Set 3: Linear Least Squares

Kyle A. Gallivan

Department of Mathematics

Florida State University

Foundations of Computational Math 1

Fall 2012

Simple Example

Arithmetic Mean: $\bar{\eta} = \frac{1}{n}\sum_{i=1}^{n}\eta_i$

$$\bar{\eta} = \operatorname*{argmin}_{\beta\in\mathbb{R}} f(\beta) = \operatorname*{argmin}_{\beta\in\mathbb{R}} \sum_{i=1}^{n}(\eta_i - \beta)^2$$

$$f'(\beta) = -\sum_{i=1}^{n} 2(\eta_i - \beta) \;\text{ and }\; f'(\beta_{min}) = 0$$

$$0 = \sum_{i=1}^{n} 2(\eta_i - \beta_{min}) \;\rightarrow\; \sum_{i=1}^{n}\beta_{min} = \sum_{i=1}^{n}\eta_i \;\rightarrow\; \beta_{min} = \frac{\sum_{i=1}^{n}\eta_i}{n} = \bar{\eta}$$

Slightly Less Simple Example

1-dimensional Linear Regression:

- Given a set of points $(\xi_i, \eta_i),\ i = 1, \dots n$
- Determines a line $\eta = l(\xi) = \alpha\xi + \beta$ that minimizes $f(\alpha, \beta) = \Sigma_{i=1}^{n} (\eta_i - l(\xi_i))^2$
- Solution in terms of means, variance, covariance

$$\sigma_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (\xi_i - \bar{\xi})(\eta_i - \bar{\eta}) \qquad \gamma = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 \qquad l(\xi) = \gamma(\xi - \bar{\xi}) + \bar{\eta}$$

Matrix Least Squares Form

Arithmetic Mean: $$\min \left\| \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta \right\|_2$$

1-D Linear Regression: $$\min \left\| \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} - \begin{pmatrix} \xi_1 & 1 \\ \xi_2 & 1 \\ \vdots & \vdots \\ \xi_n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|_2$$

Overdetermined set of equations, $Ac = y \Longrightarrow$ Is there a solution?

Existence of a Solution

Lemma 3.1. *Suppose* $A \in \mathbb{R}^{n \times k}$, $k \leq n$ *has full column rank and let* $span[a_1, \ldots, a_k] = \mathcal{R}(A) = \mathcal{S}$ *then:*

- *If* $y \in \mathcal{S}$ *then there exists a unique* $c \in \mathbb{R}^k$ *such that* $Ac = y$.
- *If* $y \notin \mathcal{S}$ *then no* $c \in \mathbb{R}^k$ *exists such that* $Ac = y$.

Observations

- If $k = n$ then $c = A^{-1}y$
- We must redefine "solving a system" when $y \notin \mathcal{S}$.
- Essentially, we must generalize the notion of A^{-1} to define an operator $A^{\dagger} \in \mathbb{R}^{k \times n}$.
- The new definition must be consistent with the definitions used when $y \in \mathcal{S}$ and/or when $k = n$.

The Problem

Assume that $A \in \mathbb{R}^{n \times k}$, $k \leq n$, has column rank k and $b \in \mathbb{R}^n$. Find the vector $x \in \mathbb{R}^k$ that minimizes over all vectors in $\mathbb{R}^k$ the norm of the residual

$$\|b - Ax\|_2$$

That is,

$$x_{min} = \underset{x \in \mathbb{R}^k}{\operatorname{argmin}} \|b - Ax\|_2 = \underset{x \in \mathbb{R}^k}{\operatorname{argmin}} \|r\|_2$$

Transformation-based Solution

Assume $A \in \mathbb{R}^{n \times k}$ is full rank and $b \in \mathbb{R}^n$. Define the residual vector, r, for any x:

$$\min_x \|b - Ax\|_2 = \min_x \|r\|_2$$

Let $Q \in \mathbb{R}^{n \times n}$ be an orthogonal matrix and we have

$$\|r\|_2 = \|Qr\|_2$$

Change coordinate system of problem to

$$\min_x \|b - Ax\|_2 \to \min_x \|Qb - QAx\|_2$$

Same solution different defining coefficients.

