Numerical Methods I: Linear least squares

Georg Stadler
Courant Institute, NYU
stadler@cims.nyu.edu

September 28, 2017

Least-squares problems

Given data points/measurements

$$(t_i, b_i), \quad i = 1, \ldots, m$$

and a model function ϕ that relates t and b:

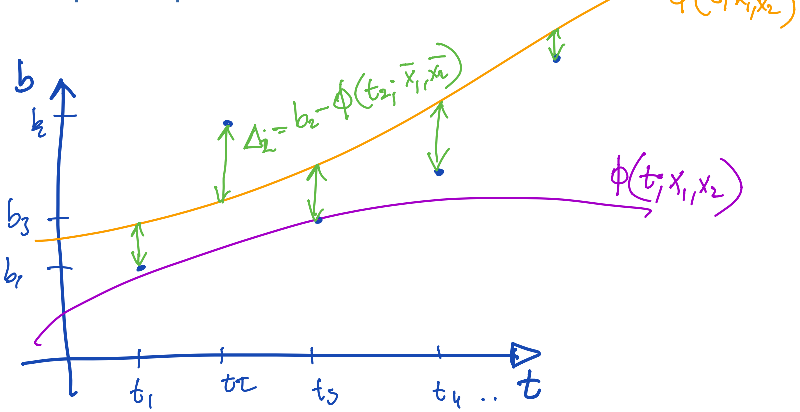
$$b = \phi(t; x_1, \ldots, x_n),$$

where $x_1, \ldots, x_n$ are model function parameters. If the model is supposed to describe the data, the deviations/errors

$$\Delta_i = b_i - \phi(t_i, x_1, \ldots, x_n)$$

should be small. Thus, to fit the model to the measurements, one must choose $x_1, \ldots, x_n$ appropriately.

Least-squares problems



Example: 1.) $\phi(t; x_1, x_2) = x_1 t^2 + x_2 \exp(t) \dots$ parameters x_1, x_2 enter linearly

"linear least squares"

2.) $\tilde{\phi}(t; x_1, x_2) = x_1^2 t^2 + \log(x_2)\, t^3$ $\quad x_1, x_2$ enter non-linearly

Least squares: Find $x_1, \ldots, x_n$ such that

$$\frac{1}{2}\sum_{i=1}^{m} \Delta_i^2 \to \min$$

From a probabilistic perspective, this corresponds to an underlying Gaussian error model.

Weighted least squares: Find $x_1, \ldots, x_n$ such that

$$\frac{1}{2}\sum_{i=1}^{m} \left(\frac{\Delta_i}{\delta b_i}\right)^2 \to \min,$$

where $\delta b_i > 0$ contain information about how much we trust the ith data point.

Alternatives to using squares:

L^1 error: Find $x_1, \ldots, x_n$ such that

$$\sum_{i=1}^{m} |\Delta_i| \to \min$$

Result can be very different, other statistical interpretation, more stable with respect to outliers.

L^∞ error: Find $x_1, \ldots, x_n$ such that

$$\max_{1 \leq i \leq m} |\Delta_i| \to \min$$

Linear least-squares

We assume (for now) that the model depends linearly on $x_1, \ldots, x_n$, e.g.:

$$\phi(t; x_1, \ldots x_n) = a_1(t)x_1 + \ldots + a_n(t)x_n$$

Ex: $\phi(t; x_1, x_2) = x_1 t^2 + x_2 \exp(t)$, data points $(t_i, b_i), i = 1, 2, 3$

$\Delta_1 = b_1 - (x_1 t_1^2 + x_2 \exp(t_1))$

$\Delta_2 = b_2 - (x_1 t_2^2 + x_2 \exp(t_2))$

$\Delta_3 = b_3 - (x_1 t_3^2 + x_2 \exp(t_3))$

$\min_{x_1, x_2} \sum_{i=1}^{3} \Delta_i^2 \longrightarrow$ linear least squares problem

unknowns: x_1, x_2 (2 unknowns), 3 equations

Linear least-squares

Choosing the least square error, this results in

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2,$$

where $\boldsymbol{x} = (x_1, \dots, x_n)^T$, $\boldsymbol{b} = (b_1, \dots, b_m)^T$, and $a_{ij} = a_j(t_i)$.

In the following, we study the overdetermined case, i.e., $m \geq n$.

<u>In example:</u>

$$A = \underbrace{\begin{bmatrix} t_1^2 & \exp(t_1) \\ t_2^2 & \exp(t_2) \\ t_3^2 & \exp(t_3) \end{bmatrix}}_{\in \mathbb{R}^{3\times 2}} \qquad x = \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\in \mathbb{R}^2}, \qquad b = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\in \mathbb{R}^3}$$

Linear least-squares

Different perspective:

Consider non-square matrices $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and rank(A) $= n$. Then the system

$$A\boldsymbol{x} = \boldsymbol{b}$$

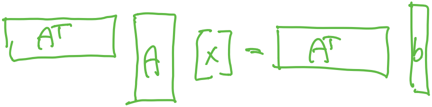
does, in general, not have a solution (more equations than unknowns). We thus instead solve a minimization problem

