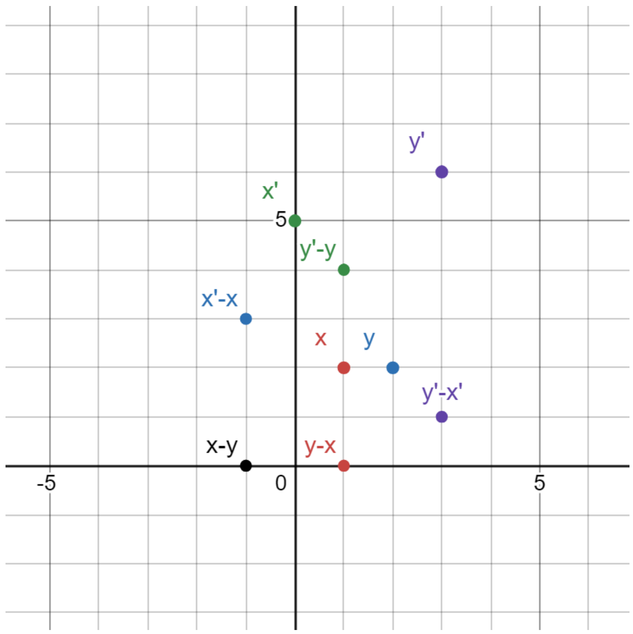
Introduction to Optimization for Computer Science — Assignment 1

Idan Drori
Ohad Sapir

Question #1

(a) *Solution:*



Let $x', y', x, y \in \mathbb{R}^d$ be such that $\langle x' - x, y - x \rangle \leq 0$ and $\langle y' - y, x - y \rangle \leq 0$.
By adding the two inequalities we get:

$$\begin{aligned} 0 &\geq \langle x' - x, y - x \rangle + \langle y' - y, x - y \rangle = \langle x' - x, y - x \rangle - \langle y' - y, y - x \rangle \\ &= \langle x' - x - y' + y, y - x \rangle = \langle x' - y', y - x \rangle + \langle y - x, y - x \rangle \\ &= -\langle y' - x', y - x \rangle + ||y - x||^2 \end{aligned}$$

So we get:

$$\langle y' - x', y - x \rangle \geq ||y - x||^2$$

And from Cauchy-Schwartz inequality:

$$\|y' - x'\| \cdot \|y - x\| \geq \langle y' - x', y - x \rangle \geq \|y - x\|^2$$

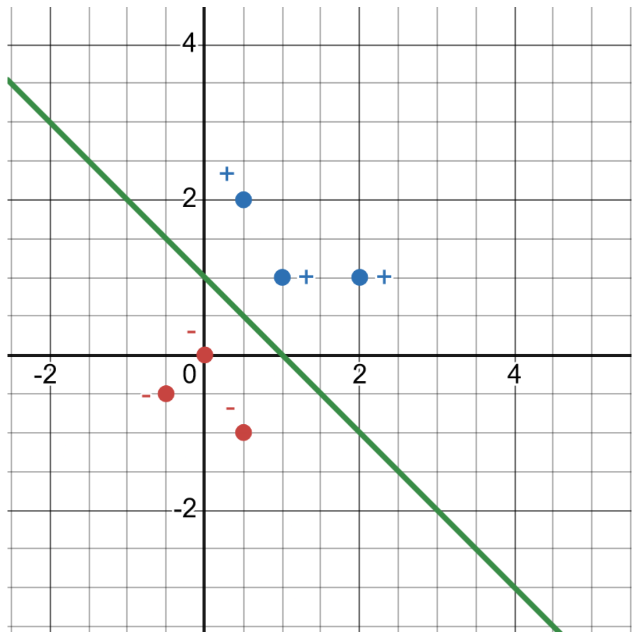
Therefore:

$$\|y' - x'\| \geq \|y - x\|$$

□

(b) *Solution:* We'll take $w = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $b = -1$, then we get

$$H_{w,b} = \{x \in \mathbb{R}^2 \mid x_2 = -x_1 + b\}$$



Let $v, u \in H_{w,b}$

$$\langle w, u - v\rangle = \langle w, u\rangle - \langle w, v\rangle = -b - (-b) = 0$$

