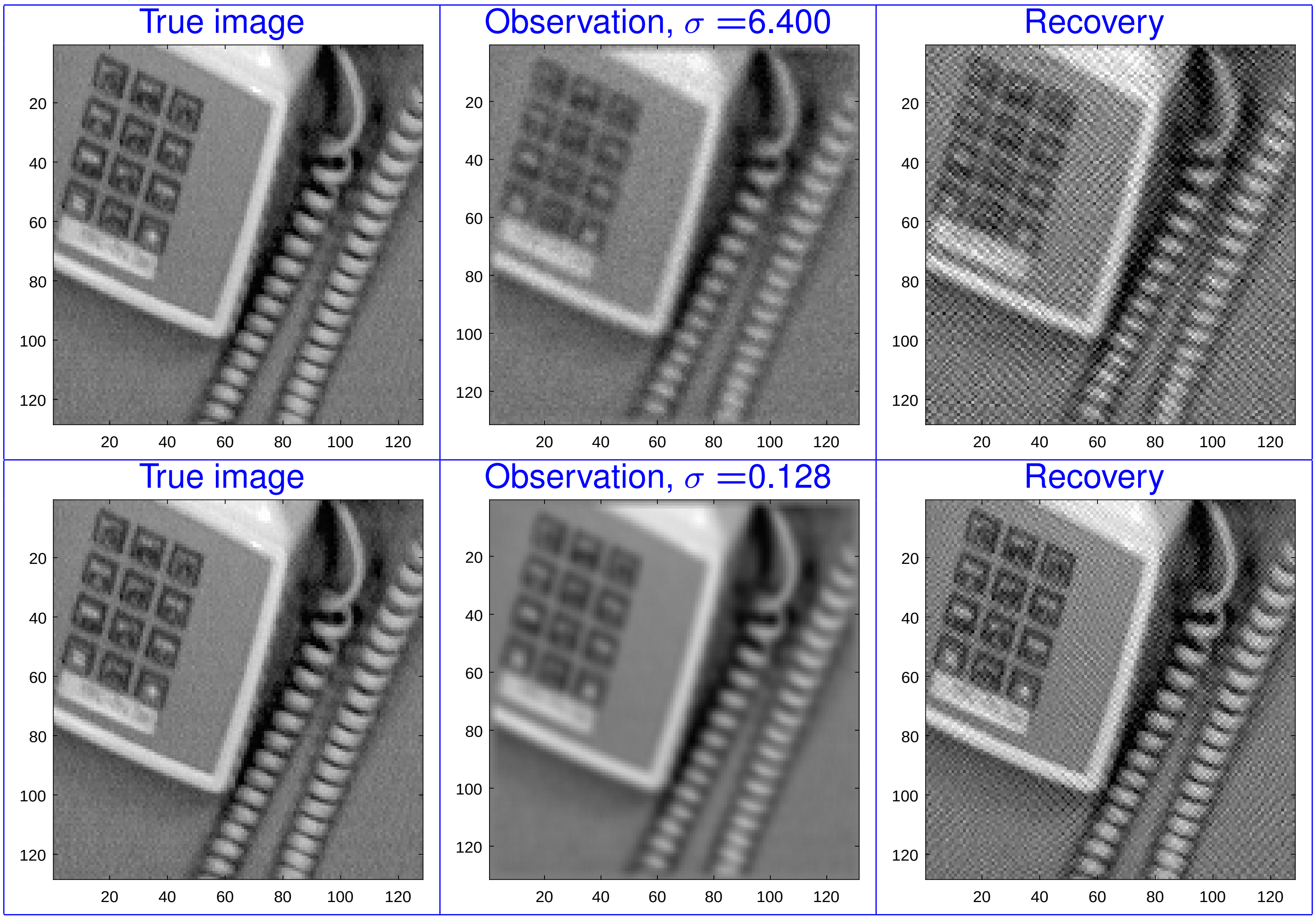

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
450
500
100
200
300
400
500


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
10
20
30
40
50
60

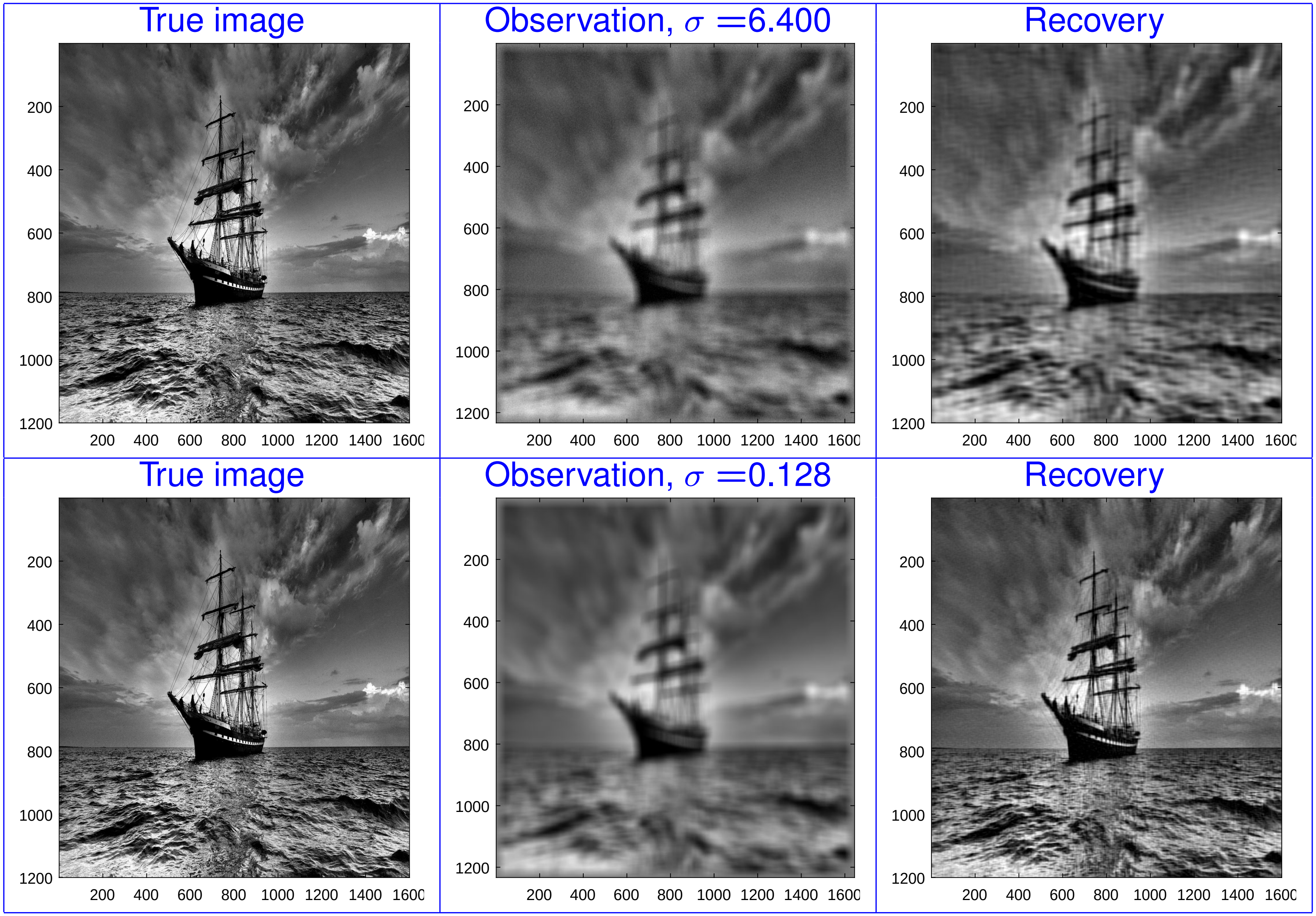


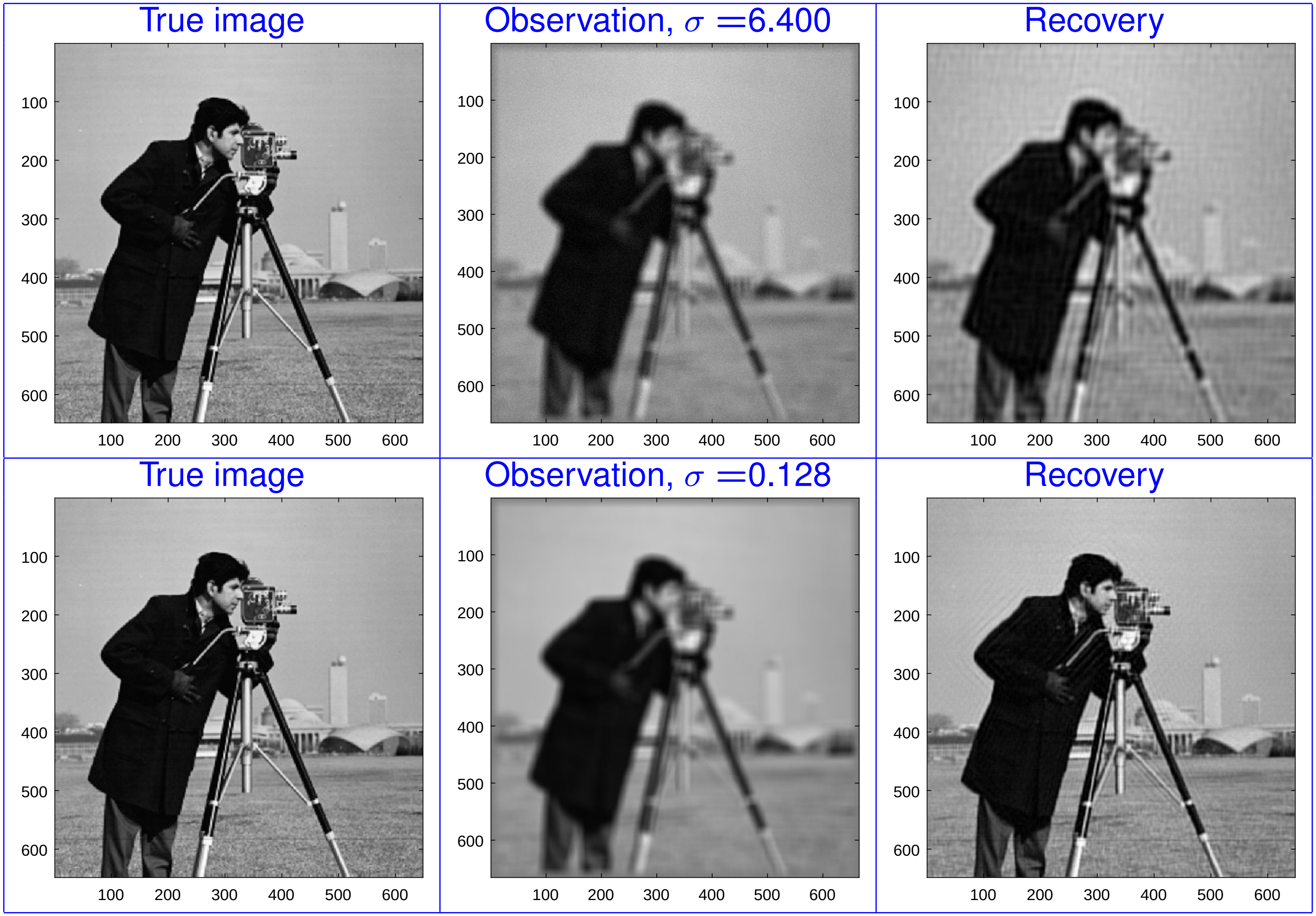
True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250

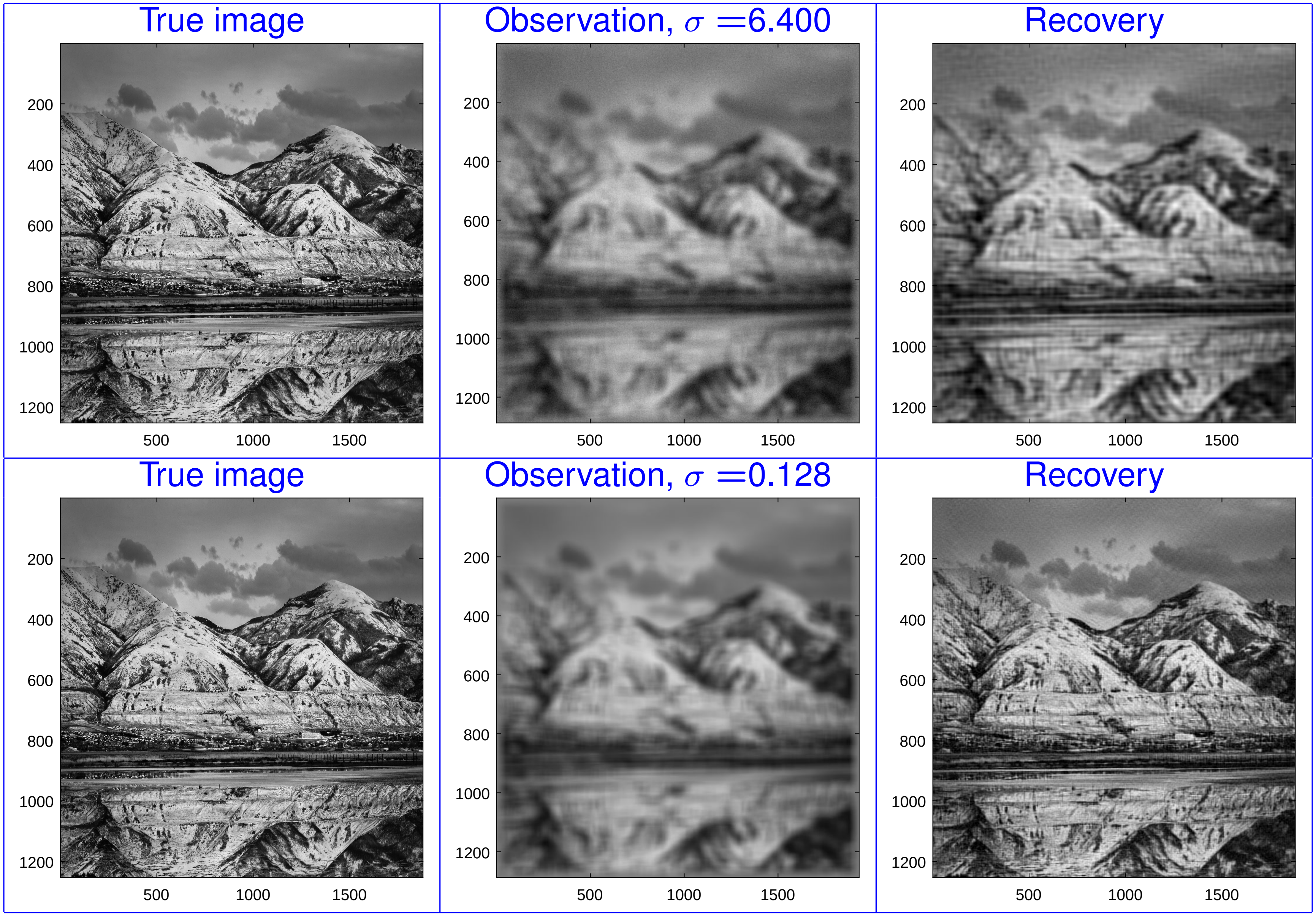


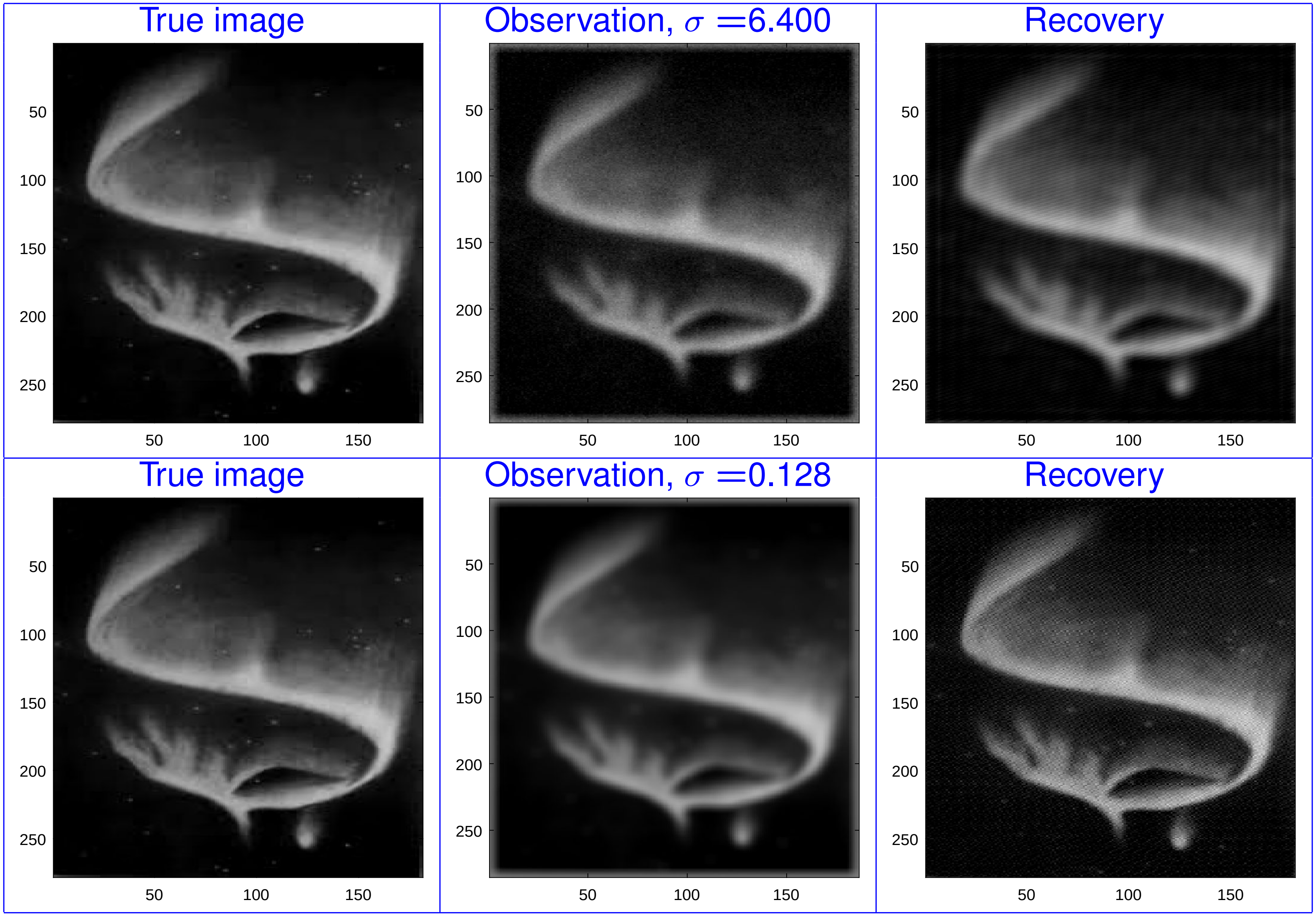
True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
20
40
60
80
100
120


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

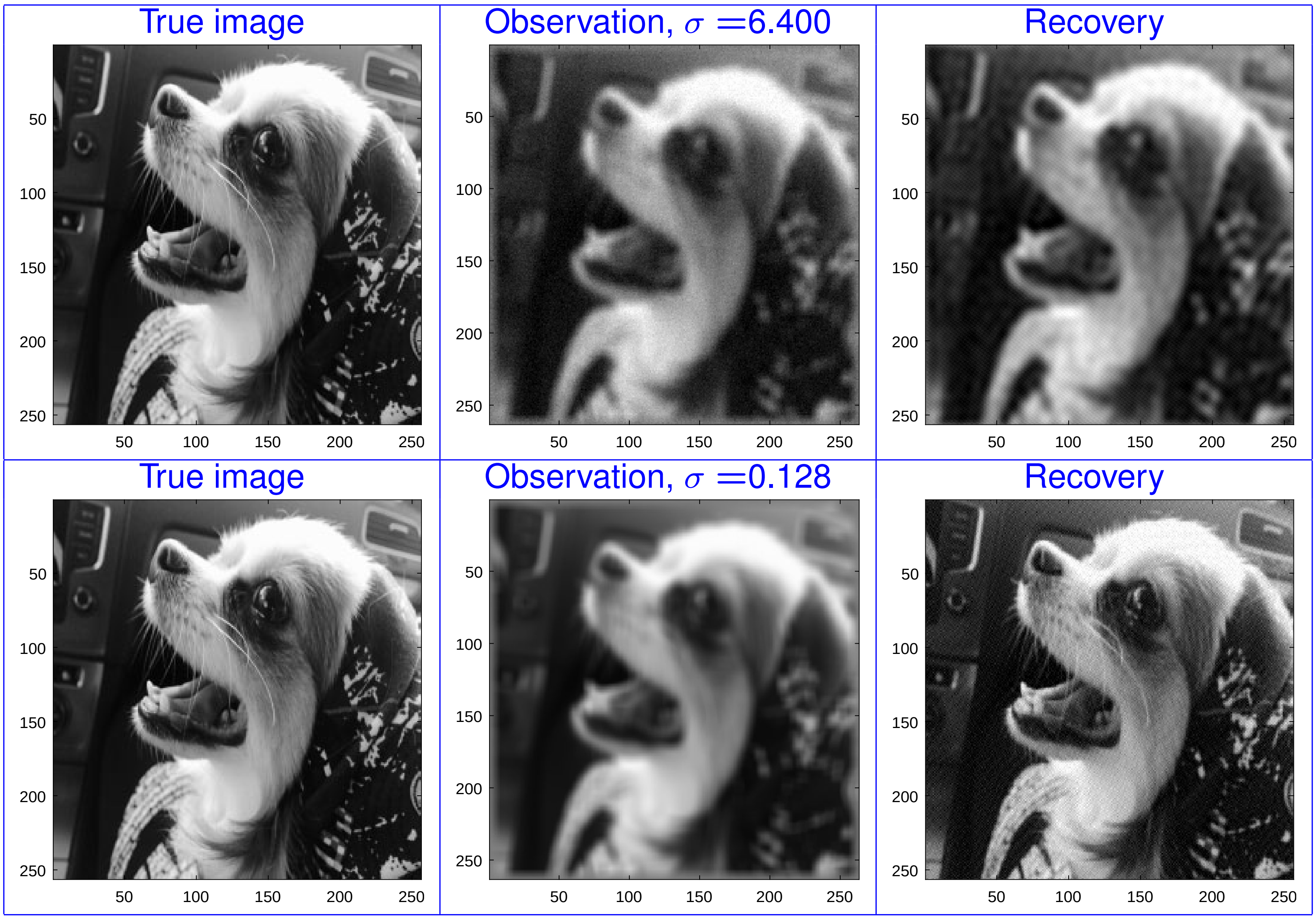


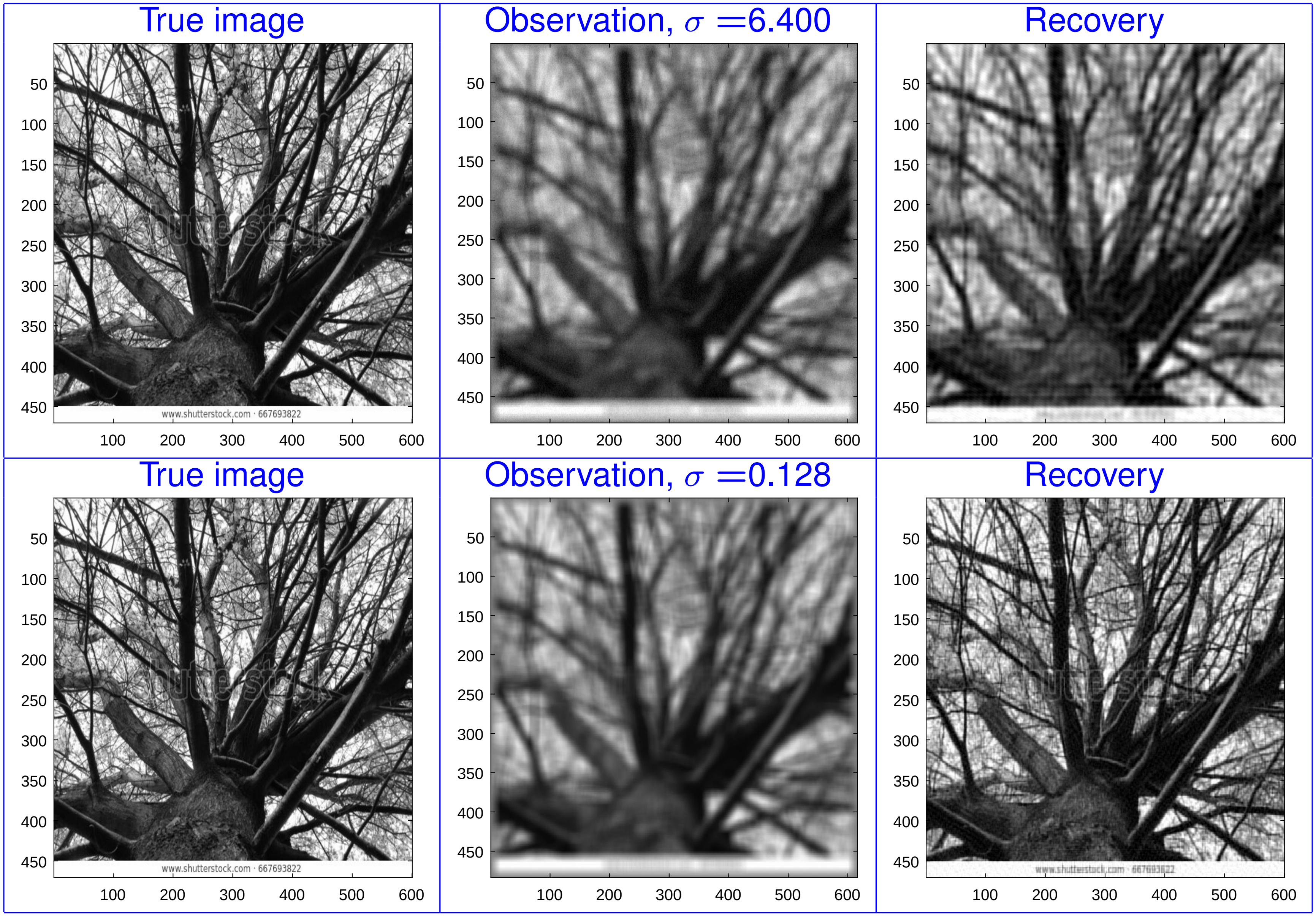
True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100
200
300
400
500
600

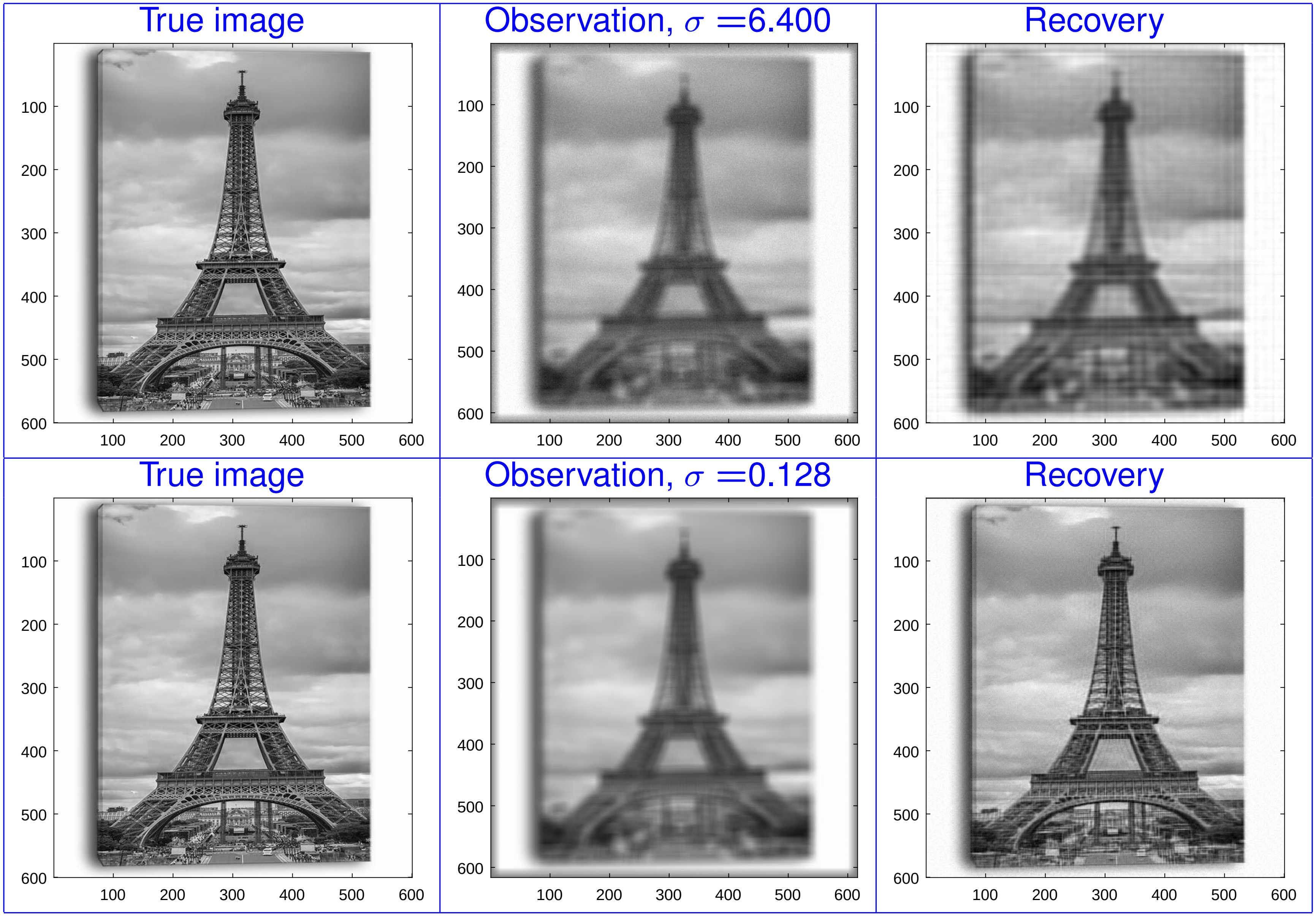

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
200
400
600
800
1000
1200
500
1000
1500

