Chapter 2
Solution of Systems of Linear Equations

2.1 Introduction

Solving systems of linear equations (or *linear systems* or, also, *simultaneous equations*) is a common situation in many scientific and technological problems. Many methods, either analytical or numerical, have been developed to solve them. A general method most used in Linear Algebra is the Gaussian Elimination, or variations of this. Sometimes they are referred to as "direct methods". Basically, it is an algorithm that transforms the system into an equivalent one but with a triangular matrix, thus allowing a simpler resolution. In many cases, though, whenever the matrix of the system has a specific structure or is sparse and the like, other methods can be more effective. They are "iterative methods", not based on a finite algorithm but on an iterative process. The two simpler methods of this kind are the Jacobi and the Gauss–Seidel methods. There is also an iterative method which transforms solving the system into minimizing a scalar function: the "Method of the Steepest Descent".

In our case, we present here a simple iterative method based on the motion of a damped harmonic oscillator in a linear force field plus a constant force field, such as the gravitational field on the surface of the Earth [13, 31]. This mechanical system evolves, under the action of time, towards the solution of the linear equations. We will apply very similar ideas to compute eigenvectors of matrices and to solve linear and nonlinear programming problems in further chapters. In the next one we will compare the performance of the mechanical method with that of the iterative methods mentioned above.

To illustrate the basic ideas, let us consider the one-dimensional case. We want to solve the (trivial, in one dimension) linear equation

$$ax = b\,. \tag{2.1}$$

To do this, we will consider a mechanical system whose solution tends to the solution of this equation. The equation of motion for a particle under a linear force

L. Vázquez and S. Jiménez, *Newtonian Nonlinear Dynamics for Complex Linear and Optimization Problems*, Nonlinear Systems and Complexity 4,
DOI 10.1007/978-1-4614-5912-5_2, © Springer Science+Business Media New York 2013

and a constant external acceleration is

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2}+\alpha\frac{\mathrm{d}x}{\mathrm{d}t}+ax=b\,, \tag{2.2}$$

where $x=x(t)$ is the one-dimensional displacement of the mass m under a dissipation ($\alpha>0$), a harmonic force with $a>0$, and a constant acceleration (b, due to a gravitational field, for instance). From the theory of linear differential equations, we know that the general solution to (2.2) can be expressed as the sum of two contributions, the *general solution of the homogeneous part*, x_h, plus a *particular solution*, x_p:

$$x(t)=x_h(t)+x_p(t)\,. \tag{2.3}$$

Moreover, since

$$\lim_{t\to\infty}x_h(t)=0,\qquad \lim_{t\to\infty}x_p(t)=\frac{b}{a}\,, \tag{2.4}$$

the solution can be expressed as a sum of a time-dependent term, $x_0(t)$, plus a constant one:

$$x(t)=x_0(t)+\frac{b}{a}\,, \tag{2.5}$$

with $x_0(t)$ decaying exponentially to zero as t goes to infinity (if both α and a are positive, as we suppose). The specific rate of that exponential decay depends on the coefficients in Eq. (2.2), but in every case, and independently of the initial conditions, the solution tends to the *asymptotically stable* constant solution given by $x(t)=b/a$.

From a mechanical point of view, the system has a total energy given by

$$E=\frac{m}{2}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2+\frac{1}{2}ax^2-bx\,, \tag{2.6}$$

with variation law

$$\frac{\mathrm{d}E}{\mathrm{d}t}=-\alpha\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2. \tag{2.7}$$

This implies, α being positive, that E will decrease under the evolution. Since E is bounded from below by the minimum value $-b^2/2a$, the system will approach asymptotically a state given by

$$\frac{\mathrm{d}x}{\mathrm{d}t}=0\,,\quad x=\frac{b}{a}\,. \tag{2.8}$$

We see that both approaches, this Mechanical one and solving directly Eq. (2.1), give the same result. In fact, the equivalence of both methods is part of Liapunov stability theory.

Remarks. 1. In the *overdamped* regime ($\alpha > 4am$), when α is large enough as compared to a and m, the dominant effect is the dissipative one and the equation of motion can be approximated by the first order equation:

$$\alpha \frac{\mathrm{d}x}{\mathrm{d}t} + ax = b\,, \tag{2.9}$$

whose solutions have the same asymptotic behaviour as those of (2.2), and tend to the constant solution $x = b/a$, independently of the initial conditions. The presence of coefficient α can be avoided rescaling time.

