SI251 - Convex Optimization, Spring 2022
Homework 1

Due on Mar 27, 2022, before class

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- You are required to write down all the major steps towards making your conclusions; otherwise you may obtain limited points ($\leq 20\%$) of the problem.
- Write your homework in English; otherwise you will get no points of this homework.
- Do your homework by yourself. Any form of plagiarism will lead to 0 point of this homework. If more than one plagiarisms during the semester are identified, we will prosecute all violations to the fullest extent of the university regulations, including but not limited to failing this course, academic probation, or expulsion from the university.
- No late submission will be accepted.
- If you have any doubts regarding the grading, you need to contact the instructor or the TAs within two days since the grade is announced.
- Handwritten assignment is acceptable, but we prefer a LaTex version.

I. **Convex Set**

1. Determine which of the following functions are convex/strictly convex:

(1) $f_1(x) = x^p, x \in (0, \infty)$. (3 points)

(2) $f_2(x) = x^x, x \in (0, \infty)$. (3 points)

(3) $f_3(x) = \tan x, x \in (-\pi/2, \pi/2)$. (3 points)

(4) $f_4(x) = x \log x, x \in (0, \infty)$.(3 points)

(5) $f_5(x) = (1 + \sqrt{x})^{-1}, \quad x \in [0, \infty)$. (This problem does not count points and the typo of domain has been fixed.) (3 points)

Solution.

(1) $f_1''(x) = p(p-1)x^{p-2} > 0$ on $(0, \infty)$, so it is strictly convex when $p > 1$ or $p < 0$, convex at $p = 0, 1$, and strictly concave when $p \in (0, 1)$. (A function is concave (resp. strictly concave) if its negative is convex (resp. strictly convex).)

(2) $f_2''(x) = x^x(1 + \log x)^2 + x^{x-1} > 0$, so it is strictly convex.

(3) $f_3''(x) = 2\sec^2 x \tan x$ is positive on $(0, \pi/2)$ but negative on $(-\pi, 0)$, so it is not strictly convex on $(-\pi/2, \pi/2)$.

(4) $f_4''(x) = 1/x > 0$ on $(0, \infty)$, so it is strictly convex.

(5) $f_5''(x) = \frac{3\sqrt{x}+1}{4(\sqrt{x}+1)^3 x^{3/2}} > 0$ on $[0, \infty)$, so it is strictly convex.

2. Solution set of a quadratic inequality. Let $C \subseteq \mathbb{R}^n$ be the solution set of a quadratic inequality,

$$C = \left\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} + \mathbf{b}^{\mathrm{T}} \mathbf{x} + c \leq 0 \right\}, \tag{1}$$

with $\mathbf{A} \in \mathbb{S}^n, \mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

(1) Show that C is convex if $\mathbf{A} \succeq 0$. (5 points)

(2) Show that the intersection of C and the hyperplane defined by $\mathbf{g}^{\mathrm{T}}\mathbf{x} + h = 0$ (where $\mathbf{g} \neq \mathbf{0}$) is convex if $\mathbf{A} + \lambda \mathbf{g}\mathbf{g}^{\mathrm{T}} \succeq \mathbf{0}$ for some $\lambda \in \mathbb{R}$. (10 points)

Solution.

(1) As we know a set is convex if and only if its intersection with an arbitrary line $\{\hat{\mathbf{x}} + t\mathbf{v} \mid t \in \mathbb{R}\}$ is convex. Insert $\hat{\mathbf{x}} + t\mathbf{v}$ into $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + c \leq 0$

$$\begin{aligned}
&(\hat{\mathbf{x}} + t\mathbf{v})^{\mathrm{T}} \mathbf{A} \hat{\mathbf{x}} + t\mathbf{v} + \mathbf{b}^{\mathrm{T}} \hat{\mathbf{x}} + t\mathbf{v} + c \\
&= \left(\mathbf{v}^{\mathrm{T}} \mathbf{A} \mathbf{v}\right) t^2 + \left(\mathbf{b}^{\mathrm{T}} \mathbf{v} + 2\hat{\mathbf{x}}^{\mathrm{T}} \mathbf{A} \mathbf{v}\right) t + c + \mathbf{b}^{\mathrm{T}} \hat{\mathbf{x}} + \hat{\mathbf{x}}^{\mathrm{T}} \mathbf{A} \hat{\mathbf{x}} \\
&= \alpha t^2 + \beta t + \gamma \leq 0
\end{aligned}$$

The intersection of C and the arbitrary line is the set defined as $\left\{\hat{\mathbf{x}} + t\mathbf{v} \mid \alpha t^2 + \beta t + \gamma \leq 0\right\}$. Since $\mathbf{A} \succeq 0$, then $\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} \geq 0$, so the above set is a continuous and closed line segment which is convex.

