Designing and Analyzing Linear Systems

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Justin Solomon

Where Are We?

- Reliable methods for factoring/solving linear systems
- Strategies for understanding when our methods will succeed

Obvious Question

So what?

Parametric Regression

$$f(\vec{x}) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

Find $\{a_1, \ldots, a_n\}$.

n Experiments

$$\vec{x}^{(k)} \mapsto y^{(k)} \equiv f(\vec{x}^{(k)})$$

$$\begin{aligned}
y^{(1)} &= f(\vec{x}^{(1)}) = a_1 x_1^{(1)} + a_2 x_2^{(1)} + \cdots + a_n x_n^{(1)} \\
y^{(2)} &= f(\vec{x}^{(2)}) = a_1 x_1^{(2)} + a_2 x_2^{(2)} + \cdots + a_n x_n^{(2)} \\
&\vdots
\end{aligned}$$

Note: The x's may have different units

Linear System for $\vec{a}$

$$\begin{pmatrix} - & \vec{x}^{(1)\top} & - \\ - & \vec{x}^{(2)\top} & - \\ & \vdots & \\ - & \vec{x}^{(n)\top} & - \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{pmatrix}$$

General Case

$$f(\vec{x}) = a_1 f_1(\vec{x}) + a_2 f_2(\vec{x}) + \cdots + a_m f_m(\vec{x})$$

$$\begin{pmatrix} f_1(\vec{x}^{(1)}) & f_2(\vec{x}^{(1)}) & \cdots & f_m(\vec{x}^{(1)}) \\ f_1(\vec{x}^{(2)}) & f_2(\vec{x}^{(2)}) & \cdots & f_m(\vec{x}^{(2)}) \\ \vdots & \vdots & \cdots & \vdots \\ f_1(\vec{x}^{(m)}) & f_2(\vec{x}^{(m)}) & \cdots & f_m(\vec{x}^{(m)}) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

f can be *nonlinear*!

Two Important Cases

$$f(x) \equiv a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

"Vandermonde system"

$$f(x) = a \cos(x + \phi)$$

Mini-Fourier

Something Fishy

Why should you have to do exactly n experiments?

What if $y^{(k)}$ is measured with error?

Overfitting

Overfitting

Representing noise rather than the underlying relationships in a statistical dataset

Interpretation of Linear Systems

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} - & \vec{r}_1^\top & - \\ - & \vec{r}_2^\top & - \\ \vdots & \cdots & \vdots \\ - & \vec{r}_n^\top & - \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \vec{r}_1 \cdot \vec{x} \\ \vec{r}_2 \cdot \vec{x} \\ \vdots \\ \vec{r}_n \cdot \vec{x} \end{pmatrix}$$

"Guess $\vec{x}$ by observing its dot products with $\vec{r}_i$'s."

What happens when $m > n$?

The observations are likely to be incompatible.

$$A\vec{x} \approx \vec{b}$$

Least Squares

$$A\vec{x} \approx \vec{b} \iff \min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2$$

Least Squares

$$A\vec{x} \approx \vec{b} \iff \min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2$$
$$\iff A^\top A\vec{x} = A^\top \vec{b}$$

Normal Equations

$$A^\top A\vec{x} = A^\top \vec{b} \overset{?}{\Longleftrightarrow} \vec{x} = (A^\top A)^{-1} A^\top \vec{b}$$

Regularization

Tikhonov regularization
("Ridge Regression;" Gaussian prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2$$

Lasso (Laplace prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_1$$

Elastic Net:

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2 + \beta\|\vec{x}\|_1$$

Regularization

Tikhonov regularization
("Ridge Regression;" Gaussian prior):

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|_2^2 + \alpha\|\vec{x}\|_2^2$$

Least Squares "In the Wild"

Example: Image alignment

$$\vec{y}_k \approx A\vec{x}_k + \vec{b}$$

$$A \in \mathbb{R}^{2\times 2}$$

$$\vec{b} \in \mathbb{R}^2$$

A Ridiculously Important Matrix

$$A^\top A$$

Properties of $A^\top A$

Symmetric

B is *symmetric* if $B^\top = B$.

Properties of $A^\top A$

Symmetric

B is *symmetric* if $B^\top = B$.

Positive (Semi-)Definite

B is *positive semidefinite* if for all $\vec{x} \in \mathbb{R}^n$, $\vec{x}^\top B\vec{x} \geq 0$. B is *positive definite* if $\vec{x}^\top B\vec{x} > 0$ whenever $\vec{x} \neq \vec{0}$.

