PURDUE

ECE 64100

Final Exam, December 11, Fall 2023

NAME ______________________________________

PUID ____

Exam instructions:

- A fact sheet is included **at the end of this exam** for your use.
- You have 120 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.(25pt) Emotional Equations

Write 75 words or less that describe your feelings about and interpretation of the following equation:

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

(You answer should be written in prose.)

Solution:

Q1: These are the detailed-balance equations for a Markov chain with stationary distribution π_i and transition probabilities $P_{i,j}$. Intuitively, it means that the rate of transitions from state i to state j equals the rate of the transitions from state j to i. If an MC is also irreducible with a finite number of states, then: a) The MC is ergodic; b) the MC has a steady state distribution of π_i; and c) the MC is reversible.

Name/PUID: ______________________________

Problem 2.(49pt) Plug-and-Play

Let $x, u \in \Re^N$, and

$$x^* = F(x^* - u^*)$$
$$x^* = H(x^* + u^*)$$

where

$$F(v) = \arg\min_{x \in \Re^N} \left\{ f(x) + \frac{1}{2\sigma^2} \|x - v\|^2 \right\} ,$$

where $f : \Re^N \to \Re$ is a continuously differentiable convex function and $H : \Re^N \to \Re$ is a denoiser assumed to be firmly non-expansive.

Furthermore, define the operator $T = (2H - I)(2F - I)$.

2a) For only this subproblem, assume that

$$H(v) = \arg\min_{x} \left\{ h(x) + \frac{1}{2\sigma^2} \|x - v\|^2 \right\} ,$$

where $h : \Re^N \to \Re$ is a continuously differentiable convex function.

Then prove that

$$\nabla_x \left\{ f(x) + h(x) \right\} \big|_{x = x^*} = 0 \ .$$

2b) Sketch a figure that illustrates the interpretation behind the equations

$$x^* = F(x^* - u^*)$$
$$x^* = H(x^* + u^*) \ .$$

2c) Explain in words the interpretation of the figure for 2b) above.

2d) Show that the operator T has a fixed point of $w_1^* = x^* - u^*$.

2e) Given the solution to the fixed point equation, Tw_1^*, find an expression for x^*.

2f) Specify a pseudo-code algorithm that uses the Mann iteration to compute w_1^*.

2g) Let $H_\theta(v)$ denote a family of functions parameterized by θ that can be used to model $H(v)$, and let $\{x_k\}_{k=0}^{K-1}$ denote a set of ground-truth images from your target application. Then describe in detail a method for estimating the parameter vector θ.

Solution:

Q2a: Since $f : \Re^N \to \Re$ is a continuously differentiable convex function and $h : \Re^N \to \Re$ is a continuously differentiable convex function, $f(x) + \frac{1}{2\sigma^2}\|x - v\|^2$ and $h(x) + \frac{1}{2\sigma^2}\|x - v\|^2$ are strictly convex.

Then substitute $v = x^* - u^*$, we have

$$x^* = F(x^* - u^*) = \arg\min_{x \in \Re^N} \left\{ f(x) + \frac{1}{2\sigma^2}\|x - x^* + u^*\|^2 \right\} \ ,$$

which leads to,

$$\begin{aligned} 0 &= \nabla_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - x^* + u^*\|^2 \right\}\Bigg|_{x=x^*} \\ &= \nabla_x f(x) + \frac{1}{\sigma^2}(x - x^* + u^*)\Bigg|_{x=x^*} \\ &= \nabla_x f(x) + \frac{1}{\sigma^2}u^*\Bigg|_{x=x^*} \ . \end{aligned}$$

So,

$$\nabla f(x^*) = \frac{-1}{\sigma^2}u^*$$

Similarly for h, substitute $v = x^* + u^*$, we have

$$x^* = H(x^* + u^*) = \arg\min_{x \in \Re^N} \left\{ h(x) + \frac{1}{2\sigma^2}\|x - x^* - u^*\|^2 \right\} \ ,$$

which leads to,

$$\nabla_x \left\{ h(x) + \frac{1}{2\sigma^2}\|x - x^* - u^*\|^2 \right\}\Bigg|_{x=x^*} = 0 \ .$$

