MATH 3795
Lecture 10. Regularized Linear Least Squares.

Dmitriy Leykekhman

Fall 2008

Goals

- Understanding the regularization.

Review.

- Consider the linear least square problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2.$$

From the last lecture:

- Let $A = U\Sigma V^T$ be the Singular Value Decomposition of $A \in \mathbb{R}^{m \times n}$ with singular values

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0$$

- The minimum norm solution is

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

- If even one singular value σ_i is small, then small perturbations in b can lead to large errors in the solution.

Regularized Linear Least Squares Problems.

- If $\sigma_1/\sigma_r \gg 1$, then it might be useful to consider the **regularized linear least squares problem** (Tikhonov regularization)

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2.$$

  Here $\lambda > 0$ is the *regularization parameter*.
- The regularization parameter $\lambda > 0$ is not known *a-priori* and has to be determined based on the problem data. See later.
- Observe that

$$\min_x \frac{1}{2}\|Ax-b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 = \min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2.$$

Regularized Linear Least Squares Problems.

- Thus

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 = \min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2. \quad (1)$$

- For $\lambda > 0$ the matrix

$$\begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} \in \mathbb{R}^{(m+n)\times n}$$

  has always full rank n. Hence, for $\lambda > 0$, the regularized linear least squares problem (1) has a unique solution.

- The normal equation corresponding to (1) are given by

$$\begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix}^T \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x = (A^T A + \lambda I)x = A^T b = \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix}^T \begin{pmatrix} b \\ 0 \end{pmatrix}^T.$$

Regularized Linear Least Squares Problems.

- SVD Decomposition: $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a 'diagonal' matrix with diagonal entries

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0.$$

- Thus the normal to (1)

$$(A^T A + \lambda I)x_\lambda = A^T b,$$

can be written as

$$(V\Sigma^T \underbrace{U^T U}_{=I} \Sigma V^T + \lambda \underbrace{I}_{=VV^T})x_\lambda = V\Sigma^T U^T b.$$

Regularized Linear Least Squares Problems.

- SVD Decomposition: $A = U\Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{m\times n}$ is a 'diagonal' matrix with diagonal entries

$$\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0.$$

- Thus the normal to (1)

$$(A^T A + \lambda I)x_\lambda = A^T b,$$

can be written as

$$(V\Sigma^T \underbrace{U^T U}_{=I} \Sigma V^T + \lambda \underbrace{I}_{=VV^T})x_\lambda = V\Sigma^T U^T b.$$

- Rearranging terms we obtain

$$V(\Sigma^T\Sigma + \lambda I)V^T x_\lambda = V\Sigma^T U^T b,$$

multiplying both sides by V^T from the left and setting $z = V^T x_\lambda$ we get

$$(\Sigma^T\Sigma + \lambda I)z = \Sigma^T U^T b.$$

Regularized Linear Least Squares Problems.

- The normal equation corresponding to (1),

$$(A^T A + \lambda I)x_\lambda = A^T b,$$

is equivalent to

$$\underbrace{(\Sigma^T \Sigma + \lambda I)}_{\text{diagonal}} z = \Sigma^T U^T b.$$

where $z = V^T x_\lambda$.

- Solution

$$z_i = \begin{cases} \frac{\sigma_i (u_i^T b)}{\sigma_i^2 + \lambda}, & i = 1, \ldots, r, \\ 0, & i = r+1, \ldots, n. \end{cases}$$

- Since $x_\lambda = Vz = \sum_{i=1}^n z_i v_i$, the solution of the regularized linear least squares problem (1) is given by

$$x_\lambda = \sum_{i=1}^r \frac{\sigma_i (u_i^T b)}{\sigma_i^2 + \lambda} v_i.$$

Regularized Linear Least Squares Problems.

- Note that

$$\lim_{\lambda \to 0} x_\lambda = \lim_{\lambda \to 0} \sum_{i=1}^{r} \frac{\sigma_i (u_i^T b)}{\sigma_i^2 + \lambda} v_i = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i = x_\dagger$$

i.e., the solution of the regularized LLS problem (1) converges to the minimum norm solution of the LLS problem as λ goes to zero.