Meaning $w \perp u - v$ for any $u, v \in H_{w,b}$ and so $w \perp H_{w,b}$.
Now we'll show that $d(x, H_{w,b}) = \frac{|w^T x + b|}{\|w\|}$. Let $x \in \mathbb{R}^d$. We've seen that to minimize $\|z - x\|$ we need to set z to be the projection of x onto $H_{w,b}$. So $z = x - \lambda w$ for some $\lambda \in \mathbb{R}$. We need to find λ such that $z \in H_{w,b}$.

$$\begin{aligned} 0 = w^T z + b &= w^T(x + \lambda w) + b \\ &= w^T x + \lambda w^T w + b \Rightarrow \lambda\|w\|^2 = -(w^T x + b) \end{aligned}$$

Solving for λ:

$$\lambda = -\frac{w^T x + b}{\|w\|^2}$$

Then we get that:

$$z = x - \frac{w^T x + b}{\|w\|^2} w$$

Calculating the distance $\|z-x\|$:

$$\begin{aligned}\|z-x\| &= \left\|x - \frac{w^Tx+b}{\|w\|^2}w - x\right\| = \left\| - \frac{w^Tx+b}{\|w\|^2}w\right\| \\ &= \left|\frac{w^Tx+b}{\|w\|^2}\right| \cdot \|w\| = \frac{|w^Tx+b|}{\|w\|^2} \cdot \|w\| \\ &= \frac{|w^Tx+b|}{\|w\|}\end{aligned}$$

□

Question #2

Let $A \in \mathbb{R}^{d\times d}$ be a symmetric matrix.

(a) *Solution:*
First, we will prove that A is PD $\iff$ $\lambda_i > 0$ for all eigenvalues $\{\lambda_i\}_{i=1}^d$ of A:
Let $A = UDU^\top$ be the spectral decomposition of A, with $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, and u_i denoting the i'th column of U.
($\implies$) For all $i \in [d]$, we have $\lambda_i = u_i^\top A u_i > 0$, since A is PD.
($\impliedby$) Let $0 \neq x \in \mathbb{R}^d$, and note that:

$$x^\top A x = x^\top U D U^\top x = \sum_{i=1}^{d} \lambda_i (U^\top x)_i^2 = \sum_{i=1}^{d} \lambda_i (u_i^\top x)^2 > 0$$

Now, the proof of the original claim:
$A \succ 0 \iff \lambda_i > 0$ for all eigenvalues $\{\lambda_i\}_{i=1}^d$ of $A \iff \lambda_i \neq 0$ and $\lambda_i \geq 0$ for all eigenvalues $\{\lambda_i\}_{i=1}^d \iff A$ is invertible and $A \succeq 0$
□

(b) *Solution:* Let $x \in \mathbb{R}^d$.

$$x^\top A x = x^\top B^\top B x \ = (Bx)^\top (Bx) = \|Bx\|^2 \geq 0$$

□

(c) *Solution:* Assume by contradiction that $A_{i,i} < 0$ for some i.
Let e_i be the i'th unit vector.

$$e_i^\top A e_i = -A_{i,i} < 0$$

in contradiction to the fact that $A \succeq 0$.
The converse is not true. For example Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, x = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

$$x^\top A x = -1$$

So $A \not\succeq 0$
□

(d) *Solution:* let $\lambda_1 \geq \lambda_2$ be the eigenvalues of A.
$(\implies)$ $A \succeq 0$, so $\lambda_1, \lambda_2 \geq 0$ and:

$$Tr(A) = \lambda_1 + \lambda_2 \geq 0$$

$$Det(A) = \lambda_1 \cdot \lambda_2 \geq 0$$

$(\impliedby)$ Assume by contradiction that $A \not\succeq 0$, meaning $\lambda_2 < 0$:
If $\lambda_1 \neq 0$ than $Tr(A) = \lambda_1 + \lambda_2 < 0$
If $\lambda_1 > 0$ than $Det(A) = \lambda_1 \cdot \lambda_2 < 0$
So $\lambda_1 \geq \lambda_2 \geq 0$ and A must be PSD.
$\square$

(e) *Solution:*
$(\implies)$ Let $A = UDU^\top$ be the spectral decomposition of A, $D = \text{diag}(\lambda_1, \ldots, \lambda_d)$. Set $D^{1/2} := \text{diag}(\sqrt{\lambda}_1, \ldots, \sqrt{\lambda}_d)$. Set $R := UD^{1/2}U^\top$.

