

Linear Inverse Problems

Ariel Cintrón-Arias

Tuesday, May 22, 2007

Inner product

$$(1, 1, 5, 7) \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = 1\beta_1 + 1\beta_2 + 5\beta_3 + 7\beta_4$$

In general,

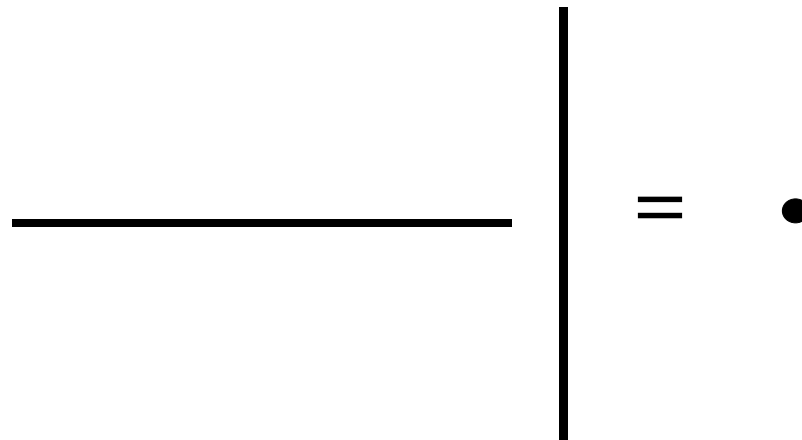
$$(u_1, \dots, u_n) \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \sum_{i=1}^n u_i \beta_i$$

row \times *column* = *scalar*

Inner product

$$(u_1, \dots, u_n) \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \sum_{i=1}^n u_i \beta_i$$

it looks like this:



Matrix vector representation of linear systems

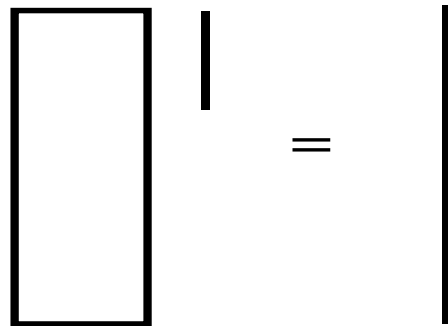
$$\begin{array}{rcl} x_{11}\beta_1 + x_{12}\beta_2 + \dots + x_{1n}\beta_n & = & v_1 \\ x_{21}\beta_1 + x_{22}\beta_2 + \dots + x_{2n}\beta_n & = & v_2 \\ \vdots & & \vdots \\ x_{m1}\beta_1 + x_{m2}\beta_2 + \dots + x_{mn}\beta_n & = & v_m \end{array}$$

Every v_i is an inner product:

$$v_i = (x_{i1}, \dots, x_{in}) \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}$$

Matrix vector representation of linear systems

$$\begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$$



Example

$$3\beta_1 + 2\beta_2 = -1$$

$$\beta_1 + \beta_2 = 2$$

$$\begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Identity matrix

In general, an n -by- n identity matrix is denoted by I :

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 1 \end{bmatrix}$$

The identity has 1 along the diagonal and 0's elsewhere. Clearly, for any matrix X , it is true that $XI = X$.

Inverse of a matrix

An n -by- n matrix X is invertible (nonsingular) if there exists X^{-1} such that $XX^{-1} = I$

$$\text{if } X = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ then } X^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Exercise #1: If $X = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$ compute X^{-1} by hand.

Exercise #2: If $X = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$ compute X^{-1} by using the MATLAB command `inv()`

Transpose of a matrix

Given a matrix X create a new matrix whose rows are the columns of X .

