ISMP2015 PITTSBURGH

A primal-dual Dikin affine scaling method for symmetric conic optimization

Ali Mohammad-Nezhad Tamás Terlaky

Department of Industrial and Systems Engineering
Lehigh University

July 15, 2015

Outline

Dikin affine scaling

Primal-dual Dikin affine scaling

Extension to symmetric cones

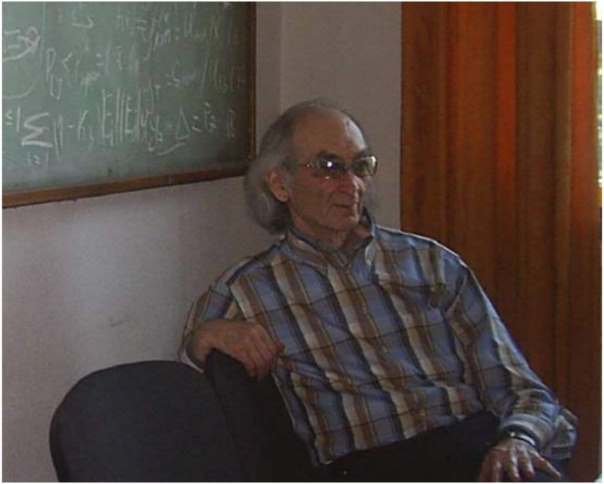
Numerical results

Outline

Semidefinite optimization

- The first idea of IPMs goes back to the work of Frisch (1955).
  - IPMs for nonlinear optimization by Fiacco and McCormic in 1960s.
  - Affine scaling direction introduced by Dikin in 1960s.
- Karmarkar (1984) revived the interest in IPMs.
- Primal-dual affine scaling methods were proposed in 1990s.
  - Monteiro, Adler, and Resende (1990).
  - Jansen, Roos, and Terlaky (1996).

Ilya I. Dikin (1936 - 2008)

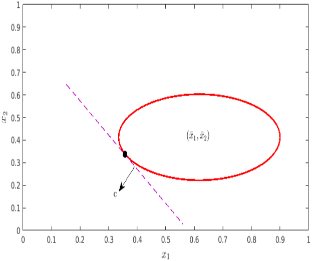


Primal affine scaling method

- Originally a primal method for linear optimization (LO).
- At each step minimizes the objective over an ellipsoid.
- The idea is to replace the nonnegativity constraint $x \geq 0$ by

$$(x - \bar{x})^T \operatorname{diag}(\bar{x})^{-2}(x - \bar{x}) \leq 1,$$

where $\bar{x}$ denotes an interior feasible solution.

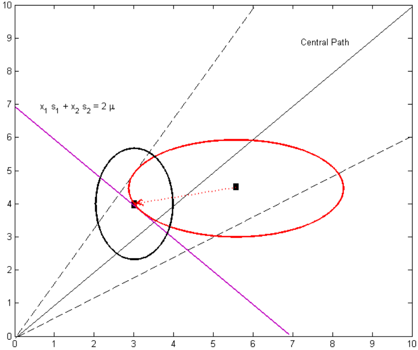


- Dikin proved that under nondegeneracy assumption, this method converges.
- The polynomial complexity of the method has still not been proved.

Outline

Illustration of the primal-dual Dikin method

- Jansen, Roos and Terlaky (1996); A primal-dual Dikin affine scaling method.
- The method has a worst-case iteration complexity $\mathcal{O}(nL)$.



Primal-dual Dikin affine scaling method

- At each step, the search directions $(\Delta x, \Delta s)$ are derived by solving

$$\begin{array}{lll} \min & x^T \Delta s + s^T \Delta x & \\ \text{s.t.} & A\Delta x & = 0, \\ & A^T \Delta y + \Delta s & = 0, \\ & \|x^{-1}\Delta x + s^{-1}\Delta s\| & \leq 1, \end{array}$$

  where $x^{-1}\Delta x$ and $s^{-1}\Delta s$ are coordinate-wise products.

