Robust Linear Optimization
and
Chance Constraints

$$\left\{\alpha^T[x;1] \equiv a^T x + b \le 0\right\},\ \alpha\text{: uncertain} \quad \textbf{(ULC)}$$
$$\Rightarrow \alpha^T[x;1] \le 0\ \forall \alpha \in \mathcal{U} \quad \textbf{(RC)}$$

♣ **Question:** *How to specify an uncertainty set?*

♠ **Answer:** *This is a modeling, heavily application-dependent, issue and as such it is beyond the scope of the RO theory.*

♠ **However:** *Sometimes we already have an uncertainty model, but a stochastic one rather than a model given in terms of an uncertainty/perturbation set.*

♠ **Claim:** *Given a stochastic uncertainty model, we can gain a lot by "translating" it into the RO paradigm.*

$$\left\{\alpha^T[x;1] \equiv a^T x + b \leq 0\right\}, \ \alpha\textbf{:} \ \text{uncertain} \qquad \textbf{(ULC)}$$

♣ **With the RO approach, we**

- *assume $\alpha = [a;b]$ to be affinely parameterized by a perturbation vector ζ:*

$$\alpha = \alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell \qquad (*)$$

- *assume that ζ runs through a given perturbation set $\mathcal{Z} \subset \mathbb{R}^L$, and*
- *require from (ULC) to be valid for all realizations of α associated with $\zeta \in \mathcal{Z}$.*

The x's satisfying the latter requirement are treated as "uncertainty-immunized."

♠ **With the Chance Constrained Stochastic Optimization approach, we also assume $(*)$, but**

- *instead of specifying the range $\mathcal{Z}$ of ζ, treat ζ as a random variable with (partially) known distribution, and*
- *associate with (ULC) the chance constraint*

$$\mathrm{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \leq \epsilon \qquad \text{(ChC)}$$

where $\epsilon \ll 1$ is a given tolerance.

The x's satisfying the latter requirement are treated as "uncertainty-immunized."

$$\left\{\alpha^T[x;1] \equiv a^T x + b \le 0\right\},\ \alpha\text{: uncertain} \quad \textbf{(ULC)}$$

$$\Rightarrow \text{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \le \epsilon \qquad \text{(ChC)}$$

♣ Chance Constraints: pro & con

Good news on chance constraints: *ignoring the consequences of "rare events," our decision-making becomes less conservative than with the worst-case oriented RO approach.*

Bad news on chance constraints: *passing from an uncertain constraint* (ULC) *to its chance constrained version makes sense under four if's as follows:*

• **If** *there are reasons to believe that uncertain data indeed are of stochastic nature*, **which not always is the case E.g., when uncertainty comes from** *measurement errors*, **even those involving randomness, it perhaps makes sense to speak about** *distribution of nominal* **(measured)** *data, given the true data*, **but not about** *distribution of the true data, given the measurements.*

$$\left\{\alpha^T[x;1] \equiv a^T x + b \leq 0\right\},\ \alpha\textbf{:}\ \text{uncertain} \quad \textbf{(ULC)}$$

$$\Rightarrow \text{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \leq \epsilon \qquad \text{(ChC)}$$

- **If** *we are smart enough to identify the underlying data distribution.*

While the latter indeed is the case in some Engineering applications (Communications, Signal Processing, Control,...), it typically is *not* **the case in "decision-making proper." Given the "curse of dimensionality" when identifying multivariate distributions from historical data,** *assigning the uncertain data a particular distribution more often than not is an act of faith rather than a solid inference from the experimental data.*

- **If** *we are satisfied with probabilistic guarantees like "with such and such x, the probability of a disaster is $\leq$ 1.e-4 (or $\leq$ 1.e-8)"*

Probabilistic guarantees usually make sense if the situation repeats itself many times. Their attractiveness in the *single-outcome* **situation is much more problematic.**

- **If** *we are smart enough to process* (ChC) *in a computationally efficient manner.*

$$\left\{\alpha^T[x;1] \equiv a^T x + b \le 0\right\},\ \alpha\text{: uncertain} \qquad \textbf{(ULC)}$$

$$\Rightarrow \left[\alpha_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] \le 0 \qquad \text{(RC)}$$

$$\Rightarrow \mathrm{Prob}\left\{\zeta : \left[\alpha_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \le \epsilon \quad \text{(ChC)}$$

♠ **We have seen that (RC) is computationally tractable whenever the (convex) perturbation set $\mathcal{Z}$ is so.**

Unfortunately, there are no similar "general tractability results" for (ChC)**; as a matter of fact,** *more often than not, chance constraints are computationally intractable.*

Reasons for intractability:

A. The Analysis problem associated with (ChC): *"given x, check whether x is feasible for* (ChC)*"* **usually is difficult: as a rule, the required probability is not available in a closed analytical form, while** *accurate* **numerical multi-dimensional integration is prohibitively time-consuming.**

Theorem **[L. Khachiyan]** *Consider the function*

$$vol(a) = \mathrm{mes}_L\{\zeta \in \mathbb{R}^L : 0 \leq \zeta_\ell \leq 1 \,\forall \ell, a^T\zeta \geq 1\}$$

where a *is an integral vector. Unless* P=NP, *no algorithm, given on input* a *and* $\delta > 0$, *is capable to compute* $\mathrm{vol}(a)$ *within accuracy* δ *in time polynomial in the bit length of* a *and in* $\ln(1/\delta)$.

$$\mathrm{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \leq \epsilon \qquad \text{(ChC)}$$

♠ **Note: One can evaluate the probability in (ChC) by Monte Carlo simulation. However, the required sample size should be of order of $1/\epsilon$ and thus is prohibitively large for small ϵ, like $\epsilon = 1.$e-6 or $\epsilon = 1.$e-8.**

Question: Should we bother about small ϵ?

Answer: Sometimes this is a must. Think about

- **reliability of the steering mechanism in your car**
- **an LO problem with 10,000 randomly perturbed hard constraints**

B. The feasible set of (ChC) typically is nonconvex, which makes problematic efficient minimization of linear objectives under (systems of) chance constraints.

$$\text{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^L \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \leq \epsilon \qquad \text{(ChC)}$$

Note: **Essentially, the only known generic case where neither one of the above difficulties A, B occurs is the case of $\zeta \sim \mathcal{N}(\mu, \Sigma)$ and $\epsilon \leq 1/2$. Here (ChC) is equivalent to**

$$[\alpha_0 + \sum_{\ell=1}^L \mu_\ell \alpha_\ell]^T [x;1] + \text{ErfInv}(\epsilon)\sqrt{[x;1]^T A \Sigma A^T [x;1]} \leq 0,$$

- $A = [\alpha_1; ...; \alpha_L]$
- $\int\limits_{\text{ErfInf}(r)}^{\infty} \frac{\mathrm{e}^{-s^2/2}}{\sqrt{2\pi}} ds \equiv r$

$$\mathrm{Prob}\left\{\zeta : \left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1] > 0\right\} \leq \epsilon \qquad \text{(ChC)}$$

Note: The body $\left[\alpha_0 + \sum_{\ell=1}^{L} \zeta_\ell \alpha_\ell\right]^T [x;1]$ **of (ChC) can be rewritten as**

$$w_0[x] + \sum_{\ell=1}^{L} \zeta_\ell w_\ell[x]$$

where $w_0[x], ..., w_L[x]$ **are** *affine* **functions of** x**.**

♣ **When (ChC) "as it is" is computationally intractable, we can replace it with its** *safe tractable convex approximation* **defined as follows:**

Definition. *A safe convex approximation of the chance constraint*

$$\mathrm{Prob}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!)$$

in variables w *is a* convex *subset* W *of the set of feasible solutions to* (!).

