CS 2601 Linear and Convex Optimization

5. Convex optimization problems (part 2)

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

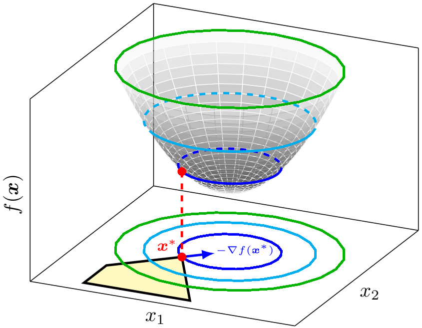
- Quadratic program and quadratically constrained QP
- Geometric program

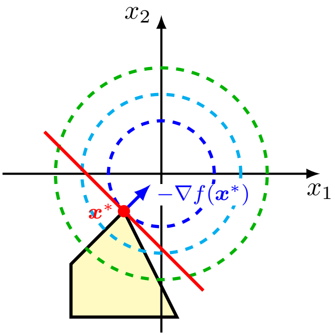
Quadratic program (QP)

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{x} \le \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

半正定

QP is convex iff $\boldsymbol{Q} \succeq \boldsymbol{O}$. Reduces to LP if $\boldsymbol{Q} = \boldsymbol{O}$.

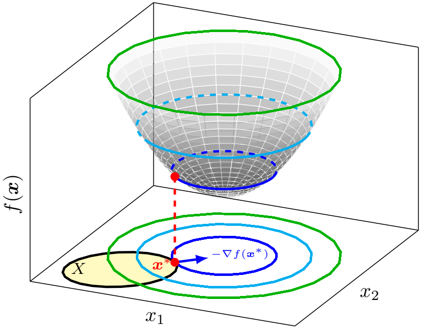


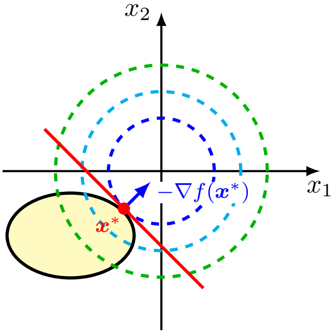


Quadratically constrained quadratic program (QCQP)

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\
\text{s.t.} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}_i\boldsymbol{x} + \boldsymbol{c}_i^T\boldsymbol{x} + \boldsymbol{d}_i \leq 0, \quad i = 1, 2, \ldots, m \\
& \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}
\end{aligned}
$$

QCQP is convex if $\boldsymbol{Q} \succeq \boldsymbol{O}$ and $\boldsymbol{Q}_i \succeq \boldsymbol{O}$, $\forall i$. Reduces to QP if $\boldsymbol{Q}_i = \boldsymbol{O}$, $\forall i$.





Example: Linear least squares regression

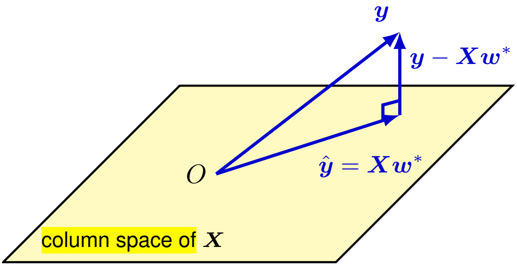
Given $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, find $\boldsymbol{w} \in \mathbb{R}^p$ s.t.

$$\min_{\boldsymbol{w}} \ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2$$

- convex QP with objective

$$f(\boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} - 2\boldsymbol{y}^T\boldsymbol{X}\boldsymbol{w} + \boldsymbol{y}^T\boldsymbol{y}$$

Geometrically, we are looking for the orthogonal projection $\hat{\boldsymbol{y}}$ of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$. Does the solution always exist?



Example: Linear least squares regression (cont'd)

By the first-order optimality condition, $\boldsymbol{w}^*$ is optimal iff

$$\nabla f(\boldsymbol{w}^*) = \mathbf{0}$$

无限制，优化条件就是梯度为0

i.e. $\boldsymbol{w}^*$ is a solution of the normal equation,

$$\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) = \mathbf{0} \iff \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{y}$$

Note. $\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}^*) = \mathbf{0}$ means precisely $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}^*$ is perpendicular to the column space of $\boldsymbol{X}$.

