Primal-Dual Geometry of Level Sets and their Explanatory Value of the Practical Performance of Interior-Point Methods for Conic Optimization

Robert M. Freund
M.I.T.

November, 2009

from papers in *SIOPT*, *Mathematics of Operations Research*, and *Mathematical Programming*

Outline

- Primal-Dual Geometry of Level Sets in Conic Optimization

- A Geometric Measure of Feasible Regions and Interior-Point Method (IPM) Complexity Theory

- Geometric Measures and their Explanatory Value of the Practical Performance of IPMs

Primal and Dual Linear Optimization Problems

$$
\begin{array}{llll}
P: \ \mathsf{VAL} := \min_x & c^T x & \qquad D: \ \mathsf{VAL} := \max_{y,z} & b^T y \\
\text{s.t.} & Ax = b & \text{s.t.} & A^T y + z = c \\
 & x \geq 0 & & z \geq 0
\end{array}
$$

$A \in \mathbb{R}^{m \times n}$

"$x \geq 0$" is "$x \in \Re^n_+$"

Primal and Dual Conic Problem

$$P: \ \text{VAL}_* := \min_x \quad c^T x \qquad\qquad D: \ \text{VAL}^* := \max_{y,z} \ b^T y$$

$$\text{s.t.} \quad Ax = b \qquad\qquad \text{s.t.} \quad A^T y + z = c$$

$$x \in C \qquad\qquad z \in C^*$$

$C \subset X$ is a regular cone: closed, convex, pointed, with nonempty interior

$C^* := \{z : z^T x \geq 0 \ \forall x \in C\}$

The Semidefinite Cone

S^k denotes the set of symmetric $k \times k$ matrices

S^k_+ denotes the set of positive semi-definite $k \times k$ symmetric matrices

S^k_{++} denotes the set of positive definite $k \times k$ symmetric matrices

"$X \succeq 0$" denotes that X is symmetric positive semi-definite

"$X \succeq Y$" denotes that $X - Y \succeq 0$

"$X \succ 0$" to denote that X is symmetric positive definite, etc.

Remark: $S^k_+ = \{X \in S^k \mid X \succeq 0\}$ is a regular convex cone. Furthermore, $(S^k_+)^* = S^k_+$, i.e., S^k_+ is self-dual.

Primal and Dual Semidefinite Optimization

$$P: \ \mathsf{VAL}_* := \mathsf{min}_x \quad c^T x \qquad\qquad D: \ \mathsf{VAL}^* := \mathsf{max}_{y,z} \quad b^T y$$

$$\text{s.t.} \quad Ax = b \qquad\qquad \text{s.t.} \quad A^T y + z = c$$

$$x \in S_+^k \qquad\qquad z \in S_+^k$$

This is slightly awkward as we are not used to conceptualizing x as a matrix.

Primal and Dual Semidefinite Optimization, again

$\min_X$	$C \bullet X$	$\max_{y,Z}$	$\sum_{i=1}^m y_i b_i$
s.t.	$A_i \bullet X = b_i \ , i = 1, \ldots, m$	s.t.	$\sum_{i=1}^m y_i A_i + Z = C$
	$X \succeq 0$		$Z \succeq 0$

Here $C \bullet X := \displaystyle\sum_{i=1}^{k} \sum_{j=1}^{k} C_{ij} X_{ij}$ is the "trace inner product,"

since $C \bullet X = \text{trace}(C^T X)$

Back to Linear Optimization

$$
\begin{array}{llr}
P: & \mathrm{VAL} := \min_x & c^T x \\
 & \text{s.t.} & Ax = b \\
 & & x \geq 0
\end{array}
\qquad
\begin{array}{llr}
D: & \mathrm{VAL} := \max_{y,z} & b^T y \\
 & \text{s.t.} & A^T y + z = c \\
 & & z \geq 0
\end{array}
$$

Meta-Lessons from Interior-Point Theory/Methods

- Linear optimization is not much more special than conic convex optimization

- A problem is ill-conditioned if VAL is finite but the primal or dual objective function level sets are unbounded

- ε-optimal solutions are important objects in their own right

- Choice of norm is important; some norms are more natural for certain settings

Meta-Lessons from Interior-Point Theory/Methods, continued

- All the important activity is in the (regular) cones

Indeed, we could eliminate the y-variable and re-write P and D as:

$$
\begin{array}{llll}
P: & \min_x & c^T x \\
 & \text{s.t.} & x - x^0 \in L \\
 & & x \geq 0
\end{array}
\qquad
\begin{array}{lll}
D: & \text{VAL} := \min_z & (x^0)^T z \\
 & \text{s.t.} & z - c \in L^\perp \\
 & & z \geq 0
\end{array}
$$

where x^0 satisfies $Ax^0 = b$ and $L = \text{null}(A)$.

But we won't.

Primal and Dual Near-Optimal Level Sets

$$P: \ \mathsf{VAL} := \min_x \quad c^T x \qquad\qquad D: \ \mathsf{VAL} := \max_{y,z} \quad b^T y$$

$$\text{s.t.} \quad \begin{array}{r} Ax = b \\ x \geq 0 \end{array} \qquad\qquad \text{s.t.} \quad \begin{array}{r} A^T y + z = c \\ z \geq 0 \end{array}$$

$$P_\varepsilon := \{x : Ax = b, x \geq 0, c^T x \leq \mathsf{VAL} + \varepsilon\}$$

$$D_\delta := \{z : \exists y \text{ satisfying } A^T y + z = c, z \geq 0, b^T y \geq \mathsf{VAL} - \delta\}$$

Level Set Geometry Measures

Let $e := (1, 1, \ldots, 1)^T$. Define for $\varepsilon, \delta > 0$:

$$R^P_\varepsilon := \max_x \ \|x\|_1 \quad \text{s.t.} \quad Ax = b, \quad x \geq 0, \quad c^T x \leq \text{VAL} + \varepsilon$$

$$r^D_\delta := \max_{y,z,r} \ r \quad \text{s.t.} \quad A^T y + z = c, \quad z \geq 0, \quad b^T y \geq \text{VAL} - \delta, \quad z \geq r \cdot e$$

R^P_ε is the norm of the largest primal ε-optimal solution

r^D_δ measures the largest distance to the boundary of $\Re^n_+$ among all dual δ-optimal solutions z

Level Set Geometry Measures, continued

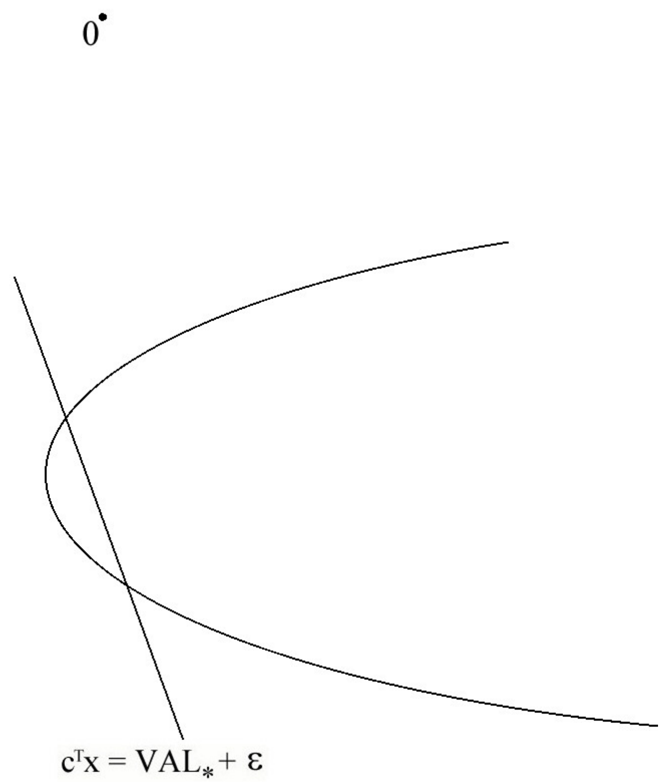
$$R_\varepsilon^P := \max_x \ \|x\|_1 \quad \text{s.t.} \quad x \in P_\varepsilon$$

$$r_\delta^D := \max_{z,r} \ r \quad \text{s.t.} \quad z \in D_\delta, \ z \geq r \cdot e$$

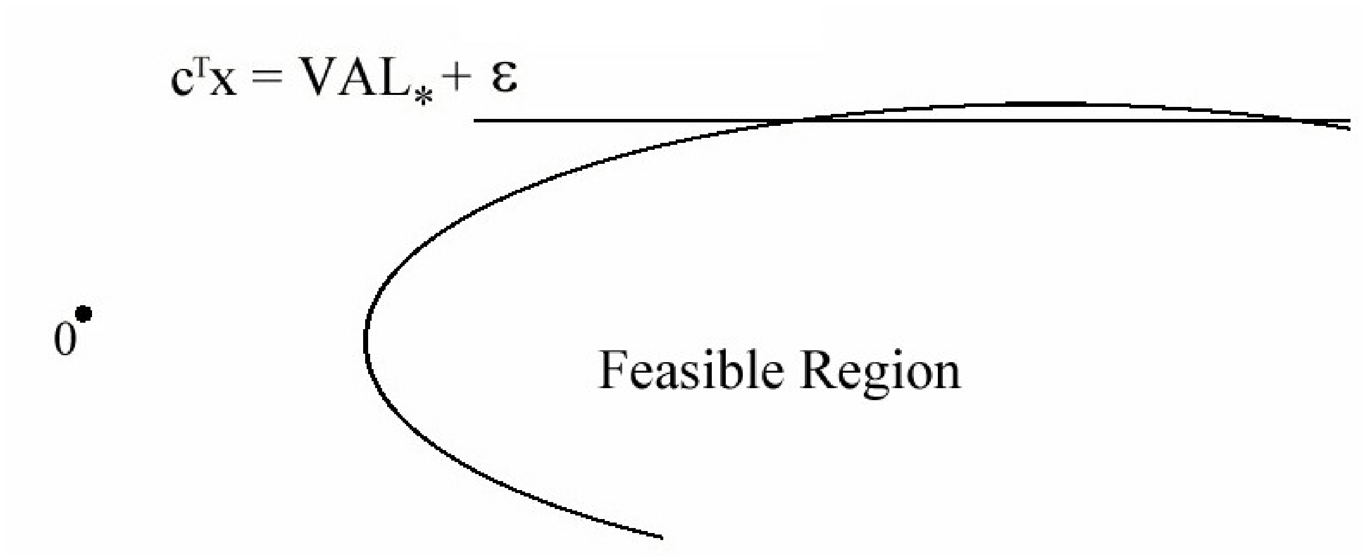
$$P_\varepsilon := \{x : Ax = b, x \geq 0, c^T x \leq \mathrm{VAL} + \varepsilon\}$$

$$D_\delta := \{z : \exists y \text{ satisfying } A^T y + z = c, z \geq 0, b^T y \geq \mathrm{VAL} - \delta\}$$

