Primal-Dual interior-point method for sum-of-log and power cones

Utkarsh Detha (MOSEK ApS)
Industrial PhD at Technical university of Denmark

SIAM Conference on Optimization(June 2, 2026)

Outline

Introduction

Conic programming

A cone program (CP) in *standard* form:

$$\begin{aligned} p_\star = \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \in \mathbf{K} \end{aligned} \tag{Primal}$$

Dual conic program:

$$\begin{aligned} d_\star = \max_{y,s} \quad & b^T y \\ \text{s.t.} \quad & A^T y + s = c \\ & s \in \mathbf{K}^* \end{aligned} \tag{Dual}$$

Assumption: $\mathbf{K}$ and $\mathbf{K}^*$ are *proper* cones and A is full row-rank.

Homogenous self-dual embedding

$$
\begin{aligned}
\min_{x,y,s,\tau,\kappa} \quad & 0 \\
\text{s.t.} \quad & Ax - b\tau && = 0 \\
& A^T y - c\tau + s && = 0 \\
& b^T y - c^T x - \kappa && = 0 \\
& x \in \mathbf{K}, \quad s \in \mathbf{K}^*, \tau \geq 0, \kappa \geq 0
\end{aligned}
$$

$$\text{where, } z = (x, \tau, s, \kappa, y)$$

- $\tau^\star > 0 \implies (\frac{x^\star}{\tau}, \frac{s^\star}{\tau}, \frac{y^\star}{\tau})$ is a primal-dual feasible solution.
- $\kappa^\star > 0 \implies b^T y^\star > 0$(primal infeasible) and/or $c^T x^\star < 0$ (dual infeasible)

The primal-dual central path

Given *strictly* feasible point z_0, points along the central path are defined for $\mu \in (0, 1]$,

$$\begin{aligned} G(z_\mu) &= \mu G(z_0) \\ s &= -\mu \nabla F(x) = \mu \tilde{s} \\ x &= -\mu \nabla F_*(s) = \mu \tilde{x} \end{aligned} \tag{1}$$

where,

$$F_*(s) = \sup_{x \in \text{int}(K)} \{-\langle x, s \rangle - F(x)\} \tag{2}$$

is the conjugate to the primal barrier.

Gradient of the conjugate barrier: computed as prescribed in Kapelevich et. al. [1]

Scaling for primal-dual methods

$W \in \mathbf{R}^{m\times m}$ s.t. $v = Wx = W^{-T}s$ and $\tilde{v} = W\tilde{x} = W^{-T}\tilde{s}$, the central path becomes,

$$\begin{aligned} G(z_\mu) &= \mu G(z_0) \\ v &= \mu\widetilde{v} \end{aligned} \tag{3}$$

- Symmetric cones: $W^T W = \nabla^2 F(w)$ for some $w \in K$ [2].
- Non-symmetric cones: Tunçel [3] prescribes generalized scaling s.t.:

$$\frac{1}{\xi\delta_F(x,s)}[\nabla^2 F(x)] \preccurlyeq W^T W \preccurlyeq \xi\delta_F(x,s)[\nabla^2 F(\tilde{x})] \tag{4}$$

for $\xi \geq 1$ and $\delta_F(x,s) = \frac{\vartheta(\mu\tilde{\mu}-1)+1}{\mu} \geq 0$

Scaling matrices for non-symmetric cones

Multi-secant generalization of Quasi-Newton update

Schnabel [4] provides a generalized BFGS update for *multiple* secant equations, i.e. the solution to:

$$\min_{T^2} \left|\left| T^2 - \mathbf{M}^{-1} \right|\right|_{W^T W}$$

$$T^2 S = X$$

$$T^2 \in \mathbb{S}^n_{++}$$

where, $S = [s, \mu\tilde{s}]$ and $X = [x, \mu\tilde{x}]$ for $x \in K$ and $s \in K^*$.

