On the Causes of Variability in IPM Iterations for Semi-Definite Programming Problems

R.M. Freund (MIT)
F. Ordóñez (USC)
K-C. Toh (NUS)

May, 2005

paper in Optimization Online
http://www.optimization-online.org/DB_HTML/2005/02/1073.html
Outline

• Variation in IPM Iterations among SDPLIB problems

• Examine Four Possible Explanations:
  – Geometric Measure of P/D Feasible/Optimal Regions
  – Condition Measure: Renegar’s $C(d)$
  – Degeneracy of Primal or Dual Optimal Solution
  – Absence of Strict Complementarity in Optimal Solution

• Computational Evaluation

• Conclusions
Variation in IPM Iterations among SDPLIB problems

SDPT3-3.1 solves SDPLIB problems in 10-60 iterations
• 92 problems

• standard equality block form, no SOCP blocks

• no linear dependent equations

• Studied 85 problems:
  
  – removed 4 infeasible problems: infd1, infd2, infp1, infp2

  – removed 3 very large problems: maxG55 (5000×5000), maxG60 (7000×7000), thetaG51 (6910×1001)

• $m : 6 − 4375$, $n : 13 − 2000$

SDPT3-3.1 default settings used throughout
Semidefinite Optimization (SDP)

\[
\begin{align*}
\min & \quad \sum_{j=1}^{n_s} \langle c_j^s, x_j^s \rangle + \langle c^l, x^l \rangle \\
\text{s.t.} & \quad \sum_{j=1}^{n_s} A_j^s \text{svec}(x_j^s) + A^l x^l = b \\
& \quad x_j^s \in K_{s_j}^s \quad \forall j \\
& \quad x^l \in \mathbb{R}^{n_l}_+ \\
\end{align*}
\]

\[m = \dim(b)\]

\[n := \sum_{j=1}^{n_s} s_j + n_l\]
SDP can be written as

\[ \min \quad \langle c, x \rangle \]
\[ \text{s.t.} \quad A(x) = b \]
\[ x \in K \]

where \( \langle c, x \rangle \) and \( A(x) \) are the linear functions in SDP and

\[ K = K^{s_1}_s \times \ldots \times K^{s_{ns}}_s \times \mathbb{R}^{nl}_+ \]

SDP dual problem is:

\[ \max \quad b^T y \]
\[ \text{s.t.} \quad A^*(y) + z = c \]
\[ z \in K^* \]
Geometric Measure $g^m$

For standard form problem with optimal value VAL and optimality tolerance $\varepsilon > 0$ define:

$$D_p := \max_{x} \|x\| \quad \text{s.t.} \quad \begin{align*}
Ax &= b \\
x &\in K \\
c^T x &\leq VAL + \varepsilon
\end{align*}$$

$D_p$ measures the size of the largest $\varepsilon$-optimal solution (among the cone variable $x$).
\( g_p := \min_x \frac{\max\{\|x\|, 1\}}{\min\{\text{dist}(x, \partial K), 1\}} \)

s.t. \quad Ax = b \\
\quad x \in K

\( g_p \) is smaller to the extent that there is a feasible solution that is not too large and that is not too close to \( \partial K \)

[F 04] Using a (theoretical) IPM that solves a primal-Phase-I followed by a primal-Phase-II, one can bound IPM iterations by

\[ O \left( \sqrt{\vartheta} \ln (D_p + g_p + ...) \right) \]
Computational complexity of solving $P$ is:

$$O\left(\sqrt{\vartheta} \ln (D_p + g_p + \ldots)\right)$$

This suggests that $D_p$ and $g_p$ might be relevant to the performance of IPMs in practice.

However, IPMs in practice are interchangeable insofar as role of primal versus dual. Therefore define $D_d$ and $g_d$ analogously for the dual problem.

Now define the aggregate measure:

$$g^m := (D_p \times D_d \times g_p \times g_d)^{1/4}$$
\(D_d := \text{maximum}_{y,z} \|z\|\]
\[\text{s.t.} \quad A^T y + z = c\]
\[z \in K^*\]
\[b^T y \geq \text{VAL} - \varepsilon\]

\(g_d := \text{minimum}_{y,z} \begin{array}{c}
\text{max}\{\|z\|, 1\} \\
\text{min}\{\text{dist}(z, \partial K^*), 1\}
\end{array}\]
\[\text{s.t.} \quad A^T y + z = c\]
\[z \in K^*\]
Computing $g^m$

\[ g^m := (D_p \times D_d \times g_p \times g_d)^{1/4} \]

$D_p$, $D_d$ are maximum norm problems, so are generally non-convex.

