EECS 127/227AT Optimization Models in Engineering
Spring 2020   Homework 5

This homework is due Friday, February 28, 2020 at 23:00 (11pm).
Self grades are due Friday, March 6, 2020 at 23:00 (11pm).

This version was compiled on 2020-02-29 11:38.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

Note*: While the problems on this homework can be completed in any order, they are most cohesive in the order in which they appear. Problems 1–3 all explore the same linear system $A\vec{x} = \vec{y}$ under different noise models; problem 1 examines the role of A's condition number in determining the system's sensitivity to different perturbations, and problems 2–3 show two different ways of using ridge regression to accommodate different models of noise. (You've already seen several ways of interpreting ridge regression in lecture; here are even more!) Problems 4–6 explore convexity and are unrelated to 1–3 or to each other.*

1. Condition number

In Lecture 8, we examined the sensitivity of solutions to linear system $A\vec{x} = \vec{y}$ (for nonsingular/invertible square matrix A) to perturbations in our measurements $\vec{y}$. Specifically, we showed that if we model measurement noise $\vec{\delta y}$ as a linear perturbation on $\vec{y}$, resulting in a linear perturbation $\vec{\delta x}$ on $\vec{x}$ — i.e., $A(\vec{x} + \vec{\delta x}) = \vec{y} + \vec{\delta y}$ — we can bound the magnitude of the solution perturbations $\vec{\delta x}$ as

$$\frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2} \le \kappa(A)\frac{\|\vec{\delta y}\|_2}{\|\vec{y}\|_2}, \tag{1}$$

where $\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$ is the condition number of A, or the ratio of A's maximum and minimum singular values. In this problem, we will establish a similar bound for perturbations on A.

(a) Consider the linear system $A\vec{x} = \vec{y}$ above, where $A \in \mathbb{R}^{n \times n}$ is invertible (i.e., square and nonsingular). Let $\delta A \in \mathbb{R}^{n \times n}$ denote a linear perturbation on matrix A generating a corresponding linear perturbation $\vec{\delta x}$ in solution $\vec{x}$, i.e.,

$$(A + \delta A)(\vec{x} + \vec{\delta x}) = \vec{y}.$$

Show that

$$\frac{\|\vec{\delta x}\|_2}{\|\vec{x} + \vec{\delta x}\|_2} \le \kappa(A)\frac{\|\delta A\|_2}{\|A\|_2}. \tag{2}$$

Solution: Rearranging the given linear system equation, we have

$$(A + \delta A)(\vec{x} + \vec{\delta x}) = y \Rightarrow A\vec{x} + A\vec{\delta x} + \delta A\vec{x} + \delta A\vec{\delta x} = \vec{y}$$

$$\Rightarrow A\vec{\delta x} + \delta A\vec{x} + \delta A\vec{\delta x} = 0 \qquad (A\vec{x} = \vec{y})$$

$$\Rightarrow A\vec{\delta x} = -\delta A(\vec{x} + \vec{\delta x})$$

$$\Rightarrow \vec{\delta x} = -A^{-1}\delta A(\vec{x} + \vec{\delta x})$$

$$\Rightarrow \|\vec{\delta x}\|_2 = \|A^{-1}\delta A(\vec{x} + \vec{\delta x})\|_2 \leq \|A^{-1}\|_2\|\delta A\|_2\|\vec{x} + \vec{\delta x}\|_2$$

$$\Rightarrow \|\vec{\delta x}\|_2 \leq \|A^{-1}\|_2\|\delta A\|_2\|\vec{x} + \vec{\delta x}\|_2 \frac{\|A\|_2}{\|A\|_2}$$

$$\Rightarrow \frac{\|\vec{\delta x}\|_2}{\|\vec{x} + \vec{\delta x}\|_2} \leq \|A\|_2\|A^{-1}\|_2 \frac{\|\delta A\|_2}{\|A\|_2}$$

$$\Rightarrow \frac{\|\vec{\delta x}\|_2}{\|\vec{x} + \vec{\delta x}\|_2} \leq \sigma_{\max}(A)\frac{1}{\sigma_{\min}(A)}\frac{\|\delta A\|_2}{\|A\|_2}$$

$$\Rightarrow \frac{\|\vec{\delta x}\|_2}{\|\vec{x} + \vec{\delta x}\|_2} \leq \kappa(A)\frac{\|\delta A\|_2}{\|A\|_2}$$

as desired.