$\rightarrow$ Dahl-Andersen algorithm chooses $\mathbf{M} = \mu\nabla^2 F(x)$

Generalized BFGS update [4]

Given (x, s) such that the following holds:

$$X^T S = \begin{bmatrix} \langle x, s \rangle & \mu \langle x, \tilde{s} \rangle \\ \mu \langle \tilde{x}, s \rangle & \mu^2 \langle \tilde{x}, \tilde{s} \rangle \end{bmatrix} = \mu \vartheta \begin{bmatrix} 1 & 1 \\ 1 & \mu \tilde{\mu} \end{bmatrix} \in \mathbb{S}^2_{++} \quad (5)$$

i.e., $\mu \tilde{\mu} > 1$ (the iterates aren't exactly on $\mathcal{C}_\mu$) then the scaling is defined as:

$$(T^{-2} =) W^T W = \mathbf{M} + \underbrace{S \left(X^T S \right)^{-1} S^T}_{\text{rank-2 update}} - \underbrace{\mathbf{M} X \left(X^T \mathbf{M} X \right)^{-1} X^T \mathbf{M}}_{\text{rank-2 downdate}} \quad (6)$$

Generalized BFGS update [4]

Given (x, s) such that the following holds:

$$X_\mu \mathbf{M} X_\mu^T = \begin{bmatrix} \langle x, \mathbf{M}x \rangle & \mu \langle x, \mathbf{M}\tilde{x} \rangle \\ \mu \langle \tilde{x}, \mathbf{M}x \rangle & \mu^2 \langle \tilde{x}, \mathbf{M}\tilde{x} \rangle \end{bmatrix} \in \mathbf{S}^2_{++} \tag{7}$$

i.e., $\mu\tilde{\mu} > 1$ (the iterates aren't exactly on $\mathcal{C}_\mu$) then the scaling is defined as:

$$(T^{-2} =) W^T W = \mathbf{M} + \underbrace{S\left(X^T S\right)^{-1} S^T}_{\text{rank-2 update}} - \underbrace{\mathbf{M}X\left(X^T \mathbf{M} X\right)^{-1} X^T \mathbf{M}}_{\text{rank-2 downdate}} \tag{8}$$

A closer look at the update

Using the spectral decomposition of $X^T S$, we re-write the update term as:

$$\frac{1}{\mu\vartheta}\underbrace{\begin{bmatrix}\frac{s-\mu\tilde{s}+\mu\lambda_+\tilde{s}}{\alpha_1} & \frac{s-\mu\tilde{s}+\mu\lambda_-\tilde{s}}{\alpha_2}\end{bmatrix}}_{\widehat{S}}\begin{bmatrix}\lambda_+ & 0\\ 0 & \lambda_-\end{bmatrix}^{-1}\begin{bmatrix}\frac{s^T-\mu\tilde{s}^T+\mu\lambda_+\tilde{s}^T}{\alpha_1}\\ \frac{s^T-\mu\tilde{s}^T+\mu\lambda_-\tilde{s}^T}{\alpha_2}\end{bmatrix}$$

where

$$\lambda_\pm = \frac{\mu\tilde{\mu}+1\pm\sqrt{(\mu\tilde{\mu}-1)^2+4}}{2}$$

Note that on $\mathcal{C}_\mu$, $\lambda_+ = 2$ and $\lambda_- = 0$.

The normalizing constants are $\alpha_\pm^2 = 1+(\lambda_\pm - 1)^2$ and on $\mathcal{C}_\mu$, $\alpha_\pm = 2$

A (slightly less) closer look at the downdate

Diagonalizing $X^T\mathbf{M}X$, we re-write the update term as:

$$\mathbf{M}X(X\mathbf{M}X^T)^{-1}X^T\mathbf{M} = V\begin{bmatrix}\Theta_+ & 0 \\ 0 & \Theta_-\end{bmatrix}^{-1} V^T$$

where,

$$\Theta_\pm = \frac{\|x\|_{\mathbf{M}}^2 + \mu^2\|\tilde{x}\|_{\mathbf{M}}^2 \pm \sqrt{\left(\|x\|_{\mathbf{M}}^2 - \mu^2\|\tilde{x}\|_{\mathbf{M}}^2\right)^2 + 4\mu^2\langle x, \tilde{x}\rangle_{\mathbf{M}}^2}}{2}$$

Note that on central path, $\Theta_+ = 2\|x\|_{\mathbf{M}}^2$ and $\Theta_- = 0$.

