ECE C143A/C243A, Spring 2021 **Final**

Department of Electrical and Computer Engineering Prof. J.C. Kao

University of California, Los Angeles TAs: T. Monsoor, S. Balla

UCLA True Bruin academic integrity principles apply.

Open: Book, Notes, PDF, CCLE, Piazza, Calculator

Closed: Other internet, communication, carrier pigeon, or other collaboration.

3:00pm-6:00pm PDT.

Tuesday, June 8, 2021.

State your assumptions and reasoning.

No credit without reasoning.

Show all work.

Name: Ziyi Yang

Signature: Ziyi Yang 杨子羿

ID#: 405064363

Problem 1	____	/ 30
Problem 2	____	/ 35
Problem 3	____	/ 20
Problem 4	____	/ 35
BONUS	____	/ 10 bonus points
Total	____	/ 120 points + 10 bonus points

1. **Topics across the entire class.** (30 points)

   (a) (5 points) True / False questions. Each question is 1 point. There is no partial credit. No justification is required for these T/F questions. If you are showing your work on a printed exam, you can circle one of **True** or **False**.

      i. Action potentials are generated at the axon hillock, which has a high density of voltage-gated sodium ion channels.

      True **False**

      ii. Consider a neuron at a resting potential of $V_r = -65$ mV, with more Na^+ ions outside the cell than inside. If a sodium-selective ion channel was opened, then the drift current (electric field force) and diffusion current (chemical equilibrium) would both drive sodium inside the cell.

      True **False**

      out + Na^+ more

      in − E chem

      iii. Myelin sheaths lead to an increase in membrane capacitance.

      True **False**

      iv. Consider recording from a Utah electrode array. It is possible for one electrode to record from more than one neuron at once.

      True **False**

      v. Planned – but not executed – reaches during a delayed reach task are more strongly represented in the primary motor cortex (M1) than in the dorsal premotor cortex (PMd).

      True **False**

(b) (13 points) Poisson processes. Consider a homogeneous Poisson process with rate $\lambda =$ 10 spikes/min.

i. (2 points) You begin observing the Poisson process. What is the probability that you have to wait at least 1 minute to see the first spike? You may leave your answer as an expression (including terms like exp), you do not have to compute the numerical value of this expression.

since this is a homogeneous Poisson process, the interspike intervals are exponentially distributed, with $t_i \sim \exp(\lambda)$ CDF: $F_T(t) = 1 - e^{-\lambda t}$

$Pr(\text{at least 1 min to see first spike}) = Pr(t_1 \geq 1)$

$= 1 - Pr(t_1 < 1)$

$= 1 - F_T(1) = e^{-10 \cdot 1} = e^{-10}$

ii. (2 points) What is the probability that you observe 5 spikes in 1 minute? You may leave your answer as an expression (including terms like exp), you do not have to compute the numerical value of this expression.

Say we observe 5 spikes in $[t, t+1]$ $\quad Pr(N(s)=n) = \dfrac{(\lambda s)^n e^{-\lambda s}}{n!}$

by restart property of homog Poisson process,

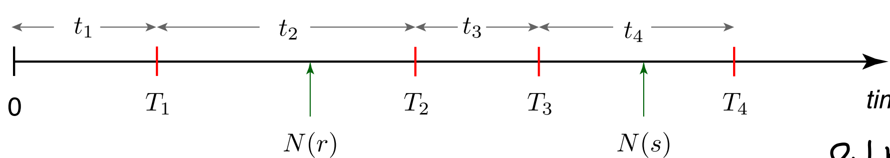
$N(t+1) - N(t) \sim \text{Poisson}(\lambda \cdot 1) \Rightarrow \text{Poisson}(10)$

so $Pr(N(t+1) - N(t) = 5) = \dfrac{10^5 \cdot e^{-10}}{5!}$

iii. (3 points) Say that you began an experiment observing this neuron. In the first minute, you observed 5 spikes. Your colleague says "Well, on average this neuron fires 10 spikes per minute, so this must influence the number of spikes we see in the next minute. We should expect to see something closer to 15 spikes in the next minute, so its rate averages out to around 10 spikes/minute. Therefore, we must have that $\mathbf{Pr}(N(2) - N(1) = 15 | N(1) = 5) > \mathbf{Pr}(N(2) - N(1) = 15)$." Is your colleague correct or incorrect? Justify your answer in no more than 3 sentences.

