Estimation techniques

March 2, 2006

Contents

1 Problem Statement

We are interested in estimating a *continuous* scalar[1] parameter $a \in \mathcal{A}$ from a vector observation $\mathbf{r}$. The observation and the parameter are related through a probabilistic mapping $p_{\mathbf{R}|A}(\mathbf{r}|a)$. a may or may not be a sample of a random variable. This leads to Bayesian and non-Bayesian estimators, respectively.

Special cases

- **Noisy observation model:**
$$\mathbf{r} = f(a, \mathbf{n})$$
where $\mathbf{n}$ (the noise component) is independent of a, and has a pdf $p_{\mathbf{N}}(\mathbf{n})$.
- **Linear model:** for a known $\mathbf{h}$,
$$\mathbf{r} = \mathbf{h}a + \mathbf{n}$$
with $\mathbf{n}$ independent of a
- **Linear Gaussian model:** linear model with $\mathbf{N} \sim \mathcal{N}(0, \Sigma_{\mathbf{N}})$, $A \sim \mathcal{N}(m_A, \Sigma_A)$

2 Bayesian Estimation Techniques

Here, $a \in A$ has a known a priori distribution $p_A(a)$. The most important Bayesian estimators are

- the MAP (maximum a posteriori) estimator
- the MMSE (minimum mean squared error) estimator
- the linear MMSE estimator

The latter two will be covered in this section. We remind that the MAP estimate is given by

$$\hat{a}_{MAP}(\mathbf{r}) = \arg\max_{a \in \mathcal{A}} p_{A|\mathbf{R}}(a|\mathbf{r}).$$

2.1 Minimum Mean Squared Error (MMSE) estimation

2.1.1 General formulation

The MMSE estimator minimizes the expected estimation error

$$\begin{aligned} \mathcal{C} &= \int_{\mathcal{A}} da \int_{\mathcal{R}} (a - \hat{a}(\mathbf{r}))^2 p_{A|\mathbf{R}}(a|\mathbf{r}) p_{\mathbf{R}}(\mathbf{r}) d\mathbf{r} \\ &= \mathbb{E}\left\{(A - \hat{a}(\mathbf{R}))^2\right\}. \end{aligned}$$

Taking the derivative (assuming this is possible) w.r.t. $\hat{a}(\mathbf{r})$ and setting the result to zero yields

$$\hat{a}(\mathbf{r}) = \int_A a \times p_{A|\mathbf{R}}(a|\mathbf{r}) da.$$

[1] We will only consider estimation of scalars. Extension to vectors is straightforward.

2.1.2 Important special case: Gaussian model

When A and $\mathbf{R}$ are jointly Gaussian, we can write

$$[A, \mathbf{R}] \sim \mathcal{N}(\mathbf{m}, \Sigma)$$

where

$$\mathbf{m} = \begin{bmatrix} m_A \\ \mathbf{m}_{\mathbf{R}} \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_A & \Sigma_{A\mathbf{R}}^T \\ \Sigma_{A\mathbf{R}} & \Sigma_{\mathbf{R}} \end{bmatrix}.$$

In that case it can be shown (after some straightforward manipulation), that

$$A|\mathbf{R} \sim \mathcal{N}\left(m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{r} - \mathbf{m}_{\mathbf{R}}), \Sigma_A - \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1} \Sigma_{A\mathbf{R}}\right)$$

so that

$$\hat{a}_{MMSE}(\mathbf{r}) \quad = \quad m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{r} - \mathbf{m}_{\mathbf{R}}) \tag{1}$$

Note that

- A posteriori (after observing $\mathbf{r}$), the uncertainty w.r.t. a is reduced from Σ_A to $\Sigma_A - \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1} \Sigma_{A\mathbf{R}} < \Sigma_A$.
- When A and $\mathbf{R}$ are independent $\Sigma_{A\mathbf{R}} = \mathbf{0}$, so that $\hat{a}_{MMSE}(\mathbf{r}) = m_A$
- For A and $\mathbf{R}$ jointly Gaussian, $\hat{a}_{MAP}(\mathbf{r}) = \hat{a}_{MMSE}(\mathbf{r})$.
- The estimator is linear in the observation

