A symmetric primal-dual algorithm for conic optimization based on the power cone

François Glineur

francois.glineur@uclouvain.be

joint work with Robert Chares
UCL/CORE (Center for Operations Research and Econometrics)
UCL/INMA (Department of Mathematical Engineering)

May 11, 2008

SIAM Conference on Optimization Boston

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Convex optimization

Nonlinear optimization

$$\min_{x \in \mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \leq 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

- Variables: finite-dimensional vector $x \in \mathbb{R}^n$
- Constraints: finite number of (in)equalities, indexed by sets I and E

Problem is convex when

- objective function f_0 is convex
- functions f_i defining inequalities $f_i(x) \leq 0$ are convex for all $i \in I$
- functions f_i defining equalities $f_i(x) = 0$ are affine for all $i \in E$

Convex optimization

Nonlinear optimization

$$\min_{x \in \mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \leq 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

- Variables: finite-dimensional vector $x \in \mathbb{R}^n$
- Constraints: finite number of (in)equalities, indexed by sets I and E

Problem is convex when

- objective function f_0 is convex
- functions f_i defining inequalities $f_i(x) \leq 0$ are convex for all $i \in I$
- functions f_i defining equalities $f_i(x) = 0$ are affine for all $i \in E$

Well-known classes of convex problems

$$\min_{x\in\mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \leq 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

1. Linear optimization (LO): f_0 and f_i are affine for all $i \in E \cup I$

$$f_i(x) = a_i^{\mathrm{T}} x - b_i$$

2. Quadratically constrained quadratic optimization (QCQO): f_0 and f_i are convex quadratic for all $i \in I$

$$f_i(x) = x^{\mathrm{T}} Q_i x + r_i^{\mathrm{T}} x + s_i \text{ with } Q_i \succeq 0$$

(equalities f_i, if present, must still be affine for $i \in E$)

2b. Convex quadratic can be rewritten using composition of squared Euclidean norm and linear (vector) function:

$$f_i(x) = \|A_i x\|^2 + (r_i^{\mathrm{T}} x + s_i) \text{ with } Q_i = A_i^{\mathrm{T}} A_i$$

Well-known classes of convex problems

$$\min_{x\in\mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \le 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

1. Linear optimization (LO): f_0 and f_i are affine for all $i \in E \cup I$

$$f_i(x) = a_i^{\mathrm{T}} x - b_i$$

2. Quadratically constrained quadratic optimization (QCQO): f_0 and f_i are convex quadratic for all $i \in I$

$$f_i(x) = x^{\mathrm{T}} Q_i x + r_i^{\mathrm{T}} x + s_i \text{ with } Q_i \succeq 0$$

(equalities f_i, if present, must still be affine for $i \in E$)

2b. Convex quadratic can be rewritten using composition of squared Euclidean norm and linear (vector) function:

$$f_i(x) = \|A_i x\|^2 + (r_i^{\mathrm{T}} x + s_i) \text{ with } Q_i = A_i^{\mathrm{T}} A_i$$

Well-known classes of convex problems

$$\min_{x\in\mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \le 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

1. Linear optimization (LO): f_0 and f_i are affine for all $i \in E \cup I$

$$f_i(x) = a_i^{\mathrm{T}} x - b_i$$

2. Quadratically constrained quadratic optimization (QCQO): f_0 and f_i are convex quadratic for all $i \in I$

$$f_i(x) = x^{\mathrm{T}} Q_i x + r_i^{\mathrm{T}} x + s_i \text{ with } Q_i \succeq 0$$

(equalities f_i, if present, must still be affine for $i \in E$)

2b. Convex quadratic can be rewritten using composition of squared Euclidean norm and linear (vector) function:

$$f_i(x) = \|A_i x\|^2 + (r_i^{\mathrm{T}} x + s_i) \text{ with } Q_i = A_i^{\mathrm{T}} A_i$$

Well-known classes of convex problems

$$\min_{x\in\mathbb{R}^n} f_0(x) \text{ such that } f_i(x) \le 0 \text{ for all } i \in I \text{ and } f_i(x) = 0 \text{ for all } i \in E$$

1. Linear optimization (LO): f_0 and f_i are affine for all $i \in E \cup I$

$$f_i(x) = a_i^{\mathrm{T}} x - b_i$$

2. Quadratically constrained quadratic optimization (QCQO): f_0 and f_i are convex quadratic for all $i \in I$

$$f_i(x) = x^{\mathrm{T}} Q_i x + r_i^{\mathrm{T}} x + s_i \text{ with } Q_i \succeq 0$$

(equalities f_i, if present, must still be affine for $i \in E$)

2b. Convex quadratic can be rewritten using composition of squared Euclidean norm and linear (vector) function:

$$f_i(x) = \|A_i x\|^2 + (r_i^{\mathrm{T}} x + s_i) \text{ with } Q_i = A_i^{\mathrm{T}} A_i$$

More classes of well-known convex problems

3. Geometric optimization (GO):
f_0 and f_i are posynomials (in exponential form) for all $i \in I$

$$f_i(x) = c_i + \sum_{j \in M_i} \exp(a_{ij}x - b_{ij})$$

Each term in the sum is the composition of exponential and affine scalar function

4. Optimization with powers: l_p-norm optimization (l_pO):
f_0 linear, f_i are affine plus sum of convex powers with affine scalar arguments for all $i \in I$

$$f_i(x) = a_{i0}x - b_{i0} + \sum_{j \in M_i} |a_{ij}x - b_{ij}|^{p_{ij}} \text{ with } p_{ij} \geq 1$$

More classes of well-known convex problems

3. Geometric optimization (GO):
f_0 and f_i are posynomials (in exponential form) for all $i \in I$

$$f_i(x) = c_i + \sum_{j \in M_i} \exp(a_{ij}x - b_{ij})$$

Each term in the sum is the composition of exponential and affine scalar function

4. Optimization with powers: l_p-norm optimization (l_pO):
f_0 linear, f_i are affine plus sum of convex powers with affine scalar arguments for all $i \in I$

$$f_i(x) = a_{i0}x - b_{i0} + \sum_{j \in M_i} |a_{ij}x - b_{ij}|^{p_{ij}} \text{ with } p_{ij} \geq 1$$

Even more classes of well-known convex problems

5. Sum-of-norm optimization (SNO):
f_0 (and f_i for all $i \in I$, if any) are convex norms with affine arguments

$$f_i(x) = \sum_{j \in M_i} \|A_{ij}x - b_{ij}\|_{p_{ij}} \text{ with } p_{ij} \geq 1$$

with $\|y\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$

6. Entropy optimization (EO):
f_0 is a sum of entropy terms, f_i are affine for all $i \in E$

$$f_0(x) = \sum_i x_i \log x_i \quad \text{(implicitly implying } x \geq 0\text{)}$$

7. Analytic centering (AC):
f_0 is a sum of logarithmic terms, f_i are affine for all $i \in I \cup E$

$$f_0(x) = -\sum_{j \in M_j} \log(a_{ij}x - b_{ij})$$

Even more classes of well-known convex problems

5. Sum-of-norm optimization (SNO):
f_0 (and f_i for all $i \in I$, if any) are convex norms with affine arguments

$$f_i(x) = \sum_{j \in M_i} \|A_{ij}x - b_{ij}\|_{p_{ij}} \text{ with } p_{ij} \geq 1$$

with $\|y\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$

6. Entropy optimization (EO):
f_0 is a sum of entropy terms, f_i are affine for all $i \in E$

$$f_0(x) = \sum_i x_i \log x_i \quad \text{(implicitly implying } x \geq 0\text{)}$$

7. Analytic centering (AC):
f_0 is a sum of logarithmic terms, f_i are affine for all $i \in I \cup E$

$$f_0(x) = -\sum_{j \in M_j} \log(a_{ij}x - b_{ij})$$

Even more classes of well-known convex problems

5. Sum-of-norm optimization (SNO):
f_0 (and f_i for all $i \in I$, if any) are convex norms with affine arguments

$$f_i(x) = \sum_{j \in M_i} \|A_{ij}x - b_{ij}\|_{p_{ij}} \text{ with } p_{ij} \geq 1$$

with $\|y\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}}$

6. Entropy optimization (EO):
f_0 is a sum of entropy terms, f_i are affine for all $i \in E$

$$f_0(x) = \sum_i x_i \log x_i \quad \text{(implicitly implying } x \geq 0)$$

7. Analytic centering (AC):
f_0 is a sum of logarithmic terms, f_i are affine for all $i \in I \cup E$

$$f_0(x) = -\sum_{j \in M_j} \log(a_{ij}x - b_{ij})$$

Properties of convex optimization

Why is it interesting to consider (or restrict yourself to) convex optimization problems?

Passive features (do not rely on knowledge of structure):

- every local minimum is a global minimum
- set of optimal solutions is convex
- optimality (KKT) conditions are necessary and sufficient (assuming some regularity condition)

Any algorithm or solver applied to a convex problem will automatically benefit from those features, even if problem is not structured

Properties of convex optimization

Why is it interesting to consider (or restrict yourself to) convex optimization problems?

Passive features (do not rely on knowledge of structure):

- ◇ every local minimum is a global minimum
- ◇ set of optimal solutions is convex
- ◇ optimality (KKT) conditions are necessary and sufficient (assuming some regularity condition)

Any algorithm or solver applied to a convex problem will automatically benefit from those features, even if problem is not structured

Properties of convex optimization

Why is it interesting to consider (or restrict yourself to) convex optimization problems?

