Covariance, Correlation, and the KL-Expansion

John Burkardt
Department of Scientific Computing
Florida State University
http://people.sc.fsu.edu/~jburkardt/presentations/...
...kl_2012.pdf

02-03 April 2012

Covariance, Correlation, and the KL-Expansion

- **Introduction**
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

Covariance, Correlation, and the KL-Expansion

- Introduction
- **Random Variables, Covariance and Correlation**
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

Random Variables: The Probability Space

Suppose we wish to model some physical process whose outcome is influenced by chance or uncertainty.

The uncertainty can be modeled with a probability space $(\Omega, \mathcal{F}, P)$:

- Ω is a space of outcomes, whose generic example is ω;
- $\mathcal{F}$ is a σ-algebra of "events" or sets of outcomes;
- P is function which defines a probability for each event;

By choosing a particular probability space for our model, we can control everything about it . . . except the actual value of the outcome ω that will be selected any time we request a sample from the space.

Random Variables: A Function of Uncertain Argument

Having modeled the uncertainty, we turn to the physical process itself.

A random variable is a function from the sample space Ω to the state space S. We'll often assume the state space is $\mathbb{R}$:

$$x : \omega \in \Omega \to x(\omega) \in S$$

An outcome ω is randomly chosen or "sampled", in accordance with the probability function P, and this produces the "observable" value $x(\omega)$.

The sampling can be repeated, inducing a stream of ω values which are observable as a sequence of values of the random variable.

If we know the probability space in use, we can derive statistical quantities about it. Otherwise, if the model is "hidden", repeated sampling can be used to estimate its structure.

Random Variables: Probability Density Function

The probability density function $p(x)$ allows us to compute the probability of the observable states:

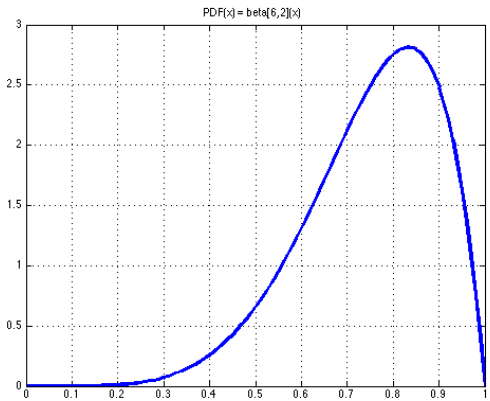
$$P(a < x < b) = \int_a^b p(x)\, dx$$

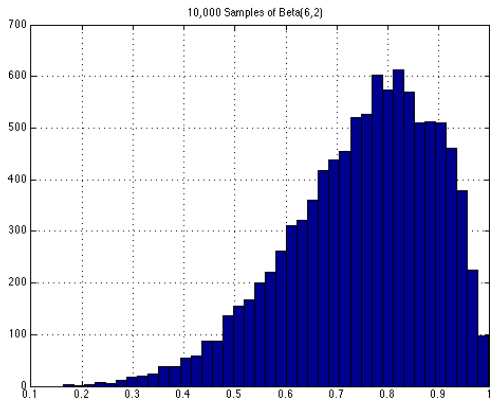
Using $\chi_{[a,b]}()$, the characteristic function of [a,b], the integral reads:

$$P(x \in [a, b]) = \int_{\mathbb{R}} \chi_{[a,b]}(x)\, p(x)\, dx$$

This second form suggests how $p(x)$ will be used within integrals to automatically account for uncertainty.

Repeated sampling of the process estimates the pdf with a histogram.





Random Variables: Probability Integrals

In the sampling approach we just discussed, we are working directly with the observed values of x, and the relative frequency of these values gives us an approximation to the probability density function $p(x)$.

But we can also look at the process as beginning in the abstract probability space, where x is actually written as a function of ω. In that case, we would integrate over the outcome space Ω, and the function $\rho(\omega)$ would represent the probability density function there, so we might wish to write

$$P(x(\omega) \in [a, b]) = \int_{\Omega} \chi_{[a,b]}(x(\omega))\, \rho(\omega)\, d\omega$$

Random Variables: Expected Value of a Variable

For a random variable $x(\omega)$, the **expected value** may be denoted by $\overline{x}$ or $\mu(x)$ or μ_x or $\mathbb{E}[x]$ or $\langle x \rangle$.

Depending on the circumstances, the expected value may be defined by:

$$\begin{aligned}
\overline{x} &= \int_{\Omega} x(\omega)\, dP \\
&= \int_{\Omega} x(\omega)\, P(d\omega) \\
&= \int_{\Omega} x(\omega)\, \rho(\omega)\, d\omega \\
&= \int_{\mathbb{R}} x\, p(x)\, dx
\end{aligned}$$

The expected value is also called the *first moment*.
Integrating x^2 gives the second moment and so on.

Random Variables: Variance of a Variable

The **variance** measures the deviation of the random variable from its expected value. It is an example of a central moment, which essentially uses $\overline{x}$ as the origin.

For a random variable $x(\omega)$, the variance may be denoted by $\text{var}(x)$ or $\sigma^2(x)$ or σ_x^2 or $\mathbb{E}[(x - \overline{x})^2]$.

Depending on the circumstances, the variance may be defined by:

$$\begin{aligned}\sigma^2(x) &= \int_\Omega (x(\omega) - \overline{x})^2 \, \rho(\omega) \, d\omega \\ &= \int_{\mathbb{R}} (x - \overline{x})^2 \, p(x) \, dx\end{aligned}$$

Random Variables: Covariance of Two Variables

The **covariance** measures the correspondence between variations from the mean in two random variables.

For random variables $x(\omega_1)$ and $y(\omega_2)$, the covariance may be denoted by $\mathrm{Cov}(x, y)$ or σ_{xy} or $\mathbb{E}[(x - \overline{x})(y - \overline{y})]$.

Depending on the circumstances, the covariance may be defined by:

$$\begin{aligned}\mathrm{Cov}(x, y) &= \int_{\Omega_1} \int_{\Omega_2} (x(\omega_1) - \overline{x})\,(y(\omega_2) - \overline{y})\,\rho(\omega_1)\,\rho(\omega_2)\,d\omega_1 d\omega_2 \\ &= \int_{\mathbb{R}^2} (x - \overline{x})\,(y - \overline{y})\,p_1(x)\,p_2(y)\,dx\,dy\end{aligned}$$

Note, in particular that $\mathrm{Cov}(x, x) = \sigma^2(x)$.

Random Variables: Correlation of Two Variables

The **correlation** of two random variables is a normalized version of the covariance, with values ranging from

- -1.0, perfectly anti-correlated,
- 0.0, uncorrelated,
- 1.0, perfectly correlated.

The correlation of two random variables may be defined by

$$\text{Corr(x,y)} = \frac{\text{Cov}(x, y)}{\sigma(x)\sigma(y)}$$

From the definition, we can see that $\text{Corr}(x, x) = 1$.

Random Variables: Covariance Matrix for Vectors

We may extend this discussion to a random value x which is an n-dimensional vector function of the outcome ω; we may distinguish this case by writing $\vec{x}(\omega)$ or $\mathbf{x}(\omega)$.

The expected value $\overline{\mathbf{x}}$ is now also a vector.

The variance $\sigma_{\mathbf{x}}^2$ is a vector:

$$\sigma_{\mathbf{x}}^2 = \int_{\mathbb{R}^n} (\mathbf{x} - \overline{\mathbf{x}})^2 \; p(\mathbf{x}) \, d\mathbf{x}$$

We can now define a covariance matrix $\mathrm{Cov}(\mathbf{x})$ whose (i, j) entry is the covariance between $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$\mathrm{Cov}(\mathbf{x})_{i,j} = \int_{\mathbb{R}^n} (\mathbf{x}_i - \overline{\mathbf{x}}_i)(\mathbf{x}_j - \overline{\mathbf{x}}_j) \, p(\mathbf{x}) \, d\mathbf{x}$$

$\mathrm{Cov}(\mathbf{x})$ is symmetric, nonnegative definite, and has diagonal $\sigma_{\mathbf{x}}^2$.

Random Variables: Correlation Matrix for Vectors

As before, the correlation matrix can be defined from the covariance matrix. Form a diagonal matrix Σ from the square roots of the variances, so that $\Sigma_{i,i} = \sigma_{\mathbf{x}}(i)$.

Then we compute the correlation matrix by:

$$\text{Corr}(\mathbf{x}) = \Sigma^{-1} \, \text{Cov}(\mathbf{x}) \, \Sigma^{-1}$$

It should be clear that the diagonal entries of Corr($\mathbf{x}$) are 1; The Cauchy-Schwarz inequality guarantees that the off-diagonal elements of the correlation matrix will lie between -1 and $+1$, and so the value of each covariance entry indicates the strength and direction of the correlation between the corresponding components.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- **A Stochastic Process**
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

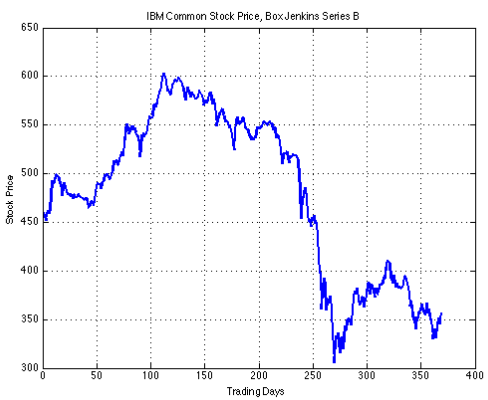
Stochastic Process: A Generalization of Random Variables

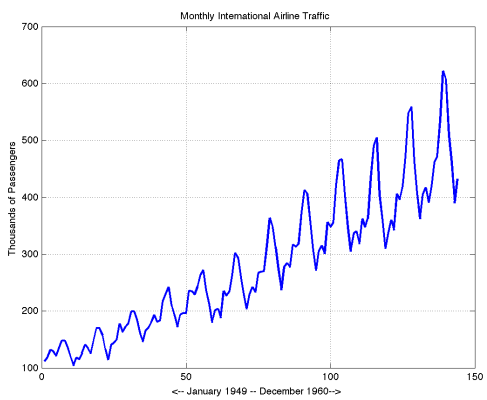
In the real world, the systems that have inherent variability or randomness are frequently time-dependent, and we can imagine describing the behavior of such a system, over time, by a highly oscillatory graph.

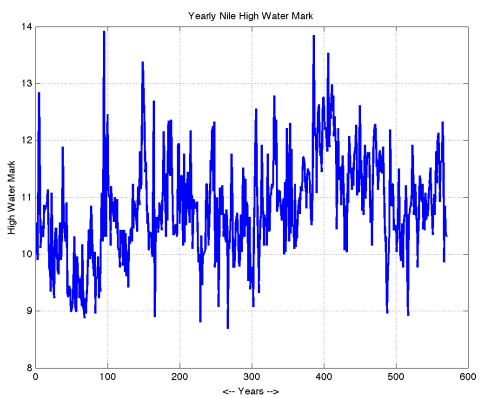
We assume that if we restarted the system, we might get a different resulting graph, but that after many restarts, we would be able to detect statistical patterns in the behavior.

Thus, it is natural to want to extend the concept of random variable or random vector to what we can think of as a sort of random function that has both explicit and implicit arguments.

This motivates the concept of a stochastic process.







Stochastic Process:

We'll think of a stochastic process as a function of both a random outcome ω and an observable parameter t, which returns a real value.

Instead of $X(t, \omega)$, the notations $X_t(\omega)$ and X_t are more common.

