Minimum Mean Squared Error Estimation: The Hilbert Space Approach

Keith M. Chugg ©

November 19, 1995

Contents

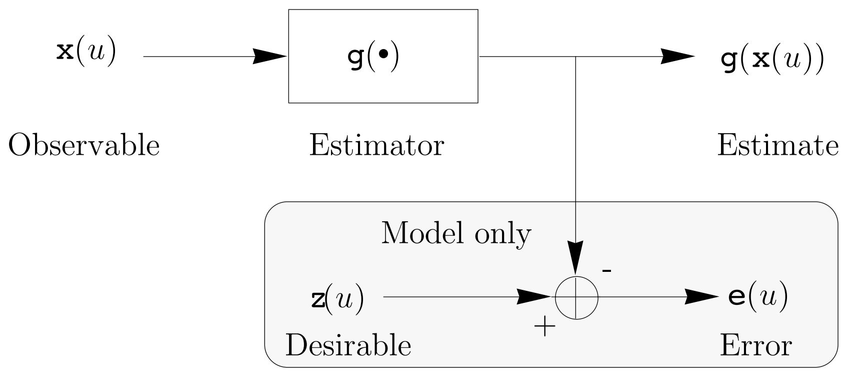

Figure 1: The general estimation problem.

1 Problem Statement

We are interested in estimating (or approximating) a *desired* random vector from another *observable* random vector. Throughout these notes we will denote the observable by $\mathbf{x}(u)$ and the desirable by $\mathbf{z}(u)$ (this is the same notation used the the supplemental notes by Scholtz). We will assume that $\mathbf{x}(u)$ is an $(m \times 1)$ complex random vector and that $\mathbf{z}(u)$ is an $(n \times 1)$ complex random vector. As an *estimate* of $\mathbf{z}(u)$, we use $\mathbf{g}(\mathbf{x}(u))$, where the *estimator* $\mathbf{g}(\cdot)$ is a deterministic mapping from $\mathcal{C}^m$ to $\mathcal{C}^n$. The problem is diagramed in Figure 1.

Typically the desired vector is not available; only the observable. The job of a good estimator is to extract information about the desirable ($\mathbf{z}(u)$) from the observable ($\mathbf{x}(u)$). In order to quantify the performance of a given estimator (i.e. what does good mean?), we must define a cost function. All of the estimation problems we will discuss in this class are concerned with the Mean-Squared-Error (MSE) cost function. For a given estimator, the MSE is defined by

$$\mathrm{MSE}(\mathbf{g}) = \mathbb{E}\left\{\|\mathbf{z}(u) - \mathbf{g}(\mathbf{x}(u))\|^2\right\} = \mathbb{E}\left\{\|\mathbf{e}(u)\|^2\right\} = \mathrm{tr}\,(\mathbf{R}_{\mathbf{e}}). \tag{1}$$

Where the error vector, $\mathbf{e}(u)$, is the difference between the desired signal and the estimate. We will denote the estimator which minimizes the MSE over some class of estimators by $\hat{\mathbf{z}}(u)$.

We will consider several *constraints* on the form of the estimator, $\mathbf{g}(\cdot)$, and solve for the best estimator by using one theorem, the Hilbert Space Projection Theorem (HSPT). This approach is rather abstract, yet fortunately we can easily check our results in each case to verify that it is indeed the best estimator.

2 The Hilbert Space Projection Theorem

In order to state the Hilbert Space Projection Theorem (HSPT), we need to define a Hilbert space.

Definition: A *Hilbert space* is a linear space with an inner product which is complete (i.e. a complete inner product space).

This is probably not too satisfying since we have not defined what "complete" means. We will define this concept when we discuss stochastic convergence later in the course - for now just consider it a property that makes taking limits easy. So a Hilbert space is just an inner product space with nice convergence properties. With this we can state the theorem which solves all the estimation problems in this course.

Theorem (HSPT): Let $\mathcal{H}$ be a Hilbert space, $\mathcal{M}$ be a closed subspace of $\mathcal{H}$, and $\boldsymbol{z} \in \mathcal{H}$. Then there is a *unique* $\hat{\boldsymbol{z}} \in \mathcal{M}$ which is closest to $\boldsymbol{z}$:

$$\|\boldsymbol{z} - \hat{\boldsymbol{z}}\| < \|\boldsymbol{z} - \boldsymbol{y}\| \quad \forall \ \boldsymbol{y} \in \mathcal{M}, \ \boldsymbol{y} \neq \hat{\boldsymbol{z}}.$$

Furthermore, a *necessary and sufficient* condition for $\hat{\boldsymbol{z}}$ to be the closest point is that it satisfy the *Orthogonality Principle:*

$$\langle \boldsymbol{z} - \hat{\boldsymbol{z}}, \boldsymbol{y} \rangle = 0 \quad \forall \ \boldsymbol{y} \in \mathcal{M}.$$

A direct result of this orthogonality condition is (see problem 4)

$$\|\boldsymbol{z} - \hat{\boldsymbol{z}}\|^2 = \|\boldsymbol{e}\|^2 = \|\boldsymbol{z}\|^2 - \|\hat{\boldsymbol{z}}\|^2.$$

Here $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ denotes the inner product defined on $\mathcal{H}$ and $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ is the associated norm (see problem 1).

A proof of this theorem is beyond the scope of this class, however the result is rather intuitive. To find the closest point from a given point to a closed subspace, we simply drop a perpendicular. This concept is illustrated in Figure 2.

The most important result to remember from the HSPT is the Orthogonality Principle; it allows use to solve for the closest point. This is demonstrated in the next subsection as well as in the solution of the MSE estimation problems.

2.1 Examples from the Prerequisites

Initially the HSPT seems very abstract, but you've actually used it many times in your previous studies. Three such applications are presented below.

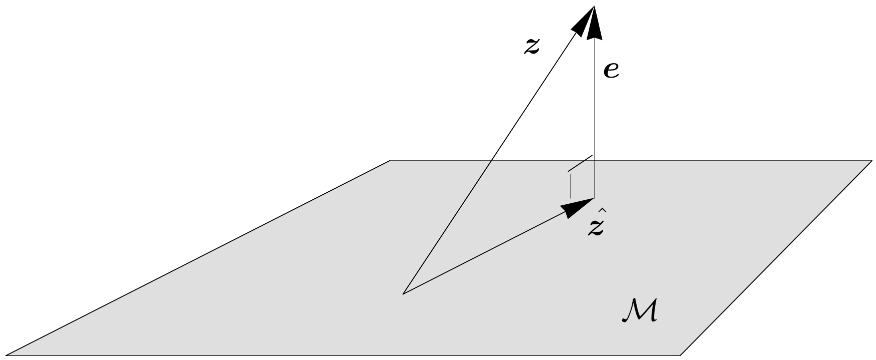

Figure 2: The HSPT - the error is orthogonal to the observation subspace.

