PURDUE

ECE 64100

Midterm Exam, November 1, Fall 2024

NAME ______________________________________

PUID ____

Exam instructions:

- A fact sheet is included **at the end of this exam** for your use.
- You have 60 minutes to work the exam.
- This is a closed-book and closed-note exam. You may not use or have access to your book, notes, any supplementary reference, a calculator, or any communication device including a cell-phone or computer.
- You may not communicate with any person other than the official proctor during the exam.

To ensure Gradescope can read your exam:

- Write your full name and PUID above and on the top of every page.
- Answer all questions in the area designated for each problem.
- Write only on the front of the exam pages.
- DO NOT run over to the next question.

Name/PUID: Key

Problem 1.(30pt) ICD update

Our goal is to minimize the MAP cost function

$$f(x) = \left\{ \frac{1}{2}\|y - Ax\|^2 + \frac{1}{2}x^t Bx \right\},$$

by performing the line search given by

$$\alpha^* = \arg\min_{\alpha} \left\{ f(x + \alpha d) \right\},$$

where d is the direction.

Problem 1a) For **gradient** descent optimization, give an expression for d.

Problem 1b) For **gradient** descent optimization, **derive** an expression for α^*?

Problem 1c) For **coordinate** descent optimization, give an expression for d.

Problem 1d) For **coordinate** descent optimization, **derive** an expression for α^*?

Problem 1e) Give a condition on d that ensures that $\alpha^* \geq 0$.

Solution:

Q1a:

$$d = -\nabla f(x) = A^t(y - Ax) - Bx$$

Q1b:

$$\begin{aligned}
0 &= \frac{df(x + \alpha d)}{d\alpha} \\
&= d^t \left[\nabla f(x + \alpha d) \right] \\
&= d^t \left[(-A)^t (y - A(x + \alpha d)) + B(x + \alpha d) \right] \\
&= d^t \left[A^t A(x + \alpha d) - A^t y + B(x + \alpha d) \right] \\
&= d^t A^t Ax + \alpha d^t A^t Ad - d^t A^t y + d^t Bx + \alpha d^t Bd \\
&= \alpha d^t A^t Ad + \alpha d^t Bd + d^t A^t Ax - d^t A^t y + d^t Bx \\
&= \alpha \left[d^t A^t Ad + d^t Bd \right] - d^t A^t (y - Ax) + d^t Bx \\
&= \alpha d^t \left[A^t A + B \right] d - d^t \left[A^t (y - Ax) - Bx \right]
\end{aligned}$$

Solving for α^* results in

$$\alpha^* = \frac{d^t \left[A^t (y - Ax) - Bx \right]}{d^t \left[A^t A + B \right] d}.$$

However, for the particular case of gradient descent $d = A^t(y - Ax) - Bx$. So we have that

$$\alpha^* = \frac{d^t d}{d^t \left[A^t A + B\right] d}$$

Q1c:

$$d = e_i$$

where $e_i \in \mathbb{R}^N$ with $[e_i]_k = \delta(i - k)$.

Q1d: From part b) we know that

$$\alpha^* = \frac{d^t \left[A^t(y - Ax) - Bx\right]}{d^t \left[A^t A + B\right] d}.$$

So by taking $d = e_i$ we have that

$$\alpha^* = \frac{(y - Ax)A_{*,i} - x^t B_{*,i}}{\|A_{*,i}\|^2 + B_{i,i}}.$$

Q1e: The condition is that

$$\begin{aligned} d^t \nabla f(x) &\leq 0 \\ d^t(-\nabla f(x)) &\geq 0 \\ d^t \left[A^t(y - Ax) - Bx\right] &\geq 0 \\ (y - Ax)^t Ad - x^t Bd &\geq 0 \end{aligned}$$

This ensures that $\alpha^* \geq 0$.

Name/PUID: ______________________________

Problem 2.(48pt) Surrogate Functions

Consider the two functions

$$\rho(x) = |x|$$
$$f_z(x) = |x - z|$$
$$g(x) = \sum_{k=0}^{K-1} |x - x_k|$$

where $x, z, x_k \in \mathbb{R}$.