My colleague is incorrect. since by restart property of homog PP,
$N(2) - N(1) \sim \text{Poisson}(\lambda \cdot 1)$ and is independent of $N(1)$.
The expected value of $N(2) - N(1)$ would still be 10 instead of 15.

iv. (6 points) Consider the Poisson process timeline shown below. Recall the rate of the Poisson process is $\lambda = 10$ spikes/min. Write the distributions of t_3, T_3, and $N(s) - N(r)$. When you write the distributions, you must include the parameters of the distribution (e.g., it is not enough to say a distribution is Poisson, you must also specify the parameter (mean) of the Poisson distribution).



0.1 min per spike
on average

$t_3 \sim \exp(0.1)$ t_3 is exponentially distributed with a mean of 0.1

$T_3 \sim \text{Poisson}(10 \cdot (t_1 + t_2 + t_3))$ T_3 is a homog PP with mean of
$10 \cdot (t_1 + t_2 + t_3)$

$N(s) - N(r) \sim \text{Poisson}(10 \cdot (s - r))$ is a homog PP with mean of
$10 \cdot (s - r)$

* in class we write $t_i \sim \exp(\lambda) \Rightarrow \exp(10)$ in this case
but since t_i is ISI, the mean should be a quantity of time,
so here I write it as $t_i \sim \exp(\frac{1}{\lambda}) \Rightarrow \exp(0.1)$

(c) (12 points) Short answer on decoding.

i. (4 points) When we showed a video of the Optimal Linear Estimator (OLE) decoder being controlled by a monkey, we observed that OLE decoder had noisy jittery movements. Give an explanation for why the OLE decoder had noisy jittery movements. Explain in no more than 4 sentences.

OLE decoder only uses the neural data at present time to predict the kinematics, which contains a lot of noise on its own. Also there are not enough parameters so the model could be underfitted.

ii. (4 points) A Wiener filter decodes neural data over a set history of P bins. In real time brain-computer interface control, what is one con of setting P to be too large? Explain in no more than 4 sentences.

Setting a very large P could increase the number of parameters by a lot, which increases the chance of overfitting. For a real-time system, this could also increase the amount of computation and slow down the system. And in reality, a movement should not depend on a neuron firing too long ago.

iii. (4 points) In HW #4, we asked you to implement a discrete decoder where the neural data (conditioned on the class) was modeled by a multivariate Gaussian distribution with class specific covariance or a Poisson distribution (naive Bayes). In class, we said that the Poisson distribution better modeled spike counts. Assume that for your dataset, when looking at any single neuron, its spike count distribution is better described by a Poisson distribution than a Gaussian distribution. Even in this setting, it's possible for a multivariate Gaussian generative model to outperform a Poisson generative model. Give one reason why this is possible. Explain in no more than 4 sentences.

As we calculate in HW4, a Poisson generative model has a linear decision boundary whereas a multivariate Gaussian model has quadratic decision boundary. This non-linearity allows the latter to fit more complex dataset and thus can have better performance.

2. **Discrete decoding using exponential family** (35 points)