Passive features (do not rely on knowledge of structure):

- ◇ every local minimum is a global minimum
- ◇ set of optimal solutions is convex
- ◇ optimality (KKT) conditions are necessary and sufficient (assuming some regularity condition)

Any algorithm or solver applied to a convex problem will automatically benefit from those features, even if problem is not structured

Properties of convex optimization

Active features (require knowledge of structure):

- existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- To use those, additional work is needed for each problem class, exploiting its specific structure
- Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- Allows to obtain an explicit and symmetric dual
- Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- To use those, additional work is needed for each problem class, exploiting its specific structure
- Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- Allows to obtain an explicit and symmetric dual
- Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- To use those, additional work is needed for each problem class, exploiting its specific structure
- Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- Allows to obtain an explicit and symmetric dual
- Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- ◇ existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- ◇ existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- ◇ To use those, additional work is needed for each problem class, exploiting its specific structure
- ◇ Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- ◇ Allows to obtain an explicit and symmetric dual
- ◇ Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- To use those, additional work is needed for each problem class, exploiting its specific structure
- Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- Allows to obtain an explicit and symmetric dual
- Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- ◇ existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- ◇ existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- ◇ To use those, additional work is needed for each problem class, exploiting its specific structure
- ◇ Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- ◇ Allows to obtain an explicit and symmetric dual
- ◇ Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- To use those, additional work is needed for each problem class, exploiting its specific structure
- Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- Allows to obtain an explicit and symmetric dual
- Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Properties of convex optimization

Active features (require knowledge of structure):

- ◇ existence of a dual problem strongly related to original problem (Lagrangean dual, with weak and (with assumption) strong duality)
- ◇ existence of dedicated algorithm with polynomial algorithmic complexity (a barrier-based interior-point method - IPM)

- ◇ To use those, additional work is needed for each problem class, exploiting its specific structure
- ◇ Reward for additional work is better understanding and ability to solve problems more efficiently (including large-scale)

Conic optimization helps us to streamline this effort:

- ◇ Allows to obtain an explicit and symmetric dual
- ◇ Helps us to write down the self-concordant barrier required for a polynomial-time interior-point algorithm

Overview

1. Introduction: convex and conic optimization

- ◇ Why convex optimization?
- ◇ Conic optimization: a standard form for convex optimization

2. Conic optimization based on the power cone

- ◇ Modelling with the power cone
- ◇ Finding duals with the power cone

3. A symmetric primal-dual algorithm

- ◇ Primal, dual and primal-dual interior-point methods
- ◇ A symmetric algorithm

Conic optimization

Generalization of linear optimization (e.g. dual form)

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

where a new ordering is used instead of $\leq$:

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

- Ordering defined by a set K:

$$a \preceq_K b \Leftrightarrow 0 \preceq_K b - a \quad \Leftrightarrow \quad b - a \in K$$

- Set K has to be a convex cone for useful properties of ordering to hold (and also: closed, solid and pointed for technical reasons)
- Conic optimization is completely equivalent to convex optimization
- The point of a conic formulation is to make it easier to benefit from active features of convex optimization (duality and algorithms)

Conic optimization

Generalization of linear optimization (e.g. dual form)

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

where a new ordering is used instead of $\leq$:

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

- Ordering defined by a set K:

$$a \preceq_K b \Leftrightarrow 0 \preceq_K b - a \quad \Leftrightarrow \quad b - a \in K$$

- Set K has to be a convex cone for useful properties of ordering to hold (and also: closed, solid and pointed for technical reasons)
- Conic optimization is completely equivalent to convex optimization
- The point of a conic formulation is to make it easier to benefit from active features of convex optimization (duality and algorithms)

Conic optimization

Generalization of linear optimization (e.g. dual form)

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

where a new ordering is used instead of $\leq$:

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

- Ordering defined by a set K:

$$a \preceq_K b \Leftrightarrow 0 \preceq_K b - a \quad \Leftrightarrow \quad b - a \in K$$

- Set K has to be a convex cone for useful properties of ordering to hold (and also: closed, solid and pointed for technical reasons)
- Conic optimization is completely equivalent to convex optimization
- The point of a conic formulation is to make it easier to benefit from active features of convex optimization (duality and algorithms)

Conic optimization

Generalization of linear optimization (e.g. dual form)

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

where a new ordering is used instead of $\leq$:

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

- Ordering defined by a set K:

$$a \preceq_K b \Leftrightarrow 0 \preceq_K b - a \quad \Leftrightarrow \quad b - a \in K$$

- Set K has to be a convex cone for useful properties of ordering to hold (and also: closed, solid and pointed for technical reasons)
- Conic optimization is completely equivalent to convex optimization
- The point of a conic formulation is to make it easier to benefit from active features of convex optimization (duality and algorithms)

Conic optimization

Generalization of linear optimization (e.g. dual form)

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \leq c$$

where a new ordering is used instead of $\leq$:

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

- Ordering defined by a set K:

$$a \preceq_K b \Leftrightarrow 0 \preceq_K b - a \quad \Leftrightarrow \quad b - a \in K$$

- Set K has to be a convex cone for useful properties of ordering to hold (and also: closed, solid and pointed for technical reasons)
- Conic optimization is completely equivalent to convex optimization
- The point of a conic formulation is to make it easier to benefit from active features of convex optimization (duality and algorithms)

Combining several cones

Considering several conic constraints

$$A_1^{\mathrm{T}} y \preceq_{K_1} c_1, A_2^{\mathrm{T}} y \preceq_{K_2} c_2, \ldots \text{ and } A_N^{\mathrm{T}} y \preceq_{K_N} c_N$$

which are equivalent to

$$c_1 - A_1^{\mathrm{T}} y \in K_1, c_2 - A_2^{\mathrm{T}} y \in K_2, \ldots \text{ and } c_N - A_N^{\mathrm{T}} y \in K_N$$

introduce the Cartesian product cone $\hat{K} = (K_1 \times K_2 \times \cdots K_N)$ to write

$$(c_1 - A_1^{\mathrm{T}} y, c_2 - A_2^{\mathrm{T}} y, \ldots, c_N - A_N^{\mathrm{T}} y) \in (K_1 \times K_2 \cdots \times K_N)$$

$$\begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} - \begin{pmatrix} A_1^{\mathrm{T}} \\ \vdots \\ A_N^{\mathrm{T}} \end{pmatrix} \succeq_{(K_1 \times K_2 \times \cdots K_N)} 0 \quad \Leftrightarrow \quad \hat{A}^{\mathrm{T}} y \preceq_{\hat{K}} \hat{c}$$

$\rightarrow$ for theory, a single cone $\hat{K}$ can be considered without loss of generality

Combining several cones

Considering several conic constraints

$$A_1^{\mathrm{T}} y \preceq_{K_1} c_1, A_2^{\mathrm{T}} y \preceq_{K_2} c_2, \ldots \text{ and } A_N^{\mathrm{T}} y \preceq_{K_N} c_N$$

which are equivalent to

$$c_1 - A_1^{\mathrm{T}} y \in K_1, c_2 - A_2^{\mathrm{T}} y \in K_2, \ldots \text{ and } c_N - A_N^{\mathrm{T}} y \in K_N$$

introduce the Cartesian product cone $\hat{K} = (K_1 \times K_2 \times \cdots K_N)$ to write

$$(c_1 - A_1^{\mathrm{T}} y, c_2 - A_2^{\mathrm{T}} y, \ldots, c_N - A_N^{\mathrm{T}} y) \in (K_1 \times K_2 \cdots \times K_N)$$

$$\begin{pmatrix} c_1 \\ \vdots \\ c_N \end{pmatrix} - \begin{pmatrix} A_1^{\mathrm{T}} \\ \vdots \\ A_N^{\mathrm{T}} \end{pmatrix} \succeq_{(K_1 \times K_2 \times \cdots K_N)} 0 \quad \Leftrightarrow \quad \hat{A}^{\mathrm{T}} y \preceq_{\hat{K}} \hat{c}$$

$\longrightarrow$ for theory, a single cone $\hat{K}$ can be considered without loss of generality

Duality for conic optimization

Problem

$$\max b^T y \text{ such that } A^T y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^T x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^T z \geq 0 \ \forall x \in K\}$$

- Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- To find the dual, only effort involved is computing the dual cone
- Potentially allows design of (symmetrical) primal-dual algorithms
- $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Duality for conic optimization

Problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \ \forall x \in K\}$$

- Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- To find the dual, only effort involved is computing the dual cone
- Potentially allows design of (symmetrical) primal-dual algorithms
- $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Duality for conic optimization

Problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \; \forall x \in K\}$$

- Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- To find the dual, only effort involved is computing the dual cone
- Potentially allows design of (symmetrical) primal-dual algorithms
- $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Duality for conic optimization

Problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \ \forall x \in K\}$$

- Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- To find the dual, only effort involved is computing the dual cone
- Potentially allows design of (symmetrical) primal-dual algorithms
- $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Duality for conic optimization

Problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \ \forall x \in K\}$$

- ◇ Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- ◇ To find the dual, only effort involved is computing the dual cone
- ◇ Potentially allows design of (symmetrical) primal-dual algorithms
- ◇ $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Duality for conic optimization

Problem

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y \preceq_K c$$

admits a nice explicit and symmetrical dual

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_{K^*} 0$$

based on the notion of dual cone

$$K^* = \{z \in \mathbb{R}^n \text{ such that } x^{\mathrm{T}} z \geq 0 \; \forall x \in K\}$$

- ◇ Weak duality always holds, strong duality holds with regularity assumption (existence of a strictly interior point)
- ◇ To find the dual, only effort involved is computing the dual cone
- ◇ Potentially allows design of (symmetrical) primal-dual algorithms
- ◇ $\hat{K} = (K_1 \times K_2 \times \cdots K_N) \Rightarrow \hat{K}^* = (K_1^* \times K_2^* \times \cdots K_N^*)$

Algorithm for conic optimization

- Interior-point methods can easily be applied to conic optimization
- Main ingredient is a good barrier function for every cone K involved: logarithmically homogeneous self-concordant barrier with parameter ν
  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K
  $$F(x) \to +\infty \text{ when } x \to \partial K$$
  - F is logarithmically homogeneous of degree ν:
  $$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$
  - F is self-concordant:
  $$\nabla^3 F(x)[h,h,h] \le 2\left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$
- Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (NESTEROV & NEMIROVSKI, 1994) (e.g. using standard short or long step path-following algorithm)

Algorithm for conic optimization

- Interior-point methods can easily be applied to conic optimization
- Main ingredient is a good barrier function for every cone K involved:

  logarithmically homogeneous self-concordant barrier with parameter ν

  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K

  $$F(x) \to +\infty \text{ when } x \to \partial K$$

  - F is logarithmically homogeneous of degree ν:

  $$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$

  - F is self-concordant:

  $$\nabla^3 F(x)[h,h,h] \le 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$

- Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (Nesterov & Nemirovski, 1994) (e.g. using standard short or long step path-following algorithm)

Algorithm for conic optimization

- ◇ Interior-point methods can easily be applied to conic optimization
- ◇ Main ingredient is a good barrier function for every cone K involved:

  logarithmically homogeneous self-concordant barrier with parameter ν

  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K

    $$F(x) \to +\infty \text{ when } x \to \partial K$$

  - F is logarithmically homogeneous of degree ν:

    $$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$

  - F is self-concordant:

    $$\nabla^3 F(x)[h,h,h] \leq 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$

- ◇ Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (Nesterov & Nemirovski, 1994) (e.g. using standard short or long step path-following algorithm)

Algorithm for conic optimization

- ◇ Interior-point methods can easily be applied to conic optimization
- ◇ Main ingredient is a good barrier function for every cone K involved:

  logarithmically homogeneous self-concordant barrier with parameter ν

  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K

    $$F(x) \to +\infty \text{ when } x \to \partial K$$

  - F is logarithmically homogeneous of degree ν:

    $$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$

  - F is self-concordant:

    $$\nabla^3 F(x)[h,h,h] \leq 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$

- ◇ Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (Nesterov & Nemirovski, 1994) (e.g. using standard short or long step path-following algorithm)

Algorithm for conic optimization

- ◇ Interior-point methods can easily be applied to conic optimization
- ◇ Main ingredient is a good barrier function for every cone K involved:

  logarithmically homogeneous self-concordant barrier with parameter ν

  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K

$$F(x) \to +\infty \text{ when } x \to \partial K$$

  - F is logarithmically homogeneous of degree ν:

$$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$

  - F is self-concordant:

$$\nabla^3 F(x)[h,h,h] \le 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$

- ◇ Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (Nesterov & Nemirovski, 1994) (e.g. using standard short or long step path-following algorithm)