Let $v = \sqrt{xs}$, $d = \sqrt{\frac{x}{s}}$, $d_x = d^{-1}\Delta x$, $d_s = d\Delta s$, and $d_y = \Delta y$.

- The scaled subproblem in the v-space is represented by

$$\begin{array}{lll} \min & v^T(d_x + d_s) & \\ \text{s.t.} & \bar{A}d_x & = 0, \\ & \bar{A}^T d_y + d_s & = 0, \\ & \|v^{-1}(d_x + d_s)\| & \leq 1, \end{array}$$

  where $\bar{A} = A\mathrm{Diag}(d)$.

Extension to semidefinite optimization (SDO)

- De Klerk, Roos, and Terlaky (1998) generalized the method to SDO.
- The search directions (D_X, D_S) are derived by solving

$$\begin{array}{lll} \min & \operatorname{trace}(V(D_X + D_S)) & \\ \text{s.t.} & \operatorname{trace}(\bar{A}_i D_X) & = 0, \quad i = 1, \ldots, m \\ & \sum_{i=1}^{m} \bar{A}_i (d_y)_i + D_S & = 0, \\ & \|V^{-\frac{1}{2}}(D_X + D_S)V^{-\frac{1}{2}}\| & \leq 1, \end{array}$$

where V is the NT scaling matrix.

D_X and D_S are defined in a similar way as in LO.

Outline

Extension to symmetric conic optimization (SCO)

- Goal: To extend the primal-dual Dikin method to SCO.
- Symmetric cone generalizes $\mathbb{R}^n_+$, $\mathcal{L}^n$, and $\mathbb{S}^n_+$.
- Euclidean Jordan Algebra (EJA) is a major tool for the analysis of SCO.
- EJA provides a unified framework for IPMs over all symmetric cones.

Symmetric cone

Let $\mathcal{J}$ be a vector space over the field of real numbers;

$\mathcal{K} \subset \mathcal{J}$ be a closed pointed convex cone;

$\mathcal{K}_+$ be the interior of $\mathcal{K}$.

$\langle x, s \rangle$ is the inner product of $x, s \in \mathcal{J}$.

The *dual* cone of $\mathcal{K}$ is

$$\mathcal{K}^* = \{s : \langle x, s \rangle \geq 0, \quad \text{for all } x \in \mathcal{K}\},$$

- $\mathcal{K}$ is *self-dual* if $\mathcal{K} = \mathcal{K}^*$.
- $\mathcal{K}$ is referred to as a *homogeneous* cone if:

  an invertible map $\mathcal{A}$ exists so that $\mathcal{A}(x) = s$ and $\mathcal{A}(\mathcal{K}) = \mathcal{K}$ for all $x, s \in \mathcal{K}_+$.
- $\mathcal{K}$ is *symmetric* if it is both self-dual and homogeneous.
- Special cases: $\mathbb{R}^n_+$, $\mathcal{L}^n$, and $\mathbb{S}^n_+$

Standard form of SCO

$$(P) \quad \min\{\langle c, x\rangle : \ Ax = b, \ x \in \mathcal{K}\},$$
$$(D) \quad \max\{b^T y : \ A^* y + s = c, \ s \in \mathcal{K}, \ y \in \mathbb{R}^m\}.$$

- $\mathcal{K}$ is a symmetric cone.
- $b \in \mathbb{R}^m$, and $c, a_i \in \mathcal{J}$ for $i = 1, ..., m$.
- $Ax = \begin{pmatrix} \langle a_1, x\rangle \\ . \\ . \\ . \\ \langle a_m, x\rangle \end{pmatrix}$; A is assumed to be full row rank.
- Interior point condition holds.

Optimality conditions and Jordan product

Optimality conditions

- Primal and dual feasibility: $Ax = b$, $A^* y + s = c$, $x, s \in \mathcal{K}$
- Optimality: $b^T y = \langle c, x \rangle \iff \langle x, s \rangle = 0$.

