Agenda

- Random Vectors
- Multivariate normal distribution
- Principal components
- Estimation of Gaussian processes
- The Wiener Filter

Random vectors

Random vectors with dependent components arise in many problems

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{pmatrix}$$

- Expectation

$$E(X) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \cdots \\ E(X_n) \end{pmatrix} = \mu$$

- Covariance matrix

$$\text{Cov}(X) = (E(X_i X_j) - E(X_i)E(X_j))_{i,j} = \Sigma$$

Properties

Take $\boldsymbol{Y} = \boldsymbol{AX}$, $\boldsymbol{A}$ is a $\boldsymbol{m}$ by $\boldsymbol{n}$ non-random matrix.

- Expectation

$$\boldsymbol{E}(\boldsymbol{Y}) = \boldsymbol{A}\boldsymbol{E}(\boldsymbol{X}).$$

- Covariance matrix

$$\text{Cov}(\boldsymbol{Y}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X})\boldsymbol{A}^{\boldsymbol{T}}$$

Multivariate normal distribution

$\boldsymbol{X}$ is a $\boldsymbol{n}$-dimensional random vector.

$\boldsymbol{X}$ is said to have a *multivariate normal distribution* if every linear combination of its components is normally distributed

$$\boldsymbol{X \sim N(\mu, \Sigma) \Leftrightarrow a^T X \sim N(a^T \mu, a^T \Sigma a)}$$

- $\boldsymbol{\mu}$ is the mean of $\boldsymbol{X}$
- $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{X}$

Theorem (Cramer) Assume $\boldsymbol{\Sigma}$ is nonsingular. This is equivalent to saying that the density of $\boldsymbol{X}$ is of the form

$$\boldsymbol{f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right).}$$

Remark: if $\boldsymbol{\Sigma}$ is singular, $\boldsymbol{X}$ does not have a density.

Properties

- Marginal distribution:

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \quad X^{(1)} = (X_1, \ldots, X_p)^T,\ X^{(2)} = (X_{p+1}, \ldots, X_n)^T$$

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

  Then

$$X^{(1)} \sim N(\mu^{(1)}, \Sigma_{11}).$$

  Marginal distribution of any subset of coordinates is multivariate normal.

- Conditional distribution: The conditional distribution of $X^{(2)}$ given $X^{(1)}$ is multivariate normal. There is a formula for the cond. mean and covariance (this is the subject of linear regression analysis).

- Independence: $\boldsymbol{X}$ is multivariate normal. If $\boldsymbol{X^{(1)}}$ and $\boldsymbol{X^{(2)}}$ are uncorrelated, i.e. Cov$(\boldsymbol{X^{(1)}}, \boldsymbol{X^{(2)}}) = \boldsymbol{\Sigma_{12}} = \boldsymbol{0}$, they are independent! (This is a very special property of the multivariate distribution.)
- $\boldsymbol{Y} = \boldsymbol{AX}$ ($\boldsymbol{A}$ non-random)

$$\boldsymbol{Y \sim N(A\mu, A\Sigma A^T).}$$

Principal components

$$X \sim N(\mu, \Sigma), \quad f(x) \propto e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}.$$

- Level sets of the density are ellipsoids

$$(x-\mu)^T \Sigma^{-1}(x-\mu) = Cste.$$

- The *principal components* are the principal axes of this ellipsoid.

Principal components

- Matrix of principal components

$$\mathbf{\Phi} = \text{Col}(\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2}, \ldots, \boldsymbol{\varphi_n})$$

- $\mathbf{\Phi}$ diagonalizes the covariance matrix

$$\mathbf{\Sigma} = \mathbf{\Phi} \boldsymbol{D} \mathbf{\Phi}^{\boldsymbol{T}}, \quad \boldsymbol{D} = \text{diag}(\boldsymbol{d_k^2}).$$

- Principal component analysis

$$\boldsymbol{X'} = \mathbf{\Phi}^{\boldsymbol{T}} \boldsymbol{X}$$

1. The coordinates $\boldsymbol{X'_1}, \ldots, \boldsymbol{X'_n}$ are uncorrelated.
2. If $\boldsymbol{X}$ is multivariate normal, the coordinates $\boldsymbol{X'_1}, \ldots, \boldsymbol{X'_n}$ are independent.

