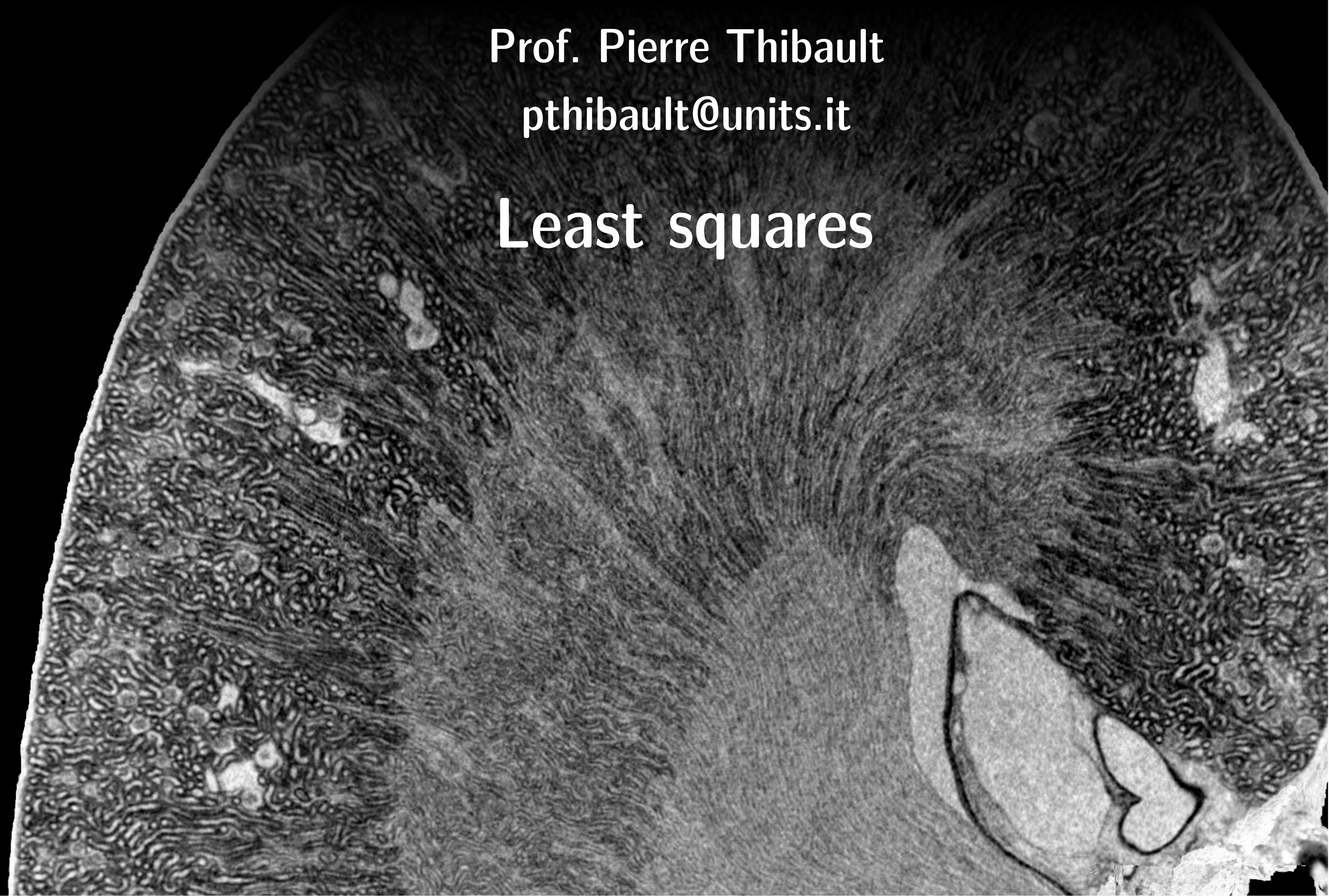
Image Processing for Physicists

Prof. Pierre Thibault
pthibault@units.it

Least squares



Overview

- General remarks on optimization
- Least squares principle
  - Application examples
- Lagrange multipliers
  - Application examples