(b) (Practice: This subpart will not be graded, but it is helpful for understanding.) Note that Eq. 1 and 2 above bound two slightly different quantities: $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2}$ and $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}+\vec{\delta x}\|_2}$, respectively. In general, we wish to establish these bounds because we want to characterize the size of $\vec{\delta x}$ under different sizes of perturbation. Which of these two bounds better serves this purpose? *Hint:* Consider different relative values of $\vec{x}$ and $\vec{\delta x}$. What happens to the bounds when $\vec{\delta x} \gg \vec{x}$?

Solution: When $\vec{\delta x}$ is small relative to $\vec{x}$, both bounds are almost equivalent, since $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}+\vec{\delta x}\|_2} \sim \frac{\|\vec{\delta x}\|_2}{\|\vec{x}\|_2}$ for small $\vec{\delta x}$. However, when $\vec{\delta x}$ is very large relative to $\vec{x}$, $\frac{\|\vec{\delta x}\|_2}{\|\vec{x}+\vec{\delta x}\|_2} \sim \frac{\|\vec{\delta x}\|_2}{\|\vec{\delta x}\|_2} = 1$ regardless of the value of $\vec{\delta x}$, so our bound in Eq. 2 tells us nothing about $\vec{\delta x}$'s size. Our bound on solution error for perturbations in $\vec{y}$ in Eq. 1 is therefore much more useful for characterizing $\vec{\delta x}$ over a wider range of perturbations than our bound on solution error for perturbations in A in Eq. 2.

2. Ridge regression for bounded output noise (1/2)

We will first solve the ridge regression problem in the case where our output measurements $\vec{y}$ are noisy and we have some bounds on this noise, as well as some specific knowledge about data matrix A.

Let square matrix $A \in \mathbb{R}^{n \times n}$ have the singular value decomposition $A = U\Sigma V^\top$, and let its smallest singular value be $\sigma_{\min}(A) > 0$.

(a) Is A invertible? If so, write the singular value decomposition of A^{-1}.

Solution: Since A is a square matrix and all of its singular values are positive, it is invertible, and the SVD of this inverse is

$$A^{-1} = V\Sigma^{-1}U^\top.$$

(b) Consider the linear equation $A\vec{x} = \vec{y}$, where $\vec{y} \in \mathbb{R}^n$ is a noisy measurement satisfying

$$\|\vec{\tilde{y}} - \vec{y}\|_2 \leq r$$

for some vector $\vec{y} \in \mathbb{R}^n$ and $r > 0$. Let $\vec{x}^*(\vec{y})$ denote the solution of $A\vec{x} = \vec{y}$. Show that

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2 \leq r} \|\vec{x}^*(\vec{\tilde{y}}) - \vec{x}^*(\vec{y})\|_2 = \frac{r}{\sigma_{\min}(A)}$$

Solution: Since A is invertible, for any $\vec{y} \in \mathbb{R}^n$, we have

$$\vec{x}^*(\vec{y}) = A^{-1}\vec{y} = V\Sigma^{-1}U^\top\vec{y}.$$

Note that

$$\{\vec{\tilde{y}} \in \mathbb{R}^n : \|\vec{\tilde{y}} - \vec{y}\|_2 \leq r\} = \{\vec{y} + \vec{u} : \vec{u} \in \mathbb{R}^n, \|\vec{u}\|_2 \leq r\},$$

and therefore,

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2 \leq r} \|\vec{x}^*(\vec{\tilde{y}}) - \vec{x}^*(\vec{y})\| = \max_{\vec{u}:\|\vec{u}\|_2 \leq r} \|\vec{x}^*(\vec{y} + \vec{u}) - \vec{x}^*(\vec{y})\|_2.$$