Such an approximation is called tractable, if W *is computationally tractable, i.e., it is given by an explicit finite system* $\mathcal{S}$ *of efficiently computable convex constraints* $f_i(w,v) \leq 0, i \leq I$ *in variables* w *and additional variables* v.

$$\mathrm{Prob}\Big\{\zeta : \overbrace{\Big[\alpha_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell \alpha_\ell\Big]^T [x;1]}^{\equiv w_0[x] + \sum_{\ell=1}^{m} \zeta_\ell w_\ell[x]} > 0\Big\} \leq \epsilon \quad \text{(ChC)}$$

$$\Rightarrow \mathrm{Prob}\{w_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!)$$

Note: *Given a safe tractable convex approximation* $f_i(w,v) \leq 0, i \leq I$ *of* (!), *the system of explicit efficiently computable convex constraints*

$$g_i(x,v) \equiv f_i(w[x],v) \leq 0,\ i \leq I$$

in variables x, v *possesses the following properties:*

- [tractability] *it is computationally tractable*
- [safety] *whenever* x *can be extended to a feasible solution to the system,* x *is feasible for the chance constraint* (ChC)*.*

Conclusion: Given a Chance Constrained LO problem with certain objective and replacing every chance constraint with its safe convex tractable approximation, we end up with an efficiently solvable convex optimization problem which is a safe approximation of the problem of interest: *every feasible solution to the approximation is feasible for the chance constrained problem.*

Safe Tractable Approximations of Scalar Chance Constraints and Robust Optimization

$$\mathrm{Prob}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \le \epsilon \qquad (!)$$

♣ *From now on, let ζ_i be with finite means.*

Observation: The feasible set W_+ of (!) is

- *conic*: $w \in W_+, t \ge 0 \Rightarrow tw \in W_+$
- *closed*
- *possesses a nonempty interior*, **specifically,**

$$e := [-1; 0; ...; 0] \in \mathrm{int} W_+.$$

Definition: *A safe convex approximation of* (!) **(i.e. a convex set** $W \subset W_+$**)** *is called normal, if it inherits the above properties of W_+, that is, it is conic, closed and $e \in \mathrm{int} W$.*

Conclusion: *A normal safe convex approximation of (!) is a closed convex cone $W \subset \mathbb{R}^{L+1}$ which is contained in the feasible set of (!) and is such that $e \in \mathrm{int} W$.*

$$\mathrm{Prob}\{w_0+\sum\nolimits_{\ell=1}^{L}\zeta_\ell w_\ell>0\}\leq\epsilon \qquad (!)$$

Conclusion: *A safe convex approximation of* (!) *is a closed convex* cone $W\subset\mathbb{R}^{L+1}$ *which is contained in the feasible set of* (!) *and is such that* $e=[-1;0;...;0]\in\mathrm{int}W$.

Facts: Let $\mathbf{K}\subset\mathbb{R}^{L+1}$ be a cone.

- **K is the dual of another cone, specifically, of its dual cone $\mathbf{K}_*=\{v\in\mathbb{R}^{L+1}:v^Tw\geq 0\,\forall w\in\mathbf{K}\}$:**

$$\mathbf{K}=\{w\in\mathbb{R}^{L+1}:w^Tv\geq 0\,\forall v\in\mathbf{K}_*\}$$

- **If $f\in\mathrm{int}\mathbf{K}$, then $V=\{v\in\mathbf{K}_*:f^Tv=1\}$ is a convex compact set and**

$$\mathbf{K}=\{w\in\mathbb{R}^{L+1}:v^Tw\geq 0\ \forall v\in V\}$$

♠ **Let W be a normal safe convex approximation of (!). Applying Facts to W in the role of $\mathbf{K}$ and e in the role of f, we get the following:**

There exists a convex compact set $V\subset\mathbb{R}^{L+1}$, *specifically, the set*

$$\begin{array}{rcl}V&=&\{v=[v_0;v_1;...;v_L]\in W_*:[-1;0;...;0]^Tv=1\}\\&=&\{v=[-1;v_1;...;v_L]\in W_*\}=\{[-1;-z]:z\in\mathcal{Z}\subset\mathbb{R}^L\}\end{array}$$

such that

$$\begin{array}{rcl}W&=&\{w\in\mathbb{R}^{L+1}:v^Tw\geq 0\,\forall v\in V\}\\&=&\{[w_0;...;w_L]\in\mathbb{R}^{L+1}:w_0+\sum_{\ell=1}^{L}z_\ell w_\ell\leq 0\,\forall z\in\mathcal{Z}\}\end{array}$$

$$\text{Prob}\{w_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!)$$

♠ **We have arrived at the following**

Proposition: *Every normal safe convex approximation W of (!) is of the Robust Counterpart form: there exists a convex compact set $\mathcal{Z} \subset \mathbb{R}^L$ such that*

$$W = \{[w_0; w_1; ...; w_L] : w_0 + \sum_{\ell=1}^{L} z_L w_L \leq 0 \ \forall z \in \mathcal{Z}\}$$

that is, W is the set of robust feasible solutions to the uncertainty-affected inequality

$$w_0 + \sum_{\ell=1}^{L} z_\ell w_\ell \leq 0,$$

in variables w, the uncertain coefficients being $z_1, ..., z_L$, and the perturbation set being $\mathcal{Z}$.

On a closest inspection,

The convex compact set $\mathcal{Z}$ associated with a normal safe convex approximation of (!) is computationally tractable whenever the approximation itself is so.

♣ **For the time being, we were speaking on the chance constraints of the form**

$$\mathrm{Prob}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon$$

tacitly assuming that *the probability distribution of ζ is fixed.* **In reality, more often than not the probability distribution P of ζ is only** *partially known* **— all we know is that** *P belongs to a given family $\mathcal{P}$ of probability distributions on $\mathbb{R}^L$*. **Whenever this is the case, it is natural to associate with randomly perturbed constraint its** *ambiguously chance constrained version*

$$\forall P \in \mathcal{P} : \mathrm{Prob}_{\zeta\sim P}\left\{ w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0 \right\} \leq \epsilon \qquad (!!)$$

Assuming from now on that ζ is "$\mathcal{P}$ uniformly summable:"

$$\sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\{\|\zeta\|\} < \infty,$$

the notions of normal/safe/convex/tractable approximation of a "usual" chance constraint word by word extend to the case of ambiguous chance constraint, and Proposition extends on this case as well.

$$p(w) := \sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^{L}\zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!!)$$

Generating-Function-Based Safe Convex Approximation of Chance Constraint

♣ **Definition:** *A generator is a convex nondecreasing function* $\gamma(\cdot)$ *on the axis such that* $\gamma(s) \to 0$ *as* $s \to -\infty$ *and* $\gamma(0) \geq 1$.

Examples: • $\gamma(s) = \max[1+s, 0]$ • $\gamma(s) = \exp\{s\}$

♠ **Observation: Let $\gamma(\cdot)$ be a generator. Then the function**

$$\Psi_P(w) = \mathbf{E}_{\zeta\sim P}\left\{\gamma(w_0 + \sum_{\ell=1}^{L}\zeta_\ell w_\ell)\right\} : \mathbb{R}^{L+1} \to \mathbb{R} \cup \{+\infty\}$$

is convex and lower semicontinuous, and this function is an upper bound on $p(w)$. Consequently, *the function*

$$\Psi(w) = \sup_{P\in\mathcal{P}} \Psi_P(w)$$

is a convex lower semicontinuous upper bound on $p(w)$.

♠ **Further, we clearly have $\alpha > 0 \Rightarrow p(w) = p(w/\alpha)$, whence $\alpha > 0 \Rightarrow \Psi(w/\alpha) \geq p(w)\ \forall w$.**

$$p(w) := \sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0+\sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \le \epsilon \qquad (!!)$$

$$\gamma(\cdot):\mathbb{R}\to\mathbb{R}_+:\ \nearrow, \text{ convex}, \gamma(-\infty)=0, \gamma(0)\ge 1$$

$$\Rightarrow \Psi(w) := \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{\gamma(w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell)\right\} \ge p(w) \quad (+)$$

Corollary. *The system*

$$\alpha\Psi(w/\alpha)-\alpha\epsilon\le 0,\ \ \alpha>0 \qquad (S)$$

in variables w,α is a system of convex constraints which is a safe convex approximation of $(!!)$: whenever w can be extended by properly chosen α to a feasible solution of (S), w is feasible for $(!!)$.

Proof. $\Psi(w)$ is convex by its origin, which, by a well-known fact of Convex Analysis, implies that $\alpha\Psi(w/\alpha)$ is convex in w,α in the domain $\alpha>0$. Thus, (S) is indeed a system of convex constraints in variables w,α.

Now let (w,α) be feasible for (S). Then $\Psi(w/\alpha)\le\epsilon$, whence $p(w/\alpha)\le\epsilon$ as well. Since $p(w)=p(w/\alpha)\le\epsilon$, w is feasible for $(!!)$, Q.E.D.

