

HW2. Math 499. Spring 2007, CSUF, Spring 2010)

Nasser M. Abbasi

Spring 2007

Contents

1	Problem	1
2	appendix	5

1 Problem

question: Consider the solution of $A\mathbf{x} = \mathbf{b} + \mathbf{n}$ where A is $m \times n$ matrix, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$ and \mathbf{n} is vector of i.i.d. Gaussian $N(0, \sigma^2 \mathbf{I}_m)$ noise vector. (i.e. noise is white Gaussian noise). Determine the best solution \vec{x}

Answer

Let us refer to the observed output (which includes the noise \mathbf{n}) as \mathbf{z} , hence we write $\mathbf{z} = \mathbf{b} + \mathbf{n}$ where \mathbf{b} is the uncontaminated output (what the observed output would be if there is no noise).

Since the noise \mathbf{n} is an additive noise to the output \mathbf{b} of the system, the since the noise has zero mean, then the mean of \mathbf{z} will be the same as the mean of \mathbf{b} . But \mathbf{b} is a deterministic signal which does not change, hence its mean is its value, hence the mean of \mathbf{z} is \mathbf{b} .

Now, \mathbf{z} is described by a probability density function PDF as follows (\mathbf{z} is in \mathbb{R}^m , hence it is m long vector)

$$\begin{aligned} \Pr(\mathbf{z}; \{\mu(\mathbf{z}), \sigma^2 \mathbf{I}_m\}) &= \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{z} - \mu(\mathbf{z})\|^2\right) \\ \Pr(\mathbf{z}; \{\mathbf{b}, \sigma^2 \mathbf{I}_m\}) &= \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{z} - \mathbf{b}\|^2\right) \end{aligned}$$

But since $A\mathbf{x} = \mathbf{b}$, then the above can be written as

$$\Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) = \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{z} - A\mathbf{x}\|^2\right)$$

Since A is a constant matrix (system is assumed to be time-invariant), hence from the above we see that the expression gives the probability of observing \mathbf{z} for a given \mathbf{x} . Hence the best estimate of \mathbf{x} would be the one which maximizes this probability. Instead of maximizing the PDF directly, we maximize its natural logarithm (a mathematical convenience trick, no more).

Now find the natural logarithm of the above quantity, and find where the result is maximum.

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \ln \Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) &= \frac{\partial}{\partial \mathbf{x}} \left(\ln \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{z} - A\mathbf{x}\|^2\right) \right) \\ &= \nabla_{\mathbf{x}} \left(\underbrace{\ln \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}}}_0 + \ln \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{z} - A\mathbf{x}\|^2\right) \right) \\ &= \nabla_{\mathbf{x}} \left(\ln \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \right) + \nabla_{\mathbf{x}} \left(-\frac{1}{2\sigma^2} \|\mathbf{z} - A\mathbf{x}\|^2 \right) \\ &= -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}} (\|\mathbf{z} - A\mathbf{x}\|^2) \end{aligned} \tag{1}$$

Start with $m = n = 1$, hence the above becomes¹

¹I will start the index at 1 to be Matlab friendly

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}} \ln \Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) &= -\frac{1}{2\sigma^2} \nabla_{x_1} (z_1^2 + a_{11}^2 x_1^2 - 2z_1 a_{11} x_1) \\
&= -\frac{1}{2\sigma^2} (2a_{11}^2 x_1 - 2a_{11} z_1) \\
&= -\frac{1}{\sigma^2} (a_{11}^2 x_1 - a_{11} z_1)
\end{aligned}$$

Set the above to zero and solve for x_1

$$\begin{aligned}
0 &= (a_{11} x_1 - z_1) \\
x_1 &= \frac{z_1}{a_{11}}
\end{aligned}$$

Hence

$$x_1 = \frac{z_1}{a_{11}}$$

This matches the least squares solution $x_1 = (A^T A)^{-1} A^T z_1 \rightarrow x_1 = (a_{11} a_{11})^{-1} a_{11} z_1 = \frac{z_1}{a_{11}}$

Now I need to do this for $m = n = 2$, and assuming that $\text{var}(z_1) = \text{var}(z_2)$. We can start from equation (1) above, shown again below