2. If we allow a to be zero in (2.2), the behaviour drastically changes and, in general, *secular* terms appear that are neither constant nor vanishing. If both, a and b, are zero, the system evolves again into a constant solution that depends now on the initial conditions. The same happens with (2.9). In the general case of many dimensions, this corresponds to undetermined systems of linear algebraic equations.

2.2 General Case

The mechanical ideas considered previously are the bases for the two families of iterative methods that we present to solve a system of linear algebraic equations. The methods are constructed choosing a finite difference method to solve either one of the associated linear differential equations, (2.2) or (2.9). In the next two sections, we will present the general case, extending the previous basic, one-dimensional, mechanical considerations to the case of a system of equations.

In this approach, we translate the problem of solving a system of linear equations into solving the equation of motion of one particle that tends asymptotically to a position which is identified with the solution of that linear system. Because of this asymptotic behaviour, we can expect to have general iterative methods which need many iterations to approach the solution, but the convergence is satisfied whenever the discretization of the differential equation of motion satisfies conditions related to the conservation and variation of the energy of the system.

Let be the linear system

$$A\vec{x} = \vec{b}\,, \tag{2.10}$$

where we assume that A is an $q \times q$ non-singular matrix (i.e., the system has a unique solution). We may associate to it Newton's equation for a linear dissipative ($\alpha > 0$) mechanical system:

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} + \alpha \frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + A\vec{x} = \vec{b} \tag{2.11}$$

and the overdamped asymptotic equation

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + A\vec{x} = \vec{b}\,. \tag{2.12}$$

For both equations, if A has a real, positive definite spectrum we have

$$\lim_{t\to\infty} \vec{x}(t) = A^{-1}\vec{b}, \tag{2.13}$$

which is the solution of the linear system (2.10). Since A is not necessarily a positive definite matrix, we may consider an equivalent problem, given by

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} + \alpha\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + M\vec{x} = \vec{v} \tag{2.14}$$

or

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + M\vec{x} = \vec{v} \tag{2.15}$$

in which:

$$M\vec{x} = \vec{v} \iff A\vec{x} = \vec{b}, \tag{2.16}$$

and such that M is positive definite. Depending on the properties of A, we may choose different possibilities for M and $\vec{v}$ (see [32]):

(1) When the spectrum of A is real and positive: $\ddot{\vec{x}} + \alpha\dot{\vec{x}} + A\vec{x} = \vec{b}$.
(2) When the spectrum of A is real and negative: $\ddot{\vec{x}} + \alpha\dot{\vec{x}} - A\vec{x} = -\vec{b}$.
(3) When the spectrum of A is real: $\ddot{\vec{x}} + \alpha\dot{\vec{x}} + AA\vec{x} = A\vec{b}$.
(4) When A has complex eigenvalues: $\ddot{\vec{x}} + \alpha\dot{\vec{x}} + A^{\mathrm{T}}A\vec{x} = A^{\mathrm{T}}\vec{b}$.

We have make use of the "dotted" notation to represent the time derivatives in a more compact way.

In order to avoid problems with the spectrum of A, which is in principle not known beforehand, we will consider in what follows

$$M = A^{\mathrm{T}}A, \quad \vec{v} = A^{\mathrm{T}}\vec{b}. \tag{2.17}$$

Although this may not be a good idea if A is ill-conditioned (see Exercise 2.5, below, and also the discussion of formula (2.7.40) in page 85 of [25]), we, thus, ensure that M is symmetric and positive definite by construction and has, therefore, a real, positive definite spectrum, and that the constant solution $\vec{x}(t) = M^{-1}\vec{v} = A^{-1}\vec{b}$ is asymptotically stable.

