

Jürgen Kusche, Annette Eicker and Ehsan
Forootan

Analysis Tools for GRACE- and Related Data Sets

Theoretical Basis

Lecture Notes

Summerschool 'Global Hydrological Cycle',
Mayschoss, Sept. 12-16, 2011

September 6, 2011

DFG Priority Program SPP1257

Mass Transport and Mass Distribution in the Earth System

Bonn University

These lecture notes were compiled on the occasion of the summerschool 'Global Hydrological Cycle', organized by DFG's priority program SPP1257 'Mass Transport and Mass Distribution in the System Earth' at September 12-16, 2011 in Mayschoss/Ahr. Our aim was to familiarize students with different background (geodesy, hydrology, oceanography, geophysics) with some mathematical concepts that are fundamental for analysing level-2 products (sets of spherical harmonic coefficients) from the GRACE mission and related geophysical data sets (model outputs in gridded form). The focus was on concepts, and technical proofs were avoided. Specific topics were filtering and basin averaging, and the application of the principal component analysis technique. Thanks for spotting typos go to Volker Klemann and Torsten Mayer-Gürr.

Jürgen Kusche, Annette Eicker and Ehsan Forootan

Bonn, September 6, 2011

Contents

1	Smoothing and Averaging of Functions on the Sphere	1
1.1	Area Averaging ('Windowing')	3
1.1.1	Exact Windowing	3
1.1.2	Exact Windowing in Spherical Harmonic Representation	4
1.2	Smoothing of Spherical Harmonic Models	5
1.2.1	Isotropic Filters	6
1.2.2	Anisotropic Filters	7
1.3	Smoothed Area Averaging	9
1.3.1	Spherical Harmonic Representation	9
1.3.2	Amplitude Damping ('Bias')	10
1.4	Filter Shape	10
1.4.1	Impulse Response	10
1.4.2	Localization	11
2	Principal Component Analysis and Related Ideas	13
2.1	Principle Component Analysis	13
2.1.1	PCA as a Data Compression Method: Mode Extraction and Data Reconstruction	13
2.1.2	Temporal PCA versus Spatial PCA	18
2.1.3	PCA of Linearly Transformed Data	19
2.1.4	PCA as a Data Whitening Method: Homogeneization	20
2.1.5	Number of Modes	21
2.1.6	PCA as a Tool for Comparing Multiple Data Sets	22
2.1.7	Rotation	23
2.2	Independent Component Analysis	25
3	Appendix	27
3.1	Spherical Harmonics	27
3.2	Spherical Coordinates	33
	References	35

Smoothing and Averaging of Functions on the Sphere

At the product level 2, GRACE data are condensed to monthly sets of fully normalized spherical harmonic coefficients. These coefficients are the outcome of a rather complex data processing scheme. For scientific analysis, users of GRACE data will have to perform a number of basic operations on the GRACE coefficients: transform dimensionless geopotential coefficients into gridded geoid heights or maps of surface density expressed through equivalent water heights, and average such maps over the surface of some hydrological catchment area or ocean basin. Moreover, one of the problems that users of the GRACE level-2 products face is the presence of increasing correlated noise ('stripes') at higher frequencies. Smoothing operators can be applied in either spatial or spectral domain in order to suppress the effect of noise in maps and area averages. The purpose of this chapter is to describe the mathematical concepts underlying these procedures.

Notation

According to [2], we write the gravitational potential at a fixed location as a function of time as

$$V(r, \theta, \lambda, t) = \frac{GM}{r} + \frac{GM}{r} \sum_{n=2}^{\bar{n}} \left(\frac{R}{r}\right)^n \sum_{m=0}^n P_{nm}(\cos \theta) (\bar{C}_{nm}(t) \cos m\lambda + \bar{S}_{nm}(t) \sin m\lambda)$$

with $\theta = \frac{\pi}{2} - \phi$, $GM = \mu$ (in [2]), and fully normalized spherical harmonic coefficients \bar{C}_{nm} and \bar{S}_{nm} . The other quantities will be explained later.

In the following, we will refer to either temporal variations in the geoid or in total water storage (TWS), the spherical harmonic coefficients of which we will denote as \bar{f}_{nm} . These quantities are related to the gravitational potential via simple spectral relations, which are valid under certain assumptions

(‘spherical approximation’, radial Earth model) that will be discussed elsewhere during the summer school. Under these assumptions, the geoid or TWS changes projected to the space domain read

$$F = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{f}_{nm} \bar{Y}_{nm}(\lambda, \theta) .$$

Total water storage change from geopotential harmonics. In this case, the common approximation is

$$\bar{f}_{nm}(t) = R \frac{\rho_e}{3} \frac{2n+1}{1+k'_n} (\bar{v}_{nm}(t) - \bar{v}_{nm})$$

with

$$\bar{v}_{nm} = \bar{C}_{nm} \quad \text{for } m \geq 0, \quad \bar{v}_{nm} = \bar{S}_{n|m|} \quad \text{for } m < 0 ,$$

and \bar{v}_{nm} either a suitable long-term mean of these (i.e. $\bar{v}_{nm} = \langle \bar{v}_{nm}(t) \rangle_{t_A}^{t_B}$) or they refer to a reference epoch t_0 .

In the above, TWS is expressed as a surface density (unit $\frac{\text{kg}}{\text{m}^2}$), the height of a water column is usually derived by scaling the above by a reference density of water, i.e. multiplication by $1/\rho_w$. Thus, ρ_w (if applied) is a reference quantity which has to be chosen as a convention (usually, $\rho_w = 1000 \frac{\text{kg}}{\text{m}^3}$ or $\rho_w = 1025 \frac{\text{kg}}{\text{m}^3}$). The average density of the Earth, ρ_e , is related to the Earth’s mass M by $M = \frac{4\pi\rho_e}{3} R^3$ and follows therefore to $\rho_e = 5517 \frac{\text{kg}}{\text{m}^3}$. Finally, the k'_n are the elastic gravity load Love numbers and follow from 1D-models of the Earth’s rheologic properties.

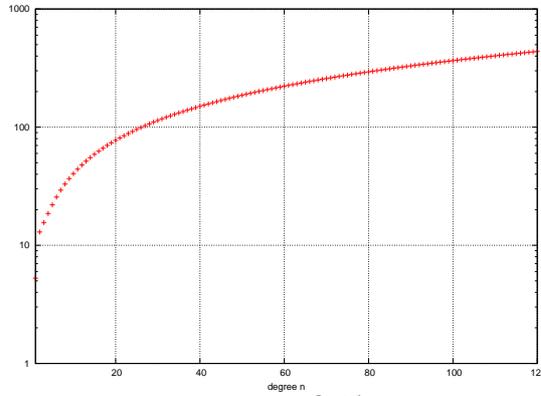


Fig. 1 Shown are the coefficients $\frac{\rho_e}{3\rho_w} \frac{2n+1}{1+k'_n}$. It is obvious that, when computing TWS harmonics from geopotential harmonics, errors in higher degrees will be amplified compared to those in low-degree coefficients

Geoid change from geopotential harmonics. In this case, the common approximation is

$$\bar{f}_{nm}(t) = R (\bar{v}_{nm}(t) - \bar{v}_{nm})$$

with the \bar{v}_{nm} as defined above.

1.1 Area Averaging ('Windowing')

In many applications one is interested in averaging a field over a certain geographical area or region or basin. This is what we call area averaging or *windowing* here.

1.1.1 Exact Windowing

Let us start with a given field F defined on the sphere Ω ,

$$F = F(\lambda, \theta) \quad (1.1)$$

expressed either in spatial domain or in spherical harmonic representation.

The area $O \subset \Omega$ can be mathematically expressed through its characteristic function

$$O = O(\lambda, \theta) = \begin{cases} 1 & (\lambda, \theta) \in O \\ 0 & (\lambda, \theta) \notin O \end{cases} \quad (1.2)$$

(1 inside the area, and 0 outside of it). Its area (size) \bar{O} is

$$\bar{O} = \int_O d\omega = \int_{\Omega} O d\omega . \quad (1.3)$$

Windowing F over O means to derive an average

$$\bar{F}_O = \frac{1}{\bar{O}} \int_O F d\omega = \frac{1}{\bar{O}} \int_{\Omega} OF d\omega \quad (1.4)$$

of F over O .

Remark. If $F = F(\lambda, \theta, t)$ is a spatio-temporal field, the average $\bar{F}_O(t)$ will be a time-series.

Remark. Usually, a polygon $O(\lambda_i, \theta_i), i = 1 \dots q$ will be used to characterize $O(\lambda, \theta)$.

Remark. In practical computations in the space domain, the integrals will be replaced by discrete sums, introducing a discretization error ϵ whose magnitude depends on the spatial grid resolution and the smoothness of both the function F and the region boundary of O . The discrete version of Eq. (1.4) can be cast as $\bar{F}_O = \mathbf{o}^T \mathbf{f}$ or $\bar{F}_O(t) = \mathbf{o}^T \mathbf{f}(t)$ if the field F is time-dependent.

Remark. If F is approximated by a spherical harmonic expansion before projecting onto a grid and evaluation of the discrete sum, an extra truncation error is introduced.

1.1.2 Exact Windowing in Spherical Harmonic Representation

Next, we consider both F and O expanded in 4π -normalized spherical harmonics \bar{Y}_{nm} .

$$F = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{f}_{nm} \bar{Y}_{nm}(\lambda, \theta) \quad (1.5)$$

$$O = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{O}_{nm} \bar{Y}_{nm}(\lambda, \theta) \quad (1.6)$$

(the overbar in \bar{f}_{nm} etc. tells that coefficients are 4π -normalized). It is immediately clear that

$$\bar{O} = \int_{\Omega} O \bar{Y}_{00} d\omega = 4\pi \bar{O}_{00} \quad (1.7)$$

The exact average of F over O is then

$$\bar{F}_O = \frac{1}{\bar{O}_{00}} \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{O}_{nm} \bar{f}_{nm} \quad (1.8)$$

or, if $\bar{o}_{nm} = \frac{\bar{O}_{nm}}{\bar{O}_{00}}$ are the region's SH coefficients further normalized to $\bar{o}_{00} = 1$,

$$\bar{F}_O = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{o}_{nm} \bar{f}_{nm} \quad (1.9)$$

Example. The first 4π -normalized coefficients of the *ocean function* are given in the table.

Table 1.1. Ocean function spherical harmonics

n	m	\bar{O}_{nm}	\bar{O}_{n-m}
0	0	$0.701227 \cdot 10^0$	$0.000000 \cdot 10^0$
1	0	$-0.176689 \cdot 10^0$	$0.000000 \cdot 10^0$
1	1	$-0.115778 \cdot 10^0$	$-0.635533 \cdot 10^{-1}$
2	0	$0.618996 \cdot 10^{-2}$	$0.000000 \cdot 10^0$
2	1	$-0.450010 \cdot 10^{-1}$	$-0.717864 \cdot 10^{-1}$
2	2	$0.471078 \cdot 10^{-1}$	$0.464998 \cdot 10^{-2}$
3	0	$-0.355365 \cdot 10^{-2}$	$0.000000 \cdot 10^0$
3	1	$0.518058 \cdot 10^{-1}$	$-0.251440 \cdot 10^{-1}$
3	2	$0.691439 \cdot 10^{-1}$	$-0.992945 \cdot 10^{-1}$
3	3	$-0.135222 \cdot 10^{-1}$	$-0.947600 \cdot 10^{-1}$

Remark. In practical computations, the spherical harmonic summation will

be evaluated up to finite degree \bar{n} , and a truncation error results. The exact average of F over O can be split into

$$\bar{F}_O = \frac{1}{\bar{O}_{00}} \sum_{n=0}^{\bar{n}} \sum_{m=-n}^n \bar{O}_{nm} \bar{f}_{nm} + \frac{1}{\bar{O}_{00}} \sum_{n=\bar{n}+1}^{\infty} \sum_{m=-n}^n \bar{O}_{nm} \bar{f}_{nm} . \quad (1.10)$$

This may be written as $\bar{F}_O = \mathbf{o}^T \mathbf{f} + \epsilon$. The second term is the truncation or omission error. It will vanish if either F or O is band-limited with degree \bar{n} , or if the high-degree components of F and O are orthogonal on the sphere in L_2 -sense.