Jordan product

- Optimal iff $x \circ s = 0$.
  - $x \circ s = 0$ is the complementarity condition.
  - $x \circ s = L(x)s$ is a bilinear map, where $L(x)$ is a symmetric matrix.
  - Properties: $L(x)e = x$, $L(x)x^{-1} = e$, and $L(x)x = x^2$.
  - Quadratic representation of x: $P(x) = 2L^2(x) - L(x^2)$.
  - $(\mathcal{J}, \circ)$ is a commutative algebra with rank r and dimension n.

Second-order cone

Example

- The second-order cone is represented as

$$\mathcal{L}^n = \{x \in \mathbf{R}^n : x_0 \geq \|x_{1:n-1}\|\}.$$

- $e_{n\times 1} = [1, 0, ..., 0]^T$
- $L(x) := \begin{bmatrix} x_0 & x_{1:n-1}^T \\ x_{1:n-1} & x_0 I_{n-1} \end{bmatrix}.$
- $x \circ s = \begin{bmatrix} x^T s \\ x_0 s_{1:n-1} + s_0 x_{1:n-1} \end{bmatrix}.$

Scaled primal-dual problem

- The NT scaling point of x and s is defined as

$$w = P(s^{-\frac{1}{2}})(P(s^{\frac{1}{2}})x)^{\frac{1}{2}}.$$

- For $x, s \in \mathcal{K}_+$, there exists a unique $w \succ_{\mathcal{K}} 0$ so that

$$v = P(w)^{-\frac{1}{2}}x = P(w)^{\frac{1}{2}}s.$$

- The scaled primal-dual system is given by

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^* d_y + d_s &= 0, \end{aligned}$$

where

$$d_x = P(w)^{-\frac{1}{2}}\Delta x, \quad d_s = P(w)^{\frac{1}{2}}\Delta s, \quad d_y = \Delta y, \quad \bar{A} = AP(w)^{\frac{1}{2}}.$$

- The complementary gap is

$$\langle x, s\rangle = \langle P(w)^{\frac{1}{2}}v, P(w)^{-\frac{1}{2}}v\rangle = \langle v, v\rangle.$$

Extension of Dikin search directions

- The generalization of the Dikin ellipsoid is written as

$$\|P(w)^{\frac{1}{2}} x^{-1} \circ P(w)^{-\frac{1}{2}} \Delta x + P(w)^{-\frac{1}{2}} s^{-1} \circ P(w)^{\frac{1}{2}} \Delta s\|_F \leq 1,$$

which is equivalent to

$$\|v^{-1} \circ (d_x + d_s)\|_F \leq 1.$$

- The Dikin search directions are derived by solving

$$\begin{array}{lll} \min & \operatorname{trace}(v \circ (d_x + d_s)) & \\ \text{s.t.} & \bar{A} d_x & = 0, \\ & \bar{A}^* d_y + d_s & = 0, \\ & \|v^{-1} \circ (d_x + d_s)\|_F & \leq 1. \end{array} \qquad (*)$$

Proximity to the central path and feasibility

- The Dikin method keeps the iterates close to the central path.
- The measure of proximity is defined as

$$\kappa(v^2) = \frac{\lambda_{\max}(v^2)}{\lambda_{\min}(v^2)},$$

  where $\lambda_{\min}(v^2)$ and $\lambda_{\max}(v^2)$ are the min and max eigenvalues of v^2.
- Note: $\kappa(v^2) \geq 1$, with equality iff $v^2 = \mu e$.

