ESE 6180: Learning, Dynamics and Control **Fall 2024**

Homework 1

Assigned: 08/28/2024 *Due: 09/13/2024*

Homework must be LaTeX'd or it will not be graded.

Grading: Each problem will be graded on a scale of 0-4. If you get 80% of the problem or more correct, and make an honest attempt at the rest, you will get 4/4. If you get 60% of the problem or more correct, you will get 3/4, etc.

Canvas: Please submit your HW as a single pdf file, with pages correctly tagged to go with each problem.

Working in groups: You are allowed to, and in fact encouraged, to discuss and work on problems with your classmates. However, each student must write up their own homework independently. Further, please make note of your collaborators in the designated spot in the homework template.

Citing references: If you referred to solutions found in published material (papers, textbooks, websites, etc.), you must cite these in your homework solutions. It is ok to use proofs that you find online for guidance, but you should indicate where and how you did so, and you should always make a first attempt at the answer on your own. Importantly, even if you are following the guidance of a proof from a paper, you must be sure to fully explain all steps, as well as fill in any missing steps.

Useful inequalities: This cheat sheet may come in handy throughout the course.

1 Matrix Norms

The $p \to q$ induced norm of a matrix M is defined as $\|M\|_{p\to q} \triangleq \sup_{\|x\|_p \leq 1} \|Mx\|_q$, where $\|\cdot\|_q$ is the ℓ_q norm on the range of M. For $M \in \mathbb{R}^{m\times n}$, show that:

(a) Any $p \to p$ induced norm is sub-multiplicative, i.e., that $\|AB\| \leq \|A\|\|B\|$.

(b) $\|A\|_{1\to 1} = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$, i.e., the maximum absolute column sum,

(c) $\|A\|_{\infty\to\infty} = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$, i.e., the maximum absolute row sum,

(d) $\|A\|_{2\to 2} = \sigma_{\max}(A)$, i.e., the maximum singular value of A,

(e) $\|A\|_{2\to 2} \leq \|A\|_F \leq \|A\|_\star$, where $\|A\|_F^2 = \operatorname{tr}\left(A^\top A\right)$ is the Frobenius norm of A and $\|A\|_\star = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i(A)$ is the nuclear or Schatten-1 norm of A,

(f) the Frobenius norm and $2 \to 2$ induced norm are both *monotone*: if $0 \preceq A \preceq B$ (i.e., A and $B - A$ are symmetric and positive semidefinite) then $\|A\| \leq \|B\|$.

2 Linear Algebra

(a) Let P and Q be two matrices.

(i) Prove that $\operatorname{tr} PQ = \operatorname{tr} QP$ whenever the products PQ and QP can be formed.

(ii) Show that if $P \succeq 0$, $Q \succeq 0$ (i.e., P and Q are symmetric and positive semidefinite), then $\operatorname{tr} PQ \geq 0$.

(iii) Use this to conclude that for $P \succ 0$, $Q \succeq 0$, $\operatorname{tr} PQ \geq \lambda_{\min}(P) \operatorname{tr} Q$.

(b) Let A, B, C be matrices. The Kronecker product of two matrices A, B is the block matrix $A \otimes B$ where the ij:th block is B rescaled by A_{ij}. The vectorization of a matrix A, denoted $\mathsf{vec}\, A$, is the block column vector in which the i:th block is the i:th column of A.

(i) Prove that $\mathsf{vec}(ABC) = (C^\mathsf{T} \otimes A)\, \mathsf{vec}\, B$.

(ii) Prove that $\operatorname{tr} A^\mathsf{T} B = \langle \mathsf{vec}\, A, \mathsf{vec}\, B\rangle$.

(c) The spectral radius of a matrix $A \in \mathbb{R}^{n\times n}$ is defined as

$$\rho(A) \triangleq \max\{|\lambda| \text{ such that } \lambda \text{ is an eigenvalue of } A\}.$$

(i) Show by example that $\rho(\cdot)$ is not a matrix norm.

(ii) Show that

$$\begin{aligned} \rho(A) < 1 \quad &\iff \quad \lim_{k\to\infty} A^k = 0 \\ \rho(A) > 1 \quad &\implies \quad \lim_{k\to\infty} \|A^k\| = \infty. \end{aligned}$$

Hint: the Jordan form of A may come in handy.