(2) The intersection of C and the hyperplane defined by $\mathbf{g}^{\mathrm{T}}\mathbf{x}+h = 0$(where $\mathbf{g} \neq \mathbf{0}$) is $\left\{\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} + c \leq 0 \mid \mathbf{g}^{\mathrm{T}}\mathbf{x} + h = 0\right\}$. Then we take a point $\hat{\mathbf{x}}$ in the above set, then $\mathbf{g}^{\mathrm{T}}\hat{\mathbf{x}} + h = 0$. Insert an arbitrary line $\hat{\mathbf{x}} + t\mathbf{v}$ into above set: $I = \left\{\hat{\mathbf{x}} + t\mathbf{v} \mid \alpha t^2 + \beta t + \gamma \leq 0, \mathbf{g}^{\mathrm{T}}\mathbf{v}t = 0\right\}, \alpha = \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v}, \beta = \mathbf{b}^{\mathrm{T}}\mathbf{v} + 2\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\mathbf{v}, \gamma = c + \mathbf{b}^{\mathrm{T}}\hat{\mathbf{x}} + \hat{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\hat{\mathbf{x}}$. When $\mathbf{g}^{\mathrm{T}}\mathbf{v} \neq 0$, then $t = 0, I = \{\hat{\mathbf{x}} \mid \gamma \leq 0\}$. I is convex whether γ is greater than 0 or not.

When $\mathbf{g}^{\mathrm{T}}\mathbf{v} = 0, I = \left\{\hat{\mathbf{x}} + t\mathbf{v} \mid \alpha t^2 + \beta t + \gamma \leq 0\right\}$. Since $\mathbf{v}^{\mathrm{T}}\left(\mathbf{A} + \lambda \mathbf{g}\mathbf{g}^{\mathrm{T}}\right)\mathbf{v} = \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} + \lambda \mathbf{v}^{\mathrm{T}}\mathbf{g}\mathbf{g}^{\mathrm{T}}\mathbf{v} = \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} = \alpha \succeq 0, I$ is convex. According to the fact that a set is convex if and only if its intersection with an arbitrary line is convex, the intersection of C and the hyperplane defined by $\mathbf{g}^{\mathrm{T}}\mathbf{x} + h = 0$ (where $\mathbf{g} \neq \mathbf{0}$) is convex.

II. **Convex Function**

1. Running average of a convex function. Suppose $f : \mathbf{R} \to \mathbf{R}$ is convex, with $\mathbf{R}_+ \subseteq \operatorname{dom} f$. Show that its running average F, defined as

$$F(x) = \frac{1}{x}\int_0^x f(t)dt, \quad \operatorname{dom} F = \mathbf{R}_{++}, \tag{2}$$

is convex. You can assume f is differentiable.(10 points)

Solution.

F is differentiable with

$$F'(x) = -\left(1/x^2\right)\int_0^x f(t)dt + f(x)/x$$

$$\begin{aligned} F''(x) &= \left(2/x^3\right)\int_0^x f(t)dt - 2f(x)/x^2 + f'(x)/x \\ &= \left(2/x^3\right)\int_0^x \left(f(t) - f(x) - f'(x)(t-x)\right)dt \end{aligned}$$

Convexity now follows from the fact that

$$f(t) \geq f(x) + f'(x)(t-x)$$

for all $x, t \in \mathbf{dom}\, f$, which implies $F''(x) \geq 0$.

Here's another (simpler?) proof. For each s, the function $f(sx)$ is convex in x. Therefore

$$\int_0^1 f(sx)ds$$

is a convex function of x. Now we use the variable substitution $t = sx$ to get

$$\int_0^1 f(sx)ds = \frac{1}{x}\int_0^x f(t)dt$$

III. **Convex Problem**

1. (*Linear Programming*) Give an explicit solution of the following LP.(10 points)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b. \end{array} \tag{3}$$

Solution.

