APPM 4510 HW1

Zane Jakobs

Note: I'm using your notation for pdfs since we aren't talking about Lie algebras, and it's pretty nice.

1

$$[X, Y] = [X][Y|X] = \frac{1}{c} e^{-\frac{x^2}{2\sigma^2} - \frac{(y-x)^2}{2\gamma^2}}$$

2

Given that $X = 2.5$, it must be the case that $Y = 2$ exactly, so

$$[Y] = \delta(y - 2),$$

where δ is the Dirac delta.

3(a)

Proof. Taking the expectation of this process at time j gives

$$\mathbb{E}[v_j] = \lambda \mathbb{E}[v_{j-1}] + \bar{\epsilon}. \tag{1}$$

In order for the desired limit to exist (which it must, since $|\lambda| < 1$ so the process is covariance-stationary), it must be the case that as $j \to \infty$, the absolute value of the series of innovations, $|\Delta V|$, must converge to 0. Therefore, as $j \to \infty$, $\mathbb{E}][v_{j+1}] \to \mathbb{E}[v_j]$, and hence, in this limit, (1) becomes

$$\lim_{j \to \infty} \left(\mathbb{E}[v_j] = \lambda \mathbb{E}[v_j] + \bar{\epsilon} \right). \tag{2}$$

This can be solved to give the desired limit,

$$\lim_{j \to \infty} \mathbb{E}[v_j] = \frac{\bar{\epsilon}}{1 - \lambda}. \qquad \square$$

3(b)

Proof. Using the same justification as above to let successive terms have equal moments in the large-time limit, we set up the system

$$\lim_{j\to\infty} \mathrm{Var}[v_j] = \lambda^2 \mathrm{Var}[v_j] + \sigma_\epsilon^2. \tag{3}$$

Then, again like with the mean, we solve to get

$$\lim_{j\to\infty} \mathrm{Var}[v_j] = \frac{\sigma_\epsilon^2}{1-\lambda^2}.$$

□

3(c)

Proof. The lag-1 autocovariance in the infinite-time limit is

$$\lim_{j\to\infty} \mathrm{Cov}[v_{j+1}, v_j] = \lim_{j\to\infty} \left(\mathbb{E}[v_{j+1}v_j] - \mathbb{E}[v_j]\mathbb{E}[v_{j+1}]\right). \tag{4}$$

Since $v_{j+1} = \lambda v_j + \epsilon_j$,

$$\mathbb{E}[v_{j+1}v_j] = \mathbb{E}[\lambda v_j^2 + \epsilon_j v_j] = \lambda \mathbb{E}[v_j^2] + \mathbb{E}[\epsilon_j]\mathbb{E}[v_j] = \lambda(\mathrm{Var}(v_j) + \mathbb{E}[v_j]^2) + \bar{\epsilon}\mathbb{E}[v_j], \tag{5}$$

where the second equality follows from the independence of the v_j and ϵ_j. We then have

$$\begin{aligned}
\lim_{j\to\infty} \mathrm{Cov}[v_{j+1}, v_j] &= \lambda\left(\frac{\sigma_\epsilon^2}{1-\lambda^2} + \frac{\bar{\epsilon}^2}{(1-\lambda)^2}\right) + \frac{\bar{\epsilon}^2}{1-\lambda} - \frac{\bar{\epsilon}^2}{(1-\lambda)^2} \\
&= \frac{(\lambda-1)\bar{\epsilon}^2}{(1-\lambda)^2} + \frac{\bar{\epsilon}^2}{1-\lambda} + \frac{\lambda\sigma_\epsilon^2}{1-\lambda^2} \\
&= \frac{\lambda\sigma_\epsilon^2}{1-\lambda^2}.
\end{aligned}$$

□

3(d)

Proof. Since we justified in 3(a) and 3(b) that the variance of v_{j+1} approaches that of v_j as $j \to \infty$, the autocorrelation in that limit is

$$\begin{aligned}
\lim_{j\to\infty} \mathrm{Corr}[v_{j+1}, v_j] &= \lim_{j\to\infty} \frac{\mathrm{Cov}[v_{j+1}, v_j]}{\left(\frac{\sigma_\epsilon^2}{1-\lambda^2}\right)^2} \\
&= \frac{\lambda}{\frac{\sigma_\epsilon^2}{1-\lambda^2}} \\
&= \frac{\lambda(1-\lambda^2)}{\sigma_\epsilon^2}.
\end{aligned}$$

□

4

$\mathrm{Rk}(\mathbf{A}) = k - 1$.

