

Feb 14: Spatial analysis of data fields

Mapping irregularly sampled data onto a regular grid

Many analysis techniques for geophysical data require the data be located at regular intervals in space and/or time. This is so for spectral analysis, digital filtering and wavelet analysis. Other techniques, such as calculating empirical orthogonal functions from a set of spatially distributed data observed at the same times (though not necessarily at regular intervals, i.e. $\Delta t \neq \text{constant}$) often require that gaps in the data be filled. This is true, for example, in the case of satellite observations where clouds partially obscure the field of view, or are subject to data drop out from instrument or algorithm failings.

In oceanography and meteorology, climatologies (e.g. seasonal or monthly means) are typically computed from a compilation of observations made at irregular locations and times, frequently with a sampling distribution that can lead to regional or temporal biases if care is not taken to recognize and address these in the mapping procedure.

These next few lectures will address techniques for producing regularly gridded maps from irregularly sampled data. These techniques have characteristics of both smoothing and filtering (removing time/space scales) and interpolation (spanning gaps in observations).

We begin by reviewing some basics of matrix and vector algebra, drawing on the “Basic Machinery” described in Chapter 3 of Wunsch (1996). [Wunsch, C., The Ocean Circulation Inverse Problem, Cambridge University Press, 442 pp., 1996.]

Matrix and vector algebra, least squares fitting via the normal equations

Topics covered:

- Review of linear algebra conventions, definitions and rules
- Weighted least squares
- Conventional least squares via the normal equations
- Least squares solution to data design matrix equation

[John's old notes \(scanned\) for linear algebra lecture](#)

Linear algebra definitions:

Matrix of M by N values:

$$\mathbf{A} = \{a_{i,j}\}, 1 \leq i \leq M, 1 \leq j \leq N$$

Vector of N values and its transpose

$$\mathbf{q} = \begin{bmatrix} q_1 \\ \vdots \\ q_i \\ \vdots \\ q_N \end{bmatrix}$$

$$\mathbf{q}^T = [q_1 \cdots q_i \cdots q_N]$$

Inner product

The **inner**, or “dot” product, of two vectors is

$$\mathbf{a}^T \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta$$

where θ is the angle (in N -dimensional space) between the two vectors.

If $\theta = 0$ then the vectors are parallel.

If $\theta = \pi/2$ then the vectors are orthogonal.

In more general terms,

$$\mathbf{a}^T \mathbf{b} = \sum_{i=1}^{i=N} a_i b_i$$

from which it follows that both vectors must be of length N (i.e. they are *conforming*) in order to compute the summation.

Basis set

Suppose we had N vectors \mathbf{e}_i , each of dimension (length) N . If it is possible to represent any arbitrary N -dimensional vector, \mathbf{f} , as a weighted sum of the N vectors, \mathbf{e}_i

$$\mathbf{f} = \sum_{i=1}^N \alpha_i \mathbf{e}_i$$

then the \mathbf{e}_i are called a *spanning set*, (or more commonly a *basis set*) because they are sufficient to span the entire N -dimensions.

To have this property, the \mathbf{e}_i must be *independent*, meaning that no single one of the \mathbf{e}_i can be represented as a weighted sum of the others excluding itself.

The coefficients α_k of the expansion can be found by solving a set of simultaneous equations describing the *projection* of \mathbf{f} onto each of the \mathbf{e}_i .

$$\mathbf{e}_k^T \mathbf{f} = \sum_{i=1}^N \alpha_i \mathbf{e}_k^T \mathbf{e}_i$$

This is easily solved in the case that the \mathbf{e}_i are mutually orthogonal and normal (have unit length), in which case we call them *orthonormal*.

Eigenvectors and eigenvalues

Matrix multiplication can be thought of as a transformation of vector \mathbf{x} into vector \mathbf{y}

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

If vector \mathbf{v} has the property that the transformation leaves its direction unchanged, then \mathbf{v} is said to be an eigenvector of matrix \mathbf{A} .

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

If \mathbf{A} is square of dimension N , there are N eigenvectors and they are orthogonal, each with a corresponding eigenvalue λ_n :

$$\mathbf{A}\mathbf{v}_n = \lambda_n \mathbf{v}_n$$

A matrix composed of the N eigenvectors, say \mathbf{Q} , satisfies

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

If \mathbf{A} is symmetric, it will have N real eigenvalues and orthonormal eigenvectors that form a basis set.