Remark. If F is truncated at degree \bar{n} , then projected into the space domain and the integral is evaluated over an exactly delineated area, the above mentioned truncation error will occur as well.

1.2 Smoothing of Spherical Harmonic Models

Smoothing or filtering a field is usually applied to suppress 'rough' or 'oscillatory' or 'noisy' components.

Consider F according to Eq. (1.1) and Eq. (1.5). A smoothed version is obtained by convolving F against a two-point kernel W with suitable properties. In the spatial domain,

$$F_W(\lambda, \theta) = \int_{\Omega} W(\lambda, \theta, \lambda', \theta') F(\lambda', \theta') d\omega \quad (1.11)$$

and in the spectral domain, in the rather general case of an arbitrarily shaped window function,

$$F_W(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{f}_{nm}^W \bar{Y}(\lambda, \theta) , \quad \bar{f}_{nm}^W = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \bar{w}_{nm}^{n'm'} \bar{f}_{n'm'} . \quad (1.12)$$

Thus, $W(\lambda, \theta, \lambda', \theta')$ describes the weighted contribution of F at point λ, θ to the windowed function F_W at point λ', θ' . In its discrete version in either spatial or spectral domain, smoothing will read $\mathbf{f}_W = \mathbf{W}\mathbf{f}$ (up to a truncation error, which we will omit in the sequel). The \bar{f}_{nm}^W are the SH coefficients of the smoothed version of F . And the $\bar{w}_{nm}^{n'm'}$ are the 4π -normalized spherical harmonic coefficients of the two-point smoothing kernel

$$W(\lambda, \theta, \lambda', \theta') = \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \bar{w}_{nm}^{n'm'} \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{n'm'}(\lambda', \theta') . \quad (1.13)$$

Or, $W(\lambda, \theta, \lambda', \theta') = \mathbf{y}^T(\lambda, \theta) \mathbf{W} \mathbf{y}(\lambda', \theta')$ with matrix \mathbf{W} containing the elements $\bar{w}_{nm}^{n'm'}$. It is obvious that for fixed λ', θ' the $\bar{w}_{nm}^{n'm'} \bar{Y}_{n'm'}(\lambda', \theta')$ are the

4π -normalized spherical harmonic coefficients of $W(\lambda, \theta)$, and vice versa; for fixed λ, θ the $\bar{w}_{nm}^{n'm'} \bar{Y}_{nm}(\lambda, \theta)$ are the SH coefficients of $W(\lambda', \theta')$. Consequently, for a given two-point kernel W ,

$$\bar{w}_{nm}^{n'm'} = \frac{1}{(4\pi)^2} \int_{\Omega} \int_{\Omega'} W(\lambda, \theta, \lambda', \theta') \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{n'm'}(\lambda', \theta') d\omega d\omega' . \quad (1.14)$$

This is the most general case of smoothing.

1.2.1 Isotropic Filters

Most filters that we encounter in the literature are isotropic, i.e. the smoothing kernel depends only on the spherical distance ψ between the two points λ, θ and λ', θ' and not on their relative orientation. A comprehensive review is [4]. For isotropic kernels, the SH coefficients of the kernel can be reduced to the Legendre coefficients of a zonal (z -symmetric) function w_n . Thus,

$$\begin{aligned} W(\psi) &= \sum_{n=0}^{\infty} (2n+1) w_n P_n(\cos \psi) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n w_n \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{nm}(\lambda', \theta') = W(\lambda, \theta, \lambda', \theta') \end{aligned} \quad (1.15)$$

(with $P_n(\cos \psi)$ being the unnormalized Legendre polynomials) or

$$\bar{w}_{nm}^{n'm'} = \delta_{nm}^{n'm'} w_n . \quad (1.16)$$

For isotropic filters, smoothing of F can be written as

$$F_W(\lambda, \theta) = \int_{\Omega} W(\psi) F(\lambda', \theta') d\omega \quad (1.17)$$

and in the spectral domain simply $\bar{f}_{nm}^W = w_n \bar{f}_{nm}$ and

$$F_W(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n w_n \bar{f}_{nm} \bar{Y}(\lambda, \theta) . \quad (1.18)$$

Example. A first example is the *boxcar filter*, which simply truncates the function F at SH degree \bar{n}

$$W_{(\bar{n})}(\psi) = \sum_{n=0}^{\bar{n}} (2n+1) P_n(\cos \psi) , \quad w_n^{(\bar{n})} = \begin{cases} 1 & \text{for } \begin{cases} n \leq \bar{n} \\ n > \bar{n} \end{cases} \\ 0 & \end{cases}$$

Example. A second example is the *Gaussian filter*, popularized by [19], for which we know an analytic expression in the spatial domain ('bell-shaped') with

$$W_d(\psi) = 2b \frac{e^{-b(1-\cos\psi)}}{1 - e^{-2b}} = \sum_{n=0}^{\infty} (2n+1)w_n^{(d)}P_n(\cos\psi)$$

with

$$b = \frac{\ln(2)}{1 - \cos \frac{d}{R}} .$$

Here, $d = R\psi_d$ is the 'half-width' radius parameter where the kernel drops from 1 at $\psi = 0$ to $\frac{1}{2}$, which is commonly used to indicate the degree of smoothing. The Legendre coefficients of the Gaussian filter are found from recursion relations,

$$w_0^{(d)} = 1 , \quad w_1^{(d)} = \left(\frac{1 + e^{-2b}}{1 - e^{-2b}} - \frac{1}{b} \right) , \quad w_{n+1}^{(d)} = -\frac{2n+1}{b}w_n^{(d)} + w_{n-1}^{(d)} .$$

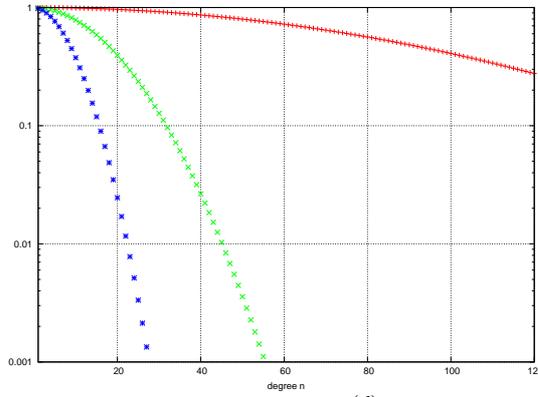


Fig. 2 Shown are the Legendre coefficients $w_n^{(d)}$ for d equal to 100 km, 500 km and 1000 km.

1.2.2 Anisotropic Filters

Anisotropic filters can be characterized into symmetric (or diagonal) filters and non-symmetric filters ([14]). A further differentiation among non-symmetric filters is possible when thinking of the coefficients $\bar{w}_{nm}^{n'm'}$ as being ordered within matrix \mathbf{W} (when we use a particular ordering scheme for the \bar{f}_{nm} , the same has to apply to the filter coefficients).

For *symmetric* filters, \mathbf{W} is diagonal and

$$\bar{w}_{nm}^{n'm'} = \delta_{nm}^{n'm'} w_{nm} . \quad (1.19)$$

Thus the smoothing kernel has the shape

$$W(\lambda, \theta, \lambda', \theta') = \sum_{n=0}^{\infty} \sum_{m=-n}^n w_{nm} \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{nm}(\lambda', \theta') . \quad (1.20)$$

It is symmetric with respect to the points λ, θ and λ', θ' .

Example. Han's filter ([11]) is of this type. They chose a Gaussian filter with 'order-dependent' smoothing radius $d(m)$,

$$w_{nm} = w_n^{(d(m))}, \quad d(m) = \frac{d_1 - d_0}{m_1} m + d_0$$

Example. The 'fan' ([21]) filter is simply a product of different Gaussian filters applied to order and degree,

$$w_{nm} = w_{nm}^{(d_1, d_2)} = w_n^{(d_1)} \cdot w_m^{(d_2)}$$

In the general case of Eq. (1.12), the filter is *non-symmetric* with respect to points λ, θ and λ', θ' and its matrix \mathbf{W} is full.

Remark. Even if \mathbf{W} is symmetric, the resulting filter would be *non-symmetric*.

Example. The DDK filter ([15], [16]). Here, the filter matrix is derived by regularization of a 'characteristic' normal equation system that involves a-priori information on the signal variance and the observation system from which we obtain the unfiltered coefficients,

$$\mathbf{W}_{(\alpha)} = \mathbf{L}_\alpha \mathbf{N} = (\mathbf{N} + \alpha \mathbf{M})^{-1} \mathbf{N}$$

or

$$\bar{w}_{nm}^{n'm'(\alpha)} = \sum_{n=0}^{\infty} \sum_{m=-n}^n L_{nm}^{n'm'(\alpha)'} N_{n'm''}^{n'm'}$$

with \mathbf{M} being an approximation to $\mathbf{C}_f = E\{\mathbf{f}\mathbf{f}^T\}$, \mathbf{N} being an approximation to $\mathbf{C}_{\hat{f}} = E\{\hat{\mathbf{f}}\hat{\mathbf{f}}^T\}$, and $L_{nm}^{n'm'(\alpha)'}$, $N_{n'm''}^{n'm'}$ the corresponding elements. In addition, α is a damping parameter. The DDK filters are non-symmetric. In [16] it was shown that the original $\mathbf{W}_{(\alpha)}$ of [15] can be safely replaced by a block-diagonal version of the matrix.

Example. The 'Swenson and Wahr' filter ([20]) is non-symmetric and it can also be represented by a block-diagonal \mathbf{W} . The idea of this filter is that an empirical model for the correlations between SH coefficients \bar{f}_{nm} of the same order and parity is formulated and then used for decorrelation.

Example. EOF filtering means one applies PCA to a time series \mathbf{f}_i of either gridded values of F or SH coefficients. A reconstruction with q modes provides

$$\mathbf{f}^{(q)} = \mathbf{E}\mathbf{I}^{(q)}\mathbf{E}^T\bar{\mathbf{f}} = \mathbf{W}^{(q)}\mathbf{f}$$

where \mathbf{E} contains the EOFs of the time series and $\mathbf{I}^{(q)}$ is a diagonal matrix with unity in the first q entries and zero otherwise. EOF filtering corresponds to application of a non-symmetric filter as well.

1.3 Smoothed Area Averaging

Now let us come back to the windowing of a spherical harmonic model F , i.e. we wish to average F over the region O . Of course we can window a smoothed version F_W of F as well, if necessary.

Another view on the same operation is as follows: In place of Eq. (1.2), we may introduce a smoothed area function O_W ,

$$O_W(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{O}_{nm}^W \bar{Y}_{nm}(\lambda, \theta) = \frac{1}{4\pi} \int_{\Omega} W(\lambda, \theta, \lambda', \theta') O(\lambda', \theta') d\omega' , \quad (1.21)$$

and we will apply O_W to the original function F

$$\bar{F}_{O_W} = \frac{1}{\bar{O}_W} \int_{\Omega} O_W F d\omega \quad (1.22)$$

(note that $\bar{O}_W = \bar{O}$ if the filter is normalized, see below). In general, the smoothing kernel W is a two-point function on the sphere, cf. Eq. (1.3).

