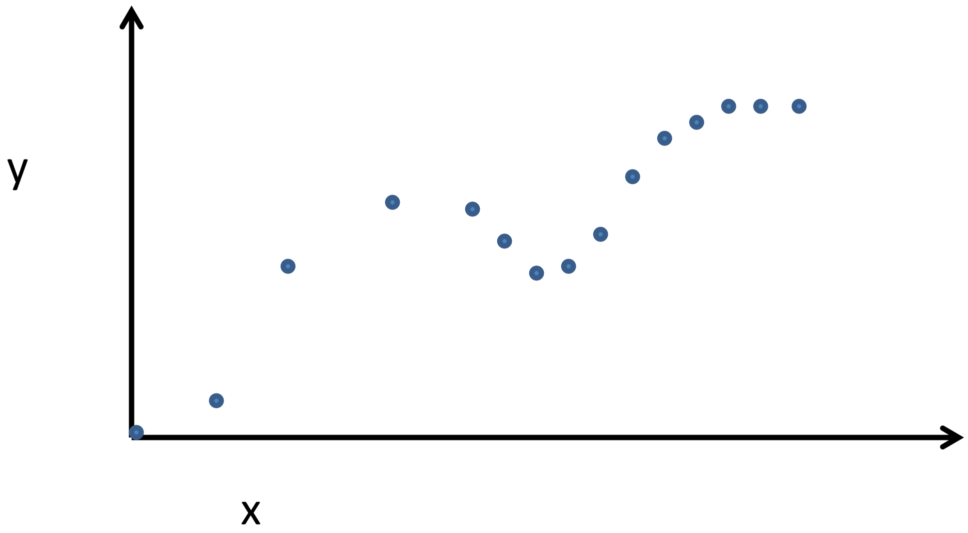
CLIMATE INTERPOLATION
MATHEMATICAL NOTES ON METHODS
PART1

Benoit Parmentier

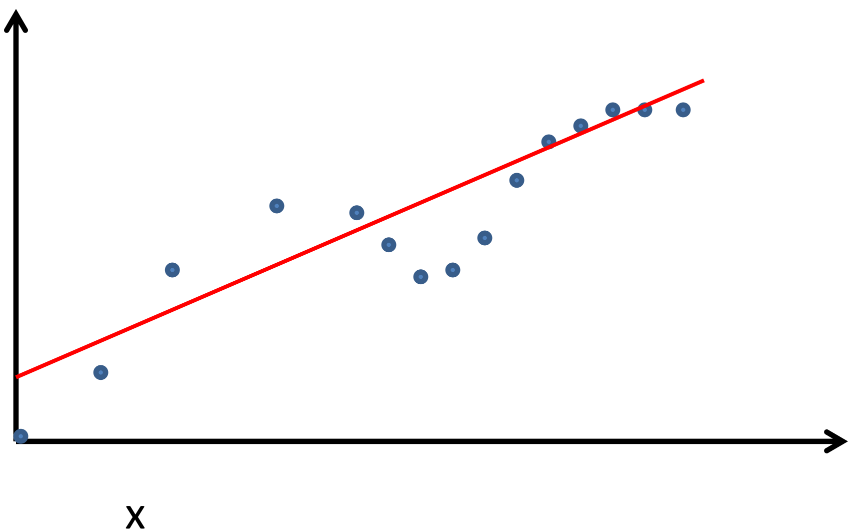
NCEAS, July 22, 2012

Notes assembled during the production of the climate interpolation review.

Estimating smooth function: univariate case



Problem
Let us assume that we have an dependent variable y and a independent variable x. We want to estimate y at unkown points given x.

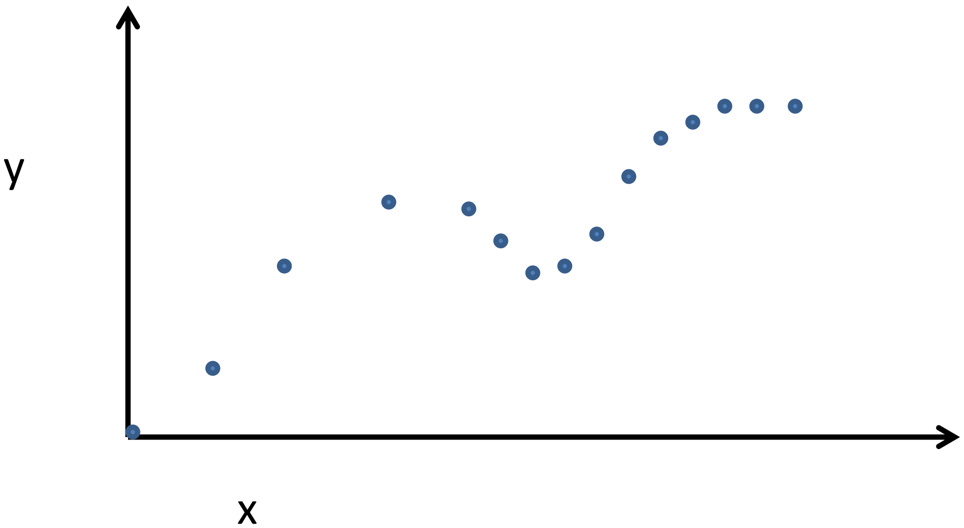


Solution
Use a straight line:

Y=f(x)
Y=ax+b

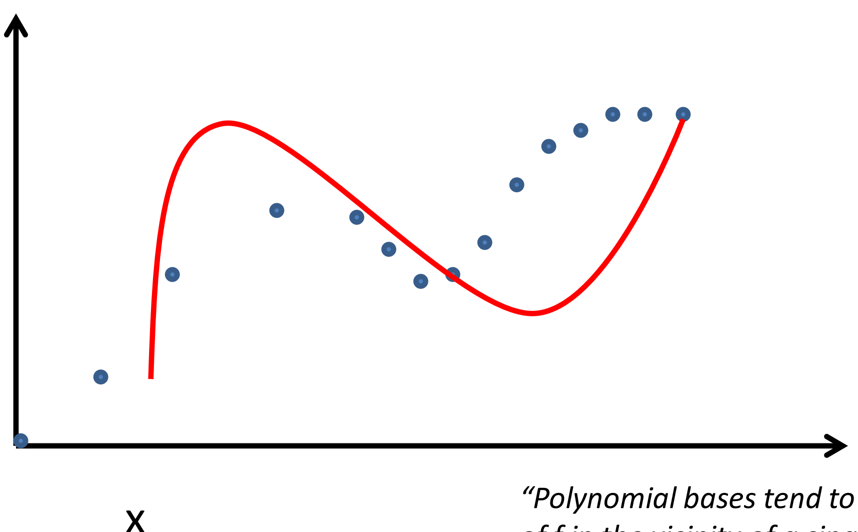
It is a however too smooth and at some location does not provide a good estimate of the value!

Estimating smooth function: univariate case



Problem

Let us assume that we have a dependent variable y and a independent variable x. We want to estimate y at unknown points given x.