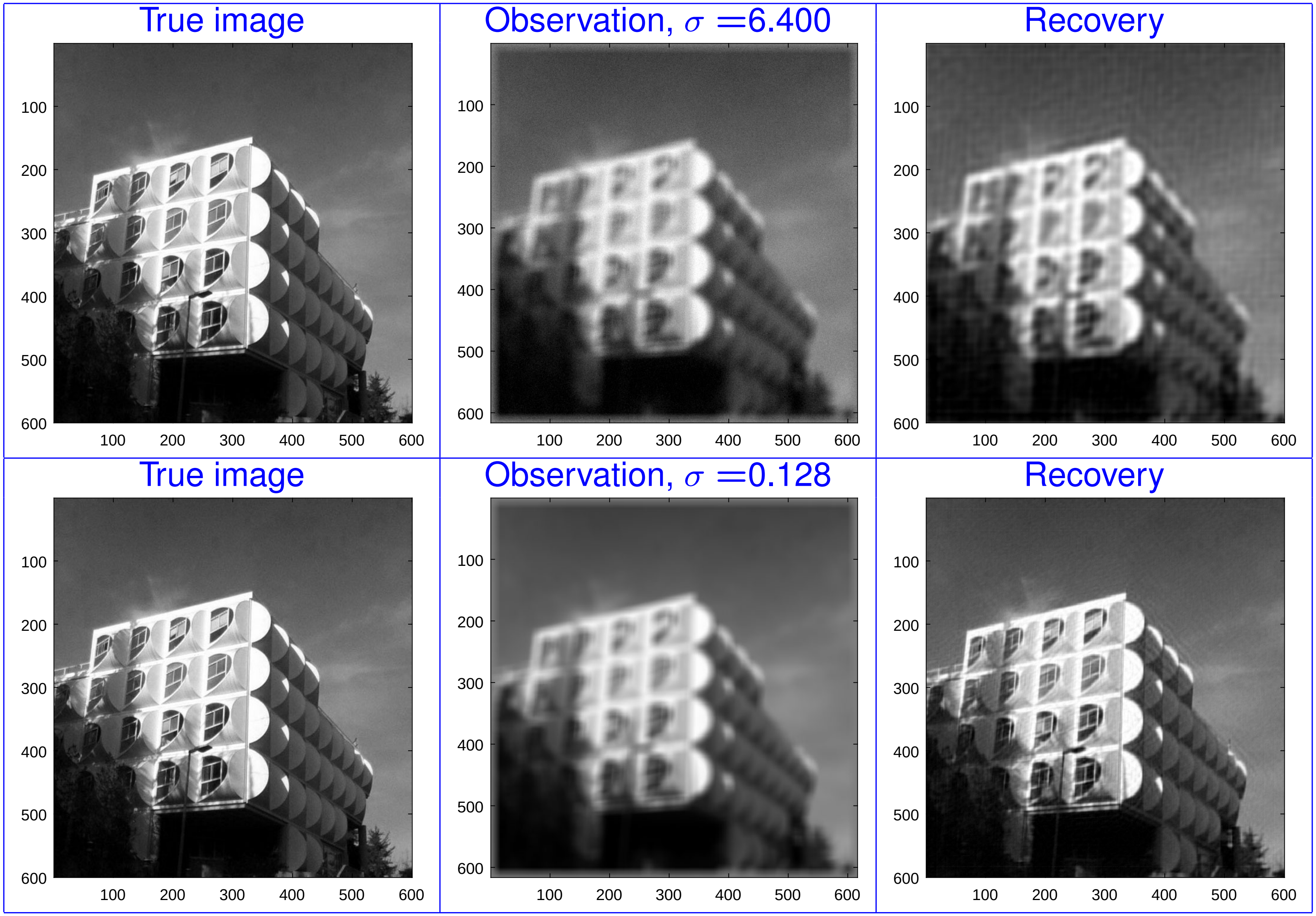

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
50
100
150

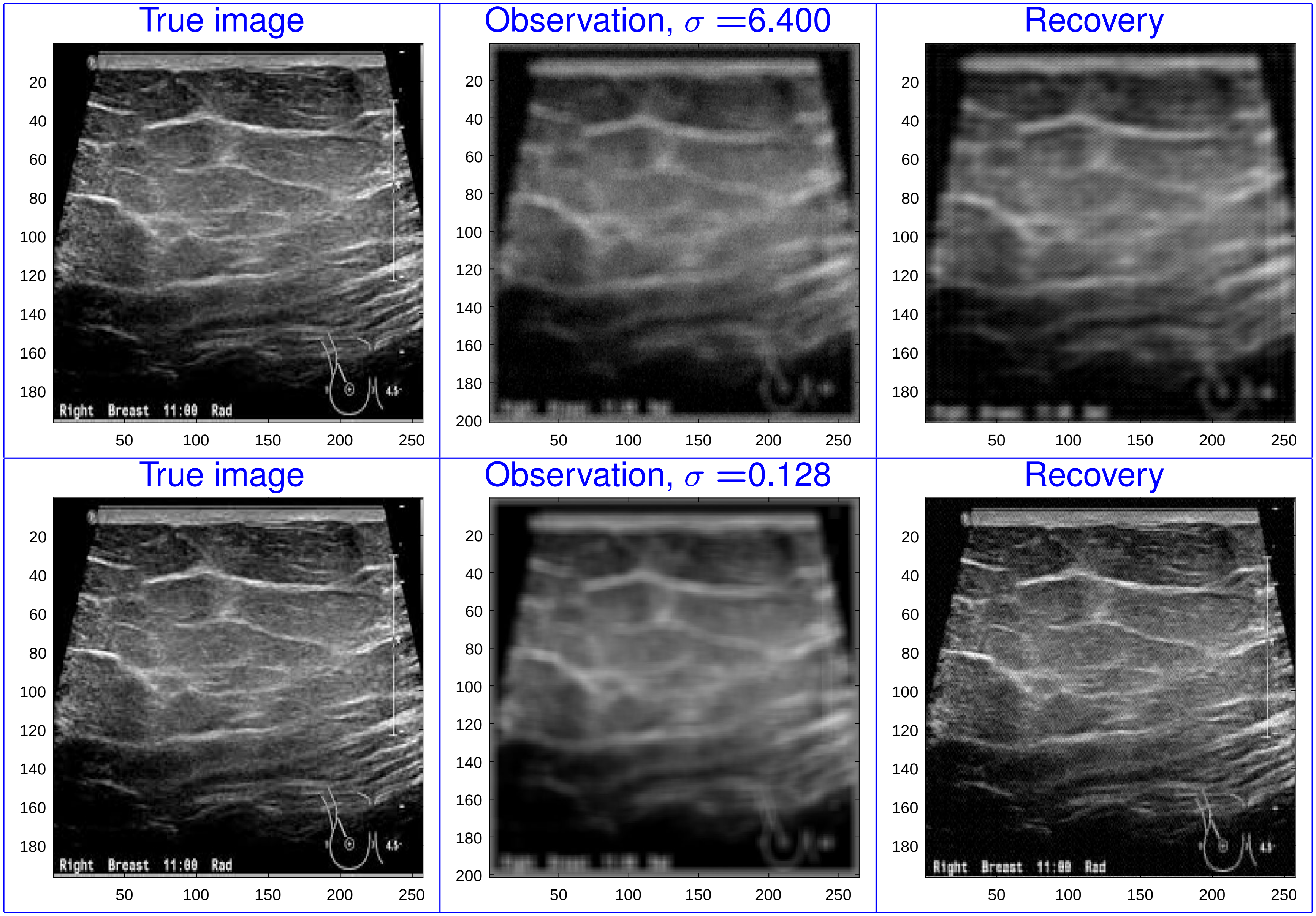
5.154

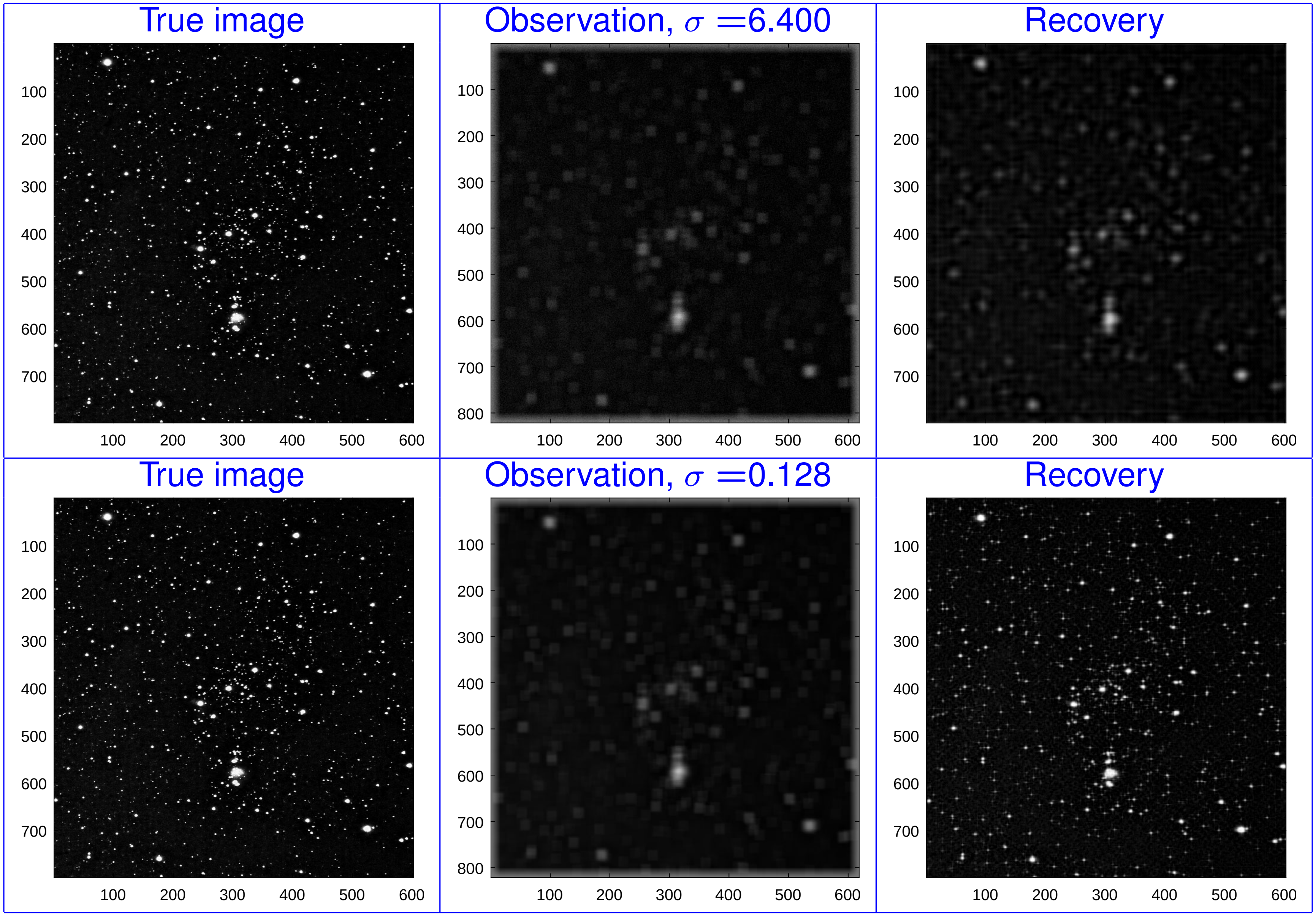

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
450
100
200
300
400
500
600

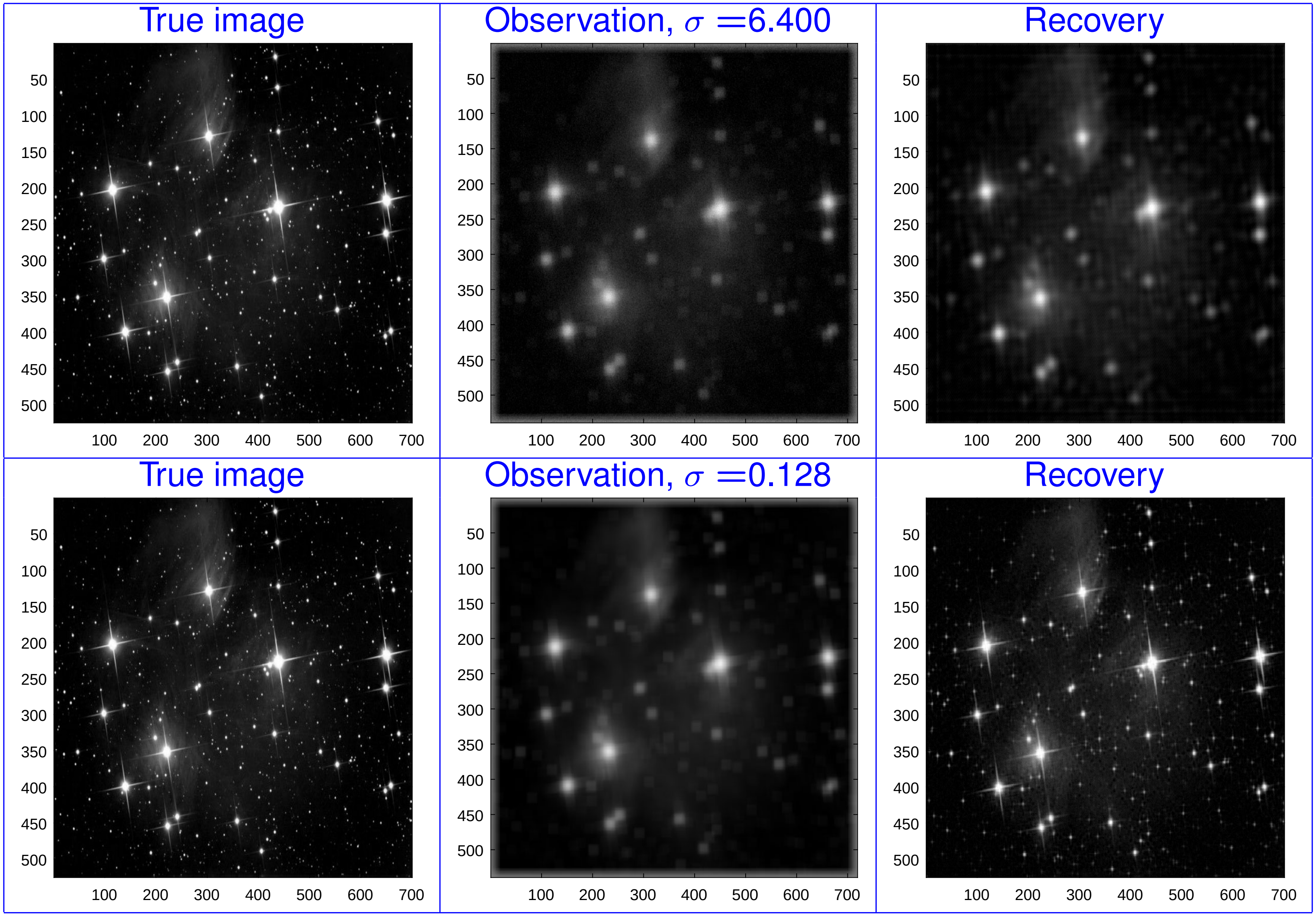

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100
200
300
400
500
600

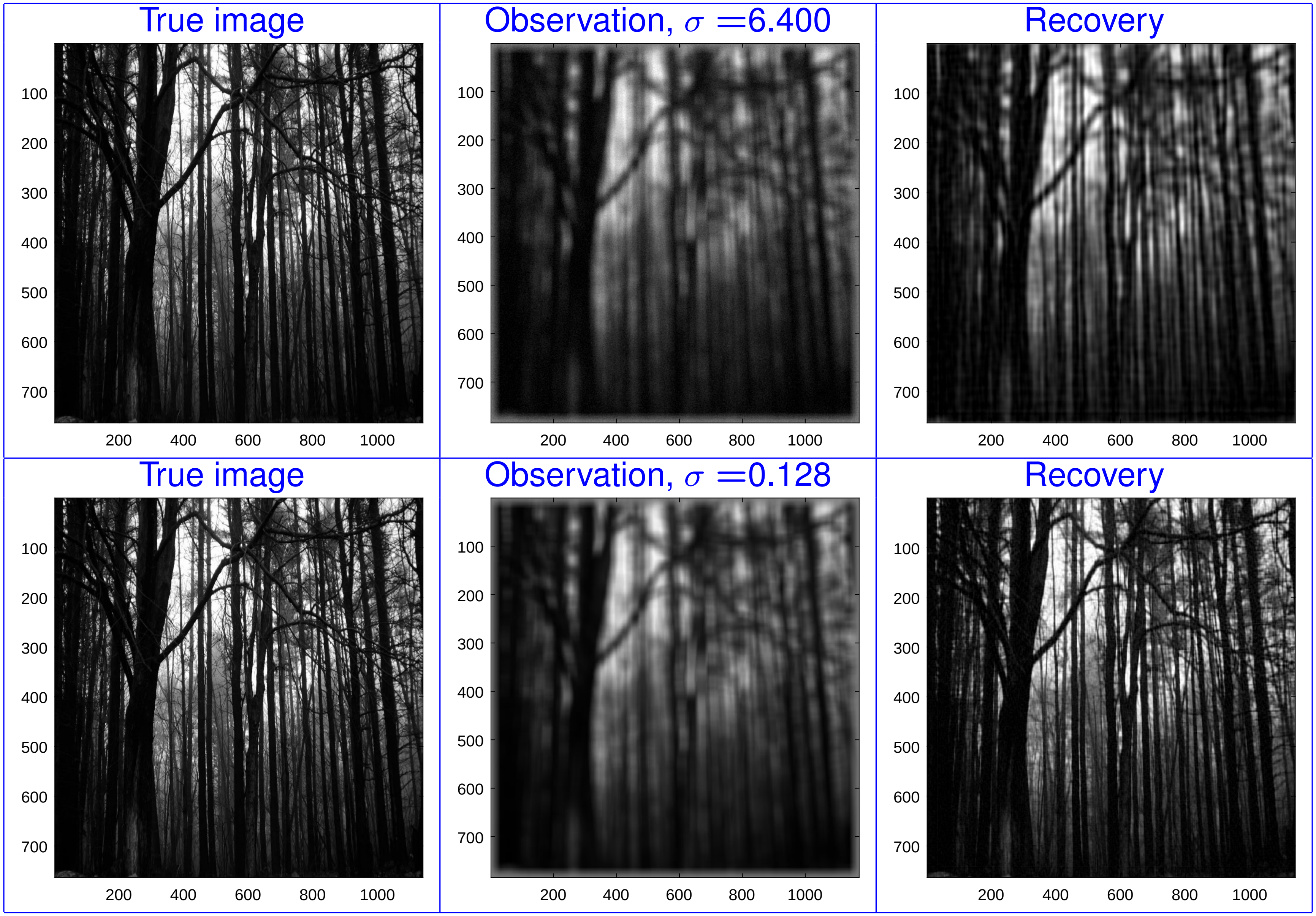

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
Right Breast 11:00 Rad

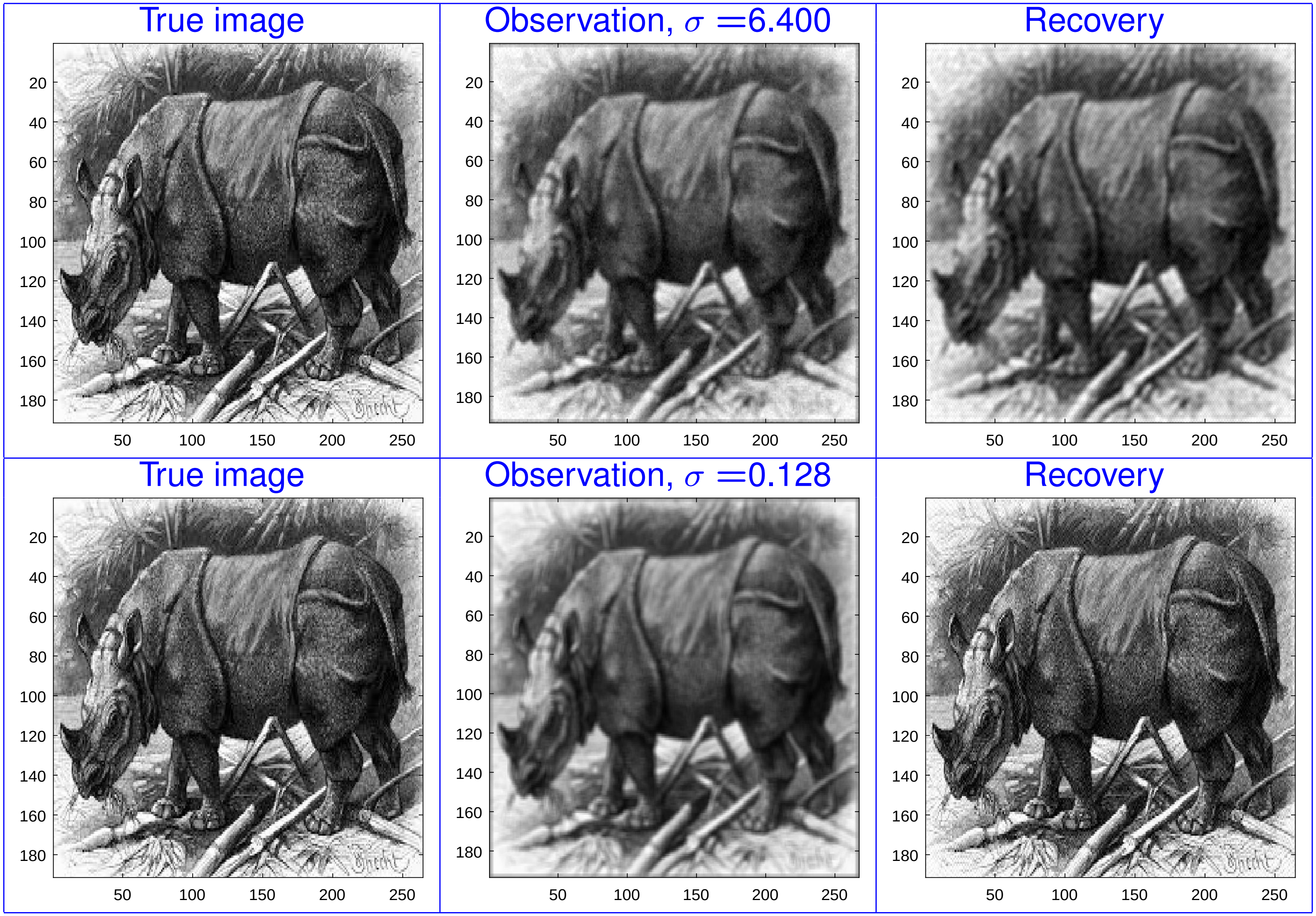

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

5.160

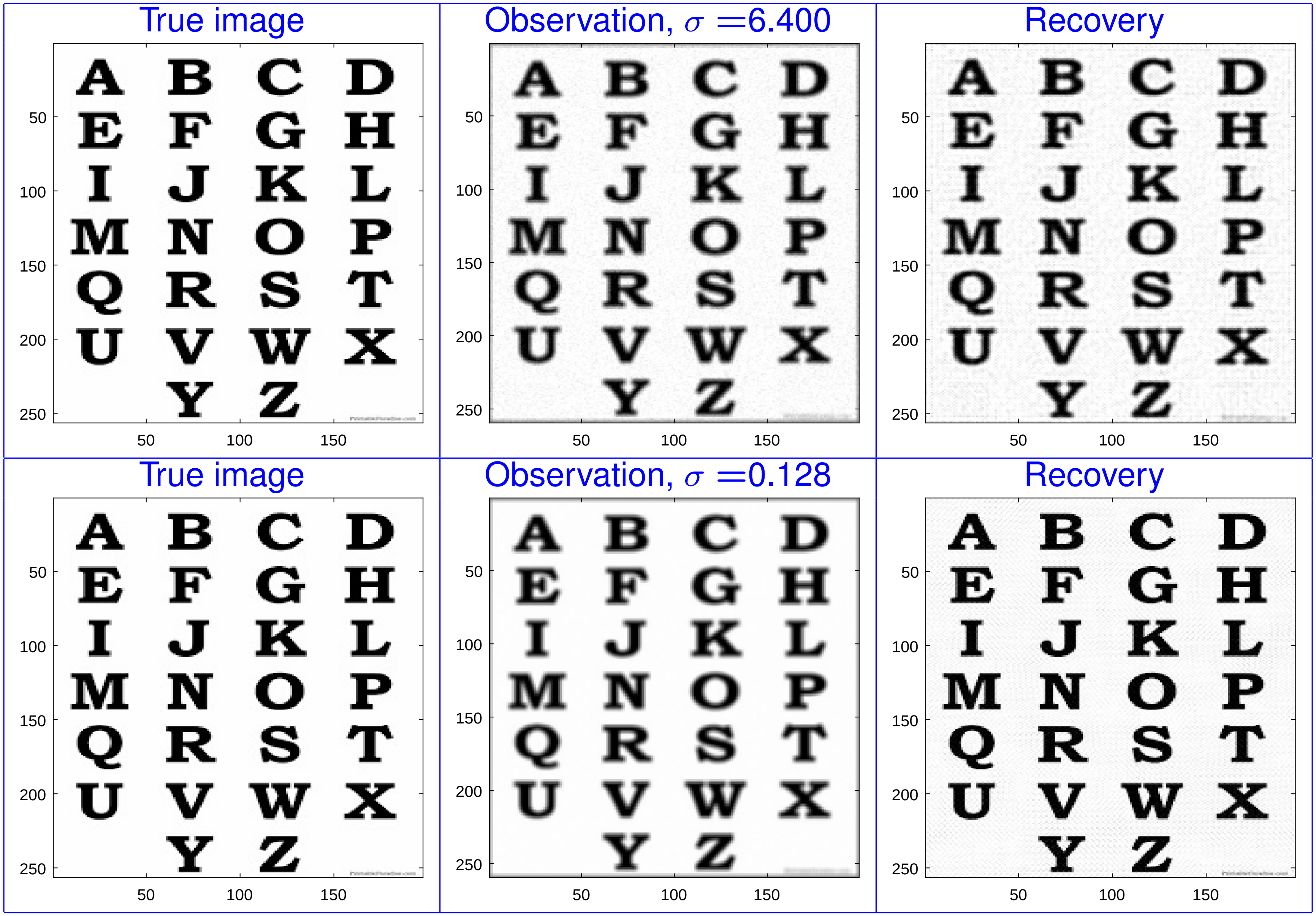

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
450
500
100
200
300
400
500
600
700


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100
200
300
400
500
600
700
200
400
600
800
1000

