1. *The Singular Value Decomposition.* Let $\boldsymbol{A}$ be $m \times n$, with $m > n$, and suppose it has singular value decomposition $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$. Let $\boldsymbol{u}_i$ be the ith column of $\boldsymbol{U}, \boldsymbol{v}_i$ be the ith column of $\boldsymbol{V}$, and σ_i the ith diagonal element of $\boldsymbol{\Sigma}$.

(a) If the rank of $\boldsymbol{A}$ is r, show that

$$\boldsymbol{A}\boldsymbol{v}_i = \begin{cases} \sigma_i \boldsymbol{u}_i, & i = 1, \dots, r \\ \boldsymbol{0}, & i = r+1, \dots, n, \end{cases}, \quad \boldsymbol{A}^T \boldsymbol{u}_i = \begin{cases} \sigma_i \boldsymbol{v}_i, & i = 1, \dots, r \\ \boldsymbol{0}, & i = r+1, \dots, m. \end{cases}$$

Solution:

Note that $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{U}(\boldsymbol{\Sigma}\boldsymbol{e}_i) = \sigma_i \boldsymbol{u}_i$ and if $r = n$ then the identity above is satisfied. Now, suppose $r < n$, then the set $\{\boldsymbol{A}\boldsymbol{v}_1, \dots, \boldsymbol{A}\boldsymbol{v}_{r+1}\} = \{\sigma_1 \boldsymbol{u}_1, \dots, \sigma_{r+1}\boldsymbol{u}_{r+1}\}$ is linearly dependent. I.e. there exists a non-trivial linear combination so that $x_1\sigma_1\boldsymbol{u}_1 + \cdots + x_{r+1}\sigma_{r+1}\boldsymbol{u}_{r+1} = \boldsymbol{0}$. But, by the mutual independence of $\boldsymbol{u}_i$, it must be that each scalar product $x_i\sigma_i = 0$. Since the singular values are positive and decreasing in index, if $\sigma_k = 0$ for $k < r$, then the dimension of the null space of A would be greater than $n - r$, violating the rank-nullity theorem. So each $\sigma_1, \dots, \sigma_r$ are strictly greater than zero. So $x_1, \dots, x_r$ are each 0, and since $(x_1, \dots, x_{r+1})$ is non-trivial, $\sigma_{r+1} = 0$. Since the singular values are positive and decreasing, $\sigma_{r+1} = \cdots = \sigma_n = 0$, and by the initial observation, $\boldsymbol{A}\boldsymbol{v}_i = 0$ for $r < i \leq n$.

A similar argument follows for the second identity on $\boldsymbol{A}^T$.

□

(b) Show that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \tag{1.20}$$

can be written $\boldsymbol{A} = \sum_{i=1}^n \boldsymbol{u}_i \sigma_i \boldsymbol{v}_i^T$ and write down the equivalent expression for $\boldsymbol{A}^\dagger$ defined in

$$\boldsymbol{A}^\dagger = \boldsymbol{V}\boldsymbol{\Sigma}^\dagger \boldsymbol{U}^T. \tag{1.21}$$

Solution:

We can write

$$\boldsymbol{\Sigma} = \sigma_1 \boldsymbol{P}_1 + \cdots + \sigma_n \boldsymbol{P}_n$$

where $\boldsymbol{P}_i$ is the $n \times m$ matrix with 1 in the (i,i)th entry and 0 otherwise. Hence $\boldsymbol{A} = \sum_{i=1}^n \sigma_i \boldsymbol{U}\boldsymbol{P}_i\boldsymbol{V}^T$. Note that the (i,j)th entry of $\boldsymbol{U}\boldsymbol{P}_i\boldsymbol{V}^T$ is the scalar product $u_{i,i}v_{j,j}$. These are precisely the entries of $\boldsymbol{u}_i\boldsymbol{v}_i^T$. We have shown $\boldsymbol{A} = \sum \boldsymbol{u}_i\sigma_i\boldsymbol{v}_i$. If we recall that the diagonal entries (up to the rank of A) of $\boldsymbol{\Sigma}^\dagger$ are $1/\sigma_i$ and zero otherwise, then a similar argument shows that

$$\boldsymbol{A}^\dagger = \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T = \sum_{i=1}^r \frac{1}{\sigma_i}\boldsymbol{v}\boldsymbol{u}^T$$

□

(c) Substitute (1.20) into

$$\boldsymbol{x}_{\text{LS}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} \tag{1.18}$$

To prove

$$\boldsymbol{x}_{\mathrm{LS}} = \sum_{i=1} \frac{\boldsymbol{u}_i^T \boldsymbol{b}}{\sigma_i} \boldsymbol{v}_i, \tag{1.22}$$

and hence that $\boldsymbol{x}_{\mathrm{LS}} = \boldsymbol{A}^\dagger \boldsymbol{b}$.

Solution:

Assuming the columns of $\boldsymbol{A}$ are linearly independent, we carry out the substitution:

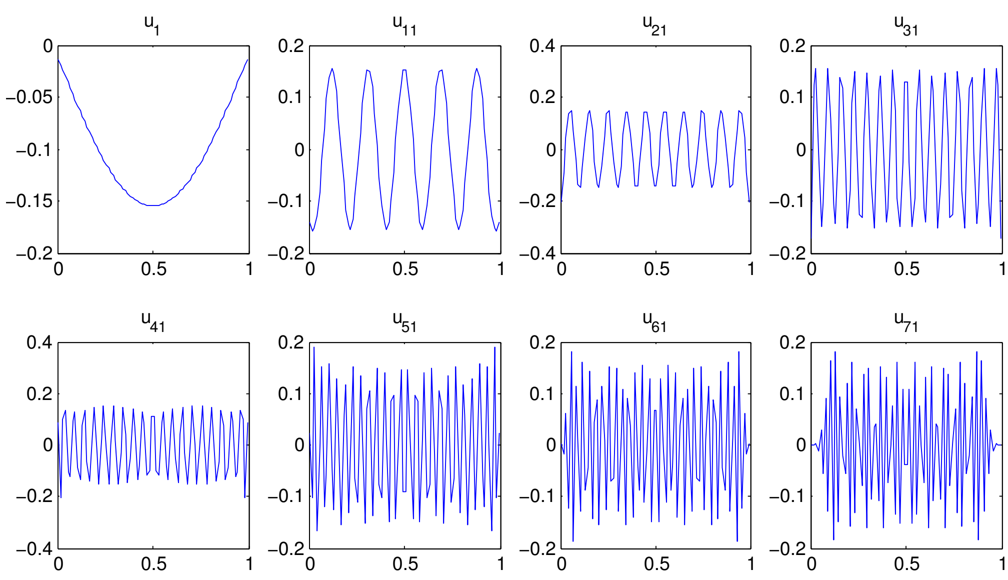
$$\begin{aligned}
\boldsymbol{x}_{\mathrm{LS}} &= (\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{b} \\
&= \boldsymbol{V}\boldsymbol{D}^{-1}\cancel{\boldsymbol{V}^T\boldsymbol{V}}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{b} && \text{where } D \text{ is the } n \times n \text{ diagonal matrix with } \sigma_i^2 \text{ on the diagonal,} \\
&= \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b} && \text{since } D\Sigma^T \text{ has } \sigma_i^{-1} \text{ as diagonal entries and 0 otherwise,} \\
&= \sum_{i=1} \frac{\boldsymbol{u}_i^T\boldsymbol{b}}{\sigma_i}\boldsymbol{v}_i.
\end{aligned}$$

□

2. *The discrete Picard condition.* For parts (a)-(d), use `Deblur1d.m`.

(a) Create plots verifying that singular vectors of the design matrix $\boldsymbol{A}$ become more oscillatory as i increases.

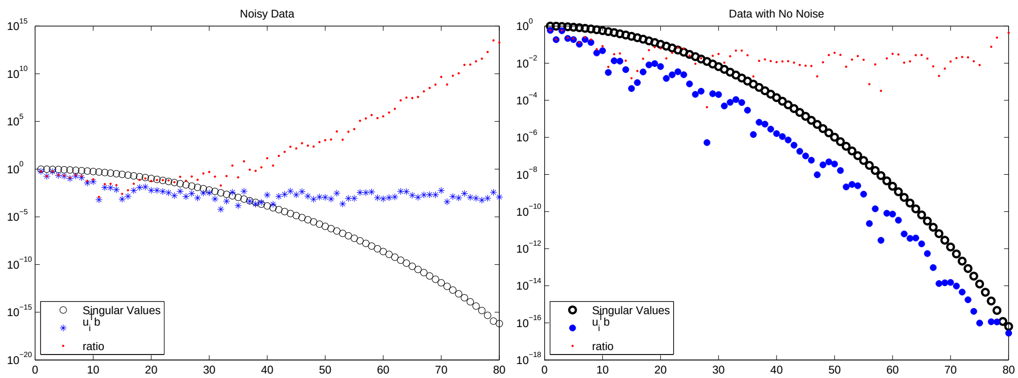
Solution:



□

(b) A Picard plot for discretized inverse problem is a plot of the values of σ_i, $|\boldsymbol{u}_i^T\boldsymbol{b}|$, and $|\boldsymbol{u}_i^T\boldsymbol{b}|/\sigma_i$. For the 1D deblurring example, create the Picard plot, first using the noise-free data and then using noisy data. Verify in both cases the *discrete Picard condition:* Ignoring the part of the Picard plot where $|\boldsymbol{u}_i^T\boldsymbol{b}|$ levels off due to either numerical round-off (in the noise-free case) or to the presence of noise in $\boldsymbol{b}$, the discrete Picard condition is satisfied if the remaining values of $|\boldsymbol{u}_i^T\boldsymbol{b}|$, on average, decay faster than σ_i.

Solution:



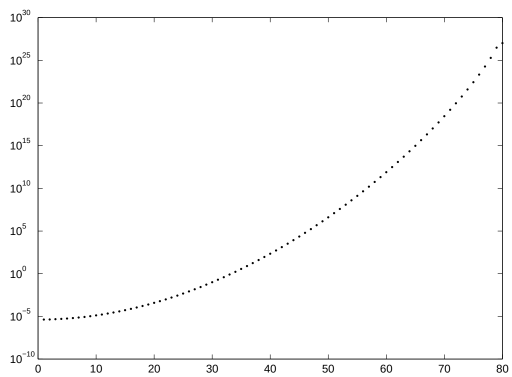
□

(c) Plot the variance values in

$$\boldsymbol{v}_i^T \boldsymbol{x}_{\text{LS}} \sim \mathcal{N}(\boldsymbol{v}_i^T \boldsymbol{x}, \sigma^3/\sigma_i^2). \tag{1.25}$$

What do they tell you about the statistical properties of the least squares solution.

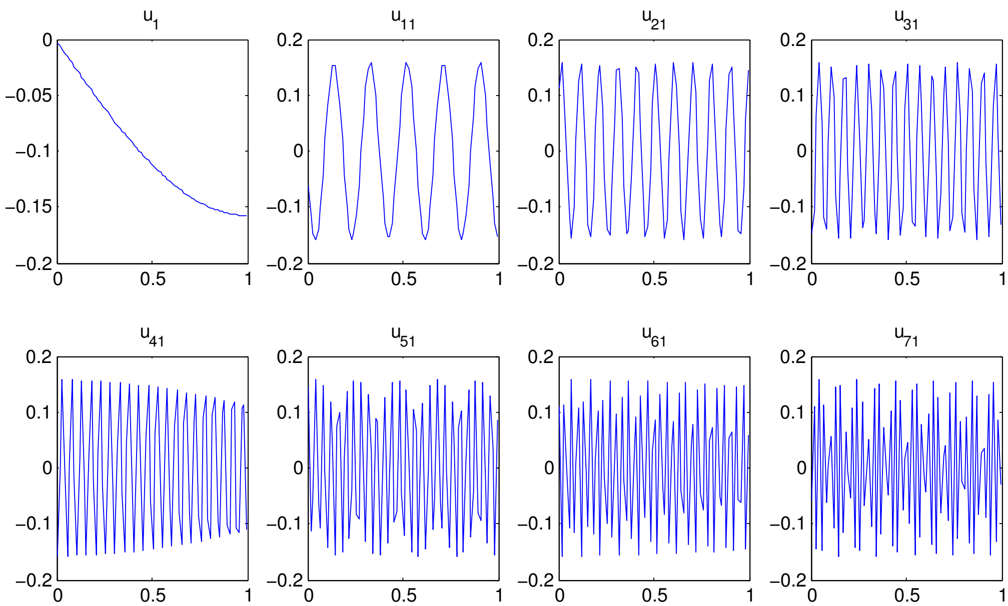
Solution:

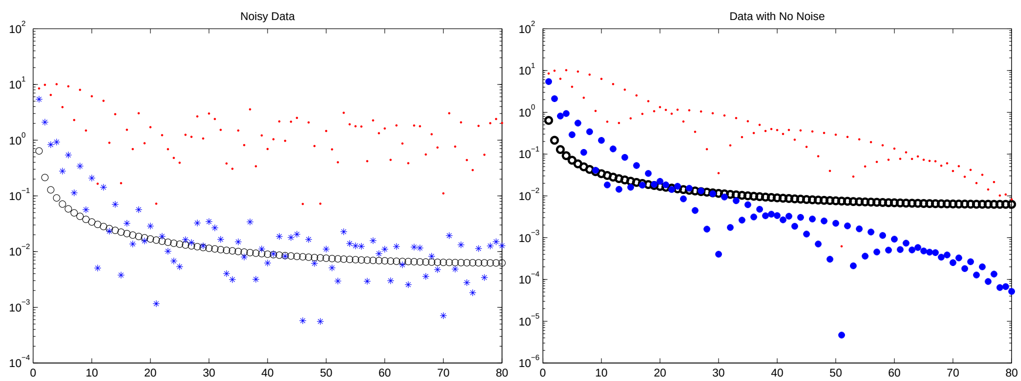


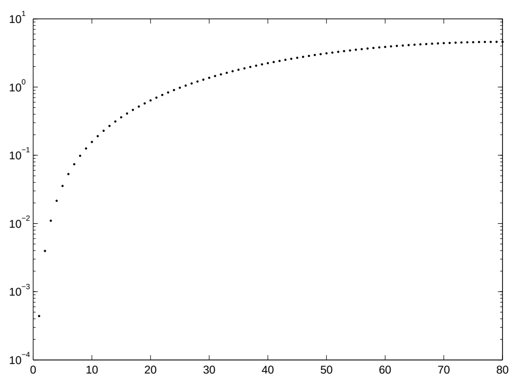
The graph above gives the variances of the least squares solution to 1-D deblurring problem projected onto the ith left singular vector in the SVD of the debluring discritization – i.e. the random vector $\boldsymbol{v}^T \boldsymbol{x}_{\text{LS}}$. The least squares solution, despite being derived from identically distributed data, has coordinates in the basis $\boldsymbol{v}_i$ whose variance grows exponentially as the index i increases as indicated by the graph above. This means that certain features in the solution (those captured by higher indexed singular vectors) are subject to *extremely high* levels of uncertainty. In fact, even if the initial noise levels are relatively quite low, any discritization refined finer than 80 or so points will still be corrupted by noise magnification by the apparent exponential growth with the singular vector index which is equal to the number of discritizaiton levels.

□

(d) Repeat (a)-(c) using `PSFrecon.m`







A similar, but less pronounced, exponential increase in the variances $\boldsymbol{v}_i^T \boldsymbol{x}_{\mathrm{LS}}$ is observed. Note that the ratio in the non-noisy data for this problem does *not* start to increase, indicating that the Picard condition is satisfied for that problem.