Problem 2a) Sketch a plot of $\rho(x)$.

Problem 2b) Sketch the best quadratic surrogate function, $\rho(x; x')$, together with the function $\rho(x)$ for $x' = 1$.

Problem 2c) Determine a general expression for the best quadratic surrogate function $\rho(x; x')$ for general choices of x and x'.

Problem 2d) Sketch a plot of $f_z(x)$ when $z = 1$.

Problem 2e) Sketch the best quadratic surrogate function, $f_z(x; x')$, together with the function $f_z(x)$ for $z = 1$ and $x' = 2$.

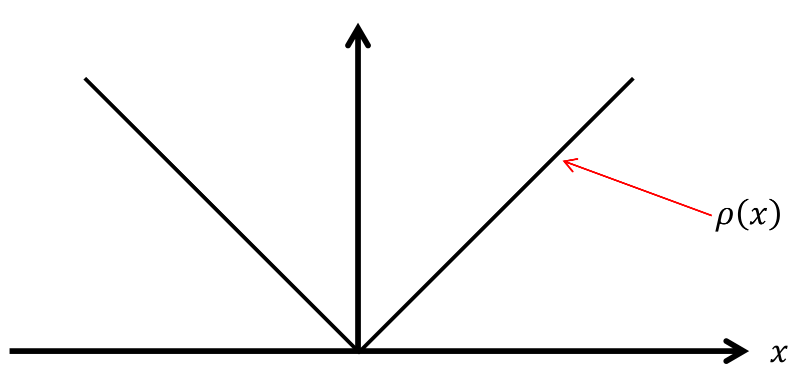
Problem 2f) Determine a general expression for the best quadratic surrogate function $f_z(x; x')$ for general choices of x, x', and z.

Problem 2g) Determine a general expression for the best quadratic surrogate function $g(x; x')$ for general choices of x, x', and x_k.

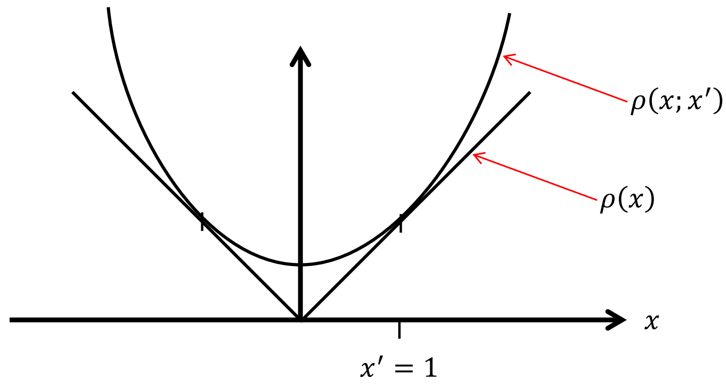
Problem 2h) Specify an iterative algorithm in terms of the surrogate function $g(x; x')$ that will converge to the global minimum of the function.

Solution:

Q2a:



Q2b:

