HW2. Math 499. Spring 2007
Independent studies course

Supervisor: Dr Angel R. Pineda, Assistant Professor Mathematics
Department California State University, Fullerton

Student: Nasser Abbasi

April 20, 2007

1 Problem

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

Answer:

Since noise $\vec{n}$ is an additive noise to the output of the system, we consider the $\vec{b}$ vector as a random signal of the same mean and variance as the noise being added to it. Hence $\vec{b}$ is described by a probability density function PDF as follows (assume $\vec{b}$ is an n long vector, i.e. a vector in an n-dimension space)

$$\Pr\left(\vec{b}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\vec{b} - \mu\left(\vec{b}\right)\right)^2\right)$$

The above gives the probability of $\vec{b}$ as a function of its mean $\mu\left(\vec{b}\right)$ and its standard deviation σ

Now, we are told this is white noise, hence the mean is zero, so the above becomes

$$\Pr\left(\vec{b}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\vec{b}\right)^2\right)$$

Now, let us parametrize the output by the input. This means we want to associate the PDF of the output by the input. We do this so that later we can ask the question of for what $\vec{x}$ is the probability of $\vec{b}$ the largest? Finding such $\vec{x}$ we give us the best solution for the observed noisy $\vec{b}$.

A parametrized random variable $\vec{b}$ by parameter $\vec{x}$ is written as follows

$$\Pr\left(\vec{b}; \vec{x}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\vec{b} - \vec{x}\right)^2\right)$$

But $A\vec{x} = \vec{b}$, hence the above becomes

$$\Pr\left(\vec{b}; \vec{x}\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(A\vec{x} - \vec{x}\right)^2\right)$$

$$\Pr\left(b_j; x_j\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left(\left(\sum_{i=1}^{n} A(j,i)\,x_i\right) - x_j\right)^2\right) \qquad j = 1:m$$

$$\Pr\left(b_j; x_j\right) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left[\left(\sum_{i=1}^{n} A(j,i)\,x_i\right)^2 + x_j^2 - 2x_j\left(\sum_{i=1}^{n} A(j,i)\,x_i\right)\right]\right)$$

Start with $m = n = 1$, hence the above becomes[1]

$$\begin{aligned}\Pr(b_1; x_1) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left[(a_{11}x_1)^2 + x_1^2 - 2x_1(a_{11}x_1)\right]\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(a_{11}^2 x_1^2 + x_1^2 - 2a_{11}x_1^2\right)\right)\end{aligned}$$

The best x_1 is the solution which maximizes the ln of the above. Hence we need to evaluate

$$\begin{aligned}\frac{\partial}{\partial x_1}\ln\left(\Pr(b_1; x_1)\right) &= \frac{\partial}{\partial x_1}\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(a_{11}^2 x_1^2 + x_1^2 - 2a_{11}x_1^2\right)\right)\right) \\ &= \frac{\partial}{\partial x_1}\left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \ln\left(\exp\left(-\frac{1}{2\sigma^2}\left(a_{11}^2 x_1^2 + x_1^2 - 2a_{11}x_1^2\right)\right)\right)\right] \\ &= \overbrace{\frac{\partial}{\partial x_1}\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\right)}^{0} + \frac{\partial}{\partial x_1}\left(-\frac{1}{2\sigma^2}\left(a_{11}^2 x_1^2 + x_1^2 - 2a_{11}x_1^2\right)\right) \\ &= \frac{\partial}{\partial x_1}\left(-\frac{1}{2\sigma^2}\left(a_{11}^2 x_1^2 + x_1^2 - 2a_{11}x_1^2\right)\right) \\ &= \frac{-1}{2\sigma^2}\left(2a_{11}^2 x_1 + 2x_1 - 4a_{11}x_1\right)\end{aligned}$$

Now set the above to zero, so we write

$$\begin{aligned}\frac{-1}{2\sigma^2}\left(2a_{11}^2 x_1 + 2x_1 - 4a_{11}x_1\right) &= 0 \\ \left(2a_{11}^2 x_1 + 2x_1 - 4a_{11}x_1\right) &= 0 \\ x_1\left(a_{11}^2 + 1 - 2a_{11}\right) &= 0\end{aligned}$$

Since $\vec{x}$ can't be zero ($A \times 0$ can't make something non-zero), hence we need to have $a_{11}^2 + 1 - 2a_{11} = 0$., Solution is $\boxed{a_{11} = 1}$

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

So, the above maximizes the $\ln \Pr\left(\vec{b};\vec{x}\right)$, hence

$$
\begin{aligned}
\Pr\left(\vec{b};\vec{x}\right) &= \Pr\left(b_1;x_1\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(a_{11}^2x_1^2+x_1^2-2a_{11}x_1\right)}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(x_1^2+x_1^2-2x_1\right)}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{x_1-x_1^2}{\sigma^2}\right)
\end{aligned}
$$

