MAT 343E, Time Series Analysis, Spring 2007-2008

Instructor Ayşe Hümeyra Bilge

The aim of this course is to give basic mathematical methods for the analysis of time series. The outline of the course is given below. Homework assignments consists of the application of the methods to selected data and they constitute an essential component of the course.

1. Metric, norm, inner products;

2. Approximation: Taylor expansion, Fourier series

3. *Presentation of Homework 1: Data collection: Download data and convert it to MATLAB format, filling gaps and removing outliers if necessary. Prepare a program that would repeat the procedure on demand.*

4. Modeling, the method of "least squares". Trend analysis, approximation by polynomials, error calculations.

5. Correlation analysis, Modeling and prediction of deterministic variations

6. *Presentation of Homework 2: Trend analysis, linear regression models and prediction for the long term trends.*

7. Spectral analysis, Sampling theorem, Fast Fourier Transform

8. Data windowing, filtering, modulation.

9. *Presentation of Homework 3: Spectral analysis of the data, modeling of periodic variations, modulation.*

10. Random variables, distributions,

11. Stochastic processes, stationary and non-stationary processes

12. *Presentation of Homework 4. Analysis of the deviations from the model. Distribution of the errors, histograms, percentiles, reliability bounds.*

13. Data driven methods: Forecast

14. Examples: Feedback, Neural networks

LECTURES 1 and 2.

1. General Properties of Time Series

- A "time series" is a set of observations. Observed data can be *continuous* or *discrete*. In the case of continuous data, one usually samples data at appropriate intervals.

- Changes in the data are broadly classified as

  - long term variations: trends with no obvious source,
  - intermediate term variations: deterministic variations that are mostly periodic or tied to other phenomena,
  - short term variations: residuals that are mostly stochastic.

- Trends in the data are slow changes that have no obvious interpretation or model. Caution: A periodic variation with a period of 100 years will appear as a trend in a 10-year observation.

- Intermediate term variations are mostly deterministic, i.e, governed by certain "laws". In many cases they are tied to physical (natural)periodic variations such as seasonal and diurnal changes. Then they are modeled in terms of trigonometric functions. In some cases, they look irregular, but they are functionally related to another physical phenomena. In this former case the model is just $y(t)$ while in the latter case it is of the form $y(t) = f(x(t))$.

- After subtracting trends and intermediate term variations from the data, whatever remains is called residual variations. In a good model, their norm should be a small percentage of the norm of the original data. Residuals can be pure noise or irregular but deterministic variations.

Referring to the graphs in textbook, on can see that

- Figure 1.5 has no intermediate and short term variations. It looks like a trend that can be modeled by polynomials or exponentials.

- Figure 1.6 looks like purely residuals with no obvious model.

- Figure 1.4 has no data trends.
- Figure 1.1 is a typical example displaying all three data components above.

2. Metric Spaces

We have seen that time series had long term, intermediate term and short term variations. In general, long and intermediate term variations are *deterministic* while the short term variations are *stochastic*. We shall start with the study of deterministic variations.

The appropriate model for deterministic variations will be obtained by viewing the time series as a point (or a vector) p in some function space and the model as a point (or vector) q in an appropriate linear subspace, in such a way that the "distance" between p and q is minimal.

This setup requires the definitions of metrics, norms, and related concepts.

We start by defining the notion of *metric* which describes the distance between elements of a *metric space*.

Definition. A metric space is a set X together with a function

$$d : X \times X \to R$$

with the following properties.

1. For all x and y, $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = 0$.
2. For all x and y $d(x, y) = d(y, x)$.
3. For all x, y and z, $d(x, z) \leq d(x, y) + d(y, z)$.

The last property is called the *triangle inequality*. The open ball centered at a point p and with radius ϵ is defined by

$$B_p(\epsilon) = \{x \in X | d(p, x) < \epsilon\}.$$

These open balls define a "topology" on X. Although there are topological spaces that do not admit any metric, most of the familiar topological spaces are metrizable.

3. Vector Spaces

Note that a metric is defined on a set, without reference to a vector space structure. On linear spaces, the *inner products* and *norms* lead to metrics. Vector spaces or linear spaces are the sets of objects called "vectors" that can be added to each other and multiplied by "scalars". Those scalars form a "field", that should be thought of a an abstraction of real numbers. For completeness we give the definition of a field.

Definition. Let F be a set on which the operations of addition and multiplication, denoted by $+$ and $\cdot$ respectively be defined. If the conditions below are satisfied, the $(F, +, \cdot)$ is called a *field*.