Regularized Linear Least Squares Problems.

- Note that

$$\lim_{\lambda\to 0} x_\lambda = \lim_{\lambda\to 0} \sum_{i=1}^{r} \frac{\sigma_i(u_i^T b)}{\sigma_i^2+\lambda} v_i = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i = x_\dagger$$

  i.e., the solution of the regularized LLS problem (1) converges to the minimum norm solution of the LLS problem as λ goes to zero.

- The representation

$$x_\lambda = \sum_{i=1}^{r} \frac{\sigma_i(u_i^T b)}{\sigma_i^2+\lambda} v_i$$

  of the solution of the regularized LLS also reveals the regularizing property of adding the term $\frac{\lambda}{2}\|x\|_2^2$ to the (ordinary) least squares functional. We have that

$$\frac{\sigma_i(u_i^T b)}{\sigma_i^2+\lambda} \approx \begin{cases} 0, & \text{if } 0 \approx \sigma_i \ll \lambda \\ \frac{u_i^T b}{\sigma_i}, & \text{if } \sigma_i \gg \lambda. \end{cases}$$

Regularized Linear Least Squares Problems.

- Hence, adding $\frac{\lambda}{2}\|x\|_2^2$ to the (ordinary) least squares functional acts as a filter. Contributions from singular values which are large relative to the regularization parameter λ are left (almost) unchanged whereas contributions from small singular values are (almost) eliminated.
- It raises an important question:

  How to choose λ?

Regularized Linear Least Squares Problems.

- Suppose that the data are $b = b_{ex} + \delta b$. We want to compute the minimum norm solution of the (ordinary) LLS with unperturbed data b_{ex}

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

but we can only compute with $b = b_{ex} + \delta b$, we don't know b_{ex}.

- The solution of the regularized least squares problem is

$$x_\lambda = \sum_{i=1}^{r} \left(\frac{\sigma_i (u_i^T b_{ex})}{\sigma_i^2 + \lambda} + \frac{\sigma_i (u_i^T \delta b)}{\sigma_i^2 + \lambda} \right) v_i.$$

Regularized Linear Least Squares Problems.

- We observed that

$$\sum_{i=1}^{r} \frac{\sigma_i (u_i^T b_{ex})}{\sigma_i^2 + \lambda} \to x_{ex} \quad \text{as} \quad \lambda \to 0.$$

- On the other hand

$$\frac{\sigma_i (u_i^T \delta b)}{\sigma_i^2 + \lambda} \approx \begin{cases} 0, & \text{if } 0 \approx \sigma_i \ll \lambda \\ \frac{u_i^T \delta b}{\sigma_i}, & \text{if } \sigma_i \gg \lambda, \end{cases}$$

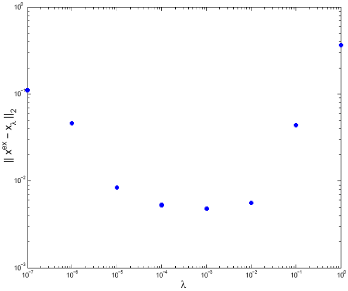
  which suggests to choose λ sufficiently large to ensure that errors δb in the data are not magnified by small singular values.

Regularized Linear Least Squares Problems.

```
% Compute A
t = 10.^(0:-1:-10)';
A = [ ones(size(t)) t t.^2 t.^3 t.^4 t.^5];

% compute exact data
xex = ones(6,1); bex = A*xex;

% data perturbation of 0.1%
deltab = 0.001*(0.5-rand(size(bex))).*bex;
b      = bex+deltab;

% compute SVD of A
[U,S,V] = svd(A); sigma = diag(S);

for i = 0:7 % solve regularized LLS for different lambda
    lambda(i+1) = 10^(-i)
    xlambda = V * (sigma.*(U'*b) ./ (sigma.^2 + lambda(i+1)))
    err(i+1) = norm(xlambda - xex);
end
```

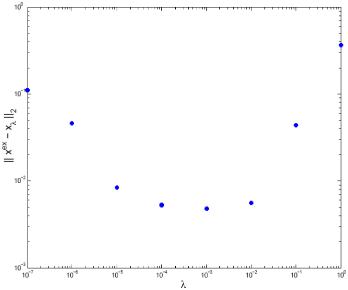
Regularized Linear Least Squares Problem.

