EECS 545 Homework 1 Solution (F21)

1. **Honor Code** (3 pts each)
   - **(a)** True
   - **(b)** True
   - **(c)** False

2. **PSD matrices**
   - **(a)** (3 points)
     For any $\boldsymbol{x}$ we have: $\boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = ||\boldsymbol{A}\boldsymbol{x}||_2^2 \geq 0$
   - **(b)** (3 points)
     For any $\boldsymbol{z}$, define $Y = (\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\boldsymbol{z}$.
     By linearity of Expectation we have:
     $\boldsymbol{z}^T\mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T]\boldsymbol{z} = \mathbb{E}[\boldsymbol{z}^T(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^T\boldsymbol{z}] = \mathbb{E}[Y^2] \geq 0$
   - **(c)** (3 points)
     Denote the Gram matrix as $\mathbf{G}$, and define $\boldsymbol{X}$ as a matrix with $\boldsymbol{x}_i$ as its columns. i.e:

     $$\boldsymbol{X} = \begin{pmatrix} | & | & | & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_d \\ | & | & | & | \end{pmatrix}$$

     Observe then $\mathbf{G} = \boldsymbol{X}^T\boldsymbol{X}$, by (a) it's PSD.

3. **Probability**
   - **(a)** (3 points)
     Beginning from the right hand side, apply the definitions of conditional expectation and expectation:

     $$E[E[X|Y]] = \sum_y \left(\sum_x xP(X = x|Y = y) \right) \cdot P(Y = y)$$

     Next use definition of conditional probability to get joint distribution:

     $$= \sum_y \sum_x xP(X = x, Y = y)$$
     $$= \sum_x x \sum_y P(X = x, Y = y)$$
     $$= \sum_x xP(X = x) = E[X]$$

     Where the last step used the definition of a marginal PDF.

(b) (3 points)
Apply definition of expectation, where the indicator function is the random variable of interest:

$$\begin{aligned} E[I[X \in \mathcal{C}]] &= 1 \cdot P(X \in \mathcal{C}) + 0 \cdot P(X \notin \mathcal{C}) \\ &= P(X \in \mathcal{C}) \end{aligned}$$

(c) (3 points)
Apply definition of expectation:

$$E[XY] = \sum_{x,y} xyP(X = x, Y = y)$$

Then using independence:

$$\begin{aligned} E[XY] &= \sum_{x,y} xyP(X = x)P(Y = y) \\ &= \left(\sum_{x} xP(X = x)\right)\left(\sum_{y} yP(Y = y)\right) \\ &= E[X]E[Y] \end{aligned}$$

(d) (3 points)
For $X, Y \in \{0, 1\}$, note that $E[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(X = 1)$. Similarly, $E[Y] = p(Y = 1)$.
Also observe that $E[XY] = \sum_{x,y \in \{0,1\}} xyP(X = x, Y = y) = P(X = 1, Y = 1)$.
Then $E[XY] = E[X]E[Y]$ implies $P(X = 1, Y = 1) = P(X = 1)P(Y = 1)$.
To finish, take complements on events, showing independence for the remaining three situations, e.g.:

$$\begin{aligned} P(X = 0, Y = 0) &= 1 - P(X = 1 \cup Y = 1) \\ &= 1 - P(X = 1) - P(Y = 1) + P(X = 1, Y = 1) \\ &= 1 - P(X = 1) - P(Y = 1) + P(X = 1)P(Y = 1) \\ &= (1 - P(X = 1))(1 - P(Y = 1)) \\ &= P(X = 0)P(Y = 0) \end{aligned}$$

It is also sufficient to explain that if two events are independent, then their complements are also independent.

4. Gaussian level sets (3 pts each)

(a) Observe $\boldsymbol{U}$ is the rotation matrix that rotates vectors counter-clockwise $\frac{-\pi}{4}$ with respect to the axis (equivalently, $\frac{\pi}{4}$ rotation clockwise), with $\boldsymbol{U}^T$ as its inverse.
Thus $\boldsymbol{\Sigma}^{-1} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^T$.
Define $\boldsymbol{x}' = \boldsymbol{U}^T(\boldsymbol{x} - \mu)$. Then in terms $\boldsymbol{x}'$, we can write:

$$1 = (\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = (\boldsymbol{x}')^T(\boldsymbol{\Lambda})^{-1}\boldsymbol{x}' = \frac{(x_1')^2}{5} + (x_2')^2$$

which is a Ellipse with lengths 1 and $\sqrt{5}$ for the semi-minor and semi-major axis respectively.