♠ A simple technical exercise allows to strengthen Corollary to the following

Proposition. *The convex constraint*

$$G(w) := \inf_{\alpha>0}\{\alpha\Psi(w/\alpha)-\alpha\epsilon\}\le 0$$

is a safe convex approximation of $(!!)$.

$$p(w) := \sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0 + \textstyle\sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \le \epsilon \qquad (!!)$$

$$\gamma(\cdot) : \mathbb{R} \to \mathbb{R}_+ : \nearrow, \text{ convex}, \ \gamma(0) \ge 1, \ \gamma(-\infty) = 0$$

$$\Rightarrow \Psi(w) := \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{\gamma(w_0 + \textstyle\sum_{\ell=1}^{L} \zeta_\ell w_\ell)\right\} \ge p(w) \quad (+)$$

Corollary of Proposition: *Let $\Psi^+(\cdot)$ be a convex upper bound on $\Psi(\cdot)$. Then the system of convex constraints in variables w, α*

$$\alpha\Psi^+(w/\alpha) - \alpha\epsilon \le 0, \ \ \alpha > 0 \qquad (S)$$

and the convex constraint

$$G^+(w) := \inf_{\alpha>0} \left\{\alpha\Psi^+(w/\alpha) - \alpha\epsilon\right\} \le 0$$

are safe convex approximations of (!!).

♣ **Example:** Assume that ζ_ℓ are known to be independent zero mean and taking values in $[-1,1]$, **or, equivalently,** *$\mathcal{P}$ is comprised of product-type distributions P with zero mean marginals P_ℓ supported on $[-1,1]$.*

♠ **Let us apply the above approximation scheme with the generator** $\gamma(s)=\exp\{s\}$**. When** $P=P_1\times\ldots\times P_L\in\mathcal{P}$**, we have**

$$\begin{array}{l}\Psi_P(w)=\mathbf{E}_{\zeta\sim P_1\times\ldots\times P_L}\{\exp\{w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell\}\}\\ =\exp\{w_0\}\prod\limits_{\ell=1}^L\mathbf{E}_{\zeta_\ell\sim P_\ell}\{\exp\{\zeta_\ell w_\ell\}\}.\end{array}$$

Lemma: *Let Q be a zero mean distribution supported on $[-1,1]$. Then*

$$\int\exp\{ts\}dQ(s)\leq\cosh(t)\leq\exp\{t^2/2\}.\qquad(*)$$

Proof. The inequality $\cosh(t)\leq\exp\{t^2/2\}$ **is evident. To prove that** $\int\exp\{ts\}dQ(s)\leq\cosh(t)$**, let** $f(s)=\exp\{ts\}-s\sinh(t)$**. Since** Q **is with zero mean and is supported on** $[-1,1]$**, we have**

$$\begin{array}{l}\int\exp\{ts\}dQ(s)=\int f(s)dQ(s)\\ \leq\max\limits_{-1\leq s\leq 1}f(s)=\max\limits_{s=\pm1}f(s)=\cosh(t).\end{array}$$

Note: *When Q is uniform on $\{-1;1\}$, the first inequality in $(*)$ becomes equality.*

♠ **We see that we are in the situation**

$$\Psi(w) = \exp\{w_0\} \prod_{\ell=1}^{L} \cosh(w_\ell)$$
$$\leq \Psi^+(w) := \exp\{w_0 + \tfrac{1}{2}\textstyle\sum_{\ell=1}^{L} w_\ell^2\}.$$

$\Rightarrow$ **For our $\mathcal{P}$, the safe convex approximation of the chance constraint**

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon$$

reads

$$\inf_{\alpha>0} \underbrace{\alpha\left[\exp\{\alpha^{-1} w_0 + \frac{\alpha^{-2}}{2}\sum\nolimits_{\ell=1}^{L} w_\ell^2\} - \epsilon\right]}_{f(\alpha)} \leq 0. \qquad (*)$$

♡ **Assuming** $w_1^2 + ... + w_L^2 > 0$**, we have** $f(\alpha) \to \infty$ **as** $\alpha \to +\infty$ **and as** $\alpha \to +0$

$\Rightarrow$ $(*)$ **is equivalent to** $\exists \alpha > 0 : \alpha^{-1} w_0 + \frac{\alpha^{-2}}{2} \sum\limits_{\ell=1}^{L} \omega_\ell^2 \leq \ln(\epsilon)$**,**

or, which is the same, to

$$w_0 + \sqrt{2\ln(1/\epsilon)}\sqrt{\sum\nolimits_{\ell=1}^{L} w_\ell^2} \leq 0. \qquad (+)$$

♡ **When** $w_1 = ... = w_L = 0$**,** $(*)$ **also is equivalent to** $(+)$**.**

♠ **Bottom line:** *When ζ_ℓ are independent zero mean random variables taking values in $[-1,1]$, the conic quadratic inequality*

$$w_0 + \sqrt{2\ln(1/\epsilon)}\sqrt{\sum\nolimits_{\ell=1}^L w_\ell^2} \leq 0. \qquad (+)$$

is a safe tractable approximation of the chance constraint

$$\mathrm{Prob}\{w_0 + \sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \leq \epsilon$$

♠ **Observation:** *The constraint $(+)$ is the RC*

$$w_0 + \sum_{\ell=1}^L a_\ell w_\ell \leq 0 \ \forall a \in \mathcal{U}$$

of the uncertain constraint

$$w_0 + \sum_{\ell=1}^L a_\ell w_\ell \leq 0$$

in variables w, with the ball

$$\mathcal{U} = \{a : \|a\|_2 \leq \sqrt{2\ln(1/\epsilon)}\}$$

in the role of the uncertainty set.

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \le \epsilon \qquad (!!)$$

♣ **Discussion. The simplest RO-based way to safely approximate** (!!) **is as follows:**

- **We choose as the uncertainty set $\mathcal{U}$ a convex compact set $\mathcal{U}^\epsilon$ which** “(1 − ϵ)-*supports*” **all distributions from $\mathcal{P}$:**

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\{\zeta \not\in \mathcal{U}^\epsilon\} \le \epsilon.$$

- **We set $\mathcal{U} = \mathcal{U}^\epsilon$, thus ensuring that**

$$\begin{array}{c} w_0 + \sum_{\ell=1}^{L} a_\ell w_\ell \le 0\,\forall a \in \mathcal{U}^\epsilon \\ \Rightarrow \forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{w_0 + \sum_{\ell=1}^{L}\zeta_\ell w_\ell > 0\right\} \le \epsilon. \end{array}$$

♠ **Note:** The perturbation set

$$\mathcal{U} = \{a : \|a\|_2 \leq \Omega := \sqrt{2\ln(1/\epsilon)}\}$$

yielded by our approximation scheme is, for ϵ fixed and L large, *incomparably smaller* than any $(1-\epsilon)$-support $\mathcal{U}^\epsilon$ of the distributions from $\mathcal{P}$.

For example, with $\epsilon = 1.\text{e-}6$ **we have** $\Omega = 5.26$**, and**

- **When** $L > \Omega^2 = 27.63$ **and** P **is the uniform distribution on the vertices of** $[-1,1]^L$**, we have** $P\{\zeta \in \mathcal{U}\} = 0$**;**
- **When** $\epsilon < 1/2$**, the Euclidean diameter of every** $(1-\epsilon)$**-support of the distributions from** $\mathcal{P}$ **is at least** $2\sqrt{L}$**, while the Euclidean diameter of** $\mathcal{U}$ **is just** $2\Omega \approx 10.51$**;**
- **The ratio of the volume of an** $(1-\epsilon)$**-support of distributions from** $\mathcal{P}$ **to the volume of** $\mathcal{U}$ **exponentially grows with** L **when** $L \geq 60$**; for** $L = 256$ **this ratio is as large as** $2 \cdot 10^{44}$**.**

$$p(w) := \sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!!)$$

$$\gamma(\cdot) : \mathbb{R} \to \mathbb{R}_+ : \ \nearrow, \text{convex}, \gamma(-\infty) = 0, \gamma(0) \geq 1$$

$$\Rightarrow \Psi(w) := \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{\gamma(w_0 + \sum_{\ell=1}^L \zeta_\ell w_\ell)\right\} \qquad (+)$$

♣ **Assume that a** *convex upper bound* $\Psi^+(\cdot)$ **on** $\Psi(\cdot)$ **used in the safe convex approximation**

$$\alpha\Psi^+(w/\alpha) - \alpha\epsilon \leq 0, \ \alpha > 0$$

of (!!) *is lower semicontinuous and possesses the following properties***:**

- **(i)** $0 \in \mathrm{int}\mathrm{Dom}\,\Psi^+$
- **(ii)** $\Psi^+(0) \geq 1$, $\Psi^+(\cdot) \geq 0$ **(this property is automatic)**
- **(iii)** $\infty > \Psi^+(-t, 0, ..., 0) \to 0$ **as** $0 \leq t \to \infty$

Theorem: *Under the latter assumption, the convex inequality*

$$G^+(w) := \inf_{\alpha>0}\{\alpha\Psi^+(w/\alpha) - \alpha\epsilon\} \leq 0 \qquad (G)$$

is a normal safe convex approximation of (!!). *This approximation is tractable, provided that* $\Psi^+(\cdot)$ *is efficiently computable.*

$$\sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0+\sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \le \epsilon \qquad (!!)$$

$$\gamma(\cdot):\mathbb{R}\to\mathbb{R}_+ : \ \nearrow, \text{convex}, \gamma(-\infty)=0, \gamma(0)\ge 1$$

$$\Psi^+(w) := \sup_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{\gamma(w_0+\sum_{\ell=1}^L \zeta_\ell w_\ell)\right\} \qquad (+)$$

♣ **Assume that a** *Fenchel-type representation*

$$\Psi^+(w) = \sup_z \left\{w^T[Bz+b]-\psi(z)\right\},$$

of Ψ^+ **is available, where** $\psi(\cdot)$ **is a convex lower semicontinuous function** *with bounded level sets* $\{z:\psi(z)\le c\}$, $c\in\mathbb{R}$.