So,

$$\nabla h(x^*) = \frac{1}{\sigma^2}u^*$$

Therefore,

$$\nabla_x \left\{ f(x) + h(x) \right\}\big|_{x=x^*} = \frac{-1}{\sigma^2}u^* + \frac{1}{\sigma^2}u^* = 0 \ .$$

Q2b:

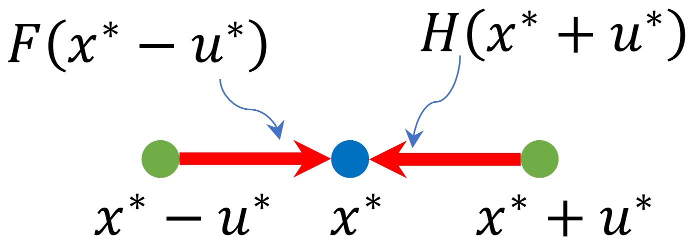

Figure 1: Illustration of Plug-and-Play

Q2c: $F(\cdot)$ is the forward model agent, and $H(\cdot)$ is the prior model agent. At the solution (x^*, u^*), the two agents are in equilibrium since they share the same output point each agent has an opposite but equal offset.

Q2d: Define

$$w_1^* = x^* - u^*$$
$$w_2^* = x^* + u^* \ .$$

Then we have that

$$(2F - I)w_1^* = 2\frac{w_1^* + w_2^*}{2} - w_1 = w_2^* \tag{1}$$
$$(2H - I)w_2^* = 2\frac{w_1^* + w_2^*}{2} - w_2 = w_1^* \ . \tag{2}$$

So substituting (**??**) into (**??**) results in

$$(2H - I)(2F - I)w_1^* = w_1^* \ ,$$

which is the desired result of $Tw_1^* = w_1^*$. So then $w_1^* = x^* - u^*$ is a fixed point of T.

Q2e: We have a fixed point solution so that $Tw_1^* = w_1^*$. We also know that

$$(2F - I)w_1^* = 2\frac{w_1^* + w_2^*}{2} - w_1 = w_2^* \ . \tag{3}$$

We also know that

$$x^* = \frac{w_1^* + w_2^*}{2} \ .$$

So then we have that

$$x^* = \frac{w_1^* + w_2^*}{2} = \frac{w_1^* + (2F - I)w_1^*}{2} = F(w_1^*) \ .$$

Q2f:

For $\rho \in (0,1)$, the Mann algorithm given by

$$\begin{aligned}&\textbf{Initialize} \quad w \\ &\textbf{Repeat}\{ \\ &\qquad w \leftarrow (1-\rho)w + \rho T w \\ &\}\end{aligned}$$

Since F and H are both firmly non-expansive, then $2F-I$ and $2H-I$ are non-expansive. This then implies that $T=(2F-I)(2H-I)$ is non-expansive. Therefore, if a fixed point exists, then the Mann iterations must converge to a fixed point.

Q2g:

To estimate θ with a set of ground-truth images $\{x_k\}_{k=0}^{K-1}$. Generate training pairs (x_k, z_k) where

$$z_k = x_k + w_k$$

where x_k is a typical image that is expected in the application and w_k is independent whit noise with variance σ^2.

Then the denoising agent $H_\theta(z)$ can be trained to minimize the loss function given by

$$L(\theta) = \sum_k \|x_k - H_\theta(z_k)\|^2 .$$

Name/PUID: ______________________________

Problem 3.(35pt) EM Algorithm

Let (X, Y) denote random objects from the exponential family of distributions denoted by $p_\theta(x, y)$ where $\theta \in \Omega$. Furthermore, let $T(X, Y)$ denote the natural sufficient statistics for the distribution where the maximum likelihood estimate has the form

$$\hat{\theta} = f(T(X, Y)) \ .$$

Also, let $l(\theta) = \log p_\theta(y)$ denote the log likelihood of the observations.

