# On empirical orthogonal functions and variational methods

Ingemar Holmström

The Swedish Meteorological and Hydrological Institute

Norrköping, Sweden

# ON EMPIRICAL ORTHOGONAL FUNCTIONS AND VARIATIONAL METHODS

by

INGEMAR HOLMSTRÖM

## 1. Coordinate transformation

One of the characteristic properties of eof:s is best demonstrated by a comparison with regression in a very simple linear case. Let $x_n$ and $y_n$ be pairs of observed variables where mean values have been substracted out. Plotting the data in a xy-diagram we may obtain something like fig. 1.

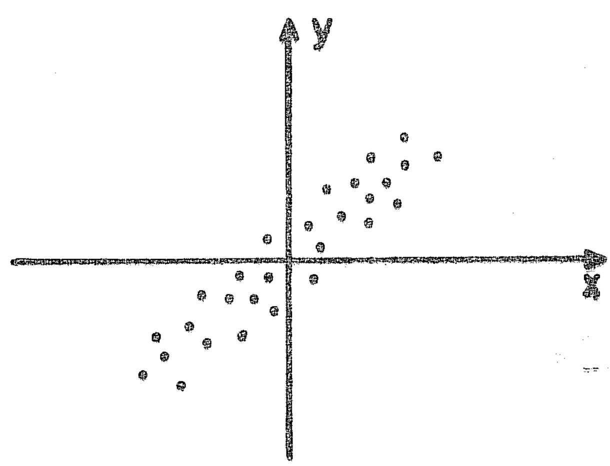


Fig. 1

Depending on the purpose for which we wish to use the data we treat them in two essentially different ways. If we wish to establish a statistical relationship by which y may be calculated if x is known we use a regression method. Assuming a linear dependance

$$y = kx + r$$

we utilize a part of the sample in order to determine k so that the variance of the residual r becomes a minimum, see fig 2a.

The regression coefficient k is found to be

(1) $$k = \mathrm{tg}\alpha = \frac{\overline{xy}}{\overline{x^2}}$$

where the horizontal bar as usual represents an averaging over all the data utilized and where $\alpha$ is the angle between the x-axis and the regression line.

On the other hand if we wish to analyze variations of x and y in order to relate these to external factors or, if we wish to compress as much as possible of the information into one single variable, we use an eof analysis. In the linear, two-dimensional case we introduce a new coordinate system, see fig 2b

(2) $$\zeta = x \cos \beta + y \sin \beta \qquad \eta = y \cos \beta - x \sin \beta$$

where $\beta$ is the angle of rotation and where we now minimize the variance of $\eta$, using the data in the sample

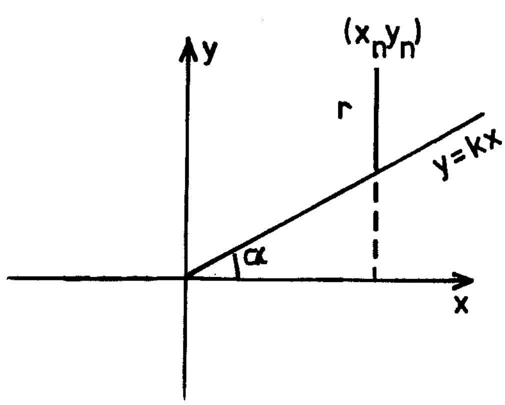


Fig. 2a

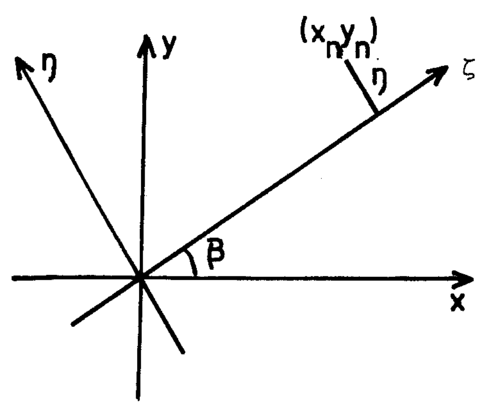


Fig. 2b

In this case we find

(3) $$\mathrm{tg}2\beta = \frac{2\overline{xy}}{\overline{x^2} - \overline{y^2}}$$

so that the angles $\alpha$ and $\beta$ are different.

Since

$$\overline{x^2} + \overline{y^2} = \overline{\zeta^2} + \overline{\eta^2}$$

the minimization of the variance in $\eta$ is equivalent to an optimization of the variance of $\zeta$. We have thus transferred as much as possible of the information to the $\zeta$-direction. In a multi-dimensional case the information in certain directions will be very small and in many applications negligible. The necessary number of degrees of freedom in calculations may then be considerably reduced. Furthermore it is easily shown from (2) and (3) that in the sample

$$\overline{\zeta\eta} = 0$$

or in other words that the information in $\zeta$ is uncorrelated to the information in $\eta$. The implication is that if the variations in x and y are due to external factors with linear responses, then the factors causing variations along $\zeta$ are uncorrelated to the factors causing variations along $\eta$. The transformation into eof therefore provides a mean to differentiate between several external influences.