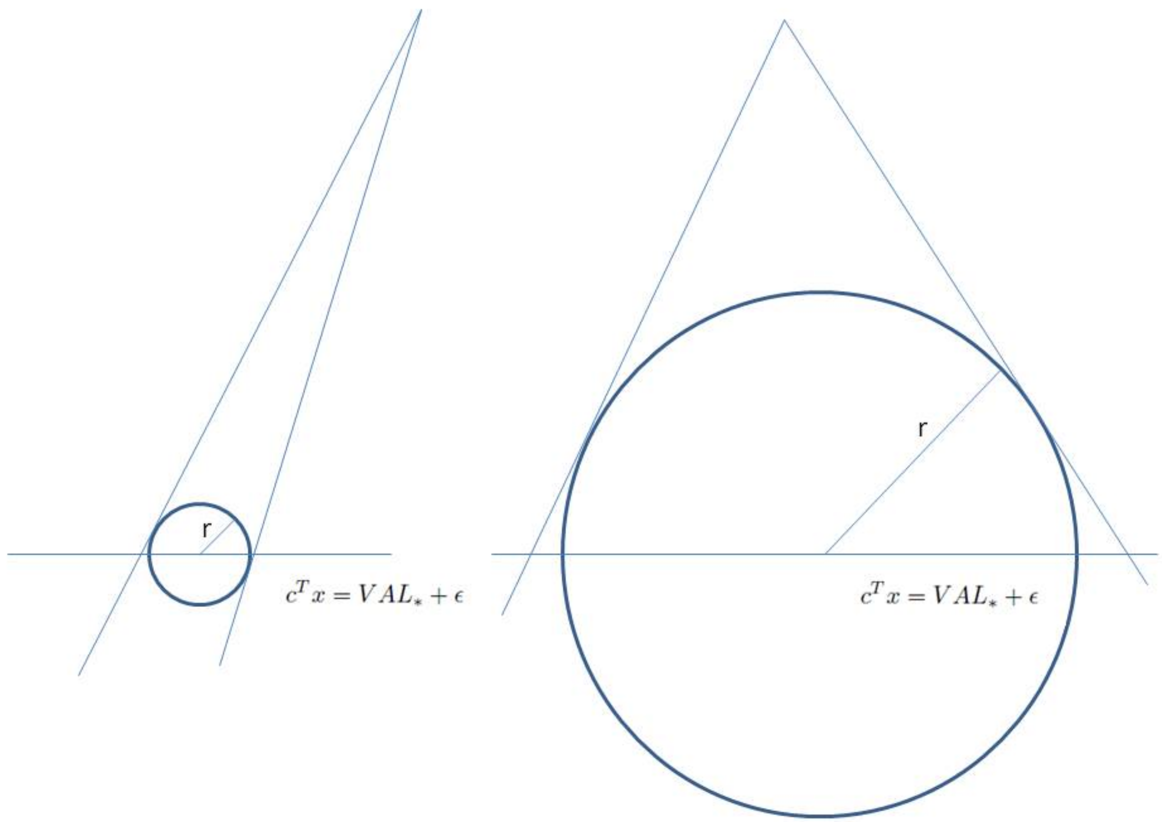
R^P_ε Measures Large Near-Optimal Solutions



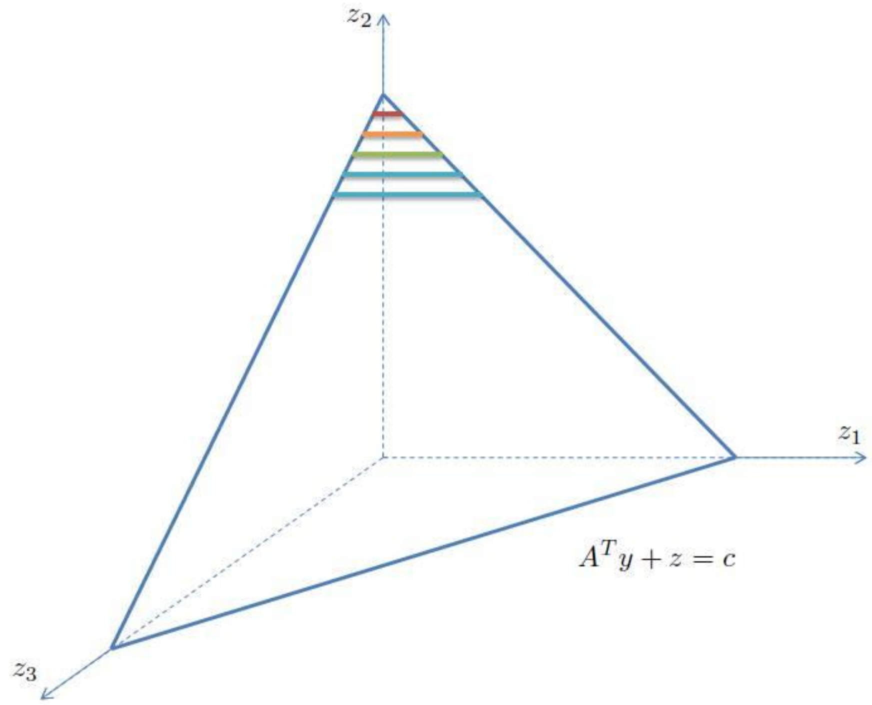
R_ε^P Measures Large Near-Optimal Solutions



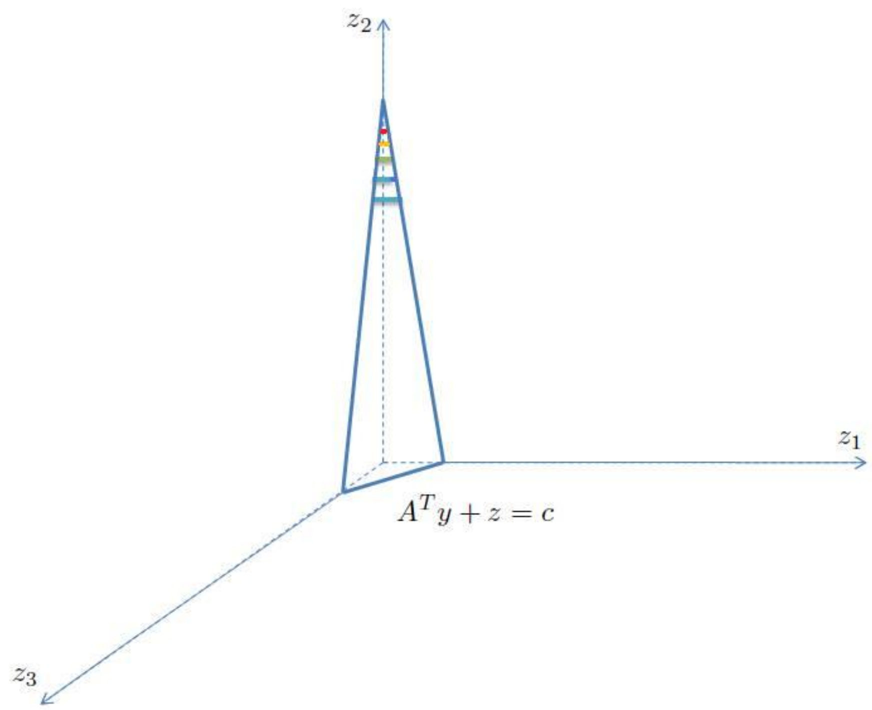
r_δ^D Measures Nicely Interior Near-Optimal Solutions



r_δ^D Measures Nicely Interior Near-Optimal Solutions



r_δ^D Measures Nicely Interior Near-Optimal Solutions



Main Result: R^P_ε and r^D_δ are Reciprocally Related

$$R^P_\varepsilon := \ \max_x \ \|x\|_1 \qquad\qquad r^D_\delta := \ \max_{z,r} \ r$$

$$\text{s.t.} \quad x \in P_\varepsilon \qquad\qquad\qquad \text{s.t.} \quad z \in D_\delta$$
$$z \geq r \cdot e$$

Main Theorem: Suppose VAL is finite. If R^P_ε is positive and finite, then

$$\min\{\varepsilon, \delta\} \leq R^P_\varepsilon \cdot r^D_\delta \leq \varepsilon + \delta \ .$$

Otherwise $\{R^P_\varepsilon, \ r^D_\delta\} = \{\infty, \ 0\}$.∎

Comments

$$\min\{\varepsilon, \delta\} \leq R^P_\varepsilon \cdot r^D_\delta \leq \varepsilon + \delta$$

- R^P_ε, r^D_δ each involves primal and dual information

- each inequality can be tight (and cannot be improved)

- setting $\delta = \varepsilon$, we obtain $\varepsilon \leq R^P_\varepsilon \cdot r^D_\varepsilon \leq 2\varepsilon$, showing these two measures are inversely proportional (to within a factor of 2)

Comments, continued

$$\min\{\varepsilon, \delta\} \leq R_\varepsilon^P \cdot r_\delta^D \leq \varepsilon + \delta$$