Interpretation

Suppose $d_1^2 \geq d_2^2 \geq \ldots \geq d_n^2$

- First principal component φ_1, $\|\varphi_1\| = 1$:

$$\text{Var}(X^T u) \leq \text{Var}(X^T \varphi_1), \forall u, \ \|u\| = 1,$$

  i.e. projection with maximal variance.

- Second principal component φ_2, $\|\varphi_2\| = 1$:

$$\text{Var}(X^T u) \leq \text{Var}(X^T \varphi_2), \forall u, \ \|u\| = 1, u \perp \varphi_1.$$

- Etc.

Gaussian processes

$\boldsymbol{X} = (\boldsymbol{X_0}, \ldots, \boldsymbol{X_{n-1}})$ is a Gaussian process if $\boldsymbol{X}$ is multivariate normal.

Examples

- White noise: $\boldsymbol{X_t}$ i.i.d. $\boldsymbol{N(0, \sigma^2)}$
- Auto-regressive processes
  1. First order

$$\boldsymbol{X_0 = Z_0}, \quad \boldsymbol{X_t = \rho X_{t-1} + Z_t}, \quad \boldsymbol{Z_t \, i.i.d. \, N(0,1)}$$

  2. Second order

$$\boldsymbol{X_0 = Z_0}, \quad \boldsymbol{X_t = \alpha X_{t-1} + \beta X_{t-2} + Z_t}, \quad \boldsymbol{Z_t \, i.i.d. \, N(0,1)}$$

  3. ...
- Stationary processes: Cov$(\boldsymbol{X_t}, \boldsymbol{X_s}) = \boldsymbol{\gamma(t-s)}$.

Karhunen Loeve Decomposition

Decomposition into principal components

- $(X_t), t = 1, \ldots, n$ is a Gaussian Process $X \sim N(0, \Sigma)$.
- Analysis

$$X' = \Phi^T X, \quad \text{Cov}(X') = D = \text{diag}(d_k^2)$$

- Synthesis

$$X = \Phi X', \quad X'_k = d_k Z_k,$$

  with Z white noise, Z_k i.i.d. $N(0, 1)$.

- Karhunen-Loeve decomposition

$$X_t = \sum_{k=1}^{n} d_k Z_k \varphi_k(t).$$

Representation of Gaussian stochastic processes as superpositions of **independent components**

1. Analysis: Find the orthonormal eigenfunctions ϕ_k and the independent components Z_k.
2. Synthesis: Synthesize the process from the independent components using the orthonormal eigenfunctions.

Example: Stationary process on the circle

- Covariance $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}(s,t) = \gamma(s \ominus t), \quad 0 \le s,t < n$$

where $s \ominus t = s - t \bmod n$

- Orthonormal eigenfunctions are sines and cosines (n even)

$$\begin{aligned}
\phi_1(t) &= 1/\sqrt{n} \\
\phi_{2k}(t) &= \sqrt{2/n}\cos(2\pi kt/n), \quad k = 1,2,\ldots,n/2-1 \\
\phi_{2k+1}(t) &= \sqrt{2/n}\sin(2\pi kt/n), \quad k = 1,2,\ldots,n/2-1 \\
\phi_n(t) &= (-1)^t/\sqrt{n}
\end{aligned}$$

Estimation of Gaussian Processes

$$Y_t = X_t + Z_t, \quad 1 \leq t \leq n$$

- Y observed
- X as before
- Z white noise, Z_t i.i.d. $N(0, \sigma^2)$
- Z independent of X

Problem: Recover X from data Y. Estimate $\hat{X} = g(Y)$

Find $\hat{X}$ such that MSE is minimum

$$MSE(X, \hat{X}) = E\|X - \hat{X}\|_2^2 = E \sum_{t=0}^{T-1} (X_t - \hat{X}_t)^2$$

The solution is given by

$$g(Y) = E\{X|Y\}$$

This is the classical regression problem.