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{x}} \ln \Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) &= -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}} (\|\mathbf{z} - A\mathbf{x}\|^2) \\
-2\sigma^2 \frac{\partial}{\partial \mathbf{x}} \ln \Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) &= \nabla_{\mathbf{x}} (\|\mathbf{z} - A\mathbf{x}\|^2) \\
&= \nabla_{\mathbf{x}} \left(\left\| \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} - \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\|_2^2 \right) \\
&= \nabla_{\mathbf{x}} \left(\left\| \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} - \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} \right\|_2^2 \right) \\
&= \nabla_{\mathbf{x}} \left(\left\| \begin{pmatrix} z_1 - (a_{11}x_1 + a_{12}x_2) \\ z_2 - (a_{21}x_1 + a_{22}x_2) \end{pmatrix} \right\|_2^2 \right) \\
&= \nabla_{\mathbf{x}} ([z_1 - (a_{11}x_1 + a_{12}x_2)]^2 + [z_2 - (a_{21}x_1 + a_{22}x_2)]^2)
\end{aligned}$$

Let

$$\begin{aligned}
f(x_1, x_2) &= [z_1 - (a_{11}x_1 + a_{12}x_2)]^2 + [z_2 - (a_{21}x_1 + a_{22}x_2)]^2 \\
&= [z_1^2 + (a_{11}x_1 + a_{12}x_2)^2 - 2z_1(a_{11}x_1 + a_{12}x_2)] + [z_2^2 + (a_{21}x_1 + a_{22}x_2)^2 - 2z_2(a_{21}x_1 + a_{22}x_2)] \\
&= [z_1^2 + a_{11}^2 x_1^2 + a_{12}^2 x_2^2 + 2a_{11}a_{12}x_1x_2 - 2z_1a_{11}x_1 - 2z_1a_{12}x_2] + \\
&\quad [z_2^2 + a_{21}^2 x_1^2 + a_{22}^2 x_2^2 + 2a_{21}a_{22}x_1x_2 - 2z_2a_{21}x_1 - 2z_2a_{22}x_2]
\end{aligned}$$

Then

$$\nabla_{\mathbf{x}} (f(x_1, x_2)) = \begin{pmatrix} \frac{\partial f(x_1, x_2)}{\partial x_1} \\ \frac{\partial f(x_1, x_2)}{\partial x_2} \end{pmatrix}$$

so

$$\begin{aligned}
\frac{\partial f(x_1, x_2)}{\partial x_1} &= [0 + 2a_{11}^2 x_1 + 0 + 2a_{11}a_{12}x_2 - 2z_1a_{11} - 0] + [0 + 2a_{21}^2 x_1 + 0 + 2a_{21}a_{22}x_2 - 2z_2a_{21} - 0] \\
&= 2a_{11}^2 x_1 + 2a_{11}a_{12}x_2 - 2z_1a_{11} + 2a_{21}^2 x_1 + 2a_{21}a_{22}x_2 - 2z_2a_{21} \\
&= x_1 (2a_{11}^2 + 2a_{21}^2) + x_2 (2a_{11}a_{12} + 2a_{21}a_{22}) - 2z_1a_{11} - 2z_2a_{21}
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial f(x_1, x_2)}{\partial x_2} &= [0 + 0 + 2a_{12}^2 x_2 + 2a_{11}a_{12}x_1 - 0 - 2z_1a_{12}] + [0 + 0 + 2a_{22}^2 x_2 + 2a_{21}a_{22}x_1 - 0 - 2z_2a_{22}] \\ &= 2a_{12}^2 x_2 + 2a_{11}a_{12}x_1 - 2z_1a_{12} + 2a_{22}^2 x_2 + 2a_{21}a_{22}x_1 - 2z_2a_{22} \\ &= x_1(2a_{11}a_{12} + 2a_{21}a_{22}) + x_2(2a_{12}^2 + 2a_{22}^2) - 2z_1a_{12} - 2z_2a_{22}\end{aligned}$$

Hence we obtain that

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}} \ln \Pr(\mathbf{z}; \{A\mathbf{x}, \sigma^2 \mathbf{I}_m\}) &= -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}}(f(x_1, x_2)) \\ &= -\frac{1}{2\sigma^2} \begin{pmatrix} x_1(2a_{11}^2 + 2a_{21}^2) + x_2(2a_{11}a_{12} + 2a_{21}a_{22}) - 2z_1a_{11} - 2z_2a_{21} \\ x_1(2a_{11}a_{12} + 2a_{21}a_{22}) + x_2(2a_{12}^2 + 2a_{22}^2) - 2z_1a_{12} - 2z_2a_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\end{aligned}$$