We distinguish three possibilities.
-(a) The problem is infeasible ($b \notin \mathcal{R}(A)$). The optimal value is ∞.

-(b) The problem is feasible, and c is orthogonal to the nullspace of A. We can decompose c as

$$c = A^T\lambda + \hat{c}, \quad A\hat{c} = 0.$$

($\hat{c}$ is the component in the nullspace of A; $A^T\lambda$ is orthogonal to the nullspace.) If $\hat{c} = 0$, then on the feasible set the objective function reduces to a constant:

$$c^T x = \lambda^T A x + \hat{c}^T x = \lambda^T b$$

The optimal value is $\lambda^T b$. All feasible solutions are optimal.

-(c) The problem is feasible, and c is not in the range of A^T ($\hat{c} \neq 0$). The problem is unbounded ($p^\star = -\infty$). To verify this, note that $x = x_0 - t\hat{c}$ is feasible for all t; as t goes to infinity, the objective value decreases unboundedly. In summary,

$$p^\star = \begin{cases} +\infty & b \notin \mathcal{R}(A) \\ \lambda^T b & c = A^T\lambda \text{ for some } \lambda \\ -\infty & \text{otherwise.} \end{cases}$$

2. (*Semidefinite Programming*) Consider $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}_0 + x_1\boldsymbol{A}_1 + \cdots + x_n\boldsymbol{A}_n$, where the vector $\boldsymbol{x} \in \mathbb{R}^n$ and the matrix $\boldsymbol{A}_i \in \mathbb{S}^m$, for $i = 0, 1, \cdots, n$. Let $\lambda_1(\boldsymbol{x}) \geq \cdots \geq \lambda_m(\boldsymbol{x})$ denotes the eigenvalues of $\boldsymbol{A}(\boldsymbol{x})$. Equivalently reformulate the following problems as SDPs.

(a) $\min_{\boldsymbol{x}} \quad \lambda_1(\boldsymbol{x}) - \lambda_m(\boldsymbol{x})$. (10 points)

Solution.

$\lambda_1(\boldsymbol{x}) \leq t_1$ if and only if $\boldsymbol{A}(\boldsymbol{x}) \preceq t_1 \boldsymbol{I}$ and $\lambda_m(\boldsymbol{x}) \geq t_2$ if and only if $\boldsymbol{A}(\boldsymbol{x}) \succeq t_2 \boldsymbol{I}$, so we have

$$\begin{array}{ll} \underset{\boldsymbol{x}, t_1, t_2}{\operatorname{minimize}} & t_1 - t_2 \\ \text{subject to} & t_2 \boldsymbol{I} \preceq \boldsymbol{A}(\boldsymbol{x}) \preceq t_1 \boldsymbol{I}. \end{array}$$

(b) $\min_{\boldsymbol{x}} \quad \sum_{i=1}^m |\lambda_i(\boldsymbol{x})|$. (10 points)

Solution.

Method I: Suppose $\boldsymbol{A}(\boldsymbol{x})$ has eigenvalue decompositon $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$. Let $\boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}^+ - \boldsymbol{A}^- = \boldsymbol{Q}^+\boldsymbol{Q}^T - \boldsymbol{Q}^-\boldsymbol{Q}^T \cdot \boldsymbol{\Lambda}$ is divided into two parts: $\boldsymbol{\Lambda}^+$and $\boldsymbol{\Lambda}^- \cdot \lambda_i(\boldsymbol{x}) \geq 0$ are in the $\boldsymbol{\Lambda}^+$and $-\lambda_i(\boldsymbol{x}) \geq 0$ are in the $\boldsymbol{\Lambda}^-$. Thus $\boldsymbol{A}^+ \succeq 0$ and $\boldsymbol{A}^- \succeq 0$. $\sum_{i=1}^m |\lambda_i(\boldsymbol{x})|$ is equivalent to trace $(\boldsymbol{A}^+)$ + trace $(\boldsymbol{A}^-)$

$$\begin{aligned} \underset{\boldsymbol{x}, \boldsymbol{A}^+, \boldsymbol{A}^-}{\operatorname{minimize}} & \operatorname{trace}\left(\boldsymbol{A}^+\right) + \operatorname{trace}\left(\boldsymbol{A}^-\right) \\ \text{subject to } & \boldsymbol{A}(\boldsymbol{x}) = \boldsymbol{A}^+ - \boldsymbol{A}^- \\ & \boldsymbol{A}^+ \succeq 0 \\ & \boldsymbol{A}^- \succeq 0 \end{aligned}$$