Theorem: *The set*

$$\mathcal{Z} = \{z:\psi(z)+\epsilon\le 0\}$$

is a closed convex nonempty compact set, and

$$\left\{w:\inf_{\alpha>0}\alpha[\Psi^+(w/\alpha)-\epsilon]\le 0\right\} = \left\{w:[Bz+b]^Tw\le 0\ \forall z\in\mathcal{Z}\right\}.$$

The Least Conservative Implementation: Conditional Value At Risk

$$\begin{array}{rl} \forall (P\in\mathcal{P}): \mathrm{Prob}_{\zeta\sim P}\{w_0+\sum_{\ell=1}^L \zeta_\ell w_\ell>0\}\leq\epsilon & (!!)\\ \gamma(\cdot):\ \nearrow, \text{convex}, \gamma(-\infty)=0, \gamma(0)\geq 1 & \\ \Rightarrow \Psi(w)=\sup\limits_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\left\{\gamma\left(w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell\right)\right\} & \\ \Rightarrow G(w):=\inf\limits_{\alpha>0}\alpha[\Psi(w/\alpha)-\epsilon]\leq 0 & (\mathrm{Appr}) \end{array}$$

♣ **Question:** *What is the best choice of* $\gamma(\cdot)$ *?*

♠ **Answer:** *As far as the conservatism of* (Appr) *is concerned, the best choice of* $\gamma(\cdot)$ *is*

$$\gamma(s)=\max[1+s,0]$$

$$[\textit{or } \gamma(s)=\max[1+\alpha s,0] \textit{ with } \alpha>0]$$

Indeed, when $\gamma(0)>1$, the conservatism is reduced by

$$\gamma(\cdot)\leftarrow\gamma(\cdot)/\gamma(1).$$

Assuming $\gamma(0)=1$ and setting $\alpha=\gamma'(+0)$, we get $\alpha>0$ and $\gamma(s)\geq 1+\alpha s$. Since $\gamma(s)\geq 0$ (as a generator), we get $\gamma(s)\geq\bar\gamma(s)=\max[1+\alpha s,0]$. $\bar\gamma(\cdot)$ is a legitimate generator, and since $\gamma(\cdot)\geq\bar\gamma(\cdot)$, passing from γ to $\bar\gamma$ can only reduce the conservatism of (Appr).

$$\forall (P\in\mathcal{P}): \mathrm{Prob}_{\zeta\sim P}\{w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell>0\}\leq\epsilon \qquad (!!)$$

$$\gamma(\cdot):\ \nearrow,\ \text{convex},\ \gamma(-\infty)=0,\gamma(0)\geq 1$$

$$\Rightarrow \Psi(w)=\sup_{P\in\mathcal{P}}\mathbf{E}_{\zeta\sim P}\left\{\gamma\left(w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell\right)\right\}$$

$$\Rightarrow G(w):=\inf_{\alpha>0}\alpha[\Psi(w/\alpha)-\epsilon]\leq 0 \qquad \text{(Appr)}$$

♠ **On a closest inspection, when $\mathcal{P}=\{P\}$ the approximation** (Appr) **associated with** $\gamma(s)=\max[1+s,0]$ **reads**

$$\underbrace{\min_a\left[a+\frac{1}{\epsilon}\mathbf{E}\left\{\max[w_0-a+\sum_{\ell=1}^L\zeta_\ell w_\ell,0]\right\}\right]}_{=:\mathrm{CVaR}_\epsilon[w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell]}\leq 0$$

This is the well-known *Conditional Value at Risk* **safe convex approximation of** (!!) **originating from Rockafellar et al.**

♠ **Bad news on the** CVaR **approximation:** *This approximation typically is computationally intractable.*

Two basic exceptions are:

- $\zeta\sim\mathcal{N}(a,Q)$ — **of no interest: here** (!!) **by itself is an explicit convex constraint on** w**, and no approximations are necessary**
- ζ **takes moderately many values** $\zeta^1,...,\zeta^N$ **with known probabilities** $\pi_1,...,\pi_N$**.**

♠ **When ζ takes moderately many values $\zeta^1, ..., \zeta^N$ with known probabilities $\pi_1, ..., \pi_N$, the CVaR approximation of the chance constraint reads**

$$\min_a \left\{ a + \frac{1}{\epsilon} \sum_i \pi_i \max \left[w_0 - a + \sum_{\ell=1}^{L} \zeta_\ell^i w_\ell, 0 \right] \right\} \leq \epsilon$$

♡ **The RC form of this approximation is**

$$w_0 + \sum_{\ell=1}^{L} a_\ell w_\ell \leq 0 \ \forall a \in \mathcal{U} = \left\{ \sum_{i=1}^{N} u_i \zeta^i : \begin{array}{c} 0 \leq u_i \leq \frac{\pi_i}{\epsilon} \\ \sum_i u_i = 1 \end{array} \right\}$$

“Tractable Case”: Bernstein Approximation

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!!)$$

♣ **Assume that**

Brn.1 Random perturbations $\zeta_1, ..., \zeta_L$ are independent, and their distributions P_ℓ belong to given families $\mathcal{P}_\ell$ of probability distributions on the axis;

Brn.2 We are smart enough to point out upper convex and lower semicontinuous *upper bounds* Φ_ℓ^+ on logarithmic moment-generating functions of distributions from $\mathcal{P}_\ell$:

$$\Phi_\ell^+(t) \geq \ln\left(\mathbf{E}_{\zeta_\ell\sim P_\ell}\{\exp\{t\zeta_\ell\}\}\right) \qquad \forall P_\ell \in \mathcal{P}_\ell$$

♠ Taking $\gamma(s) = \exp\{s\}$ and applying our approximation scheme, we get

$$\begin{array}{l}\forall (P = P_1 \times ... \times P_L \in \mathcal{P}) : \mathbf{E}_{\zeta\sim P}\{\exp\{w_0 + \sum_{\ell=1}^L \zeta_\ell w_\ell\}\} \\ = \exp\{w_0\} \prod_{\ell=1}^{L} \mathbf{E}_{\zeta_\ell\sim P_\ell}\{\exp\{\zeta_\ell w_\ell\}\} \\ \leq \Psi^+(w) = \exp\{w_0 + \sum_{\ell=1}^L \Phi_\ell^+(w_\ell)\} =: \exp\{w_0 + \Phi^+(w)\}, \\ \Phi^+(w) = \sum_{\ell=1}^L \Phi_\ell^+(w_\ell)\end{array}$$

$$p(w) := \sup_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0 + \sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \leq \epsilon \qquad (!!)$$

$\mathcal{P}: \zeta_1 \sim P_1, ..., \zeta_L \sim P_L$ are independent, $P_\ell \in \mathcal{P}_\ell$ Brn.1

$\Phi_\ell^+(t) \geq \sup_{P_\ell\in\mathcal{P}_\ell} \int \exp\{ts\} dP_\ell(s),\ \ell = 1, ..., L$ Brn.2

$$\Rightarrow \left\{\begin{array}{l} \forall P \in \mathcal{P}: \\ \mathbf{E}_{\zeta\sim P}\{\exp\{w_0 + \sum\limits_{\ell=1}^L \zeta_\ell w_\ell\}\} \leq \exp\{w_0 + \Phi^+(w)\} \quad (*) \\ \Phi^+(w) = \sum_{\ell=1}^L \Phi_\ell^+(w_\ell) \end{array}\right.$$

♠ $(*)$ **implies that** $\ln(p(w)) \leq w_0 + \Phi^+(w)$ **and therefore**

$$\forall \alpha > 0 : \ln(p(w)) \leq w_0/\alpha + \Phi^+(w/\alpha)$$

Applying "in logarithmic scale" the reasoning which led us to our general approximation scheme, we arrive at the result as follows.