1. For all x, y in F, the addition is commutative, i.e., $x + y = y + x$.
2. For all x, y, z in F, the addition is associative, i.e., $x + (y + z) = (x + y) + z$.
3. There is a unique element 0, called the *zero element* such that for all x in F, the equality $x + 0 = x$ holds.
4. For each x in F, there is a unique element $-x$, denoted as the *additive inverse of* x, such that the equality $x + (-x) = 0$ is satisfied.
5. The multiplication operation is commutative, that is for each x, y in F, $xy = yx$.
6. The multiplication is associative, that is for every x, y, z, in F, $x(yz) = (xy)z$.
7. There is a unique element 1 called the multiplicative identity of F, such that For each nonzero element x in F, $x \cdot 1 = x$.
8. For each nonzero x in F, there is a unique element denoted by x^{-1} and called the multiplicative inverse of x, such that $x \cdot x^{-1} = 1$.
9. The multiplication is distributive over addition, that is for x, y, z in F, $x(y + z) = xy + xz$.

It is easy to see that real numbers, rational numbers, complex numbers satisfy these properties hence they are fields.

Vector spaces are sets of objects on which an "addition operation" and a "multiplication by scalars" is defined.

Definition. A vector space V over a field F consists of the following

i. A set V, called the set of "vectors", together with a binary operation

$$+ : V \times V \to V,$$

ii. A field F, and a scalar product operation $\bullet$,

$$\bullet : V \times V \to F,$$

with the following properties.

1. The addition operation is commutative, associative, there is an additive identity and, each element has an additive inverse. The zero element of the vector space is denoted by $\mathbf{0}$ and the additive inverse of an element x is denoted by $(-x)$. The properties above are summarized as

$$x + y = y + x, \quad x + (y + z) = (x + y) + z,$$

$$x + \mathbf{0} = \mathbf{0} + x = \mathbf{0}, \quad x + (-x) = (-x) + x = \mathbf{0}.$$

2. Let the elements of the field F be denoted by and its multiplicative identity be denoted by 1. The multiplication by scalars satisfy the following properties.

   i. $1x = x$,

   ii. $(ab)x = a(bx)$,

   iii. $a(x + y) = ax + ay$,

   iv. $(a + b)x = ax + bx$.

We give below typical examples of vector spaces.

Examples.

- Let F be any field. Then F^n is a vector space. R^n and C^n are the most common examples.
- Matrices over a field form a vector space.
- Functions from any *set* to a field F form a vector space. Here the addition of two functions is defined point-wise as the addition of their values, i.e.,

  $$(f + g)(x) = f(x) + g(x).$$

  It is easy to check that the conditions listed above are all satisfied.

- A subset of the function spaces defined above is the set of all polynomials.
- Polynomials of degree less than or equal to n also form a vector space.

Subspaces. Let V be a vector space over a field F and W be a subset of V. If the set W is itself a vector space with respect to the operations of vector addition and the multiplication by scalars in V, then it is called a vector subspace (linear subspace) of V.

It can be seen that the conditions such as associativity and commutativity etc. are always inherited from V and on only needs to check that W is closed under the vector space operations of V. More precisely we have the following theorem.

Theorem. *Let W be a nonempty subset of V. Then W is a subspace of V if and only if for each x, y in W and each a in F, $ax + y$ is in W.*

It can be seen that the only subspaces of R^2 are the lines through the origin. Similarly, linear subspaces of R^3 are the planes through the origin. Similarly, polynomials are linear subspaces of function spaces.

Now we give the definitions of "linear combination", and subspaces "spanned by a set of vectors".

Linear combination. Let $(x_1, \ldots, x_n)$ be a set of vectors and $c_1, \ldots, c_n$ be scalars. If a vector y can be written as a sum

$$y = c_1x_1 + c_2x_2 + \ldots + c_nx_n,$$

then y is said to be a linear combination of the vectors $x_1, \ldots, x_n$.

Subspace spanned by a set. Let $S = (x_1, \ldots, x_n)$ be a set of vectors. It can be shown that the set of all linear combinations of S is a linear subspace, called the *linear span of S*.

Bases and dimensions.

We start by defining linear independence, then we define a basis for a vector space V as a linearly independent set whose span is V. Finally we define the dimension of a vector space as the number of elements in a basis.

Definition. Let V be a vector space over s field F and S be a set of distinct vectors $S = (x_1, x_2, \ldots, x_n)$. If the vector equation

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n = 0$$

has a solution for scalars $c_1, c_2, \ldots, c_n$ not all zero, then the set $(x_1, x_2, \ldots, x_n)$ is called *linearly dependent*. Otherwise, if the equation above implies that all the c_i's are zero, then the set is called linearly independent.

Checking linear dependence or linear independence in F^n is straightforward, but it can be tricky in function spaces. Here are some problems.

Problem.