2.1.3 Example: Linear Gaussian Model

Problem: *Derive the MMSE estimator for the Linear Gaussian Model*
Solution: *We know that*

$$\mathbf{R} = \mathbf{h}A + \mathbf{N}$$

with $A \sim \mathcal{N}(m_A, \Sigma_A)$, and $\mathbf{N} \sim \mathcal{N}(0, \Sigma_{\mathbf{N}})$ with A and $\mathbf{N}$ independent. We also know that the MMSE estimator is given by:

$$\hat{a}_{MMSE}(\mathbf{r}) = m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{r} - \mathbf{m}_{\mathbf{R}}).$$

Let us compute $\mathbf{m}_{\mathbf{R}}$, $\Sigma_{\mathbf{R}}$ and $\Sigma_{A\mathbf{R}}$.

$$\begin{aligned} \mathbf{m}_{\mathbf{R}} &= \mathbb{E}\{\mathbf{R}\} \\ &= \mathbf{h} m_A \end{aligned}$$

$$\begin{aligned} \Sigma_{A\mathbf{R}} &= \mathbb{E}\{(\mathbf{R} - \mathbf{m}_{\mathbf{R}})(A - m_A)\} \\ &= \mathbb{E}\{(\mathbf{h}A + N - \mathbf{h}m_A)(A - m_A)\} \\ &= \mathbb{E}\{(\mathbf{h}(A - m_A) + \mathbf{N})(A - m_A)\} \\ &= \mathbf{h}\mathbb{E}\left\{(A - m_A)^2\right\} + \mathbb{E}\{\mathbf{N}(A - m_A)\} \\ &= \mathbf{h}\Sigma_A \end{aligned}$$

and

$$\begin{aligned} \Sigma_{\mathbf{R}} &= \mathbb{E}\left\{(\mathbf{R} - \mathbf{m}_{\mathbf{R}})(\mathbf{R} - \mathbf{m}_{\mathbf{R}})^T\right\} \\ &= \mathbb{E}\left\{(\mathbf{h}(A - m_A) + \mathbf{N})(\mathbf{h}(A - m_A) + \mathbf{N})^T\right\} \\ &= \mathbf{h}\mathbb{E}\left\{(A - m_A)^2\right\}\mathbf{h}^T + \mathbb{E}\left\{\mathbf{N}\mathbf{N}^T\right\} \\ &= \mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}. \end{aligned}$$

This leads to

$$\begin{aligned}\hat{a}_{MMSE}(\mathbf{r}) &= m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{r}-\mathbf{m}_{\mathbf{R}}) \\ &= m_A + \Sigma_A \mathbf{h}^T \left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}\right)^{-1}(\mathbf{r}-\mathbf{h}m_A)\end{aligned}$$

2.2 Linear estimation

In some cases, it is preferred to have an estimator which is a linear function of the observation:

$$\hat{a}(\mathbf{r}) = \mathbf{b}^T\mathbf{r} + c$$

so that $\hat{a}(\mathbf{r})$ is obtained through an affine transformation of the observation. Clearly, when A and $\mathbf{R}$ are jointly Gaussian, the MMSE estimator is a linear estimator with

$$\mathbf{b}^T = \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}$$

and

$$c = m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{m}_{\mathbf{R}}).$$

2.3 Linear MMSE Estimator

When A and $\mathbf{R}$ are *not* jointly Gaussian, we can still use (1) as an estimator. This is known as the Linear MMSE (LMMSE) estimator:

$$\hat{a}_{LMMSE}(\mathbf{r}) = m_A + \Sigma_{A\mathbf{R}}^T \Sigma_{\mathbf{R}}^{-1}(\mathbf{r}-\mathbf{m}_{\mathbf{R}})$$

Theorem: We wish to find an estimator with the following properties

- the estimator must be linear in the observation
- $\mathbb{E}\{\hat{a}(\mathbf{r})\} = m_A$ (this can be interpreted as a form of unbiasedness)
- the estimator has the minimum MSE among all linear estimators