1.3.1 Spherical Harmonic Representation

In case of Eq. (1.15), i.e. W is isotropic, the smoothed area function can be written as

$$O_W(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{O}_{nm}^W \bar{Y}_{nm}(\lambda, \theta) \quad (1.23)$$

with

$$\bar{O}_{nm}^W = \bar{w}_n \bar{O}_{nm} . \quad (1.24)$$

The smoothed area average is found in the spectral domain as

$$\bar{F}_{O_W} = \frac{1}{\bar{O}_W} \sum_{n=0}^{\infty} \sum_{m=-n}^n w_n \bar{O}_{nm} \bar{f}_{nm} . \quad (1.25)$$

The choice $w_0 = \bar{w}_0^{00} = 1$ ('filter normalization') guarantees that

$$\frac{1}{4\pi} \int_{\Omega} O_W d\omega = \bar{O}_W = \bar{O}_0 = \frac{1}{4\pi} \int_{\Omega} O d\omega . \quad (1.26)$$

I.e. the 'area' of the smoothly varying window O_W equals to the area of O .

But, the smoothing kernel will inevitably 'leak' energy beyond the original region. I.e.

$$\frac{1}{4\pi} \int_{\Omega} O_W d\omega = \frac{1}{4\pi} \int_O O_W d\omega + \frac{1}{4\pi} \int_{\Omega/O} O_W d\omega . \quad (1.27)$$

The above can be transferred to the more general case of non-isotropic smoothing without any problem.

1.3.2 Amplitude Damping ('Bias')

For a given area O , windowing or smoothing will decrease the amplitude of the average \bar{F}_W with respect to the original average \bar{F} . What causes this reduction is best understood by explicitly writing down the 'reduction factor' $\beta_{O,W,F}$, which we define as

$$\beta_{O,W,F} = \frac{\bar{F}_{O_W}}{\bar{F}_O} \quad (1.28)$$

and which is specific for a certain area O , a certain window kernel W , and an input function F . For an isotropic smoothing kernel,

$$\beta_{O,W,F} = \frac{\bar{O}_W \int_{\Omega} O_W F d\omega}{\int_{\Omega} O F d\omega} = \frac{1}{w_0} \frac{\sum_{n=0}^{\infty} \sum_{m=-n}^n w_n \bar{O}_{nm} \bar{f}_{nm}}{\sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{O}_{nm} \bar{f}_{nm}} \quad (1.29)$$

and for $w_0 = 1$

$$\beta_{O,W,F} = 1 - \frac{1}{\bar{O}_{00} \bar{F}_O} \sum_{n=1}^{\infty} \sum_{m=-n}^n (1 - w_n) \bar{O}_{nm} \bar{f}_{nm} \quad (1.30)$$

The reduction factor clearly depends on the basin shape, the filter coefficients, and the signal itself.

Example. For $w_0 = 1$ and $F = c$, where c is a constant (i.e. the signal is constant over the whole sphere), β is exactly one, i.e. no damping occurs at all.

Example. For $w_0 = 1$ and $F = c \cdot O(\lambda, \theta)$ (the signal is constant over the area O , and exactly zero outside), the damping factor becomes (considering $\int_{\Omega} O^2 d\omega = \int_{\Omega} O d\omega = \bar{O}_{00}$)

$$\beta_{O,W,c \cdot O} = 1 - \frac{1}{c \cdot \bar{O}_{00}} \sum_{n=1}^{\infty} \sum_{m=-n}^n (1 - w_n) \bar{O}_{nm}^2$$

Example. In ([15]), a 'standard damping factor' ('scaling bias') is defined for smoothing a constant signal over a spherical cap area, and numbers are provided for Gaussian and DDK filters of different degree of smoothing and at different geographical latitudes.

1.4 Filter Shape

1.4.1 Impulse Response

For comparing smoothing kernels in the spatial domain, it is helpful to map a kernel's impulse response. This can be best understood when we imagine an

area O shrinks to a point on the sphere. By letting the basin function degrade to a Dirac function (we want to see the smoothing effect for a particular location λ', θ'), we obtain

$$O^\delta(\lambda, \theta) = \delta^{\lambda', \theta'}(\lambda, \theta) = \begin{cases} \infty & \text{for } \lambda' = \lambda, \theta' = \theta \\ 0 & \text{otherwise} \end{cases} \quad (1.31)$$

and

$$\bar{O}_{nm}^\delta = \frac{1}{4\pi} \int_{\Omega} \delta^{\lambda', \theta'}(\lambda'', \theta'') \bar{Y}_{nm}(\lambda'', \theta'') d\omega = \bar{Y}_{nm}(\lambda', \theta') . \quad (1.32)$$

Remark. Eq. (1.32) is very helpful in practical applications, since one only has to compute the $\bar{Y}_{nm}(\lambda', \theta')$. Or, with the spherical harmonic representation of the Dirac,

$$O^\delta(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{Y}_{nm}(\lambda', \theta') \bar{Y}_{nm}(\lambda, \theta) . \quad (1.33)$$

Consequently, the impulse response of the most general non-isotropic two-point kernel W will be

$$O_W^\delta(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \bar{w}_{nm}^{n'm'} \bar{Y}_{n'm'}(\lambda', \theta') \bar{Y}_{nm}(\lambda, \theta) . \quad (1.34)$$

And for an isotropic kernel

$$O_W^\delta(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n w_n \bar{Y}_{nm}(\lambda', \theta') \bar{Y}_{nm}(\lambda, \theta) . \quad (1.35)$$

1.4.2 Localization

The localization of an isotropic smoothing kernel can be best measured by its 'half-with' radius, i.e. the distance $d = R\psi_d$ where the kernel drops from 1 at $\psi = 0$ to $\frac{1}{2}$

$$W(\psi_d) = \frac{1}{2} . \quad (1.36)$$

For non-isotropic kernels, measuring the localization is more difficult. Unlike with isotropic kernels, it will depend on the particular location λ', θ' . There, one might compute the half-with radius in two directions - North and East.

Following [17] and [1], [15] introduced the variance σ_W of the squared normalized window function $W(\lambda, \theta)$ at location λ', θ' as a single measure for its localization properties. The variance is the second centralized moment of a probability density function defined on the sphere; it is an integral measure for the spreading about the expectation and it is independent of introducing a particular coordinate system on the sphere.

We suppose with [1] that W^2 has been normalized,

$$\int_{\Omega} W^2(\lambda', \theta') d\omega = \sum_{n=0}^{\infty} \sum_{m=-n}^n \left(\bar{w}_{nm}^{n' m'} \bar{Y}_{n' m'}(\lambda', \theta') \right)^2 = 1. \quad (1.37)$$

The integration in the first term applies to λ, θ . Normalization is required in order to interpret W as a probability density function. The expectation in the space domain is introduced ([1]) via

$$\boldsymbol{\mu}_W = \int_{\Omega} \mathbf{e} W^2 d\omega \quad (1.38)$$

where $\mathbf{e} = (\sin \theta \cos \lambda, \sin \theta \sin \lambda, \cos \theta)^T$ is the unit vector pointing from the origin to a location on Ω . If $W^2(\lambda, \theta)$ (for given λ', θ') is thought to represent a surface density distribution, $\boldsymbol{\mu}_W$ points to its center of mass (which is inside of Ω).

As the unit vector can be represented through the unnormalized degree-1 spherical harmonics

$$\mathbf{e} = (Y_{11}, Y_{1-1}, Y_{10})^T \quad (1.39)$$

we can write the components of $\boldsymbol{\mu}_W$ as

$$\mu_{W;x} = \int_{\Omega} W^2 Y_{11} d\omega = (W^2)_{11} \quad \mu_{W;y} = (W^2)_{1-1} \quad \mu_{W;z} = (W^2)_{10}. \quad (1.40)$$

The variance of $W(\lambda, \theta)$ is introduced in the usual fashion, i.e. as the expectation of $(\mathbf{e} - \boldsymbol{\mu}_W)^2$

$$\sigma_W^2 = \int_{\Omega} (\mathbf{e} - \boldsymbol{\mu}_W)^2 W^2 d\omega \quad (1.41)$$

Because of $(\mathbf{e} - \boldsymbol{\mu}_W)^2 = 1 + (\boldsymbol{\mu}_W)^2 - 2\mathbf{e}^T \boldsymbol{\mu}_W$ and $\int -2\mathbf{e}^T \boldsymbol{\mu}_W W^2 d\omega = -2(\boldsymbol{\mu}_W)^2$, the variance is simply

$$\sigma_W^2 = 1 - (\boldsymbol{\mu}_W)^2 = 1 - \sum_{m=-1}^1 ((W^2)_{1m})^2. \quad (1.42)$$

and its computation requires only the computation of the degree-1 harmonics of W^2 .

The degree-1 harmonics of W^2 may be computed directly, involving the *Clebsch-Gordon* coefficients, or simply by projecting the normalized W^2 onto a grid and subsequent spherical harmonic analysis.

Principal Component Analysis and Related Ideas

Products of geodetic observing systems (GRACE, altimetry) and geophysical modelling are most often represented in form of time series of spatial maps (total water storage, sea level anomalies, . . .). The user of these products will often find a few spatial pattern dominating the variability within these maps. Identifying these pattern can aid in physical interpretation, comparison of different data sets, and removing unnecessary small-scale signals or noise. Eigenspace techniques as the principal component method are among the most popular analysis techniques supporting these objectives. The purpose of this chapter is to describe the mathematical concepts behind the principle component analysis (PCA), to introduce some alternative formulations, and to make the reader aware of some of the many choices to be made by the analyst.

2.1 Principle Component Analysis

2.1.1 PCA as a Data Compression Method: Mode Extraction and Data Reconstruction

Sampling spatio-temporal fields can lead to huge amounts of data. For example, a field observed or modelled on a $1^\circ \times 1^\circ$ grid, with a time step of one day, provides already more than $23 \cdot 10^6$ data elements for one year of data. It is now a challenging task to reduce the dimensionality of the data vector and to identify the most important patterns explaining the variability of the system. The *Empirical Orthogonal Functions (EOF)* technique, also called *Principal Component Analysis (PCA)*, has become one of the most widely used methods. General references are [6] and [7]. In pattern analysis, PCA is also known as Karhunen-Loeve transform or Hotelling transform.

PCA has been used extensively to extract individual dominant *modes* of the data variability, while simultaneously suppressing those modes connected with

low variability and therefore reducing the number of data efficiently. The physical interpretability of the obtained pattern (i.e. in terms of independent physical processes) is, however, a point of discussion as the obtained modes are by definition orthogonal in space and time and this is not necessarily so in reality.