- exchanging the roles of P and D

- how to prove

Comments, continued

$$R^P_\varepsilon := \max_x \ \|x\|_1 \quad \text{s.t.} \quad x \in P_\varepsilon \qquad\qquad r^D_\delta := \max_{z,r} \ r \quad \text{s.t.} \quad z \in D_\delta, \ z \geq r \cdot e$$

$$\varepsilon \leq R^P_\varepsilon \cdot r^D_\delta \leq 2\varepsilon$$

"The maximum norms of the primal objective level sets are almost exactly inversely proportional to the maximum distances to the boundary of the dual objective level sets"

Relation to LP Non-Regularity Property

Standard LP Non-Regularity Property: If VAL is finite, the set of primal optimal solutions is unbounded iff every dual feasible z lies in the boundary of $\Re^n_+$. ∎

$$
\begin{array}{rll}
R_\varepsilon^P := & \max_x & \|x\|_1 \\
& \text{s.t.} & x \in P_\varepsilon
\end{array}
\qquad
\begin{array}{rll}
r_\delta^D := & \max_{z,r} & r \\
& \text{s.t.} & z \in D_\delta \\
& & z \geq r \cdot e
\end{array}
$$

In our notation, this is $R_\varepsilon^P = \infty$ iff $r_\delta^D = 0$, which is the second part of the **Main Theorem**

Relation to LP Non-Regularity, continued

$$R_\varepsilon^P := \text{max}_x \ \ \|x\|_1 \qquad\qquad r_\delta^D := \text{max}_{z,r} \ \ r$$

$$\text{s.t.} \quad x \in P_\varepsilon \qquad\qquad\qquad\qquad \text{s.t.} \quad z \in D_\delta$$

$$z \geq r \cdot e$$

The first part of the main theorem is: if R_ε^P is finite and positive, then

$$\text{min}\{\varepsilon, \delta\} \leq R_\varepsilon^P \cdot r_\delta^D \leq \varepsilon + \delta$$

This then is a generalization to nearly-non-regular problems, where R_ε^P is finite and r_δ^D is non-zero

Question about Main Result

$$R^P_\varepsilon := \max_x \ \|x\|_1 \quad \text{s.t.} \quad x \in P_\varepsilon \qquad\qquad r^D_\delta := \max_{z,r} \ r \quad \text{s.t.} \quad z \in D_\delta, \ z \ge r \cdot e$$

Q: Why the $\|\cdot\|_1$ norm?

A: Because $f(x) := \|x\|_1$ is a **linear** function on the cone $\Re^n_+$. The linearity gives R^P_ε nice properties. If $\|\cdot\|$ is not linear on $\Re^n_+$ then we have to slightly weaken the main theorem as we will see . . . .

Primal and Dual Conic Problem

$$P: \ \text{VAL}_* := \min_x \quad c^T x \qquad\qquad D: \ \text{VAL}^* := \max_{y,z} \quad b^T y$$

$$\text{s.t.} \quad Ax = b \qquad\qquad\qquad\qquad \text{s.t.} \quad A^T y + z = c$$

$$x \in C \qquad\qquad\qquad\qquad\qquad z \in C^*$$

$C \subset X$ is a regular cone: closed, convex, pointed, with nonempty interior

$C^* := \{z : z^T x \geq 0 \ \forall x \in C\}$

Primal and Dual Level Sets

$$P: \ \mathsf{VAL}_* := \min_x \quad c^T x \qquad\qquad D: \ \mathsf{VAL}^* := \max_{y,z} \quad b^T y$$

$$\text{s.t.} \quad Ax = b \qquad\qquad \text{s.t.} \quad A^T y + z = c$$

$$x \in C \qquad\qquad z \in C^*$$

$$P_\varepsilon := \{x : Ax = b, x \in C, c^T x \le \mathsf{VAL}_* + \varepsilon\}$$

$$D_\delta := \{z : \exists y \text{ satisfying } A^T y + z = c, z \in C^*, b^T y \ge \mathsf{VAL}^* - \delta\}$$

Level Set Geometry Measures

Fix a norm $\|x\|$ for the space of the x variables.

The dual norm is $\|z\|_* := \max\{z^T x : \|x\| \le 1\}$ for the z variables.

Define for $\varepsilon, \delta > 0$:

$$
\begin{array}{rll}
R_\varepsilon^P := & \max_x & \|x\| \\
& \text{s.t.} & Ax = b \\
& & x \in C \\
& & c^T x \le \mathrm{VAL}_* + \varepsilon
\end{array}
\qquad
\begin{array}{rll}
r_\delta^D := & \max_{y,z} & \mathrm{dist}_*(z, \partial C^*) \\
& \text{s.t.} & A^T y + z = c \\
& & z \in C^* \\
& & b^T y \ge \mathrm{VAL}^* - \delta
\end{array}
$$

$\mathrm{dist}_*(z, \partial C^*)$ denotes the distance from z to ∂C^* in the dual norm $\|z\|_*$

Level Set Geometry Measures, continued

$$R_\varepsilon^P := \max_x \ \|x\| \quad \text{s.t.} \quad \begin{array}{l} Ax = b \\ x \in C \\ c^T x \le \mathsf{VAL}_* + \varepsilon \end{array}$$

$$r_\delta^D := \max_{y,z} \ \mathrm{dist}_*(z, \partial C^*) \quad \text{s.t.} \quad \begin{array}{l} A^T y + z = c \\ z \in C^* \\ b^T y \ge \mathsf{VAL}^* - \delta \end{array}$$

R_ε^P is the norm of the largest primal ε-optimal solution

r_δ^D measures the largest distance to the boundary of C^* among all dual δ-optimal solutions z

Level Set Geometry Measures, continued

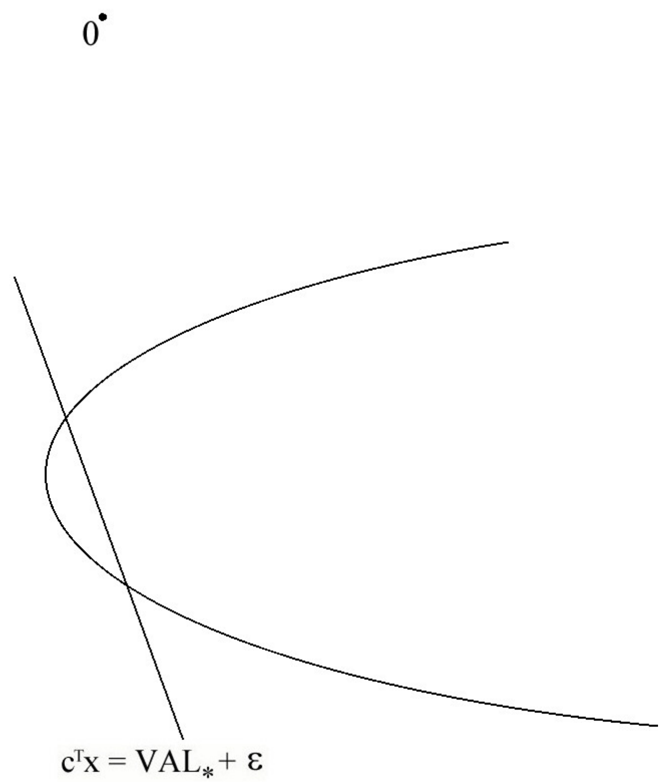
$$R^P_\varepsilon := \ \text{max}_x \ \ \|x\| \qquad\qquad r^D_\delta := \ \text{max}_z \ \ \text{dist}_*(z, \partial C^*)$$

$$\text{s.t.} \quad x \in P_\varepsilon \qquad\qquad\qquad\qquad \text{s.t.} \quad z \in D_\delta$$

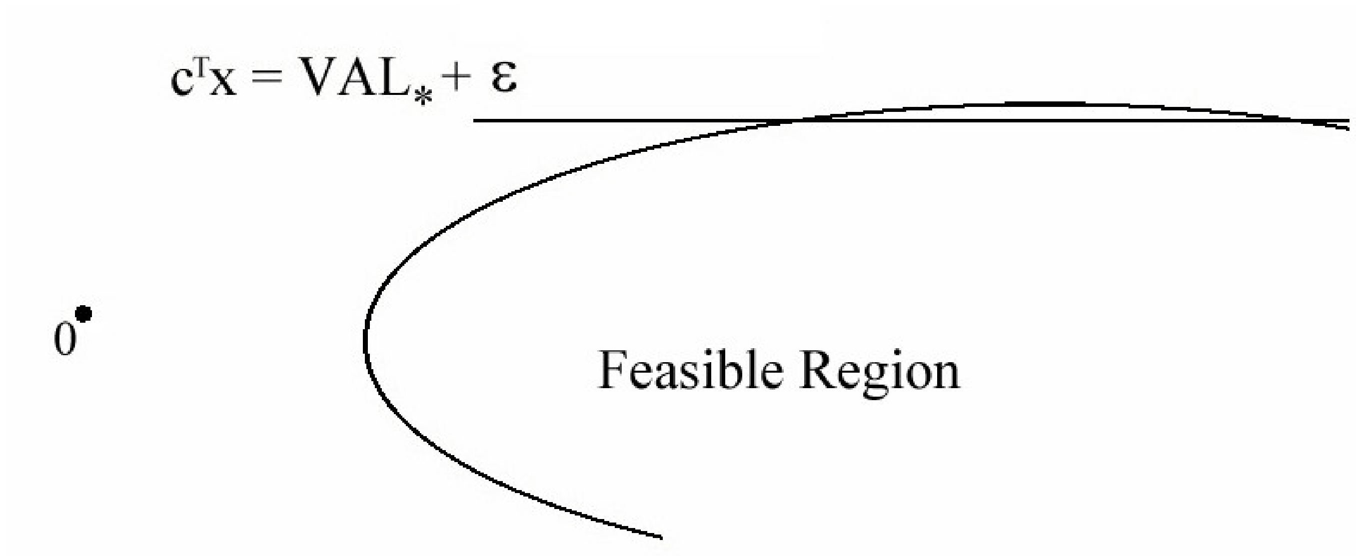
$P_\varepsilon := \{x : Ax = b, x \in C, c^T x \le \text{VAL}_* + \varepsilon\}$

