EE 641 Final Exam

December 9, Fall 2019

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)

Let $f(x)$ and $q(x;x')$ both be continuously differentiable and convex function of x, such that $\forall x', x \in \mathbb{R}^N$,

$$f(x') = q(x';x') \quad (1)$$

$$f(x) \leq q(x;x') \; . \quad (2)$$

Then using the initial state $x^{(k)}$, we can compute an updated state, $x^{(k+1)}$, using the following iteration.

$$x^{(k+1)} = \arg\min_{x\in\mathbb{R}^N} \left\{ q(x;x^{(k)}) \right\} \quad (3)$$

where $\forall x$,

$$q(x^{(k+1)};x^{(k)}) \leq q(x;x^{(k)}) \; .$$

a) Prove that $f(x^{(k+1)}) \leq f(x^{(k)})$.

b) Draw a figure illustrating why a) is true.

c) Prove that if $x^{(k+1)} = x^{(k)}$, then $\forall x$, $f(x^{(k+1)}) \leq f(x)$.

Solution:

Part a):

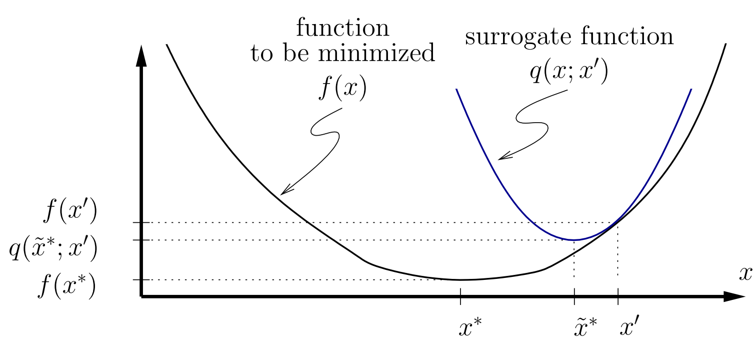
We know that $q(x^{(k)};x^{(k)}) = f(x^{(k)})$ and $q(x^{(k+1)};x^{(k)}) \geq f(x^{(k+1)})$. Since $x^{(k+1)}$ is chosen such that

$$x^{(k+1)} = \arg\min_{x\in\mathbb{R}^N} \{q(x;x^{(k)})\} \; ,$$

we also have $q(x^{(k+1)};x^{(k)}) \leq q(x^{(k)};x^{(k)})$. Combining these 2 steps, we get:

$$f(x^{(k+1)}) \leq q(x^{(k+1)};x^{(k)}) \leq q(x^{(k)};x^{(k)}) = f(x^{(k)}) \; .$$

Part b):



Plots illustrating the relationship between the true function to be minimized in a line search, $f(x)$, and the surrogate function, $q(x;x')$. Equation (1) requires that the functions be equal at $x = x'$, and equation (2) requires that the surrogate function upper bounds the true function everywhere.

Part c): We first prove that $\nabla f(x') = \nabla q(x';x')$ by contradiction. So assume that

$$d = \nabla \left[f(x') - q(x';x') \right] \neq 0$$

Then there must exist and $\epsilon > 0$, so that $f(x' + \epsilon d) - q(x' + \epsilon d; x') > 0$. But this contradicts our assumption that q is an upper bound on f. So this proves the $\nabla f(x') = \nabla q(x';x')$. Next, if $x^{(k+1)} = x^{(k)}$, then $x^{(k)}$ is a global minimum of $q(x;x^{(k)})$. This means that

$$\nabla q(x;x^{(k)})\Big|_{x=x^{(k)}} = 0 \ .$$

Using the previous result implies that

$$\nabla f(x)|_{x=x^{(k)}} = \nabla q(x;x^{(k)})\Big|_{x=x^{(k)}} = 0 \ .$$

Since f is continuously differentiable and convex, this means that $x^{(k)}$ must be a global minimum of f, which proves the result.

Problem 2. (25pt)

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

$$\hat{x} = \arg\min_{x} \{ f(x) + h(x) \} \ ,$$

where $f:\mathbb{R}^N \to \mathbb{R}$ and $h:\mathbb{R}^N \to \mathbb{R}$ are both convex functions on $\mathbb{R}^N$.

a) Use variable splitting to derive a constrained optimization problem that is equivalent to this problem.

