ECE 275A Homework #4 Solutions – Fall 2011

Homework Solutions

1. (a) We have

$$
\begin{aligned}
\ell(x) &= x^H \Pi x - 2\mathrm{Re}\, x^H B y + y^H W y \\
&= x^H \Pi x - x^H B y - y^H B^H x + y^H W y \\
&= x^H \Pi x - x^H \Pi\, \Pi^{-1} B y - y^H B^H \Pi^{-1} \Pi x + y^H W y \\
&= \left(x - \Pi^{-1} B y\right)^H \Pi \left(x - \Pi^{-1} B y\right) + y^H W y - y^H B^H \Pi^{-1} B y \\
&= \left(x - \Pi^{-1} B y\right)^H \Pi \left(x - \Pi^{-1} B y\right) + y^H \left(W - B^H \Pi^{-1} B\right) y
\end{aligned}
$$

Thus for all x

$$\ell(x) \geq y^H \left(W - B^H \Pi^{-1} B\right) y$$

with equality if and only if $x = \Pi^{-1} B y$. Thus we have proved that

$$\hat{x} = \Pi^{-1} B y = \arg\min_x \ell(x)$$

$$\ell(\hat{x}) = y^H \left(W - B^H \Pi^{-1} B\right) y = \min_x \ell(x)$$

(b) It is straightforward to apply this result to the full column-rank, weighted least-squares problem.

$$
\begin{aligned}
\ell(x) &= \|y - Ax\|_W^2 = (y - Ax)^H W (y - Ax) \\
&= x^H \underbrace{A^H W A}_{\Pi} x - x^H \underbrace{A^H W}_{B} y - y^H \underbrace{W A}_{B^H} x + y^H W y \\
&= x^H \Pi x - x^H B y - y^H B^H x + y^H W y \\
&= x^H \Pi x - 2\mathrm{Re}\, x^H B y + y^H W y
\end{aligned}
$$

With A full column rank and $W = W^H > 0$, the matrix Π is Hermitian and full rank. Thus the weighted least-squares estimate of x is

$$\hat{x} = \Pi^{-1} B y = \left(A^H W A\right)^{-1} A^H W y$$

with optimal (minimal) least-squares cost

$$\ell(\hat{x}) = y^H \left(W - B^H \Pi^{-1} B\right) y = y^H \left(W - WA(A^H W A)^{-1} A^H W\right) y$$

Comment.

Suppose that

$$\langle y_1, y_2 \rangle = y_1^H W y_2 \quad \text{and} \quad \langle x_1, x_2 \rangle = x_1^H x^2 \quad \text{(i.e., } \Omega = I\text{)}$$

Then

$$A^* = A^H W$$
$$A^+ = (AA^*)^{-1}A^* = (A^H W A)^{-1}A^H W$$
$$P_{\mathcal{R}(A)} = AA^+ = A(A^H W A)^{-1}A^H W$$
$$P_{\mathcal{N}(A^*)} = I - P_{\mathcal{R}(A)}$$

This shows that the optimal cost can be rewritten as

$$\begin{aligned}\ell(\hat{x}) &= y^H W (I - P_{\mathcal{R}(A)})y \\ &= y^H W P_{\mathcal{N}(A^*)} y \\ &= \langle y, P^2_{\mathcal{N}(A^*)} \rangle \\ &= \langle P^*_{\mathcal{N}(A^*)} y, P_{\mathcal{N}(A^*)} \rangle \\ &= \langle P_{\mathcal{N}(A^*)} y, P_{\mathcal{N}(A^*)}\end{aligned}$$

or

$$\ell(\hat{x}) = \|P_{\mathcal{N}(A^*)} y\|^2 = \|P_{\mathcal{N}(A^*)} y\|_W^2$$

What is the optimal cost if $y \in \mathcal{R}(A)$? Does this make sense?