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2.$$

2-norm, Euclidean norm

The minimum $\bar{\boldsymbol{x}}$ of this optimization problem is characterized by the normal equations:

$$A^T A \bar{\boldsymbol{x}} = A^T \boldsymbol{b}.$$



Linear least-squares: normal equations

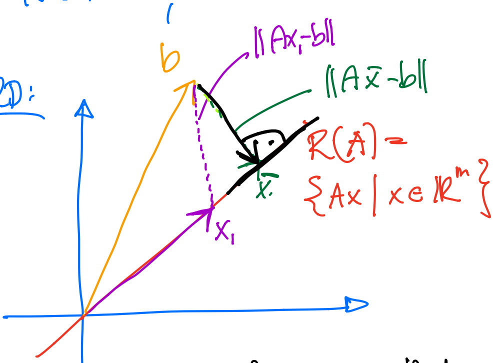
$$\min_{x \in \mathbb{R}^m} \|Ax-b\|_2^2 \qquad A \in \mathbb{R}^{m\times n}, \quad b \in \mathbb{R}^n$$

Geometrical sketch: in 2D:

$$\bar{x} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \|Ax-b\|_2^2$$

$$\Updownarrow$$

$$A\bar{x}-b \perp R(A)$$



Note: Such a characterization is only possible for norms that have a corresponding inner product $\langle \cdot, \cdot \rangle$.

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Linear least-squares: normal equations

Lemma (3.4), V, $\langle \cdot, \cdot \rangle$ finite-dim. vector space, $U \subset V$ subspace

$U^\perp$ is orthogonal complement, i.e. $U^\perp = \{ v \in V \mid \langle v, u \rangle = 0 \text{ for all } u \in U \}$

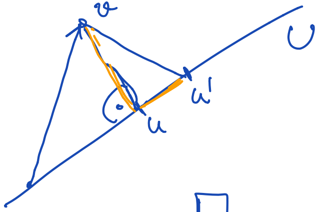
$v \in V$

Then: $\operatorname{argmin}_{u' \in U} \| u' - v \| = u \iff v - u \in U^\perp$

Proof: Let $u \in U$ such that $v - u \in U^\perp$. Then for $u' \in U$:

$$\| v - u' \|^2 = \| v - u \|^2 + 2 \underbrace{\langle v - u, u - u' \rangle}_{=0} + \| u - u' \|^2$$

$$= \| v - u \|^2 + \| u - u' \|^2 \geq \| v - u \|^2$$

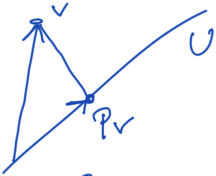


$\square$.

$\longrightarrow$ ✓

Linear least-squares: normal equations

This unique $u \in U$ is called the orthogonal projection of v onto U, $u = Pv$



P is linear

Thm: $\bar{x} = \underset{x \in \mathbb{R}^n}{\arg\min} \|Ax - b\|^2 \iff A^T A x = A^T b$

In particular, minimum is unique if $\operatorname{rank}(A) = n$ since then $A^T A$ is invertible.

Proof: $V = \mathbb{R}^m$, $U = R(A) \subset V$ subspace

$\|b - Ax\| = \min \iff \langle b - Ax, Ax' \rangle = 0 \quad \forall x' \in \mathbb{R}^n$

$\iff \langle A^T(b - Ax), x' \rangle = 0 \quad \forall x' \in \mathbb{R}^n$

$\iff A^T A x = A^T b$ $\square$.

Linear least-squares problems–QR factorization

Solving the normal equations

$$A^T A\bar{\boldsymbol{x}} = A^T\boldsymbol{b}$$

requires: $n\overset{m}{\boxed{\quad\quad\quad}}$ $\boxed{\ }$

- computing $A^T A$ (which is $O(mn^2)$)
- condition number of $A^T A$ is square of condition number of A; (problematic for the Choleski factorization)

$B \in \mathbb{R}^{n\times n}$

$\|B\|_2 = |\lambda_{max}(B)|$

$\|B^{-1}\|_2 = |\lambda_{max}(B^{-1})| = \frac{1}{|\lambda_{min}(B)|}$

Claim:

$$\kappa_2(A^TA) = \|A^TA\|_2 \|(A^TA)^{-1}\|_2 = \kappa_2(A)^2$$

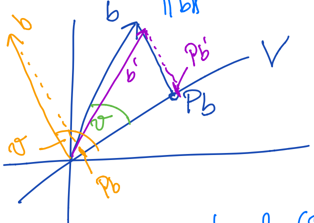
Proof:

$$\kappa_2(A)^2 \overset{\text{Def}}{=} \frac{\max_{\|x\|_2=1} \|Ax\|^2}{\min_{\|x\|_2=1} \|Ax\|^2} = \frac{\max_{\|x\|_2=1} \langle Ax, Ax\rangle}{\min_{\|x\|_2=1} \langle Ax, Ax\rangle}$$

$$= \frac{\max_{\|x\|=1} \langle A^TAx, x\rangle}{\min_{\|x\|=1} \langle A^TAx, x\rangle} = \frac{\lambda_{max}(A^TA)}{\lambda_{min}(A^TA)} = \kappa_2(A^TA)$$

