# PURDUE

ECE 64100

Midterm Exam, November 10, Fall 2023

NAME ______________________________________

PUID __________________________________________

**Exam instructions:**

- A fact sheet is included **at the end of this exam** for your use.
- You have 50 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:** ________________________________

**Problem 1.**(22pt) Emotional Equations

Write 50 words or less that describe your feelings and interpretation of the following equation:

$$F(z) = \arg\min_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - z\|^2 \right\}$$

**Solution:**

**Q1:**

*my solution:*

This is the typical form for a proximal map. The function $F$ typically takes inputs from $\Re^N$ and maps them back to the same space, $\Re^N$. Intuitively, $F$ returns a vector which moves towards the minimum of the function $f(x)$, but remains in the proximity of $z$. Hence its name.

*ChatGPT prose solution:*

The equation signifies seeking a point 'x' that minimizes a combination of 'f(x)' and the squared distance between 'x' and 'z,' with 'F(z)' as the result. It's like finding an optimal solution that balances 'f(x)' and proximity to 'z,' often used in optimization or signal processing.

*ChatGPT poem solution:*

In search of 'x,' a quest unfolds so bright,
To minimize 'f(x)' and closeness, we take flight.
'F(z)' emerges, a beacon in the math's night,
Balancing 'f(x)' and 'z's proximity, our guiding light.
In optimization's realm, it finds its might,
A signal's secret, hidden in plain sight.

**Name/PUID:** ______________________________

**Problem 2.**(36pt) MAP Estimation

Consider the following MAP estimator $\hat{x} = \arg\min_x f(x)$

$$f(x) = \frac{1}{2\sigma_w^2}\|y - Ax\|^2 + \frac{1}{2\sigma_x^2}\|x\|^2 \ ,$$

where $\lambda = \frac{\sigma_w^2}{\sigma_x^2}$.

**Problem 2a)** Write the forward model corresponding to this MAP estimator.

**Problem 2b)** Calculate the gradient, $\nabla f(x)$.

**Problem 2c)** Calculate a closed form expression for the MAP estimate, $\hat{x}$.

**Problem 2d)** Give the pseudo-code for the gradient descent algorithm.

**Problem 2e)** Calculate the ICD update for the $i^{th}$ pixel, $x_i$.

**Problem 2f)** Give the pseudo-code for the ICD algorithm.

**Solution:**

**Q2a:** The associated forward model is given by

$$\begin{aligned}
Y &= Ax + W \\
X &\perp\!\!\!\perp W \\
X &\sim N(0, \sigma_x^2 I) \\
W &\sim N(0, \sigma_w^2 I)
\end{aligned}$$

**Q2b:** The gradient is given by

$$\nabla f(x) = -\frac{1}{\sigma_w^2}A^t(y - Ax) + \frac{1}{\sigma_x^2}x$$

**Q2c:** For the MAP estimate, $\nabla f(x) = 0$, so we have that

$$\begin{aligned}
0 &= -A^t(y - A\hat{x}) + \lambda\hat{x} \\
0 &= -A^t y + A^t A\hat{x} + \lambda\hat{x} \\
0 &= -A^t y + (A^t A + \lambda)\hat{x} \\
\hat{x} &= (A^t A + \lambda)^{-1}A^t y
\end{aligned}$$

**Q2d:** The gradient descent update is given by $x \leftarrow \alpha \nabla f(x)$; so the pseudo-code is given by

Initialize $x$

Repeat {

$x \leftarrow x + \beta \left[ A^t (y - Ax) + \lambda x \right]$

}

where $\beta = \alpha \sigma_w^2$.

**Q2e:** Define $\epsilon_j \in \Re^p$ such that $[\epsilon_j]_i = \delta(j-i)$., and assume the update of the $i^{th}$ pixel is given by $x_i \leftarrow x_i + \alpha$. Then we can calculate the value of $\alpha$ as

$$
\begin{aligned}
0 &= \left[ \nabla f(x + \alpha \epsilon_i) \right]_i \\
0 &= \left[ -(y - Ax - \alpha A_{*,i})^t A + \lambda (x + \alpha \epsilon_i) \right]_i \\
0 &= -(y - Ax - \alpha A_{*,i})^t A_{*,i} + \lambda (x_i + \alpha) \\
0 &= -(y - Ax)^t A_{*,i} + \lambda x_i + \alpha (\|A_{*,i}\|^2 + \lambda)
\end{aligned}
$$

So the optimal value of $\alpha$ is given by

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

**Q2f:**

The gradient descent update is given by $x \leftarrow \alpha \nabla f(x)$; so the pseudo-code is given by

Initialize $x$

Repeat {

For each pixel $i$ {

$\alpha \leftarrow \frac{(y-Ax)^t A_{*,i} - \lambda x_i}{\|A_{*,i}\|^2 + \lambda}$

$x_i \leftarrow x_i + \alpha$

}

}

This is not a required part of the answer, but the fast version is given by

Initialize $x$

Initialize $e \leftarrow y - Ax$

Repeat {

For each pixel $i$ {

$\alpha \leftarrow \frac{e^t A_{*,i} - \lambda x_i}{||A_{*,i}||^2 + \lambda}$

$x_i \leftarrow x_i + \alpha$

$e \leftarrow e - \alpha A_{*,i}$

}

}

**Name/PUID:** ____________________________________

**Problem 3.**(42pt) Surrogate Functions

Consider the Poisson distributed random variable, $Y$, with distribution

$$p_\theta(y) = \frac{\theta^y e^{-\theta}}{y!} \ ,$$

where $\theta \in [0, \infty)$ parameterizes the distribution and $y \in [0, 1, 2, \ldots]$ is a non-negative integer.