A probabilistic generative model for classification comprises class-conditional densities $P(\mathbf{y}|C_k)$ and class priors $P(C_k)$, where $\mathbf{y} \in \mathbb{R}^D$ denotes the firing rates of D neurons and $k = 1, 2, \ldots, K$ denotes the class of the reach the monkey was performing (e.g., left, right, up, down, etc.). In homework 4, we derived the maximum likelihood parameter estimates for the gaussian class-specific and shared covariance model. In this problem, we will set up the optimization problems for finding the parameters for a more general model known as the exponential family of distribution. In this model, the class-conditional densities are given by

$$P(\mathbf{y}|C_k) = h(\mathbf{y})g(\lambda_k)\exp\left(\lambda_k^T \mathbf{y}\right)$$

where $h(\cdot)$ and $g(\cdot)$ are known deterministic functions. Like before, we assume the following class priors

$$P(C_k) = \pi_k, \quad k = 1, 2, \cdots, K$$

(a) (10 points) Assume we have N pairs of independent training samples

$$(\mathbf{y}_1, t_1), (\mathbf{y}_2, t_2), \cdots, (\mathbf{y}_N, t_N)$$

where $\mathbf{y}_i \in \mathbb{R}^D$ and $t_i \in \{1, 2, \cdots, K\}$. Write the log-likelihood of observing the training data under the exponential class conditional distribution.

$$\begin{aligned}
L(\theta) &= P((y_1, t_1), (y_2, t_2), \ldots, (y_N, t_N) | \theta) \\
&= \prod_{i=1}^{N} P(y_i, t_i | \theta) \quad \text{by indep} \\
&= \prod_{i=1}^{N} P(t_i | \theta) P(y_i | t_i, \theta) \\
&= \prod_{k=1}^{K} \prod_{i \in C_k} P(C_k) P(y_i | C_k) \\
&= \prod_{k=1}^{K} \prod_{i \in C_k} \pi_k h(y_i) g(\lambda_k) e^{\lambda_k^T y_i}
\end{aligned}$$

$$\log L(\theta) = \sum_{k=1}^{K} \sum_{i \in C_k} \log \pi_k + \log h(y_i) + \log g(\lambda_k) + \lambda_k^T y_i$$

(Additional space for 2a)

(b) (5 points) Based on your answer to part a, write the parameter set θ along with the dimensionality of each parameter.

$$\Theta = \{ \underbrace{\lambda_1, \lambda_2, \ldots, \lambda_K}_{\text{each } \in \mathbb{R}^D}, \underbrace{\pi_1, \pi_2, \ldots, \pi_K}_{\text{each } \in \mathbb{R}^1} \} = \{ \{\lambda_k\}_{k=1}^K, \{\pi_k\}_{k=1}^K \}$$

(c) (10 points) You can find the optimal parameters θ^* by maximizing the log-likelihood of observing the training data. Write down the optimization problem for finding λ_k. You should simplify the optimization problem as much as possible.

To be clear: you do not have to solve the optimization problem, i.e., you do not have to take the derivative and set it equal to zero. You simply have to state what (simplified) expression you need to maximize with respect to λ_k.

$$\log \mathcal{L}(\Theta) = \sum_{k=1}^{K} \sum_{i \in C_k} \log \pi_k + \log h(y_i) + \overbrace{\log g(\lambda_k) + \lambda_k^T y_i}^{\text{contains } \lambda_k}$$

$$\lambda_k^* = \arg\max_{\lambda_k} \left(\sum_{i \in C_k} [\log(g(\lambda_k)) + \lambda_k^T y_i] \right)$$

(d) (10 points) During decoding, we get a test sample $\mathbf{y}_{test}$ and determine what class $\hat{k}$, it belongs to by solving the optimization problem

$$\hat{k} = \arg\max_{k} \; a_k(\mathbf{y}_{test})$$

i. (5 points) Write the expression for $a_k(\mathbf{y}_{test})$ under this exponential model.

$$a_k(y_{test}) = P(C_k \mid y_{test})$$

$$= \frac{P(C_k)\,P(y_{test} \mid C_k)}{P(y_{test})}$$

Since $P(y_{test})$ is not related to k, we could just drop this term

$$\Rightarrow \; a_k(y_{test}) = P(C_k)\,P(y_{test} \mid C_k)$$

$$= \pi_k \, h(y_{test}) \, g(\lambda_k) \, e^{\lambda_k^T y_{test}}$$

So $\hat{k} = \arg\max_{k} \left(\pi_k \, h(y_{test}) \, g(\lambda_k) \, e^{\lambda_k^T y_{test}}\right)$