Method II: Similarly to ℓ_1 norm of vector, we have

$$\begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{Y}}{\operatorname{minimize}} & \operatorname{trace}(\boldsymbol{Y}) \\ \text{subject to} & \boldsymbol{Y} + \boldsymbol{A}(\boldsymbol{x}) \succeq 0 \\ & \boldsymbol{Y} - \boldsymbol{A}(\boldsymbol{x}) \succeq 0 \end{array}$$

IV. Lagrange Duality

1. Give the optimal solution of the following convex optimization problem. (15 points)

$$\begin{array}{ll} \operatorname{minimize} & -\sum_{i=1}^n \log\left(\alpha_i + x_i\right) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array} \tag{4}$$

where $\alpha_i > 0$. This problem arises in information theory, in allocating power to a set of n communication channels. The variable x_i represents the transmitter power allocated to the i th channel, and $\log\left(\alpha_i + x_i\right)$ gives the capacity or communication rate of the channel, so the problem is to allocate a total power of one to the channels, in order to maximize the total communication rate.

Solution.

Introducing Lagrange multipliers $\lambda^* \in \mathbf{R}^n$ for the inequality constraints $x^\star \succeq 0$, and a multiplier $\nu^* \in \mathbf{R}$ for the equality constraint $\mathbf{1}^T x = 1$, we obtain the KKT conditions

$$x^\star \succeq 0, \quad 1^T x^\star = 1, \quad \lambda^\star \succeq 0, \quad \lambda_i^\star x_i^\star = 0, \quad i = 1, \ldots, n,$$
$$-1/\left(\alpha_i + x_i^\star\right) - \lambda_i^\star + \nu^\star = 0, \quad i = 1, \ldots, n.$$

We can directly solve these equations to find $x^\star, \lambda^\star$, and $\nu^\star$. We start by noting that $\lambda^\star$ acts as a slack variable in the last equation, so it can be eliminated, leaving

$$x^\star \succeq 0, \quad \mathbf{1}^T x^* = 1, \quad x_i^\star\left(\nu^\star - 1/\left(\alpha_i + x_i^\star\right)\right) = 0, \quad i = 1, \ldots, n,$$
$$\nu^* \geq 1/\left(\alpha_i + x_i^*\right), \quad i = 1, \ldots, n.$$

If $\nu^* < 1/\alpha_i$, this last condition can only hold if $x_i^\star > 0$, which by the third condition implies that $\nu^\star = 1/\left(\alpha_i + x_i^\star\right)$. Solving for $x_i^\star$, we conclude that $x_i^\star = 1/\nu^\star - \alpha_i$ if $\nu^\star < 1/\alpha_i$. If $\nu^\star \geq 1/\alpha_i$, then $x_i^\star > 0$ is impossible, because it would imply $\nu^* \geq 1/\alpha_i > 1/\left(\alpha_i + x_i^\star\right)$, which violates the complementary slackness condition. Therefore, $x_i^\star = 0$ if $\nu^* \geq 1/\alpha_i$. Thus we have

$$x_i^\star = \begin{cases} 1/\nu^\star - \alpha_i & \nu^\star < 1/\alpha_i \\ 0 & \nu^\star \geq 1/\alpha_i \end{cases}$$

or, put more simply, $x_i^\star = \max\left\{0, 1/\nu^\star - \alpha_i\right\}$. Substituting this expression for $x_i^\star$ into the condition $\mathbf{1}^T x^* = 1$ we obtain

$$\sum_{i=1}^n \max\left\{0, 1/\nu^\star - \alpha_i\right\} = 1$$

The lefthand side is a piecewise-linear increasing function of $1/\nu^*$, with breakpoints at α_i, so the equation has a unique solution which is readily determined.

