Reducing the Solution Time for Convex Optimization Problems by Pre-Conditioning Transformations

Robert M. Freund
Alexandre Belloni
MIT

January, 2006

(paper available online)
Motivation

- Reduce the number of interior-point method (IPM) iterations needed to solve convex optimization problems
  - convert badly-behaved problems to well-behaved problems
  - what do “badly-behaved” and “well-behaved” mean?

- Key enabling technologies are:
  - Random sampling on convex bodies
  - Projective transformations
We know how to do at least two types of random sampling in $\mathbb{R}^d$:

- compute a random vector on the unit sphere $S^{d-1} \subset \mathbb{R}^d$

- compute a random point uniformly distributed on a convex body $S$ given an initial point $v^0 \in \text{int}S$ and a membership oracle for $S$
Uniform Vector on the Sphere
Uniform Vector on a Convex Body
Outline

• Homogeneous Conic System and Interior-Point Method (IPM)
• Polar Sets and Elementary Projective Transformations
• Deep Points and the Symmetry Function
• Hit-and-Run Algorithm for Computing a Deep Point
• Complexity Theory
• Projective Transformation Theorem
• Probabilistic “Pre-Conditioner” Algorithm
• Computational Results and Next Steps
Homogeneous Conic Linear System

Given a closed convex cone $C$, and $A \in \mathbb{R}^{m \times n}$, solve:

$$F : \begin{cases} 
Ax &= 0 \\
 x &\in C \\
 x &\neq \{0\}
\end{cases}$$
We assume that $C$ is a regular cone ($\text{int}C \neq \emptyset$ and $C$ contains no line)

We also assume herein that $F$ is feasible
Three Important Regular Convex Cones

Nonnegative Orthant:  \( C = \mathbb{R}^n_+ := \{ x \in \mathbb{R}^n : x_j \geq 0 , \ j = 1, \ldots , n \} \)

Semi-definite Cone:  \( C = S^k_{+} := \{ X \in S^k_{+} : v^t X v \geq 0 \ \forall v \in \mathbb{R}^k \} \)

Second-Order Cone:  \( Q^k = \{ x \in \mathbb{R}^k : \| (x_2, \ldots , x_k) \|_2 \leq x_1 \} \)

\( C = Q := Q^{k_1} \times \cdots \times Q^{k_q} \)

These three cones form the backbone of modern convex optimization theory and methods.
Standard Conic Feasibility Problem

\[
F : \left\{ \begin{array}{l}
Ax = 0 \\
x \in C \\
x \neq 0
\end{array} \right.
\]

Standard form conic feasibility problem is a special case of \( F \):

\[
\left\{ \begin{array}{l}
\bar{A}x = \bar{b} \\
x \in C
\end{array} \right.
\]

where \( C \leftarrow C \times \mathbb{R}_+ \), \( A \leftarrow [\bar{A}, -\bar{b}] \)
Dual Cones

Let $C$ be a convex cone

$C^* = \{ s : s^T x \geq 0 \text{ for all } x \in C \}$ is the dual cone of $C$

<table>
<thead>
<tr>
<th>$C$</th>
<th>$C^*$</th>
<th>Complexity Value $\vartheta_C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{R}_+^n$</td>
<td>$\mathbb{R}_+^n$</td>
<td>$n$</td>
</tr>
<tr>
<td>$S_+^{k \times k}$</td>
<td>$S_+^{k \times k}$</td>
<td>$k$</td>
</tr>
<tr>
<td>$Q_{k_1} \times \cdots \times Q_{k_q}$</td>
<td>$Q_{k_1} \times \cdots \times Q_{k_q}$</td>
<td>$2q$</td>
</tr>
</tbody>
</table>
Normalized Homogeneous Conic System

\[ F : \begin{cases} 
Ax &= 0 \\
 x &\in C \\
x &\neq 0 
\end{cases} \]

Pick \( \bar{s} \in \text{int} C^* \)

\[ x \in C, \ x \neq 0 \iff x \in C \text{ and } \bar{s}^T x > 0 \]

We write \( F \) equivalently as:

\[ F : \begin{cases} 
Ax &= 0 \\
 \bar{s}^T x &= 1 \\
x &\in C 
\end{cases} \]
Interior-Point Method for Solving $F$

\[
F : \begin{cases}
    Ax &= 0 \\
    \bar{s}^T x &= 1 \\
    x &\in C
\end{cases}
\]