Solution

Use a polynomial function

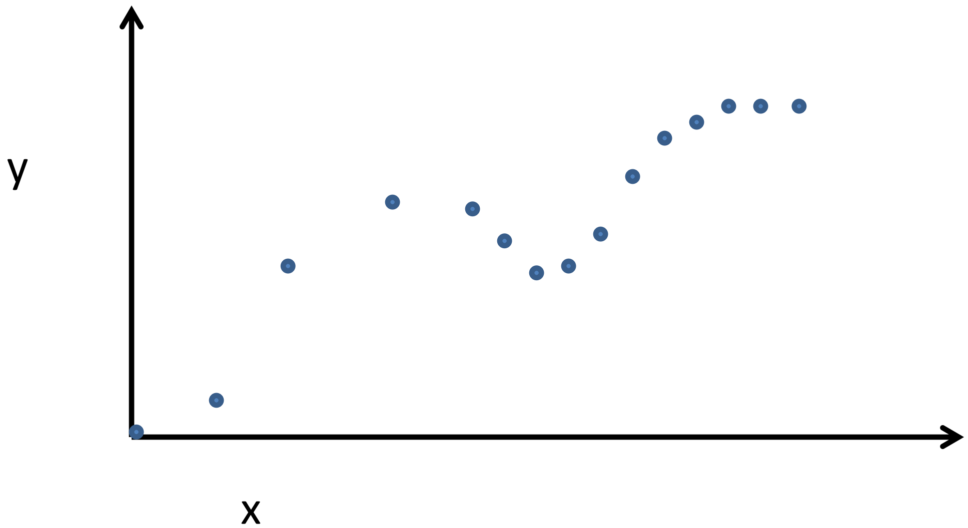
Y=f(x)

Y=a0+a1x1+a2x2+a2x3

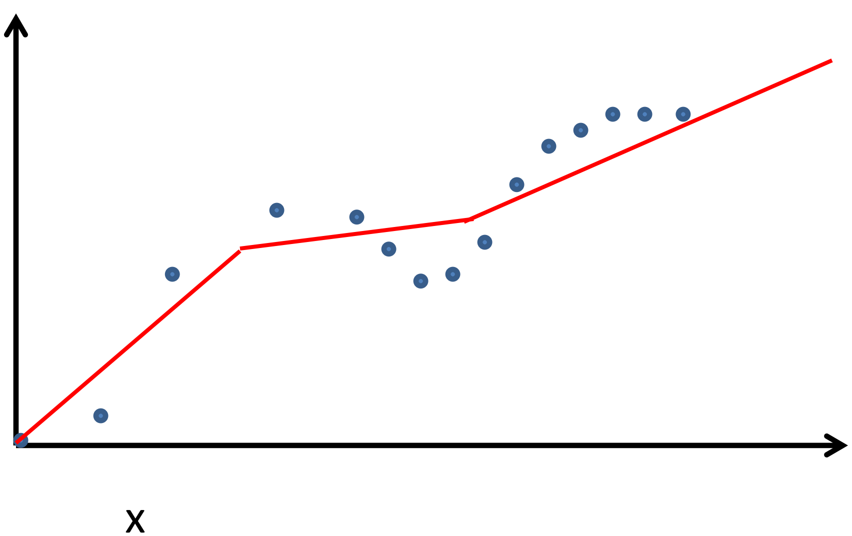
The polynomial is good at specific locations but not good at other to capture the relationship.

"Polynomial bases tend to be very useful for situations in which interest focuses on properties of f in the vicinity of a single point, but when the question of interest relate f over its whole domain…, the polynomial bases have problems." Wood 2006

Estimating smooth function: univariate case



Problem
Let us assume that we have an dependent variable y and a independent variable x. We want to estimate y at unkown points given x.



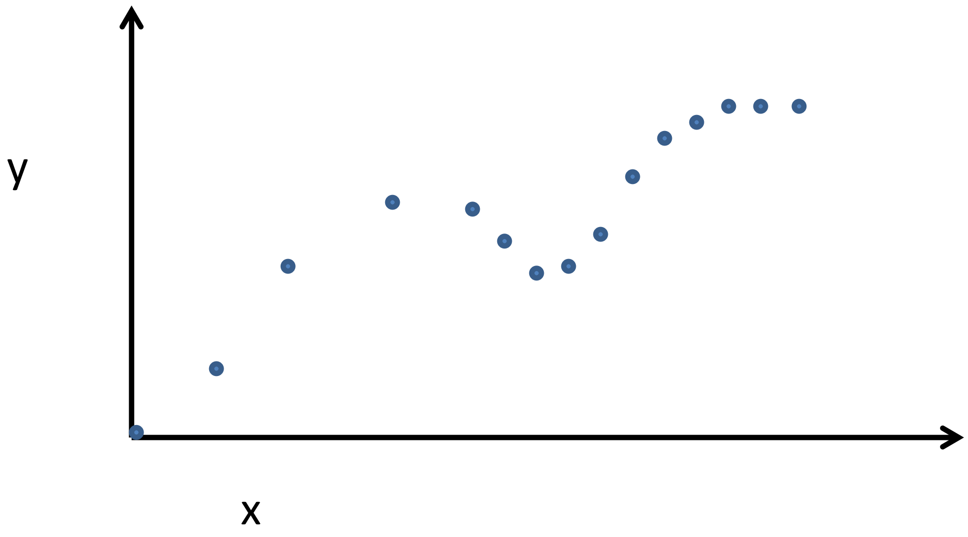
Solution
Use a linear piecewise function…

Y=f(x)
Y=y1+y2+y3

Y1=a0+a1x for xϵ [x*1,x*2]
Y2=a2+a3x for xϵ [x*2,x*3]
Y3=Y2=a2+a3x for xϵ [x*2,x*3]

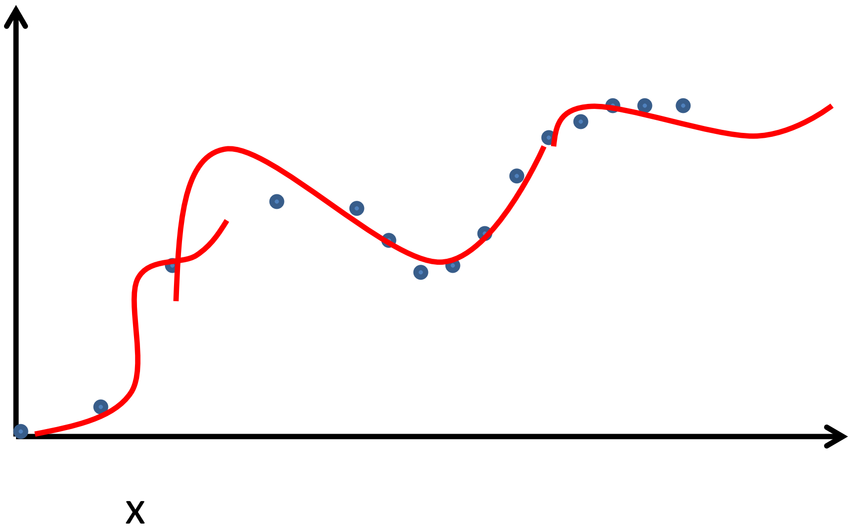
The piecewise linear function provides a good fit but is not smooth i.e. around specific knots it is varies a lot.