Thus $\mathcal{C}$ is simply a Ellipse specified above rotated clockwise $\frac{\pi}{4}$ and translated by vector $\mu = [-2\ 3]^T$

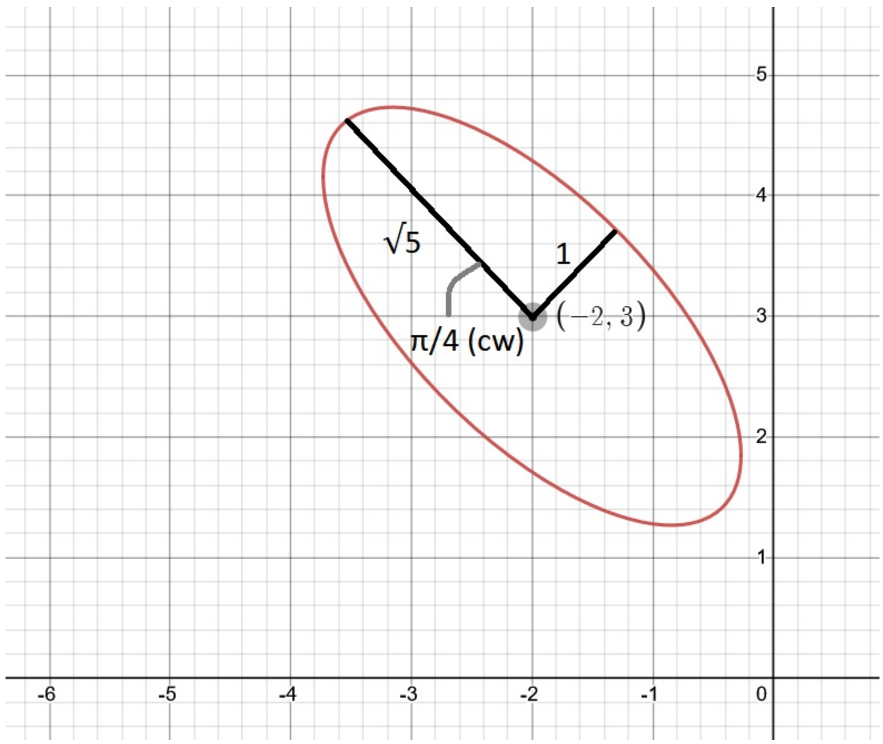

Figure 1: P3(a) plot

(b) [Method 1] Define $\boldsymbol{X'} = \boldsymbol{U}^T(\boldsymbol{X} - \mu)$, then by property[1] of Gaussian distribution $\boldsymbol{X'} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma'})$, where $\boldsymbol{\Sigma'} = \boldsymbol{U}^T \Sigma \boldsymbol{U} = \boldsymbol{\Lambda}$, as $\boldsymbol{\Lambda}$ is diagonal X_1' and X_2' are independent Gaussian random variables with the same mean 0 and variance 5 and 1 respectively

Define $R = (\boldsymbol{X'})^T \boldsymbol{\Lambda}^{-1} \boldsymbol{X'} = \frac{(X_1')^2}{5} + (X_2')^2$, notice that R follows a chi-square distribution with 2 degrees of freedom[2].

Observe $\{\boldsymbol{x} \in \mathcal{C}\} = \{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \leq 1\}$

Thus $\Pr(\boldsymbol{X} \in \mathcal{C}) = \Pr(R \leq 1) = F(1) = 0.3934$ where F denote the cdf of chi-square distribution with 2 degrees of freedom.

