EE263 Review session 6

- Least norm solutions
- Least norm vs. least squares
- General norm minimization
- Multiobjective least squares
- Regularized least squares

Least-norm solutions of undetermined equations

- consider

$$y = Ax$$

with a fat and full rank matrix A.

- set of all solutions has form

$$x = x_p + Qz$$

where the columns of Q form the basis for null(A).

- one choice of x_p is

$$x_{\mathsf{ln}} = A^T(AA^T)^{-1}y$$

- in fact, x_{ln} minimizes $\|x\|$ among all solutions of $y = Ax$. i.e., it is the optimal solution of the following.

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = y \end{array}$$

- for any x satisfying $Ax = y$,

$$\begin{aligned}(x - x_{\mathsf{ln}})^T x_{\mathsf{ln}} &= (x - x_{\mathsf{ln}})^T A^T (AA^T)^{-1} y \\ &= (Ax - Ax_{\mathsf{ln}})^T (AA^T)^{-1} y \\ &= 0\end{aligned}$$

therefore

$$\|x\|^2 = \|x - x_{\mathsf{ln}} + x_{\mathsf{ln}}\|^2 = \|x - x_{\mathsf{ln}}\|^2 + \|x_{\mathsf{ln}}\|^2 \geq \|x_{\mathsf{ln}}\|^2$$

- the least norm problem is equivalent to

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = y \end{array}$$

- introducing the Lagrangian

$$L(x, \lambda) = x^T x + \lambda^T (Ax - y)$$

- optimality conditions are,

$$\nabla_x L = 2x + A^T \lambda = 0$$
$$\nabla_\lambda L = Ax - y = 0$$

from the first condition $x_{\mathrm{ln}} = -\frac{1}{2} A^T \lambda$, and

$$AA^T \lambda = -2y \qquad \Longrightarrow \qquad \lambda = -2(AA^T)^{-1} y$$

so as before

$$x_{\mathrm{ln}} = A^T (AA^T)^{-1} y$$

Least-norm vs. least-squares

least norm solution

- for a *fat full rank* matrix A,
- $A^\dagger = A^T(AA^T)^{-1}$ is called the pseudo inverse of A
- $A^\dagger$ is a right inverse of A
- $A^\dagger y$ minimizes $\|x\|$ subject to $Ax = y$
- $I - A^T(AA^T)^{-1}A = I - A^\dagger A$ gives projection onto null(A)

least squares solution

- for a *skinny full rank* matrix A,
- $A^\dagger = (A^TA)^{-1}A^T$ is called the pseudo inverse of A
- $A^\dagger$ is a left inverse of A
- $A^\dagger y$ minimizes $\|Ax - y\|$
- $A(A^TA)^{-1}A^T = AA^\dagger$ gives projection onto range(A)

General norm minimization with equality constraints

- consider problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{subject to} & Cx = d \end{array}$$

which is equivalent to,

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax - b\|^2 \\ \text{subject to} & Cx = d \end{array}$$

- Lagrangian is

$$L(x, \lambda) = \frac{1}{2}\|Ax - b\|^2 + \lambda^T(Cx - d)$$

- optimality conditions are,

$$\nabla_x L = A^T Ax - A^T b + C^T \lambda = 0 \qquad \nabla_\lambda L = Cx - d = 0$$

in matrix form,

$$\begin{bmatrix} A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^T b \\ d \end{bmatrix}$$

Multiobjective least squares

- we want $J_1 = \|Ax - y\|^2$ small
  and also $J_2 = \|Fx - g\|^2$ small
- usually the objectives are *competing*
- sometimes referred to as *waterbed effect*
- so instead design via the following single objective,

$$\text{minimize} \qquad \|Ax - y\|^2 + \mu\|Fx - g\|^2$$

  with $\mu \geq 0$.

  in fact, this reduces to a usual least squares problem.

- how?

$$\|Ax - y\|^2 + \mu\|Fx - g\|^2 = \left\| \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix} x - \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix} \right\|^2$$
$$= \|\tilde{A}x - \tilde{y}\|^2$$

where

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix} \qquad \tilde{y} = \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix}$$

hence solution is (assuming $\tilde{A}$ full rank)

$$x = \tilde{A}^\dagger \tilde{y}$$
$$= (A^T A + \mu F^T F)^{-1}(A^T y + \mu F^T g)$$

- common example: *regularized least squares* (Tychonov regularization)

Regularized least squares

- with $F = I$, $g = 0$, the objective function is,

$$\text{minimize} \qquad \|Ax - y\|^2 + \mu\|x\|^2$$

  and the minimizer is

$$x = (A^T A + \mu I)^{-1} A^T y$$

- for $\mu > 0$, works for any A (no restrictions on shape, rank...)

regularized LS $\rightarrow$ least squares

- regularized LS

$$\text{minimize} \qquad \|Ax - y\|^2 + \mu\|x\|^2$$

the minimizer is

$$x = (A^T A + \mu I)^{-1} A^T y$$

- suppose A is *skinny and full rank*.
as $\mu \longrightarrow 0$, the minimizer converges to

$$x = (A^T A)^{-1} A^T y$$

which is the *least squares* solution.

regularized LS $\rightarrow$ least norm

- regularized LS

$$\text{minimize} \qquad \|Ax - y\|^2 + \mu\|x\|^2$$

the minimizer is

$$x = (A^T A + \mu I)^{-1} A^T y$$

- suppose A is *fat and full rank*.
as $\mu \rightarrow 0$, does the minimizer converge to the following?

$$x = (A^T A)^{-1} A^T y$$

NO WAY!
then what does x converge to as $\mu \rightarrow 0$?

ans: $x = A^T (AA^T)^{-1} y$

push-through identity

- suppose AB is square and $I + AB$ is invertible, then

$$(I + AB)^{-1}A = A(I + BA)^{-1}$$

limit behavior as $\mu \to 0$

- if A is skinny and full rank, as $\mu \to 0$,

$$(A^TA + \mu I)^{-1}A^T \to (A^TA)^{-1}A^T$$

- if A is fat and full rank, $\mu \to 0$,

$$(A^TA + \mu I)^{-1}A^T \to A^T(AA^T)^{-1}$$