$$R^2 = UD^{1/2}U^\top UD^{1/2}U^\top = UD^{1/2}D^{1/2}U^\top = UDU^\top = A$$

$(\impliedby)$

$$x^\top Ax = x^\top RRx \ = (Rx)^\top(Rx) = \|Rx\|^2 \geq 0$$

$\square$

Question #3

Calculate the gradients and Hessians of the following functions $f : \mathbb{R}^d \to \mathbb{R}$

(a) $f(x) = a^T x$
Solution:

$$f(x) = a^T x = \sum_{i=1}^{d} a_i x_i$$

Taking the partial derivative with respect to some x_k:

$$\frac{\partial}{\partial x_k}\left(\sum_{i=1}^{d} a_i x_i\right) = a_k$$

From here we get:

$$\nabla f(x) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} = a$$

Because the partial derivatives don't depend on any x_k, we'll get that for every $k, j = 1, ..d$:

$$\frac{\partial^2 f}{\partial x_k \partial x_j}(x) = 0$$

And so the Hessian matrix will simply be the zero matrix:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & ... & \frac{\partial^2}{\partial x_1 \partial x_d} f(x) \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2}{\partial x_d \partial x_1} f(x) & \frac{\partial^2}{\partial x_d \partial x_2} f(x) & ... & \frac{\partial^2}{\partial x_d \partial x_d} f(x) \end{pmatrix} = \begin{pmatrix} 0 & 0 & ... & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & ... & 0 \end{pmatrix}$$

□

(b) $f(x) = x^T A x$, where $A \in \mathbb{R}^{d \times d}$ is symmetric.
Solution:

$$f(x) = x^T A x = \sum_{j=1}^{d} \sum_{i=1}^{n} a_{ij} x_i x_j$$

And so, according to the product rule we get that the partial derivative with respect to x_k would be:

$$\begin{aligned} \frac{\partial f}{\partial x_k}(x) &= \frac{\partial}{\partial x_k}(\sum_{j=1}^{d} \sum_{i=1}^{d} a_{ij} x_i x_j) \\ &= \frac{\partial}{\partial x_k}(x_1 \sum_{i=1}^{d} a_{i1} x_1 + ... + x_k \sum_{i=1}^{d} a_{ik} x_i + ... + x_d \sum_{i=1}^{d} a_{id} x_i) \\ &= x_1 a_{k1} + ... + (\sum_{i=1}^{d} a_{ik} x_i + x_k a_{kk}) + ... + x_d a_{kd} \\ &= \sum_{j=1}^{d} a_{kj} x_j + \sum_{i=1}^{d} a_{ik} x_i \\ &= A_{k*} x + (A_{*k})^T x = (A_{k*} + (A_{*k})^T) x \end{aligned}$$

Where A_{k*} is the k-th row of A, and A_{*k} is the k-th column of A. Then:

$$\begin{aligned} \nabla f(x) &= \begin{pmatrix} (A_{1*} + (A_{*1})^T) x \\ . \\ . \\ . \\ (A_{d*} + (A_{*d})^T) x \end{pmatrix} \\ &= \begin{pmatrix} A_{1*} + (A_{*1})^T \\ . \\ . \\ . \\ A_{d*} + (A_{*d})^T \end{pmatrix} x \\ &= \left(\begin{pmatrix} A_{1*} \\ . \\ . \\ . \\ A_{d*} \end{pmatrix} + \begin{pmatrix} (A_{*1})^T \\ . \\ . \\ . \\ (A_{*d})^T \end{pmatrix} \right) x \\ &= (A + A^T) x \end{aligned}$$

A is symmetric then $A + A^T = 2A$, and so we get:

$$\nabla f(x) = 2Ax$$

For the Hessian, we saw that the partial derivatives are given by:

$$\frac{\partial f}{\partial x_k}(x) = \sum_{j=1}^{d} a_{kj}x_j + \sum_{i=1}^{d} a_{ik}x_i$$

And so the second partial derivatives are of the form:

$$\frac{\partial^2 f}{\partial x_k \partial x_k'}(x) = a_{k'k} + a_{kk'}$$

Then the Hessian is of the form:

$$\begin{aligned}
\nabla^2 f(x) &= \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x) & \dots & \frac{\partial^2}{\partial x_1 \partial x_d} f(x) \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2}{\partial x_d \partial x_1} f(x) & \frac{\partial^2}{\partial x_d \partial x_2} f(x) & \dots & \frac{\partial^2}{\partial x_d \partial x_d} f(x) \end{pmatrix} \\
&= \begin{pmatrix} a_{11} + a_{11} & a_{21} + a_{12} & \dots & a_{d1} + a_{1d} \\ \vdots & & \ddots & \vdots \\ a_{1d} + a_{d1} & a_{2d} + a_{d2} & \dots & a_{dd} + a_{dd} \end{pmatrix} \\
&= A + A^T
\end{aligned}$$

