Scaling points and reach for non-self-scaled barriers

Roland Hildebrand

Laboratoire Jean Kuntzmann / CNRS

ISMP18
July 6, 2018

Outline

Conic optimization

- Barriers
- Symmetric cones
- Scaling points

Scaling points and reach

- Scaling points as orthogonal projections
- Structures on primal-dual product
- Reach property

Conic programs

Definition

A conic program over a regular convex cone $K \subset \mathbb{R}^n$ is an optimization problem of the form

$$\min_{x \in K} \langle c, x \rangle : \quad Ax = b.$$

every convex program can be transformed into a conic program

the dual program

$$\max_{s=-(A^T z - c) \in K^*} \langle b, z \rangle$$

is a conic program over the dual cone

primal-dual methods solve both problems simultaneously

Logarithmically homogeneous barriers

Definition (Nesterov, Nemirovski 1994)

Let $K \subset \mathbb{R}^n$ be a regular convex cone. A (self-concordant logarithmically homogeneous) barrier on K is a smooth function $F : K^o \to \mathbb{R}$ on the interior of K such that

- $F(\alpha x) = -\nu \log \alpha + F(x)$ (logarithmic homogeneity)
- $F''(x) \succ 0$ (convexity)
- $\lim_{x \to \partial K} F(x) = +\infty$ (boundary behaviour)
- $|F'''(x)[h, h, h]| \leq 2(F''(x)[h, h])^{3/2}$ (self-concordance)

for all tangent vectors h at x.
The homogeneity parameter ν is called the barrier parameter.

the Hessian F'' defines a Riemannian metric on the interior K^o of K

Dual barrier

Theorem (Nesterov, Nemirovski 1994)

Let $K \subset \mathbb{R}^n$ be a regular convex cone and $F : K^o \to \mathbb{R}$ a barrier on K with parameter ν. Then the Legendre transform

$$F^*(p) = \sup_{x \in K}(\langle x, -p \rangle - F(x))$$

is a barrier on K^* with parameter ν.

the map $\mathcal{D} : x \mapsto p = -F'(x)$ is an isometry between K^o and $(K^*)^o$ with respect to the Hessian metrics defined by F'', $(F^*)''$

we have $\langle x, \mathcal{D}(x) \rangle = \nu$

Central path

consider the affine subspace $\mathcal{A} = \{(x, s) \,|\, Ax = b,\ \ s = c - A^T z\}$

the intersection $\mathcal{A} \cap (K \times K^*)$ is the set of primal-dual feasible pairs

the set $\{(x, s) \in \mathcal{A} \cap (K \times K^*)^o \,|\, \exists\ \mu > 0 :\ \ s = \mu \mathcal{D}(x)\}$ is called the central path and can be parameterized by μ
note $\langle x, s \rangle = \mu\nu$ on the central path

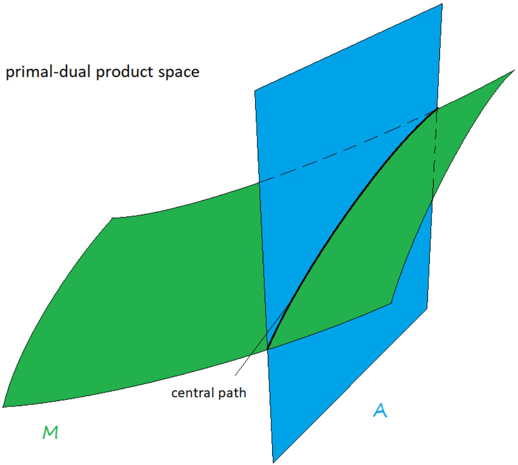
the conditions

$$(x, s) \in \mathcal{A} \cap (K \times K^*), \quad \langle x, s \rangle = 0$$

are sufficient for optimality
hence the central path tends to an optimal solution for $\mu \to 0$

path-following methods make discrete steps in the vicinity of the central path while advancing towards the solution

Geometric interpretation



$M = \{(x,s) \mid \exists\, \mu > 0 : \ \mu^{-1}s = \mathcal{D}(\mu^{-1}x)\} = \mathbb{R}_{++} \times \Gamma(\mathcal{D})$
dim $\mathcal{A} = n$, dim $M = n+1$

Symmetric cones

Definition

A commutative algebra J satisfying the condition

$$(x \bullet x) \bullet (x \bullet y) = x \bullet ((x \bullet x) \bullet y)$$

for all $x, y \in J$ is called a Jordan algebra.
A Jordan algebra is Euclidean if $\sum_{k=1}^{n} x_k \bullet x_k = 0$ implies $x_k = 0$ for all $k = 1, \ldots, n$.

