EE 641 Midterm Exam

October 18, Fall 2019

Name: **Key**

Instructions

The following is an in-class closed-book exam.

- This exam contains 3 problems worth a total of 100 points.
- You may not use any notes, textbooks, or calculators.
- You are allowed up to 55 minutes to complete the exam.

Good luck.

Problem 1. (32pt) Let $X \sim N(0, R)$ where R is a $p \times p$ symmetric positive-definite matrix. Further define the precision matrix, $B = R^{-1}$, and use the notation

$$B = \begin{bmatrix} 1/\sigma^2 & A \\ A^t & C \end{bmatrix} ,$$

where $A \in \mathbb{R}^{1\times(p-1)}$ and $C \in \mathbb{R}^{(p-1)\times(p-1)}$.

a) Calculate the marginal density of X_1, the first component of X, given the components of the matrix R.

b) Calculate the conditional density of X_1 given all the remaining components, $Y = [X_2, \cdots, X_p]^t$.

c) What is the conditional mean and covariance of X_1 given Y?

Solution:

Part a) Since $X \sim N(0, R)$, we know that $E[X_1] = 0$ and $E[X_1^2] = R_{1,1}$, and therefore, we have that

$$p(x_1) = \frac{1}{\sqrt{2\pi R_{1,1}}} \exp\left\{-\frac{1}{2R_{1,1}} x_1^2\right\}$$

Part b) We know that

$$p(x) = \frac{1}{z} \exp\left\{-\frac{1}{2} x^t B x\right\}$$

So then we know that

$$\log p(x_1, y) = -\frac{1}{2}\left(\frac{x_1^2}{\sigma^2} + 2x_1 A y + y^t C y\right) + const$$

We also know that

$$\log p(x_1|y) = \log p(x_1, y) - \log p(y)$$

So therefore, we can add any function of $z(y)$ for now, and we will be able to calculate it in the end by normalizing the probability.

$$\begin{aligned} \log p(x_1|y) &= -\frac{1}{2}\left(\frac{x_1^2}{\sigma^2} + 2x_1 A y + y^t C y\right) + z(y) \\ &= -\frac{1}{2\sigma^2}\left(x_1 + \sigma^2 A y\right)^2 + z'(y) \end{aligned}$$

Part c) From the solution of part b), we see that the conditional probability has the form

$$p(x_1|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x_1 - \mu(y)\right)^2\right\} ,$$

where the conditional mean is given by $\mu(y) = -\sigma^2 A y$ and the conditional variance is given by σ^2.

Problem 2. (32pt) Consider the optimization problem

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \left\{ ||y - Ax||_\Lambda^2 + x^t Bx \right\}$$

where A is a nonsingular $N \times N$ matrix, B is a positive-definite $N \times N$ matrix, and Λ is a diagonal and positive-definite matrix.

a) Derive a closed form expression for the solution.

b) Calculate an expression for the gradient descent update using step size $\mu \geq 0$.

c) Calculate an expression for the coordinate descent update.

Solution:

Part a) Define the MAP cost function given by

$$f(x) = ||y - Ax||_\Lambda^2 + x^t Bx$$

The we have that

$$\nabla f(x) = -2A^t \Lambda (y - Ax) + 2Bx$$

We know that $\nabla f(\hat{x}) = 0$, so we have that

$$\begin{aligned}
0 &= -2A^t \Lambda (y - A\hat{x}) + 2B\hat{x} \\
0 &= -A^t \Lambda y + \left(A^t \Lambda A + B\right) \hat{x} \\
\left(A^t \Lambda A + B\right) \hat{x} &= A^t \Lambda y \\
\hat{x} &= \left(A^t \Lambda A + B\right)^{-1} A^t \Lambda y
\end{aligned}$$

Part b) Using pseudo-code notation, we have that

$$\begin{aligned}
x &\leftarrow x - \frac{\mu}{2} \nabla f(x) \\
x &\leftarrow x + \mu \left[A^t \Lambda (y - Ax) - Bx\right]
\end{aligned}$$

Part c) We can compute the update for the i^{th} pixel by solving for α such that

$$\left[\nabla f\left(x + \alpha \epsilon_i\right)\right]_i = 0 = -2A_{*,i}^t \Lambda (y - A[x + \alpha \epsilon_i]) + 2B[x + \alpha \epsilon_i]$$

where ϵ_i is a vector that is 1 for the i^{th} component, and 0 elsewhere. So this results in

$$\begin{aligned}
0 &= -2A_{*,i}^t \Lambda (y - A[x + \alpha \epsilon_i]) + 2B_{i,*}[x + \alpha \epsilon_i] \\
0 &= -2A_{*,i}^t \Lambda (e + \alpha A_{*,i}) + 2B_{i,*}[x + \alpha \epsilon_i] \\
2A_{*,i}^t \Lambda e - 2B_{i,*} x &= \alpha 2 \left(A_{*,i}^t \Lambda A_{*,i} + B_{i,i}\right) \\
\alpha &= \frac{A_{*,i}^t \Lambda e - B_{i,*} x}{A_{*,i}^t \Lambda A_{*,i} + B_{i,i}}
\end{aligned}$$

So the ICD update is given by

$$x_i \leftarrow x_i + \frac{A_{*,i}^t \Lambda e - B_{i,*} x}{A_{*,i}^t \Lambda A_{*,i} + B_{i,i}}$$

where $e = y - Ax$.