Hence

$$x_1(2a_{11}^2 + 2a_{21}^2) + x_2(2a_{11}a_{12} + 2a_{21}a_{22}) - 2z_1a_{11} - 2z_2a_{21} = 0$$

and

$$x_1(2a_{11}a_{12} + 2a_{21}a_{22}) + x_2(2a_{12}^2 + 2a_{22}^2) - 2z_1a_{12} - 2z_2a_{22} = 0$$

so, solve for x_1, x_2 . Write as $Ax = b$ and solve:

$$\begin{pmatrix} (2a_{11}^2 + 2a_{21}^2) & (2a_{11}a_{12} + 2a_{21}a_{22}) \\ (2a_{11}a_{12} + 2a_{21}a_{22}) & (2a_{12}^2 + 2a_{22}^2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2z_1a_{11} + 2z_2a_{21} \\ 2z_1a_{12} + 2z_2a_{22} \end{pmatrix}$$

Solve for

$$\begin{aligned}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} (2a_{11}^2 + 2a_{21}^2) & (2a_{11}a_{12} + 2a_{21}a_{22}) \\ (2a_{11}a_{12} + 2a_{21}a_{22}) & (2a_{12}^2 + 2a_{22}^2) \end{pmatrix}^{-1} \begin{pmatrix} 2z_1a_{11} + 2z_2a_{21} \\ 2z_1a_{12} + 2z_2a_{22} \end{pmatrix} \\ &= \frac{\begin{pmatrix} (2a_{12}^2 + 2a_{22}^2) & -(2a_{11}a_{12} + 2a_{21}a_{22}) \\ -(2a_{11}a_{12} + 2a_{21}a_{22}) & (2a_{11}^2 + 2a_{21}^2) \end{pmatrix} \begin{pmatrix} 2z_1a_{11} + 2z_2a_{21} \\ 2z_1a_{12} + 2z_2a_{22} \end{pmatrix}}{(2a_{11}^2 + 2a_{21}^2)(2a_{12}^2 + 2a_{22}^2) - (2a_{11}a_{12} + 2a_{21}a_{22})^2} \\ &= \frac{\begin{pmatrix} (2a_{12}^2 + 2a_{22}^2)(2z_1a_{11} + 2z_2a_{21}) - (2a_{11}a_{12} + 2a_{21}a_{22})(2z_1a_{12} + 2z_2a_{22}) \\ -(2a_{11}a_{12} + 2a_{21}a_{22})(2z_1a_{11} + 2z_2a_{21}) + (2a_{11}^2 + 2a_{21}^2)(2z_1a_{12} + 2z_2a_{22}) \end{pmatrix}}{(2a_{11}^2 + 2a_{21}^2)(2a_{12}^2 + 2a_{22}^2) - (2a_{11}a_{12} + 2a_{21}a_{22})^2} \\ &= \frac{\begin{pmatrix} 4a_{11}z_1a_{22}^2 + 4a_{12}^2a_{21}z_2 - 4a_{11}a_{12}z_2a_{22} - 4z_1a_{12}a_{21}a_{22} \\ 4z_1a_{12}a_{21}^2 + 4a_{11}^2z_2a_{22} - 4a_{11}z_1a_{21}a_{22} - 4a_{11}a_{12}a_{21}z_2 \end{pmatrix}}{4a_{11}^2a_{22}^2 - 8a_{11}a_{12}a_{21}a_{22} + 4a_{12}^2a_{21}^2} \\ &= \begin{pmatrix} \frac{4a_{11}z_1a_{22}^2 + 4a_{12}^2a_{21}z_2 - 4a_{11}a_{12}z_2a_{22} - 4z_1a_{12}a_{21}a_{22}}{4a_{11}^2a_{22}^2 - 8a_{11}a_{12}a_{21}a_{22} + 4a_{12}^2a_{21}^2} \\ \frac{4z_1a_{12}a_{21}^2 + 4a_{11}^2z_2a_{22} - 4a_{11}z_1a_{21}a_{22} - 4a_{11}a_{12}a_{21}z_2}{4a_{11}^2a_{22}^2 - 8a_{11}a_{12}a_{21}a_{22} + 4a_{12}^2a_{21}^2} \end{pmatrix}\end{aligned}$$

so

$$\begin{aligned}x_1 &= \frac{4a_{11}z_1a_{22}^2 + 4a_{12}^2a_{21}z_2 - 4a_{11}a_{12}z_2a_{22} - 4z_1a_{12}a_{21}a_{22}}{4a_{11}^2a_{22}^2 - 8a_{11}a_{12}a_{21}a_{22} + 4a_{12}^2a_{21}^2} \\ &= \left(\frac{z_1a_{22} - a_{12}z_2}{a_{11}a_{22} - a_{12}a_{21}} \right)\end{aligned}$$

and

$$\begin{aligned} x_2 &= \frac{4z_1a_{12}a_{21}^2 + 4a_{11}^2z_2a_{22} - 4a_{11}z_1a_{21}a_{22} - 4a_{11}a_{12}a_{21}z_2}{4a_{11}^2a_{22}^2 - 8a_{11}a_{12}a_{21}a_{22} + 4a_{12}^2a_{21}^2} \\ &= \left(\frac{a_{11}z_2 - z_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}} \right) \end{aligned}$$