Algorithm for conic optimization

- ◇ Interior-point methods can easily be applied to conic optimization
- ◇ Main ingredient is a good barrier function for every cone K involved:

  logarithmically homogeneous self-concordant barrier with parameter ν

  - $F : \operatorname{int} K \mapsto \mathbb{R}$ is convex and three times differentiable
  - F is a barrier for cone K

    $$F(x) \to +\infty \text{ when } x \to \partial K$$

  - F is logarithmically homogeneous of degree ν:

    $$F(tx) = F(x) - \nu \log t \text{ for all } x \in \operatorname{int} K \text{ and } t > 0$$

  - F is self-concordant:

    $$\nabla^3 F(x)[h,h,h] \leq 2 \left(\nabla^2 F(x)[h,h]\right)^{\frac{3}{2}} \text{ for all } x \in \operatorname{int} K \text{ and } h \in \mathbb{R}^n$$

- ◇ Once a good barrier is known, design of a polynomial-time algorithm can be completely straightforward (Nesterov & Nemirovski, 1994) (e.g. using standard short or long step path-following algorithm)

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

3. $K = \mathbb{S}^n_+$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}^{n_l}_+ \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}^{m_1}_+ \times \cdots \times \mathbb{S}^{m_s}_+$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \le x_0\}$$

3. $K = \mathbb{S}^n_+$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}^{n_l}_+ \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}^{m_1}_+ \times \cdots \times \mathbb{S}^{m_s}_+$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

3. $K = \mathbb{S}_+^n$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}_+^{n_l} \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}_+^{m_1} \times \cdots \times \mathbb{S}_+^{m_s}$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \le x_0\}$$

3. $K = \mathbb{S}^n_+$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}^{n_l}_+ \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}^{m_1}_+ \times \cdots \times \mathbb{S}^{m_s}_+$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

3. $K = \mathbb{S}_+^n$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}_+^{n_l} \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}_+^{m_1} \times \cdots \times \mathbb{S}_+^{m_s}$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

3. $K = \mathbb{S}_+^n$ (positive semidefinite matrices) for semidefinite optimization

- Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}_+^{n_l} \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}_+^{m_1} \times \cdots \times \mathbb{S}_+^{m_s}$$

- Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- Many solvers available for problems involving these cones

Structured conic optimization

Following 3 cones are (by far) the most commonly used

1. $K = \mathbb{R}_+$ is the standard ordering, leading to linear optimization
2. $K = \mathbb{L}^n$ leads to second-order cone optimization (including QCQO)

$$\mathbb{L}^n = \{(x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \mid \sqrt{x_1^2 + \cdots + x_n^2} \leq x_0\}$$

3. $K = \mathbb{S}^n_+$ (positive semidefinite matrices) for semidefinite optimization

- ◇ Modelling typically combines constraints defined from several of these cones

$$K = \mathbb{R}^{n_l}_+ \times \mathbb{L}^{n_1} \times \cdots \times \mathbb{L}^{n_q} \times \mathbb{S}^{m_1}_+ \times \cdots \times \mathbb{S}^{m_s}_+$$

- ◇ Many problems from various domains (e.g. mechanical and electrical engineering, finance) can be modelled using these cones
- ◇ Many solvers available for problems involving these cones

Self-scaled cones

- The three cones $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$ are self-dual: $K^* = K$
- Each corresponding conic problem admits a dual of the same type (the dual of a LO problem is a LO problem, the dual of a SDO problem is a SDO, etc.)
- Moreover, these three cones are also homogeneous: they admit self-scaled barriers that allow the design of symmetric primal-dual interior-point methods (NESTEROV & TODD, 1997)
- Unfortunately, there exists essentially no other cone that is both homogeneous and self-dual
- Not all convex problems can be modelled using only $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$

Self-scaled cones

- The three cones $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$ are self-dual: $K^* = K$
- Each corresponding conic problem admits a dual of the same type (the dual of a LO problem is a LO problem, the dual of a SDO problem is a SDO, etc.)
- Moreover, these three cones are also homogeneous: they admit self-scaled barriers that allow the design of symmetric primal-dual interior-point methods (Nesterov & Todd, 1997)
- Unfortunately, there exists essentially no other cone that is both homogeneous and self-dual
- Not all convex problems can be modelled using only $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$

Self-scaled cones

- The three cones $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$ are self-dual: $K^* = K$
- Each corresponding conic problem admits a dual of the same type (the dual of a LO problem is a LO problem, the dual of a SDO problem is a SDO, etc.)
- Moreover, these three cones are also homogeneous: they admit self-scaled barriers that allow the design of symmetric primal-dual interior-point methods (NESTEROV & TODD, 1997)

- Unfortunately, there exists essentially no other cone that is both homogeneous and self-dual
- Not all convex problems can be modelled using only $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$

Self-scaled cones

- The three cones $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$ are self-dual: $K^* = K$
- Each corresponding conic problem admits a dual of the same type (the dual of a LO problem is a LO problem, the dual of a SDO problem is a SDO, etc.)
- Moreover, these three cones are also homogeneous: they admit self-scaled barriers that allow the design of symmetric primal-dual interior-point methods (Nesterov & Todd, 1997)

- Unfortunately, there exists essentially no other cone that is both homogeneous and self-dual
- Not all convex problems can be modelled using only $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$

Self-scaled cones

- The three cones $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$ are self-dual: $K^* = K$
- Each corresponding conic problem admits a dual of the same type (the dual of a LO problem is a LO problem, the dual of a SDO problem is a SDO, etc.)
- Moreover, these three cones are also homogeneous: they admit self-scaled barriers that allow the design of symmetric primal-dual interior-point methods (Nesterov & Todd, 1997)

- Unfortunately, there exists essentially no other cone that is both homogeneous and self-dual
- Not all convex problems can be modelled using only $\mathbb{R}_+$, $\mathbb{L}^n$ and $\mathbb{S}^n_+$

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Well-known convex problems and self-scaled cones

1. Linear optimization: OK with cone $\mathbb{R}_+$
2. Quadratically constrained quadratic optimization: OK with cone $\mathbb{L}^n$
3. Geometric optimization: cannot be modelled exactly, approximation with several $\mathbb{L}^n$ (size of model increases with accuracy required)
4. Optimization of p-powers: OK when $p = 1$ ($\mathbb{R}_+$) or $p = 2$ ($\mathbb{L}^n$), possible but complicated when p is rational with several $\mathbb{L}^n$ (size of model increases with size of numerator/denominator of p)
5. Sum-of-norm optimization: OK when $p = 2$ with $\mathbb{L}^n$, cannot be modelled directly when $p \neq 2$
6. Entropy optimization: cannot be modelled directly
7. Analytic centering: cannot be modelled directly (but a modified algorithm can be used)

In this talk, we introduce the power cone to model all of these exactly and design a symmetric primal-dual algorithm for it

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

- first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

- then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

- $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

- first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

- then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

- $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

- first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

- then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

  which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

- $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

- ◇ first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

- ◇ then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

- ◇ $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

◇ first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

◇ then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

◇ $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

How to create a convex cone from a function

Let $f : C \subseteq \mathbb{R} \mapsto \mathbb{R}$ a univariate (closed) convex function

In order to create a convex cone with f, we can

- ◇ first create a convex set from function f: take the epigrah

$$\operatorname{epi} f = \{(x, y) \mid f(x) \leq y\}$$

- ◇ then create a convex cone from $\operatorname{epi} f$: take the closed conic hull

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(v, z) \in \operatorname{epi} f \times \mathbb{R}_{++} \mid \frac{v}{z} \in \operatorname{epi} f\}$$

which can equivalently be written as

$$\operatorname{cone} \operatorname{epi} f = \operatorname{cl}\{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\frac{x}{z}) \leq \frac{y}{z}\}$$

- ◇ $\operatorname{cone} \operatorname{epi} f = K(f)$ is therefore a closed 3-dimensional convex cone

The power cone

Choose $f : x \mapsto |x|^{\frac{1}{\alpha}}$ with α a real parameter satisfying $0 \le \alpha \le 1$

- f is a closed convex function for every $0 \le \alpha \le 1$
- The corresponding convex cone $\mathcal{K}_\alpha$ is called the power cone

$$\mathcal{K}_\alpha = \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{++} \mid \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z}\}$$

where the defining condition can be rewritten

$$\left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z} \Leftrightarrow \frac{|x|}{z} \le (\frac{y}{z})^\alpha \Leftrightarrow |x| \le y^\alpha z^{1-\alpha}$$

and finally

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le y^\alpha z^{1-\alpha}\}$$

The power cone

Choose $f : x \mapsto |x|^{\frac{1}{\alpha}}$ with α a real parameter satisfying $0 \le \alpha \le 1$

- f is a closed convex function for every $0 \le \alpha \le 1$
- The corresponding convex cone $\mathcal{K}_\alpha$ is called the power cone

$$\mathcal{K}_\alpha = \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{++} \mid \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z}\}$$

where the defining condition can be rewritten

$$\left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z} \Leftrightarrow \frac{|x|}{z} \le \left(\frac{y}{z}\right)^\alpha \Leftrightarrow |x| \le y^\alpha z^{1-\alpha}$$

and finally

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le y^\alpha z^{1-\alpha}\}$$

The power cone

Choose $f : x \mapsto |x|^{\frac{1}{\alpha}}$ with α a real parameter satisfying $0 \leq \alpha \leq 1$

- f is a closed convex function for every $0 \leq \alpha \leq 1$
- The corresponding convex cone $\mathcal{K}_\alpha$ is called the power cone

$$\mathcal{K}_\alpha = \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{++} \mid \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

where the defining condition can be rewritten

$$\left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z} \Leftrightarrow \frac{|x|}{z} \leq \left(\frac{y}{z}\right)^\alpha \Leftrightarrow |x| \leq y^\alpha z^{1-\alpha}$$

and finally

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq y^\alpha z^{1-\alpha}\}$$

The power cone

Choose $f : x \mapsto |x|^{\frac{1}{\alpha}}$ with α a real parameter satisfying $0 \leq \alpha \leq 1$

- f is a closed convex function for every $0 \leq \alpha \leq 1$
- The corresponding convex cone $\mathcal{K}_\alpha$ is called the power cone

$$\mathcal{K}_\alpha = \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{++} \mid \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

where the defining condition can be rewritten

$$\left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z} \Leftrightarrow \frac{|x|}{z} \leq \left(\frac{y}{z}\right)^\alpha \Leftrightarrow |x| \leq y^\alpha z^{1-\alpha}$$

and finally

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq y^\alpha z^{1-\alpha}\}$$

The power cone

Choose $f : x \mapsto |x|^{\frac{1}{\alpha}}$ with α a real parameter satisfying $0 \le \alpha \le 1$

- f is a closed convex function for every $0 \le \alpha \le 1$
- The corresponding convex cone $\mathcal{K}_\alpha$ is called the power cone

$$\mathcal{K}_\alpha = \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{++} \mid \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z}\}$$

where the defining condition can be rewritten

$$\left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \le \frac{y}{z} \Leftrightarrow \frac{|x|}{z} \le (\frac{y}{z})^\alpha \Leftrightarrow |x| \le y^\alpha z^{1-\alpha}$$

and finally

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le y^\alpha z^{1-\alpha}\}$$