Why? We want to choose $g(Y) = (g_1(Y), \ldots, g_n(Y))$ such that $E\|X - T(Y)\|^2$ is minimum. In other words, for each t, we want to minimize

$$\int_{x \in \mathbf{R}^n} \int_{y \in \mathbf{R}^n} (x_t - g_t(y))^2 p_{X,Y}(x, y)\, dxdy$$

where p is the joint distribution of the vectors $x = (x_1, \ldots x_n)$ and $y = (y_1, \ldots y_n)$.

$$\int_{y \in \mathbf{R}^n} \int_{x \in \mathbf{R}^n} (x_t - y)^2 p(x, y)\, dxdy = \int_{y \in \mathbf{R}^n} \left[\int_{x \in \mathbf{R}^n} (x_t - g_t(y))^2 p(x|y)\, dx \right] dy$$

The solution is to minimize the integrand for each value of y.

$$g_t(y) = \text{argmin}_\mu \int_{x \in \mathbf{R}^n} (x_t - \mu)^2 p(x|y)\, dx$$

and the solution is

$$g_t(y) = \int_{x \in \mathbf{R}^n} x_t p(x|y)\, dx = E(X_t|Y = y).$$

Wiener's Filter

In our setup, MSE Estimation Problem:

$$\text{How to get} \quad \boldsymbol{E\{X|Y\}}$$

Solution:

$$\boldsymbol{E\{X|Y\} = \sum_{k=1}^{\infty} w_k \langle Y, \phi_k \rangle \phi_k}$$

Wiener Filter weights

$$\boldsymbol{w_k} = \frac{\text{Signal Power}}{\text{Signal+ Noise Power}} = \frac{\boldsymbol{d_k^2}}{\boldsymbol{d_k^2 + \sigma^2}}$$

Sequence Space View

$$Y_t = X_t + Z_t$$

Look at the problem in the basis of principal components Change of basis:

$$\begin{aligned} \langle Y, \phi_k \rangle &= \langle X, \phi_k \rangle + \langle Z, \phi_k \rangle \\ Y'_k &= X'_k + Z'_k \end{aligned}$$

- Z'_k Gaussian white noise sequence, Z'_k i.i.d. $N(0, \sigma^2)$
- σ noise level
- $X'_k = \langle X, \phi_k \rangle$ coordinates of X
  - The X'_k's are i.i.d. $N(0, d_k^2)$
  - The X'_k's are independent of the Z'_k's

If we come up with an estimator $\hat{\boldsymbol{X}}' = \boldsymbol{g}(\boldsymbol{Y}')$ for $\boldsymbol{X}'$, then one can estimate $\boldsymbol{X}$ via $\hat{\boldsymbol{X}} = \boldsymbol{\Phi}\hat{\boldsymbol{X}}'$.

We have the isometry

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|^2 = \|\hat{\boldsymbol{X}}' - \boldsymbol{Y}'\|^2$$

and, of course,

$$\boldsymbol{E}\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|^2 = \boldsymbol{E}\|\hat{\boldsymbol{X}}' - \boldsymbol{Y}'\|^2$$

In short, this is the same estimation problem and the solution is given by

$$\hat{\boldsymbol{X}}' = \boldsymbol{E}\{\boldsymbol{X}'|\boldsymbol{Y}'\}$$

The solution is now easy to calculate because

$$E(X'_k|Y' = y') = E(X'_k|Y'_k = y'_k)$$

From

$$\begin{array}{ccccc} Y'_k & = & X'_k & + & Z'_k \\ N(0, d_k^2 + \sigma^2) & & N(0, d_k^2) & & N(0, \sigma^2) \end{array}$$