In general “dist($x, \partial K$)” is not efficiently computable

A judicious choice of norms allows us to compute each of these four quantities efficiently via a single SDP. For $x \in K$ define:

\[ \|x\| := \sum_{j=1}^{n_s} \|\lambda(x_j^s)\|_1 + \|x^l\|_1 \]

$\lambda(x)$ is the vector of eigenvalues of $x$
Geometry Measure Results

$g^m$ was computed for 85 SDPLIB problems:

<table>
<thead>
<tr>
<th></th>
<th>$D_p$</th>
<th>$D_d$</th>
<th>$g_p$</th>
<th>$g_d$</th>
<th>$g^m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite</td>
<td>85</td>
<td>53</td>
<td>53</td>
<td>85</td>
<td>53</td>
</tr>
<tr>
<td>$\infty$</td>
<td>-</td>
<td>32</td>
<td>32</td>
<td>-</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
<td>85</td>
</tr>
</tbody>
</table>

62% of problems have finite $g^m$

The pattern in the table is no coincidence . . .

$$g_p = \infty \iff D_d = \infty \quad \text{and} \quad g_d = \infty \iff D_p = \infty$$
IPM iterations versus $\log(g^m)$

$\text{CORR}(\log(g^m), \text{IPM Iterations}) = 0.896$ (53 problems)
$C(d)$: Renegar’s Condition Measure

$$\min \langle c, x \rangle$$

s.t. \quad A(x) = b

$$x \in K$$

Renegar’s condition measure is:

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

where

- $d = (A, b, c)$, and $\|d\| = \max\{\|A\|, \|b\|, \|c\|_*\}$
- $\rho_P(d) = \min\{\|\Delta d\| : d + \Delta d \text{ is primal infeasible}\}$
- $\rho_D(d) = \min\{\|\Delta d\| : d + \Delta d \text{ is dual infeasible}\}$
Using a (theoretical) IPM that solves a primal-phase-I followed by a primal-Phase-II, one can bound IPM iterations by

\[ O\left( \sqrt{\vartheta} \ln (C(d) + \ldots) \right) \]

This suggests that \( C(d) \) might explain the variability in IPM iterations among SDPLIB suite
Computing $C(d) = \frac{\|d\|}{\min\{\rho_P(d), \rho_D(d)\}}$

[FV 99]:

$$\rho_P(d) = \min_{z \in K^*} \max_{u \geq 0} \left\{ \| A^*(y) + z \|_*, |b^t y - u| \right\}$$

$$\|y\|_* = 1$$

$$\rho_D(d) = \min_{x \in K} \max_{g \geq 0} \left\{ \| A(x) \|, |\langle c, x \rangle + g| \right\}$$

$$\|x\| = 1$$

These two non-convex programs can be solved efficiently with a judicious choice of norms . . . .
Define the norms:

- \( \|y\|_* := \|y\|_{\infty} \)

- \( \|x\| := \sum_{j=1}^{n_s} \|\lambda(x_j^s)\|_1 + \|x^l\|_1 \)
  
  - \( \lambda(x) \) is the vector of eigenvalues of \( x \)

Using these norms, computing \( \rho_P(d), \rho_D(d) \) reduces to solving \( 2m + 1 \) SDPs.

Using these two norms, \( \|A\| \) cannot be computed efficiently in theory, but can be estimated efficiently in practice.
Condition Measure Results

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

Still computing $\rho_p(d)$ for 5 problems: control11, equalG51, maxG32, theta6, thetaG11 (each takes more than a week)

<table>
<thead>
<tr>
<th>$\rho_D(d)$</th>
<th>$\rho_P(d)$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>32</td>
</tr>
<tr>
<td>&gt; 0</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>80</td>
</tr>
</tbody>
</table>

- 60% are well-posed
- 40% are almost primal infeasible
One can prove that for problems with linear independent equations:

- $g_p = \infty \iff D_d = \infty \iff \rho_P(d) = 0$

- $g_d = \infty \iff D_p = \infty \iff \rho_D(d) = 0$

- $g_p \leq 3nC(d), \ g_d \leq 3nC(d)$

- $D_p \leq C(d)^2 + C(d) \|c\|_*, \ D_d \leq C(d)^2 + C(d) \|b\|$
IPM iterations versus log \((C(d))\).

\[
\text{CORR}(\log(C(d)), \text{IPM Iterations}) = 0.630 \quad (48 \text{ problems})
\]
Degeneracy and Complementary Slackness of Optimal Solutions

[Alizadeh et al. 98] If the optimal solution is primal and dual non-degenerate and satisfies complementary slackness, then IPMs will exhibit $Q$-quadratic convergence close to the optimal solution.