The sum-of-logarithms cone

$$\mathbf{K}_{\log} = \mathrm{cl}\left\{(x, y, z) \in \mathbb{R}^m_{++} \times \mathbb{R}_{++} \times \mathbb{R} : y \sum_{i=1}^{m} \log(x_i/y) \geq z\right\} \quad (9)$$

An *LHSC* barrier, with $\vartheta = m + 2$ is,

$$\mathbf{F}(x, y, z) = -\log\left(y \sum_{i \in \mathcal{I}} \log\left(x_i/y\right) - z\right) - \log(y) - \sum_{i=1}^{m} \log(x_i) \quad (10)$$

$\implies$ decomposes into m exponential cones of dimension 3 each

The general power cone

Given $\sum_{i=1}^{m} \alpha_i = 1$, the general power cone [5] is defined as:

$$\mathbf{K}_{\alpha}^{(m,n)} = \left\{ (x_{\mathcal{I}}, x_{\mathcal{J}}) \in \mathbf{R}_{+}^{m} \times \mathbf{R}^{n} : \prod_{i \in \mathcal{I}} x_i^{\alpha_i} \geq ||x_{\mathcal{J}}||_2 \right\}$$

An *LHSC* barrier, with $\vartheta = m + 1$:

$$F(x) = -\log \left(\prod_{i \in \mathcal{I}} x_i^{2\alpha_i} - ||x_{\mathcal{J}}||_2^2 \right) - \sum_{i \in \mathcal{I}} (1 - \alpha_i) \log(x_i)$$

$\Longrightarrow$ decomposes into $m - 1$ power cones of dimension 3 each

Computing the search directions

Computing search directions

The system to solve at each iteration:

$$\begin{bmatrix} 0 & A & -b \\ -A^T & 0 & c \\ b^T & -c^T & 0 \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \\ \Delta \tau \end{bmatrix} - \begin{bmatrix} 0 \\ \Delta s \\ \Delta \kappa \end{bmatrix} = \begin{bmatrix} r_p \\ r_d \\ r_g \end{bmatrix}$$

$$W \Delta x + W^{-T} \Delta s = r_{xs}$$

$$\tau \Delta \kappa + \kappa \Delta \tau = r_{\tau\kappa}$$

which is reduced to:

$$\begin{bmatrix} W^T W & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & \tau^{-1}\kappa \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta \tau \end{bmatrix} = \begin{bmatrix} r_d + W^T r_{xs} \\ r_p \\ r_g + \tau^{-1} r_{\tau\kappa} \end{bmatrix}$$

Exploiting the structure of the scaling matrices

$$\begin{bmatrix} -\mathbf{D} & A^T & U & V & -c \\ A & 0 & 0 & 0 & -b \\ U^T & 0 & \mathbf{R}^{-1} & 0 & 0 \\ V^T & 0 & 0 & -\mathbf{S}^{-1} & 0 \\ -c^T & b^T & 0 & 0 & \tau^{-1}\kappa \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta t_u \\ \Delta t_v \\ \Delta \tau \end{bmatrix} = \begin{bmatrix} -r_d - W^T r_{xs} \\ r_p \\ 0 \\ 0 \\ r_g + \tau^{-1} r_{\tau\kappa} \end{bmatrix} \tag{11}$$

and now, it is Schur complements all the way down...

Dense Schur complement systems

$$\implies \begin{bmatrix} A\mathbf{D}^{-1}A^T & A\mathbf{D}^{-1}U & A\mathbf{D}^{-1}V \\ U^T\mathbf{D}^{-1}A^T & \mathbf{R}^{-1} + U^T\mathbf{D}^{-1}U & U^T\mathbf{D}^{-1}V \\ V^T\mathbf{D}^{-1}A^T & V^T\mathbf{D}^{-1}U & -\mathbf{S}^{-1} + V^T\mathbf{D}^{-1}V \end{bmatrix}$$

$$\implies \begin{bmatrix} \mathbf{R}^{-1} + U^T\mathbf{D}^{-1}U - \overline{U}^T A\mathbf{D}^{-1}U & U^T\mathbf{D}^{-1}V - \overline{U}^T A\mathbf{D}^{-1}V \\ V^T\mathbf{D}^{-1}U - \overline{V}^T A\mathbf{D}^{-1}U & -\mathbf{S}^{-1} + V^T\mathbf{D}^{-1}V - \overline{V}^T A\mathbf{D}^{-1}V \end{bmatrix}$$

- Solve using dense Cholesky factorization
- Improve solution accuracy using pre-conditioned conjugate gradient
- Static regularization in the last Schur complement system (+ PCG/IR)

The search directions of DA algorithm

$$\underbrace{\overbrace{W^{-T}\Delta s + W\Delta x = -v}^{\text{Affine}} + \gamma\mu\tilde{v}}_{\text{Affine + Centering}} - W^{-T}\eta \tag{12}$$