Image Processing Problems

- Image processing problems can be formulated as linear/nonlinear equations

  data : y long vector that contains all measured quantities

  variables: x independent variables auxiliary information e.g. pixel coordinates

- In many cases "true" solution does not exist (random noise!) or is hard to calculate $\boxed{\text{(inverse problem)}}$

  known variables

  model $$y = M(x; \beta)$$ model parameters "forward model"

- Find "best-guess" approximation

  goal: estimate the "best" parameters that fit with the data $\hat{\beta}$

  β_0 β_1

  $y = mx + b$

- Need understanding of "approximation"

- Need understanding of "best" approximation

Estimation

- Estimator and Estimate

estimate : $\hat{\beta}$

estimator : function $\{y\} \longrightarrow \hat{\beta}$

"inverse" of the model M

- Cost function $f(y; x, \beta)$
  - Measures how well our estimate compares to the original

$$\hat{\beta} := \min_{\beta} f$$

→ Find Minima of cost function

→ Optimization theory

Least squares principle

- Problem formulation

model : $y = M(x;\beta)$

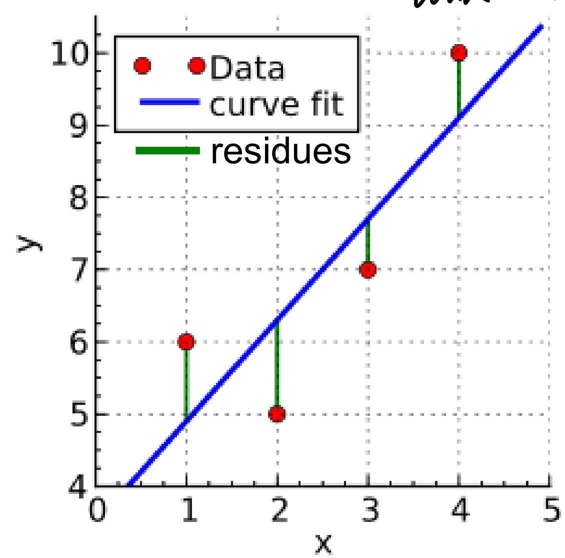
residue : $y_i - M(x_i;\beta)$ i: index of a measurement

cost function : $S(y;x,\beta) = \sum_i r_i^2$

$= \sum_i |y_i - M(x_i;\beta)|^2$ Euclidean distance between measurement and model

- Basic idea: minimize squared residues

$\hat{\beta} = \min_{\beta} S$

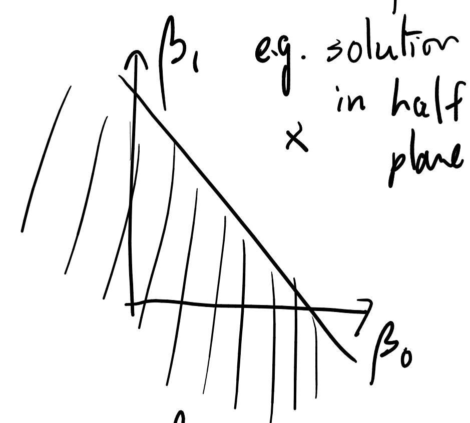


Optimization

- Find minimum/maximum of objective function (in our case: the cost function)

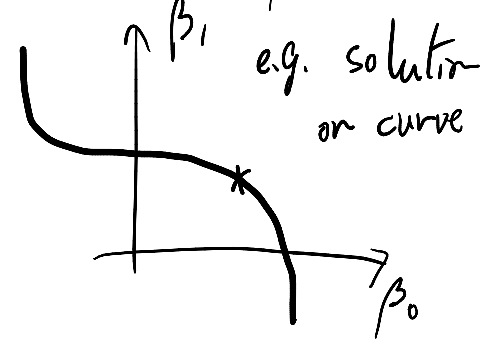
- Inequality constraints ← There might be additional information about β