We are given a $\vartheta$-logarithmically homogeneous self-concordant barrier $f(\cdot)$ for $C$: 

\[
f(tx) = f(x) - \vartheta \ln(t) \quad \text{for} \quad x \in \text{int}\, C
\]
Logarithmically-Homogeneous Barrier Calculus

\( f(\cdot) \) is a \( \vartheta \)-logarithmically homogeneous self-concordant barrier for \( C \):

- \( f^*(s) := -\inf_{x \in \text{int} C} \{ s^T x + f(x) \} \) is a \( \vartheta \)-logarithmically homogeneous barrier for \( C^* \)
- \( \bar{x} \in \text{int} C \iff -\nabla f(\bar{x}) \in \text{int} C^* \)
- \( \bar{s} \in \text{int} C^* \iff -\nabla f^*(\bar{s}) \in \text{int} C \)
- \( -\nabla f(\bar{x})^T \bar{x} = \vartheta \)
A Simple Interior-Point Problem for Solving $F$

- Choose $\bar{s} \in \text{int}C^*$,

$$ F : \begin{cases} 
    Ax &= 0 \\
    \bar{s}^T x &= 1 \\
    x &\in C
\end{cases} $$

- Define $\bar{x} \leftarrow -\frac{1}{\vartheta} \nabla f^*(\bar{s}) \in \text{int}C$ and consider the problem:

**OP:** $\theta^* := \max_{x,\theta} \theta$

$$ Ax + (A\bar{x})\theta = 0 $$

$$ \bar{s}^T x = 1 $$

$$ x \in C $$

**Note:** We control the choice of $\bar{s}$; we define $\bar{x}$ using $\bar{s}$.
Interior-Point Problem, continued

$$\text{OP}_\mu : \max_{x, \theta} -\mu f(x) + \theta$$

$$Ax + (A\bar{x})\theta = 0$$

$$\bar{s}^T x = 1$$

• $(x, \theta) = (\bar{x}, -1)$ is the **analytic center** of the feasible region of OP

• We follow the central path until we compute $(x, \theta)$ with $\theta \geq 0$
Computational Experiments

100 randomly generated ill-conditioned LP feasibility problems for each \((m, n)\)

Solved using SDPT3

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Average Iterations Before</th>
<th>Average Time Before</th>
</tr>
</thead>
<tbody>
<tr>
<td>m n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 500</td>
<td>8.48</td>
<td>0.54</td>
</tr>
<tr>
<td>500 2500</td>
<td>9.30</td>
<td>2.43</td>
</tr>
<tr>
<td>1000 5000</td>
<td>9.69</td>
<td>22.94</td>
</tr>
</tbody>
</table>
# Computational Experiments

100 randomly generated ill-conditioned LP feasibility problems for each \((m, n)\)

Solved using SDPT3

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Average Iterations</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>n</td>
<td>Before</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>8.48</td>
</tr>
<tr>
<td>500</td>
<td>2500</td>
<td>9.30</td>
</tr>
<tr>
<td>1000</td>
<td>5000</td>
<td>9.69</td>
</tr>
</tbody>
</table>

Decrease 47% Decrease 32%

**Remark:** IPM iterations decreases on all 300 problems
Given a closed convex set $S \subset \mathbb{R}^d$ with $0 \in S$, the *polar* of $S$ is:

$$S^\circ := \{ y \in \mathbb{R}^d : y^T x \leq 1 \text{ for all } x \in S \}$$
Suppose that $S \subset \mathbb{R}^d$ is convex and $0 \in \text{int}S$

An elementary projective transformation of $S \subset \mathbb{R}^d$ is given by

$$w = T(y) = \frac{y}{1 - \hat{v}^T y}$$

where $\hat{v} \in \text{int}S^\circ$
Translation, Polarity, and Projective Transformation

\[
y = \frac{w}{1 + \hat{v}^T w}
\]

Projective Transform

\[
w = \frac{y}{1 - \hat{v}^T y}
\]

Translation

\[
\hat{S} = (S^o - \hat{v})^o
\]

Polarity

\[
S^o
\]
A convex body is a compact convex set with nonempty interior.
Deep Points in a Convex Body

$S \subset \mathbb{R}^m$ is a convex body

Which points are more/most inside, i.e., “deep” in $S$?
Deep Points in a Convex Body