**Problem 3a)** Calculate a closed from expression for the negative log likelihood, $-\log p_\theta(y)$.

**Problem 3b)** Then calculate $f(\theta) = -\log p_\theta(y) + C(y)$, so that $f(\theta)$ has the simplest form.

**Problem 3c)** Sketch $f(\theta)$ for $y = 1$ and $\theta > 0$.

**Problem 3d)** Prove that $f(\theta)$ is a convex function for $\theta > 0$.

**Problem 3e)** Calculate the maximum likelihood estimate of $\hat{\theta} = T(Y)$ as a function of $Y$.

**Problem 3f)** Calculate closed form expressions for the first and second derivatives of $f(\theta)$ for $\theta > 0$ given by

$$g(\theta) = \frac{d}{d\theta} f(\theta)$$
$$h(\theta) = \frac{d^2}{d\theta^2} f(\theta) \ .$$

**Problem 3g)** Calculate the functions $a(\theta')$ and $b(\theta')$ so that $f(\theta; \theta')$ with the form

$$f(\theta; \theta') = \frac{a(\theta')}{2} (\theta - \theta')^2 + b(\theta') (\theta - \theta') \ ,$$

is a surrogate function for $f(\theta)$ for $\theta \geq 1$.

**Solution:**

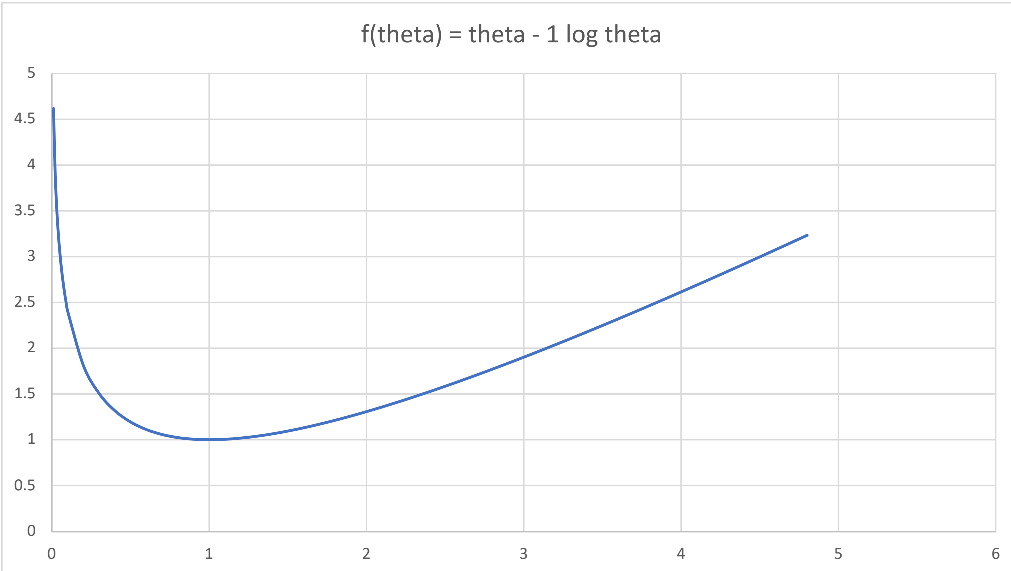
**Q3a:** The negative log likelihood is given by

$$-\log p_\theta(y) = -y \log \theta + \theta + \log y! \ .$$

**Q3b:** The negative log likelihood can be simplified by dropping the term that does not depend on $\theta$.

$$f(\theta) = \theta - y \log \theta \ .$$

**Q3c:** Plot of $f(x)$ below




**Q3d:** Since the second derivative for $\theta > 0$ is given by

$$\frac{d^2 f(\theta)}{d\theta^2} = \frac{y}{\theta^2} > 0 \ ,$$

we know that $f(\theta)$ must be strictly convex.

**Q3e:** The first derivative of $f(\theta)$ is given by

$$\frac{df(\theta)}{d\theta} = 1 - \frac{y}{\theta} \ .$$

So setting the derivative to zero yields,

$$\hat{\theta} = T(y) = y \ .$$

**Q3f:** The first and second derivatives are given by

$$\frac{df(\theta)}{d\theta} = 1 - \frac{y}{\theta}$$
$$\frac{d^2 f(\theta)}{d\theta^2} = \frac{y}{\theta^2} \ .$$

**Q3g:** Let $g(\theta)$ and $h(\theta)$ denote the first and second derivatives of $f(\theta)$. Then in order for $f(\theta;\theta')$ to be a surrogate function, it is enough to meet the following two conditions.

$$b(\theta') = g(\theta') = 1 - \frac{y}{\theta'}$$
$$a(\theta') = \max_{\theta \in [1,\infty)} h(\theta') = \max_{\theta > 1} \frac{y}{\theta^2} = y \ .$$

So the surrogate function is given by

$$f(\theta;\theta') = y(\theta - \theta')^2 + \left(1 - \frac{y}{\theta'}\right)(\theta - \theta') \ .$$

# 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 $\Delta$, 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 $f : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ 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 $f : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ 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:**
$d^{(k)} = -\nabla f(x^{(k)})$
$\alpha$ 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**

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

**Miscellaneous**

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