Note that the optimal error (which must be orthogonal to the range of A) is

$$\hat{e} = y - \hat{y} = y - P_{\mathcal{R}(A)} y = (I - P_{\mathcal{R}(A)})y = P_{\mathcal{N}(A^*)} y$$

Therefore the optimal cost can also be written as

$$\ell(\hat{x}) = \|\hat{e}\|^2 = \|\hat{e}\|_W^2$$

showing that the optimal least-squares error is the minimal residual error "power".

2. As in lecture, define the derivative with respect to a vector to be a *row operator*,

$$\frac{\partial}{\partial x} = \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n} \right).$$

An equivalent statement to Entry 1 in Table E.1, would be an identity involving the *undefined expression* $\frac{\partial}{\partial x} x^T A$ (the derivative of a row vector with respect to a vector), which is not sensible within *our framework* as it has been developed so far. To make this expression *sensible,* we would have to expand our framework to include new operations and/or new objects in addition to scalars, vectors, row-vectors, and matrices. The usual extension is the standard tensor calculus, which we will not discuss.[1]

[1]The derivative of a row-vector (a *covector* in tensor calculus parlance) by a vector is not a matrix which is a rank-2 tensor known as a $(1,1)$-tensor, but a *different* type of rank-2 tensor known as a $(0,2)$-tensor.

The equivalent identities to Equation (E.4) and Entry 2 of Table E.1,

$$\frac{\partial}{\partial x}\, c^T x = c^T \quad \text{and} \quad \frac{\partial}{\partial x}\, Ax = A$$

are both easily proved at the component level. We can also prove Entry 2 from Equation (E.4) in two ways. Either apply (E.4) component-wise to Ax or note that for an arbitrary vector d, we have $\frac{\partial}{\partial x} d^T Ax = d^T A$.

One can also easily prove the equivalent statement to Entry 4 at the component level. We have

$$\frac{\partial}{\partial x_k} \sum_i \sum_j a_{ij}\, x_i\, x_j = \sum_{i \neq k} a_{ik}\, x_i + \sum_{j \neq k} a_{kj}\, x_j + 2\, a_{kk}\, x_k = \sum_i a_{ik}\, x_i + \sum_j a_{kj}\, x_j$$

or

$$\frac{\partial}{\partial x}\, x^T Ax = x^T A + x^T A^T$$

which is the result to be proved. By setting $A = I$ we therefore have the equivalent result to Entry 3,

$$\frac{\partial}{\partial x}\, x^T x = \frac{\partial}{\partial x} \|x\|^2 = 2x^T \,.$$

By assuming that A is symmetric we obtain the equivalent result to Entry 5,

$$\frac{\partial}{\partial x}\, x^T Ax = \frac{\partial}{\partial x}\, \|x\|_A^2 = 2x^T A \,.$$

Finally, the chain rule for differentiating $z = z(y(x))$ with respect to x (which is equivalent to Entry 6),

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x} \,,$$

(equivalently, in terms of jacobian matrices, $J_{z \circ y} = J_z\, J_y$), follows from the component-level expression given at the very top of page 897 in Moon.

3. Note that Entry 2 in Table E.1 is not really well-defined within the vector-matrix framework developed in Moon & Stirling as they have not explained what it means to take the derivative of a *row vector* by a column vector.[2] To do so in a consistent manner

[2]In class we define how to take the derivative of a column vector $f(x)$ with respect to a column vector x as the action of the (*row*) partial derivative operator $\frac{\partial}{\partial x}$ on the vector $f(x)$. Similarly, Moon & Stirling also define the derivative of a column vector respect to a column vector—this is done right at the outset of Appendix E. Thereafter, all additional identities provided by Moon & Stirling should be consistent with that definition. According to Identities 1 and 2 of Table E.1, the derivative of the *column vector* $A^T x$ (note the transpose on A here) with respect to the column vector x and the derivative of the *row vector* $x^T A$ with respect to the column vector x *both* give the *same* mathematical object A. Since row and column vectors generally correspond to different mathematical objects, this result requires some explanation or

requires either extending the vector-matrix results of Appendix E or introducing tensor-like concepts.[3]

Also note that the last entry in Table E.1 uses an egregious notation in the parenthetical comment. A better notation is $z = z(y)$ and $y = y(x)$, in which case the given identity (which is proved on page 897 of Moon) makes sense.