Transformation-based Solution

- What coordinate system, i.e., what constraints on Qb and QA?
- k degrees of freedom $\rightarrow$ What $k \times k$ system defines x_{min}?
- How do we determine/construct Q?

Transformation-based Solution

Compute $H \in \mathbb{R}^{n \times n}$, an orthogonal matrix, and $R \in \mathbb{R}^{k \times k}$ a nonsingular upper triangular matrix such that

$$HA = \begin{pmatrix} R \\ 0 \end{pmatrix} \quad \text{which in turn defines} \quad Hb = \begin{pmatrix} c \\ d \end{pmatrix}$$

- Transformation introduces structure with 0 values.
- compresses basis into k directions with $O(k^2)$ scalars

Transformation-based Solution

$$\|r\|_2^2 = \|Hr\|_2^2$$

$$\left\| \begin{pmatrix} c \\ d \end{pmatrix} - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right\|_2^2 = \left\| \begin{pmatrix} c - Rx \\ d \end{pmatrix} \right\|_2^2 = \|c - Rx\|_2^2 + \|d\|_2^2.$$

- Minimized when $Rx = c$.
- $x_{min} = R^{-1}c$ is unique.

Elementary Orthogonal Matrices

Theorem 3.2. *Given* $x \in \mathbb{R}^n$ *with* $\|x\|_2 \neq 0$,

$$Q = I + \alpha x x^T \text{ is orthogonal } \leftrightarrow \alpha = -\frac{2}{\|x\|_2^2} \text{ or } \alpha = 0$$

Note that

$$Q = I + \alpha x x^T = I - 2uu^T$$

where $\|u\|_2 = 1$

Q is called an elementary reflector.

Householder Reflector

Theorem 3.3. *Given a vector* $v \in \mathbb{R}^n$, *and* $\gamma = \pm\|v\|_2$, *if*

$$x = v + \gamma e_1 \;\; \textit{and} \;\; \alpha = -\frac{2}{\|x\|_2^2}.$$

then

$$Hv = (I + \alpha x x^T)v = -\gamma e_1$$

The sign of γ is chosen as $sgn(e_1^T v)$ for numerical reasons.

Example

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \gamma = 2, \quad x = \begin{pmatrix} 3 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \|x\|_2^2 = 12, \quad \alpha = -\frac{1}{6}$$

$$H = \frac{1}{6} \begin{pmatrix} -3 & -3 & -3 & -3 \\ -3 & 5 & -1 & -1 \\ -3 & -1 & 5 & -1 \\ -3 & -1 & -1 & 5 \end{pmatrix}$$

$$Hv = -2e_1$$

Householder Reflector

Easy to generalize to introducing 0's in positions $i+1, \cdots, n$ leaving elements $1, \cdots, i-1$ untouched and element i an updated nonzero value:

$$H_i v = \begin{pmatrix} I_{i-1} & 0 \\ 0 & \tilde{H} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ -e_1\gamma \end{pmatrix}$$

$$\tilde{H} = I_{n-i+1} + \alpha x x^T, \quad x \in \mathbb{R}^{n-i+1}$$

$$H_i = \begin{pmatrix} I_{i-1} & 0 \\ 0 & I_{n-i+1} \end{pmatrix} + \alpha \begin{pmatrix} 0 \\ x \end{pmatrix} \begin{pmatrix} 0 & x^T \end{pmatrix}$$

Transformation-based Factorization

Assume $A \in \mathbb{R}^{n \times k}$ with $n \geq k$ and full rank. H_i are Householder reflectors. $R \in \mathbb{R}^{k \times k}$ is upper triangular.

$$H_k H_{k-1} \cdots H_2 H_1 A = B$$

$$B = \begin{pmatrix} R \\ 0 \end{pmatrix}$$

$$HA = B \rightarrow A = H^T B$$

Transformation-based Factorization

- $H = H_k H_{k-1} \cdots H_2 H_1$ is an $n \times n$ orthogonal matrix.
- Note that H and H^T do not have a simple structure like L in the LU factorization.
- H is never computed explicitly, the individual parameters for the H_i are stored.