Then the estimator is the LMMSE estimator.
Proof: Given a generic linear estimator

$$\hat{a}(\mathbf{r}) = \mathbf{b}^T\mathbf{r} + c \tag{2}$$

we would like to find $\mathbf{b}$ and c such that the resulting estimator is unbiased and has minimum variance. Clearly

$$\mathbb{E}\{\hat{a}(\mathbf{r})\} = \mathbf{b}^T\mathbf{m}_{\mathbf{R}} + c$$

so that $\mathbf{b}$ and c have to satisfy

$$\begin{aligned}\mathbf{b}^T\mathbf{m}_{\mathbf{R}} + c &= m_A \\ c &= m_A - \mathbf{b}^T\mathbf{m}_{\mathbf{R}} \\ c^2 &= m_A^2 + \mathbf{b}^T\mathbf{m}_{\mathbf{R}}\mathbf{m}_{\mathbf{R}}^T\mathbf{b} - 2m_A\mathbf{b}^T\mathbf{m}_{\mathbf{R}}\end{aligned}$$

We also know that

$$\begin{aligned}\Sigma_{\mathbf{R}} &= \mathbb{E}\left\{(\mathbf{R}-\mathbf{m}_{\mathbf{R}})(\mathbf{R}-\mathbf{m}_{\mathbf{R}})^T\right\} \\ &= \mathbb{E}\left\{\mathbf{R}\mathbf{R}^T\right\} - \mathbf{m}_{\mathbf{R}}\mathbf{m}_{\mathbf{R}}^T\end{aligned}$$

and

$$\begin{aligned}\Sigma_{A\mathbf{R}} &= \mathbb{E}\{(\mathbf{R}-\mathbf{m}_{\mathbf{R}})(A-m_A)\} \\ &= \mathbb{E}\{\mathbf{R}A\} - \mathbf{m}_{\mathbf{R}}m_A.\end{aligned}$$

Let us now look at the variance of the estimation error (the estimation error is zero-mean):

$$\begin{aligned}
V &= \mathbb{E}\left\{(\hat{a}(\mathbf{R}) - A)^2\right\} \\
&= \mathbb{E}\left\{\left(\mathbf{b}^T\mathbf{R} + c - A\right)^2\right\} \\
&= \mathbf{b}^T\mathbb{E}\left\{\mathbf{R}\mathbf{R}^T\right\}\mathbf{b} + \mathbb{E}\left\{A^2\right\} + c^2 + 2c\mathbf{b}^T\mathbf{m}_R - 2cm_A - 2\mathbf{b}^T\mathbb{E}\left\{\mathbf{R}A\right\} \\
&= \mathbf{b}^T\left(\Sigma_{\mathbf{R}} + \mathbf{m}_{\mathbf{R}}\mathbf{m}_{\mathbf{R}}^T\right)\mathbf{b} + \Sigma_A + m_A^2 + c^2 + 2c\left(\mathbf{b}^T\mathbf{m}_R - m_A\right) - 2\mathbf{b}^T\left(\Sigma_{A\mathbf{R}} + \mathbf{m}_{\mathbf{R}}m_A\right) \\
&= \mathbf{b}^T\left(\Sigma_{\mathbf{R}} + \mathbf{m}_{\mathbf{R}}\mathbf{m}_{\mathbf{R}}^T\right)\mathbf{b} + \Sigma_A + m_A^2 - c^2 - 2\mathbf{b}^T\left(\Sigma_{A\mathbf{R}} + \mathbf{m}_{\mathbf{R}}m_A\right) \\
&= \mathbf{b}^T\Sigma_{\mathbf{R}}\mathbf{b} + \Sigma_A - 2\mathbf{b}^T\Sigma_{A\mathbf{R}}
\end{aligned}$$

where we have used the fact that $c^2 = m_A^2 + \mathbf{b}^T\mathbf{m}_{\mathbf{R}}\mathbf{m}_{\mathbf{R}}^T\mathbf{b} - 2m_A\mathbf{b}^T\mathbf{m}_{\mathbf{R}}$. Taking derivative w.r.t. $\mathbf{b}$ and equating the result to zero, yields[2]

$$2\Sigma_{\mathbf{R}}\mathbf{b} - 2\Sigma_{A\mathbf{R}} = 0$$

and thus (since $\Sigma_{\mathbf{R}}^{-1}$ is symmetric)

$$\mathbf{b} = \Sigma_{\mathbf{R}}^{-1}\Sigma_{A\mathbf{R}} \tag{3}$$

and

$$\begin{aligned}
c &= m_A - \mathbf{b}^T\mathbf{m}_{\mathbf{R}} \\
&= m_A - \Sigma_{A\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{m}_{\mathbf{R}}
\end{aligned} \tag{4}$$