Modelling with the power cone

Each well-known class of structured convex problems presented earlier can be formulated as conic optimization based on the power cone

1. Linear optimization: OK since $\mathcal{K}_\alpha$ contains two copies of $\mathbb{R}_+$
2. Optimization of p-powers: OK since

$$(x, y, 1) \in \mathcal{K}_\alpha \Leftrightarrow |x| \leq y^\alpha \Leftrightarrow |x|^{\frac{1}{\alpha}} \leq y \Leftrightarrow |x|^p \leq y$$

   when choosing $p = \frac{1}{\alpha} \geq 1$
   $\rightarrow$ use dual to formulate l_p-norm optimization
3. Optimization of p-norms: note that for $z \in \mathbb{R}^n$, the epigraph of the norm

$$\|z\|_p \leq t \quad \Leftrightarrow \quad |z_1|^p + |z_2|^p \cdots |z_n|^p \leq t^p \text{ and } t \geq 0$$

   (with $p \geq 1$) is not separable and cannot be formulated using a combination of epigraphs of convex powers $|z_i|^p \leq x$

Modelling with the power cone

Each well-known class of structured convex problems presented earlier can be formulated as conic optimization based on the power cone

1. Linear optimization: OK since $\mathcal{K}_\alpha$ contains two copies of $\mathbb{R}_+$
2. Optimization of p-powers: OK since

$$(x, y, 1) \in \mathcal{K}_\alpha \Leftrightarrow |x| \leq y^\alpha \Leftrightarrow |x|^{\frac{1}{\alpha}} \leq y \Leftrightarrow |x|^p \leq y$$

   when choosing $p = \frac{1}{\alpha} \geq 1$
   $\rightarrow$ use dual to formulate l_p-norm optimization
3. Optimization of p-norms: note that for $z \in \mathbb{R}^n$, the epigraph of the norm

$$\|z\|_p \leq t \quad \Leftrightarrow \quad |z_1|^p + |z_2|^p \cdots |z_n|^p \leq t^p \text{ and } t \geq 0$$

   (with $p \geq 1$) is not separable and cannot be formulated using a combination of epigraphs of convex powers $|z_i|^p \leq x$

Modelling with the power cone

Each well-known class of structured convex problems presented earlier can be formulated as conic optimization based on the power cone

1. Linear optimization: OK since $\mathcal{K}_\alpha$ contains two copies of $\mathbb{R}_+$
2. Optimization of p-powers: OK since

$$(x, y, 1) \in \mathcal{K}_\alpha \Leftrightarrow |x| \leq y^\alpha \Leftrightarrow |x|^{\frac{1}{\alpha}} \leq y \Leftrightarrow |x|^p \leq y$$

when choosing $p = \frac{1}{\alpha} \geq 1$
$\rightarrow$ use dual to formulate l_p-norm optimization

3. Optimization of p-norms: note that for $z \in \mathbb{R}^n$, the epigraph of the norm

$$\|z\|_p \leq t \quad \Leftrightarrow \quad |z_1|^p + |z_2|^p \cdots |z_n|^p \leq t^p \text{ and } t \geq 0$$

(with $p \geq 1$) is not separable and cannot be formulated using a combination of epigraphs of convex powers $|z_i|^p \leq x$

Modelling with the power cone

Each well-known class of structured convex problems presented earlier can be formulated as conic optimization based on the power cone

1. Linear optimization: OK since $\mathcal{K}_\alpha$ contains two copies of $\mathbb{R}_+$
2. Optimization of p-powers: OK since

$$(x, y, 1) \in \mathcal{K}_\alpha \Leftrightarrow |x| \leq y^\alpha \Leftrightarrow |x|^{\frac{1}{\alpha}} \leq y \Leftrightarrow |x|^p \leq y$$

   when choosing $p = \frac{1}{\alpha} \geq 1$
   $\rightarrow$ use dual to formulate l_p-norm optimization
3. Optimization of p-norms: note that for $z \in \mathbb{R}^n$, the epigraph of the norm

$$\|z\|_p \leq t \quad \Leftrightarrow \quad |z_1|^p + |z_2|^p \cdots |z_n|^p \leq t^p \text{ and } t \geq 0$$

   (with $p \geq 1$) is not separable and cannot be formulated using a combination of epigraphs of convex powers $|z_i|^p \leq x$

Modelling with the power cone (II)

3. However, we can use the following trick:

$$
\begin{aligned}
& |z_1|^p + |z_2|^p \cdots |z_n|^p \le t^p \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p + |\frac{z_2}{t}|^p \cdots |\frac{z_n}{t}|^p \le 1 \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p \le \frac{x_1}{t}, \ \cdots, \ |\frac{z_n}{t}|^p \le \frac{x_n}{t} \text{ and } \frac{x_1}{t} + \cdots + \frac{x_n}{t} = 1 \\
\Leftrightarrow \quad & (x_1, t, y_1) \in K_{\frac{1}{p}}, \ \cdots, \ (x_n, t, y_n) \in K_{\frac{1}{p}} \text{ and } x_1 + \cdots + x_n = t
\end{aligned}
$$

which can be modelled with a conic optimization problem involving n power cones ; therefore, sum-of-norm optimization is OK

4. In particular, this construction with $\alpha = \frac{1}{2}$ gives the standard second-order cone

   $\rightarrow$ quadratically constrained quadratic optimization is OK

5. The remaining problems (geometric, entropy, analytic centering) involve logarithms or exponentials and seem out of reach for the power cone

Modelling with the power cone (II)

3. However, we can use the following trick:

$$
\begin{aligned}
& |z_1|^p + |z_2|^p \cdots |z_n|^p \le t^p \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p + |\frac{z_2}{t}|^p \cdots |\frac{z_n}{t}|^p \le 1 \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p \le \frac{x_1}{t}, \ \cdots, \ |\frac{z_n}{t}|^p \le \frac{x_n}{t} \text{ and } \frac{x_1}{t} + \cdots + \frac{x_n}{t} = 1 \\
\Leftrightarrow \quad & (x_1, t, y_1) \in K_{\frac{1}{p}}, \ \cdots, \ (x_n, t, y_n) \in K_{\frac{1}{p}} \text{ and } x_1 + \cdots + x_n = t
\end{aligned}
$$

which can be modelled with a conic optimization problem involving n power cones ; therefore, sum-of-norm optimization is OK

4. In particular, this construction with $\alpha = \frac{1}{2}$ gives the standard second-order cone
   $\rightarrow$ quadratically constrained quadratic optimization is OK

5. The remaining problems (geometric, entropy, analytic centering) involve logarithms or exponentials and seem out of reach for the power cone

Modelling with the power cone (II)

3. However, we can use the following trick:

$$
\begin{aligned}
& |z_1|^p + |z_2|^p \cdots |z_n|^p \leq t^p \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p + |\frac{z_2}{t}|^p \cdots |\frac{z_n}{t}|^p \leq 1 \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p \leq \frac{x_1}{t}, \ \cdots, \ |\frac{z_n}{t}|^p \leq \frac{x_n}{t} \text{ and } \frac{x_1}{t} + \cdots + \frac{x_n}{t} = 1 \\
\Leftrightarrow \quad & (x_1, t, y_1) \in K_{\frac{1}{p}}, \ \cdots, \ (x_n, t, y_n) \in K_{\frac{1}{p}} \text{ and } x_1 + \cdots + x_n = t
\end{aligned}
$$

which can be modelled with a conic optimization problem involving n power cones ; therefore, sum-of-norm optimization is OK

4. In particular, this construction with $\alpha = \frac{1}{2}$ gives the standard second-order cone

   $\rightarrow$ quadratically constrained quadratic optimization is OK

5. The remaining problems (geometric, entropy, analytic centering) involve logarithms or exponentials and seem out of reach for the power cone

Modelling with the power cone (II)

3. However, we can use the following trick:

$$\begin{aligned}
& |z_1|^p + |z_2|^p \cdots |z_n|^p \le t^p \\
\Leftrightarrow\ & |\frac{z_1}{t}|^p + |\frac{z_2}{t}|^p \cdots |\frac{z_n}{t}|^p \le 1 \\
\Leftrightarrow\ & |\frac{z_1}{t}|^p \le \frac{x_1}{t},\ \cdots,\ |\frac{z_n}{t}|^p \le \frac{x_n}{t} \text{ and } \frac{x_1}{t} + \cdots + \frac{x_n}{t} = 1 \\
\Leftrightarrow\ & (x_1, t, y_1) \in K_{\frac{1}{p}},\ \cdots,\ (x_n, t, y_n) \in K_{\frac{1}{p}} \text{ and } x_1 + \cdots + x_n = t
\end{aligned}$$

which can be modelled with a conic optimization problem involving n power cones ; therefore, sum-of-norm optimization is OK

4. In particular, this construction with $\alpha = \frac{1}{2}$ gives the standard second-order cone

   $\rightarrow$ quadratically constrained quadratic optimization is OK

5. The remaining problems (geometric, entropy, analytic centering) involve logarithms or exponentials and seem out of reach for the power cone

Modelling with the power cone (II)

3. However, we can use the following trick:

$$\begin{aligned}
& |z_1|^p + |z_2|^p \cdots |z_n|^p \le t^p \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p + |\frac{z_2}{t}|^p \cdots |\frac{z_n}{t}|^p \le 1 \\
\Leftrightarrow \quad & |\frac{z_1}{t}|^p \le \frac{x_1}{t}, \; \cdots, \; |\frac{z_n}{t}|^p \le \frac{x_n}{t} \text{ and } \frac{x_1}{t} + \cdots + \frac{x_n}{t} = 1 \\
\Leftrightarrow \quad & (x_1, t, y_1) \in K_{\frac{1}{p}}, \; \cdots, \; (x_n, t, y_n) \in K_{\frac{1}{p}} \text{ and } x_1 + \cdots + x_n = t
\end{aligned}$$

   which can be modelled with a conic optimization problem involving n power cones ; therefore, sum-of-norm optimization is OK

4. In particular, this construction with $\alpha = \frac{1}{2}$ gives the standard second-order cone
   $\rightarrow$ quadratically constrained quadratic optimization is OK
5. The remaining problems (geometric, entropy, analytic centering) involve logarithms or exponentials and seem out of reach for the power cone

Modelling with the power cone (III)

5. However, we can use the following well-known limit

$$\lim_{\alpha \to 0^+} (1 + \alpha x)^{\frac{1}{\alpha}} = e^x \ ,$$

valid for any real x, to obtain the exponential function

Letting $x = z + \alpha x'$ (a linear transformation), the definition of the power cone becomes:

$$y^\alpha z^{1-\alpha} \geq |z + \alpha x'| \iff y^\alpha z^{-\alpha} \geq \left|1 + \alpha \frac{x'}{z}\right| \iff \frac{y}{z} \geq \left|1 + \alpha \frac{x'}{z}\right|^{\frac{1}{\alpha}}$$

$$\iff \left|1 + \alpha \left(\frac{x'}{z}\right)\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}$$

which, when taking the limit $\alpha \to 0$, gives the condition

$$\exp\left(\frac{x'}{z}\right) \leq \frac{y}{z}$$

Modelling with the power cone (III)