Problem 3. (36pt) For the following problem, consider the MAP cost function given by

$$f(x) = \frac{1}{2}||y - Ax||^2_\Lambda + \sum_{\{s,r\}\in\mathcal{P}} b_{s,r}\rho(x_s - x_r) \ . \tag{1}$$

where $y \in \Re^N$, $x \in \Re^N$, $A \in \Re^{N\times N}$ has rank N, Λ is positive-definite, and $\rho(\Delta)$ is a positive convex function of Δ. Also, define the subevel set $\mathcal{A}_\alpha$ to be

$$\mathcal{A}_\alpha = \left\{x \in \mathbb{R}^N : f(x) \leq \alpha\right\} \ ,$$

and define the inverse image of a set $S \subset \Re$ to be

$$f^{-1}(S) = \{x \in \Re^N : f(x) \in S\} \ .$$

For this problem you can use the following theorems:

T1: A set in $\Re^N$ is compact if and only if it is closed and bounded.

T2: If f is a continuous function and S is closed, then the inverse images $f^{-1}(S)$ is closed.

a) Prove that for all $\alpha \in \mathbb{R}$ the subevel set $\mathcal{A}_\alpha$ is closed.
b) Prove that there exists an $\alpha \in \mathbb{R}$ such that the subevel set $\mathcal{A}_\alpha$ is non-empty and compact.
c) Prove there exists a MAP estimate, $\hat{x}$, so that $\forall x \in \Re^N$, $f(\hat{x}) \leq f(x)$.
d) Prove that the MAP estimate is unique.

Solution:

Part a) To prove this, notice the following three facts.
First, $A_\alpha = f^{-1}(S)$ where $S = (-\infty, \alpha]$. Second, $S = (-\infty, \alpha]$ is closed set. Third, since f is the sum of convex functions, it must be convex; and since all convex functions are continuous, f must be continuous.
So then using these three facts, and applying theorem T1, we see that $A_\alpha = f^{-1}(S)$ must be a closed set.

Part b) Since we know that $\mathcal{A}_\alpha$ is closed, we only need to show that there exists an α such that $\mathcal{A}_\alpha$ is bounded and non-empty in order to prove that it is compact.
First, select any α such that

$$\alpha > f(0) \ ,$$

where $x = 0$ denotes the vector with all elements of zero. Then $\mathcal{A}_\alpha \neq \emptyset$.
Next notice that

$$f(x) \ = \ \frac{1}{2}||y - Ax||^2_\Lambda + \sum_{\{s,r\}\in\mathcal{P}} b_{s,r}\rho(x_s - x_r)$$

$$
\begin{aligned}
&\geq \frac{1}{2}||y - Ax||_\Lambda^2 \\
&\geq \frac{\lambda_{min}}{2}||y - Ax||^2 \\
&\geq \frac{\lambda_{min}}{2}\{||y - Ax||\}^2 \\
&\geq \frac{\lambda_{min}}{2}\{||Ax|| - ||y||\}^2 \\
&\geq \frac{\lambda_{min}}{2}\{\beta_{min}||x|| - ||y||\}^2
\end{aligned}
$$

where $\lambda_{min} > 0$ is the minimum eigenvalue of Λ, and $\beta_{min} > 0$ is the smallest singular value A.

Next define the ball $B_r = \{x \in \Re^N : ||x|| \leq r\}$ where

$$
r = \frac{1}{\beta_{min}}\left(\sqrt{\frac{2\alpha}{c_{min}}} + ||y||\right) .
$$

Then for all $x \notin B_r$, we have that

$$
\begin{aligned}
f(x) &\geq \frac{c_{min}}{2}\{\beta_{min}||x|| - ||y||\}^2 \\
&\geq \frac{c_{min}}{2}\{\beta_{min}r - ||y||\}^2 \\
&= \alpha \ ,
\end{aligned}
$$

and therefore it must be that $x \notin \mathcal{A}_\alpha$. In other word, $\bar{B}_r \subset \bar{\mathcal{A}}_\alpha$, which in turn implies that $\mathcal{A}_\alpha \subset B_r$. Then since B_r is bounded, $\mathcal{A}_\alpha$ must also be bounded. Then since $\mathcal{A}_\alpha$ is bounded and closed, it must be compact.

Part c) Since $f(x)$ is a continuous function on the non-empty compact set $\mathcal{A}_\alpha$, it must take on a global minimum $x^* \in \mathcal{A}_\alpha$.

Pick any $x \in \Re^N$. Then either $x \in \mathcal{A}_\alpha$ or $x \notin \mathcal{A}_\alpha$. If $x \in \mathcal{A}_\alpha$, then by the fact that x^* is a global minimum, we know that $f(x^*) \leq f(x) \leq \alpha$. If $x \notin \mathcal{A}_\alpha$, then by the fact that x^* is a global minimum, we know that $f(x^*) \leq \alpha \leq f(x)$. So therefore, x^* is a global minimum of $f(x)$ over $x \in \Re^N$.

Part d) We will prove this by contradiction. First notice that $f(x)$ is strictly convex since it is the sum of a strictly convex and convex function. Assume that there exists $x' \neq x^*$ such that $f(x') = f(x^*)$. Define $\bar{x} = \frac{x'+x^*}{2}$, then since $f(x)$ is strictly convex, we have that

$$
f(\bar{x}) < \frac{f(x') + f(x^*)}{2} .
$$

But this can not be possible since $f(x^*)$ is a global minimum. So therefore, x^* must be the unique global minimum.