This suggests degeneracy and existence of complementary slackness might affect performance of IPMs in practice.
Primal Degeneracy

Let $x = (x_{1}^{s}, \ldots, x_{n_{s}}^{s})$ be primal feasible

rank($x_{j}^{s}$) = $r_{j}$, with $D_{r_{j}}$ diagonal matrix of positive eigenvalues

$$x_{j}^{s} = [Q_{j}^{(1)}Q_{j}^{(2)}] \begin{bmatrix} D_{r_{j}} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_{j}^{(1)} \\ Q_{j}^{(2)} \end{bmatrix}$$

$x$ is primal nondegenerate if the matrix

$$B(x) := \begin{bmatrix} A_{1}^{s}(Q_{1}^{(1)} \otimes Q_{1}^{(1)}), & A_{1}^{s}(Q_{1}^{(1)} \otimes Q_{1}^{(2)}), \\
\cdots, & \\
A_{n_{s}}^{s}(Q_{n_{s}}^{(1)} \otimes Q_{n_{s}}^{(1)}), & A_{n_{s}}^{s}(Q_{n_{s}}^{(1)} \otimes Q_{n_{s}}^{(2)}) \end{bmatrix}$$

has full row rank
Dual Degeneracy

Let $y$ and $z = (z_1^s, \ldots, z_{n_s}^s)$ be dual feasible pair with

$$\text{rank}(z_j^s) = \tilde{r}_j,$$

with $\tilde{D}_{\tilde{r}_j}$ diagonal matrix of positive eigenvalues

$$z_j^s = \begin{bmatrix} \tilde{Q}_j^{(1)} \tilde{Q}_j^{(2)} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & \tilde{D}_{\tilde{r}_j} \end{bmatrix} \begin{bmatrix} \tilde{Q}_j^{(1)} \\ \tilde{Q}_j^{(2)} \end{bmatrix}$$

$(y, z)$ is dual nondegenerate if the matrix

$$\tilde{B}(y, z) := \begin{bmatrix} A_1^s (\tilde{Q}_1^{(1)} \otimes \tilde{Q}_1^{(1)}) , \ldots , A_{n_s}^s (\tilde{Q}_{n_s}^{(1)} \otimes \tilde{Q}_{n_s}^{(1)}) \end{bmatrix}$$

has full column rank.

Let $N$ denote the number of columns of $\tilde{B}(y, z)$
Strict Complementarity

Let \( x = (x_1^s, \ldots, x_n^s) \) and \( y \) and \( z = (z_1^s, \ldots, z_n^s) \) be primal and dual feasible with

\[
\text{rank}(x_j^s) = r_j
\]

\[
\text{rank}(z_j^s) = \tilde{r}_j
\]

\[
\langle x_j^s, z_j^s \rangle = 0
\]

Then strict complementarity holds if

\[
r_j + \tilde{r}_j = s_j, \quad j = 1, \ldots, n_s
\]
Degeneracy and Strict Complementarity Results

Determined the degeneracy of the optimal solution for 68 of 85 SDPLIB problems.

<table>
<thead>
<tr>
<th></th>
<th>Dual</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Degenerate</td>
<td>Nondegenerate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primal</td>
<td>Degenerate</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Nondegenerate</td>
<td>15</td>
<td>43</td>
<td>58</td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>45</td>
<td></td>
<td>68</td>
</tr>
</tbody>
</table>

63% of the 68 SDPLIB problems are nondegenerate

All problems exhibit strict complementarity except three: control10, hinf12, truss6
Degeneracy Measure

Recall that a solution pair \(x\) and \((y, z)\) is

- primal nondegenerate if \(B(x)\) has full row rank (\(=m\))

- dual nondegenerate if \(\tilde{B}(y, z)\) has full column rank (\(=N\))

