AMATH 731: Applied Functional Analysis Fall 2014

Banach's fixed point theorem and iterative solutions of systems of linear equations

(These notes are to be inserted after Section 3.9 of the AMATH 731 Course Notes, p. 45)

Banach's fixed point theorem has been of great importance in the study of iterative solutions of systems of linear algebraic equations. It can yield sufficient conditions for convergence as well as error bounds. We shall be concerned with the solution of a system of n linear equations having the form

$$x = Cx + b. \tag{1}$$

Here, $x = (x_1, x_2, \cdots, x_n)$, C is a given $n \times n$ real matrix, and $b = (b_1, b_2, \cdots, b_n)$ a given vector.

The following discussion can be presented in the context of either complete metric spaces or Banach spaces. We begin with a metric space approach. In what follows, our complete metric space/Banach space X will be $\mathbf{R}^n$ equipped with the sup metric/norm, i.e.,

$$d(x, z) = \| x - z \|_\infty = \max_{1 \le i \le n} |x_i - z_i|, \quad x, z \in X. \tag{2}$$

On X we shall define $T : X \to X$ by

$$y = Tx = Cx + b. \tag{3}$$

Clearly, a solution to the linear system (1) is a fixed point of T. We determine sufficient conditions for T to be a contraction on X. It is not too difficult to show that

$$d(Tx, Tz) \le K d(x, z), \tag{4}$$

where

$$K = \max_j \sum_{k=1}^{n} |c_{jk}|. \tag{5}$$

This coincides with our recent result (Course Notes, Section 3.8) for the operator norm $\| C \|$ of the linear operator C under the sup norm. If $K < 1$, then T is a contraction on X, from which we have the following result.

Theorem 1 *If the system of linear equations in (1) satisfies*

$$\sum_{k=1}^{n} |c_{jk}| < 1, \quad j = 1, 2, \cdots, n, \tag{6}$$

then it has precisely one solution x. This solution can be obtained as the limit of the iteration sequence $\{x^{(0)}, x^{(1)}, x^{(2)}, \cdots\}$, where $x^{(0)} \in X$ is arbitrary and

$$x^{(n+1)} = Cx^{(n)} + b, \quad n = 0, 1, \cdots. \tag{7}$$

Furthermore, the error bounds are given by

$$d(x^{(m)}, x) \le \frac{K}{1-K}\, d(x^{(m-1)}, x^{(m)}) \le \frac{K^m}{1-K}\, d(x^{(0)}, x^{(1)}). \tag{8}$$

Condition (6) is sufficient for convergence. It is called a *row sum criterion*. If the sup metric were replaced by another metric, another condition (operator norm for C) would be obtained.

Just to give an idea of how this result is related to practical iteration methods, a system of n linear equations in n unknowns is usually written in the form

$$Ax = c, \tag{9}$$

where A is an $n \times n$ matrix. Assuming that $\det A \neq 0$, many iterative methods express A in the form

$$A = B - G, \tag{10}$$

where B is a nonsingular matrix. Then Eq. (9) becomes

$$Bx = Gx + c, \tag{11}$$

or

$$x = B^{-1}(Gx + c), \tag{12}$$

which has the form of (1) with the correspondence

$$C = B^{-1}G, \quad b = B^{-1}c. \tag{13}$$

This formulation leads naturally to the two standard methods of *Jacobi iteration* and *Gauss-Seidel iteration*. For a brief but readable discussion, see E. Kreyszig, *Introductory Functional Analysis with Applications*, Section 5.2, "Systems of Linear Equations," p. 307-312.

Let us now return to the iterative solution of Eq. (1) from a Banach space/operator norm perspective. The Lipschitz factor K in Eq. (5) is also the operator norm of C, i.e., $K = \| C \|$. Indeed, Eq. (4) can be rewritten as

$$\| Tx - Tz \| = \| Cx - Cz \| \leq \| C \| \| x - z \| . \tag{14}$$

The fixed point relation $x = Cx + b$ may be rewritten as

$$x = (I - C)^{-1}b. \tag{15}$$

If $K = \| C \| < 1$, then, as in Section 3.8, we can use the series expansion for $(I - C)^{-1}$ to give

$$x = (I + C + C^2 + C^3 + \cdots)b. \tag{16}$$

Two remarks here:

1. Note how this expression for x satisfies the equation $Cx + b = x$.
2. Note that this expression holds meaning only in a Banach space setting and not in a metric space one.

The result in (16) may be connected to the iteration sequence $x^{(n)} \to x$ of Eq. (7) as follows: If we start with an arbitrary $x^{(0)} \in X$, then

$$\begin{aligned} x^{(1)} &= Cx^{(0)} + b \\ x^{(2)} &= Cx^{(1)} + b = C^2x^{(0)} + Cb + b \\ x^{(3)} &= Cx^{(2)} + b = C^3x^{(0)} + C^2b + Cb + b \quad \cdots \quad . \end{aligned} \tag{17}$$

In general, we have

$$x^{(n)} = (I + C + C^2 + \cdots + C^{n-1})b + C^n x^{(0)}, \quad n = 1, 2, \cdots. \tag{18}$$

This sequence converges to x in Eq. (16), independent of the choice of $x^{(0)}$, since

$$\| C^n x^{(0)} \| \leq \| C \|^n \| x^{(0)} \| \to 0 \quad \text{as} \quad n \to \infty. \tag{19}$$

Note that in the special case $x^{(0)} = 0$, the result in Eq. (18) becomes

$$x^{(n)} = (I + C + C^2 + \cdots + C^{n-1})b, \quad n = 1, 2, \cdots, \tag{20}$$

which are the nth partial sums to the *Neumann series* given as "f_n" at the end of Section 3.9, p. 45, of the Course Notes.