SI251 - Convex Optimization homework 1

Deadline: 2022-10-09 23:59:59

1. You can use Word, Latex or handwriting to complete this assignment. If you want to submit a handwritten version, scan it clearly.
2. The **report** has to be submitted as a PDF file to Gradescope, other formats are not accepted.
3. The submitted file name is **student_id+your_student_name.pdf**.
4. Late policy: You have 4 free late days for the quarter and may use up to 2 late days per assignment with no penalty. Once you have exhausted your free late days, we will deduct a late penalty of 25% per additional late day. Note: The timeout period is recorded in days, even if you delay for 1 minute, it will still be counted as a 1 late day.
5. You are required to follow ShanghaiTech's academic honesty policies. You are not allowed to copy materials from other students or from online or published resources. Violating academic honesty can result in serious sanctions.

Any plagiarism will get Zero point.

1 Convex sets

1. (15 pts) Please prove that the following sets are convex:

   1) **(Ellipsoids)** $\left\{x | \sqrt{(x-x_c)^T P (x-x_c)} \leq r\right\} \quad (x_c \in \mathbb{R}^n, r \in \mathbb{R}, P \succeq 0)$;

   2) **(Symmetric positive semidefinite matrices)** $S_+^{n\times n} = \left\{P \in S^{n\times n} | P \succeq 0\right\}$;

   3) The set of points closer to a given point than a given set, i.e.,

   $$\left\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\right\},$$

   where $S \in R^n$.

   Solution:

   1) Define $\|x\|_P = \sqrt{x^T P x}$. As this set is closed, it suffices to show midpoint convexity. Pick $x, y \in S$. We have $\|x - x_c\|_P \leq r$ and $\|y - x_c\|_P \leq r$. Then

   $$\begin{aligned}\left\|\frac{x+y}{2} - x_c\right\|_P &= \left\|\frac{1}{2}x - \frac{1}{2}x_c + \frac{1}{2}y - \frac{1}{2}x_c\right\|_P \\ &\leq \frac{1}{2}\|x - x_c\|_P + \frac{1}{2}\|y - x_c\|_P \\ &\leq \frac{1}{2}r + \frac{1}{2}r = r.\end{aligned}$$

   2) Let $A \succeq 0, B \succeq 0$ and $\lambda \in [0,1]$. Pick $y \in \mathbb{R}^n$, then

   $$y^T(\lambda A + (1-\lambda)B)y = \lambda y^T A y + (1-\lambda) y^T B y \geq 0.$$

   3) This set is convex because it can be expressed as

   $$\bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\},$$

   i.e., an intersection of halfspaces. (For fixed y, the set $\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$ is a halfspace).

2. (15 pts) Let C be a nonempty subset of $\mathbb{R}^n$, and let λ_1 and λ_2 be positive scalars. Show that if C is convex, then $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$. Show by example that this need not be true when C is not convex.

   Hint: A vector x in $\lambda_1 C + \lambda_2 C$ is of the form $x = \lambda_1 x_1 + \lambda_2 x_2$, where $x_1, x_2 \in C$.

   Solution: We always have $(\lambda_1 + \lambda_2)C \subset \lambda_1 C + \lambda_2 C$, even if C is not convex. To show the reverse inclusion assuming C is convex, note that a vector x in $\lambda_1 C + \lambda_2 C$ is of the form $x = \lambda_1 x_1 + \lambda_2 x_2$, where $x_1, x_2 \in C$. By convexity of C, we have

   $$\frac{\lambda_1}{\lambda_1 + \lambda_2} x_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_2 \in C,$$

   and it follows that

   $$x = \lambda_1 x_1 + \lambda_2 x_2 \in (\lambda_1 + \lambda_2)C,$$

   so $\lambda_1 C + \lambda_2 C \in (\lambda_1 + \lambda_2)C$. For a counterexample when C is not convex, let C be a set in $\mathbb{R}^n$ consisting of two vectors, 0 and $x \neq 0$, and let $\lambda_1 = \lambda_2 = 1$. Then C is not convex, and $(\lambda_1 + \lambda_2)C = 2C = \{0, 2x\}$, while $\lambda_1 C + \lambda_2 C = C + C = \{0, x, 2x\}$, showing that $(\lambda_1 + \lambda_2)C \neq \lambda_1 C + \lambda_2 C$.