Estimating smooth function: univariate case



Problem

Let us assume that we have an dependent variable y and a independent variable x. We want to estimate y at unkown points given x.



Solution

Use a polynomial function

Y=f(x)

Y=s(x)

Where s(x) is a piecewise polynomial…

ESTIMATING THE SMOOTH FUNCTION

f(x) = Σ ai bi(x) or $$f(x) = \sum_{j=1}^{m} \alpha_j b_j(x)$$

The function (polynomial) that we want to find can be expressed a sum of basis function.

$$\mathbf{X} = \begin{bmatrix} b_1(x_1) & b_2(x_1) & .\,. & b_m(x_1) \\ b_1(x_2) & b_2(x_2) & .\,. & b_m(x_2) \\ b_1(x_3) & b_2(x_3) & .\,. & b_m(x_3) \\ . & . & .\,. & . \\ . & . & .\,. & . \\ b_1(x_n) & b_2(x_n) & .\,. & b_m(x_n) \end{bmatrix} = \begin{bmatrix} \mathbf{b}(x_1)^T \\ \mathbf{b}(x_2)^T \\ \mathbf{b}(x_3)^T \\ . \\ . \\ \mathbf{b}(x_n)^T \end{bmatrix}$$

$$f(x) = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + \alpha_5 x^4$$

Find f(x) using the least square criterion:

$$\left\| \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_5 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix} \right\|^2 \quad (\text{i.e. } \|\mathbf{X}\boldsymbol{\alpha} - \mathbf{y}\|^2)$$

Thus, we are trying to describe the shape of the relationship between the response (Tmx,PCRCP) decomposing the function in a sum of basis function. Taken together these functions for a space with mathematical properties…

SPLINES AND PIECE-WISE POLYNOMIALS

Given a set of point n, it is possible to demonstrate that we can always fit a polynomial of degree n passing through every point. The coefficient of this polynomial form a VANDERMONDE matrix. This matrix as interesting properties for estimation.

→However, we are interested in fitting piece-wise polynomials of lower degree than n because of "numerical"? instability and overshooting.

Reasoning:

- Using every point as a basis makes this a large system with a large space dimension, we can use a subset of them, the knots.
- We are interested in capturing global trends rather than every detail hence we do not need to go through every point, this is a regression problem.
- Piece wise polynomial are flexible and can capture both local and global charactersitic of the relationship.
- To ensure continuity, we must enforce certain constraint at connecting points or "knots".
- In some case, the knots are not actual points but are derived from the distribution of the variable.

THIN PLATE SPLINE

- A surface introduced in geometric design by Duchon 1976.
- Given K points in locations x, y with values z, TPS is the surface the passes through the point with 2(k+3) parameters.
- The parameters are: k points and 6 affine motion parameters (3 scalings and 3 rotations)??
- There are many surfaces that can pass through a set of given k points!!

SMOOTHING THIN PLATE SPLINE (STPS)

- A surface that passes through k points but with a regularization so that the solution is unique.
- STPS is the function/surface that minimizes the area of the surface. It is the most efficient in term of the material used to fit a thin plate of metal/plastic through the set of point.
- In one dimension, it is equivalent to minimizing the variation or the curve i.e. its bending energy?? There are other additional form of energy possible…(see wiki).

REGRESSION SMOOTHING THIN PLATE SPLINE

- A surface that does not necessarily pass through k points but is smooth at the knots.

REGRESSION SMOOTHING THIN PLATE SPLINE

It minimizes the objective function with fidelity and bending energy criteria.

The solution is a cubic in one dimension and a TPS function in 2D.

The surface does not go through all the point but “not too far” reflecting a minimization of error or residuals as well as bending energy.

The cubic spline which corresponds to a sum of piece wise cubic polynomial can be restated as another cubic polynomial (monomial??) and a kernel function (the biharmonic spline). It reflects a global and local component!!!

“The cubic smoothing spline is more difficult to generalize to two or higher dimension: the so-called thin-plate spline is one example. P.
Laplacian penalty

“Another generalization is known as multivariate tensor product splines. These are also useful for generalizing univariate regression splines. The basic idea is to construct two-dimensional basis functions by multiplying together one dimensional basis functions in the respective predictors.

KNOTS PROBLEM IN REGRESSION SPLINES

Given a set of k points, the regression splines fits a model that does not pass through every point.

Regression splines passes through a subset of representative points called “knots” that form a small basis set for the all set of point.

→The problem is to find this set of knots (i.e. representative points) from the all set of k points. (Wood 2003)

“The model is typically fitted as a linear or generalized linear model without imposing a wiggliness penalty. The covariate points that are used to obtain the reduced basis are known as the knots of the regression spline. The number of knots controls the flexibility of the model, but unfortunately their location also tends to have a marked effect on the fitted model (see for example Hastie and Tibshirani 1990)”.

→ Some of the problems with knot placement can be partially alleviated by abandoning pure regression splines in favour of regression splines (e.g. Wahba (1980) and Parker and Rice (1985)) where the required penalty is that that associated with regression spline basis.

PENALIZED REGRESSION SPLINES

The splines does not depend on the number of knots but the number of knots need to be chosen so that there are close to the number of degree of freedom.

Actual model degree of freedom is controlled by lamda.

NOTE THAT IN THIS CASE THE DEGREE OF FREEDOM RELATES TO THE NUMBER OF BASIS NECESSARY TO REPRESENT THE DATA POINTS!!! IT DESCRIBES THE INTRA RELATION OF THESE POINTS!!!

RADIAL BASIS FUNCTION

RBF play a central role in interpolation. RBF are function that have an argument which depends on the distance to a reference point (often called center).

Sum of radial basis function can represent/approximate a function.

RBF are used in many context such as Kriging , Splines or Neural Network (RBFN see Lin et al. 2008) to estimate weights.

Frequently used are:

Gaussian RBF: $\phi(r) = \exp(-\varepsilon r)$

Inverse Quadratic: $1/(1+ (\varepsilon r)^2)$

Bi-harmonic spline: $\phi(r)$: $r^2 \ln(r)$

Polyharmonic spline: $\phi(r)$: $r^k \ln(r)$

As function approximation:
$Y(x) = \Sigma\, a_i\, \phi(r)$
$r = || x - x_i ||$ with different x_i centers
r is a distance function such as the Euclidean or other forms

http://en.wikipedia.org/wiki/Radial_basis_function

RADIAL BASIS FUNCTION

Expressing polynomial basis in a form of Cubic Radial Basis function simplifies the problem of estimation because it leads to a four banded matrix with specific properties? Is there a term that shows the trend? Appears so see references

Radial basis functions are basis function that express…

A radial basis function interpolator can be written as

$$F^*(x) = \sum_{i=1}^{n} b_i g(x, x_i) + \sum_{j=0}^{p} a_j f_j(x)$$

Myers 1999

Let $g(x,y)$ be a real valued function defined on $R_k \times R_k$. $g(x,y)$ is positive definite if for any points $x_1, \ldots, x_n$ and any coefficients $\lambda_1, \ldots, \lambda_n$ the quadratic form

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j g(x_i, x_j) \tag{7}$$

is positive (except when all the coefficients are zero). Of course then $g(x_i, x_j)$ generates a positive definite matrix which is invertible. Let $f_0(x), \ldots, f_p(x)$

Where g(x,xi) is a kernl function
And fj(x) are linearly independent functions
The parameters ai and bi need to be
Determined.