$$
\begin{array}{ll}
p(w) := \sup\limits_{P\in\mathcal{P}} \mathrm{Prob}_{\zeta\sim P}\{w_0+\sum_{\ell=1}^L \zeta_\ell w_\ell > 0\} \leq \epsilon & (!!)\\
\mathcal{P}: \zeta_1\sim P_1,...,\zeta_L\sim P_L \text{ are independent, } P_\ell\in\mathcal{P}_\ell & \text{Brn.1}\\
\Phi_\ell^+(t) \geq \sup_{P_\ell\in\mathcal{P}_\ell} \int \exp\{ts\}dP_\ell(s),\ \ell=1,...,L & \text{Brn.2}\\
\Rightarrow \Phi^+(w) = \sum_{\ell=1}^L \Phi_\ell^+(w_\ell) &
\end{array}
$$

Theorem: *In the case of Brn.1-2, (!!) admits safe convex approximation given by the convex constraint*

$$H^+(w) := \inf_{\alpha>0} \alpha\left[w_0/\alpha + \Phi^+(w/\alpha) + \ln(1/\epsilon)\right] \leq 0$$

This approximation is tractable, provided that the lower semicontinuous convex functions $\Phi_\ell^+(\cdot)$ are efficiently computable.
Assuming $0 \in \mathrm{int}\mathrm{Dom}\,\Phi^+$, the approximation is normal and thus is of the RC form:

$$H^+(w) \leq 0 \Leftrightarrow w_0 + a^T[w_1;...;w_l] \leq 0\,\forall a\in\mathcal{U},$$

where $\mathcal{U}$ is a nonempty convex compact set. When, in addition,

$$\Phi^+(w) = \sup_z \left[[w_1;...;w_L]^T[Bz+b] - \phi(z)\right]$$

for a lower semicontinuous convex function $\phi(\cdot)$ with bounded level sets, one can take

$$\mathcal{U} = \{Bz+b : \phi(z) \leq \ln(1/\epsilon)\}.$$

$$\forall (P\in\mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{w_0+\sum_{\ell=1}^{L}\zeta_\ell w_\ell>0\right\}\leq\epsilon \quad (!!)$$

Example 1: $\zeta_\ell\sim\mathcal{N}(\mu_\ell,\sigma_\ell^2)$ are independent; we know lower and upper bounds $\underline{\mu}_\ell$, $\overline{\mu}_\ell$ on μ_ℓ and upper bounds $\overline{\sigma}_\ell$ on σ_ℓ.

Building Bernstein approximation of (!!)**:**

$$\begin{array}{rl}
& \mathcal{P}_\ell=\{\mathcal{N}(\mu,\sigma^2):\underline{\mu}_\ell\leq\mu\leq\overline{\mu}_\ell,\sigma\leq\overline{\sigma}_\ell\}\\
& \xi\sim\mathcal{N}(\mu,\sigma^2)\Rightarrow\ln\left(\mathbf{E}\{\exp\{t\xi\}\}\right)=\mu t+\frac{\sigma^2}{2}t^2\\
\Rightarrow & \Phi_\ell(t):=\sup_{P_\ell\in\mathcal{P}_\ell}\ln\left(\mathbf{E}_{\zeta_\ell\sim P_\ell}\{\exp\{t\zeta_\ell\}\}\right)\\
& =\max[\underline{\mu}_\ell t,\overline{\mu}_\ell t]+\frac{\overline{\sigma}_\ell^2}{2}t^2\\
& =\max\limits_{0\leq\lambda\leq1}\left\{[\lambda\underline{\mu}_\ell+(1-\lambda)\overline{\mu}_\ell]t+\frac{\overline{\sigma}_\ell^2}{2}t^2\right\}\\
& =\max\limits_{0\leq\lambda\leq1}\max\limits_{v}\left\{vt-\frac{1}{2\overline{\sigma}_\ell^2}[\lambda\underline{\mu}_\ell+(1-\lambda)\overline{\mu}_\ell-v)^2\right\}\\
& =\max\limits_{v}\left\{vt-\min\limits_{0\leq\lambda\leq1}\frac{1}{2\overline{\sigma}_\ell^2}[\lambda\underline{\mu}_\ell+(1-\lambda)\overline{\mu}_\ell-v]^2\right\}\\
& =\max\limits_{v}\{vt-\phi_\ell(v)\}\\
& \phi_\ell(v)=\frac{1}{2\overline{\sigma}_\ell^2}\cdot\left\{\begin{array}{ll}(v-\overline{\mu}_\ell)^2, & v\geq\overline{\mu}_\ell\\ 0, & v\in[\underline{\mu}_\ell,\overline{\mu}_\ell]\\ (v-\underline{\mu}_\ell)^2, & v\leq\underline{\mu}_\ell\end{array}\right.\\
& =\frac{\mathrm{dist}^2(v,[\underline{\mu}_\ell,\overline{\mu}_\ell])}{2\overline{\sigma}_\ell^2}
\end{array}$$

♠ **The Bernstein approximation is**

$$\begin{array}{l}\inf\limits_{\alpha>0}\alpha\left\{w_0/\alpha+\sum_{\ell=1}^{L}\left[\max[\underline{\mu}_\ell w_\ell/\alpha,\overline{\mu}_\ell w_\ell/\alpha]+\frac{\overline{\sigma}_\ell^2}{2}(w_\ell/\alpha)^2\right]+\ln(1/\epsilon)\right\}\leq 0\\ \Leftrightarrow\ w_0+\sum_{\ell=1}^{L}\max[\underline{\mu}_\ell w_\ell,\overline{\mu}_\ell w_\ell]\\ \qquad\qquad\qquad+\sqrt{2\ln(1/\epsilon)}\sqrt{\sum_{\ell=1}^{L}\overline{\sigma}_\ell^2 w_\ell^2}\leq 0\end{array}$$

♠ **The uncertainty set $\mathcal{U}$ participating in the RC representation of the Bernstein approximation is**

$$\mathcal{U}=\{a\in\mathbb{R}^L:\sum_{\ell=1}^{L}\frac{\mathrm{dist}^2(a_\ell,[\underline{\mu}_\ell,\overline{\mu}_\ell])}{2\overline{\sigma}_\ell^2}\leq\ln(1/\epsilon)\};$$

this is the arithmetic sum of the box

$$\{\underline{\mu}\leq a\leq\overline{\mu}\}$$

and the ellipsoid

$$\left\{a:\sqrt{\sum_{\ell=1}^{L}a_\ell^2/\overline{\sigma}_\ell^2}\leq\sqrt{2\ln(1/\epsilon)}\right\}$$

♡ **When $\underline{\mu}_\ell=\overline{\mu}_\ell=0$, we end up with ellipsoidal uncertainty set.**

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{w_0 + \sum\nolimits_{\ell=1}^{L} \zeta_\ell w_\ell > 0\right\} \leq \epsilon \qquad (!!)$$

Example 2: ζ_ℓ are independent zero mean supported on $[-1,1]$ **(by "scaling"** $\zeta_\ell \leftarrow \mu_\ell + \sigma_\ell \zeta_\ell$ **this covers the case of independent** ζ_ℓ **with known means and known finite ranges). We have**

$$\begin{array}{rl} & \mathcal{P}_\ell = \{P_\ell : \text{ zero mean and supported on } [-1,1]\} \\ \Rightarrow & \Phi_\ell(t) := \sup\limits_{P_\ell \in \mathcal{P}_\ell} \ln\left(\int \exp\{ts\} dP_\ell(s)\right) = \ln(\cosh(t)) \\ & = \sup\limits_v \{vt - \phi(v)\}, \\ & \phi(v) = \sup\limits_t \{vt - \ln(\cosh(t))\} \\ \Rightarrow & \phi(v) = \frac{1}{2}[(1+v)\ln(1+v) + (1-v)\ln(1-v)], \\ & \mathrm{Dom}\,\phi = [-1,1]. \end{array}$$

$\Rightarrow$ **The Bernstein approximation of** (!!) **is**

$$w_0 + \inf_{\alpha>0} \alpha\left[\sum\nolimits_{\ell=1}^{L} \ln(\cosh(w_\ell/\alpha)) + \ln(1/\epsilon)\right] \leq 0.$$