Obtain

$$E(X'_k|Y'_k = y'_k) = \frac{d_k^2}{d_k^2 + \sigma^2}\, y'_k.$$

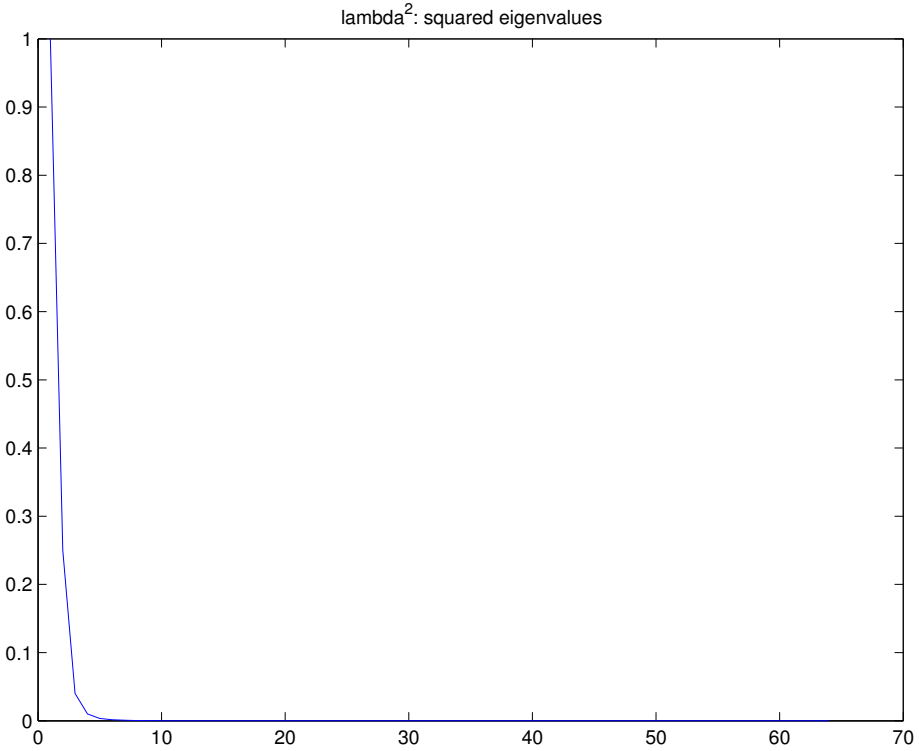
Conclusion

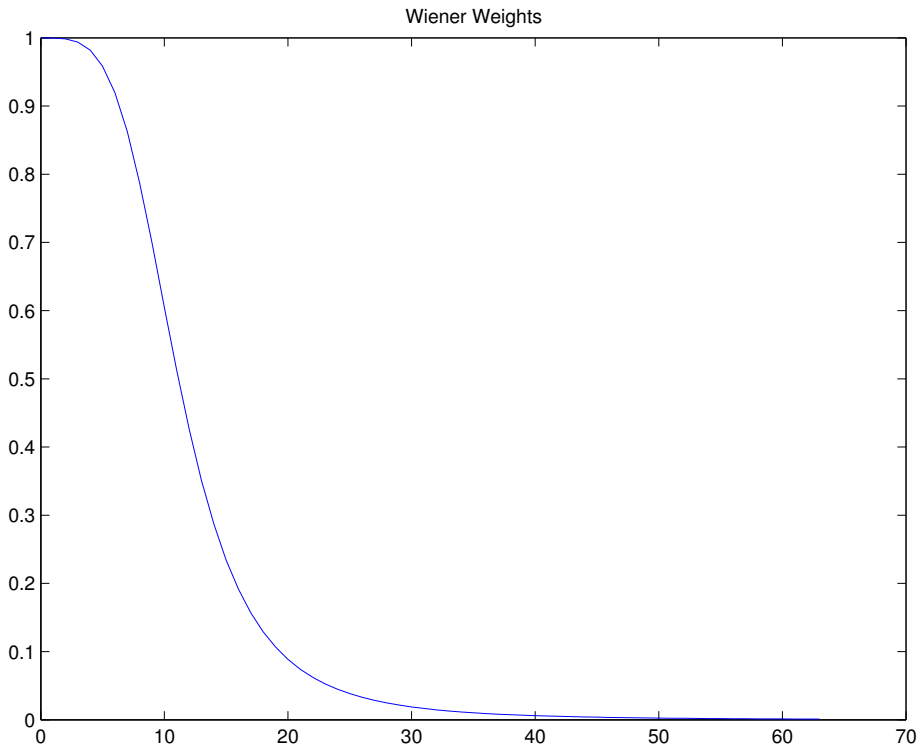
$$E(X'|Y') = WY', \quad W = \text{diag}(w_k)$$

and therefore

$$E(X|Y) = \Phi W \Phi^T Y \quad \Leftrightarrow \quad E(X_t|Y) = \sum_{k=1}^{n} w_k \langle Y, \phi_k \rangle \phi_k(t).$$