(iii) Use the previous result to prove Gelfand's formula: $\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}$. *Hint: define the matrices*

$$A_{\pm\varepsilon} \triangleq \frac{A}{\rho(A) \pm \varepsilon}.$$

(iv) Show that $\rho(A) < 1$ if and only if there exists $X \succ 0$ such that

$$A^\top X A - X \prec 0.$$

(d) Let M be a trilinear form acting on $\mathbb{R}^{d\times d\times d}$ via $M[x, y, z] = \sum_{ijk} M_{ijk} x_i y_j z_k$. The operator norm of M is $\|M\|_{\mathsf{op}} \triangleq \sup_{\|x\|,\|y\|,\|z\|\le 1} \left|\sum_{ijk} M_{ijk} x_i y_j z_k\right|$ where $\|\cdot\|$ is the Euclidean norm. Prove that the vector $x^\mathsf{T} M x$ with components $(x^\mathsf{T} M x)_i \triangleq \sum_{jk} M_{ijk} x_j x_k$ satisfies

$$\|x^\mathsf{T} M x\| \le \|M\|_{\mathsf{op}} \|x\|^2.$$

3 Control Theory

We first recall some definitions:

- Let $\ell_2^n(-\infty, \infty)$ be the space of square integrable sequences of $\mathbb{R}^n$-valued vectors, i.e.,

$$\ell_2^n(-\infty, \infty) \triangleq \{\vec{u} = (\dots, u_{-1}, u_0, u_1, \dots) \mid u_k \in \mathbb{R}^n, \sum_{k=-\infty}^{\infty} \|u_k\|_2^2 < \infty\}.$$

  $\ell_2^n(-\infty, \infty)$ can be equipped with the inner-product $\langle \vec{u}, \vec{v}\rangle \triangleq \sum_{k=-\infty}^{\infty} u_k^\top v_k$, which induces the norm $\|\vec{u}\|_2^2 \triangleq \sum_{k=-\infty}^{\infty} \|u_k\|_2^2$.

- $\ell_2^n(-\infty, 0]$ and $\ell_2^n[0, \infty)$ are the restrictions of $\ell_2^n(-\infty, \infty)$ to signals supported on $(-\infty, 0]$ and $[0, \infty)$, respectively.

- Consider the operator $\Psi_c : \ell_2^p(-\infty, 0] \to \mathbb{R}^n$ defined by

$$\vec{u} \mapsto \sum_{k=-\infty}^{0} A^{-k} B u(k).$$

This operator can be though of as the response to a system described by

$$x(t+1) = Ax(t) + Bu(t),\; x(-\infty) = 0$$

to an input $\vec{u} \in \ell_2^n(-\infty, 0]$, where the output vector is $x(1)$. Note here we assume $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^p$.

- Recall that (A, B) is controllable if there exists $k \in \mathbb{N}$ such that $\begin{bmatrix} B & AB & A^2B & \cdots & A^{k-1}B \end{bmatrix}$ has full row rank.

(a) Show that if (A, B) is controllable and $\rho(A) < 1$, then the matrix $X_c \triangleq \Psi_c \Psi_c^\dagger$ is nonsingular. Here, $M^\dagger$ denotes the adjoint of a linear operator M.

(b) Show that for any $x(1) \in \mathbb{R}^n$, the input $u^\star \triangleq \Psi^\dagger X_c^{-1} x(1)$ is the element of minimum norm in the set

$$\{\vec{u} \in \ell_2^p(-\infty, 0] \,|\, \Psi_c \vec{u} = x(1)\}.$$

Hint: consider the operator $P \triangleq \Psi_c^\dagger X_c^{-1} \Psi_c$. *What kind of operator is this?*

(c) Show that the following is true:

$$\{\Psi_c \vec{u} \,|\, \vec{u} \in \ell_2^p(-\infty, 0],\; \|\vec{u}\|_2 \le 1\} = \{X_c^{1/2} x \,|\, x \in \mathbb{R}^n,\; \|x\|_2 \le 1\}.$$

(d) Define the ellipsoid

$$\mathcal{E}_c \triangleq \{X_c^{1/2} x \,|\, x \in \mathbb{R}^n,\; \|x\|_2 \le 1\}.$$

In light of the previous results, how should we interpret the principal axes of $\mathcal{E}_c$?