Grem Matrix

Case I. $\boldsymbol{X}$ has full column rank, i.e. $\operatorname{rank} \boldsymbol{X} = p$

- $\boldsymbol{X}^T\boldsymbol{X} \succ \boldsymbol{O}$
- unique solution

$$\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Example: Linear least squares regression (cont'd)

Example. Solve

$$\min_{\boldsymbol{w}} \ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2$$

with

$$\boldsymbol{X} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}.$$

Solution. The normal equation is

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{y}$$

with

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

Since $\boldsymbol{X}$ has full column rank,

$$\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix}$$

Example: Linear least squares regression (cont'd)

Case II. $\operatorname{rank} \boldsymbol{X} = r < p$. WLOG assume the first r columns are linearly independent, i.e. Without Loss of Generality

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$$

where $\boldsymbol{X}_1 \in \mathbb{R}^{n \times r}$ and $\operatorname{rank} \boldsymbol{X}_1 = r$.

Claim. There is a solution $\boldsymbol{w}^*$ with the last $p - r$ components being 0.

- $\boldsymbol{X}$ and $\boldsymbol{X}_1$ have the same column space
- If $\boldsymbol{w}_1^*$ solves

$$\min_{\boldsymbol{w}_1 \in \mathbb{R}^r} \|\boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{w}_1\|$$

  then $\boldsymbol{w}^* = \begin{bmatrix} \boldsymbol{w}_1^* \\ \boldsymbol{0} \end{bmatrix}$ solves $\min_{\boldsymbol{w} \in \mathbb{R}^p} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|$
- $\boldsymbol{w}_1^* = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{y}$

Question. Is the solution unique in this case?

A. $\operatorname{rank} \boldsymbol{X} < p \implies \exists \boldsymbol{w}_0 \neq \boldsymbol{0}$ s.t. $\boldsymbol{X}\boldsymbol{w}_0 = \boldsymbol{0}$, so $\boldsymbol{w}^* + \boldsymbol{w}_0$ is also a solution.

Example: Linear least squares regression (cont'd)

Example Solve $\min_{\boldsymbol{w}} \ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2$ with

$$\boldsymbol{X} = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 3 \\ 2 \\ 2 \end{bmatrix}.$$

Solution. Note $\operatorname{rank} \boldsymbol{X} = 2 < 3$.

- Let

$$\boldsymbol{X}_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

- By the previous example,

$$\boldsymbol{w}_1^* = (\boldsymbol{X}_1^T \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1^T \boldsymbol{y} = (1.5, 2)^T$$

  is a solution to $\min_{\boldsymbol{w}_1 \in \mathbb{R}^2} \|\boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{w}_1\|^2$.
- $\boldsymbol{w}^* = (1.5, 2, 0)^T$ is a solution to $\min_{\boldsymbol{w} \in \mathbb{R}^3} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|^2$.

Example: Linear least squares regression (cont'd)

Example (cont'd). The normal equation to the original problem is

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{w} = \boldsymbol{X}^T\boldsymbol{y}$$

where

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 4 & 0 & 4 \\ 0 & 1 & -1 \\ 4 & -1 & 5 \end{bmatrix}, \quad \boldsymbol{X}^T\boldsymbol{y} = \begin{bmatrix} 6 \\ 2 \\ 4 \end{bmatrix}$$

- Note $\boldsymbol{X}^T\boldsymbol{X}$ is not invertible, so we cannot use the formula[1] $\boldsymbol{w}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$
- The solution $\boldsymbol{w}^* = (1.5, 2, 0)^T$ satisfies the normal equation.
- The normal equation has infinitely many solutions given by

$$\boldsymbol{w} = (1.5, 2, 0)^T + \alpha(-1, 1, 1)^T, \quad \alpha \in \mathbb{R}.$$

All of them are solutions to the least squares problem.

[1] This formula still applies if we use the so-called pseudo inverse of $\boldsymbol{X}^T\boldsymbol{X}$.

General unconstrained QP

Minimize quadratic function with $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ s.t. $\boldsymbol{Q} \succeq \boldsymbol{O}$,

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

By the first-order condition, the solutions satisfy

$$\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0}$$

Case I. $\boldsymbol{Q} \succ \boldsymbol{O}$. There is a unique solution $\boldsymbol{x}^* = -\boldsymbol{Q}^{-1}\boldsymbol{b}$.

Case II. $\det \boldsymbol{Q} = 0$ and $\boldsymbol{b} \in$ column space of $\boldsymbol{Q}$.[2] There are infinitely many solutions. (why?)

Case III. $\det \boldsymbol{Q} = 0$ and $\boldsymbol{b} \notin$ column space of $\boldsymbol{Q}$. There is no solution, and $f^* = -\infty$.

---

[2]This is equivalent to $-\boldsymbol{b} \in$ column space of $\boldsymbol{Q}$, the latter being just another way to say $\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0}$ has a solution.

General unconstrained QP (cont'd)

To understand why $f^* = -\infty$ in case III, first assume $\boldsymbol{Q}$ is diagonal.