We require that, for any fixed t, the process is square integrable:

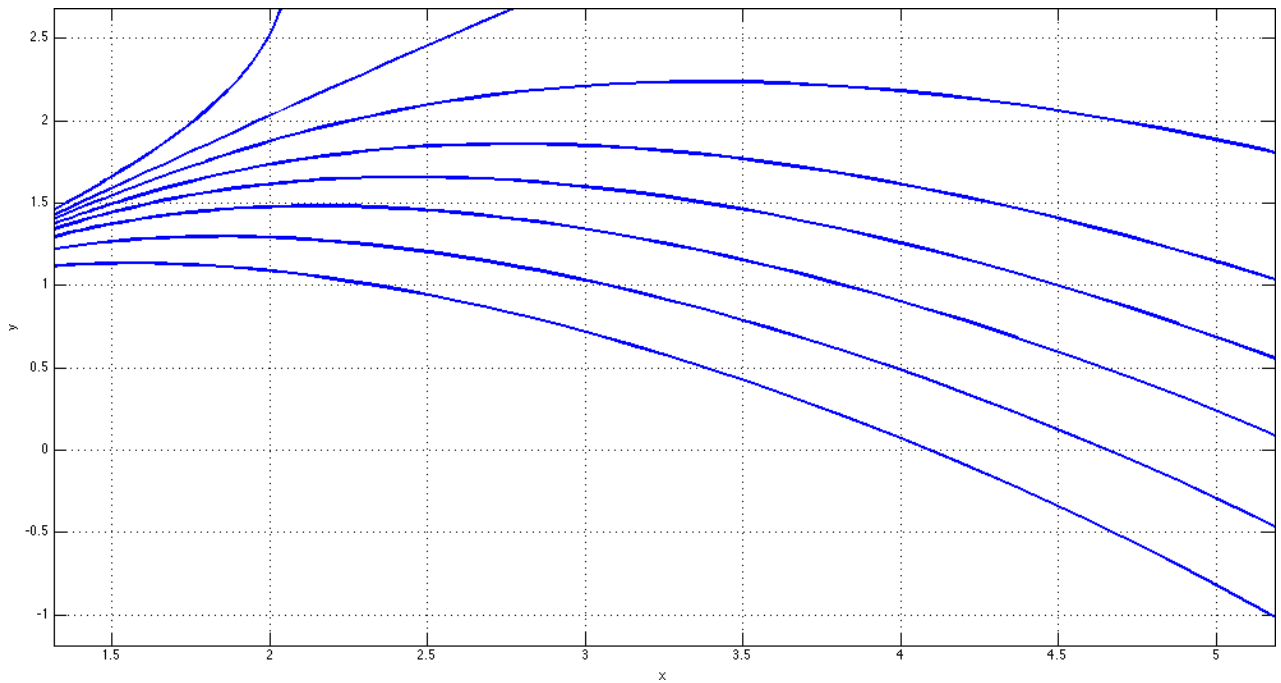
$$\int_\Omega X_t(\omega)^2 \, \rho(\omega) \, d\omega < \infty$$

The value t is sometimes called the *index* of the stochastic process. We usually expect that t is a continuously varying quantity satisfying $-\infty < a \leq t \leq b < \infty$. It is also possible to choose t to be a finite or infinite discrete index set.

Stochastic Process: Only the Initial Data Uncertain

A typical realization of a stochastic process will be a path in time, selected by the value ω, and indexed by t.

A natural example would involve a differential equation. If only the initial condition is subject to variation, then we can imagine samples of the process evolving smoothly in time, giving us a bundle of paths.



Stochastic Process: Uncertainty at All Times

In general, we will be interested in processes in which the evolution in time includes uncertainty, so that knowing the initial state is not enough to follow the path.

Stochastic Process: Associated Quantities

Let X_t be a stochastic process.

The mean of X_t, denoted $\mu_X(t)$, is:

$$\mu_X(t) = \int_\Omega X(t,\omega)\, d\omega$$

The variance of X_t, denoted $\mathrm{var}_X(t)$, is:

$$\mathrm{var}_X(t) = \int_\Omega (X(t,\omega) - \mu_X(t))^2\, d\omega$$

The covariance of X_t, denoted $\mathrm{K}_X(t,s)$, is:

$$K_X(s,t) = \int_\Omega (X(s,\omega) - \mu_X(s))\,(X(t,\omega) - \mu_X(t))\, d\omega$$

or, if $\mu_X(t) \equiv 0$,

$$K_X(s,t) = \int_\Omega X(s,\omega)\, X(t,\omega)\, d\omega$$

Stochastic Process: The Covariance Function

$K_X(t, s)$, the covariance of X_t, has the following properties:

- $K_X(,)$ is continuous in s and t;
- $K_X(,)$ is symmetric: $K_X(s, t) = K_X(t, s)$;
- $K_X(,)$ is positive semidefinite: $K_X(t, t) \geq 0$;

Stochastic Process: The Eigenvalue Problem

If we define an operator T_{K_X} for any function $f \in L^2[a, b]$, by

$$T_{K_X}(f) = \int_a^b K_x(s, t)\, f(s)\, ds$$

then we can pose the eigenvalue problem: find eigenvalues λ_i and eigenfunctions $\phi_i(t)$ so that

$$T_{K_X}(\phi_i) = \int_a^b K_x(s, t)\, \phi_i(s)\, ds = \lambda_i \phi_i(t)$$

The sequence of solutions $\phi_i(t), i = 1 \ldots \infty$ can be arranged as an orthonormal basis for the space $L^2[a, b]$. It is natural to arrange the sequence so that the corresponding sequence of eigenvalues is descending.

Stochastic Process: The Representation of X_t

Mercer's theorem now tells us that the solutions of the eigenvalue problem provide a representation for the original stochastic process

$$X_t = \mu_X(t) + \sum_{i=1}^{\infty} \alpha_i \, \phi_i(t)$$

where the coefficients α can be determined by

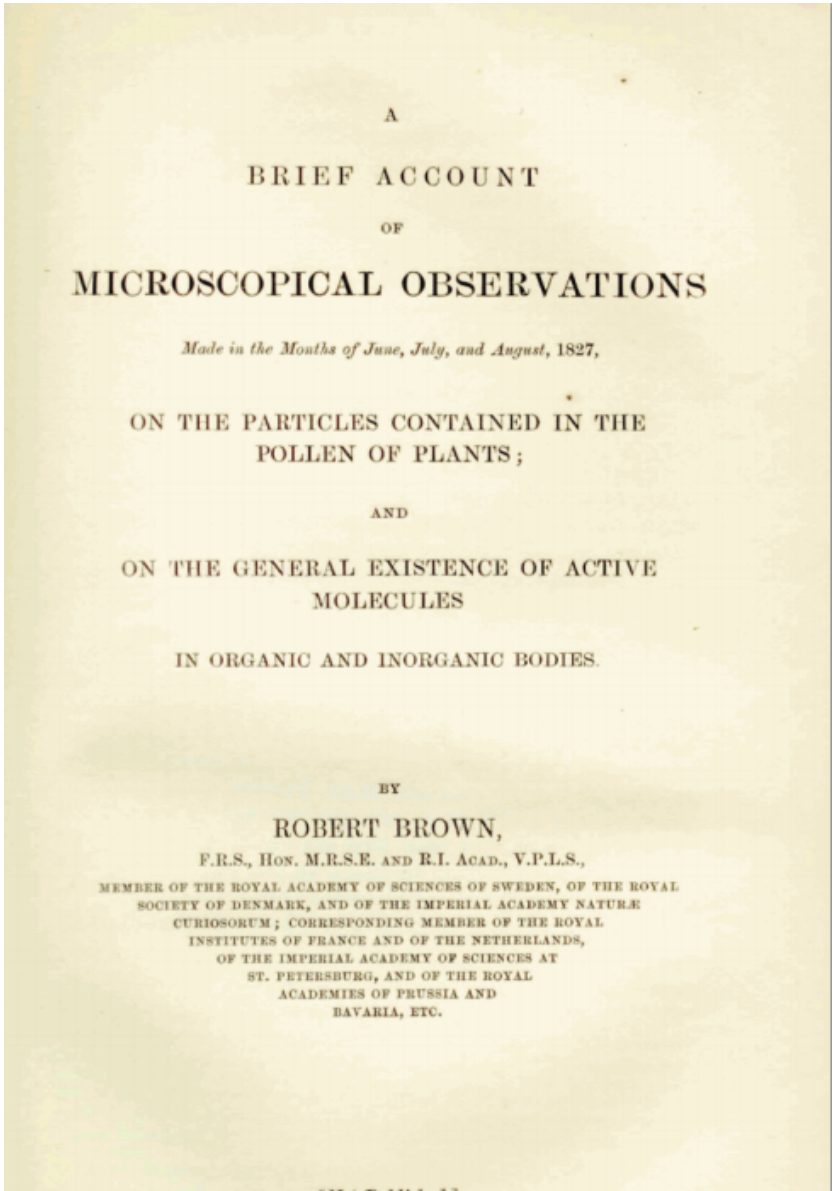
$$\alpha_i = \int_a^b (X_t - \mu_X(t)) \, \phi_i(t) \, dt$$

The variables α_i have zero mean, are uncorrelated, and have variance λ_i.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- **Brownian Motion**
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

BROWN: An Idealization of Observed Random Motion



A

BRIEF ACCOUNT

OF

MICROSCOPICAL OBSERVATIONS

Made in the Months of June, July, and August, 1827,

ON THE PARTICLES CONTAINED IN THE POLLEN OF PLANTS;

AND

ON THE GENERAL EXISTENCE OF ACTIVE MOLECULES

IN ORGANIC AND INORGANIC BODIES.

BY

ROBERT BROWN,

F.R.S., Hon. M.R.S.E. and R.I. Acad., V.P.L.S.,

MEMBER OF THE ROYAL ACADEMY OF SCIENCES OF SWEDEN, OF THE ROYAL SOCIETY OF DENMARK, AND OF THE IMPERIAL ACADEMY NATURÆ CURIOSORUM; CORRESPONDING MEMBER OF THE ROYAL INSTITUTES OF FRANCE AND OF THE NETHERLANDS, OF THE IMPERIAL ACADEMY OF SCIENCES AT ST. PETERSBURG, AND OF THE ROYAL ACADEMIES OF PRUSSIA AND BAVARIA, ETC.

Brownian motion is a mathematical idealization inspired by observation of the incessant, irregular motion of pollen particles in liquid.

The idealized impulses may be thought of as a kind of white noise, uncorrelated forces varying randomly in direction and strength over time.

But physics tells us that the path of the particle must be continuous. As time progresses, the particle's position is essentially the integral of the white noise.

BROWN: Mathematical Properties

We write W_t (in honor of Norbert Wiener) to represent a typical instance of Brownian motion, evaluated at time t.

A Brownian motion is assumed to satisfy the following requirements:

1. $W_0 = 0$;
2. $W_{t2} - W_{t1} \sim \mathcal{N}(0, t2 - t1)$ for $0 \leq t1 \leq t2$;
3. $W_{t2} - W_{t1}$ and $W_{t4} - W_{t3}$ independent if $0 \leq t1 \leq t2 \leq t3 \leq t4$;
4. W_t is (almost surely) continuous in t.

The particle starts at 0, and its distance from the origin at time t has variance t, which means its typical distance is actually of the order of $\sqrt{t}$. That is the same kind of behavior one sees in the simpler mathematical model of the *random walk*.