$\pi_k h(y_{test}) g(\lambda_k) e^{\lambda_k^T y_{test}}$

ii. (5 points) Use the expressions for $a_m(\mathbf{y}_{test})$ and $a_n(\mathbf{y}_{test})$ to derive the decision boundary between C_m and C_n. Comment on the functional form of the decision boundary (e.g., is it linear, quadratic, or some other functional form).

along decision boundary between C_m and C_n,

$$a_m(y_{test}) = a_n(y_{test})$$

$$\pi_m \cancel{h(y_{test})} g(\lambda_m) e^{\lambda_m^T y_{test}} = \pi_n \cancel{h(y_{test})} g(\lambda_n) e^{\lambda_n^T y_{test}}$$

$$e^{(\lambda_m^T - \lambda_n^T) y_{test}} = \frac{\pi_n g(\lambda_n)}{\pi_m g(\lambda_m)}$$

$$(\lambda_m - \lambda_n)^T y_{test} - \underbrace{\ln \frac{\pi_n g(\lambda_n)}{\pi_m g(\lambda_m)}}_{\text{constant}} = 0$$

this is a linear decision boundary

3. **Inference in probabilistic graphical models** (20 points)

In this problem, you will compute some probabilities for an inference task in a probabilistic graphical model. Consider the directed acyclic graph shown below

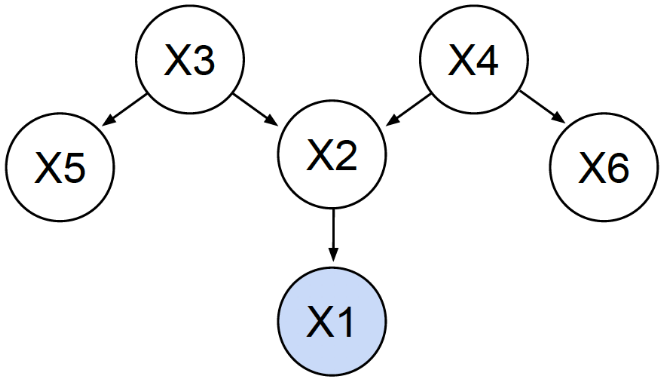


Figure 1: A directed acyclic graph over the variables $X1, \cdots, X6$. Note that $X1$ is observed (which is denoted by the fact that it's shaded in) and the remaining variables are unobserved.

This problem continues on the next page.

(a) (10 points) Compute the posterior probability $P(X2 = 2|X1 = 1)$ from the following local probabilities:

$$P(X1 = 1|X2 = 2) = 0.5$$
$$\sum_{X3,X4} P(X2 = 2|X3, X4)P(X3)P(X4) = 0.25$$
$$\sum_{X2} P(X1 = 1|X2) \sum_{X3,X4} P(X3)P(X4)P(X2|X3, X4) = 0.7$$

from the graph we can see $X_3 \perp\!\!\!\perp X_4$, so

$$P(X_3, X_4) = P(X_3)P(X_4)$$

by second eqn,

$$\sum_{X_3,X_4} P(X_2=2|X_3,X_4)P(X_3)P(X_4) = 0.25$$

$$\sum_{X_3,X_4} P(X_2=2|X_3,X_4)P(X_3,X_4) = 0.25$$

$$\sum_{X_3,X_4} P(X_2=2,X_3,X_4) = 0.25$$

$$P(X_2=2) = 0.25$$

then by third eqn,

$$\sum_{X_2} P(X_1=1|X_2) \sum_{X_3X_4} P(X_2,X_3,X_4) = 0.7$$

$$\sum_{X_2} P(X_1=1|X_2)P(X_2) = 0.7$$

$$\sum_{X_2} P(X_1=1,X_2) = 0.7$$

$$P(X_1=1) = 0.7$$

thus $P(X_2=2|X_1=1) = \dfrac{P(X_2=2)P(X_1=1|X_2=2)}{P(X_1=1)}$

$$= \frac{0.25 \cdot 0.5}{0.7} = 0.179$$

(b) (10 points) Compute the posterior probability $P(X3 = 4|X1 = 1)$ from the following local probabilities:

$$P(X3 = 4) = 0.6$$
$$\sum_{X2} P(X1 = 1|X2) \sum_{X4} P(X2|X3 = 4, X4)P(X4) = 0.4$$

from the graph we have, $P(X_2, X_3, X_4) = P(X_2|X_3, X_4)P(X_3)P(X_4)$

by second eqn, $P(X_1, X_2, X_3, X_4) = P(X_1|X_2)P(X_2, X_3, X_4)$

pull the summation over X_4 to the front since $P(X_1|X_2)$ does not contain X_4

swap the order of summation since there is no difference

$$\sum_{X_4}\sum_{X_2} P(X_1=1|X_2) \frac{P(X_2, X_3=4, X_4)}{\underbrace{\cancel{P(X_3=4)}}_{0.6}} = 0.4$$

$$\sum_{X_4}\sum_{X_2} P(X_1=1, X_2, X_3=4, X_4) = 0.24$$

$$P(X_1=1, X_3=4) = 0.24 = P(X_3=4|X_1=1)\overset{0.7 \text{ from (a)}}{\cancel{P(X_1=1)}}$$

$$\Rightarrow P(X_3=4|X_1=1) = 0.343$$

4. **Continuous decoding.** (35 points)
Consider the following linear dynamical system:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{q}_k$$

where $\mathbf{x} \in \mathbb{R}^m$ is the state variable and $\mathbf{y} \in \mathbb{R}^n$ is the observation. We learned the Kalman filter is a recursive solution to this system when $\mathbf{w}_k$ and $\mathbf{q}_k$ are zero-mean Gaussian random variables. In this problem we will perform some analysis for the case when $\mathbf{w}_k$ and $\mathbf{q}_k$ are Gaussian noises with non-zero means. Particularly, $\mathbf{w}_k \sim \mathcal{N}(\mathbf{a}, \mathbf{W})$ and $\mathbf{q}_k \sim \mathcal{N}(\mathbf{c}, \mathbf{Q})$.

Hints: Note, these hints are for both question 4 and the bonus. Let random variable $\mathbf{u}$ *have mean* $\mu = \mathbb{E}[\mathbf{u}]$ *and random variable* $\mathbf{v}$ *have mean* $\nu = \mathbb{E}[\mathbf{v}]$*. Then, the following facts hold:*

- *The covariance between* $\mathbf{u}$ *and* $\mathbf{v}$ *is:*

$$\Sigma_{\mathbf{uv}} = \mathbf{cov}(\mathbf{u}, \mathbf{v}) = \mathbb{E}[(\mathbf{u} - \mu)(\mathbf{v} - \nu)^T]$$

- *The covariance of* $\mathbf{u}$ *is:*

$$\Sigma_{\mathbf{uu}} = \mathbf{cov}(\mathbf{u}, \mathbf{u}) = \mathbb{E}[(\mathbf{u} - \mu)(\mathbf{u} - \mu)^T]$$

Finally, recall that $\mathbb{E}[f(\mathbf{u})] = f(\mathbb{E}[\mathbf{u}])$ *for any linear function* f*.*

(a) (10 points) Consider the state process:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k \quad \text{when} \quad \mathbf{w}_k \perp\!\!\!\perp \mathbf{x}_{k-1}$$

Find the mean (μ) and covariance (Σ) for the distribution of $\mathbf{x}_k|\mathbf{x}_{k-1}$. The final expressions you derive must be in terms of $\mathbf{A}, \mathbf{a}, \mathbf{W}$ and $\mathbf{x}_{k-1}$ only.