2.1.1 Linear Estimation of a Deterministic Vector from Another

The space of all n-dimensional complex vectors, $\mathcal{C}^n$, is a Hilbert space, with inner product defined by the standard complex dot product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^\dagger \mathbf{x}. \tag{2}$$

Consider the problem of estimating one vector in $\mathcal{C}^n$, $\mathbf{z}$, by a scalar multiple of another, $\mathbf{x}$. This is the problem of minimizing $\|\mathbf{z} - \alpha \mathbf{x}\|$ over all $\alpha \in \mathcal{C}$. Thus we are trying to find the closest point in $\mathcal{M}$ to $\mathbf{z}$, where

$$\mathcal{M} = \{\alpha \mathbf{x} : \alpha \in \mathcal{C}\}. \tag{3}$$

The HSPT theorem says that there is a closest point in $\mathcal{M}$, namely $\hat{\mathbf{z}} = \alpha_{\mathrm{opt}} \mathbf{x}$. To find the closest point (i.e. solve for α_{opt}) we apply the Orthogonality Principle

$$(\alpha \mathbf{x})^\dagger (\mathbf{z} - \alpha_{\mathrm{opt}} \mathbf{x}) = 0 \ \ \forall \ \ \alpha \in \mathcal{C} \tag{4}$$

Therefore the optimal value of α satisfies

$$\alpha_{\mathrm{opt}} \mathbf{x}^\dagger \mathbf{x} = \mathbf{x}^\dagger \mathbf{z}. \tag{5}$$

If $\mathbf{x} \neq \mathbf{0}$, then

$$\alpha_{\mathrm{opt}} = \frac{\mathbf{x}^\dagger \mathbf{z}}{\mathbf{x}^\dagger \mathbf{x}} = \frac{\langle \mathbf{z}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle}, \tag{6}$$

so that the closest point is

$$\hat{\mathbf{z}} = \left(\frac{\mathbf{x}^\dagger \mathbf{z}}{\mathbf{x}^\dagger \mathbf{x}} \right) \mathbf{x} = \left(\frac{\mathbf{x}^\dagger \mathbf{z}}{\|\mathbf{x}\|} \right) \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{7}$$

You should recall this result from your linear algebra class.

2.1.2 Linear Estimation of a Deterministic Vector from Several Others

Again we will consider the application of the HSPT to the Hilbert Space of $\mathcal{C}^n$. This time we will try to approximate a given n-dimensional vector $\mathbf{b}$ by a linear combination of m other vectors in $\mathcal{C}^n$, $\mathbf{a}_1, \mathbf{a}_2 \ldots \mathbf{a}_m$. So in this case our subspace is

$$\mathcal{M} = \left\{ \mathbf{y} = \sum_{i=1}^{m} x_i \mathbf{a}_i : x_i \in \mathcal{C} \quad i = 1, 2, \ldots m \right\} \tag{8}$$
$$= \{\mathbf{y} = \mathbf{A}\mathbf{x} : \mathbf{x} \in \mathcal{C}^m\}, \tag{9}$$

where the $(n \times m)$ matrix $\mathbf{A}$ is defined by

$$\mathbf{A} = \left[\begin{array}{cccc} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_m \end{array} \right]. \tag{10}$$

You should recognize this as the standard problem of finding $\mathbf{x}$ so that $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ is minimized (i.e. the linear least-squares problem). Applying the Orthogonality Principle yields

$$(\mathbf{A}\mathbf{x})^\dagger (\mathbf{b} - \mathbf{A}\mathbf{x}_{\text{opt}}) = 0 \quad \forall \ \mathbf{x} \in \mathcal{C}^m. \tag{11}$$

This simplifies to

$$\mathbf{x}^\dagger [\mathbf{A}^\dagger (\mathbf{b} - \mathbf{A}\mathbf{x}_{\text{opt}})] = 0 \quad \forall \ \mathbf{x} \in \mathcal{C}^m, \tag{12}$$

which is true if and only if (see problem 5)

$$\mathbf{A}^\dagger \mathbf{A}\mathbf{x}_{\text{opt}} = \mathbf{A}^\dagger \mathbf{b}. \tag{13}$$

The relation in (13) is known as the "normal equations."

If the matrix $\mathbf{A}^\dagger \mathbf{A}$ is invertible, then we have

$$\mathbf{x}_{\text{opt}} = (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{b} \tag{14}$$
$$\hat{\mathbf{z}} = \mathbf{A}(\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{b}. \tag{15}$$

You should recognize this as the standard least-squares solution.

2.1.3 Fourier Series Expansion

The most powerful aspect of the Hilbert space approach is its generality. This allows us to solve problems in abstract spaces as well as concrete spaces like $\mathcal{C}^n$. In this example we consider the Hilbert space $\mathcal{L}_2[0, T]$, the space of all square integrable functions on the interval $[0, T]$. A point, $\boldsymbol{x}$, in $\mathcal{L}_2[0, T]$ actually represents an entire function

$$\boldsymbol{x} = \{x(t) : t \in [0, T]\}. \tag{16}$$

$\mathcal{L}_2[0, T]$ is the space of all functions which have finite energy, (i.e. they are square integrable)

$$\boldsymbol{x} \in \mathcal{L}_2[0, T] \Rightarrow \quad \int_0^T |x(t)|^2 dt < \infty. \tag{17}$$

The inner product for this Hilbert space is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int_0^T x(t) y^*(t) dt. \tag{18}$$

Given a point $\boldsymbol{z} \in \mathcal{L}_2[0,T]$, what is the closest function in the span of $\{\boldsymbol{e}_k\}_{k=-n}^n$, where $\boldsymbol{e}_k = \{e^{j2\pi \frac{k}{T} t} : t \in [0,T]\}$? Again, the HSPT solves this problem. The subspace in this case is

$$\mathcal{M} = \left\{ \boldsymbol{y} = \sum_{k=-n}^{n} c_k \boldsymbol{e}_k : c_k \in \mathcal{C} \;\; k = -n, \ldots n-1, n \right\}. \tag{19}$$

Denoting the optimal values of the coefficients by Z_k (i.e. $\hat{\boldsymbol{z}} = \sum_{k=-n}^n Z_k \boldsymbol{e}_k$), we can apply the Orthogonality Principle

$$\left\langle \boldsymbol{z} - \textstyle\sum_{k=-n}^n Z_k \boldsymbol{e}_k, \sum_{k=-n}^n c_k \boldsymbol{e}_k \right\rangle = 0 \quad \forall \;\; c_k \in \mathcal{C}, \;\; k = -n, \ldots n-1, n. \tag{20}$$

Simplifying,

$$\sum_{k=-n}^{n} c_k^* \langle \boldsymbol{z}, \boldsymbol{e}_k \rangle = \sum_{k=-n}^{n} \sum_{l=-n}^{n} c_k^* Z_l \langle \boldsymbol{e}_l, \boldsymbol{e}_k \rangle, \tag{21}$$

which holds for all choices of the c_k's. This simplifies because the the complex exponentials are orthogonal

$$\langle \boldsymbol{e}_l, \boldsymbol{e}_k \rangle = \int_0^T e^{j2\pi \frac{(l-k)}{T} t} dt \tag{22}$$

$$= T\delta_K(l-k) = \begin{cases} 0 & l \neq k \\ T & l = k \end{cases} \tag{23}$$

Using this, the Orthogonality Principle reduces to

$$Z_k = \frac{\langle \boldsymbol{z}, \boldsymbol{e}_k \rangle}{\langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle} = \frac{1}{T} \int_0^T z(t) e^{-j2\pi \frac{k}{T} t} dt. \tag{24}$$

So the best approximation is

$$\hat{z}(t) = \sum_{k=-n}^{n} Z_k e^{j2\pi \frac{k}{T} t}, \tag{25}$$

which are the first $2n+1$ terms of the Fourier Series. Thus we can view the Fourier Series as a result based on the HSPT.

