4. Regularized least squares

- Regularized least squares
- Tikhonov regularization

Regularized least squares

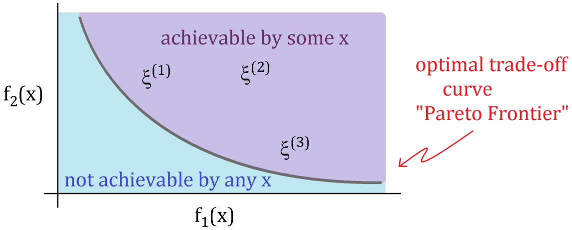
Multi-objective least-squares

Many problems need to balance competing objectives, e.g.

- make $f_1(x) = \|Ax - b\|_2^2$ small
- make $f_2(x) = \|Fx - g\|_2^2$ small

Can make $f_1(x)$ or $f_2(x)$ small, but not both.

Example: $\xi^{(i)} = \left(f_1(x^{(i)}), f_2(x^{(i)})\right)$

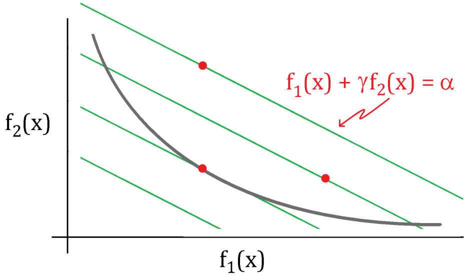


Weighted-sum objective

- Weighted-sum objective gives a Pareto optimal solution:

$$f_1(x) + \gamma f_2(x) = \|Ax - b\|^2 + \gamma \|Fx - g\|^2$$

- parameter $\gamma \geq 0$ defines relative weight between objectives
- points where $f_1(x) + \gamma f_2(x) = \alpha$ correspond to a line with slop $-\gamma$



Example: Signal denoising

- Suppose we observe noisy measurements of a signal:

$$b = \hat{x} + w \quad \text{with} \quad \hat{x} \in \mathbf{R}^n \quad \text{signal}, \quad w \in \mathbf{R}^n \quad \text{noise}$$

- Naive least squares fits noise perfectly

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \tfrac{1}{2}\|x - b\|^2$$

- Suppose we have prior information that the signal is "smooth"
- Then we might balance fit against smoothness

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \underbrace{\tfrac{1}{2}\|x - b\|^2}_{f_1(x)} \; + \; \underbrace{\tfrac{1}{2}\gamma \sum_{i=1}^{n-1} (x_i - x_{i+1})^2}_{f_2(x)}$$

where $f_2(x)$ "encourages" smoothness of the solution x

Example: Signal denoising

- Define the finite difference matrix

$$D = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \in \mathbf{R}^{n-1 \times n}$$

so that $\sum_{i=1}^{n-1} (x_i - x_{i+1})^2 = \|Dx\|^2$ $\qquad \|Ax - \hat{b}\|_2^2$

- Resulting least-squares objective:

$$\|x - b\|_2^2 + \gamma \|Dx\|^2 = \left\| \overbrace{\begin{bmatrix} I \\ \sqrt{\gamma} D \end{bmatrix}}^{A} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2$$

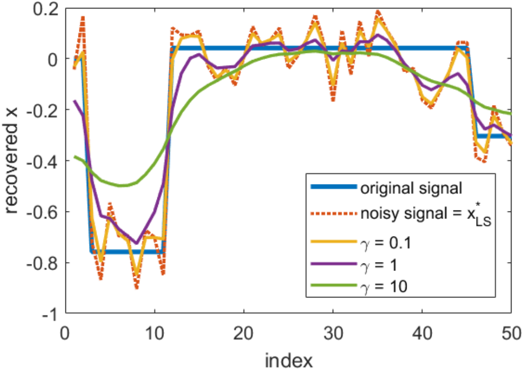
- Normal equations

$$(I + \gamma D^T D)x = b$$

$$A^T A x = A^T b$$

Example: Signal denoising

Demo



In homework, will discover a much better penalty function

Regularized least squares (aka Tikhonov)

- General form

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\|Ax - b\|^2 + \frac{\gamma}{2}\|Dx\|^2, \quad \gamma \geq 0$$

$$\|a\|_2^2 + \|b\|_2^2 = \left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|_2^2$$

  - $\|Dx\|_2^2$ is the **regularization penalty term**
  - $\gamma \geq 0$ is the **regularization parameter**

- Equivalent expression for objective

$$\frac{1}{2}\|Ax - b\|_2^2 + \frac{\gamma}{2}\|Dx\|^2 = \left\| \begin{bmatrix} A \\ \sqrt{\frac{\gamma}{2}} D \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2$$

- Normal equations

$$(A^T A + \frac{\gamma}{2} D^T D)x = A^T b$$

Singular value decomposition.

The SVD of a matrix $A \in \mathbb{R}^{m \times n}$ is

$$A = U \Sigma V^T$$

where $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{m \times n}$ is "diagonal".

find x s.t. $(A^T A x = A^T b)$ and x has least norm

If A is a rank r matrix:

1. $u_1, \dots, u_r$ is a basis for column space
2. $u_{r+1}, \dots, u_m$ is a basis for $N(A^T)$.
3. $v_1, \dots, v_r$ is a basis for row space $R(A^T)$
4. $v_{r+1}, \dots, v_n$ is a basis for $N(A)$.

Regularized least squares.

$$\min_{x \in \mathbb{R}^n} \quad \|Ax-b\|_2^2 + \gamma \|x\|_2^2 \quad \text{—— ①}$$

The normal equation: $(A^TA + \gamma I)x = A^Tb$.

How does γ affect the solution to ① ?

Let x^*_γ be the minimizer to ①. Express x^*_γ using b, U, V, Σ.

Regularized least squares using SVD.

The normal equation implies

$$(A^TAx + \gamma Ix) = A^Tb$$

$$\Leftrightarrow (V\Sigma^TU^TU\Sigma V^Tx + \gamma VV^Tx) = V\Sigma^TU^Tb$$

$$\Leftrightarrow \Sigma^T\Sigma V^Tx + \gamma V^Tx = \Sigma^TU^Tb$$

$$\Leftrightarrow \Sigma^T\Sigma z + \gamma z = \Sigma^TU^Tb$$

where $z := V^Tx \in \mathbb{R}^n$

so, $$z_i = \begin{cases} \dfrac{\sigma_i u_i^Tb}{\sigma_i^2+\gamma}, & i = 1, \dots, r \\ 0, & i = r+1, \dots, n \end{cases}$$

Regularized least squares using SVD.

Note that $z = V^T x \Rightarrow x = Vz$

So, $$x^*_\gamma = \sum_{i=1}^{r} \frac{\sigma_i u_i^T b}{\sigma_i^2 + \gamma} v_i$$

and $$\lim_{\gamma \to 0} x^*_\gamma = \lim_{\gamma \to 0} \sum_{i=1}^{r} \frac{\sigma_i u_i^T b}{\sigma_i^2 + \gamma} v_i$$

$$= \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

is the minimum norm solution to the linear least squares problem. $\min_x \|Ax - b\|_2^2$