$D_\delta := \{z : \exists y \text{ satisfying } A^T y + z = c, z \in C^*, b^T y \ge \text{VAL}^* - \delta\}$

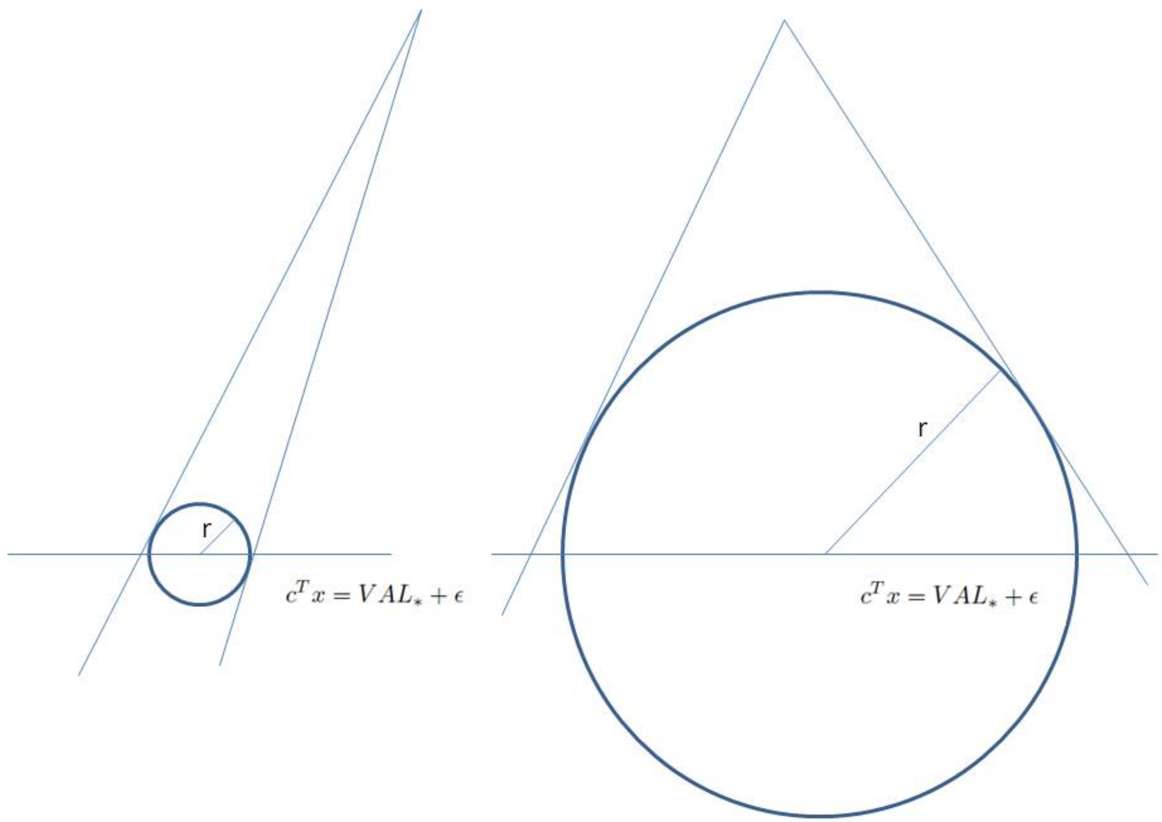
R^P_ε Measures Large Near-Optimal Solutions



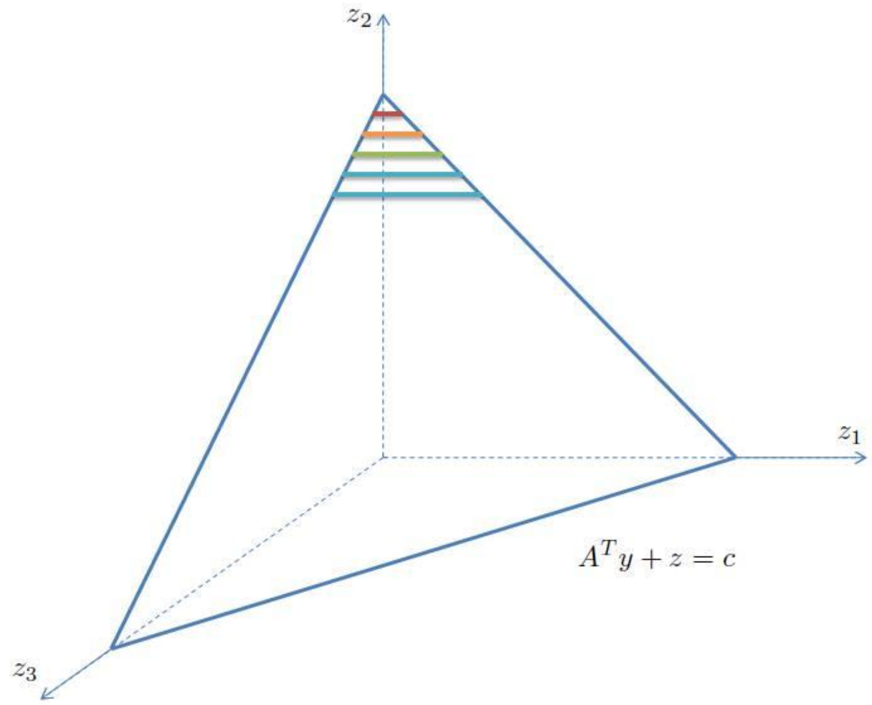
R^P_ε Measures Large Near-Optimal Solutions



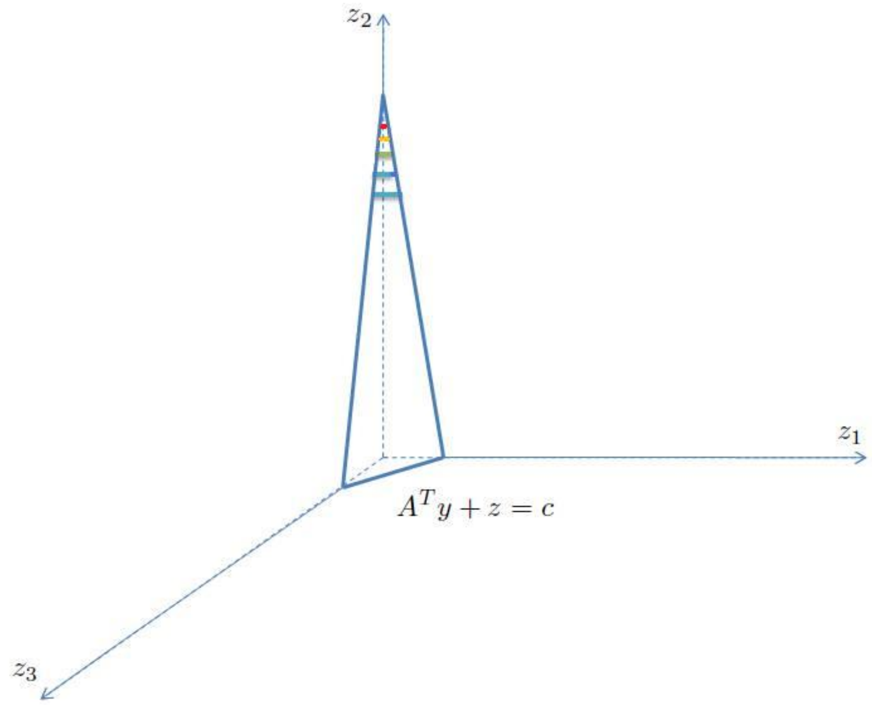
r_δ^D Measures Nicely Interior Near-Optimal Solutions



r_δ^D Measures Nicely Interior Near-Optimal Solutions



r_δ^D Measures Nicely Interior Near-Optimal Solutions



Main Result, Again: R^P_ε and r^D_δ are Reciprocally Related

$$R^P_\varepsilon := \ \max_x \ \|x\| \qquad\qquad r^D_\delta := \ \max_z \ \mathrm{dist}_*(z, \partial C^*)$$

$$\text{s.t.} \quad x \in P_\varepsilon \qquad\qquad\qquad \text{s.t.} \quad z \in D_\delta$$

Main Theorem: Suppose VAL_* is finite. If R^P_ε is positive and finite, then

$$\tau_{C^*} \cdot \min\{\varepsilon, \delta\} \le R^P_\varepsilon \cdot r^D_\delta \le \varepsilon + \delta \ .$$

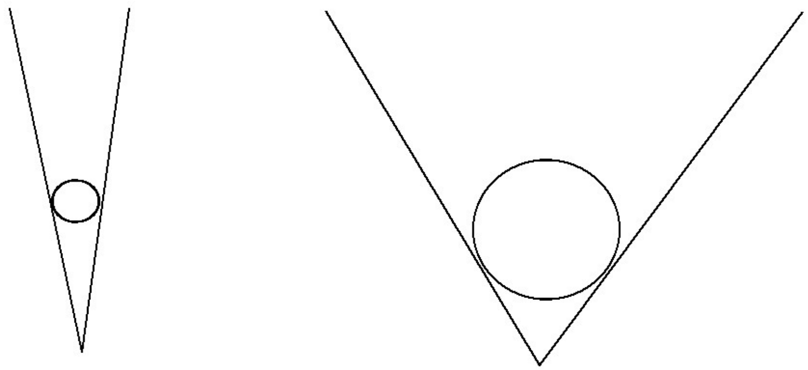
If $R^P_\varepsilon = 0$, then $r^D_\delta = \infty$. If $R^P_\varepsilon = \infty$ and VAL^* is finite, then $r^D_\delta = 0$.∎

Here τ_{C^*} denotes the width of the cone C^* . . . .