The linear transformation (2) may easily be generalized to the multidimensional case. Let $f_n(t)$ be a given set of N observed functions in the interval (O,T). Forming

(4) $\Sigma c_n f_n(t) = r(t)$

we wish to determine the coefficients $c_n$ so that the variance of the residual r(t) over the interval (O,T) has an extreme value (minimum). If the set is linearly dependent the residual will be zero for all t but this will not influence the calculation.

In order to exclude the trivial solution, $c_n = 0$ for all n, we have to add the constraint

(5) $\Sigma c_n^2 = \text{constant} \neq 0$

and using a Lagrange multiplier, λ, we arrive at the extreme value problem

$$\frac{\partial}{\partial c_k}\{\int_0^T [\sum_{n=1}^{N} c_n f_n(t)]^2 dt - \lambda \sum_{n=1}^{N} c_n^2\} = 0$$

which leads to the following set of equations.

(6) $\sum_{n=1}^{N} c_n \overline{f_n f_k} = \lambda c_k \quad k = 1, 2, \ldots, N.$

The covariance matrix $\overline{f_n f_k}$ is symmetric and we thus have N eigenvalues $\lambda_i$ and corresponding to these also N eigenvectors $c_{in}$ with the orthogonality condition

$$\sum_{n=1}^{N} c_{in} c_{jn} = \delta_{ij} \sum_{n=1}^{N} c_{in}^2$$

where (5) is the normalizing condition.

When solving the matrix equations (6) one usually orders the eigenvectors $c_{in}$ according to decreasing eigenvalue which corresponds to decreasing value of the variance of the residual r(t) in (4). A vanishing eigenvalue therefore indicates that the set $f_n(t)$ is linearly dependent and thus has less than N degrees of freedom. A very small eigenvalue indicates similarly a degree of freedom that is almost non-existent in the functions $f_n(t)$ and therefore may be neglected.

In the above derivation (and also in the following) we have for convenience assumed the observed functions $f_n(t)$ to be continuous. Nothing is, however, changed if the functions are known only at discrete points, it only means that integrals are replaced by summations.

## 2. Series expansion

A different approach to eof can be made using variational calculus. We then consider the following series expansion of the given functions $f_n(t)$ discussed in the previous paragraph

$$(7) \quad f_n(t) = \sum_{i=1}^{N} \alpha_i(t) g_{in}$$

where $\alpha_i(t)$ and $g_{in}$ are so far undetermined functions and coefficients. We now require the series to converge as rapidly as possible taking all N functions $f_n(t)$ into consideration. This means that each term should take up as much as possible of the total variance in all $f_n(t)$ and denoting by $\alpha(t)g_n$ a general term (dropping index i) we have the variational problem

$$\int_o^T \sum_{n=1}^{N} [f_n(t) - \alpha(t)g_n]^2 dt = \text{minimum}$$

Varying here $\alpha(t)$ and $g_n$ we have

$$\int_o^T \sum_{n=1}^{N} [f_n(t) - \alpha(t)g_n][\alpha(t)\delta g_n + g_n\delta\alpha(t)]dt = 0$$

and since $\delta\alpha(t)$ and $\delta g_n$ are independent and arbitrary we obtain the following two equations

$$(8) \quad \int f_n(t)\alpha(t)dt = g_n \int \alpha^2(t)dt$$

$$(9) \quad \sum_{n=1}^{N} f_n(t)g_n = \alpha(t)\sum_{n=1}^{N} g_n^2 = \alpha(t)$$

where in (9) we have applied the normalizing condition

$$(10) \quad \sum_{n=1}^{N} g_n^2 = 1$$

Elimination of $\alpha(t)$ between (8) and (9) now gives the following matrix equation

$$(11) \quad \sum_{n=1}^{N} g_n\overline{f_n f_k} = g_k\overline{\alpha^2} = \lambda g_k$$

which is identical to equation (6) in the previous paragraph.

The coefficients $g_n$ and $c_n$ are thus the same if we apply the same normalizing conditions (5) and (10). Furthermore if we compare (9) and (4) we see that

$$\alpha(t) = r(t)$$

and that the eigenvalue corresponds to the variance of the residual r(t) in (4).

Finally it should be noted that if in (8) and (9) we eliminate $g_n$ instead of $\alpha(t)$ we arrive at the following integral equation of the Fredholm type

$$(12)\quad \int_0^T \alpha(t')\sum_{n=1}^{N} f_n(t)f_n(t')dt' = \alpha(t)\int\alpha^2(t)dt$$

where the kernel again is symmetric and the eigenfunctions thus are orthogonal. The integral equation (12) also has the same eigenvalues as the matrix equation (11).