Consider the $n \times 1$ data vector \mathbf{y} , given for p time epochs t_i ,

$$\mathbf{y}_i = \begin{pmatrix} y_{1;i} \\ y_{2;i} \\ \vdots \\ y_{n;i} \end{pmatrix} \quad i = 1 \dots p. \quad (2.1)$$

Typically, \mathbf{y}_i contains the values of an observed or modelled field in n locations (the nodes of a two-dimensional grid or a set of discrete scattered observation sites; but the \mathbf{y}_i could also contain n spherical harmonic coefficients), at time t_i . We will assume that the data are centered, i.e. the time average per node $\frac{1}{p} \sum_{i=1}^p y_{j;i}$ is already reduced from the observations $y_{j;i}$, or

$$\frac{1}{p} \sum_{i=1}^p y_{j;i} = 0 \quad (2.2)$$

Another way to look at eq. (2.1) is to decompose the data vector $\mathbf{y}_i = \mathbf{I}\mathbf{y}_i$ according to the individual locations,

$$\mathbf{y}_i = y_{1;i} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + y_{2;i} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \dots + y_{n;i} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = y_{1;i} \mathbf{u}_1 + y_{2;i} \mathbf{u}_2 + \dots + y_{n;i} \mathbf{u}_n. \quad (2.3)$$

The basis vectors \mathbf{u}_j are independent of time, orthogonal, normalized with respect to the standard scalar product $(\mathbf{a}, \mathbf{b}) = \mathbf{a}^T \mathbf{b}$, and they are each associated with an individual location. One may interpret the original observations $y_{j;i}$ as coordinates in an ‘‘observation space’’ with regard to the trivial unit basis \mathbf{u}_j , in an n -dimensional vector space. Clearly, this interpretation suggests that other bases and other coordinates might be useful as well. The following will lead to a different choice of basis.

We collect all \mathbf{y}_i in the $n \times p$ data matrix \mathbf{Y} (assuming in what follows that the data is complete in the sense that for every location j there exists a data value $y_{j;i}$ for any epoch t_i). With other words, we assume for every location in the set there exists an uninterrupted time series of observations. The data matrix is then

$$\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) = \begin{pmatrix} y_{1;1} & y_{1;2} & \cdots & y_{1;p} \\ y_{2;1} & y_{2;2} & \cdots & y_{2;p} \\ \vdots & \vdots & & \vdots \\ y_{n;1} & y_{n;2} & \cdots & y_{n;p} \end{pmatrix}. \quad (2.4)$$

Its rows contain the time series per location, whereas its columns contain the entire data from all locations per time epoch.

We might be weighting the data matrix, e.g taking the individual accuracy of the data at different locations into account, or according to the latitude of the nodes. In this case, the homogenized data matrix becomes $\bar{\mathbf{Y}} = \mathbf{Y}\mathbf{G}$, where $\mathbf{G}\mathbf{G}^T = \mathbf{P}$ is the weight matrix.

The $n \times n$ signal covariance matrix \mathbf{C} contains the variances and covariances (i.e. second central moments) of the data viewed as time series per location. From the data samples \mathbf{y}_i , it can be estimated (empirically) as

$$\mathbf{C} = \frac{1}{p}\mathbf{Y}\mathbf{Y}^T = \frac{1}{p} \begin{pmatrix} \sum_{i=1}^p y_{1;i}^2 & \sum_{i=1}^p y_{1;i}y_{2;i} & \cdots & \sum_{i=1}^p y_{1;i}y_{n;i} \\ \sum_{i=1}^p y_{2;i}y_{1;i} & \sum_{i=1}^p y_{2;i}^2 & \cdots & \sum_{i=1}^p y_{2;i}y_{n;i} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^p y_{n;i}y_{1;i} & \sum_{i=1}^p y_{n;i}y_{2;i} & \cdots & \sum_{i=1}^p y_{n;i}^2 \end{pmatrix}, \quad (2.5)$$

or using the weighted matrix $\bar{\mathbf{Y}}$ instead. Note that the signal covariance matrix $\mathbf{C}' = \frac{1}{n}\mathbf{Y}^T\mathbf{Y}$, in contrast, contains the spatial variance and covariances of the data viewed as a function of position, for any t_i : there the sum extends over the n locations. Adding all the n individual variances from the time series provides what is often called the total variance,

$$\Delta^2 = \frac{1}{p} \sum_{j=1}^n \left(\sum_{i=1}^p y_{j;i}^2 \right) = \text{trace}(\mathbf{C}). \quad (2.6)$$

An alternative way to decompose the data vector is given by the eigenvalue decomposition of the signal covariance matrix \mathbf{C}

$$\mathbf{C} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T \quad (2.7)$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the n eigenvalues λ_i , and the columns of the orthogonal $n \times n$ matrix \mathbf{E} contain the corresponding eigenvectors \mathbf{e}_i . The sum of all eigenvalues equals to the matrix trace, and therefore to the total variance

$$\sum_{j=1}^n \lambda_j = \Delta^2. \quad (2.8)$$

We assume the eigenvalues and eigenvectors are ordered according to the magnitude of the eigenvalues; i.e. λ_1 is the largest one. Then, one can state that each eigenvalue ‘‘explains’’ a fraction

$$\eta_j = \frac{\lambda_j}{\Delta^2} \quad (2.9)$$

of the total variance, with the first eigenvalue explaining the largest part and so on. The eigenvalues of $\mathbf{C} = \frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ equal to $\frac{1}{\sqrt{p}}$ times the singular values of the data matrix \mathbf{Y} . The SVD of the data matrix can be written

$$\mathbf{Y} = \mathbf{E}\mathbf{\Delta}\bar{\mathbf{D}} , \quad (2.10)$$

where of course now

$$\lambda_j = \frac{1}{p}\Delta_j^2 \quad (2.11)$$

We will come back later to the $n \times n$ diagonal matrix $\mathbf{\Delta}$ and the $n \times p$ orthogonal matrix $\bar{\mathbf{D}}$.

Principle component analysis replaces the basis \mathbf{u}_j by the eigenvectors \mathbf{e}_j of \mathbf{C} as the vector basis for representing the original observations \mathbf{y}_i . One has to adopt a convention about the scaling of the eigenvectors, and in what follows we will assume they are normalized,

$$\mathbf{e}_j^T \mathbf{e}_j = 1 , \quad (2.12)$$

and their first entry is positive

$$e_{1;j} > 0 \quad (2.13)$$

just as it was the case for the original basis \mathbf{u}_j . In the same way as the \mathbf{u}_j can be associated with a discrete version of a delta function (they point exactly at the j -th data location with a value of one there, and zero values otherwise), the \mathbf{e}_j can be viewed as discrete version of a function which describes common pattern in the entire data. They are called empirical orthogonal functions (EOFs) or simply 'modes'. The first EOF \mathbf{e}_1 contains thus the dominant pattern (that is, if λ_1 is distinctly larger than the other eigenvalues). If the original data is provided on two-dimensional gridded locations, it is common to visualize the corresponding EOFs on this grid. Then, the principal component representation of the $n \times 1$ data vector at t_i , $i = 1, \dots, p$ is

$$\mathbf{y}_i = d_{1;i}\mathbf{e}_1 + d_{2;i}\mathbf{e}_2 + \dots + d_{n;i}\mathbf{e}_n = \sum_{j=1}^n d_{j;i}\mathbf{e}_j = \mathbf{E}\mathbf{d}_i \quad (2.14)$$

where the "principal components" (PCs) or PC scores $d_{j;i}$ are determined from projecting the original data onto the new basis

$$d_{j;i} = \mathbf{e}_j^T \mathbf{y}_i . \quad (2.15)$$

The $d_{j;i}$ can be viewed upon as time series, $i = 1 \dots p$, whereas the index j points at the pattern \mathbf{e}_j where the time series is associated with. Or,

$$\mathbf{d}_i = \mathbf{E}^T \mathbf{y}_i . \quad (2.16)$$

Since $\mathbf{E}^T \mathbf{E} = \mathbf{I}$, this can be written as $\mathbf{d}_i = (\mathbf{E}^T \mathbf{E})^{-1} \mathbf{E}^T \mathbf{y}_i$ as well. As the ordering is according to the magnitude of the eigenvalues, it is often sufficient to compute only a few, say \bar{n} , of the $d_{j;i}$. The reconstructed data will then still exhibit the largest part of the total variability:

$$\bar{\mathbf{y}}_i = \sum_{j=1}^{\bar{n}} d_{j;i} \mathbf{e}_j . \quad (2.17)$$

By this construction, EOFs constitute (normalized) spatial patterns whose amplitude evolution is given by the corresponding PC. The EOF itself does not change in time.

Remark. In other words, PCA decomposes the original data into time-invariant ('standing') spatial pattern, which are scaled by the corresponding time-variable PC. Therefore, PCA is not suitable for discovering *propagating pattern* in the data, since those will be distributed over several standing modes in the analysis.

Remark. Since the data are assumed as centered, one may say that PCA makes use of the second central moments of the data (only) to decorrelate them.

Remark. From the point of view of estimation theory, Eq. (2.5) assumes that the data are perfectly centered. In practice, one will probably compute and remove the sample mean of the time series. Then, in Eq. (2.5), one might use $\frac{1}{p-1}$ in place of $\frac{1}{p}$ in order to unbiasedly estimate the second central moments. It does *not* matter for the computation of the EOFs and the PCs, since the EOFs will be normalized (Eq. 2.12) anyway and the PCs follow from the normalized EOFs and the data.

Remark. The reconstructed data, Eq. (2.17) can be expressed by

$$\bar{\mathbf{y}}_i = \mathbf{E} \mathbf{I}^{(\bar{n})} \mathbf{E}^T \mathbf{y}_i$$

where $\mathbf{I}^{(\bar{n})}$ is a diagonal matrix with unity in the first \bar{n} entries and zero otherwise, I.e., decomposition and partial reconstruction can be viewed as a linear operation (in first order at least).

From Eq. (2.14), it is clear that the data matrix \mathbf{Y} is referred to the EOFs by

$$\mathbf{Y} = \mathbf{E} \mathbf{D} , \quad (2.18)$$

where the rows of \mathbf{D} now contain the PCs for all EOFs (e.g., the first row contains the temporal evolution of the first EOF), and the columns of \mathbf{D} contain the PC vectors \mathbf{d}_i (each vector contains the temporal amplitude of all EOFs for one particular epoch). With other words, we write

$$\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_p) = \begin{pmatrix} d_{1;1} & d_{1;2} & \dots & d_{1;p} \\ d_{2;1} & d_{2;2} & \dots & d_{2;p} \\ \vdots & \vdots & & \vdots \\ d_{n;1} & d_{n;2} & \dots & d_{n;p} \end{pmatrix}.$$

Then, for the total variance

$$\Delta^2 = \text{trace}(\mathbf{E}\mathbf{D}\mathbf{D}^T\mathbf{E}^T) = \text{trace}(\mathbf{D}^T\mathbf{D}) = \sum_{j=1}^n \sum_{i=1}^p d_{j;i}^2. \quad (2.19)$$

The aim of PCA is to find a linear combination of the original data nodes that explains the maximum variability (variance) of the data. This means, we are searching for the mode \mathbf{e} such that $\mathbf{Y}\mathbf{e}$ has maximum variance. The variance of the centered time series $\mathbf{Y}\mathbf{e}$ is

$$\frac{1}{p}(\mathbf{Y}\mathbf{e})^T(\mathbf{Y}\mathbf{e}) = \frac{1}{p}\mathbf{e}^T\mathbf{C}\mathbf{e}. \quad (2.20)$$

Usually we require \mathbf{e} to be normalized. The task is then to maximize Eq. (2.20) subject to $\mathbf{e}^T\mathbf{e} = 1$. The solution to this problem is the eigenvalue problem $\mathbf{C}\mathbf{e} = \lambda\mathbf{e}$, with eigenvectors \mathbf{e}_i and eigenvalues λ_i as introduced earlier.

However, the data vectors \mathbf{y}_i will contain a random error, and such will the eigenvalues and eigenvectors derived from the data matrix. This has to be considered in particular if eigenvalues are close to each other.

2.1.2 Temporal PCA versus Spatial PCA

PCA as described above is sometimes called *temporal PCA*, since it departs from the correlations between time series of data (which are contained in the $n \times n$ covariance matrix \mathbf{C}). On the other hand, it is perfectly valid to consider, for the same data set, the spatial correlations and build the $p \times p$ spatial covariance matrix $\mathbf{C}' = \frac{1}{n}\mathbf{Y}^T\mathbf{Y}$, or

$$\mathbf{C}' = \frac{1}{n}\mathbf{Y}^T\mathbf{Y} = \frac{1}{n} \begin{pmatrix} \sum_{j=1}^n y_{j;1}^2 & \sum_{j=1}^n y_{j;1}y_{j;2} & \dots & \sum_{j=1}^n y_{j;1}y_{j;p} \\ \sum_{j=1}^n y_{j;2}y_{j;1} & \sum_{j=1}^n y_{j;2}^2 & \dots & \sum_{j=1}^n y_{j;2}y_{j;p} \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^n y_{j;p}y_{j;1} & \sum_{j=1}^n y_{j;p}y_{j;2} & \dots & \sum_{j=1}^n y_{j;p}^2 \end{pmatrix}. \quad (2.21)$$

In fact, if $p \ll n$, storing \mathbf{C}' requires much less memory space compared to storing \mathbf{C} .

PCA based upon \mathbf{C}' is called spatial PCA. Of course, temporal and spatial PCA are closely related: \mathbf{C} and \mathbf{C}' are of different dimension but they share the same eigenvalues (apart from a factor that depends only on n and p).

An eigenvalue decomposition (and comparison with the decomposition of \mathbf{C}) reveals

$$\mathbf{C}' = \frac{p}{n} \bar{\mathbf{D}}^T \mathbf{\Lambda} \bar{\mathbf{D}} \quad (2.22)$$

where we have $\bar{\mathbf{D}} = \mathbf{\Delta}^{-1} \mathbf{E}^T \mathbf{Y} = \mathbf{\Delta}^{-1} \mathbf{D}$.

It is thus obvious that the k -th EOF of the spatial PCA (k -th column of $\bar{\mathbf{D}}^T$) corresponds to the k -th PCs of the temporal PCA. Alternatively, this can be seen as follows: From

$$\mathbf{C} \mathbf{e}_j = \lambda_j \mathbf{e}_j$$

follows

$$\frac{n}{np} \mathbf{Y}^T \mathbf{Y} \mathbf{Y}^T \mathbf{e}_j = \lambda_j \mathbf{Y}^T \mathbf{e}_j$$

and the eigenvectors of \mathbf{C}' can be read off as $\mathbf{Y}^T \mathbf{e}_j$. Thus

$$\mathbf{E}' = \mathbf{Y}^T \mathbf{E} = \mathbf{D} = \mathbf{\Delta} \bar{\mathbf{D}} .$$

2.1.3 PCA of Linearly Transformed Data

It is interesting to consider the PCA of a set of linearly transformed $m \times 1$ data vectors

$$\mathbf{z}_i = \mathbf{A} \mathbf{y}_i, \quad i = 1 \dots p \quad (2.23)$$

with $m \times n$ matrix \mathbf{A} . Again p is the number of time epochs. The number of data nodes m might be larger, equal or less than n .

Example. The original data might contain spherical harmonic coefficients of a field, and the transformed data contain gridded values. In this case $m > n$ is not uncommon. Matrix \mathbf{A} contains the spherical harmonics for each given coefficient evaluated for each grid node.

Example. The original data contain values on a global grid of certain spacing. We ask in how far the EOFs and PCs on a local subgrid, i.e. for some region of the globe, will differ from those evaluated from the global data set. In this case, $m < n$ and the matrix \mathbf{A} equals to the identity matrix, with its rows removed for all nodes that are not present in the local subgrid.

Obviously the transformed data matrix is $\mathbf{Z} = \mathbf{A} \mathbf{Y}$. Furthermore we have

$$\mathbf{C}_z = \frac{1}{p} \mathbf{Z} \mathbf{Z}^T = \frac{1}{p} \mathbf{A} \mathbf{Y} \mathbf{Y}^T \mathbf{A}^T = \mathbf{A} \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T \mathbf{A}^T \quad (2.24)$$

where \mathbf{E} and $\mathbf{\Lambda}$ contain the eigenvectors and eigenvalues of the original data covariance matrix. Obviously, the eigenvectors and eigenvalues of \mathbf{C}_z will differ from those of \mathbf{C} , meaning that both the EOFs and the PCs of the transformed data will differ from those of the original data (unless in some special cases).

Let μ_i be the eigenvalues of $\mathbf{A}^T \mathbf{A}$. For the eigenvalues of $\mathbf{C}_z = \mathbf{A} \mathbf{C} \mathbf{A}^T$, which equal to the eigenvalues of $\mathbf{C} \mathbf{A}^T \mathbf{A}$, the following inclusion holds ([5])

$$\lambda_i^z \in [\min(\mu_i) \cdot \min(\lambda_i), \max(\mu_i) \cdot \max(\lambda_i)] \quad (2.25)$$

This illustrates clearly, how the spectrum of the transformed data is widened by the spectrum of $\mathbf{A}^T \mathbf{A}$.

2.1.4 PCA as a Data Whitening Method: Homogeneization

Obviously, one can interpret the PCs as a 'whitened' version of the original data. To make this clear, we will consider instead of

$$\mathbf{d}_i = \mathbf{E}^T \mathbf{y}_i \quad (2.26)$$

the homogenized PCs $\bar{d}_{j;i} = \frac{1}{\sqrt{\lambda_i}} d_{j;i}$, or

$$\bar{\mathbf{d}}_i = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{d}_i = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \mathbf{y}_i = \bar{\mathbf{E}}^T \mathbf{y}_i . \quad (2.27)$$

Here, we have introduced the column-by-column scaled matrix $\bar{\mathbf{E}} = \mathbf{E} \mathbf{\Lambda}^{-\frac{1}{2}}$.

Remark. It is clear by now that the homogenized PCs $\bar{\mathbf{d}}_i$ are the column vectors of the SVD matrix $\bar{\mathbf{D}}$.

The scaled EOFs are not of unit length anymore, but still orthogonal,

$$\bar{\mathbf{E}}^T \bar{\mathbf{E}} = \mathbf{\Lambda}^{-1} . \quad (2.28)$$

The signal covariance matrix of the original data \mathbf{y}_i is $\mathbf{C} = \mathbf{E} \mathbf{\Lambda} \mathbf{E}^T$, thus the covariance of the PCs will be

$$\mathbf{C}_d = \mathbf{E}^T \mathbf{E} \mathbf{\Lambda} \mathbf{E} \mathbf{E}^T = \mathbf{\Lambda} , \quad (2.29)$$

or, for clarity,

$$\sum_{i=1}^p d_{j;i}^2 = \lambda_j .$$

And the signal covariance of the homogenized PCs will be

$$\mathbf{C}_{\bar{d}} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}^T \mathbf{E} \mathbf{\Lambda} \mathbf{E} \mathbf{E}^T \mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I} . \quad (2.30)$$

From the last two expressions, it is obvious that the PCs and the homogenized PCs are uncorrelated, with the latter ones also being of unit variance. Therefore, the (homogenized) PCA is often viewed as a data whitening transformation. The homogenized EOFs are directly obtained from applying the rescaling to the original eigenvectors

$$\bar{\mathbf{e}}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{e}_j , \quad (2.31)$$

of the data.

2.1.5 Number of Modes

In many applications of PCA, we will avoid to retain all n modes, but rather use a subset of \bar{n} *dominant* ones. The reasoning can be different: We may want to compress the data, or we may want to get rid of those modes that supposedly contain noise. Or, PCA is just considered as a preprocessing and we will subsequently apply e.g. rotation on the dominant modes. Let $\bar{\mathcal{J}} = \{j_1, j_2, \dots, j_{\bar{n}}\}$ denote the index set of all modes to be retained, i.e.

$$\bar{\mathbf{y}}_i = \sum_{j \in \bar{\mathcal{J}}} d_{j;i} \mathbf{e}_j .$$

A rule that determines $\bar{\mathcal{J}}$ is called a *selection rule*.

It has been suggested by Eq. (2.9) that each eigenvalue of the data covariance explains a certain fraction of the total variance Δ^2 , Eq. (2.6). This indicates that the strategy to choose a reasonable subset of modes could simply be

$$\bar{\mathcal{J}} = \{j \mid \sum_{j \in \bar{\mathcal{J}}} \eta_j > \epsilon\} .$$

This strategy is by far the most often followed one, with a typical threshold value of 0.9.

A selection rule (*North's rule*) that is often considered goes back to [18]. It is based on the perception that the data \mathbf{y}_i represent independent realizations or samples of a random field with unknown stochastic moments. From these realizations, one will be able to reconstruct the *true* covariance \mathbf{C}' only up to an error that depends on \mathbf{C}' and the number n of data realizations. With other words, \mathbf{C} as computed through Eq. (2.5) will be considered as a stochastic quantity being contaminated by an error whose covariance can be estimated from \mathbf{C} and n . Therefore, the eigenvalues and eigenvectors of \mathbf{C} have to be considered as stochastic as well. [18] proceed to show that 'typical' errors of neighbouring eigenvalues and eigenvectors will then be

$$\delta\lambda_j = \sqrt{\frac{2}{n}} \lambda_j + \dots \quad \delta\mathbf{e}_j = \frac{\delta\lambda_j}{\lambda_k - \lambda_j} \mathbf{e}_k + \dots$$

'Neighbouring' means that λ_k is the eigenvalue numerically closest to λ_j . This selection rule says that if the 'typical' error of an eigenvalue is comparable to the difference of this eigenvalue to its neighbour, then the 'typical' error of the corresponding EOF will be of the size of the neighbouring EOF itself. One will then tend to disregard this mode in the reconstruction. Or,

$$\bar{\mathcal{J}} = \{j \mid \delta\lambda_j < |\lambda_k - \lambda_j| = \min_{i \neq j} |\lambda_i - \lambda_j|\} .$$

Several other selection rules have been proposed since then, based on different principles. More recently, Monte Carlo methods have been applied frequently to test the statistical significance of modes.

2.1.6 PCA as a Tool for Comparing Multiple Data Sets

We are often interested in comparing multiple data sets, e.g. satellite-derived vs. modelled, or different model output data sets. Several statistical algorithms allow to derive correlation measures, similarities and joint pattern and so on. Here, we will only focus on the application of the PCA as described before in such a situation.

Consider the $n \times 1$ vector \mathbf{y} , given for p time epochs t_i , and extracted from M different data sets, or

$$\mathbf{y}_i^{(m)} = \begin{pmatrix} y_{1;i}^{(m)} \\ y_{2;i}^{(m)} \\ \vdots \\ y_{n;i}^{(m)} \end{pmatrix} \quad i = 1 \dots p, \quad m = 1 \dots M, \quad (2.32)$$

which we may recast in a 'super data matrix'

$$\mathbf{X} = \left(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(m)} \right). \quad (2.33)$$

If all data vectors are considered as equally good, bias-free (i.e. centered free of errors), and describing the same phenomena apart from unavoidable data/model errors, i.e. as independent realizations of the same data vector, one may simply compute the covariance matrix

$$\mathbf{C} = \frac{1}{pM} \mathbf{X} \mathbf{X}^T \quad (2.34)$$

and go on as described before.