On the Width of a Cone

Let K be a convex cone with nonempty interior

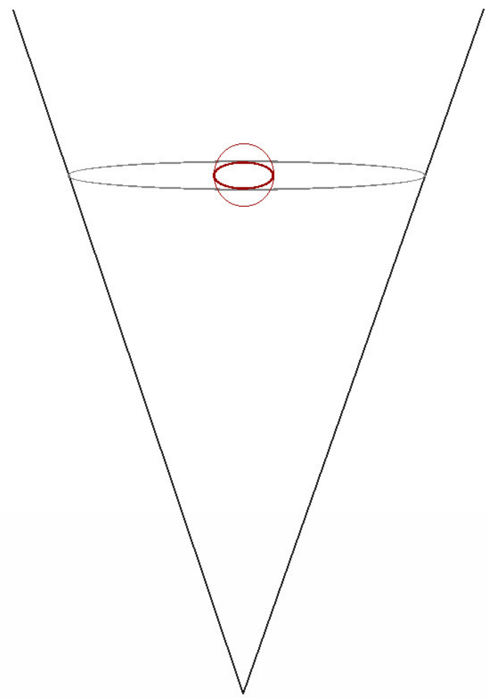
$$\tau_K := \max_x \{\text{dist}(x, \partial K) : x \in K, \|x\| \leq 1\}$$



If K is a regular cone, then $\tau_K \in (0,\ 1]$

τ_K generalizes Goffin's "inner measure"

A Cone with small Width τ_K



$\tau_K << 1$

Equivalence of Norm Linearity and Width of Polar Cone

Proposition: Let K be a regular cone. The following statements are equivalent:

- $\tau_{K^*} \geq \alpha$, and

- there exists $\bar{w}$ for which

$$\alpha \bar{w}^T x \leq \|x\| \leq \bar{w}^T x \qquad \text{for all } x \in K \ \blacksquare$$

Corollary: $\tau_{K^*} = 1$ implies $f(x) := \|x\|$ is linear on K $\blacksquare$

Main Result, Again: R^P_ε and r^D_δ are Reciprocally Related

$$R^P_\varepsilon := \ \max_x \ \ \|x\| \qquad\qquad r^D_\delta := \ \max_z \ \ \mathrm{dist}_*(z, \partial C^*)$$

$$\text{s.t.} \quad x \in P_\varepsilon \qquad\qquad\qquad \text{s.t.} \quad z \in D_\delta$$

Main Theorem: Suppose VAL_* is finite. If R^P_ε is positive and finite, then

$$\tau_{C^*} \cdot \min\{\varepsilon, \delta\} \le R^P_\varepsilon \cdot r^D_\delta \le \varepsilon + \delta \ .$$

If $R^P_\varepsilon = 0$, then $r^D_\delta = \infty$. If $R^P_\varepsilon = \infty$ and VAL^* is finite, then $r^D_\delta = 0$.▮

Comments

$$\tau_{C^*} \cdot \min\{\varepsilon, \delta\} \leq R^P_\varepsilon \cdot r^D_\delta \leq \varepsilon + \delta$$

- R^P_ε, r^D_δ each involves primal and dual information

- each inequality can be tight (and cannot be improved)

- many naturally arising norms have $\tau_{C^*} = 1$

- setting $\delta = \varepsilon$, we obtain $\varepsilon \leq R^P_\varepsilon \cdot r^D_\varepsilon \leq 2\varepsilon$, showing these two measures are inversely proportional (to within a factor of 2)

Application: Robust Optimization [J.Vera]

Amended format:

$$
\begin{array}{llll}
P: \ z^*(b) := \max_x & c^T x & D: \ \min_y & b^T y \\
\text{s.t.} & b - Ax \in K & \text{s.t.} & A^T y = c \\
& & & y \in K^*
\end{array}
$$

For a given tolerance $\varepsilon > 0$, what is the limit on the size of a perturbation Δb so that $|z^*(b + \Delta b) - z^*(b)| \le \varepsilon$?

Application: Robust Optimization, continued

$$
\begin{array}{llll}
P: & z^*(b) := \max_x & & c^T x \\
 & & \text{s.t.} & b - Ax \in K
\end{array}
\qquad
\begin{array}{lll}
D: & \min_y & b^T y \\
 & \text{s.t.} & A^T y = c \\
 & & y \in K^*
\end{array}
$$

Theorem [Vera]: Let $\varepsilon > 0$ and Δb satisfy:

$$\|\Delta b\| \le \tau_K \left(\frac{\varepsilon}{R_\varepsilon^D} \right) \ .$$

Then $|z^*(b + \Delta b) - z^*(b)| \le \varepsilon$.∎

The result says that $\tau_K \cdot \varepsilon / R_\varepsilon^D$ is the required bound on the perturbation of the RHS needed to guarantee a change of no more than ε in the value of the problem.

τ_K for Self-Scaled Cones

Nonnegative Orthant: $K = K^* = I\!\!R^n_+$, define $\|x\|_p := \sqrt[p]{\sum_{j=1}^n |x_j|^p}$

Then $\tau_K = n^{(1/p-1)}$, whereby $\tau_K = 1$ for $p = 1$

Semidefinite Cone: $K = K^* = S^k_+$,

Define $\|X\|_p := \|\lambda(X)\|_p := \sqrt[p]{\sum_{j=1}^k |\lambda_j(X)|^p}$

Then $\tau_K = k^{(1/p-1)}$, whereby $\tau_K = 1$ for $p = 1$

Second-Order Cone: $K = K^* = \{x \in \Re^n : \|(x_1, \ldots, x_{n-1})\|_2 \leq x_n\}$

Define $\|x\| := \max\{\|(x_1, \ldots, x_{n-1})\|_2, |x_n|\}$, then $\tau_K = 1$

Some Relations with Renegar's Condition Number

For $\varepsilon \le \|c\|_*$ it holds that:

$$R^P_\varepsilon \le \; C^2(d) + C(d)\frac{\varepsilon}{\|c\|_*}$$

$$r^P_\varepsilon \;\; \ge \;\; \frac{\varepsilon \tau_C}{3\|c\|_*(C^2(d) + C(d))}$$

Outline, again

-

- A Geometric Measure of Feasible Regions and Interior-Point Method (IPM) Complexity Theory

-

Primal and Dual Conic Problem

$$P: \ \text{VAL}_* := \min_x \ c^T x \quad \text{s.t.} \ Ax = b, \ x \in C$$

$$D: \ \text{VAL}^* := \max_{y,z} \ b^T y \quad \text{s.t.} \ A^T y + z = c, \ z \in C^*$$

$C \subset X$ is a regular cone: closed, convex, pointed, with nonempty interior

$C^* := \{z : z^T x \geq 0 \ \forall x \in C\}$

Geometric Measure of Primal Feasible Region

$$G^P := \text{minimum}_x \quad \max\left\{\frac{\|x\|}{\text{dist}(x, \partial C)}, \|x\|, \frac{1}{\text{dist}(x, \partial C)}\right\}$$

$$\begin{array}{ll} \text{s.t.} & Ax = b \\ & x \in C \end{array}$$

G^P is smaller to the extent that there is a feasible solution that is not too large and that is not too close to ∂C

G^P is smaller if the primal has a "well-conditioned" feasible solution

Geometric Complexity Theory of Conic Optimization

$$G^P := \text{minimum}_x \quad \max\left\{\frac{\|x\|}{\text{dist}(x, \partial C)}, \|x\|, \frac{1}{\text{dist}(x, \partial C)}\right\}$$

$$\begin{array}{ll} \text{s.t.} & Ax = b \\ & x \in C \end{array}$$

[F 04] Using a (theoretical) interior-point method that solves a primal-Phase-I followed by a primal-Phase-II, one can bound the IPM iterations to compute an ε-optimal solution by

$$O\left(\sqrt{\vartheta_C}\left(\ln\left(R_\varepsilon^P\right) + \ln\left(G^P\right) + \ln(1/\varepsilon)\right)\right)$$

Geometric Complexity Theory of Conic Optimization

Computational complexity of solving primal problem P is:

$$O\left(\sqrt{\vartheta_C}\left(\ln\left(R_\varepsilon^P\right)+\ln\left(G^P\right)+\ln(1/\varepsilon)\right)\right)$$

Is this just a pretty theory?

Are R_ε^P and G^P correlated with the performance of IPMs on conic problems in practice, say from the SDPLIB suite of SDP problems?

IPMs in practice are interchangeable insofar as role of primal versus dual. Therefore let us replicate the above theory for the dual problem.

Geometric Measure of Dual Feasible Region

$$G^D := \text{minimum}_{y,z} \quad \max\left\{\frac{\|z\|_*}{\text{dist}_*(z,\partial C^*)}, \|z\|_*, \frac{1}{\text{dist}_*(z,\partial C^*)}\right\}$$

$$\begin{array}{ll} \text{s.t.} & A^T y + z = c \\ & z \in C^* \end{array}$$

G^D is smaller to the extent that there is a feasible dual solution that is not too large and that is not too close to ∂C^*

G^D is smaller if the dual has a "well-conditioned" feasible solution

Geometric Complexity Theory of Conic Optimization

$$
\begin{array}{lll}
G^D := \text{minimum}_z & \max\left\{\dfrac{\|z\|_*}{\text{dist}_*(z,\partial C^*)}, \|z\|_*, \dfrac{1}{\text{dist}_*(z,\partial C^*)}\right\} \\
\text{s.t.} & A^T y + z = c \\
 & z \in C^*
\end{array}
$$

[F 04] Using a (theoretical) interior-point method that solves a dual-Phase-I followed by a dual-Phase-II, one can bound the IPM iterations to compute an ε-optimal solution by

$$
O\left(\sqrt{\vartheta_C}\left(\ln\left(R_\varepsilon^D\right) + \ln\left(G^D\right) + \ln(1/\varepsilon)\right)\right)
$$

Aggregate Geometry Measure for Primal and Dual

Define the aggregate geometry measure:

$$G^A := \left(R_\varepsilon^P \times R_\varepsilon^D \times G^P \times G^D\right)^{1/4}$$

G^A aggregates the primal and dual level set measures and the primal and dual feasible region geometry measures

Let us compute G^A for the SDPLIB suite and see if G^A is correlated with IPM iterations.