Since we can write the differences between the estimates as

$$\vec{x}^*(\vec{y} + \vec{u}) - \vec{x}^*(\vec{y}) = A^{-1}(\vec{y} + \vec{u}) - A^{-1}\vec{y} = A^{-1}\vec{u} = V\Sigma^{-1}U^\top\vec{u},$$

we obtain

$$\max_{\vec{u}:\|\vec{u}\|_2 \leq r} \|\vec{x}^*(\vec{y} + \vec{u}) - \vec{x}^*(\vec{y})\|_2 = \max_{\vec{u}:\|\vec{u}\|_2 \leq r} \|V\Sigma^{-1}U^\top\vec{u}\|_2$$

$$= \max_{\vec{u}:\|\vec{u}\|_2 \leq r} \|\Sigma^{-1}\vec{u}\|_2$$

where the last equality follows from the fact that U and V are orthonormal matrices. The matrix Σ^{-1} is a diagonal matrix of entries are the inverse of those in Σ, and thus

$$\max_{\vec{u}:\|\vec{u}\|_2 \leq r} \|\Sigma^{-1}\vec{u}\|_2 = r\sigma_{\max}(\Sigma^{-1}) = \frac{r}{\sigma_{\min}(\Sigma)} = \frac{r}{\sigma_{\min}(A)}$$

as desired.

(c) What happens if the smallest singular value of A is very close to zero? Why is this problematic for finding our solution vector $\vec{x}^*$?

Solution: In part (b), we showed that a perturbation of magnitude r on the measurement can change our estimate $\vec{x}^*$ by up to $\frac{r}{\sigma_{\min}(A)}$. If $\sigma_{\min}(A)$ is very small, the estimate can change by a large amount even if the measurements are only slightly perturbed. We say in this instance that our solution is very "sensitive" to perturbations in $\vec{y}$.

(d) Now assume that we find optimal value $\vec{x}^*$ via ridge regression, i.e., we compute

$$\vec{x}_\lambda^*(\vec{\tilde{y}}) = \arg\min_{\vec{x}} \|A\vec{x} - \vec{\tilde{y}}\|_2^2 + \lambda\|\vec{x}\|_2^2$$

for some chosen value $\lambda \in \mathbb{R}$. Compute $\vec{x}_\lambda^*(\vec{\tilde{y}})$, our optimal solution vector (now parameterized by λ), by solving this optimization problem. You may use the solution from class/the last HW for this part.

Solution: By setting the gradient of $\langle A\vec{x} - \vec{y}, A\vec{x} - \vec{y}\rangle + \lambda\langle\vec{x}, \vec{x}\rangle$ to zero, we obtain:

$$2A(A\vec{x} - \vec{y}) + 2\lambda\vec{x} = 2\left[(A^\top A + \lambda I)\vec{x} - A^\top\vec{y}\right] = 0,$$

and since this is the point at which $\vec{x} = \vec{x}_\lambda^*(\vec{y})$, we have

$$\vec{x}_\lambda^*(\vec{y}) = (A^\top A + \lambda I)^{-1}A^\top\vec{y}.$$

Alternatively, we can note that

$$\|A\vec{x} - \vec{y}\|_2^2 + \lambda\|\vec{x}\|_2^2 = \left\|\begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix}\vec{x} - \begin{bmatrix} \vec{y} \\ \vec{0} \end{bmatrix}\right\|_2^2,$$

then using the traditional least squares solution, we have

$$\vec{x}_\lambda^*(\vec{y}) = \left(\begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix}^\top \begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix}\right)^{-1}\begin{bmatrix} A \\ \sqrt{\lambda}I \end{bmatrix}^\top\begin{bmatrix} \vec{y} \\ \vec{0} \end{bmatrix} = (A^\top A + \lambda I)^{-1}A^\top\vec{y}.$$

(e) Show that for all $\lambda > 0$,

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2\leq r}\|\vec{x}_\lambda^*(\vec{\tilde{y}}) - \vec{x}_\lambda^*(\vec{y})\|_2 \leq \frac{r}{2\sqrt{\lambda}}.$$

How does the value of λ affect the sensitivity of your solution $\vec{x}_\lambda^*(\vec{y})$ to noise in $\vec{y}$?

Hint: For every $\lambda > 0$, we have

$$\max_{\sigma>0}\frac{\sigma}{\sigma^2 + \lambda} = \frac{1}{2\sqrt{\lambda}}.$$

(You need not show this; this optimization can be solved by setting the derivative of the objective function to 0 and solving for σ.)