the symmetric cones (self-dual homogeneous) can be represented exactly as the cones of squares $K = \{x \bullet x \,|\, x \in J\}$ of Euclidean Jordan algebras

Automorphisms and duality

for every invertible $w \in J$ the map

$$P(w) : x \mapsto 2w \bullet (w \bullet x) - (w \bullet w) \bullet x$$

is a self-adjoint automorphism of K

the duality $\mathcal{D}$ is represented by the inverse: $\mathcal{D}(x) = x^{-1}$

in particular, the central path condition $s = \mu\mathcal{D}(x)$ becomes

$$x \bullet s = \mu \cdot e$$

with e the identity element in J

Example: semi-definite matrix cone

$$X \bullet Y = \frac{XY + YX}{2}, \quad e = I, \quad \mathcal{D}(X) = X^{-1}$$

Self-scaled barriers

Definition

Let $K \subset \mathbb{R}^n$ be a regular convex cone, let K^* be its dual cone, let F be a self-concordant barrier on K with parameter ν, and let F^* be the dual barrier on K^*. Then F is called *self-scaled* if for every $x, w \in K^o$ we have

$$s = F''(w)x \in \operatorname{int} K^*, \qquad F^*(s) = F(x) - 2F(w) - \nu.$$

A cone K admitting a self-scaled barrier is called *self-scaled cone*.

Hauser, Güler, Lim, Schmieta 1998 – 2002:

- self-scaled cone $\Leftrightarrow$ symmetric cone
- self-scaled barriers on products are sums of self-scaled barriers on irreducible components
- self-scaled barriers on irreducible cones are log-determinants

Scalings

let F be a self-scaled barrier on a symmetric cone

for every $(x, s) \in (K \times K^*)^o$ there exists a unique scaling point $w \in K^o$ such that

$$F''(w)x = s$$

equivalently, there exists a self-adjoint automorphism $A = P(w^{-1})$ of K with induced automorphism $B = A^{-T} = P(w)$ of K^* such that

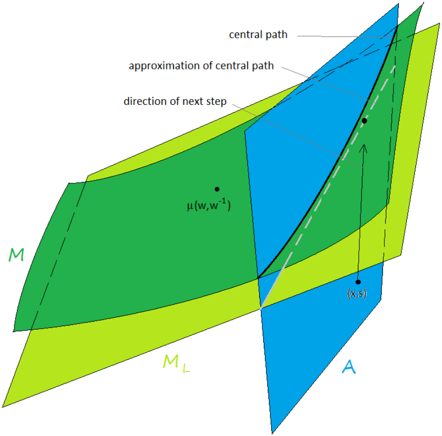
$$B(s) = A(x)$$

Nesterov-Todd type methods proceed from one primal-dual iterate (x, s) to the next by solving a linearized version of the system

$$[P(w^{-1})](x) \bullet [P(w)](s) = \mu \cdot e$$

while staying in $\mathcal{A} \cap (K \times K^*)^o$

Geometric interpretation



M_L is a linear approximation of $M = \mathbb{R}_{++} \times \Gamma(\mathcal{D})$ at $\mu(w, w^{-1})$ (equivalently at (w, w^{-1}))

Generalization to non self-scaled barriers

the geometric interpretation works independently of the self-scaled property

provided we find an adequate generalization of the scaling point w corresponding to a primal-dual pair (x, s)

[Tuncel 2001] defines the scaling point for general barriers via the property (see also [Nesterov 2006])

$$F''(w)x = s$$

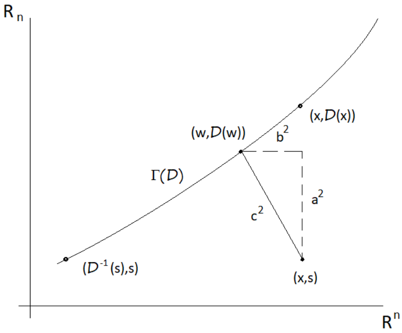
and proves existence

in [Nesterov 2006] w appears in a different context:

- scaling point is found from primal iterate
- primal-dual pair (x, s) is found from scaling point

Scaling point as nearest point

in order for the linear approximation to be accurate the scaling pair $(w, \mathcal{D}(w))$ has to be close to the current iterate



minimizer in the product metric on $(K \times K^*)^o$ is the geodesic mean

- consistent with definition for self-scaled barriers
- difficult to compute in the general case

Product of dual pair of spaces

Is there a better choice of a metric in $\mathbb{R}^n \times \mathbb{R}_n$?

