1 Iterative methods for linear systems

In this section we want to solve the linear system

$$AX = B \tag{1}$$

by a fixed point iteration-like method. The following example illustrates one way this can be done.

Example Solve the system

$$\begin{aligned} 5x - y &= 3, \\ x + 4y &= 9. \end{aligned}$$

we can write this system as

$$\begin{aligned} x &= \frac{3+y}{5}, \\ y &= \frac{9-x}{4} \end{aligned}$$

and use the iterations

$$x_{k+1} = \frac{3+y_k}{5}, \tag{2}$$

$$y_{k+1} = \frac{9-x_k}{4}. \tag{3}$$

We then start with an initial guess (x_0, y_0), plug it in the above equations to compute the next iteration (x_1, y_1) and so on. Notice that we can put the above equations in matrix form as

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{9}{4} \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{5} \\ -\frac{1}{4} & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix}.$$

with $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ one gets $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{9}{4} \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{21}{20} \\ \frac{21}{10} \end{bmatrix}, \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1.02 \\ 1.9875 \end{bmatrix}, \begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0.9975 \\ 1.995 \end{bmatrix}, \ldots$. The true solution is $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, which is well approximated by $\begin{bmatrix} x_4 \\ y_4 \end{bmatrix}$.

Just as we worried about convergence of the fixed point iterations for functions of one variable, we worry about convergence of iterations for linear systems. For example, by exchanging the two equations of the previous example and doing the same iterations again we get

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} -3 \\ 9 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} -39 \\ 24 \end{bmatrix}, \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} -99 \\ 204 \end{bmatrix}.$$

These iterations do not appear to be converging to the solution at all. In what follows we will formalize the iterative methods and look at the question of convergence.

The iterative method used in the previous discussiotn can be set up as follows.

1.1 Jacobi Iterations

For solving the linear system

$$AX = B$$

we write the matrix A in the form

$$A = D + T$$

where D is the diagonal part of A and $T = A - D$ is the upper and lower triangular parts of A. Then we have

$$(D + T)\, X = B$$

or

$$DX = B - TX$$

so that

$$X = D^{-1}\,(B - TX)$$

and the Jacobi method finds a solution by the iterations

$$\begin{aligned} X_{k+1} &= D^{-1}\,(B - TX_k) \\ &= D^{-1}B - D^{-1}TX_k \end{aligned} \tag{4}$$

provided two things: first that the diagonal elements of A are nonzero and second that the above iterations will converge. In the previous example we have

$$\begin{aligned} X_{k+1} &= \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & -\frac{1}{4} \end{bmatrix} \left(\begin{bmatrix} 3 \\ 9 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} X_k \right) \\ &= \begin{bmatrix} \frac{3}{5} \\ \frac{9}{4} \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{5} \\ -\frac{1}{4} & 0 \end{bmatrix} X_k \end{aligned}$$

as before. For the system with the switched equations we have

$$X_{k+1} = \begin{bmatrix} -3 \\ 9 \end{bmatrix} + \begin{bmatrix} 0 & -4 \\ -5 & 0 \end{bmatrix} X_k.$$

To see why the first iterative scheme converges and the second on did not let us consider a general linear iterative method of the form

$$X_{k+1} = SX_k + B.$$

The matrix S is called the iteration matrix. For simplicity we will assume that $B = 0$. Therefore, we will have the iterative scheme

$$X_{k+1} = SX_k$$

Starting with an initial guess X_0, we can easily show that

$$X_1 = SX_0, X_2 = S^2X_0, \ldots, X_k = S^kX_0.$$

Assuming further that the matrix S is diagonalizable, we can write it in the form

$$S = P\Lambda P^{-1}$$

where is a diagonal matrix that contains the eigenvalues of S. Therefore,

$$S^k = P\Lambda^k P^{-1}.$$

Thus, if one of the eigenvalues, say λ_j of S has magnitude larger than one, then $\left|\lambda_j^k\right| \to \infty$ as $k \to \infty$. This shows that in order for the iterative scheme to converge we must have

$$|\lambda_j| < 1, j = 1, 2, \ldots, n$$

(the case one or more λ_j has $|\lambda_j| = 1$is sort of an undetermined case, we may or may not have convergence.) Applying this criterion to the previous two iterations give:

Eigenvalues of the first iteration matrix $\begin{bmatrix} 0 & \frac{1}{5} \\ -\frac{1}{4} & 0 \end{bmatrix}$ are $\pm\frac{1}{2\sqrt{5}}i$ which have magnitude $\frac{1}{2\sqrt{5}} < 1$, while the eigenvaules of the second iteration matrix $\begin{bmatrix} 0 & -4 \\ -5 & 0 \end{bmatrix}$ are $\pm\sqrt{20}$, both havind magnitude greater than one.