Solution:

Using our solution from part (d), we have that

$$\vec{x}_\lambda^*(\vec{\tilde{y}}) - \vec{x}_\lambda^*(\vec{y}) = (A^\top A + \lambda I)^{-1}A^\top(\vec{\tilde{y}} - \vec{y})$$

and thus

$$\begin{aligned}\|\vec{x}_\lambda^*(\vec{\tilde{y}}) - \vec{x}_\lambda^*(\vec{y})\|_2 &= \|(A^\top A + \lambda I)^{-1}A^\top(\vec{\tilde{y}} - \vec{y})\|_2 \\ &\leq \|(A^\top A + \lambda I)^{-1}A^\top\|_2\|(\vec{\tilde{y}} - \vec{y})\|_2 \\ &= \sigma_{\max}\left((A^\top A + \lambda I)^{-1}A^\top\right)\|(\vec{\tilde{y}} - \vec{y})\|_2.\end{aligned}$$

This value is maximized when $\|\vec{\tilde{y}} - \vec{y}\|_2 = r$, so we can write new bound

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}}-\vec{y}\|_2\leq r}\|\vec{x}_\lambda^*(\vec{\tilde{y}}) - \vec{x}_\lambda^*(\vec{y})\|_2 \leq \sigma_{\max}\left((A^\top A + \lambda I)^{-1}A^\top\right)r.$$

To simplify this bound further, we need to compute the largest singular value of $(A^\top A + \lambda I)^{-1}A^\top$. Plugging in our decomposition $A = U\Sigma V^\top$, we first compute

$$A^\top A + \lambda I = V\Sigma U^\top U\Sigma V^\top + \lambda I = V\Sigma^2 V^\top + \lambda I = V\Sigma^2 V^\top + \lambda VV^\top = V(\Sigma^2 + \lambda I)V^\top,$$

which leads to

$$(A^\top A + \lambda I)^{-1} A^\top = V(\Sigma^2 + \lambda I)^{-1} V^\top V \Sigma U^\top = V(\Sigma^2 + \lambda I)^{-1} \Sigma U^\top.$$

This is the SVD decomposition of $(A^\top A + \lambda I)^{-1} A^\top$, and its singular values are the diagonal elements of $(\Sigma^2 + \lambda I)^{-1}\Sigma$. Given the singular values of A, $\{\sigma_i\}_{i=1}^n$, the singular values of $(\Sigma^2 + \lambda I)^{-1}\Sigma$ are

$$\left\{ \frac{\sigma_i}{\sigma_i^2 + \lambda} \right\}_{i=1}^n,$$

all of which we know are smaller than $\frac{1}{2\sqrt{\lambda}}$ from the given hint. We can then rewrite our bound above as

$$\max_{\vec{\tilde{y}}:\|\vec{\tilde{y}} - \vec{y}\|_2 \leq r} \|\vec{x}_\lambda^*(\vec{\tilde{y}}) - \vec{x}_\lambda^*(\vec{y})\|_2 \leq \sigma_{\max}\left((A^\top A + \lambda I)^{-1} A^\top \right) r \leq \frac{r}{2\sqrt{\lambda}}$$

as desired.

The larger we choose our λ, the tighter our bound on the deviation of our noisy solution $\vec{x}_\lambda^*(\vec{\tilde{y}})$ from our true solution $\vec{x}_\lambda^*(\vec{y})$. In other words, if the regularization parameter λ is large enough, **small perturbations in the measurement cannot change the estimate by a large amount**.

3. **Ridge regression for data matrix noise (2/2)**

Next, we will solve the ridge regression problem in the case where our data matrix A is noisy and we know some properties of this noise.

Consider the standard least-squares problem

$$\min_{\vec{x}} \ \|A\vec{x} - \vec{y}\|_2^2,$$

in which the data matrix $A \in \mathbb{R}^{m \times n}$ is noisy. We model this noise by assuming that each row $\vec{a}_i^\top \in \mathbb{R}^n$ has the form $\vec{a}_i = \vec{\hat{a}}_i + \vec{u}_i$, where the noise vector $\vec{u}_i \in \mathbb{R}^n$ has zero mean and covariance matrix $\sigma^2 I_n$, with $\sigma \in \mathbb{R}$ a measure of the size of the noise. Therefore, now the matrix A is a function of the random variable $U = (\vec{u}_1, \ldots, \vec{u}_m)$, which we denote by A_U. We will use $\hat{A}$ to denote the matrix with rows $\vec{\hat{a}}_i^\top$, $i = 1, \ldots, m$. To account for this noise, we replace the standard least squares formulation above with