Hence

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \left(\frac{z_1a_{22} - a_{12}z_2}{a_{11}a_{22} - a_{12}a_{21}} \right) \\ \left(\frac{a_{11}z_2 - z_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}} \right) \end{pmatrix} \quad (2)$$

is the least squares error. To validate

$$\begin{aligned} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \left(\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right)^{-1} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^T \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= \left(\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \right)^{-1} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}^2 + a_{21}^2 & a_{11}a_{12} + a_{21}a_{22} \\ a_{11}a_{12} + a_{21}a_{22} & a_{12}^2 + a_{22}^2 \end{pmatrix}^{-1} \begin{pmatrix} a_{11}z_1 + a_{21}z_2 \\ z_1a_{12} + z_2a_{22} \end{pmatrix} \\ &= \begin{pmatrix} \frac{(a_{12}^2 + a_{22}^2)}{(a_{11}a_{22} - a_{12}a_{21})^2} & -\frac{a_{11}a_{12} + a_{21}a_{22}}{(a_{11}a_{22} - a_{12}a_{21})^2} \\ -\frac{a_{11}a_{12} + a_{21}a_{22}}{(a_{11}a_{22} - a_{12}a_{21})^2} & \frac{(a_{11}^2 + a_{21}^2)}{(a_{11}a_{22} - a_{12}a_{21})^2} \end{pmatrix} \begin{pmatrix} a_{11}z_1 + a_{21}z_2 \\ z_1a_{12} + z_2a_{22} \end{pmatrix} \\ &= \begin{pmatrix} \frac{(z_1a_{22} - a_{12}z_2)}{a_{11}a_{22} - a_{12}a_{21}} \\ \frac{a_{11}z_2 - z_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}} \end{pmatrix} \quad (3) \end{aligned}$$

Compare equations (2) and (3) above, they are the same. OK, confirmed.

2 appendix

Some definitions:

σ : standard deviation of a distribution or random variable.

VAR: Variance. The square of σ or σ^2 , the larger the variance, larger the spread that is.

$E(x_i)$: Expected value of random variable = $\mu(x_i)$, its mean.
 $= \sum_{\text{all } x} x P(x)$. \rightarrow probability of x

Covariance: measures how much one random variable varies together with another random variable.

$$\begin{aligned} \text{Cov}(X, Y) &= E\left[\underbrace{(x - E(x))}_{\text{error in } x} \underbrace{(y - E(y))}_{\text{error in } y} \right] \\ &= E\left[(x - \mu(x)) (y - \mu(y)) \right] \\ &= E(xy) - E(x)E(y). \end{aligned}$$

note, $\text{Cov}(X, Y) = 0$ if X, Y are linearly independent random variables.

(1b) Weighted Least squares normal equations

$$(A^T W W^T A) \hat{X}_w = A^T W W^T b$$

where $W W^T = \text{Cov}^{-1} = C$

So linearly independent observation vector b_i , $\text{Cov}(b_i, b_j)$

$$= \begin{bmatrix} E[(b_1 - E(b_1))(b_1 - E(b_1))] & 0 & 0 & \dots & 0 \\ 0 & E[(b_2 - E(b_2))(b_2 - E(b_2))] & & & \\ \vdots & & & & \\ 0 & & & & E[(b_m - E(b_m))(b_m - E(b_m))] \end{bmatrix}$$

Since mean is given as zero. Then $i.e. E(b_i) = 0$

$$\text{Cov}(b) = \begin{bmatrix} E(b_1^2) & 0 & 0 & \dots & 0 \\ \vdots & E(b_2^2) & & & \\ 0 & & E(b_3^2) & & \\ & & & \dots & \\ 0 & & & & E(b_m^2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & & & \\ & \sigma_2^2 & & & \\ & & \dots & & \\ & & & & \sigma_m^2 \end{bmatrix}$$

Expected value of square of observation is its variance

Since $\text{Cov}(b)$ is diagonal, then $C = \text{Cov}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \frac{1}{\sigma_2^2} & \\ & & \dots \\ & & & \frac{1}{\sigma_m^2} \end{bmatrix}$

hence $\hat{x}_w = (A^T C A)^{-1} A^T C b$

Solution
when noise is
present

Since A is known, C can be found since σ^2 is given for the noise, and b is known (this is the observation itself) then we can calculate the above to find \hat{x}_w . #