Example. $n = 3$, $\boldsymbol{Q} = \text{diag}\{\lambda_1, \lambda_2, 0\}$ with $\lambda_1, \lambda_2 > 0$, $\boldsymbol{b} = (b_1, b_2, b_3)^T$, $c = 0$.

$$f(\boldsymbol{x}) = \left(\frac{\lambda_1}{2}x_1^2 + b_1 x_1\right) + \left(\frac{\lambda_2}{2}x_2^2 + b_2 x_2\right) + b_3 x_3$$

The column space of $\boldsymbol{Q}$ is

$$\text{span}\left\{\begin{bmatrix}\lambda_1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\\lambda_2\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}\right\} = \{(x_1, x_2, 0)^T : x_1, x_2 \in \mathbb{R}\}$$

So $\boldsymbol{b} \notin$ column space of $\boldsymbol{Q} \iff b_3 \neq 0$.

Since $f(\boldsymbol{x})$ is affine in x_3, it is unbounded below, so $f^* = -\infty$.

General unconstrained QP (cont'd)

不是对角化可以让他对角化

When $\boldsymbol{Q}$ is non-diagonal,

- Diagonalize $\boldsymbol{Q}$ by an orthogonal matrix $\boldsymbol{U}$, so

$$\boldsymbol{Q} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T, \quad \text{where } \boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_n\}$$

- Let $\boldsymbol{x} = \boldsymbol{U}\boldsymbol{y}$ and $\boldsymbol{b} = \boldsymbol{U}\tilde{\boldsymbol{b}}$. Then

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{U}^T\boldsymbol{Q}\boldsymbol{U}\boldsymbol{y} + \tilde{\boldsymbol{b}}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{y} + c = \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y} + \tilde{\boldsymbol{b}}^T\boldsymbol{y} + c \triangleq g(\boldsymbol{y})$$

- Minimizing $f(\boldsymbol{x})$ is equivalent to minimizing $g(\boldsymbol{y})$.
  - $\det \boldsymbol{Q} = 0 \iff \det \boldsymbol{\Lambda} = 0$; $\quad \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{0} \iff \boldsymbol{\Lambda}\boldsymbol{y} + \tilde{\boldsymbol{b}} = \boldsymbol{0}$
- In case III, $\exists i_0$ s.t. $\lambda_{i_0} = 0$ but $\tilde{b}_{i_0} \neq 0$, so $g(\boldsymbol{y})$ is affine in y_{i_0} and hence unbounded below,

$$g(\boldsymbol{y}) = \sum_{i \neq i_0} \left(\frac{1}{2}\lambda_i y_i^2 + \tilde{b}_i y_i \right) + \tilde{b}_{i_0} y_{i_0} + c \implies f^* = g^* = -\infty$$

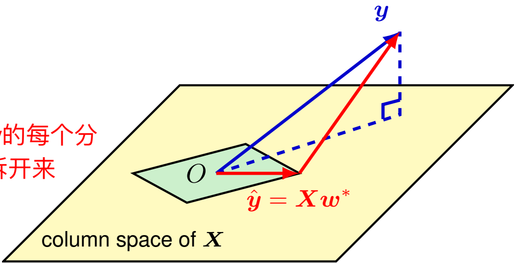
Example: Lasso

Lasso (Least Absolute Shrinkage and Selection Operator)

Given $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, $t > 0$,

$$\begin{array}{ll} \min\limits_{\boldsymbol{w}} & \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \\ \mathrm{s.t.} & \|\boldsymbol{w}\|_1 \le t \end{array}$$

把w的每个分
量拆开来

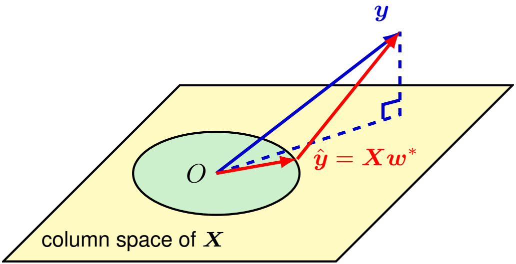


- convex problem? yes
- QP? no, but can be converted to QP
- optimal solution exists? yes
  - compact feasible set
- optimal solution unique?
  - yes if $n \ge p$ and $\boldsymbol{X}$ has full column rank ($\boldsymbol{X}^T\boldsymbol{X} \succ \boldsymbol{O}$, strictly convex)
  - no in general, e.g. $p > n$ and t is large enough for unconstrained optima to be feasible

Example: Ridge regression

Given $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n\times p}$, $t > 0$,