2 Convex functions

3. (12 pts) Let $C \subset \mathbb{R}^n$ be convex and $f : C \to R^\star$. Show that the following statements are equivalent:

   (a) $\text{epi}(f)$ is convex.

   (b) For all points $x_i \in C$ and $\{\lambda_i | \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1, i = 1, 2, \cdots, n\}$, we have
$$f\Big(\sum_{i=1}^n \lambda_i x_i\Big) \leq \sum_{i=1}^n \lambda_i f(x_i).$$

   (c) For $\forall x, y \in C$ and $\lambda \in [0, 1]$,
$$f\Big((1-\lambda)x + \lambda y\Big) \leq (1-\lambda) f(x) + \lambda f(y).$$

   Solution: To see that (a) implies (b) we note that, for all $i = 1, 2, \cdots, n$, $(f(x_i), x_i) \in \text{epi}(f)$. Since this latter set is convex, we have that
$$\sum_{i=1}^n \lambda_i \left(f\left(x_i\right), x_i\right) = \left(\sum_{i=1}^n \lambda_i f\left(x_i\right), \sum_{i=1}^n \lambda_i x_i\right) \in \text{epi}(f),$$
   which, in turn, implies that
$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f\left(x_i\right).$$
   This establishes (b). It is obvious that (b) implies (c). So it remains only to show that (c) implies (a) in order to establish the equivalence. To this end, suppose that $\left(z_1, x_1\right), \left(z_2, x_2\right) \in \text{epi}(f)$ and take $0 \leq \lambda \leq 1$. Then
$$(1-\lambda)\left(z_1, x_1\right) + \lambda\left(z_2, x_2\right) = \left((1-\lambda) z_1 + \lambda z_2, (1-\lambda) x_1 + \lambda x_2\right),$$
   and since $f\left(x_1\right) \leq z_1$, $f\left(x_2\right) \leq z_2$, $(1-\lambda) > 0$, and $\lambda > 0$, so that
$$(1-\lambda) f\left(x_1\right) + \lambda f\left(x_2\right) \leq (1-\lambda) z_1 + \lambda z_2.$$
   Hence, by the assumption (c), $f\left((1-\lambda) x_1 + \lambda x_2\right) \leq (1-\lambda) z_1 + \lambda z_2$, which shows this latter point is in $\text{epi}(f)$.

4. (18 pts) Please determine whether the following functions are convex, concave or none of those, and give a detailed explanation for your choice.

   1)
$$f_1(x_1, x_2, \cdots, x_n) = \begin{cases} -(x_1 x_2 \cdots x_n)^{\frac{1}{n}}, & \text{if } x_1, \cdots, x_n > 0 \\ \infty & \text{otherwise;} \end{cases}$$

   2) $f_2(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$, on $\mathbb{R}^2_{++}$;

   3) $f_3(x, u, v) = -\log(uv - x^T x)$ on $\textbf{dom} f = \{(x, u, v) | uv > x^T x, \ \ u, v > 0\}$.

Solution:

1) Denote $X = \mathrm{dom}\,(f_1)$. It can be seen that f_1 is twice continuously differentiable over X and its Hessian matrix is given by

$$\nabla^2 f_1(x) = \frac{f_1(x)}{n^2} \begin{bmatrix} \frac{1-n}{x_1^2} & \frac{1}{x_1 x_2} & \cdots & \frac{1}{x_1 x_n} \\ \frac{1}{x_2 x_1} & \frac{1-n}{x_2^2} & \cdots & \frac{1}{x_2 x_n} \\ & & \vdots & \\ \frac{1}{x_n x_1} & \frac{1}{x_1 x_2} & \cdots & \frac{1-n}{x_n^2} \end{bmatrix}$$

for all $x = (x_1, \ldots, x_n) \in X$. From this, direct computation shows that for all $z = (z_1, \ldots, z_n) \in \mathbb{R}^n$ and $x = (x_1, \ldots, x_n) \in X$, we have

$$z' \nabla^2 f_1(x) z = \frac{f_1(x)}{n^2} \left(\left(\sum_{i=1}^{n} \frac{z_i}{x_i} \right)^2 - n \sum_{i=1}^{n} \left(\frac{z_i}{x_i} \right)^2 \right)$$

