PURDUE

ECE 64100

Final Exam, December 9, Fall 2024

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: <u>Key</u>

Problem 0.(6pt) Your name

Problem 1.(48pt) Maximum Likelihood Estimate

Let $X_1, \ldots X_N$ be i.i.d. samples from the family of distributions $P_\theta\{X_n = m\} = \theta_m$ for $m = 0, \ldots, M-1$ and $\theta \in \Omega$ such that

$$\Omega = \{\theta \in \mathbb{R}^M : \theta_m \geq 0 \text{ and } \sum_{m=0}^{M-1} \theta_m = 1.0\} \ .$$

Furthermore, define

$$N_m = \sum_{n=1}^{N} \delta(x_n = m) \ ,$$

and in order to make things easier, assume the technical condition that that for all m, $N_m \neq 0$.

1a) Prove that Ω is a convex set.

1b) Is Ω an open or closed set? Justify your answer.

1c) Prove that Ω a bounded set.

1d) What name is given to the set Ω. (Hint: Starts with an "s".)

1e) Derive an expression for the negative log likelihood, $l(\theta) = -\log p_\theta(x)$ where $x = [x_1, \ldots x_N]$.

1f) Prove that $l(\theta)$ is a strictly convex function.

1g) Is the maximum likelihood estimate unique? Justify your answer.

1h) Derive an expression for the maximum likelihood estimate of θ given $X = x$.

Solution:

Q1a: Let $x, y \in \Omega$ and let $\lambda \in [0,1]$. Then let $z = \lambda x + (1-\lambda) y$, then

$$z_n = \lambda x_n + (1-\lambda) y_n \in [0,1] \ ,$$

and also

$$\begin{aligned} \sum_{n=0}^{M-1} z_n &= \sum_{n=0}^{M-1} (\lambda x_n + (1-\lambda) y_n) && (1) \\ &= \lambda \left(\sum_{n=0}^{M-1} x_n \right) + (1-\lambda) \sum_{n=0}^{M-1} y_n && (2) \\ &= \lambda 1 + (1-\lambda) 1 && (3) \\ &= 1 \ . && (4) \end{aligned}$$

So therefore, $z \in \Omega$.

Q1b: Ω is a closed set because the boundaries of the set are included in Ω.

Q1c: In order to show that the set is bounded, we need to show there there exists a $T \in \mathbb{R}$ such that for all $x \in \Omega$, $\|x\| < T$.
Select $T = M + 1$, then since each component of $x_n \leq 1$, we have that $\|x\| \leq M < T$. So therefore, Ω is bounded.

Q1d: The set Ω is know as the "simplex" set.

Q1e:

$$\begin{aligned} l(\theta) &= -\log P_\theta\{X = x\} && (5) \\ &= -\log \prod_{n=1}^{N} \theta_{x_n} && (6) \\ &= -\log \prod_{n=0}^{M-1} \theta_m^{N_m} && (7) \\ &= -\sum_{n=0}^{M-1} N_m \log \theta_m \ , && (8) \end{aligned}$$

where

$$N_m = \sum_{n=1}^{N} \delta(x_n = m) \ .$$

Q1f: $l(\theta)$ must be strictly convex because it is a sum of strictly convex function in each variable. Note that $N_m \neq 0$, so every term of the sum is strictly convex.

Q1g: Yes, because the function being minimized is a strictly convex function on a convex set.

Q1h:

$$\hat{\theta}_m = \frac{N_m}{N} \ .$$

Name/PUID: ______________________________

Problem 2.(60pt) Plug-and-Play for Poisson Observations

Let $Y \sim \text{Pois}(X)$ where $X \in \mathbb{R}^{+N}$ and $X \sim p(x)$. [1]

Furthermore, let $\hat{H}(x)$ be a MMSE denoiser that is designed to remove additive white Gaussian noise with variance σ^2.

2a) Derive an expression for the conditional probability $p(y|x)$.

2b) Derive an expression for the negative log likelihood of $f(x) = -\log p(y|x)$.

2c) Give an explicit expression for the forward model proximal map

2d) Give an explicit expression for the prior model proximal map.

2e) What alternate interpretation can be given for H? Be specific.

2f) Are H and $\hat{H}$ the same? How do they differ?

2g) Describe how $\hat{H}$ can be designed using training data?

2h) Specify the Plug-and-Play (PnP) algorithm in terms of forward model proximal map, F, and the denoiser $\hat{H}$.

2i) Specify conditions on $\hat{H}$ that guarantee convergence of the PnP algorithm?

2j) If the PnP algorithm converges, then what conditions hold at convergence?

[1] Here the notation $Y \sim \text{Pois}(x)$ means each component of Y_n is i.i.d. with a Poisson distribution of mean x_n.

Solution:

Q2a:

$$p(y|x) = \prod_{n=1}^{N} \frac{x_n^{y_n} e^{-x_n}}{y_n!} \tag{9}$$

Q2b:

$$\begin{aligned}
f(x) &= -\log p(y|x) && (10)\\
&= -\sum_{n=1}^{N} \{y_n \log x_n - x_n - \log y_n!\} && (11)\\
&= \sum_{n=1}^{N} \{x_n - y_n \log x_n + \log y_n!\} && (12)
\end{aligned}$$

Q2c:

$$\begin{aligned}
F(z) &= \arg\min_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - z\|^2 \right\} && (13)\\
&= \arg\min_x \left\{ \sum_{n=1}^{N} \{x_n - y_n \log x_n + \log y_n!\} + \frac{1}{2\sigma^2}\|x - z\|^2 \right\} && (14)\\
&= \arg\min_x \left\{ \sum_{n=1}^{N} \{x_n - y_n \log x_n\} + \frac{1}{2\sigma^2}\|x - z\|^2 \right\} && (15)\\
& && (16)
\end{aligned}$$

Q2d:

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

Q2e: If we rewrite the proximal map as

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

then we can observe that the first term can be interpreted as $-\log p(v|x)$ where

$$p(v|x) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left\{ -\frac{1}{2\sigma^2}\|v - x\|^2 \right\} ,$$

So H can be interpreted as the MAP estimate of X given V where

$$V = X + W ,$$

where $W \sim N(0, \sigma^2 I)$ and $X \sim p(x)$.

Q2f: No, H is the MAP denoiser, but $\hat{H}$ is the approximate MMSE denoiser created from training data.

Q2g: Take a set of example training images, $x^{(k)}$ for $k = \{0, \ldots, K-1\}$. Then for each k, add independent white Gaussian noise.

$$v^{(k)} = x^{(k)} + \sigma W \ ,$$

where $W \sim N(0, I)$.

Then fit an estimator $H_\theta(x)$ to the training data via the optimization

$$\hat{\theta} = \arg\min_\theta \left\{ \sum_{k=0}^{K-1} \|v^{(k)} - x^{(k)}\|^2 \right\} \ .$$

Then use $\hat{H} = H_{\hat{\theta}}$ as the approximate MMSE denoiser.

Q2h: The algorithm is given by

$$
\begin{aligned}
&\textbf{Initialize} \quad v, u \\
&\textbf{Repeat}\{ \\
&\qquad x \leftarrow F(v - u) \\
&\qquad v \leftarrow \hat{H}(x + u) \\
&\qquad u \leftarrow u + (x - v) \\
&\}
\end{aligned}
$$

Q2i: Any of the following are sufficient answers:

1) A sufficient condition for convergence is that $T = (2F - I)(2\hat{H} - I)$ is non-expansive.

2) This is in turn guaranteed by if $\hat{H}$ is firmly non-expansive.

Q2j: Let x^* and u^* denote the converged values, then at convergence the following equilibrium holds.

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

Name/PUID: ________________________________

Problem 3.(42pt) EM Algorithm for Exponential 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 $Y_n \sim \frac{1}{\lambda_m} e^{-y/\lambda_m}$ be conditionally independent random variables given $X_n = m$, and let $\theta = \{\pi_0, \lambda_0, \cdots, \pi_{M-1}, \lambda_{M-1}\}$ parameterize the joint distribution.

Problem 3a) Calculate $p_\theta(x, y)$, an expression for the joint probability density of $\{X_n, Y_n\}_{n=1}^N$.

Problem 3b) Calculate $l(\theta)$, an expression for the negative log likelihood from the measurements $\{X_n, Y_n\}_{n=1}^N$.

Problem 3c) Calculate $\hat{\pi}_m$, the maximum likelihood estimate of π_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 3d) Calculate $\hat{\lambda}_m$, the maximum likelihood estimate of λ_m given $\{X_n, Y_n\}_{n=1}^N$.

Problem 3e) Use Bayes' rule to calculate an expression for $f(m|y_n) = P\{X_n = m | Y_n = y_n\}$.

Problem 3f) Specify the E-step of the EM algorithm for the estimation of θ for this specific problem.

Problem 3g) Specify the M-step of the EM algorithm for the estimation of θ for this specific problem.

Solution:

Q3a: We first calculate the joint probability density of each $\{X_n, Y_n\}$ pairs given by

$$p(x_n, y_n) = p(y_n \mid x_n)\,\pi_{x_n} = \frac{1}{\lambda_{x_n}} \exp\{-y_n/\lambda_{x_n}\}\pi_{x_n}.$$

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

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

Q3b: 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:

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

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

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

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

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

This is the negative log likelihood from the measurements.

Q3c: 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

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

Q3d: The natural sufficient statistics for θ given (X, Y) are

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

Therefore, the ML estimate of π_m is

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

Q3e: The posterior probability calculated by Bayes' rule,

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

Q3f:

The E-step:

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

$$f_n(m) = \frac{\frac{1}{\lambda_m} e^{-y_n/\lambda_m} \pi_m}{\sum_{m=0}^{M-1} \frac{1}{\lambda_m} e^{-y_n/\lambda_m} \pi_m}$$

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

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

Q3g:

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 4.(48pt) Reversible Markov Chains

Let $W_n \in \{0, \ldots, M-1\}$ be a homogeneous Markov chain with transition probabilities given by

$$Q_{i,j} = \begin{cases} 1/3 & \text{if } |(i-j)\text{mod}_M| \leq 1 \\ 0 & \text{otherwise} \end{cases} .$$

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

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

and let $p(x) = \frac{1}{Z}\exp\{-u(x)\}$ be the associated Gibbs distribution.

Problem 4a) Does Q represent an irreducible Markov chain? Prove your answer.

Problem 4b) Does Q represent an aperiodic Markov chain? Prove your answer.

Problem 4c) Does Q represent an ergodic Markov chain? Prove your answer.

Problem 4d) Does Q represent a reversible Markov chain? Prove your answer.

Problem 4e) What is the stationary distribution for Q? Prove your answer.

Problem 4f) Specify the Metropolis algorithm for generating a Markov chain, X_n, that samples from the distribution $p(x)$ using proposal distribution $Q_{i,j}$.

Problem 4g) Prove that the Markov chain X_n is irreducible and aperiodic.

Problem 4h) Prove that the Markov chain X_n satisfies the detailed balance equations with the stationary distribution $p(x)$.

Solution:

Q4a: Yes, since after M iterations, it is possible to get between any two states.

Q4b: Yes, since the MC is irreducible, all the states communicate. Given any state, $Q_{i,i} = 1/3 > 0$, so no state can be periodic.

Q4c: Yes, since the MC is irreducible, aperiodic, and finite, it must be ergodic.

Q4d: Choose $\pi = [1/M, 1/M, \cdots, 1/M]$. Then if $|i - j| \leq 1$ we have that

$$\begin{aligned} p_i Q_{i,j} &= \frac{1}{M} Q_{i,j} & (17) \\ &= \frac{1}{M} \frac{1}{3} & (18) \\ &= \frac{1}{M} Q_{j,i} & (19) \\ &= p_j Q_{j,i} \ . & (20) \end{aligned}$$

If $|i - j| > 1$, then $Q_{i,j} = Q_{j,i} = 0$, so we also have that

$$p_i Q_{i,j} = 0 = p_j Q_{j,i} \ . \tag{21}$$

So the detailed balance equations hold. So the MR is reversible.

Q4e: 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.

Q4f: The algorithm is given by

Initialize x

Repeat$\{$

$W \sim p(j) = Q_{x,j}$

$\Delta U \leftarrow u(W) - u(x)$

$\alpha \leftarrow \min\{1, \exp(-\Delta U)\}$

With probability α

$x \leftarrow W$

$\}$

Q4g: Again, it is always possible to get between any two states in M steps. So the MC X_n is irreducible.
In a single step, it is always possible to remain in the same state, so it must be aperiodic.

Q4h: 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.
First notice that $\forall i, j$, $Q_{i,j} = Q_{j,i}$.
Also define the acceptance probability as

$$\alpha(i,j) = \min\{1, \exp(-[u(j) - u(i)])\} \ .$$

Then let $\pi_i = \frac{1}{Z}\exp\{-u(i)\}$; assume that $i \neq j$; and without loss of generality, that $u(j) \geq u(i)$. Then we have that

$$\begin{aligned}
\pi_i P_{i,j} &= \frac{1}{Z}\exp\{-u(i)\}\left[\alpha(i,j)Q_{i,j}\right] && (22)\\
&= \frac{1}{Z}\exp\{-u(i)\}\left[\exp\{-[u(j) - u(i)]\}Q_{i,j}\right] && (23)\\
&= \frac{1}{Z}\exp\{-u(j)\}Q_{i,j} && (24)\\
&= \frac{1}{Z}\exp\{-u(j)\}\alpha(j,i)Q_{j,i} && (25)\\
&= \pi_j P_{j,i} \ . && (26)
\end{aligned}$$

Also, for $i = j$, the detail balance equation is trivially true.
So we see that $\pi_i = \frac{1}{Z}\exp\{-u(i)\}$ solves the detailed balance equations.

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 $q(w|x^{(k)})$