RBF must be positive definitive functions to be useful in interpolation…

TPS is the function that passes through the points and minimizes the bending energy (i.e. integral of the second derivative).SO WE CAN EXPRESS THE POLYNOMIAL USED FOR INTERPOLATION AS A LAGRANGE POLYNOMIAL. THIS POLYNOMIAL IS ABLE TO INVERT AND VANDERMONDE MATRIX. A VANDERMONDE MATRIX IS A SPECIAL MATRIX…http://en.wikipedia.org/wiki/Lagrange_polynomial

SMOOTHING THIN PLATE SPLINES: OPTIMIZATION PROBLEM

$$\mathbf{O\ (f(x))} = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda J_m(f)$$

Penalty for too much roughness…

m=order of the partial derivative, Nb of covariates
n=number of points in the dataset

Fit measure

Smoothness parameter

Roughness measure

O (f(x)) is a "objective functional", the solution of the optimization problem is not a scalar but a function!!!

→ Differential equation problem of the Lagrange-Euler type (see Mitas et al.2009)

This is the solution…

$$f(x) = \sum_{j=1}^{M} a_j \phi_j(x) + \sum_{i=1}^{n} b_i \psi(r_i)$$

With phi x being a monomial of order m

||x-xi ||= R(x)

Psi: radial basis function RBF

SMOOTHING CUBIC SPLINE: OPTIMIZATION PROBLEM

In one dimension

The expression of the interpolating spline $\sigma(x)$ relative to n data points x_α is (Ahlberg, Nilson, and Walsh, 1967).

$$\sigma(x) = a_0 + a_1 x + \sum_{\alpha=1}^{n} b^\alpha |x - x_\alpha|^3. \qquad (7)$$

Dubrule 1983, shows that the solution to the optimization problem is a trend function + an RBF function.

The coefficients a_0, a_1, and b^α are determined by the system:

$$\begin{cases} \sum_\alpha b^\alpha = 0 \\ \sum_\alpha b^\alpha x_\alpha = 0 \\ \sigma(x_\alpha) = z(x_\alpha) \quad (\forall \alpha \in \{1, \ldots, n\}). \end{cases} \qquad (8)$$

For the smoothing spline, the expression of $\sigma(x)$ is the same, with the following conditions:

$$\begin{cases} \sum_\alpha b^\alpha = 0 \\ \sum_\alpha b^\alpha x_\alpha = 0 \\ \sigma(x_\alpha) + \dfrac{b^\alpha}{w_\alpha^{\,2}} = \sigma(x_\alpha) + \dfrac{b^\alpha}{\rho} S_\alpha^{\,2} \\ \qquad = y(x_\alpha) \quad (\forall \alpha \in \{1, \ldots, n\}). \end{cases} \qquad (9)$$

In one dimension, the solution to the objective functional is a cubic splines function. This function has two part: a first degree monomial and a RBF (kernel).

THIN PLATE SPLINE: OPTIMIZATION PROBLEM

In two dimensions

For the interpolating spline, we have (Duchon, 1975):

$$\sigma(x) = \sigma(u, v) = a_0 + a_1 u + a_2 v + \sum_{\alpha=1}^{n} b^{\alpha} r_{\alpha}^{2} \log r_{\alpha} \quad (10)$$

where

$$r_{\alpha}^{2} = (u - u_{\alpha})^2 + (v - v_{\alpha})^2$$

and a_0, a_1, a_2, and the b^{α} are given by the system:

$$\begin{cases} \sum_{\alpha} b^{\alpha} = 0 \\ \sum_{\alpha} b^{\alpha} u_{\alpha} = \sum_{\alpha} b^{\alpha} v_{\alpha} = 0 \\ \sigma(x_{\alpha}) = z(x_{\alpha}) \quad (\forall \alpha \in \{1, \ldots, n\}). \end{cases} \quad (11)$$

The expression is the same for smoothing splines, but the system is:

$$\begin{cases} \sum_{\alpha} b^{\alpha} = 0 \\ \sum_{\alpha} b^{\alpha} u_{\alpha} = \sum_{\alpha} b^{\alpha} v_{\alpha} = 0 \\ \sigma(x_{\alpha}) + \dfrac{b^{\alpha}}{w_{\alpha}^{2}} = \sigma(x_{\alpha}) + \dfrac{b^{\alpha}}{\rho} S_{\alpha}^{2} \\ \quad = y(x_{\alpha}) \quad (\forall \alpha \in \{1, \ldots, n\}). \end{cases} \quad (12)$$

LINKS BETWEEN KRIGING AND TPS: DUAL PROBLEM

Huntchinson and Gessler 1994

$$f(\boldsymbol{x}) = \sum_{j=1}^{M} a_j \phi_j(\boldsymbol{x}) + \sum_{i=1}^{n} b_i \psi(r_i)$$