Problem 3a) Give a general formula for the Q-function, $Q(\theta, \theta')$, in the EM update.

Problem 3b) Sketch a figure that illustrates the critical property of the Q-function, $Q(\theta, \theta')$, in relation to the log likelihood function, $l(\theta)$.

Problem 3c) Write a mathematical expression that specifies the critical property shown in your sketch above.

Problem 3d) Specify in pseudo-code the general EM algorithm in terms of the Q-function.

Problem 3e) Specify in pseudo-code the general EM algorithm in terms of the function f.

Solution:

Q3a: A general formulation for Q-function,

$$Q\left(\theta; \theta'\right) = \mathbb{E}\left[\log p(y, X \mid \theta) \mid Y = y, \theta'\right]$$

General EM update:
E-step : $Q\left(\theta; \theta^{(k)}\right) = \mathbb{E}\left[\log(p(y, X \mid \theta)) \mid Y = y, \theta^{(k)}\right]$
M-step : $\theta^{(k+1)} = \arg\max_{\theta \in \Omega} Q\left(\theta; \theta^{(k)}\right)$

Q3b: $Q\left(\theta; \theta'\right)$ is a surrogate function for maximizing the log likelihood function.

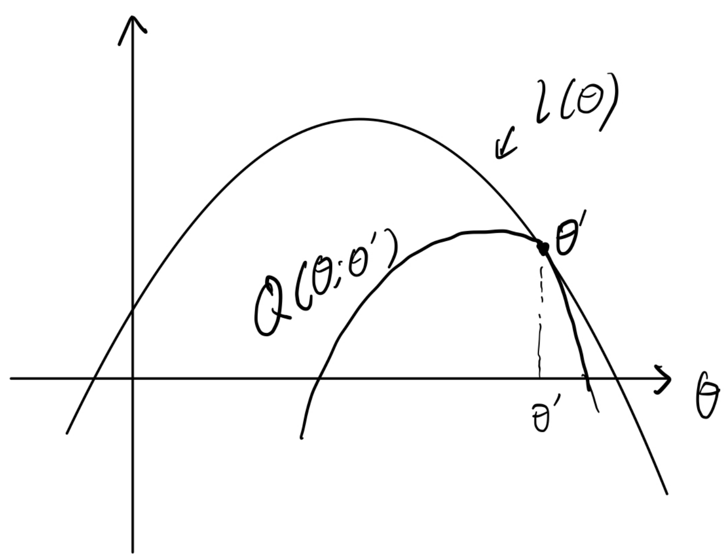

Figure 2: Illustration of surrogate function

Q3c: A mathematical expression that specifies the critical property of surrogate function,

$$l(\theta) \geq Q\left(\theta; \theta'\right) - Q\left(\theta'; \theta'\right) + l\left(\theta'\right)$$

Q3d:

EM Algorithm Pseudocode:

Initialize: $\theta^{(0)}$ and set $k = 0$

Repeat until convergence:

E-step: $Q\left(\theta; \theta^{(k)}\right) = \mathbb{E}\left[\log(p(y, X \mid \theta)) \mid Y = y, \theta^{(k)}\right]$

M-step: $\theta^{(k+1)} = \arg\max_{\theta \in \Omega} Q\left(\theta; \theta^{(k)}\right)$

$k \leftarrow k + 1$

Q3e:

EM Algorithm Pseudocode:

Initialize: $\theta^{(0)}$ and set $k = 0$

Repeat until convergence:

E-step: Compute the expected value of the sufficient statistics

$\bar{T}(X, Y) = \mathbb{E}[T(X, Y)|\theta^{(k)}]$

M-step: $\theta^{(k+1)} = f(\bar{T}(X, Y))$

$k \leftarrow k + 1$

Name/PUID: ____

Problem 4.(42pt) EM Algorithm for Poisson Observations

Let X_n for $n = 1, \cdots, N$ be a series of i.i.d. multinomial random variables with distribution $P\{X_n = m\} = \pi_m$, and let $Y_n \sim \text{Poisson}\{\lambda_m\}$ be conditionally independent random variables given X_n, and let $\theta = \{\pi_0, \lambda_0, \cdots, \pi_{M-1}, \lambda_{M-1}\}$ parameterizes the joint distribution.