So setting $\Pr\left(\vec{b};\vec{x}\right)=1$

$$
\begin{aligned}
1 &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{x_1-x_1^2}{\sigma^2}\right) \\
\sqrt{2\pi\sigma^2} &= \exp\left(\frac{x_1-x_1^2}{\sigma^2}\right) \\
\ln\sqrt{2\pi\sigma^2} &= \frac{x_1-x_1^2}{\sigma^2} \\
\sigma^2\ln\sqrt{2\pi\sigma^2} &= x_1-x_1^2 \\
x_1^2-x_1+\sigma^2\ln\sqrt{2\pi\sigma^2} &= 0
\end{aligned}
$$

Solution is: $\boxed{x_1=\frac{1}{2}+\frac{1}{2}\sqrt{1-2\sigma^2\ln 2-2\sigma^2\ln\pi-2\sigma^2\ln\sigma^2}}$

or $\boxed{x_1=\frac{1}{2}-\frac{1}{2}\sqrt{1-2\sigma^2\ln 2-2\sigma^2\ln\pi-2\sigma^2\ln\sigma^2}}$

Lets check: if noise variance is zero, then $n=0$ since it is zero mean, hence $b=Ax$, and lets say $b=1$, then $x_1=\frac{1}{A}=\frac{1}{a_{11}}=1$

Plug in $\sigma^2=0$ in the above, we obtain $x_1=\frac{1}{2}+\frac{1}{2}\sqrt{1}=1$ (the solution $x_1=0$ is not used, does not make sense).

Now for the 2D case: $n=m=2$

$$
\Pr\left(b_j;x_j\right)=\frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}}\exp\left(-\frac{1}{2\sigma^2}\left(\left(\sum_{i=1}^{n}A\left(j,i\right)x_i\right)^2+x_j^2-2x_j\left(\sum_{i=1}^{n}A\left(j,i\right)x_i\right)\right)\right)\qquad j=1:2
$$

so

$$
\begin{aligned}
\Pr\left(b_1;x_1\right) &= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[\left(a_{11}x_1+a_{12}x_2\right)^2+x_1^2-2x_1\left(a_{11}x_1+a_{12}x_2\right)\right]\right)\\
&= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[\left(a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2\right)+x_1^2-2\left(a_{11}x_1^2+a_{12}x_1x_2\right)\right]\right)\\
&= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right]\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\Pr\left(b_2;x_2\right) &= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[\left(a_{21}x_1+a_{22}x_2\right)^2+x_2^2-2x_2\left(a_{21}x_1+a_{22}x_2\right)\right]\right)\\
&= \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[a_{21}^2x_1^2+a_{22}^2x_2^2+2a_{21}a_{22}x_1x_2+x_2^2-2a_{21}x_2x_1-2a_{22}x_2^2\right]\right)
\end{aligned}
$$

Maximize $\ln\Pr\left(b_j;x_j\right)$, for $j=1$ we have

$$
\begin{aligned}
\ln\Pr\left(b_1;x_1\right) &= \ln\frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right]\right)\\
&= \ln\frac{1}{2\pi\sigma^2}+\left(-\frac{1}{2\sigma^2}\left[a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right]\right)
\end{aligned}
$$

take the derivative with respect to x_1 and set it to zero

$$
\begin{aligned}
&\frac{\partial}{\partial x_1}\ln\left(\Pr\left(b_1;x_1\right)\right)=0\\
0 &= \frac{\partial}{\partial x_1}\left(-\frac{1}{2\sigma^2}\left[a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right]\right)\\
&= \frac{\partial}{\partial x_1}\left(a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right)\\
&= 2a_{11}^2x_1+2a_{11}a_{12}x_2+2x_1-4a_{11}x_1-2a_{12}x_2\\
&= a_{11}^2x_1+a_{11}a_{12}x_2+x_1-2a_{11}x_1-a_{12}x_2
\end{aligned}
$$

take the derivative with respect to x_2 and set it to zero

$$
\begin{aligned}
&\frac{\partial}{\partial x_2}\ln\left(\Pr\left(b_1;x_1\right)\right)=0\\
&= \frac{\partial}{\partial x_2}\left(a_{11}^2x_1^2+a_{12}^2x_2^2+2a_{11}a_{12}x_1x_2+x_1^2-2a_{11}x_1^2-2a_{12}x_1x_2\right)\\
&= 2a_{12}^2x_2+2a_{11}a_{12}x_1-2a_{12}x_1\\
&= a_{12}^2x_2+a_{11}a_{12}x_1-a_{12}x_1
\end{aligned}
$$