General solution

$$\sum_{i=1}^{n} b_i \phi_j(\boldsymbol{x}_i) = 0 \quad j = 1, \ldots, M$$

With constraint

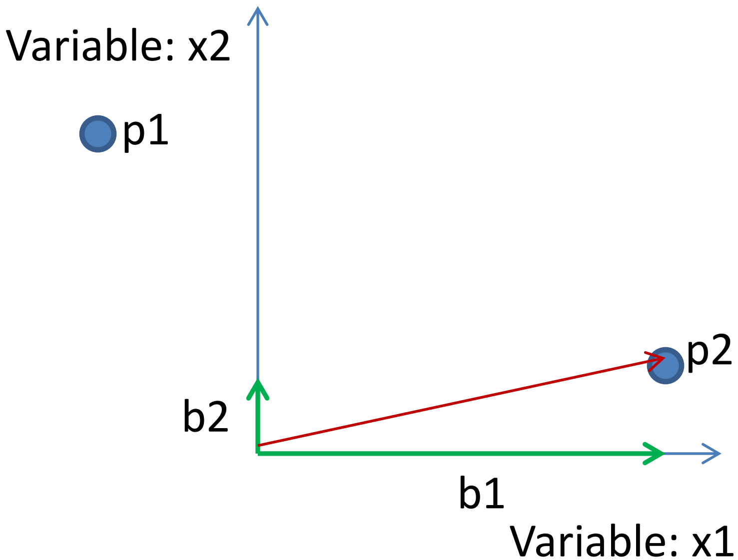
F(x1)=a1x1+a2x2+a3x3+

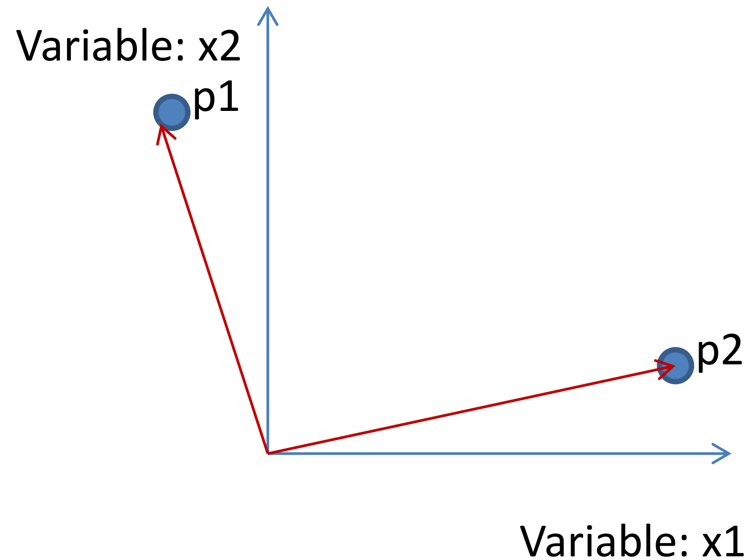
$$f(\boldsymbol{x}) = \sum_{j=1}^{M} a_j \phi_j(\boldsymbol{x}) + \sum_{i=1}^{n} b_i \psi(r_i)$$

Where m= number of covariates
N= number of observations/points/knots

In short, the optimization of the functional can be recasted into another functional which makes explicit the link between kriging and TPS. This can be done because of the duality property and the Riesz theorem function analysis/functional algebra.

KEY IDEA USED IN PENALIZED LEAST SQUARE: DUALITY AND CHANGE OF BASIS





p2= b1***e1** + b2***e2**

p1: point/observation
X1: axis of reference
p2: vector associated to point p2
e1: vector basis for x1, standard Euclidean basis
e2: vector basis for x2, standard Euclidean basis

Change of basis:
The components of p2 form a new basis while the unit basis of the variables are the points!!
b1= e1***p1** + e2***p2**

P2=(a1,a2)
P1=(b1,b2)

We can reverse the role of observation and variable without changing the structure of the space!!! This is just a change of perspective.

KRIGING: THE MANY KINDS…

Simple Kriging
Mean is known and not modeled (→ second order stationariy)

Ordinary kriging
Mean is not known and modeled (→weak stationarity)

Universal kriging
Mean not known and modeled as a drift.

Regression Kriging

The interpolated surface is then constructed using statistical conditions of unbiasedness and minimum variance. In its dual form (Hutchinson and Gessler 1993; Matheron 1971) the universal Kriging interpolation function can be written as

$$F(\mathbf{r}) = T(\mathbf{r}) + \sum_{j=1}^{N} \lambda_j \, C(\mathbf{r} - \mathbf{r}_j) \qquad (5)$$

Mitas and Mitasovas 1999.

KED AND RK MATRICES (HENGL ET AL. 2009)

There is equivalence between RK and KED when regression uses GLS estimates for the trend…Details of the derivation
In Hengl 2009:38

It can be demonstrated that both KED and RK algorithms give exactly the same results (Stein, 1999; Hengl et al., 2007a). Start from KED where the predictions are made as in ordinary kriging using $\hat{z}_{\mathrm{KED}}(\mathrm{s}_0) = \lambda_{\mathrm{KED}}^{\mathrm{T}} \cdot \mathbf{z}$. The KED kriging weights ($\lambda_{\mathrm{KED}}^{\mathrm{T}}$) are obtained by solving the system (Wackernagel, 2003, p.179):

$$\begin{bmatrix} \mathrm{C} & \mathrm{q} \\ \mathrm{q}^{\mathrm{T}} & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_{\mathrm{KED}} \\ \phi \end{bmatrix} = \begin{bmatrix} \mathrm{c}_0 \\ \mathrm{q}_0 \end{bmatrix}$$

$$\hat{\beta}_{\mathrm{GLS}} = \left(\mathrm{q}^{\mathrm{T}} \cdot \mathrm{C}^{-1} \cdot \mathrm{q}\right)^{-1} \cdot \mathrm{q}^{\mathrm{T}} \cdot \mathrm{C}^{-1} \cdot \mathrm{z}$$

KED AND RK EQUIVALENCE (HENGL ET AL. 2009)