♡ **The RC form of the Bernstein approximation is**

$$w_0 + a^T[w_1;...;w_L] \leq 0 \ \forall a \in \mathcal{U}^{\mathrm{Entr}}$$

$$\mathcal{U}^{\mathrm{Entr}} = \left\{a : -1 \leq a_\ell \leq 1, \right.$$
$$\left. \sum_{\ell=1}^{L}[(1-a_\ell)\ln(1-a_\ell) + (1+a_\ell)\ln(1+a_\ell)] \leq 2\ln(1/\epsilon)\right\}.$$

$$\forall(P\in\mathcal{P}):\mathrm{Prob}_{\zeta\sim P}\left\{w_0+\sum_{\ell=1}^{L}\zeta_\ell w_\ell>0\right\}\leq\epsilon\ \ (!!)$$

$$\mathcal{P}:\ \text{products of zero mean } P_\ell \text{ supported on } [-1,1]$$

♡ **The Entropy approximation of (!!) is**

$$w_0+a^T[w_1;...;w_L]\leq 0\ \forall a\in\mathcal{U}^{\rm Entr}$$

$$\mathcal{U}^{\rm Entr}=\left\{a:-1\leq a_\ell\leq 1,\right.$$
$$\left.\sum_{\ell=1}^{L}[(1-a_\ell)\ln(1-a_\ell)+(1+a_\ell)\ln(1+a_\ell)]\leq 2\ln(1/\epsilon)\right\}.$$

or, which is the same,

$$w_0+\inf_{\alpha>0}\alpha\left[\sum_{\ell=1}^{L}\ln(\cosh(w_\ell/\alpha))+\ln(1/\epsilon)\right]\leq 0.$$

♡ **We have** $(1-v)\ln(1-v)+(1+v)\ln(1+v)\geq v^2$**,** $-1\leq v\leq 1$**, and thus**

$$\mathcal{U}^{\rm Entr}\subset\mathcal{U}^{\rm BallBox}=\left\{a:\|a\|_\infty\leq 1,\|a\|_2\leq\sqrt{2\ln(1/\epsilon)}\right\}.$$

Replacing $\mathcal{U}^{\rm Entr}$ **with** $\mathcal{U}^{\rm BallBox}$**, we arrive at a slightly more conservative, as compared to the Entropy one,** *Ball-Box approximation* **of** (!!)**:**

$$w_0+a^T[w_1;...;w_L]\leq 0\ \forall a\in\mathcal{U}^{\rm BallBox}$$
$$\Updownarrow$$
$$\exists u:w_0+\|u\|_1+\sqrt{2\ln(1/\epsilon)}\|[w_1-u_1;...;w_L-u_L]\|_2\leq 0.$$

$$\forall(P\in\mathcal{P}):\ \mathrm{Prob}_{\zeta\sim P}\left\{w_0+\sum_{\ell=1}^L\zeta_\ell w_\ell>0\right\}\leq\epsilon\quad(!!)$$
$$\mathcal{P}:\ \text{products of zero mean } P_\ell \text{ supported on } [-1,1]$$

♡ **Further extending $\mathcal{U}^{\mathrm{BallBox}}$ to the ball**

$$\mathcal{U}^{\mathrm{Ball}}=\{a:\|a\|_2\leq\sqrt{2\ln(1/\epsilon)}\},$$

we arrive at a more conservative, as compared to the the Ball-Box one, *Ball approximation* **of** (!!)**:**

$$w_0+a^T[w_1;...;w_L]\leq 0\ \forall a\in\mathcal{U}^{\mathrm{Ball}}$$
$$\Updownarrow$$
$$w_0+\sqrt{2\ln(1/\epsilon)}\sqrt{\sum_{\ell=1}^L w_\ell^2}\leq 0$$

which we have obtained from the Bernstein approximation scheme when replacing the best possible under circumstances functions $\Phi_\ell(t)=\ln(\cosh(t))$ with their upper bounds $\frac{1}{2}t^2$.

♡ **Observing that $\|z\|_1\leq\sqrt{\dim z}\|z\|_2$, we can extend $\mathcal{U}^{\mathrm{BallBox}}$ to the set $\mathcal{U}^{\mathrm{Budget}}=\{a\in\mathbb{R}^L:\|a\|_\infty\leq 1,\|a\|_1\leq\sqrt{2\ln(1/\epsilon)L}\}$, thus arriving at the** *Budgeted approximation* **of** (!!)**:**

$$w_0+a^T[w_1;...;w_L]\leq 0\ \forall a\in\mathcal{U}^{\mathrm{Budget}}$$
$$\Updownarrow$$
$$\exists u:w_0+\|u\|_1+\sqrt{2\ln(1/\epsilon)L}\|[w_1-u_1;...;w_L-u_L]\|_\infty\leq 0$$

(purely LO approximation).

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{w_0 + \sum_{\ell=1}^{L} \zeta_\ell w_\ell > 0\right\} \leq \epsilon \quad (!!)$$
$$\mathcal{P}: \ \text{products of zero mean } P_\ell \text{ supported on } [-1,1]$$

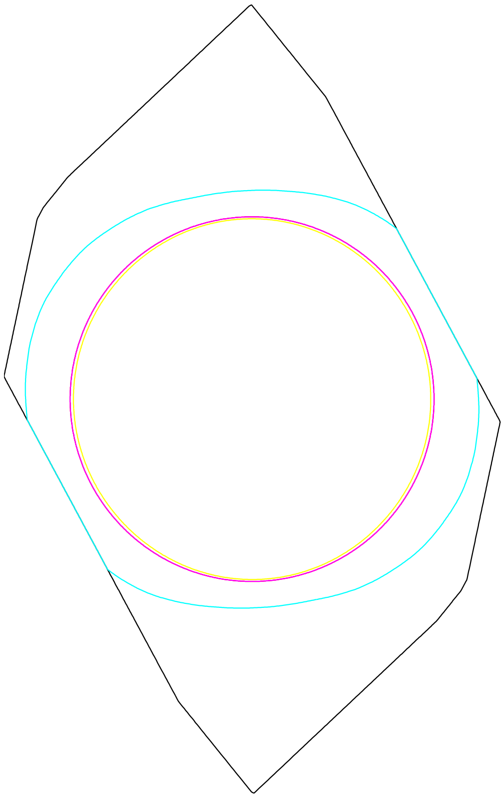
♡ **Finally, extending** $\mathcal{U}^{\mathrm{Entr}}$ **to the box** $\mathcal{U}^{\mathrm{Box}} = \{a : \|a\|_\infty \leq 1\}$**, we get the most conservative** *Box approximation* **of** (!!)**:**

$$w_0 + a^T[w_1; ...; w_L] \leq 0 \ \forall a \in \mathcal{U}^{\mathrm{Box}}$$
$$\Updownarrow$$
$$w_0 + \sum_{\ell=1}^{L} |w_\ell| \leq 0$$

This approximation uses only the domain information on ζ_ℓ **and ignores completely the fact that** ζ_ℓ **are independent with zero means.**

$$\forall(P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{w_0 + \sum_{\ell=1}^{L}\zeta_\ell w_\ell > 0\right\} \leq \epsilon \quad (!!)$$

$\mathcal{P}$: products of zero mean P_ℓ supported on $[-1, 1]$



Example 2: Random 2D central cross-sections of perturbation sets corresponding to various approximations, $L = 256$

- **black: Box approximation**
- **cyan: Budgeted approximation**
- **red: Ball approximation**
- **magenta: Ball-Box approximation**
- **yellow: Entropy approximation**

Illustration: Portfolio selection

There are $L = 200$ assets with independent random yearly returns r_ℓ. It is known that

- For $\ell \leq 199$, return r_ℓ has expectation $\mu_\ell = 1.05 + 0.3\frac{200-\ell}{199}$ and varies in $[\mu_\ell - \sigma_\ell, \mu_\ell + \sigma_\ell]$, $\sigma_\ell = 0.05 + 0.6\frac{200-\ell}{199}$;
- For $\ell = 200$, $r_\ell \equiv 1.05$ ["money in the bank"].