1. Let $f_1 = (1 + x)$, $f_2 = (1 - x)$, $f_3 = (1 - x^2)$, $f_4 = (1 + x^2)$. Is this set linearly independent?

2. $f_1 = \sin(x)$, $f_2 = \cos(x)$, $f_3 = \sin(2x)$, $\cos(2x)$. Is this set linearly independent?

3. $f_1 = \sin^2(x)$, $f_2 = \cos^2(x)$, $f_3 = \sin(2x)$, $\cos(2x)$. Is this set linearly independent?

In you think that a set is linearly independent, write down a linear combination of the given functions. If they are equal to the zero function, they are zero at each point. Thus there are n functions int he set, evaluate them at n points to get a set of n algebraic equations. If these imply that all coefficients are zero, then you have proved that the given set is linear independent. But if these equations do not imply that all are zero, this may either be due to our bad choice of points, or the set is linearly dependent. For example try to evaluate the last two examples at multiples of 2π. Never conclude linear dependence by checking a few points. For proving linear dependence, you should use identities to prove that the linear combination vanishes without all coefficients being zero.

We can now define the basis of a vector space.

Definition. A basis for a vector space is a linearly independent set of vectors that span the given vector space.

Remarks.

- If V has a basis consisting of a finite number of vectors, then it is called a *finite dimensional vector space*. The basis of a vector space is not unique, but every basis has the same number of elements. This common number is called the dimension of V.

- If V is not finite dimensional, then V is called an *infinite dimensional vector space*. Every vector space, finite or infinite dimensional has a basis. For an infinite dimensional vector space, a set is a basis, if every

vector an be written as a finite linear combination of the basis elements. Bases of infinite dimensional vector spaces are not very useful. Instead one works with "complete sets", that allow every vector to be expanded as an infinite series in terms of a sequence of functions.

Coordinates. If $\mathcal{B} = (e_1, \ldots, e_n)$ is an ordered basis and a vector v is written (uniquely) as

$$v = a_1 e_1 + \ldots + a_n e_n,$$

then the scalars a_i are called the components of v with respect to the basis $\mathcal{B}$.

4. Normed Spaces

The notion of "norm" is an abstraction for the length of a vector in R^n.

Definition. A "norm" on a vector space V is a map

$$\| \ \| : V \to R,$$

with the follow

1. For all x in V, $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$.
2. For all $x$in V and all a in F, $\|ax\| = |a| \ \|x\|$, where $|a|$ is the absolute value of a.
3. For all x, y, z in V, $\|x + y\| \leq \|x\| + \|y\|$.

On a normed vector space we can define a distance function by

$$d(v, w) = \|v - w\|.$$

One can show that this gives in fact a metric.

The standard norm on R^n is

$$\|x\| = \sqrt{x_1^2 + \ldots + x_n^2}$$

and this gives the metric

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}.$$

5. Inner Product Spaces

The well-known dot product on R^2 or R^3 is the prototype of the concept of inner product on vector spaces. For simplicity we assume that the field F is the real numbers R, because the definitions are slightly different for the complex case. An inner product on a vector space V over R is a map

$$(\ , \ ) : V \times V \to R$$

satisfying the following properties

1. For all x, y in V, $(x, y) = (y, x)$
2. For all x, y, z in V and a in R, $(ax + y, z) = a(x, z) + (y, z)$.
3. $\|(x, y)| \leq (x, x)^{1/2}(y, y)^{1/2}$.

The last property is known as the Schwarz inequality. A vector space with an inner product is called an inner product space.

Given an inner product space V, we can define a norm on V by

$$\|x\| = (x, x)^{1/2}.$$

The standard inner product on R^n is

$$(x, y) = sqrtx_1y_1 + \ldots + x_ny_n,$$

which can also be expressed in terms of matrices as

$$(x, y) = x^t y = y^t x.$$

Another common inner product space, is the vector space V of square integrable functions on an interval (a, b). If f and g are in V, then

$$(f, g) = \int_a^b f(t)g(t)\ dt.$$

If x, y are elements of an inner product space we say that they are orthogonal if $(x, y) = 0$. A set $\{e_1, \ldots, e_n\}$ is called orthogonal if (e_i, e_j) is nonzero only when $i = j$. It is called orthonormal if

$$(e_i, e_j) = \delta_{ij}.$$

Given a vector v and a subspace W of V, we say that v is orthogonal to W if v is orthogonal to every vector in W.

Given a subspace W of V and a vector x not in W, the orthogonal projection of x onto W is a vector y in W, such that the vector $v = x - y$ is orthogonal to W.

Although it is common to use standard bases in many vector space computations, in many instances the choice of special bases allows the simplification of certain proofs. In these lines, there is an important procedure for the construction of orthonormal sets of vectors, called the "Gram-Schimdt Orthogonalization Procedure". In this procedure, one starts with any given set of vectors and obtains an othonormal set. We skip the details of this procedure, but we shall use this result, by saying for example that "we take an orhonormal basis for a given subspace".