$$\min_{\vec{x}} \ \mathbb{E}_U\{\|A_U\vec{x} - \vec{y}\|_2^2\},$$

where $\mathbb{E}_U$ denotes the expected value with respect to the random variable U. Show that this problem can be written as

$$\min_{\vec{x}} \ \|\hat{A}\vec{x} - \vec{y}\|_2^2 + \lambda\|\vec{x}\|_2^2,$$

where $\lambda \geq 0$ is some regularization parameter, which you will determine. In other words, show that regularized least-squares can be interpreted as a way to take into account uncertainties in the matrix A, in the expected value sense.

Hint 1: Compute the expected value of $((\vec{\hat{a}}_i + \vec{u}_i)^\top \vec{x} - y_i)^2$, for a specific row index i.

Hint 2: Trace trick: We can use the fact that trace of a scalar is equal to the scalar and write,

$$\mathbb{E}_U\{\vec{u}_i^\top B \vec{u}_i\} = \mathbb{E}_U\{\text{trace}(\vec{u}_i^\top B \vec{u}_i)\}$$
$$= \mathbb{E}_U\{\text{trace}(\vec{u}_i \vec{u}_i^\top B)\},$$

for any matrix B.

Solution: Considering a fixed row index i, define $r_i \doteq y_i - \vec{a}_i^\top \vec{x}$. We consider the expected value

$$\mathbb{E}_u\{(\vec{a}_i + \vec{u}_i)^\top \vec{x} - y_i\}^2 = \mathbb{E}\left\{(\vec{u}_i^\top \vec{x})^2 - 2r_i(\vec{u}_i^\top \vec{x}) + r_i^2\right\}$$
$$= \mathbb{E}\left\{(\vec{u}_i^\top \vec{x})^2\right\} + r_i^2$$
$$= \sigma^2 \|\vec{x}\|_2^2 + r_i^2,$$

where we have used $\mathbb{E}_u\{u_i\} = 0$ in the second line and the expression

$$\mathbb{E}_u\{\vec{u}_i^\top x\}^2 = \mathbb{E}\{\vec{x}^\top \vec{u}_i \vec{u}_i^\top \vec{x}\} = \mathbb{E}\left\{\text{trace}(\vec{u}_i \vec{u}_i^\top \vec{x} \vec{x}^\top)\right\}$$
$$= \text{trace}\left(\mathbb{E}\left\{\vec{u}_i \vec{u}_i^\top \vec{x} \vec{x}^\top\right\}\right) = \text{trace}\left(\mathbb{E}\left\{\vec{u}_i \vec{u}_i^\top\right\} \vec{x} \vec{x}^\top\right)$$
$$= \text{trace}\left(\sigma^2 I \vec{x} \vec{x}^\top\right) = \sigma^2 \text{trace}(\vec{x} \vec{x}^\top) = \sigma^2 \vec{x}^\top \vec{x}.$$

Summing over $i = 1, \ldots, m$, we obtain,

$$\mathbb{E}_U\{\|A_U \vec{x} - \vec{y}\|_2^2\} = \mathbb{E}\left\{\sum_{u=1}^m ((\vec{a}_i + \vec{u}_i)^\top \vec{x} - y_i)^2\right\}$$
$$= \sum_{i=1}^m \mathbb{E}\left\{((\vec{a}_i + \vec{u}_i)^\top \vec{x} - y_i)^2\right\}$$
$$= \sum_{i=1}^m \left(r_i^2 + \sigma^2 \|\vec{x}\|_2^2\right)$$
$$= \left\|\hat{A}\vec{x} - \vec{y}\right\|_2^2 + \lambda \|\vec{x}\|_2^2,$$

for $\lambda = m\sigma^2$.

4. Visualizing rank-1 matrices

In this problem, we explore the effect of rank constraints on the convexity of matrix sets.

First, consider the set of all 2×2 matrices with diagonal elements $(1, 2)$, which we can write explicitly as

$$S_{2\times 2} = \left\{\begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix} \middle| x, y \in \mathbb{R}\right\}.$$

(a) Is set $S_{2\times 2}$ convex? If so, provide a proof, and if not, provide a counterexample.