$$\mu = E[X_k|X_{k-1}] = E[AX_{k-1} + W_k|X_{k-1}]$$
$$= AE[X_{k-1}|X_{k-1}] + \cancel{E[W_k|X_{k-1}]} \quad E[W_k]$$
$$= AX_{k-1} + a$$

$$\Sigma = \text{Cov}(X_k, X_k|X_{k-1}) = E[(X_k - E[X_k|X_{k-1}])(X_k - E[X_k|X_{k-1}])^T]$$
$$= E[(W_k - a)(W_k - a)^T] = \text{Cov}(W_k, W_k) = W$$

(b) (10 points) Consider just the observation process:

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{q}_k$$

Suppose $\mathbf{x}_k \perp\!\!\!\perp \mathbf{q}_k$ and $\mathbf{x}_k \sim \mathcal{N}(\mu_{\mathbf{x}_k}, \Sigma_{\mathbf{x}_k\mathbf{x}_k})$. Find the mean ($\mu$) and covariance ($\Sigma$) for the distribution of $\mathbf{y}_k|\mathbf{q}_k$. The final expressions you derive must be in terms of $\mathbf{C}, \mathbf{c}, \mathbf{Q}, \mu_{\mathbf{x}_k}, \Sigma_{\mathbf{x}_k\mathbf{x}_k}$ and $\mathbf{q}_k$ only.

$$\mu = E[y_k|q_k] = E[Cx_k + q_k|q_k]$$
$$= C\underbrace{E[x_k|\cancel{q_k}]}_{E[x_k]} + E[q_k|q_k]$$
$$= C\mu_{x_k} + q_k$$

$$\Sigma = \mathrm{Cov}(y_k, y_k|q_k) = E[(y_k - E[y_k|q_k])(y_k - E[y_k|q_k])^T]$$
$$= E[(Cx_k - C\mu_{x_k})(Cx_k - C\mu_{x_k})^T]$$
$$= E[C(x_k - \mu_{x_k})(x_k - \mu_{x_k})^T C^T]$$
$$= C\,E[(x_k - \mu_{x_k})(x_k - \mu_{x_k})^T]\,C^T = C\Sigma_{x_k x_k} C^T$$

(c) (15 points) By appending a 1 to the state variables $\mathbf{x}$ and the observation vector $\mathbf{y}$, i.e. with the new variables $\tilde{\mathbf{x}} = [\mathbf{x}^\mathsf{T}\ 1]^T$ and $\tilde{\mathbf{y}} = [\mathbf{y}^\mathsf{T}\ \ 1]^T$, we write a new linear dynamical system:

$\tilde{x} \in \mathbb{R}^{m+1}$

$\tilde{y} \in \mathbb{R}^{n+1}$

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_{k-1} + \tilde{\mathbf{w}}_k$$
$$\tilde{\mathbf{y}}_k = \tilde{\mathbf{C}}\tilde{\mathbf{x}}_k + \tilde{\mathbf{q}}_k$$

In this dynamical system, $\tilde{\mathbf{w}}_k$ and $\tilde{\mathbf{q}}_k$ are zero-mean random variables.

i. (5 points) What are the dimensions of $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{w}}_k$ and $\tilde{\mathbf{q}}_k$?

$\tilde{A} : (m+1)\times(m+1)$ $\quad \tilde{w}_k : \in \mathbb{R}^{m+1}$

$\tilde{C} : (n+1)\times(m+1)$ $\quad \tilde{q}_k : \in \mathbb{R}^{n+1}$

ii. (10 points) Determine $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{w}}_k$, and $\tilde{\mathbf{q}}_k$ such that this dynamical system is equivalent to the original dynamical system in this problem (with non-zero mean noise terms $\mathbf{w}_k \sim \mathcal{N}(\mathbf{a}, \mathbf{W})$ and $\mathbf{q}_k \sim \mathcal{N}(\mathbf{c}, \mathbf{Q})$). You must write $\tilde{\mathbf{A}}$, $\tilde{\mathbf{C}}$, $\tilde{\mathbf{w}}_k$, and $\tilde{\mathbf{q}}_k$ solely in terms of $\mathbf{A}, \mathbf{C}, \mathbf{a}, \mathbf{c}, \mathbf{W}, \mathbf{Q}, \mathbf{w}_k$ and $\mathbf{q}_k$.