3. *Sampling the least squares solution.*

(a) Let $\boldsymbol{v}$ be a random n-vector with mean $\boldsymbol{\mu} = E[\boldsymbol{v}]$ and covariance

$$\boldsymbol{C} = \text{cov}(\boldsymbol{v}) = E\big[(\boldsymbol{v} - \boldsymbol{\mu})(\boldsymbol{v} - \boldsymbol{\mu})^T\big].$$

Use the linearity of E to show that if $\boldsymbol{B}$ is an $m \times n$ matrix, the mean and covariance of $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{v}$ are $\boldsymbol{B}\boldsymbol{\mu}$ and $\boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T$, respectively.

Solution:

Directly from linearity we have that $E[\boldsymbol{B}\boldsymbol{v}] = \boldsymbol{B}E[\boldsymbol{v}] = \boldsymbol{B}\boldsymbol{\mu}$. Using this result, the covariance is given by

$$\begin{aligned}
\text{cov}[\boldsymbol{B}\boldsymbol{v}] &= E\big[(\boldsymbol{B}\boldsymbol{v} - \boldsymbol{B}\boldsymbol{\mu})(\boldsymbol{B}\boldsymbol{v} - \boldsymbol{B}\boldsymbol{\mu})^T\big] \\
&= E\big[\boldsymbol{B}(\boldsymbol{v} - \boldsymbol{\mu})(\boldsymbol{v} - \boldsymbol{\mu})^T \boldsymbol{B}^T\big] \\
&= \boldsymbol{B}E\big[(\boldsymbol{v} - \boldsymbol{\mu})(\boldsymbol{v} - \boldsymbol{\mu})^T\big]\boldsymbol{B}^T \\
&= \boldsymbol{B}\boldsymbol{C}\boldsymbol{B}^T
\end{aligned}$$

□

(b) Use part (a) to prove

$$\boldsymbol{x}_{\text{LS}} \sim \mathcal{N}(\boldsymbol{x}, \sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}).$$

Hint: First show that $\boldsymbol{x}_{\text{LS}} = \boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$.

Solution:

We assume that a linear transformation of a normal random vector is also distributed normally.

Since $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$, we have that $\boldsymbol{x}_{\text{LS}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A}x + (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{\epsilon} = \boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{\epsilon}$. Recall that $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, so by linearity and part(a), $E[\boldsymbol{x}_{\text{LS}}] = \boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{0} = \boldsymbol{x}$ and $\text{cov}(\boldsymbol{x}_{\text{LS}}) = \boldsymbol{0} + (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\sigma^2\boldsymbol{I}\boldsymbol{A}((\boldsymbol{A}^T\boldsymbol{A})^{-1})^T = \sigma^2\boldsymbol{A}^T\boldsymbol{A}$ since $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ is symmetric. I.e. $(\boldsymbol{A}^T\boldsymbol{A})(\boldsymbol{A}^T\boldsymbol{A})^{-1} = I$ impies $((\boldsymbol{A}^T\boldsymbol{A})^{-1})^T(\boldsymbol{A}^T\boldsymbol{A}) = I^T = I = (\boldsymbol{A}^T\boldsymbol{A})(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ if and only if $((\boldsymbol{A}^T\boldsymbol{A})^{-1})^T = (\boldsymbol{A}^T\boldsymbol{A})$.

□

(c) In the m-file `TwoVarTest.m`, samples of $\boldsymbol{x}_{\text{LS}}$ are computed by first sampling $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \sigma^2\boldsymbol{I})$ and then computing the corresponding least squares solution $\boldsymbol{x}_{\text{LS}}$ via the normal equation. Add a line of code that also samples $\boldsymbol{x}_{\text{LS}}$ using the approach in part (b), then plot the two collections of samples together in different colors to verify that the sample distributions look roughly the same. Also compare their sample mean and covariance matrices using MATLAB's `mean` and `cov` functions.