The third requirement means that the increments or changes over disjoint time intervals have no influence on each other.

BROWN: Expected Value and Variance

Given the mathematical properties, what are the statistical characteristics of W_t?

By condition 1:

$$E(W_t) = \mu(W_t) = \langle W_t \rangle = 0$$

By condition 2:

$$\begin{aligned} \text{Var}(W_t) = \sigma^2_{W_t} &= \langle W_t - E(W_t), W_t - E(W_t) \rangle \\ &= E(\,(W_t - E(W_t))^2\,) \\ &= E(W_t^2 - 2W_t\,E(W_t) + E^2(W_t)) \\ &= E(W_t^2) \\ &= t \text{ because } W_t \sim \mathcal{N}(0, t) \end{aligned}$$

BROWN: The Covariance

By condition 3:

$$
\begin{aligned}
\mathrm{Cov}(W_s, W_t) &= \langle W_s - E(W_s), W_t - E(W_t) \rangle \\
&= E(\,(W_s - E(W_s)\,(W_t - E(W_t))\,) \\
&= E(W_s W_t - W_s\,E(W_t) - W_t\,E(W_s) + E(W_s)E(W_t)) \\
&= E(W_s W_t) \\
&= E(W_s(W_s + (W_t - W_s)))\,) \text{ assume } s < t, \\
&= E(W_s^2) + E(W_s\,(W_t - W_s))) \\
&= E(W_s^2) + \cancelto{0}{E(W_s\,(W_t - W_s))}, \text{ by independent increments} \\
&= s \text{ or, in general, } \min(s, t).
\end{aligned}
$$

BROWN: The Correlation Function

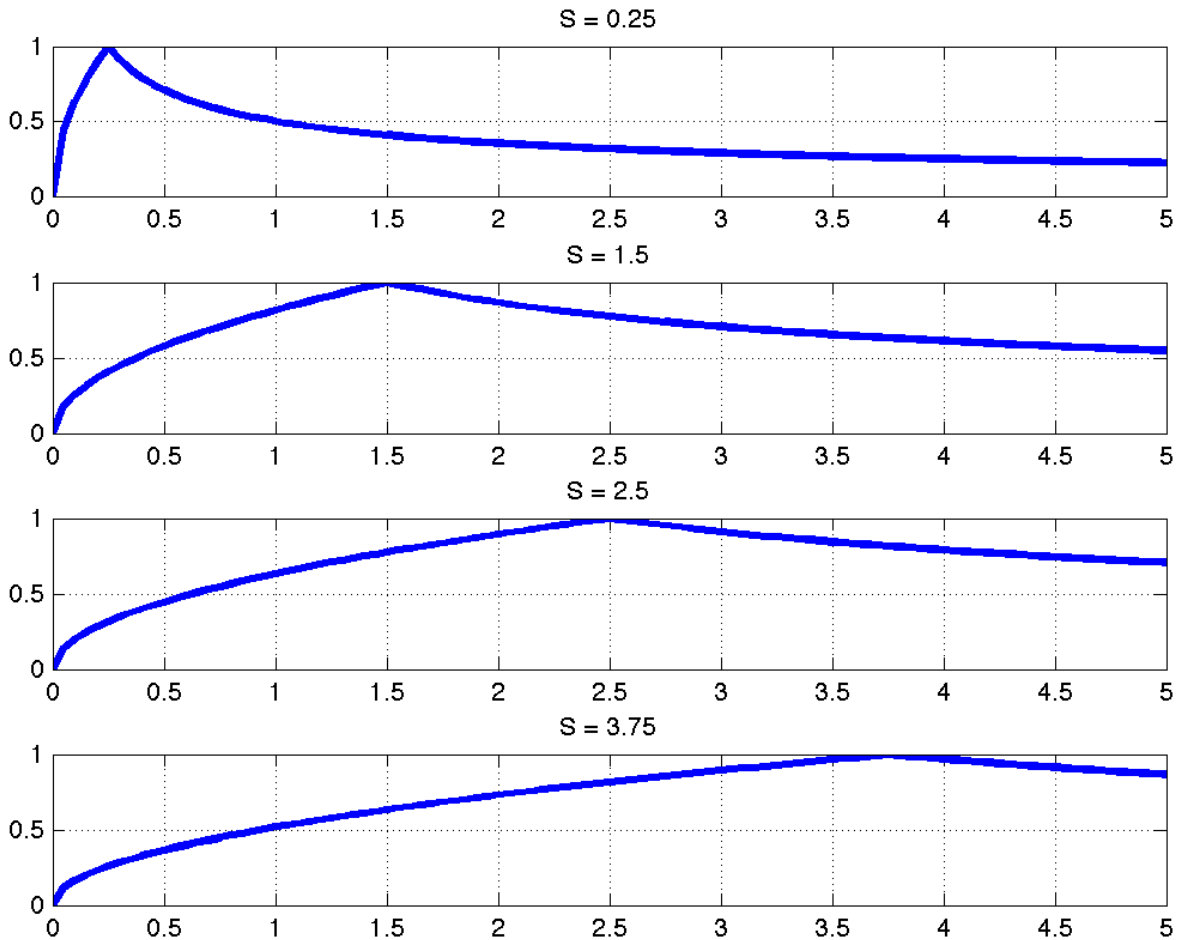
The correlation is:

$$\begin{aligned} \text{Cor}(W_s, W_t) &= \frac{\text{Cov}(W_s, W_t)}{\sigma_{W_s}\sigma_{W_t}} \\ &= \frac{\min(s,t)}{\sqrt{st}} \\ &= \sqrt{\frac{\min(s,t)}{\max(s,t)}} \end{aligned}$$

which obviously stays between 0 and 1. As the argument s gets large, the correlation $Cor(W_s, W_t)$ stays near 1 for a much broader interval. W_s tends to get large, so changes become relatively less significant, and hence the correlation is stronger.

BROWN: "Slices" of the Brownian Correlation Function

For a given value of s, the correlation function $Cor(s,t)$ is:
0 at $t = 0$; 1 at $t = s$; 0 at $t = \infty$.
The function cannot be rewritten as a function of $|s - t|$, so it is called an anisotropic correlation function.



BROWN: Simulating Brownian Motion

The requirement $W_{t2} - W_{t1} \sim \mathcal{N}(0, t2 - t1)$ for $0 \leq t1 \leq t2$ gives us an immediate method of simulating Brownian motion.

To simulate values at equal time steps Δt, we need increments ΔW distributed like $\mathcal{N}(0, \Delta t)$ – meaning they should be proportional to $\sqrt{\Delta t}$.

1. $w(0) = 0$
2. $w(i+1) = w(i) + \sqrt{\Delta t} * n(0, 1)$

We can precompute a vector of n normal values R, let G be a matrix 0 above the main diagonal and 1 on or below it, and write

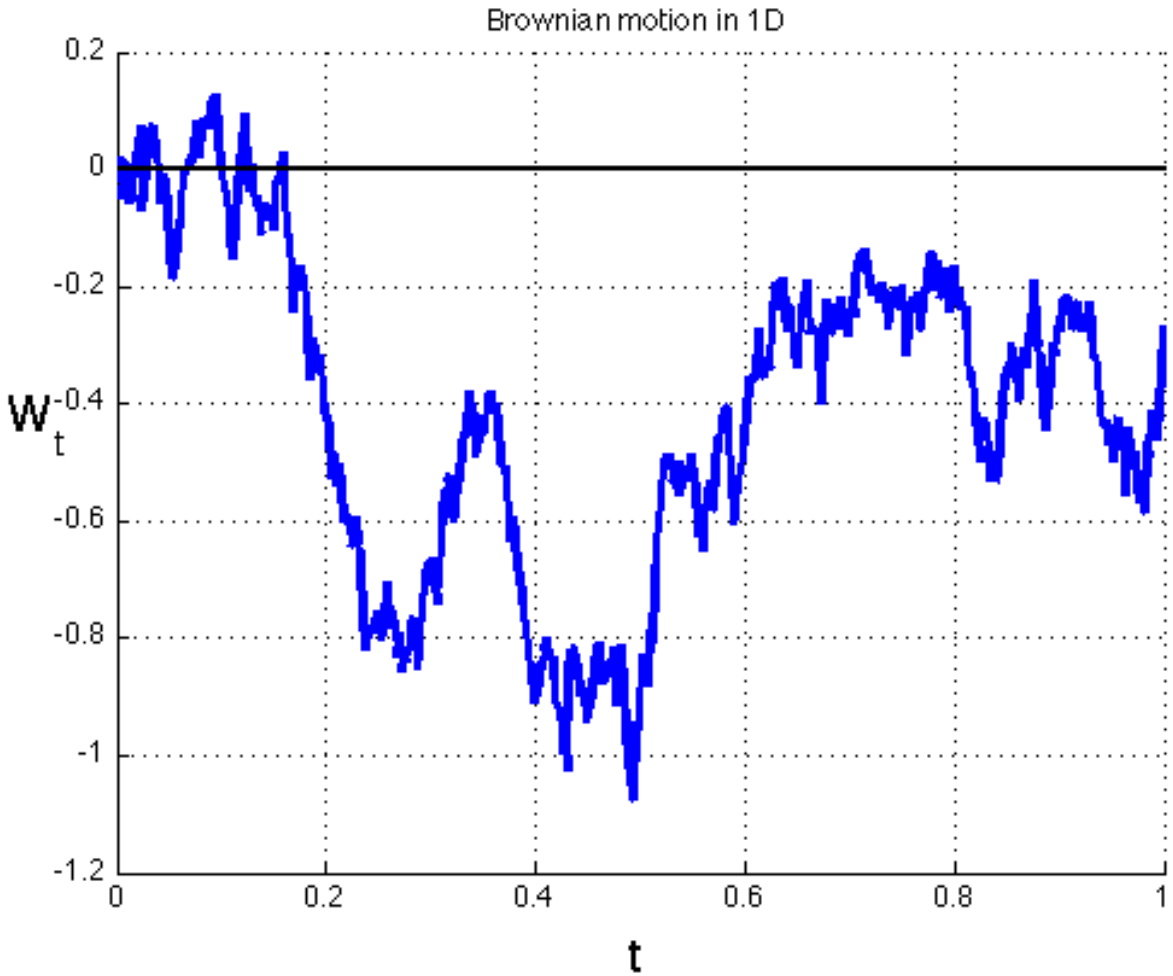
$$W = \sqrt{\Delta t} * G * R$$

which indicates how the independent values R are transformed into the correlated values W.