V. **Disciplined Convex Program and CVX**

1. Implementation of hard-margin SVM.(15 points)

We know from [1] that for given input data $\boldsymbol{X} = \{\boldsymbol{X}_1, \dots, \boldsymbol{X}_N\}$ and learning objective $\boldsymbol{y} = \{y_1, \dots, y_N\}$, hard-margin SVM is an algorithm for solving maximum-margin hyperplane in linear separable problems with the constraints that the distance between the sample points and the decision boundary is not less than 1. The hard-margin SVM problem can be equivalently reformulated to a quadratic convex problem, i.e.

$$\begin{aligned} &\max_{\boldsymbol{w},b} \frac{2}{\|\boldsymbol{w}\|} && \Longleftrightarrow && \min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \\ &\text{s.t. } y_i\left(\boldsymbol{w}^\top \boldsymbol{X}_i + b\right) \geq 1 && && \text{s.t. } \quad y_i\left(\boldsymbol{w}^\top \boldsymbol{X}_i + b\right) \geq 1, \end{aligned} \tag{5}$$

where $\boldsymbol{w}$ denotes the normal vector of hyperplane and b denotes the intercept.

In this problem, you are required to train a hard-margin SVM to to **classify "Setosa" and "Versicolour"** in the IRIS dataset (The dataset of "Virginica" is not needed for this problem).

Remarks: (Important!)

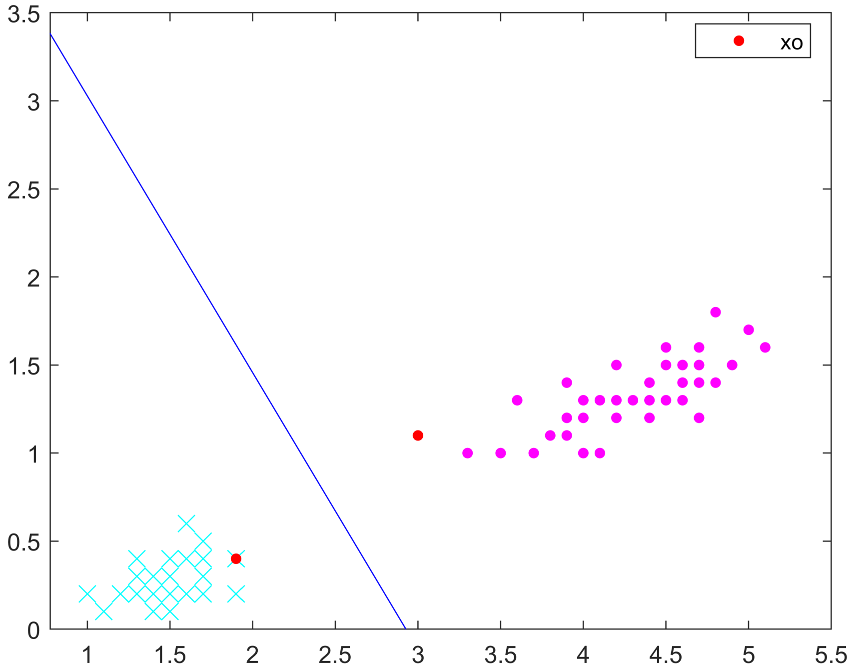
- Make sure that you are adapting the CVX toolbox to optimize the hard-margin SVM.
- Draw the support vector in your simulation results.
- The simulation results need to be presented in this assignment.
- Submit your code.

```
load fisheriris
variety = ~strcmp(species,'virginica');

X = meas(variety,3:4);
y = species(variety);
gscatter(X(:,1),X(:,2),y);

variety = strcmp(y , 'setosa');
y = double(variety);
ind = find(y == 0);
y(ind) = -1;
% train SVM
n = size(X,2);
cvx_begin
    variable w(n);
    variable b;
    minimize( 1/2*norm(w) );
    subject to
        y.*( X * w + b) -1 >= 0;
cvx_end
% find support vector
out = y.*( X * w + b);
out=round(out*100)/100;
ind = find( out == 1);

% draw support vector
figure,gscatter(X(:,1),X(:,2),y,'mc','.x',[15,10]);
hold on
plot([0 -b/(w(1))],[-b/(w(2)) 0],'b')
hold on
for i = 1:length(ind)
    gscatter(X(ind(i),1),X(ind(i),2),'xo');
end
```



References

[1] "Support vector machine." https://en.wikipedia.org/wiki/Support-vector_machine#Hard-margin.