Math Review Notes—Convex Optimization

Gregory Faletto

1 Convex Optimization

These are my notes from taking EE 588 at USC and the textbook *Convex Optimization* (Boyd and Vandenberghe) 7th printing.

Need to cover:

- Update rules for optimization problems (e.g. gradient descent, be able to write down gradient, etc.)
- Know which algorithms are useful in which settings
- Homework-like problems from first part of class (no proofs though) (Boyd homework is good practice)
- Understand how to derive algorithms
- Understand how to calculate gradients, proximal functions, etc.
- Understand examples, how to run algorithms
- Only conceptual thing: duality question (write down dual)
- Formulate problems as convex optimization problems

Do not need to cover:

- ADMM
- Proofs from 2nd half of class (rates of convergence, etc.)
- Coding

1.1 Convex Functions

Theorem 1. Jensen's Inequality: f is convex if and only if

$$\frac{f(a)+f(b)}{2} \geq f\left(\frac{a+b}{2}\right)$$

for all $a, b \in \mathbf{dom}(f)$.

1.2 Schur Complement Trick

1.2.1 Definition

For a matrix $X \in \boldsymbol{S}^n$ partitioned as

$$X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

the Schur complement is (if $\mathbf{det}(A) \neq 0$)

$$S = C - B^T A^{-1} B$$

The Schur complement has two useful properties in convex analysis.

Theorem 2. (a) $X \succ 0$ if and only if $A \succ 0$ and $S \succ 0$.

(b) If $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$.

1.2.2 The Trick

Suppose we are trying to express a problem as a semidefinite program (SDP); that is, in the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \ldots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

where $G, F_1, \ldots, F_n \in \boldsymbol{S}^k$ and $A \in \mathbb{R}^{p \times n}$. If we have a constraint of the form $c^T F(x)^{-1} c \leq t$ where $F(x)$ is symmetric and positive definite and $t \in \mathbb{R}$, by Theorem 2(b) we can write

$$c^T F(x)^{-1} c \leq t \iff \begin{bmatrix} F(x) & c \\ c^T & t \end{bmatrix} \succeq 0$$

in order to get our constraint in the form required for an SDP.

1.2.3 Example 1: Last Year's Final, Question 2(b)

Suppose we have the constraints

$$\begin{gathered} Ax + b \geq 0 \\ \frac{(c^T x)^2}{d^T x} \leq t \end{gathered}$$

which we would like to express in an SDP. By Theorem 2(b) we can write

$$\frac{(c^T x)^2}{d^T x} \leq t \iff d^T x - (c^T x)^T t^{-1} c^T x \geq 0 \iff \begin{bmatrix} t & c^T x \\ c^T x & d^T x \end{bmatrix} \succeq 0$$

Since

$$Ax + b \geq 0 \iff \mathbf{diag}(Ax + b) \succeq 0$$

we can finally write our constraints as

$$\begin{bmatrix} \mathbf{diag}(Ax+b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \succeq 0$$

1.2.4 Example 2: Last Year's Final, Question 4(b)

Suppose we have the constraints

$$\begin{aligned} Ax + b &\geq 0 \\ \frac{(c^T x)^2}{d^T x} &\leq t \end{aligned}$$

which we would like to express in an SDP. By Theorem 2(b) we can write

$$\frac{(c^T x)^2}{d^T x} \leq t \iff d^T x - (c^T x)^T t^{-1} c^T x \geq 0 \iff \begin{bmatrix} t & c^T x \\ c^T x & d^T x \end{bmatrix} \succeq 0$$

Since

$$Ax + b \geq 0 \iff \mathbf{diag}(Ax+b) \succeq 0$$

we can finally write our constraints as

$$\begin{bmatrix} \mathbf{diag}(Ax+b) & 0 & 0 \\ 0 & t & c^T x \\ 0 & c^T x & d^T x \end{bmatrix} \succeq 0$$

1.3 Duality

Theorem 3. Slater's condition/constraint qualification: Strong duality holds for a convex problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\ & Ax = b \end{array}$$

if it is strictly feasible, i.e., there exists at least one x in the domain of f_0 such that $f_i(x) < 0,\ i = 1, 2, \ldots, m$, $Ax = b$.

1.4 MLE estimates

For linear estimates with iid noise

$$y_i = a_i^T x + v_i, i = 1, \dots, m$$

where a is observed and $x \in \mathbb{R}^n$ are the parameters to be estimated, the likelihood function is

$$p_x(y) = \prod_{i=1}^{m} \Pr(v_i = y_i - a_i^T x \mid x)$$

Therefore the log likelihood function is:

$$\ell_x(y) = \sum_{i=1}^{m} \log[\Pr(v_i = y_i - a_i^T x \mid x)]$$

1.5 Practice Final (2017 Final)

(1) (a) Strictly convex. Multiply by x/x (allowed in this case since $x > 0$) to get $\frac{x^2}{x+1}$ which is a quadratic over linear, which is convex in $\mathbb{R}^{++}$ according to CVX rules.