3 Minimum MSE Estimation Using the HSPT

The first step in applying the HSPT to the estimation problem stated in Section 1 is to identify random vectors with a Hilbert space. We do this by first defining the Hilbert space of finite variance random variables and then extending the notion to random vectors.

The collection of all random variables defined on a given sample space, $\mathcal{U}$, forms a Hilbert space. We will denote this space by $\mathcal{W}_{\mathcal{U}}$, or just $\mathcal{W}$ for simplicity. A point in this space, $\boldsymbol{x}$, actually represents an entire random variable (a function)

$$\boldsymbol{x} \in \mathcal{W} \;\Rightarrow\; \boldsymbol{x} = \{x(u) : u \in \mathcal{U}\}. \tag{26}$$

The inner product for this Hilbert space and the implied norm are (see problem 1)

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{W}} = \mathbb{E}\left\{x(u)y^*(u)\right\} \tag{27a}$$

$$\|\boldsymbol{x}\|_{\mathcal{W}} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathcal{W}}} = \sqrt{\mathbb{E}\left\{|x(u)|^2\right\}}. \tag{27b}$$

This is the reason that $x(u)$ and $y(u)$ are referred to as orthogonal if $\mathbb{E}\left\{x(u)y^*(u)\right\} = 0$; in this case $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal in $\mathcal{W}$. This definition for the norm means that equality of two elements in $\mathcal{W}$ is almost sure equality of the random variables

$$\boldsymbol{x} = \boldsymbol{y} \quad \Longleftrightarrow \quad x(u) \overset{\text{as}}{=} y(u) \quad \Longleftrightarrow \quad \mathrm{P{\scriptsize R}}\left\{x(u) = y(u)\right\} = 1. \tag{28}$$

Since the space of all finite variance random variables is a Hilbert space, it is not surprising that the space of all n-dimensional second order random vectors is also a Hilbert space. We will denote this space as $\mathcal{W}_{\mathcal{U}}^n$ or just $\mathcal{W}^n$. A point in this space represents n random variables, each with finite second moments

$$\boldsymbol{x} \in \mathcal{W}^n \;\Rightarrow\; \boldsymbol{x} = \{\mathbf{x}(u) : u \in \mathcal{U}\}, \tag{29}$$

where

$$\mathbf{x}(u) = \begin{bmatrix} x(u,1) \\ x(u,2) \\ \vdots \\ x(u,n) \end{bmatrix}, \tag{30}$$

with $\mathbb{E}\left\{|x(u,i)|^2\right\} < \infty$ for $i = 1, 2 \ldots n$. The inner product for this Hilbert space and the implied norm are (see problem 2)

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{W}^n} = \mathbb{E}\left\{\mathbf{y}^\dagger(u)\mathbf{x}(u)\right\} = \mathrm{tr}\left(\mathbf{R}_{\mathbf{xy}}\right) \tag{31}$$

$$\|\boldsymbol{x}\|_{\mathcal{W}^n} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathcal{W}^n}} = \sqrt{\mathbb{E}\left\{\|\mathbf{x}(u)\|^2\right\}} = \sqrt{\mathrm{tr}\left(\mathbf{R}_{\mathbf{x}}\right)}. \tag{32}$$

Note that we use the notation $\|\mathbf{x}(u)\|^2 = \mathbf{x}^\dagger(u)\mathbf{x}(u)$, which is a random variable, to denote the standard Euclidean norm of the random vector $\mathbf{x}(u)$. This should not be confused with $\|\boldsymbol{x}\|_{\mathcal{W}^n}$, for $\boldsymbol{x} \in \mathcal{W}^n$, which is a deterministic scalar.[1] Once again equivalence of two points in $\mathcal{W}^n$ implies that the random vectors are equal almost surely.

Now that we can view second order random vectors as points in a Hilbert space, we can easily solve several MSE estimation problems. In each problem we will have the following framework:

[1]Scholtz uses the notation $|\mathbf{x}(u)|^2 = \mathbf{x}^\dagger(u)\mathbf{x}(u)$ in an attempt to avoid confusion. I prefer my notation because it emphasize the fact that for a fixed u, $\mathbf{x}(u)$ is a point in $\mathcal{C}^n$.

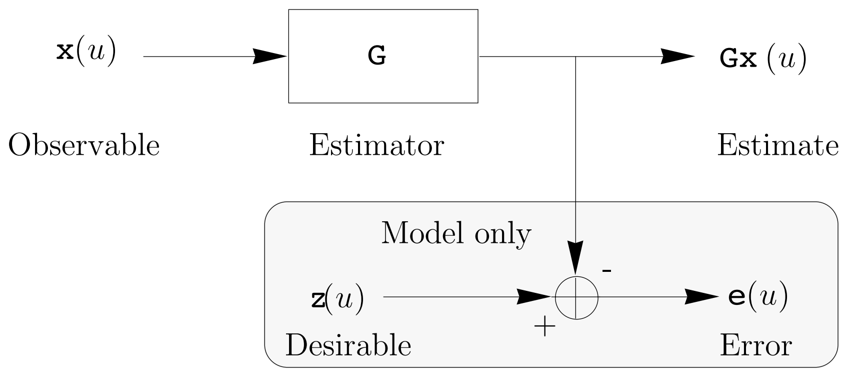

Figure 3: The linear estimation problem.

- We wish to estimate $\mathbf{z}(u)$ by some mapping of $\mathbf{x}(u)$, $\mathbf{g}(\mathbf{x}(u))$. We will view $\mathbf{z}(u)$ and $\mathbf{x}(u)$ as two points, $\boldsymbol{z}$ and $\boldsymbol{x}$, in the Hilbert space $\mathcal{W}^n$.
- The best estimator is the one that minimizes the MSE, namely $\mathbb{E}\left\{\|\mathbf{e}(u)\|^2\right\} = \|\boldsymbol{e}\|^2_{\mathcal{W}^n}$.
- We will solve this problem by noting that the solution is the point in

$$\mathcal{M} = \{\boldsymbol{y} \in \mathcal{W}^n : \mathbf{y}(u) = \mathbf{g}(\mathbf{x}(u))\}$$

which is closest to $\boldsymbol{z}$. In other words we want to choose $\mathbf{g}(\cdot)$ such that $\|\boldsymbol{z} - \boldsymbol{y}\|_{\mathcal{W}^n}$ is minimized over all $\boldsymbol{y} \in \mathcal{M}$. The solution is then obtained by applying the HSPT, specifically the Orthogonality Principle. The form of the solution depends on the *constraints* placed of the form of $\mathbf{g}(\cdot)$, and thus $\mathcal{M}$.

In the following subsections we solve several constrained MSE estimation problem following the above recipe.

