MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering and Computer Science

6.241: Dynamic Systems—Fall 2007

Homework 3 Solutions

Exercise 3.2 i) We would like to minimize the 2-norm of u, i.e., $\|\underline{u}\|_2^2$. Since y_n is given as

$$y_n = \sum_{i=1}^{n} h_i u_{n-1}$$

we can rewrite this equality as

$$y_n = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix} \begin{bmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{bmatrix}$$

We want to find the $\underline{u}$ with the smallest 2-norm such that

$$\bar{y} = A\underline{u}.$$

where we assume that A has a full rank (i.e. $h_i \neq 0$ for some i, $1 \leq i \leq n$). Then, the solution reduces to the familiar form:

$$\hat{u} = A'(AA')^{-1}\bar{y}.$$

By noting that $AA' = \sum_{i=1}^{n} h_i^2$, we can obtain $\hat{u}_j$ as follows;

$$\hat{u}_j = \frac{h_j \bar{y}}{\sum_{i=1}^{n} h_i^2}, \quad \text{for j} = 0, 1, \cdots, \text{n} - 1.$$

ii) a) Let's introduce e as an error such that $y_n = \overline{y} - e$. It can also be written as $\overline{y} - y_n = e$. Then now the quantity we would like to minimize can be written as

$$r(\overline{y} - y_n)^2 + u_0^2 + \cdots + u_{n-1}^2$$

where r is a positive weighting parameter. The problem becomes to solve the following minimization problem :

$$\hat{u} = arg \min_u \sum_{i=1}^{n} u_i^2 + re^2 = arg \min_u (\|\underline{u}\|_2^2 + r\|e\|_2^2),$$

from which we see that r is a weight that characterizes the tradeoff between the size of the final error, $\bar{y} - y_n$, and energy of the input signal, $\underline{u}$.

In order to reduce the problem into the familiar form, i.e, $\|y - Ax\|$, let's augment $\sqrt{r}e$ at the bottom of $\underline{u}$ so that a new augmented vector, $\underline{\tilde{u}}$ is

$$\underline{\tilde{u}} = \begin{bmatrix} \underline{u} \\ \cdots \\ \sqrt{r}e \end{bmatrix},$$

This choice of $\tilde{\underline{u}}$ follows from the observation that this is the $\tilde{\underline{u}}$ that would have $\|\tilde{\underline{u}}\|_2^2 = \|\underline{u}\|_2^2 + re^2$, the quantity we aim to minimize .

Now we can write $\overline{y}$ as follows

$$\overline{y} = \begin{bmatrix} A & \vdots & \frac{1}{\sqrt{r}} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \cdots \\ \sqrt{r}e \end{bmatrix} = \tilde{A}\tilde{\underline{u}} = A\underline{u} + e = y_n + e.$$

Now, $\hat{u}$ can be obtained using the augmented A, $\tilde{A}$, as

$$\hat{u} = \tilde{A}'(\tilde{A}\tilde{A}')^{-1}\overline{y} = \begin{bmatrix} A' \\ \frac{1}{\sqrt{r}} \end{bmatrix} \left[AA' + \frac{1}{r} \right] \overline{y}.$$

By noting that

$$AA' + \frac{1}{r} = \sum_{i=1}^{n} h_i^2 + \frac{1}{r},$$

we can obtain $\hat{u}_j$ as follows

$$\hat{u}_j = \frac{h_j \overline{y}}{\sum_{i=1}^{n} h_i^2 + \frac{1}{r}} \quad \text{for } \ \mathrm{j} = 0, \cdots, \mathrm{n} - 1.$$

ii) b) When $r = 0$, it can be interpreted that the error can be anything, but we would like to minimize the input energy. Thus we expect that the solution will have all the $u_i{}'s$ to be zero. In fact, the expression obtained in ii) a) will be zero as $r \to 0$. On the other hand, the other situation is an interesting case. We put a weight of ∞ to the final state error, then the expression from ii) a) gives the same expression as in i) as $r \to \infty$.