6. Approximations and Orthogonal Projections

A continuous time series recorded over a finite time interval is a function $f(t)$ on some interval $[a, b]$, hence it is an element of a function space. In the analysis of the long and intermediate term variations, we want to fit a model to the data, that is we want to approximate the recorded function by a small number of simple functions. These simple functions which constitute our "model functions" form a finite dimensional subspace and the model for the data is a linear combination of these functions or a vector in this subspace.

If we work with sampled data over a finite time interval, the data is now an vector in a finite dimensional space R^n for large n. The model will be again an element in a finite dimensional subspace. In both cases we want to approximate an arbitrary vector by an element of some finite dimensional subspace, by minimizing the norm of the difference. The explicit expression of this norm will be different in each case.

We will prove that for any inner product the norm of the error will be minimal, precisely when the difference between the data and the model will be orthogonal to the subspace spanned by the modal functions.

Proposition. *Let V be an inner product space, W be a finite subspace of V and y be an element of V that is not in W. Then the best approximation to y in W is the orthogonal projection of y on W.*

Proof. Let $\{e_1, \ldots, e_N\}$ be an orthonormal basis for W. Let x be the vector

in W that is the candidate for the approximate model. With respect to the basis above, this vector can be written as

$$x = a_1 e_1 + a_2 e_2 + \ldots + a_N e_N,$$

and the aim is to minimize the norm , or the norm square of the difference

$$z = y - x = y - \sum_{i=1}^{N} a_i e_i.$$

The norm of the difference depends on the parameters a_i, $i = 1, \ldots, n$. Denoting this difference by E, we have

$$\begin{aligned} E(a_1, \ldots, a_N) &= \left(y - \sum_{i=1}^{N} a_i, e_i, y - \sum_{i=1}^{N} a_i, e_i \right) \\ &= (y, y) - 2 \sum_{i=1}^{N} a_i (y, e_i) + \sum_{i,j=1}^{N} a_i a_j (e_i, e_j) \\ &= (y, y) - 2 \sum_{i=1}^{N} a_i (y, e_i) + \sum_{i=1}^{N} a_i^2. \end{aligned}$$

We recall that the last inequality is obtained from the orthonormality of the e_i's. To minimize E, we also recall that the minimum of a quadratic function occurs at points where the gradient vector is zero and the Hessian matrix (the matrix of second partial derivatives) is positive definite. The gradient of E is $\partial E / \partial a_i$ and equating these to zero we obtain

$$a_i = (y, e_i), \quad i = 1, \ldots, N.$$

The Hessian matrix for E is

$$H = \begin{pmatrix} \frac{\partial^2 E}{\partial a_1^2} & \frac{\partial^2 E}{\partial a_1 \partial a_2} & \cdots & \frac{\partial^2 E}{\partial a_1 \partial a_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial a_1 \partial a_n} & \frac{\partial^2 E}{\partial a_2 \partial a_n} & \cdots & \frac{\partial^2 E}{\partial a_n^2} \end{pmatrix}$$

just the identity matrix,

hence the conditions above give the minimum of E. With this choice of the model x, the difference $z = y - x$ is orthogonal to W since $(z, e_i) = 0$ for each i. □

As an application, we can prove that the Fourier coefficients give the best approximation to a square integrable function in terms of trigonometric functions. Here V is the vector space functions that are square integrable on the interval $[0, 2\pi]$, W is the subspace spanned by $\sin(k\pi t)$ and $\cos(k\pi t)$, for $k = 0, \ldots, N$. The inner product is given by

$$\begin{aligned}(y(t), \sin(k\pi t)) &= \int_0^{2\pi} y(t)\ \sin(k\pi t)\ dt, \\ (y(t), \cos(k\pi t)) &= \int_0^{2\pi} y(t)\ \cos(k\pi t)\ dt,\end{aligned}$$

The expression above seems very easy, one obtains the coefficients simply by taking the inner product of the observed data with the model functions e_i's. The tricky point is that the e_i's are assumed to be an orthonormal set, thus if we start with an arbitrary set of functions, we should put first obtain from them an orthonormal set. Thus, the formula above is useful only when we have in hand an orthonormal set of model functions, such as the trigonometric functions in the example above.

In the general case, the matrix formulation discussed below is more useful.

7. Matrix Formulation of the Approximation Problem

In this section we re-derive the result above in terms of matrices. This derivation will lead to a practical formula and an algorithm for model construction.It is important that you follow the computations in this section, in order to convert indexed expression to matrix equations.