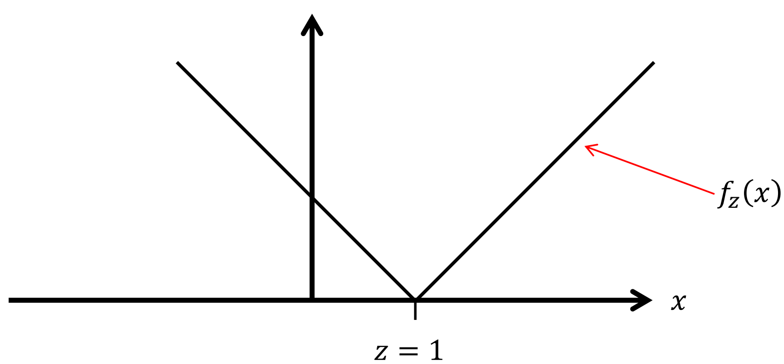


Q2c: When $x' \neq 0$, then the surrogate function has the form

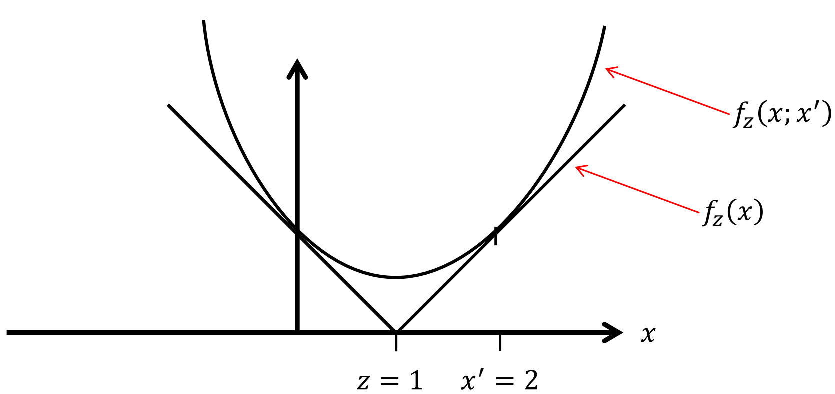
$$\rho(x;x') = \frac{\rho'(x')}{2x'}x^2 = \frac{1}{2|x'|}x^2.$$

However, the surrogate function does not exist when $x' = 0$

Q2d:



Q2e:



Q2f:

First notice that

$$f_z(x) = \rho(x - z).$$

So then we know that

$$\begin{aligned} f_z(x; x') &= \rho'(x - z; x' - z) \\ &= \frac{1}{2|x' - z|}(x - z)^2 \end{aligned}$$

Again, the surrogate function does not exist when $x' - z = 0$

Q2g:

First notice that

$$g(x) = \sum_{k=0}^{K-1} f_{x_k}(x).$$

So then we know that

$$g(x; x') = \sum_{k=0}^{K-1} \frac{1}{2|x' - x_k|}(x - x_k)^2$$

Again, the surrogate function does not exist when $x' - x_k = 0$ for any k.

Q2h:

$x' \leftarrow 0$

Repeat {

$x' \leftarrow \arg\min_x g(x; x')$

}

Name/PUID: ______________________________

Problem 3.(24pt) Proximal Maps

Define the cost function

$$f_v(x) = u(x) + \frac{1}{2\sigma^2}\|x - v\|^2,$$

and the associated proximal map as

$$F(v) = \arg\min_{x\in\mathbb{R}^N} f_v(x)$$

where $u : \mathbb{R}^N \to \mathbb{R} \cup \infty$ is a continuously differentiable proper closed convex function.

Problem 3a) Provide the outline of a proof that the function $f_v : \mathbb{R}^N \to \mathbb{R} \cup \infty$ takes on a minimum value.

(Hint: This proof is a little tricky, so you can just provide the outline of a proof. You can use the theorem that any continuous function on a compact set (i.e., closed and bounded set) takes on its minimum value.)

Problem 3b) Using the result of 3a above, prove that the function $f_v : \mathbb{R}^N \to \mathbb{R} \cup \infty$ takes on a **unique** minimum?

Problem 3c) Let $u(x; y) = \frac{1}{2}\|y - Ax\|_\Lambda^2$ so that

$$F(v; y) = \arg\min_{x\in\mathbb{R}^N} \left\{ \frac{1}{2}\|y - Ax\|_\Lambda^2 + \frac{1}{2\sigma^2}\|x - v\|^2 \right\}.$$

Then show that the proximal map has the explicit form

$$F(v; y) = v + \left(\frac{1}{\sigma^2} I + A^t \Lambda A\right)^{-1} A^t \Lambda \left(y - Av\right) .$$

Problem 3d) Let $F(v; y)$ be a proximal map with the form

$$F(v; y) = \arg\min_{x\in\mathbb{R}^N} \left\{ \frac{1}{2}\|y - Ax\|_\Lambda^2 + \frac{1}{2\sigma^2}\|x - v\|^2 \right\}.$$

Then show that $F(v; Y)$ has an interpretation as the MAP estimate of X given Y under the assumption that $Y = AX + W$ where X and W are independent Gaussian random vectors with $X \sim N(v, \sigma^2 I)$ and $W \sim N(0, \Lambda^{-1})$.