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

c) Write out an expression for the two proximal map functions $F(x)$ and $H(x)$ corresponding the the function $f(x)$ and $h(x)$.

d) Write out the ADMM algorithm for solving this problem in terms of the proximal maps $F(x)$ and $H(x)$.

Solution:

Part a):

This is equivalent to the constrained optimization problem given by

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

Part b):

The augmented Lagrangian function is given by

$$L(x, v; u) = f(x) + h(v) + \frac{a}{2} \| x - y + u \|^2$$

The iterative algorithm is given by

initialize u=0

Repeat {

$(\hat{x}, \hat{v}) \leftarrow \arg\min_{x,v \in \mathbb{R}^N} L(x, v; u)$

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

}

Part c):

The two proximal functions are

$$
\begin{aligned}
F(v) &= \arg\min_{x\in\mathbb{R}^N} \left\{ f(x) + \frac{a}{2}\|x-v\|^2 \right\} \\
H(v) &= \arg\min_{x\in\mathbb{R}^N} \left\{ h(x) + \frac{a}{2}\|x-v\|^2 \right\} \ .
\end{aligned}
$$

Part d):

The ADMM algorithm for this problem is given by

initialize u=0

initialize v=0

Repeat {

$x \leftarrow F(v-u)$

$v \leftarrow H(x+u)$

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

}

Problem 3. (25pt)

Let $\{X_n\}_{n=1}^N$ be i.i.d. random variables with distribution

$$P\{X_n = m\} = \pi_m \ ,$$

where $\sum_{m=0}^{M-1} \pi_m = 1$. Also, let Y_n be conditionally independent random variables given X_n, with Poisson conditional distribution

$$p(y_n|x_n = m) = \frac{\lambda_m^{y_n} e^{-\lambda_m}}{y_n!} \ .$$

a) Write out the density function for the vector Y.

b) What are the natural sufficient statistics for the complete data (X, Y)?

c) Give an expression for the ML estimate of the parameter

$$\theta = (\pi_0, \lambda_0, \cdots, \pi_{M-1}, \lambda_{M-1}) \ ,$$

given the complete data (X, Y).

d) Give the EM update equations for computing the ML estimate of the parameter $\theta = (\pi_0, \lambda_0, \cdots, \pi_{M-1}, \lambda_{M-1})$ given the incomplete data Y.

Solution:

Part a): We first calculate the distribution of each Y_n given by

$$p(y_n) = \sum_{m=0}^{M-1} p(y_n|x_n = m)\pi_m = \sum_{m=0}^{M-1} \frac{\lambda_m^{y_n} e^{-\lambda_m}}{y_n!}\pi_m \ .$$

Since the Y_n are independent, we have that

$$p(y) = \prod_{n=1}^{N} \left\{ \sum_{m=0}^{M-1} \frac{\lambda_m^{y_n} e^{-\lambda_m}}{y_n!}\pi_m \right\}$$

Part b): The natural sufficient statistics for θ given (X, Y) are

$$\begin{aligned} N_m &= \sum_{N=1}^{N} \delta(X_n = m) \\ b_m &= \sum_{N=1}^{N} Y_n \delta(X_n = m) \end{aligned}$$

for $m = \{0, \cdots, M-1\}$.

Part c): Given the complete data, the ML estimate is given by

$$\begin{aligned} \hat{\pi}_m &= \frac{N_m}{N} \\ \hat{\lambda}_m &= \frac{b_m}{N_m} \end{aligned}$$

Part d): To compute the E-step, we first calculate the posterior probability

$$f_n(m) = P\{X_n = m | Y_n = y_n\} = \frac{\frac{\hat{\lambda}_m^{y_n} e^{-\hat{\lambda}_m}}{y_n!}\hat{\pi}_m}{\sum_{m=0}^{M-1} \frac{\hat{\lambda}_m^{y_n} e^{-\hat{\lambda}_m}}{y_n!}\hat{\pi}_m} \; .$$

Then $m = \{0, \cdots, M-1\}$ calculate

$$\begin{aligned} \hat{N}_m &\leftarrow \sum_{N=1}^{N} f_n(m) \\ \hat{b}_m &\leftarrow \sum_{N=1}^{N} Y_n \delta(X_n = m) \end{aligned}$$

To compute the M-step, calculate

$$\begin{aligned} \hat{\pi}_m &\leftarrow \frac{\hat{N}_m}{N} \\ \hat{\lambda}_m &\leftarrow \frac{\hat{b}_m}{\hat{N}_m} \end{aligned}$$

And repeat untiled converged.