$$g(\beta) \leq 0$$



- Equality constraints

$$h(\beta) = 0$$



Typically taken into account using Lagrange multipliers

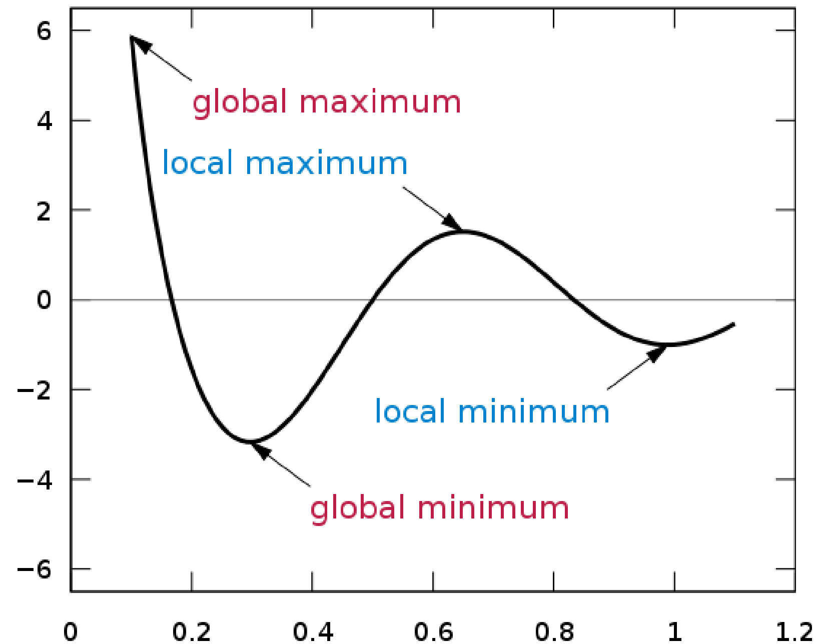
- Standard: minimization problem (negation of maximization problem)

Global/Local Minima/Maxima

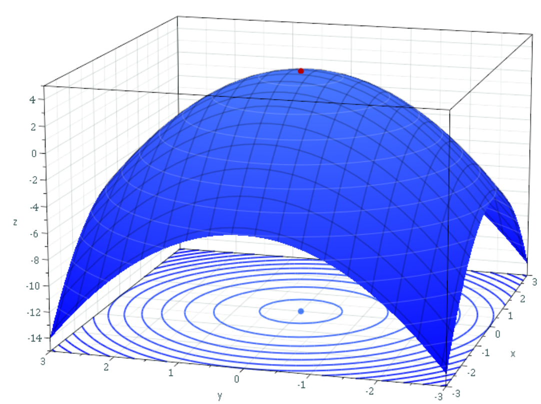
- Find extremal point of function

$$\frac{\partial S}{\partial \beta} = 0 \rightarrow \text{local optimum}$$

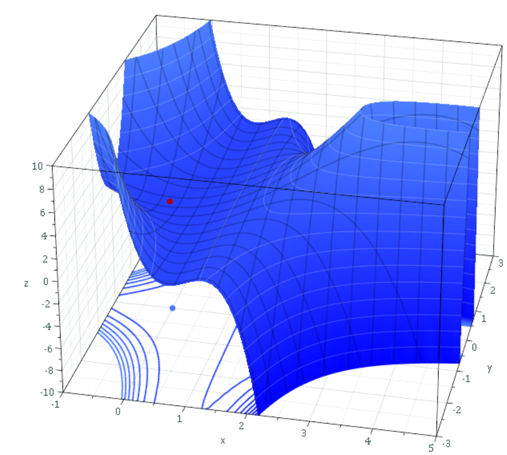
$$\nabla_\beta S = 0$$



convex function



non-convex function



- Convex problems:

$\rightarrow$ local minimum is also global minimum

- All linear problems are convex!