Substitution of (3) and (4) into (2) gives us the final result

$$\hat{a}(\mathbf{r}) = m_A + \Sigma_{A\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}(\mathbf{r} - \mathbf{m}_{\mathbf{R}})$$

which is clearly the desired result. QED

2.4 Orthogonality Condition

2.4.1 Principle

The orthogonality condition is useful in deriving (L-)MMSE estimators.
Theorem: for LMMSE estimation, the estimation error is (statistically) orthogonal to the observation when $\mathbf{m}_{\mathbf{R}} = \mathbf{0}$ and $m_A = 0$:
Proof:

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{R}(\hat{a}(\mathbf{R}) - A)\right\} &= \mathbb{E}\left\{\mathbf{R}\left(\Sigma_{A\mathbf{R}}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{R} - A\right)\right\} \\
&= \mathbb{E}\left\{\mathbf{R}\mathbf{R}^T\right\}\left(\Sigma_{\mathbf{R}}^{-1}\Sigma_{A\mathbf{R}}\right) - \mathbb{E}\left\{\mathbf{R}A\right\} \\
&= \Sigma_{\mathbf{R}}\left(\Sigma_{\mathbf{R}}^{-1}\Sigma_{A\mathbf{R}}\right) - \Sigma_{A\mathbf{R}} \\
&= \mathbf{0}.
\end{aligned}$$

QED.

2.4.2 Example: Linear Model

Problem: *Verify the orthogonality condition assuming* $m_A = 0$.
Solution: *For* $m_A = 0$, *we need to verify that* $\mathbb{E}\left\{\mathbf{R}(\hat{a}(\mathbf{R}) - A)\right\} = \mathbf{0}$, *we see that we need to verify that*

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{R}\hat{a}(\mathbf{R})\right\} &= \mathbb{E}\left\{\mathbf{R}A\right\} \\
&= \mathbf{h}\Sigma_A
\end{aligned}$$

[2] To see this requires some knowledge of vector calculus: differentiation w.r.t. $\mathbf{x}$ gives us $\mathbf{x}^T\mathbf{A}\mathbf{x} \rightarrow 2\mathbf{A}\mathbf{x}$ and $\mathbf{x}^T\mathbf{y} \rightarrow \mathbf{y}$.

$$
\begin{aligned}
\mathbb{E}\left\{\mathbf{R}\hat{a}\left(\mathbf{R}\right)\right\} &= \mathbb{E}\left\{\mathbf{R}\left(\Sigma_A\mathbf{h}^T\left(\mathbf{h}\Sigma_A\mathbf{h}^T+\Sigma_{\mathbf{N}}\right)^{-1}\mathbf{R}\right)\right\} \\
&= \mathbb{E}\left\{\mathbf{R}\mathbf{R}^T\right\}\left(\mathbf{h}\Sigma_A\mathbf{h}^T+\Sigma_{\mathbf{N}}\right)^{-1}\Sigma_A\mathbf{h} \\
&= \left(\mathbf{h}\Sigma_A\mathbf{h}^T+\Sigma_{\mathbf{N}}\right)\left(\mathbf{h}\Sigma_A\mathbf{h}^T+\Sigma_{\mathbf{N}}\right)^{-1}\Sigma_A\mathbf{h} \\
&= \Sigma_A\mathbf{h}
\end{aligned}
$$

3 Non-Bayesian Estimation Techniques

The above techniques cannot be applied when we do not consider A to be a random variable. Another approach is called for, where we would like unbiased estimators with small variances. The most important non-Bayesian estimators are

- the ML (maximum likelihood) estimator
- the BLU (best linear unbiased) estimator
- the LS (least squares) estimator

The latter two will be covered in this section. All expectations are taken for a given value of a. We remind that the ML estimate is given by

$$\hat{a}_{ML}\left(\mathbf{r}\right) = \arg\max_{a\in\mathcal{A}} p_{\mathbf{R}|A}\left(\mathbf{r}\left|a\right.\right).$$

3.1 The BLU estimator

As before, we consider a generic linear estimator:

$$\hat{\mathbf{a}}\left(\mathbf{r}\right) = \mathbf{b}^T\mathbf{r} + c.$$

To obtain an unbiased estimator, we need to assume that $c = 0$ and that

$$\mathbb{E}\left\{\mathbf{R}\right\} = \mathbf{h}a$$

for some known vector $\mathbf{h}$. We introduce

$$
\begin{aligned}
\Sigma_{\mathbf{R}} &= \mathbb{E}\left\{\left(\mathbf{R}-\mathbb{E}\left\{\mathbf{R}\right\}\right)\left(\mathbf{R}-\mathbb{E}\left\{\mathbf{R}\right\}\right)^T\right\} \\
&= \mathbb{E}\left\{\mathbf{R}\mathbf{R}^T\right\} - a^2\mathbf{h}\mathbf{h}^T
\end{aligned}
$$

which is assumed to be known to the estimator. Our goal is to find a $\mathbf{b}$ that leads to an unbiased estimator with minimal variance for all a.