Exercise 3.3 This problem is similar to Example 3.4, except now we require that $\dot{p}(T) = 0$. We can derive, from $x(t) = \ddot{p}(t)$, that $p(t) = x(t) * tu(t) = \int_0^t (t - \tau)x(\tau)\mathrm{d}\tau$ where $*$ denotes convolution and $u(t)$ is the unit step, defined as 1 when $t > 0$ and 0 when $t < 0$. (One way to derive this is to take $x(t) = \ddot{p}(t)$ to the Laplace domain, *taking into account initial conditions*, to find the transfer function $H(s) = P(s)/X(s)$, hence the impulse response, $h(t)$ such that $p(t) = x(t) * h(t)$). Similarly, $\dot{p}(t) = x(t) * u(t) = \int_0^t x(\tau)\mathrm{d}\tau$. So, $y = p(T) = \int_0^T (T - \tau)x(\tau)\mathrm{d}\tau$ and $0 = \dot{p}(T) = x(t) * u(t) = \int_0^T x(\tau)\mathrm{d}\tau$. You can check that $< g(t), f(t) >= \int_0^T g(t)f(t)\mathrm{d}\tau$ is an inner product on the space of continuous functions on $[0, T]$, denoted by $C[0, T]$, which we are searching for $x(t)$. So, we have that $y = p(T) =< (T - t), x(t) >$ and $0 = \dot{p}(T) =< 1, x(t) >$. In matrix form,

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} < T - t, x(t) > \\ < 1, x(t) > \end{bmatrix} = \prec \begin{bmatrix} T - t & 1 \end{bmatrix}, x(t) \succ$$

where $\prec .,. \succ$ denotes the Grammian, as defined in chapter 2. Now, in chapter 3, it was shown that the minimum length solution to $y = \prec A, x \succ$, is $\hat{x} = A \prec A, A \succ^{-1} y$. So, for our problem,

$$\hat{x} = \begin{bmatrix} T - t & 1 \end{bmatrix} \prec \begin{bmatrix} T - t & 1 \end{bmatrix}, \begin{bmatrix} T - t & 1 \end{bmatrix} \succ^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix}.$$

Where, using the definition of the Grammian, we have that:

$$\prec \begin{bmatrix} T - t & 1 \end{bmatrix}, \begin{bmatrix} T - t & 1 \end{bmatrix} \succ = \begin{bmatrix} < T - t, T - t > & < T - t, 1 > \\ < 1, T - t > & < 1, 1 > \end{bmatrix}.$$

Now, we can use the definition for inner product to find the individual entries, $< T-t, T-t >= \int_0^T (T-t)^2 \mathrm{d}t = T^3/3$, $< T-t, 1 >= \int_0^T (T-t)\mathrm{d}t = T^2/2$, and $< 1, 1 >= T$. Plugging these in, one can simplify the expression for $\hat{x}$ and obtain $\hat{x}(t) = \frac{12y}{T^2}[\frac{1}{2} - \frac{t}{T}]$ for $t \in [0, T]$.

Alternatively, we have that $x(t) = \ddot{p}(t)$. Integrating both sides and taking into account that $p(0) = 0$ and $\dot{p}(0) = 0$, we have $p(t) = \int_0^t \int_0^{t_1} x(\tau)\mathrm{d}\tau \ \mathrm{d}t_1 = \int_0^t f(t_1)\mathrm{d}t_1$. Now, we use the integration by parts formula, $\int_0^t u \ \mathrm{d}v = uv|_0^t - \int_0^t v \ \mathrm{d}u$, with $u = f(t_1) = \int_0^{t_1} x(\tau)\mathrm{d}\tau$, and $\mathrm{d}v = \mathrm{d}t_1$; hence $\mathrm{d}u = \mathrm{d}f(t_1) = x(t_1)\mathrm{d}t_1$ and $v = t_1$. Plugging in and simplifying we get that $p(t) = \int_0^t \int_0^{t_1} x(\tau)\mathrm{d}\tau \ \mathrm{d}t_1 = \int_0^t (t-\tau)x(\tau)\mathrm{d}\tau$. Thus, $y = p(T) = \int_0^T (T-\tau)x(\tau)\mathrm{d}\tau =< T-t, x(t) >$. In addition, we have that $0 = \dot{p}(T) = \int_0^T x(\tau)\mathrm{d}\tau =< 1, x(t) >$. That is, we seek to find the minimum length $x(t)$ such that

$$\begin{aligned} y &= < T-t, x(t) > \\ 0 &= < 1, x(t) > . \end{aligned}$$

Recall that the minimum length solution $\hat{x}(t)$ must be a linear combination of $T - t$ and 1, i.e., $\hat{x}(t) = a_1(T - t) + a_2$. So,