Wiener's Weights





Example

$$
\begin{aligned}
\phi_1(t) &= 1/\sqrt{n} \\
\phi_{2k}(t) &= \sqrt{2/n}\cos(2\pi kt/n), \quad k = 1, 2, \ldots, n/2 - 1 \\
\phi_{2k}(t) &= \sqrt{2/n}\sin(2\pi kt/n), \quad k = 1, 2, \ldots, n/2 - 1 \\
\phi_n(t) &= (-1)^t/\sqrt{n}
\end{aligned}
$$

Synthesize a Gaussian process:

$$X = \sum_k d_k Z_k \phi_k(t)$$

```
n = 64;
k = 0:n-1;
d = 1./(1 + k.^2);
figure; plot(d.^2)

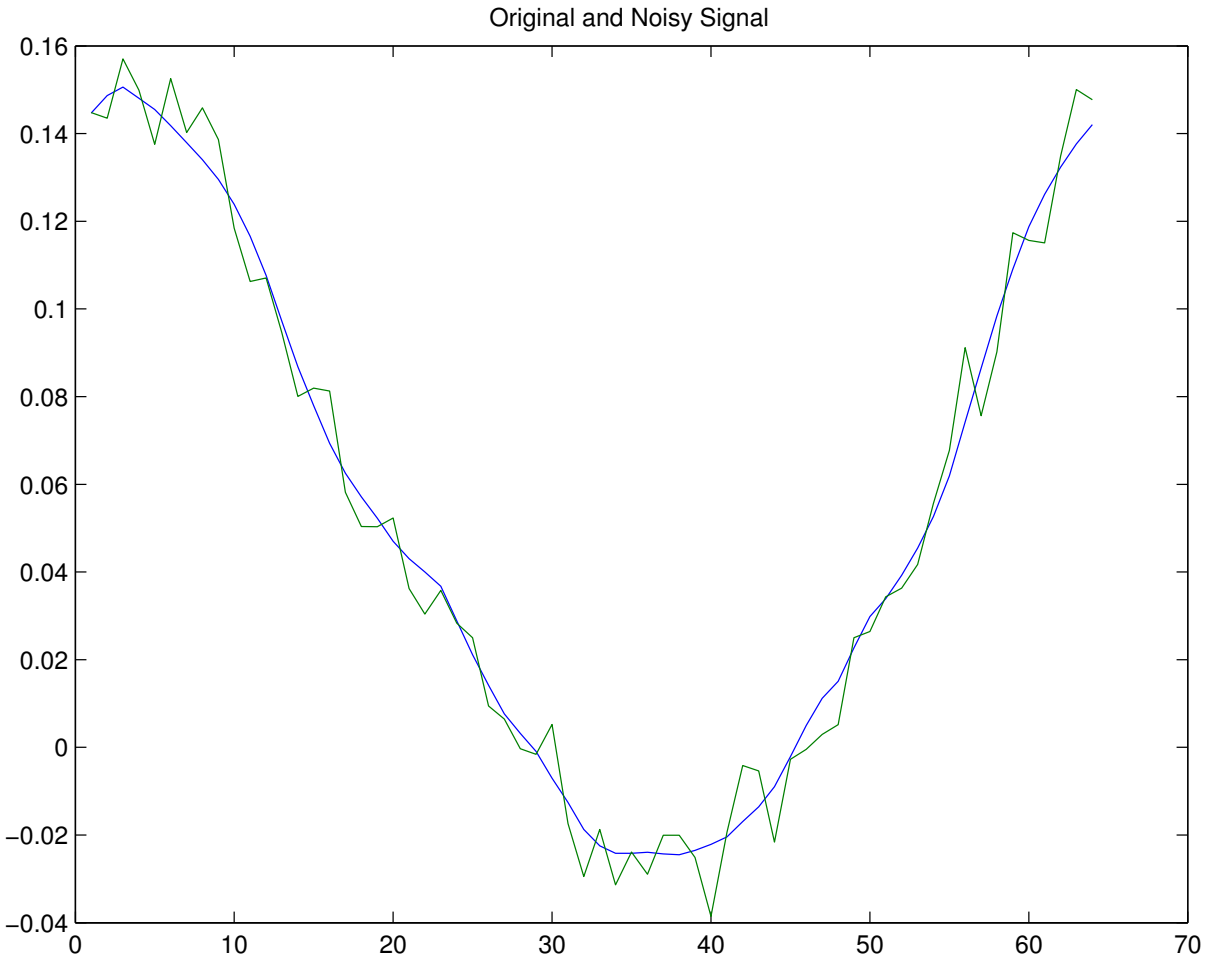
Z = randn(1,n);
Xp = lambda.*Z;
X = inv_my_dct(Xp);
```