neither the vector space $\mathbb{R}^n$ nor its dual $\mathbb{R}_n$ carry a canonical metric

the product $\mathbb{R}^n \times \mathbb{R}_n$ has a lot more structure

- flat pseudo-Riemannian metric
  $G((x,p);(y,q)) = \frac{1}{2}(\langle x,q\rangle + \langle y,p\rangle)$
- $\mathrm{dist}((x,p);(y,q)) = \langle x-y, p-q\rangle$
- symplectic form $\omega((x,p);(y,q)) = \frac{1}{2}(\langle x,q\rangle - \langle y,p\rangle)$

$\mathbb{R}^n \times \mathbb{R}_n$ is a flat para-Kähler space form

Duality graph as Lagrangian submanifold

let $\mathcal{D}$ be the duality map of a self-concordant barrier with parameter ν

- the duality graph $\Gamma(\mathcal{D})$ is a Lagrangian submanifold of $\mathbb{R}^n \times \mathbb{R}_n$
- the metric on $\Gamma(\mathcal{D})$ equals ν times the submanifold metric induced by $\mathbb{R}^n \times \mathbb{R}_n$
- the curvature of $\Gamma(\mathcal{D})$ is globally bounded by $\sqrt{\nu}$

similar assertions hold when passing to the product $\mathbb{R}P^{n-1} \times \mathbb{R}P_{n-1}$ of projective spaces

Consistency

the scaling pair $\mu(w, w^{-1})$ defined by the equation

$$F''(w)x = s$$

is indeed the nearest point on $\mu \cdot \Gamma(\mathcal{D})$ in the pseudo-Riemannian metric of the para-Kähler space:

$$\min_{w \in K^o} \langle x - \mu w, s + \mu F'(w) \rangle$$

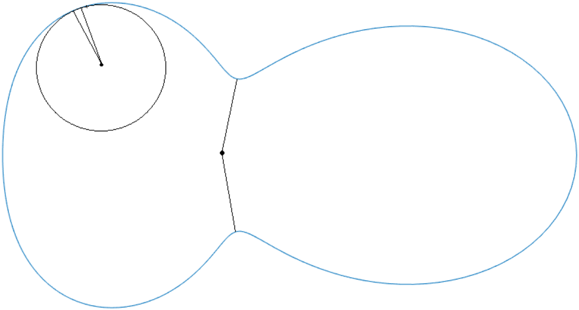
differentiating with respect to w gives the first order condition

$$-\mu(s + \mu F'(w)) + \mu F''(w)(x - \mu w) = 0$$

highlighted terms cancel by $F''(w)w = -F'(w)$

a similar minimization problem considered already in [Nesterov, Todd 1997]

Existence of nearest point



obstacles for the existence of a nearest point:

- global: points far away on the submanifold are close in ambient space
- local: curvature of the manifold

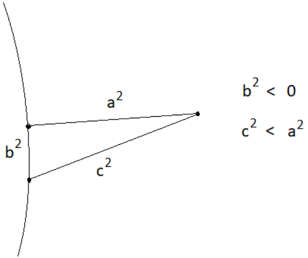
Reach property

Definition (Federer 1959)

Let $A \subset E$ be a subset of a Euclidean space.
A unique closest point of A is a point $x \in E$ such that there exists a unique point $a \in A$ with $||x - a|| = d(x, A)$.
The reach of a point $a \in A$ is the largest $r \geq 0$ such that the open ball $B_r^o(a)$ around a consists of unique closest points.
The reach of A is the infimum over $a \in A$ of the reach of a.

- A as infinite reach if and only if A is closed convex
- smooth compact connected submanifolds have positive reach
- the reach of a is continuous on A
- for smooth manifolds A the inverse of the reach is bounded from below by the curvature of A
- can be generalized to subsets of Riemannian manifolds

Reach in pseudo-Riemannian space forms



Definition

Let $M \subset \mathcal{M}$ be negative definite of maximal dimension.
A unique closest point of M is a point $x \in \mathcal{M}$ such that there exists a unique point $z \in M$ with $(a; x) = \inf_{z' \in M} d(x, z')$.
The reach of a point $z \in M$ is the largest $r \geq 0$ such that the open ball $B_r^o(z)$ around z in the normal submanifold to M at z consists of unique closest points.
The reach of M is the infimum over $z \in M$ of the reach of z.

Main result

Theorem
Let $K \subset \mathbb{R}^n$ be a regular convex cone and F a self-concordant barrier on K with parameter ν.
The corresponding Lagrangian submanifold $\Gamma(\mathcal{D}) \subset \mathbb{R}^n \times \mathbb{R}_n$ has reach $\nu^{-1/2}$.
The corresponding Lagrangian submanifold in $\mathbb{R}P^{n-1} \times \mathbb{R}P_{n-1}$ has reach $\arccos \sqrt{\frac{\nu-1}{\nu}}$.

in particular, in a tube of corresponding radius scaling points defined via the nearest point on the graph $\Gamma(\mathcal{D})$ exist and are unique

Thank you