ADVANCES IN CONVEX OPTIMIZATION: CONIC PROGRAMMING

Arkadi Nemirovski

Georgia Institute of Technology
and
Technion – Israel Institute of Technology

- Convex Programming – "solvable case" in Optimization
- Revealing structure of a convex program: Conic Programming
- Exploiting structure of a convex program: Polynomial Time Interior Point Methods
- Conic Quadratic and Semidefinite Programming: expressive abilities and applications

Convex Programming – “Solvable Case” in Optimization

• **Mathematical Programming** is about solving optimization problems of the form

$$\mathrm{Opt} = \min_{x\in\mathbb{R}^n} \{f(x) : g_i(x) \leq 0,\ 1 \leq i \leq m\}$$

with “good enough” (usually C^1) objective $f(\cdot)$ and constraints $g_i(\cdot)$.

• **MP is primarily operational:** while the descriptive issues (existence/uniqueness/characterization of a solution) are of definite importance, **the major goal is to approximate an optimal solution numerically**

⇒ The primary role in MP Theory is played by investigating **complexity** of generic MP problems and developing **efficient** solution algorithms.

$$\text{Opt} = \min_{x\in\mathbb{R}^n} \{f(x) : g_i(x) \leq 0,\, 1 \leq i \leq m\}$$

- **In late 1970's it was understood that**

— Convex Programming (f and g_i are convex) is "computationally tractable": under mild computability and boundedness assumptions, generic Convex Programming problems admit provably efficient solution algorithms.

— In contrast to this, typical generic nonconvex problems seem to be intractable: no provably efficient algorithms for these problems are known, and, unless P=NP, no such algorithms exist.

• **A generic convex problem:** **a family $\mathcal{P}$ of instances**

$$\mathrm{Opt}(p) = \min_{x\in\mathbb{R}^{n(p)}} \left\{ f_p(x) : g_{i,p}(x) \leq 0,\ 1 \leq i \leq m(p) \right\} \qquad (p)$$

such that

- **within $\mathcal{P}$, an instance p can be identified by its data vector** $\mathrm{Data}(p) \in \mathbb{R}^{N(p)}$

- **all instances $p \in \mathcal{P}$ are convex.**

Example: Linear Programming. **The objective and the constraints in (p) are affine functions of x, and**

$$\mathrm{Data}(p) = \left(m(p), n(p), \text{coefficients of } f_p, g_{1,p}, \ldots, g_{m(p),p} \right).$$

$$\mathcal{P} = \left\{ \mathrm{Opt}(p) = \min_{x \in \mathbb{R}^{n(p)}} \left\{ f_p(x) : g_{i,p}(x) \leq 0,\, 1 \leq i \leq m(p) \right\} \right\}$$

- **A solution algorithm for a generic problem $\mathcal{P}$: a code $\mathcal{B}$ for a Real Arithmetic computer which, given on input**
  - **the data $\mathrm{Data}(p)$ of an instance $p \in \mathcal{P}$,**
  - **a required accuracy $\epsilon > 0$,**

produces in finitely many operations of precise Real Arithmetics

- **either an ϵ-solution x_ϵ: $f_p(x_\epsilon) \leq \mathrm{Opt}(p) + \epsilon$ & $g_{i,p}(x_\epsilon) \leq \epsilon \,\forall i$,**
- **or a correct claim that p is infeasible/below unbounded.**

$$\mathcal{P} = \left\{ \mathrm{Opt}(p) = \min_{x\in\mathbb{R}^{n(p)}} \left\{ f_p(x) : g_{i,p}(x) \leq 0,\ 1 \leq i \leq m(p) \right\} \right\}$$

• **A solution algorithm is <u>efficient</u> ($\equiv$ polynomial time), if the # of operations is bounded by**

$$\mathbf{Poly}(\underbrace{\dim \mathrm{Data}(p)}_{\mathrm{Size}(p)}, \underbrace{\log\left(\frac{\mathrm{Size}(p) + \|\mathrm{Data}(p)\|_\infty}{\epsilon}\right)}_{\mathrm{Digits}(p,\epsilon)}).$$

• **Theorem.** Let $\mathcal{P}$ be a generic convex problem with instances

$$\mathrm{Opt}(p) = \min_{x\in\mathbb{R}^{n(p)}} \left\{ f_p(x) : g_{i,p}(x) \leq 0, i \leq m(p), \|x\|_\infty \leq 1 \right\} \qquad (p)$$

normalized by the requirement

$$\forall (x \in \mathbb{R}^{n(p)}, \|x\|_\infty \leq 1) : |f_p(x)| \leq 1,\ |g_{i,p}(x)| \leq 1,\ 1 \leq i \leq m(p).$$

There exists an explicit algorithm (Ellipsoid Method) which finds an ϵ-solution to (p), $0 < \epsilon < 1$, or detects correctly that (p) is infeasible, by computing $(0.2\epsilon/n(p))$-accurate approximations to the values and the subgradients of $f_p, g_{i,p}$ along $3n^2(p)\ln(2n(p)/\epsilon)$ successively generated search points, with additional $O(1)n(p)(n(p)+m(p))$ arithmetic operations per search point.