As before, we can use the existence of a decreasing energy for (2.14):

$$\begin{aligned} E(\vec{x}) &= \frac{1}{2}\frac{\mathrm{d}\vec{x}^{\mathrm{T}}}{\mathrm{d}t}\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + \frac{1}{2}\vec{x}^{\mathrm{T}}M\vec{x} - \vec{x}^{\mathrm{T}}\vec{v}, \\ &= \frac{1}{2}\left\|\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\right\|^2 + \frac{1}{2}\vec{x}^{\mathrm{T}}M\vec{x} - \vec{x}^{\mathrm{T}}\vec{v}, \end{aligned} \tag{2.18}$$

where $\|\cdot\|$ is the Euclidean vector norm, or 2-norm. The variation law is

$$\frac{\mathrm{d}E(\vec{x})}{\mathrm{d}t} = -\alpha\left\|\frac{\mathrm{d}\vec{x}}{\mathrm{d}t}\right\|^2. \tag{2.19}$$

Once the evolution equation is chosen, we integrate it with a numerical scheme in order to approach the asymptotically stable solution. Since we have exchanged our problem for an equivalent one, we must keep in mind that our method should never be more time-consuming than tackling the original problem: we look thus for an explicit, finite-difference method to solve the equations. We have chosen the forward Euler method. We will call it the *damped* method when applied to (2.14) and the *overdamped* one when applied to (2.15).

2.3 Damped Method

A suitable, finite difference scheme to solve (2.14) is the following [22, 29]:

$$\frac{\vec{x}_{n+1} - 2\vec{x}_n + \vec{x}_{n-1}}{\tau^2} + \alpha \frac{\vec{x}_{n+1} - \vec{x}_{n-1}}{2\tau} + M\vec{x}_n = \vec{v}, \tag{2.20}$$

where τ is the mesh-size of the time variable and $\vec{x}_n$ denotes the position at time $t = n\tau$. We have chosen it since it is the simplest finite difference method to solve (2.14). The scheme is a two-step recursion, as it should since it represents a second order equation, and here we need two initial values to start the computations. Since we are looking for an asymptotically stable solution, we may choose, for instance, $\vec{x}_1 = \vec{x}_0$ and $\vec{x}_0 \neq \vec{0}$ arbitrary. To actually perform the numerical computations, we may express the scheme as

$$\vec{x}_{n+1} = \frac{1}{1 + \frac{\tau\alpha}{2}} \left[\left(2I - \tau^2 M\right) \vec{x}_n - \left(1 - \frac{\tau\alpha}{2}\right) \vec{x}_{n-1} + \tau^2 \vec{v} \right], \tag{2.21}$$

where I denotes the identity matrix of the appropriate order. If we multiply (2.20) on the left by $(\vec{x}_{n+1} - \vec{x}_{n-1})^{\mathrm{T}}/2\tau$, and rearrange terms, we obtain

$$\frac{E_{n+1} - E_n}{\tau} = -\alpha \left(\frac{\vec{x}_{n+1} - \vec{x}_{n-1}}{2\tau} \right)^2, \tag{2.22}$$

where

$$E_{n+1} \equiv \frac{1}{2} \left(\frac{\vec{x}_{n+1} - \vec{x}_n}{\tau} \right)^2 + \frac{1}{2} \vec{x}_n^{\mathrm{T}} M \vec{x}_{n+1} - \frac{1}{2} \left(\vec{x}_{n+1}^{\mathrm{T}} \vec{v} + \vec{x}_n^{\mathrm{T}} \vec{v} \right) \tag{2.23}$$

is the discrete counterpart of the energy (2.18).

As we can see, the variation of the discrete energy associated with the difference equation (2.20) is similar to that in (2.19), and its decreasing character depends only on the sign of α and not on the solution. This property guarantees the convergence of the numerical solution to that of the system (2.10), provided the value of τ is small enough, unless it has no solution. If that is the case, Eq. (2.14) has a linear component that grows linearly in time and the solution of (2.20) does not converge to a constant.

Equation (2.20) has two arbitrary parameters, α and τ. We will see in what follows how to choose them in order to optimize the computations.

Although a single equation such as (2.21) is more accurate, computation-wise, for the sake of the analysis we translate (2.20) into a system of two equations. Keeping in mind the Mechanical analogy we define:

$$\vec{p}_n = \frac{\vec{x}_{n+1} - \vec{x}_n}{\tau}, \tag{2.24}$$

which is a consistent discrete representation of the momentum (since $m = 1$). With this and (2.20) the scheme becomes

$$\begin{cases} \left(\frac{\alpha}{2} I + \tau M\right) \vec{x}_{n+1} + \left(1 + \frac{\tau\alpha}{2}\right) \vec{p}_{n+1} = \frac{\alpha}{2} \vec{x}_n + \vec{p}_n + \tau \vec{v}, \\ \vec{x}_{n+1} = \vec{x}_n + \tau \vec{p}_n, \end{cases} \tag{2.25}$$