Problem 4a) Calculate $p_\theta(x, y)$, an expression for the joint probability density of $\{X_n, Y_n\}_{n=1}^N$.

Problem 4b) Calculate $l(\theta)$, an expression for the negative log likelihood from the measurements $\{X_n, Y_n\}_{n=1}^N$.

Problem 4c) Calculate $\hat{\pi}_m$, the maximum likelihood estimate of π_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 4d) Calculate $\hat{\lambda}_m$, the maximum likelihood estimate of λ_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 4e) Use Bayes' rule to calculate an expression for $f(m|y_n) = P\{X_n = m|Y_n = y_n\}$.

Problem 4f) Specify in pseudo-code the EM algorithm for the estimation of θ for this specific problem.

Solution:

Q4a: We first calculate the joint probability density of each $\{X_n, Y_n\}$ pairs given by

$$p(x_n, y_n) = p(y_n \mid x_n)\, \pi_{x_n} = \frac{\lambda_{x_n}^{y_n} e^{-\lambda_{x_n}}}{y_n!} \pi_{x_n}.$$

Since the X_n are independent and the Y_n are independent, we have that

$$p(x, y) = \prod_{n=1}^{N} \left\{ \frac{\lambda_{x_n}^{y_n} e^{-\lambda_{x_n}}}{y_n!} \pi_{x_n} \right\}$$

Q4b: The negative log likelihood is given by $l(\theta) = -\log p(x, y)$. In order to calculate the negative logarithm of the given probability function $p(x, y)$, we apply the logarithm to the product. The negative logarithm of a product becomes the sum of the negative logarithms of the individual terms.

1. Apply the negative logarithm to the product:

$$-\log p(x, y) = -\log \left(\prod_{n=1}^{N} \frac{\lambda_{x_n}^{y_n} e^{-\lambda_{x_n}}}{y_n!} \pi_{x_n} \right)$$

2. Convert the logarithm of a product into a sum of logarithms:

$$-\log p(x, y) = -\sum_{n=1}^{N} \log \left(\frac{\lambda_{x_n}^{y_n} e^{-\lambda_{x_n}}}{y_n!} \pi_{x_n} \right)$$

3. Apply the logarithm properties to the terms inside the sum:

$$-\log p(x, y) = -\sum_{n=1}^{N} \left(y_n \log \lambda_{x_n} - \lambda_{x_n} - \log(y_n!) + \log \pi_{x_n} \right)$$

This is the negative log likelihood from the measurements.

Q4c: The natural sufficient statistics for θ given (X, Y) are

$$N_m = \sum_{n=1}^{N} \delta(X_n = m)$$
$$b_m = \sum_{n=1}^{N} Y_n \delta(X_n = m)$$

Therefore, the ML estimate of π_m is

$$\hat{\pi}_m = \frac{N_m}{N} \; .$$

Q4d: The natural sufficient statistics for θ given (X, Y) are

$$N_m = \sum_{n=1}^{N} \delta(X_n = m)$$
$$b_m = \sum_{n=1}^{N} Y_n \delta(X_n = m)$$

Therefore, the ML estimate of π_m is

$$\hat{\lambda}_m = \frac{b_m}{N_m} \ .$$

Q4e: The posterior probability calculated by Bayes' rule,

$$\begin{aligned} f(m|y_m, \theta) &= P\{X_n = m \mid Y_n = y_n\} \\ &= \frac{P\{Y_n = y_n \mid X_n = m\} P\{X_n = m\}}{\sum_{m=0}^{M-1} P\{Y_n = y_n \mid X_n = m\} P\{X_n = m\}} \\ &= \frac{\frac{\lambda_m^{y_n} e^{-\lambda_m}}{y_n!}\pi_m}{\sum_{m=0}^{M-1} \frac{\lambda_m^{y_n} e^{-\lambda_m}}{y_n!}\pi_m} \\ &= \frac{\lambda_m^{y_n} e^{-\lambda_m}\pi_m}{\sum_{m=0}^{M-1} \lambda_m^{y_n} e^{-\lambda_m}\pi_m} \end{aligned}$$

Q4f:

Assume we start with initiate parameter estimates of $\hat{\theta} = \{\hat{\lambda}_m, \hat{\pi}_m\}_{m=0}^{M-1}$.

