Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 12

Second-Order Cone Models

Each problem that I solved became a rule which served afterwards to solve other problems.

René Descartes

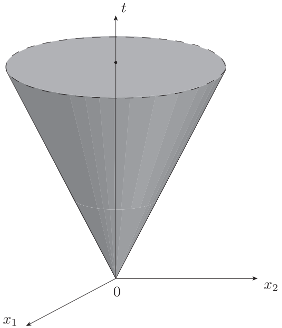
Outline

Introduction

- Second-order cone programming (SOCP) is a generalization of linear and quadratic programming that allows for affine combinations of variables to be constrained inside a special convex set, called a *second-order cone*.
- The SOCP model includes as special cases LPs, as well as problems with convex quadratic objective and constraints.
- SOCP models are particularly useful in geometry problems, approximation problems, as well as in probabilistic (chance-constrained) approaches to linear optimization problems in which the data is affected by random uncertainty.

The second-order cone

- The second-order cone (SOC) in $\mathbb{R}^3$ is the set of vectors (x_1, x_2, t) such that $\sqrt{x_1^2 + x_2^2} \leq t$. Horizontal sections of this set at level $\alpha \geq 0$ are disks of radius α.



- In arbitrary dimension: an $(n+1)$-dimensional SOC is the following set:

$$\mathcal{K}_n = \{(x, t),\, x \in \mathbb{R}^n,\, t \in \mathbb{R} : \ \|x\|_2 \leq t\}. \tag{1}$$

The rotated second-order cone

- The rotated second-order cone in $\mathbb{R}^{n+2}$ is the set

$$\mathcal{K}_n^r = \left\{(x, y, z), x \in \mathbb{R}^n, y \in \mathbb{R}, z \in \mathbb{R} : x^\top x \le 2yz,\ y \ge 0,\ z \ge 0\right\}.$$

- The rotated second-order cone in $\mathbb{R}^{n+2}$ can be expressed as a linear transformation (actually, a rotation) of the (plain) second-order cone in $\mathbb{R}^{n+2}$, since

$$\|x\|_2^2 \le 2yz,\ y \ge 0,\ z \ge 0 \iff \left\|\begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y-z) \end{bmatrix}\right\|_2 \le \frac{1}{\sqrt{2}}(y+z). \qquad (2)$$

That is, $(x, y, z) \in \mathcal{K}_n^r$ if and only if $(w, t) \in \mathcal{K}_n$, where

$$w = (x, (y-z)/\sqrt{2}), \quad t = (y+z)/\sqrt{2}.$$

- Constraints of the form $\|x\|_2^2 \le 2yz$, as appearing in (2), are usually referred to as *hyperbolic* constraints.

Standard SOC constraint

- The standard format of a second-order cone constraint on a variable $x \in \mathbb{R}^n$ expresses the condition that $(y, t) \in \mathcal{K}_m$, with $y \in \mathbb{R}^m$, $t \in \mathbb{R}$, where y, t are some affine functions of x.
- These affine functions can be expressed as $y = Ax + b$, $t = c^\top x + d$, hence the condition $(y, t) \in \mathcal{K}_m$ becomes

$$\|Ax + b\|_2 \leq c^\top x + d, \tag{3}$$

  where $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $d \in \mathbb{R}$.
- For example, the quadratic constraint

$$x^\top Q x + c^\top x \leq t, \quad Q \succeq 0$$

  can be expressed in conic form as

$$\left\| \begin{bmatrix} \sqrt{2} Q^{1/2} x \\ t - c^\top x - 1/2 \end{bmatrix} \right\|_2 \leq t - c^\top x + 1/2.$$

Second-order cone programs

- A second-order cone program is a convex optimization problem having linear objective and SOC constraints. When the SOC constraints have the standard form (3), we have a SOCP in *standard inequality form*:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{s.t.:} \quad & \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \quad i = 1, \ldots, m, \end{aligned} \tag{4}$$

  where $A_i \in \mathbb{R}^{m_i,n}$ are given matrices, $b_i \in \mathbb{R}^{m_i}$, $c_i \in \mathbb{R}^n$ are vectors, and d_i are given scalars.
- SOCPs are representative of a quite large class of convex optimization problems. Indeed, LPs, convex QPs, and convex QCQPs can all be represented as SOCPs.

Linear programs as SOCPs

The linear program (LP) in standard inequality form

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.:} \quad & a_i^\top x \leq b_i, \quad i = 1, \ldots, m, \end{aligned}$$

can be readily cast in SOCP form as

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.:} \quad & \|C_i x + d_i\|_2 \leq b_i - a_i^\top x, \quad i = 1, \ldots, m, \end{aligned}$$

where $C_i = 0$, $d_i = 0$, $i = 1, \ldots, m$.

Quadratic programs as SOCPs

The quadratic program (QP)

$$\begin{aligned} \min_{x} \quad & x^\top Q x + c^\top x \\ \text{s.t.:} \quad & a_i^\top x \leq b_i, \qquad i = 1, \dots, m, \end{aligned}$$

where $Q = Q^\top \succeq 0$, can be cast as an SOCP as

$$\begin{aligned} \min_{x,y} \quad & c^\top x + y \\ \text{s.t.:} \quad & \left\| \begin{bmatrix} 2Q^{1/2}x \\ y - 1 \end{bmatrix} \right\|_2 \leq y + 1, \\ & a_i^\top x \leq b_i, \qquad i = 1, \dots, m. \end{aligned}$$

Quadratic-constrained quadratic programs as SOCPs

The convex quadratic-constrained quadratic program (QCQP)

$$\begin{aligned} \min_x \quad & x^\top Q_0 x + a_0^\top x \\ \text{s.t.:} \quad & x^\top Q_i x + a_i^\top x \le b_i, \quad i = 1, \ldots, m, \end{aligned}$$

with $Q_i = Q_i^\top \succeq 0$, $i = 0, 1, \ldots, m$, can be cast as an SOCP as

$$\begin{aligned} \min_{x,t} \quad & a_0^\top x + t \\ \text{s.t.:} \quad & \left\| \begin{bmatrix} 2Q_0^{1/2} x \\ t - 1 \end{bmatrix} \right\|_2 \le t + 1, \\ & \left\| \begin{bmatrix} 2Q_i^{1/2} x \\ b_i - a_i^\top x - 1 \end{bmatrix} \right\|_2 \le b_i - a_i^\top x + 1, \quad i = 1, \ldots, m. \end{aligned}$$

Sums and maxima of norms

- The problem

$$\min_x \quad \sum_{i=1}^{p} \|A_i x - b_i\|_2,$$

where $A_i \in \mathbb{R}^{m,n}$, $b_i \in \mathbb{R}^m$ are given data, can be readily cast as an SOCP by introducing auxiliary scalar variables $y_1, \ldots, y_p$ and rewriting the problem as

$$\begin{aligned} \min_{x,y} \quad & \sum_{i=1}^{p} y_i \\ \text{s.t.:} \quad & \|A_i x - b_i\|_2 \leq y_i \quad i = 1, \ldots, p. \end{aligned}$$

- Similarly, the problem

$$\min_x \quad \max_{i=1,\ldots,p} \|A_i x - b_i\|_2$$

can be cast in SOCP format as

$$\begin{aligned} \min_{x,y} \quad & y \\ \text{s.t.:} \quad & \|A_i x - b_i\|_2 \leq y \quad i = 1, \ldots, p. \end{aligned}$$

Inventory control

Classic inventory control model of Harris (1913):

$$\min_{x>0}\; hx + \frac{cd}{x},$$

where

- x is the order quantity (to be determined);
- h is the annual cost of holding one unit in stock;
- c is the charge for a delivery, and d is the annual demand.

Multi-item extension

$$\min_x \sum_{i=1}^{n} h_i x_i + \frac{c_i d_i}{x_i} \;\;:\;\; b^T x \le b_0, \;\; l \le x \le u,$$

where

- $x \in \mathbb{R}^n$ is the order quantity vector;
- $h, c, d \in \mathbb{R}^n$ correspond to holding, delivery costs, and demand;
- $b_0, b \in \mathbb{R}^n$ correspond to space constraints;
- $l, u \in \mathbb{R}^n_{++}$ correspond to bounds on vector x.

SOCP model

We introduce slack variables to model the fractional part:

$$\min_{x,y} \sum_{i=1}^{n} h_i x_i + c_i d_i y_i \ : \ b^T x \leq b_0, \ \ l \leq x \leq u, \ \ y_i x_i \geq 1, \ \ 1 \leq i \leq n.$$

As seen in page 6, the hyperbolic constraints on $y, x \in \mathbb{R}^n_{++}$ can be equivalently expressed as a n second-order cone constraint in 3D:

$$\left\| \begin{pmatrix} 2 \\ yi - x_i \end{pmatrix} \right\|_2 \leq y_i + x_i, \ \ i = 1, \ldots, n. \tag{5}$$

Hence, the above problem is an SOCP.