- **Corollary.** **Under**

  Computability Assumption: Given the data Data(p) **of an instance** $p \in \mathcal{P}$**, a tolerance** $\delta \in (0,1)$**, and** $x \in \mathbb{R}^{n(p)}$**,** $\|x\|_\infty \leq 1$**, the values and subgradients of** $f_p, g_{i,p}$ **at** x **can be computed within accuracy** δ **in** **Poly**(Size(p), Digits(p, δ)) **operations**

$\mathcal{P}$ **admits a polynomial time solution algorithm.**

• A convex problem always has a lot of structure (otherwise, how could we know that the problem is convex?)

• "Universal" polynomial time algorithms, like the Ellipsoid method, are black box oriented: they utilize detailed a priori knowledge of the structure and the data of a convex problem for the only purpose to compute the objective and the constrains at a point.

$\Rightarrow$ Poor (although polynomial time) performance: the arithmetic cost of accuracy digit is at least $O(n^4)$, which makes impossible to solve in realistic time problems with just few hundreds of variables...

• Question: How to reveal and to utilize the structure of a convex problem?

• Answer (the best known so far): to use conic reformulations of convex problems.

Conic Reformulation of a Convex Program

• **When passing from a Linear Programming program**

$$\min_{x\in\mathbb{R}^n}\left\{c^Tx : b - Ax \le 0\right\}$$

to convex ones, the traditional way is to replace linear objective c^Tx and linear left hand sides of the constraints with convex functions.

• **A more productive way is to pass from the coordinate-wise vector inequality $u \le v \Leftrightarrow v - u \in \mathbb{R}^n_+$ in $b - Ax \le 0$ to a more general vector inequality**

$$u \le_{\mathbf{K}} v \quad \Leftrightarrow \quad v - u \in \mathbf{K}$$

[$\mathbf{K} \subset \mathbb{R}^n$: convex pointed closed cone, $\mathrm{int}\,\mathbf{K} \neq \emptyset$]

thus arriving at convex programs in the conic form:

$$\min_{x\in\mathbb{R}^n}\left\{c^Tx : b - Ax \le_{\mathbf{K}} 0\right\}$$

• (c, A, b): **data** • $\mathbf{K}$: **structure**

• **Conic problem:**

$$\min_{x \in \mathbb{R}^n} \left\{ c^T x : Ax - b \geq_{\mathbf{K}} 0 \right\}$$

• **Every convex problem can be reformulated equivalently as a conic one. However: a general convex cone has no more structure than a general convex function. So what is the point?**

Fact: "Nearly all" interesting for applications convex problems are covered by just 3 generic conic problems:

• **Linear Programming**: $\mathbf{K}$ is a nonnegative orthant:

$$\min_x \left\{ c^T x : Ax - b \geq 0 \right\} \qquad \textbf{(LP)}$$

• **Conic Quadratic Programming**: $\mathbf{K}$ is a direct product of Lorentz cones $\mathbf{L}^n = \{x \in \mathbb{R}^n : x_n \geq \left(\sum_{i=1}^{n-1} x_i^2\right)^{1/2}\}$:

$$\min_x \left\{ c^T x : \|A_i x - b_i\|_2 \leq c_i^T x - d_i,\ 1 \leq i \leq m \right\} \qquad \textbf{(CQP)}$$

• **SemiDefinite Programming**: $\mathbf{K}$ is a direct product of semidefinite cones $\mathbf{S}_+^n = \{X = X^T \in \mathbf{R}^{n\times n} : X \succeq 0 [\Leftrightarrow x^T X x \geq 0 \forall x]\}$:

$$\min_x \left\{ c^T x : A_0 + \sum_{i=1}^n x_i A_i \succeq 0 \right\} \qquad \textbf{(SDP)}$$

Note: **LP** $\subset$ **CQP** $\subset$ **SDP**

• Good news about Conic Programming, especially LP/CQP/SDP:

- Fully symmetric and "algorithmic" duality allowing for instructive processing of conic programs "on paper" and heavily utilized by solution algorithms

- Existence of theoretically and practically powerful algorithms — Polynomial Time Interior Point Methods

- Extremely powerful "expressive abilities" of CQP/SDP
  $\Rightarrow$ huge spectrum of applications

Conic Duality

• **Duality in MP is about building lower bounds on the optimal value in an optimization program, i.e., about certifying negative statements "there is no feasible solution with the value of the objective $< ...$"**