For instance, if $X = \begin{bmatrix} 5 & 0 & 2 \\ 5 & 0 & 2 \\ 5 & 0 & 2 \\ 5 & 0 & 2 \\ 5 & 0 & 2 \\ 5 & 0 & 2 \end{bmatrix}$,

then the transpose $X^T = \begin{bmatrix} 5 & 5 & 5 & 5 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$

Exercise #3: In MATLAB type in

$$X = \begin{bmatrix} 5 & 1 & 3 \\ 3 & 1 & 25 \\ -1 & 5 & 2 \\ 0 & 0 & 2 \\ -1 & -1 & -2 \end{bmatrix}$$

- Compute the transpose by typing X' .
- What are the dimensions of X ? Use the command `size(X)`
- Create another rectangular matrix X . For instance, use `X=rand(300,5)`. Is this matrix X nonsingular? Try `inv(X)???`
- Is the matrix $X^T X$ rectangular?? Type in `X'*X`. Type in `size(X'*X)`. Does $(X^T X)^{-1}$ exist? Try `inv(X'*X)??`

While a rectangular matrix X does not have an inverse, the matrix $X^T X$ is square and sometimes $(X^T X)^{-1}$ exists.

Suppose we are interested in solving $X^T X \beta = v$ for β . Whenever $(X^T X)^{-1}$ exists, the solution is

$$\hat{\beta} = (X^T X)^{-1} v$$

Exercise #4:

In MATLAB type in `xx =` $\begin{bmatrix} 0.8350 & 0.6670 \\ 0.3330 & 0.2660 \end{bmatrix}$

Using `v = [0.168; 0.067]`, find `betahat = inv(xx)*v`.

Using `vtilde = [0.168; 0.066]`, find `betatilde = inv(xx)*vtilde`.

Compare `betahat` and `betatilde`.

Just as it is illustrated in the previous exercise, even though sometimes $(X^T X)^{-1}$ exists, the solution $\hat{\beta} = (X^T X)^{-1} X^T v$ may be quite sensitive to **variations** in v .

This feature is related to algebraic properties of the matrix $X^T X$ and can be identified by a quantity called the **condition number**. Before we define it we need to recall the definition of a matrix norm.

The 1-norm of a matrix $M \in \mathbb{R}^{m \times n}$ is defined by

$$\|M\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |M_{ij}|$$

The **condition number** of a nonsingular matrix M is defined by

$$\kappa(M) = \|M\|_1 \|M^{-1}\|_1$$

Ill-conditioning

- Ill-conditioning: An invertible matrix M is called ill-conditioned if its condition number $\kappa(M)$ is large.
- Well-conditioning: Matrices with small condition numbers are said to be well-conditioned.

Exercise #5 In MATLAB the condition number is computed by using the command `cond`.

Using the matrix `xx` from exercise # 4, compute `cond(xx, 1)`

Is the matrix `xx` well-conditioned? or is it ill-conditioned?

Remark: Solving $X^T X \beta = v$ for β , requires knowledge about `cond($X^T X$, 1)`.

Linear least-squares

Let's talk about the scalar case first. Suppose we observe y and have reason to think it is the output of a linear model $f(\beta)$.

In order to account for discrepancies between reality and the model we include an error ε and write

$$y = f(\beta) + \varepsilon$$

Given the observation y we want to find the $\hat{\beta}$ that gives the smallest distance between the model and the observation. In fact, to ensure the minimization of only positive discrepancies we solve the following problem:

$$\min_{\beta} (y - f(\beta))^2$$

Example

Suppose $f(\beta_1, \beta_2) = 3\beta_1 + \beta_2$.

Let's solve

$$\min_{(\beta_1, \beta_2)} (y - 3\beta_1 - \beta_2)^2$$

The minimizer $(\hat{\beta}_1, \hat{\beta}_2)$ can be found from solving the optimality conditions:

$$\frac{\partial}{\partial \beta_1} (y - 3\beta_1 - \beta_2)^2 = 0$$

$$\frac{\partial}{\partial \beta_2} (y - 3\beta_1 - \beta_2)^2 = 0$$

These equations reduce to

$$18\beta_1 + 6\beta_2 = 6y$$

$$6\beta_1 + 2\beta_2 = 2y$$

Example

In matrix-vector form we write

$$\begin{bmatrix} 18 & 6 \\ 6 & 2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 6y \\ 2y \end{bmatrix}$$

The solution $(\hat{\beta}_1, \hat{\beta}_2)$ is then found by

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 18 & 6 \\ 6 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 6y \\ 2y \end{bmatrix}$$

Exercise # 6:

Using MATLAB find `cond([18 6; 6 2],1)` and `inv([18 6; 6 2])`.