Unbiased

An unbiased estimator must satisfy $\mathbb{E}\left\{\hat{a}\left(\mathbf{r}\right)\right\} = a$, for all $a \in \mathcal{A}$, so that

$$\mathbf{b}^T\mathbf{h} = 1.$$

Variance of Estimation Error

The variance of the estimation error is given by

$$\begin{aligned} V_a &= \mathbb{E}\left\{(\hat{a}(\mathbf{R}) - a)^2\right\} \\ &= \mathbb{E}\left\{\left(\mathbf{b}^T\mathbf{r} - a\right)^2\right\} \\ &= \mathbb{E}\left\{\left(\mathbf{b}^T(\mathbf{r} - \mathbf{h}a)\right)^2\right\} \\ &= \mathbf{b}^T\mathbb{E}\left\{(\mathbf{r} - \mathbf{h}a)^2\right\}\mathbf{b} \\ &= \mathbf{b}^T\Sigma_{\mathbf{R}}\mathbf{b}. \end{aligned}$$

This leads us to the following optimization problem: find $\mathbf{b}$ that minimizes V_a, subject to $\mathbf{b}^T\mathbf{h} = 1$. Using a Lagrange multiplier technique, we find we need to minimize

$$\mathbf{b}^T\Sigma_{\mathbf{R}}\mathbf{b} - \lambda\left(\mathbf{b}^T\mathbf{h} - 1\right).$$

This leads to

$$\mathbf{b} = \Sigma_{\mathbf{R}}^{-1}\mathbf{h}\lambda.$$

With the constraint $\mathbf{b}^T\mathbf{h} = 1$ giving rise to

$$\lambda\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h} = 1$$

so that

$$\mathbf{b} = \Sigma_{\mathbf{R}}^{-1}\mathbf{h}\left(\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\right)^{-1}$$

and

$$\begin{aligned} V_a &= \left(\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\right)^{-1}\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\Sigma_{\mathbf{R}}\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\left(\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\right)^{-1} \\ &= \left(\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\right)^{-1}. \end{aligned}$$

3.1.1 Example: Linear Gaussian Model

Problem: *derive the BLU estimator for the Linear Gaussian Model*
Solution:

$$\hat{a}_{BLU}(\mathbf{r}) = \left(\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{h}\right)^{-1}\mathbf{h}^T\Sigma_{\mathbf{R}}^{-1}\mathbf{r}$$

where

$$\begin{aligned} \Sigma_{\mathbf{R}} &= \mathbb{E}\left\{(\mathbf{R} - \mathbf{m}_{\mathbf{R}})(\mathbf{R} - \mathbf{m}_{\mathbf{R}})^T\right\} \\ &= \mathbb{E}\left\{(\mathbf{R} - \mathbf{h}a)(\mathbf{R} - \mathbf{h}a)^T\right\} \\ &= \mathbb{E}\left\{(\mathbf{h}a + N - \mathbf{h}a)(\mathbf{h}a + \mathbf{N} - \mathbf{h}a)^T\right\} \\ &= \Sigma_{\mathbf{N}}. \end{aligned}$$

Hence

$$\hat{a}_{BLU}(\mathbf{r}) = \left(\mathbf{h}^T\Sigma_{\mathbf{N}}^{-1}\mathbf{h}\right)^{-1}\mathbf{h}^T\Sigma_{\mathbf{N}}^{-1}\mathbf{r}$$

3.2 The LS estimator

If we refer back to our problem statement, the LS estimator tries to find a that minimizes the distance the between the observation and the 'reconstructed' noiseless observation:

$$d = \mathbb{E}\left\{\left\|\mathbf{r} - h\left(a, \mathbf{0}\right)\right\|^2\right\}.$$