The E-step:

For $n = 1, \cdots, N$ and $m = 0, \cdots, M-1$ calculate the posterior probability

$$f_n(m) = \frac{\hat{\lambda}_m^{y_n} e^{-\hat{\lambda}_m}\hat{\pi}_m}{\sum_{m=0}^{M-1} \hat{\lambda}_m^{y_n} e^{-\hat{\lambda}_m}\hat{\pi}_m}$$

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

$$\hat{N}_m = \sum_{n=1}^{N} f_n(m)$$
$$\hat{b}_m = \sum_{n=1}^{N} Y_n f_n(m)$$

The M-step:

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

$$\hat{\pi}_m = \frac{\hat{N}_m}{N}$$
$$\hat{\lambda}_m = \frac{\hat{b}_m}{\hat{N}_m}$$

And repeat until converged.

Name/PUID: ____________________________________

Problem 5.(49pt) Markov Chains and Stochastic Sampling

Let $u : \Omega \rightarrow \Re$ be an energy function defined on $\Omega = \{0, \cdots M-1\}$ such that

$$Z = \sum_{x \in \Omega} u(x) \ ,$$

and let

$$p(x) = \frac{1}{Z} \exp\{u(x)\} \ ,$$

be its associated Gibbs distribution.

Further define the distribution

$$q(i|j) = \begin{cases} 1/3 & \text{if } 0 = mod_M(i-j) \\ 1/3 & \text{if } 1 = mod_M(i-j) \\ 1/3 & \text{if } M-1 = mod_M(i-j) \end{cases} \ .$$

Also, let Z_n denote the random process generated by $Z_n \sim q(i|Z_{n-1})$.

Problem 5a) Prove that Z_n is a Markov chain.

Problem 5b) Prove that Z_n irreducible.

Problem 5c) Prove that Z_n aperiodic.

Problem 5d) Prove that Z_n is ergodic.

Problem 5e) Find the stationary distribution for Z_n.

Problem 5f) Is Z_n reversible? Justify your answer.

Problem 5g) Specify a pseudo-code algorithm for generating samples from the Gibbs distribution, $p(x)$, using the Metropolis Algorithm.

Solution:

Q5a: Since the probability distribution of Z_n only depends on Z_{n-1}, it is a Markov chain.

Q5b: It is enough to show that for all i, j there is a finite sequence of states, $(z_0, \cdots, z_n)$ with $z_0 = i$ and $z_n = j$ that occurs with probability greater than 0. To do this, observe that just choose n so that $z_0 = i, z_1 = mod_M(i+1), z_1 = mod_M(i+2), \cdot, z_n = j$. Then this sequence occurs with probability $(1/3)^n > 0$.

Q5c: Since $P_{0,0} > \frac{1}{3} > 0$ and the Markov chain is irreducible by part Q2.2 above, then the Markov chain is aperiodic.

Q5d: Since the Markov chain has i) finite state; ii) is irreducible; iii) is aperiodic, then it must be ergodic.

Q5e: Choose $\pi = [1/M, 1/M, \cdots, 1/M]$. Then π_i satisfies the full balance equations given by

$$\pi P = \pi.$$

So this must be the stationary distribution of the ergodic Markov chain.

Q5f: The Markov chain is reversible. In order to prove this we need to show that

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

for all state pairs, i, j.

Case 1: Assume $(i - j) mod_M \in \{-1, 0, 1\}$, then we have that

$$\begin{aligned} \pi_i P_{i,j} &= \pi_j P_{j,i} \\ \frac{1}{M}\frac{1}{3} &= \frac{1}{M}\frac{1}{3} , \end{aligned}$$

so the detailed balance equations hold.

