ECE 555 **Assignment # 2**

Issued: January 27 Due: February 3, 2011

Solutions

Exercises:

3 Consider the linear state space model on $\mathbb{R}^2$,

$$X(t+1) = AX(t) + \mathcal{E}(t+1)$$

where the eigenvalues of A are strictly less than one. It is assumed that $\boldsymbol{\mathcal{E}}$ is i.i.d., with zero-mean and finite variance. A course on linear systems theory will tell you that for any matrix $Q > 0$, there is a solution to the discrete-time *Lyapunov equation* $A^{\mathrm{T}}MA = M - Q$, where the 2×2 matrix M is positive definite.

(i) Obtain a solution to Poisson's equation with $c(x) = x^{\mathrm{T}}Qx$.

(ii) Obtain a solution to Foster's criterion using the Lyapunov function $V(x) = a\log(1 + h(x))$, where $a > 0$ is a constant, and h is your solution to (i).

Solution: (i) We use the zero mean assumption to conclude that a quadratic will solve the equation:

$$V(x) = x^T Mx + b,$$

with $b \geq 0$ arbitrary, and $M > 0$ to be chosen.

$$\begin{aligned}
\mathcal{D}V(x) &= \mathsf{E}[X(t+1)^T MX(t+1) - X(t)^T MX(t) | X(t) = x] \\
&= \mathsf{E}[(Ax + \mathcal{E}(1))^T (Ax + \mathcal{E}(1)) - x^T Mx] \\
&= (x^T A^T MAx + \sigma_M^2 - x^T Mx),
\end{aligned}$$

where $\sigma_M^2 = \mathsf{E}[\mathcal{E}(1)^T M\mathcal{E}(1)]$. Hence from the Lyapunov equation $A^{\mathrm{T}}MA = M - Q$ we conclude that

$$\mathcal{D}V(x) = -\|x\|^2 + \sigma_M^2$$

We obtain a solution to Poisson's equation with $\eta = \sigma_M^2$.

(ii) For simplicity we assume that $h(x) = x^T Mx$ in (i) is obtained with $Q = I$. That is, $b = 0$ and $A^{\mathrm{T}}MA = M - I$.

With $V(x) = a\log(1 + h(x))$ we have,

$$\mathcal{D}V(x) = a\mathsf{E}[\log(1 + h(X(t+1)) - \log(1 + h(X(t)))|X(t) = x].$$

Jensen's Inequality gives the upper bound:

$$\mathcal{D}V(x) \leq a\log(\mathsf{E}[1 + h(X(t+1))|X(t) = x]) - V(x)$$

Applying (i) we obtain

$$\mathcal{D}V(x) \leq a\big(\log(1 + h(x) - \|x\|^2 + \sigma_M^2)\big) - V(x)$$

Concavity of the log gives

$$\log(a+b) \le \log(a) + \frac{b}{a}, \quad \text{for all } a, b.$$

Hence,

$$\begin{aligned}
\mathcal{D}V(x) &\le a\Big(\log(1+h(x))) + \frac{-\|x\|^2 + \sigma_M^2}{1+h(x)}\Big) - V(x) \\
&= a\frac{-\|x\|^2 + \sigma_M^2}{1+h(x)}
\end{aligned}$$

Choose $\varepsilon > 0$ such that $\|x\|^2 \ge \varepsilon h(x) = \varepsilon x^T M x$ for all x. Then, the right hand side of the previous bound can be manipulated to obtain

$$\begin{aligned}
\mathcal{D}V(x) &\le a\left(\frac{-\varepsilon h(x) + \sigma_M^2}{1+h(x)}\right) \\
&= a\left(\frac{-\varepsilon(1+h(x)) + \varepsilon + \sigma_M^2}{1+h(x)}\right) \\
&\le -\frac{\varepsilon}{2}a \qquad\qquad \text{provided } 1+h(x) \ge 2\left(\frac{\varepsilon + \sigma_M^2}{\varepsilon}\right).
\end{aligned}$$

Conclusion: Foster's criterion holds provided $a \ge \frac{2}{\varepsilon}$.

4 Suppose that R is a *non-negative* $N \times N$ matrix: This means that $R_{ij} \ge 0$ for each $1 \le i \le N$. Suppose that s is a column vector, ν is a row vector, satisfying $\nu s > 0$, and $R \ge s \otimes \nu$. That is, $R_{ij} \ge s_i \nu_j$ for each i, j. Finally, suppose that $r > 0$ is a solution to the equation,

$$\nu G s = 1 \quad \text{where} \quad G = \sum_{k=0}^{\infty} r^{-k-1}(R - s\otimes\nu)^k \tag{1}$$

Using the ideas of the lecture on January 27, show that $\mu = \nu G$ and $h = Gs$ are left and right eigenvectors of R:

$$\mu R = r\mu, \qquad Rh = rh$$

Notes: This construction is part of the celebrated theory of Perron & Frobenius — begun approximately one hundred years ago. The scalar r is the *PerronFrobenius eigenvalue*.

