Lecture 6: Linear and nonlinear least-squares problems; the Gauss-Newton method

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Linear and nonlinear least-squares problems

- a way to solve overdetermined (linear and nonlinear) systems of equations:

  $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m \geq n$; $r(x) = 0$ or $r(x) \approx 0$.

$$\Downarrow$$

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \sum_{j=1}^{m} [r_j(x)]^2 = \frac{1}{2} \|r(x)\|^2. \qquad \text{(L/N LS)}$$

  $\Longrightarrow$ unconstrained optimization problems with special structure.

- often, computationally cheaper to solve if structure is exploited:

  $\longrightarrow$ "simplify" damped Newton's method to exploit this structure.

- many applications: data fitting, data assimilation for weather forecasting, climate modelling, etc.

Data fitting application

Times $t_j \quad \longrightarrow \quad y_j$, $j = \overline{1,m}$, measurements.

Model: $\Phi(x,t)$, continuous in t; parameters $x \in \mathbb{R}^n$, $n < m$.

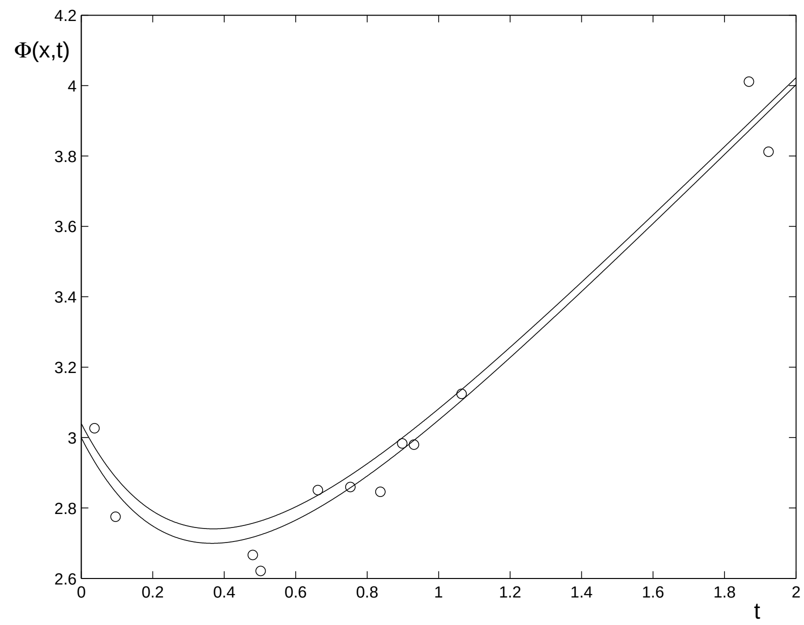
Find x: $\Phi(x,t_j)$ “close to” y_j, $j = \overline{1,m}$;

Choice of model: $\Phi(x,t) = x_1 + x_2 t + e^{-x_3 t}$, where $x = (x_1, x_2, x_3) \in \mathbb{R}^3$.

$$\min_{x \in \mathbb{R}^3} \frac{1}{2} \sum_{j=1}^{m} (\Phi(x,t_j) - y_j)^2.$$

$\longrightarrow x^*$.

Optimal model: $\Phi(x^*,t)$.



- In (NLS), let $r_j(x) := \Phi(x,t_j) - y_j$, $j = \overline{1,m}$: residuals.

The Linear Least-Squares (LLS) problem

- $r(x) := Jx + r$, $\forall\, x \in \mathbb{R}^n$; $J \in \mathbb{R}^{m\times n}$, $r \in \mathbb{R}^m$, $m \geq n$.

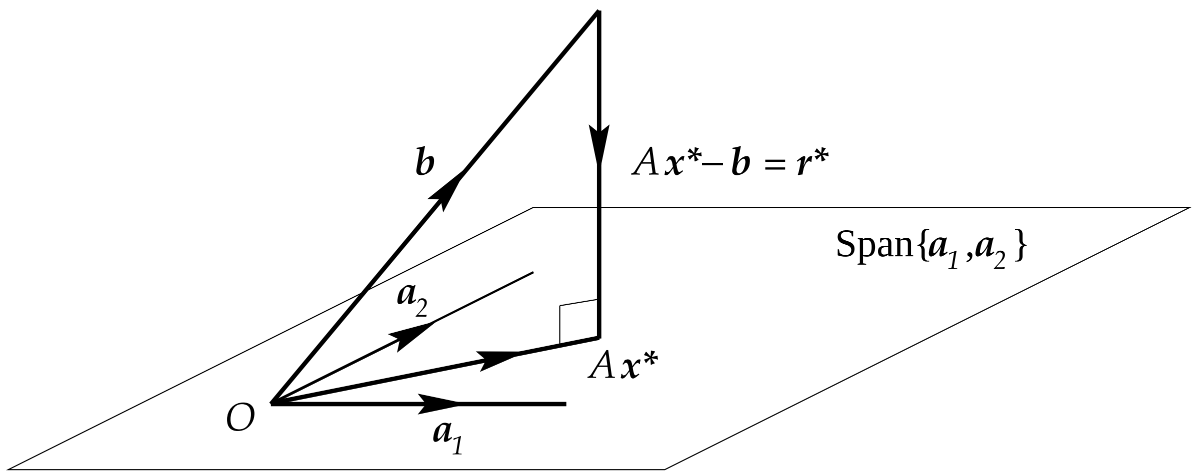
$$\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2}\|Jx + r\|^2. \qquad \text{(LLS)}$$

- f convex quadratic; (global) minimizer x^* of f == solution of linear system (normal equations)

$$J^T(Jx^* + r) = 0 \iff J^T Jx^* = -J^T r.$$

Geometrical interpretation:

- $r(x) = Ax - b$.
LLS: find orthogonal projection of b onto the subspace/plane determined by the columns of A.