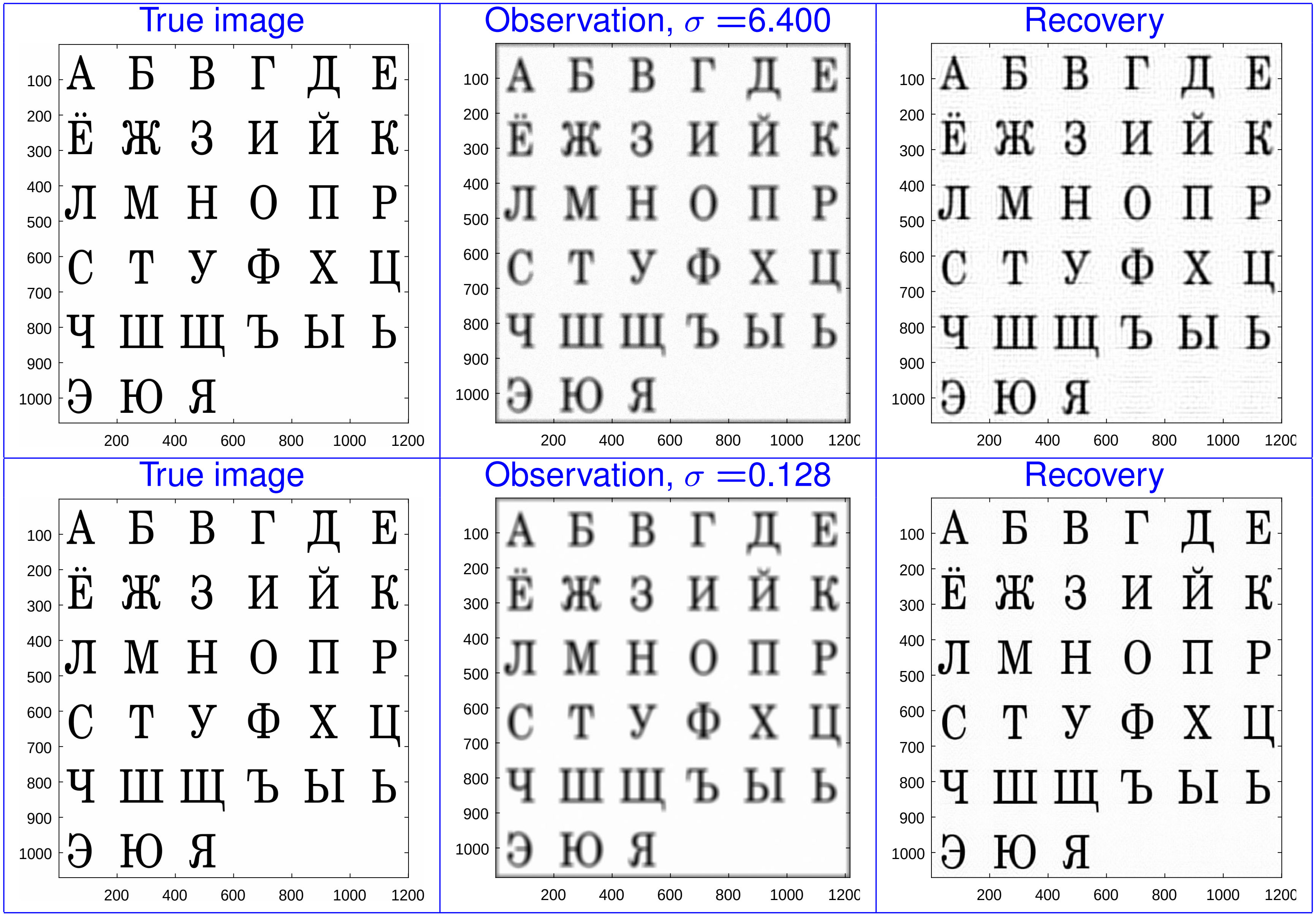
5.162

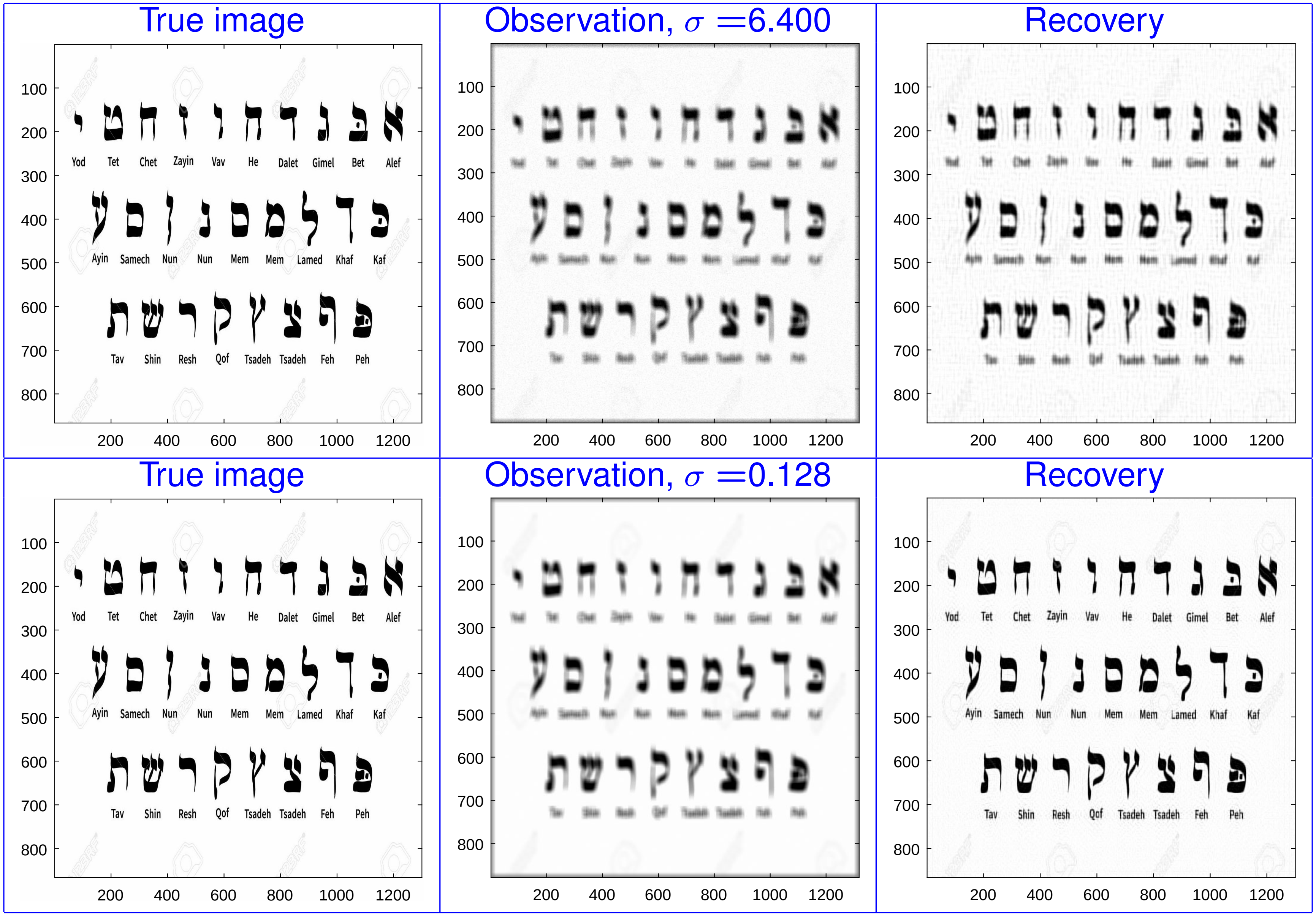

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
20
40
60
80
100
120
140
160
180
50
100
150
200
250

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery



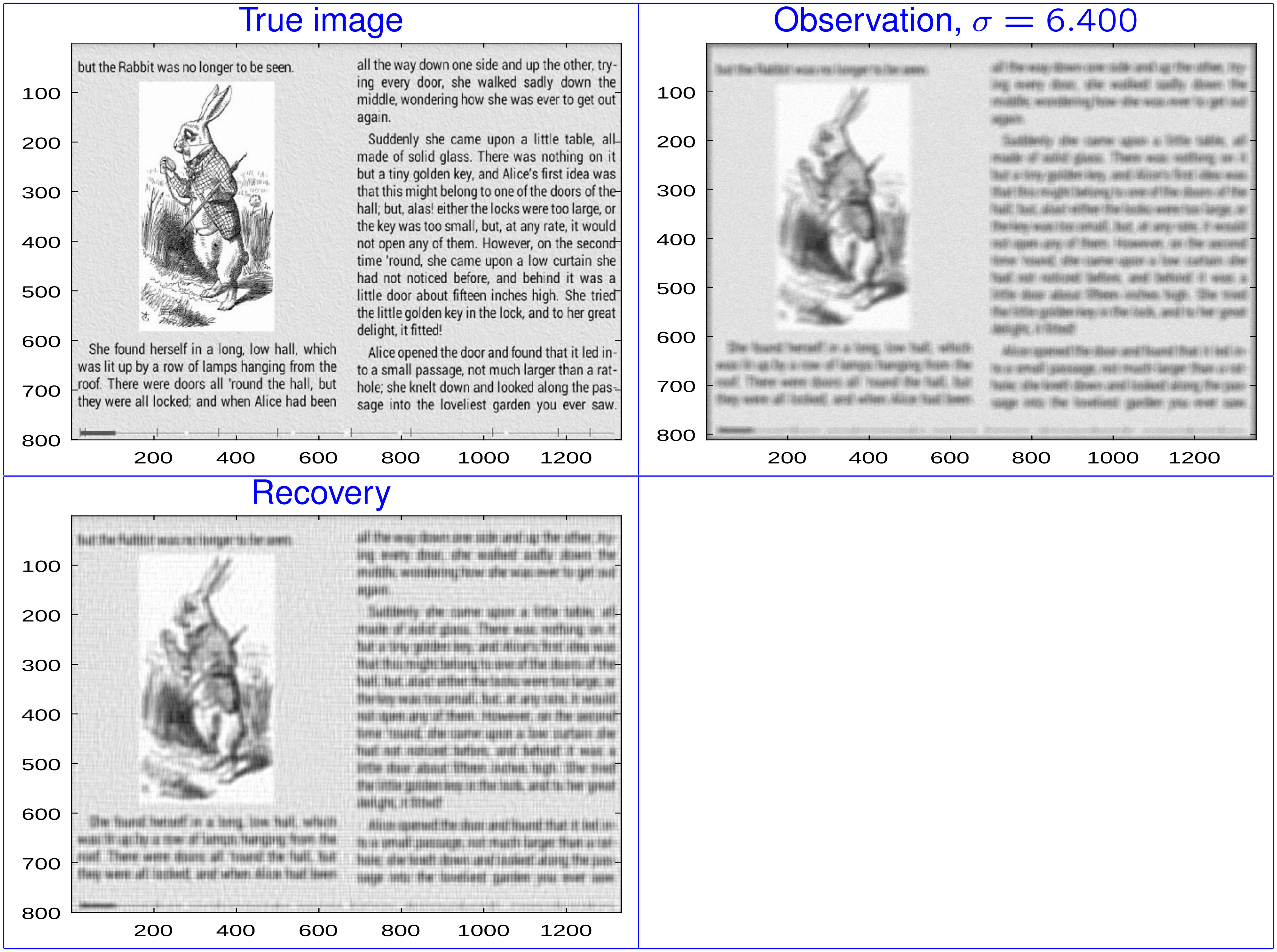
True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

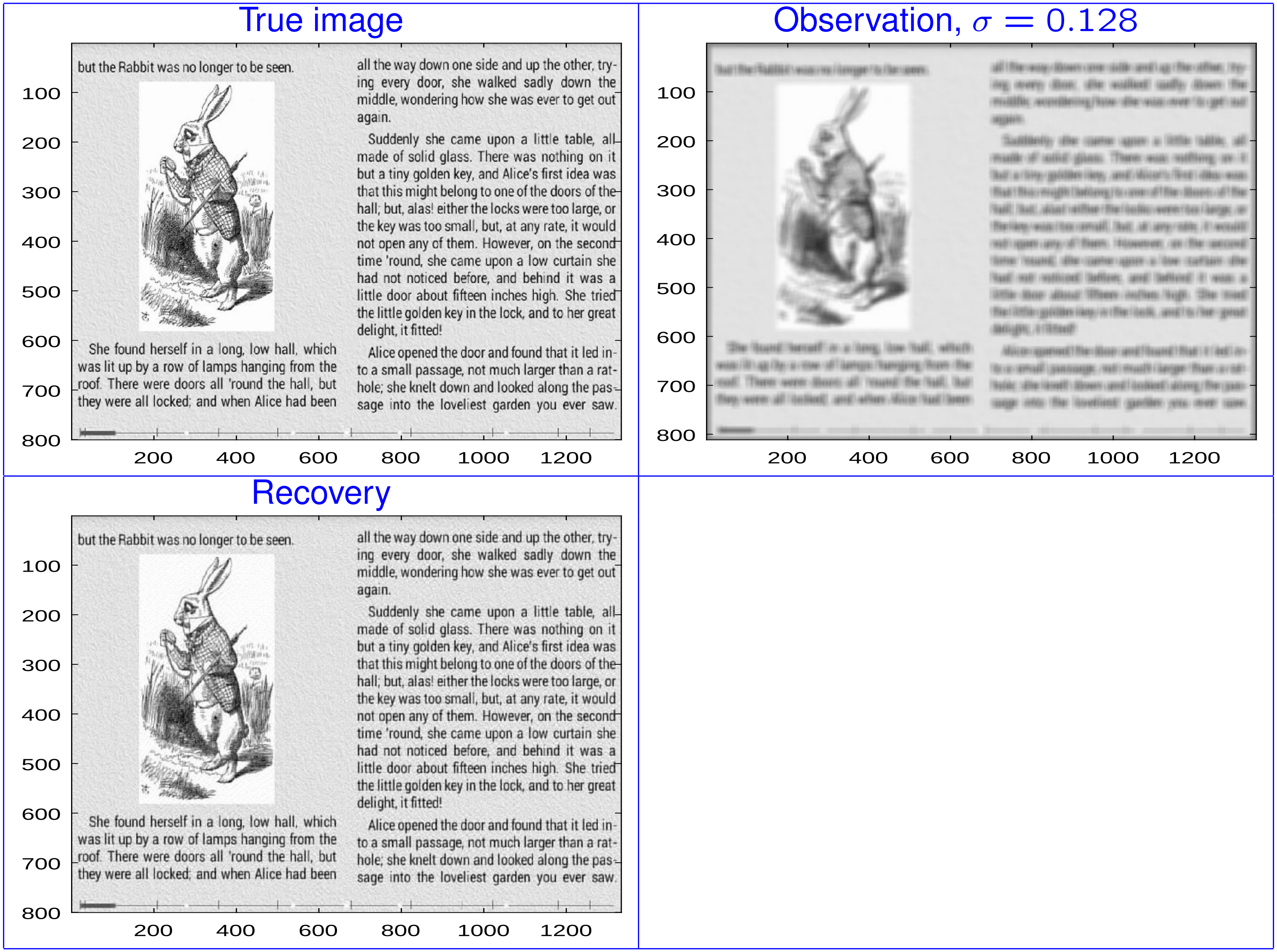

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100 200 300 400 500 600 700 800 900 1000
200 400 600 800 1000 1200

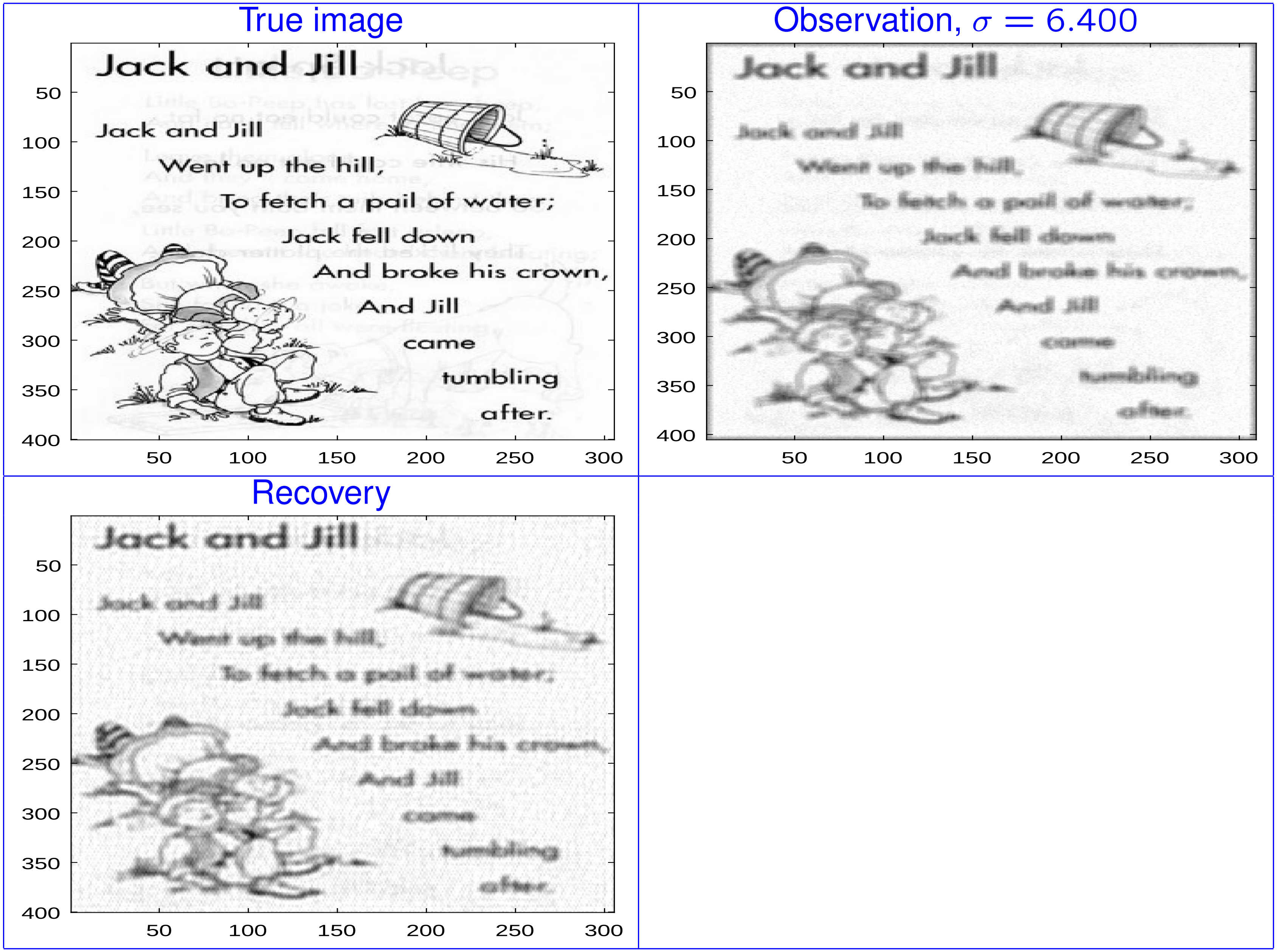


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

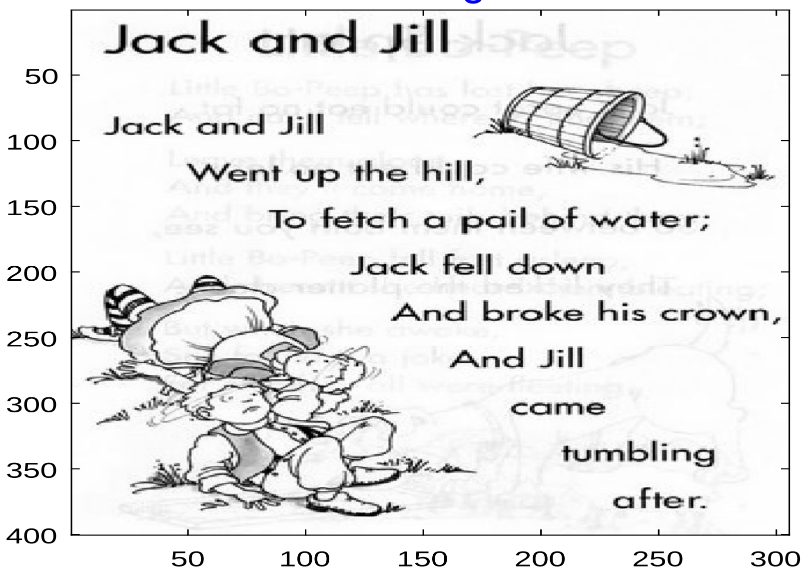

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100
200
300
400
500
600
700
800
200
400
600
800
1000
1200
1 234567 890128 >


True image
but the Rabbit was no longer to be seen.
She found herself in a long, low hall, which was lit up by a row of lamps hanging from the roof. There were doors all 'round the hall, but they were all locked; and when Alice had been
all the way down one side and up the other, trying every door, she walked sadly down the middle, wondering how she was ever to get out again.
Suddenly she came upon a little table, all made of solid glass. There was nothing on it but a tiny golden key, and Alice's first idea was that this might belong to one of the doors of the hall; but, alas! either the locks were too large, or the key was too small, but, at any rate, it would not open any of them. However, on the second time 'round, she came upon a low curtain she had not noticed before, and behind it was a little door about fifteen inches high. She tried the little golden key in the lock, and to her great delight, it fitted!
Alice opened the door and found that it led into a small passage, not much larger than a rat-hole; she knelt down and looked along the passage into the loveliest garden you ever saw.
Observation, σ = 6.400
Recovery
100
200
300
400
500
600
700
800
200
400
600
800
1000
1200

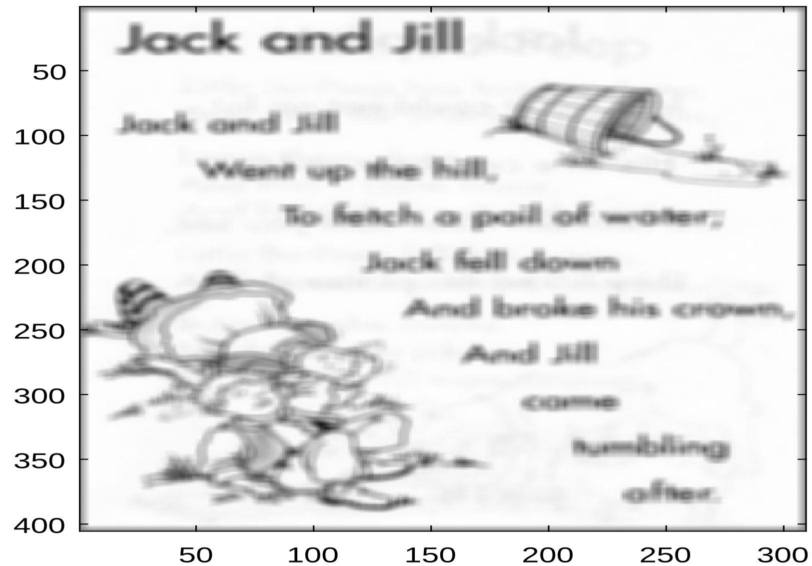

True image
Observation, $\sigma = 0.128$
Recovery
100
200
300
400
500
600
700
800
200
400
600
800
1000
1200
but the Rabbit was no longer to be seen.
She found herself in a long, low hall, which was lit up by a row of lamps hanging from the roof. There were doors all 'round the hall, but they were all locked; and when Alice had been
all the way down one side and up the other, trying every door, she walked sadly down the middle, wondering how she was ever to get out again.
Suddenly she came upon a little table, all made of solid glass. There was nothing on it but a tiny golden key, and Alice's first idea was that this might belong to one of the doors of the hall; but, alas! either the locks were too large, or the key was too small, but, at any rate, it would not open any of them. However, on the second time 'round, she came upon a low curtain she had not noticed before, and behind it was a little door about fifteen inches high. She tried the little golden key in the lock, and to her great delight, it fitted!
Alice opened the door and found that it led into a small passage, not much larger than a rat-hole; she knelt down and looked along the passage into the loveliest garden you ever saw.


True image
Observation, $\sigma = 6.400$
Recovery
Jack and Jill
Jack and Jill
Went up the hill,
To fetch a pail of water;
Jack fell down
And broke his crown,
And Jill
came
tumbling
after.
50
100
150
200
250
300
350
400

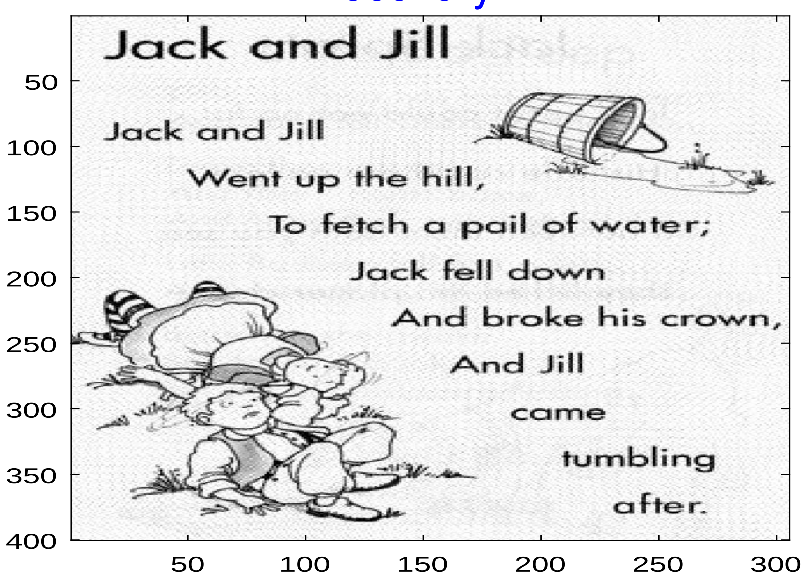
True image

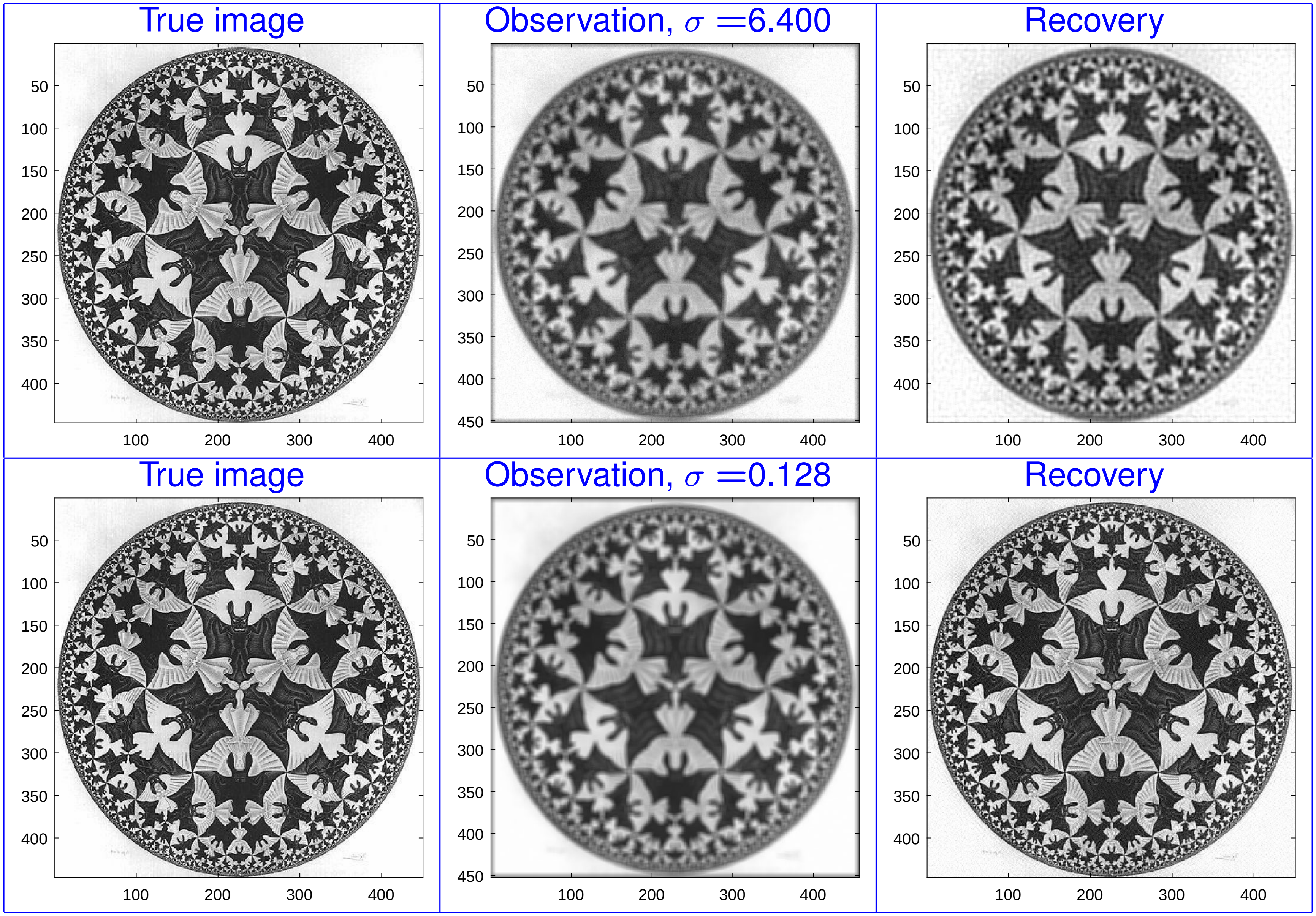


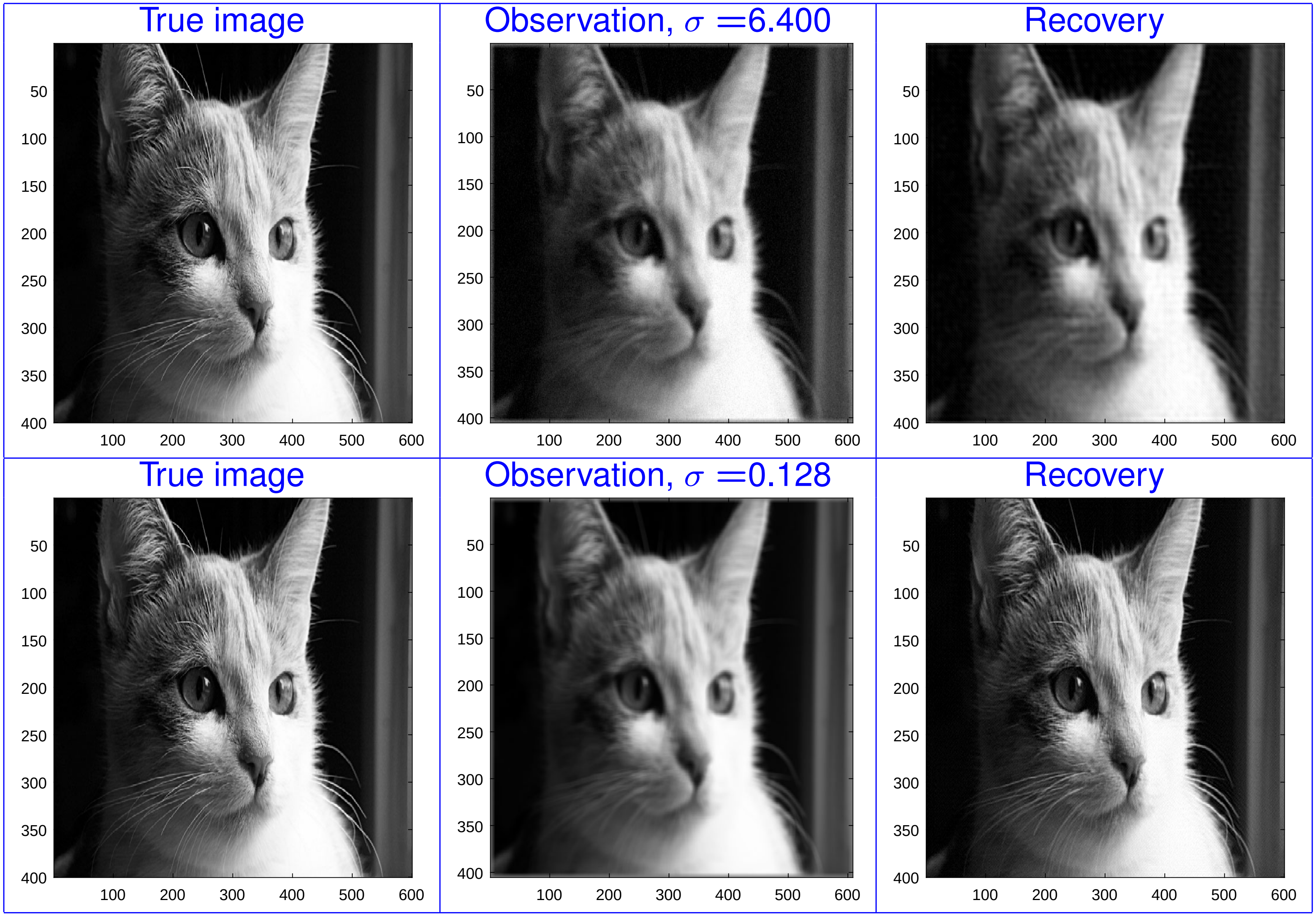
Observation, $\sigma = 0.128$



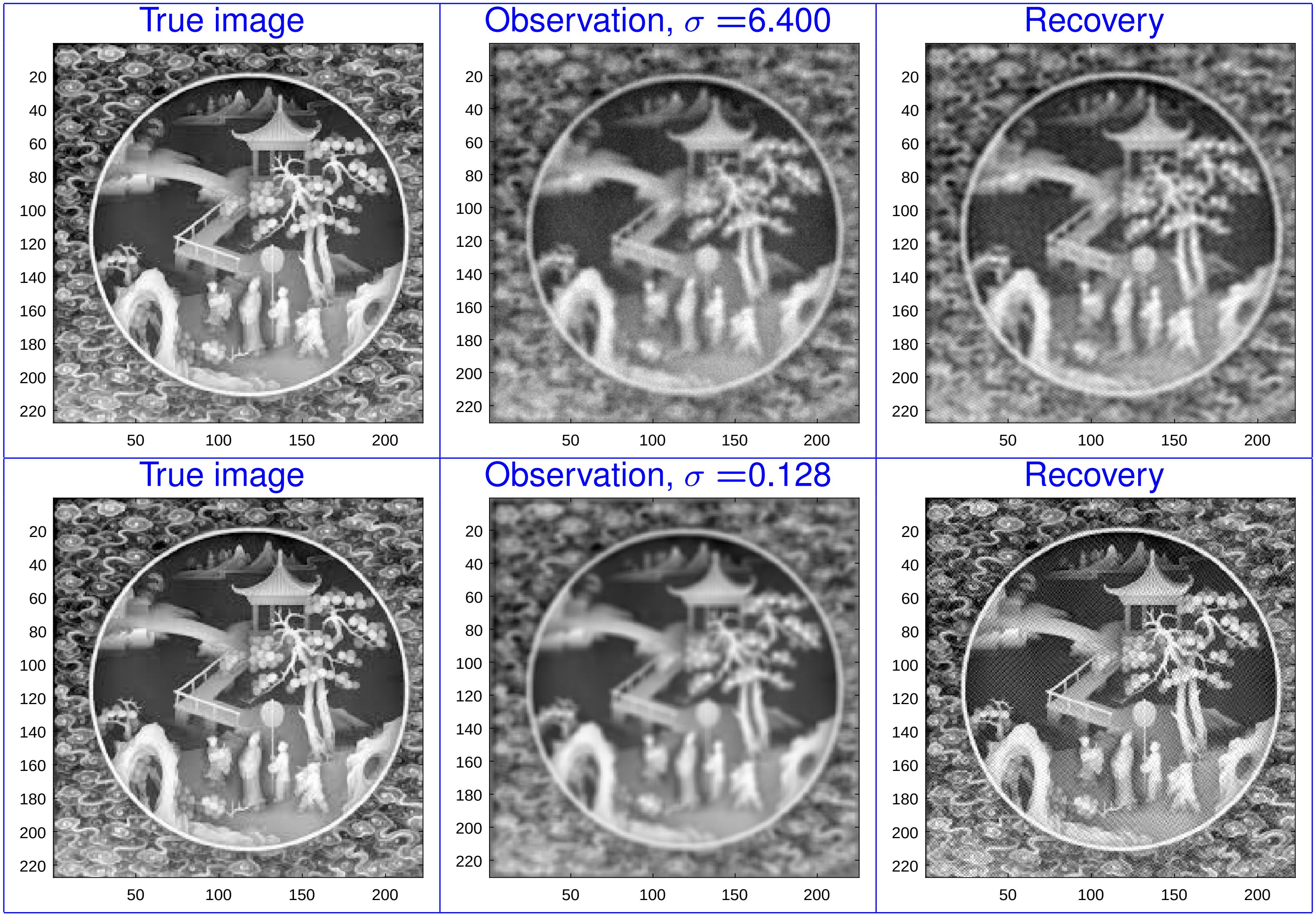
Recovery

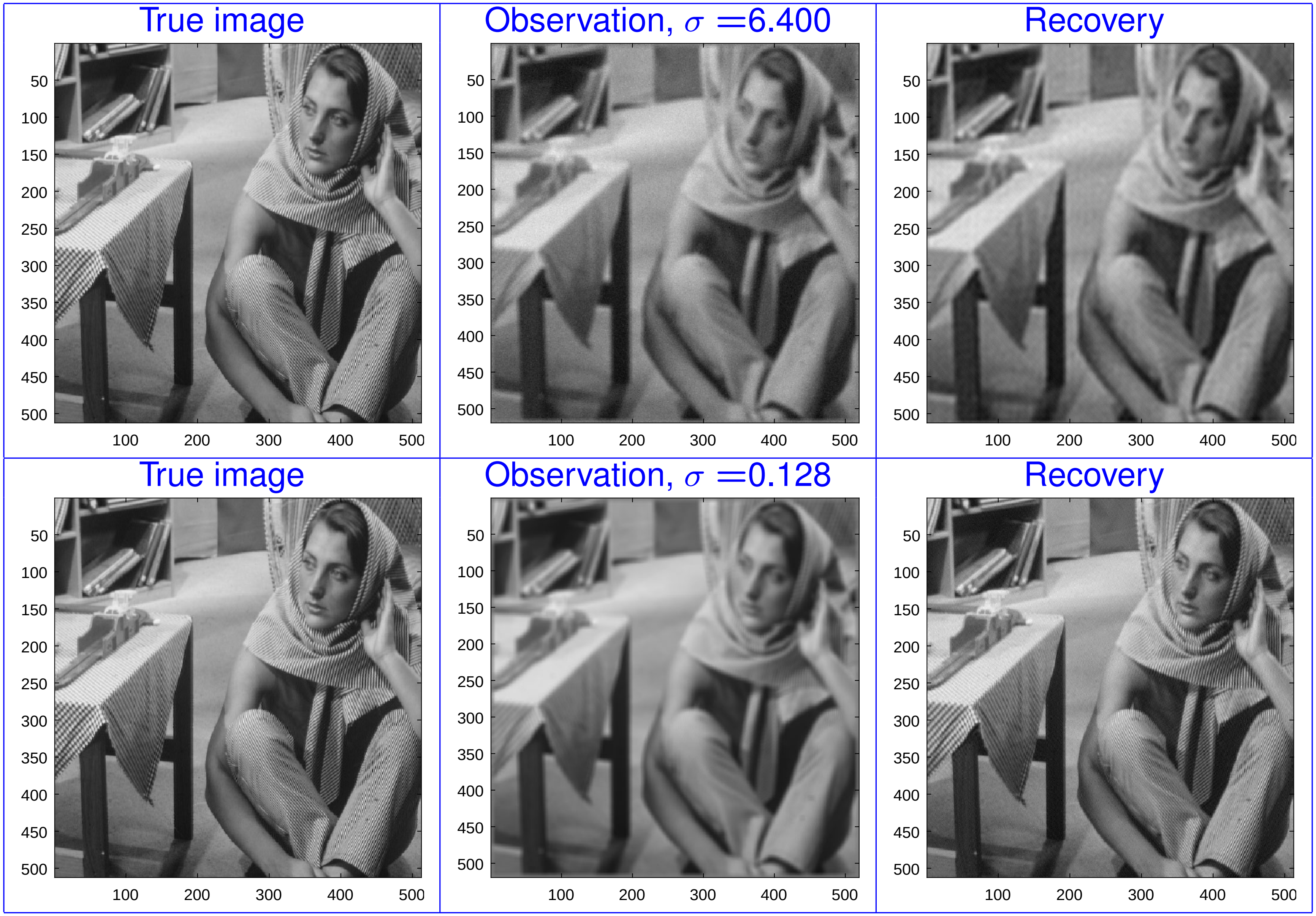



True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
450
100
200
300
400

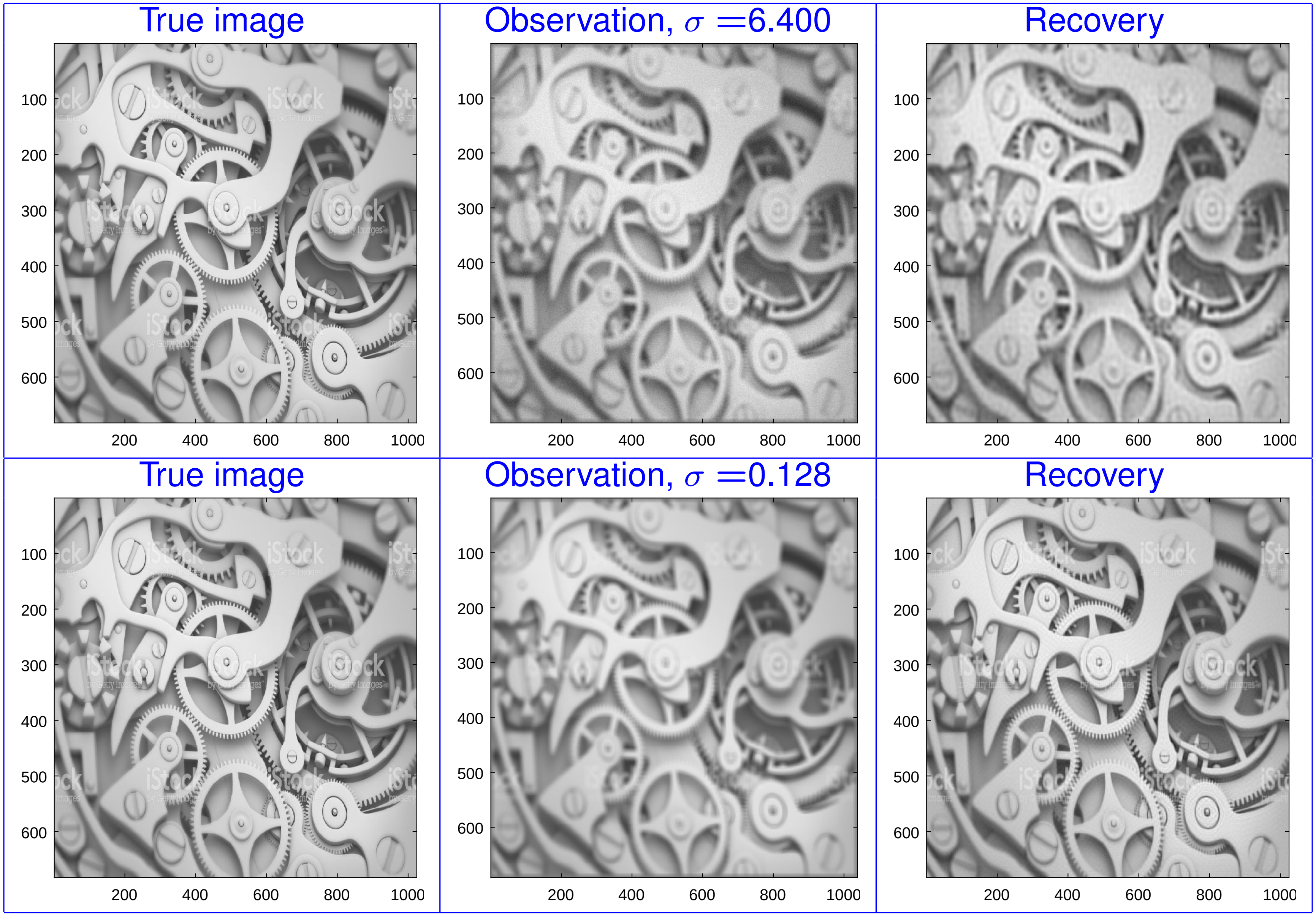


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
100
200
300
400
500
600

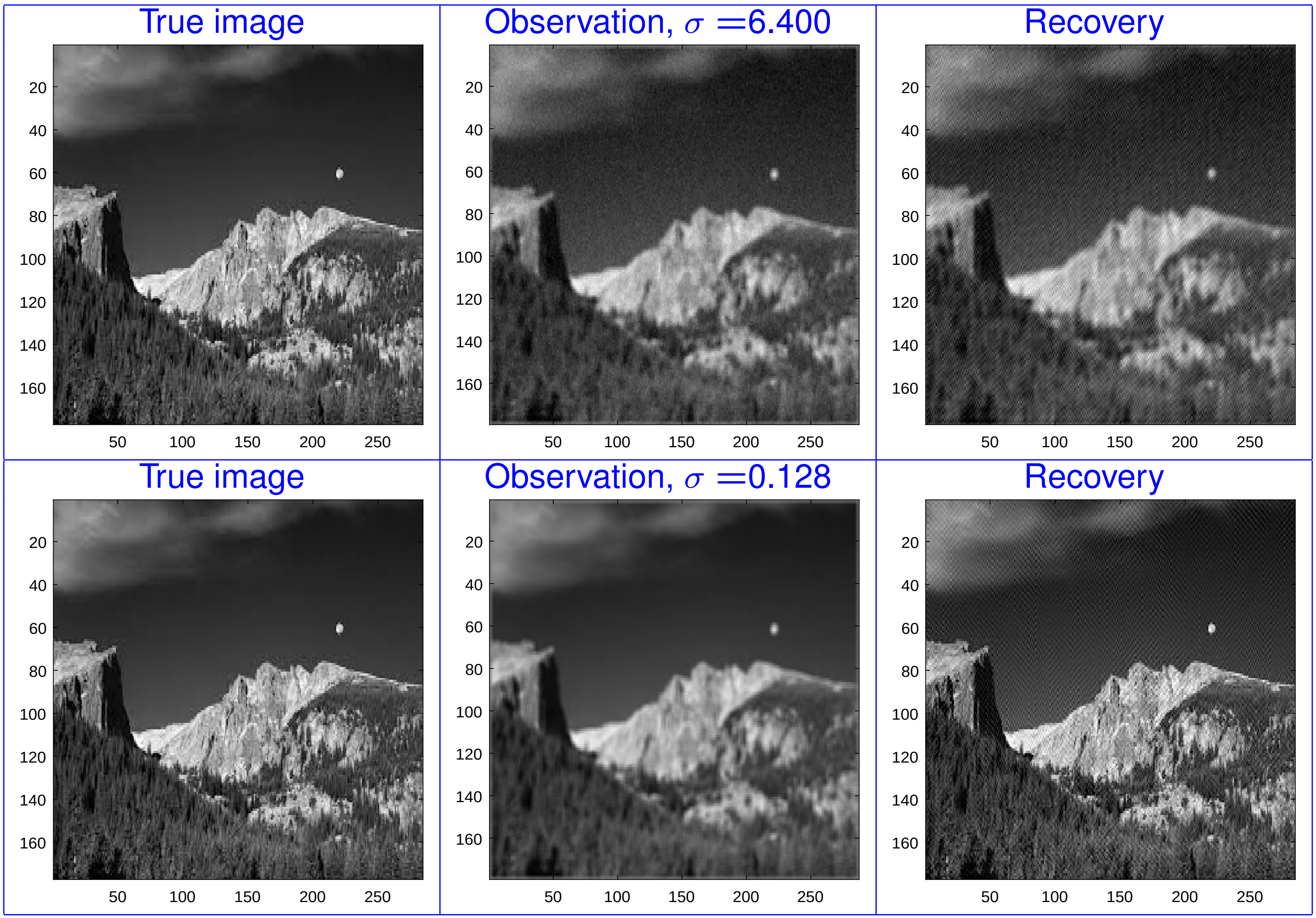

True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

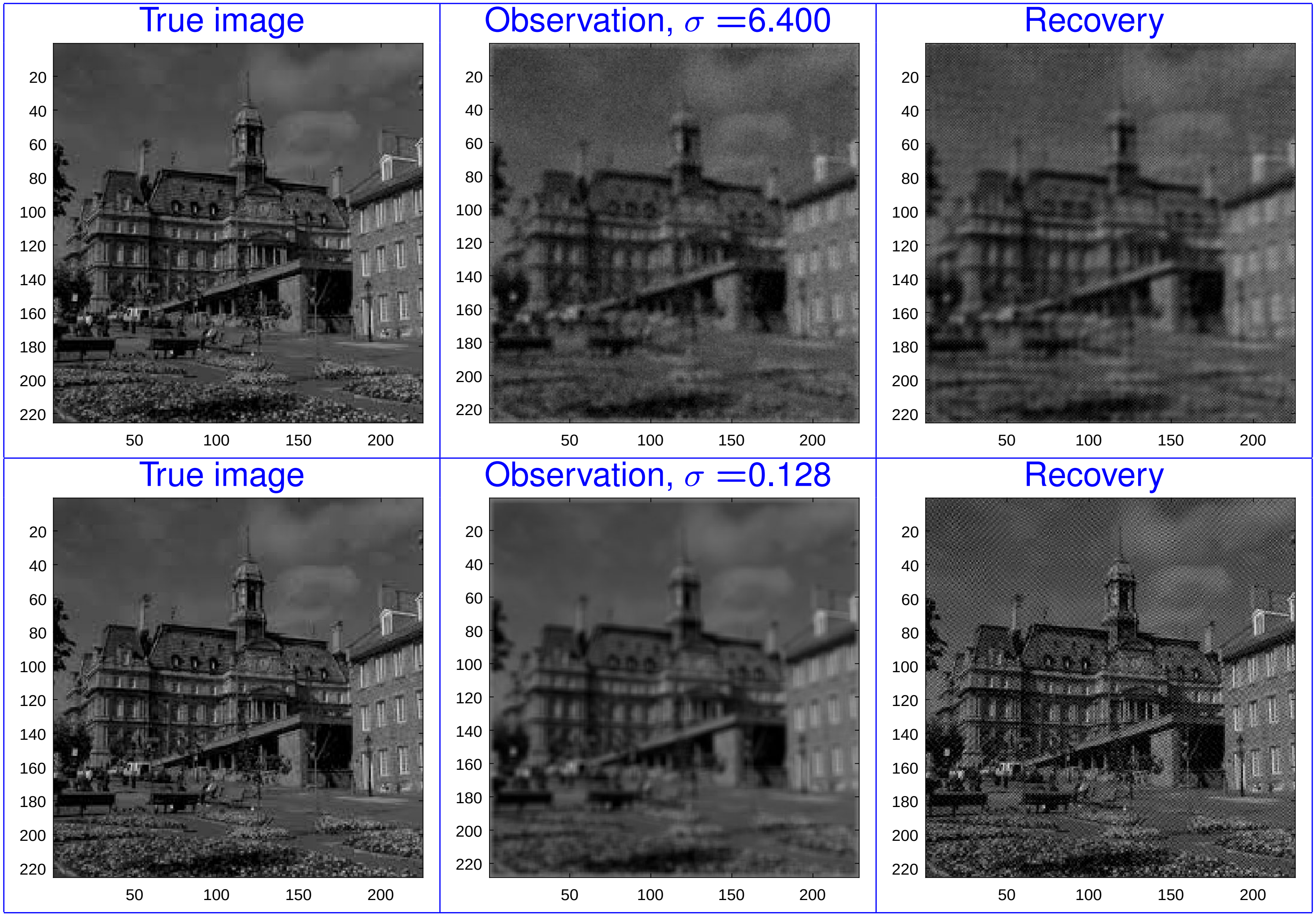


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
50
100
150
200
250
300
350
400
450
500
100
200
300
400
500


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
100
200
300
400
500
600
200
400
600
800
1000

5.177


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery
20
40
60
80
100
120
140
160
50
100
150
200
250


True image
Observation, σ =6.400
Recovery
True image
Observation, σ =0.128
Recovery

.

ESTIMATING SIGNALS IN MONOTONE GENERALIZED LINEAR MODELS

- *Generalized Linear Model*
- *Developing tools*
  - *Variational inequalities with monotone operators*
- *Sample Average Approximation estimate*
- *Stochastic Approximation estimate*
- *Illustrations*
- *Variation: Multi-State Spatio-Temporal Processes*

What the story is about

♣ **Ultimate Goal:** To recover *unknown* signal $x \in \mathbb{R}^n$ from observations

$$\omega^K = (\omega_1, ..., \omega_K)$$

given by

Generalized Linear Model: $\omega_k = (y_k, \eta_k)$, where

— ω_k, $k = 1, ..., K$, are i.i.d.

— the common distribution P of *regressors* η_k is independent of signal x

— the joint distribution of *label* $y_k \in \mathbb{R}^m$ and *regressor* $\eta_k \in \mathbb{R}^{n \times m}$ depends solely on signal x, and

$$\mathbf{E}_{|\eta_k}\{y_k\} = \psi(\eta_k^T x)$$

• $\psi(\cdot) : \mathbb{R}^m \to \mathbb{R}^m$: known *link function* • $\mathbf{E}_{|\eta_k}\{\cdot\}$: conditional, given η_k, expectation over y_k

• We assume that a priori information on signal x reduces to $x \in \mathcal{X}$, for a given convex compact set $\mathcal{X} \subset \mathbb{R}^n$.

$$\boxed{\{(y_k,\eta_k)\}_{k\leq K} \text{ i.i.d., } \mathbf{E}_{|\eta_k}\{y_k\} = \psi(\eta_k^T x),\ x \in \mathcal{X}\ ??\ \Rightarrow ??\, x}$$

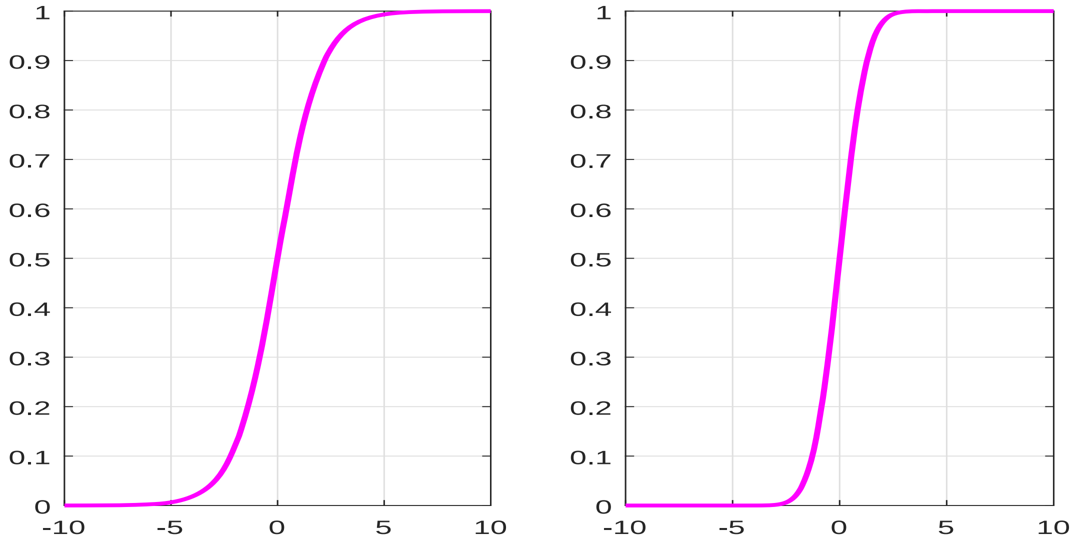
Examples of GLM's:

Linear model: $\psi(s) \equiv s$. *Assuming additive signal- and regressor-independent noise*, the problem becomes to recover signal x from observations (y_k, η_k) , $k \leq K$, where regressors η_k are i.i.d. with independent of x distribution,

$$y_k = \eta_k^T x + \xi_k,$$

and ξ_k, $k \leq K$, are independent of η_k i.i.d. zero mean observation noises.

♣ *Linear model admits "special treatment" which was our previous subject.*



logit (left) and probit (right) link functions

Logit model (Logistic regression): $m = 1$, $\psi(s) = \exp\{s\}/(1 + \exp\{s\})$, $\eta_k \in \mathbb{R}^n$, $1 \leq k \leq K$, are i.i.d.. Given η_k, y_k takes value 1 with probability $\psi(\eta_k^T x)$ and value 0 with complementary probability.

Probit model: Exactly as Logistic Regression, but with the cdf of the standard Gaussian distribution in the role of link: $\psi(s) = \Phi(s) := \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{s} \exp\{-t^2/2\}dt$.

♣ *Both Logit and Probit models are widely used in Regression Analysis with binary dependent variables.*

Signal Recovery in GLM

The standard signal recovery in GLM model is given by *Maximum Likelihood* (ML).
♠ Assuming the conditional, signal x and regressor η given, distribution of the label y to have density $p(y, \eta^T x)$ w.r.t. some reference measure, *the conditional by the sequence of regressors log-likelihood of the sequence of labels as a function of candidate signal* z *is* $\sum_{k=1}^K \ln(p(y_k, \eta_k^T z))$. *The ML estimate* $\widehat{x}$ *of the signal underlying observations is obtained by maximizing log-likelihood in* $z \in \mathcal{X}$.

- *In Linear model* with Gaussian noise the ML estimate is given by Least Squares:

$$\widehat{x} \in \operatorname*{Argmin}_{z \in \mathcal{X}} \sum_{k=1}^K \|y_k - \eta_k^T z\|_2^2$$

- *In Logit model* the ML estimate is

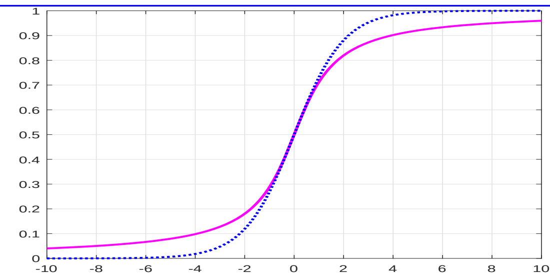
$$\widehat{x} \in \operatorname*{Argmin}_{z \in \mathcal{X}} \sum_{k=1}^K \left[\ln\left(1 + \exp\{\eta_k^T z\}\right) - y_k \eta_k^T z\right]$$

- *In Probit model* the ML estimate is

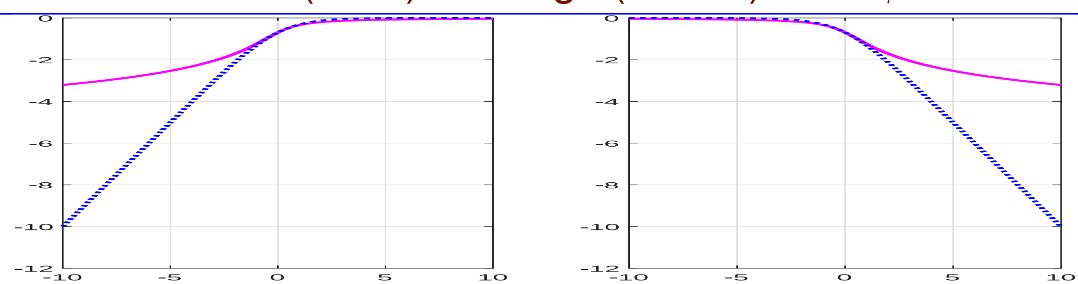
$$\widehat{x} \in \operatorname*{Argmin}_{z \in \mathcal{X}} \sum_{k=1}^K \left[-\ln\left(1 - \Phi(\eta_k^T z)\right) - y_k \ln\left(\frac{\Phi(\eta_k^T z)}{1 - \Phi(\eta_k^T z)}\right)\right]$$

In all these cases likelihood maximization (which we convert to minimizing minus log-likelihood) happens to be convex, and thus efficiently solvable, problem.

However: Minimizing minus log-likelihood in GLM can be a *nonconvex* problem. For example, this happens when the link function $\psi(s) = \exp\{s\}/(1 + \exp\{s\})$ in Logit model is replaced with $\psi(s) = \frac{1}{2} + \frac{1}{\pi}\,\mathrm{atan}(s)$:



atan (solid) and logit (dotted) links ψ



$\ln(\psi)$ (left) and $\ln(1-\psi)$ (right) for atan (solid) and logit (dotted) links
[with binary labels, $\ln(\psi)$, $\ln(1-\psi)$ must be concave to make log-likelihood concave]

$$\boxed{\{(y_k,\eta_k)\}_{k\leq K} \text{ i.i.d., } \mathbf{E}_{|\eta_k}\{y_k\} = \psi(\eta_k^T x),\ x\in\mathcal{X}\ ??\ \Rightarrow ??\, x}$$

Common wisdom is to recover x by minimizing minus log-likelihood by Newton method and to hope for the better.

With non-concave log-likelihood, this approach can fail...

Question: *Can we do better?*

Answer: Yes! Under monotonicity assumption on the link function, there exists an alternative to Maximum Likelihood computationally efficient signal recovery with provably reasonably good performance.

♠ **Monotonicity assumption**, in nutshell, requires from $\psi(\cdot) : \mathbb{R}^m \to \mathbb{R}^m$ to be *monotone*:

$$\langle \psi(s) - \psi(s'), s - s' \rangle \geq 0\ \forall s, s' \in \mathbb{R}^m$$

Motivation: Recovering signal x from noisy observations hardly can be easier than recovering $w = \eta^T x$ from *noiseless* observation

$$y = \psi(w). \tag{$*$}$$

Monotonicity of ψ is, basically, the weakest general-type structural assumption which ensures computational tractability of the square system of nonlinear equations $(*)$.

Executive Summary on Variational Inequalities with Monotone Operators

Definition: *Let $X \subset \mathbb{R}^N$ be a closed convex set and $G : X \to \mathbb{R}^N$ be a vector field. G is called monotone on X, if*

$$\langle G(y) - G(y'), y - y' \rangle \geq 0 \ \forall y, y' \in X. \qquad (*)$$

If $(*)$ can be strengthened to

$$\langle G(y) - G(y'), y - y' \rangle \geq \alpha \|y - y'\|_2^2 \ \forall y, y' \in X, \qquad [\alpha > 0]$$

G is called strongly monotone, with modulus α, on X.

Examples:

A. Univariate ($N = 1$) monotone vector fields on closed convex subset X of $\mathbb{R}$ are exactly non-decreasing real-valued functions on X.

B. If $f : X \to \mathbb{R}$ is convex differentiable on X, the gradient field $G(x) = \nabla f(x)$ of f is monotone on X. The same holds true for (any) subgradient field of convex function $f : X \to \mathbb{R}$, provided that subdifferential of f at every point $x \in X$ is nonempty.

C. Let $X = U \times V$, and $f(u, v)$ be differentiable on X convex in $u \in U$ and concave in $v \in V$ function. Then the vector field

$$G(u, v) = [\nabla_u f(u, v); -\nabla_v f(u, v)]$$

is monotone on X. The same holds true when smoothness of f is weakened to Lipschitz continuity, and ∇_u, ∇_v are replaced with respective partial sub- and supergradients.

Fact: *Let $G : X \to \mathbb{R}^N$ be continuously differentiable vector field on a closed convex subset X, $\mathrm{int}\, X \neq \emptyset$, of $\mathbb{R}^N$. G is monotone on X iff the symmeterized Jacobian*

$$J_{\mathrm{s}}[G](x) := \frac{1}{2}\left[\frac{\partial G(x)}{\partial x} + \left[\frac{\partial G(x)}{\partial x}\right]^T\right]$$

is positive semidefinite for all $x \in X$. G is strongly monotone with modulus $\alpha > 0$ on X iff $J_{\mathrm{s}}[G](x) \succeq \alpha I_N$, $x \in X$.

Variational Inequality $\mathrm{VI}(G, X)$ *associated with closed convex set X and a monotone on X vector field G reads*

$$\textit{find } z_* \in X : \langle G(z), z - z_* \rangle \geq 0 \;\forall z \in X \qquad (*)$$

Vectors $z_* \in X$ satisfying $(*)$ are called weak solutions to $\mathrm{VI}(G, X)$. A strong solution to $\mathrm{VI}(G, X)$ is a point $z_* \in X$ such that

$$\langle G(z_*), z - z_* \rangle \geq 0 \;\forall z \in X.$$

- *A strong solution is a weak one, since by monotonicity $\langle G(z), z - z_* \rangle \geq \langle G(z_*), z - z_* \rangle,\; z, z_* \in X$. The inverse is true provided that G is continuous on X.*

Note: If $z_* \in X$ is a zero of $G(\cdot)$: $G(z_*) = 0$, then z_* clearly is a strong solution to $\mathrm{VI}(G, X)$. Strong solution is a "substitution" of zero of G - it can exist when G does not vanish at any point of X, And a weak solution is a "substitution" of a strong one: for a monotone G, weak solution does exist whenever X is convex compact set. When G is monotone *and continuous* on X, weak and strong solutions are the same.

$X \subset \mathbb{R}^m$: closed and convex $\quad G : X \to \mathbb{R}^m$: monotone on X

Weak solution to $\mathrm{VI}(G, X)$: $\quad z_* \in X$ such that $\langle G(z), z - z_* \rangle \geq 0 \, \forall z \in X$

Strong solution to $\mathrm{VI}(G, X)$: $\quad z_* \in X$ such that $\langle G(z_*), z - z_* \rangle \geq 0 \, \forall z \in X$

Facts:

- Weak solutions to $\mathrm{VI}(G, X)$ form a closed convex subset of X; this set is nonempty, *provided X is bounded*.
- When G is a subgradient field of continuous convex function $f : X \to \mathbb{R}$, weak solutions to $\mathrm{VI}(G, X)$ are exactly the minimizers of f on X. More generally, when G is the monotone vector field associated with continuous convex-concave $f(u, v) : X = U \times V \to \mathbb{R}$, the weak solutions to $\mathrm{VI}(G, X)$ are exactly the saddle points of f on $U \times V$.

Fact: *Approximating weak solutions to Monotone Variational Inequalities is computationally tractable task – all basic algorithms of convex minimization admit "VI versions."*

Let us define *inaccuracy* $\mathrm{Res}(x|G,X)$ of a candidate solution $z \in X$ to the VI

$$\text{find } z_* \in X : \langle G(z), z - z_* \rangle \geq 0 \;\forall z \in X$$

as

$$\mathrm{Res}(z|G,X) = \sup_{y \in X} \langle G(y), z - y \rangle,$$

so that $\mathrm{Res}(z|G,X) \geq 0$ and $\mathrm{Res}(z|G,X) = 0$ iff z is a weak solution to $\mathrm{VI}(G,X)$.

Fact: *Approximating weak solutions to Monotone Variational Inequalities is computationally tractable task – all basic algorithms of convex minimization admit "VI versions"*

For example, *assuming that*

— X is closed convex set contained in a given $\|\cdot\|_2$-ball of radius R and containing ball of a given radius $r > 0$,

— G is monotone on X and $\|G(x)\|_2 \leq V$, $x \in X$, for some known V,

for every $\epsilon \in (0, VR)$, *a solution* $z \in X$ *with* $\mathrm{Res}(z|G, X \leq \epsilon)$ *can be found*

- **by Ellipsoid method** – *in* $O(1)N^2 \ln\left(\frac{NVR}{\epsilon} \cdot \frac{R}{r} + 1\right)$ *iterations*, with the computational effort per iteration dominated by the necessity
  - (a) to check whether a point belongs to X, and if not - to separate the point from X by a linear form,
  - (b) to compute the value of G at a point of X, and
  - (c) to perform, on the top of (a), (b), $O(N^2)$ additional arithmetic operations
- **by Subgradient Descent** – *in* $O(1)\frac{V^2R^2}{\epsilon^2}$ *iterations*, with computational effort per iteration dominated by the necessity to compute metric projection of a point onto X and the value of G at a point;
- **by Mirror Prox** – *in* $O(1)\frac{LR^2}{\epsilon}$ *iterations*, provided G is Lipschits continuous, with constant L, on X, with the same iteration complexity as for Subgradient Descent.

Strongly Monotone Variational Inequalities

$$\boxed{\text{find } z_* \in X : \langle G(z), z - z_* \rangle \geq 0 \ \forall z \in X \qquad (\mathrm{VI}(G, X))}$$

Fact: *Let G be strongly monotone, with modulus $\alpha > 0$, on convex compact set X. Then the weak solution z_* to $\mathrm{VI}(G, X)$ is unique, and for every $z \in X$ it holds*

$$\begin{array}{ll} (a) & \alpha\|z - z_*\|_2^2 \leq \langle G(z), z - z_* \rangle \\ (b) & \alpha\|z - z_*\|_2^2 \leq 4\mathrm{Res}(z|G, X) \end{array}$$

Indeed, setting $z_t = z_* + t(z - z_*)$, for $0 < t < 1$ we have

$$\langle G(z), z - z_t \rangle \geq \alpha\|z - z_t\|_2^2 + \langle G(z_t), z - z_t \rangle$$

by strong monotonicity, and $\langle G(z_t), z - z_t \rangle = \frac{1-t}{t}\langle G(z_t), z_t - z_* \rangle \geq 0$.

$\Rightarrow \langle G(z), z - z_t \rangle \geq \alpha\|z - z_t\|_2^2 \forall t \in (0, 1)$

$\Rightarrow [t \to +0]$ (a).

Next, by (a) applied to $z_{\frac{1}{2}}$ in the role of z, $\langle G(z_{\frac{1}{2}}), z_{\frac{1}{2}} - z_* \rangle \geq \frac{\alpha}{4}\|z - z_*\|_2^2$

$\Rightarrow \mathrm{Res}(z|G, X) \geq \langle G(z_{\frac{1}{2}}), z - z_{\frac{1}{2}} \rangle = \langle G(z_{\frac{1}{2}}), z_{\frac{1}{2}} - z_* \rangle \geq \frac{\alpha}{4}\|z - z_*\|_2^2$

$\Rightarrow (b)$.

$$\boxed{\{(y_k,\eta_k)\}_{k\leq K} \text{ i.i.d., } \mathbf{E}_{|\eta_k}\{y_k\}=\psi(\eta_k^Tx),\ x\in\mathcal{X}\ ??\ \Rightarrow ??\, x}$$

Main Observation: *Under* (slightly strengthened, see below) *Monotonicity Assumption*

"ψ is continuous and monotone on $\mathbb{R}^m$"

the signal x underlying observations in GLM is the unique weak solution to a Variational Inequality $\mathrm{VI}(G,\mathcal{X})$ with strongly monotone on $\mathcal{X}$ vector field G.

Indeed, given GLM, let P be the distribution of regressors η_k, and let

$$F(z)=\mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^Tz)\}$$

Observe that for fixed $\eta\in\mathbb{R}^{n\times m}$, $z\mapsto F_\eta(z):=\eta\psi(\eta^Tz)$ is a vector field on $\mathbb{R}^n$ *and this field is monotone and continuous along with ψ:*

$$z,z'\in\mathbb{R}^N\Rightarrow\langle\eta\psi(\eta^Tz)-\eta\psi(\eta^Tz'),z-z'\rangle=\langle\psi(\eta^Tz)-\psi(\eta^Tz'),\eta^Tz-\eta^Tz'\rangle\geq 0.$$

Under mild regularity assumptions, monotonicity and continuity are preserved when taking expectation w.r.t. η. Assuming from now on that

— the distribution P of η has finite moments of all orders, and

— $\psi(\cdot):\mathbb{R}^m\to\mathbb{R}^m$ is monotone, continuous, and with polynomial growth at infinity,

the vector field F is well defined, continuous, and monotone.

$$\boxed{\begin{array}{c}\{(y_k,\eta_k)\}_{k\leq K}\text{ i.i.d., }\mathbf{E}_{|\eta_k}\{y_k\}=\psi(\eta_k^Tx),\ x\in\mathcal{X}\ ??\Rightarrow ??\,x\\ F(z)=\mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^Tz)\}\end{array}}$$

Let us make

Assumption A: *The monotone vector field F is strongly monotone, with modulus $\alpha > 0$, on $\mathcal{X}$.*

It is immediately seen that a simple *sufficient* condition for Assumption A is strong monotonicity of ψ on bounded subsets of $\mathbb{R}^m$ *plus* positive definiteness of the second order moment matrix $\mathbf{E}_{\eta\sim P}\{\eta\eta^T\}$ *plus* compactness of $\mathcal{X}$.

Observe that

A: *Underlying observations signal x is zero of continuous and monotone vector field*

$$G(z) = F(z) - F(x) : \mathcal{X} \to \mathbb{R}^n;$$

under Assumption **A**, *G is strongly monotone, with modulus $\alpha > 0$, on $\mathcal{X}$.*

B. *For every fixed $z \in \mathcal{X}$ and every k, observation (y_k, η_k) induces unbiased estimate*

$$G_{y_k,\eta_k}(z) = \eta_k\psi(\eta_k^Tz) - \eta_k y_k.$$

of $G(z)$.

Indeed,

$$\mathbf{E}_{y,\eta}\left\{\eta\psi(\eta^Tz) - \eta y\right\} = \mathbf{E}_{\eta\sim P}\left\{\eta\psi(\eta^Tz) - \eta\mathbf{E}_{|\eta}\{y\}\right\} = \mathbf{E}_{\eta\sim P}\left\{\eta\psi(\eta^Tz) - \eta\psi(\eta^Tx)\right\} = F(z) - F(x)$$

$\{(y_k, \eta_k)\}_{k\le K}$ i.i.d., $\mathbf{E}_{\vert\eta_k}\{y_k\} = \psi(\eta_k^T x)$, $x \in \mathcal{X}$?? $\Rightarrow$?? x
$F(z) = \mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^T z)\}$: strongly monotone with modulus $\alpha > 0$ on $\mathcal{X}$, $G(z) = F(z) - F(x)$ $\mathbf{A}$: x is the unique weak solution to $\mathrm{VI}(G, \mathcal{X})$ $\mathbf{B}$: Observable vector fields $G_{y_k,\eta_k}(z) = \eta_k\psi(\eta_k^T z) - \eta_k y_k$ are unbiased estimates of vector field $G(z)$

Conclusion: *We can recover x via solving $\mathrm{VI}(G, \mathcal{X})$ by an algorithm capable to work with unbiased stochastic estimates of $G(\cdot)$ instead of the actual values of G.*

$\{(y_k,\eta_k)\}_{k\leq K}$ i.i.d., $\mathbf{E}_{\|\eta_k}\{y_k\}=\psi(\eta_k^T x)$, $x\in\mathcal{X}$?? $\Rightarrow$?? x
$F(z)=\mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^T z)\}$: strongly monotone with modulus $\alpha>0$ on $\mathcal{X}$, $G(z)=F(z)-F(x)$ $\mathbf{A}$: x is the unique weak solution to $\mathrm{VI}(G,\mathcal{X})$ $\mathbf{B}$: Observable vector fields $G_{y_k,\eta_k}(z)=\eta_k\psi(\eta_k^T z)-\eta_k y_k$ are unbiased estimates of vector field $G(z)$

♠ There are two basic approaches to solving "stochastic" monotone VI:

Sample Average Approximation: *Approximate the "vector field of interest" $G(x)$ by its empirical approximation*

$$G_{\omega^K}(z)=\frac{1}{K}\sum\nolimits_{k=1}^K\left[\eta_k\psi(\eta_k^T z)-\eta_k y_k\right]$$

which is monotone along with ψ, find a weak solution $\widehat{x}(\omega^K)$ to $\mathrm{VI}(G_{\omega^K},\mathcal{X})$ *and take $\widehat{x}$ as the SAA estimate of x.*

Stochastic Approximation: *Run stochastic analogy of the simplest First Order algorithm for solving deterministic monotone VI's – the Stochastic Approximation*

$$z_k=\mathrm{Proj}_{\mathcal{X}}\left[z_{k-1}-\gamma_k G_{y_k,\eta_k}(z_{k-1})\right],\ k=1,2,...,K$$

- $\mathrm{Proj}_{\mathcal{X}}[z]=\mathrm{argmin}_{y\in\mathcal{X}}\|y-z\|_2$: metric projection onto $\mathcal{X}$
- $z_0\in\mathcal{X}$ (arbitrary) deterministic starting point
- $\gamma_k>0$: deterministic stepsizes

Sample Average Approximation Estimate

$\omega^K = \{\omega_k = (y_k, \eta_k)\}_{k\le K}$ i.i.d., $\mathbf{E}_{\vert\eta_k}\{y_k\} = \psi(\eta_k^T x)$, $x \in \mathcal{X}$?? $\Rightarrow$?? x $F(z) = \mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^T z)\}$: strongly monotone with modulus $\alpha > 0$ on $\mathcal{X}$, $G(z) = F(z) - F(x)$
$\Rightarrow \qquad G_{\omega^K}(z) = \frac{1}{K}\sum_{k=1}^K \left[\eta_k\psi(\eta_k^T z) - \eta_k y_k\right] : \mathbf{E}_{\omega^K\sim P_x^K}\{G_{\omega^K}(z)\} = G(z)$
$\Rightarrow \qquad \widehat{x}_{\mathsf{SAA}}(\omega^K) \in \mathcal{X} : \langle G_{\omega^K}(z), z - \widehat{x}_{\mathsf{SAA}}(\omega^K)\rangle \ge 0\ \forall z \in \mathcal{X}$

♠ There exists rather sophisticated theoretical performance analysis of SAA recovery, resulting, under mild assumptions, in tight non-asymptotic upper bounds on the recovery error $\mathbf{E}\{\|\widehat{x}(\omega^K) - x\|_2^2\}$.

♠ Assume that the link function ψ (which we have assumed to be a continuous monotone vector field on $\mathbb{R}^m$) is the gradient field of a (automatically convex) continuously differentiable function Ψ:

$$\psi(s) = \nabla\Psi(s).$$

Note: The assumption definitely holds true when ψ is univariate, as in Logit and Probit models.

Observation: When $\psi = \nabla\Psi$, the SAA $G_{\omega^K}(z)$ is the gradient field of a continuously differentiable convex function as well:

$$G_{\omega^K}(z) = \nabla_z\left[\mathcal{G}_{\omega^K}(z) := \frac{1}{K}\sum_{k=1}^K\left[\Psi(\eta_k^T z) - z^T\eta_k y_k\right]\right]$$

$\Rightarrow$ *The SAA estimate* $\widehat{x}_{\mathsf{SAA}}(\omega^K)$ *minimizes* $\mathcal{G}_{\omega^K}(z)$ *over* $\mathcal{X}$.

Examples:

Linear model $\psi(s) \equiv s = \nabla_s \frac{s^T s}{2}$. In this case, the SAA estimate reduces to Least Squares:

$$\widehat{x}_{\mathrm{SAA}}(\omega^K) \in \mathop{\mathrm{Argmin}}_{z\in\mathcal{X}} \frac{1}{2K}\sum\nolimits_{k=1}^{K}\|y_k - \eta_k^T z\|_2^2$$

Note: For linear model with regressor- and signal-independent Gaussian noise:

$$y_k = \eta_k^T x + \xi_k,\ 1 \le k \le K$$

[noises $\xi_k \sim \mathcal{N}(0, \sigma^2 I)$ are independent of regressors and of each other]

the SAA estimate is the same as the ML one.

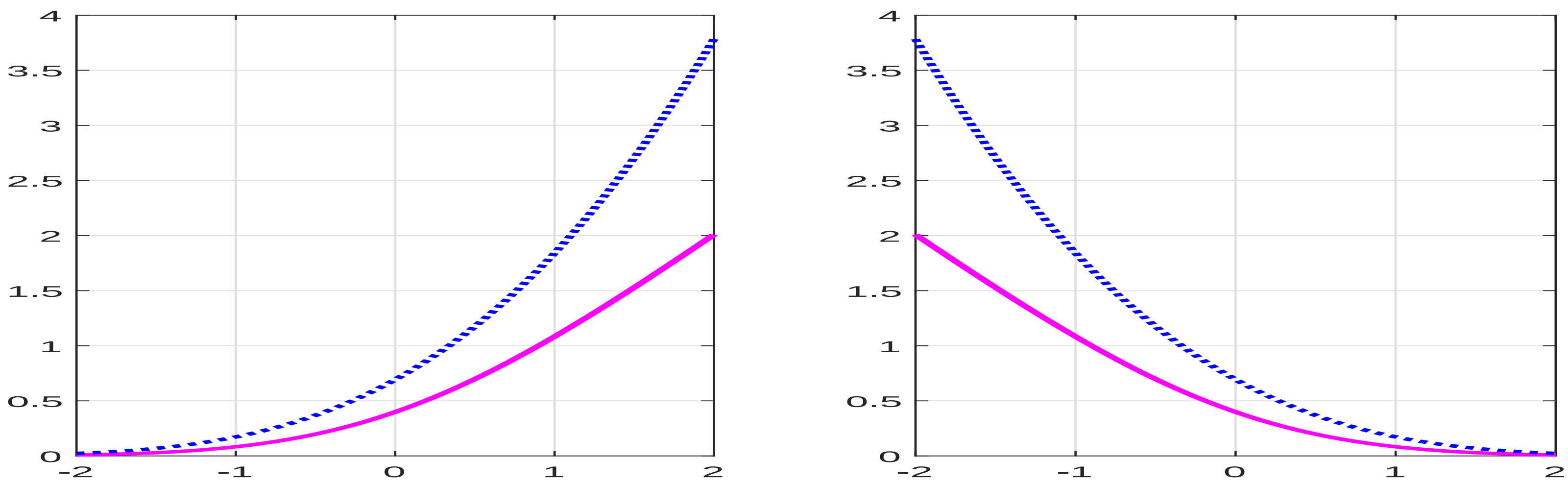
Logit model $\psi(s) = \exp\{s\}/(1+\exp\{s\})$. The SAA estimate is

$$\widehat{x}_{\mathrm{SAA}}(\omega^K) \in \underset{z\in\mathcal{X}}{\mathrm{Argmin}}\, \frac{1}{K}\sum_{k=1}^{K}\left[\ln(1+\exp\{\eta_k^T z\}) - y_k\eta_k^T z\right]$$

and *happens* to be the same as the ML estimate.

Probit model $\psi(s) = \Phi(s) = \mathrm{Prob}_{\xi\sim\mathcal{N}(0,1)}\{\xi \le s\}$. Here

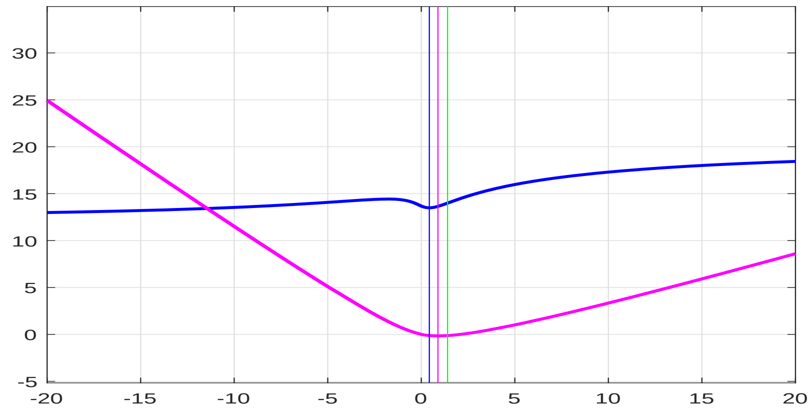
$$\begin{array}{rcl} \widehat{x}_{\mathrm{SAA}}(\omega^K) & \in & \underset{z\in\mathcal{X}}{\mathrm{Argmin}}\, \frac{1}{K}\sum_{k=1}^{K}\Big[\overbrace{(\eta_k^T z)\Phi(\eta_k^T z) + (2\pi)^{-1/2}\exp\{-(\eta_k^T z)^2/2\} - y_k\eta_k^T z}^{A_{y_k}(\eta_k^T z)}\Big] \\ \widehat{x}_{\mathrm{ML}}(\omega^K) & \in & \underset{z\in\mathcal{X}}{\mathrm{Argmin}}\, \frac{1}{K}\sum_{k=1}^{K}\Big[\underbrace{y_k \ln\left((1-\Phi(\eta_k^T z))/\Phi(\eta_k^T z)\right) - \ln(1-\Phi(\eta_k^T z))}_{B_{y_k}(\eta_k^T z)}\Big] \end{array}$$



A_0, B_0 $\qquad$ A_1, B_1

Note: In the above GLM's, finding ML estimates happened to be efficiently solvable convex problems. It is *not* so in general.

