5 Least Squares

slide 28

goal: Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^n$, determine a "reasonable" solution to

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{5.1}$$

.

Remark 5.1 *For $m > n$, problem (5.1) is* overdetermined *so one cannot expect existence of a classical solution. For $m < n$, problem (5.1) is* underdetermined *so one cannot expect uniqueness.* ■

A reasonable approach is to minimize the residual $\mathbf{b} - \mathbf{A}\mathbf{x}$ in some norm of interest. The ℓ^2-norm $\|\cdot\|_2$ is particularly convenient as we will later see.

Definition 5.2 (least squares solution) $\mathbf{x} \in \mathbb{R}^n$ *is called a* least squares solution *of* $\mathbf{A}\mathbf{x} = \mathbf{b}$*, if it solves the following minimization problem:*

$$\textit{Find } \mathbf{x} \in \mathbb{R}^n \ \textit{s.t.} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 = \min \{ \|\mathbf{b} - \mathbf{A}\mathbf{y}\|_2 \,|\, \mathbf{y} \in \mathbb{R}^n \} \tag{5.2}$$

Although a theory for general $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be developed, we consider, in the interest of simplicity and brevity, in the present section only the case that $\mathbf{A}$ has full rank. That is, if $m \geq n$, then $\mathbf{A}$ has n linearly independent columns and if $m \geq n$, then $\mathbf{A}$ has m linearly independent rows.

Example 5.3 *The* `matlab` *command* `polyfit` *actually solves a least squares problem: given $n+1$ data points (x_i, y_i), $i = 0, \ldots, n$ and $m \leq n$, the coefficients $(a_j)_{j=0}^m$ of the polynomial $\pi(x) := \sum_{j=0}^m a_j x^j$ are found such that $\sum_{i=0}^n (\pi(x_i) - y_i)^2$ is minimized.* `matlab` *actually uses the technique based on the QR-factorization described below.* ■

5.1 Method of the normal equations

goal: derive a linear system of equations for the solution $\mathbf{x}$ of (5.2).
To that end, let $\mathbf{x} \in \mathbb{R}^n$ be the solution of (5.2) and let $\mathbf{v} \in \mathbb{R}^n$ be arbitrary but fixed. Define

$$\begin{aligned} \pi :\ & \mathbb{R} \to \mathbb{R} \\ t \mapsto\ & \|\mathbf{b} - \mathbf{A}(\mathbf{x} + t\mathbf{v})\|_2^2 = \|\mathbf{b} - \mathbf{A}\mathbf{x} - t\mathbf{A}\mathbf{v}\|_2^2 = \langle \mathbf{b} - \mathbf{A}\mathbf{x}, \mathbf{b} - \mathbf{A}\mathbf{x} \rangle_2 - 2t \langle \mathbf{b} - \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{v} \rangle_2 + t^2 \|\mathbf{A}\mathbf{v}\|_2^2 \end{aligned}$$

π is (as a function of t) a quadratic polynomials and has, by the choice of $\mathbf{x}$, a minimum at $t = 0$ $\big($choose $\mathbf{y} = \mathbf{x} + t\mathbf{v}$ in (5.2)$\big)$. Hence,

$$0 = \pi'(0) = 2\langle \mathbf{b} - \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{v} \rangle_2 = 2\mathbf{v}^\top \mathbf{A}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}).$$

Since $\mathbf{v} \in \mathbb{R}^n$ is arbitrary, we conclude that

$$0 = \mathbf{v}^\top \mathbf{A}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) \quad \forall \mathbf{v} \in \mathbb{R}^n \qquad \Rightarrow \quad \mathbf{A}^\top (\mathbf{b} - \mathbf{A}\mathbf{x}) = 0 \ \in \mathbb{R}^n.$$

Hence, $\mathbf{x}$ satisfies the *normal equations*

$$\mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b} \tag{5.3}$$

The normal equations (5.3) are a necessary condition for solutions $\mathbf{x}$ of (5.2). They are also sufficient: By tracing back the above steps, one observes that, if $\mathbf{x}$ solves (5.3) then for every fixed $\mathbf{v}$ the polynomial $t \mapsto \|\mathbf{b}-\mathbf{A}(\mathbf{x}+t\mathbf{v})\|_2^2$ has a minimum at $t=0$. Since also $\|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 = \pi(0) \leq \pi(1) = \|\mathbf{b}-\mathbf{A}(\mathbf{x}+\mathbf{v})\|_2^2$ for every $\mathbf{v}$, one concludes $\|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 \leq \|\mathbf{b}-\mathbf{A}\mathbf{y}\|_2^2 \quad \forall \mathbf{y} \in \mathbb{R}^n$. We have thus proved:

Theorem 5.4 $\mathbf{x} \in \mathbb{R}^n$ *solves (5.2), if and only if it solves (5.3).*

In many applications the square system (5.3) is solvable and thus an option to solve the least squares problem.