And because A is symmetric, we get:

$$\nabla^2 f(x) = 2A$$

□

(c) $f(x) = \|x\|^2$.
Solution:

$$f(x) = \|x\|^2 = \sum_{i=1}^{d} x_i^2$$

And so we get that the partial derivative is of the form:

$$\frac{\partial f}{\partial x_k} = 2x_k$$

Then the gradient is:

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_d \end{pmatrix} = 2x$$

The second partial derivative will be:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{cases} 2 & i = j \\ 0 & i \neq j \end{cases}$$

And so we'll get that the Hessian is:

$$\nabla^2 f(x) = 2 \cdot I_d = \begin{pmatrix} 2 & 0 & 0 & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{pmatrix}$$

$\square$

(d) $f(x) = \sum_{i=1}^{d} x_i \ln x_i$, where $f : \mathbb{R}^d_{++} \to \mathbb{R}$, and $\mathbb{R}^d_{++} := \{x \in \mathbb{R}^d \mid x_i > 0 \ \forall i \in [d]\}$
Solution:

$$\frac{\partial f}{\partial x_k} = \ln x_k + x_k \cdot \frac{1}{x_k} = \ln x_k + 1$$

And so the gradient will be:

$$\nabla f(x) = \begin{pmatrix} \ln x_1 + 1 \\ \ln x_2 + 1 \\ \vdots \\ \ln x_d + 1 \end{pmatrix}$$

The second partial derivative will be of the form:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{cases} \frac{1}{x_i} & i = j \\ 0 & i \neq j \end{cases}$$

Then the Hessian will be:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{1}{x_1} & 0 & \dots & 0 \\ 0 & \frac{1}{x_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{x_d} \end{pmatrix}$$

$\square$

(e) $f(x) = \ln\left(\sum_{i=1}^{d} e^{x_i}\right)$. *Solution:*

$$\frac{\partial f}{\partial x_k}(x) = \frac{e^{x_k}}{\sum_{i=1}^{d} e^{x_i}}$$

And the the gradient is:

$$\nabla f(x) = \begin{pmatrix} \frac{e^{x_1}}{\sum_{i=1}^{d} e^{x_i}} \\ \frac{e^{x_2}}{\sum_{i=1}^{d} e^{x_i}} \\ \vdots \\ \frac{e^{x_d}}{\sum_{i=1}^{d} e^{x_i}} \end{pmatrix} = \frac{1}{\sum_{i=1}^{d} e^{x_i}} \begin{pmatrix} e^{x_1} \\ e^{x_2} \\ \vdots \\ e^{x_d} \end{pmatrix}$$

The second derivative is of the form:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \begin{cases} \frac{e^{x_i} \cdot \sum_{k=1}^{d} e^{x_k} - e^{-2x_i}}{(\sum_{k=1}^{d} e^{x_k})^2} & i = j \\ \frac{-e^{x_i + x_j}}{(\sum_{k=1}^{d} e^{x_k})^2} & i \neq j \end{cases} = \frac{e^{x_i}}{(\sum_{k=1}^{d} e^{x_k})^2} \cdot \begin{cases} \sum_{k \neq i} e^k & i = j \\ -e^{x_j} & i \neq j \end{cases}$$

So the Hessian is of the form:

$$\nabla^2 f(x) = \frac{1}{(\sum_{k=1}^{d} e^{x_k})^2} \cdot \begin{pmatrix} e^{x_1} \cdot \sum_{k=2}^{d} e^{x_k} & -e^{x_1+x_2} & -e^{x_1+x_3} & ... & -e^{x_1+x_d} \\ -e^{x_1+x_2} & e^{x_2} \cdot \left(e^{x_1} + \sum_{k=3}^{d} e^{x_k} \right) & & ... & -e^{x_2+x_d} \\ \vdots & & & \ddots & \vdots \\ -e^{x_1+x_d} & & & ... & e^{x_d} \cdot \sum_{k=1}^{d-1} e^{x_k} \end{pmatrix}$$

$\square$