$S \subset \mathbb{R}^m$ is a convex body

Which points are more/most inside, i.e., “deep” in $S$?
Deep Points and the Symmetry Function

$$\text{sym}(\bar{x}, S) := \max \{ t \mid y \in S \Rightarrow \bar{x} - t(y - \bar{x}) \in S \}$$

originally Minkowski, re-introduced for IPMs by Renegar, Nesterov & Nemirovskii
$S \subset \mathbb{R}^m$ is a convex body

$\text{sym}(x, S) \in [0, 1]$ for all $x \in S$

$\text{sym}(x, S) = 0$ for $x \in \partial S$

$\text{sym}(x, S)$ is invariant under invertible affine transformation of $S$
Symmetry Function, continued

$S \subset \mathbb{R}^m$ is a convex body

There exists $x^* \in S$ for which $\text{sym}(x^*, S) \geq 1/m$
Symmetry Function, continued

$S \subset \mathbb{R}^m$ is a convex body

$\text{sym}(x^*, S) = 1$ if $S$ is symmetric about $x^*$
Invariance of $\text{sym}(0, \cdot)$ under Polarity

$S \subset \mathbb{R}^m$ is a convex body

**Proposition:** If $0 \in S$, then $\text{sym}(0, S) = \text{sym}(0, S^\circ)$
$S \subset \mathbb{R}^m$ is a convex body

$\text{sym}(\cdot, S)$ is not invariant under projective transformation
Symmetry Function, continued

\[
S \subset \mathbb{R}^m
\]

The center of mass \( \mu \) of \( S \) satisfies \( \text{sym}(\mu, S) \geq 1/m \)
Deterministically Computing a Deep Point

\( S \subset \mathbb{R}^m \)

<table>
<thead>
<tr>
<th>Point</th>
<th>Symmetry Guarantee</th>
<th>Computational Burden for ( S ) polyhedral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic Center</td>
<td>( 1/(3\vartheta + 1) )</td>
<td>( \approx \text{LP} )</td>
</tr>
<tr>
<td>Löwner-John center</td>
<td>( 1/m )</td>
<td>( \approx \text{SDP} )</td>
</tr>
<tr>
<td>Center of Mass</td>
<td>( 1/m )</td>
<td>#P-Hard (deterministic)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \approx ) polynomial-time (stochastic)</td>
</tr>
</tbody>
</table>
Computing a Deep Point in $S$ with “High Probability”

$S \subset \mathbb{R}^m$

Let $v^1, \ldots, v^M \sim U[S]$, and define $\hat{v} = \frac{1}{M} \sum_{i=1}^{M} v^i$
Computing a Deep Point in $G$ with “High Probability”

$S \subset \mathbb{R}^m$

**Theorem:** Let $M = \lceil 5m \left( \ln \left( \frac{2}{\delta} \right) \right)^2 \rceil$. Then $\text{sym}(\hat{v}, S) \geq \frac{1}{2m}$ with probability $p \geq 1 - \delta$. 
The Hit-and-Run Algorithm
On the Complexity of Hit-and-Run on $S$

(Lovász, Simonovits, Kannan, Vempala, ...)

Let $B(c, \rho)$ denote the Euclidean ball centered at $c$ with radius $\rho$. Suppose that $r, R$ satisfy:

$$B(0, r) \subset S \subset B(0, R).$$

Starting at $v^0 = 0 \in S$, the hit-and-run random walk will approximate a uniform distribution on $S$ in at most

$$O\left(d^4 \ln \left(\frac{R}{\varepsilon \cdot \text{dist}_2(0, \partial S)}\right)\right)$$

iterations.
Back to Homogeneous Conic Feasibility Problem

\[ F : \begin{cases} \ Ax &= 0 \\ x &\in C \\ x &\neq 0 \end{cases} \]
Normalized Homogeneous Conic System

\[
F : \begin{cases}
    Ax &= 0 \\
    x &\in C \\
    x &\neq 0
\end{cases}
\]

Pick \( \bar{s} \in \text{int}C^* \)

\( x \in C, x \neq 0 \iff x \in C \text{ and } \bar{s}^T x > 0 \)

We write \( F \) equivalently as:

\[
F : \begin{cases}
    Ax &= 0 \\
    \bar{s}^T x &= 1 \\
    x &\in C
\end{cases}
\]
Image Set $H$

\[ F : \begin{cases} 
Ax &= 0 \\
\bar{s}^T x &= 1 \\
x &\in C 
\end{cases} \]