Solution: Set $S_{2\times 2}$ is indeed convex. In general, a set S is convex if for all elements $s_1, s_2 \in S$, and $\lambda \in [0, 1]$,

$$\lambda s_1 + (1 - \lambda)s_2 \in S.$$

Considering $s_1 = \begin{bmatrix} 1 & x_1 \\ y_1 & 2 \end{bmatrix}, s_2 = \begin{bmatrix} 1 & x_2 \\ y_2 & 2 \end{bmatrix} \in S_{2\times 2}$, we have

$$\lambda s_1 + (1-\lambda)s_2 = \lambda \begin{bmatrix} 1 & x_1 \\ y_1 & 2 \end{bmatrix} + (1-\lambda) \begin{bmatrix} 1 & x_2 \\ y_2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1\lambda + 1(1-\lambda) & \lambda x_1 + (1-\lambda)x_2 \\ \lambda y_1 + (1-\lambda)y_2 & 2\lambda + 2(1-\lambda) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \lambda x_1 + (1-\lambda)x_2 \\ \lambda y_1 + (1-\lambda)y_2 & 2 \end{bmatrix}$$

$$\in S_{2\times 2}.$$

(b) Suppose we now wish to define $S_{2\times 2}^{(1)} \subset S_{2\times 2}$, the set of all rank-1 matrices in $S_{2\times 2}$. Write out conditions on x and y (i.e. equation constraints that x and y must satisfy) to define $S_{2\times 2}^{(1)}$ explicitly.

Solution: If a matrix $s = \begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix}$ is an element of $S_{2\times 2}^{(1)}$, then its columns must be linearly dependent, i.e.,

$$\begin{bmatrix} x \\ 2 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ y \end{bmatrix}$$

for some $\alpha \in \mathbb{R}$. Computing each row equality, we have $x = \alpha$ and $2 = \alpha y \Rightarrow y = \frac{2}{\alpha}$, and thus

$$S_{2\times 2}^{(1)} = \left\{ \begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix} \Big| x = \alpha, y = \frac{2}{\alpha}, \alpha \in \mathbb{R} \right\}$$

$$= \left\{ \begin{bmatrix} 1 & x \\ y & 2 \end{bmatrix} \Big| x, y \in \mathbb{R}, y = \frac{2}{x} \right\}$$

or alternatively

$$S_{2\times 2}^{(1)} = \left\{ \begin{bmatrix} 1 & \alpha \\ \frac{2}{\alpha} & 2 \end{bmatrix} \Big| \alpha \in \mathbb{R} \right\}.$$

Note that this is equivalent to

$$S_{2\times 2}^{(1)} = \left\{ \begin{bmatrix} 1 & \frac{1}{\beta} \\ 2\beta & 2 \end{bmatrix} \Big| \beta \in \mathbb{R} \right\},$$

which we can see by defining $\beta \doteq \frac{1}{\alpha}$.

(c) Is set $S_{2\times 2}^{(1)}$ convex? If so, provide a proof, and if not, provide a counterexample. Plot the x–y curve described by constraints you found in the earlier part and observe its shape.

Hint: Any linear function applied to a convex set generates another convex set[1], and the function that maps set $S_{2\times 2}^{(1)}$ to variables (x, y) is linear.[2]

Solution: Set $S_{2\times 2}^{(1)}$ is **not** convex. You can observe this by plotting the x–y curve:

---

[1] You are asked to prove this in problem 6(a).

[2] We can show this directly from the definition of linearity: define function $f : S_{2\times 2}^{(1)} \to \mathbb{R}^2$ that maps each set element s to its corresponding off-diagonal values (x, y). Then for any two elements $s_1, s_2 \in S_{2\times 2}^{(1)}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, we have $f(\alpha_1 s_1 + \alpha_2 s_2) = \alpha_1 f(s_1) + \alpha_2 f(s_2)$.
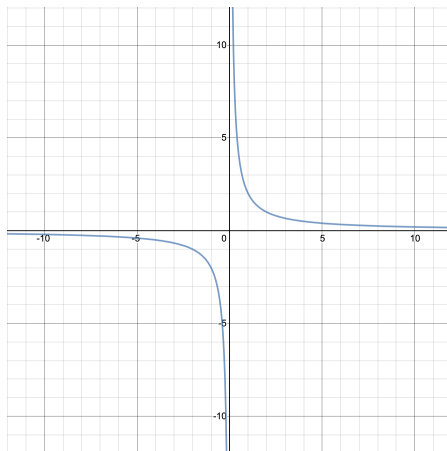
which is a linear projection of $S_{2 \times 2}^{(1)}$ and is not convex.