3.1 Linear Minimum MSE (LMMSE) Estimation

The first constraint that we consider is that $\mathbf{g}(\cdot)$ must be linear. In this case we can represent $\mathbf{g}(\cdot)$ by matrix multiplication

$$\mathbf{y}(u) = \mathbf{g}(\mathbf{x}(u)) \qquad \text{linear} \Rightarrow \ \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u), \tag{33}$$

where $\mathbf{G}$ is an $(m \times n)$ matrix (remember $\mathbf{z}(u)$ is $(n \times 1)$ and $\mathbf{x}(u)$ is $(m \times 1)$. The block diagram for the LMMSE estimation problem is shown in Figure 3.

In the abstract Hilbert space model the subspace of all linear estimators based on the observation $\boldsymbol{x}$ is

$$\mathcal{M}_L = \{\boldsymbol{y} \in \mathcal{W}^n : \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u)\}. \tag{34}$$

The HSPT says the point in $\mathcal{M}_L$ which is closest to $\boldsymbol{z}$ is the unique $\hat{\boldsymbol{z}}$ which satisfies the Orthogonality Principle

$$\langle \boldsymbol{z} - \hat{\boldsymbol{z}}, \boldsymbol{y} \rangle_{\mathcal{W}^n} = 0 \quad \forall \;\; \boldsymbol{y} \in \mathcal{W}^n. \tag{35}$$

Equivalently, this means that there is an optimal choice for $\mathbf{G}$, $\mathbf{G}_{\text{opt}}$, so that $\hat{\mathbf{z}}(u) = \mathbf{G}_{\text{opt}}\mathbf{x}(u)$ is the LMMSE estimate of $\mathbf{z}(u)$ based on the observation $\mathbf{x}(u)$. The Orthogonality Principle can thus be rewritten as

$$\begin{aligned}
\mathbb{E}\left\{(\mathbf{G}\mathbf{x}(u))^\dagger(\mathbf{z}(u) - \mathbf{G}_{\text{opt}}\mathbf{x}(u))\right\} &= \operatorname{tr}\left(\mathbb{E}\left\{(\mathbf{z}(u) - \mathbf{G}_{\text{opt}}\mathbf{x}(u))(\mathbf{G}\mathbf{x}(u))^\dagger\right\}\right) && (36)\\
&= \operatorname{tr}\left((\mathbf{R}_{\mathbf{zx}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}})\mathbf{G}^\dagger\right) = 0, && (37)
\end{aligned}$$

which must hold for any choice of $\mathbf{G}$. Since this holds for arbitrary $\mathbf{G}$, it follows that a necessary and sufficient condition for $\mathbf{G}_{\text{opt}}$ to be the LMMSE estimator is (see problem 5)

$$\mathbf{R}_{\mathbf{zx}} = \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}. \tag{38}$$

This equation is known as the Wiener-Hopf equation. If $\mathbf{R}_{\mathbf{x}}$ is nonsingular, then the choice of $\mathbf{G}_{\text{opt}}$ is unique

$$\mathbf{G}_{\text{opt}} = \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}, \tag{39}$$

so that the LMMSE estimate is

$$\hat{\mathbf{z}}(u) = \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}(u). \tag{40}$$

The minimum MSE is also given by the HSPT

$$\begin{aligned}
\text{MSE}(\mathbf{G}_{\text{opt}}) &= \|\boldsymbol{z}\|^2_{\mathcal{W}^n} - \|\hat{\boldsymbol{z}}\|^2_{\mathcal{W}^n} && (41)\\
&= \operatorname{tr}(\mathbf{R}_{\mathbf{z}}) - \operatorname{tr}\left(\mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}\mathbf{G}^\dagger_{\text{opt}}\right) && (42)\\
&= \operatorname{tr}(\mathbf{R}_{\mathbf{z}}) - \operatorname{tr}\left(\mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{x}}(\mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1})^\dagger\right) && (43)\\
&= \operatorname{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{xz}}\right). && (44)
\end{aligned}$$

3.1.1 Direct Verification of the LMMSE Estimate

Since we haven't presented a proof of the HSPT, you may be weary of trusting the results. One nice aspect of the HSPT is that the results are simple to check once the answer is known.

Let $\mathbf{G}$ be any $(n \times m)$ matrix, then

$$\begin{aligned}
\text{MSE}(\mathbf{G}) &= \mathbb{E}\left\{\|\mathbf{z}(u) - \mathbf{G}\mathbf{x}(u)\|^2\right\} && (45)\\
&= \mathbb{E}\left\{\|(\mathbf{z}(u) - \mathbf{G}_{\text{opt}}\mathbf{x}(u)) + (\mathbf{G}_{\text{opt}} - \mathbf{G})\mathbf{x}(u)\|^2\right\} && (46)\\
&= \mathbb{E}\left\{\|(\mathbf{z}(u) - \mathbf{G}_{\text{opt}}\mathbf{x}(u))\|^2\right\} + \operatorname{tr}\left((\mathbf{G}_{\text{opt}} - \mathbf{G})\mathbf{R}_{\mathbf{x}}(\mathbf{G}_{\text{opt}} - \mathbf{G})^\dagger\right) && (47)\\
&\quad +2\Re\left\{\operatorname{tr}\left((\mathbf{R}_{\mathbf{zx}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}})(\mathbf{G}_{\text{opt}} - \mathbf{G})^\dagger\right)\right\}. && (48)
\end{aligned}$$

By design the cross term is zero (i.e. we choose $\mathbf{G}_{\rm opt}$ according to (38)). Thus

$$\begin{aligned} \mathrm{MSE}(\mathbf{G}) &= \mathrm{MSE}(\mathbf{G}_{\rm opt}) + \mathrm{tr}\left((\mathbf{G}_{\rm opt}-\mathbf{G})\mathbf{R}_{\mathbf{x}}(\mathbf{G}_{\rm opt}-\mathbf{G})^{\dagger}\right) && (49)\\ &\geq \mathrm{MSE}(\mathbf{G}_{\rm opt}), && (50)\end{aligned}$$

since $\mathrm{tr}\left((\mathbf{G}_{\rm opt}-\mathbf{G})\mathbf{R}_{\mathbf{x}}(\mathbf{G}_{\rm opt}-\mathbf{G})^{\dagger}\right) \geq 0$ follows from the fact that $\mathbf{R}_{\mathbf{x}}$ is non-negative definite (see problem 5). If $\mathbf{R}_{\mathbf{x}}$ is positive definite, then the choice of $\mathbf{G}_{\rm opt}$ is unique and the inequality in (50) is strict for $\mathbf{G} \neq \mathbf{G}_{\rm opt}$.

The value of $\mathrm{MSE}(\mathbf{G}_{\rm opt})$ is also easily calculated directly

$$\begin{aligned} \mathrm{MSE}(\mathbf{G}_{\rm opt}) &= \mathbb{E}\left\{\|\mathbf{z}(u)-\mathbf{G}_{\rm opt}\mathbf{x}(u)\|^2\right\} && (51)\\ &= \mathbb{E}\left\{(\mathbf{z}(u)-\mathbf{G}_{\rm opt}\mathbf{x}(u))^{\dagger}\mathbf{z}(u)\right\} - \underbrace{\mathbb{E}\left\{(\mathbf{z}(u)-\mathbf{G}_{\rm opt}\mathbf{x}(u))^{\dagger}\mathbf{G}_{\rm opt}\mathbf{x}(u)\right\}}_{=0} && (52)\\ &= \mathrm{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{R}_{\mathbf{zx}}\mathbf{G}_{\rm opt}^{\dagger}\right) && (53)\\ &= \mathrm{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{xz}}\right). && (54)\end{aligned}$$

The cross term is zero in the above as a result of the orthogonality principle (check it!).