Pivoting for SPD C

Goal:

Solve $C\vec{x} = \vec{d}$ for symmetric positive definite C.

$$C = \begin{pmatrix} c_{11} & \vec{v}^\top \\ \vec{v} & \tilde{C} \end{pmatrix}$$

Forward Substitution

$$E = \begin{pmatrix} {}^{1}\!/\!\sqrt{c_{11}} & \vec{0}^\top \\ \vec{r} & I_{(n-1)\times(n-1)} \end{pmatrix}$$

Symmetry Experiment

Try post-multiplication:

$$ECE^\top$$

Review: What's New?

- Positive definite $\implies$ existence of $\sqrt{c_{11}}$
- Symmetry $\implies$ apply E to both sides

Cholesky Factorization

$$C = LL^\top$$

Observation about Cholesky

$$L = \begin{pmatrix} L_{11} & 0 & 0 \\ \vec{\ell}_k^\top & \ell_{kk} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix}$$

$$\Downarrow$$

$$LL^\top = \begin{pmatrix} \times & \times & \times \\ \vec{\ell}_k^\top L_{11}^\top & \vec{\ell}_k^\top \vec{\ell}_k + \ell_{kk}^2 & \times \\ \times & \times & \times \end{pmatrix}$$

Observation about Cholesky

$$\ell_{kk} = \sqrt{c_{kk} - \|\ell_k\|_2^2}$$

$$L_{11}\vec{\ell}_k = \vec{c}_k$$

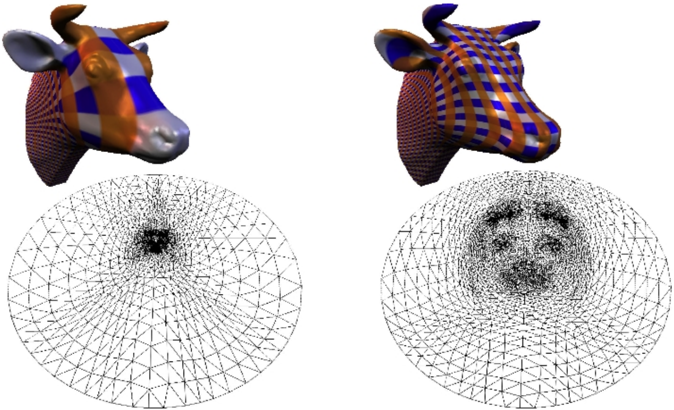
Interpretation of Cholesky

<philosophy>

What is $\vec{x}^\top C\vec{x}$?

</philosophy>

Harmonic Parameterization



http://www.mpi-inf.mpg.de/departments/d4/areas/meshproc/rz_Kowz.jpg

Storing Sparse Matrices

Want $O(n)$ storage if we have $O(n)$ nonzeros!

Examples:

- List of triplets `(r,c,val)`
- For each row `r`, `matrix[r]` holds a dictionary `c`$\mapsto$`A[r][c]`

Fill

$$\begin{pmatrix} \times & \times & \times & \times & \times \\ 0 & \times & 0 & 0 & 0 \\ 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & \times & 0 \\ 0 & 0 & 0 & 0 & \times \end{pmatrix} \Rightarrow \begin{pmatrix} \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times \end{pmatrix}$$

Avoiding Fill

- Common strategy: Permute rows/columns

- Mostly heuristic constructions

  Minimizing fill in Cholesky is NP-complete!

- Alternative strategy:
  Avoid Gaussian elimination altogether

  Lots more in a few weeks!

Banded Matrices

$$\begin{pmatrix} \times & \times & & & \\ \times & \times & \times & & \\ & \times & \times & \times & \\ & & \times & \times & \times \\ & & & \times & \times \end{pmatrix}$$

- Storage?
- Solving?

Cyclic Matrices

$$\begin{pmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{pmatrix}$$

- Storage?
- Solving?

Preview

$$
\begin{aligned}
\text{cond } A^\top A &= \|A^\top A\| \|(A^\top A)^{-1}\| \\
&\approx \|A^\top\| \|A\| \|A^{-1}\| \|(A^\top)^{-1}\| \\
&= \|A\|^2 \|A^{-1}\|^2 \\
&= (\text{cond } A)^2 \\
&☹
\end{aligned}
$$

▸ Next