Let y be an n-element time series, i.e., $y \in R^n$, and f_i be model functions,given as n-vectors, for $i = 1, \ldots, N$. Usually, n is much larger than N, that is we work with a small number of model functions. Theoretically, it is in general possible to take $N = n$, but you will see that in practical applications, certain matrices become nearly singular as N gets large. The rule would be to increase N gradually and stop at the point where you start getting unwanted results.

Let's agree on denoting the labels of the model functions with the letters i, j etc. and their vector indices by the Greek letters α, β, etc. Thus, y^α is the α'th element of the time series y and f_i^α is the α'th element of the time series f_i. Writing vector indices as superscripts is a convention in tensor

analysis. With this convention, y^α is an $n \times 1$ column vector, f_i's are $n \times 1$ column vectors for each i. Putting these together as the columns of a matrix, we obtain an $n \times N$ matrix F whose i'th column is the time series for the model function f_i. The element in the α'th row, i'th column of the matrix F is

$$(F)_{\alpha i} = f_i^\alpha.$$

An arbitrary vector in the subspace spanned by the f_i's is written as

$$x = \sum_{i=1}^{N} a_i \ f_i.$$

Note that x is an n vector, that we write componentwise as

$$x^\alpha = \sum_{i=1}^{N} a_i \ f_i^\alpha.$$

Now, you should convince yourself that this is the matrix equation

$$x = Fa$$

where a is the column vector of size N consisting of the coefficients a_i's. Then, the error is given by

$$E = \|y - x\| = \|y - Fa\| = (y - Fa, y - Fa) = (y - Fa)^t (y - Fa)$$

where the superscript t denotes the transpose. Thus

$$E = y^t y - 2y^t Fa + a^t F^t Fa.$$

In component form this is equivalent to

$$E = \sum_{\alpha=1}^{n} (y^\alpha)^2 - 2 \sum_{\alpha=1}^{n} \sum_{i=1}^{N} y^\alpha \ f_i^\alpha \ a_i + \sum_{\alpha=1}^{n} \sum_{i=1}^{N} \sum_{j=1}^{N} (f_i^\alpha \ a_i)(f_j^\alpha \ a_j).$$

Then $\partial E / \partial a_k = 0$ gives

$$\begin{aligned} \frac{\partial E}{\partial a_k} &= -2 \sum_{\alpha=1} N y^\alpha f_k^\alpha + \sum_{\alpha=1}^{n} \sum_{j=1}^{N} f_k^\alpha f_j^\alpha a_j + \sum_{\alpha=1}^{n} \sum_{i=1}^{N} f_i^\alpha a_i f_k^\alpha \\ &= -2 \sum_{\alpha=1} N y^\alpha f_k^\alpha + 2 \sum_{\alpha=1}^{n} \sum_{i=1}^{N} f_k^\alpha f_i^\alpha a_i \\ &= -2 \left(F^t y \right)_k + 2 \left(F^t F a \right)_k \end{aligned}$$

Setting this equal to zero we obtain the matrix equation

$$F^t F a = F^t y.$$

In order to solve for a, we should invert the square $N \times N$ matrix $F^t F$. Thus

$$a = (F^t F)^{-1}\ F^t\ y.$$

As discussed at the end of the previous section, if we take too many model functions, this matrix will be close to a singular matrix. The formula above gives the coefficients for the model. We can summarize this as an algorithm.

Algorithm. Finding the best approximation to y in terms of the model functions f_i's, $i = 1, \ldots, N$.

- Express the time series as a column vector, $y \in R^n$.
- Built the model functions as column vectors of the same size $f_i \in R^n$, $i = 1, \ldots, N$.
- Place the model functions of an $n \times N$ matrix, $F = [f_1, \ldots, f_N]$.
- Solve the matrix equation $F^t F a = F^t y$ as

  $$a = (F^t F)^{-1} F^t y.$$

- If $F^t F$ is singular or nearly singular, modify the model functions and repeat.
- The model is $x = Fa$. You may skip the previous step and write directly

  $$x = F(F^t F)^{-1} F^t y.$$

  You may do this if you are not interested in seing the parameters a_i's.
- Plot y and x to see whether the model is satisfactory.
- For a quantitative measure of the goodness of the model, compute the absolute error,

  $$E = \left[(y - x)^t (y - x)\right]^{1/2}.$$

  But the relative error

  $$E_r = \frac{\left[(y - x)^t (y - x)\right]^{1/2}}{\left[y^t y - x\right]^{1/2}}$$

and the percentage error

$$E_p = \frac{\left[(y-x)^t(y-x)\right]^{1/2}}{\left[y^t y - x\right]^{1/2}} \times 100$$

are more useful

8. Practical Problems

- **Data collection:** In practice, finding and converting data to a usable format is the most difficult part. You should be able to download files and convert them to text format. Then rearrange them in an array or vector.
- Data may have non-numeric entries. You should interpret and remove these non-numeric values. It is preferable to do this by an automated procedure.
- There may be missing values in the regular data. You should have a strategy for filling missing values. The simplest way is to assign the next or previous non-trivial data or better, their average.
- There may be "bad" data, either because of a problem in the data collection or in the instrumentation. It may also happen that whatever you call bad data is due to an unusual event. At any rate, there may be data points that you want disregard. These are called outliers in the data. You should also have a criterion for removing outliers.
- Always plot the data to decide on your modeling strategy.
- At this stage you are asked to remove the trends (i) by using standard MATLAB functions, (ii) by repeating this with your own functions.