For one possible counterexample, consider $s_1 = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}, s_2 = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \in S_{2 \times 2}^{(1)}, \lambda = \frac{1}{2}$.
Then

$$\begin{aligned} \lambda s_1 + (1 - \lambda)s_2 &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} + (1 - \frac{1}{2}) \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\frac{1}{2} \\ \frac{1}{2} & 2 \end{bmatrix} \\ &\notin S_{2 \times 2}^{(1)}. \end{aligned}$$

(d) In this class, we will sometimes pose optimization problems in which we optimize over sets of matrices. Since low-dimensional models are often easier to interpret, it would be nice to impose rank constraints on these solution matrices. Suppose we wish to solve the optimization problem

$$\min_{A \in S_{2 \times 2}^{(1)}} \|A\|_F^2$$

which is equivalent to

$$\min_{A \in S_{2 \times 2}} \|A\|_F^2$$
$$\text{s.t.} \quad \mathrm{rk}(A) = 1.$$

Is this optimization problem convex?

Solution: While optimizing $\|A\|_F^2$ over the set of all $A \in \mathbb{R}^{2 \times 2}$ is convex, because $S_{2 \times 2}^{(1)}$ is not convex, this optimization problem is **not** convex because the domain of the function is not convex.

5. Properties of convex functions

In this exercise, we examine convexity and what it represents graphically.

(a) In what region between $[0, 2\pi]$ is $\sin(x)$ a convex function? In what region between $[0, 2\pi]$ is $\sin(x)$ a concave function? Give a region between $[0, 2\pi]$ where $\sin(x)$ is neither convex nor concave.

Solution: The function $\sin(x)$ is convex (in fact, strictly convex) between $[\pi, 2\pi]$; similarly, it is concave (in fact, strictly concave) between $[0, \pi]$. It is non-convex and non-concave for any interval between $[0, 2\pi]$ that is not a subset of the two aforementioned intervals. Note that our interval could even be disjoint!

(b) Plot $\sin(x)$ between $[0, 2\pi]$. For each of the 3 intervals defined above in part (a), draw a chord to illustrate graphically on what regions the function is convex, concave, and neither convex nor concave.

Solution:
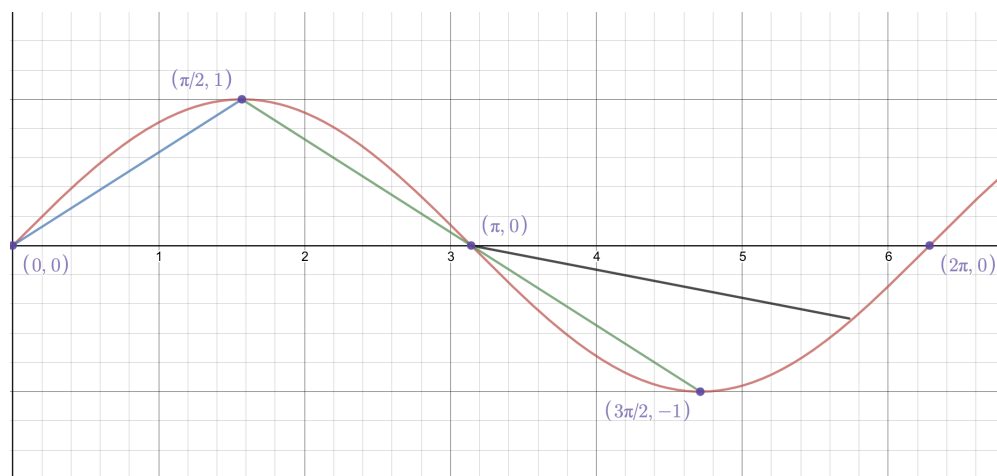
In the region $[0, \pi]$, the function is concave and all chords (e.g., the *blue* chord above) lie below the function. In the region $[\pi, 2\pi]$, the function is convex and all chords (e.g., the *black* chord above) lie above the function. When considering the full region $[0, 2\pi]$, or any region that is not a subset of the two regions above, chords (like the example *green* chord above) do not lie strictly above or strictly below the function.