Note that this quadratic form is nonnegative for all $z \in \mathbb{R}^n$ and $x \in X$, since $f_1(x) < 0$, and for any real numbers $\alpha_1, \ldots, \alpha_n$, we have

$$(\alpha_1 + \cdots + \alpha_n)^2 \leq n \left(\alpha_1^2 + \cdots + \alpha_n^2 \right)$$

in view of the fact that $2\alpha_j \alpha_k \leq \alpha_j^2 + \alpha_k^2$. Hence, $\nabla^2 f_1(x)$ is positive semidefinite for all $x \in X$, and it follows from the second-order conditions that f_1 is convex.

2) The Hessian is

$$\begin{aligned} \nabla^2 f_2(x) &= \begin{bmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & (1-\alpha)(-\alpha) x_1^{\alpha} x_2^{-\alpha-1} \end{bmatrix} \\ &= \alpha(1-\alpha) x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} -1/x_1^2 & 1/x_1 x_2 \\ 1/x_1 x_2 & -1/x_2^2 \end{bmatrix} \\ &= -\alpha(1-\alpha) x_1^{\alpha} x_2^{1-\alpha} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix}^T \\ &\preceq 0. \end{aligned}$$

Hence, f_2 is concave.

3) We can express f_3 as

$$f_3(x, u, v) = -\log u - \log \left(v - x^T x / u \right).$$

The first term is convex. The function $v - x^T x/u$ is concave because v is linear and $x^T x/u$ is convex on $\{(x, u) \mid u > 0\}$. Therefore the second term in f_3 is convex: it is the composition of a convex decreasing function $-\log t$ and a concave function. Therefore f_3 is convex.

3 Convex optimization problems

5. (15 pts) **Robust quadratic programming**. In the lecture, we have learned about robust linear programming as an application of second-order cone programming. Now we will consider a similar robust variation of the convex quadratic program

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax \preceq b. \end{array}$$

For simplicity, we assume that only the matrix P is subject to errors, and the other parameters (q, r, A, b) are exactly known. The robust quadratic program is defined as

$$\begin{array}{ll} \text{minimize} & \sup_{P \in \mathcal{E}} \left((1/2)x^T P x + q^T x + r\right) \\ \text{subject to} & Ax \prec b \end{array}$$

where $\mathcal{E}$ is the set of possible matrices P.

For each of the following sets $\mathcal{E}$, express the robust QP as a convex problem in a standard form (e.g., QP, QCQP, SOCP, SDP).

(a) A finite set of matrices: $\mathcal{E} = \{P_1, ..., P_K\}$, where $P_i \in S_+^n, i = 1, ..., K$.

(b) A set specified by a nominal value $P_0 \in S_+^n$ plus a bound on the eigenvalues of the deviation $P - P_0$:

$$\mathcal{E} = \{P \in \mathbf{S}^n \mid -\gamma I \preceq P - P_0 \preceq \gamma I\}$$

where $\gamma \in \mathbf{R}$ and $P_0 \in \mathbf{S}_+^n$.

(c) An ellipsoid of matrices:

$$\mathcal{E} = \left\{ P_0 + \sum_{i=1}^{K} P_i u_i \mid \|u\|_2 \le 1 \right\}.$$

You can assume $P_i \in \mathbf{S}_+^n, i = 0, \ldots, K$.

Solution:

(a) The objective function is a maximum of convex function, hence convex.
We can write the problem as

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & (1/2)x^T P_i x + q^T x + r \le t, \quad i = 1, \ldots, K \\ & Ax \preceq b, \end{array}$$

which is a QCQP in the variable x and t.