If we suspect that different sensors or models see different phenomena, which is to say the data are not coming from the same p.d.f., one may of course apply PCA on each data set independently. This provides M covariance matrices $\mathbf{C}^{(m)}$. A comparison is then hampered by the fact that each data set will be represented in its own basis $\mathbf{e}_j^{(m)}$. To facilitate comparison, one may project all data sets onto the basis derived from \mathbf{C} or from one of the data sets (maybe the one we trust most), say. from $\mathbf{C}^{(m^*)}$. This is, we compare the data sets on the level of principle componets with a joint basis,

$$\mathbf{d}_i^{(m)} = \mathbf{E}^T \mathbf{y}_i^{(m)} \quad (2.35)$$

or

$$\mathbf{d}_i^{(m)} = \mathbf{E}^{(m^*)T} \mathbf{y}_i^{(m)}. \quad (2.36)$$

2.1.7 Rotation

Rotated EOF is a technique which attempts to overcome some common shortcomings of PCA. For example, the *mathematical* constraints (orthogonality of EOFs *and* uncorrelatedness of PCs) of PCA, in connection with the dependence of the computation domain (see 'PCA of Linearly Transformed Data') may render the modes found in data difficult to interpret. *Physical modes* may not necessarily be orthogonal and thus leak into several different mathematical modes in PCA. REOF is a technique which sacrifices either orthogonality of the EOFs *or* uncorrelatedness of the PCs, while adding new optimization criteria that seek to find physically plausible modes.

Rotated homogenized EOFs

An understanding of the idea of REOF starts with the observation that, viewed as a whitening transformation, PCA with the basis vectors $\bar{\mathbf{e}}_j$ is *not unique*. To see this, the data vectors \mathbf{y}_i , with covariance \mathbf{C} are expressed by

$$\mathbf{y}_i = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\bar{\mathbf{d}}_i . \quad (2.37)$$

It is possible to replace the $\bar{\mathbf{d}}_i$ by any set of $i = 1, \dots, p$ rotated $n \times 1$ homogenized PCs,

$$\bar{\mathbf{r}}_i = \mathbf{V}\bar{\mathbf{d}}_i \quad (2.38)$$

with $n \times n$ orthogonal matrix \mathbf{V} , i.e. $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. Then,

$$\mathbf{C}_{\bar{\mathbf{r}}} = \mathbf{V}\mathbf{V}^T = \mathbf{I} . \quad (2.39)$$

We have

$$\bar{\mathbf{d}}_i = \mathbf{V}^T\bar{\mathbf{r}}_i \quad (2.40)$$

and

$$\mathbf{y}_i = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T\bar{\mathbf{r}}_i = \bar{\mathbf{E}}\boldsymbol{\Lambda}\mathbf{V}^T\bar{\mathbf{r}}_i . \quad (2.41)$$

It is obvious that the data covariance $\mathbf{C} = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T(\mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T)^T = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^T$ does not depend on \mathbf{V} . Hence, the transform $\mathbf{y}_i = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T\bar{\mathbf{r}}_i = \bar{\mathbf{E}}\boldsymbol{\Lambda}\mathbf{V}^T\bar{\mathbf{r}}_i$ with rescaled and rotated PCs whitens the data as good as the original homogenized PCs. The rotated basis vectors (or rotated EOFs) are now the column vectors of $\bar{\mathbf{F}} = \mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T = \bar{\mathbf{E}}\boldsymbol{\Lambda}\mathbf{V}^T$.

We have seen in Eq. (2.40) that the rotated homogenized PCs have diagonal and equal covariance, just as the original homogenized PCs,

$$\mathbf{C}_{\bar{\mathbf{r}}} = \mathbf{C}_{\bar{\mathbf{d}}} = \mathbf{I} .$$

The PCs, viewed as time series per EOF, are uncorrelated and they do not lose this property when an arbitrary orthogonal rotation is applied to the EOFs. The rotated homogenized EOFs, however, will not be orthogonal anymore,

$$\bar{\mathbf{F}}^T\bar{\mathbf{F}} = \mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{E}^T\mathbf{E}\boldsymbol{\Lambda}^{\frac{1}{2}}\mathbf{V}^T = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T . \quad (2.42)$$

Rotated EOFs

On the other hand, one can define rotated EOFs by straight application of an orthogonal matrix \mathbf{V} to the EOFs \mathbf{E} ,

$$\mathbf{F} = \mathbf{E}\mathbf{V}^T \quad (2.43)$$

i.e. without homogeneizing the PCs first. The data is then represented through rotated PCs,

$$\mathbf{y}_i = \mathbf{F}^T \mathbf{r}_i . \quad (2.44)$$

In this case, the rotated EOFs remain orthogonal, since

$$\mathbf{F}^T \mathbf{F} = \mathbf{V} \mathbf{E}^T \mathbf{E} \mathbf{V}^T = \mathbf{I} . \quad (2.45)$$

But now, the $i = 1, \dots, p$ rotated PCs

$$\mathbf{r}_i = \mathbf{V} \mathbf{d}_i = \mathbf{F}^T \mathbf{y}_i \quad (2.46)$$

lose the property of being uncorrelated since

$$\mathbf{C}_r = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T . \quad (2.47)$$

In summary, by rotation either the orthogonality of the EOFs or the uncorrelatedness of the PCs will be destroyed.

Rotation principles

So far, nothing has been said regarding the particular choice of an orthogonal matrix \mathbf{V} in EOF and PC rotation. All orthogonal \mathbf{V} are able to reproduce the data, whereas only for $\mathbf{V} = \mathbf{I}$ both orthogonality in space and time can be preserved. Which one (in space or time) we sacrifice by rotation, depends upon application to homogeneized or original EOFs and PCs.

In REOF, one usually specifies an optimization criterion $\mathcal{F}(\mathbf{V})$ in terms of rotated EOFs or rotated PCs, to be met subject to the condition $\mathbf{V}\mathbf{V}^T = \mathbf{I}$. In other words, an orthogonal $n \times n$ matrix has $\frac{n(n-1)}{2}$ degrees of freedom and these have to be chosen such as to optimize $\mathcal{F}(\mathbf{V})$.

When we have

$$\mathbf{F} = \mathbf{E}\mathbf{V}^T$$

with elements $f_{j;i}$ of the j th rotated EOF, the following family of VARIMAX criteria is in use

$$\mathcal{F}(\mathbf{V}) = \sum_{i=1}^n \left(\sum_{j=1}^n f_{j;i}^4 - \frac{\gamma}{n} \left(\sum_{j=1}^n f_{j;i}^2 \right)^2 \right) . \quad (2.48)$$

The quantity inside the summation is proportional to the variance of the square of the rotated EOFs \mathbf{f}_j (for $\gamma = 1$). This variance will be big if some values $f_{j;i}$ are close to 1 and many are near 0. Consequently, it is often claimed that the varimax rotation attempts to 'simplify' the patterns by localizing the 'regions of action'.

In practice, one will rotate only the first \bar{n} EOFs corresponding to the largest singular values, then the above reads

$$\mathcal{F}(\mathbf{V}) = \sum_{i=1}^{\bar{n}} \left(\sum_{j=1}^n f_{j;i}^4 - \frac{\gamma}{n} \left(\sum_{j=1}^n f_{j;i}^2 \right)^2 \right). \quad (2.49)$$

2.2 Independent Component Analysis

We follow [12]. Consider orthogonal EOF rotation with homogenized PCs, i.e

$$\bar{\mathbf{r}}_i = \mathbf{V}\bar{\mathbf{d}}_i \quad (2.50)$$

for $i = 1, \dots, p$ time steps. Collecting the $n \times 1$ vectors of homogenized PCs in $n \times$ matrices $\bar{\mathbf{D}}$ and $\bar{\mathbf{R}}$, this is

$$\bar{\mathbf{R}} = \mathbf{V}\bar{\mathbf{D}}, \quad (2.51)$$

and the rotated EOFs will be

$$\bar{\mathbf{F}} = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{V}. \quad (2.52)$$

For any orthogonal \mathbf{V} the rotated homogenized PCs $\bar{\mathbf{r}}_i$ are uncorrelated and of unit variance, i.e. as a time series in i

$$\sum_{i=1}^p \bar{r}_{j;i}^2 = 1 \quad j = 1, \dots, n$$

$$\sum_{i=1}^p \bar{r}_{j;i}\bar{r}_{k;i} = 0 \quad j \neq k$$

In [12] it is suggested to choose \mathbf{V} such that the $\bar{\mathbf{r}}_i$ are close to being *independent*.

Independence is stronger than uncorrelatedness, and defining (and testing) it requires to involve higher moments of the pdf of the $\bar{r}_{j;i}$. Different criteria are in use in the literature on *Independent Component Analysis (ICA)*.

The line of reasoning in [12] is as follows. If $\bar{r}_{j;i}$ and $\bar{r}_{k;i}$ are independent, then the time series of the squares $\bar{r}_{j;i}^2, \bar{r}_{k;i}^2$ should be uncorrelated (after centering), or

$$\sum_{i=1}^p \left(\bar{r}_{j;i}^2 - \frac{1}{p} \sum_{l=1}^p \bar{r}_{j;l}^2 \right) \left(\bar{r}_{k;i}^2 - \frac{1}{p} \sum_{l=1}^p \bar{r}_{j;l}^2 \right) = 0 \quad j \neq k .$$

This can be written in matrix notation. Let \odot denote the Hadamard matrix product, i.e.

$$\bar{\mathbf{R}} \odot \bar{\mathbf{R}} = \begin{pmatrix} r_{1;1}^2 & r_{1;2}^2 & \cdots & r_{1;p}^2 \\ r_{2;1}^2 & r_{2;2}^2 & \cdots & r_{2;p}^2 \\ \vdots & \vdots & & \vdots \\ r_{n;1}^2 & r_{n;2}^2 & \cdots & r_{n;p}^2 \end{pmatrix}$$

and let $\mathbf{H} = \mathbf{H}^2$ be the $p \times p$ centering matrix (with $\mathbf{i} = (1, 1, \dots, 1)^T$)

$$\mathbf{H} = \mathbf{I} - \frac{1}{p} \mathbf{i} \mathbf{i}^T .$$

Then, for independent time series $\bar{r}_{j;i}$ the (empirical) covariance matrix of the centered squares

$$\mathbf{C}_{\mathbf{r}^2} = \frac{1}{p} ((\bar{\mathbf{R}} \odot \bar{\mathbf{R}}) \mathbf{H}) ((\bar{\mathbf{R}} \odot \bar{\mathbf{R}}) \mathbf{H})^T = \frac{1}{p} (\bar{\mathbf{R}} \odot \bar{\mathbf{R}}) \mathbf{H} (\bar{\mathbf{R}} \odot \bar{\mathbf{R}})^T \quad (2.53)$$

must be diagonal.

In other words, an ICA approach can be constructed by defining an objective function $\mathcal{F}(\mathbf{V})$ that penalizes off-diagonal elements of $\mathbf{C}_{\mathbf{r}^2}$. ICA will then seek a rotation matrix \mathbf{V} through optimization of $\mathcal{F}(\mathbf{V})$.

Remark. The above idea ([12]) makes use of fourth statistical moments, but other moments may be used for defining an objective function as well.

Appendix

3.1 Spherical Harmonics

Spherical harmonic series

It is common to represent real-valued phenomena on the sphere as spherical harmonic series

$$F(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=0}^n (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\cos \theta) \quad (3.1)$$

with longitude λ , colatitude θ , the spherical harmonic degree n and order m , where $n \geq m \geq 0$, the spherical harmonic coefficients C_{nm} and S_{nm} , and the associated Legendre functions of the first kind P_{nm} .