Nonzero Pattern

$$H_1 A = A^{(1)}$$

$$H_1 \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} = \begin{pmatrix} *' & *' & *' & *' \\ 0 & *' & *' & *' \\ 0 & *' & *' & *' \\ 0 & *' & *' & *' \\ 0 & *' & *' & *' \\ 0 & *' & *' & *' \end{pmatrix}$$

Nonzero Pattern

$$
\begin{aligned}
H_2 A^{(1)} &= A^{(2)} \\
H_2 \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix} &= \begin{pmatrix} * & * & * & * \\ 0 & *' & *' & *' \\ 0 & 0 & *' & *' \\ 0 & 0 & *' & *' \\ 0 & 0 & *' & *' \\ 0 & 0 & *' & *' \end{pmatrix}
\end{aligned}
$$

Nonzero Pattern

$$
\begin{aligned}
H_3 A^{(2)} &= A^{(3)} \\
H_3 \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix} &= \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & *' & *' \\ 0 & 0 & 0 & *' \\ 0 & 0 & 0 & *' \\ 0 & 0 & 0 & *' \end{pmatrix}
\end{aligned}
$$

Nonzero Pattern

$$H_4 A^{(3)} = A^{(4)}$$

$$H_4 \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \\ 0 & 0 & 0 & * \end{pmatrix} = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & *' \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$A^{(4)}$ is in upper trapezoidal form as desired.

Geometry of Linear Least Squares

At the minimizer $x_{min} = R^{-1}c$ and $\|r_{min}\|_2^2 = \|d\|_2^2$

$$\|r\|_2^2 = \|b - Ax\|_2^2 = \|Hb - HAx\|_2^2 = \|Hr\|_2^2$$

$$\left\| \begin{pmatrix} c \\ d \end{pmatrix} - \begin{pmatrix} R \\ 0 \end{pmatrix} x \right\|_2^2 = \left\| \begin{pmatrix} c - Rx \\ d \end{pmatrix} \right\|_2^2 = \|c - Rx\|_2^2 + \|d\|_2^2.$$

$$Hr_{min} = \begin{pmatrix} c - Rx_{min} \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ d \end{pmatrix}$$

Geometry of Linear Least Squares

$$Hr_{min} = \begin{pmatrix} 0 \\ d \end{pmatrix} \to r_{min} = H^T \begin{pmatrix} 0 \\ d \end{pmatrix}$$

$$r_{min}^T A = \begin{pmatrix} 0^T & d^T \end{pmatrix} HA = \begin{pmatrix} 0^T & d^T \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix} = 0^T$$

Inner product of r_{min} with each column of A is 0.

Geometry of Linear Least Squares

Theorem 3.4. *Assume that $A \in \mathbb{R}^{n\times k}$, $k \leq n$, has column rank k and $b \in \mathbb{R}^n$. If*

$$x_{min} = \underset{x\in\mathbb{R}^k}{\operatorname{argmin}} \|b - Ax\|_2$$

minimizes the least squares cost function then

$$r_{min} = b - Ax_{min}$$

$$r_{min}^T A = 0^T$$

That is, the residual is orthogonal to $\mathcal{R}(A)$.