$$\begin{aligned} y &= < T-t, a_1(T-t) + a_2 > &&= a_1 \int_0^T (T-t)^2 \mathrm{d}t + a_2 \int_0^T (T-t)\mathrm{d}t &&= a_1 \tfrac{T^3}{3} + a_2 \tfrac{T^2}{2} \\ 0 &= < 1, a_1(T-t) + a_2 > &&= \int_0^T (a_1(T-t) + a_2)\mathrm{d}t &&= a_1 \tfrac{T^2}{2} + a_2 T. \end{aligned}$$

This is a system of two equations and two unknowns, which we can rewrite in matrix form:

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix},$$

So,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix}^{-1} \begin{bmatrix} y \\ 0 \end{bmatrix}.$$

Exercise 4.1 Note that for any $v \in C^m$, (show this!)

$$\|v\|_\infty \leq \|v\|_2 \leq \sqrt{m}\|v\|_\infty. \tag{1}$$

Therefore, for $A \in C^{m\times n}$ with $x \in \mathbb{C}^n$

$$\|Ax\|_2 \leq \sqrt{m}\|Ax\|_\infty \to \text{for } x \neq 0, \tfrac{\|Ax\|_2}{\|x\|_2} \leq \sqrt{m}\tfrac{\|Ax\|_\infty}{\|x\|_2}.$$

But, from equation (1), we also know that $\frac{1}{\|x\|_\infty} \geq \frac{1}{\|x\|_2}$. Thus,

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \frac{\sqrt{m}\|Ax\|_\infty}{\|x\|_2} \leq \frac{\sqrt{m}\|Ax\|_\infty}{\|x\|_\infty} \leq \sqrt{m}\|A\|_\infty, \tag{2}$$

Equation (2) must hold for all $x \neq 0$, therefore

$$\max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2 \leq \sqrt{m}\|A\|_\infty.$$

To prove the lower bound $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2$, reconsider equation (1):

$$\|Ax\|_\infty \leq \|Ax\|_2 \rightarrow \text{for } x \neq 0, \frac{\|Ax\|_\infty}{\|x\|_2} \leq \frac{\|Ax\|_2}{\|x\|_2} \leq \|A\|_2 \rightarrow \frac{\sqrt{n}\|Ax\|_\infty}{\|x\|_2} \leq \frac{\sqrt{n}\|Ax\|_2}{\|x\|_2} \leq \sqrt{n}\|A\|_2. \quad (3)$$

But, from equation (1) for $x \in C^n$, $\frac{\sqrt{n}}{\|x\|_2} \geq \frac{1}{\|x\|_\infty}$. So,

$$\frac{\|Ax\|_\infty}{\|x\|_\infty} \leq \frac{\sqrt{n}\|Ax\|_\infty}{\|x\|_2} \leq \sqrt{n}\|A\|_2$$

for all $x \neq 0$ including x that makes $\frac{\|Ax\|_\infty}{\|x\|_\infty}$ maximum, so,

$$\max_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \|A\|_\infty \leq \sqrt{n}\|A\|_2,$$

or equivalently,

$$\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2.$$

Exercise 4.5 Any $m \times n$ matrix A, it can be expressed as

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V',$$

where U and V are unitary matrices. The "Moore-Penrose inverse", or *pseudo-inverse* of A, denoted by A^+, is then defined as the $n \times m$ matrix

$$A^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U'.$$

a) Now we have to show that A^+A and AA^+ are symmetric, and that $AA^+A = A$ and $A^+AA^+ = A^+$. Suppose that Σ is a diagonal invertible matrix with the dimension of $r \times r$. Using the given definitions as well as the fact that for a unitary matrix U, $U'U = UU' = I$, we have

$$\begin{aligned} AA^+ &= U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V'V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U' \\ &= U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} I \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U' \\ &= U \begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} U', \end{aligned}$$

which is symmetric. Similarly,

$$\begin{aligned}
A^+A &= V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U'U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V' \\
&= V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} I \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V' \\
&= V\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} V'
\end{aligned}$$

which is again symmetric.
The facts derived above can be used to show the other two.

$$\begin{aligned}
AA^+A &= (AA^+)A = U\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} U'A \\
&= U\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} U'U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V' \\
&= U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V' \\
&= A.
\end{aligned}$$