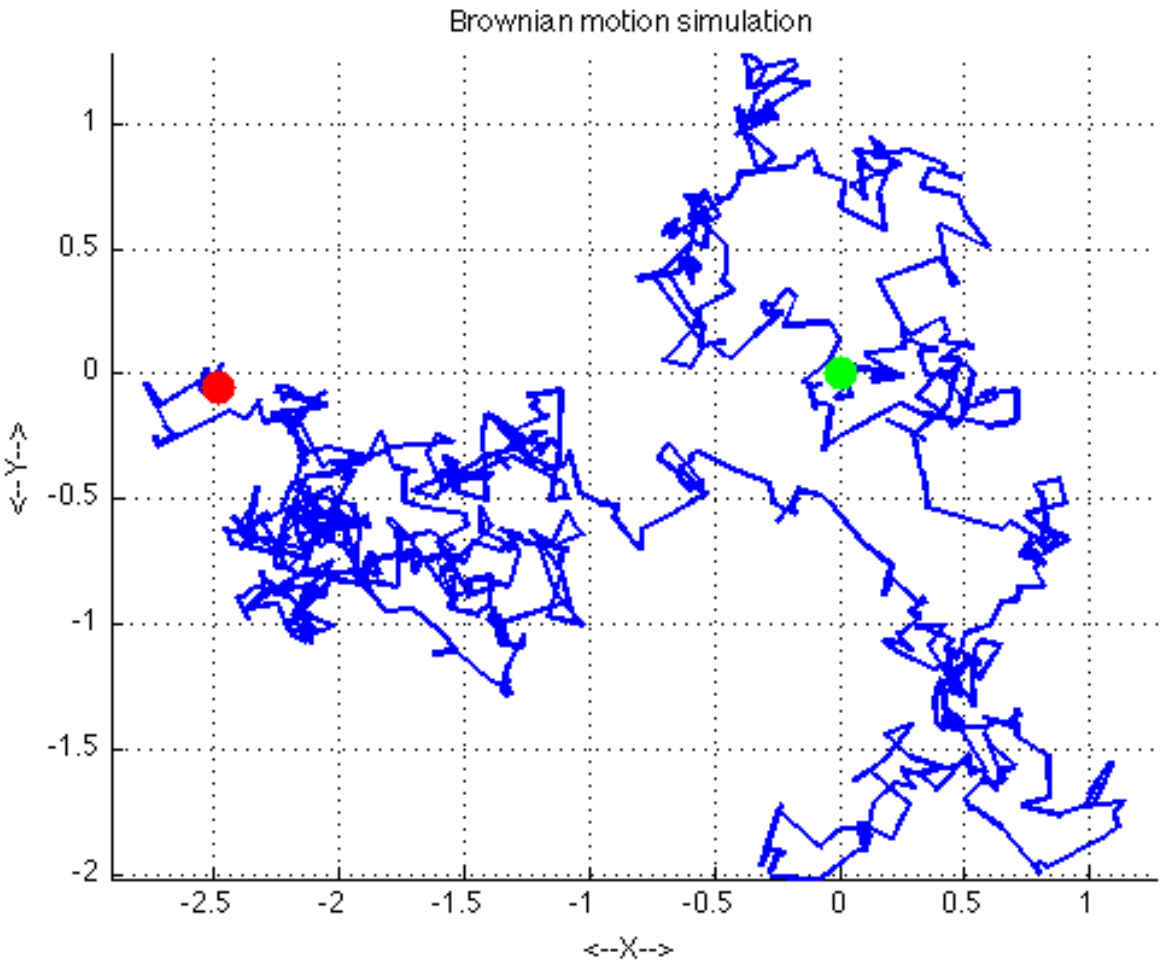
BROWN: MATLAB Simulation

```
%
%  Time step.
%
  dt = t / n;
%
%  We want N+1 successive positions.
%
  x(1:n+1) = zeros ( n + 1 );
%
%  We take N steps.
%
  s(1:n) = sqrt ( dt ) * randn ( n );
%
%  Each position is the cumulative sum of all previous steps.
%
  x(2:n+1) = cumsum ( s(1:n) );
```

BROWN: Simulation of 1D Brownian Motion



BROWN: Simulation of 2D Brownian Motion



A 2D Brownian motion can be generated from the 1D motion by choosing a uniform random direction for each step.

BROWN: Representation of Brownian Motion

Let us restrict the time argument t to the interval $[0, 1]$.
A Brownian motion W_t may be written as

$$W_t = \int_0^t \xi(s)ds = \sum_{i=0}^{\infty} \alpha_i \int_0^t \psi_i(s)ds$$

where $\{\psi_i\}_{i=0}^{\infty}$ is a complete orthonormal basis for $L^2[0, 1]$ and $\xi(s)$ is the formal time derivative $\dot{W}_s$.

The coefficients in the expansion may be determined by

$$\alpha_i = \int_0 \xi(s)\psi_i(s)ds$$

and will be independent and Gaussian.

Conversely, inserting any chosen set of independent Gaussian coefficients $\{\alpha_i\}_{i=0}^{\infty}$ into the above expansion will produce a Brownian motion.

BROWN: Representation of Brownian Motion

In particular, Brownian motion over $0 \leq t \leq 1$ can be represented as an infinite sine series expansion, with Gaussian coefficients ($\sim \mathcal{N}(0,1)$):

$$W_t = \sqrt{2} \sum_{i=0}^{\infty} \alpha_i \frac{\sin((i+\frac{1}{2})\pi t)}{(i+\frac{1}{2})\pi}$$

where

$$\alpha_i = \sqrt{2} \int_0^1 \xi(t) \sin((i+\frac{1}{2})\pi t) dt$$

BROWN: Representation of Brownian Motion

This representation suggests another way to simulate Brownian motion. This time, we don't discretize in space, but rather in frequency.

Computationally, we may truncate such an expansion at a given frequency, obtaining a representation that can be evaluated at any point in the interval.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- **Isotropic Correlation Functions**
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

ISO-COR

Let us consider a function $a(:)$, whose argument might be a scalar time t, or a spatial variable $x \in \mathbb{R}^n$. Suppose that, for our function $a(:)$, the correlation depends only on the **distance** between the two arguments, that is, on $|x - y|$ for scalars, or $||x - y||$ for vector arguments.

Defining $\rho(x, y) \equiv ||x - y||$, we are asserting that, for some function $c : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned} \text{Corr}(a_x, a_y) &= \frac{E((a_x - \mu_x)(a_y - \mu_y))}{\sigma_x \sigma_y} \\ &= c(||x - y||) \\ &= c(\rho(x, y)) \end{aligned}$$

In this case, we say the covariance is isotropic.

ISO-COR

Based on our understanding of physical law, it is natural to assume that many common phenomena are related in precisely this way. Thus, the gravitational attraction between two given masses depends only on their distance, not on their position.

If we assume that the stochastic problem we are interested in has an underlying isotropic covariance function, this results in a great deal of simplification in the analysis and simulation of the process.

The definition of $c()$ guarantees that

- $c()$ is symmetric with respect to x and y;
- $c()$ is bounded between -1 and $+1$;
- $c(0) = 1$

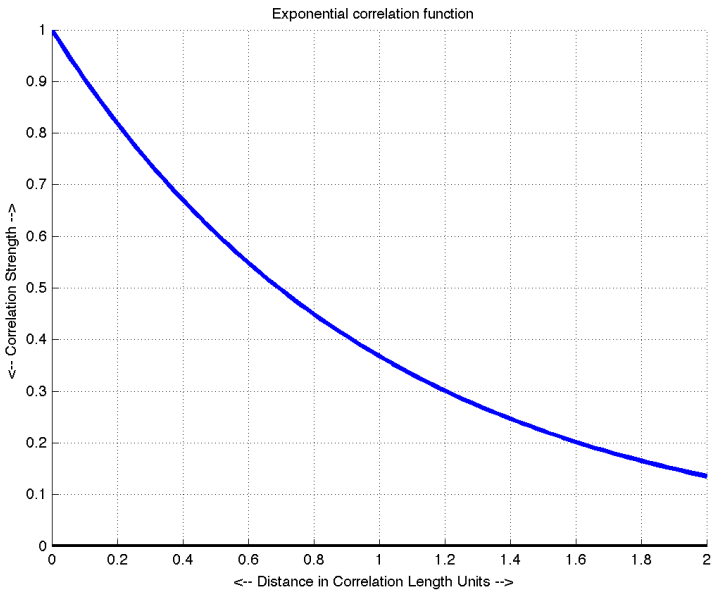
Once we have picked a function c, it is natural to immediately define a parameter called the "correlation length", which allows us to define a natural scale between our physical distances and the natural scale for c.

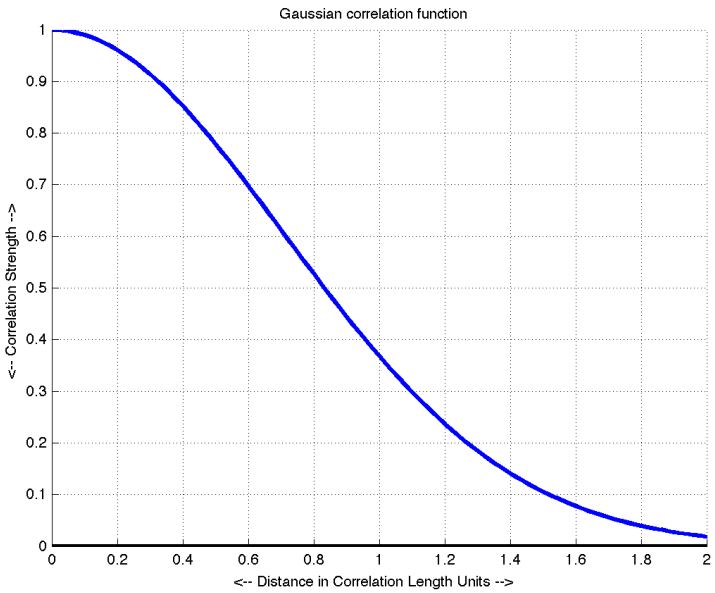
ISO-COR: Sample Correlation Functions

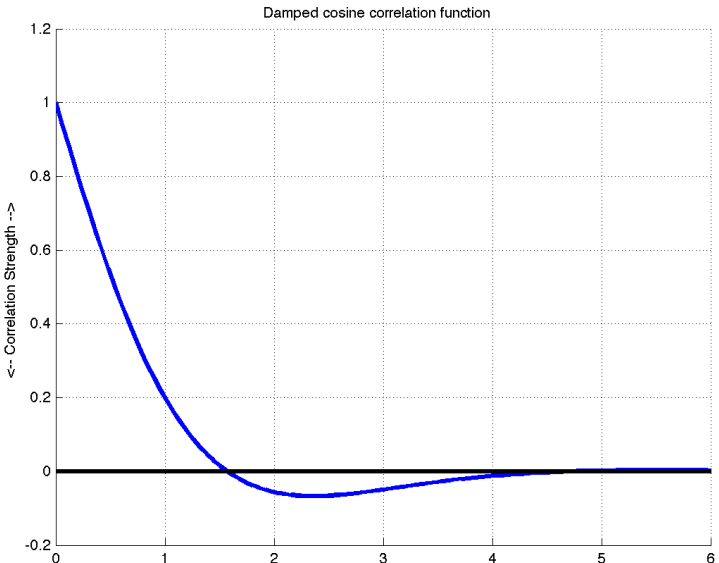
There are a few popular correlation functions based on correlation length. We assume ρ_0 is a given correlation length, define $\rho = ||x - y||$, and let $\hat{\rho} \equiv \frac{\rho}{\rho_0}$.