(b) For given x, the supremum of $x^T \Delta P x$ over $-\gamma I \le \Delta P \le \gamma I$ is given by

$$\sup_{-\gamma I \preceq \Delta P \preceq \gamma I} x^T \Delta P x = \gamma x^T x$$

Therefore we can express the robust QP as

$$\begin{array}{ll} \text{minimize} & (1/2)x^T \left(P_0 + \gamma I\right) x + q^T x + r \\ \text{subject to} & Ax \preceq b \end{array}$$

which is a QP.

(c) For given x, the quadratic objective function is

$$\begin{aligned}&(1/2)\left(x^T P_0 x + \sup_{\|u\|_2 \leq 1} \sum_{i=1}^K u_i \left(x^T P_i x\right)\right) + q^T x + r \\ &= (1/2) x^T P_0 x + (1/2)\left(\sum_{i=1}^K \left(x^T P_i x\right)^2\right)^{1/2} + q^T x + r.\end{aligned}$$

This is a convex function of x : each of the functions $x^T P_i x$ is convex since $P_i \succeq 0$. The second term is a composition $h\left(g_1(x), \ldots, g_K(x)\right)$ of $h(y) = \|y\|_2$ with $g_i(x) = x^T P_i x$. The functions g_i are convex and nonnegative. The function h is convex and, for $y \in \mathbf{R}_+^K$, nondecreasing in each of its arguments. Therefore the composition is convex. The resulting problem can be expressed as

$$\begin{array}{ll}\text{minimize} & (1/2) x^T P_0 x + \|y\|_2 + q^T x + r \\ \text{subject to} & (1/2) x^T P_i x \leq y_i, \quad i = 1, \ldots, K \\ & Ax \preceq b\end{array}$$

which can be further reduced to a SOCP

$$\begin{array}{ll}\text{minimize} & u + t \\ \text{subject to} & \left\|\left[\begin{array}{c} P_0^{1/2} x \\ 2u - 2q^T x - 1/4 \end{array}\right]\right\|_2 \leq 2u - 2q^T x + 1/4 \\ & \left\|\left[\begin{array}{c} P_i^{1/2} x \\ 2y_i - 1/4 \end{array}\right]\right\|_2 \leq 2y_i + 1/4, \quad i = 1, \ldots, K \\ & \|y\|_2 \leq t \\ & Ax \preceq b.\end{array}$$

The variables are x, u, t , and $y \in \mathbf{R}^K$. Note that if we square both sides of the first inequality, we obtain

$$x^T P_0 x + \left(2u - 2q^T x - 1/4\right)^2 \leq \left(2u - 2q^T x + 1/4\right)^2,$$

i.e., $x^T P_0 x + 2q^T x \leq 2u$. Similarly, the other constraints are equivalent to $(1/2) x^T P_i x \leq y_i$.

6. (25 pts) **Group-SVM Estimator**. SVM (Support Vector Machine) is a well-known classification model in the field of machine learning. The key idea of it is to find a hyperplane that can maximize the margin between points from different classes (a.k.a. support vectors). Mathematically, SVM is to solve the following optimization problem:

$$\begin{aligned}\min_{\boldsymbol{w}, b, \xi_i} \ & \frac{1}{2}\|\boldsymbol{w}\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \ & y_i\left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b\right) \geqslant 1 - \xi_i \\ & \xi_i \geqslant 0, i = 1, 2, \ldots, m.\end{aligned}$$

From another perspective, $\frac{1}{2}\|\boldsymbol{w}\|^2$ can also be viewed as a regularization term. Perhaps, you have come up with some ideas to replace ℓ_2-norm regularization term with some other regularization term, e.g. ℓ_1-norm, ℓ_∞-norm. That is what we want to discuss in this homework.

