Least-Squares Fitting of Model Parameters to Experimental Data

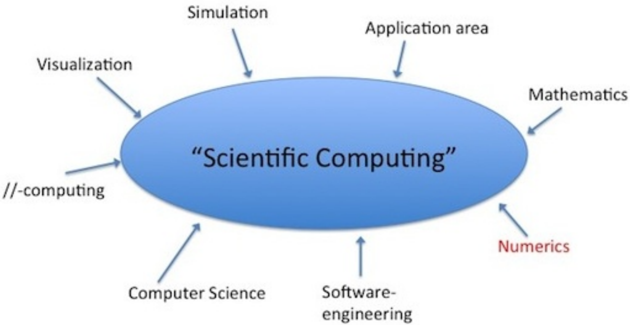
Inge Söderkvist

Div. of Mathematical Sciences, Dept of Engineering Sciences and Mathematics, LTU, room E193

Outline of talk

- What about *Science* and *Scientific Computing*
- Linear least-squares
- Nonlinear least-squares
- Separable least-squares (the variable projection method)
- Rigid body movements

Scientific Computing



Goal

Computational tools for obtaining accurate solutions in short time

Least-squares fitting

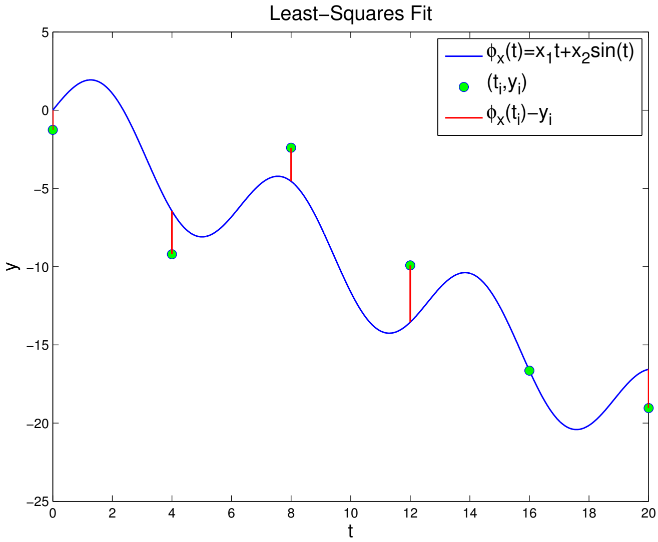
Given:

- Data points (t_i, y_i), $i = 1, \ldots, m$.
- A model function $\phi_x(t)$ that depends on some unknown parameters $x \in \mathbb{R}^n$, $m > n$.

Task:

Find parameters x by solving

$$\min_x \sum_{i=1}^{m} (\phi_x(t_i) - y_i)^2.$$


Least−Squares Fit
ϕx(t)=x1t+x2sin(t)
(ti,yi)
ϕx(ti)−yi
y
t
5
0
−5
−10
−15
−20
−25
0
2
4
6
8
10
12
14
16
18
20

Least-squares fitting, cont

- Nice statistical properties when $y_i = \phi_x(t_i) + e_i, \quad e_i$ comes from the exponential family.
- Implies easy computations compared to other measures.
- Sensitive to outliers.
- Many variants exists: error in variables, total lsqr, weighted lsqr, constrained lsqr, regularized lsqr,....

Formulated as a constrained problem

$$\begin{aligned} \min_{x} \quad & \sum_{i=1}^{m} (\tilde{y}_i - y_i)^2, \\ \text{s.t.} \quad & \phi_x(t_i) - \tilde{y}_i = 0, \; i = 1, \ldots, m. \end{aligned}$$

Linear least-squares

- $\phi_x(t)$ depends linearly on x. (Ex: $\phi_x(t) = x_1 t + x_2 \sin t$)
- We solve

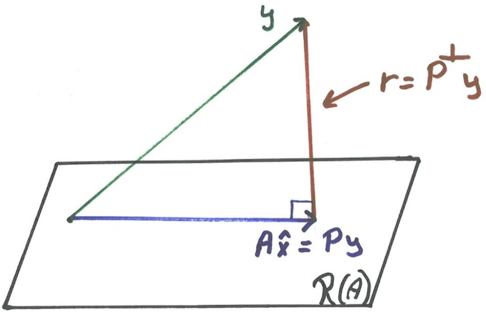
$$\min_x \|Ax - y\|_2^2, \text{ (Ex: } A = \begin{bmatrix} t_1 & \sin(t_1) \\ \vdots & \vdots \\ t_m & \sin(t_m) \end{bmatrix})$$