Orthonormal vectors

Orthonormal vectors satisfy the property:

$$\mathbf{e}_k^T \mathbf{e}_i = \delta_{ik}$$

where δ_{ik} is the Kronecker delta:

$$\delta_{ik} = 1 \text{ if } i = k, \text{ (normal) and } \delta_{ik} = 0 \text{ if } i \neq k \text{ (orthogonal).}$$

Then $\mathbf{e}_k^T \mathbf{f} = \sum_{i=1}^N \alpha_i \delta_{ki} = \alpha_k$ is the *projection* of \mathbf{f} onto basis vector \mathbf{e}_k and we have easily solved for the coefficients α_k .

Matrix multiplication

Matrix multiplication is $C_{ij} = \sum_{p=1}^P A_{ip} B_{pj}$ i^{th} row of A times j^{th} column of B

which requires the dimensions be *conformable* $M \times N \sim M \times P \ P \times N$ (The requirement that matrix operations be conformable is your friend in Matlab.)

We write $\mathbf{C} = \mathbf{BA}$

Matrix operation rules:

$\mathbf{AB} \neq \mathbf{BA}$ multiplication *is not* commutative

$\mathbf{ABC} = (\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ multiplication *is* associative

$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ the expansion of transpose product

$\text{trace}(\mathbf{A}) = \sum_{i=1}^N a_{ii}$ is sum of diagonal elements

A **symmetric** matrix has the property $\mathbf{A} = \mathbf{A}^T$ so the product $\mathbf{A}^T \mathbf{A}$ is the dot product of all rows of the matrix with themselves.

The **identity** matrix is symmetric:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ so each element } I_{ij} = \delta_{ij}$$

The **inverse** of a matrix \mathbf{A} is denoted \mathbf{A}^{-1} and defined such that $\mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$

It follows that

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$

Norm or length

The length or *norm* of a vector can be defined in many ways, but the conventional l_2 norm is defined

$$\|\mathbf{f}\| = (\mathbf{f}^T \mathbf{f})^{1/2} = \sqrt{\sum_{i=1}^N f_i^2}$$

The Cartesian distance between two vectors is

$$\|\mathbf{a} - \mathbf{b}\| = \sqrt{(\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b})} = \left[(x_a - x_b)^2 + (y_a - y_b)^2 \right]^{1/2}$$

Sometimes the distance between two vectors is weighted

$$\|\mathbf{c}\| = \sqrt{\sum_{i=1}^n c_i W_{ii} c_i} = (\mathbf{c}^T \mathbf{W} \mathbf{c})^{1/2}$$

where in order to be useful the weighting matrix would usually be symmetric and positive definite.

Differentiation

Consider a *scalar*, J , (a single number, not a vector) that is the product

$$J = \mathbf{r}^T \mathbf{q} = \mathbf{q}^T \mathbf{r} \quad (\text{so the vectors must be conformable})$$

Differentiating this scalar with respect to the vector \mathbf{q} produces a vector *gradient* as the result

$$\frac{\partial}{\partial \mathbf{q}} (\mathbf{q}^T \mathbf{r}) = \frac{\partial}{\partial \mathbf{q}} (\mathbf{r}^T \mathbf{q}) = \mathbf{r}$$

much like the differentiation using the product rule for any two variables r and q .

$$\frac{\partial}{\partial q}(rq) = r$$

For a quadratic form where the scalar J may be written:

$$J = \mathbf{q}^T \mathbf{A} \mathbf{q} \quad (\text{this requires the matrix } \mathbf{A} \text{ be } N \times N)$$

we get

$$\frac{\partial J}{\partial \mathbf{q}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{q}$$

much like the differentiation of a quadratic product Aq^2

$$\frac{\partial}{\partial q}(Aq^2) = 2Aq$$

Most spatial analysis of data that entails fitting or smoothing data to fit some statistical or dynamical model, involves some form of weighted least squares fitting.

Least squares fitting

In simple least squares fitting of a set of observations to a linear function, or linear regression, what is assumed is that a set of observations y_i can be described by a “model”

$$\begin{aligned} y(t) &= \theta(t) + n(t) \\ &= a + bt + n(t) \end{aligned}$$

Here, $n(t)$ is the measurement noise and is the source of the misfit between the observations and the “model”.

We can write this as a matrix equation:

$$\mathbf{E} \mathbf{x} + \mathbf{n} = \mathbf{y}$$

where

$$\mathbf{E} = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_M \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \mathbf{n} = \begin{pmatrix} n_1 \\ \vdots \\ n_M \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix}$$

Having zero error would be exceptional, so in general the parameters a and b will represent a best possible fit of the model to the observations.