Linear least squares

- Problem formulation

$$y = M(x;\beta)$$

$$y = X \cdot \beta$$

matrix known

parameters

$M \times N$

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} x_{00} & x_{01} & \cdots \\ \vdots & & \\ & & \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}$$

- Minimize cost function (least squares)

$$S = \sum_i \left| y_i - (X\beta)_i \right|^2$$

$$= \sum_i \left| y_i - \sum_j x_{ij}\beta_j \right|^2$$

quadratic function in β_j

$\Downarrow$

minimization reduces to solving a linear problem.

Example: Expectation value

- Given a set of N random numbers, find an estimate for the expectation value of the underlying probability distribution

y_i : data $\qquad E(y) = \mu$

μ is the parameter to optimize over (our "β")

$$S = \sum_i (y_i - \mu)^2$$

$$\frac{\partial S}{\partial \mu} = 2 \sum_i (\mu - y_i) = 0$$

$$N\mu = \sum_i y_i \qquad \mu = \frac{1}{N} \sum_i y_i = \langle y \rangle$$

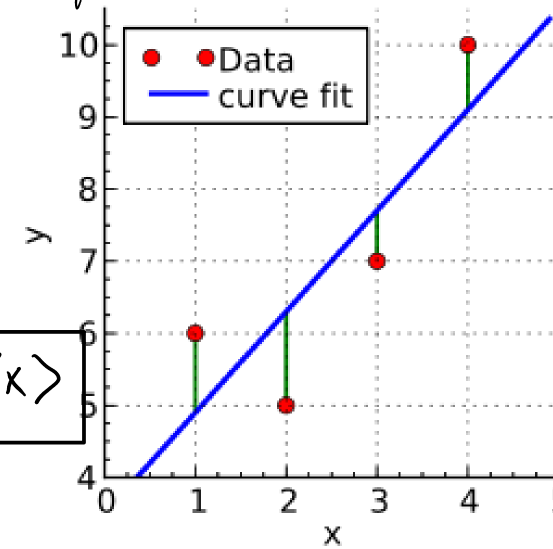
mean value

* the mean value of a set of measurements is the least squares estimator for the expectation value.

Example: Linear regression

- Given a set of measurements, find the parameters of a linear regression model

y_i : data model $y_i = m x_i + b$ $(m, b) = \beta$



$$S(m,b) = \sum_i |y_i - m x_i - b|^2$$

$$\frac{\partial S}{\partial b} = 2 \sum_i (m x_i + b - y_i) = 0$$

$$m \sum_i x_i + N b = \sum_i y_i \Rightarrow \boxed{b = \langle y \rangle - m \langle x \rangle}$$

$$\frac{\partial S}{\partial m} = 2 \sum_i x_i (m x_i + b - y_i) = 0$$

$$m \sum_i x_i^2 + b \sum_i x_i - \sum_i x_i y_i = 0 \Rightarrow \boxed{m \langle x^2 \rangle + b \langle x \rangle = \langle x y \rangle}$$

$$m = \frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x \rangle^2 - \langle x^2 \rangle}$$

$$b = \langle y \rangle - \langle x \rangle \left(\frac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle x \rangle^2 - \langle x^2 \rangle} \right)$$

$$= \frac{\langle y \rangle \langle x^2 \rangle + \langle x \rangle \langle xy \rangle}{\langle x^2 \rangle - \langle x \rangle^2}$$

Example: Deconvolution

- Problem

original image : f (β)

measured image : g (y)

convolution kernel : h (x)

model: $g = h * f$

Naive solution : take F.T. and divide

$G = H \cdot F \Rightarrow F = G/H$

noise!

in reality : measurement is $G + N$

$\Rightarrow F = G/H + N/H$ ← can become dominant at high spatial frequencies

cost function :

$$S = E\left[\sum_i |f_i - (w * g)_i|^2 \right]$$

solution : $W = \frac{H^*}{|H|^2 + \frac{P_N}{P_s}}$

Wiener filter

Original Blurred Wiener filtered

General linear least squares

Solve $y = X\beta$ by minimizing $S(\beta) = \sum_i |y_i - (X\beta)_i|^2$

$$\frac{\partial S}{\partial \beta_j} = 0 \quad \Rightarrow \quad \sum_{ij} X_{ij} \beta_j X^*_{ij} = \sum_i X^*_{ij} y_j$$