Outline, again

-

-

- Geometric Measures and their Explanatory Value of the Practical Performance of IPMs

Semi-Definite Programming (SDP)

• broad generalization of LP

• emerged in early 1990's as the most significant computationally tractable generalization of LP

• independently discovered by Alizadeh and Nesterov-Nemirovskii

• applications of SDP are vast, encompassing such diverse areas as integer programming and control theory

Partial List of Applications of SDP

• LP, Convex QP, Convex QCQP

• tighter relaxations of IP ($\leq$ 12% of optimality for MAXCUT)

• static structural (truss) design, dymamic truss design, antenna array filter design, other engineered systems problems

• control theory

• shape optimization, geometric design, volume optimization problems

• D-optimal experimental design, outlier identification, data mining, robust regression

• eigenvalue problems, matrix scaling/design

• sensor network localization

• optimization or near-optimization with large classes of non-convex polynomial constraints and objectives (Parrilo, Lasserre, SOS methods)

• robust optimization methods for standard LP, QCQP

IPM Set-up for SDP

$$\begin{array}{lll} \min_X & C \bullet X & \\ \text{s.t.} & A_i \bullet X = b_i \ , i = 1, \ldots, m \\ & X \succeq 0 \end{array} \qquad \begin{array}{ll} \max_{y,Z} & \sum_{i=1}^m y_i b_i \\ \text{s.t.} & \sum_{i=1}^m y_i A_i + Z = C \\ & Z \succeq 0 \end{array}$$

IPM for SDP, Central Path

$$\text{Central path:} \quad X(\mu) := \operatorname{argmin}_X \; C \bullet X - \mu \sum_{j=1}^{n} \ln(\lambda_j(X))$$

$$\text{s.t.} \quad A_i \bullet X = b_i \;, i = 1, \ldots, m$$
$$X \succ 0$$

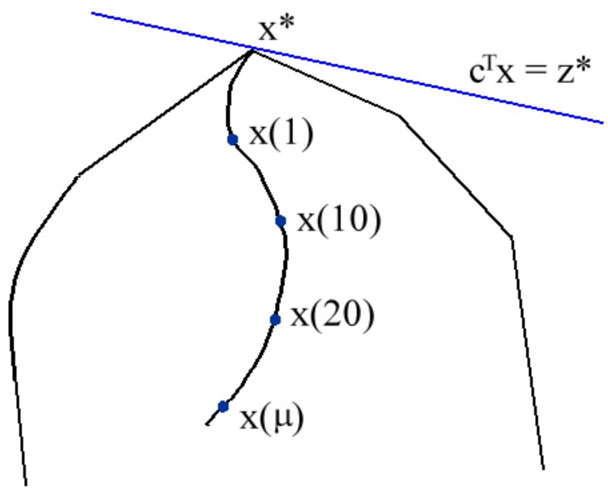
$$\text{Central path:} \quad X(\mu) := \operatorname{argmin}_X \; C \bullet X - \mu \ln(\det(X))$$

$$\text{s.t.} \quad A_i \bullet X = b_i \;, i = 1, \ldots, m$$
$$X \succ 0$$

IPM for SDP, Central Path, continued

Central path: $\quad X(\mu) := \ \text{argmin}_X \ \ C \bullet X + \mu \ln(\det(X))$

$$\begin{array}{ll} \text{s.t.} & A_i \bullet X = b_i \ , i = 1, \ldots, m \\ & X \succ 0 \end{array}$$



IPM for SDP, Central Path, continued

Central path: $\quad X(\mu) := \ \mathrm{argmin}_X \ \ C \bullet X + \mu \ln(\det(X))$

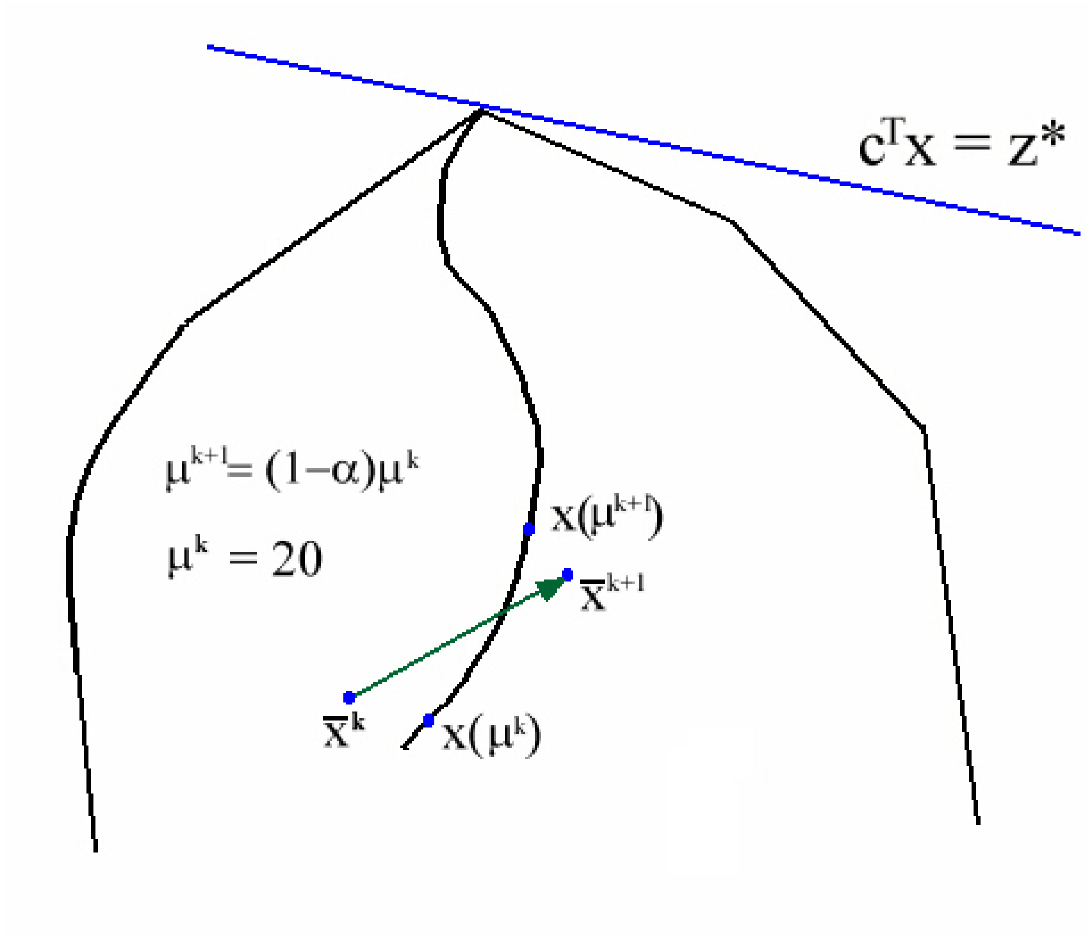
$$\begin{array}{ll} \text{s.t.} & A_i \bullet X = b_i \ , i = 1, \ldots, m \\ & X \succ 0 \end{array}$$

Optimality gap property of the central path:

$$C \bullet X(\mu) - \mathrm{VAL}_* \leq n \cdot \mu$$

Algorithm strategy: trace the central path $X(\mu)$ for a decreasing sequence of values of $\mu \searrow 0$

IPM Strategy for SDP



IPM for SDP: Computational Reality

- 1991-94 - Alizadeh, Nesterov and Nemirovski - IPM theory for SDP
- 1996 - software for SOCP, SDP - 10-60 iterations on SDPLIB suite, typically $\sim$30 iterations
- Each IPM iteration is expensive to solve:

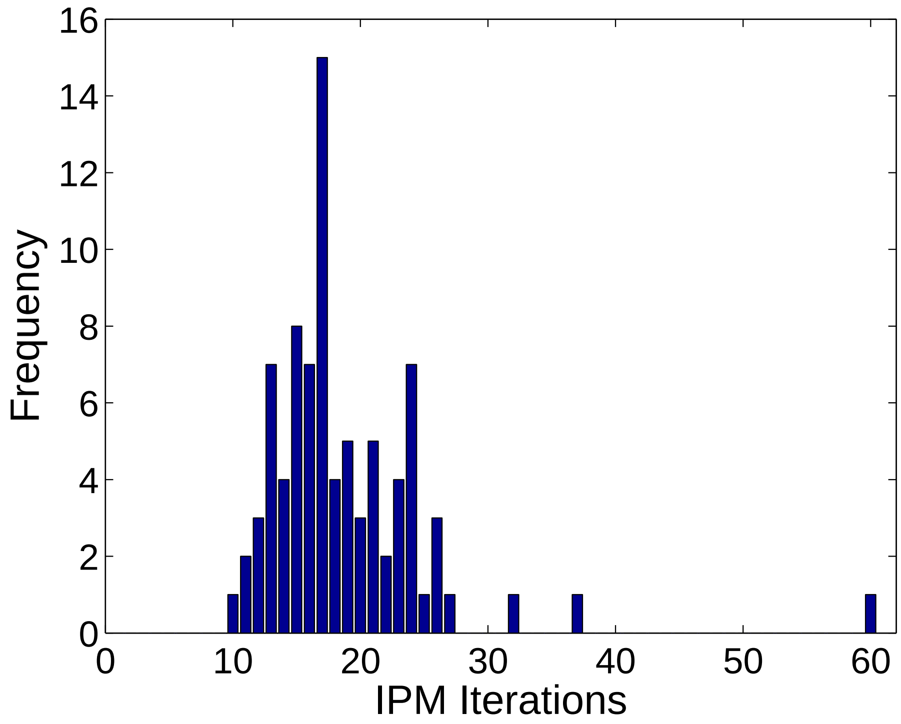
$$\begin{pmatrix} H(x^k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$$

- $O(n^6)$ work per iteration, managing sparsity and numerical stability are tougher bottlenecks
- most IPM computational research since 1996 has focused on work per iteration, sparsity, numerical stability, lower-order methods, etc.

