EE 641 Final Exam

Fall 2018

Name: Key

Instructions

- This exam contains 4 problems worth a total of 100 points.
- You may have up to 120 minutes to take the exam.
- You may not use any notes, textbooks, or calculators.
- Answer questions precisely and completely. Credit will be subtracted for vague answers.

Good luck.

Problem 1. (25pt)

Consider the MAP cost function given by

$$f(x; y) = ||y - Ax||^2 + ||x||_1$$

where x and y are vectors in $\mathbb{R}^N$, $||x||_1$ is the L_1 norm of x, and $A \in \mathbb{R}^{N\times N}$ is a full rank matrix. Furthermore, we say that, x^*, is a fixed point of the ICD algorithm if a full ICD update using an initial value of x^* produces an updated value of x^*.

a) Prove that $f(x; y)$ is a strictly convex function of x.

b) Prove that this function takes on a local minimum which is also its unique global minimum.

c) Calculate a closed form expression for the ICD update.
(Hint: The solution uses a shrinkage function.)

d) If x^* is a fixed point of the ICD algorithm, then is it is a global minimum of $f(x; y)$? Justify your answer.

a) $\nabla ||y - Ax||^2 = 2(Ax - y)^t A$

$= x^t A^t A - y^t A$

$\nabla\nabla ||y - Ax||^2 = \underbrace{2A^t A}$ ⇐ Hessian

$|A| > 0 \Rightarrow$ positive ~~non negative~~ definite

$\Rightarrow$ ~~convex~~ strictly convex

$||x||_1 = \sum_{i=1}^{N} |x_i|$ ⇐ sum of convex functions $\Rightarrow$ convex

$f(x; y) =$ strictly convex function + convex function

$\Rightarrow$ strictly convex

Name: ______________________________

b) $f(0) = \|y\|^2$ Since $|A| > 0$ then,

$\exists r_0$ s.t. $\forall x$ s.t. $\|x\| \geq r_0$

the $f(x) > \|y\|^2$

Therefore, $f(x)$ takes on its infemum on the set $\|x\| \leq r_0$

$\Rightarrow$ Since $\|x\| \leq r_0$ is a compact set,

$\exists x^*$ s.t. $\|x^*\| \leq r_0$ and

$$f(x^*) \leq f(x) \quad \forall x \in \{\|x\| \leq r_0\}$$

$$\Rightarrow f(x^*) \leq f(x) \quad \forall x \in \mathbb{R}^n$$

Since $f(x)$ is strictly convex

x^* must be unique.

c) $\nabla f(x) = 2A^t \underbrace{(Ax - y)}_{e}$

$$[\nabla f(x)]_i = \frac{\partial f}{\partial x_i} = 2e^t \underbrace{A_{*i}}_{i\text{th row of } A} = \mathcal{O}_1$$

$$\frac{\partial^2 f}{\partial x_i^2} = 2\|A_{*i}\|^2 = \mathcal{O}_2$$

Name: ______________________

Let x_i' be the updated value of x_i, then

$$f(x_i') = a + \theta_1(x_i' - x_i) + \frac{\theta_2}{2}(x_i' - x_i)^2 + |x_i'|$$

$$= a' + \frac{\theta_2}{2}\left(x_i' - \left(x_i - \frac{\theta_1}{\theta_2}\right)\right) + |x_i'|$$

$$x_i^* = \arg\min_{x_i'} \left\{ \frac{\theta_2}{2}\left(x_i' - \left(x_i - \frac{\theta_1}{\theta_2}\right)\right)^2 + |x_i'| \right\}$$

$$= \arg\min_{x_i'} \left\{ \frac{1}{2}\left(x_i' - \left(x_i - \frac{\theta_1}{\theta_2}\right)\right)^2 + \frac{1}{\theta_2}|x_i'| \right\}$$

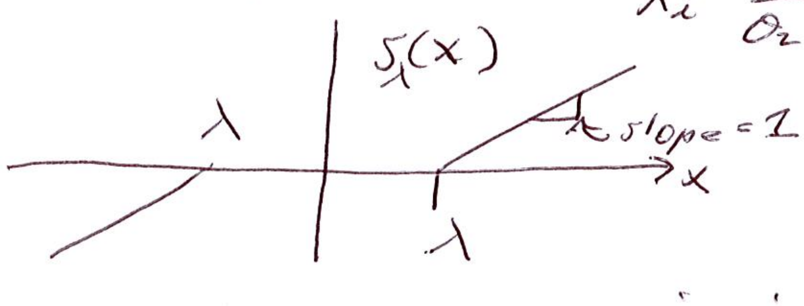
$$= S_{1/\theta_2}\left(x_i - \frac{\theta_1}{\theta_2}\right)$$

where

$$\frac{1}{\theta_2} = \frac{1}{2\|A_{*i}\|^2}$$

$$x_i - \frac{\theta_1}{\theta_2} = x_i + \frac{(y-Ax)^T A_{*i}}{\|A_{*i}\|^2}$$

shrinkage function



d) Yes, this is a global minimum.