Now take the derivative with respect to x_2 of $\Pr(b_2; x_2)$ and set it to zero

$$\frac{\partial}{\partial x_2}\ln\left(\Pr\left(b_2; x_2\right)\right) = 0$$

$$\begin{aligned}
0 &= \frac{\partial}{\partial x_2}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\left[a_{21}^2x_1^2 + a_{22}^2x_2^2 + 2a_{21}a_{22}x_1x_2 + x_2^2 - 2a_{21}x_2x_1 - 2a_{22}x_2^2\right]\right) \\
&= \frac{\partial}{\partial x_2}\left(-\frac{1}{2\sigma^2}\left[a_{21}^2x_1^2 + a_{22}^2x_2^2 + 2a_{21}a_{22}x_1x_2 + x_2^2 - 2a_{21}x_2x_1 - 2a_{22}x_2^2\right]\right) \\
&= \frac{\partial}{\partial x_2}\left(a_{21}^2x_1^2 + a_{22}^2x_2^2 + 2a_{21}a_{22}x_1x_2 + x_2^2 - 2a_{21}x_2x_1 - 2a_{22}x_2^2\right) \\
&= \left(2a_{22}^2x_2 + 2a_{21}a_{22}x_1 + 2x_2 - 2a_{21}x_1 - 4a_{22}x_2\right) \\
&= a_{22}^2x_2 + a_{21}a_{22}x_1 + x_2 - a_{21}x_1 - 2a_{22}x_2
\end{aligned}$$

Now take the derivative with respect to x_1 of $\Pr(b_2; x_2)$ and set it to zero

$$\begin{aligned}
0 &= \frac{\partial}{\partial x_1}\left(a_{21}^2x_1^2 + a_{22}^2x_2^2 + 2a_{21}a_{22}x_1x_2 + x_2^2 - 2a_{21}x_2x_1 - 2a_{22}x_2^2\right) \\
&= 2a_{21}^2x_1 + 2a_{21}a_{22}x_1 - 2a_{21}x_2 \\
&= a_{21}^2x_1 + a_{21}a_{22}x_1 - a_{21}x_2
\end{aligned}$$

Hence we need to solve 4 equations

$$\begin{aligned}
a_{11}^2x_1 + a_{11}a_{12}x_2 + x_1 - 2a_{11}x_1 - a_{12}x_2 &= 0 \\
a_{22}^2x_2 + a_{21}a_{22}x_1 + x_2 - a_{21}x_1 - 2a_{22}x_2 &= 0 \\
a_{21}^2x_1 + a_{21}a_{22}x_1 - a_{21}x_2 &= 0 \\
a_{12}^2x_2 + a_{11}a_{12}x_1 - a_{12}x_1 &= 0
\end{aligned}$$

or

$$\begin{aligned}
x_1\left(a_{11}^2 - 2a_{11} + 1\right) + x_2\left(a_{11}a_{12} - a_{12}\right) &= 0 \\
x_1\left(a_{21}a_{22} - a_{21}\right) + x_2\left(a_{22}^2 + 1 - 2a_{22}\right) &= 0 \\
\left(a_{21}^2 + a_{21}a_{22}\right)x_1 - a_{21}x_2 &= 0 \\
a_{12}^2x_2 + \left(a_{11}a_{12} - a_{12}\right)x_1 &= 0
\end{aligned}$$

4 equations, 4 unknowns, solve for $a_{11}, a_{12,}a_{21,}a_{22}$, use these to find $\vec{x}$ as we did for $n = 1$

(verify if I am on the right track before continuing)

I know looked at the textbook by Kay, I think I need to use 3.29 on page 45, but can't now figure how to do that, since I have no way to fit this problem into that equation. Need more time.

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 about it.

$E(x_i)$: Expected value of random variable $= \mu(x_i)$, its mean.

$$= \sum_{\text{all } x} x P(x) \quad \to \text{probability of } x$$

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

$$cov(x,y) = E[\overbrace{(x - E(x))}^{\text{error in } x} \overbrace{(y - E(y))}^{\text{error in } y}]$$

$$= E[(x - \mu(x))(y - \mu(y))]$$

$$= E(xy) - E(x)E(y)$$

note, $cov(x,y) = 0$ 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 $WW^T = Cov^{-1} = C$

So Linearly independent observation vector b, $Cov(b_i, b_j)$

$$= \begin{bmatrix} E[\overbrace{(b_1 - E(b_1))}^{e_1}\overbrace{(b_1 - E(b_1))}^{e_1}] & 0 & 0 \cdots 0 \\ 0 & E[(b_2 - E(b_2))(b_2 - E(b_2))] & \\ 0 & & \ddots \\ 0 & & E(b_m - E(b_m))(\ ) \end{bmatrix}$$

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

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

Expected value of square of the observation is its variance since cov(b) is diagonal, then $C = Cov^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots & \\ & & & \frac{1}{\sigma_m^2} \end{bmatrix}$

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

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 $\hat{x}_w$. #