When $h\left(a, \mathbf{0}\right) = \mathbf{h}a$, we find that

$$\begin{aligned} d &= \mathbb{E}\left\{\left\|\mathbf{r} - \mathbf{h}a\right\|^2\right\} \\ &= \mathbb{E}\left\{\mathbf{r}^T\mathbf{r}\right\} + a^2\mathbf{h}^T\mathbf{h} - 2a\mathbf{r}^T\mathbf{h} \end{aligned}$$

Minimizing w.r.t. a gives us

$$2a\mathbf{h}^T\mathbf{h} - 2\mathbf{r}^T\mathbf{h} = 0$$

and finally

$$\hat{a}_{LS}\left(\mathbf{r}\right) = \left(\mathbf{h}^T\mathbf{h}\right)^{-1}\mathbf{h}^T\mathbf{r}.$$

3.3 Orthogonality condition

The reconstruction error is orthogonal to the parameter:

$$\begin{aligned} \left(\mathbf{r} - \mathbf{h}\hat{a}_{LS}\left(\mathbf{r}\right)\right)^T a &= \left(\mathbf{r} - \mathbf{h}\left(\mathbf{h}^T\mathbf{h}\right)^{-1}\mathbf{h}^T\mathbf{r}\right)^T a \\ &= \left(\mathbf{r} - \mathbf{r}\right)^T a \\ &= \mathbf{0} \end{aligned}$$

4 Extension to vector estimation

When

$$\mathbf{r} = \mathbf{H}\mathbf{a} + \mathbf{n}$$

with $\mathbf{r}$ and $\mathbf{n}$ $N \times 1$ vectors, $\mathbf{H}$ a $N \times M$ matrix, and $\mathbf{a}$ an $M \times 1$ vector. The MMSE, LS and BLU estimates are then given by

$$\begin{aligned} \hat{\mathbf{a}}_{LMMSE}\left(\mathbf{r}\right) &= \mathbf{m_A} + \Sigma_\mathbf{A}\mathbf{H}^T\left(\mathbf{H}\Sigma_\mathbf{A}\mathbf{H}^T + \Sigma_\mathbf{N}\right)^{-1}\left(\mathbf{r} - \mathbf{H}\mathbf{m_A}\right) \\ &= \mathbf{m_A} + \left(\mathbf{H}^T\Sigma_\mathbf{N}^{-1}\mathbf{H} + \Sigma_\mathbf{A}^{-1}\right)^{-1}\mathbf{H}^T\Sigma_\mathbf{N}^{-1}\left(\mathbf{r} - \mathbf{H}\mathbf{m_A}\right) \end{aligned}$$

$$\hat{\mathbf{a}}_{LS}\left(\mathbf{r}\right) = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{r}$$

$$\hat{\mathbf{a}}_{BLU}\left(\mathbf{r}\right) = \left(\mathbf{H}^T\Sigma_\mathbf{N}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\Sigma_\mathbf{N}^{-1}\mathbf{r}$$

5 Further problems

5.1 Problem 1: relations for Linear Gaussian Models

How are MMSE, LS and BLU related? Consider the scalar case. The estimators are given by

$$\hat{a}_{MMSE}\left(\mathbf{r}\right) = m_A + \Sigma_A\mathbf{h}^T\left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_\mathbf{N}\right)^{-1}\left(\mathbf{r} - \mathbf{h}m_A\right)$$

$$\hat{a}_{LS}\left(\mathbf{r}\right) = \left(\mathbf{h}^T\mathbf{h}\right)^{-1}\mathbf{h}^T\mathbf{r}$$

$$\hat{a}_{BLU}\left(\mathbf{r}\right) = \left(\mathbf{h}^T\Sigma_\mathbf{N}^{-1}\mathbf{h}\right)^{-1}\mathbf{h}^T\Sigma_\mathbf{N}^{-1}\mathbf{r}$$

5.1.1 No a priori information: MMSE and LS

When $m_A = 0$ and $\Sigma_A^{-1} = 0$ we get

$$\begin{aligned}
\hat{a}_{MMSE}(\mathbf{r}) &= \mathbf{h}^T \left(\mathbf{h}\mathbf{h}^T\right)^{-1} \mathbf{r} \\
&= \left(\mathbf{h}\mathbf{h}^T\right)^{-1} \mathbf{h}^T \mathbf{r} \\
&= \hat{a}_{LS}(\mathbf{r})
\end{aligned}$$