Solution:

Q3a:

1. Since $u(x)$ is proper, we know that there $\exists x \in \mathbb{R}^N$ s.t., $f_v(x) < \infty$.

2. Then prove $\exists \alpha \in \mathbb{R}$ such that $A = \{x \in \mathbb{R}^N : f_v(x) \leq \alpha\}$ is closed and bounded.

3. Then by the stated theorem, we know that $\exists x^* \in \mathbb{R}^N$ such that x^* is the global minimum of f_v on A.

4. Then show that x^* must be the minimum of f_v on $\mathbb{R}^N$.

Q3b:

Since f_v is strictly convex, x^* must be the unique minimum of f_v on $\mathbb{R}^N$.

Q3c:

Let $z = x - v$. Then we need to minimize the function

$$\begin{aligned}\tilde{f}(z) &= \frac{1}{2}\|y - A(z+v)\|_\Lambda^2 + \frac{1}{2\sigma^2}\|z\|^2 \\ &= \frac{1}{2}\|(y - Av) - Az\|_\Lambda^2 + \frac{1}{2\sigma^2}\|z\|^2 \\ &= \frac{1}{2}\|\epsilon - Az\|_\Lambda^2 + \frac{1}{2\sigma^2}\|z\|^2\end{aligned}$$

where $\epsilon = y - Av$.

The solution to this minimization problem is given by

$$\begin{aligned}z^* &= \left(\frac{1}{\sigma^2}I + A^t\Lambda A\right)^{-1} A^t\Lambda\epsilon \\ &= \left(\frac{1}{\sigma^2}I + A^t\Lambda A\right)^{-1} A^t\Lambda(y - Av).\end{aligned}$$

Then since $z^* = x^* - v$, we have that

$$\begin{aligned}x^* &= z^* + v \\ &= v + \left(\frac{1}{\sigma^2}I + A^t\Lambda A\right)^{-1} A^t\Lambda(y - Av).\end{aligned}$$

Q3d:

Consider the cost function of

$$f(x) = \frac{1}{2}\|y - Ax\|_\Lambda^2 + \frac{1}{2\sigma^2}\|x - v\|^2.$$

We can interpret the left-hand term as the forward model term $-\log p(y|x)$, and the right-hand term as the prior model $-\log p(x)$.

Then the forward model has the form $Y = AX + W$ where $W \sim N(0, \Lambda^{-1})$, and the prior model term corresponds to $X \sim N(v, \sigma^2 I)$.

ECE641 Fact Sheet

Maximum Likelihood (ML) Estimator (Frequentist)

$$\hat{\theta} = \arg\max_{\theta\in\Omega} p_\theta(Y) = \arg\max_{\theta\in\Omega} \log p_\theta(Y)$$
$$0 = \nabla_\theta p_\theta(Y)|_{\theta=\hat{\theta}}$$
$$\hat{\theta} = T(Y)$$
$$\bar{\theta} = \mathbb{E}_\theta[\hat{\theta}]$$
$$\text{bias}_\theta = \bar{\theta} - \theta \qquad \text{var}_\theta = \mathbb{E}_\theta[(\hat{\theta}-\bar{\theta})^2]$$
$$\text{MSE} = \mathbb{E}_\theta[(\hat{\theta}-\theta)^2] = \text{var}_\theta + (\text{bias}_\theta)^2$$

For $Y = AX + W$, where X and W are independent zero mean Gaussian distributed with R_X and R_W, respectively. Then the ML estimate is find by maximizing $log(p_{y/x}(y/x))$:

$$\hat{X}_{ML} = (A^t R_W^{-1} A)^{-1} A^t R_W^{-1} y$$

Maximum A Posteriori (MAP) Estimator

$$\begin{aligned}\hat{X}_{MAP} &= \arg\max_{x\in\Omega} p_{x|y}(x|Y)\\ &= \arg\max_{x\in\Omega} \log p_{x|y}(x|Y)\\ &= \arg\min_{x\in\Omega}\{-\log p_{y|x}(y|x) - \log p_x(x)\}\end{aligned}$$