- Affine direction tries to reduce μ
- Centering direction improves centrality (γ is the centering parameter)
- η is the higher-order correction

$$-\frac{1}{2}\nabla^3 F(x)[\Delta x_a, \nabla^2 F(x)^{-1}\Delta s_a] \tag{13}$$

Algorithm outline

Key steps in the primal-dual algorithm for non-symmetric cones [6]

1. Starting point: $x_0 = s_0 = -\nabla F(x_0)$
2. Scaling info: compute $\mathbf{D}^{-1}$, $\mathbf{R}^{-1}$, $\mathbf{S}^{-1}$, U and V.
3. Pure affine step + Corrector evaluation
4. Affine step-length to boundary (α_a) and centering parameter γ
5. Combined step and largest step within $N\beta$ neighbourhood.
6. Check termination and loop back to step 2

Numerical results

A simple sum-of-logarithms cone program

$$
\begin{aligned}
\max \quad & z \\
\text{s.t.} \quad & Ax = b \\
& \sum x = 1 \\
& (x, 1, z) \in \mathbf{K}_{\log}
\end{aligned}
\tag{14}
$$

A simple sum-of-logarithms cone program

dim(K)	solver	μ	**Status**	**Iter**
5	N-D	5.2e-08	Optimal	9
	N-D N/C	1.3e-07	Optimal	12
	MSK	2.0e-10	Optimal	9
20	N-D	9.6e-08	Optimal	13
	N-D N/C	2.5e-07	NearOpt	14
	MSK	2.4e-11	Optimal	11
80	N-D	8.2e-08	Optimal	14
	N-D N/C	5.0e-06	Unknown	15
	MSK	1.2e-10	Optimal	11
200	N-D	7.9e-08	Optimal	38
	N-D N/C	4.1e-07	NearOpt	31
	MSK	9.1e-11	Optimal	12

SONC decomposability of polynomials

- Non-negativity of a polynomial can be certified if it can be written as a sum of non-negative circuit polynomials (SONC)
- Papp [7] describe a two-phase approach to determine the SONC bound of a polynomial
- Following results are based on models generated in phase-1 from the SdW problem set
- Each non-negative circuit corresponds to a power cone in the conic program. All the power cones are of dimension n (number of variables in the polynomial)

SONC-decomposability (phase-1) of SdW problem set

dim(K)	num(K)	solver	μ	**Status**	**Iter**
5	413	N-D	2.6e-11	Optimal	28
		MSK	1.3e-12	Optimal	21
8	393	N-D	1.1e-08	NearOpt	50
		MSK	3.0e-09	Optimal	48
11	393	N-D	1.3e-11	Optimal	32
		MSK	3.4e-10	Optimal	20
21	383	N-D	7.6e-09	NearOpt	46
		MSK	5.4e-10	Optimal	55
30	375	N-D	1.4e-10	Optimal	29
		MSK	5.9e-12	Optimal	17
41	275	N-D	2.4e-10	Optimal	22
		MSK	1.1e-12	Optimal	19

References

[1] Lea Kapelevich et al. "Computing Conjugate Barrier Information for Nonsymmetric Cones". en. In: *Journal of Optimization Theory and Applications* (Aug. 2022).

[2] Yu. E. Nesterov and M. J. Todd. "Primal-Dual Interior-Point Methods for Self-Scaled Cones". In: *SIAM Journal on Optimization* 8.2 (May 1998). Publisher: Society for Industrial and Applied Mathematics, pp. 324–364.

[3] Levent Tunçel. "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form". en. In: *Foundations of Computational Mathematics* 1.3 (July 2001), pp. 229–254.

[4] Robert B. Schnabel. *Quasi-Newton Methods Using Multiple Secant Equations ; CU-CS-247-83*. http://id.loc.gov/vocabulary/iso639-2/eng.

[5] Robert Chares. “Cones and interior point algorithms for structured convex optimization involving powers and exponentials”. PhD. Universite catholique de Louvain Ecole Polytechnique de Louvain, 2009.

[6] Joachim Dahl and Erling D. Andersen. “A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization”. en. In: *Mathematical Programming* 194.1 (July 2022), pp. 341–370.

[7] Dávid Papp. “Duality of sum of nonnegative circuit polynomials and optimal SONC bounds”. In: *Journal of Symbolic Computation* 114 (Jan. 2023), pp. 246–266.