where I is the $q \times q$ identity matrix. Let us write this in matrix and vector form as:

$$N_+ \vec{Y}_{n+1} = N_- \vec{Y}_n + \vec{W}, \tag{2.26}$$

with the block matrices and stacked vectors:

$$N_+ = \left(\begin{array}{c|c} \frac{\alpha}{2} I + \tau M & \left(1 + \frac{\tau\alpha}{2}\right) I \\ \hline I & O \end{array}\right), \quad N_- = \left(\begin{array}{c|c} \frac{\alpha}{2} I & I \\ \hline I & \tau I \end{array}\right), \tag{2.27}$$

$$\vec{Y}_{n+1} = \begin{pmatrix} \vec{x}_{n+1} \\ \vec{p}_{n+1} \end{pmatrix}, \quad \vec{Y}_n = \begin{pmatrix} \vec{x}_n \\ \vec{p}_n \end{pmatrix}, \quad \vec{W} = \begin{pmatrix} \tau \vec{v} \\ \vec{0} \end{pmatrix}, \tag{2.28}$$

where O is the $q \times q$ null matrix.

It can be easily seen that matrix N_+ is invertible unless $\tau\alpha = -2$, which cannot occur since α and τ are both positive. Thus we have an iterative process that can be written formally as

$$\vec{Y}_{n+1} = (N_+)^{-1} N_- \vec{Y}_n + (N_+)^{-1} \vec{W}. \tag{2.29}$$

A sufficient condition to ensure the convergence of this process for any initial values is to have all eigenvalues of the iteration matrix

$$N \equiv (N_+)^{-1} N_- \tag{2.30}$$

of modulus strictly less than 1. Let us compute those eigenvalues. From the characteristic equation, we have (in block form)

$$\lambda \text{ is eigenvalue of } N \iff \left|\begin{array}{c|c} (1-\lambda)\frac{\alpha}{2} I - \lambda \tau M & \left[1 - \lambda\left(1 + \frac{\tau\alpha}{2}\right)\right] I \\ \hline (1-\lambda) I & \tau I \end{array}\right| = 0$$

(and dealing with columns to get an upper triangular block matrix:)

$$\iff \left| \begin{array}{c|c} -\lambda\tau M + (1-\lambda)\left[\frac{\alpha}{2} - \frac{1}{\tau} + \frac{\lambda}{\tau}\left(1+\frac{\alpha\tau}{2}\right)\right] I & \left[1-\lambda\left(1+\frac{\tau\alpha}{2}\right)\right] I \\ \hline O & \tau I \end{array} \right| = 0$$

$$\iff \left| M - \frac{1-\lambda}{\lambda\tau}\left[\frac{\alpha}{2} - \frac{1}{\tau} + \frac{\lambda}{\tau}\left(1+\frac{\alpha\tau}{2}\right)\right] I \right| = 0$$

$$\iff \frac{1-\lambda}{\lambda\tau}\left[\frac{\alpha}{2} - \frac{1}{\tau} + \frac{\lambda}{\tau}\left(1+\frac{\alpha\tau}{2}\right)\right] \quad \text{is an eigenvalue of } M,$$

$$\iff \left(1+\frac{\alpha\tau}{2}\right)\lambda^2 + (\mu\tau^2 - 2)\lambda + \left(1-\frac{\alpha\tau}{2}\right) = 0, \tag{2.31}$$

where μ is any eigenvalue of M. Thus, for every eigenvalue μ of M, we get two eigenvalues of N:

$$\lambda_\pm(\mu,\tau,\alpha) = \frac{2-\mu\tau^2 \pm \tau\sqrt{\mu^2\tau^2 - 4\mu + \alpha^2}}{2+\alpha\tau}. \tag{2.32}$$

If we want the fastest convergence rate, we should look for values of α and τ such that $|\lambda|$ be as small as possible and, in any case, strictly less than 1: in this way our iterative process will be convergent.