The primal-dual Dikin method

Input

An interior feasible feasible solution (x^0, y^0, s^0)

Parameters

Proximity measure $\tau > 1$ so that $\kappa(x^0 \circ s^0) \leq \tau$
Steplength α with default value $\frac{1}{\tau\sqrt{r}}$
Accuracy parameter ϵ

$x := x^0, \;\; s := s^0$

Repeat

Obtain $(\Delta x, \Delta s)$ by solving (*)
Set $x := x + \alpha \Delta x$
Set $s := s + \alpha \Delta s$

Until $\mathrm{trace}(x \circ s) \leq \epsilon$

Iteration complexity of the Dikin method

Theorem

Let $\epsilon > 0$, $\alpha = \frac{1}{\tau\sqrt{r}}$ and $\tau > 1$ so that $\kappa(x^0 \circ s^0) \leq \tau$.

The method terminates after at most $\lceil \tau r \log \frac{\mathrm{trace}(x^0 \circ s^0)}{\epsilon} \rceil$ iterations.

It yields a feasible solution (x, s) such that $\kappa(v^2) \leq \tau$ and $\mathrm{trace}(x \circ s) \leq \epsilon$.

Outline

Numerical results

- We compare the Dikin affine scaling method and SeDuMi solver.
- 13 second-order conic problems from Dimacs library.
- SeDuMi was modified to implement the Dikin affine scaling method.
- Centering steps are introduced as safeguard.

Numerical results

	Dikin-type affine scaling				SeDuMi		
Instance	CPU/Iter	Obj	relinf	Cent	CPU/Iter	Obj	relinf
nql30n	3.6/34	-0.94602	4.1E-08	0	0.9/15	-0.94602	1.3E-08
nql30o	7.2/42	0.94603	5.2E-08	1	3.4/19	0.94603	4.0E-09
nql60n	12.7/34	-0.93505	8.2E-08	0	2.9/14	-0.93505	1.7E-08
nql60o	44.3/61	0.93513	7.6E-09	3	18.8/22	0.93516	1.7E-08
nql180n	447.7/86	-0.92772	3.0E-06	5	37/12	-0.92749	3.2E-06
nb	4.7/47	-0.05070	3.3E-09	1	1.5/20	-0.050703	4.7E-12
nb-L1	4.4/32	-13.01220	4.8E-05	0	2.1/18	-13.0122	1.0E-09
nb-L2	11.0/48	-1.62890	5.9E-05	4	3.0/16	-1.6289	1.8E-09
nb-L2-B	3.8/34	-0.10257	1.4E-12	0	1.6/16	-0.102569	2.7E-11
qssp30n	11.8/79	-6.49660	4.5E-08	12	1.3/20	-6.4966	1.5E-08
qssp30o	40.8/77	6.50280	7.0E-04	11	11.4/21	6.5182	0.0021
qssp60n	82.7/125	-6.56280	1.7E-06	30	7.7/27	-6.56269	1.2E-08
qssp60o	332.5/109	6.59330	1.2E-03	27	11.5/3	127.7891	0.0284

Conclusion

- Generalized the primal-dual Dikin affine scaling method for SCO.
- Established polynomial complexity of the method.
- EJA provides a powerful tool for the analysis of the method for SCO.
- The method is viable by the numerical results.

Thank you for your attention

Any questions?

Euclidean Jordan algebra (EJA)

Definition

Let $\mathcal{J}$ be a vector space over $\mathbb{R}^n$ with bilinear map $(x, y) \to x \circ y$.

$(\mathcal{J}, \circ)$ is referred to as EJA if for all $x, y \in \mathcal{J}$

1. $x \circ y = y \circ x$,
2. $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$ where $x^2 = x \circ x$,
3. there exists an inner product so that $\langle x \circ y, s \rangle = \langle x, y \circ s \rangle$.

There exists a unique element e such that $x \circ e = e \circ x = x$ for all $x \in \mathcal{J}$.

- EJA is commutative over the field of real numbers.
- EJA is power associative; that is $x^{p+q} = x^p \circ x^q$.

Cone of squares

Definition

The cone of squares of $(\mathcal{J}, \circ)$ is defined as

$$\mathcal{K}(\mathcal{J}) = \{x^2 : x \in \mathcal{J}\},$$

where $x^2 = x \circ x$.