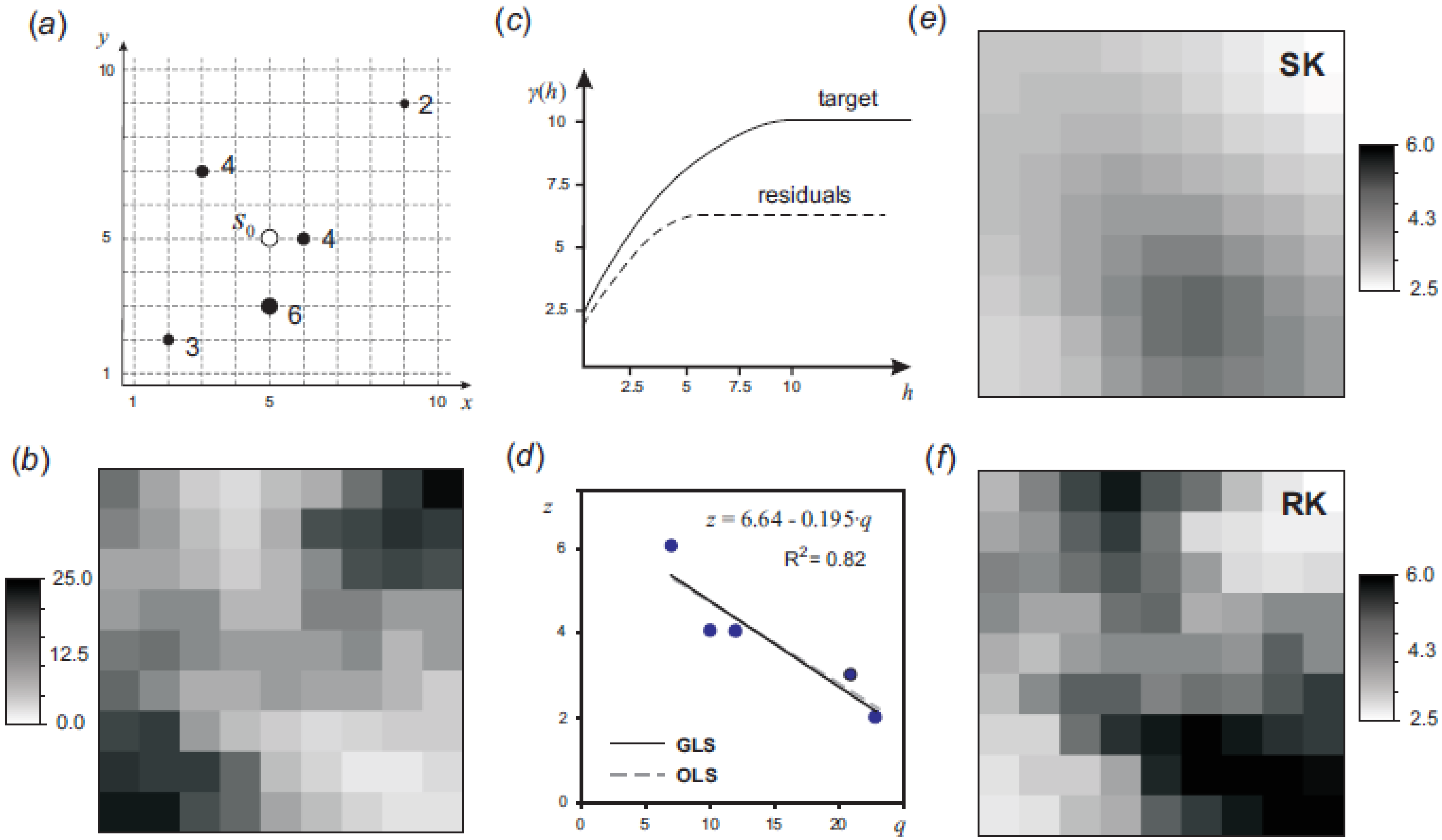


Fig. 2.6: Comparison of ordinary kriging and regression-kriging using a simple example with 5 points (Burrough and McDonnell, 1998, p.139–141): (a) location of the points and unvisited site; (b) values of the covariate q; (c) variogram for target and residuals, (d) OLS and GLS estimates of the regression model and results of prediction for a 10×10 grid using ordinary kriging (e) and regression-kriging (f). Note how the RK maps reflects the pattern of the covariate.

OLS and GLS estimates of the regression step in Regression Kriging (RK) are almost identical in practice….

KERNEL FUNCTIONS AND SMOOTHING SPLINES BASES I

Smoothing functions splines are related to kernel functions and can be understood as acting as moving average filters acting on data points.

"For a smoother with symmetric smoother matrix S, the eigendecomposition of S can be used to describe its behaviour. This is much like the use of a transfer function to describe a linear filter for time series."

"The transfer function is a convenient tool both for describing the action of a filter, and for designing one.p.59 Hastie

KERNEL FUNCTIONS AND SMOOTHING SPLINES BASES II

Add figures Hastie and Tibshirani p.58 and p.29 for the equivalent kernel….

Reproducing Kernel Hilbert Space (Hastie and Tibshirani 1990)

This allows to recast the minimizing functional in terms of functional of a kernel and function basis??

Q is reproducing kernels that provides bases for representing the solution.

INVERSE DISTANCE WEIGHTING

$$F(\mathbf{r}) = \sum_{i=1}^{m} w_{\mathrm{i}}\, z(\mathbf{r}_i) = \frac{(\sum_{i=1}^{m} z(\mathbf{r}_i) \;/\; |\mathbf{r}-\mathbf{r}_i|^p}{\sum_{j=1}^{m} 1/|\,\mathbf{r}-\mathbf{r}_j|^p}$$

Where r is a vector of observation

Introduced by Sheppard 1968 in GIS/spatial analysis.

→ Sum of the weights must be equal to one. $\hat{z}(\mathbf{s}_0) = \lambda_0^{\mathrm{T}} \cdot \mathbf{z}$ Hengl 2009

Technically IDW has a kernel function that is an inverse power of p. P can be fitted from the data.

Other kernel functions can be used such as exponential decays.

GEOGRAPHICALLY WEIGHTED REGRESSION

becomes

GWR works by dividing the study areas in subregions where local regression models are run. When the region is equal an observation the model is fully localized. Observations are also weighted by distance using a Kernel function.

-When region used for estimation overlap there may be some problem (see Griffith et al.)

In many ways GWR is similar to LOESS p.29 Hastie and Tibshirani 1990

As such the coefficient of regressions are calculated by incorporating the weights such that

OLS $$\hat{\beta} = (X^T X)^{-1} X^T y$$

WLS $$\hat{\beta} = (X^T W X)^{-1} X^T W y$$

White paper Fortherigham et al.

where X is the design matrix containing independent variables and a column 1 and y is the dependent variable.

VARIATIONAL APPROACH TO INTERPOLATION (MITAS AND MITASOVA 1999)

The variational approach to interpolation and approximation is based on the assumption that the interpolation function should pass through (or close to) the data points and, at the same time should be as smooth as possible. The two requirements are combined into a single condition of minimizing the sum of the deviations from the measured points and the smoothness semi-norm of the spline function.

$$\sum_{j=1}^{N} |z_j - F(\mathbf{r}_j)|^2\, w_j + w_0 I(F) = minimum$$

$$F(\mathbf{r}) = T(\mathbf{r}) + \sum_{j=1}^{N} \lambda_j\, R(\mathbf{r}, \mathbf{r}_j)$$

The solution to the variational problem is a function composed of T(x) and R(x).
The solution depends on I(F) which is the smoothness semi-norm.

I(F) → can be bivariate smoothness normed with squares of the second derivatives
→ can include higher order derivatives
→ can include the first order derivative (membrane term)
→ can include the first derivative (membrane term) and higher orders (RST)

There are at least two deficiencies of the TPS function: (1) the plate stiffness causes the function to overshoot in regions where data create large gradients; (2) the second derivatives diverge in the data points, causing difficulties in surface geometry analysis.

REGULARIZED LINEAR SPLINES WITH TENSIONS

Table 1 Examples of bivariate spline functions, their corresponding smoothness seminorms and Euler–Lagrange equations.

Method	*I(F)*	*Euler–Lagrange Eq.*
Membrane	$\int [F_x^2 + F_y^2]\, d\mathbf{r}$	harmonic
Minimum curvature[a]	$\int [F_{xx}^2 + F_{yy}^2]\, d\mathbf{r}$	biharmonic modified
Thin plate spline[b]	$\int [F_{xx}^2 + F_{yy}^2 + 2F_{xy}^2]\, d\mathbf{r}$	biharmonic
Thin plate spline+tension[c]	$\int [\varphi^2 [F_x^2 + F_y^2] + [F_{xx}^2 + \ldots]\, d\mathbf{r}$	harmonic+biharmonic
Regular thin plate spline[c]	$\int [F_{xx}^2 + \ldots] + \tau^2[F_{xxx}^2 + \ldots]\, d\mathbf{r}$	biharmonic+6th-order
Regular spline with tension[d]	$\Sigma_{mn}\, c_{mn}(\varphi) \int [F_{x^n y^m}]^2\, d\mathbf{r}$	all even orders