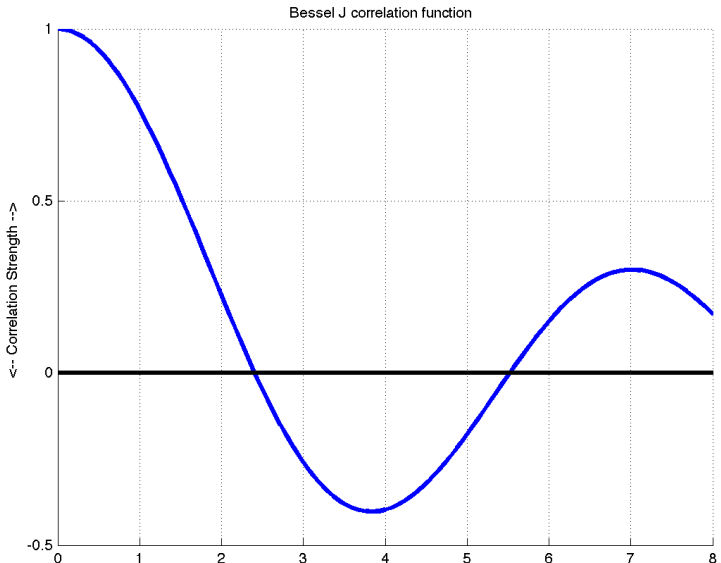
Bessel J	$c(\rho) = J_0(\hat{\rho})$
Bessel K	$c(\rho) = \hat{\rho} K_1(\hat{\rho})$
Constant	$c(\rho) = 1$ everywhere;
Damped Cosine	$c(\rho) = e^{-\hat{\rho}} * \cos(\hat{\rho})$;
Damped Sine	$c(\rho) = \frac{\sin(\hat{\rho})}{\hat{\rho}}$
Exponential	$c(\rho) = e^{-\hat{\rho}}$
Gaussian	$c(\rho) = e^{-\hat{\rho}^2}$
Linear	$c(\rho) = \max(1 - \hat{\rho}, 0)$;
White noise	$c(\rho) = 1$ if $\rho = 0$, 0 otherwise;

ISO-COR: Correlation Function Plots









ISO-COR: Computing a Correlated Path or Sample Vector

Our correlation function describes the statistical behavior of sample functions $f(\rho)$. We might suppose that $f(\rho)$ is a spatially varying thermal conductivity in a material, a porosity in an underground strata, or the temporally varying strength of an electrical signal.

To do computations, we need to create samples or realizations of $f()$, and these must be generated in a way that reflects the underlying correlation function.

Typically, we want to create a discrete vector of values sampled at equal intervals. In this case, the correlation matrix has a simple structure:

$$
C = \begin{matrix}
1 & c(dx) & c(2dx) & c(3dx) & \ldots \\
c(dx) & 1 & c(dx) & c(2dx) & \ldots \\
c(2dx) & c(dx) & 1 & c(dx) & \ldots \\
c(3dx) & c(2dx) & c(dx) & 1 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{matrix}
$$

ISO-COR: Computing a Correlated Path or Sample Vector

To generate our data:

1. construct the correlation matrix C;
2. compute the eigendecomposition $C = Q * D * Q'$;
3. compute the square root matrix $S = Q * \sqrt{D} * Q'$;
4. compute N normal independent values R;
5. set $X = S * R$;

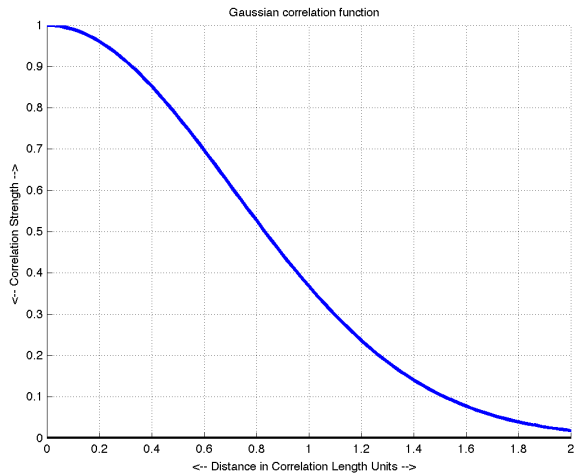
The symmetry of C implies that Q is orthogonal.
The computed X values will be a sample with the appropriate statistics.

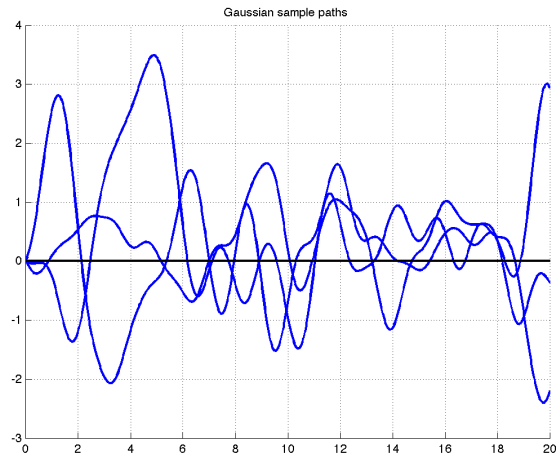
ISO-COR: Computing a Correlated Path or Sample Vector

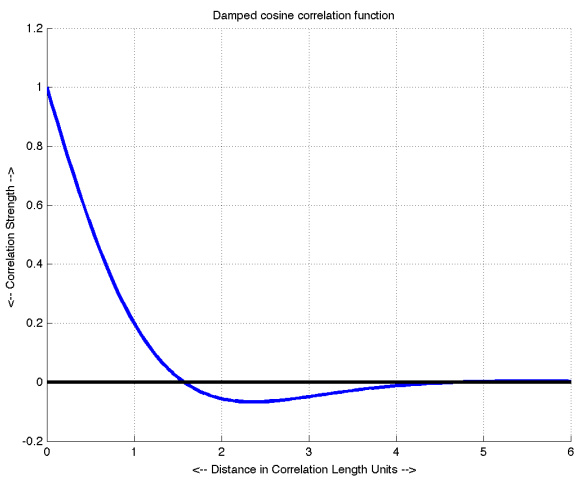
```
rho_vec = linspace ( 0.0, rhomax, n );

cor_vec = correlation ( n, rho_vec, rho0 );
cor_vec = [ cor_vec(n:-1:2)', cor_vec(1:n)' ];
cor = zeros ( n, n );
for i = 1 : n
  cor(i,1:n) = cor_vec(n+1-i:2*n-i);
end

[ v, d ] = eig ( cor );
d = max ( d, 0.0 );

sqrt_d = sqrt ( d );
sqrt_cor = v * sqrt_d * v';

r = randn ( n, 1 );

x = sqrt_cor * r;
```

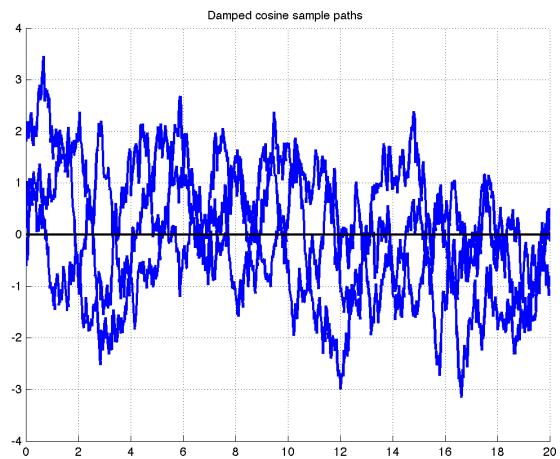
ISO-COR: Examples of Correlation Plots and Paths

```
Gaussian Correlation      |  Gaussian Paths
--------------------------+-----------------
Damped Cosine Correlation |  Damped Cosine Paths
```









ISO-COR: An FFT Approach for 2D and 3D Data

For higher dimensions, generating a sample 2D matrix or 3D block of correlated data means that the typical amount of data is $m * n$ or $l * m * n$, and the correlation matrix will require the square of this much data, making the eigenvalue decomposition approach expensive or impractical.

An alternative approach represents the data in terms of the Fast Fourier Transform. For a 2D array with Gaussian correlation, we arrive at a representation in which the typical FFT coefficient will have the form:

$$\hat{\psi}(\kappa_j, \lambda_k) = c\, e^{-(\kappa_j^2 + \lambda_k^2)}\, e^{2\pi i \phi_{j,k}}$$

where $\phi_{j,k}$ is a uniform random value. Simulating a sample field then simply involves choosing the random values $\phi_{j,k}$ and taking the inverse FFT; this process works in a similar fashion for 1D vectors or 3D blocks.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- **SVD for the Discrete Case**
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- Conclusion

SVD: The Singular Value Decomposition

Every (real) m by n matrix A has a singular value decomposition:

$$A = U S V^T$$

where

- U is an m by m orthogonal matrix ($U^T U = I$);
- S is an m by n diagonal matrix with nonnegative entries;
- V is an n by n orthogonal matrix;

The diagonal entries of S, called the singular values of A, are chosen to appear in descending order, and are equal to the square roots of the nonzero eigenvalues of AA^T or $A^T A$

SVD: Facts About the Singular Value Decomposition

r, the number of nonzero diagonal elements in S, is the rank of A; very small nonzeros may indicate numeric singularities.

The i-th diagonal element of S is the i-th largest eigenvalue of AA^T (and also of AA^T). Hence, we may write this value as $\sqrt{\lambda_i}$.

Let u_i and v_i^T be the i-th columns of U and V^T. Then A maps the i-th column of V^T to the i-th column of U.

The columns of U and V provide a *singular value expansion* of A:

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T$$

SVD: Facts About the Singular Value Decomposition

If we use all r terms, the singular value expansion is exact.

But let Ak represent the sum of just the first k terms of the expansion. Then Ak is a matrix of rank k, the sum of k rank-1 outer products. Of all rank k matrices, Ak is the best approximation to A in two senses:

Minimum L2 norm:

$||A - Ak||_2 \equiv$square root of maximum eigenvalue of $(A - Ak)^T(A - Ak)$

$||A - Ak||_2^2 = s_{k+i}^2 = \lambda_{k+1}$

Minimum Frobenius (sum of squares) norm:

$$||A - Ak||_F \equiv \sqrt{\sum_{i,j}(A_{i,j} - Ak_{i,j})^2}$$

$$||A - Ak||_F^2 = \sum_{k+1}^{r} s_i^2 = \sum_{k+1}^{r} \lambda_i$$

SVD: Using the Facts

U and V are natural bases for the input and output of A.

In the natural bases, the SVD shows that multiplying by A is simply stretching the i-th component by s_i:

$$x = \sum_{i=1}^{r} v_i^T * c_i \implies y = A * x = \sum_{i=1}^{r} u_i * (s_i * c_i)$$

The relative size of the singular values indicates the importance of each column.

The singular value expansion produces an optimal, indexed family of reduced order models of A.

SVD: Using the Facts

The singular value decomposition is the discrete version of the Karhunen-Loeve (KL) expansion that is typically applied to stochastic processes that produce, for any time t, a field of values varying spatially with x.

Since it's easier to understand discrete problems, let's prepare for the KL expansion by looking at how the SVD is used with a set of data.

Let us re-imagine the columns of our discrete data as being n snapshots in discrete time indexed by j. Each snapshot will record m values in a "space" indexed by i.

SVD

If we pack our data into a single matrix A, then $A_{i,j}$ means the measurement at position i and time j.

It is reasonable to expect correlation in this data; the "neighbors" of $A_{i,j}$, in either space or time, might tend to have similar values.

Moreover, the overall "shape" of the data for one time or one spatial coordinate might be approximately repeated elsewhere in the data.

This is exactly the kind of behavior the SVD can detect.

		<-- "Space" -->						
	1890	1	12	12	33	29	22	3	0
	1891	0	31	23	44	18	13	1	0
"Time"	1892	0	23	44	25	17	17	13	1
	1893	1	30	49	37	15	23	10	1
V	1894	0	30	18	74	9	5	0	2

SVD - Snowfall at Michigan Tech

We have a data file of the monthly snowfall in inches, over 121 winters at Michigan Tech. We'll think of the months as the "space" dimension.