$X^\dagger X$ is square and positive definite

$(X^\dagger X)^{-1}$ exists!

$$X^\dagger X \beta = X^\dagger y$$

$$\beta = \underbrace{(X^\dagger X)^{-1} X^\dagger}_{} y$$

Moore-Penrose pseudo-inverse

the "best" inverse in the least square sense

$(X^\dagger X)^{-1} X^\dagger$ can be <u>very</u> costly to compute!

Numerically, there exist methods to solve $\min_\beta S(\beta)$ without computing the pseudo-inverse

e.g. numpy.linalg.lstsq

"linear algebra" "least square"

General linear least squares

Another example: fitting a 2D plane in an image

model: $\underline{P = A + B\cdot i + C\cdot j}$

- i, j pixel indices (x)
- A, B, C parameters (β)
- input image I (y)

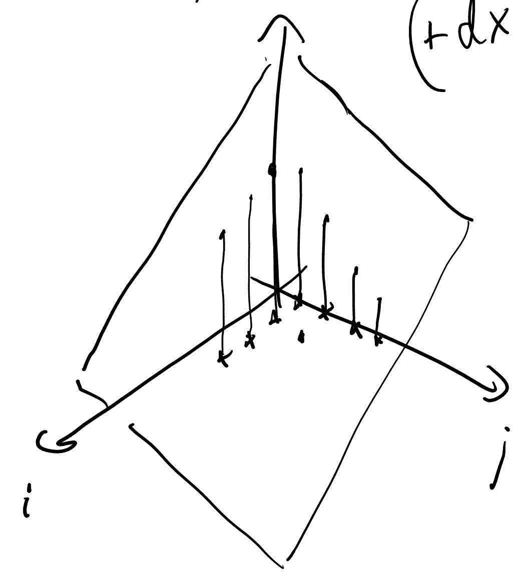
model: $y = X\beta$

$$\begin{pmatrix} I(0,0) \\ I(1,0) \\ I(2,0) \\ \vdots \\ I(N,0) \\ I(0,1) \\ \vdots \\ I(N,M) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & \vdots & 0 \\ 1 & \vdots & \vdots \\ 1 & \vdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & N & M \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