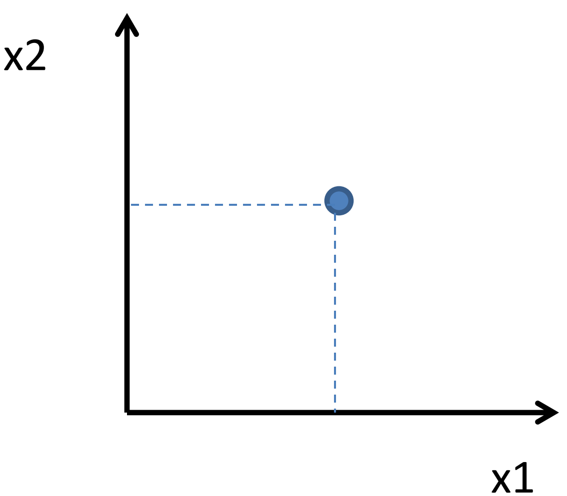
[a] Briggs 1974, Duchon 1975, Hutchinson and Bischof 1983, Wahba 1990
[b] Franke 1985, Hutchinson 1989
[c] Mitas and Mitasova 1988
[d] Mitasova et al 1995; Mitas and Mitasova 1997

"The problem of overshoots can be eliminated by adding the first order derivatives into the seminorm I(F), leading to TPS with tension (Franke 1985; Hutchinso 1989; Mitas and Mitasov 1988). Change of the tension tunes the surface from a stiff plate into an elastic membrane . "

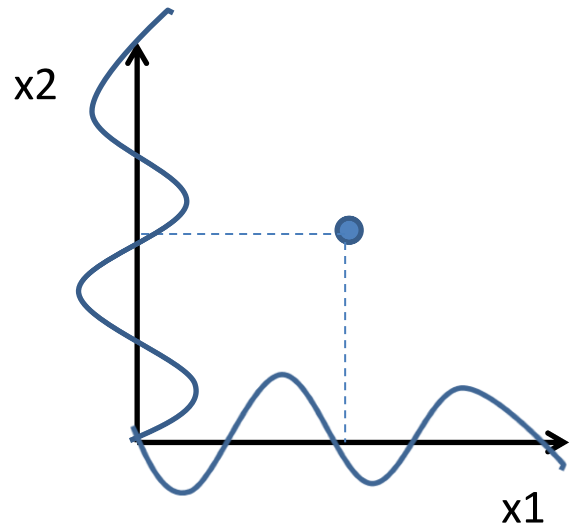
Additional notes on mathematical concepts background slides

BASIS FUNCTION

A basis function:*" is an element of a particular basis for a function space. Every continuous function in the function space can be represented as a linear combination of basis functions, just as every vector in a vector space can be represented as a linear combination of basis vectors." wikipedia http://en.wikipedia.org/wiki/Basis_function*



Vector space with
axis being vectors



Function space with
axis being functions

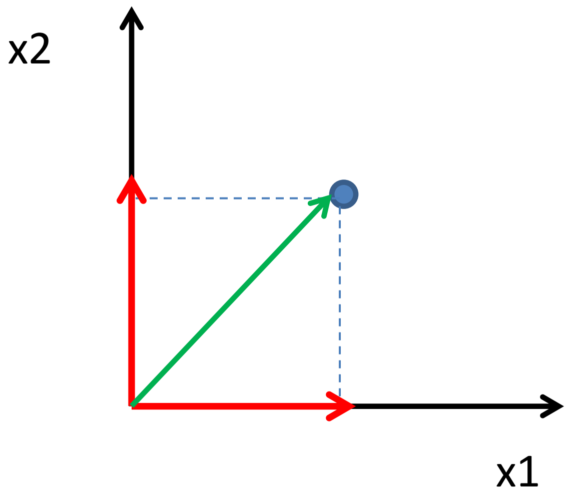
A point can be reference using coordinates representing how "much" of each axis/var there is

Basis functions are orthogonal so their inner product is equal to zero and they are square integrables.

VECTOR SPACE-NORMED SPACE AND ALGEGRA

A mathematical structure or object that allow for certain operations on element such that the results of these operations are within the structure.

Element: vector and scalar
Operations: addition, multiplication
Other properties: commutativity, associativity etc. sometime with norm



Vector space with
axis being vectors

OBJECTS AND STRUCTURE IN MATHEMATIC: RELEVANT TO PROBLEMS IN ALGEBRA

number set→ group→ring→field (corpus)→vector space→functional space

Level of abstraction

Set: collection of object

Algebraic structure: set+finite operations

group: set + more complex operations (rules of combination, closure)

Ring: set + more generalized operation for number and matrices

Field:…
Vector space
Functional space
field-> corps commutatif (in French).

VECTOR BASIS AND DEGREE OF FREEDOMS

In regression splines the degree of freedom is equal to the number of bases used. It also relates to knots.

Constraints reduce the number of freedoms.

In classical statistics degree of freedoms relate to the number of observations…

DUAL SPACE AND TRANSPOSE OF MATRIX

Restating the wikipedia in the context of our problem:

ai = coefficients of the sum of basis used to represent the function form of the relationship between x and y

B(x) = is the function of x

y= Σai*B(xi)

This is the polynomial basis function= yi
The sum of yi in the vicinity gives an appximation of Y

Φ(ri) : is the kernel function associated to ai
it is function in the reproducing kernel Hilbert space

The gist of the idea is to restate the minimization problem such that the kernel Φ(ri) is what is sought to solve the penalized least square objective. Given the polynomial basis or yi values what are the coefficients? The coefficients are in effect equivalent to kernel function resulting in weights applied to neighboring observations???

DUAL SPACE AND TRANSPOSE OF MATRIX

Idea, the argument becomes the function and the function the argument. This is similar to having the variable becoming the observation and the observation becoming the variable. In the TPS context, finding the function versus finding the coefficients…

http://en.wikipedia.org/wiki/Dual_space

Transpose of a linear map If $f : V \to W$ is a linear map, then the *transpose* (or *dual*) $f^* : W^* \to V^*$ is defined byfor every $\varphi \in W^*$. The resulting functional $f^*(\varphi)$ is in V^*, and is called the *pullback* of φ along f.The following identity holds for all $\varphi \in W^*$ and $v \in V$:where the bracket [•,•] on the left is the duality pairing of V with its dual space, and that on the right is the duality pairing of W with its dual. This identity characterizes the transpose,[6] and is formally similar to the definition of the adjoint.The assignment $f \mapsto f^*$ produces an injective linear map between the space of linear operators from V to W and the space of linear operators from W^* to V^*; this homomorphism is an isomorphism if and only if W is finite-dimensional. If $V = W$ then the space of linear maps is actually an algebra under composition of maps, and the assignment is then an antihomomorphism of algebras, meaning that $(fg)^* = g^*f^*$. In the language of category theory, taking the dual of vector spaces and the transpose of linear maps is therefore a contravariant functor from the category of vector spaces over F to itself. Note that one can identify $(f^*)^*$ with f using the natural injection into the double dual.If the linear map f is represented by the matrix A with respect to two bases of V and W, then f^* is represented by the transpose matrix A^T with respect to the dual bases of W^* and V^*, hence the name. Alternatively, as f is represented by A acting on the left on column vectors, f^* is represented by the same matrix acting on the right on row vectors. These points of view are related by the canonical inner product on $\mathbf{R}n$, which identifies the space of column vectors with the dual space of row vectors.