Define the degeneracy measure

\[
\gamma = \max \left\{ 1 - \frac{\text{row rank} B(x)}{m}, 1 - \frac{\text{col rank} \tilde{B}(y, z)}{N} \right\}
\]
\[ \gamma = \max \left\{ 1 - \frac{\text{row rank}B(x)}{m}, 1 - \frac{\text{col rank} \tilde{B}(y, z)}{N} \right\} \]

- \( \gamma = 0 \) if solution is primal and dual nondegenerate
- \( \gamma > 0 \) if solution is primal or dual degenerate
- \( \gamma \rightarrow 1 \) as degeneracy increases
CORR(\(\gamma\), IPM Iterations) = 0.100 

(68 problems)
Non-strict complementarity Measure

Recall that a solution pair $x$ and $(y, z)$ such that $\langle x, z \rangle = 0$ is strictly complimentary if

$$x^s + z^s \succ 0 \quad \text{and} \quad x^l + z^l > 0$$

The solution is close to non-strict complementarity if either:

an eigenvalue of $x^s + z^s \sim 0$ or a coordinate of $x^l + z^l \sim 0$

Note also that for $(x(\mu), y(\mu), z(\mu))$ on the central path, with $\lambda_x$ and $\lambda_z$ eigenvalues of $x^s(\mu)$ and $z^s(\mu)$ we have

$$\lambda_x + \lambda_z \succeq 2\sqrt{\mu}I \quad \text{and} \quad x^l(\mu) + z^l(\mu) \geq 2\sqrt{\mu}e$$
\[ \lambda_x + \lambda_z \geq 2\sqrt{\mu}I \quad \text{and} \quad x^l(\mu) + z^l(\mu) \geq 2\sqrt{\mu}e \]

For an approximate optimal solution \((x, y, z)\) define

\[ w = \frac{1}{2\sqrt{\mu}}(x + z) \text{ with } \mu = \frac{1}{n}\langle x, z \rangle \text{ and } \lambda_w \text{ eigenvalues of } w^s \]

\[ T^s = \{ j : (\lambda_w)_j \leq T \} \text{ and } T^l = \{ j : w^l_j \leq T \} \]

Measure of non-strict complementarity

\[ \kappa = \frac{-1}{|T^s| + |T^l|} \left( \sum_{j \in T^s} \log((\lambda_w)_j) + \sum_{j \in T^l} \log(w^l_j) \right) \]

If \( \kappa \to 0 \) the problem tends to have a non-strictly complimentary optimal solution
IPM iterations versus non-strict complementarity measure $\kappa$

$\text{CORR}(\kappa, \text{IPM Iterations}) = 0.423$ (85 problems)
Comparing $C(d)$, $g^m$, $\gamma$, and $\kappa$

Sample Correlations (number of problems in bold):

<table>
<thead>
<tr>
<th></th>
<th>Iterations</th>
<th>log($C(d)$)</th>
<th>log($g^m$)</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iterations</td>
<td>1.000</td>
<td>$0.630 \ (48)$</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>log($C(d)$)</td>
<td>$0.896 \ (53)$</td>
<td>$0.831 \ (48)$</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>log($g^m$)</td>
<td>0.100 \ (68)</td>
<td>0.030 \ (38)</td>
<td>-0.000 \ (42)</td>
<td>1.000</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.423 \ (85)</td>
<td>$0.631 \ (48)$</td>
<td>$0.708 \ (53)$</td>
<td>-0.256 \ (68)</td>
</tr>
</tbody>
</table>
Some Conclusions

- 62% of 85 SDPLIB problems have finite geometry measure $g^m$
- CORR(log $(g^m)$, IPM Iterations) = 0.896 among the SDPLIB problems with finite geometry measure $g^m$
- 32 of 80 SDPLIB problems are almost primal infeasible, i.e. $C(d) = +\infty$
- CORR(log $(C(d))$, IPM Iterations) = 0.630 among the 42 problems with finite $C(d)$
- 63% of 68 SDPLIB problems are primal and dual non-degenerate
- $\gamma$ is not correlated with IPM iterations
- $\kappa$ is moderately correlated with IPM iterations.
IPM Iterations versus $n$

Scatter Plot of IPM iterations and $n := \sum_{j=1}^{n_s} s_j + n_l$
IPM Iterations versus $m$

Scatter Plot of IPM iterations and $m$