$$\begin{pmatrix} x_k \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} A & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0\,0\cdots 0 & 0 \end{pmatrix}}_{\tilde{A}} \begin{pmatrix} x_{k-1} \\ 1 \end{pmatrix} + \underbrace{\begin{pmatrix} w_k \\ 1 \end{pmatrix}}_{\tilde{w}_k}$$

likewise for $\tilde{C}$ and $\tilde{q}_k$

mxm with same entries as in A

a col of 0's

$$\tilde{A} = \left.\underbrace{\begin{pmatrix} \boxed{A}_{m \times m} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \\ 0\,0\cdots 0 & 0 \end{pmatrix}}_{m+1}\right\} m+1 \qquad \tilde{w}_k = \begin{pmatrix} w_k \\ 1 \end{pmatrix}$$

pad a 1 vertically to w_k

a row of 0's

same as in C, nxm

col of 0's

$$\tilde{C} = \left.\underbrace{\begin{pmatrix} \boxed{C}_{n \times m} & \begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix} \\ 0\,0\cdots 0 & 0 \end{pmatrix}}_{m+1}\right\} n+1 \qquad \tilde{q}_k = \begin{pmatrix} q_k \\ 1 \end{pmatrix}$$

row of 0's

Vertically pad a 1 to q_k

5. **Bonus** (10 points) For the observation process from the previous problem:

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{q}_k$$

(a) (8 points) When $\mathbf{x}_k \not\perp\!\!\!\perp \mathbf{q}_k$, what is the mean ($\mu$) and covariance ($\Sigma$) for the distribution of $\mathbf{y}_k$? The final expressions you derive must be in terms of $\mathbf{C}, \mathbf{c}, \mathbf{Q}, \mu_{\mathbf{x}_k}, \Sigma_{\mathbf{x}_k\mathbf{x}_k}, \Sigma_{\mathbf{q}_k\mathbf{x}_k}$ and $\Sigma_{\mathbf{x}_k\mathbf{q}_k}$ only.

$$\mu = E[y_k] = E[Cx_k + q_k]$$
$$= CE[x_k] + E[q_k]$$
$$= C\mu_{x_k} + c$$

$$\Sigma = Cov(y_k, y_k) = E[(y_k - \mu)(y_k - \mu)^T]$$
$$= E[(Cx_k + q_k - C\mu_{x_k} - c)(Cx_k + q_k - C\mu_{x_k} - c)^T]$$
$$= E[(C(x_k - \mu_{x_k}) + (q_k - c))(C(x_k - \mu_{x_k}) + (q_k - c))^T]$$
$$= E[C(x_k - \mu_{x_k})(x_k - \mu_{x_k})^T C^T + C(x_k - \mu_{x_k})(q_k - c)^T + (q_k - c)(x_k - \mu_{x_k})^T C^T$$
$$+ (q_k - c)(q_k - c)^T]$$
$$= C\Sigma_{x_k}C^T + C\Sigma_{x_k q_k} + \Sigma_{q_k x_k}C^T + Q$$

(b) (2 points) If $\mathbf{x}_k \perp\!\!\!\perp \mathbf{q}_k$, can your answer for the previous part (a) be further simplified? What will be the parameters for the distribution of $\mathbf{y}_k$ in this case?

Hint: If the random variables $\mathbf{u}$ *and* $\mathbf{v}$ *are independent, then* $\mathbb{E}[\mathbf{uv}] = \mathbb{E}[\mathbf{u}]\mathbb{E}[\mathbf{v}]$.

Yes, if $x_k \perp\!\!\!\perp q_k$, $E[(x_k - \mu_{x_k})(q_k - c)^T] = E[x_k - \mu_{x_k}]E[q_k - c] = 0$

likewise $E[(q_k - c)(x_k - \mu_{x_k})^T] = 0$

thus $\Sigma = C\Sigma_{x_k}C^T + Q$, $\mu = C\mu_{x_k} + c$ stays the same