For $Y = AX + W$, where X and W are independent zero mean Gaussian distributed with R_X and R_W, respectively. Then the MAP or equivalently MMSE estimate is:

$$\hat{X}_{MAP} = (A^t R_W^{-1} A + R_X^{-1})^{-1} A^t R_W^{-1} y$$

Power Spectral Density (zero-mean WSS Gaussian process)

1D DTFT:

$$S_X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} R(n) e^{-j\omega n}$$

2D DSFT:

$$S_X(e^{j\omega_1}, e^{j\omega_2}) = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} R(m,n) e^{-j\omega_1 m - j\omega_2 n}$$

Causal Gaussian Models

$\sigma_n^2 \triangleq \mathbb{E}[\mathcal{E}_n^2]$, $\hat{X} = HX$, $\mathcal{E} = (I-H)X = AX$, $\mathbb{E}[\mathcal{E}\mathcal{E}^t] = \Lambda$, $\Lambda = \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_N^2\}$

$p_x(x) = |det(A)| p_{\mathcal{E}}(Ax)$, $|det(A)| = 1$,
$R_X = (A^t \Lambda^{-1} A)^{-1}$

1-D Gaussian AR models:

- Toeplitz $H_{i,j} = h_{i-j}$
- Circulant $H_{i,j} = h_{(i-j) mod N}$
- P^{th} order IIR filter $X_n = \mathcal{E}_n + \sum_{i=1}^P X_{n-i} h_i$, $R_{\mathcal{E}}(i-j) = \mathbb{E}[\mathcal{E}_i \mathcal{E}_j] = \sigma_c^2 \delta_{i-j}$
- $R_X(n) * (\delta_n - h_n) * (\delta_n - h_{-n}) = R_{\mathcal{E}}(n) = \sigma_c^2 \delta_n$, $S_X = \frac{\sigma_c^2}{|1-H(\omega)|^2}$

2-D Gaussian AR:

- $\mathcal{E}_s = X_s - \sum_{r\in W_p} h_r X_{s-r}$,
- Toeplitz block Toeplitz $H_{mN+k, nN+l} = h_{m-n, k-l}$

Non-causal Gaussian Models

- $\sigma_n^2 \triangleq \mathbb{E}[\mathcal{E}_n^2 | X_i, i \neq n]$, $B_{i,j} = \frac{1}{\sigma_i^2}(\delta_{i-j} - g_{i,j})$, $\sigma_n^2 = (B_{n,n})^{-1}$, $g_{n,i} = \delta_{n-i} - \sigma_n^2 B_{n,i}$ (homogeneous: $g_{i,j} = g_{i-j}, \sigma_i^2 = \sigma_{NC}^2$)
- $G_{i,j} = g_{i,j}$, $\Gamma = \text{diag}\{\sigma_1^2, \sigma_2^2, ..., \sigma_N^2\}$, $B = \Gamma^{-1}(I-G)$, $\Gamma = \text{diag}(B)^{-1}$, $G = I - \Gamma B$, $\mathbb{E}[\mathcal{E}_n X_{n+k}] = \sigma_{NC}^2 \delta_k$
- $R_X(n) * (\delta_n - g_n) * (\delta_n - g_{-n}) = R_{\mathcal{E}}(n) = \sigma_{NC}^2(\delta_n - g_n)$, $S_X = \frac{\sigma_{NC}^2}{1-G(\omega)}$, $R_X(n) * (\delta_n - g_n) = \sigma_{NC}^2 \delta_n$
- **Relationship b/w AR and GMRF:** $\sigma_{NC}^2 = \frac{\sigma_c^2}{1+\sum_{n=1}^P h_n^2}$, $g_n = \delta_n - \frac{(\delta_n - h_n)*(\delta_n - h_{-n})}{1+\sum_{n=1}^P h_n^2} (= \frac{\rho}{1+\rho^2}(\delta_{n-1} + \delta_{n+1}),\ P = 1)$

Surrogate Function

Our objective is to find a surrogate function $\rho(\Delta;\Delta')$, to the potential function $\rho(\Delta)$.