Image set is $H := \{ Ax \in \mathbb{R}^m : x \in C \ , \ \bar{s}^T x = 1 \}$

Note that $0 \in H$ if and only if $F$ is feasible.
The behavior of $F$ depends on how deep 0 lies in $H$.
The behavior of $F$ depends on how deep 0 lies in $H$.

well-conditioned but harder to solve
The behavior of $F$ depends on how deep $0$ lies in $H$. 

ill-conditioned
The behavior of $F$ depends on how deep 0 lies in $H$
Interior-Point Method (IPM) for Solving $F$

$$F: \begin{cases} \ Ax &= 0 \\ \overline{s}^T x &= 1 \\ x &\in C \end{cases}$$

**Theorem.** A suitably-tuned standard IPM will compute a solution of $F$ in at most

$$\left[ 9\sqrt{\vartheta} \ln \left( 11\vartheta \left( 1 + \frac{1}{\text{sym}(0, H)} \right) \right) \right]$$

iterations of Newton’s method.

$\text{sym}(0, H)$ measures how deep 0 is in $H$
IPM Complexity Bound, continued

- A standard IPM will compute a solution of $F$ in at most
  \[9\sqrt{\vartheta} \ln \left(11\vartheta \left(1 + \frac{1}{\text{sym}(0, H)}\right)\right)\] iterations of Newton’s method.

- depends only on $\vartheta$ and $\text{sym}(0, H)$

- dependence on $\vartheta$ not generally observed in practice

- dependence on $\text{sym}(0, H)$ is observed in practice

- Is this the end of the story?
Projective Transformation of $H$ to Improve $\text{sym}(0, H)$

$$w = T(y) := \frac{y}{1 - \hat{v}^T y}$$

Can we conveniently construct $\hat{v}$ for which $T(y)$ improves the symmetry of the transformed image set?
Projective Transformation of $H$ to Improve $\text{sym}(0, H)$
Projective Transformation of $H$ to Improve sym$(0, H)$

This is equivalent to translating $H^\circ$ by $\hat{v}$
Main Theorem: \( \text{sym}(0, \hat{H}) = \text{sym}(\hat{\nu}, H^\circ) \)
Main Theorem: \( \text{sym}(0, \hat{H}) = \text{sym}(\hat{v}, H^{\circ}) \)

Proof uses two fact:
- invariance of \( \text{sym}(0, \cdot) \) under polarity, and
- equivalence of translation and projective transformation under polarity
Projective Transformation of $H$ to Improve $\text{sym}(0, H)$

$$y = \frac{w}{1+\hat{\nu}^tw}$$

$$w = \frac{y}{1-\hat{\nu}^ty}$$

$H^{\circ}$ is transformed to $H^{\circ \circ}$ through projective polarity twice, followed by a translation.
Conceptual Procedure for Transforming $F$

**Step 0.** Given $\bar{s} \in \text{int}C^*$ define $F$: \[
\begin{cases}
Ax = 0 \\
\bar{s}^T x = 1 \\
x \in C
\end{cases}
\]

**Step 1.** Construct $H^\circ$

**Step 2.** Compute a point $\hat{v}$ deep in $H^\circ$ (sym($\hat{v}, H^\circ$) $>> 0$)

**Step 3.** Projectively transform $F$ using $T(y) = \frac{y}{1-\hat{v}^T y}$

To make the procedure practical, we need:

- explicit/convenient form for $H^\circ$
- some simple formulas for projective transformations
- method to conveniently compute a deep point $\hat{v} \in H^\circ$
Explicit form for $H^\circ$

**Proposition:** Suppose that $0 \in H := \{ Ax : \bar{s}^T x = 1, \ x \in C \}$. Then $H^\circ = \{ v : \bar{s} - A^T v \in C^* \}$

Note that $0 \in \text{int} H^\circ$

Note that $H^\circ$ is typically convenient to work when $C$ is self-scaled
\[ \hat{v} \in \text{int}H^\circ = \{v : \bar{s} - A^Tv \in C^*\} \text{ induces PT:} \]

\[ T(y) := \frac{y}{1 - \hat{v}^Ty} \]