- $\mathcal{K}(\mathcal{J})$ is a closed pointed convex cone with nonempty interior.

Theorem

A cone is symmetric iff it is the cone of squares of some EJA.

Rank and characteristic polynomial

Let r be the smallest integer so that $\{e, x, x^2, \cdots, x^r\}$ is linearly dependent.

- r is the degree of x denoted by $\deg(x)$ for $x \in \mathcal{J}$.
- The rank of $\mathcal{J}$ is defined as $\mathrm{rk}(\mathcal{J}) = \max_{x \in \mathcal{J}}\{\deg(x)\}$.
- x is regular if $\deg(x) = \mathrm{rk}(\mathcal{J})$.

Suppose that x is a regular element of $\mathcal{J}$.

- The characteristic polynomial of x is given by

$$\lambda^r - a_1(x)\lambda^{r-1} + ... + (-1)^r a_r(x),$$

where $a_1(x), ..., a_r(x)$ are real numbers so that

$$x^r - a_1(x)x^{r-1} + ... + (-1)^r a_r(x)e = 0.$$

Eigenvalues, trace and determinant

Definition

Let $\lambda_1, \ldots, \lambda_r$ be the roots of the characteristic polynomial.

$\lambda_1, \ldots, \lambda_r$ are called the eigenvalues of $x \in \mathcal{J}$.

- $\text{trace}(x) = \lambda_1 + ... + \lambda_r$.
- $\det(x) = \lambda_1 \lambda_2 ... \lambda_r$.
- $\langle x, y \rangle = \text{trace}(x \circ y)$.
- $\|x\|_F = \sqrt{\lambda_1^2 + \cdots + \lambda_r^2}$.
- $\|x\|_2 = \max_i |\lambda_i|$.

Spectral decomposition

Definition

A Jordan frame is a set of elements $\{q_1, \cdots, q_k\}$ of $\mathcal{J}$ so that

- q_i cannot be represented by the sum of two other elements.
- $q_i^2 = q_i$ for $i = 1, ..., k$.
- $q_i \circ q_j = 0$ for all $i \neq j$.
- $q_1 + \cdots + q_k = e$.

Theorem

Let $\mathcal{J}$ be an EJA with rank r.

Each $x \in \mathcal{J}$ can be uniquely represented as

$$x = \lambda_1 q_1 + \cdots + \lambda_r q_r,$$

where the eigenvalues are real numbers.

Second-order cone

Example

Let $x \in \mathcal{L}^n$.

- It can be shown that

$$x^2 - 2x_0 x + (x_0^2 - \|x_{1:n-1}\|^2)e = 0.$$

- $r = 2$ for this EJA.
- The spectral decomposition is given by

$$x = \lambda_1 q_1 + \lambda_2 q_2,$$

where

$$\lambda_1 = x_0 - \|x_{1:n-1}\|, \quad \lambda_2 = x_0 + \|x_{1:n-1}\|,$$
$$q_1 = \frac{1}{2}\begin{pmatrix} 1 \\ -\frac{x_{1:n-1}}{\|x_{1:n-1}\|} \end{pmatrix}, \quad q_2 = \frac{1}{2}\begin{pmatrix} 1 \\ \frac{x_{1:n-1}}{\|x_{1:n-1}\|} \end{pmatrix}.$$

Spectral decomposition

- In particular:

$$x^{\frac{1}{2}} = \lambda_1^{\frac{1}{2}} q_1 + \cdots + \lambda_k^{\frac{1}{2}} q_r,$$

$$x^{-1} = \lambda_1^{-1} q_1 + \cdots + \lambda_k^{-1} q_r.$$

- x is invertible if $\lambda_1, \ldots, \lambda_r$ are nonzero.
- $x \in \mathcal{K}$ ($\mathcal{K}_+$) if $\lambda_1, \ldots, \lambda_r$ are nonnegative (positive).