Maximum Curvature Method

Assume the surrogate function of the form

$$\rho(\Delta;\Delta') = \alpha_1 \Delta + \frac{\alpha_2}{2}(\Delta - \Delta')^2$$

where $\alpha_1 = \rho'(\Delta')$ and $\alpha_2 = \max_{\Delta\in\mathbb{R}} \rho''(\Delta)$.

Symmetric Bound Method

Assume that potential function is bounded by symmetric and quadratic function of Δ, then the surrogate function is

$$\rho(\Delta;\Delta') = \frac{\alpha_2}{2}\Delta^2$$

which results in the following symmetric bound surrogate function:

$$\rho(\Delta;\Delta') = \begin{cases} \frac{\rho'(\Delta')}{2\Delta'}\Delta^2 & \text{if} \quad \Delta' \neq 0 \\ \frac{\rho''(0)}{2}\Delta^2 & \text{if} \quad \Delta' = 0 \end{cases}$$

Review of Convexity in Optimization

Definition A.6. ***Closed, Bounded, and Compact Sets***

Let $\mathcal{A} \subset \mathbb{R}^N$, then we say that $\mathcal{A}$ is:

- **Closed** if every convergent sequence in $\mathcal{A}$ has its limit in $\mathcal{A}$.
- **Bounded** if $\exists M$ such that $\forall x \in \mathcal{A}$, $\|x\| < M$.
- **Compact** if $\mathcal{A}$ is both closed and bounded.

Definition A.11. ***Closed Functions***

We say that function $f : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\}$ is **closed** if for all $\alpha \in \mathbb{R}$, the sublevel set $\mathcal{A}_\alpha = \{x \in \mathbb{R}^N : f(x) \leq \alpha\}$ is closed set.

Theorem A.6. ***Continuity of Proper, Closed, Convex Functions***

Let $f : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper convex function. Then f is closed if and only if it is lower semicontinuous.

Optimization Methods:

Gradient Descent: $x^{(k+1)} = x^{(k)} - \beta\nabla f(x^{(k)})$

Gradient Descent with Line Search:

$d^{(k)} = -\nabla f(x^{(k)})$

α solves the equation : $0 = \frac{\partial f(x^{(k)}+\alpha d^{(k)})}{\partial \alpha} = [\nabla f(x^{(k)} + \alpha d^{(k)})]^t d^{(k)}$.

Update: $x^{(k+1)} \leftarrow x^k + \alpha\frac{\|d^{(k)}\|^2}{\|d^{(k)}\|_Q^2}d^{(k)}$ where $Q = A^t\Lambda A + B$

Coordinate Descent :

$\alpha = \frac{(y-Ax)^t\Lambda A_{*,s} - x^t B_{*,s}}{\|A_{*,s}\|_\Lambda^2 + B_{s,s}}$ (for $Y|X \sim N(AX, \Lambda^{-1})$)

$x_s \leftarrow x_s + \frac{(y-Ax)^t A_{*,s} - \lambda(x_s - \Sigma_{r\in\partial s} g_{s-r} x_r)}{\|A_{*,s}\|^2 + \lambda}$, $\lambda = \frac{\sigma^2}{\sigma_x^2}$

Pairwise quadratic form identity

$x^t Bx = \sum_{s\in S} a_s x_s^2 + \frac{1}{2}\sum_{s\in S}\sum_{r\in S} b_{s,r}|x_s - x_r|^2$, $a_s = \sum_{r\in S} B_{s,r}$, $b_s = -B_{s,r}$

Miscellaneous

For any invertible matrix A, 1. $\frac{\partial |A|}{\partial A} = |A|A^{-1}$ 2. $\frac{\partial tr(BA)}{\partial A} = B$ 3. $tr(AB) = tr(BA)$