Proof. By the definition of a basis, the $\mathbf{v}_i$ form a basis for $\mathbb{R}^k$, thus the rank of $\mathbf{V}$ is $\mathrm{Rk}(\mathbf{V}) = k$. Subtracting $\frac{1}{k}\mathbf{1}\mathbf{1}^T$ from the $k \times k$ identity gives a matrix, each of whose rows has one element that is 1 greater than the element in the row above/below it, allowing $k-1$ of those rows to be put into a linear combination that is equal to the k-th row, thus the matrix $\mathbf{I} - \frac{1}{k}\mathbf{1}\mathbf{1}^T$ must have rank $k-1$. Since a matrix product is one representation for composition of linear transformations, we note that $\mathbf{A}$ can be viewed as the composition of a rank $k-1$ transformation and a rank k transformation, and thus cannot have a rank greater than $k-1$ (in simpler language, a transformation that is composed of other transformations is "no more invertible" than its "least invertible" individual part). It is also true that $\mathbf{A}$ can be constructed by "gluing" the $k \times k$ submatrix of $\mathbf{V}$ that spans $\mathbb{R}^k$ multiplied against $\mathbf{I} - \frac{1}{k}\mathbf{1}\mathbf{1}^T$, with the result of the product of $\mathbf{I} - \frac{1}{k}\mathbf{1}\mathbf{1}^T$ and the remaining $(n-k) \times k$ submatrix of $\mathbf{V}$, and applying an (invertible) permutation matrix. The first product is the product of a full-rank matrix (of rank k) and a rank-deficient (rank $k-1$) matrix, and thus has rank $k-1$. The second product, when "glued" with the first, cannot add to the rank of $\mathbf{A}$, since these $n-k$ rows were originally linear combinations of the other k and they were composed with the exact same transformation as those other k, so those original linear combinations are preserved (that is, if row $k+3$ of $\mathbf{V}$ were equal to the sum of rows $k-5$ and $k-143$ of $\mathbf{V}$, the same will be true of $\mathbf{A}$). Thus, $\mathrm{Rk}(\mathbf{A}) = k-1$ exactly. □

5

Newton's method entails the repeated computation of $\mathbf{J}$, the Jacobian of $J(\mathbf{x})$ (Note: since J is a scalar function, its Jacobian would probably be more appropriately called its gradient, and is a vector, not a matrix), the Hessian $\mathcal{H}$ (NOT the same as $\mathbf{H}(\mathbf{x})$, solution of the equation

$$\mathcal{H} \cdot (\delta\mathbf{x}) = -\mathbf{J} \tag{6}$$

for $\delta\mathbf{x}$, and updating $\mathbf{x} \to \mathbf{x} + \gamma\delta\mathbf{x}$ for some step size $\gamma \in (0, 1]$ (the choice of a γ could be replaced with a line search for the optimal step size). We can compute $\mathbf{J}$ as

$$\begin{aligned}
\mathbf{J} &= \frac{\partial J}{\partial \mathbf{x}^T} \\
&= 2\mathbf{x}^T\mathbf{C}^{-1} + \epsilon^2 \frac{\partial}{\partial \mathbf{x}^T}\left(\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{x})^T\mathbf{y} + \mathbf{H}(\mathbf{x})^T\mathbf{H}(\mathbf{x})\right) \\
&= 2\mathbf{x}^T\mathbf{C}^{-1} - \epsilon^2 \left(\mathbf{y}^T\frac{\partial \mathbf{H}(\mathbf{x})}{\partial \mathbf{x}^T} + \frac{\partial \mathbf{H}(\mathbf{x})^T}{\partial \mathbf{x}^T}\mathbf{y} - \frac{\partial \mathbf{H}(\mathbf{x})^T}{\partial \mathbf{x}^T}\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{x})^T\frac{\partial \mathbf{H}(\mathbf{x})}{\partial \mathbf{x}^T}\right).
\end{aligned}$$