SDPLIB Suite

- `http://www.nmt.edu/~sdplib/`
- 92 problems
- standard equality block form, SDP variables and LP variables
- no linear dependent equations
- Work with 85 problems:
  - removed 4 infeasible problems: `infd1`, `infd2`, `infp1`, `infp2`
  - removed 3 very large problems: `maxG55` (5000×5000), `maxG60` (7000×7000), `thetaG51` (6910×1001)
- $m : 6 - 4375$, $n : 13 - 2000$

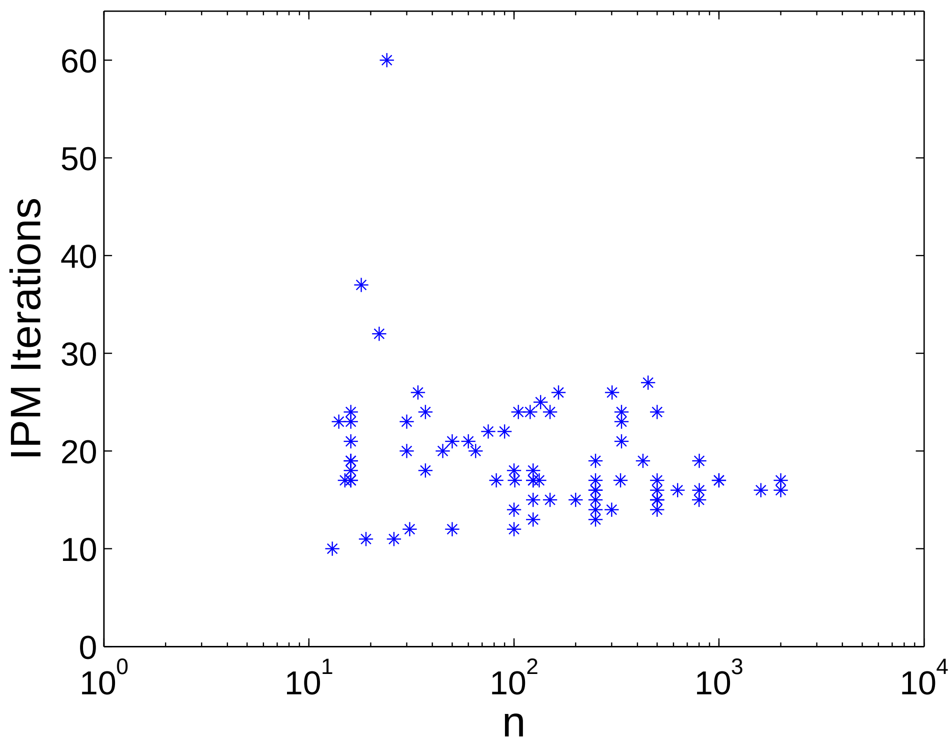
Histogram of IPM Iterations for SDPLIB problems



SDPT3-3.1 default settings used throughout

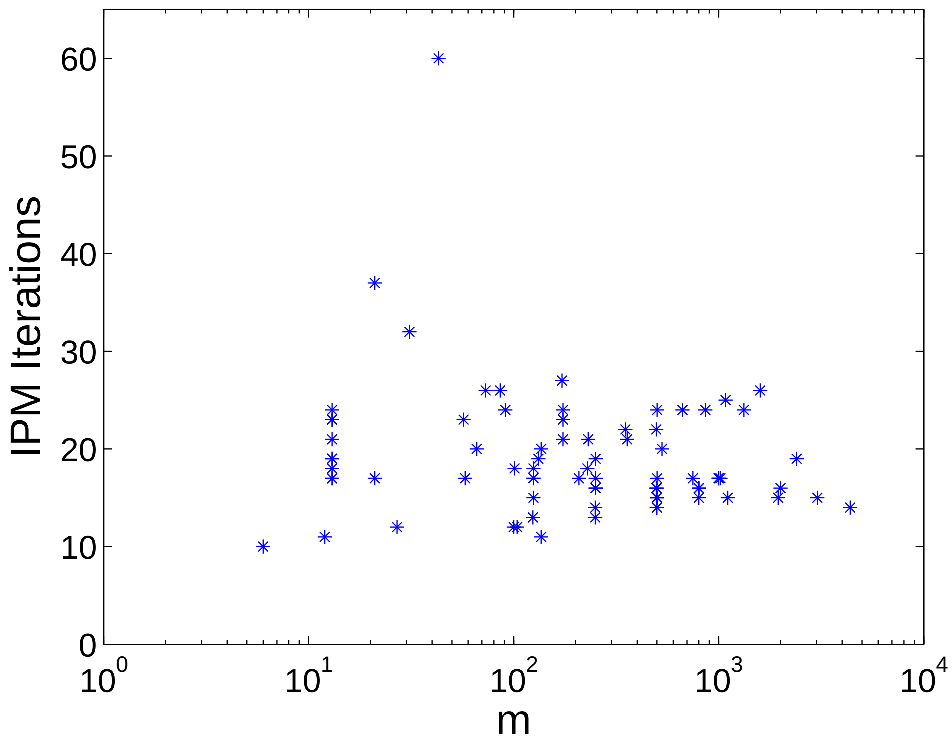
SDPT3-3.1 solves the 85 SDPLIB problem instances in 10-60 iterations

IPM Iterations versus n



Scatter Plot of IPM iterations and $n := n_s + n_l$

IPM Iterations versus m



Scatter Plot of IPM iterations and m

Computing the Aggregate Geometry Measure G^A

$$G^A := \left(R_\varepsilon^P \times R_\varepsilon^D \times G^P \times G^D\right)^{1/4}$$

R_ε^P, R_ε^D are *maximum* norm problems, so are generally non-convex.

Computing G^P, G^D involves working with "dist$(x, \partial C)$, dist$_*(z, \partial C^*)$" which is not efficiently computable in general

A judicious choice of norms allows us to compute all four quantities efficiently via one associated SDP for each quantity.

Geometry Measure Results

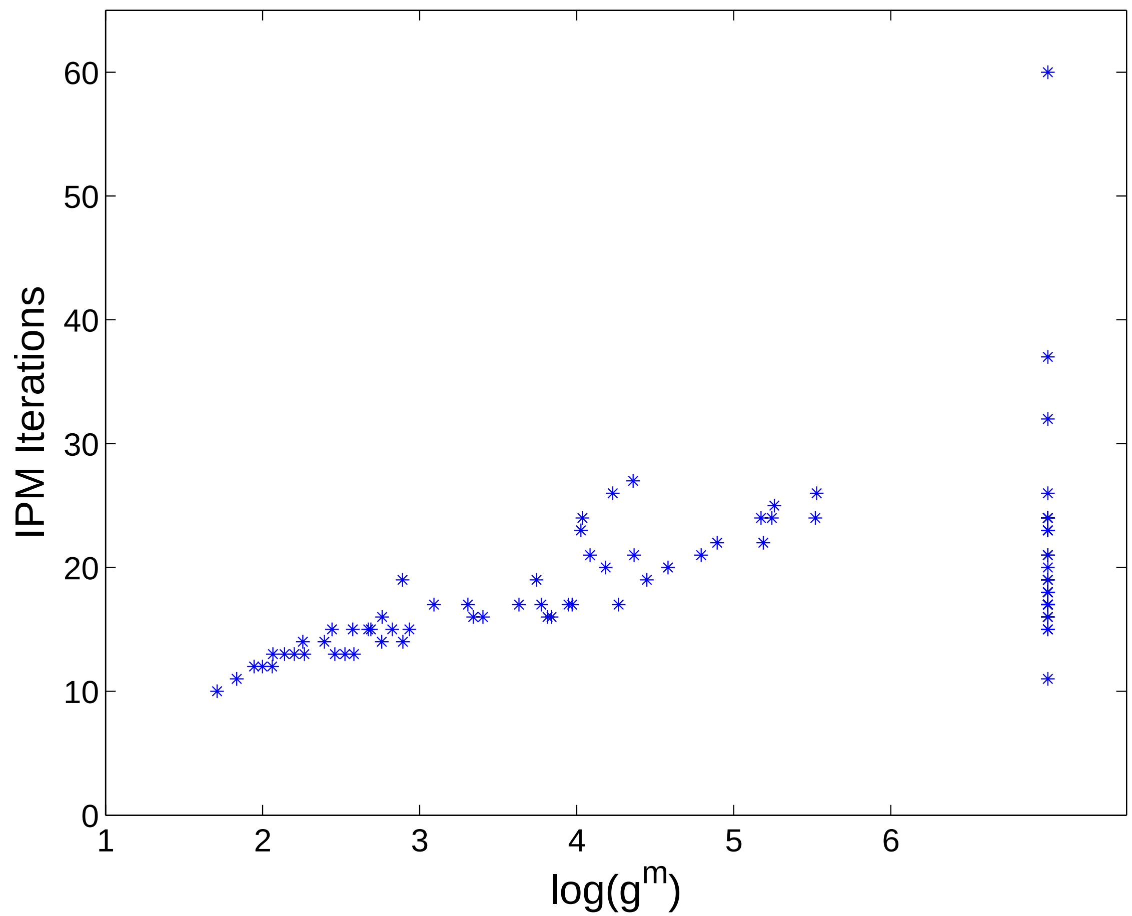
G^A was computed for 85 SDPLIB problems:

	G^A	R^P_ε	R^D_ε	G^P	G^D
Finite	53	85	53	53	85
Infinite	32	-	32	32	-
Total	85	85	85	85	85

62% of problems have finite G^A

The pattern in the table is no coincidence . . .

$$G^P = \infty \iff R^D_\varepsilon = \infty \quad \text{and} \quad G^D = \infty \iff R^P_\varepsilon = \infty$$



IPM iterations versus $\log(G^A)$

$$\mathrm{CORR}\left(\log(G^A),\ \text{IPM Iterations}\right) = 0.901 \qquad\qquad \text{(53 problems)}$$

What About Other Behavioral Measures?