Theorem 5.5 *Let $m \geq n$ and let the columns of $\mathbf{A}$ be linearly independent. Then $\mathbf{A}^\top\mathbf{A}$ is an invertible matrix, and the unique solution of (5.3) is the unique solution of (5.2).*

Proof: If the columns of $\mathbf{A}$ are linearly independent, then $\mathbf{A}^\top\mathbf{A}\mathbf{y} = 0$ implies $\mathbf{y} = 0$ (Exercise!). Since $\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{n\times n}$ is a square matrix, it is invertible. Thus (5.3) is uniquely solvable. By Theorem 5.4 the problem (5.2) is uniquely solvable. □

5.2 least squares using QR-factorizations

A problem often encountered when solving the least squares problem (5.2) using the normal equations (5.3) is that the matrix $\mathbf{A}^\top\mathbf{A}$ is ill-conditioned, i.e., $\kappa(\mathbf{A}^\top\mathbf{A})$ is very large. In many applications, therefore, one solves (5.2) using the QR-factorization of $\mathbf{A}$ in spite of the increased cost.[1]

5.2.1 QR-factorization

Definition 5.6 (orthogonal matrix) *A matrix $\mathbf{Q} \in \mathbb{R}^n$ is an* orthogonal matrix*, if $\mathbf{Q}^{-1} = \mathbf{Q}^\top$.*

Example 5.7 *In $\mathbb{R}^3$, reflections at a plane, rotations, or permutations matrices:*

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & & \\ & \cos\theta & \sin\theta \\ & -\sin\theta & \cos\theta \end{pmatrix}, \quad \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$$

■

Orthgonal matrices realize transformations of $\mathbb{R}^n$ that preserve a) (euklidean) length and b) angles:

Exercise 5.8 *Let $\mathbf{Q}$ be an orthogonal matrix. Show:*

[1]In the typically setting of $m \gg n$, the cost based on QR-factorization is $2mn^2$ versus mn^2 for the method based on the normal equations.

(a) $\|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ *for all* $\mathbf{x} \in \mathbb{R}^n$.

(b) $\mathbf{x}^\top \mathbf{y} = ((\mathbf{Q})\mathbf{x})^\top (\mathbf{Q}\mathbf{y})$ *for all* $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$.

(c) $\|\mathbf{Q}\|_2 = \|\mathbf{Q}^\top\|_2 = 1$ *and conclude* $\kappa(\mathbf{Q}) = 1$ *(with respect to* $\|\cdot\|_2$*).*

(d) The columns of $\mathbf{Q}$ *have length* 1 *and are pairwise orthogonal.* ∎

We have

Theorem 5.9 *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with linearly independent columns. Then* $\mathbf{A}$ *can be written as* $\mathbf{A} = \mathbf{Q}\mathbf{R}$*, where* $\mathbf{Q} \in \mathbb{R}^{m\times m}$ *is orthogonal and* $\mathbf{R}$ *"upper triangular" matrix in the sense that* $(\mathbf{R}_{ij} = 0$ *for* $i > j$.

Proof: Applying the Gram-Schmidt orthogonalization process to the vectors $\mathbf{A}_{:,1}, \mathbf{A}_{:,2}, \ldots, \mathbf{A}_{:,n}$ yields the first n columns of $\mathbf{Q}$ as well as $\mathbf{R}$. The remaining $m-n$ columns of $\mathbf{Q}$ have to be selected such that the $\mathbf{Q}$ is orthogonal. □

Remark 5.10 *There are several algorithms to compute the QR-factorization of* $\mathbf{A}$*. Their cost is* $O(m^2 n)$*. In* `matlab`*, QR-factorization is realized with* `qr`*, in* `python` *as* `numpy.linalg.qr`*.* ∎

Remark 5.11 *If* $m \geq n$ *and if* $\mathbf{A}$ *has full rank (i.e., the columns of* $\mathbf{A}$ *are linearly independent), then the first* n *columns of* $\mathbf{Q}$ *are an orthogonal basis of the range of* $\mathbf{Q}$*, i.e., the span of the first* n *columns of* $\mathbf{Q}$ *is the span of the columns of* $\mathbf{A}$. ∎

finis 8.DS

5.2.2 Solving least squares problems with QR-factorization

Let $\mathbf{A} = \mathbf{Q}\mathbf{R}$ where $\mathbf{Q} \in \mathbb{R}^{m\times m}$ is orthogonal and $\mathbf{R} \in \mathbb{R}^{m\times n}$ is upper triangular. We assume $m \geq n$. We partition

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}^\star \\ 0 \end{pmatrix}, \qquad \mathbf{R}^* \in \mathbb{R}^{n\times n} \quad \text{upper triangular.}$$

If we assume that the columns of $\mathbf{A}$ are linearly independent, then the diagonal entries of the matrix $\mathbf{R}^*$ are non-zero, i.e., $\mathbf{R}^*$ is invertible (since the columns of $\mathbf{R}$ are linearly independent). We partition $\mathbf{Q}^\top \mathbf{b}$ as

$$\mathbf{Q}^\top \mathbf{b} = \begin{pmatrix} \mathbf{b}^* \\ \widetilde{\mathbf{b}} \end{pmatrix}, \qquad \mathbf{b}^* = (\mathbf{Q}^\top \mathbf{b})([1:n]) \in \mathbb{R}^n, \qquad \widetilde{\mathbf{b}} = (\mathbf{Q}^\top \mathbf{b})([n+1:m]) \in \mathbb{R}^{m-n},$$

We observe that for arbitrary $\mathbf{y}$ we have

$$\|\mathbf{A}\mathbf{y} - \mathbf{b}\|^2 = \|\mathbf{Q}\mathbf{R}\mathbf{y} - \mathbf{b}\|^2 = \|\mathbf{Q}(\mathbf{R}\mathbf{y} - \mathbf{Q}^\top \mathbf{b})\|^2 \overset{\text{Ex. 5.8}}{=} \|\mathbf{R}\mathbf{y} - \mathbf{Q}^\top \mathbf{b}\|^2 = \|\mathbf{R}^*\mathbf{y} - \mathbf{b}^*\|_2^2 + \|\widetilde{\mathbf{b}}\|_2^2.$$

This expression is minimized for the choice $\mathbf{y} = (\mathbf{R}^*)^{-1}\mathbf{b}^*$. We have thus arrived at the following way to compute the minimizer: