Student Name: Guangyu Lin

Collaboration Statement:

Total hours spent: 2 days

I discussed ideas with these individuals:

- Office Hour with TA
- TODO
- ...

I consulted the following resources:

- bishop test book
- TODO
- ...

By submitting this assignment, I affirm this is my own original work that abides by the course collaboration policy.

Links: [HW3 instructions] [collab. policy]

Contents

1a: Problem Statement

Define $\Sigma = LL^T$. Show the following:

$$|\det(L^{-1})| = \frac{1}{(\det\Sigma)^{\frac{1}{2}}} \tag{1}$$

1a: Solution

Since we know $\Sigma = LL^T$, we can substitute into the equation and we can get the left side as $\frac{1}{(detLL^T)^{\frac{1}{2}}}$ Then by hint v, the left side can be rewrite as $\frac{1}{((detL)(detL^T))^{\frac{1}{2}}}$ By hint vi we know $det(L) = det(L^T)$ so the equation becomes $\frac{1}{(detL)^{2*\frac{1}{2}}}$ which is just $\frac{1}{|(det(L))|}$. Now by hint iv we can say $|det(L^{-1})| = \frac{1}{|det(L)|}$.

1b: Problem Statement

Show that the pdf of x is given by:

$$p(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{(\det\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \tag{2}$$

1b: Solution

The general function is $p(x) = f(S(x))|det(J_s(x))|$ where $S(x) = L^{-1}(x-m)$, $f(u) = (2\pi)^{-\frac{D}{2}}e^{-\frac{1}{2}u^Tu}$ and $J_s(x) = L^{-1}$
it turns out $p(x) = (2\pi)^{-\frac{D}{2}}e^{-\frac{1}{2}(L^{-1}(x-m))^T(L^{-1}(x-m))} * |det(L^{-1})|$
Now we can simplify this term $(L^{-1}(x-m))^T(L^{-1}(x-m))$
By hint vii we know $(L^{-1}(x-m))^T = (x-m)^T(L^{-1})^T$ therefore we can rewrite the previous term as $(x-m)^T(L^{-1})^T * L^{-1}(x-m)$
By hint ii we know $(L^{-1})^T = (L^T)^{-1}$ so we have $(x-m)^T(L^T)^{-1} * L^{-1}(x-m)$
By hint iii we know $(LL^T)^{-1} = (L^T)^{-1} * L^{-1}$ so we can write the term as $(x-m)^T(LL^T)^{-1}(x-m)$
we defined $\Sigma = LL^T$ so the term becomes $(x-m)T\Sigma^{-1}(x-m)$
and the function becomes $p(x) = (2\pi)^{-\frac{D}{2}}e^{-\frac{1}{2}(x-m)^T\Sigma^{-1}(x-m)} * |det(L^{-1})|$

from 1a we know $|\det(L^{-1})| = \frac{1}{(\text{det}\Sigma)^{\frac{1}{2}}}$ therefore $p(x) = (2\pi)^{-\frac{D}{2}} e^{-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)} * \frac{1}{det(\Sigma)^{\frac{1}{2}}}$
rerange the formula and we can get $p(x) = \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{det(\Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)}$

1c: Problem Statement

Complete the Python code below, to show how to turn samples from a standard 1D Gaussian, via NumPy's 'randn()' into a sample from a multivariate Gaussian.x

1c: Solution

```
import numpy as np

def sample_from_mv_gaussian(mu_D, Sigma_DD, random_state=np.random):
  ''' Draw sample from multivariate Gaussian

  Args
  ----
  mu_D : 1D array, size D
    Mean vector
  Sigma_DD : 2D array, shape (D, D)
    Covariance matrix. Must be symmetric and positive definite.

  Returns
  -------
  x_D : 1D array, size D
    Sampled value of Gaussian with provided mean and covariance
  '''
  D = mu_D.size
  L_DD = np.linalg.cholesky(Sigma_DD) # compute L from Sigma
  # GOAL: draw each entry of u_D from standard Gaussian
  u_D = random_state.randn(D) # TODO FIXME use random_state.randn(...)
  # GOAL: Want x_D ~ Gaussian(mean = m_D, covar=Sigma_DD)
  x_D = L_DD @ u_D + m_D # TODO FIXME transform u_D into x_D
  return x_D
```

2a: Problem Statement

Show that the Metropolis-Hastings transition distribution $\mathcal{T}$ satisfies detailed balance with respect to the target distribution p^*.

That is, show that:

$$p^*(a)\mathcal{T}(b|a) = p^*(b)\mathcal{T}(a|b) \tag{3}$$

for all possible $a \neq b$, where a, b are any two distinct values of the random variable.

2a: Solution

To show $p^*(a)\mathcal{T}(b|a) = p^*(b)\mathcal{T}(a|b)$ we can show whether $\frac{p^*(a)\mathcal{T}(b|a)}{p^*(b)\mathcal{T}(a|b)}$ equal to 1 or not.

Now we can substite$\mathcal{T}$ by definition and we can get $\frac{const \cdot \tilde{p}(a)Q(b|a)min(1, \frac{\tilde{b}Q(a|b)}{\tilde{p}(a)Q(b|a)}}{const \cdot \tilde{p}(b)Q(a|b)min(1, \frac{\tilde{a}Q(b|a)}{\tilde{p}(b)Q(a|b)}}$

we know a, b are two different random variables that drawn from the same distribution, so we can just cancel the constant term in the fraction and we have $\frac{\tilde{p}(a)Q(b|a)min(1,\frac{\tilde{p}(b)Q(a|b)}{\tilde{p}(a)Q(b|a)})}{\tilde{p}(b)Q(a|b)min(1,\frac{\tilde{p}(a)Q(b|a)}{\tilde{p}(b)Q(a|b)})}$ now we can rewrite this term as $\frac{\tilde{p}(a)Q(b|a)}{\tilde{p}(b)Q(a|b)} \cdot \frac{min(1,\frac{\tilde{p}(b)Q(a|b)}{\tilde{p}(a)Q(b|a)})}{min(1,\frac{\tilde{p}(a)Q(b|a)}{\tilde{p}(b)Q(a|b)})}$ by the hint we know $\frac{min(1,\frac{\tilde{p}(b)Q(a|b)}{\tilde{p}(a)Q(b|a)})}{min(1,\frac{\tilde{p}(a)Q(b|a)}{\tilde{p}(b)Q(a|b)})}$ is just $\frac{\tilde{p}(b)Q(a|b)}{\tilde{p}(a)Q(b|a)}$ now the term becomes $\frac{\tilde{p}(a)Q(b|a)}{\tilde{p}(b)Q(a|b)} \cdot \frac{\tilde{p}(b)Q(a|b)}{\tilde{p}(a)Q(b|a)}$ and turns out it is just 1. So we can say $p^*(a)\mathcal{T}(b|a) = p^*(b)\mathcal{T}(a|b)$.

3a: Problem Statement

(See diagram on 3a Instructions web page)

You start at Medford/Tufts station, and take 1000 steps. What is your probability distribution over ending this journey at each of the 7 stations? Report as a vector (use order of nodes in the diagram, small to large). Round to 3 decimal places.

3a: Solution

```
x = np.array([1,0,0,0,0,0,0])
p = np.array([[0,1,0,0,0,0,0], #probability of each station towards the next
              [0.5,0,0.5,0,0,0,0],
              [0,0.5,0,0.5,0,0,0],
              [0,0,0.5,0,0.5,0,0],
              [0,0,0,0.5,0,0,0.5],
              [0,0,0,0,0,0,1],
              [0,0,0,0,0.5,0.5,0]])
for i in range(1000): # take 1000 steps
    x = x @ p

print(np.round(x,3))
```

The result is [0.167 0.000 0.333 0.000 0.333 0.167 0.000]

3b: Problem Statement

Is there a unique stationary distribution for this Markov chain? If so, explain why. If not, explain why not.

3b: Solution

There is no unique stationary distribution for this Markov chain since it is not ergodic which needs to satisfy both irreducible and aperiodic. By looking at the probability distribution of each station, we see some station has 0 probability to reach and it violates the definition that in every time step we can reach all stations. For example, taking 2000 or 3333 steps will generated completely different results. After T_0, there are at least one state that $p(k|i) = 0$ where i is the initial state and k means the state after T_0, so this is not a unique stationary distribution.