5.1.2 Uncorrelated noise: BLU and LS

When $\Sigma_{\mathbf{N}} = \sigma^2 \mathbf{I}$,

$$\begin{aligned}
\hat{a}_{BLU}(\mathbf{r}) &= \left(\mathbf{h}^T\mathbf{h}\right)^{-1} \mathbf{h}^T \mathbf{r} \\
&= \hat{a}_{LS}(\mathbf{r})
\end{aligned}$$

5.1.3 Zero-noise: MMSE and LS

For the MMSE estimator, when $\Sigma_{\mathbf{N}} = \mathbf{0}$,

$$\begin{aligned}
\hat{a}_{MMSE}(\mathbf{r}) &= m_A + \mathbf{h}^T \left(\mathbf{h}\mathbf{h}^T\right)^{-1} (\mathbf{r} - \mathbf{h} m_A) \\
&= \mathbf{h}^T \left(\mathbf{h}\mathbf{h}^T\right)^{-1} \mathbf{r} \\
&= \hat{a}_{LS}(\mathbf{r}) .
\end{aligned}$$

5.2 Problem 2: orthogonality

Problem: Derive the LMMSE estimator, starting from the orthogonality condition.
Solution: We introduce $Q = A - m_A$, and $\mathbf{S} = \mathbf{R} - \mathbf{m}_{\mathbf{R}} = \mathbf{R} - \mathbf{h} m_A$, and find the LMMSE estimate of Q from the observation $\mathbf{S}$. The orthogonality principle then tells us that

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{S}\hat{q}(\mathbf{S})\right\} &= \mathbb{E}\left\{\mathbf{S}Q\right\} \\
&= \mathbf{h}\Sigma_A
\end{aligned}$$

with, for some (as yet undetermined) $\mathbf{b}$:

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{S}\hat{q}(\mathbf{S})\right\} &= \mathbb{E}\left\{\mathbf{S}\mathbf{b}^T\mathbf{S}\right\} \\
&= \left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}\right) \mathbf{b}
\end{aligned}$$

so that

$$\mathbf{b} = \left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}\right)^{-1} \Sigma_A \mathbf{h}.$$

Hence

$$\hat{q}_{LMMSE}(\mathbf{s}) = \Sigma_A \mathbf{h}^T \left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}\right)^{-1} \mathbf{s}.$$

The estimate of A is then given by

$$\begin{aligned}
\hat{a}_{LMMSE}(\mathbf{r}) &= \hat{q}_{LMMSE}(\mathbf{r} - \mathbf{m}_{\mathbf{R}}) + m_A \\
&= m_A + \Sigma_A \mathbf{h}^T \left(\mathbf{h}\Sigma_A\mathbf{h}^T + \Sigma_{\mathbf{N}}\right)^{-1} (\mathbf{r} - \mathbf{h} m_A) .
\end{aligned}$$

5.3 Practical example: multi-antenna communication

In this problem, we show how LMMSE and LS can be used when ML or MAP leads to algorithms which are too complex to implement.

5.3.1 Problem

We are interested in the following problem: we transmit a vector of n_T iid data symbols $\mathbf{a} \in \mathcal{A} = \Omega^{n_T}$ over an AWGN channel. Here n_T is the number of transmit antennas. The receiver is equipped with n_R receive antennas. We can write the observation as

$$\mathbf{r} = \mathbf{H}\mathbf{a} + \mathbf{n}$$

where $\mathbf{H}$ is a (known) $n_R \times n_T$ channel matrix and $\mathbf{N} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n_R}\right)$. In any practical communication scheme, Ω is a finite set of size M (e.g., BPSK signaling with $\Omega = \{-1, +1\}$). The symbols are iid, uniformly distributed over Ω.

1. Determine the ML and MAP estimates of $\mathbf{a}$. Determine the complexity of the receiver (number of operations to estimate $\mathbf{a}$) as a function of n_T.
2. How can we use LMMSE and LS be used to estimate $\mathbf{a}$? We assume $\Sigma_{\mathbf{A}} = \mathbf{I}_{n_T}$ and $\mathbf{m}_{\mathbf{A}} = \mathbf{0}$. Determine the complexity of the resulting estimators as a function of n_T.