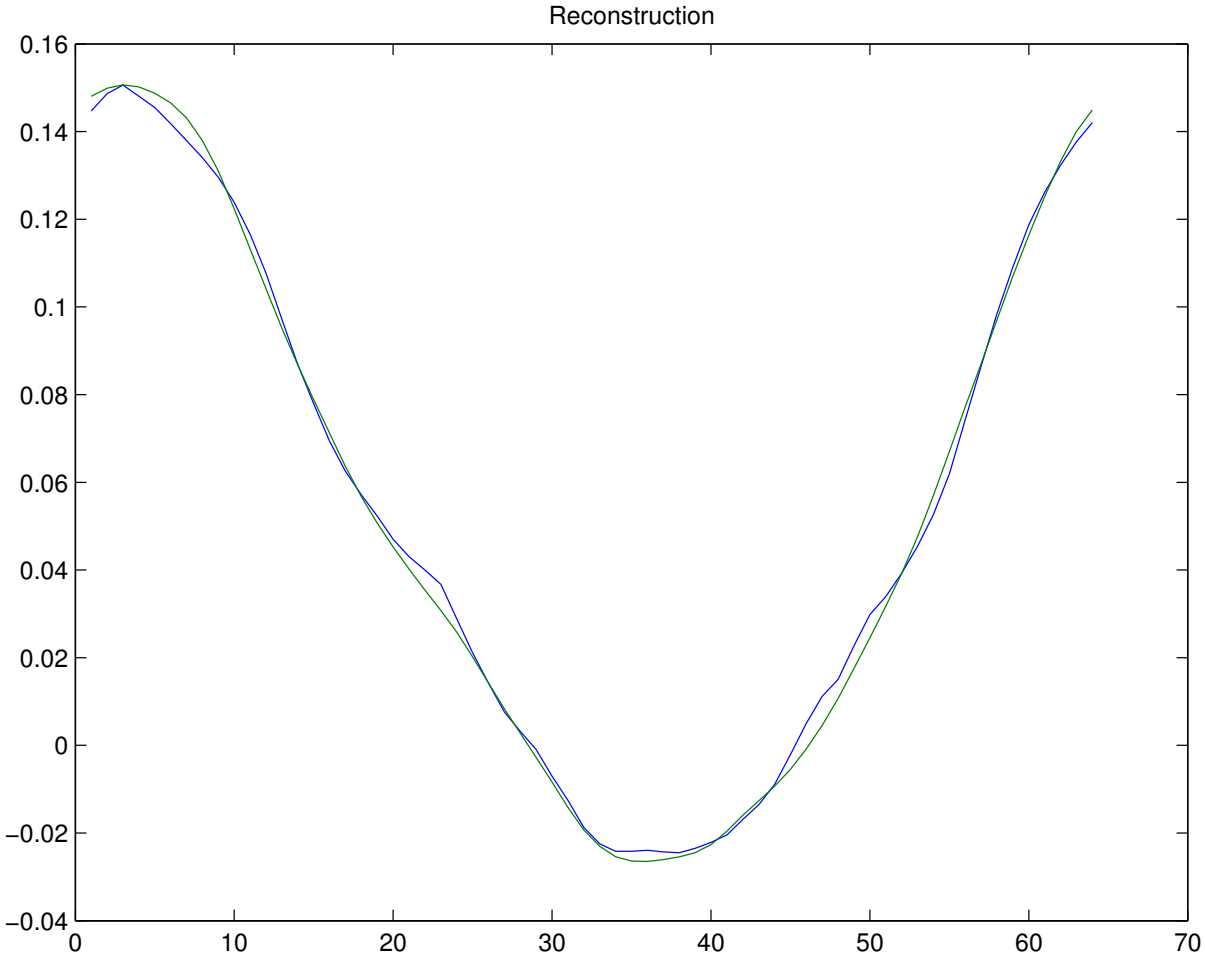
Synthesize noisy data:

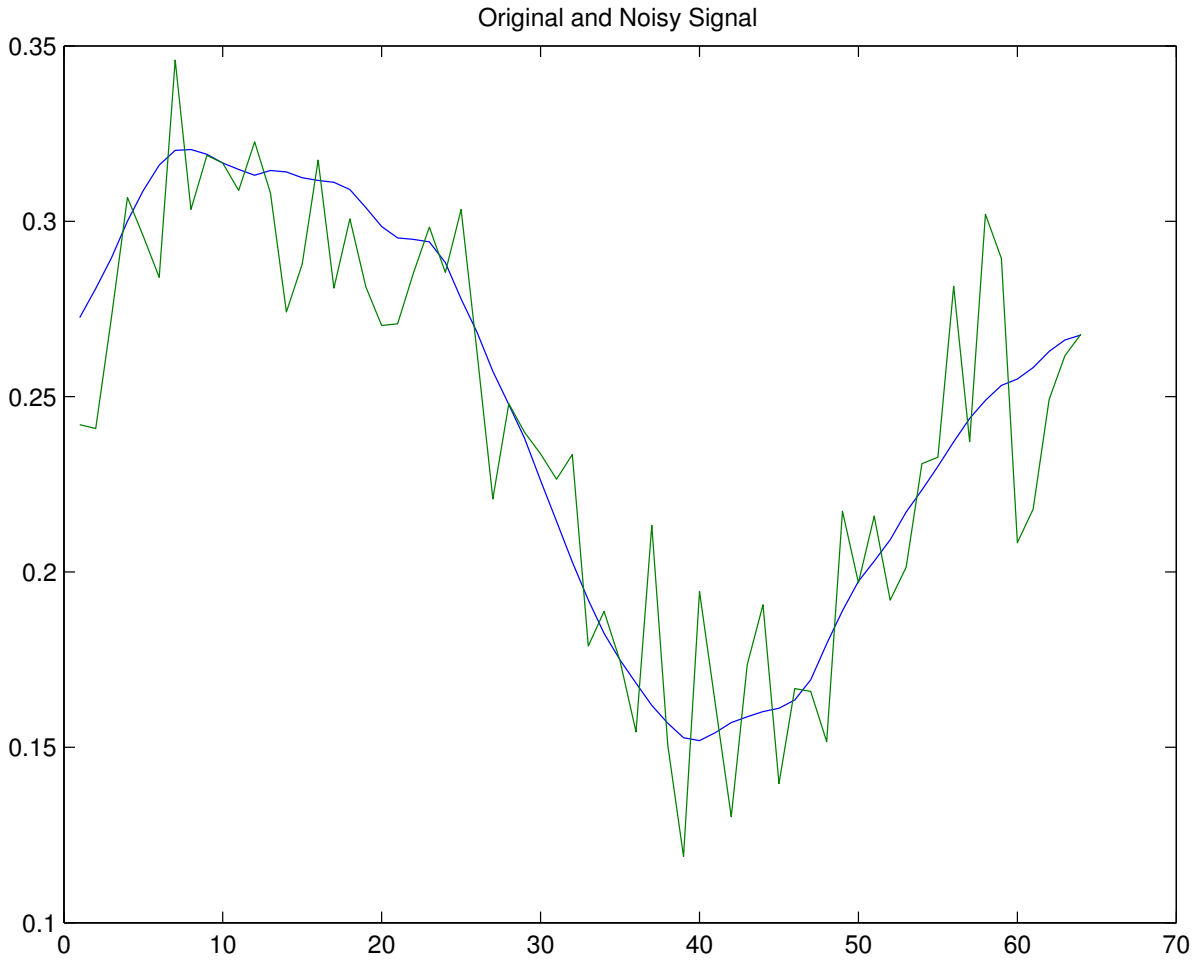
```
sigma = .003*sqrt(n);
Y = X + sigma*randn(1,n);
```

