CSC320 week 2 tutorial note

Zhaowei Liu

January 17, 2017

System of linear equations is common in this course (and outside the course). An example is the matting equations:

$$\begin{bmatrix} 1 & 0 & 0 & -R_k \\ 0 & 1 & 0 & -G_k \\ 0 & 0 & 1 & -B_k \\ 1 & 0 & 0 & -R'_k \\ 0 & 1 & 0 & -G'_k \\ 0 & 0 & 1 & -B'_k \end{bmatrix} \begin{bmatrix} R_0 \\ G_0 \\ B_0 \\ \alpha_0 \end{bmatrix} = \begin{bmatrix} R_\Delta \\ G_\Delta \\ B_\Delta \\ R'_\Delta \\ G'_\Delta \\ B'_\Delta \end{bmatrix}$$

All such linear systems can be written in the form

$$A\boldsymbol{x} = \boldsymbol{b}$$

where A is a $m \times n$ matrix, $\boldsymbol{x}$ is a (column) vector with n entries, and $\boldsymbol{b}$ is a (column) vector with m entries. Generally A and $\boldsymbol{b}$ are known (e.g. in the matting problem), and we need to solve for $\boldsymbol{x}$. If $m = n$ and A is invertible, we have $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. But this is not the case for many problems, like the matting problem, where $m > n$. Such system of equations is called overdetermined system. In general, there is no solution which satisfies all constraints, so we need to define the solution in some other way.

Before talking about how to solve overdetermined systems, let's take a different view of the equations in the system, using the above matting equations as examples. The matting equations in matrix form can be rewritten into 6 equations:

$$\begin{cases} R_0 - \alpha R_k = R_\Delta \\ G_0 - \alpha G_k = G_\Delta \\ B_0 - \alpha B_k = B_\Delta \\ R_0 - \alpha R'_k = R'_\Delta \\ G_0 - \alpha G'_k = G'_\Delta \\ B_0 - \alpha B'_k = B'_\Delta \end{cases}$$

and every equation in the linear system corresponds to an observation. For example, in the first equation $R_0 - \alpha R_k = R_\Delta$, R_Δ is what we observe and its value is determined by R_0, α and R_k (although the values of R_0 and α are unknown). However, the observed values may not be the true values. Actually it's very unlikely that the observed values are the true values in most problems. So we generally model the true values to be the observed values plus some error terms (or noise), called $\boldsymbol{e}$. Then the linear system becomes

$$A\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{e}$$

where $\boldsymbol{b}$ is the observed values and $\boldsymbol{e}$ is a vector of length n.

One way to define the solution ($\boldsymbol{x}^*$) to an overdetermined system is the solution which minimizes the norm of the error term.

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} ||\boldsymbol{e}|| = \underset{\boldsymbol{x}}{\operatorname{argmin}} ||A\boldsymbol{x} - \boldsymbol{b}|| = \underset{\boldsymbol{x}}{\operatorname{argmin}} ||A\boldsymbol{x} - \boldsymbol{b}||^2$$

Define a new function $e(\boldsymbol{x})$ as follows:

$$e(\boldsymbol{x}) = ||A\boldsymbol{x} - \boldsymbol{b}||^2 = (A\boldsymbol{x} - \boldsymbol{b})^T(A\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{x}^T A^T A\boldsymbol{x} - \boldsymbol{x}^T A^T \boldsymbol{b} - \boldsymbol{b}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{b}$$

Take the derivative on e with respect to $\boldsymbol{x}$:

$$\frac{d}{d\boldsymbol{x}} e(\boldsymbol{x}) = 2A^T A\boldsymbol{x} - 2A^T\boldsymbol{b}$$

Set the derivative to zero, we can get:

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

If $A^T A$ is invertible (this must be true if the system is overdetermined):

$$\boldsymbol{x} = (A^T A)^{-1} A^T\boldsymbol{b}$$

pseudo-inverse can be interpreted as, when applied to `b` yield solution that minimizes error

Assuming $A \in \mathbb{R}^{m\times n}$, we define the "pseudo-inverse" of a matrix A as:

$$A^{-1} := (A^T A)^{-1} A^T$$

when $A^T A$ is invertible. Note that if A is a square matrix and is invertible, its pseudo-inverse is the same as its inverse (this is also why we use the inverse notation $^{-1}$ for pseudo-inverse, although some people prefer to other notations). A^{-1} is called pseudo-inverse since it has similar propoties as matrix inverse. For example:

$$A^{-1}A = I, \text{ note that } AA^{-1} \text{ may not be identity matrix}$$

$$AA^{-1}A = A$$

$$A^{-1}AA^{-1} = A^{-1}$$

$$(AA^{-1})^T = AA^{-1}$$

Calculating the pseudo-inverse directly from the definition is computationally expensive, since calculating the inverse of $A^T A$ is necessary. We want to find more efficient algorithms for pseudo-inverse. So we will introduce SVD (singular value decomposition).

Assume $M \in \mathbb{R}^{m\times n}$, then M can be written as the product of 3 matrices:

$$M_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$$

where the subscripts describe the dimensions of the matrices. Furthermore we have U, V to be orthogonal and Σ to be diagonal with only non-negative entries on diagonal. Such way of decomposition is calles SVD.

A square matrix Q of real number entries is called orthogonal if we pick any two of the columns $\boldsymbol{v}_i, \boldsymbol{v}_j$, we can get

$$\boldsymbol{v}_i^T\boldsymbol{v}_j = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

A nice property of orthogonal matrix Q is $Q^{-1} = Q^T$.

Now we claim the pseudo-inverse of matrix A is:

$$A^{-1} = V\Sigma^{-1}U^T$$

where $A = U\Sigma V^T$ is the SVD of A, and Σ^{-1} is the pseudo-inverse of Σ. Since Σ is a diagonal matrix, its pseudo-inverse can be calculated by inverting every non-zero terms on the diagonal and then transpose the matrix. Thus we don't need to calculating the inverse of $A^T A$ to get the pseudo-inverse. SInce there is efficient algorithm to calculate SVD (e.g. Golub-Reinsch algorithm), we can efficiently calculate the pseudo-inverse of a matrix.

Finally we need to prove $A^{-1} = V\Sigma^{-1}U^T$. Recall that A^{-1} is defined as $(A^T A)^{-1}A^T$, and $A = U\Sigma V^T$.

$$\begin{aligned} A^{-1} = (A^T A)^{-1}A^T &= ((U\Sigma V^T)^T U\Sigma V^T)^{-1}(U\Sigma V^T)^T \\ &= (V\Sigma^T U^T U\Sigma V^T)^{-1}V\Sigma^T U^T \\ &= (V\Sigma^T\Sigma V^T)^{-1}V\Sigma^T U^T \\ &= (V^T)^{-1}\Sigma^{-1}(\Sigma^T)^{-1}V^{-1}V\Sigma^T U^T \\ &= V\Sigma^{-1}(\Sigma^T)^{-1}\Sigma^T U^T \\ &= V\Sigma^{-1}U^T \end{aligned}$$