We want to distribute \$1 between the assets in order to maximize the Value-at-0.5%-Risk of the portfolio in a year from now, that is, we want to solve the chance constrained problem

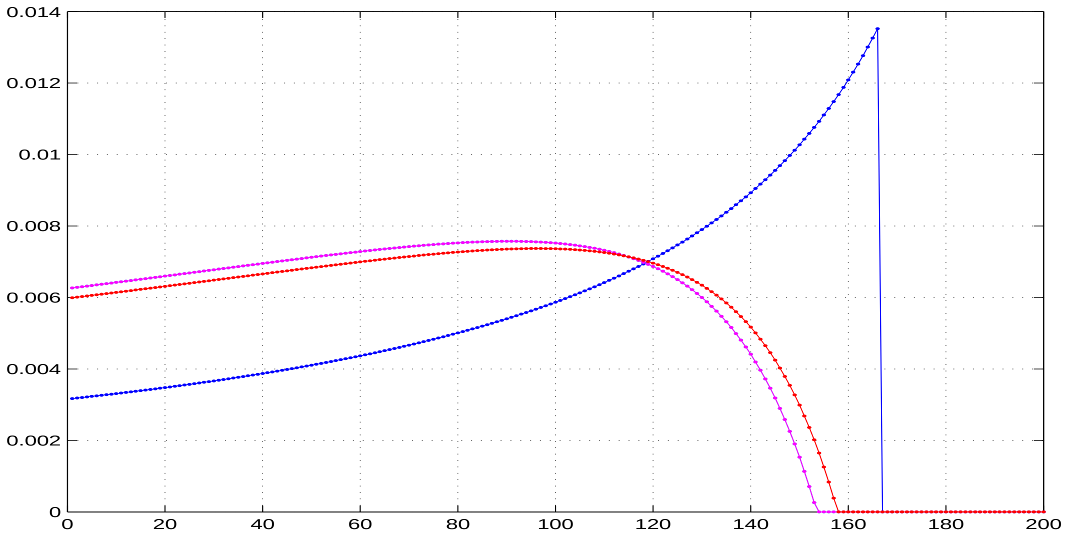
$$\max_{t,x}\left\{t : \begin{array}{l} \mathrm{Prob}\left\{\sum_{\ell=1}^{L} r_\ell x_\ell < t\right\} \leq \epsilon = 0.005 \\ x \geq 0, \sum_\ell x_\ell = 1 \end{array}\right\}$$

♠ **Setting $r_\ell = \mu_\ell + \sigma_\ell \zeta_\ell$, the random variables ζ_ℓ are independent zero mean and take values in $[-1, 1]$. The problem of interest now reads**

$$\max_{t,x}\left\{t : \begin{array}{l} x \geq 0, \sum_\ell x_\ell = 1 \\ \mathrm{Prob}\{w_0[x,t] + \sum_{\ell=1}^{n} \zeta_\ell w_\ell[x,t] > 0\} \leq \epsilon \end{array}\right\},$$

$$w_0[x,t] = t - \sum_{\ell=1}^{L} \mu_\ell x_\ell,\ \ w_\ell[x,t] = -\sigma_\ell x_\ell$$

Replacing the chance constraint with its safe tractable approximation and solving the resulting convex program, we get a *feasible* **suboptimal solution to the problem of interest.**



Distribution of $1 between assets for various approximations of the Portfolio Selection problem

- **blue: Budgeted approximation**
- **magenta: Ball and Ball-Box approximations**
- **red: Entropy approximation**

The Box approximation (not shown on the plot) leads to the solution "keep all money in the bank."

Approx.	Box	Budgeted	Ball	Ball-Box	Entropy
Opt. Val.	1.0500	1.1012	1.1200	1.1200	1.1209

Beyond the Scope of Affinely Perturbed Chance Constraints with Independent Perturbations

♣ **As far as** *tractable* **approximations are concerned, our construction imposes on ζ_ℓ the requirements to** *be independent* **and to** *enter affinely* **the body of the chance constraint.**

Consider a different situation, where

- **the random perturbations ζ_ℓ enter the body of chance constraint** *quadratically* **(decision variables still enter it linearly), and**
- **we have** *partial* **information on the** *marginal distributions* **P_ℓ of ζ_ℓ, on their** *covariances***, and on the** *domain* **of ζ, but no more than that.**

♠ **Thus, from now on our chance constraint is**

$$\begin{array}{l}\forall (P\in\mathcal{P}):\\ \mathrm{Prob}_{\zeta\sim P}\{\underbrace{\zeta^TU[x]\zeta+2\zeta^Tv[x]+w[x]}_{=\mathrm{Tr}(W[x]Z[\zeta])}>0\}\leq\epsilon\end{array}$$

$$W[x]=\left[\begin{array}{c|c}U[x]&v[x]\\ \hline v^T[x]&w[x]\end{array}\right],\ Z[\zeta]=\left[\begin{array}{c|c}\zeta\zeta^T&\zeta\\ \hline \zeta^T&1\end{array}\right]$$

where $W[x]$ is affine in x.

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{\mathrm{Tr}(WZ[\zeta]) > 0\right\} \leq \epsilon \qquad (!!)$$

As about $\mathcal{P}$, we assume that this family is comprised of all distributions on $\mathbb{R}^L$ such that

- **the marginal distributions P_ℓ belong to known families $\mathcal{P}_\ell$ of probability distributions on the axis,**
- **the covariance matrix $V = \mathbf{E}\{Z[\zeta]\}$ of ζ is known to belong to a given closed convex subset $\mathcal{V}$ of the positive semidefinite cone,**
- **ζ is supported on a set S given by a finite list of quadratic constraints:**

$$\mathrm{Tr}(A_i Z[\zeta]) \leq 0, \ \ i = 1, ..., I.$$

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{\mathrm{Tr}(WZ[\zeta]) > 0\right\} \le \epsilon \qquad (!!)$$

$$\mathcal{P} : \zeta_\ell \sim P_\ell \in \mathcal{P}_\ell,\ \mathbf{E}_{\zeta\sim P}\{Z[\zeta]\} \in \mathcal{V}$$

$$\forall P \in \mathcal{P} : \mathrm{supp}(P) \subset S = \{z : \mathrm{Tr}(A_i Z[z]) \le 0,\ i \le I\}$$

♣ **The idea** of our approximation scheme inspired by Bertsimas, Popescu and Sethuraman, is as follows:

• *We build a mechanism which produces pairs* $(f(z) : \mathbb{R}^L \to \mathbb{R}, \lambda > 0)$ *such that*

A. $f(z) \ge 0$ *on* S

B. $f(z) \ge \lambda$ *whenever* $z \in S$ *and* $\mathrm{Tr}(WZ[z]) > 0$

C. *We can point out an upper bound* $\Psi[f]$ *on* $\sup\limits_{P\in\mathcal{P}} \mathbf{E}_{\zeta\sim P}\{f(\zeta)\}$.

Clearly, for a pair $(f(\cdot), \lambda)$ **produced by our mechanism for every** $P \in \mathcal{P}$ **we have**

$$\Psi[f] \ge \int f(z)dP(z) = \int\limits_S f(z)dP(z) \ge \int\limits_{\substack{z\in S:\\ \mathrm{Tr}(WZ[z])>0}} f(z)dP(z)$$

$$\ge \lambda \mathrm{Prob}_{\zeta\sim P}\{\mathrm{Tr}(WZ[\zeta]) > 0\},$$

so that *the condition*

$$\Psi[f] \le \lambda\epsilon \qquad (+)$$

is sufficient for the validity of (!!)**.**

We impose on the "free parameters" of our construction the requirement to ensure (+)**, thus arriving at a safe approximation of** (!!)**.**

♠ Implementation:

C: The simplest way to ensure the possibility to bound from above $\sup\limits_{P\in\mathcal{P}}\int f(z)dP(z)$ **is to stick to "simple"** f**, e.g., those of the form**

$$f(z)=\sum_{\ell=1}^{L} f_\ell(z_\ell)+\mathrm{Tr}(FZ[z]), \qquad (*)$$

which allows to take

$$\Psi[f]=\sum_{\ell=1}^{L}\sup_{P_\ell\in\mathcal{P}_\ell}\int f_\ell(s)dP_\ell(s)+\sup_{V\in\mathcal{V}}\mathrm{Tr}(FV).$$

A: The simplest way to ensure that a function f **of the form** $(*)$ **is** ≥ 0 **on** $S=\{z:\mathrm{Tr}(A_iZ[z])\leq 0, i\leq I\}$ **is to ensure that**

$$f_\ell(s)\geq a_\ell+2b_\ell s+c_\ell s^2,\ 1\leq\ell\leq L \qquad (a)$$

and that the quadratic form of z

$$\sum_{\ell=1}^{L}[a_\ell+2b_\ell z_\ell+c_\ell z_\ell^2]+\mathrm{Tr}(FZ[z])+\sum_{i=1}^{I}\mu_i\mathrm{Tr}(A_iZ[z])$$

where the parameters μ_i *are* ≥ 0, **is nonnegative everywhere, which amounts to the matrix inequality**

$$\left[\begin{array}{c|c}\mathrm{Diag}\{c_1,...,c_L\} & b\\ \hline b^T & \sum_\ell a_\ell\end{array}\right]+F+\sum_{i=1}^{I}\mu_iA_i\succeq 0$$

$$f(z)=\sum_{\ell=1}^{L} f_\ell(z_\ell)+\mathrm{Tr}(FZ[z]), \qquad (*)$$

B: **The simplest way to ensure that a function** f **of the form** $(*)$ **is** $\geq \lambda$ **on** $S=\{z:\mathrm{Tr}(A_iZ[z])\leq 0, i\leq I, \mathrm{Tr}(WZ[z])>0\}$ **is to ensure that**

$$f_\ell(s)\geq p_\ell+2q_\ell s+r_\ell s^2,\ 1\leq\ell\leq L \qquad (b)$$

and that the quadratic form of z

$$\begin{array}{r}\sum_{\ell=1}^{L}[p_\ell+2q_\ell z_\ell+r_\ell z_\ell^2]+\mathrm{Tr}(FZ[z])+\sum_{i=1}^{I}\nu_i\mathrm{Tr}(A_iZ[z])\\ -\mathrm{Tr}(WZ[z])-\lambda\end{array}$$

where the parameters ν_i *are* ≥ 0**, is nonnegative everywhere, which amounts to the matrix inequality**

$$\left[\begin{array}{c|c}\mathrm{Diag}\{r_1,...,r_L\} & q\\ \hline q^T & \sum_\ell p_\ell-\lambda\end{array}\right]+F-W+\sum_{i=1}^{I}\nu_iA_i\succeq 0$$

$$\forall (P \in \mathcal{P}) : \mathrm{Prob}_{\zeta\sim P}\left\{\mathrm{Tr}(WZ[\zeta]) > 0\right\} \leq \epsilon \qquad (!!)$$
$$\mathcal{P} : \zeta_\ell \sim P_\ell \in \mathcal{P}_\ell,\ \mathbf{E}_{\zeta\sim P}\{Z[\zeta]\} \in \mathcal{V}$$
$$\forall P \in \mathcal{P} : \mathrm{supp}(P) \subset S = \{z : \mathrm{Tr}(A_i Z[z]) \leq 0,\ i \leq I\}$$

♠ **We have arrived at the following**

Theorem: *Whenever variables* W, λ, $\{a_\ell, b_\ell, c_\ell, p_\ell, q_\ell, r_\ell\}_{\ell=1}^L$, $\{\mu_i, \nu_i\}_{i=1}^I$, $F \in \mathbf{S}^{L+1}$, θ *satisfy the system of convex constraints*

$$(a)\quad \left[\begin{array}{c|c} \mathrm{Diag}\{c_1, ..., c_L\} & b \\ \hline b^T & \sum_\ell a_\ell \end{array}\right] + F + \sum_{i=1}^I \mu_i A_i \succeq 0$$

$$(b)\quad \left[\begin{array}{c|c} \mathrm{Diag}\{r_1, ..., r_L\} & q \\ \hline q^T & \sum_\ell p_\ell - \lambda \end{array}\right] + F - W + \sum_{i=1}^I \nu_i A_i \succeq 0$$

$$(c)\quad \left[\begin{array}{c|c} \lambda & 1 \\ \hline 1 & \theta \end{array}\right] \succeq 0 \text{ [says that } \lambda > 0]$$

$$(d)\quad \sum_{\ell=1}^L \sup_{P_\ell \in \mathcal{P}_\ell} \int \max[a_\ell + 2b_\ell s + c_\ell s^2, p_\ell + 2q_\ell s + r_\ell s^2] dP_\ell(s) + \max_{V \in \mathcal{V}} \mathrm{Tr}(FV) \leq \lambda\epsilon$$

$$(e)\quad \mu_i \geq 0,\ \nu_i \geq 0$$

Thus, this system is a safe convex approximation of (!!). *This approximation is tractable, provided that the suprema in* (d) *are efficiently computable.*

How it works? Portfolio Selection revisited.

There are L assets with random yearly returns $r_\ell = 1 + \mu_\ell + \sigma_\ell\zeta_\ell$, $1 \leq \ell \leq L$, where $\mu_\ell \geq 0$ and $\sigma_\ell \geq 0$ are known expected gains and their variabilities, and ζ_ℓ are random perturbations taking values in $[-1, 1]$. Given partial information on the distribution of $\zeta = [\zeta_1; ...; \zeta_L]$, we want to distribute \$1 between the assets in order to maximize the guaranteed value-at-ϵ-risk of the profit $\sum_\ell [\mu_\ell + \sigma_\ell\zeta_\ell] x_\ell$.

♡ In the experiments, we set $L = 16,\ \epsilon = 0.005$,

$$\mu_\ell = 0.001 + 0.9\frac{\ell - 1}{L - 1},\ \sigma_\ell = \left(0.9 + 0.2\frac{\ell - 1}{L - 1}\right)\mu_\ell$$

and have considered 3 families of distributions $\mathcal{P}$:

$\mathcal{P}_1$: ζ_ℓ are independent and take values ±1 with probabilities 0.5

$\mathcal{P}_2$: ζ_ℓ take values ±1 with probabilities 0.5, and $|\mathbf{E}\{\zeta_\ell\zeta_{\ell'}\}| \leq \rho = 0.012$ when $\ell \neq \ell'$

$\mathcal{P}_3$: ζ_ℓ take values ±1 with probabilities 0.5.

♡ We have carried out 4 experiments:

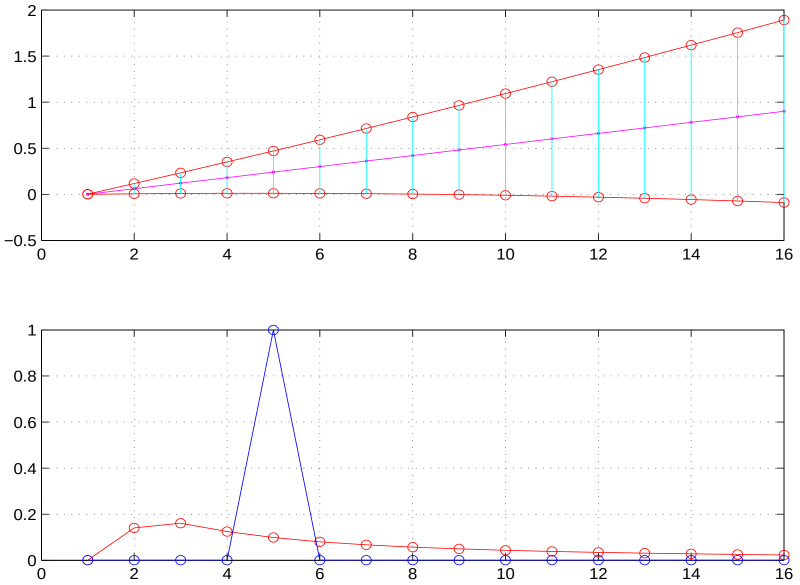
I: $\mathcal{P} = \mathcal{P}_1$, Bernstein approximation scheme;

II: $\mathcal{P} = \mathcal{P}_2$, $\mathcal{V} = \mathcal{V}_{0.012}$ (positive semidefinite matrices with unit diagonal and off-diagonal entries from $[-0.012, 0.012]$, except for all-zero off-diagonal entries in row and column #17), domain information given by $\zeta_\ell^2 \leq 1$, $1 \leq \ell \leq L$;

III: $\mathcal{P} = \mathcal{P}_3$, $\mathcal{V} = \mathcal{V}_1$ (positive semidefinite matrices with unit diagonal and zero off-diagonal entries in row and column # 17), same domain information as in II;

IV: $\mathcal{P} = \mathcal{P}_3$, no variance and domain information (i.e., $\mathcal{V}$ is comprised of all positive semidefinite matrices).

Results:



Upper plot: the expected profits μ_ℓ and variabilities σ_ℓ. Magenta: μ_ℓ; red: the possible profits $\mu_\ell \pm \sigma_\ell$.
Lower plot: investments in the resulting portfolios. Red: experiment **I**; blue: experiments **II** – **IV**.

Experiment	**I**	**II**	**III**	**IV**
Opt. Val.	0.0445	0.0112	0.0103	0.0103

Remarks:

- **Portfolios in experiments II – IV ("put everything in the asset with the largest guaranteed profit $\mu_\ell - \sigma_\ell$") are *exactly optimal* for the corresponding ambiguously chance constrained problems, and in fact – for their unambiguous versions corresponding to the distribution**

$$P_* = 0.988 \cdot \mathrm{Uniform}(\{-1;1\}^L) + 0.012 \cdot \mathrm{Uniform}\{-\mathbf{1};\mathbf{1}\}$$

- **With the true distribution being P_*, the "red" portfolio, based on the hypothesis of independent returns, has *negative* Value-at-0.5%-Risk.**
- **In order to distinguish, with reliability 0.99, between the distributions $\mathrm{Uniform}(\{-1;1\}^L)$ and P_* from historical data, you need to observe yearly returns for more than 290 years!**