Year	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Tot
1890	1	12	12	33	29	22	3	0	112
1891	0	31	23	44	18	13	1	0	130
1892	0	23	44	25	17	17	13	1	140
1893	1	30	49	37	15	23	10	1	166
1894	0	30	18	74	9	5	0	2	138
....	...	...	...	...	...	...	...	...	...
2006	6	6	27	38	37	20	31	0	165
2007	0	21	40	55	32	24	14	0	186
2008	0	17	70	85	27	5	15	0	219
2009	3	4	87	39	19	0	0	0	152
2010	0	26	33	72	18	13	18	0	180

http://www.mtu.edu/alumni/favorites/snowfall/snowfall.html

SVD

To analyze our data, we consider each of the 121 snowfall records, starting with x^{1890}, as a column of 8 numbers, and form the m=8 by n=121 matrix A:

$$A = \left[x^{1890}|x^{1891}|...|x^{2010}\right]$$

Now we determine the SVD decomposition $A = USV^T$.

The columns of U are an orthogonal set of "spatial" behaviors or modes (typical behavior in a fixed year over a span of months).

The columns of V are typical behaviors or modes in a fixed month over a span of years. In both cases, the most important behaviors are listed first.

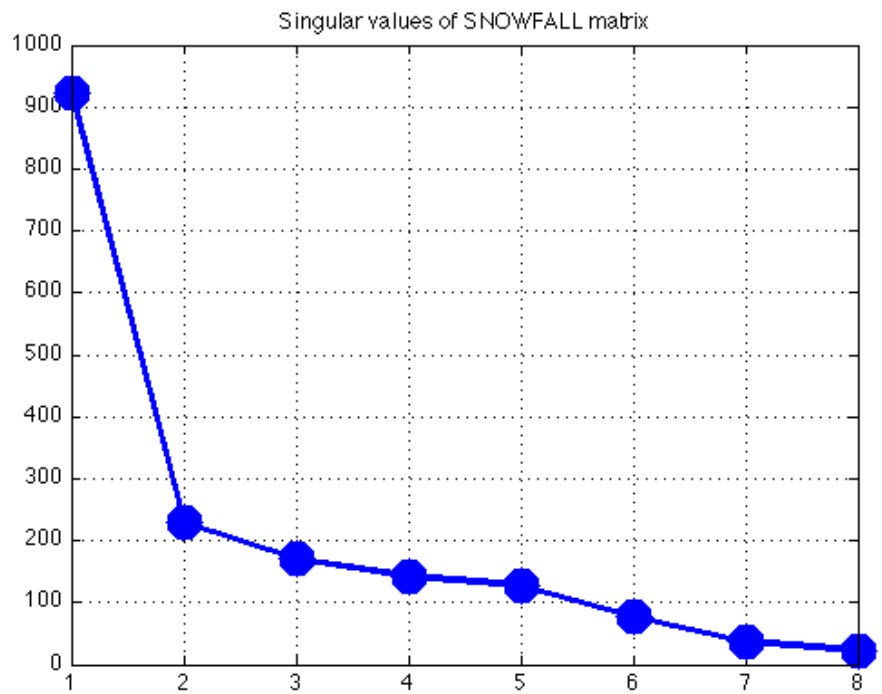
The diagonal matrix S contains the "importance" or "energy" or signal strength associated with each behavior.

SVD: The 9 Singular Values

The S data shows the relative importance of the first two modes is:

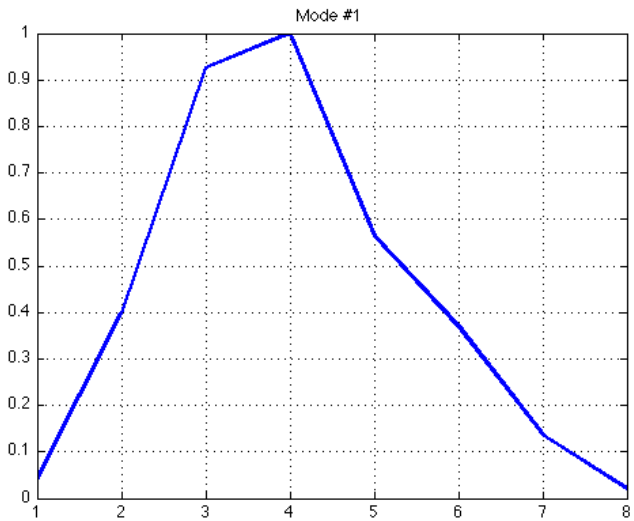
$$\frac{s1^2}{\sqrt{\sum_{i=1}^{8} s_i^2}} = 0.87 \quad \frac{s2^2}{\sqrt{\sum_{i=1}^{8} s_i^2}} = 0.05$$

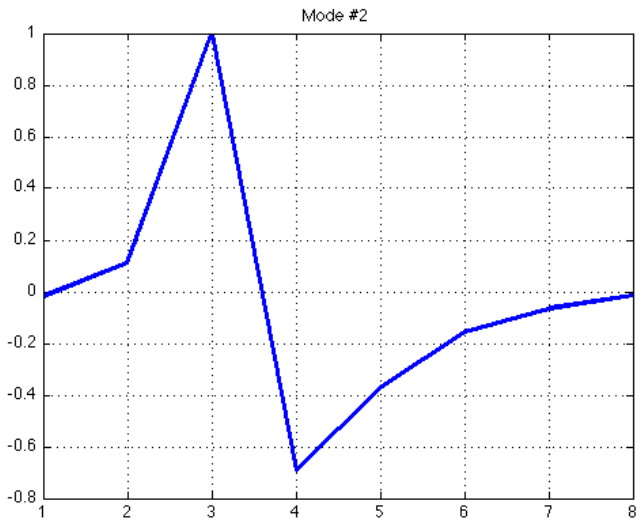
The first pair of modes, u_1 and v_1, by itself, can approximate the entire dataset with a relative accuracy of 87%.

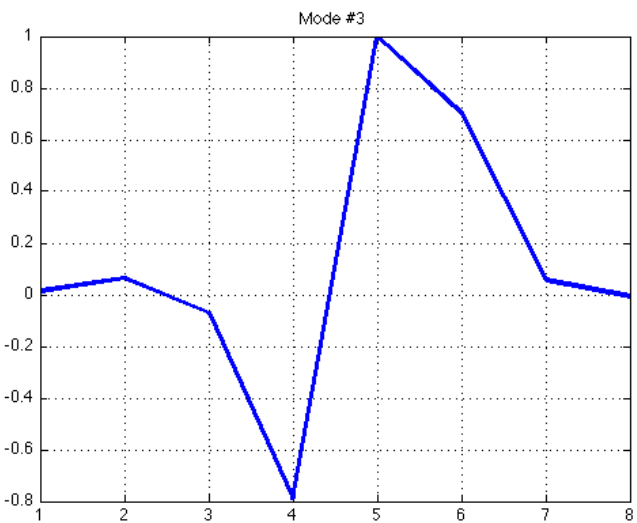


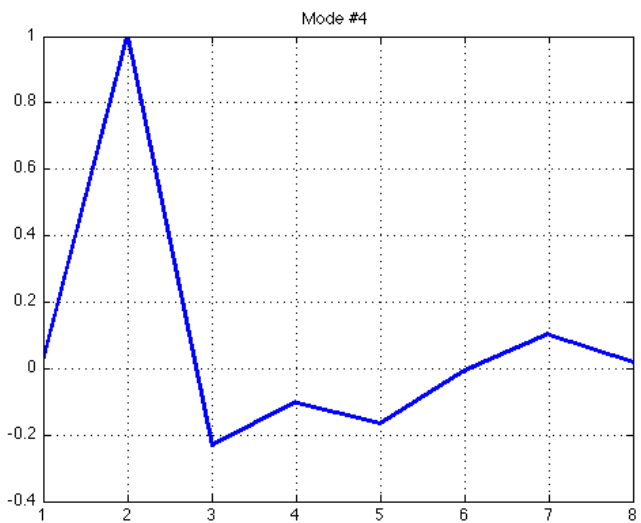
SVD: Four Strongest Snowfall Modes For a Year

1 December/January High (DOMINANT)	2 More December, less later
3 February High, less January	4 More November snow



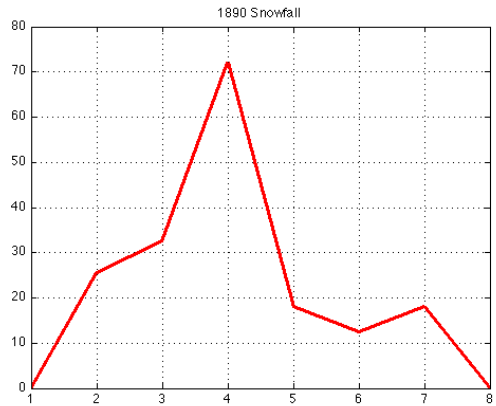


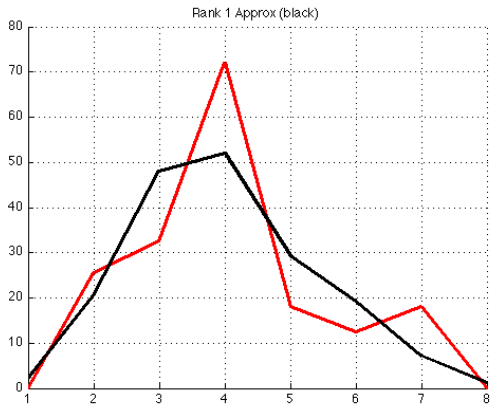


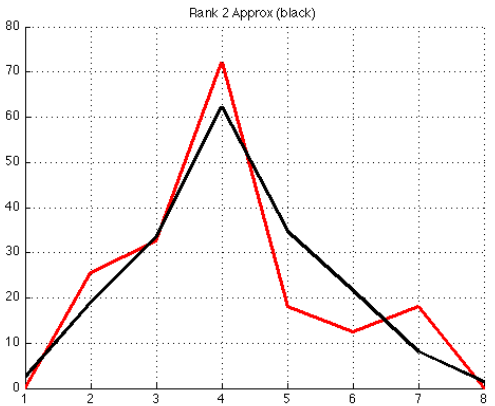


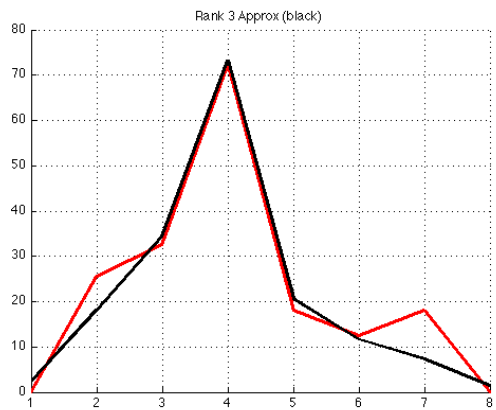
SVD: Approximating 2010-2011 Snowfall

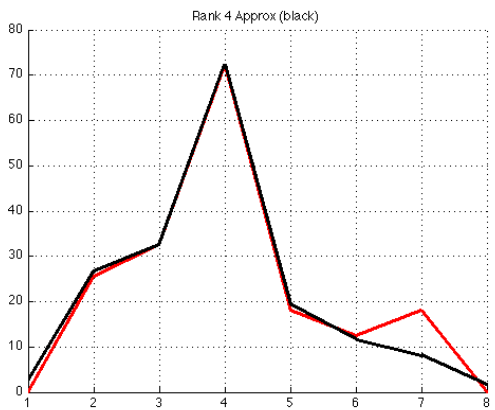
2010-2011 Data | 1 Mode | 2 Modes
---
3 Modes | 4 Modes | 5 Modes