(b) Not convex, it is convex for $x \geq -1$, but there is a boundary problem at $x = -1$. Note that Jensen's inequality (Theorem 1)

$$\frac{f(a) + f(b)}{2} \geq f\left(\frac{a+b}{2}\right)$$

is violated because

$$\frac{f(-1.3) + f(-0.9)}{2} = \frac{2.3 + 0}{2} = 1.15 \leq 2.2 = f(-1.1) = f\left(\frac{-1.3 + -0.9}{2}\right)$$

(c)

(d)

$$f(x) = \sup \log\left(\frac{p(t)}{q(t)}\right) = \sup\{\log p(t) - \log q(t)\} = \sup\{\log\left(\sum_{i=1}^{n} \exp(x_i \sin(it))\right) - \sum_{i=1}^{n} x_i \sin(it)\}$$

(e) The proximal mapping is

$$\text{prox}_{\mathcal{R}}(z) = \arg\min_y \frac{1}{2}\|z - y\|_2^2 + \mathcal{R}(y) = \arg\min_y \frac{1}{2}\sum_{i=1}^{n}(z_i - y_i)^2 + \sum_{i=1}^{n} w_i|y_i|$$

$$= \arg\min_y \frac{1}{2}\sum_{i=1}^{n}\left[(z_i - y_i)^2 + w_i|y_i|\right]$$

Taking the gradient of the inside quantity with respect to y, we have

$$\nabla(y) = \begin{pmatrix} \frac{1}{2} \cdot 2(z_1 - y_1) + \mathbf{sign}(y_1) w_1 \\ \frac{1}{2} \cdot 2(z_2 - y_2) + \mathbf{sign}(y_2) w_2 \\ \vdots \\ \frac{1}{2} \cdot 2(z_n - y_n) + \mathbf{sign}(y_n) w_n \end{pmatrix} = \begin{pmatrix} z_1 - y_1 + \mathbf{sign}(y_1) w_1 \\ z_2 - y_2 + \mathbf{sign}(y_2) w_2 \\ \vdots \\ z_n - y_n + \mathbf{sign}(y_n) w_n \end{pmatrix}$$

Setting equal to 0, we have

$$y = \begin{pmatrix} z_1 \pm w_1 \\ z_2 \pm w_2 \\ \vdots \\ z_n \pm w_n \end{pmatrix}$$

(2) (a) The constraint is convex (affine). The denominator is affine. Since $c^T x = x^T c$, the numerator

$$(c^T x)^2 = (c^T x)(c^T x) = x^T c c^T x = x^T (c c^T) x$$

is convex since cc^T is positive semidefinite.

(b) We start by using the epigraph trick to transform the problem:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \dfrac{(c^T x)^2}{d^T x} \leq t \\ & Ax + b \geq 0 \end{array}$$

We are trying to express this problem as a semidefinite program (SDP); that is, in the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \ldots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

where $G, F_1, \ldots, F_n \in \boldsymbol{S}^k$ and $A \in \mathbb{R}^{p \times n}$. The first constraint

$$\frac{(c^T x)^2}{d^T x} \leq t$$

can be expressed in the form

$$(c^T x)^2 \leq t d^T x \iff (c^T x c^T - t d^T) x \leq 0$$

We have a constraint

$$Ax + b \geq 0$$

which can be expressed in the form

$$Ax \geq -b$$

$$c^T F(x)^{-1} c \leq t$$

where $F(x)$ is symmetric and positive definite and $t \in \mathbb{R}$, by Theorem 2(b) we can write

$$c^T F(x)^{-1} c \le t \iff \begin{bmatrix} F(x) & c \\ c^T & t \end{bmatrix} \succeq 0$$

in order to get our constraint in the form required for an SDP.

(3) (a) Yes, g is convex over $\mathcal{X}$ since it is quadratic over linear.

(b) The only points satisfying the constraint have $x_1 = 0$. Therefore the primal optimal value (the only feasible value) is $e^0 = \boxed{1}$.

(c) Lagrangian:

$$L(x, \lambda) = e^{-x_1} + \lambda(x_1^2/x_2)$$

The Lagrangian obtains its minimum value of 0 when $x_2 = x_1^3$ and $x_1 \to \infty$. Thus, its dual function ($g(\lambda) = \min_x L(x, \lambda)$) is

$$g(\lambda) = 0$$

The dual problem is then

$$\boxed{\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & \lambda \ge 0 \end{array}}$$

(d) The optimal value of the dual problem is 0. Strong duality does not hold since the optimum of the dual problem is less than the optimum of the primal problem. We can also tell this because Slater's Condition (Theorem 3) is violated; that is, there is no (x_1, x_2) that is strictly feasible since x_1 must equal 0, which is on the boundary of the feasible region.