- What seems to be the problem??
- We had one observation y and intended to estimate two parameters $(\hat{\beta}_1, \hat{\beta}_2)$.
- Perhaps more observations are welcome and needed!!!.
- What if we had two observations (y_1, y_2) and try to estimate two parameters $(\hat{\beta}_1, \hat{\beta}_2)$?
- What if we had 20 observations (y_1, \dots, y_{20}) and try to estimate two parameters $(\hat{\beta}_1, \hat{\beta}_2)$?

We now consider a more general formulation with more observations y than parameters β :

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

In matrix-vector form:

$$y = X\beta + \varepsilon$$

where $m > n$ (more observations than parameters) and
 $y \in \mathbb{R}^m$ (column vector), $\varepsilon \in \mathbb{R}^m$ (column vector)
 $\beta \in \mathbb{R}^n$ (column vector), $X \in \mathbb{R}^{m \times n}$ (rectangular matrix)

How can we formulate the minimization problem in higher dimensions?

Goal: Find $\hat{\beta}$ that gives the “smallest distance” between the model and the observations.

We need to measure the distance of a vector. Let's use the 2-norm of a vector.

The 2-norm of a vector v is the inner product of itself:

$$v^T v = (v_1, \dots, v_m) \begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ v_m \end{bmatrix} = \sum_{i=1}^m v_i^2 := \|v\|_2^2$$

Using the 2-norm we can write the distance between the model and the observations:

$$\|y - X\beta\|_2^2 = (y - X\beta)^T (y - X\beta)$$

The smallest distance is given by the solution to:

$$\min_{\beta} \|y - X\beta\|_2^2 = \min_{\beta} (y - X\beta)^T (y - X\beta)$$

The minimizer $\hat{\beta}$ is found from the optimality conditions:

$$\frac{\partial}{\partial \beta} (y - X\beta)^T (y - X\beta) = 0$$

Computing the partial derivatives $\frac{\partial}{\partial \beta} (y - X\beta)^T (y - X\beta)$ is analogous to calculating the derivative of a square function, i.e. $\frac{d}{dz} (z^2) = 2z$.

Doing a tiny bit of matrix algebra

$$\begin{aligned}
 & \frac{\partial}{\partial \beta} (y - X\beta)^T (y - X\beta) \\
 &= -X^T (y - X\beta) + [(y - X\beta)^T (-X)]^T \\
 &= -X^T (y - X\beta) - X^T (y - X\beta) \\
 &= -2X^T (y - X\beta) \\
 &= -2(X^T y - X^T X\beta)
 \end{aligned}$$

Therefore, solving the optimality conditions $\frac{\partial}{\partial \beta} (y - X\beta)^T (y - X\beta) = 0$ is equivalent to solving $X^T X\beta = X^T y$ for β .

If $(X^T X)^{-1}$ exists (and preferably with small $\text{cond}(X^T X, 1)$) then $\hat{\beta} = (X^T X)^{-1} X^T y$

In summary:

Given observations $(y_1, \dots, y_m)^T$ and assuming

$$y = X\beta + \varepsilon$$

The parameter that gives the smallest distance between the model and the observations is

$$\hat{\beta} := \operatorname{Argmin}_{\beta} \|y - X\beta\|_2^2 = (X^T X)^{-1} X^T y$$

Some remarks

- The parameters estimates $\hat{\beta}$ depend on whether or not $\text{cond}(X^T X, 1)$ is large.
- How do we know if the parameters estimates are good, fine or bad estimates? Uncertainty in the estimation is related to the measurement error...
(to be addressed in an upcoming tutorial titled “Statistical View of Linear Least Squares” by Dr. J. Crooks and S. Luna-Gomez)
- An alternative way of computing $\hat{\beta}$ is to define a cost function, say $L(\beta) = (y - X\beta)^T (y - X\beta)$ and apply numerical optimization algorithms such as `fminsearch`...
(you will learn how to use this function during the lectures “Solving the Vibrating Beam: Inverse Problem” by Dr. W. LeFew as well as in “Solving the Vibrating Beam: Optimization” by K. Sutton).