A geodesic interior-point method for linear optimization over symmetric cones

Frank Permenter

Toyota Research Institute
We consider problems of form

\[
\begin{align*}
\text{minimize} & \quad \langle c, x \rangle \\
\text{subject to} & \quad Ax = b \\
& \quad x \in K
\end{align*}
\]

where $$K$$ is a symmetric cone, i.e.,

$$K = \{x^2 : x \in \text{Euclidean Jordan Algebra}\}$$

Example $$K$$:

- Nonnegative orthant: $$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0\}$$
- Second-order cone: $$\mathbb{L}^{n+1} := \{(x_0, x) \in \mathbb{R} \times \mathbb{R}^n : \|x\| \leq x_0\}$$
- Semidefinite cone: $$\mathbb{S}^n_+ := \{VV^T : V \in \mathbb{R}^{n \times n}\}$$
Euclidean Jordan algebras behave like symmetric matrices

A Euclidean Jordan algebra $\mathcal{J}$ is a commutative algebra over real inner-product space satisfying

$$\langle x \circ y, z \rangle = \langle y, x \circ z \rangle, \quad (x \circ y) \circ x^2 = x \circ (y \circ x^2)$$

Examples:

- $\mathbb{R}^n$ with $[x \circ y]_i = x_iy_i$
- $\mathbb{R} \times \mathbb{R}^n$: with $(x_0, x) \circ (y_0, y) := (x_0y_0 + x^T y, x_0x + y_0x)$.
- Symm. matrices with $X \circ Y := \frac{1}{2}(XY + YX)$

Has all the “nice” properties of symmetric matrices:

- An identity $e$, i.e., $e \circ x = x$.
- Spectral decomp. $x = \sum_{i=1}^n \lambda_i e_i$ with $\lambda_i \in \mathbb{R}$.
- ”Matrix” functions: $f(x) := \sum_{i=1}^n f(\lambda_i) e_i$ for any $f : \mathbb{R} \to \mathbb{R}$.
Euclidean Jordan algebras: invented for quantum mechanics.

- 1933: introduced as generalization of density matrices.
- 1934: fully classified (Jordan/Von Neumann/Wigner)
  - The LP, SDP, SOCP algebras (previous slide).
  - The Hermitian matrices with complex and quaternion entries
  - The 27 dim. exceptional algebra.
- A “failed” theory: the exceptional algebra only “new” object.

Euclidean Jordan algebras: a tool for optimization.

- Parallel theory of self-scaled cones: Nesterov/Todd (1990s)
- Self-scaled ⇔ symmetric: Guler.
- A successful theory: elegant software, short papers.
A primal-dual pair of symmetric cone programs:

\[
\begin{align*}
\text{min.} \quad & \langle c, x \rangle \\
\text{subj. to} \quad & Ax = b, \ x \in K \
\end{align*}
\]

\[
\begin{align*}
\text{max.} \quad & b^T y \\
\text{subj. to} \quad & s = c - A^T y, \ s \in K^* \\
\end{align*}
\]

- Feasible \((x, s)\) optimal if comp. slackness \(x \circ s = 0\) holds.
- Central-path \((\hat{x}(\mu), \hat{s}(\mu))\) solves \(\mu\)-perturbed optimality cons.

\[
x \circ s = \mu e, \\
Ax = b, \ s = c - A^T y
\]

\[
x \in K \quad s \in K
\]

- Interior-point methods approximate \((\hat{x}(\mu_i), \hat{s}(\mu_i))\) for \(\mu_i \to 0\).
# Initialize to centered-points

\[ s, x \leftarrow \hat{s}(\mu_0), \hat{x}(\mu_0) \]

**while** \( \mu > \mu_f \) **do**

- Decrease \( \mu \)
- **# Apply to central path eqs.**
  - \( x, s \leftarrow \text{NewtonsMethod}(s, x; \mu) \)

**end**

- IPMs differ by choice of comp. slackness condition:

\[
  x \circ s = e \iff Mx \circ M^{-T} s = e \quad \text{if } M(\mathcal{K}) = \mathcal{K}
\]

- “\( \sqrt{n}\)-bound”: \( O(\|e\|) \) convergence if \( M \) picked from \( (x_i, s_i) \)

- Iterates updated inside fixed *subspaces*

\[
A(x_i - x_{i+1}) = 0, \quad s_i - s_{i+1} \in \text{range } A^T,
\]
We present a new IPM for symmetric cone optimization.