5.3.2 Solution

Solution - part 1

The MAP and ML estimators are considered to be optimal in the case of estimating discrete parameters, in a sense of minimizing the error probability.

$$\begin{aligned}
\hat{\mathbf{a}}_{ML}(\mathbf{r}) &= \arg\max_{\mathbf{a}\in\Omega^{n_T}} p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}\,|\mathbf{a}) \\
&= \arg\max_{\mathbf{a}\in\Omega^{n_T}} \log p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}\,|\mathbf{a}) \\
&= \arg\max_{\mathbf{a}\in\Omega^{n_T}} -\frac{1}{\sigma^2}\|\mathbf{r}-\mathbf{H}\mathbf{a}\|^2 \\
&= \arg\min_{\mathbf{a}\in\Omega^{n_T}} \|\mathbf{r}-\mathbf{H}\mathbf{a}\|^2
\end{aligned}$$

$$\begin{aligned}
\hat{\mathbf{a}}_{MAP}(\mathbf{r}) &= \arg\max_{\mathbf{a}\in\Omega^{n_T}} p_{\mathbf{A}|\mathbf{R}}(\mathbf{a}\,|\mathbf{r}) \\
&= \arg\max_{\mathbf{a}\in\Omega^{n_T}} p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}\,|\mathbf{a})\frac{p_{\mathbf{A}}(\mathbf{a})}{p_{\mathbf{R}}(\mathbf{r})} \\
&= \arg\max_{\mathbf{a}\in\Omega^{n_T}} p_{\mathbf{R}|\mathbf{A}}(\mathbf{r}\,|\mathbf{a}) \\
&= \hat{\mathbf{a}}_{ML}(\mathbf{r})
\end{aligned}$$

Complexity: for each $\mathbf{a} \in \Omega^{n_T}$, we need to compute $\|\mathbf{r}-\mathbf{H}\mathbf{a}\|^2$. Hence, the complexity is exponential in the number of transmit antennas.

Special case:

Let us assume $\Omega = \{-1, +1\}$ and $n_T = 1$, $n_R > 1$. Then

$$\hat{a}_{ML}(r) = \arg\max_{a\in\Omega}\left(a\mathbf{h}^T\mathbf{r}\right)$$

In this case we combine the observations from multiple receive antennas. Each observation is weighted with the channel gain. This means that when the channel is unreliable on a given antenna (so that h_k is small), this has only a small contribution to our decision statistic.

Solution - part 2

Let us temporarily forget that $\mathbf{a}$ lives in a discrete space. We can then introduce the LMMSE and LS estimates as follows:

$$\hat{\mathbf{a}}_{LMMSE}(\mathbf{r}) = \mathbf{H}^T \left(\mathbf{H}\mathbf{H}^T + \sigma^2 \mathbf{I}_{n_R}\right)^{-1} \mathbf{r}$$

$$\hat{\mathbf{a}}_{LS}(\mathbf{r}) = \left(\mathbf{H}^T \mathbf{H}\right)^{-1} \mathbf{H}^T \mathbf{r}.$$

Note that $\hat{\mathbf{a}}_{LMMSE}(\mathbf{r})$ and $\hat{\mathbf{a}}_{LS}(\mathbf{r})$ may not belong to Ω^{n_T}. Hence, we need to quantize these estimates (mapping the estimate to the closest point in Ω^{n_T}); this can be done on a symbol-per-symbol basis. Generally the LS estimator will have poor performance when $\mathbf{H}\mathbf{H}^T$ is close to singular. Note that the LMMSE estimator requires knowledge of σ^2, while the MAP and ML estimators do not.

Complexity: now the complexity is of the order $n_R \times n_T$ (since any matrix not depending on $\mathbf{r}$ can be pre-computed by the receiver), which is linear in the number of transmit antennas.

Special case: -=-=-Comment: there was a mistake in the lecture. These are the correct results-=-=-

Let us assume that we use more transmit that receive antennas: $n_T > n_R$. In that case the the $n_R \times n_R$ matrix $\mathbf{H}\mathbf{H}^T$ is necessarily singular, so that LS estimation will not work. The LMMSE estimator requires the inversion of $\mathbf{H}\mathbf{H}^T + \sigma^2 \mathbf{I}_{n_R}$. When $\sigma^2 \neq 0$, this matrix is always invertible. This is a strange situation, where noise is actually helpful in the estimation process.