However, it would not be a global minimum if $f(x) = \|y-Ax\|^2 + \|Bx\|_1$ for $B \neq$ diagonal

Problem 2. (25pt)

Consider the general problem of convex optimization with a positivity constraint given by

$$\hat{x} = \arg \min_{x \in \mathbb{R}^{+N}} f(x) \ ,$$

where $f : \mathbb{R}^N \to \mathbb{R}$ is a convex function, and $\mathbb{R}^{+N}$ represents the N-dimensional set of non-negative real numbers.

In order to remove the constraint, we may define the proper, closed, convex function

$$g(x) = \begin{cases} 0 & \text{if } x \in \mathbb{R}^{+N} \\ \infty & \text{if } x \notin \mathbb{R}^{+N} \end{cases} \ .$$

Then the minimum is given by the solution to the unconstrained optimization problem

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \{f(x) + g(x)\} \ . \tag{1}$$

Using this formulation, do the following.

a) Use variable splitting to derive a constrained optimization problem that is equivalent to equation (1).

b) Formulate the augmented Lagrangian for this constrained optimization problem and give the iterative algorithm for solving the augmented Lagrangian problem.

c) Use the ADMM approach to formulate an iterative algorithm for solving the augmented Lagrangian.

d) Simplify the expressions for the ADMM updates and give the general simplified ADMM algorithm for implementing positivity constraints in convex optimization problems.

a) $(\hat{x}, \hat{v}) = \arg\min_{x=v} \left(f(x) + g(v) \right)$

b) $(\hat{x}, \hat{v}) = \arg\min_{x,v} \left(f(x) + g(v) + \frac{1}{2a} \|x - v + u\|^2 \right)$

Name: ______________________

repeat {

$$(\hat{x}, \hat{v}) \leftarrow \underset{(x,v)}{\arg\min} \{ f(x) + g(v) + \|x - v + u\|^2 \}$$

$$u \leftarrow u + (\hat{x} - \hat{v})$$

}

c) initialize $\hat{v}$; $u = 0$

repeat {

$$\hat{x} \leftarrow \underset{x}{\arg\min} \{ f(x) + \|x - \hat{v} + u\|^2 \}$$

$$\hat{v} \leftarrow \underset{v}{\arg\min} \{ g(v) + \|\hat{x} - v + u\|^2 \}$$

$$u \leftarrow u + (\hat{x} - \hat{v})$$

}

d) $$\hat{v} \leftarrow \underset{v}{\arg\min} \{ g(v) + \|v - \hat{x} - u\|^2 \}$$

Name: ______________________

$$\hat{v} = \arg\min_{v} \left\{ g(v) + \| v - (\hat{x} + u) \|^2 \right\}$$

$$= \text{clip}(\hat{x} + u, [0, \infty))$$

Initialize $\hat{v} = 0$ $u = 0$

repeat {

$\hat{x} \Leftarrow \arg\min_{x} \left\{ f(x) + \| x - \hat{v} + u \|^2 \right\}$

$\hat{v} \Leftarrow \text{clip}(\hat{x} + u, [0, \infty])$

$u \Leftarrow u + (\hat{x} - \hat{v})$

}

Problem 3. (25pt)

Let $\{X_n\}_{n=1}^N$ be i.i.d. random variables with $P\{X_n = i\} = \pi_i$ for $i = 0, \cdots, M-1$. Also, assume that $Y_n \in \mathbb{R}^p$ are conditionally independent Gaussian random vectors given X_n and that the conditional distribution of Y_n given X_n is distributed as $N(\mu_{x_n}, \gamma_{x_n})$.

a) Give an expression for the maximum likelihood estimates of the parameters $\{\pi_i, \mu_i, \gamma_i\}_{i=0}^{M-1}$ given the complete data $\{X_n, Y_n\}_{n=1}^N$.

b) Give an expression for the posterior distribution of X_n given $\{Y_n\}_{n=1}^N$.

c) Give an expression for the expectation and maximization steps of the EM algorithm for estimating the parameters $\{\pi_i, \mu_i, \gamma_i\}_{i=0}^{M-1}$ from the observations $\{Y_n\}_{n=1}^N$.