Key idea: update \((s_i, x_i)\) using geodesics of \(K\) instead of subspaces such that complementarity is maintained.

\[
\begin{align*}
Ax_i &= b, \\
s_i &= c - A^T y_i, \\
\underbrace{x_i \circ s_i = \mu_i e}_{\text{existing algs}} &\text{ and } \underbrace{x_i \circ s_i = \mu_i e}_{\text{this talk}} \quad \forall \text{ iters. } i
\end{align*}
\]

Features: half the variables, primal-dual symmetric, pure Newton, affine invar., simple proofs, complexity matches state-of-art.

Remainder of talk:

Part I: The special-case of linear programming
- Log-space transformation of central-path
- A log-space IPM and \(O(\sqrt{n})\) complexity.

Part II: The generalization to symmetric cones
- From log-space to geodesics
- A geodesic IPM and \(O(\|e\|)\) complexity.

Part III: Computational results and software

minimize \( c^T x \)
subject to \( Ax = b \)
\( x \geq 0, \) i.e., \( x \in \mathbb{R}^n_+ \)
For LP, we track primal-dual central path in log-domain

Our approach: reformulate central-path conditions

\[ x_i s_i = \mu, \quad Ax = b, \quad s = c - A^T y, \quad x \geq 0, \quad s \geq 0, \]

in the log domain

\[ A(\sqrt{\mu} \exp v) = b, \quad \sqrt{\mu} \exp(-v) = c - A^T y, \]

and solve with Newton’s method.

Properties:

- Uses comp. slackness \( x_i s_i = \mu \) to reduce log parameters:

\[ v := \log \frac{1}{\sqrt{\mu}} x = - \log \frac{1}{\sqrt{\mu}} s \]

- \( v \)-parameterization ensures \( s_i x_i = \mu \).
- Primal-dual symmetric.
- Seems unanalyzed in the literature!!!
A log-space interior-point method for LP ($\mathcal{K} = \mathbb{R}^n_+$).

# Init to log of centered point.
$v \leftarrow \hat{v}(\mu_0)$

while $\mu > \mu_f$ do
    $\mu \leftarrow \exp(-2t)\mu$
    # Newton on log domain central-path eqs.
    for $i = 1, 2, \ldots, m$ do
        $v \leftarrow v + d(v, \mu)$
    end
end

$x \leftarrow \sqrt{\mu} \exp v$, $s \leftarrow \sqrt{\mu} \exp -v$

**Theorem (Main Result)**

If $(m, t)$ satisfies $2^m \geq c_1 |\log \epsilon|$ and $q(t) = c_2 \frac{1}{n}$, then

- At most $c_3 m \sqrt{n} \log \frac{\mu_0}{\mu_f}$ Newton steps execute.
- $\| \log x - \log \hat{x}(\mu) \| \leq \epsilon$ and $\| \log s - \log \hat{s}(\mu) \| \leq \epsilon$

where $q(u) := 2(\cosh u - 1)$ and $c_1, c_2, c_3$ known constants.
To update $\mu$, must estimate distance to new centered point

Natural distance is $\delta(a, b) := \|a - b\|_2$. Unfortunately,

- Hard to bound $\delta(v, \hat{v}(\mu))$ after Newton update of $v$.
- $\delta(\hat{v}(\mu_1), \hat{v}(\mu_2))$ unknown in general.