In practical problems where the matrix A has a very large dimension, in the order of thousands, it is not feasible of course to compute the eigenvalues of the iteration matrix in order to check for convergence. We would like to be able to tell whether we have convergence or not by inspecting the matrix A itself. For this we introduce the following definition.

Definition 1 *An $n \times n$ matrix A is called strictly diagonally dominant if*

$$|a_{kk}| > \sum_{\substack{j=1, \\ j \neq k}}^{n} |a_{kj}|, \ k = 1, 2, , \ldots, n.$$

Theorem 2 *If the matrix A of the linear system (1) is strictly diagonally dominant, then the system has a unique solution P and the Jacobi iterations (4) will converge to P for any choice of the initial guess P_0.*

It remains to clarify what is meant here by the convergence of a sequence of vectors.

1.2 Convergence of sequences of vectors

A sequence of vectors $\{X_k\}_{k=1}^{\infty}$ will be said to converge to a vector X if the *distance* between X_k and X goes to zero as $k \to \infty$. For the distance between two vecots P and Q we can take the Euclidean norm

$$\|P - Q\| = \left(\sum_{k=1}^{n} |p_k - q_k|^2 \right)^{1/2}.$$

Clearly, if $\|P - Q\| = 0$, then $P = Q$. Therefore, convergence of sequences of vectors means that the "sequence" of norms $\|X_k - X\|$ converges to zero as $k \to \infty$. The disadvantage of the Euclidean norm, however, is that it is more costly to compute. Instead, we use the norm

$$\|P - Q\|_1 = \sum_{k=1}^{n} |p_k - q_k| .$$

This norm satisfies the general requirements of a norm; namely:

1. $\|P\|_1 \geq 0$, and $\|P\|_1 = 0$ if and only if $P = 0$,
2. $\|cP\|_1 = |c| \, \|P\|_1$ for all $c \in \mathbb{R}$,
3. $\|P + Q\|_1 \leq \|P\|_1 + \|Q\|_1$.

1.3 The Gauss-Seidel Iterations

An alternative method to the Jacobi iteration is known as the Gauss-Seidel method. The idea behind the method is to update any component of an iterations as soon as it is computed. Thus, for the iterations (2), (3), we use x_{k+1} once it is computed. The equations become

$$\begin{aligned} x_{k+1} &= \frac{3 + y_k}{5}, \\ y_{k+1} &= \frac{9 - x_{k+1}}{4}. \end{aligned}$$

In matrix form, this system becomes

$$\begin{bmatrix} 1 & 0 \\ \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} \\ \frac{9}{4} \end{bmatrix} + \begin{bmatrix} 0 & \frac{1}{5} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \end{bmatrix}.$$

after 3 iterations we get $\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0.999 \\ 2.0002 \end{bmatrix}$, so the convergence is accelerated by this updating.

The Gauss Seidel iterative method for the general linear system (1) can be put in matrix form as follows. First we write the matrix A into diagonal, upper triangular and lower triangular form as

$$A = D + L + U.$$

The linear system then may be written as

$$(D + L + U)\, X = B.$$

Sending the upper triangular part to the right hand side and multiplying by D^{-1} gives