- The solution $\hat{x} = (A^T A)^{-1} A^T y = A^\dagger y$ satisfies

$$A\hat{x} = Py, \qquad \|A\hat{x} - y\|_2 = \|r\|_2 = \|P^\perp y\|_2$$

where P and $P^\perp$ are matrices that projects onto $\mathcal{R}(A)$ and onto $\mathcal{R}(A)^\perp$, respectively.

Projections



Let $A = QR,\ \ Q \in \mathrm{R}^{m\times n},\ Q^TQ = I,\ R \in \mathrm{R}^{n\times n}$ (qr-decomp)

$$A\hat{x} = Py \Leftrightarrow QR\hat{x} = QQ^Ty \Leftrightarrow R\hat{x} = Q^Ty$$

Numerically stable computations ! (e.g., $x = A\backslash y$ in Matlab)

Linear least-squares in matlab

How to solve $\min_x \|Ax - y\|_2$ in matlab ??

1. $x = A\backslash y$ (QR-factorization of A)
2. $x = pinv(A) * y$ (SVD of A)
3. $x = (A' * A)\backslash(A' * y)$ (Cholesky-factorisation of $A^T A$)
4. $x = inv(A' * A) * (A' * y)$ (Avoid !!)

Nonlinear least-squares

- $\phi_x(t)$ depends nonlinearly on x, (e.g., $\phi_x(t) = e^{x_1 t} + \sin(x_2 t)$)
- Define $f_i(x) = \phi_x(t_i) - y_i$.
- We minimize

$$F(x) = \frac{1}{2}\sum_{i=1}^{m} f_i^2(x) = \frac{1}{2}\|f(x)\|_2^2$$

using iterative methods, e.g., the Gauss-Newton method.

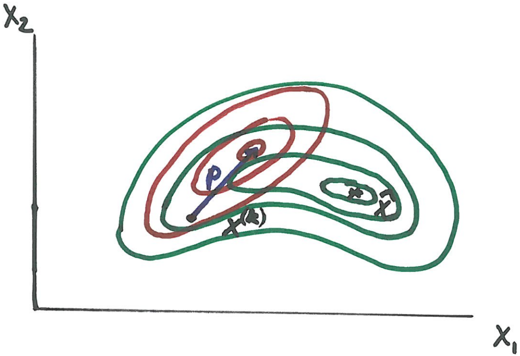
Newtons Method

Newtons method for or solving $\min_x F(x)$ or $\nabla_x F(x) = 0$:

$k = 0$
guess $x^{(0)}$
Repeat
    Compute Hessian matrix $H = \nabla^2 F(x^{(k)}) \in \mathrm{R}^{n \times n}$,
    Compute gradient $g = \nabla F(x^{(k)}) \in \mathrm{R}^n$
    Solve the linear system of equations $Hp = -g$
    Compute step length β
    $x^{(k+1)} = x^{(k)} + \beta\, p$
    $k = k + 1$
Until convergence

Based on $F(x^{(k)} + p) \approx F(x^{(k)}) + p^T g + \frac{1}{2} p^T H p$.

Geometry Newtons method



$$F(x) = k$$

$$F(x^{(k)}) + p^T g + \frac{1}{2} p^T H p = k$$

Gauss-Newton method

Here: $F(x) = \frac{1}{2}\sum_{i=1}^{m} f_i^2(x) = \frac{1}{2}\|f(x)\|_2^2$

$H = (J^T J + \sum_i (f_i \nabla_x^2 f_i)), \;\; g = J^T f \in \mathrm{R}^n,$

where $J = \nabla f(x^{(k)}) \in \mathrm{R}^{m\times n}$ is the Jacobian matrix

Hence:

$$\begin{aligned} p_{New} &= H^{-1}g = -(J^T J + \sum_i (f_i \nabla_x^2 f_i))^{-1} J^T f \\ p_{GN} &= -(J^T J)^{-1} J^T f \end{aligned}$$

Note: p_{GN} solves $\min_p \|f(x^{(k)}) + Jp\|_2$ (e.g., $p = -J\backslash f$ in Matlab)