[Method 2] You can identify $\boldsymbol{\Sigma}^{-1} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^T$, where $\boldsymbol{\Lambda}^{-1/2} = diag(\frac{1}{\sqrt{5}}, 1)$. Then define $\boldsymbol{X'} = \boldsymbol{A}(\boldsymbol{X} - \mu)'$ where $\boldsymbol{A} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^T$. Invoke the linear transformation property of Gaussian distribution, we can conclude $\boldsymbol{X'} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{I}$ is the identity matrix. Hence, $R = (\boldsymbol{X'})^T \boldsymbol{X'}$ follows a Chi-square distribution with 2 degrees of freedom.

5. Unconstrained Optimization (3 points each)

(a) Let f, g be convex functions.

---

[1] if $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for any matrix $\boldsymbol{A}$, $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{x} - \mathbf{b}) \sim \mathcal{N}(\boldsymbol{A}(\boldsymbol{\mu} - \mathbf{b}), \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$

[2] A random variable Y follows a chi-square distribution if $Y = X_1^2 + X_2^2$ for X_1, X_2 that are independent Gaussian random variables with mean 0 and variance 1

Take two arbitrary points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$ and $t \in (0,1)$, let $\boldsymbol{x} = t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2$, then by convexity of f and g we have:

$$f(\boldsymbol{x}) = f\big(t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2\big) \leq tf(\boldsymbol{x}_1) + (1-t)f(\boldsymbol{x}_2) \tag{1}$$

$$g(\boldsymbol{x}) = g\big(t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_2\big) \leq tg(\boldsymbol{x}_1) + (1-t)g(\boldsymbol{x}_2) \tag{2}$$

sum Eq(1) and Eq(2) we have:

$$\begin{aligned}
(f+g)(\boldsymbol{x}) &= f(\boldsymbol{x}) + g(\boldsymbol{x}) \\
&\leq tf(\boldsymbol{x}_1) + tg(\boldsymbol{x}_1) + (1-t)f(\boldsymbol{x}_2) + (1-t)g(\boldsymbol{x}_2) \\
&= t(f+g)(\boldsymbol{x}_1) + (1-t)(f+g)(\boldsymbol{x}_2)
\end{aligned}$$

(b) Let f, g be convex twice continuously differentiable functions. Then by the linearity of differentiation the Hessian of $f+g$ is given by

$$\nabla^2(f+g) = \nabla^2 f + \nabla^2 g.$$

By Property 7 from the notes on unconstrained optimization, $\nabla^2 f, \nabla^2 g$ are PSD. Since the sum of two PSD matrices is clearly PSD, the Hessian of $f+g$ is PSD so $f+g$ is convex (also by property 7).

(c) Let $\boldsymbol{x} = [x_1, \ldots, x_d]^T$, A_{ij} denote the (i,j)-th entry of matrix $\boldsymbol{A}$, and b_i denote the i-th entry of $\mathbf{b}$ then the quadratic function $f(\boldsymbol{x})$ can be written explicitly as:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \mathbf{b}^T\boldsymbol{x} + c = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} A_{ij}x_i x_j + \sum_{i=1}^{d} b_i x_i + c\,.$$

Applying the definition of the Hessian matrix, the (k,ℓ)-th entry of $\nabla^2 f(x)$ is given by:

$$\begin{aligned}
\big[\nabla^2 f(\boldsymbol{x})\big]_{k,l} &= \frac{\partial^2 f(\boldsymbol{x})}{\partial x_k \partial x_\ell} \\
&= \frac{\partial^2}{\partial x_k \partial x_\ell}\left\{\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} A_{ij}x_i x_j + \sum_{i=1}^{d} b_i x_i + c\right\} \\
&= \frac{\partial}{\partial x_k}\left\{\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} A_{ij}\frac{\partial}{\partial x_\ell}x_i x_j + \sum_{i=1}^{d}\frac{\partial}{\partial x_\ell}b_i x_i\right\} \\
&= \frac{\partial}{\partial x_k}\left\{\sum_{i=1}^{d} A_{i\ell}x_i + b_\ell\right\} \\
&= A_{k\ell}\,,
\end{aligned}$$

thus the Hessian of f is $\boldsymbol{A}$. The function f is convex when $\boldsymbol{A}$ is positive semi-definite, and strictly convex if $\boldsymbol{A}$ is positive definite.