Arithmetic Mean

$$\min_{\beta \in \mathbb{R}} \|y - e\beta\|_2 = \min_{\beta \in \mathbb{R}} \left\| \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \beta \right\|_2 \rightarrow \beta_{min} = \bar{\eta} = \frac{\sum_{i=1}^{n} \eta_i}{n}$$

$$r_{min}^T e = (y - e\bar{\eta})^T e = \sum_{i=1}^{n} \eta_i - \bar{\eta} \sum_{i=1}^{n} 1 = \sum_{i=1}^{n} \eta_i - \bar{\eta} n = 0$$

Orthogonal Complements

Definition 3.1. If $\mathcal{S}$ is a subspace of $\mathbb{R}^n$ with dimension k then

$$\mathcal{S}^\perp = \{x \in \mathbb{R}^n \mid \forall y \in \mathcal{S} \quad < x, y >= x^T y = 0\}$$

Lemma 3.5. *If $\mathcal{S}$ is a subspace of $\mathbb{R}^n$ and $x \in \mathbb{R}^n$ then there exist two unique vectors $u \in \mathcal{S}$ and $v \in \mathcal{S}^\perp$ such that $x = u + v$. Equivalently, $\mathbb{R}^n = \mathcal{S} \oplus \mathcal{S}^\perp$.*

Orthogonal Complements

Proof. Let $x = u + v = \tilde{u} + \tilde{v}$.

$$x - x = 0 = (u - \tilde{u}) - (\tilde{v} - v) = d_u - d_v \rightarrow d_u = d_v$$

$$d_u = (u - \tilde{u}) \in \mathcal{S} \text{ and } d_v = (\tilde{v} - v) \in \mathcal{S}^\perp \rightarrow d_u^T d_v = 0$$

$$\text{Since } \|d_u - d_v\|_2^2 = \|d_u\|_2^2 + \|d_v\|_2^2 - 2d_u^T d_v$$

$$0 = \|d_u\|_2^2 + \|d_v\|_2^2 + 0 \rightarrow d_u = d_v = 0$$

$$\therefore u = \tilde{u} \text{ and } v = \tilde{v}$$

$\square$

Geometry of Linear Least Squares

Theorem 3.6. *Assume that $A \in \mathbb{R}^{n\times k}$, $k \leq n$, has column rank k and $b \in \mathbb{R}^n$ with $b = b_1 + b_2$, where $b_1 \in \mathcal{R}(A)$ and $b_2 \in \mathcal{R}^\perp(A)$. If*

$$x_{min} = \underset{x\in\mathbb{R}^k}{\operatorname{argmin}} \|b - Ax\|_2$$

minimizes the least squares cost function then

$$b_1 = Ax_{min}$$

$$r_{min} = b - Ax_{min} = b_2 \in \mathcal{R}^\perp(A)$$

The linear map $A^\dagger : \mathbb{R}^n \to \mathbb{R}^k$ that maps $b \to x_{min}$ is called the generalized (or pseudo) inverse of A. It is one-to-one and onto for $\mathcal{R}(A) \leftrightarrow \mathbb{R}^k$.

Householder Reflectors and Projections

Let $A \in \mathbb{R}^{n \times k}$ have full rank, $H \in \mathbb{R}^{n \times n}$ be an orthogonal matrix and $R \in \mathbb{R}^{k \times k}$ a nonsingular upper triangular matrix such that

$$HA = \begin{pmatrix} R \\ 0 \end{pmatrix} \text{ and } Hb = \begin{pmatrix} c \\ d \end{pmatrix}$$

$$H^T = \begin{bmatrix} Q & Q_\perp \end{bmatrix}, \quad Q \in \mathbb{R}^{n \times k} \text{ and } Q_\perp \in \mathbb{R}^{n \times n-k}$$

Householder Reflectors and Projections

We have

$$A = QR \text{ and } b = Qc + Q_\perp d = b_1 + b_2$$

Therefore, given a Householder reflector factorization of H we can compute

- an orthonormal basis for $\mathcal{R}(A)$
- the projection of b onto $\mathcal{R}(A)$,
- and the projection of b onto $\mathcal{R}^\perp(A)$

by applying the reflectors to a series of vectors.