How well does Renegar's "condition measure" $C(d)$ explain the practical performance of IPMs on the SDPLIB suite?

$C(d)$: Renegar's Condition Measure

$$
(P_d): \quad \begin{array}{lrcl} \min & c^T x & & \\ \text{s.t.} & Ax & = & b \\ & x & \in & C \end{array}
\qquad\qquad
(D_d): \quad \begin{array}{lrcl} \max & b^T y & & \\ \text{s.t.} & A^T y + z & = & c \\ & z & \in & C^* \end{array}
$$

- $d = (A, b, c)$ is the data for the instance P_d and D_d

- $\|d\| = \max\{\|A\|, \|b\|, \|c\|_*\}$

Distance to Primal and Dual Infeasibility

$$(P_d): \quad \begin{array}{lrcl} \min & c^T x & & \\ \text{s.t.} & Ax & = & b \\ & x & \in & C \end{array} \qquad\qquad (D_d): \quad \begin{array}{lrcl} \max & b^T y & & \\ \text{s.t.} & A^T y + z & = & c \\ & z & \in & C^* \end{array}$$

Distance to primal infeasibility:

$$\rho_P(d) = \min \left\{ \|\Delta d\| : P_{d+\Delta d} \ \text{ is infeasible} \right\}$$

Distance to dual infeasibility:

$$\rho_D(d) = \min \left\{ \|\Delta d\| : D_{d+\Delta d} \ \text{ is infeasible} \right\}$$

The condition measure is:

$$C(d) = \frac{\|d\|}{\min\{\rho_P(d), \rho_D(d)\}}$$

$C(d)$: Renegar's Condition Measure

$$C(d) = \frac{\|d\|}{\min\{\rho_P(d), \rho_D(d)\}}$$

In theory, $C(d)$ has been shown to be connected to:

- bounds on sizes of feasible solutions and aspect ratios of inscribed balls in feasible regions
- bounds on sizes of optimal solutions and objective values
- bounds on rates of deformation of feasible regions as data is modified
- bounds on deformation of optimal solutions as data is modified
- bounds on the complexity of a variety of algorithms

[Renegar 95] Using a (theoretical) IPM that solves a primal-phase-I followed by a primal-Phase-II, one can bound IPM iterations needed to compute an ε-optimal solution by

$$O\left(\sqrt{\vartheta_C}\,(\ln\left(C(d)\right) + \ln(1/\varepsilon))\right)$$

$\vartheta_C = n_s + n_l$ for SDP

Is $C(d)$ just really nice theory?

Is $C(d)$ correlated with IPM iterations among SDPLIB problems?

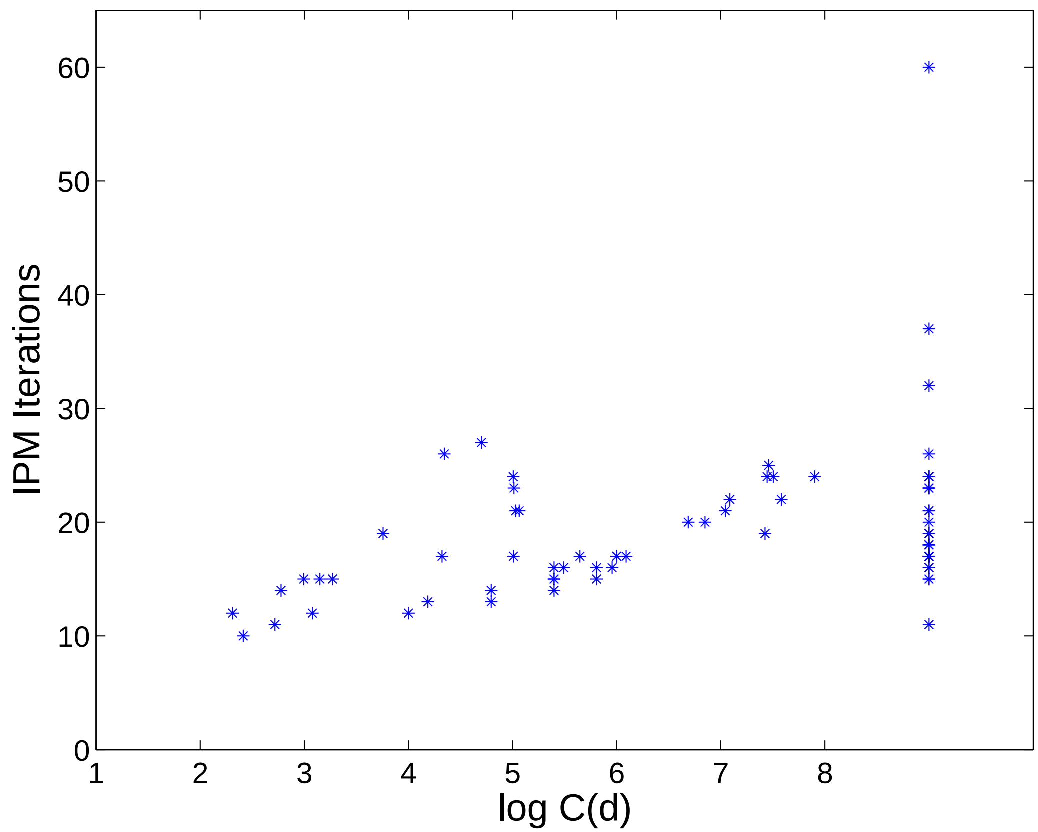
Condition Measure Results

Computed $C(d)$ for 80 (out of 85) problems:

Unable to compute $\rho_p(d)$ for 5 problems: `control11`, `equalG51`, `maxG32`, `theta6`, `thetaG11` ($m =$ 1596, 1001, 2000, 4375, 2401, respectively)

		$\rho_D(d)$		
		0	> 0	Total
$\rho_P(d)$	0	0	32	32
	> 0	0	48	48
	Total	0	80	80

- 60% are well-posed
- 40% are almost primal infeasible

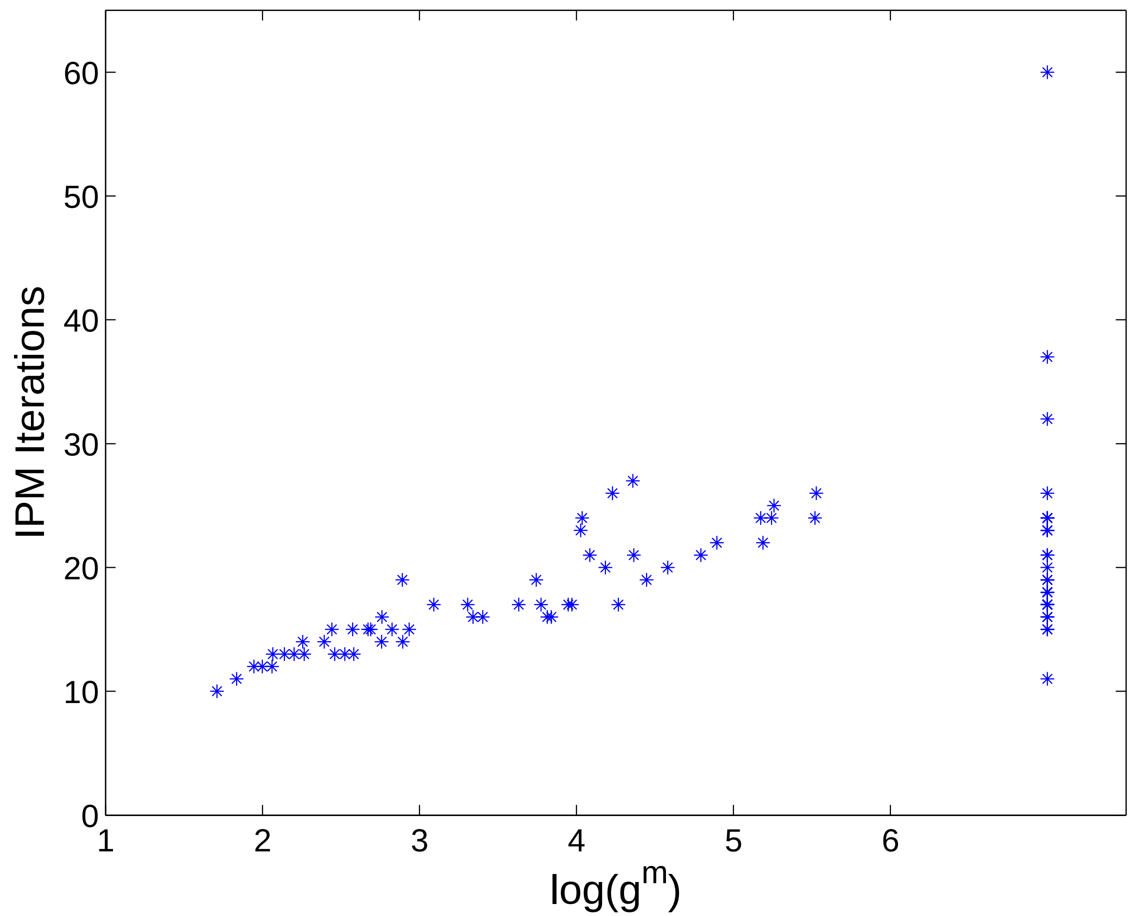


IPM iterations versus $\log(C(d))$.

$\mathrm{CORR}(\log(C(d)),\ \text{IPM Iterations}) = 0.630$ (48 problems)

Some Conclusions

- 62% of 85 SDPLIB problems have finite aggregate geometry measure G^A
- CORR($\log\left(G^A\right)$, IPM Iterations) $= 0.901$ among the SDPLIB problems with finite geometry measure G^A
- 32 of 80 SDPLIB problems are almost primal infeasible, i.e. $C(d) = +\infty$
- CORR($\log\left(C(d)\right)$, IPM Iterations) $= 0.630$ among the 42 problems with finite $C(d)$



IPM iterations versus $\log(G^A)$

$\mathrm{CORR}\left(\log(G^A),\ \text{IPM Iterations}\right) = 0.901$ (53 problems)