$$\left(I + D^{-1}L\right) X = D^{-1}B - D^{-1}UX,$$

which leads to the iterative method

$$\left(I + D^{-1}L\right) X_{k+1} = D^{-1}B - D^{-1}UX_k. \tag{5}$$

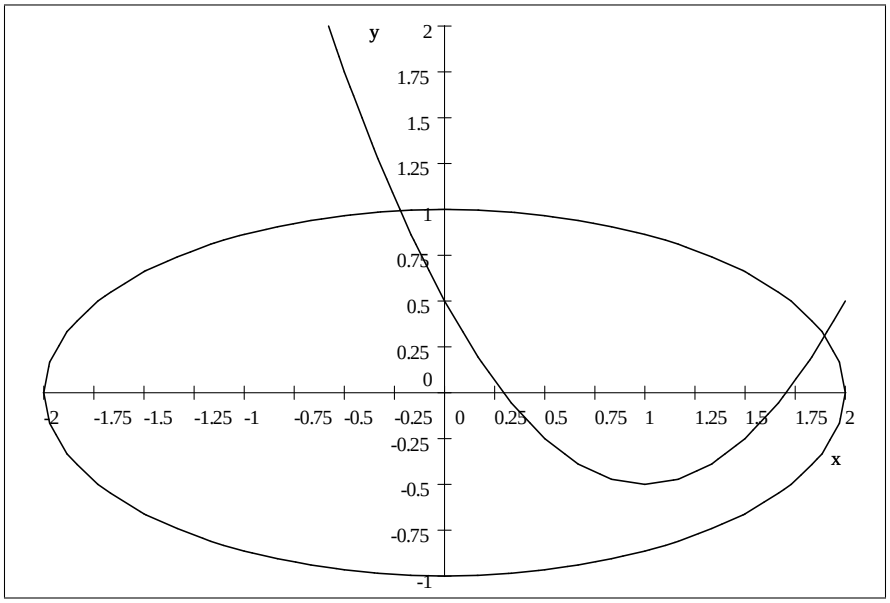
Observe that X_{k+1} is obtained by a forward sweep cycle since $\left(I + D^{-1}L\right)$ is lower triangular with ones along the diagonal.

Theorem 3 *If the matrix A of the linear system (1) is strictly diagonally dominant, then the system has a unique solution P and the Gauss-Seidel iterations (5) will converge to P for any choice of the initial guess P_0.*

2 Iterative Methods for Nonlinear Systems

Suppose we have a system of nonlinear equations, such as

$$\begin{aligned} x^2 - 2x - y + 1/2 &= 0, \\ x^2 + 4y^2 - 4 &= 0. \end{aligned}$$



that we want to solve simultaneously. The first rquation represents a parabola and the second one represents an ellipse. The two curves intersect in two points

as shown in the figure. Therefore, the system has two solutions. To develop numerical methods for the solution of such systems we need to use functions of several variables. The system can be written as

$$G(\mathbf{x}) = \mathbf{0}, \tag{6}$$

where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and

$$G(\mathbf{x}) = G(x, y) = \begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix} = \begin{bmatrix} x^2 - 2x - y + 1/2 \\ x^2 + 4y^2 - 4 \end{bmatrix}.$$

More generally we will be considering functions G whose domains are subsets of $\mathbb{R}^N$ and whose ranges are in $\mathbb{R}^N$, that is $G : D \subset \mathbb{R}^N \to \mathbb{R}^N$. Such functions will be represented as

$$G(\mathbf{x}) = \begin{bmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_N(\mathbf{x}) \end{bmatrix}$$

and we will use the notation $G'(\mathbf{x})$ to denote their total derivatives:

$$G'(\mathbf{x}) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_N} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_N} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial g_N}{\partial x_1} & \frac{\partial g_N}{\partial x_2} & \cdots & \frac{\partial g_N}{\partial x_N} \end{bmatrix}.$$

For example, G' for the function G above is

$$G'(\mathbf{x}) = \begin{bmatrix} 2x - 2 & -1 \\ 2x & 2y \end{bmatrix}.$$

We will also need to be able to assign a norm to an $N \times N$ matrix A in order to be able to talk about its "size". One such norm is the so called maximum row sum norm, which is defined by

$$\|A\| = \max_{1 \le i \le N} \left\{ \sum_{j=1}^{N} |a_{ij}| \right\}.$$

For example, the norm of the matrix

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

is calculated as

$$\|A\| = \max\{4, 14, 15\} = 15.$$

We will also be interested in finding solutions to the fixed point equation

$$\mathbf{x} = G(\mathbf{x}). \tag{7}$$

2.1 Fixed Point Iterations

The fixed point iterations for the system (7) take the form

$$\mathbf{x}_{k+1} = G(\mathbf{x}_k) \tag{8}$$

with an initial guess $\mathbf{x}_0$.