Problem 4. (25pt)

Let X_n be a sequence of multivariate random vectors that form a homogeneous Markov Chain. More specifically, for each n, let $X_n = (X_{n,0}, X_{n,1}, \cdots, X_{n,M-1})$ where $X_{n,m} \in \{0, \cdots, K-1\}$.

Furthermore, let $p(x) > 0$ be any probability density function defined over the set $x \in \{0, \cdots, K-1\}^M$, and let $p_m(x_m | x_i \text{ for } i \neq m) > 0$ be its associated conditional density functions.

Given these definitions, the rule for generating X_n given X_{n-1} is given by:

Step 1: Generate a uniformly distributed random variable J on the set $\{0, \cdots, M-1\}$.

Step 2: Generate an independent random variable $W \sim p_J(x_J | X_i \text{ for } i \neq J)$.

Step 3: For $i \neq J$, set $X_{n,i} \leftarrow X_{n-1,i}$; and for $i = J$, set $X_{n,i} \leftarrow W$.

a) Show that the Markov chain has a finite number of states.

b) Show that the Markov chain is irreducible.

c) Show that the Markov chain is aperiodic.

d) Prove that Markov chain is ergodic with asymptotic distribution $p(x)$.

e) Intuitively, why does this make sense?

Solution:

Part a):

The total number of states for the Markov chain is $K^M < \infty$.

Part b):

Let z and z' be any two states, then it is enough to show that

$$P\{X_M = z' | X_0 = z) > 0 \ .$$

We can do this by having a sequence of states in which only one component of the state changes with each step. More specifically,

$$x_{t,i} = \begin{cases} z'_i & \text{if } i < t \\ z_i & \text{if } i \geq t \ . \end{cases}$$

Then

$$\begin{aligned} P\{X_t = x_t | X_{t-1} = x_{t-1}) &= P\{J = t\} p_t(x_{t,t} | x_{t-1,i} \text{ for } i \neq t) \\ &\geq \frac{1}{M} p_t(x_{t,t} | x_{t-1,i} \text{ for } i \neq t) \\ &> 0 \ , \end{aligned}$$

which implies that

$$P\{X_M = z'|X_0 = z) \geq \prod_{t=1}^{M} P\{X_t = x_t|X_t = x_{t-1}) > 0 \ .$$

So the Markov chain is irreducible.

Part c):

Notice that for all states z and all j,

$$\begin{aligned} P\{X_t = z|X_{t-1} = z) &\geq P\{J = t\}\, p_j(z_j|z_i \text{ for } i \neq j) \\ &= \frac{1}{M}\, p_j(z_j|z_i \text{ for } i \neq j) \\ &> 0 \ . \end{aligned}$$

So therefore, each state is aperiodic. Also, since all states communicate, if any state is aperiodic, then they must all be aperiodic.

Part d):

Since the Markov chain is aperiodic, and irreducible with a finite number of states then the Markov chain must be ergodic, and any solution to the FBE must be the unique stationary distribution of the Markov chain.

So therefore, it is enough to show that $p(x)$ solves the FBE.

If $X_0 \sim p(x)$, then

$$\begin{aligned} P\{X_1 = x_1\} &= \sum_{i=0}^{M} \frac{1}{M} p_t(x_{1,t}|x_{1,i} \text{ for } i \neq t) p(x_{1,i} \text{ for } i \neq j) \\ &= p_t(x_{1,t}|x_{1,i} \text{ for } i \neq t) p(x_{1,i} \text{ for } i \neq j) \\ &= p(x_1) \ . \end{aligned}$$

So the FBE hold for $P\{X_1 = x_1\} = p(x_1)$, which is the stationary distribution of the MC.

Part e):

Intuitive, if you replace a randomly selected component of the state vector with a random variable chosen from the desired conditional distribution given its neighbors, then you would expect that it should eventually settle down the the distribution $p(x)$.