9. Deterministic Models

In previous sections we have seen how to represent data using a mathematical model. In many cases, we analyze a process generated by or closely related to some other processes. For example, the variations in the critical frequency of the ionosphere are due to the variations in Sun's electromagnetic field. There are various measures of this electromagnetic activity; the smoothed sunspot numbers is one of these. A qualitative overview shows that the critical frequency have nearly 11 year periods. We may choose to model it by harmonics of 11 year sinusoids, or we may use a linear model in terms of the sunspot numbers. In another application if we observe that the mail traffic have a seasonal variation we may decide either to use harmonics of sinusoidal variations of 6 month periods or use a model where the amount of mail traffic is linearly related to the number of students enrolled.

In all cases, we have to decide whether there is really a correlation between these events. The proper way to answer this question is to use statistical techniques, that will be discussed later in the course. At this stage we use two qualitative tools that will guide us in our decisions.

The first tool is the scatter plot of the data: from the appearance of the graph we decide whether there is a deterministic relationship and if so, decide on the model. To make the scatter plot, first organize both variables as time series over a common interval and with identical sampling rates. For example, it may happen that the data for the critical frequency is given as hourly values over a 40 year period while sunspot numbers are given as monthly values over 140 years. One has first arrange both data as time series over a common interval, taking care in converting data matrices to data vectors. Then one has to obtain data with identical sampling rates. This is done either by using "spline interpolation" on coarsely sampled data, or "smoothing" the finely sampled data. For the purposes of the present coarse, it is sufficient to smooth data by taking averages.

After having arranged both data as time series over a common interval and with identical sampling rates, say in vectors X and Y, simply plot Y versus X with the $plot(X, Y), '*'$ command. This will give a scatter plot. If the data points are agglomerated along a line of some other curve, you may decide on using the variable X as a model for the variable Y. We note that at present this is only a qualitative observation.

In practice one should prefer to use modeling by lower order functions because they are more stable for prediction purposes. For example a linear model may not be as good as a third order model over a time period $[t_1, t_2]$,

but in general gives better results for predicting the function over $[t_2, t_3]$.

In practice, whether you use a mathematical or physical model for your data, the computational procedure is the same, you put the column vectors for the model functions in the matrix F as described above and build your model for the data as before.

10. The Fourier Series

Let f be a real valued function over an interval I and $\|\cdot\|$ be a norm on the functions defined on I. f is called "square integrable" on I, denoted as $f \in L^2(I)$, if

$$\int_I \|f(t)\| \, dt$$

is defined.

We have seen that if we have a sequence of functions, ϕ_n, $n = 0, 1, \ldots$ orthonormal on interval I, we can define a series

$$\sum_{n=0}^{N} c_n \phi_n$$

where

$$c_n = \int_I f(t)\phi_n(t) \, dt$$

and the series above was the best approximation to $f(t)$ in the finite dimensional space spanned by the ϕ_n's.

We now let N go to infinity, and obtain the infinite sum, or the series

$$\sum_{n=0}^{\infty} c_n \phi_n \qquad (*)$$

where the c_n's are defined as above. The series $(*)$ is called the Fourier series generated by f. The following theorem states the relation of the series to f.

Theorem. *Let $\{\phi_0, \phi_1, \ldots\}$ be orthonormal on I, assume that $f \in L^2(I)$ and let the Fourier series generated by f be*

$$f \sim \sum_{n=0}^{\infty} c_n \phi_n.$$

Then

i. *The series* $\sum |c_n|^2$ *converges and satisfies the inequality*

$$\sum_{n=0}^{\infty} |c_n|^2 \leq \|f\|^2 \qquad \text{Bessel}'\text{s inequality.}$$

ii. *The equation*

$$\sum_{n=0}^{\infty} |c_n|^2 = \|f\|^2 \qquad \text{Parseval}'\text{s formula}$$

holds if and only if we also have

$$lim_{N\to\infty}\|f - S_N\| = 0, (**)$$

where the S_N*'s are the partial sums defined by*

$$S_N(t) = \sum_{k=0}^{N} c_k \phi_k(t).$$

For the proof of this theorem, we refer to (T. Apostol, MAthematical Analysis, Addison-Wesley, 2.Ed. 1974). We can see that the Fourier series for f, which is itself a function, represents f in the sense that the limit in $(**)$ is zero. This is in a sense a weak condition, since we are taking the limit of a sequences of numbers.