4 Statistics and Probability

(a) Let $X = \frac{1}{n}\sum_{i=1}^n X_i$, where X_i is iid Bernoulli(p). Show that

$$\mathbf{P}[X \ge (1+\varepsilon)\mathbf{E}[X]] \le \exp\left(-\frac{np\varepsilon^2}{2+\varepsilon}\right) \quad \forall \varepsilon > 0$$

and

$$\mathbf{P}[X \le (1-\varepsilon)\mathbf{E}[X]] \le \exp\left(-\frac{np\varepsilon^2}{2}\right) \quad \forall \varepsilon \in (0,1).$$

Conclude that $\mathbf{P}[|X - \mathbf{E}[X]| \ge \varepsilon \mathbf{E}[X]] \le 2\exp\left(-\frac{np\varepsilon^2}{3}\right)$ for all $\varepsilon \in (0,1)$.

(b) We consider fixed design linear regression with scalar responses. The terminology fixed design refers to the fact that the X_i below are non-random.

The setup is as follows:

- For $i \in [n]$ there exists $\theta_\star \in \mathbb{R}^d$ such that $Y_i = \theta_\star^\mathsf{T} X_i + W_i$.
- The noise variables W_i are independent, mean zero and with variance σ^2.
- Define $\Phi^\mathsf{T} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$. $\Phi \in \mathbb{R}^{n \times d}$ is often called the design (matrix).
- We assume that $\Phi^\mathsf{T}\Phi$ is invertible. This guarantees existence and uniqueness of the least squares estimator.
- The least squares estimator is

$$\widehat{\theta} \in \operatorname*{argmin}_{\theta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n |Y_i - \theta^\mathsf{T} X_i|^2.$$

- The excess risk is defined as

$$\texttt{ER}(\theta) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\left[|Y_i - \theta^\top X_i|^2\right] - \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\left[|Y_i - \theta_\star^\top X_i|^2\right]$$

and compares how well a predictor θ does against the ground truth $\theta_\star$.

(i) Show that the least squares estimator can be written $(\Phi^\top\Phi)^{-1}\Phi^\top Y_{1:n}$.

(ii) Show that $\Phi\widehat{\theta}$ is the orthogonal projection of $Y_{1:n}$ onto the column space of Φ. In other words, show that $\Phi\widehat{\theta}$ is the closest point to $Y_{1:n}$ in the column space of Φ (closest is measured by the Euclidean metric on $\mathbb{R}^n$).

(iii) Prove that the excess risk can be written as:

$$\texttt{ER}(\theta) = \frac{1}{n}\mathbf{E}\left[\|\Phi(\theta - \theta_\star)\|^2\right].$$

(iv) Prove the bias-variance decomposition. Namely, show that:

$$\texttt{ER}(\theta) = \|\Phi(\mathbf{E}[\theta] - \theta_\star)\|^2 + \mathbf{E}\left[\|\Phi(\theta - \mathbf{E}[\theta])\|^2\right].$$

(v) Prove that the excess risk in the fixed design setting satisfies

$$\texttt{ER}(\widehat{\theta}) = \frac{\sigma^2 d}{n}.$$

.

(c) Fill in the details for the proof in the lecture notes establishing control of the moments of sub-Gaussians. Recall also that the Gamma function is defined as

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1}e^{-u}\mathrm{d}u \qquad x > 0.$$

For this exercise, you are not allowed to use Stirling's Formula/Approximation but rather each estimate must be established from first principles.

(i) for every integer n: evaluate the integral to show that $\Gamma(n) = n! \leq n^n$.

(ii) For every integer $n : \Gamma(n+1/2) \leq 1 + \Gamma(n+1) \leq 2(n+1)^{n+1}$. Optional: Can you prove a tighter estimate?

(iii) For every odd integer p: $2p\left(\frac{p+1}{2}\right)^{(p+1)/2} \leq 4ep^{p/2}$. *Hint*: maximize $2p^{3/2}\left(\frac{p+1}{2p}\right)^{(p+1)/2}$. Optional: Can you prove a tighter estimate?

(iv) For every positive random variable X, show that $\mathbf{E}X = \int_0^\infty \mathbf{P}(X > s)ds$.