Case 2: Assume $(i - j) mod_M \notin \{-1, 0, 1\}$, then we have that

$$\begin{aligned} \pi_i P_{i,j} &= \pi_j P_{j,i} \\ \frac{1}{M} 0 &= \frac{1}{M} 0 , \end{aligned}$$

so again the detailed balance equations hold.

Since the detailed balance equations always hold, the Markov chain is reversible.

Q5g:

Metropolis Algorithm for Gibbs Sampling:

Initialize: $x_0 \sim \text{Uniform}(\Omega)$

For $n = 0$ **to** $N - 1$ **do:**

1. Sample $\omega \sim q(i|x_n)$
2. Compute acceptance probability: $\alpha = \min\left(1, \exp\{-(u(\omega) - u(x_n))\}\right)$
3. Accept or reject:

   Generate $r \sim \text{Uniform}(0, 1)$

   If $r < \alpha$ then $x_{n+1} = \omega$ else $x_{n+1} = x_n$

End For

ECE641 Fact Sheet

Probability Background

Total Probability
$P(A) = \sum_n P(A|B_n)P(B_n)$

Total Probability for Conditional Probabilities
$P(A|C) = \sum_n P(A|B_n, C)P(B_n|C)$

Bayes' Rule
$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

Conditional Joint Probability
$P(A, B|C) = P(A|B, C)P(B|C)$

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

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

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

Miscellaneous

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

Plug and Play

(non-expansive map)

(CE equations)

$x^* = F(x^* - u^*)$

$x^* = H(x^* + u^*)$

(Douglas-Rachford algorithm)

set $\rho \in (0,1)$

initialize w_1

repeat{

$w_1' \leftarrow T w_1$

$w_1 \leftarrow (1-\rho)w_1' + \rho w_1$

}

return w_1

Note that here $w_1 = x - u$, $w_2 = x + u$, and $x = \frac{w_1+w_2}{2}$, so then $(2F - I)w_1 = w_2$. And, $T = (2H - I)(2F - I)$.

(Convergence of Douglas-Rachford algorithm)

When F and H are proximal maps of proper closed convex functions f and h then Douglas-Rachford algorithm converges to both the CE solution and the MAP estimate.

EM algorithm

General EM Algorithm:
E-step : $Q(\theta;\theta^{(k)}) = \mathbb{E}[\log(p(y,X|\theta))|Y=y,\theta^{(k)}]$
M-step : $\theta^{(k+1)} = \arg\max_{\theta\in\Omega} Q(\theta;\theta^{(k)})$
(ML estimate for Gaussian mixture)
$\log p(y,x|\theta) = \Sigma_{n=1}^{N}\log p(y_n,x_n|\theta) = \Sigma_{n=1}^{N}\Sigma_{m=0}^{M-1}\delta(x_n - m)\{\log p(y_n|\mu_m,\sigma_m) + \log\pi_m\}$

(Exponential Family)
A family of density functions $p_\theta(y)$ for y and θ is said to be a exponential family if there exists functions $\eta(\theta)$, $s(y)$, and $d(\theta)$ and natural statistic $T(y)$ such that $p_\theta(y) = exp\{\langle\eta(\theta),T(y)\rangle + d(\theta) + s(y)\}$

(sufficient statistic)
$T(Y)$ is a sufficient statistic for the family of distributions $p_\theta(y)$ if the density functions can be written in the form $p_\theta(y) = h(y)g(T(y),\theta)$ where g and h are any two functions.