EUCLIDEAN AND HILBERT SPACE

Hilbert space is a generalization of a vector space which extends a vector algebra (i.e. vector space with addition, scalar multiplication and vector multiplication) to Euclidean space with 2 and 3 dimensions n dimensions where n is infinite.

It is a vector algebra with an inner product allowing the notion of metric for distances (i.e. geometry). It is a generalization of the Euclidean space.*

Hilbert space is an infinite dimensional function space equipped with a norm where distance and directions are meaningful.

This concept is important because in splines, the basis function "live" or form a Hilbert space of degree K corresponding to the number of knots…. (see Wood 2003, Wahba 1990)

EUCLIDEAN SPACE:

a real affine space with inner product. A affine space can be seen as vector space without origin. It is defined more technically in mathematics.

"Affine space is nothing more than a vector space whose origin we try to forget about, by adding translations to the linear maps" Marcel Berger (Wikipedia) http://en.wikipedia.org/wiki/Affine_space.

Euclidean space is flat!! It adheres to planar geometry.

REPRODUCING KERNEL HILBERT SPACE

Add info…

It is a space that associate a kernel to the inner product of function...

NORMS AND SEMI-NORMS

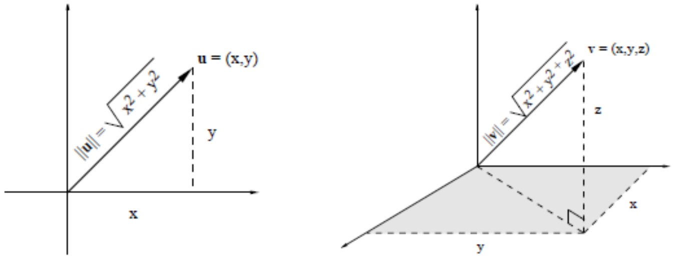


FIGURE 5.1.1

This measure of length,

$$\|\mathbf{u}\| = \sqrt{x^2 + y^2} \quad \text{and} \quad \|\mathbf{v}\| = \sqrt{x^2 + y^2 + z^2},$$

Euclidean Vector Norm

For a vector $\mathbf{x}_{n\times 1}$, the *euclidean norm* of $\mathbf{x}$ is defined to be

- $\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = \sqrt{\mathbf{x}^T\mathbf{x}}$ whenever $\mathbf{x} \in \Re^{n\times 1}$,
- $\|\mathbf{x}\| = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \sqrt{\mathbf{x}^*\mathbf{x}}$ whenever $\mathbf{x} \in \mathcal{C}^{n\times 1}$.

Meyer et al. 2001
http://matrixanalysis.com/DownloadChapters.html

Norms are defined by inner products and are used to define geometrical concepts in the mathematical space.

Norms are function that assign length to vectors. All norms must be strictly positive.

Semi-norms are can assign zero length in contrast to norms…

In the optimization problem for smoothing splines, the “roughness criterion” is defined as a norm or semi norm (Mitas 1999).

$$\sum_{j=1}^{N} |z_j - F(\mathbf{r}_j)|^2\, w_j + w_0 I(F) = minimum \qquad (6)$$

Where I(F)
Is a semi-norm…

REPRODUCING KERNEL HILBERT SPACE

Hilbert space has basis that are functions.

A point P in an Hilbert space has coordinates: (a1,a2,…,an) with n being infinity

P is (a1*f1, a2*f2, etc.…) → P has certain amount “a1” of basis f1, and a certain amount “a2” of basis f2 etc. The collection of coefficient (a1,a2,...) form a vector.

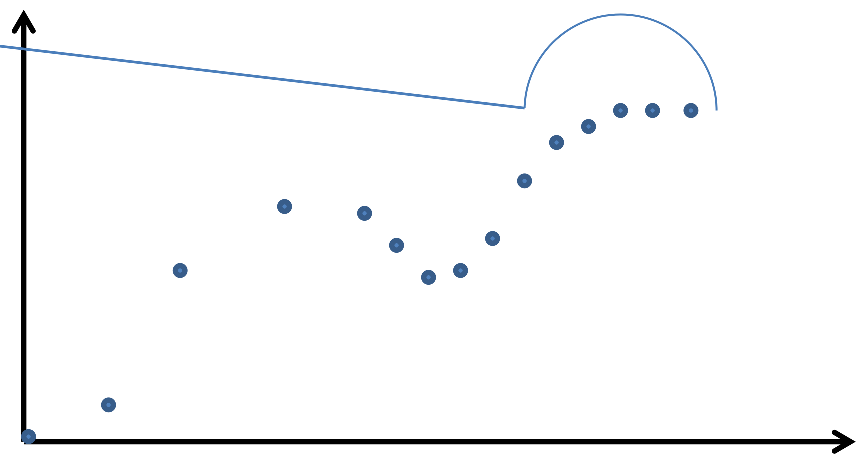
The coordinates themselves can be seen as being a function!! Why not change our view point in which case the coordinates are the basis and the functions are the coefficients!!!

In the context of regression, instead of searching for the function y=f(x) we search for
y0 = Σai*yi

Where y0 could be tmax at location x0

ai are weights dependent on x!!!
ai=R(xi) where R is the kernel function in reproducing kernel hilbert space.

Since yi is also dependent on x, we have s sum of product or inner product of function that converges to the solution y0!!



OPTIMIZATION DUAL AND PRIMAL FORMS

The linear case[*Linear programming problems are optimization problems in which the objective function and the constraints are all linear. In the primal problem, the objective function is a linear combination of n variables. There are m constraints, each of which places an upper bound on a linear combination of the n variables. The goal is to maximize the value of the objective function subject to the constraints. A solution is a vector (a list) of n values that achieves the maximum value for the objective function.In the dual problem, the objective function is a linear combination of the m values that are the limits in the m constraints from the primal problem. There are n dual constraints, each of which places a lower bound on a linear combination of m dual variables.*](http://en.wikipedia.org/wiki/Dual_problem)

http://en.wikipedia.org/wiki/Dual_problem

Possible link to the context of TPS optimization:

In this context, the goal is to find the function that minimizes the objective function (penalized least square). The constraints are on the coefficients bi of the function basis.

Thus this is the primal form and turning it around we can maximize the objective function from the dual form problem: the variable is now the set of coefficient and the constraints are the function. If we associate the coefficient to a kernel then we have a function too!

REGULARIZATION THEORY

Regularization is used in mathematic to solve ill-posed problems. Problems are ill-posed when unique solutions do not exist or the solution is unstable (ie small variation have a large impact on the value).

Solutions may be found by stating additional assumption that translate into mathematical constraints. The problem is regularized or stabilized.

Tikhonov regularization

Variational regularization

Total variation of differentiable functions The total variation of a differentiable function can be expressed as an integral involving the given function instead of as the supremum of the functionals of definitions **1.1** and **1.2** **wikipwdia**

http://en.wikipedia.org/wiki/Tikhonov_regularization