(c) Show that for all $x \in [0, \frac{\pi}{2}]$,
$$\frac{2}{\pi}x \leq \sin x \leq x.$$

Solution: From part (a), we know that $\sin(x)$ is concave on $[0, \frac{\pi}{2}]$, and thus every value lies below every tangent and above every chord that can be defined in the region.

In the region $[0, \frac{\pi}{2}]$, $\sin(x)$ can therefore be upper bounded by its tangent at 0 (the identity function $f(x) = x$) and lower bounded by the chord between $(0, \sin(0))$ and $(\pi/2, \sin(\pi/2))$ (the linear function $\frac{2}{\pi}x$).

Note that we could establish different upper and lower bounds as well; all values of $\sin(x)$ lie below any tangent line of the function, and values within the span of a chord lie above that chord.

6. **Convexity**

(a) Show the conservation of convexity through affine transformation, i.e., prove that if $S \subseteq \mathbb{R}^n$ is convex, then the image of S under an affine function f,

$$f(S) = \{f(\vec{x}) \mid \vec{x} \in S\},$$

is convex.

Solution: Let $\vec{y}_1, \vec{y}_2 \in f(S)$. This implies there exist $\vec{x}_1, \vec{x}_2 \in S$ such that $\vec{y}_1 = A\vec{x}_1 + \vec{b}$ and $\vec{y}_2 = A\vec{x}_2 + \vec{b}$.
We want to show that $\lambda \vec{y}_1 + (1 - \lambda)\vec{y}_2 \in f(S)$ for $0 \leq \lambda \leq 1$.
Since S is convex we have $\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2 \in S$. Further $A(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2) + \vec{b} = \lambda \vec{y}_1 + (1-\lambda)\vec{y}_2$.
This shows that $\lambda \vec{y}_1 + (1 - \lambda)\vec{y}_2 \in f(S)$.

(b) Show that a function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph, defined as $\text{epi}(f) = \{(\vec{x}, t) \mid \vec{x} \in \text{dom}(f), f(\vec{x}) \leq t\}$, is convex.

Solution: Let's prove the backward direction first. Recall,

$$\text{epi}(f) = \{(\vec{x}, t) \mid \vec{x} \in \text{dom}(f), \ f(\vec{x}) \leq t\}$$

If $\text{epi}(f)$ is a convex set, then for all $(\vec{x}_1, t_1), (\vec{x}_2, t_2) \in \text{epi}(f)$ we have that $\lambda(\vec{x}_1, t_1) + (1 - \lambda)(\vec{x}_2, t_2) \in \text{epi}(f)$ for $\lambda \in [0, 1]$. In particular, we let $t_1 = f(x_1)$ and $t_2 = f(x_2)$. By definition of the epigraph, we have $(\lambda \vec{x}_1 + (1 - \lambda)\vec{x}_2, \lambda f(\vec{x}_1) + (1 - \lambda)f(\vec{x}_2)) \in \text{epi}(f)$.
Therefore, $f(\lambda \vec{x}_1 + (1-\lambda)\vec{x}_2) \leq \lambda f(\vec{x}_1) + (1-\lambda)f(\vec{x}_2)$. Since this holds for all $\vec{x}_1, \vec{x}_2 \in \text{dom}(f)$, we have that f is convex.

To show the forward direction, for $\text{epi}(f)$ to be a convex set, we need to satisfy the definition of convex set S

$$\lambda(\vec{x}_1, t_1) + (1 - \lambda)(\vec{x}_2, t_2) \in S$$

for $\lambda \in [0, 1]$ and $(\vec{x}_1, t_1), (\vec{x}_2, t_2) \in S$. But from convexity of f we have that

$$\begin{aligned} f(\lambda \vec{x}_1 + (1 - \lambda)\vec{x}_2) &\leq \lambda f(\vec{x}_1) + (1 - \lambda)f(\vec{x}_2) \\ &\leq \lambda t_1 + (1 - \lambda)t_2 \end{aligned}$$

which shows $\lambda(\vec{x}_1, t_1) + (1 - \lambda)(\vec{x}_2, t_2) \in \text{epi}(f)$.

7. **Homework process**
   Whom did you work with on this homework? List the names and SIDs of your group members.