Image as a 1D vector

$f(x,y) = a + bx + cy$

$(+ dx^2 + exy + fy^2)$



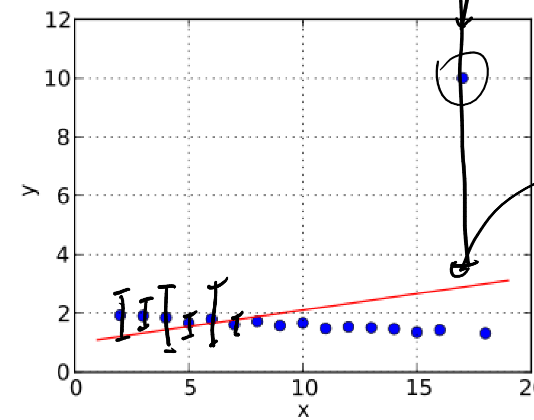
Weighted least squares

outlier

- Problem: sensitivity to outliers

weight

$$S = \sum_i w_i r_i^2$$



least square solution

w_i often related to uncertainty

$$w_i = \frac{1}{\sigma_i^2}$$

- Solution: penalize problematic values using weights

$$\|\ \ \|^2 = \sum_i |\ \ |^2$$

$$\hat{\beta} = \min_{\beta} \| w^{\frac{1}{2}} (X\beta - y) \|^2$$

$$S = \sum_i (\sqrt{w_i}\, r_i)^2$$

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

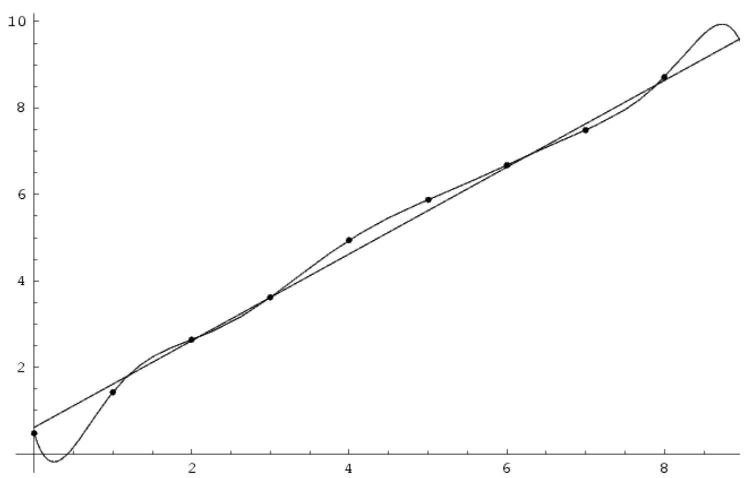
$$w = \begin{pmatrix} w_0 & & 0 \\ 0 & w_1 & \\ 0 & 0 & w_2 \\ & & & \ddots \end{pmatrix}$$

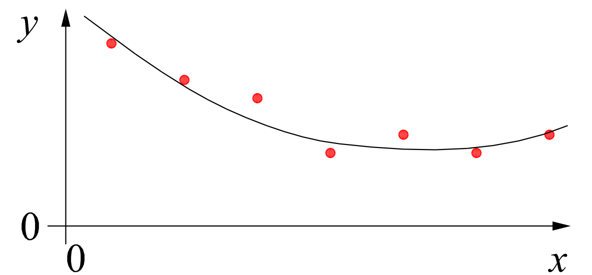
Solving least squares problems

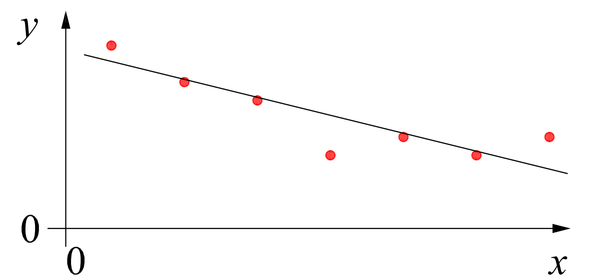
- Many approaches to solution exist
  - Pseudo inverse
  - Singular value decomposition (SVD)
  - QR decomposition
  - Iterative methods
  - ...
- Choice depends on
  - Robustness
  - Speed
  - Memory consumption
  - ...

Overfitting & ill-defined problems

- Guess can only be as good as the underlying model
- Too complicated models can lead to too complicated solutions







- Simultaneous optimization of model and its parameters
- Need *regularization* ← a way to reduce the effective number of degrees of freedom by imposing constraints between the parameters through additional terms in the cost function.

Lagrange multipliers

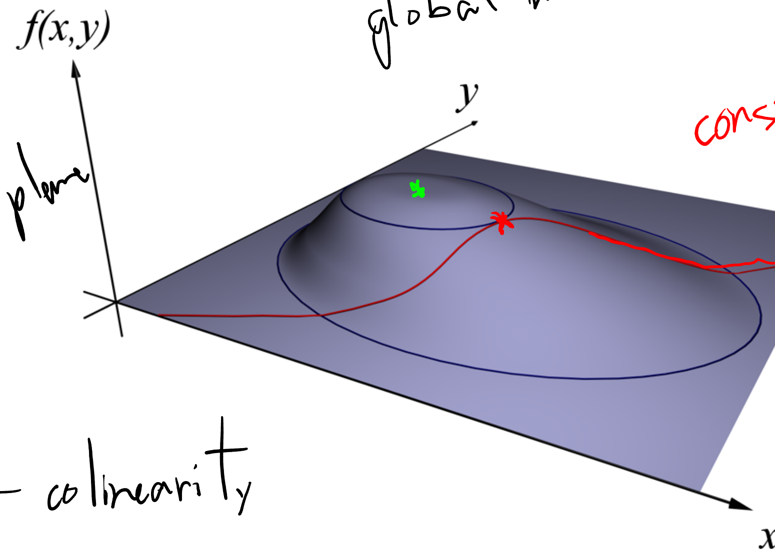
- Optimization under equality constraints