A converse to part (i) of the theorem above is the Riesz-Fisher theorem, which states the series (*) defines a function whose Fourier series representation is the series we started with.

The theorems above do not say anything on whether the series is convergent at some t, and even if it converges, whether the limit is equal or not to $f(t)$.

Now we assume that f is continuous over $[0, 2\pi]$, have period 2π, and let ϕ_n's be sinusoidal functions. Then if

$$f(t) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nt) + b_n \sin(nt),$$

the the sequence of partial sums S_N's converge in the mean to f, that is the sequence of numbers

$$\int_{|} f(t) - S_N(t)|^2 \, dt$$

converges to zero. The Parsevals's formula is reduced to

$$\frac{1}{\pi}\int_0^{2\pi} \|f(t)\|^2 \, dt = \frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2),$$

and if the Fourier series for f converges for some t, then its limit is equal to $f(t)$.

Using the formulas

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin(x) = \frac{e^{ix} - e^{-ix}}{2i},$$

we can write the Fourier series for f as

$$f \sim \sum_{n=-\infty}^{\infty} c_n e^{int}$$

where c_n's are complex numbers.

If f is periodic on some interval $[a, b]$ the definitions can be modified by scalings. If f is not continuous, then there are various sufficient conditions on the convergence and the limit of the Fourier series for f, that will lead to the fact that at a point of jump discontinuity, the limit will be the mean of the right and left hand side limits of f.

If f is defined and continuous on some finite interval, but it is not periodic, we can work with the periodic extension of f and use the results above. In order to ensure periodic extension, we first redefine $f(b)$ as $f(b) = f(a)$. This redefinition may bring in a jump discontinuity in the extended function even if f is continuous, but there is no serious problem in dealing with the Fourier series.

The main problem is the case where f is already defined for all t and is not periodic. For this case we should define the Fourier integral.

11. The Fourier Transform

The representation of functions by integrals is a commonly used technique in analysis. We may define

$$g(y) = \int_{-\infty}^{\infty} K(x, y) f(x) \, dx.$$

The function $K(x, y)$ is called the kernel of the transformation. You have probably seen the Laplace transform

$$F(s) = \int_0^\infty e^{-st} f(t)\ dt.$$

Here we shall assume that $f(t)$ belongs to the function spaces $L^1(R)$ and $L^2(R)$, that is it integrable and square integrable on R.

The Fourier transform of a function $f(t)$ is defined by

$$F(\omega) = \int_\infty^\infty e^{-i\omega t} f(t)\ dt.$$

The Fourier integral theorem describes in which sense and under which conditions the function $F(\omega)$ represents $f(t)$. We only state that at points of continuity of f, we have the inversion formula

$$f(t) = \lim_{\alpha \to \infty} \frac{1}{2\pi} \int_{-\alpha}^{\alpha} F(\omega) e^{i\omega t}\ d\omega.$$

the functions $f(t)$ and $F(\omega)$ are called Fourier transform pairs, t is called the "time domain" variable and "ω" is called the frequency domain variable.

An important tool in the analysis of integral transforms is the so-called convolution integral defined by

$$h(x) = f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) g(x - \tau)\ d\tau = \int_{-\infty}^{\infty} f(x - \tau) g(\tau)\ d\tau.$$

We can think of the convolution integral as keeping $f(\tau)$ fixed, shifting $g(-\tau)$ by x, multiplying and integrating them, to get the value of the function $h(x) = f(x) * g(x)$ at the point x. The importance of this integral comes from the fact that multiplication in the time domain corresponds to convolution in the frequency domain and vice-versa. That is

$$\mathcal{F}\left(f(t)g(t)\right) = F(\omega) * G(\omega),$$

and

$$\mathcal{F}^{-1}\left(F(\omega)G(\omega)\right) = f(t) * g(t).$$

As the Fourier transform and its inverse are defined by integrals, they are linear operators.

We now study various problems associated with finding periodicities in data observed for a finite time and sampled at equally spaces intervals.

12. Spectral Analysis

Spectral analysis of data is the determination of periodic variations in the data. The "Fourier Series", the "Fourier Transform"', the "Discrete Fourier Transform" and the "Fast Fourier Transform" are the main ingredients of spectral analysis.