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \\ \text{s.t.} \quad & \|\boldsymbol{w}\|_2^2 \le t \end{aligned}$$



- convex problem? yes
- QCQP? yes
- optimal solution exists? yes
  - compact feasible set
- optimal solution unique?
  - yes if $n \ge p$ and $\boldsymbol{X}$ has full column rank ($\boldsymbol{X}^T\boldsymbol{X} \succ \boldsymbol{O}$, strictly convex)
  - no in general

Example: SVM

Linearly separable case

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}\|\boldsymbol{w}\|^2$$

quadratic function

$$\text{s. t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1, \quad i = 1, 2, \ldots, m$$

affine function

Soft margin SVM

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s. t.} \quad y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, m$$

$$\boldsymbol{\xi} \geq \mathbf{0}$$

Equivalent unconstrained form

This is not linear, so no longer a QP problem.

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{n}(1 - y_i b - y_i\boldsymbol{x}_i^T\boldsymbol{w})^+$$

Outline

- Quadratic program and quadratically constrained QP
- Geometric program

Geometric program

A monomial is a function $f : \mathbb{R}^n_{++} = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > \boldsymbol{0}\} \to \mathbb{R}$ of the form

单项式

$$f(\boldsymbol{x}) = \gamma x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

for $\gamma > 0$, $a_1, \ldots, a_n \in \mathbb{R}$. A posynomial is a sum of monomials,

正项式

$$f(\boldsymbol{x}) = \sum_{k=1}^{p} \gamma_k x_1^{a_{k1}} x_2^{a_{k2}} \cdots x_n^{a_{kn}}$$

A geometric program (GP) is an optimization problem of the form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s. t.} \quad & g_i(\boldsymbol{x}) \leq 1, \quad i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 1, \quad j = 1, \ldots, r \end{aligned}$$

where $f, g_i, i = 1, \ldots, m$ are posynomials and $h_j, j = 1, \ldots, r$ are monomials. The constraint $\boldsymbol{x} > \boldsymbol{0}$ is implicit.

Geometric program (cont'd)

GP is nonconvex (why?)

$$\min_{\boldsymbol{x}} \quad \sum_{k=1}^{p_0} \gamma_{0k} x_1^{a_{0k1}} x_2^{a_{0k2}} \cdots x_n^{a_{0kn}} \qquad \text{Not convex}$$

$$\text{s. t.} \quad \sum_{k=1}^{p_i} \gamma_{ik} x_1^{a_{ik1}} x_2^{a_{ik2}} \cdots x_n^{a_{ikn}} \le 1, \quad i = 1, \ldots, m \qquad \text{Not convex}$$

$$\eta_j x_1^{c_{j1}} x_2^{c_{j2}} \cdots x_n^{c_{jn}} = 1, \quad j = 1, \ldots, r \qquad \text{Not affine}$$

By $y_i = \log x_i$, $b_{ik} = \log \gamma_{ik}$, $d_j = \log \eta_j$, GP can be formulated as

$$\min_{\boldsymbol{y}} \quad \log \left(\sum_{k=1}^{p_0} e^{\boldsymbol{a}_{0k}^T \boldsymbol{y} + b_{0k}} \right) \qquad \text{Log–sum funcition}$$

$$\text{s. t.} \quad \log \left(\sum_{k=1}^{p_i} e^{\boldsymbol{a}_{ik}^T \boldsymbol{y} + b_{ik}} \right) \le 0, \quad i = 1, \ldots, m$$

$$\boldsymbol{c}_j^T \boldsymbol{y} + d_j = 0, \quad j = 1, \ldots, r \qquad \text{affine}$$

This is convex by the convexity of log-sum-exp (soft max) functions

Geometric program (cont'd)

Example. Let $u = \log x$, $v = \log y$, $w = \log z$.

$$
\begin{aligned}
\min_{x,y,z>0} \quad & x^{-1}y + xz \\
\text{s. t.} \quad & 2x^{-1} \leq 1 \\
& \frac{1}{3}x \leq 1 \\
& x^2y^{-1/2} + 3y^{1/2}z^{-1} \leq 1 \\
& x^2y^{-1}z^{-2} = 1
\end{aligned}
$$

is equivalent to

$$
\begin{aligned}
\min_{u,v,w} \quad & \log(e^{-u+v} + e^{u+w}) \\
\text{s. t.} \quad & \log 2 - u \leq 0 \\
& -\log 3 + u \leq 0 \\
& \log(e^{2u-\frac{1}{2}v} + e^{\log 3+\frac{1}{2}v-w}) \leq 0 \\
& 2u - v - w = 0
\end{aligned}
$$