Also,

$$\begin{aligned}
A^+AA^+ &= (A^+A)A^+ = V\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} V'A^+ \\
&= V\begin{pmatrix} I_{r\times r} & 0 \\ 0 & 0 \end{pmatrix} V'V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U' \\
&= V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U' \\
&= A^+.
\end{aligned}$$

b) We have to show that when A has full column rank then $A^+ = (A'A)^{-1}A'$, and that when A has full row rank then $A^+ = A'(AA')^{-1}$. If A has full column rank, then we know that $m \geq n$, $rank(A) = n$, and

$$A = U\begin{pmatrix} \Sigma_{n\times n} \\ 0 \end{pmatrix} V'.$$

Also, as shown in chapter 2, when A has full column rank, $(A'A)^{-1}$ exists. Hence

$$\begin{aligned}
(A'A)^{-1}A' &= \left(V\begin{pmatrix} \Sigma' & 0 \end{pmatrix} U'U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V'\right)^{-1} V\begin{pmatrix} \Sigma' & 0 \end{pmatrix} U' \\
&= \left(V\Sigma'\Sigma V'\right)^{-1} V\begin{pmatrix} \Sigma' & 0 \end{pmatrix} U' \\
&= V(\Sigma'\Sigma)^{-1}V'V\begin{pmatrix} \Sigma' & 0 \end{pmatrix} U' \\
&= V(\Sigma'\Sigma)^{-1}\begin{pmatrix} \Sigma' & 0 \end{pmatrix} U' \\
&= V\begin{pmatrix} \Sigma^{-1} & 0 \end{pmatrix} U' \\
&= A^+.
\end{aligned}$$

Similarly, if A has full row rank, then $n \geq m$, $rank(A) = m$, and

$$A = U \begin{pmatrix} \Sigma_{m\times m} & 0 \end{pmatrix} V'.$$

It can be proved that when A has full row rank, $(A'A)^{-1}$ exists. Hence,

$$\begin{aligned} A'(AA')^{-1} &= V \begin{pmatrix} \Sigma' \\ 0 \end{pmatrix} U' \left(U \begin{pmatrix} \Sigma & 0 \end{pmatrix} V'V \begin{pmatrix} \Sigma' \\ 0 \end{pmatrix} U' \right)^{-1} \\ &= V \begin{pmatrix} \Sigma' \\ 0 \end{pmatrix} U' \left(U\Sigma\Sigma'U' \right)^{-1} \\ &= V \begin{pmatrix} \Sigma' \\ 0 \end{pmatrix} U'U(\Sigma\Sigma^{-1})U' \\ &= V \begin{pmatrix} \Sigma^{-1} \\ 0 \end{pmatrix} U' \\ &= A^+. \end{aligned}$$

c) Show that, of all x that minimize $\|y - Ax\|_2$, the one with the smallest length $\|x\|_2$ is given by $\hat{x} = A^+y$. If A has full row rank, we have shown in chapter 3 that the solution with the smallest length is given by

$$\hat{x} = A'(AA')^{-1}y,$$

and from part (b), $A'(AA')^{-1} = A^+$. Therefore

$$\hat{x} = A^+y.$$

Similary, it can be shown that the pseudo inverse is the solution for the case when a matrix A has a full column rank (compare the results in chapter 2 with the expression you found in part (b) for A^+ when A has full column rank).

Now, let's consider the case when a matrix A is rank deficient, *i.e.*, $rank(A) = r < \min(m, n)$ where $A \in C^{m\times n}$ and is thus neither full row or column rank. Suppose we have a singular value decomposition of A as

$$A = U\Sigma V',$$

where U and V are unitary matrices. Then the norm we are minimizing is

$$\|Ax - y\| = \|U\Sigma V'x - y\| = \|U(\Sigma V'x - U'y)\| = \|\Sigma z - U'y\|,$$

where $z = V'x$, since $\|\cdot\|$ is unaltered by the orthogonal transformation, U. Thus, x minimizes $\|Ax - y\|$ if and only if z minimizes $\|\Sigma z - c\|$, where $c = U'y$. Since the rank of A is r, the matrix Σ has the nonzero singular values $\sigma_1, \sigma_2, \cdots, \sigma_r$ in its diagonal entries. Then we can rewrite $\|\Sigma z - c\|^2$ as follows:

$$\|\Sigma z - c\|^2 = \sum_{i=1}^{r} (\sigma_i z_i - c_i)^2 + \sum_{i=r+1}^{n} c_i^2.$$

It is clear that the minimum of the norm can be achived when $z_i = \frac{c_i}{\sigma_i}$ for $i = 1, 2, \cdots, r$ and the rest of the z_i's can be chosen arbitrarily. Thus, there are infinitely many solutions $\hat{z}$ and the solution with the minimum norm can be achieved when $z_i = 0$ for $i = r+1, r+2, \cdots, n$. Thus, we can write this $\hat{z}$ as

$$z = \Sigma_1 c,$$

where

$$\Sigma_1 = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

and Σ_r is a square matrix with nonzero singular values in its diagonal in decreasing order. This value of z also yields the value of x of minimal 2 norm since V is a unitary matrix.
Thus the solution to this problem is

$$\hat{x} = Vz = V\Sigma_1 c = V\Sigma_1 U'y = A^+ y.$$

It can be easily shown that this choice of A^+ satisfies all the conditions, or definitions, of pseudo inverse in a).

Exercise 4.6. a) Suppose $A \in C_n^m$ has full column rank. Then QR factorization for A can be easily constructed from SVD:

$$A = U \begin{pmatrix} \Sigma_n \\ 0 \end{pmatrix} V'$$

where Σ_n is a $n \times n$ diagonal matrix with singular values on the diagonal. Let $Q = U$ and $R = \Sigma_n V'$ and we get the QR factorization. Since Q is an orthogonal matrix, we can represent any $Y \in C_p^m$ as

$$Y = Q \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Next

$$\|Y - AX\|_F^2 = \|Q \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} - Q \begin{pmatrix} R \\ 0 \end{pmatrix} X\|_F^2 = \|Q \begin{pmatrix} Y_1 - RX \\ Y_2 \end{pmatrix}\|_F^2$$

Denote

$$D = \begin{pmatrix} Y_1 - RX \\ Y_2 \end{pmatrix}$$

and note that multiplication by an orthogonal matrix does not change Frobenius norm of the matrix:

$$\|QD\|_F^2 = tr(D'Q'QD) = tr(D'D) = \|D\|_F^2$$

Since Frobenius norm squared is equal to sum of squares of all elements, square of the Frobenius norm of a block matrix is equal to sum of the squares of Frobenius norms of the blocks:

$$\|\begin{pmatrix} Y_1 - RX \\ Y_2 \end{pmatrix}\|_F^2 = \|Y_1 - RX\|_F^2 + \|Y_2\|_F^2$$

Since Y_2 block can not be affected by choice of X matrix, the problem reduces to minimization of $\|Y_1 - RX\|_F^2$. Recalling that R is invertible (because A has full column rank) the solution is

$$X = R^{-1} Y_1$$

b) Evaluate the expression with the pseudoinverse using the representations of A and Y from part a):

$$(A'A)^{-1} A'Y = \left(\begin{bmatrix} R' & 0 \end{bmatrix} Q'Q \begin{bmatrix} R \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} R' & 0 \end{bmatrix} Q'Q \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = R^{-1} (R')^{-1} \begin{bmatrix} R' & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = R^{-1} Y_1$$

From 4.5 b) we know that if a matrix has a full column rank, $A^+ = (A'A)^{-1} A'$, therefore both expressions give the same solutions.

c)

$$\|Y - AX\|_F^2 + \|Z - BX\|_F^2 = \| \begin{pmatrix} Y \\ Z \end{pmatrix} - \begin{pmatrix} A \\ B \end{pmatrix} X\|_F^2$$

Since A has full column rank, $\begin{pmatrix} A \\ B \end{pmatrix}$ also has full column rank, therefore we can apply results from parts a) and b) to conclude that

$$X = \left(\begin{pmatrix} A \\ B \end{pmatrix}' \begin{pmatrix} A \\ B \end{pmatrix} \right)^{-1} \begin{pmatrix} A \\ B \end{pmatrix}' \begin{pmatrix} Y \\ Z \end{pmatrix} = (A'A + B'B)^{-1} (A'Y + B'Z)$$

6.241J / 16.338J Dynamic Systems and Control
Spring 2011