Example: $y_k = \mathrm{atan}(\eta x) + 3\xi_k$ with i.i.d. regressors $\eta_k \sim \mathcal{N}(0,1)$ and independent of regressors i.i.d. noises $\xi_k \sim \mathcal{N}(0,1)$. With $\mathcal{X} = [-20, 20]$, $K = 20$, this is what can happen:



- Magenta curve: graph of the objective Ψ_{SAA} to be minimized on $\mathcal{X}$ to get the SAA estimate
- Blue curve: graph of the objective Ψ_{ML} to be minimized on $\mathcal{X}$ to get the ML estimate
- Abscissae of vertical segments:

  — green: true signal ≈ 1.4047

  — magenta: $\widehat{x}_{\text{SAA}} \approx 0.8910$ – minimizer of Ψ_{SAA}

  — blue: *local* minimizer ≈ 0.4300 of Ψ_{ML}; the global minimizer of Ψ_{ML} on $\mathcal{X}$ is $\widehat{x}_{\text{ML}} = -20$

Note: *With one-dimensional signal, the ML estimate can be computed by "brute force." With multidimensional signal, potential nonconvexity of minus log-likelihood can result in severe computational difficulties. For the SAA estimate, computational tractability is "built in."*

Stochastic Approximation Estimate

$$z_k = \text{Proj}_{\mathcal{X}}\left[z_{k-1} - \gamma_k G_{y_k,\eta_k}(z_{k-1})\right], \; k = 1, 2, ..., K$$

- $\text{Proj}_{\mathcal{X}}[z] = \text{argmin}_{y\in\mathcal{X}} \|y - z\|_2$: metric projection onto $\mathcal{X}$
- $z_0 \in \mathcal{X}$ (arbitrary) deterministic starting point
- $\gamma_k > 0$: deterministic stepsizes

♠ The *basic* performance analysis for the SA estimate is as follows. Let us augment Assumption A with

Assumption B: *For some $M < \infty$ and for every signal $x \in \mathcal{X}$, denoting by P_x the common distribution of observations $\omega_k = (y_k, \eta_k)$, $k \leq K$, stemming from signal x, one has*

$$\mathbf{E}_{(y,\eta)\sim P_x}\left\{\|\eta y\|_2^2\right\} \leq M^2 \;\forall x \in \mathcal{X}.$$

$\{(y_k,\eta_k)\sim P_x\}_{k\leq K}$ i.i.d., $\mathbf{E}_{|\eta_k}\{y_k\}=\psi(\eta_k^T x)$, $x\in\mathcal{X}$?? $\to$?? x

A: $F(z)=\mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^T z)\}$: strongly monotone with modulus $\alpha>0$ on $\mathcal{X}$, $G(z)=F(z)-F(x)$

B: $\mathbf{E}_{(y,\eta)\sim P_x}\{\|\eta y\|_2^2\}\leq M^2\ \forall x\in\mathcal{X}$

$$G_{y,\eta}(z)=\eta\psi(\eta^T z)-\eta y$$
$$z_k=\mathrm{Proj}_{\mathcal{X}}\left[z_{k-1}-\gamma_k G_{y_k,\eta_k}(z_{k-1})\right],\ 1\leq k\leq K$$

Simple standard fact: *Under Assumptions* **A**, **B** *and with stepsizes*

$$\gamma_k=\frac{1}{(k+1)\alpha},\ 1\leq k\leq K, \qquad (*)$$

whatever be signal $x\in\mathcal{X}$ underlying observations $\omega_k=(y_k,\eta_k)$, for the SA iterates z_k it holds

$$\mathbf{E}_{\omega^k\sim P_x^k}\left\{\|z_k-x\|_2^2\right\}\leq\frac{4M^2}{(k+1)\alpha^2},\ 1\leq k\leq K \qquad (!)$$

$[P_x^k$: distribution of observation $\omega^k=(\omega_1,...,\omega_k)$, the signal being $x]$

Good news: *Typically, the $O(1/k)$-rate of convergence established in (!) is the best rate allowed by Statistics.*

Another good news: *Error bound (!) is non-asymptotic and is governed by the true modulus of strong monotonicity α of F and the true "magnitude of uncertainty" M.*

Not so good news: *To ensure (!), we need to use stepsizes $(*)$ with α lower-bounding the true modulus of strong monotonicity of F on $\mathcal{X}$. Overestimating this modulus could completely destroy (!).*

$F(z) = \mathbf{E}_{\eta\sim P}\left\{\eta\psi(\eta^T z)\right\} : \mathcal{X} \to \mathbb{R}^n \,\&\, \langle F(z) - F(z'), z - z'\rangle \geq \alpha\|z - z'\|_2^2 \,\&\, G(z) = F(z) - F(x) \quad \Rightarrow$ $\langle G(z), z - x\rangle \geq \alpha\|z - x\|^2,\ z \in \mathcal{X}$	(a)
$G_{y,\eta}(z) = \eta\psi(\eta^T y) - \eta y \,\&\, \mathbf{E}_{(y,\eta)\sim P_x}\{G_{y,\eta}(z)\} = G(z),\ z \in \mathcal{X}$	(b)
$\omega_k = (y_k, \eta_k) \sim P_x$ i.i.d. $\mathbf{E}_{\lvert\eta_k}\{y_k\} = \psi(\eta_k^T x)$	(c)
$\mathbf{E}_{(y,\eta)\sim P_w}\left\{\|\eta y\|_2^2\right\} \leq M^2,\ w \in \mathcal{X}$	(d)

Proof of Standard Fact:

• Claim: from (b)-(d) it follows that

$$\forall (x, z \in \mathcal{X}) : \|F(z)\| \leq M \;\&\; \mathbf{E}_{(y,\eta)\sim P_x}\{\|G_{y,\eta}(z)\|_2^2 \leq 4M^2 \qquad (e)$$

Indeed, denoting by P the distribution of regressors (it is independent of the signal), we have

$$\forall (x \in \mathcal{X}) : M^2 \geq \mathbf{E}_{(y,\eta)\sim P_x}\left\{\|\eta y\|_2^2\right\} = \underbrace{\mathbf{E}_{\eta\sim P}\left\{\mathbf{E}_{\lvert\eta}\{\|\eta y\|_2^2\}\right\} \geq \mathbf{E}_{\eta\sim P}\{\|\eta \mathbf{E}_{\lvert\eta}\{y\}\|_2^2\}}_{\text{Jensen's inequality}} = \mathbf{E}_{\eta\sim P}\left\{\|\eta\psi(\eta^T x)\|_2^2\right\}$$

$$\Rightarrow \left\{\begin{array}{l} \|F(z)\|_2 = \|\mathbf{E}_{\eta\sim P}\{\eta\psi(\eta^T z)\}\|_2 \leq \mathbf{E}_{\eta\sim P}\{\|\eta\psi(\eta^T z)\|_2\} \leq \sqrt{\mathbf{E}_{\eta\sim P}\{\|\eta\psi(\eta^T z)\|_2^2\}} \leq M \\ \mathbf{E}_{(y,\eta)\sim P_x}\{\|G_{y,\eta}(z)\|_2^2\} = \mathbf{E}_{(y,\eta)\sim P_x}\{\|\eta\psi(\eta^T z) - \eta y\|_2^2\} \leq 2\left[\mathbf{E}_{\eta\sim P}\{\|\eta\psi(\eta^T z)\|_2^2\} + \mathbf{E}_{(y,\eta)\sim P_x}\{\|\eta y\|_2^2\}\right] \leq 4M^2 \end{array}\right.$$

• Let us fix signal $x \in \mathcal{X}$ underlying observations $\omega_k = (y_k, x_k)$. Observe that by construction z_k is a deterministic function of $\omega^k = (\omega_1, ..., \omega_k)$: $z_k = Z_k(\omega^k)$. Setting $D_k(\omega^k) = \frac{1}{2}\|Z_k(\omega^k) - x\|_2^2$, we have

$$\begin{array}{rcl} D_k(\omega^k) & \leq & \frac{1}{2}\|[Z_{k-1}(\omega^{k-1}) - x] - \gamma_k G_{y_k,\eta_k}(Z_{k-1}(\omega^{k-1}))\|_2^2 \\ & = & D_{k-1}(\omega^{k-1}) - \gamma_k\langle G_{y_k,\eta_k}(Z_{k-1}(\omega^{k-1})), Z_{k-1}(\omega^{k-1}) - x\rangle + \frac{1}{2}\gamma_k^2\|G_{y_k,\eta_k}(Z_{k-1}(\omega^{k-1}))\|_2^2 \end{array}$$

Taking expectation and invoking (b), (a), (e) and the fact that $(y_k, \eta_k) \sim P_x$ are independent across k, we get

$$\begin{array}{rl} d_k := \mathbf{E}_{\omega^k\sim P_x^k}\left\{D_k(\omega^k)\right\} & \leq d_{k-1} - \gamma_k \mathbf{E}_{\omega^{k-1}\sim P_x^{k-1}}\left\{\langle G(Z_{k-1}(\omega^{k-1})), Z_{k-1}(\omega^{k-1}) - x\rangle\right\} + 2\gamma_k^2 M^2 \\ & \leq (1 - 2\alpha\gamma_k)d_{k-1} + 2\gamma_k^2 M^2. \end{array}$$

$$\boxed{\begin{array}{c} D_k(\omega^k) = \frac{1}{2}\|Z_k(\omega^k) - x\|_2^2,\ d_k := \mathbf{E}_{\omega^k \sim P_x^k}\left\{D_k(\omega^k)\right\} \leq (1 - 2\alpha\gamma_k)d_{k-1} + 2\gamma_k^2 M^2,\ 1 \leq k \leq K \\ \hline \gamma_k = \frac{1}{(k+1)\alpha} \end{array}} \qquad (!)$$

• Let us prove by induction in k that with $S = \frac{2M^2}{\alpha^2}$ for $k = 0, 1, ..., K$ it holds

$$d_k \leq \frac{S}{k+1} \qquad (*_k)$$

Base $k = 0$: Let D be $\|\cdot\|_2$-diameter of $\mathcal{X}$ and $z_\pm \in \mathcal{X}$ be such that $\|z_+ - z_-\|_2 = D$. Invoking (e) and strong monotonicity, with modulus α, of F on $\mathcal{X}$, we have

$$\alpha D^2 \leq \langle F(z_+) - F(z_-), z_+ - z_- \rangle \leq 2MD \Rightarrow D \leq \frac{2M}{\alpha} \Rightarrow d_0 \leq \frac{D^2}{2} \leq \frac{2M^2}{\alpha^2},$$

implying $(*_0)$.

Step $k-1 \Rightarrow k$: Assuming $k \geq 1$ and $(*_{k-1})$ true, note that $2\alpha\gamma_k = \frac{2}{k+1} \leq 1$. Invoking (!) and $(*_{k-1})$, we get

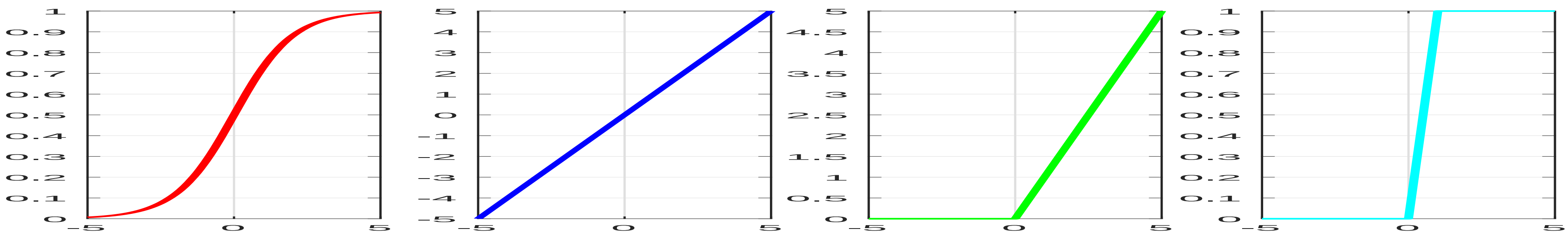
$$d_k \leq \left[1 - \frac{2}{k+1}\right]\frac{S}{k} + \frac{2M^2}{(k+1)^2\alpha^2} = S\left[\frac{1}{k}\left(1 - \frac{2}{k+1}\right) + \frac{1}{(k+1)^2}\right] = \frac{S}{k+1}\left[1 - \frac{1}{k} + \frac{1}{k+1}\right] \leq \frac{S}{k+1}.$$

Induction is complete.

• Since $d_k = \frac{1}{2}\mathbf{E}_{\omega^k \sim P_x^k}\left\{\|Z_k(\omega^k) - x\|_2^2\right\}$, $(*_k)$ reads

$$\mathbf{E}_{\omega^k \sim P_x^k}\left\{\|Z_k(\omega^k) - x\|_2^2\right\} \leq \frac{4M^2}{(k+1)\alpha^2}. \qquad \square$$

Experiment: We consider four univariate link functions:



Logit, $y \in \{0,1\}$
$\psi(s) = \frac{\exp\{s\}}{1+\exp\{s\}}$
$\mathrm{Prob}_{|\eta}\{y=1\} = \psi(\eta^T x)$

Linear, $y \in \mathbb{R}$
$\psi(s) = s$
$y \sim \mathcal{N}(\psi(\eta^T x), 1)$

Hinge, $y \in \mathbb{R}$
$\psi(s) = \max[s, 0]$
$y \sim \mathcal{N}(\psi(\eta^T x), 1)$

Ramp, $y \in \mathbb{R}$
$\psi(s) = \min[1, \max[0, s]]$
$y \sim \mathcal{N}(\psi(\eta^T x), 1)$

- In all four cases, $\mathcal{X} = \{x \in \mathbb{R}^{100} : \|x\|_2 \leq 1\}$, $\eta_k \sim \mathcal{N}(0, I_{100})$

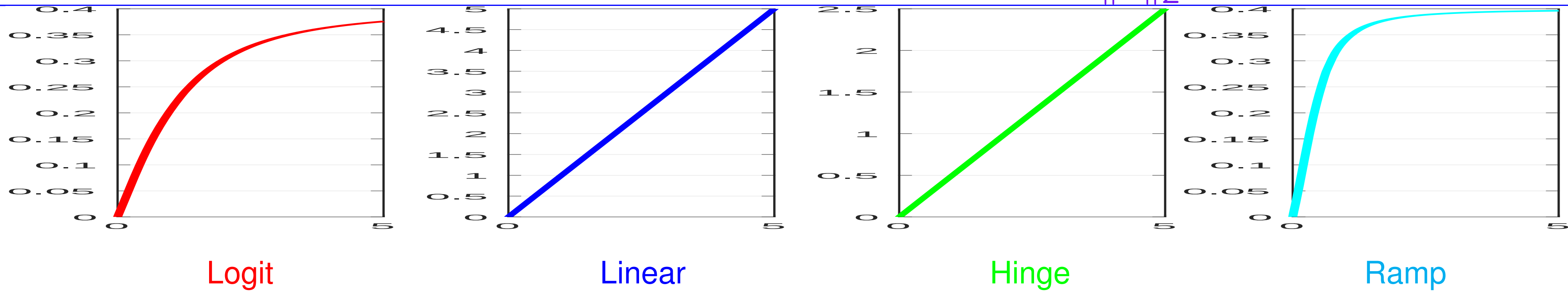
Note: *When we know in advance the common distribution P of regressors η_k, the vector field*

$$F(z) = \mathbf{E}_{\eta \sim P}\left\{\eta \psi(\eta^T z)\right\}$$

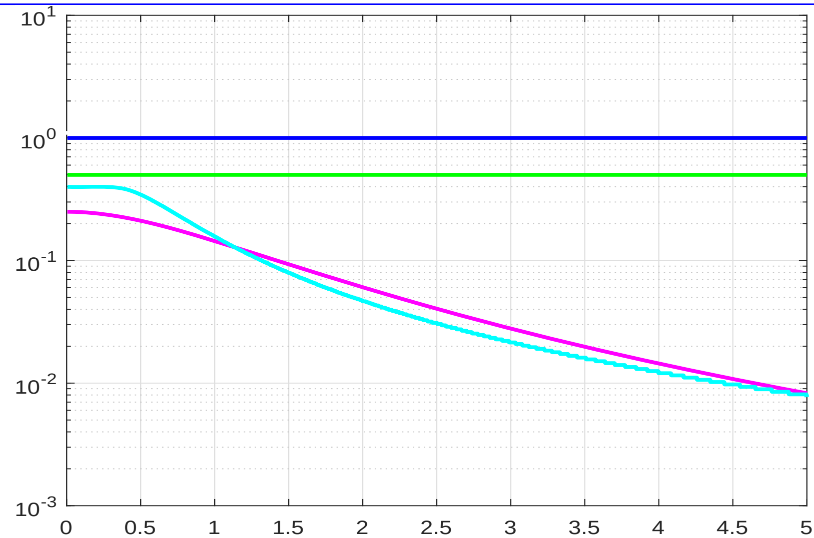
becomes known. In addition, when $P = \mathcal{N}(0, I_n)$, F becomes extremely simple:

$$F(x) = \Psi(\|x\|_2)\frac{x}{\|x\|_2},\ \Psi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} s\psi(ts)\mathrm{e}^{-s^2/2}ds$$

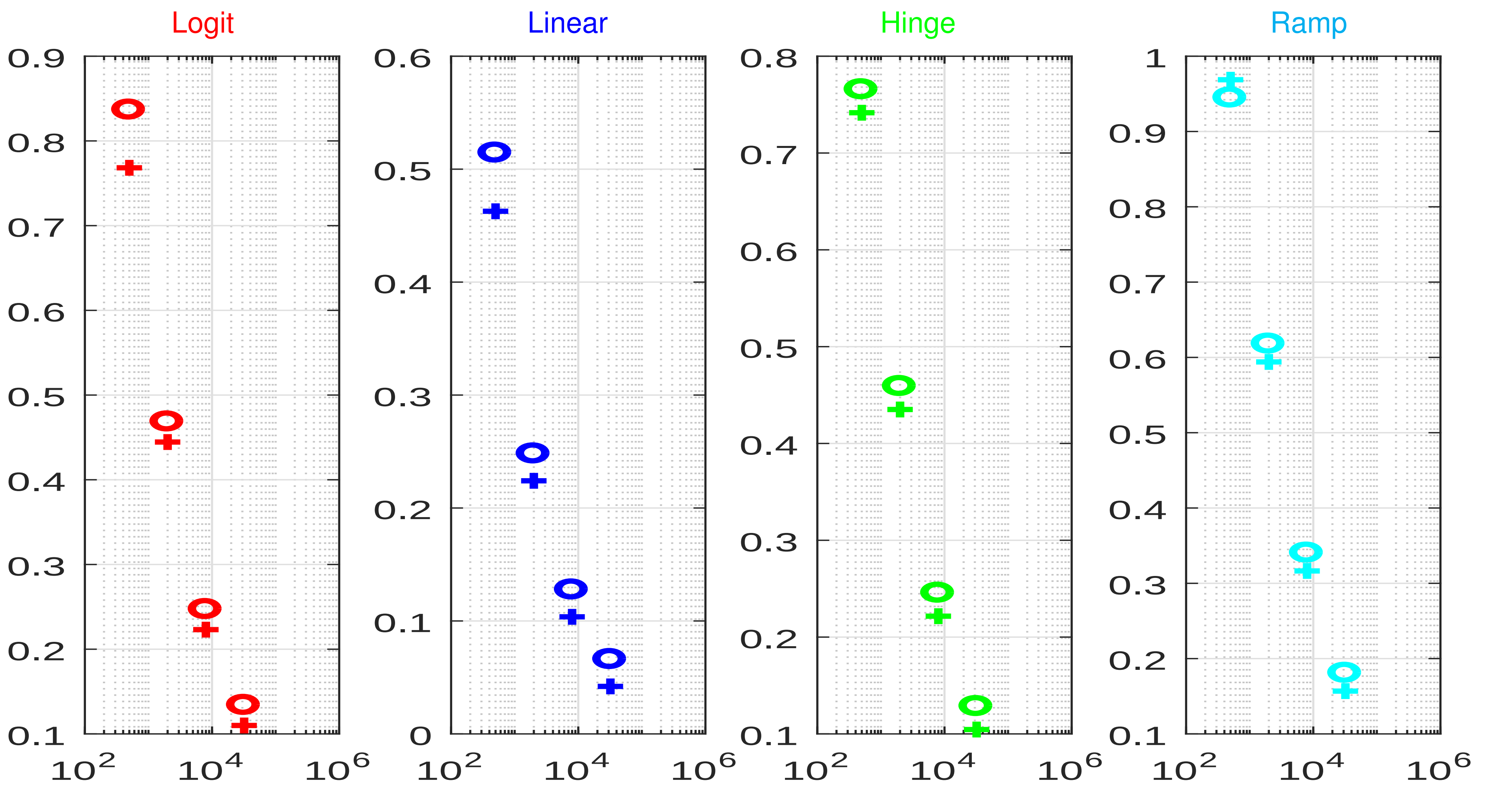
$$\eta \sim \mathcal{N}(0, I_n) \Rightarrow F(x) = \Psi(\|x\|_2)\frac{x}{\|x\|_2}$$



Functions Ψ for our four cases



Modulae of strong monotonicity of vector fields $F(\cdot)$ on $\{z : \|z\|_2 \leq R\}$ vs. R



Average $\|\cdot\|_2$-recovery errors for SA (o) and SAA (+) estimates vs $K = 500, 2000, 8000, 32000$

"Single-Observation" Case

♠ **Situation:** We observe *deterministic* sequence of regressors $\{\eta_k \in \mathbb{R}^{n\times m}\}_{k\leq K}$ and sequence of *random* labels $y^K = \{y_k \in \mathbb{R}^m\}_{k\leq K}$. The labels $y_1,...,y_K$ are independent of each other with distributions $P_{x,k}$ parameterized by unknown signal $x \in \mathcal{X} \subset \mathbb{R}^n$, and

$$\mathbf{E}_{y_k\sim P_{x,k}}\{y_k\} = \psi(\eta_k^T x),\ x \in \mathcal{X}.$$

Our goal is to recover x given $\{\eta_k\}_{k\leq K}$ and y^K.

Note: *In fact we have a single-observation GLM with deterministic regressor* η^K*, random label* y^K*, and link function* ψ^K *given by*

$$\eta^K = [\eta_1,...,\eta_K] \in \mathbb{R}^{n\times mK},\ y^K = \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix} \in \mathbb{R}^{mK},\ \psi^K([u_1;...;u_K]) = \begin{bmatrix} \psi(u_1) \\ \vdots \\ \psi(u_K) \end{bmatrix} : \mathbb{R}^{mK} \to \mathbb{R}^{mK}.$$

Indeed, we clearly have

$$\mathbf{E}_{y^K\sim P_{x,1}\times...\times P_{x,K}}\left\{y^K\right\} = \psi^K([\eta^K]^T x),\ x \in \mathcal{X}$$

$\Rightarrow$ *We can apply our machinery!*

♠ **Situation** (reworded): We are given

- a deterministic regressor matrix $\eta \in \mathbb{R}^{n\times M}$
- a convex compact signal set $\mathcal{X} \subset \mathbb{R}^n$
- a random observation ("label") $y \in \mathbb{R}^M$ with distribution P_x parameterized by signal $x \in \mathcal{X}$ in such a way that

$$\mathbf{E}_{y\sim P_x}\{y\} = \phi(\eta^T x)$$

for a given link function $\phi(\cdot) : \mathbb{R}^M \to \mathbb{R}^M$

Given y and η, we want to recover x.

Note: Under the circumstances the vector field

$$F(z) = \eta\phi(\eta^T z) : \mathbb{R}^n \to \mathbb{R}^n$$

becomes fully observable!

Assumptions:

$\mathbf{A}'$: *The vector field* $\phi(\cdot) : \mathbb{R}^M \to \mathbb{R}^M$ *is continuous and monotone, so that* $F(\cdot)$ *is continuous and monotone on* $\mathbb{R}^n$*; in addition,* F *is strongly monotone, with modulus* $\alpha > 0$*, on* $\mathcal{X}$*.*

$\mathbf{B}'$: *For some* $\sigma < \infty$ *it holds*

$$\mathbf{E}_{y\sim P_z}\left\{\|\eta[y-\phi(\eta^T z)]\|_2^2\right\} \leq \sigma^2 \;\forall z \in \mathcal{X}$$

$$\boxed{\mathbf{E}_{y\sim P_x}\{y\} = \phi(\eta^T x) \ \&\ x\in\mathcal{X} \ \&\ \mathbf{E}_{y\sim P_z}\left\{\|\eta[y-\phi(\eta^T z)]\|_2^2\right\} \le \sigma^2\,\forall z\in\mathcal{X}}$$

♠ Under the circumstances, the SAA estimate $\widehat{x}_{\textsf{SAA}}(y)$ of signal x underlying observation y is the weak solution of $\mathrm{VI}(G_y,\mathcal{X})$ with

$$G_y(z) = \eta\phi(\eta^T z) - \eta y$$

Proposition *Under Assumptions* $\mathbf{A}'$, $\mathbf{B}'$ *one has*

$$\mathbf{E}_{y\sim P_x}\left\{\|\widehat{x}_{\textsf{SAA}}(y) - x\|_2^2\right\} \le \sigma^2/\alpha^2 \ \forall x\in\mathcal{X}.$$

$$\boxed{G_y(z) = \eta\phi(\eta^T z) - \eta y: \ \alpha\text{-strongly monotone on } \mathcal{X}}$$

Proof of Proposition: Let x be the signal underlying observation, y be a realization of the observation, and let $\widehat{x} = \widehat{x}_{\mathrm{SAA}}(y)$, so that $\widehat{x}$ is a weak and therefore a strong, by $\mathbf{A}'$, solution to $\mathrm{VI}(G_y, \mathcal{X})$. It suffices to verify that

$$\|\widehat{x} - x\| \leq \alpha^{-1} \| \underbrace{\eta[y - \phi(\eta^T x)]}_{\Delta} \|_2 \qquad (!)$$

Setting $G(z) = F(z) - F(x)$, we have

$$\begin{array}{c}
G_y(z) = F(z) - \eta y = F(z) - F(x) + [F(x) - \eta y] = G(z) - \eta[y - \phi(\eta^T x)] = G(z) - \Delta;\\
\widehat{x} \text{ solves } \mathrm{VI}(G_y, \mathcal{X}) \Rightarrow 0 \leq \langle G_y(\widehat{x}), x - \widehat{x}\rangle = \langle G(\widehat{x}), x - \widehat{x}\rangle - \langle \Delta, x - \widehat{x}\rangle \Rightarrow\\
-\langle G(\widehat{x}), x - \widehat{x}\rangle \leq -\langle \Delta, x - \widehat{x}\rangle \qquad (a)\\
G(x) = 0 \Rightarrow \langle G(x), x - \widehat{x}\rangle = 0 \qquad (b)
\end{array}$$

so that

$$\alpha\|x - \widehat{x}\|_2^2 \leq \overbrace{\langle G(x) - G(\widehat{x}), x - \widehat{x}\rangle \leq -\langle \Delta, x - \widehat{x}\rangle}^{\text{by } (a),(b)} \leq \|\Delta\|_2 \|x - \widehat{x}\|_2$$
$$\Rightarrow (!) \qquad \square$$

Example: Assume that

- ϕ is continuous and strongly monotone, with modulus $\varkappa > 0$, on the entire $\mathbb{R}^M$,
- $n \times M$ regressor η is a realization of random matrix $\mathbf{H}$ with independent of each other $\mathcal{N}(0,1)$ entries,
- $y = \phi(\eta^T x) + \xi$, where $\xi \sim \mathcal{N}(0, \lambda^2 I_M)$ is independent of η,
- $M \gg n$.