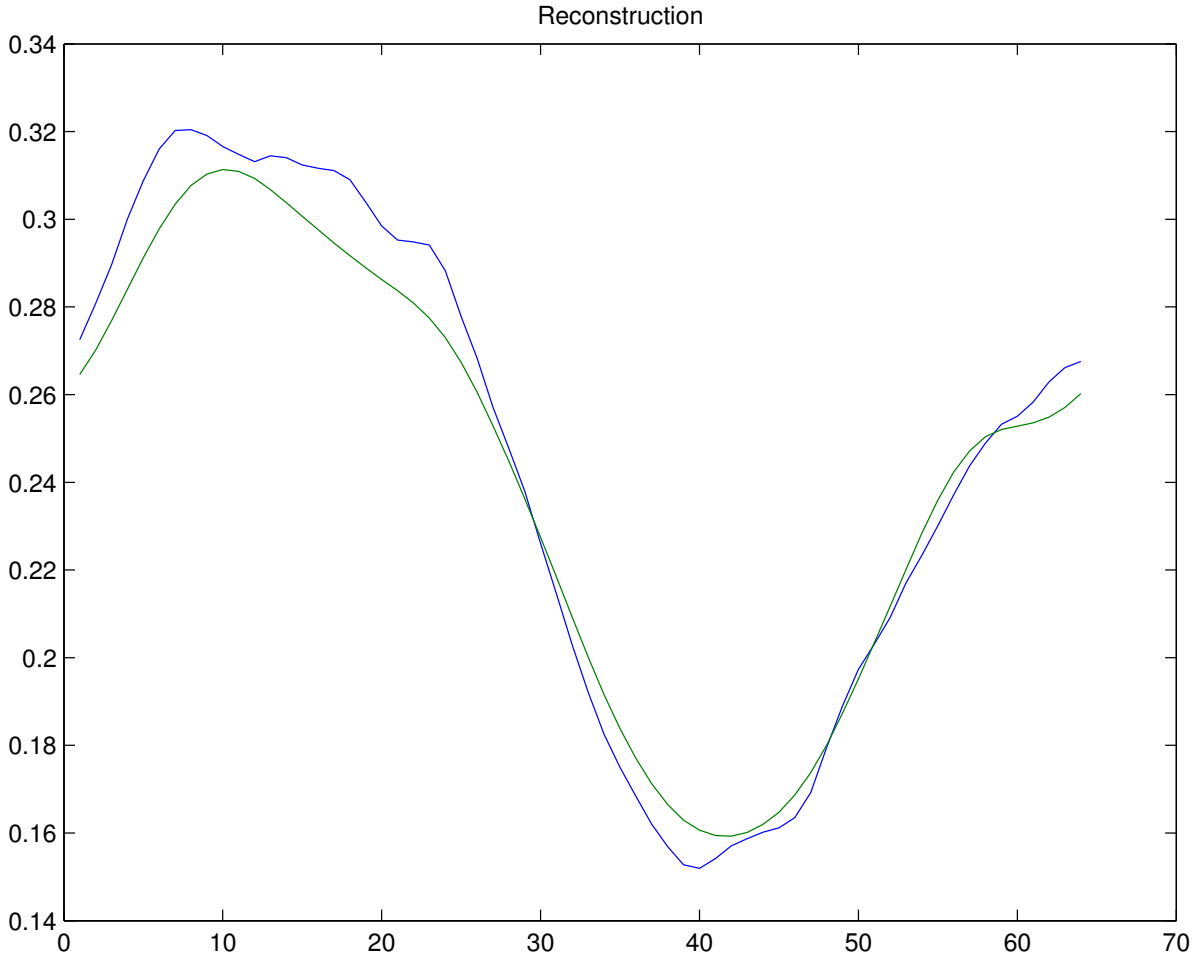
Apply Wiener's filter:

```
FY = my_dct(Y);
Weights = d.^2 ./(d.^2 + sigma.^2);
FR = FY.*Weights;
R = inv_my_dct(FR);
```


Original and Noisy Signal
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−0.02
−0.04
0
10
20
30
40
50
60
70


Reconstruction
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−0.02
−0.04
0
10
20
30
40
50
60
70


Original and Noisy Signal
0.35
0.3
0.25
0.2
0.15
0.1
0
10
20
30
40
50
60
70


Reconstruction
0.34
0.32
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0
10
20
30
40
50
60
70