Legendre polynomials and associated Legendre functions

The associated Legendre functions of degree n and order m , $n \geq m \geq 0$, can be expressed through the m -th derivatives of the Legendre polynomials of degree n , $P_n = P_{n0}$, with respect to $t = \cos \theta$,

$$P_{nm}(t) = (1 - t^2)^{m/2} \frac{d^m P_n(t)}{dt^m}, \quad (3.2)$$

which may be written as

$$P_{nm}(\cos \theta) = \sin^m \theta \frac{d^m P_n(\cos \theta)}{d(\cos \theta)^m}. \quad (3.3)$$

They fulfill the differential equation

$$(1 - t^2) \frac{d^2 P_{nm}}{dt^2} - 2t \frac{dP_{nm}}{dt} + \left(n(n+1) - \frac{m^2}{1-t^2} \right) P_{nm} = 0 \quad (3.4)$$

or

$$\frac{d}{d\theta} \left(\sin \theta \frac{dP_{nm}}{d\theta} \right) + \left(n(n+1) \sin \theta - \frac{m^2}{\sin \theta} \right) P_{nm} = 0. \quad (3.5)$$

Note that sometimes (e.g. [3]) the associated Legendre functions are defined as $P_n^m = (-1)^m P_{nm}$. The Rodrigues formula expresses the Legendre polynomials P_n of degree n through the n -th derivatives of $(1-t^2)^n = \sin^{2n} \theta$,

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n (t^2 - 1)^n}{dt^n} \quad (3.6)$$

they satisfy the differential equation

$$n(n+1)P_n - 2t \frac{dP_n}{dt} + (1-t^2) \frac{d^2 P_n}{dt^2} = 0 \quad (3.7)$$

or

$$n(n+1)P_n + \frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{dP_n}{d\theta} \right) = 0. \quad (3.8)$$

An expansion of the Legendre polynomials and associated Legendre functions

Table 3.1. Legendre polynomials and associated Legendre functions

n	m	P_{nm}
0	0	1
1	0	$\cos \theta$
1	1	$\sin \theta$
2	0	$\frac{1}{2}(3 \cos^2 \theta - 1) = \frac{1}{4}(3 \cos 2\theta + 1)$
2	1	$3 \sin \theta \cos \theta = \frac{3}{2} \sin 2\theta$
2	2	$3 \sin^2 \theta$
3	0	$\frac{1}{2}(5 \cos^3 \theta - 3 \cos \theta) = \frac{1}{8}(5 \cos 3\theta + 3 \cos \theta)$
3	1	$\sin \theta \left(\frac{15}{2} \cos^2 \theta - \frac{3}{2} \right) = \frac{3}{4} \sin \theta (5 \cos 2\theta + 3)$
3	2	$15 \sin^2 \theta \cos \theta = \frac{15}{2} \sin \theta \sin 2\theta$
3	3	$15 \sin^3 \theta$

into trigonometric series reads

$$P_{nm}(\cos \theta) = \sin^m \theta \sum_{q=0}^{\text{int}(\frac{n-m}{2})} T_{nmq} \cos^{n-m-2q} \theta, \quad (3.9)$$

where $\text{int}(x)$ means the integer part of x , and the coefficients T_{nmq} are given by ([10],[9])

$$T_{nmq} = \frac{(-1)^q (2n-2q)!}{2^n q! (n-q)! (n-m-2q)!}. \quad (3.10)$$

Relations (3.2) and (3.2) can be combined to

$$P_{nm}(t) = \frac{(1-t^2)^{m/2}}{2^n n!} \frac{d^{n+m}(t^2-1)^n}{dt^{n+m}}. \quad (3.11)$$

This is being used to define associate Legendre functions P_{nm} of negative order m ; $0 > m \geq -n$. The relation between P_{nm} and $P_{n,-m}$ is ([8])

$$P_{n,-m}(t) = (-1)^m \frac{(n-m)!}{(n+m)!} P_{nm} \quad (3.12)$$

$$P_{nm}(t) = (-1)^m \frac{(n+m)!}{(n-m)!} P_{n,-m}. \quad (3.13)$$

Alternative notations for the real-valued spherical harmonic series

There are $2n+1$ spherical harmonics of degree n . Another way to write eq. (3.1) is

$$F(\lambda, \theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^n f_{nm} Y_{nm}(\lambda, \theta) \quad (3.14)$$

with $f_{nm} = C_{nm}$ for $m \geq 0$, $f_{nm} = S_{n|m|}$ for $m < 0$, and

$$Y_{nm}(\lambda, \theta) = Y_{nm1}(\lambda, \theta) = \cos m\lambda P_{nm}(\cos \theta) \quad m \geq 0 \quad (3.15)$$

$$Y_{nm}(\lambda, \theta) = Y_{n|m|2}(\lambda, \theta) = \sin |m|\lambda P_{n|m|}(\cos \theta) \quad m < 0. \quad (3.16)$$

Integration over the unit sphere

The spherical harmonics Y_{nm} are orthogonal on the unit sphere Ω . Integrating products of spherical harmonics Y_{nm} yields

$$\int_{\Omega} Y_{nm} Y_{n'm'} d\omega = 4\pi \frac{1}{H_{nm}^2} \delta_{nn'} \delta_{mm'} \quad (3.17)$$

with

$$H_{nm} = \sqrt{(2 - \delta_{0m})(2n+1) \frac{(n-m)!}{(n+m)!}}, \quad (3.18)$$

in particular

$$H_{n0} = \sqrt{(2n+1)}. \quad (3.19)$$

Consequently,

$$\int_{\Omega} Y_{nm} d\omega = 4\pi \delta_{n0} \delta_{m0}. \quad (3.20)$$

Integrals over various products of derivatives of spherical harmonics can be found in [13].

4 π - or fully normalized spherical harmonics

It is common in geodesy to introduce 4 π - or fully normalized associated Legendre functions

$$\bar{P}_{nm} = \Pi_{nm} P_{nm} . \quad (3.21)$$

The relation between the \bar{P}_{nm} of positive order, $n \geq m \geq 0$ and those of negative order, $P_{n,-m}$, is

$$\bar{P}_{n,-m}(t) = (-1)^m \bar{P}_{nm} \quad (3.22)$$

$$\bar{P}_{nm}(t) = (-1)^m \bar{P}_{n,-m} . \quad (3.23)$$

Using the \bar{P}_{nm} of positive order, we introduce 4 π - or fully normalized spherical harmonics

$$\bar{Y}_{nm} = \Pi_{nm} Y_{nm} \quad (3.24)$$

or

$$\bar{Y}_{nm}(\lambda, \theta) = \cos m\lambda \bar{P}_{nm}(\cos \theta) \quad m \geq 0 \quad (3.25)$$

$$\bar{Y}_{nm}(\lambda, \theta) = \sin |m|\lambda \bar{P}_{n|m|}(\cos \theta) \quad m < 0 .$$

with spherical harmonic coefficients $\bar{C}_{nm} = \frac{1}{\Pi_{nm}} C_{nm}$, $\bar{S}_{nm} = \frac{1}{\Pi_{nm}} S_{nm}$, or \bar{f}_{nm} , \bar{f}_{nm1} , \bar{f}_{nm2} accordingly. By definition, these fully normalized spherical harmonics fulfill

$$\int_{\Omega} \bar{Y}_{nm} \bar{Y}_{n'm'} d\omega = 4\pi \delta_{nn'} \delta_{mm'} . \quad (3.26)$$

The addition theorem relates fully (4 π -) normalized spherical harmonics and the (un-normalized) Legendre polynomials

$$\frac{1}{2n+1} \sum_{m=-n}^n \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{nm}(\lambda', \theta') = P_n(\cos \psi) . \quad (3.27)$$

In particular,

$$\frac{1}{2n+1} \sum_{m=-n}^n \bar{Y}_{nm}^2(\lambda, \theta) = 1 . \quad (3.28)$$

Practical computation of the fully normalized spherical harmonics

In practice, the fully normalized associated Legendre functions $\bar{P}_{nm}(\cos \theta)$ are computed via recursion relations.

Example. One of the most often applied recursive algorithms for the normalized Legendre functions as a function of co-latitude θ is the following

```

    c = cos θ
    s = sin θ
    P̄00 = 1
    P̄11 = √3 · s
    do n = 2, n̄
        an = √((2n+1)/2n)
        P̄nm = an · s · P̄n-1 n-1
    end do
    do n = 1, n̄
        bn = √(2n+1)
        P̄nn-1 = bn · c · P̄n-1 n-1
    end do
    do n = 2, n̄
        do m = n, 0, -1
            cn = √((2n+1)/((n-m)(n+m)))
            dn = √(2n-1)
            en = √(((n-m-1)(n+m-1))/(2n-3))
            P̄nm = cn · (dn · c · P̄n-1 m - en · P̄n-2 m)
        end do
    end do
end do

```

Normalized complex spherical harmonics

Normalized complex spherical harmonics are introduced in different ways. Following e.g. [8] and using associated Legendre functions of positive and negative order, $n \geq m \geq -n$

$$\begin{aligned}
 \bar{Y}_{nm} &= \frac{(-1)^m}{\sqrt{4\pi}} \Xi_{nm} (\cos m\lambda + i \sin m\lambda) P_{nm}(\cos \theta) \\
 &= \frac{(-1)^m}{\sqrt{4\pi}} \Xi_{nm} \underbrace{e^{im\lambda} P_{nm}(\cos \theta)}_{\mathcal{Y}_{nm}}
 \end{aligned} \tag{3.29}$$

where

$$\Xi_{nm} = \sqrt{(2n+1) \frac{(n-m)!}{(n+m)!}} = \frac{I_{nm}}{\sqrt{2 - \delta_{0m}}} . \tag{3.30}$$

Here

$$\bar{\mathcal{Y}}_{nm} = (-1)^m \bar{\mathcal{Y}}_{n,-m}^* \quad \bar{\mathcal{Y}}_{nm}^* = (-1)^m \bar{\mathcal{Y}}_{n,-m} \quad (3.31)$$

follows from eq. (3.23) and

$$\begin{aligned} & \Xi_{nm} P_{nm}(\cos \theta) (\cos m\lambda + i \sin m\lambda) \\ &= (-1)^m \Xi_{n,-m} P_{n,-m}(\cos \theta) (\cos m\lambda - i \sin(-m\lambda)) . \end{aligned}$$

Consequently, in place of eq. (3.29) we could write

$$\begin{aligned} \bar{\mathcal{Y}}_{nm} &= \frac{(-1)^m}{\sqrt{4\pi}} \Xi_{nm} (\cos m\lambda + i \sin m\lambda) P_{nm}(\theta) \quad m \geq 0 \quad (3.32) \\ &= (-1)^m \bar{\mathcal{Y}}_{n|m|}^* \quad m < 0 . \end{aligned}$$

This is to relate complex spherical harmonics of negative order to associated Legendre functions of positive order. The $\bar{\mathcal{Y}}_{nm}$ are 1-normalized, thus

$$\int_{\Omega} \bar{\mathcal{Y}}_{nm} \bar{\mathcal{Y}}_{n'm'}^* d\omega = \delta_{nn'} \delta_{mm'} . \quad (3.33)$$

And,

$$\sum_{m=-n}^n \bar{\mathcal{Y}}_{nm}(\lambda, \theta) \bar{\mathcal{Y}}_{nm}^*(\lambda', \theta') \quad (3.34)$$

$$\begin{aligned} &= \bar{\mathcal{Y}}_{n0}(\lambda, \theta) \bar{\mathcal{Y}}_{n0}^*(\lambda', \theta') + \sum_{m=1}^n (\bar{\mathcal{Y}}_{nm}(\lambda, \theta) \bar{\mathcal{Y}}_{nm}^*(\lambda', \theta') + \bar{\mathcal{Y}}_{nm}^*(\lambda, \theta) \bar{\mathcal{Y}}_{nm}(\lambda', \theta')) \\ &= \frac{1}{4\pi} \sum_{m=-n}^n \bar{Y}_{nm}(\lambda, \theta) \bar{Y}_{nm}(\lambda', \theta') = \frac{2n+1}{4\pi} P_n(\cos \psi) . \end{aligned} \quad (3.35)$$