5. However, we can use the following well-known limit

$$\lim_{\alpha \to 0^+} (1 + \alpha x)^{\frac{1}{\alpha}} = e^x \ ,$$

valid for any real x, to obtain the exponential function
Letting $x = z + \alpha x'$ (a linear transformation), the definition of the power cone becomes:

$$y^\alpha z^{1-\alpha} \geq |z + \alpha x'| \ \Leftrightarrow \ y^\alpha z^{-\alpha} \geq \left| 1 + \alpha \frac{x'}{z} \right| \Leftrightarrow \frac{y}{z} \geq \left| 1 + \alpha \frac{x'}{z} \right|^{\frac{1}{\alpha}}$$

$$\Leftrightarrow \ \left| 1 + \alpha \left(\frac{x'}{z} \right) \right|^{\frac{1}{\alpha}} \leq \frac{y}{z}$$

which, when taking the limit $\alpha \to 0$, gives the condition

$$\exp\left(\frac{x'}{z} \right) \leq \frac{y}{z}$$

Modelling with the power cone (III)

5. However, we can use the following well-known limit

$$\lim_{\alpha \to 0^+} (1+\alpha x)^{\frac{1}{\alpha}} = e^x \ ,$$

valid for any real x, to obtain the exponential function
Letting $x = z + \alpha x'$ (a linear transformation), the definition of the power cone becomes:

$$y^\alpha z^{1-\alpha} \geq \left|z + \alpha x'\right| \ \Leftrightarrow \ y^\alpha z^{-\alpha} \geq \left|1 + \alpha \frac{x'}{z}\right| \Leftrightarrow \frac{y}{z} \geq \left|1 + \alpha \frac{x'}{z}\right|^{\frac{1}{\alpha}}$$

$$\Leftrightarrow \ \left|1 + \alpha \left(\frac{x'}{z}\right)\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}$$

which, when taking the limit $\alpha \to 0$, gives the condition

$$\exp\left(\frac{x'}{z}\right) \leq \frac{y}{z}$$

Modelling with the power cone (III)

5. However, we can use the following well-known limit

$$\lim_{\alpha \to 0^+} (1 + \alpha x)^{\frac{1}{\alpha}} = e^x \ ,$$

valid for any real x, to obtain the exponential function
Letting $x = z + \alpha x'$ (a linear transformation), the definition of the power cone becomes:

$$y^\alpha z^{1-\alpha} \geq \left| z + \alpha x' \right| \ \Leftrightarrow \ y^\alpha z^{-\alpha} \geq \left| 1 + \alpha \frac{x'}{z} \right| \Leftrightarrow \frac{y}{z} \geq \left| 1 + \alpha \frac{x'}{z} \right|^{\frac{1}{\alpha}}$$

$$\Leftrightarrow \ \left| 1 + \alpha \left(\frac{x'}{z} \right) \right|^{\frac{1}{\alpha}} \leq \frac{y}{z}$$

which, when taking the limit $\alpha \to 0$, gives the condition

$$\exp\left(\frac{x'}{z} \right) \leq \frac{y}{z}$$

Modelling with the power cone (IV)

5. Finally, the limit when $\alpha \to 0$ of the image of the cone $\mathcal{K}_\alpha$ by the linear transformation $x' = \frac{x-z}{\alpha}$ is the exponential cone:

$$\mathcal{E}_p = \{(x', y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \exp(\frac{x'}{z}) \leq \frac{y}{z}\}$$

We can now model the epigraph of exponential function (add the linear constraint $z = 1$)
$\rightarrow$ geometric optimization is OK

6. We can also model the epigraph of minus logarithm

$$(-x, y, 1) \in \mathcal{E}_p \Leftrightarrow \exp(-x) \leq y \Leftrightarrow -\log y \leq x$$

$\rightarrow$ analytic centering is OK

7. Finally, we can also model the epigraph of entropy:

$$(-x, 1, z) \in \mathcal{E}_p \Leftrightarrow \exp(-x/z) \leq 1/z \Leftrightarrow -x/z \leq -\log z \Leftrightarrow z \log z \leq x$$

$\rightarrow$ entropy optimization is OK

Modelling with the power cone (IV)

5. Finally, the limit when $\alpha \to 0$ of the image of the cone $\mathcal{K}_\alpha$ by the linear transformation $x' = \frac{x-z}{\alpha}$ is the exponential cone:

$$\mathcal{E}_p = \{(x', y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \exp\left(\frac{x'}{z}\right) \leq \frac{y}{z}\}$$

   We can now model the epigraph of exponential function (add the linear constraint $z = 1$)
   $\to$ geometric optimization is OK

6. We can also model the epigraph of minus logarithm

$$(-x, y, 1) \in \mathcal{E}_p \Leftrightarrow \exp(-x) \leq y \Leftrightarrow -\log y \leq x$$

   $\to$ analytic centering is OK

7. Finally, we can also model the epigraph of entropy:

$$(-x, 1, z) \in \mathcal{E}_p \Leftrightarrow \exp(-x/z) \leq 1/z \Leftrightarrow -x/z \leq -\log z \Leftrightarrow z \log z \leq x$$

   $\to$ entropy optimization is OK

Modelling with the power cone (IV)

5. Finally, the limit when $\alpha \to 0$ of the image of the cone $\mathcal{K}_\alpha$ by the linear transformation $x' = \frac{x-z}{\alpha}$ is the exponential cone:

$$\mathcal{E}_p = \{(x', y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \exp(\frac{x'}{z}) \leq \frac{y}{z}\}$$

   We can now model the epigraph of exponential function (add the linear constraint $z = 1$)
   $\rightarrow$ geometric optimization is OK
6. We can also model the epigraph of minus logarithm

$$(-x, y, 1) \in \mathcal{E}_p \Leftrightarrow \exp(-x) \leq y \Leftrightarrow -\log y \leq x$$

   $\rightarrow$ analytic centering is OK
7. Finally, we can also model the epigraph of entropy:

$$(-x, 1, z) \in \mathcal{E}_p \Leftrightarrow \exp(-x/z) \leq 1/z \Leftrightarrow -x/z \leq -\log z \Leftrightarrow z \log z \leq x$$

   $\rightarrow$ entropy optimization is OK

Modelling with the power cone (IV)

5. Finally, the limit when $\alpha \to 0$ of the image of the cone $\mathcal{K}_\alpha$ by the linear transformation $x' = \frac{x-z}{\alpha}$ is the exponential cone:

$$\mathcal{E}_p = \{(x', y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \exp\left(\frac{x'}{z}\right) \leq \frac{y}{z}\}$$

   We can now model the epigraph of exponential function (add the linear constraint $z = 1$)
   $\to$ geometric optimization is OK

6. We can also model the epigraph of minus logarithm

$$(-x, y, 1) \in \mathcal{E}_p \Leftrightarrow \exp(-x) \leq y \Leftrightarrow -\log y \leq x$$

   $\to$ analytic centering is OK

7. Finally, we can also model the epigraph of entropy:

$$(-x, 1, z) \in \mathcal{E}_p \Leftrightarrow \exp(-x/z) \leq 1/z \Leftrightarrow -x/z \leq -\log z \Leftrightarrow z \log z \leq x$$

   $\to$ entropy optimization is OK

Combining different types of constraints

Part of the usefulness of this framework is that it allows combinations of different types of constraints in a completely seamless way

An example: the Lambert W function, defined by $W(x) \exp W(x) = x$

From *MathWorld*: Banwell and Jayakumar (2000) showed that a W-function describes the relation between voltage, current and resistance in a diode, and Packel and Yuen (2004) applied the W-function to a ballistic projectile in the presence of air resistance. Other applications have been discovered in statistical mechanics, quantum chemistry, combinatorics, enzyme kinetics, the physiology of vision, the engineering of thin films, hydrology, and the analysis of algorithms (Hayes 2005).

Combining different types of constraints

Part of the usefulness of this framework is that it allows combinations of different types of constraints in a completely seamless way

An example: the Lambert W function, defined by $W(x) \exp W(x) = x$

From *MathWorld*: Banwell and Jayakumar (2000) showed that a W-function describes the relation between voltage, current and resistance in a diode, and Packel and Yuen (2004) applied the W-function to a ballistic projectile in the presence of air resistance. Other applications have been discovered in statistical mechanics, quantum chemistry, combinatorics, enzyme kinetics, the physiology of vision, the engineering of thin films, hydrology, and the analysis of algorithms (Hayes 2005).

An example: the Lambert W function

$W(x)$ is real for $x \geq 0$, and concave on that domain ; therefore, we can try to model the convex set defined by $0 \leq y \leq W(x)$ (intersection of its hypograph with nonnegative orthant)

$$0 \leq y \leq W(x) \Leftrightarrow 0 \leq y \exp y \leq W(x) \exp W(x) \Leftrightarrow 0 \leq y \exp y \leq x$$

which can be obtained using

- a exponential constraint $\exp\left(\frac{z}{y}\right) \leq \frac{x}{y}$ and
- a quadratic constraint $z \geq y^2$

Indeed, we can check that

$$0 \leq y \exp y = y \exp(y^2/y) \leq y \exp(z/y) \leq x$$

In summary, combining a quadratic and an exponential constraint, we have shown that

$$0 \leq y \leq W(x) \Leftrightarrow (x, y, z) \in \mathcal{E}_p \text{ and } (z, 1, x) \in \mathcal{K}_2$$

An example: the Lambert W function

$W(x)$ is real for $x \geq 0$, and concave on that domain ; therefore, we can try to model the convex set defined by $0 \leq y \leq W(x)$ (intersection of its hypograph with nonnegative orthant)

$$0 \leq y \leq W(x) \Leftrightarrow 0 \leq y \exp y \leq W(x) \exp W(x) \Leftrightarrow 0 \leq y \exp y \leq x$$

which can be obtained using

- a exponential constraint $\exp\left(\frac{z}{y}\right) \leq \frac{x}{y}$ and
- a quadratic constraint $z \geq y^2$

Indeed, we can check that

$$0 \leq y \exp y = y \exp(y^2/y) \leq y \exp(z/y) \leq x$$

In summary, combining a quadratic and an exponential constraint, we have shown that

$$0 \leq y \leq W(x) \Leftrightarrow (x, y, z) \in \mathcal{E}_p \text{ and } (z, 1, x) \in \mathcal{K}_2$$

An example: the Lambert W function

$W(x)$ is real for $x \geq 0$, and concave on that domain ; therefore, we can try to model the convex set defined by $0 \leq y \leq W(x)$ (intersection of its hypograph with nonnegative orthant)

$$0 \leq y \leq W(x) \Leftrightarrow 0 \leq y \exp y \leq W(x) \exp W(x) \Leftrightarrow 0 \leq y \exp y \leq x$$

which can be obtained using

- a exponential constraint $\exp\left(\frac{z}{y}\right) \leq \frac{x}{y}$ and
- a quadratic constraint $z \geq y^2$