We have many more data points than the 2 parameters a and b , so the problem is said to be “over determined”.

Frequently, the measure of best fit is the parameter choice that minimizes the mean squared misfit of model and data.

$$\begin{aligned} \min J &= \sum_{i=1}^M n_i^2 = \mathbf{n}^T \mathbf{n} = (\mathbf{E}\mathbf{x} - \mathbf{y})^T (\mathbf{E}\mathbf{x} - \mathbf{y}) \\ &= (\mathbf{E}\mathbf{x})^T (\mathbf{E}\mathbf{x}) - \mathbf{y}^T \mathbf{E}\mathbf{x} - (\mathbf{E}\mathbf{x})^T \mathbf{y} + \mathbf{y}^T \mathbf{y} \end{aligned}$$

Each of these terms is a scalar, so each is its own transpose.

$$\text{So, } \mathbf{y}^T \mathbf{E}\mathbf{x} = (\mathbf{y}^T \mathbf{E}\mathbf{x})^T = (\mathbf{E}\mathbf{x})^T \mathbf{y}$$

Also, $(\mathbf{E}\mathbf{x})^T \mathbf{y} = \mathbf{x}^T \mathbf{E}\mathbf{y}$ so we have

$$J = \mathbf{x}^T \mathbf{E}^T \mathbf{E}\mathbf{x} - 2\mathbf{x}^T \mathbf{E}^T \mathbf{y} + \mathbf{y}^T \mathbf{y}$$

To minimize, we differentiate with respect to \mathbf{x} and set to zero, anticipating a minimum.

$$\frac{\partial J}{\partial \mathbf{x}} = 2(\mathbf{E}^T \mathbf{E})\mathbf{x} - 2\mathbf{E}^T \mathbf{y} = 0$$

This leads to the set of normal equations

$$(\mathbf{E}^T \mathbf{E})\mathbf{x} = \mathbf{E}^T \mathbf{y}$$

Assuming the inverse of the normal equations matrix exists, the solution is

$$\mathbf{x} = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \mathbf{y}$$

No assumptions have been made about the statistical probability density functions of the errors n_i .

Let's give some consideration to how the estimated parameters of the model fit, a and b denoted by \mathbf{x} are affected by the random elements of the observations.

Assume the estimated values are unbiased, then the $\langle \mathbf{x} \rangle = \langle \mathbf{x}^{\text{true}} \rangle$ (expected values).

The uncertainty in the estimated values is described by their variance about the true mean.

$$\begin{aligned}
 P &= \left\langle \left\langle \mathbf{x} - \mathbf{x}^{\text{true}} \right\rangle \left\langle \mathbf{x} - \mathbf{x}^{\text{true}} \right\rangle^{\text{T}} \right\rangle \\
 &= \left(\mathbf{E}^{\text{T}} \mathbf{E} \right)^{-1} \mathbf{E}^{\text{T}} - \mathbf{x}^{\text{true}} \left\langle \mathbf{n} \mathbf{n}^{\text{T}} \right\rangle \mathbf{E} \left(\mathbf{E}^{\text{T}} \mathbf{E} \right)^{-1}
 \end{aligned}$$

In the special case that we have uncorrelated errors, $\left\langle \mathbf{n} \mathbf{n}^{\text{T}} \right\rangle = \sigma_n^2 \mathbf{I}$, i.e. all the observations are known with an uncertainty $\pm \sigma_n$

So the uncertainty in the parameter estimates is $P = \sigma_n^2 \left(\mathbf{E}^{\text{T}} \mathbf{E} \right)^{-1}$

and the uncertainty in the estimated derived from these parameters is

$$\begin{aligned}
 \mathbf{y}^{\text{est}} &= \mathbf{E} \mathbf{x} \\
 \mathbf{y}^{\text{est}} - \mathbf{y} &= \mathbf{n} \\
 P &= (\mathbf{n}^{\text{est}} - \mathbf{n})(\mathbf{n}^{\text{est}} - \mathbf{n})^{\text{T}} \\
 &= \sigma_n^2 (\mathbf{I} - \mathbf{E}(\mathbf{E}^{\text{T}} \mathbf{E})^{-1} \mathbf{E}^{\text{T}})
 \end{aligned}$$

(see Wunsch section 3.3 for details).

You should, at the very least, examine the residuals of the “model” fit compared to the data to see if they are randomly distributed.