Workaround: introduce a new measure $h > \delta^2$:

$$
\delta(v, \hat{v}(\mu_2)) \leq \delta(v, \hat{v}(\mu_1)) + \delta(\hat{v}(\mu_1), \hat{v}(\mu_2)) \\
\leq \sqrt{h(v, \hat{v}(\mu_1))} + \sqrt{h(\hat{v}(\mu_1), \hat{v}(\mu_2))}
$$

easy to bound known exactly
We use hyperbolic upper-bound $h(a, b)$ of log-distance

\[ h(a, b) := \sum_{i=1}^{n} q(a_i - b_i) \text{ where } q(u) := 2(\cosh u - 1) \geq u^2 \]

Key Features

- $\|a - b\|^2 \leq h(a, b) \leq q(\|a - b\|)$.
- For $k > 0$, known exactly for centered-points $\hat{v}(\mu)$ and $\hat{v}(k\mu)$
  \[ \frac{1}{n} h(\hat{v}(\mu), \hat{v}(k\mu)) = q\left(\frac{1}{2} \log k\right) \]
- A natural Lyapunov function for Newton’s method.
  \[ \{ v : h(v, \hat{v}(\mu)) \leq \frac{1}{2} \} \text{ quadratically convergences to } \hat{v}(\mu) \]

Other properties:

- $h$ almost a metric (triangle ineq fails)
- $h(v, \hat{v})$ is linear in $\exp v$ and $\exp -v$.
  \[ h(v, \hat{v}) = \langle \exp v, \exp -\hat{v} \rangle + \langle \exp -v, \exp \hat{v} \rangle - 2n \]
Proof: $h(\hat{v}(k\mu), \hat{v}(\mu)) = n(q(\frac{1}{2} \log k))$

Recall primal-dual affine constraints $Ax = b$ and $s = c - A^T y$ and let $w = \exp \hat{v}(\mu)$, $z = \exp \hat{v}(k\mu)$. Then,

$$\sqrt{\mu}w - \sqrt{k\mu}z \in \text{null } A, \quad \sqrt{\mu}w^{-1} - \sqrt{k\mu}z^{-1} \in \text{range } A^T$$

both prim. feas. \quad both dual feas.

Hence,

$$0 = \langle w - \sqrt{k}z, w^{-1} - \sqrt{k}z^{-1} \rangle$$

Rearranging and using $\langle z, z^{-1} \rangle = \langle w, w^{-1} \rangle = n$ yields:

$$\langle z, w^{-1} \rangle + \langle w, z^{-1} \rangle = n(\sqrt{k} + \frac{1}{\sqrt{k}}) = 2n(\cosh \log(\frac{1}{2} k)).$$

Claim follows by combining $q(t) := 2 \cosh(t) - 1$ with identity

$$h(\hat{v}(k\mu), \hat{v}(\mu)) = \langle z, w^{-1} \rangle + \langle w, z^{-1} \rangle - 2n$$
Let \( f(\alpha) \) denote proximity to central path as func. of step-size \( \alpha \).

\[
f(\alpha) = h(\hat{v}(\mu), v + \alpha d) := 2 \sum_{i=1}^{n} (\cosh(v_i + \alpha d_i - \hat{v}_i) - 1)
\]

**Theorem**

*If Newton dir. \( d \) satisfies \( \|d\|_\infty^2 \leq 2 \), then \( f(1) \leq \frac{1}{2} \|d\|_\infty^2 f(0) \).*

**Pf:** Taylor’s theorem \( f(\alpha) = f(0) + f'(0)\alpha + \frac{1}{2} f''(\zeta)\alpha^2 \) for \( \zeta \in [0, \alpha] \) and lemmas:

- \( f'(0) = - (f(0) + \|d\|^2) \)
- \( f''(\zeta) \leq 2\|d\|^2 + \|d\|_\infty^2 \max_{u \in \{0, \alpha\}} f(u) \)

Quadratic converg. follows by proving: \( f(0) \leq \frac{1}{2} \Rightarrow \frac{1}{2} \|d\|^2 \leq f(0) \).
Part II: Generalization: a geodesic-interior point method for symmetric cone optimization.

minimize $\langle c, x \rangle$
subject to $Ax = b$
$x \in K$
For curve \( c : [0, 1] \to \mathbb{R}_{++} \), let
\[
L(c) := \int_0^1 \| c(t)^{-1} \circ c'(t) \| dt
\]
Let \( g(t) := \exp(t \log a + (1 - t) \log b) \) for \( a, b \in \mathbb{R}_{++}^n \).