In this case, with probability rapidly approaching 1 as $M \to \infty$,

— $F(z) = \eta\phi(\eta^T z)$ is strongly monotone, with modulus $\alpha = O(1)\varkappa M$, on $\mathbb{R}^n$,

— $\mathbf{E}_{y\sim P_x}\left\{\|\eta[y-\phi(\eta^T x)]\|_2^2\right\} = \mathbf{E}_{\xi\sim\mathcal{N}(0,\lambda^2 I_M)}\left\{\|\eta\xi\|_2^2\right\} \leq \sigma^2 := O(1)\lambda^2 M n$

$\Rightarrow$ *Modulo rapidly going to 0 as $M \geq O(1)n$ grows probability of getting "pathological" η, we have*

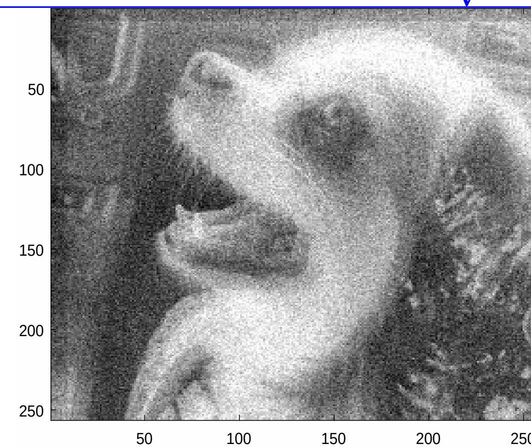
$$\mathbf{E}\left\{\|\widehat{x}_{\mathrm{SAA}}(y) - x\|_2^2\right\} \leq \frac{\sigma^2}{\alpha^2} \leq O(1)\frac{\lambda^2 n}{\varkappa^2 M}.$$

Illustration: Image reconstruction from blurred noisy observation

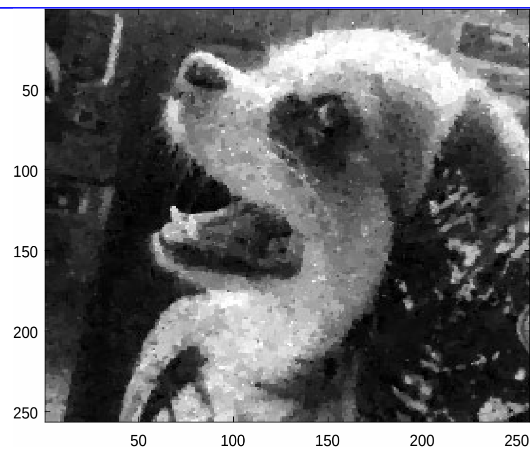
$$y = [\varkappa \star x]^{1/2} + \sigma\xi$$

$\varkappa$:	nonnegative 2D kernel, $\|\varkappa_1\|_1 = 1$	$\star$:	2D convolution
x:	2D image to be recovered	$[\cdot]^{1/2}$:	entrywise square root
ξ:	white Gaussian noise	σ:	$1.2 \approx 0.075\sqrt{\|x\|_\infty}$

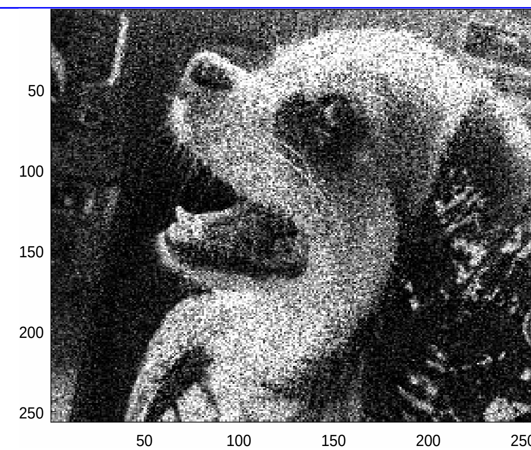
True image



Observation y



SAA recovery
$\mathcal{X}$: nonnegative part of TV ball



SAA recovery
$\mathcal{X}$: nonnegative orthant

Illustration: Tale of Two Retailers

♣ **Tale:** *There are two competing retailers, U and V, selling red herrings.*

• *A retailer creates "selling capacity" $z \in \mathbb{R}_+$ (e.g., rents some areas, summing up to z, in several stores).*

• *Denoting by u and v the selling capacities of U and V, the daily expected losses (minus profits) of the retailers are*

$$U(u,v) = pu - \frac{u}{u+v+c}D, \; V(u,v) = qv - \frac{v}{u+v+c}D,$$

• D: money volume of total daily demand • $c > 0$: total selling capacity of other retailers

• p, q: daily expences to support unit selling capacity for U and for V

Rationale: we assume that the actual demand D is split between U, V and other retailers proportionally to their selling capacities.

• *We assume that the actual capacities $(u_*, v_*) \in \mathbb{R}^2_+$ form Nash Equilibrium, meaning that*

— when V selects capacity v_*, U has no incentive to deviate from selection u_*:

$$U(u, v_*) \geq U(u_*, v_*) \, \forall u \in \mathbb{R}_+$$

— when U selects capacity u_*, V has no incentive to deviate from selection v_*:

$$V(u_*, v) \geq V(u_*, v_*) \, \forall v \in \mathbb{R}_+$$

♠ **Goal:** Given in advance

— D, c, and closed, convex and bounded set $\mathcal{X}$ known to contain "parameter of interest" $\beta := [p; q]$

— K i.i.d. unbiased observations y_k, $1 \leq k \leq K$, of (u_*, v_*)

we want to recover β.

Note: Observation noise can come, e.g., from the fact that the selling capacities of U and V are distributed among many locations, and we measure the capacities in K locations selected at random from the uniform distribution.

Executive Summary on Convex Nash Equilibria

♠ **Situation:** There are m players, i-th selecting $x_i \in X_i \neq \emptyset$.

• Losses of players are known functions $f_i(x_1, ..., x_m)$ of the vector $x = [x_1; ...; x_m] \in \mathcal{X} := X_1 \times ... \times X_m$ of their selections.

• *Nash equilibria* are points $x^* \in \mathcal{X}$ such that no one of the players has incentive to replace his choice with another one, provided that the remaining players stick to their choices. In other words, $x^* \in \mathcal{X}$ is a Nash equilibrium iff

$$\forall (i, x_i \in X_i) : f_i(x_1^*, ..., x_{i-1}^*, x_i, x_{i+1}^*, ..., x_m^*) \geq f_i(x^*).$$

♠ Nash equilibrium problem is called *convex*, if

- all X_i are nonempty closed convex sets
- for every i, $f_i(x)$ is convex in x_i and jointly concave in the collection $\{x_j : j \neq i\}$ of all remaining x_j's
- $\sum_i f_i(x)$ is convex

Example: The standard convex-concave saddle point problem

$$\min_{u \in U} \max_{v \in V} \phi(u, v)$$

on closed convex domains U, V can be thought of as Nash equilibrium problem with loss $\phi(u, v)$ of the player selecting u and loss $-\phi(u, v)$ of the player selecting v.

Fact: Consider convex Nash Equilibrium problem with continuously differentiable losses $f_i(x)$ and let us associate with it the vector field

$$F(x) = \left[\frac{\partial}{\partial x_1} f_1(x); \frac{\partial}{\partial x_2} f_2(x); ...; \frac{\partial}{\partial x_m} f_m(x)\right] : \mathcal{X} \to \mathbb{R}^m.$$

This vector field is monotone, and the weak (or, which is the same since F is continuous, strong) solutions to $\mathrm{VI}(F, \mathcal{X})$ are exactly the Nash equilibria.

Fact: *When $c > 0$, the function $\frac{s}{s+t+c} = 1 - \frac{t+c}{s+t+c}$ of nonnegative s, t is concave in s and convex in t*

$\Rightarrow$ In Tale of Two Retailers, losses of players U, V

$$U(u,v) = pu - \frac{u}{u+v+c}D,\ V(u,v) = qv - \frac{v}{u+v+c}D$$

are *convex* in the choices of the players and *concave* in the choices of their adversaries, while the sum of these losses

$$pu + qv - D\frac{u+v}{u+v+c}$$

is convex in u, v

$\Rightarrow$ *Nash equilibrium in Tale is weak≡strong solution to $\mathrm{VI}(G_\beta, \mathbb{R}^2_+)$ with monotone (in fact, strongly monotone) on $\mathbb{R}^2_+$ operator*

$$G_\beta(u,v) = \underbrace{\left[-\frac{v+c}{(u+v+c)^2}D; -\frac{u+c}{(u+v+c)^2}D\right]}_{G(u,v)} + \beta. \qquad [\beta = [p;q]]$$

Note: Field G is *not* potential – this is not the gradient field of a function!

$$G(u,v)=\left[-\frac{v+c}{(u+v+c)^2}D;-\frac{u+c}{(u+v+c)^2}D\right]:\mathbb{R}^2_+\to\mathbb{R}^2$$

Fact: The strongly anti-monotone vector field $-G$ is one-to one smooth mapping of $\mathbb{R}^2_+$ onto the domain

$$\Pi=\{[p;q]:0<p\leq\theta,p^2/\theta\leq q\leq\sqrt{\theta p}\}\qquad[\theta=D/c]$$

with smooth anti-monotone inverse mapping $\phi(p,q)$ given by explicit formula:

$$[p;q]\in\Pi,\ \phi(p,q)=\left[\begin{array}{c}\frac{cq}{p+q}\left[1+\frac{\theta}{2(p+q)}+\sqrt{\frac{\theta^2}{4(p+q)^2}+\frac{\theta}{p+q}}\right]-c\\ \frac{cp}{p+q}\left[1+\frac{\theta}{2(p+q)}+\sqrt{\frac{\theta^2}{4(p+q)^2}+\frac{\theta}{p+q}}\right]-c\end{array}\right]$$
$$\Rightarrow\phi(p,q)\in\mathbb{R}^2_+\ \&\ [p;q]+G(\phi(p,q))=0.$$

In words: *For $[p;q]\in\Pi$, $\phi(p,q)$ is the vector of selections of U and V, the cost coefficients for supporting capacities being p for U and q for V.*

Bottom line: In Tale of Two Retailers, given compact convex subset $\mathcal{X}\subset\Pi$ known to contain the vector $\beta=[p;q]$ of parameters to be recovered, identifying p, q reduces to recovering signal $\beta\in\mathcal{X}$ in GLM where

- the link function is the *monotone* vector field $\overline{\phi}\equiv-\phi:\Pi\to\mathbb{R}^2$
- the regressors η_k, $k\leq K$, are the unit 2×2 matrices
- the labels are $-y_k\in\mathbb{R}^2$, where y_k are i.i.d. unbiased observations of $[u;v]=\phi(\beta)$.

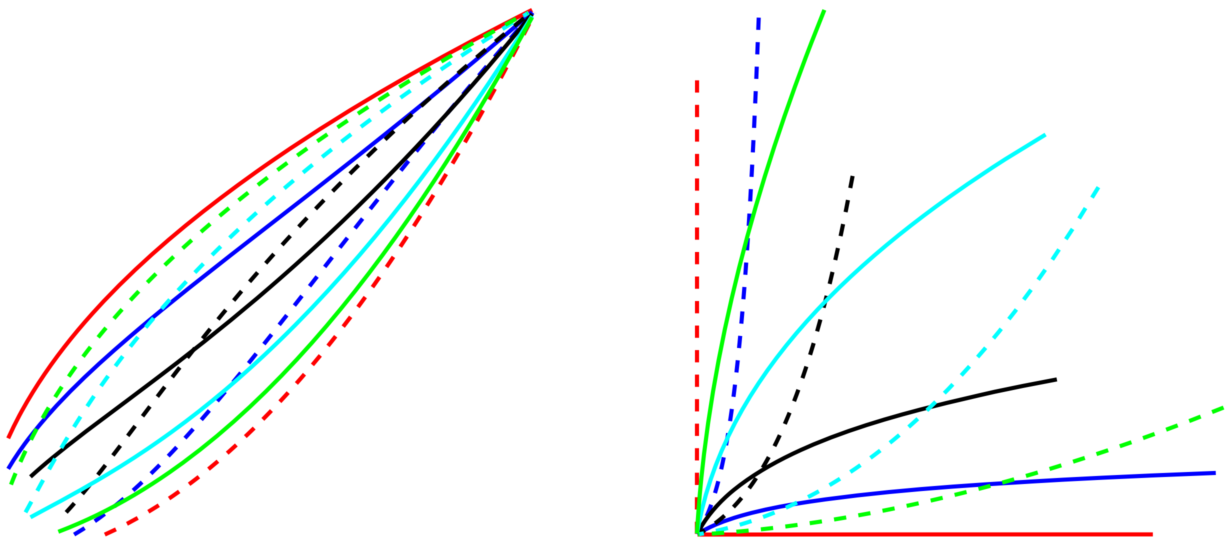
How It Works

♠ **Setup:** $D = 100$, $c = 1$

• Selling capacities of U and V are (randomly) distributed over $n = 400$ locations and are observed at $K = 40$ randomly selected locations.

• **Relative recovery errors**, data over 1000 simulations:

error	mean	median	max
$\|\beta - \widehat{\beta}\|_2 / \|\beta\|_2$	0.073	0.063	0.314



several curves in Π (left) and their ϕ-images in $\mathbb{R}_2^+$ (right)

Note: Similar Tale can be told about *any* number M of retailers.

Variation: Multi-State Spatio-Temporal Processes

[Ongoing joint research with Anatoli Juditsky, Yao Xie, and Liyan Xie, arXiv:2003.12935]

♠ Motivation: *Discrete time modeling of interconnected self-exciting processes*

• A realization of inhomogeneous *Poisson process* is an increasing sequence of positive reals $t_1 < t_2 < ...$ interpreted as times at which certain events (e.g., earthquakes or calls to a service center) happen. The process is characterized by *intensity function* $\lambda(t) \geq 0$, namely, as follows:

• What happens in time window $[t, t+h]$ is independent of what happened prior to time t, and in this window, the probability for happening

— *exactly one event* is $\lambda(t)h + \overline{o}(h)$

— *no event* is $1 - \lambda(t)h + \overline{o}(h)$

— *more than one event* is $\overline{o}(h)$.

In many respects we can think about Poisson process as about the limit, as $h \to +0$, of discrete time processes with realizations which are random sequences $\{\xi_i \in \{0,1\}, i \geq 1\}$ with independent entries ξ_i and probability of $\xi_i = 1$ equal to $\lambda(ih)h$. These discrete time processes are, basically, what we get, for small h, from realizations of Poisson process when splitting the time domain $t \geq 0$ into consecutive segments Δ_i of duration h and setting $\xi_i = 0$ or $\xi_i = 1$ depending on whether in a realization there were no, or there were, events in "time cell" Δ_i.

• A Hawkes, or *self-exciting*, process, can *informally* be thought of as a generalization of Poisson process where the intensity $\lambda(t)$ (which in Poisson process is deterministic function of t) becomes random, and an event at time τ increases $\lambda(t)$ for $t \geq \tau$ by some $\mu(t-\tau)$.

♠ What follows is motivated by the desire to get a simple "computation-friendly" discrete time model of a self-exciting process by splitting continuous time into short consecutive windows ("cells') and neglecting the chances for more than one event to occur in a cell.

• In addition, we consider several interacting processes of this type.

♠ Consider situation as follows:

• There are K locations. At time instant t (time is discrete!) location k can be at one of $M+1$ states, enumerated $0,1,...,M$; $\omega_{tk}\in\{0,1,...,M\}$ stands for the state of location k at time t. We call state 0 the *ground state*, and states $p\geq 1$ – *events* [of type] p

• Locations influence each other: location ℓ at state q at time τ contributes to the probability of event p in location k at time $t>\tau$.

We assume that *the conditional on the "history of the process" prior to time t (i.e., on the array $\omega^{t-1}=\{\omega_{\tau k}:\tau\leq t-1, 1\leq k\leq K\}$) probability $\pi_{tk}[p|\omega^{t-1}]$ of event p at location k at time t is*

$$\pi_{tk}[p|\omega^{t-1}]:=\mathrm{Prob}_{|\omega^{t-1}}\{\omega_{tk}=p\}=\beta_{kp}+\sum_{s\geq 1}\sum_{\ell\leq K}\beta^s_{k\ell}(p,\omega_{t-s,\ell})$$

• "birthrate" β_{kp}: component of $\pi_{tk}[p|\omega^{t-1}]$ independent of the history

• $\beta^s_{k\ell}(p,q)$: contribution of the event "location ℓ at time $t-s$ was in state q" to the (conditional on the history) probability of event p at location k at time t.

Clearly, the conditional on ω^{t-1} probability of ground state at time t at location k is $1-\sum_{p=1}^{M}\pi_{tk}[p|\omega^{t-1}]$

♠ We observe the process on time horizon $t\leq N$, and our goal is to recover from our observation ω^N the collection $\beta=\{\beta_{kp},\beta^s_{k\ell}(p,q)\}$ of parameters of our process.

$$\pi_{tk}[p|\omega^{t-1}] := \mathrm{Prob}_{|\omega^{t-1}}\{\omega_{tk} = p\} = \beta_{kp} + \sum_{s\geq 1}\sum_{\ell\leq K} \beta^s_{k\ell}(p, \omega_{t-s,\ell})$$

♠ We assume once for ever that *the process has finite memory:* $\beta^s_{k\ell}(p,q) = 0$ *whenever* $s > d$, where $d \geq 1$ is some known "memory depth."

$\Rightarrow$ *What matters as far as the behavior of the process on time horizon* $t = 1, 2, ..., N$ *is concerned, is the array* $\{\omega_{\tau k} : -d+1 \leq \tau \leq N, 1 \leq k \leq K\}$.

♡ From now on we slightly modify our previous notation and set

$$\begin{array}{c}\omega^t_\tau = \{\omega_{rk} : \tau \leq r \leq t, 1 \leq k \leq K\},\\ \omega^t = \omega^t_{-d+1} = \{\omega_{rk} : -d+1 \leq r \leq t, 1 \leq k \leq K\},\\ \beta = \{\beta_{kp}, \beta^s_{k\ell}(p,q) : 1 \leq k, \ell \leq K, 1 \leq s \leq d, 1 \leq p \leq M, 0 \leq q \leq M\}\end{array}$$

Assigning components of β serial numbers, we treat β as a column vector, and set $\nu = \dim \beta$.

♠ It is convenient to encode the collection of states of locations k, $1 \leq k \leq K$, at time t by KM-dimensional block vector $\overline{\omega}_t$, with K blocks of dimension M each. Vector $\overline{\omega}_t$ is defined as follows:

— when at time t in location k event p takes place, the k-th block in $\overline{\omega}_t$ is the p-th basic orth in $\mathbb{R}^M$

— when at time t location k is in the ground state 0, the k-th block in $\overline{\omega}_t$ is zero.

For example, with $K = 3$ and $M = 2$,

$$\overline{\omega}_t = [0; 1; 1; 0; 0; 0]$$

encodes the fact that at time t

— at location 1, event 2 takes place — $[0; 1]$ is the second basic orth in $\mathbb{R}^M = \mathbb{R}^2$

— at location 2, event 1 takes place — $[1; 0]$ is the first basic orth in $\mathbb{R}^M = \mathbb{R}^2$

— location 3 is in the ground state 0 — $[0; 0]$ is the zero in $\mathbb{R}^M = \mathbb{R}^2$

• Note that *not every Boolean KM-dimensional vector $\overline{\omega}$ can encode observed states of locations at time t*; to be "legitimate," every one of M-dimensional blocks in $\overline{\omega}$ must have *at most one* nonzero entry.

• Our model says that *the conditional, given* $\omega^{t-1}=\omega^{t-1}_{-d+1}$*, probability* $\pi_{tk}[p|\omega^{t-1}]$ *of event* p *at time* t *at location* k *is*

$$\pi_{tk}[p|\omega^{t-1}] := \mathrm{Prob}_{|\omega^{t-1}}\{\omega_{tk}=p\} = \beta_{kp} + \sum_{s=1}^{d}\sum_{\ell=1}^{K}\beta^{s}_{k\ell}(p,\omega_{t-s,\ell})$$

This is the same as to say that

The conditional, given ω^{t-1}*, expectation of the Boolean vector* $\overline{\omega}_t$ *is the* KM*-dimensional vector with entries* $\pi_{tk}[p|\omega^{t-1}]$, $1\leq k\leq K, 1\leq p\leq M$.

♠ We arrive at the model where

— our observation at time t is the vector $\overline{\omega}_t\in\mathbb{R}^{KM}$; this vector is Boolean, with at most one entry equal to 1 in every one of the K blocks of dimension M comprising $\overline{\omega}_t$

— we have $\mathbf{E}_{|\omega^{t-1}}\{\overline{\omega}_t\}=\eta^T(\omega^{t-1}_{t-d})\beta$ for readily given functions $\eta(\cdot)$ defined on the set $\Omega_{dKM}=\{\omega_{sk}\in\{0,1,...,M\}:1\leq k\leq K, 1\leq s\leq d\}$ and taking values in the space of $\nu\times KM$-matrices.

Note: Our model is close to the GLM model with identity link function, regressors $\eta(\omega^{t-1}_{t-d})$, and labels $y_t=\overline{\omega}_t$, the difference being in inter-dependence and non-stationarity of the regressors.

$\Rightarrow$ *We can try to recover* β *by the techniques we have developed for GLM's.*

Note: Inter-dependence of regressors makes it difficult to use SA, but the SAA approach still can be tried!

• our observation at time t is the vector $\overline{\omega}_t \in \mathbb{R}^{KM}$; this vector is Boolean, with at most one entry equal to 1 in every one of the K blocks of dimension M comprising $\overline{\omega}_t$

• we have $\mathbf{E}_{|\omega^{t-1}}\{\overline{\omega}_t\} = \eta^T(\omega_{t-d}^{t-1})\beta$ for readily given functions $\eta(\cdot)$ defined on the set of arrays $\{\omega_{sk} \in \{0,1,...,M\} : 1 \leq k \leq K, 1 \leq s \leq d\}$ and taking values in the space of $\nu \times KM$-matrices.

♠ **Assumption:** We are given a convex compact set $\mathcal{X} \subset \mathbb{R}^\nu$ which contains the vector β of parameters of the observed process and is such that

For every $x \in \mathcal{X}$ and every $\omega_{t-d}^{t-1} \in \Omega_{dKM}$ M-dimensional blocks in the KM-dimensional vector $\eta^T(\omega_{t-d}^{t-1})x$ are nonnegative with sum of entries ≤ 1:

$$\forall x \in \mathcal{X} : \left\{ \begin{array}{ll} x_{kp} + \sum_{s=1}^d \sum_{\ell=1}^K \min_{0\leq q\leq M} x_{k\ell}^s(p,q) \geq 0 \ \forall (1 \leq p \leq M, 1 \leq k \leq K) & (a) \\ \sum_{p=1}^M \left[x_{kp} + \sum_{s=1}^d \sum_{\ell=1}^K \max_{0\leq q\leq M} x_{k\ell}^s(p,q)\right] \leq 1 \ \forall (1 \leq k \leq K) & (b) \end{array}\right.$$

Motivation: p-th entry in an M-dimensional block, associated with location k, of $\eta^T(\omega_{t-d}^{t-1})\beta$ is conditional, ω^{t-1} given, probability for event p to take place in this location at time t $\Rightarrow$ *these entries must be nonnegative, and their sum over $p = 1,...,M$ should be ≤ 1.*

$\Rightarrow$ We lose nothing when restricting our attention with candidate parameter vectors x for which blocks in $\eta^T(\omega_{t-d}^{t-1})x$, for all $\omega_{t-d}^{t-1} \in \Omega_{dKM}$, are nonnegative with the sum of entries ≤ 1.

$$\mathbf{E}_{|\omega^{t-1}}\{\overline{\omega}_t\} = \eta^T(\omega_{t-d}^{t-1})\beta$$
$$\beta \in \mathcal{X}$$

♠ According to our methodology, the SAA recovery $\widehat{\beta}$ of β from observations ω^N is a solution to the variational inequality

$$\text{find } z_* \in \mathcal{X} : \langle G_{\omega^N}(z), z - z_* \rangle \geq 0 \; \forall z \in \mathcal{X} \qquad \mathrm{VI}(G_{\omega^N}, \mathcal{X})$$

given by $\mathcal{X}$ and the *affine monotone* vector field

$$G_{\omega^N}(x) = \frac{1}{N}\sum_{t=1}^{N}\left[\eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})x - \eta(\omega_{t-d}^{t-1})\overline{\omega}_t\right].$$

Note: $G_{\omega^N}(\cdot)$ is the gradient field of the quadratic function:

$$G_{\omega^N}(x) = \nabla_x \Phi_{\omega^N}(x),\; \Phi_{\omega^N}(x) := \frac{1}{2N}\sum_{t=1}^{N}\|\eta^T(\omega_{t-d}^{t-1})x - \overline{\omega}_t\|_2^2$$

⇒ *Our estimate $\widehat{\beta}$ is nothing but the Least Squares estimate:*

$$\widehat{\beta} = \widehat{\beta}_{\mathrm{LS}}(\omega^N) \in \operatorname*{Argmin}_{x\in\mathcal{X}} \Phi_{\omega^N}(x). \qquad (LS)$$

$\boxed{G_{\omega^N}(x) = \frac{1}{N}\sum_{t=1}^N \left[\eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})x - \eta(\omega_{t-d}^{t-1})\overline{\omega}_t\right] \;\; \widehat{\beta}: \text{ solution to VI}(G_{\omega^N}, \mathcal{X})}$

Towards Performance Analysis

♠ **Observation:** Consider, along with the observable vector field $G_{\omega^N}(\cdot)$, the *un*observable vector field

$$\overline{G}_{\omega^N}(x) = \frac{1}{N}\left[\sum_{t=1}^N \eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})x - \eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})\beta\right]$$

Note: $G_{\omega^N}(x) - \overline{G}_{\omega^N}(x)$ *is independent of* x *and* $\overline{G}_{\omega^N}(\beta) = 0$

$$\Rightarrow \; G_{\omega^N}(\beta) = G_{\omega^N}(\beta) - \overline{G}_{\omega^N}(\beta) = \frac{1}{N}\sum_{t=1}^N \underbrace{\eta(\omega_{t-d}^{t-1})\overbrace{\left[\eta^T(\omega_{t-d}^{t-1})\beta - \overline{\omega}_t\right]}^{\zeta_t}}_{\xi_t}$$

$$G_{\omega^N}(\beta) = \frac{1}{N}\sum_{t=1}^{N} \underbrace{\eta(\omega_{t-d}^{t-1})\overbrace{\left[\eta^T(\omega_{t-d}^{t-1})\beta - \overline{\omega}_t\right]}^{\zeta_t}}_{\xi_t}$$

Fact: Denoting by $\mathbf{E}_{|\omega^s}$ the conditional, ω^s being fixed, expectation, we have

$$\mathbf{E}_{|\omega^{t-1}}\{\xi_t\} = \eta(\omega_{t-d}^{t-1})\mathbf{E}_{|\omega^{t-1}}\{\zeta_t\} = 0$$

Indeed, $\mathbf{E}_{|\omega^{t-1}}\{\overline{\omega}_t\} = \eta^T(\omega_{t-d}^{t-1})\beta$.
Fact: $\|\zeta_t\|_\infty \leq 1$.
Indeed, the entries in $\eta^T(\omega_{t-d}^{t-1})\beta$ are probabilities, and the entries in $\overline{\omega}_t$ are zeros and ones.
Fact: It is easily seen that $\eta(\omega_{d-1}^{t-1})$ is Boolean matrix with at most one nonzero in every row
$\Rightarrow \|\xi_t\|_\infty \leq \|\zeta_t\|_\infty \leq 1$.
Corollary: *Typical value of $\|G_{\omega^N}(\beta)\|_\infty$ is of order of $1/\sqrt{N}$:*

$$\mathrm{Prob}\{\|G_{\omega^N}(\beta)\|_\infty > \gamma/\sqrt{N}\} \leq 2\nu \exp\{-\gamma^2/2\}\ \forall \gamma \geq 0.$$

Claim: $\text{Prob}\{\|G_{\omega^N}(\beta)\|_\infty > \gamma/\sqrt{N}\} \leq 2\nu\exp\{-\gamma^2/2\}\ \forall\gamma\geq 0$.Indeed, let us fix $i\leq\nu$. Given $\alpha\geq 0$, let us prove by induction in t that

$$\mathbf{E}_{\omega^t|\omega^0}\left\{\exp\{\sum\nolimits_{s=1}^t \alpha[\xi_t]_i\}\right\} \leq \exp\{\alpha^2 t/2\} \qquad (I_t)$$

<u>Base</u> $t=0$ is evident.
<u>Step</u> $t\mapsto t+1$: assuming (I_t) takes place, we have

$$\begin{array}{l}\mathbf{E}_{\omega^{t+1}|\omega^0}\left\{\sum_{s=1}^{t+1}\alpha[\xi_t]_i\right\} = \mathbf{E}_{\omega^t|\omega^0}\left\{\left[\sum_{s=1}^t\alpha[\xi_t]_i\right]\mathbf{E}_{|\omega^t}\left\{\exp\{\alpha[\xi_{t+1}]_i\}\right\}\right\}\\ \underbrace{\leq}_{(a)} \mathbf{E}_{\omega^t|\omega^0}\left\{\left[\sum_{s=1}^t\alpha[\xi_t]_i\right]\exp\{\alpha^2/2\}\right\}\underbrace{\leq}_{(b)}\exp\{\alpha^2(t+1)/2\}\end{array}$$

• (b) is given by (I_t)
• (a) is given by the following **Well known fact:** *Let ζ be zero mean random variable taking values in $[-\alpha,\alpha]$. Then $\mathbf{E}\{\exp\{\zeta\}\}\leq\exp\{\alpha^2/2\}$.*
Note: The conditional, ω^t given, distribution of $\alpha[\xi_{t+1}]_i$ is zero mean and is supported on $[-\alpha,\alpha]$, and thus obeys the premise of the Well known fact.
• $(I_t)\Rightarrow$ **Claim**: By (I_N) we have for d
$Delta\geq 0$ and $\alpha\geq 0$:

$$\text{Prob}\{\frac{1}{N}\sum\nolimits_{t=1}^N[\xi_t]_i>\Delta\}\leq\mathbf{E}\left\{\exp\{\frac{1}{N}\sum\nolimits_{t=1}^N\alpha[\xi_t]_i\}\right\}\exp\{-\alpha\Delta\}\leq\exp\{\frac{\alpha^2}{2N}-\alpha\Delta\}$$

$\Rightarrow$ [optimizing in α] $\text{Prob}\{\frac{1}{N}\sum_{t=1}^N[\xi_t]_i>\Delta\}\leq\exp\{-N\Delta^2/2\}$
$\Rightarrow$ $\text{Prob}\{\frac{1}{N}\sum_{t=1}^N[\xi_t]_i>\gamma/\sqrt{N}\}\leq\exp\{-\gamma^2/2\}$
Applying the same reasoning to $-\xi_t$ in the role of ξ_t, we get $\text{Prob}\{\frac{1}{N}\sum_{t=1}^N[\xi_t]_i<-\gamma/\sqrt{N}\}\leq\exp\{-\gamma^2/2\}$, and Claim follows from the union bound.