Indeed, we can check that

$$0 \leq y \exp y = y \exp(y^2/y) \leq y \exp(z/y) \leq x$$

In summary, combining a quadratic and an exponential constraint, we have shown that

$$0 \leq y \leq W(x) \Leftrightarrow (x, y, z) \in \mathcal{E}_p \text{ and } (z, 1, x) \in \mathcal{K}_2$$

An example: the Lambert W function

$W(x)$ is real for $x \geq 0$, and concave on that domain ; therefore, we can try to model the convex set defined by $0 \leq y \leq W(x)$ (intersection of its hypograph with nonnegative orthant)

$$0 \leq y \leq W(x) \Leftrightarrow 0 \leq y \exp y \leq W(x) \exp W(x) \Leftrightarrow 0 \leq y \exp y \leq x$$

which can be obtained using

- a exponential constraint $\exp\left(\frac{z}{y}\right) \leq \frac{x}{y}$ and
- a quadratic constraint $z \geq y^2$

Indeed, we can check that

$$0 \leq y \exp y = y \exp(y^2/y) \leq y \exp(z/y) \leq x$$

In summary, combining a quadratic and an exponential constraint, we have shown that

$$0 \leq y \leq W(x) \Leftrightarrow (x, y, z) \in \mathcal{E}_p \text{ and } (z, 1, x) \in \mathcal{K}_2$$

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Dual for cones from functions

Computing an explicit dual for all the problems mentioned can be done in a single step: simply find the dual cone of $\mathcal{K}_\alpha$

- Recall the following (adapted) definition of a conjugate function:

$$f^*(y) = \sup_x \bigl(-y^{\mathrm{T}} x - f(x)\bigr)$$

- The following results characterizes the dual of $K(f)$

$$\begin{aligned} K(f)^* &= \{(x,y,z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\tfrac{x}{z}) \le \tfrac{y}{z}\}^* \\ &= \mathrm{cl}\{(x,y,z) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R} \mid f^*(\tfrac{x}{y}) \le \tfrac{z}{y}\} \end{aligned}$$

(e.g. Rockafellar)

- This dual $(K(f))^*$ is nearly equal to $K(f^*)$
(only difference is the permutation between y and z)

Dual for cones from functions

Computing an explicit dual for all the problems mentioned can be done in a single step: simply find the dual cone of $\mathcal{K}_\alpha$

- Recall the following (adapted) definition of a conjugate function:

$$f^*(y) = \sup_x \left(-y^{\mathrm{T}} x - f(x) \right)$$

- The following results characterizes the dual of $K(f)$

$$\begin{aligned} K(f)^* &= \{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\tfrac{x}{z}) \le \tfrac{y}{z}\}^* \\ &= \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R} \mid f^*(\tfrac{x}{y}) \le \tfrac{z}{y}\} \end{aligned}$$

(e.g. Rockafellar)

- This dual $(K(f))^*$ is nearly equal to $K(f^*)$
  (only difference is the permutation between y and z)

Dual for cones from functions

Computing an explicit dual for all the problems mentioned can be done in a single step: simply find the dual cone of $\mathcal{K}_\alpha$

- Recall the following (adapted) definition of a conjugate function:

$$f^*(y) = \sup_x \big(-y^{\mathrm{T}} x - f(x)\big)$$

- The following results characterizes the dual of $K(f)$

$$\begin{aligned} K(f)^* &= \{(x, y, z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\tfrac{x}{z}) \leq \tfrac{y}{z}\}^* \\ &= \mathrm{cl}\{(x, y, z) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R} \mid f^*(\tfrac{x}{y}) \leq \tfrac{z}{y}\} \end{aligned}$$

(e.g. Rockafellar)

- This dual $(K(f))^*$ is nearly equal to $K(f^*)$ (only difference is the permutation between y and z)

Dual for cones from functions

Computing an explicit dual for all the problems mentioned can be done in a single step: simply find the dual cone of $\mathcal{K}_\alpha$

- Recall the following (adapted) definition of a conjugate function:

$$f^*(y) = \sup_x \big(-y^{\mathrm{T}} x - f(x)\big)$$

- The following results characterizes the dual of $K(f)$

$$\begin{aligned} K(f)^* \quad &= \{(x,y,z) \in C \times \mathbb{R} \times \mathbb{R}_{++} \mid f(\tfrac{x}{z}) \le \tfrac{y}{z}\}^* \\ &= \mathrm{cl}\{(x,y,z) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R} \mid f^*(\tfrac{x}{y}) \le \tfrac{z}{y}\} \end{aligned}$$

(e.g. Rockafellar)

- This dual $(K(f))^*$ is nearly equal to $K(f^*)$
(only difference is the permutation between y and z)

Dual for the power cone

- Recall that the conjugate of $x \mapsto \alpha |x|^{\frac{1}{\alpha}}$ is $x \mapsto \beta |x|^{\frac{1}{\beta}}$ with $\alpha + \beta = 1$
- Therefore the dual for this scaled version of the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \alpha \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

is

$$(\mathcal{K}_\alpha)^* = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \beta \left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y}\}$$

- However we have

$$\left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y} \Leftrightarrow \left|\frac{x}{y}\right| \leq (\frac{z}{y})^\beta \Leftrightarrow |x| \leq z^\beta y^{1-\beta} = y^\alpha z^\beta$$

so that, up to some constant factor, the dual of $\mathcal{K}_\alpha$ is equal to itself (with the same value for parameter α)

Dual for the power cone

- Recall that the conjugate of $x \mapsto \alpha |x|^{\frac{1}{\alpha}}$ is $x \mapsto \beta |x|^{\frac{1}{\beta}}$ with $\alpha + \beta = 1$
- Therefore the dual for this scaled version of the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \alpha \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

is

$$(\mathcal{K}_\alpha)^* = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \beta \left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y}\}$$

- However we have

$$\left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y} \Leftrightarrow \left|\frac{x}{y}\right| \leq (\frac{z}{y})^\beta \Leftrightarrow |x| \leq z^\beta y^{1-\beta} = y^\alpha z^\beta$$

so that, up to some constant factor, the dual of $\mathcal{K}_\alpha$ is equal to itself (with the same value for parameter α)

Dual for the power cone

- Recall that the conjugate of $x \mapsto \alpha |x|^{\frac{1}{\alpha}}$ is $x \mapsto \beta |x|^{\frac{1}{\beta}}$ with $\alpha + \beta = 1$
- Therefore the dual for this scaled version of the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \alpha \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

is

$$(\mathcal{K}_\alpha)^* = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \beta \left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y}\}$$

- However we have

$$\left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y} \Leftrightarrow \left|\frac{x}{y}\right| \leq (\frac{z}{y})^\beta \Leftrightarrow |x| \leq z^\beta y^{1-\beta} = y^\alpha z^\beta$$

so that, up to some constant factor, the dual of $\mathcal{K}_\alpha$ is equal to itself (with the same value for parameter α)

Dual for the power cone

- Recall that the conjugate of $x \mapsto \alpha |x|^{\frac{1}{\alpha}}$ is $x \mapsto \beta |x|^{\frac{1}{\beta}}$ with $\alpha + \beta = 1$
- Therefore the dual for this scaled version of the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \alpha \left|\frac{x}{z}\right|^{\frac{1}{\alpha}} \leq \frac{y}{z}\}$$

is

$$(\mathcal{K}_\alpha)^* = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid \beta \left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y}\}$$

- However we have

$$\left|\frac{x}{y}\right|^{\frac{1}{\beta}} \leq \frac{z}{y} \Leftrightarrow \left|\frac{x}{y}\right| \leq (\frac{z}{y})^\beta \Leftrightarrow |x| \leq z^\beta y^{1-\beta} = y^\alpha z^\beta$$

so that, up to some constant factor, the dual of $\mathcal{K}_\alpha$ is equal to itself (with the same value for parameter α)

Dual for the power cone

- Actually, to make $\mathcal{K}_\alpha$ exactly self-dual, one should adapt its definition as follows:

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

(or, alternatively, keep the original definition and use a different inner product for the definition of the dual cone

$$< (x, y, z), (x^*, y^*, z^*) >= xx^* + \alpha yy^* + \beta zz^* \ )$$

- Self-duality clearly also holds for Cartesian products of power cones

$$(\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N}) = (\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N})^*$$

- This implies that the class of conic optimization problems based on the power cone (that includes all the structured convex problems mentioned earlier) is closed under taking the dual (which was not the case for most of the individual classes themselves)
- $\mathcal{K}_\alpha$ is self-dual but is not homogeneous (otherwise $\rightarrow$ self-scaled !)

Dual for the power cone

- Actually, to make $\mathcal{K}_\alpha$ exactly self-dual, one should adapt its definition as follows:

$$\mathcal{K}_\alpha = \{(x,y,z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

  (or, alternatively, keep the original definition and use a different inner product for the definition of the dual cone

$$< (x,y,z),(x^*,y^*,z^*) >= xx^* + \alpha yy^* + \beta zz^* \;)$$

- Self-duality clearly also holds for Cartesian products of power cones

$$(\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N}) = (\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N})^*$$

- This implies that the class of conic optimization problems based on the power cone (that includes all the structured convex problems mentioned earlier) is closed under taking the dual (which was not the case for most of the individual classes themselves)
- $\mathcal{K}_\alpha$ is self-dual but is not homogeneous (otherwise $\longrightarrow$ self-scaled !)

Dual for the power cone

- Actually, to make $\mathcal{K}_\alpha$ exactly self-dual, one should adapt its definition as follows:

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

(or, alternatively, keep the original definition and use a different inner product for the definition of the dual cone

$$< (x, y, z), (x^*, y^*, z^*) >= xx^* + \alpha yy^* + \beta zz^* \;)$$

- Self-duality clearly also holds for Cartesian products of power cones

$$(\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N}) = (\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N})^*$$

- This implies that the class of conic optimization problems based on the power cone (that includes all the structured convex problems mentioned earlier) is closed under taking the dual (which was not the case for most of the individual classes themselves)
- $\mathcal{K}_\alpha$ is self-dual but is not homogeneous (otherwise $\longrightarrow$ self-scaled !)

Dual for the power cone

- Actually, to make $\mathcal{K}_\alpha$ exactly self-dual, one should adapt its definition as follows:

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

(or, alternatively, keep the original definition and use a different inner product for the definition of the dual cone

$$< (x, y, z), (x^*, y^*, z^*) >= xx^* + \alpha yy^* + \beta zz^* \;)$$

- Self-duality clearly also holds for Cartesian products of power cones

$$(\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N}) = (\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N})^*$$

- This implies that the class of conic optimization problems based on the power cone (that includes all the structured convex problems mentioned earlier) is closed under taking the dual (which was not the case for most of the individual classes themselves)
- $\mathcal{K}_\alpha$ is self-dual but is not homogeneous (otherwise $\longrightarrow$ self-scaled !)