Conditioning of an orthogonal projection:

Lemma: $P: \mathbb{R}^m \to V$ orthogonal projection onto $V \subset \mathbb{R}^n$.
For $b \in \mathbb{R}^n$, denote by ϑ angle between b and V
i.e. $\sin\vartheta = \dfrac{\|b - Pb\|_2}{\|b\|}$



Then the relative condition number of (P, b) w.r.t. the 2-norm is $\kappa = \dfrac{1}{\cos\vartheta}\|P\|_2$

Proof: $\|Pb\|^2 = \|b\|^2 - \|b - Pb\|^2$

$$\Rightarrow \frac{\|Pb\|^2}{\|b\|^2} = 1 - \sin^2\vartheta = \cos^2\vartheta$$

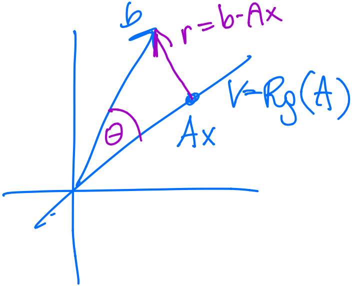
P is linear, i.e. $\kappa_{rel} = \dfrac{\|b\|_2}{\|Pb\|_2}\|P'(b)\|_2 = \dfrac{1}{\cos\vartheta}\|P\|_2$

$\square$

Solving the normal equation is equivalent to computing $P\boldsymbol{b}$, the orthogonal projection of $\boldsymbol{b}$ onto the subspace V spanned by columns of A.

Let $\boldsymbol{x}$ be the solution of the least square problem and denote the residual by $\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x}$, and

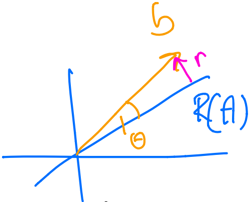
$$\sin(\theta) = \frac{\|\boldsymbol{r}\|_2}{\|\boldsymbol{b}\|_2}.$$



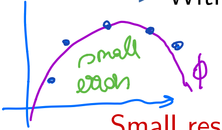
The relative condition number κ of $\boldsymbol{x}$ in the Euclidean norm is bounded by

- With respect to puerturbations in $\boldsymbol{b}$:

$$\kappa \leq \frac{\kappa_2(A)}{\cos(\theta)}$$



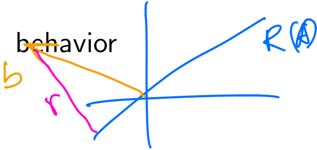
- With respect to perturbations in $\boldsymbol{A}$:

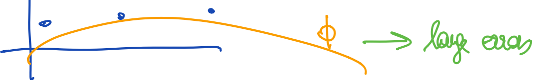


$$\kappa \leq \kappa_2(A) + \kappa_2(A)^2 \tan(\theta)$$

Small residual problems $\cos(\theta) \approx 1$, $\tan(\theta) \approx 0$: behavior similar to linear system.

Large residual problems $\cos(\theta) \ll 1$, $\tan(\theta) > 1$: behavior essentially different from linear system.





Linear least-squares problems–QR factorization

One would like to avoid the multiplication $A^T A$ and use a suitable factorization of A that aids in solving the normal equation, the QR-factorization:

$$\boxed{A} = \overbrace{\boxed{Q_1 \mid Q_2}}^{Q} \, \overbrace{\boxed{\begin{matrix} R_1 \\ \hline 0 \end{matrix}}}^{R}$$

$$A = QR = \left[Q_1, Q_2\right] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1,$$

where $Q \in \mathbb{R}^{m \times m}$ is an orthonormal matrix ($QQ^T = I$), and $R \in \mathbb{R}^{m \times n}$ consists of an upper triangular matrix and a block of zeros.

How can the QR factorization be used to solve the normal equation?

$Q^TA = Q^TQR = R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$

$$\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \min_{\boldsymbol{x}} \|Q^T(A\boldsymbol{x} - \boldsymbol{b})\|^2 = \min_{\boldsymbol{x}} \| \begin{bmatrix} \boldsymbol{b}_1 - R_1\boldsymbol{x} \\ \boldsymbol{b}_2 \end{bmatrix} \|^2,$$

$= \min_{x} \|b_1 - R_1x\|^2 + \|b_2\|^2$

where $Q^T\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix}$.

Thus, the least squares solution is $\boldsymbol{x} = R^{-1}\boldsymbol{b}_1$ and the residual is $\|\boldsymbol{b}_2\|$.

Linear least-squares problems–QR factorization

$A = QR$

How can we compute the QR factorization?

Givens rotations

Use sequence of rotations in 2D subspaces:

For $m \approx n$: $\sim n^2/2$ square roots, and $4/3n^3$ multiplications

For $m \gg n$: $\sim nm$ square roots, and $2mn^2$ multiplications

Householder reflections

Use sequence of reflections in 2D subspaces

For $m \approx n$: $2/3n^3$ multiplications

For $m \gg n$: $2mn^2$ multiplications