Properties

- \( L(g) = \| \log a - \log b \| \).
- The curve \( g(t) \) is a geodesic, i.e., it minimizes \( L(c) \) over \( c(t) \) satisfying \( c(0) = a \) and \( c(1) = b \).
- Geodesic distance \( \delta(a, b) := L(g) \) is a metric on \( \mathbb{R}_{++}^n \).
Generalization to symmetric, positive definite matrices

For curve $c : [0, 1] \rightarrow S^n_{++}$, let

$$L(c) := \int_0^1 \|c^{-1/2} c'(t) c^{-1/2}\| dt$$

For $A, B \in S^n_{++}$, define $g(t)$ via matrix exp/log:

$$g(t) := A^{1/2} \exp(tD) A^{1/2}, \quad D := \log(A^{-1/2} BA^{-1/2})$$

Properties:

- $g(t)$ is geodesic between $A$ and $B$ with length $\|D\|$.
- $\log g(t)$ a line-segment if $AD = DA$, otherwise
  $$\log(A) + Dt \neq \log g(t)$$
- Geodesic distance $\delta(A, B) := L(g)$ is a metric on $S^n_{++}$.
- $[g(t)]^{-1}$ is geodesic between $A^{-1}$ and $B^{-1}$.

Reference: Bhatia, PSD Matrices.
Let $\mathcal{K} := \{x^2 : x \in \mathcal{J}\}$. For curve $c : [0, 1] \rightarrow \text{int} \mathcal{K}$, define

$$L(c) := \int_0^1 \| Q(c(t)^{-1/2})c'(t) \| dt$$

where $Q(w)z := 2w \circ (w \circ z) - (w \circ w) \circ z$.

Properties:
- $Q(w)z$ reduces to $Z \mapsto WZW$ if $\mathcal{J} = \mathbb{S}^n$.
- The geodesics are curves of form

$$g(t) := Q(a^{1/2}) \exp td, \quad \exp d := \sum_{m=0}^{\infty} \frac{1}{m!} d^m.$$

- Geodesic distance $\delta(a, b)$ a metric on $\text{int} \mathcal{K}$.
- Inverse property: if $g(t)$ is geodesic between $a$ and $b$, then

$$[g(t)]^{-1} = \text{geodesic between } a^{-1} \text{ and } b^{-1}$$

Reference: Lim, Riemannian and Finster structures of symmetric cones.
A geodesic interior-point method for symmetric cone $\mathcal{K}$

$$
\mu \leftarrow \mu_0, \ w \leftarrow \frac{1}{\sqrt{\mu}} \hat{\chi}(\mu_0)
$$

**While** $\mu > \mu_f$ **do**

$$
\mu \leftarrow \frac{1}{k} \mu
$$

**For** $i = 1, 2, \ldots, m$ **do**

$$
w \leftarrow Q(w^{1/2}) \exp d(w, \mu)
$$

**End**

**End**

$$
x \leftarrow \sqrt{\mu} w, \ s \leftarrow \sqrt{\mu} w^{-1}
$$

Iterates joined by geodesic curve $g(t) = Q(w^{1/2}) \exp td$.

$d(w, \mu)$ defined by linearizing

$$
s = \sqrt{\mu} Q(w^{1/2}) \exp d,
$$

$$
\approx \sqrt{\mu} Q(w^{1/2})(e + d),
$$

and substituting into central-path eqs.
A geodesic interior-point method for symmetric cone $\mathcal{K}$

$$
\mu \leftarrow \mu_0, \ w \leftarrow \frac{1}{\sqrt{\mu}} \hat{x}(\mu_0) \\
\textbf{while} \ \mu > \mu_f \ \textbf{do} \\
\quad \mu \leftarrow \frac{1}{k} \mu \\
\quad \textbf{for} \ i = 1, 2, \ldots, m \ \textbf{do} \\
\quad \quad w \leftarrow Q(w^{1/2}) \exp d(w, \mu) \\
\quad \textbf{end} \\
\textbf{end} \\
x \leftarrow \sqrt{\mu} w, \ s \leftarrow \sqrt{\mu} w^{-1}
$$

Iterates joined by geodesic curve $g(t) = Q(w^{1/2}) \exp td$.

Formulae for $Q(w^{1/2}) \exp d$:

- **LP:** $\exp(\log w + d)$ — reduces to log-space IPM!
- **SDP:** $W^{1/2}(\exp D)W^{1/2}$
- **SOCP:** $(2zz^T - (\det z)R) \exp d$, where $(z_0, z_1) = w^{1/2}$,

$$R(z) = (z_0, -z_1), \quad \det z = z_0^2 - \|z_1\|^2.$$
(Almost) identical convergence result holds!

Recall LP ($\mathcal{K} = \mathbb{R}_+^n$) result on centering-steps $m$ and $\mu$-update $k$:

<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If</strong> $(m, k, \epsilon)$ <strong>satisfies condition</strong> $p(m, k, \epsilon, n) \geq 0$, <strong>then</strong></td>
</tr>
<tr>
<td>• At most $c_0 m \sqrt{n} \log \frac{\mu_0}{\mu_f}$ Newton steps execute.</td>
</tr>
<tr>
<td>• $| \log x - \log \hat{x}(\mu) | \leq \epsilon$ and $| \log s - \log \hat{s}(\mu) | \leq \epsilon$</td>
</tr>
</tbody>
</table>

For general $\mathcal{K}$, use geodesic distance $\delta$ and norm of identity $\|e\|$.

<table>
<thead>
<tr>
<th>Theorem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>If</strong> $(m, k, \epsilon)$ <strong>satisfies condition</strong> $p(m, k, \epsilon, |e|) \geq 0$, <strong>then</strong></td>
</tr>
<tr>
<td>• At most $c_0 m |e| \log \frac{\mu_0}{\mu_f} m$ Newton steps execute.</td>
</tr>
<tr>
<td>• $\delta(x, \hat{x}(\mu)) \leq \epsilon$ and $\delta(s, \hat{s}(\mu)) \leq \epsilon$</td>
</tr>
</tbody>
</table>
Our log-distance upper-bound also generalizes

For \( a, b \in \text{int} \mathcal{K} \), we have \( h(a, b) \geq \delta^2(a, b) \) for \( h(a, b) := \langle a, b^{-1} \rangle + \langle a^{-1}, b \rangle - 2\|e\|^2 \)

Same key properties hold!

- \( h \) known exactly for \( \hat{w}(k\mu) \) and \( \hat{w}(\mu) \) on central path.

\[
\begin{align*}
  h(\hat{w}(k\mu), \hat{w}(\mu)) &= \|e\|^2(q(\frac{1}{2} \log k)) \\
  q(u) &= 2(\cosh u - 1)
\end{align*}
\]

- \( h \) a natural Lyapunov function for Newton’s method.

\[
\{ w : h(w, \hat{w}(\mu)) \leq \frac{1}{2} \} \text{ is quadratically convergent}
\]

Remark: for psd cone \( \mathbb{S}_+^n \), \( h \) is symmetrized KL-divergence:

\[
h(A, B) = \text{Tr} A^{-1} B + \text{Tr} AB^{-1} - 2n
\]

\[
= KL(p_A, p_B) + KL(p_B, p_A) \quad p_M := \mathcal{N}(0, M).
\]
Part III: Computation and software.
We implement a more aggressive algorithm

\[
\textbf{while } \mu > \mu_f \textbf{ do} \\
\quad \mu \leftarrow \inf \{ \mu : h_{ub}(w, \mu) \leq \gamma \} \\
\quad w \leftarrow Q(w^{1/2}) \exp \frac{1}{\alpha} d(w, \mu) \\
\textbf{end} \\
\]

\[
x \leftarrow \sqrt{\mu} w, \ s \leftarrow \sqrt{\mu} w^{-1}
\]

- Provably converges from arbitrary initial point \( w \).
- Step-size \( \frac{1}{\alpha} \) where \( \alpha = \min\{ \frac{2}{\|d\|_\infty^2}, 1 \} \)
- \( h_{ub} \) an efficiently computable upper-bound of \( h(w, \hat{w}(\mu)) \)
- Conjectured \( O(\|e\|) \) iteration bound.
Comparison with $\|e\|$ bound on random SDPs

Red: $\|e\|$ bound. Blue: implemented.
Linear convergence of $\mu$ on typical SDPs

where $n = \|e\|^2$
Try it yourself!

www.mit.edu/~fperment/solver/

- C++ compiled to Javascript. Runs in browser!
- Python/Matlab interfaces also in development.