Dual for the power cone

- Actually, to make $\mathcal{K}_\alpha$ exactly self-dual, one should adapt its definition as follows:

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq (\frac{y}{\sqrt{\alpha}})^\alpha (\frac{z}{\sqrt{\beta}})^\beta\}$$

(or, alternatively, keep the original definition and use a different inner product for the definition of the dual cone

$$< (x, y, z), (x^*, y^*, z^*) >= xx^* + \alpha yy^* + \beta zz^* \ )$$

- Self-duality clearly also holds for Cartesian products of power cones

$$(\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N}) = (\mathcal{K}_{\alpha_1} \times \mathcal{K}_{\alpha_2} \times \cdots \times \mathcal{K}_{\alpha_N})^*$$

- This implies that the class of conic optimization problems based on the power cone (that includes all the structured convex problems mentioned earlier) is closed under taking the dual (which was not the case for most of the individual classes themselves)
- $\mathcal{K}_\alpha$ is self-dual but is not homogeneous (otherwise $\rightarrow$ self-scaled !)

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization

2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone

3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- Self-concordance of F guarantees that Newton's method can track the central path accurately
- The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}}x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}}x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}}\lambda - \mu \nabla F(x) = c$$

- An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- Self-concordance of F guarantees that Newton's method can track the central path accurately
- The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu}\log\frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- Self-concordance of F guarantees that Newton's method can track the central path accurately
- The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- Self-concordance of F guarantees that Newton's method can track the central path accurately
- The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- ◇ An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- ◇ Self-concordance of F guarantees that Newton's method can track the central path accurately
- ◇ The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- Self-concordance of F guarantees that Newton's method can track the central path accurately
- The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \frac{1}{\epsilon}\right) \text{ Newton steps}$$

How does a primal algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

Given a self-concordant barrier $F(x)$ for K, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\min c^{\mathrm{T}} x + \mu F(x) \text{ such that } Ax = b$$

whose (unique) optimal solution $x(\mu)$ obeys the following first-order conditions

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

- ◇ An interior-point scheme traces the set $\{x(\mu)\}$ (called the central path) as parameter μ tends to 0
- ◇ Self-concordance of F guarantees that Newton's method can track the central path accurately
- ◇ The best methods obtain a solution with ϵ accuracy after

$$\mathcal{O}\left(\sqrt{\nu} \log \tfrac{1}{\epsilon}\right) \text{ Newton steps}$$

Barrier function for $\mathcal{K}_\alpha$

- NESTEROV proposer a self-concordant barrier for the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta = \hat{y}^\alpha \hat{z}^\beta\}$$

with a parameter $\nu = 4$ (posing for convenience $\hat{y} = \frac{y}{\sqrt{\alpha}}$ and $\hat{z} = \frac{z}{\sqrt{\beta}}$)

$$\bar{F}_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \log \hat{y} - \log \hat{z}$$

- This can be improved: the following barrier

$$F_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \beta \log \hat{y} - \alpha \log \hat{z}$$

is self-concordant with a lower parameter $\nu = 3$ (G.-CHARES, 2008)

- For a Cartesian product $\mathcal{K}_{\alpha_1} \times \cdots \times \mathcal{K}_{\alpha_N}$ we sum each component: $F_{\alpha_1}(x_1, y_1, z_1) + \cdots + F_{\alpha_1}(x_N, y_N, z_N)$ has parameter $\nu = 3N$
- Therefore the iteration complexity to solve conic problems involving $\mathcal{K}_\alpha$ depends only on the number of cones N (not on parameter α)

$$\mathcal{O}\left(\sqrt{N} \log \frac{1}{\epsilon}\right) \text{ to obtain a solution with } \epsilon \text{ accuracy}$$

Barrier function for $\mathcal{K}_\alpha$

- NESTEROV proposer a self-concordant barrier for the power cone

$$\mathcal{K}_\alpha = \{(x,y,z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta = \hat{y}^\alpha \hat{z}^\beta\}$$

with a parameter $\nu = 4$ (posing for convenience $\hat{y} = \frac{y}{\sqrt{\alpha}}$ and $\hat{z} = \frac{z}{\sqrt{\beta}}$)

$$\bar{F}_\alpha(x,y,z) = -\log\left(\hat{y}^{2\alpha}\hat{z}^{2\beta} - x^2\right) - \log \hat{y} - \log \hat{z}$$

- This can be improved: the following barrier

$$F_\alpha(x,y,z) = -\log\left(\hat{y}^{2\alpha}\hat{z}^{2\beta} - x^2\right) - \beta \log \hat{y} - \alpha \log \hat{z}$$

is self-concordant with a lower parameter $\nu = 3$ (G.-CHARES, 2008)

- For a Cartesian product $\mathcal{K}_{\alpha_1} \times \cdots \times \mathcal{K}_{\alpha_N}$ we sum each component: $F_{\alpha_1}(x_1,y_1,z_1) + \cdots + F_{\alpha_1}(x_N,y_N,z_N)$ has parameter $\nu = 3N$
- Therefore the iteration complexity to solve conic problems involving $\mathcal{K}_\alpha$ depends only on the number of cones N (not on parameter α)

$$\mathcal{O}\left(\sqrt{N}\log\frac{1}{\epsilon}\right) \text{ to obtain a solution with } \epsilon \text{ accuracy}$$

Barrier function for $\mathcal{K}_\alpha$

- NESTEROV proposer a self-concordant barrier for the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq (\frac{y}{\sqrt{\alpha}})^\alpha (\frac{z}{\sqrt{\beta}})^\beta = \hat{y}^\alpha \hat{z}^\beta\}$$

with a parameter $\nu = 4$ (posing for convenience $\hat{y} = \frac{y}{\sqrt{\alpha}}$ and $\hat{z} = \frac{z}{\sqrt{\beta}}$)

$$\bar{F}_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \log \hat{y} - \log \hat{z}$$

- This can be improved: the following barrier

$$F_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \beta \log \hat{y} - \alpha \log \hat{z}$$

is self-concordant with a lower parameter $\nu = 3$ (G.-CHARES, 2008)

- For a Cartesian product $\mathcal{K}_{\alpha_1} \times \cdots \times \mathcal{K}_{\alpha_N}$ we sum each component: $F_{\alpha_1}(x_1, y_1, z_1) + \cdots + F_{\alpha_1}(x_N, y_N, z_N)$ has parameter $\nu = 3N$
- Therefore the iteration complexity to solve conic problems involving $\mathcal{K}_\alpha$ depends only on the number of cones N (not on parameter α)

$$\mathcal{O}\left(\sqrt{N} \log \frac{1}{\epsilon}\right) \text{ to obtain a solution with } \epsilon \text{ accuracy}$$

Barrier function for $\mathcal{K}_\alpha$

- Nesterov proposer a self-concordant barrier for the power cone

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta = \hat{y}^\alpha \hat{z}^\beta\}$$

with a parameter $\nu = 4$ (posing for convenience $\hat{y} = \frac{y}{\sqrt{\alpha}}$ and $\hat{z} = \frac{z}{\sqrt{\beta}}$)

$$\bar{F}_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \log \hat{y} - \log \hat{z}$$

- This can be improved: the following barrier

$$F_\alpha(x, y, z) = -\log\left(\hat{y}^{2\alpha} \hat{z}^{2\beta} - x^2\right) - \beta \log \hat{y} - \alpha \log \hat{z}$$

is self-concordant with a lower parameter $\nu = 3$ (G.-Chares, 2008)
- For a Cartesian product $\mathcal{K}_{\alpha_1} \times \cdots \times \mathcal{K}_{\alpha_N}$ we sum each component: $F_{\alpha_1}(x_1, y_1, z_1) + \cdots + F_{\alpha_1}(x_N, y_N, z_N)$ has parameter $\nu = 3N$
- Therefore the iteration complexity to solve conic problems involving $\mathcal{K}_\alpha$ depends only on the number of cones N (not on parameter α)

$$\mathcal{O}\left(\sqrt{N} \log \frac{1}{\epsilon}\right) \text{ to obtain a solution with } \epsilon \text{ accuracy}$$

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a dual algorithm work ?

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given a self-concordant barrier $G(s)$ for K^*, consider the family of unconstrained problem parameterized by $\mu > 0$

$$\max b^{\mathrm{T}} y - \mu G(s) \text{ such that } A^{\mathrm{T}} y + s = c$$

whose (unique) optimal solution $(y(\mu), s(\mu))$ obeys the following first-order conditions

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

- ◇ Dual algorithm works exactly like primal algorithm (only the way to write down the linear equations differs)
- ◇ For (products of) power cones, we have $K^* = K$, so that we can also take the same barrier as for the primal $G(s) = F_\alpha(s)$
- ◇ However, primal and dual iterate are completely independent: how can one compute them simultaneously and symmetrically ?

How does a primal-dual algorithm work ?

$$\min c^T x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

$$\max b^T y \text{ such that } A^T y + s = c \text{ and } s \in K^*$$

Given self-concordant barriers $F(x)$ for K and $G(s)$ for K^*, primal and dual central paths obey

$$Ax = b \text{ and } A^T \lambda - \mu \nabla F(x) = c$$

$$A^T y + s = c \text{ and } -\mu A \nabla G(s) = b$$

Those two sets of equations will coincide if we have

$$x = -\mu \nabla G(s), \quad \lambda = y \quad \text{and} \quad s = -\mu \nabla F(x)$$

This holds if the gradients $-\nabla F$ and $-\nabla G$ are inverse to each other (this uses the fact that they are homogeneous of degree -1) or equivalently if barriers F and G are conjugate to each other

How does a primal-dual algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given self-concordant barriers $F(x)$ for K and $G(s)$ for K^*, primal and dual central paths obey

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

Those two sets of equations will coincide if we have

$$x = -\mu \nabla G(s), \quad \lambda = y \quad \text{and} \quad s = -\mu \nabla F(x)$$

This holds if the gradients $-\nabla F$ and $-\nabla G$ are inverse to each other (this uses the fact that they are homogeneous of degree -1) or equivalently if barriers F and G are conjugate to each other

How does a primal-dual algorithm work ?

$$\min c^{\mathrm{T}}x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

$$\max b^{\mathrm{T}}y \text{ such that } A^{\mathrm{T}}y + s = c \text{ and } s \in K^*$$

Given self-concordant barriers $F(x)$ for K and $G(s)$ for K^*, primal and dual central paths obey

$$Ax = b \text{ and } A^{\mathrm{T}}\lambda - \mu\nabla F(x) = c$$

$$A^{\mathrm{T}}y + s = c \text{ and } -\mu A\nabla G(s) = b$$

Those two sets of equations will coincide if we have

$$x = -\mu\nabla G(s), \quad \lambda = y \quad \text{and} \quad s = -\mu\nabla F(x)$$

This holds if the gradients $-\nabla F$ and $-\nabla G$ are inverse to each other (this uses the fact that they are homogeneous of degree -1) or equivalently if barriers F and G are conjugate to each other

How does a primal-dual algorithm work ?

$$\min c^{\mathrm{T}} x \text{ such that } Ax = b \text{ and } x \succeq_K 0$$

$$\max b^{\mathrm{T}} y \text{ such that } A^{\mathrm{T}} y + s = c \text{ and } s \in K^*$$