Since in both cases we have orthogonal eigenfunctions or eigenvectors we find for the eof expansion the following two orthogonality conditions

$$\sum_{n=1}^{N} g_{in}g_{jn} = \delta_{ij}\ ;\ \int_0^T\alpha_i(t)\alpha_j(t)dt = \delta_{ij}\int_0^T \alpha_i^2(t)dt$$

where i and j are indices for eigenvectors (-functions) corresponding to different eigenvalues.

3. Generalizations

The series expansion approach to eof has in many applications the advantage that it is easier to decide beforehand on the kind of series that will give the required information. Another advantage is also that it is rather easy to generalize to problems where the conventional eof-expansion is of limited value.

As an example we consider a function f(x,y,z) given in a certain domain by continuous or discrete point observations.

To begin with there are three different possibilities for expansion

$$(13)\quad f(x,y,z) = \Sigma\alpha_n(x)A_n(y,z) = \Sigma\beta_n(y)B_n(z,x) =$$

$$= \Sigma\gamma_n(z)C_n(x,y)$$

where, depending on the character of the data and the purpose of the analysis one of these will be the best choice. If the first one is chosen we may in turn expand the functions $A_n(y,z)$ into eof. We have here again three independent variables, y, z and n, where due to a possible truncation in the first series n may be limited. The three possibilities are

$$A_n(y,z) = \sum_m p_{nm}(z)P_m(y) = \sum_m q_{nm}(y)Q_m(Z) =$$

$$= \sum_m r_{nm}R_m(z,y)$$

where the last one is trivial and the other two will lead to results similar to a corresponding treatment of the other series in (13).

In any case one will arrive at summations over two indices and this will generally complicate truncation, as is for instance the case in spectral models using spherical harmonics.

One way to overcome this difficulty is to consider an expansion of the type

$$(14)\quad f(x,y,z) = \Sigma a_n(x)b_n(y)c_n(z)$$

which no longer will lead to a conventional eof-expansion. With the requirement that a general term a(x)b(y)c(z) should take up as much as possible of the total variance we have a variational problem of the type

$$\iiint [f(x,y,z) - a(x)b(y)c(z)]^2 dxdydz = \text{minimum}$$

or, after varying a(x), b(y) and c(z) and separating equations

$$(15)\begin{cases} \iint f(x,y,z)\,a(x)\,b(y)\,dxdy = c(z) \\ \iint f(x,y,z)\,b(y)\,c(z)\,dydz = \lambda a(x) \\ \iint f(x,y,z)\,c(z)\,a(x)\,dzdx = \lambda b(y) \end{cases}$$

where we have taken

$$\int a^2(x)\,dx = 1; \quad \int b^2(y)\,dy = 1; \quad \int c^2(z)\,dz = \lambda$$

Elimination of c(z) in (15) gives the system

$$(16)\begin{cases} \int_{x'}\int_{y}\int_{y'} Q(x,x',y,y')\,a(x')\,b(y')\,b(y)\,dx'dydy' = \\ \qquad = \lambda a(x) \\ \int_{x}\int_{x'}\int_{y'} Q(x,x',y,y')\,a(x)\,a(x')\,b(y')\,dxdx'dy' = \\ \qquad = \lambda b(y) \end{cases}$$

where

$$Q(x,x'y,y') = \int_z f(x,y,z)\,f(x',y',z)\,dz$$

The system (16) may be solved by an iterative procedure starting with an initial guess on a(x) and b(y) and, when the solution is stable, calculate c(z) from the first equation in (15) and subtract the product from the function f(x,y,z). It may also be possible to arrive in an easier way to a solution by expanding in a first step the function f(x,y,z) into eof where x and y are separated from z. This has not yet been tested, nor has any expansion in the case of four or more independent variables been considered.

As we shall see later in this lecture expansions of this generalized type may be of importance in connection with certain types of regression problems.

Another generalization of the eof expansion is obtained in the case we add some kind of constraint to the expansion. An example of this where a geostrophic constraint is applied to a simultaneous expansion of wind and geopotential data will be given in a separate lecture.

Non-linearity

So far we have only been dealing with linear transformations or linear regression. The efficiency of these methods is in many applications sufficient for valuable results to be obtained but non-linearity is common in meteorology and a few words will therefore have to be said about such cases.