- computing x^*: Cholesky factorization of $J^T J$; QR or SVD of J.

A simple LLS example

Fit a line to the data $(t_i, y_i) \in \{(-1,3),(0,2),(1,0),(2,4)\}$.

- for some $x = (x_1 \ x_2)^T \in \mathbb{R}^2$, $\Phi(x,t) := x_1 + x_2 t$, $t \in \mathbb{R}$, defines a line.
- determine $x = (x_1 \ x_2)^T$ as solution of (LLS)

$$\min_{x \in \mathbb{R}^2} \sum_{i=1}^{4} \|\Phi(x,t_i) - y_i\|^2.$$

$$\Phi(x,t_i) - y_i = 0,\ i = \overline{1,4} \quad \Leftrightarrow \quad \begin{cases} x_1 & -x_2 & = & 3 \\ x_1 & & = & 2 \\ x_1 & +x_2 & = & 0 \\ x_1 & +2x_2 & = & 4. \end{cases}$$

A simple LLS example ...

Let J matrix of system; x^* LLS solution iff $J^T J x^* = J^T y$.

$$\begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 9 \\ 5 \end{pmatrix},$$

$\Leftrightarrow \; x^* = (2.2, 0.1)$ and $\Phi(x^*, t) = 2.2 + 0.1t$.

Nonlinear Least-Squares (NLS)

- $r : \mathbb{R}^n \to \mathbb{R}^m$ with $m \geq n$; r smooth.

$$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \sum_{j=1}^{m} [r_j(x)]^2 = \frac{1}{2} \|r(x)\|^2. \qquad \text{(NLS)}$$

- $r(x^*) = 0$: zero-residual pb.; $r(x^*) \neq 0$: nonzero-residual pb.

- $\nabla f(x) = J(x)^T r(x)$, where $J(x)$ Jacobian of r at x.
- $\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x) \nabla^2 r_j(x)$.

- (Damped and modified) Newton's method for minimization may be applied to f: $\quad \nabla^2 f(x) s^k = -\nabla f(x)$.

- $r_j(x^*) \approx 0$ or $\nabla^2 r_j(x^*)$ small $\Longrightarrow$ $r_j(x) \nabla^2 r_j(x)$ small when x close to x^* $\Longrightarrow$ $\nabla^2 f(x) \approx J(x)^T J(x) := \widetilde{\nabla^2} f(x)$.

- $J(x)^T J(x)$ positive semidefinite; if $J(x)$ full column rank $\Rightarrow J(x)^T J(x)$ positive definite.

Gauss-Newton method for nonlinear least-squares

- GN direction:

$$\widetilde{\nabla^2 f}(x^k)s^k = -\nabla f(x^k) \Longleftrightarrow J(x^k)^T J(x^k)s^k = -J(x^k)^T r(x^k),$$

and so s^k solves the (LLS):

$$\min_{s\in\mathbb{R}^n} \frac{1}{2}\|J(x^k)s + r(x^k)\|^2.$$

$\longrightarrow$ f approx. by convex quadratic model for each k.

- s^k descent provided $J(x^k)$ full column rank!

Gauss-Newton method for nonlinear least-squares:

Choose $\epsilon > 0$ and $x^0 \in \mathbb{R}^n$.
While $\|\nabla f(x^k)\| > \epsilon$, REPEAT:

- solve the linear system $\widetilde{\nabla^2 f}(x^k)s^k = -\nabla f(x^k)$.
- set $x^{k+1} = x^k + \alpha^k s^k$, with $\alpha^k \in (0, 1]$; $k := k + 1$.

END.

Convergence properties of Gauss-Newton method

- $\nabla f(x) = 0$ may not imply $r(x) = 0$
- (global convergence) $J(x^k)$ uniformly full-rank for all x^k (etc.) $\Longrightarrow$ $\|\nabla f(x^k)\| = \|J(x^k)^T r(x^k)\| \to 0$, $k \to \infty$.
- (local convergence) if $r(x^*) = 0$ and $J(x^*)$ full-rank (etc.) $\Longrightarrow$ $x^k \to x^*$ quadratically.

Gauss-Newton vs. Newton method:

- computational cost per iteration: N > GN.
- N direction may be ascent.
- only linear rate for GN when $r(x^*) \neq 0$.
- N & GN mthds unreliable without a linesearch (or other safeguards). Use bArmijo linesearch for example.

Gauss-Newton vs. Newton: an example

- $r : \mathbb{R} \to \mathbb{R}^2$; $r(x) := (x + 1 \quad 0.1x^2 + x - 1)^T$
- $r(x^*) = (1, \ -1)^T \neq 0 \ \longrightarrow$ nonzero residuals problem: only linear convergence asymptotically for GN.

	1	2	3	4	5	6
N	1.0	0.14	0.003	$1.5 \cdot 10^{-6}$	$4.3 \cdot 10^{-13}$	$3.1 \cdot 10^{-26}$
GN	1.0	0.13	0.014	0.0014	0.00014	0.000014