Given self-concordant barriers $F(x)$ for K and $G(s)$ for K^*, primal and dual central paths obey

$$Ax = b \text{ and } A^{\mathrm{T}} \lambda - \mu \nabla F(x) = c$$

$$A^{\mathrm{T}} y + s = c \text{ and } -\mu A \nabla G(s) = b$$

Those two sets of equations will coincide if we have

$$x = -\mu \nabla G(s), \quad \lambda = y \quad \text{and} \quad s = -\mu \nabla F(x)$$

This holds if the gradients $-\nabla F$ and $-\nabla G$ are inverse to each other (this uses the fact that they are homogeneous of degree -1) or equivalently if barriers F and G are conjugate to each other

Primal-dual algorithm for the power cone

In the case of self-scaled cones, one has $K = K^*$, the same barrier for the primal and for the dual, and that barrier is self-conjugate

$$F(x) = G(x) \quad \forall x \in K \text{ and } F^*(x) = F(x) + \text{constant } \forall x \in K$$

For example, in the linear case, $F(x) = -\log x$ is a barrier for $\mathbb{R}_+$ and

$$F^*(x) = -1 - \log x = F(x) + \text{constant} \qquad F'(x) = -1/x \Rightarrow F'(F'(x)) = x$$

Unfortunately, this is no longer true for the power cone:
it is easy to check numerically that both Nesterov's barrier and the improved barrier are not self-conjugate

(exceptions: $\mathcal{K}_{\frac{1}{2}}$ is a second-order cone, for which $-\log(2yz - x^2)$ is a self-conjugate barrier, while $\mathcal{K}_0$ and $\mathcal{K}_1$ are polyhedral and also admit a self-conjugate barrier)

Overview

1. **Introduction: convex and conic optimization**
   - ◇ Why convex optimization?
   - ◇ Conic optimization: a standard form for convex optimization
2. **Conic optimization based on the power cone**
   - ◇ Modelling with the power cone
   - ◇ Finding duals with the power cone
3. **A symmetric primal-dual algorithm**
   - ◇ Primal, dual and primal-dual interior-point methods
   - ◇ A symmetric algorithm

Decomposition of barrier $F\alpha$

Although not self-conjugate, improved barrier F_α can be decomposed as

$$\begin{aligned} F_\alpha(x,y,z) &= -\log\left(\hat{y}^{2\alpha}\hat{z}^{2\beta} - x^2\right) - \beta\log\hat{y} - \alpha\log\hat{z} \\ &= -\log\left(x^{-1}\hat{y}^{\alpha}\hat{z}^{\beta} - x\hat{y}^{-\alpha}\hat{z}^{-\beta}\right) - \log x - \log\hat{y} - \log\hat{z} \\ &= H_\alpha(x,y,z) + L(x,y,z) \end{aligned}$$

- $H_\alpha(x,y,z) = -\log\left(x^{-1}\hat{y}^{\alpha}\hat{z}^{\beta} - x\hat{y}^{-\alpha}\hat{z}^{-\beta}\right)$ is logarithmically 0-homogeneous
- $L(x,y,z) = -\log x - \log\hat{y} - \log\hat{z}$ is logarithmically 3-homogeneous ; recall that it is self-conjugate: $L^*(x,y,z) = L^*(x,y,z)$
- Moreover, simple computations show that

$$H^*_\alpha(x,y,z) = -H_\alpha(x,y,z) \text{ for all } 0 \leq \alpha \leq 1$$

(H_α is self-conjugate with an additional change of sign)

A symmetric reformulation

Consider the primal barrier subproblem

$$\min c^{\mathrm{T}} x + \mu F_\alpha(x) \text{ such that } Ax = b$$

and instead of using directly $F(x) = H_\alpha(x) + L(x)$

$$(\min c^{\mathrm{T}} x + \mu H_\alpha(x) + \mu L(x) \text{ such that } Ax = b)$$

we reformulate it as follows

$$\min c^{\mathrm{T}} x + \mu H_\alpha(x) + \mu L(x') \text{ such that } Ax = b \text{ and } x = x'$$

Note that on the feasible region, the objective function is self-concordant $\longrightarrow$ polynomial-time complexity achievable

Dual for the symmetric reformulation

$$\min c^{\mathrm{T}} x + \mu H_\alpha(x) + \mu L(x') \text{ such that } Ax = b \text{ and } x = x'$$

admits the following dual problem

$$\max b^{\mathrm{T}} y - \mu H_\alpha(-s) - \mu L(s') \text{ such that } A^{\mathrm{T}} y + y' + s = c \text{ and } y' = s'$$

where we used the fact that

$$\big(H_\alpha(x) + L(x')\big)^* = H_\alpha^*(x) + L^*(x') = H_\alpha(-x) + L(x')$$

(valid because each term involves different variables)

On the following slide, in the interest of clarity, we write all expressions in the case $\mu = 1$ (dependence on μ is always linear and easy to handle)

Primal-dual central path for the symmetric reformulation

$$\min c^{\mathrm{T}} x + H_\alpha(x) + \mu L(x') \text{ such that } Ax = b \text{ and } x = x'$$

$$\max b^{\mathrm{T}} y - H_\alpha(-s) - \mu L(s') \text{ such that } A^{\mathrm{T}} y + y' + s = c \text{ and } y' = s'$$

admit the following optimality conditions

$$Ax = b, \; x = x', \; A^{\mathrm{T}} \lambda + \lambda' - \nabla H_\alpha(x) = c, \; \lambda' = -\nabla L(x')$$

$$A^{\mathrm{T}} y + y' + s = c, \; y' = s', \; A \nabla H_\alpha(-s) = b \text{ and } \nabla H_\alpha(-s) = -\nabla L(s')$$

which coincide if we let $\lambda = y$, $\lambda' = y'$ and

$$s = -\nabla H_\alpha(x) \Leftrightarrow x = +\nabla H_\alpha(-s) \quad \text{and} \quad s' = -\nabla L(x') \Leftrightarrow x' = -\nabla L(s')$$

We have a self-dual system of optimality conditions

Symmetric form of the optimality conditions

Our optimality conditions are now self-dual, but they are not algebraically symmetric: we have to write either $s = -\mu \nabla F(x)$ or $x = -\mu \nabla G(x)$

For the "linear" logarithmic barrier $L(x, y, z) = -\log x - \log \hat{y} - \log \hat{z}$

$$s' = -\mu \nabla L(x') \Leftrightarrow x' = -\mu \nabla L(s')$$

an equivalent symmetric equation is well-known:

$$s' = -\mu(-x'^{-1}) \quad \Leftrightarrow \quad x'_i s'_i = \mu$$

(a similar equation can be written for all self-scaled cones)

Is it also possible for the "nonlinear" logarithmic component $H_\alpha(x, y, z) = -\log\big(x^{-1} \hat{y}^\alpha \hat{z}^\beta - x \hat{y}^{-\alpha} \hat{z}^{-\beta}\big)$?

$$s = -\nabla H_\alpha(x) \Leftrightarrow x = +\nabla H_\alpha(-s)$$

Symmetric form for the H_α condition

$$s = -\nabla H_\alpha(x) \Leftrightarrow x = +\nabla H_\alpha(-s)$$

Writing $x = (x, y, z)$ and $s = (x^*, y^*, z^*)$
and using again the notation $\hat{y} = \frac{y}{\sqrt{\alpha}}$, $\hat{y'} = \frac{y'}{\sqrt{\alpha}}$, etc.
these conditions can be rewritten as

$$\begin{array}{rcl} -x_i x_i^* & = & K \\ \hat{y}_i \hat{y}_i^* & = & K \\ \hat{z}_i \hat{z}_i^* & = & K \end{array}$$

with

$$K = \frac{x^{-1} \hat{y}^\alpha \hat{z}^\beta - x^{*-1} \hat{y}^{*\alpha} \hat{z}^{*\beta}}{x^{-1} \hat{y}^\alpha \hat{z}^\beta + x^{*-1} \hat{y}^{*\alpha} \hat{z}^{*\beta}}$$

which is completely symmetric under taking the dual
$\rightarrow$ the optimality conditions can also be written in an algebraically symmetric way

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \leq t$ constraint)

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \big(\frac{y}{\sqrt{\alpha}}\big)^\alpha \big(\frac{z}{\sqrt{\beta}}\big)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \leq t$ constraint)

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \leq t$ constraint)

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \leq t$ constraint)

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \leq \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \leq t$ constraint)

Concluding remarks

Modelling with the power cone

In conclusion, the family of self-dual 3-dimensional cones $\mathcal{K}_\alpha$

$$\mathcal{K}_\alpha = \{(x, y, z) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \mid |x| \le \left(\frac{y}{\sqrt{\alpha}}\right)^\alpha \left(\frac{z}{\sqrt{\beta}}\right)^\beta\}$$

- can model a very large class of structured convex problems (with the notable exception of semidefinite optimization)
- enables their resolution with powerful interior-point methods
- allows the easy computation of their dual problems

Convex problems covered include linear, quadratic, second-order cone, quadratically constrained, geometric, l_p-norm, sum-of-norm, entropy optimization and others, as well as any combinations of these

Potential drawback: conic modelling sometimes require the introduction of some additional variables (e.g. $\|x\|_p \le t$ constraint)

Concluding remarks

Algorithms for power cone optimization

- Standard primal or dual interior-point algorithms can be applied to power cone optimization problems, using the following self-concordant barrier with parameter 3

$$F_\alpha(x, y, z) = -\log\Big(\big(\frac{y}{\sqrt{\alpha}}\big)^{2\alpha}\big(\frac{z}{\sqrt{\beta}}\big)^{2\beta} - x^2\Big) - \beta \log y - \alpha \log z$$

- A completely symmetric primal-dual formulation has been proposed which for which the optimality conditions can also be written in an algebraically symmetric way

  $\rightarrow$ design a completely primal-dual symmetric interior-point method

  (still need a rigorous proof of polynomial-time complexity)

Concluding remarks

Algorithms for power cone optimization

◇ Standard primal or dual interior-point algorithms can be applied to power cone optimization problems, using the following self-concordant barrier with parameter 3

$$F_\alpha(x,y,z) = -\log\Big(\big(\frac{y}{\sqrt{\alpha}}\big)^{2\alpha}\big(\frac{z}{\sqrt{\beta}}\big)^{2\beta} - x^2\Big) - \beta\log y - \alpha\log z$$

◇ A completely symmetric primal-dual formulation has been proposed which for which the optimality conditions can also be written in an algebraically symmetric way

$\rightarrow$ design a completely primal-dual symmetric interior-point method

(still need a rigorous proof of polynomial-time complexity)

Concluding remarks

Algorithms for power cone optimization

- Standard primal or dual interior-point algorithms can be applied to power cone optimization problems, using the following self-concordant barrier with parameter 3

$$F_\alpha(x,y,z) = -\log\Big(\big(\frac{y}{\sqrt{\alpha}}\big)^{2\alpha} \big(\frac{z}{\sqrt{\beta}}\big)^{2\beta} - x^2 \Big) - \beta \log y - \alpha \log z$$

- A completely symmetric primal-dual formulation has been proposed which for which the optimality conditions can also be written in an algebraically symmetric way

  $\rightarrow$ design a completely primal-dual symmetric interior-point method

  (still need a rigorous proof of polynomial-time complexity)