Facility location problems

- Consider the problem of locating a warehouse to serve a number of service locations. The design variable is the location of the warehouse, $x \in \mathbb{R}^2$, while the service locations are given by the vector $y_i \in \mathbb{R}^2$, $i = 1, \ldots, m$.
- One possible location criterion is to determine x so as to minimize the maximum distance from the warehouse to any location. This amounts to consider a minimization problem of the form

$$\min_x \max_{i=1,\ldots,m} \|x - y_i\|_2,$$

which is readily cast in SOCP form as follows:

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.:} \quad & \|x - y_i\|_2 \le t, \quad i = 1, \ldots, m. \end{aligned}$$

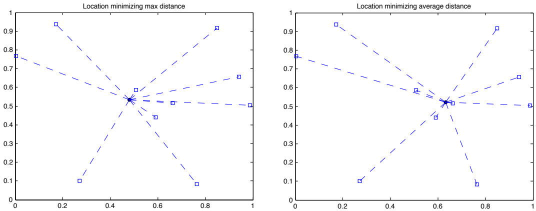
Facility location problems

- An alternative location criterion, which is a good proxy for the average transportation cost, is the *average* distance from the warehouse to the facilities:

$$\min_x \frac{1}{m}\sum_{i=1}^m \|x - y_i\|_2,$$

which can be cast as the SOCP

$$\begin{array}{ll} \min\limits_{x,t} & \frac{1}{m}\sum_{i=1}^m t_i \\ \text{s.t.:} & \|x - y_i\|_2 \le t_i, \quad i = 1, \ldots, m. \end{array}$$



Square-root LASSO

$$p^* := \min_w \; \|X^T w - y\|_2 + \lambda \|w\|_1$$

where

- $X \in \mathbb{R}^{m \times n} = [a_1, \ldots, a_n]$ is the data matrix, with $a_i \in \mathbb{R}^m$ the vector that corresponds to feature i;
- $y \in \mathbb{R}^m$ is a response vector;
- $\lambda > 0$ is a sparsity-inducing parameter;
- $w \in \mathbb{R}^n$ is the vector of regression coefficients.

Above is an SOCP (why?)

Examplar selection

We are given a data matrix $X = [x_1, \ldots, x_m] \in \mathbb{R}^{m \times n}$, with $x_i \in \mathbb{R}^m$ the data points. We seek to find a subset of data points $\{x_j\}_{j \in \mathcal{J}}$, with $\mathcal{J} \subseteq \{1, \ldots, m\}$ having a low number of elements, such that

$$\forall\, i \in \{1, \ldots, n\} \;:\; x_i \approx \sum_{j \in \mathcal{J}} x_j w_{ij}$$

In other words, every data point can be accurately represented as a linear combination of the same few data points. We can write the above condition as $X \approx XW^\top$, with the matrix W having a lot of its *columns* entirely zero.

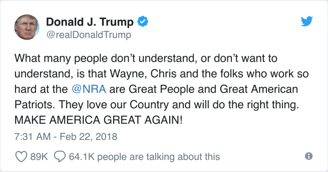
The problem can be modeled as an SOCP:

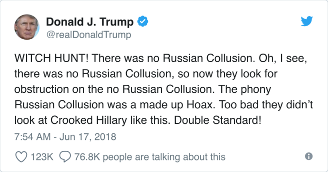
$$\min_{W=[w_1,\ldots,w_m]} \|X - XW^T\|_F \;:\; \sum_{j=1}^{m} \|w_j\|_2 \leq \kappa,$$

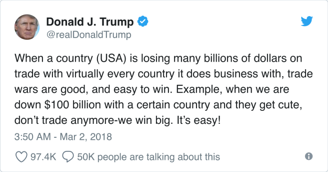
where a small value of the parameter $\kappa > 0$ encourages many columns in W to be entirely zero. The indices j of the non-zero columns form the set $\mathcal{J}$.

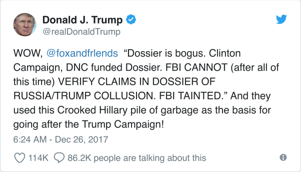
Example

Examplars for D. Trump's tweets (by G. Cheng, A. Askari)









4 of 10 exemplars computed from $10,000$ randomly subsampled tweets from Donald Trump's tweet corpus as of 9/26/18. We remove all twitter handles, numbers, urls, punctuation, and English stop-words as specified by sklearn, and use a Count Vectorizer to represent the text data as a matrix.

Problems involving powers of variables

SOCP can model some problems that involve powers of variables. This is useful for example in the context of a variant of least-squares:

$$p^* := \min_w \|X^T w - y\|_2 + \lambda \sum_{i=1}^n |w_i|^\alpha,$$

where $\alpha > 1$ is the ratio of two integers. A typical choice is $\alpha \in (1, 2)$, so as to achieve a good trade-off between encouraging sparsity and robustness to noise.

Take $\alpha = 3/2$ for example. The condition $t \geq u^{3/2}$ for $u \geq 0$, is equivalent to the existence of $v \geq 0$ such that

$$vt \geq u^2, \quad u \geq v^2.$$

The above problem with $\alpha = 3/2$ can thus be modeled as an SOCP (with rotated cone constraints):

$$p^* = \min_{w,v,u,t} \|X^T w - y\|_2 + \lambda \sum_{i=1}^n t_i \quad : \quad \begin{array}{l} t \geq 0, \quad u \geq |w|, \\ v_i t_i \geq u_i^2, \quad u_i \geq v_i^2, \quad i = 1, \ldots, n. \end{array}$$