A typical systems application uses $R_{ij} = e^{f_i} P_{ij}$, where P is a transition matrix, and f is envisioned as a function on the state space. The eigenvalue r is a value of the moment generating function found in risk-sensitive control and in large deviations theory for Markov chains. *Why*? Because in this case,

$$R^n_{ij} = \mathsf{E}\big[\exp(\sum_{t=0}^{n-1} f(X(t)))\mathbf{1}(X(n)=j) \mid X(0)=i\big]$$

Solution: We just prove the result for h — the eigenvector property for μ is proved using the same arguments.

We have the following identity from the definition of G in (1):

$$r^{-1}(R - s \otimes \nu)G = \sum_{k=1}^{\infty} r^{-k-1}(R - s \otimes \nu)^k$$

That is, $r^{-1}(R - s \otimes \nu)G = G - r^{-1}I$. Multiplying on the right by s gives,

$$r^{-1}(R - s \otimes \nu)Gs = Gs - r^{-1}s$$

Since $\nu Gs = 1$ this means that $r^{-1}(RGs - s) = Gs - r^{-1}s$. Canceling common terms and substituting $h = Gs$ proves the result.

5 Suppose that there exists a function s and a probability distribution ν on X satisfying $P = s \otimes \nu$ (that is, $P(x, y) = s(x)\nu(y)$ for each $x, y \in \mathsf{X}$.) Show that $s \equiv 1$, ν is an invariant measure, and hence $\boldsymbol{X}$ is i.i.d. with marginal distribution ν.

Solution: First, show $s \equiv 1$: Since ν is a probability measure,

$$1 = P(x, \mathsf{X}) = s(x)\nu(\mathsf{X}) = s(x).$$

Next, $\boldsymbol{X}$ is iid:

$$\begin{aligned} \mathsf{P}\left(X(t+1) = x | X(t) = x_0, X(t-1) = x_1, \ldots\right) &= \mathsf{P}(X(t+1) = x | X(t) = x_0) \\ &= \nu(x), \qquad \text{independent of } x_0, x_1, \ldots. \end{aligned}$$

This implies that $\boldsymbol{X}$ is iid, with marginal ν.

6 Suppose that the minorization condition holds, $P \geq s \otimes \nu$, with $s(x) \equiv \epsilon > 0$ constant. This is a form of *Doeblin's condition*[1]. Observe that $Q(x, y) := [P(x, y) - \epsilon\nu(y)]/(1 - \epsilon)$ is a transition matrix (let's stick to a countable state space for simplicity). Doeblin's condition can be equivalently expressed $P = \epsilon 1 \otimes \nu + (1 - \epsilon)Q$, or

$$P(x, y) = \epsilon\nu(y) + (1 - \epsilon)Q(x, y), \qquad x, y \in \mathsf{X}$$

Show that $G(x, \mathsf{X}) \leq \epsilon^{-1}$ for all x, where the potential kernel is given by,

$$G = \sum_{k=0}^{\infty} (P - s \otimes \nu)^k.$$

Is there an invariant probability measure?

Solution: This is establish as in lecture:

$$Gs = \sum_{k=0}^{\infty} (P - s \otimes \nu)^k s \leq 1 \quad \text{everywhere.}$$

[1]Wolfgang Doeblin was a *remarkable person. Just 10 years ago*, his sealed manuscript was opened which revealed that he developed the theory of stochastic differential equations while fighting the Germans in WW2! His formulation predates Ito, and is also simpler.

Since $s(x) = \varepsilon$ for all x, this means

$$\varepsilon G(x, \mathsf{X}) \leq 1 \text{ for all } x.$$

We have seen in lecture that $\mu = \nu G$ is invariant. It also has finite mass:

$$\mu(\mathsf{X}) = \nu G \mathbf{1} \leq \varepsilon^{-1} \nu \mathbf{1} = \varepsilon^{-1}.$$

So $\pi = \mu/\mu(\mathsf{X})$ is an invariant probability.

Note: This condition is somewhat stronger than Doeblin's condition. The minorization with s constant implies that the chain is uniformly ergodic:

$$\lim_{n\to\infty} \sup_{A} \sup_{x} |P^n(x, A) - \pi(A)| = 0$$

There is a simple proof based on coupling – See CTCN, or ask me!