3.2 Affine Minimum MSE (AMMSE) Estimation

In general, the LMMSE estimate is *biased*; that is $\mathbb{E}\{\hat{\mathbf{z}}(u)\} \neq \mathbf{m}_{\mathbf{z}}$ (see problem 6). If the estimator, $\mathbf{g}(\cdot)$ is constrained only to be affine, then this bias can be eliminated. An affine mapping has the representation

$$\mathbf{y}(u) = \mathbf{g}(\mathbf{x}(u)) \qquad \text{affine} \Rightarrow \ \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u) + \mathbf{c}, \tag{55}$$

where $\mathbf{G}$ is an $(m \times n)$ matrix and $\mathbf{c}$ is $(n \times 1)$. The block diagram for the AMMSE estimation problem is shown in Figure 4.

In the abstract Hilbert space model, the subspace of all affine estimators based on the observation $\boldsymbol{x}$ is

$$\mathcal{M}_A = \{\boldsymbol{y} \in \mathcal{W}^n : \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u) + \mathbf{c}\}. \tag{56}$$

Application of the HSPT implies that there is an optimal choice for $\mathbf{G}$ and $\mathbf{c}$, namely $\mathbf{G}_{\rm opt}$ and $\mathbf{c}_{\rm opt}$, so that $\hat{\mathbf{z}}(u) = \mathbf{G}_{\rm opt}\mathbf{x}(u) + \mathbf{c}_{\rm opt}$ is the AMMSE estimate of $\mathbf{z}(u)$ based on the observation $\mathbf{z}(u)$. The Orthogonality Principle implies

$$\mathbb{E}\left\{(\mathbf{G}\mathbf{x}(u)+\mathbf{c})^{\dagger}(\mathbf{z}(u)-\mathbf{G}_{\rm opt}\mathbf{x}(u)-\mathbf{c}_{\rm opt})\right\} = 0 \quad \forall \ \mathbf{G}, \ \mathbf{c}. \tag{57}$$

In particular, we can take $\mathbf{G} = \mathbf{O}$, so that

$$\mathbf{c}^{\dagger}(\mathbf{m}_{\mathbf{z}} - \mathbf{G}_{\rm opt}\mathbf{m}_{\mathbf{x}} - \mathbf{c}_{\rm opt}) = 0 \quad \forall \ \mathbf{c}. \tag{58}$$

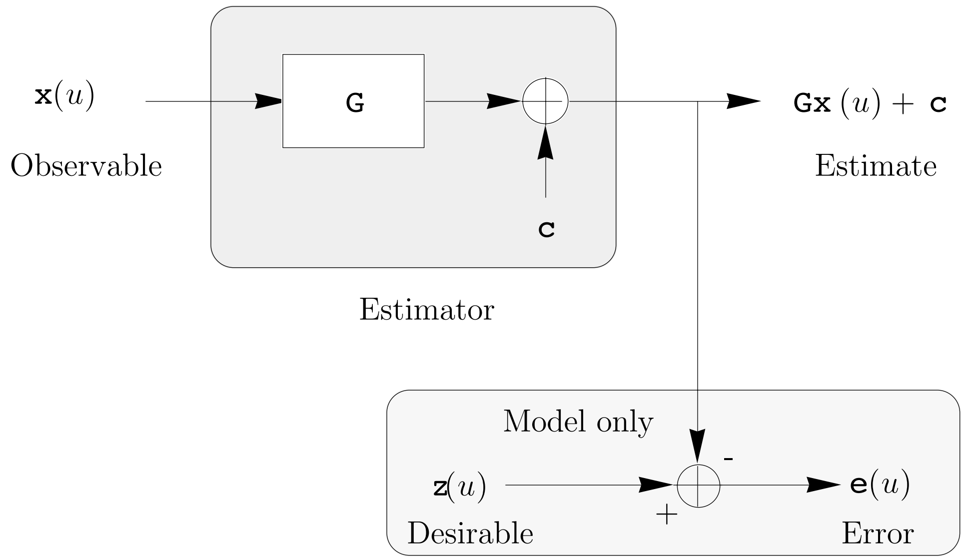

Figure 4: The affine estimation problem.

Since $\mathbf{c}$ is arbitrary, it follows that $\mathbf{m_z} - \mathbf{G}_{\text{opt}}\mathbf{m_x} - \mathbf{c}_{\text{opt}} = \mathbf{0}$, or

$$\mathbf{c}_{\text{opt}} = \mathbf{m_z} - \mathbf{G}_{\text{opt}}\mathbf{m_x}. \tag{59}$$

Substituting this back into (57) yields

$$\mathbb{E}\left\{(\mathbf{Gx}(u) + \mathbf{c})^{\dagger}(\mathbf{z}_0(u) - \mathbf{G}_{\text{opt}}\mathbf{x}_0(u))\right\} = 0 \quad \forall \;\; \mathbf{G},\ \mathbf{c}. \tag{60}$$

Taking $\mathbf{c} = -\mathbf{Gm_x}$ yields

$$\mathbb{E}\left\{(\mathbf{Gx}_0(u))^{\dagger}(\mathbf{z}_0(u) - \mathbf{G}_{\text{opt}}\mathbf{x}_0(u))\right\} = 0 \quad \forall \;\; \mathbf{G}, \tag{61}$$

which is can be viewed as a *linear* estimation problem (i.e. the LMMSE estimate of $\mathbf{z}_0(u)$ based on $\mathbf{x}_0(u)$). It follows that $\mathbf{G}_{\text{opt}}$ satisfies the Wiener-Hopf equation for affine estimators

$$\mathbf{K_{zx}} = \mathbf{G}_{\text{opt}}\mathbf{K_x}. \tag{62}$$

If $\mathbf{K_x}$ is invertible, then the AMMSE estimate is

$$\hat{\mathbf{z}}(u) = \mathbf{K_{zx}}\mathbf{K_x}^{-1}(\mathbf{x}(u) - \mathbf{m_x}) + \mathbf{m_z}, \tag{63}$$

and the corresponding minimum MSE is

$$\text{MSE}(\mathbf{G}_{\text{opt}}; \mathbf{c}_{\text{opt}}) = \text{tr}\left(\mathbf{K_z} - \mathbf{K_{zx}}\mathbf{K_x}^{-1}\mathbf{K_{xz}}\right). \tag{64}$$

3.3 Causal LMMSE Estimation

The class linear of linear estimators can be further constrained so that the estimator represents a causal linear operator; that is we can require $\mathbf{G}$ to be lower triangular. This restriction only makes most sense when the observation and the desired vector are the same dimension and both correspond to the same physical time set; therefore we will assume that $m = n$ for the causal constraint.

In this case the $\mathbf{g}(\cdot)$ has the following constraint

$$\mathbf{y}(u) = \mathbf{g}(\mathbf{x}(u)) \qquad \text{linear, causal} \Rightarrow \ \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u), \tag{65}$$

where $\mathbf{G}$ is an $(n \times n)$ lower triangular matrix.