A typical example where non-linearity is important is the eof-expansion of a pressure or geopotential field. If deviations $\phi'(x,y,t)$ from a mean value are considered we have

$$\phi'(x,y,t) = \Sigma\alpha_n(t) f_n(x,y)$$

and since $\alpha_n(t)$ will take negative as well as positive values it follows that low and high geopotential areas will be expanded by the same horizontal structure functions. Our general experience is, however, that the horizontal scale of lows is smaller than that of highs and in an optimized expansion we should therefore have the structure functions $f_n(x,y)$ depending also on $\alpha$ or, in other words, replaced by a function $g_n(x,y,\alpha)$. Assuming the influence of $\alpha$ on $g_n(x,y,\alpha)$ to be small we may take

$$g_n(x,y,\alpha) \simeq g_n(x,y,0) + \alpha\left(\frac{\partial g_n}{\partial \alpha}\right)_{\alpha=0}$$

which will give second order products in the expansion.

The variational approach will lead to a system of integral equations that will be very difficult to solve.

In general the introduction of non-linearity into eof-expansion results in a considerable increase in the difficulties of obtaining solutions and we lack suitable methods for this purpose. An example will show this. If in fig. 1 we instead had a distribution of the kind shown in fig. 3

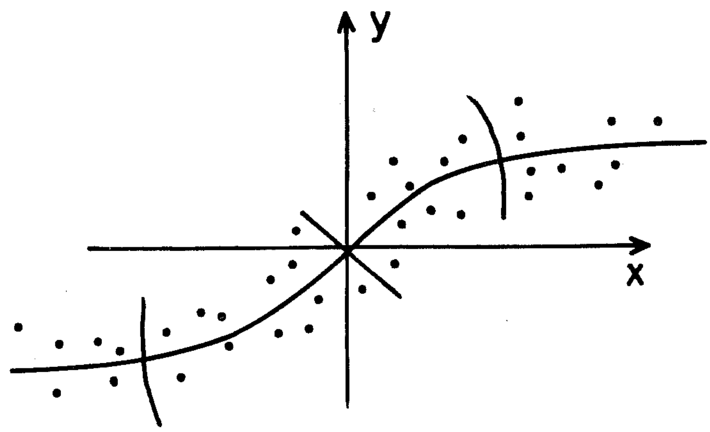


Fig. 3

the linear transformation would not be very efficient. Instead we should introduce a new orthogonal coordinate system

$$\xi = \psi(x,y); \quad \eta = \phi(x,y)$$

where as before we minimize the variance of $\eta$ but where we have to add a constraint on the variance of the curvature of $\xi$ since otherwise we shall obtain a broken line joining all the points $(x_n y_n)$. The curvature is a complicated differential expression and carrying out the variation we immediately get exceedingly awkward non-linear differential equations to solve.

Almost the same difficulty is encountered if we try to establish a regression equation $y = f(x) + r$ with the constraint that the variance of the curvature, given a certain weight, should be minimized at the same time as the variance of the residual r.

One may therefore have to satisfy oneself by an approximate approach, where the variance of first and second order derivatives are minimized. Let $y = g(x)$ be a curve consisting of linear segments joining all observed points. We may then require

$$(17)\quad \int_a^b \{[f(x) - g(x)]^2 + \mu_1 (\frac{df}{dx})^2 + \mu_2 (\frac{d^2 f}{dx^2})^2\} dx = \min$$

where $(a,b)$ is the interval for $x$ and $\mu_1$ and $\mu_2$ are weighting factors. Variation of $f$ gives

$$\mu_2 \frac{d^4 f}{dx^4} - \mu_1 \frac{d^2 f}{dx^2} + f = g$$

with the natural boundary conditions

$$\mu_1 \frac{df}{dx} - \mu_3 \frac{d^3 f}{dx^3} = 0 \quad \frac{d^2 f}{dx^2} = 0 \quad \text{for } x = a,b$$

if we do not prescribe $f$ and $df/dx$ respectively at the boundaries.

The solution is then easily obtained and may be shown to result in damping at an increased rate of higher wave numbers, the rate of damping depending on the choice of $\mu_1$ and $\mu_2$.

A similar approximate approach may be used for the coordinate transformation leading to a "non-linear" eof expansion.

In the case of non-linear regression it is worth while to mention a method that in certain cases may give interesting results. Let $a(t)$ be the predictand and $b_1(t)$ and $b_2(t)$ suitable predictors.

The regression equation is then of the form

$$a(t) = f[b_1(t), b_2(t)] + r$$

where f is an undetermined function

which we may expand into eof, taking

$$f[b_1(t),\ b_2(t)] = \Sigma\alpha_n(b_1)h_n(b_2)$$

The calculation of $\alpha_n(b_1)$ and $h_n(b_2)$ will depend on the kernel

$$\int a(b_1,b_2)a(b_1',b_2)db_2$$

which will necessitate a mapping of a at gridpoints of $b_1$ and $b_2$ in phase space. The analysis required for this purpose may be based on an equation similar to (17) but with x-derivatives replaced by two-dimensional nabla-operators.

The method may also be extended to three or more predictors if the expansion indicated in the preceeding paragraph is used.