(e) Now for the primal problem, instead of $x_1 = 0$, we have

$$\frac{x_1^2}{x_2} \le u \iff x_1^2 \le u x_2 \implies -\sqrt{u x_2} \le x_1 \le \sqrt{u x_2}$$

Since e^{-x_1} is minimized as $x_1 \to \infty$, our optimal solution is $x_2 \to \infty, x_1 = \sqrt{u x_2} \to \infty$ yielding a primal optimal value of $\boxed{0}$. For the dual problem, we have

$$L(x, \lambda) = e^{-x_1} + \lambda\left(\frac{x_1^2}{x_2} - u\right)$$

Dual function ($g(\lambda) = \min_x L(x, \lambda)$):

$$\frac{x_1^2}{x_2} - u = 0 \implies x_2 = \frac{x_1^2}{u}$$

and let $x_1 \to -\infty$ to yield

$$g(\lambda) = 0$$

The dual problem is then

$$\boxed{\text{maximize} \quad 0}$$

with optimal value 0, so there is no longer a duality gap. We can also tell this because Slater's Condition (Theorem 3) is satisfied; that is, there exists an (x_1, x_2) which is strictly feasible (say $(x_1, x_2) = (\sqrt{u}, 10)$).

(4) (a) Yes, the set is convex. If $(u_i, v_i) = \boldsymbol{u}_i$, each

$$\sqrt{(x-u_i)^2 + (y-v_i)^2} = \|\boldsymbol{x} - \boldsymbol{u}_i\|_2$$

is convex in $\boldsymbol{x}$. Therefore the function

$$\sum_{i=1}^{k} \|\boldsymbol{x} - \boldsymbol{u}_i\|_2$$

is convex. For any fixed d, this set is a sublevel set of this function, which is convex since the function is convex.

(b) This is a feasibility problem:

$$\begin{array}{ll} \text{find} & \boldsymbol{x} \\ \text{subject to} & \displaystyle\sum_{i=1}^{k} \|\boldsymbol{x} - \boldsymbol{u}_i\| \leq d \\ & \displaystyle\sum_{i=1}^{j} \|\boldsymbol{x} - \boldsymbol{v}_i\| \leq e \end{array}$$

or

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \displaystyle\sum_{i=1}^{k} \|\boldsymbol{x} - \boldsymbol{u}_i\| \leq d \\ & \displaystyle\sum_{i=1}^{j} \|\boldsymbol{x} - \boldsymbol{v}_i\| \leq e \end{array}$$

for two sets of points in $\mathbb{R}^2$ $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$. We would like to express these constraints as matrix inequalities in order to have an SDP. To do this, first rewrite the problem as

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \|\boldsymbol{x} - \boldsymbol{u}_i\| \leq t_i, i = 1, \ldots, k \\ & \|\boldsymbol{x} - \boldsymbol{v}_i\| \leq s_i, s = 1, \ldots, j \\ & \mathbf{1}^T t \leq d \\ & \mathbf{1}^T s \leq e \end{array}$$

Then note that we can use the Schur trick:

$$(\boldsymbol{x} - \boldsymbol{u}_i)^T I (\boldsymbol{x} - \boldsymbol{u}_i) \leq t_i \iff \begin{bmatrix} I & \boldsymbol{x} - \boldsymbol{u}_i \\ (\boldsymbol{x} - \boldsymbol{u}_i)^T & t_i \end{bmatrix} \succeq 0$$

and write the optimization problem as an SDP:

$$\boxed{\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & \begin{bmatrix} I & \boldsymbol{x}-\boldsymbol{u}_i \\ (\boldsymbol{x}-\boldsymbol{u}_i)^T & t_i \end{bmatrix} \succeq 0, i=1,\ldots,k \\ & \begin{bmatrix} I & \boldsymbol{x}-\boldsymbol{v}_i \\ (\boldsymbol{x}-\boldsymbol{v}_i)^T & s_i \end{bmatrix} \succeq 0, s=1,\ldots,j \\ & \mathbf{1}^T t \leq d \\ & \mathbf{1}^T s \leq e \end{array}}$$

(5) (a) To minimize the MSE:

$$\mathcal{L}(z) = \sum_r (y_r - |a_r^T x|^2)^2$$

For MLE estimate:

$$p_x(y) = \prod_{r=1}^m \Pr(w_r = y_r - (a_r^T x)^2 \mid x) = \frac{1}{(y_r - (a_r^T x)^2)!} \cdot \exp\big(-(a_r^T x)^2\big) \cdot (a_r^T x)^{2[y_r - (a_r^T x)^2]}$$

Therefore the log likelihood function is:

$$\ell_x(y) = \sum_{i=1}^m \log[\Pr(y_i - a_i^T x \mid x)] = \sum_{i=1}^m \log\left[\frac{1}{(y_r - (a_r^T x)^2)!} \cdot \exp\big(-(a_r^T x)^2\big) \cdot (a_r^T x)^{2[y_r - (a_r^T x)^2]}\right]$$

$$= \sum_{i=1}^m \log\left[\frac{1}{(y_r - (a_r^T x)^2)!}\right] - (a_r^T x)^2 + 2[y_r - (a_r^T x)^2] \cdot \log\big[(a_r^T x)\big]$$

(b) b

(c) c

(d) d

(e) e