In the abstract Hilbert space model the subspace of all linear causal estimators based on the observation $\boldsymbol{x}$ is

$$\mathcal{M}_{LC} = \{\boldsymbol{y} \in \mathcal{W}^n : \mathbf{y}(u) = \mathbf{G}\mathbf{x}(u) \ \mathbf{G} \text{ lower triangular}\}. \tag{66}$$

The Orthogonality Principle implies

$$\mathbb{E}\left\{(\mathbf{G}\mathbf{x}(u))^\dagger(\mathbf{z}(u) - \mathbf{G}_{\text{opt}}\mathbf{x}(u))\right\} = 0 \quad \forall \ \text{causal } \mathbf{G}, \tag{67}$$

which simplifies to

$$\operatorname{tr}\left((\mathbf{R}_{\mathbf{zx}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}})\mathbf{G}^\dagger\right) = 0 \quad \forall \ \text{causal } \mathbf{G}. \tag{68}$$

Since $\mathbf{G}$ is an arbitrary lower triangular matrix, it follows that (see problem 5)

$$\mathbb{C}\left\{\mathbf{R}_{\mathbf{zx}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}\right\} = \mathbf{O}, \tag{69}$$

or

$$\mathbb{C}\left\{\mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}\right\} = \mathbb{C}\left\{\mathbf{R}_{\mathbf{zx}}\right\} \tag{70}$$

where $\mathbb{C}\{\cdot\}$ is the "causal part operator"

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad \Rightarrow \quad \mathbb{C}\{\mathbf{A}\} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{71}$$

If $\mathbf{R}_{\mathbf{x}}$ is invertible, then the Cholesky factorization can be performed, so that

$$\mathbf{R}_{\mathbf{x}} = \mathbf{H}\mathbf{H}^\dagger, \tag{72}$$

where $\mathbf{H}$ is an invertible, lower triangular matrix. Since $\mathbf{G}$ is an arbitrary lower triangular matrix, it can be replaced in (68) by $\mathbf{F}\mathbf{H}^{-1}$, where $\mathbf{F}$ is any lower triangular matrix, resulting in

$$\operatorname{tr}\left((\mathbf{R}_{\mathbf{zx}}\mathbf{H}^{-\dagger} - \mathbf{G}_{\text{opt}}\mathbf{H})\mathbf{F}^\dagger\right) = 0 \quad \forall \ \text{causal } \mathbf{F}. \tag{73}$$

Since $\mathbf{F}$ is an arbitrary causal matrix, we arrive at

$$\mathbb{C}\left\{\mathbf{R}_{\mathbf{zx}}\mathbf{H}^{-\dagger} - \mathbf{G}_{\text{opt}}\mathbf{H}\right\} = \mathbf{O}, \tag{74}$$

or

$$\mathbf{G}_{\text{opt}} = \mathbb{C}\left\{\mathbf{R}_{\mathbf{zx}}\mathbf{H}^{-\dagger}\right\}\mathbf{H}^{-1}. \tag{75}$$

The corresponding minimum MSE is

$$\text{MSE}(\mathbf{G}_{\text{opt}}) = \text{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}\mathbf{G}_{\text{opt}}^{\dagger}\right). \tag{76}$$

The best causal affine estimator follows in a similar manner, the details are left to the student.

3.4 Unconstrained MMSE Estimation

Up to this point we have considered only constrained MMSE estimation problems. The question considered in this section is: Can we find the optimal estimator $\mathbf{g}(\cdot)$ with no constraints imposed? The answer is yes, but the cost of the optimal MMSE estimator is a *complete* statistical description of the observation and desirable.

Before stating the result, it is useful to review conditional expectation. The *conditional expectation mapping* of $\mathbf{z}(u)$ given $\mathbf{x}(u)$ is defined by

$$\mathbf{g}_{\mathbf{z}(u)|\mathbf{x}(u)}(\mathbf{x}) = \mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}\right\} = \int_{\mathcal{C}^n} \mathbf{z} f_{\mathbf{z}(u)|\mathbf{x}(u)}(\mathbf{z}|\mathbf{x}) d\mathbf{z}. \tag{77}$$

Notice that the conditional expectation mapping is a *deterministic mapping* from $\mathcal{C}^m$ to $\mathcal{C}^n$ (or $\mathcal{R}^m$ to $\mathcal{R}^n$ if $\mathbf{x}(u)$ and $\mathbf{z}(u)$are real random vectors). When we apply the this function to the m-dimensional random vector, $\mathbf{x}(u)$, we obtain an n-dimensional random vector denoted by

$$\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\} = \mathbf{g}_{\mathbf{z}(u)|\mathbf{x}(u)}(\mathbf{x}(u)). \tag{78}$$

We will refer to this random vector as the *conditional expectation* of $\mathbf{z}(u)$ given $\mathbf{x}(u)$. The distinction between the conditional expectation mapping and the conditional expectation is often not emphasized, since it should be clear from the context which is intended. In order to avoid confusion we will clearly distinguish between the two.

The unconstrained MMSE estimation problem is the problem stated in Figure 1, the subspace of unconstrained estimators is

$$\mathcal{M}_U = \{\boldsymbol{y} \in \mathcal{W}^n : \mathbf{y}(u) = \mathbf{g}(\mathbf{x}(u))\}. \tag{79}$$

The Orthogonality Principle requires that $\mathbf{g}_{\text{opt}}(\cdot)$ satisfies

$$\mathbb{E}\left\{(\mathbf{g}(\mathbf{x}(u)))^{\dagger}(\mathbf{z}(u) - \mathbf{g}_{\text{opt}}(\mathbf{x}(u)))\right\} = 0. \tag{80}$$

Conditioning on $\mathbf{x}(u)$ implies

$$\mathbb{E}_{\mathbf{x}(u)}\left\{\mathbb{E}\left\{\left[\mathbf{g}(\mathbf{x}(u))\right)^{\dagger}\right]\left[\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\} - \mathbf{g}_{\text{opt}}(\mathbf{x}(u))\right]\Big|\,\mathbf{x}(u)\right\}\right\} = 0, \tag{81}$$

since $\mathbb{E}_{\mathbf{x}(u)}\{\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\}\} = \mathbb{E}\left\{\mathbf{z}(u)\right\}$ and $\mathbb{E}_{\mathbf{x}(u)}\{\mathbf{g}(\mathbf{x}(u))|\mathbf{x}(u)\} = \mathbb{E}\left\{\mathbf{g}(\mathbf{x}(u))\right\}$ for any deterministic function $\mathbf{g}(\cdot)$. This must hold for all choices of $\mathbf{g}(\cdot)$; in particular it holds for $\mathbf{g}(\mathbf{x}) = \mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}\right\} - \mathbf{g}_{\text{opt}}(\mathbf{x})$. Substituting yields

$$\mathbb{E}_{\mathbf{x}(u)}\left\{\|\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\} - \mathbf{g}_{\text{opt}}(\mathbf{x}(u))\|^2\right\} = 0, \tag{82}$$

which implies that $\mathbf{g}_{\text{opt}}(\mathbf{x}) = \mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}\right\}$. So the optimal estimator is the conditional expectation mapping and the MMSE estimate is the conditional expectation

$$\hat{\mathbf{z}}(u) = \mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\}. \tag{83}$$

The associated MMSE is

$$\text{MSE}\Big(\mathbf{g}_{\text{opt}}\Big) = \text{tr}\left(\mathbf{R}_{\mathbf{z}}\right) - \mathbb{E}\left\{\|\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\}\|^2\right\}. \tag{84}$$

4 Summary of MMSE Estimation Results

Since the subspace of unconstrained estimators includes linear and affine estimator, one can never do worse with the conditional expectation than with a linear or affine estimator. A Venn diagram of the estimator subspaces is shown in Figure 5. From this diagram it follows that a linear estimator is never better than an affine estimator, which in turn is never better than the conditional expectation. The linear estimator is never worse than the causal linear estimator, with a similar result holding for affine estimators.