$\max_{x,y} f$ constraint $g(x,y) = c$ ← a curve in the (x,y) plane

↓

$\nabla f = 0$

instead of $\nabla f = 0$ $\nabla f = -\lambda \nabla g$ ← colinearity



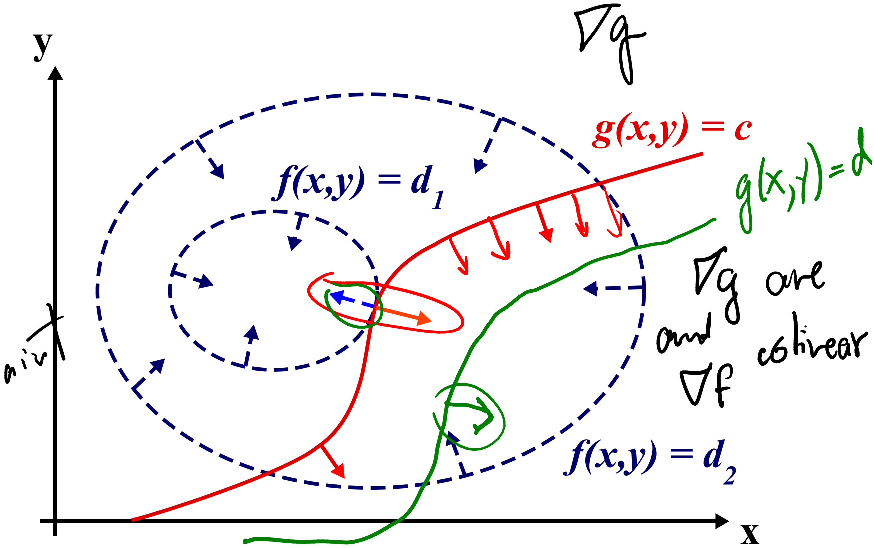
$$\Rightarrow \quad \mathcal{L} = f + \lambda\left(g(x,y) - c\right)$$

$$\nabla \mathcal{L} = \nabla f + \lambda \nabla g = 0$$

maximum

$$\frac{\partial \mathcal{L}}{\partial \lambda} = g(x,y) - c = 0$$

← satisfaction of the constraint



Tikhonov Regularization

Linear least squares:

$$S = \|y - X\beta\|^2 \quad \left(= \sum_i |y_i + (X\beta)_i|^2 \right) \qquad \|\beta\|^2 = \sum_i \beta_i^2$$

add a term

$$S' = \|y - X\beta\|^2 + \underbrace{\lambda \|\beta\|^2}_{\text{raises the total cost if any of the } \beta_i \text{ becomes large}}$$

avoids "explosion" of parameters. Used to make numerical problems well-conditioned

$$\nabla S' = 0 \quad \rightarrow \quad \beta \left(X^\dagger X + \lambda \mathbb{1} \right)^{-1} X^\dagger y$$

Nonlinear least squares

- If possible: linearize

not a linear function in β

$$S = \sum_i \left| y_i - f(x_i;\beta) \right|^2$$

initial guess

e.g. $\rightarrow$ $f(x_i;\beta) \simeq a(x_i) + b(x_i)(\beta - \beta_0)$

- Linearization not possible? $\longrightarrow$ iterative solution, brute force search, etc...