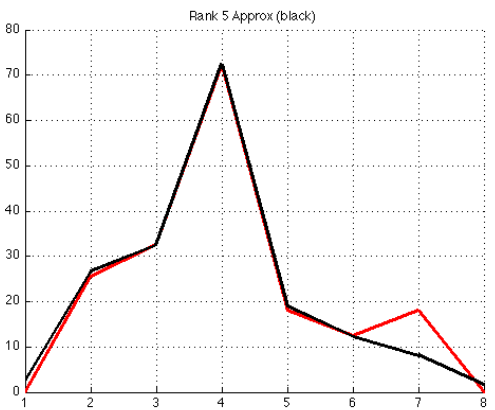








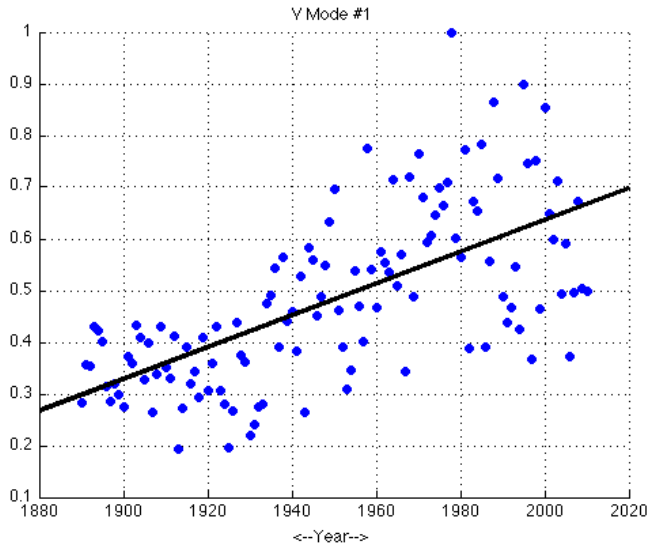


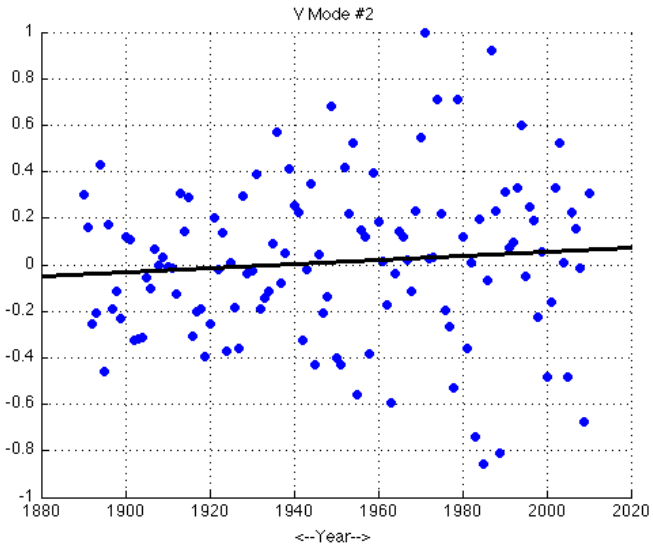


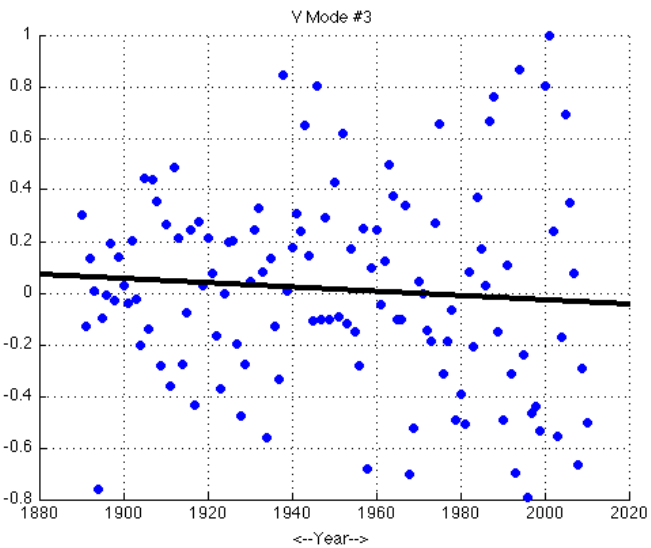
The same kind of approximating is occurring for all 121 sets of data!

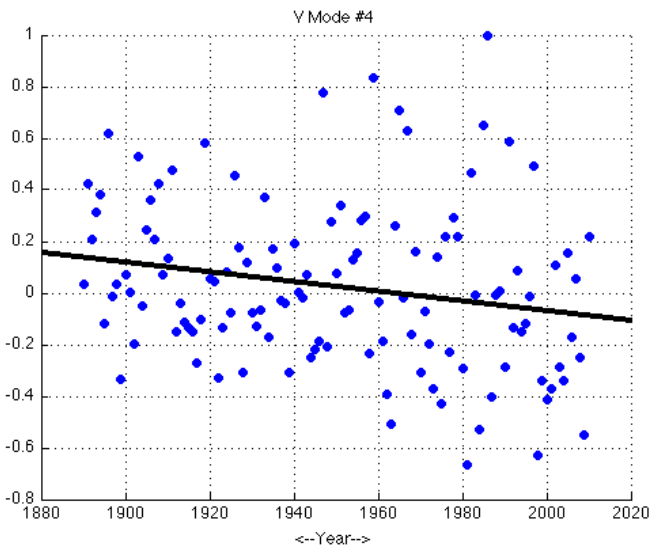
SVD: Four Strongest "Time" Modes

The linear regression line suggests the "December/January High" pattern (upper left) is steadily gaining importance over the years.



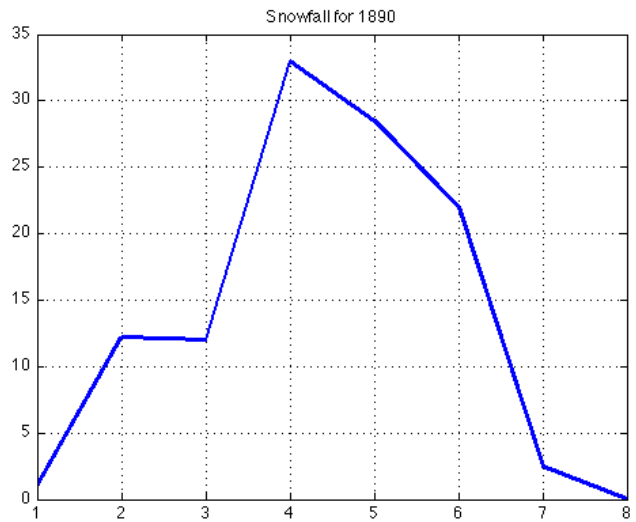


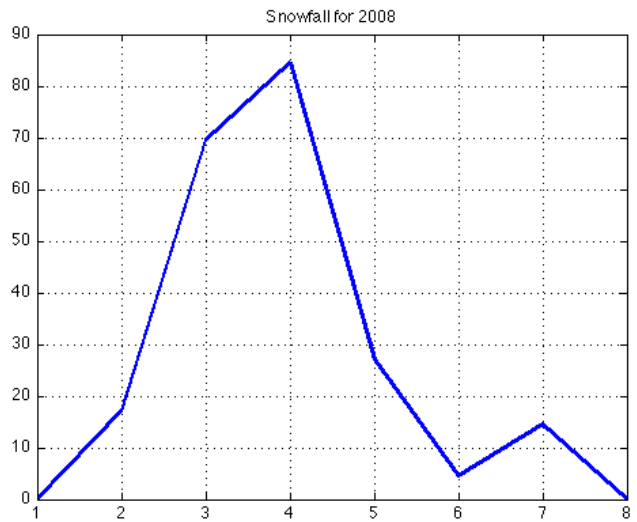




SVD: Typical Old and New Snowfall Patterns

To see how heaviest snowfall is coming earlier, compare the 1890 January/February style snowfall with the 2008 December/January style:





SVD: Conclusions

Data gathered at discrete places and times is easier to understand than the corresponding continuous cases.

The SVD shows how underlying patterns and correlations can be detected, and represented as a sum of the form

$$A = \sum_{i=1}^{r} \sqrt{\lambda_i}\, u_i\, v_i^T$$

where the λ values represent a strength, the u's represent variation in space, and v variation in time.

The structure of the u and v vectors suggests something about the preferred modes of the system, and the size of the λ coefficients allows us to understand the relative important of different modes, and to construct reduced order models if we wish.

SVD: Conclusions

Given that our data was stored in A, we may think of the matrices AA^T and A^TA as a form of a covariance matrix.

The singular values $\sqrt{\lambda_i}$ are the square roots of eigenvalues of both these matrices.

U contains eigenvectors of the "spatial" covariance matrix AA^T.

V contains eigenvectors of the "temporal" covariance matrix A^TA.

Very similar statements will hold for the continuous case.

SVD: Computational Implementations

```
F77 LAPACK:    call dgesvd ( jobu, jobvt, m, n, a, lda, &
               s, u, ldu, vt, ldvt, work, lwork, info )

C/C++ GSL:     gsl_linalg_SV_decomp ( A, V, S, work );

Java JAMA:     SingularValueDecomposition ( A );
               A.getU(); A.getS(); A.getV();

Mathematica:   { u, s, v } = SingularValueDecompostion ( a )

Matlab:        [ u, s, v ] = svd ( A );

Python numpy: u, s, vh = svd ( a, full_matrices=1, compute_uv=1)
```

F77 Lapack: http://www.netlib.org/lapack/
C/C++ GSL: http://www.gnu.org/software/gsl/
Java JAMA: http://math.nist.gov/javanumerics/jama/
Python numpy: http://numpy.scipy.org/

.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- **The Karhunen-Loeve Expansion**
- Orthogonal Polynomials
- Conclusion

KL:

Mercer's theorem applied to the case in which we had a stochastic process which, at every time t, took on a real value x. We now suppose the more general case in which at every time, the value of the process is a function, that is, a random field...

KL:

Suppose that $X(t,\omega)$ is a stochastic process with mean $\mu_X(t)$ and covariance $\text{cov}_X(t,s)$. Then there are numbers λ_i, functions $\psi_i(t)$, and random variables $y_i(\omega)$ such that

$$X(t,\omega) = \mu_X(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \psi_i(t)\, y_i(\omega)$$

Here,

- λ_i are (eigenvalues)?;
- $\psi_i(t)$ are (random field)?;
- $y_i(\omega)$ are (eigenfunction)?

KL:

The KL expansion:

$$X(t,\omega) = \mu_X(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, \psi_i(t)\, y_i(\omega)$$

breaks the process up into digestible pieces:

- *mean value* in $\mu_X(t)$;
- *weights* in λ_i;
- *time variation* in $\psi_i(t)$;
- *stochastic dependence* in $y_i(\omega)$.

KL:

Because the random variables $y_i(\omega)$ are uncorrelated, the variance of X is just the sum of the eigenvalues:

$$\text{var}(X) = \sum_{i=1}^{\infty} \lambda_i$$

KL: Mean + Covariance $\rightarrow$ Stochastic Process

One way to look at the KL representation is that all the information in a square-integrable stochastic process with continuous covariance is contained in the mean and the covariance.