The results of the estimation problems considered are listed below

Constraint	Estimate	Min. MSE
Linear	$\hat{\mathbf{z}}(u) = \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{x}(u)$	$\text{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{R}_{\mathbf{zx}}\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{xz}}\right)$
Linear Causal	$\hat{\mathbf{z}}(u) = \mathbb{C}\left\{\mathbf{R}_{\mathbf{zx}}\mathbf{H}^{-\dagger}\right\}\mathbf{H}^{-1}\mathbf{x}(u)$	$\text{tr}\left(\mathbf{R}_{\mathbf{z}} - \mathbf{G}_{\text{opt}}\mathbf{R}_{\mathbf{x}}\mathbf{G}_{\text{opt}}^{\dagger}\right)$
Affine	$\hat{\mathbf{z}}(u) = \mathbf{K}_{\mathbf{zx}}\mathbf{K}_{\mathbf{x}}^{-1}\mathbf{x}_0(u) + \mathbf{m}_{\mathbf{z}}$	$\text{tr}\left(\mathbf{K}_{\mathbf{z}} - \mathbf{K}_{\mathbf{zx}}\mathbf{K}_{\mathbf{x}}^{-1}\mathbf{K}_{\mathbf{xz}}\right)$
Affine Causal	$\hat{\mathbf{z}}(u) = \mathbb{C}\left\{\mathbf{K}_{\mathbf{zx}}\mathbf{H}^{-\dagger}\right\}\mathbf{H}^{-1}\mathbf{x}_0(u) + \mathbf{m}_{\mathbf{z}}$	$\text{tr}\left(\mathbf{K}_{\mathbf{z}} - \mathbf{G}_{\text{opt}}\mathbf{K}_{\mathbf{x}}\mathbf{G}_{\text{opt}}^{\dagger}\right)$
Unconstrained	$\hat{\mathbf{z}}(u) = \mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\}$	$\text{tr}\left(\mathbf{R}_{\mathbf{z}}\right) - \mathbb{E}\left\{\|\mathbb{E}\left\{\mathbf{z}(u)|\mathbf{x}(u)\right\}\|^2\right\}$

In the table $\mathbf{H}$ is the Cholesky factor for $\mathbf{R}_{\mathbf{x}}$ in the linear causal estimator and the Cholesky factor for $\mathbf{K}_{\mathbf{x}}$ in the affine causal estimator. Also the results in the table assume that all the required inverses exist; if this is not so, the Wiener-Hopf equations must be solved directly (see problem 13)

5 Problems

1. This problem develops concepts related to inner product spaces. An *inner product* maps two elements of a linear space into a scalar, denoted by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. An inner product allows us to measure "angles" between elements of the space. An inner product must satisfy the following properties ($\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are arbitrary elements of the linear space and α is a complex scalar)

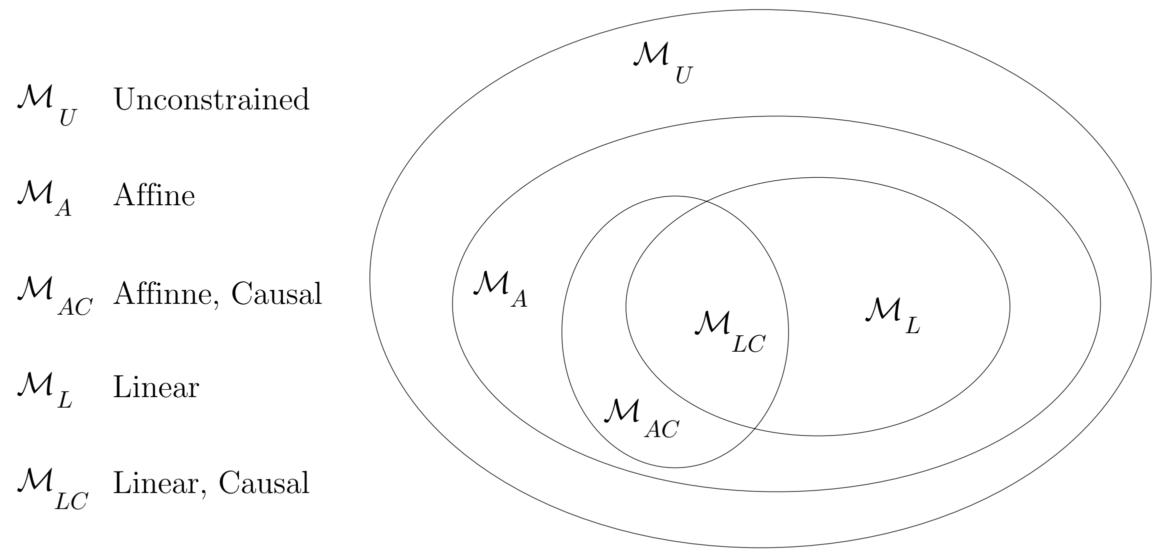

Figure 5: Venn diagram of estimator subspaces for various constraints

(IP1) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = (\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^*$ (Hermitian Symmetry)

(IP2) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$ (Additivity)

(IP3) $\langle \alpha \boldsymbol{x}, \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ (Homogeneity in first argument)

(IP4) if $\boldsymbol{x} \neq \boldsymbol{0}$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$ (Positivity).

(a) Prove the Cauchy-Schwartz inequality

$$|\langle \boldsymbol{z}, \boldsymbol{x} \rangle|^2 \leq \langle \boldsymbol{x}, \boldsymbol{x} \rangle \langle \boldsymbol{z}, \boldsymbol{z} \rangle.$$

<u>Hint:</u> For any α, $\langle \boldsymbol{z} - \alpha \boldsymbol{x}, \boldsymbol{z} - \alpha \boldsymbol{x} \rangle \geq 0$. Minimize this quantity with respect to α and show that the minimum value is still non-negative.

(b) A norm allows the measurement of "length" in the linear space. A norm maps an element of the linear space into a non-negative scalar (it's length), denoted by $\|\boldsymbol{x}\|$. A norm must satisfy the following four properties

(N1) $\|\boldsymbol{x}\| \geq 0$ (Positivity)

(N2) $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ (Subadditivity)

(N3) $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$ (Positive Homogeneity)

(N4) $\|\boldsymbol{x}\| = 0$ implies $\boldsymbol{x} = \boldsymbol{0}$ (Positive Definite).

Show that any inner product defines a norm by $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

<u>Hint:</u> For (N2), start with $\|\boldsymbol{x} + \boldsymbol{y}\|^2$, and use the fact that $\Re\{\alpha\} \leq |\alpha|$ for any α along with Cauchy-Schwartz.