Gauss Newtons Method

Gauss-Newtons method for or solving $\|f(x)\|_2$:

$k = 0$
guess $x^{(0)}$

Repeat

Compute $J = \nabla f(x^{(k)}) \in \mathrm{R}^{m \times n}$
Compute $f = f(x^{(k)}) \in \mathrm{R}^{m}$
Solve the linear least squares problem $\min_p \|f(x^{(k)}) + Jp\|_2$.
Compute step length β
$x^{(k+1)} = x^{(k)} + \beta\, p$
$k = k + 1$

Until convergence

Based on $\|f(x^{(k)} + p)\|_2 \approx \|f(x^{(k)}) + Jp\|_2$.

$F(x^{(k)} + p) \approx F(x^{(k)}) + p^T J^T f + \frac{1}{2} p^T J^T J p$.

Properties of Gauss-Newtons method

- Only first order derivatives are needed
- p_{GN} is a descent direction if J has full rank.
- Asymptotic rate of convergence is linear (quadratic if $f(\hat{x}) == 0$)

$$\lim_{k\to\infty} \frac{\|x^{(k+1)} - \hat{x}\|}{\|x^{(k)} - \hat{x}\|} = \frac{\|f(\hat{x})\|}{\rho}, \ \rho = \text{ radius of curvature.}$$

- Converges to local minima
- Global properties ????

Difficulties

- How to find initial guess $x^{(0)}$?
- How to deal with ill-conditioned Jacobians ?

$$p_{LM} = -(J^T J + \lambda I)^{-1} J^T f \text{ (Levenberg-Marquardt)}$$

$$\min_x (1-\lambda) \|f(x)\|_2^2 + \lambda \text{ regularization term}$$

- How to deal with slow rate of convergence ?

  Use higher order derivatives (Newtons method)

Separable least-squares

Example: $\phi_x(t) = x_1 e^{x_3 t} + x_2 e^{x_4 t} = a_1 e^{b_1 t} + a_2 e^{b_2 t}$.

- ϕ depends linearly on $a_1, \ldots, a_{n_1}$
- ϕ depends nonlinearly on $b_1, \ldots, b_{n_2}$
- The least-squares problem is

$$\min_{a,b} \frac{1}{2} \|y - A(b)a\|_2^2,$$

Example:

$$A(b) = \begin{bmatrix} e^{b_1 t_1} & e^{b_2 t_1} \\ \vdots & \vdots \\ e^{b_1 t_m} & e^{b_2 t_m} \end{bmatrix}$$

The variable projection method

For a given vector b the solution $\hat{a}$ satisfies

$$\|y - A(b)\hat{a}\|_2 = \|P(b)^\perp y\|_2, \qquad (P(b)^\perp \text{ projects onto } \mathcal{R}(A(b))^\perp)$$

The original problem has been transformed to

Nonlinear projected problem

$$\min_b \frac{1}{2}\|P(b)^\perp y\|_2^2,$$

The variable projection method

$$\min_b \frac{1}{2}\|P(b)^{\perp}y\|_2^2,$$

- The dimension of the problem is n_2 instead of $n_1 + n_2$
- Can be solved with standard methods (GN,New). How to obtain derivatives ??
- Asymptotic convergence with GN is almost the same as when applying GN on the original problem (Ruhe,Wedin, 1980).
- Global properties ???

How to find derivatives

Notation: $C_i = \frac{\partial C(b)}{\partial b_i}, \quad C_{ij} = \frac{\partial^2 C(b)}{\partial b_i \partial b_j}$.

$$P_i^\perp = -P^\perp A_i A^\dagger - (P^\perp A_i A^\dagger)^T,\ i = 1, \dots, n_2 \,.$$

(Golub, Pereyea, 1973). Last term can be ignored.

$$P_{ij}^\perp = -B_{ij} - B_{ij}^{\mathrm{T}},\ i, j = 1, \dots, n_2,\ \text{where}$$

$$B_{ij} = -P_j A_i A^\dagger + P^\perp A_{ij} A^\dagger + P^\perp A_i A^\dagger ((A_j A^\dagger)^{\mathrm{T}} P^\dagger - A_j A^\dagger)$$

(Borges 2009).