Solution:

An example of the the requisite line of code is:

```
xx_LS = repmat(x,1,nsamp) + sigma*inv(A'*A)*A'*randn(2,nsamp);
```

Below, we compare the means and covariance matrices of the two distributions,

```
>> mean(x_LS'),mean(xx_LS')

ans =

    1.2887    0.7118


ans =

    0.7682    1.2383

>> cov(x_LS'), cov(xx_LS')

ans =

   47.0491  -47.0134
  -47.0134   46.9989


ans =

   50.1726  -50.1840
  -50.1840   50.2155
```

Note that the mean of each solution is relatively far from `[1 1]'`. The initial `x_LS` was computed via MATLAB's \ operator, and produces similarly biased results as when using the `inv` function on the standard normal data via the approach in part (b). Moreover, when using the backslash operator to calculate the mean via the approach in part (b), we get this bizarre result

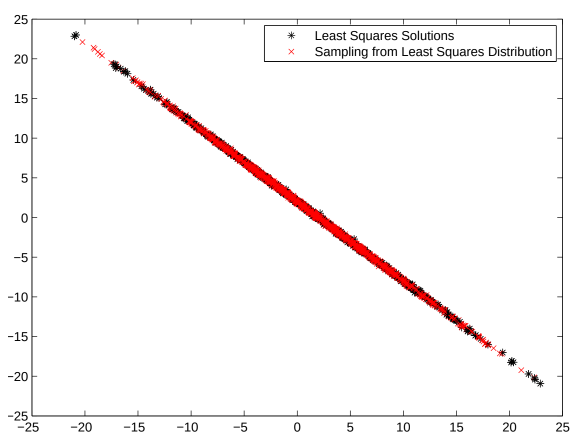
```
>> xx_LS = repmat(x,1,nsamp) + sigma*(A'*A)\(A'*randn(2,nsamp));
>> mean(xx_LS')

ans =

  -24.9983   26.8798
```

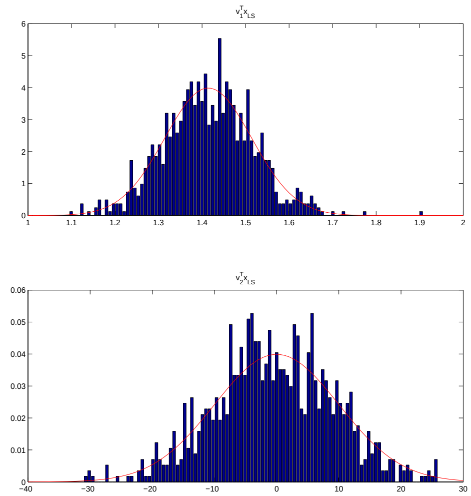
From which, I can only assume that something other than LU decomposition is happening. In any case, solving for $\boldsymbol{x}$ in $(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{x} = \boldsymbol{b}$ is ill-conditioned, and this may be a bi-product of this.

Below, is the distribution of both methods overlayed on each other.



□

(d) Create normalized histograms (using the `hist` function) from the samples of $\boldsymbol{v}_1^T \boldsymbol{x}_{\mathrm{LS}}$ and $\boldsymbol{v}_2^T \boldsymbol{x}_{\mathrm{LS}}$ computed in `TwoVarTest.m` and then graph the resulting sample probabilty densities together with the analytic true normal densities given by (1.25).



The code to produce these is given below

```
norm_curve = @(x,sig,mu)( 1/(sqrt(2*pi)*sig)*exp(-(x-mu).^2/(2*sig^2)) );
figure(4)
  subplot(2,1,1), [n1,c1] = hist(v1'*x_LS,100);
  bar(c1,n1/sum(n1*(c1(2)-c1(1)))); % Normalized Histogram
  hold on;
  t = get(gca(),'xlim')*[1-(0:.01:1);0:.01:1]; % This makes a linspace from the current axes' xlim
  plot(t,norm_curve(t,(sigma/s1),v1'*x),'r-'); % Plot normal curve
  title('v_1^Tx_{LS}')

  subplot(2,1,2), [n2,c2] = hist(v2'*x_LS,100);
  bar(c2,n2/sum(n2*(c2(2)-c2(1)))), % Normalized Histogram
  hold on;
  t = get(gca(),'xlim')*[1-(0:.01:1);0:.01:1]; % This makes a linspace from the current axes' xlim
  plot(t,norm_curve(t,(sigma/s2),v2'*x),'r-'); % Plot normal curve
  title('v_2^Tx_{LS}')
```