Now, a variant of SVM called Group SVM will be introduced to you. Group SVM considers that parameters of SVM should be divided into groups, and the maximum of parameters in the same group should not be too large. Here is the formula of Group SVM.

$$\begin{aligned} \min_{\boldsymbol{w},b,\xi_i} \ & \lambda \sum_{g=1}^{G} \|\boldsymbol{w_g}\|_\infty + \sum_{i=1}^{m} \xi_i \\ \text{s.t.} \ & y_i \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b\right) \geqslant 1 - \xi_i \\ & \xi_i \geqslant 0, i = 1, 2, \ldots, m. \end{aligned}$$

where $\boldsymbol{w} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_G)$. (Note: This does not mean $\boldsymbol{w} \in \mathbb{R}^G$)

(a) Please reformulate the problem as a convex problem.

(b) Please derive the dual form of this problem.

(c) Use CVX to solve the problem that you have derived and plot the result of it (need to upload your code as well). Hint: you can generate the dataset via using two-dimensional Gaussian distributions which have different means.

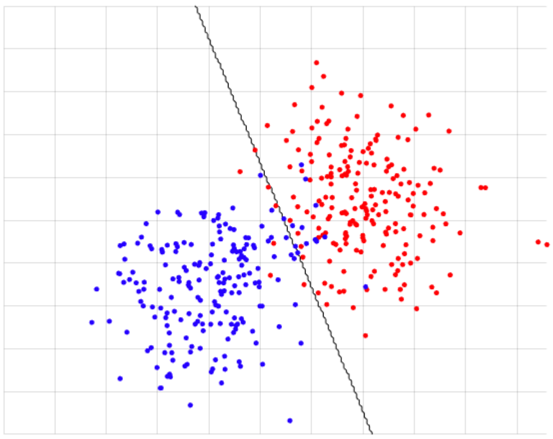

Figure 1: Example

(d) As is known to us, kernel method can be applied to SVM for the extension in nonlinear scenarios. Furthermore, is kernel method available to Group SVM as well? if your answer is yes, please derive the kernelised Group SVM. Otherwise, please illustrate your reason why kernel method cannot be applied to Group SVM.

Solution:

(a) Introduce variables $\boldsymbol{w} = \boldsymbol{w}^+ - \boldsymbol{w}^-, \boldsymbol{w}^+ \geq 0, \boldsymbol{w}^- \geq 0$ and $\boldsymbol{v} \geq 0 \in \mathbb{R}^G$. Then, the equivalent convex problem is

$$
\begin{aligned}
\min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^m, b\in\mathbb{R} \\ \boldsymbol{w}^-, \boldsymbol{w}^-\in\mathbb{R}^p, \boldsymbol{v}\in\mathbb{R}^G}} \quad & \sum_{i=1}^m \xi_i + \lambda \sum_{g=1}^G v_g \\
\text{s.t. } \; & \xi_i + y_i \mathbf{x}_i^T \boldsymbol{w}^+ - y_i \mathbf{x}_i^T \boldsymbol{w}^- + y_i b \geq 1 \quad i \in [m] \\
& v_g - w_j^+ - w_j^- \geq 0 \quad j \in \mathcal{I}_g, g \in [G] \\
& \boldsymbol{\xi} \geq 0, \boldsymbol{w}^+ \geq 0, \boldsymbol{w}^- \geq 0, \boldsymbol{v} \geq 0.
\end{aligned}
$$

where $\mathcal{I}_g$ is the gth subset.

(b) Introduce dual variables $\boldsymbol{\pi} \in \mathbb{R}^m$. Use KKT optimality condition, we can derive the dual form of the primal problem

$$
\begin{aligned}
\max_{\boldsymbol{\pi}\in\mathbb{R}^m} \quad & \sum^m \pi_i \\
\text{s.t.} \quad & \sum_{j\in\mathcal{I}_g} \left| \sum_{i=1}^m y_i x_{ij} \pi_i \right| \leq \lambda \quad g \in [G] \\
& \mathbf{y}^T \boldsymbol{\pi} = 0 \\
& 0 \leq \pi_i \leq 1 \quad i \in [m].
\end{aligned}
$$

(c) See Group_SVM.

(d) No, since the items in Group-SVM that contain x cannot be denoted as an inner product.