• **For conic problems, Fenchel-Lagrange duality becomes fully symmetric and "algorithmic":**

$$(P): \quad \mathrm{Opt}(P) = \min_x \{c^T x : \overbrace{Ax - b}^{\xi} \geq_{\mathbf{K}} 0\}$$

$$\Updownarrow$$

$$\min_\xi \left\{ e^T \xi : \xi \in [\mathcal{L} - b] \bigcap \mathbf{K} \right\}$$

$$\left[e : A^T e = c, \mathcal{L} = \mathrm{Im} A \right]$$

$\Downarrow$ **[F.–L. Duality]**

$$\max_\lambda \left\{ b^T \lambda : \lambda \in [\mathcal{L}^\perp + e] \bigcap \mathbf{K}_* \right\}$$

$$\Updownarrow$$

$$(D): \quad \mathrm{Opt}(D) = \max_\lambda \left\{ b^T \lambda : A^T \lambda = c, \lambda \geq_{\mathbf{K}_*} 0 \right\}$$

$$\left[\mathbf{K}_* = \{\lambda : \lambda^T \xi \geq 0 \ \forall \xi \in \mathbf{K}\} \right]$$

$$\text{Opt}(P) = \min_x \{c^T x : Ax - b \geq_{\mathbf{K}} 0\} \qquad (P)$$
$$\text{Opt}(D) = \max_\lambda \{b^T\lambda : A^T\lambda = c, \lambda \geq_{\mathbf{K}_*} 0\} \quad (D)$$

Conic Duality Theorem:

• [Symmetry] **Conic duality is fully symmetric: (D) is a conic problem, and its dual is (equivalent to) (P)**

• [Weak Duality] $\text{Opt}(D) \leq \text{Opt}(P)$

• [Strong Duality] **If one of the problems (P), (D) is strictly feasible and bounded, then the other problem is solvable, and $\text{Opt}(D) = \text{Opt}(P)$.**

In particular, if both (P), (D) are strictly feasible, then both are solvable with equal optimal values, and a primal-dual feasible pair (x, λ) is primal-dual optimal iff

$$c^T x - b^T\lambda = 0 \qquad \Leftrightarrow \qquad [Ax - b]^T\lambda = 0.$$

• **Conic Duality**

$$\text{Opt}(P) = \min_x \{c^T x : Ax - b \geq_{\mathbf{K}} 0\} \qquad (P)$$

$$\text{Opt}(D) = \max_\lambda \{b^T \lambda : A^T \lambda = c, \lambda \geq_{\mathbf{K}_*} 0\} \quad (D)$$

is a special case of Lagrange Duality: If convex problem

$$\text{Opt}(\text{Pr}) = \min_x \{f(x) : g_i(x) \leq 0, 1 \leq i \leq m\}$$

is strictly feasible and bounded, then its Lagrange dual

$$\text{Opt}(\text{Dl}) = \max_{\lambda \geq 0} L(\lambda), \quad L(\lambda) \equiv \inf_x \{f(x) + \sum_i \lambda_i g_i(x)\}$$

is solvable, and $\text{Opt}(\text{Pr}) = \text{Opt}(\text{Dl})$**.**

In contrast to the Lagrange Duality, Conic Duality is

- **fully symmetric – (D) "remembers" (P).**

- **completely algorithmic – passing from (P) to (D) is a purely mechanical process.**

• Algorithmic nature of Convex Duality makes it a powerful tool for instructive analytical – "on paper" – processing conic programs.

Example: Truss Topology Design. A truss is a mechanical construction, like electric mast, railroad bridge, or Eiffel Tower, comprised of thin elastic bars linked to each other at nodes.
In a TTD problem, one is given

- a finite 2D/3D nodal set,
- a set of tentative bars – allowed pair connections of nodes,
- a set of loading scenarios – collections of forces acting at the nodes,

and looks for a construction of a given weight which is the most stiffest w.r.t. the scenario loads.

• **Stiffness of a truss w.r.t. a load is measured by compliance – the potential energy capacitated in the truss as a result of its deformation under the load (the less is compliance, the better).**

Mathematically, the TTD problem is

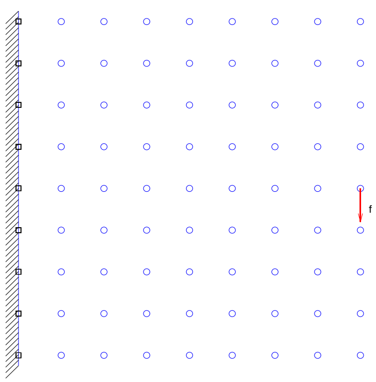
$$\min_{\tau,t_i}\left\{\tau : \left[\begin{array}{c|c} 2\tau & f_\ell^T \\ \hline f_\ell & \sum_{i=1}^N t_i b_i b_i^T \end{array}\right] \succeq 0,\ 1 \le \ell \le K, t \ge 0, \sum_i t_i \le W\right\}$$

• t_i: **bar volumes**

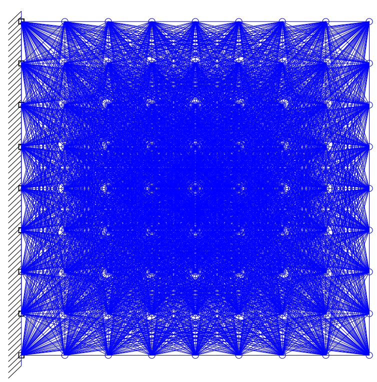
• $f_\ell \in \mathbb{R}^M$: **loads (**M: **total # of nodal degrees of freedom)**

• τ: **upper bound on the worst-case, w.r.t. loads** f_ℓ, $1 \le \ell \le K$, **compliance**

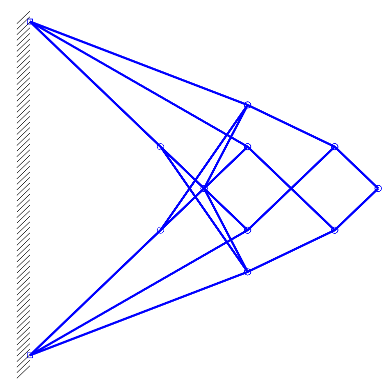
• **In TTD, one starts with a "dense" nodal grid and allows for all pair connections of nodes by bars. At the optimum, most of the bars get zero volume, thus revealing the optimal topology:**



9×9 **nodal set** **(**$M = 144$**) and loading scenario**



81 tentative nodes and 2,039 tentative bars



optimal console 12 nodes, 32 bars

• **In order to capture topology design, one should work with dense grids (M of order of few thousands)**
⇒ **The design dimension $N = O(M^2)$ of the TTD is in the range of millions...**

• **Cure: Semidefinite Duality. In the dual of TTD, most of the variables can be eliminated analytically, which results in the problem of dimension $\approx MK << N = O(M^2)$:**

$$\min_{\alpha,v,\gamma}\left\{-2\sum_\ell f_\ell^T v_\ell + W\gamma : \begin{array}{c}\left[\begin{array}{c|ccc}\gamma & b_i^T v_1 & \dots & b_i^T v_K\\ \hline b_i^T v_1 & \alpha_1 & & \\ \vdots & & \ddots & \\ b_i^T v_K & & & \alpha_K\end{array}\right]\succeq 0\ \forall i\\ 2\sum_\ell \alpha_\ell = 1\end{array}\right\}$$

• **Taking dual to the (processed!) dual of TTD, we end up with instructive (and unexpected) equivalent bar-stress reformulation of the TTD problem:**

$$\min_{\tau,t_i}\left\{\tau:\ \left[\begin{array}{c|c}2\tau & f_\ell^T\\ \hline f_\ell & \sum_{i=1}^N t_ib_ib_i^T\end{array}\right]\succeq 0\,\forall\ell,\ t\geq 0,\sum_i t_i\leq W\right\}$$

$$\Rightarrow\ \min_{\alpha,v,\gamma}\left\{-2\sum_\ell f_\ell^Tv_\ell+W\gamma:\ \begin{array}{c}\left[\begin{array}{c|ccc}\gamma & b_i^Tv_1 & \dots & b_i^Tv_K\\ \hline b_i^Tv_1 & \alpha_1 & & \\ \vdots & & \ddots & \\ b_i^Tv_K & & & \alpha_K\end{array}\right]\succeq 0\,\forall i\\ 2\sum_\ell \alpha_\ell=1\end{array}\right\}$$

$$\Rightarrow\ \min_{\tau,t,q}\left\{\tau:\ \sum_i\frac{q_{i\ell}^2}{2t_i}\leq\tau,\ \sum_i q_{i\ell}b_i=f_\ell\,\forall\ell,\ t\geq 0,\sum_i t_i\leq W\right\}$$

Exploiting Structure of a Convex Program: Polynomial Time Interior Point Methods

1979: polynomial time solvability of LP (Khachiyan) via the Ellipsoid Method

1984: the first IPM for LP (Karmarkar): theoretical efficiency + practical performance competitive with the one of the Simplex Method

1986: first polynomial path-following IPMs for LP (Renegar, Gonzaga): improved complexity bounds + transparent construction with potential for nonlinear extensions

"Interior Point Revolution" (mid-1980's – late 1990's):

- nonlinear extensions: general theory of IPMs in Convex Programming
- advanced theory (Nesterov & Todd, 1997-98) of IPMs for conic problems on symmetric cones (LP/CQP/SDP)

Polynomial Time IPMs: Path-Following Scheme

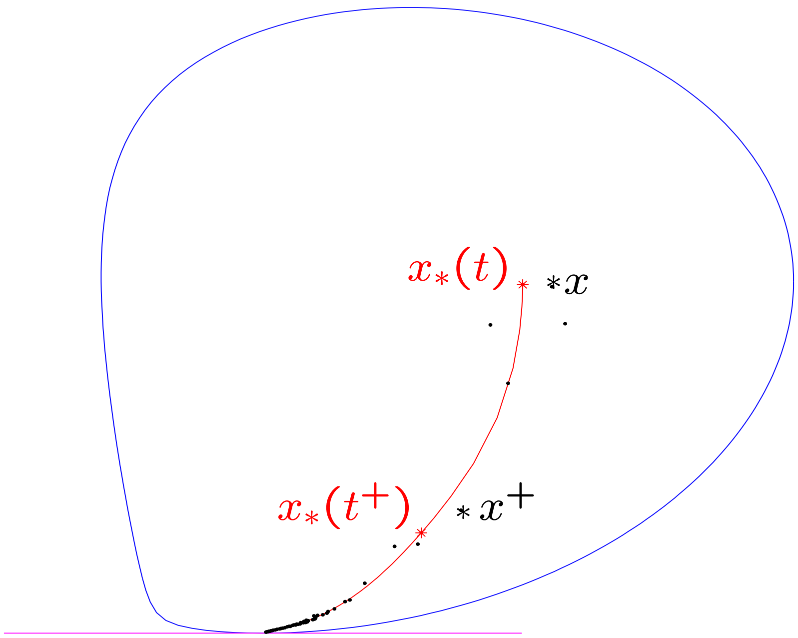
• **Path-Following Scheme (Fiacco & McCormic, 1968)** for solving convex program

$$\min_x \left\{ c^T x : x \in G \right\}$$

• Equip G with a **barrier** – a C^2 function $F : \mathrm{int}\, G \to \mathbb{R}$ with $F''(\cdot) \succ 0$ and closed level sets $\{x \in \mathrm{int}\, G : F(x) \le a\}$;

• Trace the **path** $x_*(t) = \mathrm{argmin}_{x \in \mathrm{int}\, G} F_t(x), \quad F_t(x) = tc^T x + F(x)$ as the **penalty parameter** $t \to \infty$:

Given (x, t) with x close to $x_*(t)$,
• **replace t with $t^+ > t$;**
• **minimize $F_{t^+}(\cdot)$ by Newton method, x being the starting point, until a point x^+ close to $x_*(t^+)$ is built;**
• **replace $(x, t) \leftarrow (x^+, t^+)$ and loop**



• **It was discovered in late 1980's that the path-following scheme becomes polynomial when specific self-concordant barriers are used:**

> **Let $G \subset \mathbb{R}^n$ be a convex domain. A $\mathbf{C}^3$ convex function $F : \mathrm{int}\, G \to \mathbb{R}$ is called a ϑ-self-concordant barrier for G, if F is a barrier for G and $\forall (x \in \mathrm{int}\, G, h \in \mathbb{R}^n)$:**
>
> **A.** [self-concordance] $\left|D^3F(x)[h,h,h]\right| \leq 2\left(D^2F(x)[h,h]\right)^{3/2}$
>
> **B.** [s.-c.b. quantification] $|DF(x)[h]| \leq \vartheta^{1/2}\left(D^2F(x)[h,h]\right)^{1/2}$

Interpretation: $D^2F(x)$ **defines a local Euclidean metrics**

$$\|h\|_x = \left(D^2F(x)[h,h]\right)^{1/2}.$$

A, B mean that $D^2F(\cdot)$ and $F(\cdot)$ are Lipschitz continuous w.r.t. this local metrics with constants 2 and $\vartheta^{1/2}$, respectively.

Theorem. Let $G \subset \mathbb{R}^n$ be a closed convex domain not containing lines, $c \in \mathbb{R}^n$ be such that the level sets $\{x \in G : c^T x \leq a\}$ are bounded, and F be a ϑ-s.-c.b. for G. Then

(i) The path $x_*(t) = \operatorname*{argmin}_{\text{int}\, G} F_t(x)$, $F_t(x) = tc^T x + F(x)$, $t > 0$, is well-defined

(ii) Let us say that (x, t) is close to the path, if $t > 0$ and

$$\lambda(x,t) \equiv \left([\nabla F_t(x)]^T [\nabla^2 F_t(x)]^{-1} \nabla F_t(x)\right)^{1/2} \leq 0.1.$$

Given (x_0, t_0) close to the path, consider the recurrence

$$\begin{bmatrix} t_{i-1} \\ x_{i-1} \end{bmatrix} \mapsto \begin{bmatrix} t_i &=& \exp\{0.1/\sqrt{\vartheta}\} t_{i-1} \\ x_i &=& x_{i-1} - \frac{1}{1+\lambda(x_{i-1},t_i)} [\nabla^2 F_{t_i}(x_{i-1})]^{-1} \nabla F_{t_i}(x_{i-1}) \end{bmatrix}$$

Then all (x_i, t_i) are well-defined and close to the path, and

$$\forall i : c^T x_i - \min_G c^T x \leq 2\vartheta t_i^{-1} = 2\vartheta t_0^{-1} \exp\{-0.1 i/\sqrt{\vartheta}\}.$$

Thus, every $O(1)\sqrt{\vartheta}$ steps add an accuracy digit.

• Conclusion: When we are smart enough to equip the feasible domain G of a convex problem $\min\limits_{x\in G} c^T x$ with an efficiently computable ϑ-s.-c.b. F with not too large ϑ, we get a polynomial time IPM for solving the problem.

Note: Every convex domain $G \subset \mathbb{R}^n$ admits $O(n)$-s.-c.b. E.g., when G is a pointed cone, we can set

$$F(x) = O(1) \log \int\limits_{G_*} \exp\{-x^T \xi\} d\xi$$

- "Good" – efficiently computable – s.-c.b.'s are known for a wide variety of "basic" convex domains
- All standard convexity-preserving operations can be equipped with simple rules to combine good s.-c.b.'s for the operands into a good s.-c.b. for the result.

$\Rightarrow$ Essentially, the entire Convex Programming is within the grasp of polynomial time IPMs.

• **The Interior Point constructions become maximally flexible as applied to conic problems on cones with many symmetries, most notably on homogeneous self-dual cones, which covers LP/CQP/SDP. The related theory is intrinsically linked to the theory of Euclidean Jordan Algebras.**

In LP/CQP/SDP, one uses the self-concordant barriers as follows:

$\mathbf{K}$	$F_{\mathbf{K}}$	ϑ
$\mathbb{R}_+$	$-\ln(x)$	1
$\mathbf{L}^n$	$-\ln(x_n^2 - \sum_{i=1}^{n-1} x_i^2)$	2
$\mathbf{S}_+^n$	$-\ln\det X$	n
$\mathbf{K}_1 \times ... \times \mathbf{K}_m$	$F_{\mathbf{K}_1}(x^1) + ... + F_{\mathbf{K}_m}(x^m)$	$\sum_i \vartheta(F_{\mathbf{K}_i})$

and solves simultaneously the problem of interest and its dual ("primal-dual IPMs").

• Primal-dual LP/CQP/SDP IPMs underly the best known so far complexity bounds for these generic problems and, in addition, allow for

- on-line adjustable "long step" path-tracing policies
  $\Rightarrow$ in practice, much faster convergence than for the "off-line" worst-case-oriented penalty updating rule, with no risk to violate the theoretical complexity bound

- an elegant way to initialize path-tracing

- building infeasibility/unboundedness certificates,...

• Practical performance of primal-dual IPMs for LP/CQP/SDP is usually much better than the one predicted by the worst-case-oriented theoretical complexity analysis.

Challenge: **On extremely large-scale CQP/SDP problems ($10^4 - 10^6$ design variables), IPMs become too time-consuming. What to do?**

• **At present, the best we can do in the extremely large-scale case is to use computationally cheap methods with (nearly) dimension-independent polynomial in $1/\epsilon$ (not in $\ln(1/\epsilon)$!) complexity.**

Expressive Abilities and Applications of CQP/SDP

Fact: Let $\mathcal{F}$ be a family of cones closed w.r.t. taking direct products and passing from a cone to its dual.
There exists a simple notion of an "$\mathcal{F}$-representation" of a convex function/set such that

• Problem $\min\limits_x \{f(x) : g_i(x) \leq 0, 1 \leq i \leq m, x \in X\}$ can be easily reduced to a conic program on a cone from $\mathcal{F}$, provided that we know $\mathcal{F}$-representations of f, g_i, X.

• $\mathcal{F}$-representations admit a simple "calculus" which shows that the results of all standard convexity-preserving operations with functions/sets are $\mathcal{F}$-representable, provided that the operands are so.

• Calculus of conic representations is independent of $\mathcal{F}$ and fully algorithmic: $\mathcal{F}$-representation of the result of an operation is readily given by $\mathcal{F}$-representations of the operands.

• **Applying Calculus of conic representations to "elementary" LP/CQP/ SDP-representable functions/sets, we get a possibility to recognize convex problems which can be converted to LP/CQP/SDP.**

(CQP) / **(SDP)**

$$\min c^T x \text{ subject to}$$

$$(a): \quad Ax = b,\ x \ge 0$$

$$(b): \quad \left(\sum_{i=1}^{8} |x_i|^3\right)^{1/3} \le x_2^{1/7} x_3^{2/7} x_4^{3/7} + 2 x_1^{1/5} x_5^{2/5} x_6^{1/5}$$

$$5x_2 \ge x_1^{-1/2} x_2^{-2} + 2 x_2^{-1/3} x_3^{-3} x_4^{-5/8}$$

$$\begin{bmatrix} x_2 & x_1 & & \\ x_1 & x_4 & x_3 & \\ & x_3 & x_6 & x_3 \\ & & x_3 & x_8 \end{bmatrix} \preceq 5I$$

$$(c): \quad \begin{bmatrix} x_1 & x_2 - x_1 & x_3 - x_2 & x_4 - x_3 \\ x_2 - x_1 & x_2 & x_3 - x_2 & x_4 - x_3 \\ x_3 - x_2 & x_3 - x_2 & x_3 & x_4 - x_3 \\ x_4 - x_3 & x_4 - x_3 & x_4 - x_3 & x_4 \end{bmatrix} \succeq 0$$

$$x_1 + x_2 \sin(\phi) + x_3 \sin(2\phi) + x_4 \sin(4\phi) \ge 0,\ 0 \le \phi \le \pi/2$$

- **The problem can be systematically converted into SDP**

$$\min_{z \in \mathbb{R}^{98}} \{ d^T z : A_0 + \sum_i z_i A_i \succeq 0 \}$$

- **Removing constraints (c), the problem becomes a CQP**

$$\min_{w \in \mathbb{R}^{48}} \{ e^T w : \|P_i w + p_i\|_2 \le q_i^T w + r_i, 1 \le i \le m \}$$

and can be reduced, in a polynomial time fashion, to LP.

Expressive Abilities of CQP

$$\min_x \left\{ c^T x : \|A_i x - b_i\|_2 \leq c_i^T x - d_i,\ 1 \leq i \leq m \right\} \qquad \textbf{(CQP)}$$

- **Sample of CQP-representable functions/sets:**
  - $\|\cdot\|_p$, $p \in \mathbf{Q}$
    $\Rightarrow$ **Approximation in $\|\cdot\|_p$**
  - **convex quadratic forms**
    $\Rightarrow$ **Convex Quadratic Programming**
  - **power monomials** $-x_1^{p_1} x_2^{p_2} \ldots x_n^{p_n}$, $x \geq 0$ ($p_i \in \mathbf{Q}_+, \sum_i p_i \leq 1$),
    power monomials $x_1^{-p_1} x_2^{-p_2} \ldots x_n^{-p_n}$, $x > 0$ ($p_i \in \mathbf{Q}_+$)
    $\Rightarrow$ **Geometric Programming in algebraic form**
  - $f(x,t) = x^T (B^T \mathrm{Diag}\{t\} B)^{-1} x$, $t \in \mathbb{R}^n_{++}$
    $\Rightarrow$ **Truss Topology/Electric Circuit Design**

Expressive abilities of LP

Theorem. The Lorentz cones admit fast polyhedral approximation. Specifically, for every $\epsilon \in (0, 0.1)$ and every n, one can point out

- a polyhedral cone $\mathbf{P} \subset \mathbb{R}^{\rfloor 2n \ln(1/\epsilon) \lfloor}$ given by an explicit system of $\rfloor 5n \ln(1/\epsilon) \lfloor$ homogeneous linear inequalities, and
- an explicit linear mapping $\mathcal{M} : \mathbb{R}^{\rfloor 2n \ln(1/\epsilon) \lfloor} \to \mathbb{R}^n$

such that $\mathcal{M}(\mathrm{P})$ is in-between L^n and the "$(1+\epsilon)$-extension" of L^n:

$$\{\underbrace{(v,t)}_{\in \mathbb{R}^n} : \|v\|_2 \le t\} = \mathbf{L}^n \subset \mathcal{M}(\mathbf{P}) \subset \mathbf{L}^n_\epsilon := \{\underbrace{(v,t)}_{\in \mathbb{R}^n} : \|v\|_2 \le (1+\epsilon)t\}$$

$\Rightarrow$ CQP can be reduced, in a polynomial time fashion, to LP.

Expressive Abilities of SDP

$$\min_x \{c^T x : \sum_i x_i A_i \succeq B\} \qquad \textbf{(SDP)}$$

- **Sample of SDP-representable functions/sets:**
  - **All CQP-representable functions/sets**
  - **Symmetric functions of eigenvalues of symmetric matrices/singular values of rectangular matrices**

Theorem. **Let $f(x) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be symmetric and SDP-representable. Then $F(X) = f(\lambda_1(X), ..., \lambda_n(X)) : \mathbf{S}^n \to \mathbb{R} \cup \{+\infty\}$ is SDP-representable as well.**

• The cones of (coefficients of) univariate algebraic/trigonometric polynomials of a given degree nonnegative on a given segment

<u>Theorem.</u> For a segment $\Delta \subset \mathbb{R}$, the sets

$$\{(A_0, ..., A_d) \in (\mathbf{S}^n)^{d+1} : A_0 + tA_1 + ... + t^d A_d \succeq 0 \,\forall t \in \Delta\}$$

are SDP-representable with explicit SDP representations.

$\Rightarrow$ Minimization of a univariate algebraic/trigonometric polynomial over a segment is an SDP program.

• Due to its tremendous expressive abilities, SDP has a wide variety of applications, including those in

- Relaxations of difficult combinatorial problems
- Ellipsoidal approximations of convex sets
- Statistics
- Robust Control
- Structural Design
- Communications
- Signal Processing,...