Theorem 4 *Suppose $C \subset \mathbb{R}^N$ is closed and bounded, $G : C \to C$ and $\|G'(\mathbf{x})\| < 1$ for alll $\mathbf{x} \in C$. Then G has a unique fixed point $\mathbf{p} \in C$. Furthermore, the iterations (8) converge to $\mathbf{p}$ for any choice of the initial guess $\mathbf{x}_0 \in C$.*

Example Let

$$G(\mathbf{x}) = G(x, y) = \begin{bmatrix} \frac{x^2 - y + 1/2}{2} \\ \frac{-x^2 - 4y^2 + 8y + 4}{8} \end{bmatrix}.$$

Then

$$G'(\mathbf{x}) = \begin{bmatrix} x & -1/2 \\ -1/4x & -y + 1 \end{bmatrix}.$$

The norm $\|G'(\mathbf{x})\|$ is

$$\|G'(\mathbf{x})\| = \max\left\{ |x| + 1/2, \frac{|x|}{4} + |y - 1| \right\}.$$

The condition $\|G'(\mathbf{x})\| < 1$ means that we must have

$$|x| + 1/2 < 1$$

and

$$\frac{|x|}{4} + |y - 1| < 1.$$

The first inequality gives

$$-1/2 < x < 1/2$$

and the second one gives

$$|y - 1| < 1 - \frac{|x|}{4},$$

which means

$$\begin{aligned} |y - 1| &< \min\left(1 - \frac{|x|}{4}\right) \\ &= 1 - \max\frac{|x|}{4} = 1 - \frac{1}{8} = \frac{7}{8}. \end{aligned}$$

Therefore

$$\frac{1}{8} < y < \frac{15}{8}.$$

Thus, for all $\mathbf{x}$ inside the rectangle $C = [-1/2, 1/2] \times [1/8, 1]$, we have $\|G'(\mathbf{x})\| < 1$. Using the techniques of For all $\mathbf{x} \in C$,

$$\begin{aligned}
\frac{x^2 - y + 1/2}{2} &< \frac{1/4 - 1/8 + 1/2}{2} = \frac{5}{16} < 1/2, \\
\frac{x^2 + y + 1/2}{2} &> \frac{-1 + 1/2}{2} = -\frac{1}{4} > -1/2, \\
\frac{-x^2 - 4y^2 + 8y + 4}{8} &< \frac{-4 + 8 + 4}{8} = 1, \\
\frac{-x^2 - 4y^2 + 8y + 4}{8} &> \frac{-1/4 - 4 + 1 + 4}{8} = \frac{-1/4 + 1}{8} > 1/8.
\end{aligned}$$

It follows that $G : C \to C$. Thus, by the above theorem, G has a unique fiexed point in C and we can use the fixed point iterations with $\mathbf{x}_0 = (0, 1/2)$ to obtain it. Performing the calculations we get

$$\mathbf{x}_1 = (0, 0.8750), \mathbf{x}_2 = (-0.1875, 0.9922), \cdots, \mathbf{x}_5 = (-0.2222, 0.9938).$$

2.2 Newton Method

Newton method for solving the system

$$G(\mathbf{x}) = \mathbf{0}$$

has the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k - G'(\mathbf{x}_k) G(\mathbf{x}_k).$$

Examlpe Let

$$G(\mathbf{x}) = \begin{bmatrix} x^2 - 2x - y + 1/2 \\ x^2 + 4y^2 - 4 \end{bmatrix}.$$

Then

$$G'(\mathbf{x}) = \begin{bmatrix} 2x - 2 & -1 \\ 2x & 2y \end{bmatrix}.$$

Starting with the initial guess $\mathbf{x}_0 = (0, 1/2)$ and iterating we get

$$\mathbf{x}_1 = (2., 1/2), \mathbf{x}_2 = (1.9167, 0.3333), \cdots, \mathbf{x}_4 = (1.9007, 0.3112).$$