The Jacobian matrix, J, of $r = P(b)^{\perp} y$

Let

- $P_i^{\perp} \approx -P^{\perp} A_i A^{\dagger}$
- $Q \in \mathrm{R}^{m \times n_1}$ be an orthogonal base for $\mathcal{R}(A(b))$.
- $\hat{a} = A^{\dagger} y$

$$J(:, i) = -P_i^{\perp} y = -P^{\perp} A_i A^{\dagger} y = -(I - QQ^T) A_i \hat{a}$$

Examples of separable nonlinear least squares problems

- Sums of exponentials
- NURBS
- Rigid body movements

Sum of exponentials

$$\phi_{a,b}(t) = \sum_{i=0}^{N} a_i e^{b_i t}$$

- Literature reports global benefits using the variable projection method.
  (see e.g., Golub ,Pereyra 2003 and references therein)

NURB-curves

$$c(t) = \frac{\sum_{i=0}^{n} B_i(t) w_i p_i}{\sum_{i=0}^{n} B_i(t) w_i} = \phi_{a,b}(t) = \frac{\sum_{i=0}^{n} B_i(t) a_i}{\sum_{i=0}^{n} B_i(t) b_i},$$

- The variable projection method is most often slower than using the unseparated formulation. (see Bergström, Söderkvist, 2012)

Rigid body movements

$$\min_{R,d} \quad \sum_{i=1}^{m} \| R p_i + d - q_i \|_2^2$$

$$s.t. \quad R^T R = I, \;\; det(R) = 1$$

- Separable least-squares problem with special structure. No standard iteration is needed.
- Applications in orthopedics, robotics, computer vision, reverse engineering, automatic shape verification, etc..
- *Determining the movements of the skeleton using well-configured markers* by I Söderkvist, PÅ Wedin, Journal of biomechanics, 1993 . Over 600 citations

Rigid body movements, cont.

The orthogonal Procrustes problem

$d = \overline{q} - R\overline{p}$

$$\min_{R \in \Omega} \|RA - B\|_F$$

where $A = (p_1 - \overline{p}, \ldots, p_m - \overline{p}), \quad B = (q_1 - \overline{q}, \ldots, q_m - \overline{q})$

Solved using SVD, $U\Sigma V^T = BA^T \in \mathrm{R}^{3\times 3}, \quad R = UV^T$

- Efficient and stable computational solution procedure.
- Needs coordinates of a ordered set of landmarks.
- $(R = U \, diag(1, 1, -1) \, V^T$ if $det(UV^T) = -1)$

Procrustes

PROCRUSTES, also called POLYPEMON or DAMASTES, in Greek legend, a robber dwelling in the neighbourhood of Eleusis, who was slain by Thesus (q.v). He had an iron bed (or according to some accounts, two beds) on which he compelled his victims to lie, stretching or cutting off their legs to make them fit the bed's length. The bed of Procrustes"has become proverbial for inflexibility. (Encyclopædia Britannica. Vol 18, 1964)

Condition numbers

Assume that

- $B = RA, \;\; A, B \in R^{3\times n}, \;\; R \in \Omega$
- A and B have the singular values $\sigma_1 \leq \sigma_2 \leq \sigma_3 \leq 0$, where $\sigma_2 > 0$.

Let the orthogonal matrix $R + \Delta R$ be the solution to the perturbed problem

$$\min_{(R+\Delta R)\in\Omega} \|(R + \Delta R)(A + \Delta A) - (B + \Delta B)\|.$$

Condition numbers

A first order bound of $\|\Delta R\|$ is given by

$$\lim_{\substack{\varepsilon_A \to 0 \\ \varepsilon_B \to 0}} \sup_{\substack{\|\Delta A\| \leq \varepsilon_A \\ \|\Delta B\| \leq \varepsilon_B}} \frac{\|\Delta R\|}{\|\Delta A\| + \|\Delta B\|} = \frac{\sqrt{2}}{(\sigma_2^2 + \sigma_3^2)^{\frac{1}{2}}}$$

- The square sum of the distances of the landmarks to the closest straight line equals $\sigma_2^2 + \sigma_3^2$.

Conclusions

- Computational methods for solving least squares problem are useful in modelling.
- Numerical optimization and linear algebra provide important basic tools.
- The structure of the problem can often be utilized to gain efficiency and robustness.

Thanks for your attention !