Define \( \hat{s} := \bar{s} - A^T\hat{v} \). Under \( T(y) \), the transformed feasibility problem is simply:

\[
\hat{F} : \begin{cases} 
Ax & = 0 \\
\hat{s}^Tx & = 1 \\
x & \in C 
\end{cases}
\]

and the transformed image set is simply:

\[ \hat{H} = \{v : \hat{s} - A^Tv \in C^*\} \]
Computing a deep Point in $H^\circ$

$$H^\circ = \{ v : \bar{s} - A^T v \in C^* \}$$

Use Hit-and-Run random walk starting at $v^0 = 0$ and try to compute approximate center of mass

Let $v^1, \ldots, v^M \sim U[S]$, and define $\hat{v} = \frac{1}{M} \sum_{i=1}^{M} v^i$
Practical Procedure for Transforming $F$

**Step 0.** Given $\bar{s} \in \text{int}C^*$ define $F$: 

$$
\begin{align*}
Ax &= 0 \\
\bar{s}^T x &= 1 \\
x &\in C
\end{align*}
$$

**Step 1.** Use Hit-and-Run to compute a deep point $\hat{v} \in H^\circ := \{v : \bar{s} - A^T v \in C^*\}$

**Step 3.** Compute $\hat{s} := \bar{s} - A^T \hat{v}$

**Step 4.** Construct the Problem $\hat{F}$: 

$$
\begin{align*}
Ax &= 0 \\
\hat{s}^T x &= 1 \\
x &\in C
\end{align*}
$$

The transformed image set is $\hat{H} := \{Ax \in \mathbb{R}^m : x \in C \ , \ \hat{s}^T x = 1\}$, and $\text{sym}(0, \hat{H}) = \text{sym}(\hat{v}, H^\circ)$
Suppose we compute $\hat{v} \in \text{int}H^\circ$ with $\text{sym}(\hat{v}, H^\circ) \geq \alpha$

- A standard IPM will compute a solution of $\hat{F}$ in at most

$$\left\lceil 9\sqrt{\vartheta} \ln \left(11\vartheta \left(1 + \frac{1}{\alpha}\right)\right) \right\rceil$$

iterations of Newton’s method

- If $\alpha \geq \frac{1}{O(m^\kappa)}$ for fixed $\kappa$, then IPM is strongly-polynomial time

- Symmetry also bounds the geometric behavior, see paper for details
Computational Experiments

100 randomly generated ill-conditioned LP feasibility problems for each $(m, n)$

Solved using SDPT3

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Average Iterations</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>n</td>
<td>Before</td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>8.48</td>
</tr>
<tr>
<td>500</td>
<td>2500</td>
<td>9.30</td>
</tr>
<tr>
<td>1000</td>
<td>5000</td>
<td>9.69</td>
</tr>
</tbody>
</table>
Computational Experiments

100 randomly generated ill-conditioned LP feasibility problems for each \((m, n)\)

Solved using SDPT3

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Average Iterations</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>m  n</td>
<td>Before</td>
<td>After</td>
</tr>
<tr>
<td>100 500</td>
<td>8.48</td>
<td>4.10</td>
</tr>
<tr>
<td>500 2500</td>
<td>9.30</td>
<td>5.17</td>
</tr>
<tr>
<td>1000 5000</td>
<td>9.69</td>
<td>5.20</td>
</tr>
</tbody>
</table>

Decrease 47% Decrease 32%

Remark: IPM iterations decreases on all 300 problems
Computation: Iterations vs. Number of Steps

The graph shows the relationship between the number of iterations (on the y-axis) and the number of steps of the random walk (on the x-axis, in a logarithmic scale). Two lines are plotted: the original median and the preconditioned median. The graph indicates that as the number of steps increases, the number of iterations decreases.
Computation: Iterations vs. Number of Steps

The graph illustrates the relationship between Iterations (on the y-axis) and the Number of Steps of the Random Walk (on the x-axis). The plot compares two conditions:
- Original 90% Band
- Preconditioned 90% Band

The number of steps is shown on a log scale, ranging from $10^6$ to $10^3$. The graph shows a downward trend in Iterations as the Number of Steps increases, with the Preconditioned 90% Band generally showing a lower number of iterations compared to the Original 90% Band.
Computations: Time vs. Number of Steps

![Graph showing the relationship between running time (in seconds) and the number of steps of a random walk (on a log scale). The graph compares the running times for the original median and the preconditioned median.](image)
Computations: Time vs. Number of Steps
Next Steps

- Further adaptation of PT method directly to optimization
- Work with linear algebra of systems to reduce work per iteration
- Application of PT method to polynomial sum-of-squares feasibility problems