(c) A distance function (also referred to as a metric) measures the distance between points of a (not necessarily Linear) space. The distance between two points is denoted $d(\boldsymbol{x}, \boldsymbol{y})$; it must satisfy the following properties

(D1) $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ (Positivity)

(D2) $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$ (Symmetry)

(D3) $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$ (Triangle Inequality)

(D4) $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ implies $\boldsymbol{x} = \boldsymbol{y}$ (Strict Positivity).

Show that any norm induces a distance defined by $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$.

(d) Show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{W}} = \mathbb{E}\{x(u)y^*(u)\}$ defines an inner product on the linear space $\mathcal{W}$, the space of all finite variance random variables.

- Why does equality in this space correspond to almost sure equality of the random variables?
- What is the Cauchy-Schwartz inequality in this case?
- What is the induced norm and distance for this space?

2. Show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{W}^n} = \mathbb{E}\left\{\mathbf{y}^\dagger(u)\mathbf{x}(u)\right\}$ defines an inner product on the linear space $\mathcal{W}^n$, the space of all second order random vectors of dimension n.

   - Why does equality in this space correspond to almost sure equality of the random vectors?
   - What is the Cauchy-Schwartz inequality in this case?
   - What is the induced norm and distance for this space?
   - Why is $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{W}^n} = \operatorname{tr}(\mathbf{R}_{\mathbf{xy}})$?

3. Show that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \int_0^T x(t)y^*(t)dt$ defines an inner product on the linear space $\mathcal{L}_2[0, T]$, the space of all square integrable functions on $[0, T]$ discussed in Section 2.1.3.

   - What is the Cauchy-Schwartz inequality in this case?
   - What is the induced norm and distance for this space?
   - What does equality mean in this space? Is there a physical interpretation?

4. Show that
$$\|\boldsymbol{z} - \hat{\boldsymbol{z}}\|^2 = \|\boldsymbol{e}\|^2 = \|\boldsymbol{z}\|^2 - \|\hat{\boldsymbol{z}}\|^2,$$
in the HSPT.

5. This problem fills in some facts used in the derivations of the MMSE estimators

   (a) Let $\mathbf{a} \in \mathcal{C}^n$ such that
$$\mathbf{b}^\dagger\mathbf{a} = 0 \quad \forall \ \mathbf{b} \in \mathcal{C}^n.$$
   Show that $\mathbf{a} = \mathbf{0}$.

(b) Let $\mathbf{G}$ be an arbitrary $(n \times m)$ matrix and $\mathbf{R_x}$ be the correlation matrix of a an $(m \times 1)$ random vector. Show that

$$\operatorname{tr}\left(\mathbf{G}\mathbf{R_x}\mathbf{G}^\dagger\right) \geq 0 \quad \forall \;\; \mathbf{G} \neq \mathbf{O}.$$

Show that the inequality is strict if $\mathbf{R_x}$ is invertible.

(c) Let $\mathbf{A}$ be an $(n \times m)$ matrix such that $\operatorname{tr}(\mathbf{AB}) = 0$ for any $(m \times n)$ matrix $\mathbf{B}$. Show that this implies $\mathbf{A} = \mathbf{O}$.

(d) Let $\mathbf{A}$ be an $(n \times n)$ matrix such that $\operatorname{tr}(\mathbf{AB}) = 0$ for any lower triangular $(n \times n)$ matrix $\mathbf{B}$. Show that this implies $\mathbb{C}\{\mathbf{A}\} = \mathbf{O}$.

6. An estimate is called unbiased if $\mathbb{E}\{\hat{\mathbf{z}}(u)\} = \mathbb{E}\{\mathbf{z}(u)\}$; otherwise the estimate biased. Show that the LMMSE estimator is biased in general and that the AMMSE estimator is unbiased. Under what condition is the LMMSE estimator unbiased?

7. Consider estimating $z(u)$ from of $x(u)$, where both are finite second moment random variables.

   (a) Show that the unconstrained minimum MSE estimate is unbiased; that is $\mathbb{E}\{\hat{z}(u)\} = \mathbb{E}\{z(u)\}$.

   (b) Show that $\mathbb{E}\{z(u)\hat{z}(u)\} = \mathbb{E}\{(\hat{z}(u))^2\}$, where $\hat{z}(u)$ is the unconstrained MMSE estimate.

8. Verify directly that the conditional expectation mapping is the unconstrained MMSE estimator.

9. Which estimator performs better: the LMMSE estimator or the Causal AMMSE estimator?

10. The observation, $\mathbf{x}(u)$, and desirable, $\mathbf{z}(u)$, are given by

$$\begin{aligned} \mathbf{x}(u) &= w(u)\mathbf{x}_1 + v(u)\mathbf{x}_2 \\ \mathbf{z}(u) &= v(u)\mathbf{z}_1, \end{aligned}$$

where

$$\mathbf{x}_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 8 \end{bmatrix} \quad \mathbf{z}_1 = \begin{bmatrix} 16 \\ 8 \end{bmatrix}$$

and the joint probability of $w(u)$ and $v(u)$ is given by

		$w(u)$	
		0	1
$v(u)$	0	$\frac{2}{8}$	$\frac{1}{8}$
	1	$\frac{2}{8}$	$\frac{3}{8}$

.

Find the minimum MSE estimator of $\mathbf{z}(u)$ based on $\mathbf{x}(u)$ and the corresponding minimum MSE for the following constraints:

(a) Linear

(b) Linear and Causal

(c) Affine

(d) Affine and Causal

(e) Unconstrained

(f) Rank the performance of these estimators.

11. Let $z(u)$ and $x(u)$ be jointly Gaussian real random variables. Denote the variance of $z(u)$ and $x(u)$ by σ_z^2 and σ_x^2, respectively and denote their means by m_z and m_x. Let the normalized covariance coefficient, κ_{zx}, be defined as[2]

$$\kappa_{zx} = \frac{\mathbb{E}\left\{(z(u) - m_z)(x(u) - m_x)\right\}}{\sigma_z \sigma_x}.$$

What is the unconstrained MMSE estimate of $z(u)$ based on the observation $x(u)$ (in terms of the given parameters)? What is the best affine estimator? What is the associated value of the minimum MSE for both of these estimators?

12. If $z(u)$ and $x(u)$ are statistically independent random variables, what is the best unconstained MMSE estimator of $z(u)$ based on the observation $x(u)$? Explain.

13. Consider finding the LMMSE estimate of $\mathbf{z}(u)$ based on $\mathbf{x}(u)$ when both are zero mean and $\mathbf{K}_{\mathbf{x}}$ is singular.

(a) Show that $\mathbf{G}_{\text{opt}} = \mathbf{K}_{\mathbf{zx}}\mathbf{K}_{\mathbf{x}}^{\mathrm{I}}$ (the pseudo-inverse) satisfies the Wiener-Hopf equation (and thus provides the LMMSE solution).

[2] Usually this is referred to as the normalized correlation coefficient and denoted by ρ, but the notation used here is more consistent with the definitions used in EE562a.

(b) The pseudo-inverse provides only one possible solution to the Wiener-Hopf equation. Describe some other solutions.

(c) The HSPT states that $\hat{\mathbf{z}}(u)$ should be unique, but for singular $\mathbf{K_x}$ $\mathbf{G}_{\text{opt}}$ is not unique. Explain this apparent contradiction.