Permanent challenge: Extending the scope of applications
– building SDP models for various problems of Engineering and Management

• <u>Example:</u> **Semidefinite Relaxations of difficult problems**

A (nonconvex) quadratically constrained quadratic problem

$$\begin{array}{l}\mathrm{Opt} = \max\limits_x \{f_0(x) : f_i(x) \leq 0,\, 1 \leq i \leq m\},\\ f_i(x) = x^T A_i x + 2b_i^T x + c_i\end{array} \qquad (*)$$

can be NP-hard. E.g., $x_j^2 = x_j \Leftrightarrow x_j \in \{0, 1\}$.

• **Passing to the matrix variable** $X = \begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix}$, $(*)$ **becomes**

$$\max_X \left\{ \mathrm{Tr}(\mathcal{A}_0 X) : \begin{array}{c} \mathrm{Tr}(\mathcal{A}_i X) \leq 0,\, 1 \leq i \leq m, \\ X_{11} = 1, X \succeq 0 \\ \mathrm{Rank}(X) = 1 \end{array} \right\} \qquad \left[\mathcal{A}_i = \begin{bmatrix} c_i & b_i^T \\ b_i & A_i \end{bmatrix}\right]$$

Eliminating the "troublemaking" rank constraint, we arrive at the SDP relaxation of $(*)$

$$[\mathrm{Opt} \leq] \quad \mathrm{SDP} = \max_X \left\{ \mathrm{Tr}(\mathcal{A}_0 X) : \begin{array}{c} \mathrm{Tr}(\mathcal{A}_i X) \leq 0, 1 \leq i \leq m, \\ X \succeq 0, X_{11} = 1 \end{array} \right\}$$

$$\text{Opt} = \max_x \left\{ f_0(x) : f_i(x) \le 0,\ 1 \le i \le m \right\},\ f_i(x) = x^T A_i x + 2b_i^T x + c_i$$

$$\Downarrow$$

$$[\text{Opt} \le]\ \text{SDP} = \max_X \left\{ \text{Tr}(\mathcal{A}_0 X) : \begin{array}{c} \text{Tr}(\mathcal{A}_i X) \le 0, 1 \le i \le m, \\ X \succeq 0, X_{11} = 1 \end{array} \right\}$$

- **Interpretation:** **In the relaxation, we maximize the expected value of the original objective over random solutions satisfying at average the original constraints.**

• **In good cases, SDP relaxations yield provably tight bounds.**

Example: **It is NP-hard to compute**

$$\begin{array}{rcl} \text{Opt} & = & \max_x\{x^TAx : \|x\|_\infty \leq 1\} \equiv \max\limits_x \left\{x^TAx : x_i^2 \leq 1,\, 1 \leq i \leq n\right\} \\ & \leq & \text{SDP} = \max_X \{\text{Tr}(AX) : X_{ii} \leq 1, 1 \leq i \leq n,\, X \succeq 0\} \end{array}$$

even when 4% accuracy is sought. However:

• A **is diagonal-dominated with** $A_{ij} \leq 0$ **for** $i \neq j$

$\Rightarrow$ Opt $\leq$ SDP $\leq 1.1382\,$Opt **[Goemans & Williamson, '95]**

• $A \succeq 0 \Rightarrow$ Opt $\leq$ SDP $\leq \frac{\pi}{2}$Opt **[Nesterov, '98]**

$\Rightarrow$ **Tight approximations of matrix norms**

When $p > 2 > r \geq 1$, SDP yields a computable upper bound on the (computationally intractable!) matrix norm $\|A\|_{pr} = \max\{\|Ax\|_r : \|x\|_p \leq 1\}$ tight within factor $\theta(p,r) \leq \frac{3\pi}{6\sqrt{3}-2\pi} = 2.2936...$ [cf. the Grothendieck inequality ('53) dealing with $p = \infty, r = 1$; here the constant can be improved to $\frac{\pi}{2\ln(1+\sqrt{2})} \approx 1.7822...$]

• $\forall A$: $\mathrm{Opt} \leq \mathrm{SDP} \leq O(1)\ln(n+1)\mathrm{Opt}$ **(valid with the unit box in $\mathbb{R}^n$ replaced by intersection of n centered at the origin ellipsoids in $\mathbb{R}^m$).**

Challenge: Processing CQP/SDP with uncertain data

• More often than not, the data in real life problems are uncertain – not known exactly when problem is solved. When "immunizing" solutions against data uncertainty, one ends up with

- **semi-infinite** conic constraints $\forall([A,b]\in\mathcal{U}): Ax-b\geq_{\mathbf{K}} 0$ ("uncertain-but-bounded" data perturbations), or
- **chance** conic constraints $\text{Prob}_{[A,b]\sim P}\{[A,b]: Ax-b\not\geq_{\mathbf{K}} 0\}\leq\epsilon$ (stochastic data perturbations).

Fact: When $\mathbf{K}$ is a Lorentz or a Semidefinite cone, the above constraints typically are computationally intractable

• **How to build "reasonably tight" computationally tractable sufficient conditions for the validity of semi-infinite and the chance conic constraints?**