Example: Image registration

- Problem formulation: estimate the parameters of a transform s.t. the difference between original and distorted image is minimal

$$S = \sum_{i,j} \left| T(i,j) - B(i+i_0, j+j_0) \right|^2 = \sum_{i,j} T(i,j)^2 + \overbrace{\sum_{i,j} B(i+i_0, j+j_0)^2}^{*} - 2\underbrace{\sum_{i,j} T(i,j)\, B(i+i_0, j+j_0)}_{\text{cross-correlation}}$$

i_0, j_0 are the parameters to optimize

$$* : \sum_{i,j} M(i,j)\, B(i+i_0, j+j_0)$$

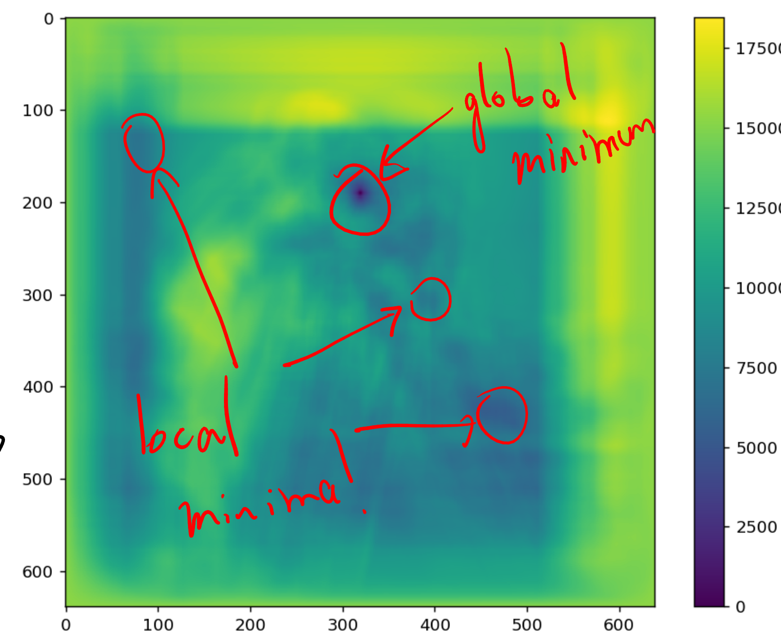
Mask

Base image

Template

at what (i,j) does the template best match the base image?

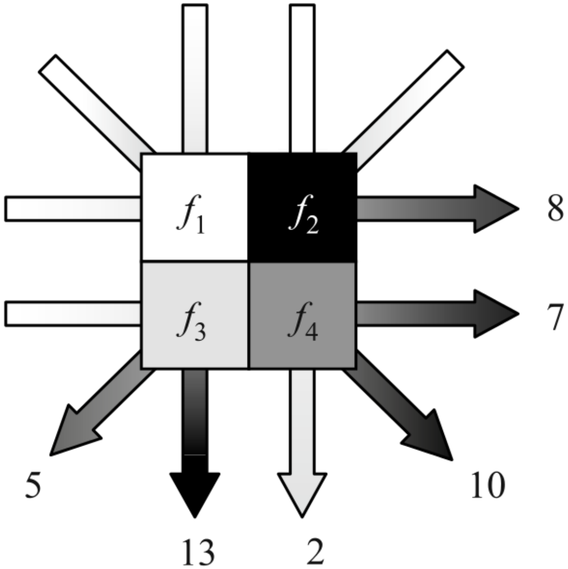
Distance map $= S$



Iterative solutions

- Move towards optimum in steps
  - Gradient descent
  - Newtons method
  - Gauss-Newton algorithm
  - Conjugate gradients
  - ...
- Projection onto constrain sets

Tomography revisited



Algebraic reconstruction techniques

Algebraic reconstruction techniques

Algebraic reconstruction techniques

Algebraic reconstruction techniques

Summary

- Approximate solutions can be found using estimation
- Approximation quality can be quantified by cost function
- Optimum solution is found by minimizing the cost function
- Least square estimator minimizes squared residues
- Lagrange multipliers can be used to implement additional constraints
- Iterative schemes allow solution of hard problems