a)

$$N_i = \sum_{n=1}^{N} \delta(X_n = i)$$

$$b_i = \sum_{n=1}^{N} Y_n \, \delta(X_n = i)$$

$$S_i = \sum_{n=1}^{N} Y_n Y_n^t \, \delta(X_n = i)$$

$$\hat{\pi}_i = \frac{N_i}{N}$$

$$\hat{\mu}_i = \frac{b_i}{N_i}$$

$$\hat{R}_i = \frac{S_i}{N_i} - \left(\frac{b_i}{N_i}\right)^2$$

Name: ______________________________

b) $p(x_n = k | y_n) = f(k|y_n)$

$$= \frac{p(y_n | x_n = k)\, p(x_n = k)}{\sum_k p(y_n | x_n = k)\, p(x_n = k)}$$

$$f(k|y) = \frac{\frac{1}{z} \exp\left\{-\frac{1}{2} \|y - \mu_k\|^2_{R_k}\right\} \pi_k}{\sum_k \frac{1}{z} \exp\left\{-\frac{1}{2} \|y - \mu_k\|^2_{R_k}\right\} \pi_k}$$

c)

$$\bar{N}_i = \sum_{n=1}^{N} f(i|y_n)$$

$$\bar{b}_i = \sum_{n=1}^{N} y_n f(i|y_n)$$

$$\bar{S}_i = \sum_{n=1}^{N} y_n y_n^t f(i|y_n)$$

$$\hat{\pi}_i^{(k+1)} = \frac{\bar{N}_i}{N}$$

$$\hat{\mu}_i^{(k+1)} = \frac{\bar{b}_i}{\bar{N}_i}$$

$$\hat{R}_i^{(k+1)} = \frac{\bar{S}_i}{\bar{N}_i} - \left(\frac{\bar{b}_i}{\bar{N}_i}\right)^2$$

Problem 4. (25pt)

Consider the random variable X with density

$$p(x) = \frac{1}{\sigma z(p)} \exp\left\{ -\frac{1}{p\sigma^p}|x|^p \right\} ,$$

with $p = 1.2$ and $\sigma = 1$. Consider the case of a Metropolis simulation algorithm for sampling from the distribution of $p(x)$ with the proposals generated as $W \leftarrow X^k + Z$ where $Z \sim N(0,1)$.

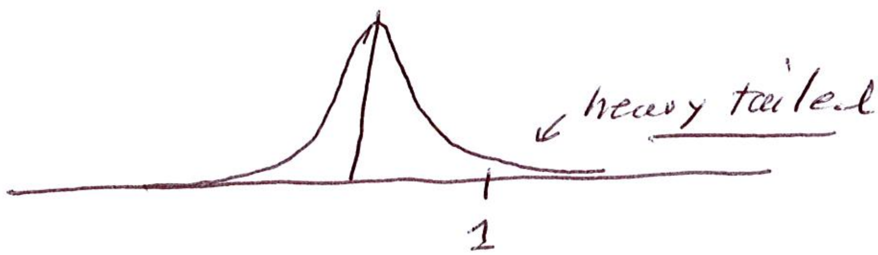
a) Sketch the density function for $p(x)$.

b) Give an expression for the proposal distribution $q(w|x)$ and show that the proposal distribution obeys the symmetry condition given by $q(w|x) = q(x|w)$.

c) Derive an expression for the acceptance probability α.

d) Write out the Metropolis algorithm in psuedo-code for generating samples from the distribution $p(x)$.

a)



b) $$q(w|x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\|w-x\|^2 \right\}$$

$$q(w|x) = q(x|w)$$

c) $$\alpha = \min\left\{1, e^{-\Delta E}\right\}$$

$$\Delta E = \frac{1}{p\sigma^p}\left(|w|^p - |x|^p\right)$$

$$= \frac{1}{p}\left(|w|^p - |x|^p\right)$$

Name: ____________________

$$\alpha = \min\left\{1, e^{-\frac{1}{p}(|w|^p - |x|^p)}\right\}$$

d) initialize x^0; $k \leftarrow 0$

repeat {

Generate $w \leftarrow x^k + z$

$z \sim N(0,1)$

$\alpha \leftarrow \min\left\{1, e^{-\frac{1}{p}(|w|^p - |x|^p)}\right\}$

With prob α

Accept w

$x^{k+1} \leftarrow w$

else

reject w

$x^{k+1} \leftarrow x^k$

k++

}