For completeness, the second (as it stands, not well defined) entry in Table E.1 can be replaced by Equation E.3.

Note that with the Cartesian coordinate system assumption that $\Omega = I$, one can easily convert the row-vector derivative identities derived in the previous homework problem into the corresponding column-vector derivative identities by simply transposing the identities (possible with some minor renaming).

Alternatively, one can prove the column-vector derivative identities directly. From Moon & Stirling's Identity E.3 (which is proved in the book), Entry 1 of the table easily follows by noting that for an arbitrary vector c we have

$$\nabla_x \, c^T A x = \nabla_x \, (A^T c)^T x = A^T c$$

which (since c is arbitrary) yields the desired identity $\nabla_x \, Ax = A^T$.

Entry 4 is proved at the component level using the same argument given in the previous homework problem. Entries 3 and 5 of Table E.1 are just special cases of Entry 4.

4. See the Lecture Viewgraphs.

5. In the following Let $H = [h_1, \cdots, h_n]$.

   (a) Note that *as stated* this situation does *not* involve a linear mapping H and therefore there is *no* domain space and *no* codomain space.[4] What we do have, according to the problem statement, is an ambient Hilbert space $\mathcal{Y} = \mathbb{C}^m$ and an n-dimensional Hilbert subspace of this space, $\mathcal{H} \subset \mathcal{Y}$, given by

$$\mathcal{H} = \text{span}\,\{h_1, \cdots, h_n\} \ ,$$

interpretation. *In essence, the distinction between a row vector (an object which transforms covariantly) and a column vector (an object which transforms contravariantly) is being ignored in Moon & Stirling (which is ok in a Cartesian coordinate system).* In Moon & Stirling, the distinction between matrix A as a bilinear functional of two column vectors and as a bilinear functional on a row vector and a column vector has been blurred.

[3] The identity, properly interpreted, is not false. Without further elaboration, we just don't quite know what it means. Indeed, it can formally, and easily, be proved as follows: $\nabla_x x^T Ac = \nabla_x c^T A^T x = (A^T c)^T = c^T A$, which is true for *all* c, and therefore $\nabla_x x^T A = A$.

[4] The concepts of domain and codomain presuppose the existence of a mapping.

where h_i, $i = 1, \cdots, n$ are linearly independent (and therefore form a basis for $\mathcal{H}$).

The optimal approximation of x, $\hat{x} = \theta_1 h_1 + \cdots \theta_n h_n$, in the subspace spanned by $h_1, \cdots, h_n$ is determined from the orthogonality condition,

$$x - \hat{x} = x - (\theta_1 h_1 + \cdots \theta_n h_n) = x - H\theta \perp \text{Span}\{h_1, \cdots, h_n\} = \mathcal{R}(H)\,.$$

This condition is equivalent to the requirement that

$$\langle x - H\theta, h_i \rangle = (x - H\theta)^H C^{-1} h_i = 0\,, \quad \text{for } i = 1, \cdots, n\,,$$

or, equivalently,

$$(x - H\theta)^H C^{-1} [h_1, \cdots, h_n] = (x - H\theta)^H C^{-1} H = 0\,.$$

This yields the normal equations,

$$H^H C^{-1} H \hat{\theta} = H^H C^{-1} x\,,$$

so that,

$$\widehat{x} = H\hat{\theta} = H(H^H C^{-1} H)^{-1} H^H C^{-1} x = H H^{+} x = P_{\mathcal{R}(H)} x\,.$$

(b) This derivation should be standard for you by now.

(c) Here we need just need to expand the loss function $J = (x - H\theta)^H C^{-1} (x - H\theta)$ into separate terms and then make the identification with the terms of the resulting quadratic form exactly as done in the solution to Problem 1 given above.