This suggests that we can create a model stochastic process simply by choosing the mean and covariance; naturally, we will prefer a covariance function for which the eigenvalues and eigenfunctions can be easily determined, and which can be adjusted to suit our physical problem.

KL: Mean + Covariance $\rightarrow$ Stochastic Process

Now we will consider how, given a mean $\mu(t)$ and covariance $K(s,t)$, we can use the KL expansion to express a corresponding stochastic process.

The form of the expansion requires that we determine the Mercer eigenvalues λ_i and the corresponding eigenfunctions $\psi_i(t)$. To express a particular realization, we choose coefficients z_i which are uncorrelated and of unit variance.

Putting this together, we arrive at:

$$X_t = X(t,\omega) = \mu(t) + \sum_{i=0}^{\infty} \sqrt{\lambda_i}\psi_i(t)z_i(\omega)$$

KL: Computation

Suppose that our kernel function is the Brownian kernel and our interval is [0,1]:

$$K(s,t) = \min(s,t)$$

We need to determine the eigenvalues and eigenfunctions of the operator

$$(Tf)(t) = \int_0^{+1} K(s,t)f(s)ds$$

KL: Computation

For $i = 0, \ldots, \infty$, the eigenvalues are:

$$\lambda_i = \frac{1}{(i + \frac{1}{2})^2 \pi^2}$$

The eigenvectors are:

$$\psi_i(t) = \sqrt{2} \sin((i + \frac{1}{2})\pi t)$$

KL: Computation

Suppose that, over the interval [-1,+1], our kernel function $K(s,t)$ is the Gaussian isotropic correlation function

$$K(s,t) = e^{-(s-t)^2}$$

We need to determine the eigenvalues and eigenfunctions of the operator

$$(Tf)(t) = \int_{-1}^{+1} K(s,t)f(s)ds$$

KL: Computation

For $i = 0, \ldots, \infty$, the eigenvalues are:

$$\lambda_i = ?$$

The eigenvectors are:

$$\psi_i(t) = ?$$

KL: Computation

Suppose that, over the interval [-1,+1], our kernel function $K(s,t)$ is the exponential isotropic correlation function

$$K(s,t) = e^{-|s-t|}$$

We need to determine the eigenvalues and eigenfunctions of the operator

$$(Tf)(t) = \int_{-1}^{+1} K(s,t)f(s)ds$$

KL: Computation

Let v_j and w_k be the sequences of solutions, in ascending order, of

$$1 - v\tan(v) = 0$$
$$w + \tan(w) = 0$$

Then, for $i = 1, \ldots, \infty$, the eigenvalues are

$$\lambda_i = \begin{cases} \frac{2}{(1+v_j)^2} & \text{if } i = 2*j+1 \\ \frac{2}{(1+w_k)^2} & \text{if } i = 2*k \end{cases}$$

The eigenvectors are

$$\psi_i(t) = \begin{cases} \cos(v_j t)/\sqrt{1 + \frac{\sin(2v_j)}{2v_j}} & \text{if } i = 2*j+1 \\ \sin(w_k t)/\sqrt{1 - \frac{\sin(2w_k)}{2w_k}} & \text{if } i = 2*k \end{cases}$$

KL: Computation

```
mu =
psi =
```

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- **Orthogonal Polynomials**
- Conclusion

ORTHO:

Orthogonal polynomials play several roles in stochastic computations:

- quadrature rules for collocation methods;
- natural basis for approximation;
- a basis for the polynomial chaos expansion.

The Askey scheme organizes as many as 13 commonly used orthogonal polynomial families, including the well known cases of **Hermite**, **Laguerre** and **Legendre** polynomials.

The choice of the polynomial family depends in part on the domain and the weight function.

For multidimensional problems, product polynomials can be constructed in the usual way, and different polynomial families may be chosen for different dimensions.

ORTHO: Quadrature

To see how quadrature is needed, suppose we begin with the deterministic problem

$$-\nabla \cdot (a(\vec{x})\nabla u(\vec{x})) = f(\vec{x})$$

$\mathbf{a}(\vec{x})$ is the diffusivity, $\mathbf{f}(\vec{x})$ a source term. A finite element approach to the deterministic problem integrates the equation against various test functions $\mathbf{v_i}(\vec{x})$:

$$\int_D a(\vec{x})\,\nabla u(\vec{x}) \cdot \nabla v_i(\vec{x})\,d\vec{x} = \int_D f(\vec{x}) v_i(\vec{x})\,d\vec{x}$$

ORTHO: Quadrature

Now suppose that we assume a stochastic component to the diffusivity, with a Gaussian distribution $\rho(\omega) = e^{\frac{-\omega^2}{2}}$. Then to compute the finite element coefficients of the expected value, we need to solve:

$$\int_\Omega \int_D a(\vec{x};\omega)\,\nabla u(\vec{x};\omega)\cdot\nabla v_i(\vec{x})\,d\vec{x}\,\rho(\omega)\,d\omega = \int_\Omega \int_D f(\vec{x}) v_i(\vec{x};\omega)\,d\vec{x}\,\rho(\omega)\,d\omega$$

I WENT WRONG HERE I THINK.

ORTHO: Approximation, Inner Product and Norm

Orthogonal polynomials can be used to approximate a function.

We will suppose we are interested in functions which are square integrable with respect to the Hermite weight $e^{\frac{-x^2}{2}}$.

For two functions f and g, we can define the inner product

$$(f, g) = \int_{\mathbb{R}} f(x)\, g(x)\, e^{\frac{-x^2}{2}}\, dx$$

from which we can define the norm

$$||f||^2 = (f, f)$$

As long as $||f|| < \infty$, an orthogonal polynomial approximation exists.

ORTHO: The Orthonormal Family

For our weight and region, the appropriate family is the (probabilist's) Hermite polynomials $he(n, x)$. The first few elements are:

$$\begin{aligned}
\text{he}(0, x) &= 1 \\
\text{he}(1, x) &= x \\
\text{he}(2, x) &= x^2 - 1 \\
\text{he}(3, x) &= x^3 - 3x \\
\text{he}(4, x) &= x^4 - 6x^2 + 3 \\
\text{he}(5, x) &= x^5 - 10x^3 + 15x
\end{aligned}$$

The $\text{he}(n, x)$ are pairwise orthogonal, with norm $\sqrt{2\pi}n!$. Dividing each polynomial by its norm defines an orthonormal family $\psi_n(x)$ so that:

$$(\psi_i(x), \psi_j(x)) = \int_{\mathbb{R}} \psi_i(x)\, \psi_j(x)\, e^{\frac{-x^2}{2}}\, dx = \delta_{i,j}$$

ORTHO: Approximation

Let f be any function with $||f|| < \infty$, and construct the series of generalized Fourier coefficients:

$$c_i = (f, \psi_i(x)) = \int_{\mathbb{R}} f(x)\, \psi_i(x)\, e^{\frac{-x^2}{2}}\, dx$$

Then we have the following sequence of approximations to f:

$$f_n(x) = \sum_{i=0}^{n} c_i\, \psi_i(x)$$

Within the space of polynomials up to degree n, this approximation is the best possible. Moreover, it can be shown that

$$\lim_{n \to \infty} ||f - f_n|| = 0$$

so that the approximation sequence is convergent.

Covariance, Correlation, and the KL-Expansion

- Introduction
- Random Variables, Covariance and Correlation
- A Stochastic Process
- Brownian Motion
- Isotropic Correlation Functions
- SVD for the Discrete Case
- The Karhunen-Loeve Expansion
- Orthogonal Polynomials
- **Conclusion**

CONCLUSION: References

Volker **Barthelmann**, Erich **Novak**, Klaus **Ritter**,
High Dimensional Polynomial Interpolation on Sparse Grids,
Advances in Computational Mathematics, 12(4) 2000, p273-288.

John **Burkardt**, Max **Gunzburger**, Clayton **Webster**,
Reduced Order Modeling of Some Nonlinear Stochastic PDE's,
IJNAM, 4(3-4) 2007, p368-391.

Thomas **Gerstner**, Michael **Griebel**,
Numerical Integration Using Sparse Grids,
Numerical Algorithms, 18(3-4), 1998, p209-232.

Andreas **Klimke**, Barbara **Wohlmuth**,
Algorithm 847: SPINTERP: Piecewise Multilinear Hierarchical Sparse Grid Interpolation in MATLAB,
ACM Transactions on Mathematical Software, 31,(4), 2005, p561-579.

Petter Abrahamsen,
A Review of Gaussian Random Fields and Correlation Functions,
Norwegian Computing Center, 1997.

Sergey **Smolyak**,
Quadrature and Interpolation Formulas for Tensor Products of Certain Classes of Functions,
Doklady Akademii Nauk SSSR, 4, 1963, p240-243.

SVD - REJECTED SLIDE

Since correlations indicate that some information is redundant, or less important, we seek a way of looking at this data which emphasizes the information, or the strongest signal that is present.

Especially if the data contains random noise, we want to be able to accept the strong signals, and ignore the noise.

Suppose we form an m by n matrix X, using x^j for the j-th column. Each column is a vector, and so it has a direction and a norm. Norms are easy to understand, so let's divide them out, and concentrate on the direction information, replacing x^j by $\hat{x}^j = x^j / ||x^j||$. If we divide each column by its norm, that will make it easier to see which vectors have the same direction. In particular, the dot product of two vectors will now be a number between -1 and +1, representing the correlation between the two direction vectors. In fact, if $\hat{X}$ is the matrix formed from just the directions, we have

$$\hat{C} = \hat{X}' \hat{X}$$

SVD - REJECTED SLIDE

Since correlations indicate that some information is redundant, or less important, we seek a way of looking at this data which emphasizes the information, or the strongest signal that is present.

Especially if the data contains random noise, we want to be able to accept the strong signals, and ignore the noise.

BROWN:

Consider a small particle at position $\vec{x}(t)$, moving with a velocity $\vec{v} = \frac{d\vec{x}}{dt}$ through a liquid with a viscosity λ, leading to the mathematical model:

$$\frac{d^2\vec{x}}{dt^2} = -\lambda \frac{d\vec{x}}{dt}$$

which predicts that the velocity will exponentially decay to zero. Instead, we notice that the particle reaches a state where it seems to be constantly jostled by relatively small forces of random direction and magnitude, causing it to trace out a slow, irregular path through the liquid.

We hypothesize that the motion is described by the *Langevin equation*:

$$\frac{d^2\vec{x}}{dt^2} = -\lambda \frac{d\vec{x}}{dt} + \vec{\eta}(t)$$

where $\vec{\eta}(t)$ represents the small, incessant, stochastic force.

BROWN:

The quantity $\vec{\eta}(t)$ must do two things:

- be a good approximation to the physical observations;
- be mathematically, statistically, and computationally tractable.

We assume there is no preferred direction, so that

$$E(\vec{\eta}(t)) = \langle \vec{\eta}(t) \rangle = \vec{0}$$

and no correlation between values at different times:

$$\langle \vec{\eta}(t_1)\, \vec{\eta}(t_2) \rangle = \delta(t_2 - t_1)$$

It is common to simplify the equation, dropping the viscosity term so we can focus on the stochastic term:

$$\frac{d^2\vec{x}}{dt^2} = \vec{\eta}(t)$$

BROWN

A mathematical model of Brownian motion begins with notation: W_t, we mean the location of xxx

- zero expectation
- continuity
- independent increments

$W_t - W_0 \sim \mathcal{N}(0, 1)$