Iteration: Class Notes: Nov. 15, 2018

The simplest matrix equation to solve is $Gx = b$, where G is the identity matrix, $G = I$, since then $x = b$. One can take advantage of this if G is near the identity matrix.
Say $G = I - A$ where A is "small" and write the equation $(I - A)x = b$ as

$$x = Ax + b.$$

As our initial approximation we can try anything, say,$x^{(0)} = b$, then let

$$\begin{aligned} x^{(1)} =& Ax^{(0)} + b = Ab + b \\ x^{(2)} =& Ax^{(1)} + b = A(Ab + b) + b = A^2b + Ab + b, \end{aligned}$$

so

$$\begin{aligned} x^{(k+1)} =& Ax^{(k)} + b \\ =& A^kb + A^{k-1}b + \cdots + Ab + b \end{aligned} \tag{1}$$

In order to discuss the convergence of these successive approximations we need to review our discussion of norm of vectors and matrices.

Review: Some Matrix Norms
Square ($n \times n$ matrix $A = (a_{ij})$.

First, some vector norms. Say $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

$$\|x\|_1 = \sum_k |x_k|, \qquad \|x\|_2 = \Big[\sum_k |x_k|^2\Big]^{1/2}, \qquad \|x\|_\infty = \max_k |x_k|.$$

With every vector norm is an associated matrix norm:

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \max_{x\neq 0} \frac{\|Ax.\|}{\|x\|}$$

Thus $\|A\|$ is the maximum A can stretch a unit vector: $\|Ax\| \le \|A\|\|x\|$

For $\|A\|_1$, and $\|A\|_\infty$, it is not difficult to obtain

$\|A\|_1 = \max_j \sum_i |a_{ij}|$, largest absolute column sum

$\|A\|_\infty = \max_i \sum_j |a_{ij}|$, largest absolute row sum

$\|A\|_2$ is more complicated:

$$\|A\|_2^2 = \max_{\|x\|_2=1} \|Ax\|_2^2 = \max_{\|x\|_2=1} \langle Ax,\, Ax\rangle = \max_{\|x\|_2=1} \langle x,\, A^*Ax\rangle = \lambda_{\max},$$

where $\lambda_{\max}$ is the largest eigenvalue of the matrix A^*A (it is positive semi-definite). [1]

[1] A related norm (important in numerical analysis) is the *Frobenius norm*: $\|A\|_F = \big[\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2\big]^{1/2}$.

EXAMPLE:

$$A = \begin{pmatrix} 7 & 2 & 0 \\ 3 & 5 & -1 \\ 0 & 5 & -6 \end{pmatrix}$$

$\|A\|_1 = \max\{10,\ 12,\ 7\} = 12, \qquad \|A\|_\infty = \max\{9,\ 9,\ 11\} = 11,$

All of these special matrix norms norms have the additional *sub-multicative* property $\|AB\| \le \|A\|\|B\|$. In particular $\|A^k\| \le \|A\|^k$ for $k = 1, 2, \ldots$. Consequently, if $\|A\| < 1$, then $\|A^k\| \to 0$. This is a basic reason that matrices A with $\|A\| < 1$ are so useful. For instance, the simple algebraic identity

$$(I - A)(I + A + A^2 + \cdots + A^k) = I - A^{k+1}$$

shows that if $\|A\| < 1$ then $I - A$ is invertible and

$$\|(I - A)(I + A + A^2 + \cdots + A^k) - I\| = \|A^{k+1}\| \le \|A\|^{k+1} \to 0.$$

so

$$I + A + A^2 + \cdots + A^k \to (I - A)^{-1}.$$

This proves the convergence of the iteration in equation (1) and is the essence of many iteration methods in mathematics and its applications. We will see one typical example just below.

[REMARK: The standard example $A = \left(\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right)$ has the property that $\|A\|_1 = 2$ but $A^k \to 0$; indeed, $A^2 = 0$.]

EXAMPLE: We solve $\begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/3 & 1 & 1/2 \\ 1/2 & 1/3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11/18 \\ 11/18 \\ 11/18 \end{pmatrix}$. The matrix on the left has the form $I - A$ where

$$A = -\begin{pmatrix} 0 & 1/2 & 1/3 \\ 1/3 & 0 & 1/2 \\ 1/2 & 1/3 & 0 \end{pmatrix}$$

Both $\|A\|_1 = \|A\|_\infty = 5/6 < 1$ so the iterations of equation (1) converge. Sample output (these should be column vectors):

$$x^{10} = \begin{pmatrix} 0.27950 \\ 0.27950 \\ 0.27950 \end{pmatrix} \qquad x^{40} = \begin{pmatrix} 0.33311 \\ 0.33311 \\ 0.33311 \end{pmatrix} \qquad x^{80} = \begin{pmatrix} 0.33333 \\ 0.33333 \\ 0.33333 \end{pmatrix}$$

Here the convergence is quite slow.

Jacobi Iteration: Diagonally Dominant Matrices

One rarely has matrices that are already of the form $I - A$ where A is small. The class of diagonally dominant matrices is almost as simple and arises much more frequently.

A is *(strictly) row diagonally dominant* if

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad \text{for all } i = 1, 2, \ldots, n.$$

A is *(strictly) column diagonally dominant* if

$$|a_{jj}| > \sum_{i \neq j} |a_{ij}| \quad \text{for all } j = 1, 2, \ldots, n.$$

The above example A is row diagonally dominant but not column diagonally dominant.

Consider the equation $Ax = b$ where A is row diagonally dominant. Write A as

$$A = D + A_0 = \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix} + \begin{pmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & 0 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & 0 & \cdots & a_{3n} \\ & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & 0 \end{pmatrix},$$

where D is the diagonal of A and A_0 is the rest of A. Because A is (strictly) diagonally dominant, none of its diagonal elements are zero. Multiplying by D^{-1} we can write $Ax = b$ as

$$(I + D^{-1}A_0)x = D^{-1}b$$

To understand $D^{-1}A_0$, recall that multiplication on the left by a diagonal matrix multiplies the rows (multiplying on the right multiplies the columns). [2] Thus,

$$D^{-1}A_0 = \begin{pmatrix} 0 & a_{12}/a_{11} & a_{13}/a_{11} & \cdots & a_{1n}/a_{11} \\ a_{21}/a_{22} & 0 & a_{23}/a_{22} & \cdots & a_{2n}/a_{22} \\ a_{31}/a_{33} & a_{32}/a_{33} & 0 & \cdots & a_{3n}/a_{33} \\ & \vdots & & \ddots & \vdots \\ a_{n1}/a_{nn} & a_{n2}/a_{nn} & a_{n3}/a_{nn} & \cdots & 0 \end{pmatrix}$$

But since A was row diagonally dominant, $|a_{11}| > |a_{12}| + |a_{13}| \cdots + |a_{1n}|$. Thus the absolute sum of the first row of $D^{-1}A_0$ is less than 1. Similarly the absolute sum of the each row of $D^{-1}A_0$ is less than 1. This shows that $\|D^{-1}A_0\|_\infty < 1$ so the procedure of equation (1) can be used.

An important virtue of the decomposition $A = D + A_0$ is that D is a diagonal matrix which is almost as easy to work with as the identity matrix.

---

[2] If $C := \begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$ and $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ then $CA = \begin{pmatrix} a_{11}x & a_{12}x \\ a_{21}y & a_{22}y \end{pmatrix}$ and $AC = \begin{pmatrix} a_{11}x & a_{12}y \\ a_{21}x & a_{22}y \end{pmatrix}$.