Markov Chains

Parameter Estimation for Markov Chains: $N_j = \delta(X_0 - j)$, $K_{i,j} = \Sigma_{n=1}^{N}\delta(X_n - j)\delta(X_{n-1} - i)$
$\log(p(x)) = \sum_{j\in\Omega}\{N_j\log(\tau_j) + \sum_{i\in\Omega}K_{i,j}\log(P_{i,j})\}$
ML Estimate $\hat{\tau_j} = N_j$ and $\hat{P_{i,j}} = \frac{K_{i,j}}{\sum_{j\in\Omega}K_{i,j}}$
Marginal density at any time n: $\pi^{(n)} = \pi^{(0)}P^n$ and $\pi^{(\infty)} = \pi^{(0)}P^\infty$
Log likelihood of HMM (MAP Estimate):
$\hat{x} = \arg\max_{x\in\Omega^N}\{\log\tau_{x_0} + \Sigma_{n=1}^{N}\{\log f(y_n|x_n) + \log P_{x_{n-1},x_n}\}\}$
State Sequence Estimation and Dynamic Programming:
$L(j,n) = \max_{x_{>n}}\{\log p(y_{>n},x_{>n}|x_n = j)\}$ and $L(j,N) = 0$
$L(i,n-1) = \max_{j\in\Omega}\{\log f(y_n|j) + \log P_{i,j} + L(j,n)\}$
$\hat{x_0} = \arg\max_{j\in\Omega}\{\log\tau_j + L(j,0)\}$
$\hat{x_n} = \arg\max_{j\in\Omega}\{\log P_{\hat{x_{n-1}},j} + \log f(y_n|j) + L(j,n)\}$
State Probability and the Forward-Backward Algorithm:
$\alpha_n(j) = p(x_n = j, y_n, y_{<n})$ $\beta_n(j) = p(y_{>n}|x_n = j)$
$p(x_{n-1} = i, x_n = j|y) = \frac{\alpha_{n-1}(i)P_{i,j}f(y_n|j)\beta_n(j)}{p(y)}$
$\alpha_n(j) = \sum_{i\in\Omega}\alpha_{n-1}(i)P_{i,j}f(y_n|j)$
$\beta_n(i) = \sum_{j\in\Omega}P_{i,j}f(y_{n+1}|j)\beta_{n+1}(j)$

(Irreducible Markov Chain). A discrete-time, discrete-space homogeneous Markov chain is said to be irreducible if for all states $i,j\in\Omega$, i and j communicate.

(Communicating States). States $i,j\in\Omega$ of a discrete-time, discrete-space homogeneous Markov chain are said to communicate if there exists integers $m > 0$ and $n > 0$ such that $[P^m]_{i,j} > 0$ and $[P^n]_{j,i} > 0$.

(period of state) State $i\in\Omega$ of a discrete-time, discrete-space homogeneous Markov chain has period $d(i) = gcd\{n\in\mathbb{N}_+|[P^n]_{i,i} > 0\}$.
State i is aperiodic if $d(i) = 1$ and periodic if $d(i) > 1$.
Ergodic MC : $\pi_j = lim_{n\to\infty}[P^n]_{i,j} > 0$

(detailed balance equations)
$\pi_i P_{i,j} = \pi_j P_{j,i}$
$\sum_{i\in\Omega}\pi_i = 1$

(full balance equations)
$\pi^\infty = \pi^\infty P$ or $\pi_j = \sum_{i\in\Omega}\pi_i P_{i,j}$
$\sum_{i\in\Omega}\pi_i = 1$

Stochastic Sampling

(inverse transform sampling)
$X \leftarrow F^{-1}(U)$ where $U \leftarrow Rand([0,1])$ and $F^{-1}(u) = \inf\{x|F(x) \geq u\}$ generates a sample from random variable X with CDF $F(x) = P\{X \leq x\}$

(Metropolis algorithm)

initialize X^0

for k from 0 to $K-1\{$

$U \leftarrow Rand([0,1])$

$W \leftarrow Q^{-1}(U|X^{(k)})$

$\alpha \leftarrow \min\{1, e^{-[u(W)-u(X^{(k)})]}\}$

$U \leftarrow Rand([0,1])$

if $U < \alpha$ then $X^{(k+1)} \leftarrow W$

else $X^{(k+1)} \leftarrow X^{(k)}$

$\}$

Note: where $Q^{-1}(\cdot|x^{(k)})$ is the inverse CDF corresponding to proposal density $q(w|x^{(k)})$