- The error $\|x_{ex} - x_\lambda\|_2$ for different values of λ (loglog-scale):



Regularized Linear Least Squares Problem.

- The error $\|x_{ex} - x_\lambda\|_2$ for different values of λ (loglog-scale):



- For this example $\lambda \approx 10^{-3}$ is a good choice for the regularization parameter λ. However, we could only create this figure with the knowledge of the desired solution x_{ex}.
- How can we determine a $\lambda \geq 0$ so that $\|x_{ex} - x_\lambda\|_2$ is small without knowledge of x_{ex}.
- One approach is the **Morozov discrepancy principle**.

Regularized Linear Least Squares Problem.

- Suppose $b = b_{ex} + \delta b$. We do not know the perturbation δb, but we assume that we know its size $\|\delta b\|$.
- Suppose the unknown desired solution x_{ex} satisfies $Ax_{ex} = b_{ex}$.
- Hence,

$$\|Ax_{ex} - b\| = \|Ax_{ex} - b_{ex} - \delta b\| = \|\delta b\|.$$

- Since the exact solution satisfies $\|Ax_{ex} - b\| = \|\delta b\|$ we want to find a regularization parameter $\lambda \geq 0$ such that the solution x_λ of the regularized least squares problem satisfies

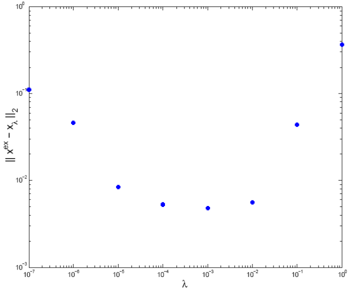
$$\|Ax_\lambda - b\| = \|\delta b\|$$

This is **Morozov's discrepancy principle.**

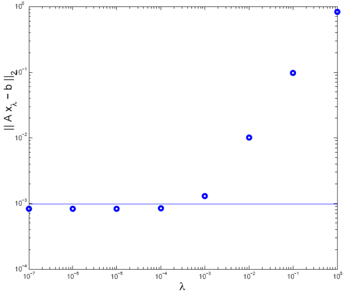
Regularized Linear Least Squares Problem.

Let's see now this works for the previous Matlab example.

The error $\|x_{ex} - x_\lambda\|_2$ for different values of λ (log-log-scale)



The residual $\|Ax_\lambda - b\|_2$ and $\|\delta b\|_2$ for different values of λ (log-log- scale)



Regularized Linear Least Squares Problem.

- Morozov's discrepancy principle: Find $\lambda \geq 0$ such that

$$\|Ax_\lambda - b\| = \|\delta b\|$$

Regularized Linear Least Squares Problem.

- Morozov's discrepancy principle: Find $\lambda \geq 0$ such that

$$\|Ax_\lambda - b\| = \|\delta b\|$$

- To compute $\|Ax_\lambda - b\|$ for given $\lambda \geq 0$ we need to solve a regularized linear least squares problem

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 = \min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$$

to get x_λ and then we have to compute $\|Ax_\lambda - b\|$.

Regularized Linear Least Squares Problem.

- Morozov's discrepancy principle: Find $\lambda \geq 0$ such that

$$\|Ax_\lambda - b\| = \|\delta b\|$$

- To compute $\|Ax_\lambda - b\|$ for given $\lambda \geq 0$ we need to solve a regularized linear least squares problem

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 + \frac{\lambda}{2}\|x\|_2^2 = \min_x \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2$$

  to get x_λ and then we have to compute $\|Ax_\lambda - b\|$.
- Let $f(\lambda) = \|Ax_\lambda - b\| - \|\delta b\|$. Finding $\lambda \geq 0$ such that

$$f(\lambda) = 0$$

  is a **root finding problem**. We will discuss in the future how to solve such problems. In this case f maps a scalar λ into a scalar $f(\lambda) = \|Ax_\lambda - b\| - \|\delta b\|$, but the evaluation of f requires the solution of a regularized LLS problems and can be rather expensive.