Proof of Well known fact: Let ζ be zero mean random variable supported on $[-\alpha,\alpha]$. For every γ we have

$$\mathbf{E}\{e^{\zeta}\} = \mathbf{E}\{e^{\zeta} - \gamma\zeta\} \leq \max_{-\alpha\leq s\leq\alpha}[e^s - \gamma s] = \max\left[e^{\alpha} - \gamma\alpha, e^{-\alpha} + \gamma\alpha\right]$$

where the concluding equality is due to the convexity of $e^s - \gamma s$ in s.
Setting $\gamma = \frac{\exp\{\alpha\}-\exp\{-\alpha\}}{2\alpha}$ we get

$$\mathbf{E}\{e^{\zeta}\} \leq \cosh(\alpha) \leq \exp\{\alpha^2/2\},$$

(to arrive at the concluding inequality, compare coefficients of the power series for $\cosh(s)$ and $\exp\{s^2/2\}$ and note that $\frac{1}{(2k)!} \leq \frac{1}{2^k k!}$, $k = 1, 2, ...$). $\square$

$$\boxed{G_{\omega^N}(x) = \underbrace{\frac{1}{N}\sum_{t=1}^{N}\eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})}_{A[\omega^N]}\,x - \underbrace{\frac{1}{N}\sum_{t=1}^{N}\eta(\omega_{t-d}^{t-1})\overline{\omega}_t}_{a[\omega^N]}}$$

Fact: *Typical value of $\|G_{\omega^N}(\beta)\|_\infty$ is of order of $1/\sqrt{N}$:*

$$\mathrm{Prob}\{\|G_{\omega^N}(\beta)\|_\infty > \gamma/\sqrt{N}\} \leq 2\nu\exp\{-\gamma^2/2\}\ \forall \gamma \geq 0.$$

♠ We can use Fact to design *online* upper bound on the recovering error.

• Given $\nu\times\nu$ matrix $A \succeq 0$, let us set

$$\vartheta_p[A] = \max\left\{s : x^TAx \geq s\|x\|_p^2\right\} \qquad [1 \leq p \leq \infty]$$

For example, $\vartheta_2[A]$ is the minimal eigenvalue of A.

• **Observation:** $A \succeq 0 \Rightarrow x^TAx \geq \frac{1}{2}\left[\vartheta_p[A]\|x\|_p^2 + \vartheta_r[A]\|x\|_r^2\right] \geq \sqrt{\vartheta_p[A]\vartheta_r[A]}\|x\|_p\|x\|_r.$

• **Fact:** *The Least Squares recovery $\widehat{\beta} = \widehat{\beta}(\omega^N)$ satisfies the bound*

$$\|\widehat{\beta}(\omega^N) - \beta\|_p \leq \|G_{\omega^N}(\beta)\|_\infty/\sqrt{\vartheta_1[A[\omega^N]]\vartheta_p[A[\omega^N]]}.$$

As a result, the recovery error admits online probabilistic bound: for every $\epsilon \in (0,1)$ one has

$$\mathrm{Prob}\left\{\|\widehat{\beta} - \beta\|_p \leq \frac{\sqrt{2\ln(2\nu/\epsilon)}}{\sqrt{N\vartheta_1[A[\omega^N]]\vartheta_p[A[\omega^N]]}}\ \forall p \in [1,\infty]\right\} \leq \epsilon.$$

Fact: $\text{Prob}\{\|G_{\omega^N}(\beta)\|_\infty > \gamma/\sqrt{N}\} \leq 2\nu\exp\{-\gamma^2/2\}\ \forall\gamma\geq 0.$

$$\Rightarrow \text{Prob}\{\|G_{\omega^N}(\beta)\|_\infty \leq \sqrt{2\ln(2\nu/\epsilon)/N}\} \geq 1-\epsilon \qquad (*)$$

Claim: The Least Squares recovery $\widehat{\beta}=\widehat{\beta}(\omega^N)$ satisfies the bound

$$\|\widehat{\beta}(\omega^N)-\beta\|_p \leq \|G_{\omega^N}(\beta)\|_\infty/\sqrt{\vartheta_1[A[\omega^N]]\vartheta_p[A[\omega^N]]}. \qquad (!)$$

As a result, the recovery error admits online probabilistic bound: for every $\epsilon\in(0,1)$ one has

$$\text{Prob}\left\{\|\widehat{\beta}-\beta\|_p \leq \frac{\sqrt{2\ln(2\nu/\epsilon)}}{\sqrt{N\vartheta_1[A[\omega^N]]\vartheta_p[A[\omega^N]]}}\ \forall p\in[1,\infty]\right\} \leq \epsilon.$$

Proof: The probabilistic bound follows from (!) in view of $(*)$.

To demonstrate (!), let us fix ω^N and set $\widehat{\beta}=\widehat{\beta}(\omega^N)$, $A=A[\omega^N]$, $G(\cdot)=G_{\omega^N}(\cdot)$, $\Delta=\widehat{\beta}-\beta$.

- $G(\cdot)$ is affine $\Rightarrow G(\widehat{\beta})=G(\beta)+A\Delta$
- $\widehat{\beta}$ is weak≡strong solution to $\text{VI}(G,\mathcal{X}) \Rightarrow \langle G(\widehat{\beta}),\beta-\widehat{\beta}\rangle \geq 0$

$\Rightarrow \langle G(\beta)+A\Delta,-\Delta\rangle \geq 0$

$\Rightarrow \sqrt{\vartheta_1[A]\vartheta_p[A]}\|\Delta\|_1\|\Delta\|_p \leq \langle\Delta,A\Delta\rangle \leq \langle G(\beta),\Delta\rangle \leq \|G(\beta)\|_\infty\|\Delta\|_1$

$\Rightarrow \sqrt{\vartheta_1[A]\vartheta_p[A]}\|\Delta\|_1\|\Delta\|_p \leq \|G(\beta)\|_\infty\|\Delta\|_1.$ $\square$

$$\boxed{\vartheta_p[A] = \max\left\{s : x^T A x \geq s\|x\|_p^2\,\forall x\right\} = \min_x\{x^T A x : \|x\|_p = 1\}}$$

How to compute $\vartheta_p[A]$?

Given $\nu \times \nu$ matrix $A \succeq 0$, the computation of $\vartheta_p[A]$ is easy in the trivial case of degenerate A (in which case $\vartheta_p[A] = 0$).

When $A \succ 0$, computing $\vartheta_p[A]$ is easy when

A. $p = \infty$: $\vartheta_\infty[A] = \min_x \left\{x^T A x : \|x\|_\infty = 1\right\} = \min_{1\leq s\leq\nu} \min_x \left\{x^T A x : \|x\|_\infty \leq 1, x_s = 1\right\}$

B. $p = 2$: $\vartheta_2[A]$ is the minimal eigenvalue of A.

C. When $1 \leq p < 2$, computing $\vartheta_p[A]$ *exactly* seems to be difficult. However, *when $1 \leq p \leq 2$, $\vartheta_p[A]$ admits efficiently computable lower bound tight within the factor $\frac{\pi}{2}$.*

Indeed, $\vartheta_p[A]$ is the largest ρ such that the ellipsoid $\{x : x^T A x \leq 1\}$ is contained in the unit ball $\{x : \|x\|_p \leq 1\}$ of $\|\cdot\|_p$. Passing to the polars, this is the same as to say that $\vartheta_p[A]$ is the largest ρ such that the ellipsoid $\{y : y^T A^{-1} y \leq \rho^{-1}\}$ contains the unit ball of the norm $\|\cdot\|_q$, $q = p/(p-1)$, conjugate to $\|\cdot\|_p$. The bottom line is that

$$\vartheta_p[A] = \frac{1}{\max_{y:\|y\|_q \leq 1} y^T A^{-1} y}.$$

When $p \in [1, 2)$, we have $q \in (2, \infty]$ $\Rightarrow$ computing the maximum of the quadratic form $y^T A^{-1} y$ over $Y = \{y : \|y\|_q \leq 1\}$ admits semidefinite relaxation:

$$\max_{y \in Y} y^T A^{-1} y \leq \max_X \left\{ \mathrm{Tr}(A^{-1} X) : X \succeq 0, \|[X_{1,1}; X_{2,2}, \ldots; X_{\nu,\nu}]\|_{q/2} \leq 1 \right\}. \qquad (*)$$

By a version of Nesterov's $\pi/2$ Theorem, semidefinite relaxation, as applied to upper-bounding maximum of a *positive semidefinite* quadratic form over a set given by convex constraints on the *squares* of variables, as is the case in $(*)$, is tight within the factor $\pi/2$

$\Rightarrow$ *The quantity*

$$\frac{1}{\max_X \left\{ \mathrm{Tr}(A^{-1} X) : X \succeq 0, \|[X_{1,1}; X_{2,2}, \ldots; X_{\nu,\nu}]\|_{q/2} \leq 1 \right\}}$$

is an efficiently computable tight within the factor $\pi/2$ lower bound on $\vartheta_p[A]$.

Maximum Likelihood Recovery

♠ Consider spatio-temporal process with K locations, $M+1$ states (ground state 0 and events $1,2,...,M$) and memory depth d and assume that the vector of parameters of this process

$$\beta=\{\beta_{kp},\beta^s_{k\ell}(p,q):1\leq k,\ell\leq K,1\leq s\leq d,1\leq p\leq M,0\leq q\leq M\}\in\mathbb{R}^\nu$$

is known to belong to a given convex compact set $\mathcal{X}\subset\mathbb{R}^\nu$ such that for some $\varsigma>0$ and all $x\in\mathcal{X}$ one has

$$\begin{array}{rcll}\varsigma&\leq&x_{kp}+\sum_{s=1}^d\sum_{\ell=1}^K\min_{0\leq q\leq M}x^s_{k\ell}(p,q)\;\forall 1\leq k\leq K,1\leq p\leq M&(a)\\ 1-\varsigma&\geq&\sum_{p=1}^K\left[x_{kp}+\sum_{s=1}^d\sum_{\ell=1}^K\max_{0\leq q\leq M}x^s_{k\ell}(p,q)\right]\;\forall 1\leq k\leq K&(b)\end{array}$$

♠ Assume that the conditional, ω^{t-1} given, random states ω_{tk} of locations k at time t are independent across k.

$\Rightarrow$ *The conditional, ω^{t-1} given, minus log-likelihood of collection of states $\omega_t = \{\omega_{tk} : 1 \leq k \leq K\}$ at time t is $\sum_{k=1}^{K} \psi^k_{\omega_{tk}}(\omega^{t-1}, \beta)$,*

$$\psi^k_{\omega_{tk}}(\omega^{t-1}, \beta) = \left\{ \begin{array}{ll} -\ln([\eta^T(\omega^{t-1}_{t-d})\beta]_{kp}) & , \omega_{tk} = p \in \{1, ..., M\} \\ -\ln\left(1 - \sum_{p=1}^{M}[\eta^T(\omega^{t-1}_{t-d})\beta]_{kp}\right) & , \omega_{tk} = 0 \end{array} \right.$$

$\Rightarrow$ *Maximizing the conditional, given ω^0, likelihood of observation ω^N we arrive at the Maximum Likelihood estimate*

$$\widehat{\beta}_{\mathrm{ML}}(\omega^N) \in \mathop{\mathrm{Argmin}}_{x \in \mathcal{X}} \Psi_{\omega^N}(x) := \frac{1}{N} \sum_{t=1}^{N} \sum_{k=1}^{K} \psi^k_{\omega_{tk}}(\omega^{t-1}, x) \qquad (ML)$$

♠ **Note:** Optimization problem in (ML) is convex and therefore efficiently solvable!

$$\widehat{\beta}_{\mathsf{ML}}(\omega^N) \in \operatorname*{Argmin}_{x\in\mathcal{X}} \Psi_{\omega^N}(x) := \frac{1}{N}\sum\nolimits_{t=1}^{N}\sum\nolimits_{k=1}^{K}\psi^k_{\omega_{tk}}(\omega^{t-1},x) \qquad (ML)$$

♠ Solving convex optimization problem in (ML) is equivalent to solving VI$(G_{\omega^N},\mathcal{X})$ with

$$G_{\omega^N}(x) = \nabla\Psi_{\omega^N}(x).$$

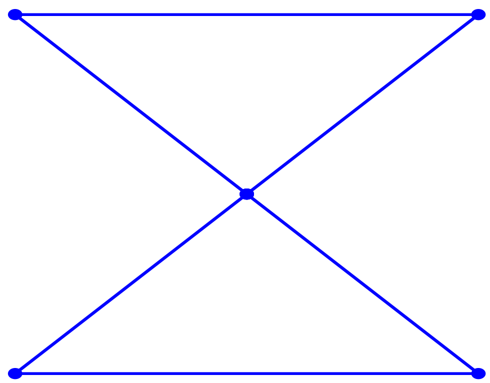
♡ **Note:** $G_{\omega^N}(\cdot)$ is monotone vector field on $\mathcal{X}$.

♡ **Note:** On a closer inspection, *typical value of $G_{\omega^N}(\beta)$ is of order of $1/\sqrt{N}$:*

$$\mathrm{Prob}\{\|G_{\omega^N}(\beta)\|_\infty > \gamma\ominus/\sqrt{N}\} \leq 2\nu\exp\{-\gamma^2/2\}\ \forall\gamma\geq 0,$$

with $\ominus$ (which was just 1 for the LS recovery) depending on ς.

How It Works: Recovering Network Structure



- $K = 5$ locations, $M = 2$ events, memory depth $d = 8$
- *It is known in advance* that state $q \in \{0, 1, 2\}$ in location ℓ contributes to the probability of event $p \in \{1, 2\}$ in location k at a later time only when $q \geq p$
- Interacting locations – neighbors in the network: k, ℓ *are not adjacent* $\Rightarrow \beta^s_{k\ell}(p, q) = 0$
- **Note:** When recovering the parameters of the process, we do *not* know the underlying network and act as if all pairs of locations were interacting.
- Our ultimate goal is to recover the network underlying the process we observe.

• Restrictions on $\mathcal{X}$:

— nonnegativity of all components of β & $\beta^s_{k\ell}(p,q)=0$ when $p>q$

— natural restriction $\sum_{p=1}^M \left[\beta_{kp}+\sum_{s=1}^d\sum_{\ell=1}^K \max_{0\leq q\leq m}\beta^s_{k\ell}(p,q)\right]\leq 1$, $k\leq K$

— $\beta^s_{k\ell}(p,q)$ should be nonincreasing and convex in s.

$\Rightarrow$ *the dimension of β is 610*

• Time horizon $N=60,000$ (not as large as it looks — we need to recover 610 parameters!)

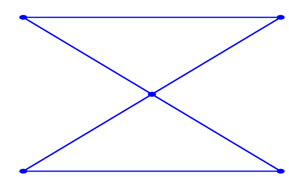
♠ Quality of recovery:

	$\|\cdot\| = \|\cdot\|_1$	$\|\cdot\| = \|\cdot\|_2$	$\|\cdot\| = \|\cdot\|_\infty$
$\|\beta - \widehat{\beta}_{\rm ML}\|$	0.9612(19.3%)	0.0600(15.5%)	0.0145(27.0%)
$\|\beta - \widehat{\beta}_{\rm LS}\|$	1.0272(20.7%)	0.0642(16.6%)	0.0145(26.9%)

In parentheses: $\|\beta - \widehat{\beta}\|$ in percents of $\|\beta\|$

♠ Network recovery:

$k \backslash \ell$	1	2	3	4	5
1	0.066	0.044	0.047	0.003	0.005
2	0.042	0.049	0.056	0.009	0.005
3	0.044	0.040	0.056	0.045	0.048
4	0.000	0.002	0.048	0.060	0.043
5	0.003	0.007	0.047	0.044	0.059



Uniform norms of collections of *recovered* interaction coefficients for locations k, ℓ

♠ Recovering frequency of events:

location	event #1	event #2
1	0.058/0.058/0.059	0.043/0.043/0.042
2	0.060/0.059/0.060	0.042/0.042/0.041
3	0.079/0.079/0.078	0.050/0.048/0.051
4	0.059/0.059/0.060	0.042/0.041/0.040
5	0.061/0.062/0.061	0.042/0.041/0.041

blue: in observations red: in simulations with $\beta \leftarrow \widehat{\beta}_{\rm ML}$ cyan: in simulations with $\beta \leftarrow \widehat{\beta}_{\rm LS}$

♠ *Given* ω^N *and* t*,* the most natural error measure for a candidate estimate $\widehat{\beta}(\omega^N)$ is the *prediction error*

$$\Delta_{\|\cdot\|}[\widehat{\beta}|t] = \|\eta^T(\omega_{t-d}^{t-1})[\widehat{\beta} - \beta]\|$$

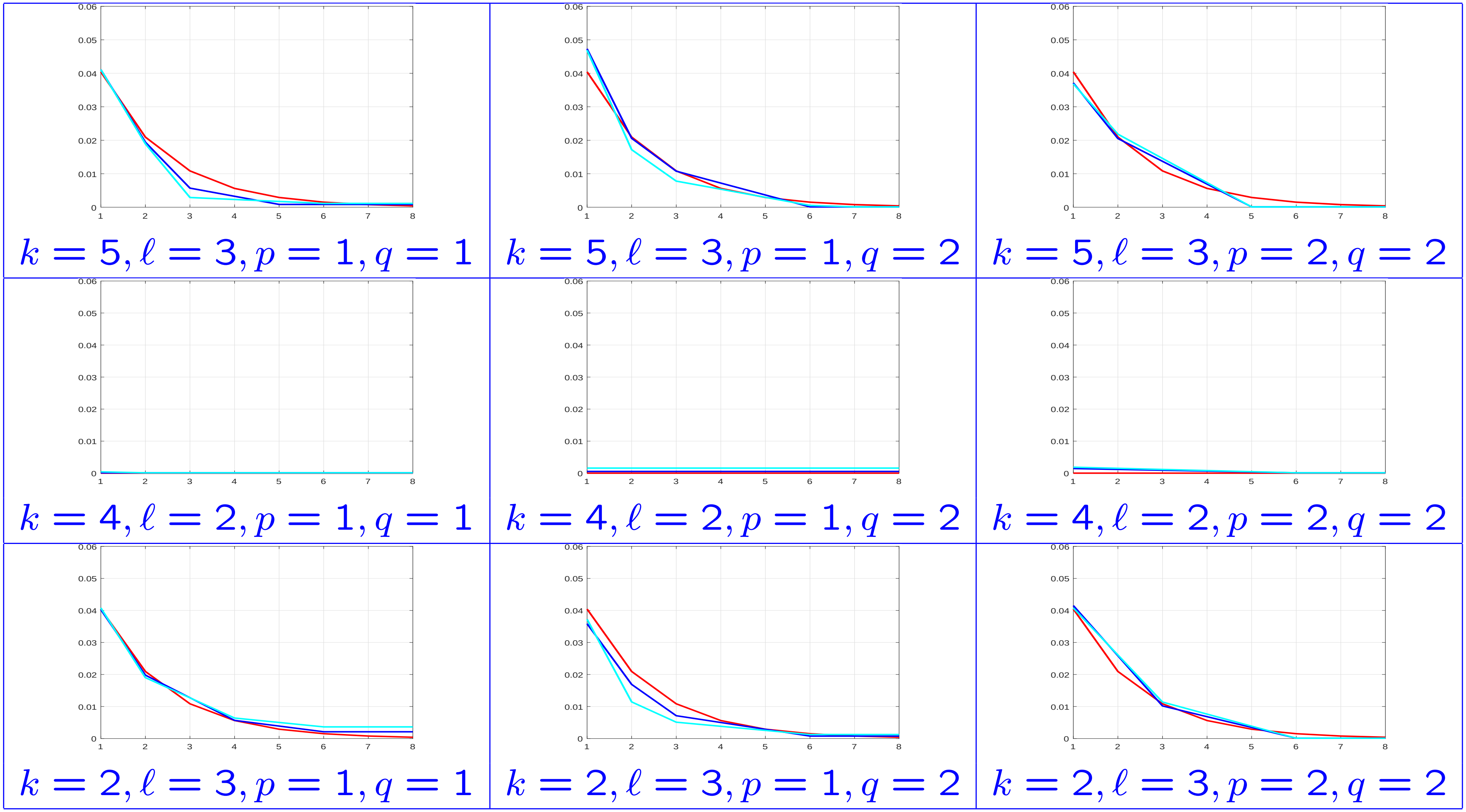
— deviation of the vector of probabilities of various events in various locations at time t as predicted by $\widehat{\beta}$ from the vector of true, under our model, probabilities.

• Here is the statistics of prediction error in our experiment:

recovery	$\|\cdot\| = \|\cdot\|_1$	$\|\cdot\| = \|\cdot\|_2$	$\|\cdot\| = \|\cdot\|_\infty$
$\widehat{\beta}_{\text{LS}}$	0.1339(5.54%) 0.0315(5.84%)	0.0545(6.60%) 0.0127(7.01%)	0.0370(8.23%) 0.0083(10.1%)
$\widehat{\beta}_{\text{ML}}$	0.1396(5.78%) 0.0298(5.52%)	0.0502(6.09%) 0.0120(6.60%)	0.0387(8.61%) 0.0077(9.48%)

red: $\max\limits_{t\leq N} \Delta_{\|\cdot\|}[\widehat{\beta}|t]$ red, %: $\max\limits_{t\leq N} \Delta_{\|\cdot\|}[\widehat{\beta}|t] / \max\limits_{t\leq N} \|\eta^T(\omega_{t-d}^{t-1})\beta\|$

cyan: $\frac{1}{N}\sum_{t\leq N} \Delta_{\|\cdot\|}[\widehat{\beta}|t]$ cyan, %: $\sum_{t\leq N} \Delta_{\|\cdot\|}[\widehat{\beta}|t] / \sum_{t\leq N} \|\eta^T(\omega_{t-d}^{t-1})\beta\|$



Sample of recoveries of $\beta^s_{k\ell}(p,q)$ vs. s

blue: β red: $\widehat{\beta}_{\mathrm{ML}}$ cyan: $\widehat{\beta}_{\mathrm{LS}}$

♠ Self-Exciting:

location	frequency of pairs of events at consecutive times
1	0.0190/0.0186/0.0191/0.0102
2	0.0189/0.0191/0.0183/0.0103
3	0.0282/0.0266/0.0277/0.0167
4	0.0181/0.0178/0.0179/0.0102
5	0.0199/0.0194/0.0198/0.0107

blue: observation

red: simulation with $\beta \leftarrow \widehat{\beta}_{\mathrm{ML}}$

cyan: simulation with $\beta \leftarrow \widehat{\beta}_{\mathrm{LS}}$

green: frequency of pairs for events independent across time

Extension: Nonlinear Link

♠ Let us identify $K\times M$ array $\{y_{kp}:1\leq k\leq K,1\leq p\leq M\}$ with KM-dimensional block vector with k-th block being $[y_{k1};y_{k2};...;y_{kM}]$. With this interpretation, $K\times M$ array $\phi(z)=\{\phi_{kp}(z):1\leq k\leq K,1\leq p\leq M\}$ of functions depending on KM-dimensional vector z becomes a *vector field*

$$\phi(z):\mathbb{R}^{KM}\to\mathbb{R}^{KM}$$

♣ Assume that we are given

A. Vector field $\phi(z)=\{\phi_{kp}(z)\}:\mathbb{R}^{KM}\to\mathbb{R}^{KM}$ and convex compact domain $\mathcal{Z}\subset\mathbb{R}^{KM}$ such that

- ϕ is continuous and monotone on $\mathcal{Z}$,
- $\forall z\in\mathcal{Z}:\phi_{kp}(z)\geq 0\,\forall k,p\ \&\ \sum_{p=1}^{M}\phi_{kp}(z)\leq 1.$

B. Memory depth d and function $\eta(\{\omega_{sk}\})$ defined on the set Ω_{dKM} of arrays $\{\omega_{sk}\in\{0,1,...,M\}:1\leq s\leq d,1\leq k\leq K\}$ and taking values in the space of $\nu\times(KM)$ matrices

C. A convex compact set $\mathcal{X}\in\mathbb{R}^{\nu}$ such that $\eta^T(\{\omega_{tk}\})x\in\mathcal{Z}$ for all $x\in\mathcal{X}$.

♣ Assume that we are given

A. Vector field $\phi(z) = \{\phi_{kp}(z)\} : \mathbb{R}^{KM} \to \mathbb{R}^{KM}$ and convex compact domain $\mathcal{Z} \subset \mathbb{R}^{KM}$ such that

- ϕ is continuous and monotone on $\mathcal{Z}$,
- $\forall z \in \mathcal{Z} : \phi_{kp}(z) \geq 0 \,\forall k, p \;\&\; \sum_{p=1}^{M} \phi_{kp}(z) \leq 1$.

B. Memory depth d and function $\eta(\{\omega_{sk}\})$ defined on the set Ω_{dKM} of arrays $\{\omega_{sk} \in \{0, 1, ..., M\} : 1 \leq s \leq d, 1 \leq k \leq K\}$ and taking values in the space of $\nu \times (KM)$ matrices

C. A convex compact set $\mathcal{X} \in \mathbb{R}^{\nu}$ such that $\eta^T(\{\omega_{tk}\})x \in \mathcal{Z}$ for all $x \in \mathcal{X}$.

♠ Given $\beta \in \mathcal{X}$ and $\omega^0_{-d+1} \in \Omega_{dKM}$, we can associate with the above data random process evolving on time horizon $t = 1, 2, ..., N$ as follows:

- the state of the process in spatio-temporal cell tk is $\omega_{tk} \in \{0, 1, ..., M\}$
- the conditional, ω^{t-1} given, probability to have $\omega_{tk} = p \in \{1, ..., M\}$ is $\phi_{kp}\left(\eta^T(\omega^{t-1}_{t-d})\beta\right)$,
- the conditional, ω^{t-1} given, probability to have $\omega_{tk} = 0$ is

$$1 - \sum_{p=1}^{M} \phi_{kp}\left(\eta^T(\omega^{t-1}_{t-d})\beta\right).$$

Note: So far we have dealt with $\phi(z) \equiv z$ and specific structure of $\eta(\cdot)$ and β.

♠ In the situation in question,

• **The role** of observable vector field $G_{\omega^N}(x)$ (which used to be the gradient field of a convex quadratic function) is played by the monotone vector field

$$G_{\omega^N}(x) = \frac{1}{N}\sum_{t=1}^{N}\left[\eta(\omega_{t-d}^{t-1})\phi\left(\eta^T(\omega_{t-d}^{t-1})x\right) - \eta(\omega_{t-d}^{t-1})\overline{\omega}_t\right],$$

where $\overline{\omega}_t \in \mathbb{R}^{KM}$ is our encoding of the collection $\{\omega_{tk} : 1 \leq k \leq K\}$ by Boolean vector

• **The role** of *un*observable vector field $\overline{G}_{\omega^N}(x)$ is played by the monotone vector field

$$\overline{G}_{\omega^N}(x) = \frac{1}{N}\sum_{t=1}^{N}\left[\eta(\omega_{t-d}^{t-1})\phi\left(\eta^T(\omega_{t-d}^{t-1})x\right) - \eta(\omega_{t-d}^{t-1})\phi(\eta^T(\omega_{t-d}^{t-1})\beta)\right],$$

for which β is a zero. As before, $G_{\omega_N}(\cdot) - \overline{G}_{\omega^N}(\cdot)$ is constant.

• **As before,** $G_{\omega^N}(\beta) = G_{\omega^N}(\beta) - \overline{G}_{\omega^N}(\beta) = \frac{1}{N}\sum_{t=1}^N \eta(\omega_{t-d}^{t-1})\left[\eta^T(\omega_{t-d}^{t-1})\beta - \overline{\omega}_t\right]$ is martingale-difference of typical magnitude of order of $1/\sqrt{N}$:

$$\mathrm{Prob}\left\{\|G_{\omega^N}(\beta)\|_\infty > \gamma\ominus/\sqrt{N}\right\} \leq 2\nu\exp\{-\gamma^2/2\}\ \forall\gamma > 0$$

$\ominus$: the maximal, over $\omega_{t-d}^{t-1} \in \Omega_{dKM}$ and $i \leq \nu$, ℓ_1-norm of i-th row in $\eta(\omega_{t-d}^{t-1})$.

• **Recommended recovery,** as before, is the solution to $\mathrm{VI}(G_{\omega^N}, \mathcal{X})$

THE END

THANK YOU AND TAKE CARE!