We have already seen the approximation of a given function in terms of sinusoidal functions with angular frequencies $k\omega$, for $k = 0, 1, \ldots, n$. This is called the "Fourier Polynomial for $f(t)$. If $f(t)$ is sinusoidal, i.e., $f(t) = \cos(\omega t)$ or $f(t) = \sin(\omega t)$, its Fourier polynomial consists of a single term.

We have then seen that if $f(t)$ is periodic or it is defined over a finite interval and extended periodically, it has a Fourier series expansion

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n t}$$

where we did not put the equality sign, since the series may not converge pointwise. In the expression above ω_n is the angular frequency of the n'th mode given by

$$\omega_n = n\frac{2\pi}{T},$$

and the Fourier coefficients are defined by

$$c_n = \frac{1}{T}\int_{t_0}^{t_0+T} f(t)\ e^{-i\omega_n t}\ dt.$$

If this series converges for each t, it defines an analytic function. This function may not be equal to $f(t)$ at certain points. Under suitable conditions, the series converges to $f(t)$ at points where $f(t)$ is continuous, and converges to $(f(t^+) - f(t^-))/2$ at a jump discontinuity.

For aperiodic $f(t)$ we have the Fourier transform pairs, defined by

$$F(\omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt, \qquad f(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t}d\omega.$$

In practice we work with data observed for a finite time interval and sampled at an appropriate rate. To deal with such data, one needs to define the "Discrete Fourier Transform" (DFT), briefly outlined below. In practice, we may use directly the Fast Fourier Transform algorithms that are available in signal processing toolboxes.

If $f(n)$, $n = 1, \ldots$ is a finite or infinite sequence, its z-transform is defines by

$$F(z) = \sum_{n=-\infty}^{\infty} f(n) z^{-n}$$

where z is a complex number. Writing $z = e^{i\omega}$ we have

$$F(re^{i\omega}) = \sum_{n=-\infty}^{\infty} f(n) r^n e^{-i\omega n}.$$

Thus the z-transform is equal to the Fourier transform when $r = 1$.

The conditions for the convergence of the z-transform is that the series above the absolute convergence od the series above, given by

$$\sum_{-\infty}^{\infty} |f(n) r^{-n}| < \infty.$$

Note that for $r > 1$, the z-transform may converge for the cases where the Fourier transform does not converge. This is the case for unit step function for example. In general the series defining the z-transform is convergent in an annular region in the complex plane defined by $r_0 < r < r_1$ where $r_0 \geq 0$ and $r_1 < \infty$. Such a power series in terms of z is called a Laurent series, because negative powers are involved. In the region of convergence, $F(z)$ is an analytic function, in particular all its derivatives are defined and continuous.

For cases where the observed sequence is of finite duration, it is possible to define what is called the "Discrete Fourier Transform" which is itself a finite sequence, rather than a continuous function.

An efficient algorithm for the computation of the Discrete Fourier Transform is the "Fast Fourier Transform" (FFT) which is nowadays a standard library function for example in MATLAB. For our purposes it will be sufficient to know basic properties.

13. Outline

- We have seen how to compute the Fourier transform pairs for a pulse of finite duration, and for a Gaussian pulse.
- The Fourier transform of sinusoidal functions are not defined in terms of regular functions. We have seen how the Dirac δ function is defined

as a distribution, and defined the Fourier transform of the sinusoidal functions in terms of the δ functions.

- We have seen how to compute convolution integrals, in particular for functions of compact support.
- The effect of finite observation time is the multiplication of the function by a rectangular window. This leads to the convolution in the frequency domain. As an example you should compute the effect of finite observation time on a pure sinusoidal function and the sum of two sinusoidal functions.
- In order to diminish the problems associated with finite observation time one can use different "windows". These are localized functions whose Fourier transforms have suppressed side lobes.
- Another application of the convolution integral is the amplitude modulation, that corresponds to the multiplication of a slowly varying function with a rapidly varying function. This brings the spectra of the slowly varying signal to the frequency of the rapidly varying signal. You should be able to describe this situation either as a convolution.
- The effect of sampling is described by the sampling theorem, which states that a signal that is band limited to ω_c can be recovered by its samples taken at a rate larger than $2\omega_c$.
- Frequencies above ω_c should be filtered prior to sampling. This may be done by averaging, which is convolution by a square wave.
- We have seen how to interpret the peaks in the Fast Fourier Transform FFT

References.

1. C. Chatfield, "The Analysis of Time Series", Chapman and Hall, 1996
2. H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis", Cambridge University Press, 1997
3. R.G. Brown, "Smoothing, Forecasting and Prediction of Discrete Time Series", Prentice-Hall, 1963(?).
4. T.W. Anderson, "The Statistical Analysis of Time Series", Wiley, 1994.

5. A. Oppenheim and R.W.Schafer, "Digital Signal Processing", Prentice-Hall 1975.