The relation between the complex $\bar{\mathcal{Y}}_{nm}$ and the real-valued \bar{Y}_{nm} is thus

$$\begin{aligned} \bar{\mathcal{Y}}_{nm} &= \frac{(-1)^m}{\sqrt{4\pi}} \frac{1}{\sqrt{2 - \delta_{0m}}} (\bar{Y}_{nm} + i \bar{Y}_{n,-m}) \quad m \geq 0 \\ \bar{\mathcal{Y}}_{nm} &= \frac{1}{\sqrt{4\pi}} \frac{1}{\sqrt{2}} (\bar{Y}_{n|m|} - i \bar{Y}_{n,-|m|}) \quad m < 0 . \end{aligned}$$

Some integrals

Some useful integrals are expressed below, using both unnormalized and fully normalized spherical harmonic representation.

$$\frac{1}{4\pi} \int_{\Omega} F d\omega = f_{00} = \bar{f}_{00} \quad (3.36)$$

$$\frac{1}{4\pi} \int_{\Omega} F^2 d\omega = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{f_{nm}^2}{P_{nm}^2} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{f}_{nm}^2 \quad (3.37)$$

$$\frac{1}{4\pi} \int_{\Omega} FG d\omega = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{f_{nm} g_{nm}}{P_{nm}^2} = \sum_{n=0}^{\infty} \sum_{m=-n}^n \bar{f}_{nm} \bar{g}_{nm} \quad (3.38)$$

$$\frac{1}{4\pi} \int_{\Omega} FY_{nm} d\omega = \frac{f_{nm}}{P_{nm}^2} = \frac{\bar{f}_{nm}}{P_{nm}} \quad (3.39)$$

$$\frac{1}{4\pi} \int_{\Omega} F\bar{Y}_{nm} d\omega = \frac{f_{nm}}{P_{nm}} = \bar{f}_{nm} \quad (3.40)$$

$$\begin{aligned} \frac{1}{4\pi} \int_{\Omega} F\bar{Y}_{nm} d\omega &= \frac{(-1)^m}{\sqrt{4\pi}} \frac{1}{\sqrt{2-\delta_{0m}}} (\bar{f}_{nm} + i\bar{f}_{n,-m}) \quad m \geq 0 \\ &= \frac{(-1)^m}{\sqrt{4\pi}} \frac{1}{\sqrt{2}} (\bar{f}_{n|m|} + i\bar{f}_{n,-|m|}) \quad m < 0 \end{aligned} \quad (3.41)$$

3.2 Spherical Coordinates

We use spherical longitude λ , co-latitude $\theta = \frac{\pi}{2} - \phi$ and radius r . 'Geodetic' coordinates can all be easily transformed to spherical coordinates.

A vector field, when represented with respect to the local basis $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\lambda$, reads

$$\mathbf{f} = f_r \mathbf{e}_r + f_\theta \mathbf{e}_\theta + f_\lambda \mathbf{e}_\lambda. \quad (3.42)$$

The gradient and the Laplace operator applied to a 3D-function $F(\lambda, \theta, r)$ in spherical coordinates are

$$\nabla F = \frac{\partial F}{\partial r} \mathbf{e}_r + \frac{1}{r} \frac{\partial F}{\partial \theta} \mathbf{e}_\theta + \frac{1}{r \sin \theta} \frac{\partial F}{\partial \lambda} \mathbf{e}_\lambda. \quad (3.43)$$

$$\begin{aligned} \Delta F = \nabla \cdot \nabla F &= \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial F}{\partial r} \right) + \frac{1}{r^2} \left(\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial F}{\partial \theta} \right) \right) \\ &\quad + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 F}{\partial \lambda^2} \\ &= \frac{\partial^2 F}{\partial r^2} + \frac{2}{r} \frac{\partial F}{\partial r} + \frac{1}{r^2} \frac{\partial^2 F}{\partial \theta^2} + \frac{1}{r^2 \tan \theta} \frac{\partial F}{\partial \theta} \\ &\quad + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 F}{\partial \lambda^2} \end{aligned} \quad (3.44)$$

The gradient of the vector field can be written as a matrix, with entries

$$\nabla \mathbf{f} = \begin{pmatrix} \frac{\partial f_r}{\partial r} & \frac{1}{r} \frac{\partial f_r}{\partial \theta} - \frac{f_\theta}{r} & \frac{1}{r \sin \theta} \frac{\partial f_r}{\partial \lambda} - \frac{f_\lambda}{r} \\ \frac{\partial f_\theta}{\partial r} & \frac{1}{r} \frac{\partial f_\theta}{\partial \theta} + \frac{f_r}{r} & \frac{1}{r \sin \theta} \frac{\partial f_\theta}{\partial \lambda} - \cot \theta \frac{f_\lambda}{r} \\ \frac{\partial f_\lambda}{\partial r} & \frac{1}{r} \frac{\partial f_\lambda}{\partial \theta} & \frac{1}{r \sin \theta} \frac{\partial f_\lambda}{\partial \lambda} + \cot \theta \frac{f_\theta}{r} + \frac{\partial f_r}{\partial r} \end{pmatrix} \quad (3.45)$$

The divergence of the vector field is

$$\begin{aligned}\nabla \cdot \mathbf{f} &= \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 f_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta f_\theta) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \lambda} f_\lambda \\ &= \frac{2}{r} f_r + \frac{\partial}{\partial r} f_r + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta f_\theta) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \lambda} f_\lambda\end{aligned}\quad (3.46)$$

For completeness, we note the strain tensor $\boldsymbol{\epsilon} = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$ and the (Cauchy) stress tensor in spherical coordinates:

$$\begin{aligned}\boldsymbol{\epsilon} &= \epsilon_{rr} \mathbf{e}_r \mathbf{e}_r^T + \epsilon_{\theta\theta} \mathbf{e}_\theta \mathbf{e}_\theta^T + \epsilon_{\lambda\lambda} \mathbf{e}_\lambda \mathbf{e}_\lambda^T \\ &\quad + \epsilon_{r\theta} (\mathbf{e}_r \mathbf{e}_\theta^T + \mathbf{e}_\theta \mathbf{e}_r^T) + \epsilon_{r\lambda} (\mathbf{e}_r \mathbf{e}_\lambda^T + \mathbf{e}_\lambda \mathbf{e}_r^T) + \epsilon_{\theta\lambda} (\mathbf{e}_\theta \mathbf{e}_\lambda^T + \mathbf{e}_\lambda \mathbf{e}_\theta^T)\end{aligned}\quad (3.47)$$

in particular

$$\epsilon_{rr} = \frac{\partial u_r}{\partial r} \quad (3.48)$$

$$\epsilon_{\theta\theta} = \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} + \frac{1}{r} u_r \quad (3.49)$$

$$\epsilon_{\lambda\lambda} = \frac{1}{r \sin \theta} \frac{\partial u_\lambda}{\partial \lambda} + \frac{1}{r} u_r + \frac{1}{r \tan \theta} u_\theta \quad (3.50)$$

$$\epsilon_{r\theta} = \frac{1}{2} \left(\frac{1}{r} \frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{1}{r} u_\lambda \right) \quad (3.51)$$

$$\epsilon_{r\lambda} = \frac{1}{2} \left(\frac{1}{r \sin \theta} \frac{\partial u_r}{\partial \lambda} + \frac{\partial u_\lambda}{\partial r} - \frac{1}{r} u_\lambda \right) \quad (3.52)$$

$$\epsilon_{\theta\lambda} = \frac{1}{2} \left(\frac{1}{r \sin \theta} \frac{\partial u_\theta}{\partial \lambda} + \frac{1}{r} \frac{\partial u_\lambda}{\partial \theta} + \frac{1}{r \tan \theta} u_\lambda \right) \quad (3.53)$$

and

$$\begin{aligned}\boldsymbol{\sigma} &= \sigma_{rr} \mathbf{e}_r \mathbf{e}_r^T + \sigma_{\theta\theta} \mathbf{e}_\theta \mathbf{e}_\theta^T + \sigma_{\lambda\lambda} \mathbf{e}_\lambda \mathbf{e}_\lambda^T \\ &\quad + \sigma_{r\theta} (\mathbf{e}_r \mathbf{e}_\theta^T + \mathbf{e}_\theta \mathbf{e}_r^T) + \sigma_{r\lambda} (\mathbf{e}_r \mathbf{e}_\lambda^T + \mathbf{e}_\lambda \mathbf{e}_r^T) + \sigma_{\theta\lambda} (\mathbf{e}_\theta \mathbf{e}_\lambda^T + \mathbf{e}_\lambda \mathbf{e}_\theta^T) .\end{aligned}\quad (3.54)$$

References

1. Freedden W (1999) Multiscale modelling of spaceborne geodata, Teubner, Stuttgart
2. Bettadpur S (2007) Gravity Recovery and Climate Experiment: Level-2 Gravity Field Product User Handbook (Rev 2.3, February 20, 2007), GRACE 327-734
3. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical recipes in C, Cambridge University Press, Cambridge, NY
4. Jekeli C (1981) Alternative methods to smooth the Earth's gravity field, Rep. No. 327, Geodetic Science, Ohio State University
5. Hackbusch W (1993) Iterative solutions of large sparse systems of equations, Springer, New York
6. Jolliffe I (1986) Principal component analysis, Springer, New York
7. Preisendorfer R (1988) Principal component analysis in meteorology and oceanography, Elsevier, Amsterdam
8. Edmonds AR (1974) Angular momentum in quantum mechanics, Princeton University Press, Princeton, NJ
9. Heiskanen WA, Moritz H (1967) Physical geodesy, Freeman and Company, San Francisco
10. Kaula WM (1966) Theory of satellite geodesy, Blaisdell, Waltham, MA
11. Han S-C, Shum CK, Jekeli C, Kuo C-Y, Wilson CR, Seo K-W (2005) Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. *Geophys J Int* 163, 18-25
12. Hannachi A, Unkel S, Trendafilov NT, Jolliffe IT (2009) Independent component analysis of climate data: A new look at EOF rotation. *Journal of Climate* 22, 2797-2812
13. Higgins TP, Kopal Z (1968) Volume integrals of the products of spherical harmonics and their application to viscous dissipation phenomena in fluids, *Astrophysics and space science* 2:352-369
14. Klees R, Revtova E, Gunter B, Ditmar P, Oudman E, Winsemius H, Savenije H (2008) The design of an optimal filter for monthly GRACE gravity models, *Geophys J I* 175(2):417-432
15. Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J Geod* (2007) 81:733749
16. Kusche J, Schmidt R, Petrovic S, Rietbroek R (2009) Decorrelated GRACE timevariable gravity solutions by GFZ, and their validation using a hydrological model, *J Geod*, 83:903-913

17. Narcowich F, Ward J (1996) Nonstationary wavelets on the m -sphere for scattered data, *Appl Comp Harm Anal*, 3:324-336
18. North G, Bell T, Cahalan R, Moeng F (1982) Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review* 110, 699–706
19. Wahr J, Molenaar M, Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J Geophys Res*, 103, B12, 30205-30229
20. Swenson S, Wahr J (2006) Post-processing removal of correlated errors in GRACE data. *Geophys. Res. Lett.* 33, L08402
21. Zhang Z-Z, Chao BF, Hsu H, Lu Y (2009) An effective filtering for GRACE timevariable gravity: Fan filter. *Geophys Res Lett*, 36, L17311