A fundamental property of the second order equation in λ (2.31) is that for any eigenvalue μ of M, we have

$$\lambda_+(\mu,\tau,\alpha)\,\lambda_-(\mu,\tau,\alpha) = \frac{2-\tau\alpha}{2+\tau\alpha}, \tag{2.33}$$

and, thus, independent of the value of μ. Since the time step τ is positive, this quantity is less than 1: if we can manage to have $\lambda_\pm$ complex, non-real, it means that both would have a modulus strictly less than 1, since in that case they will be complex conjugate and $\lambda_+\lambda_- = |\lambda_\pm|^2$. Thus the iteration would be convergent. Moreover, we may look for optimal values of τ and α in the following way: let us consider some specific eigenvalue μ. We want:

$$\begin{aligned} \lambda_+ = \bar{\lambda}_- \text{ (complex, non-real)} &\iff \mu^2\tau^2 - 4\mu + \alpha^2 \le 0 \\ &\iff \mu \in [\mu_-, \mu_+], \end{aligned} \tag{2.34}$$

where:

$$\mu_- = \frac{2-\sqrt{4-\tau^2\alpha^2}}{\tau^2}, \quad \mu_+ = \frac{2+\sqrt{4-\tau^2\alpha^2}}{\tau^2}. \tag{2.35}$$

This can be inverted to give:

$$\tau = \frac{2}{\sqrt{\mu_+ + \mu_-}}, \quad \alpha = 2\sqrt{\frac{\mu_+ \mu_-}{\mu_+ + \mu_-}}. \tag{2.36}$$

If we want this to hold for every eigenvalue μ, we may choose μ_- as the smallest eigenvalue of M, and μ_+ as the greatest: these are real positive values since we have chosen $M = A^{\mathrm{T}}A$ and, thus, symmetric and positive definite. In this way we ensure convergence of the method independently of the initial conditions.

In fact, the eigenvalues μ of M are related to the singular values σ of the original matrix A:

$$\mu = \sigma^2. \tag{2.37}$$

We may thus define σ_+ and σ_-, accordingly. Once these values are known, or equivalently μ_+ and μ_-, we compute via (2.36) τ and α and get a rough a priori estimate of the rate of convergence. From (2.32) we have

$$|\lambda| = \sqrt{\frac{2 - \tau\alpha}{2 + \tau\alpha}} = \frac{\sigma_+ - \sigma_-}{\sigma_+ + \sigma_-} \tag{2.38}$$

and we may estimate the error at iteration step n by $|\lambda|^n$. The key point in choosing the most favourable values for parameters τ and α seems, thus, to compute, or at least estimate, μ_+ and μ_-. We present in Chap. 4, and implement in Chap. 5, an iterative method that can be used to achieve this in a reasonable way. Otherwise, tentative values of τ and α may be used. A problem arises whenever μ_- is much smaller than μ_+, since $|\lambda|$ will be very close to unity and the convergence very slow.

Although a priori information about μ_+ and μ_- is not easy to obtain, at least there is a bound that can be established in a simple way: since M is a positive definite matrix, all its eigenvalues are real and strictly positive and, thus, its trace, which corresponds to the sum of all the eigenvalues, is an upper bound to $\mu_+ + \mu_-$. This provides us with the following lower bound for the optimal value of τ:

$$\tau \geq \frac{2}{\sqrt{\mathrm{tr}(M)}}, \tag{2.39}$$

In the case of two dimensions, we have in fact that

$$\mu_+ + \mu_- = \mathrm{tr}(M), \quad \mu_+ \mu_- = \det(M), \tag{2.40}$$

and we can determine the optimal values of both parameters without computing the eigenvalues:

$$\tau = \frac{2}{\mathrm{tr}(M)}, \quad \alpha = 2\sqrt{\frac{\det(M)}{\mathrm{tr}(M)}}. \tag{2.41}$$

We will use this in Example 3.1.1, in next chapter.

2.4 Overdamped Method

To simulate Eq. (2.15) we use the numerical scheme

$$\frac{\vec{x}_{n+1}-\vec{x}_n}{\tau}+M\vec{x}_n=\vec{v}, \tag{2.42}$$

or equivalently

$$\vec{x}_{n+1}=(I-\tau M)\vec{x}_n+\tau\vec{v}. \tag{2.43}$$

This corresponds to the forward Euler Method applied to the overdamped equation (2.15). If we take $\alpha=2/\tau$ in (2.21), we obtain

$$\vec{x}_{n+1}=\left(I-\frac{\tau^2}{2}M\right)\vec{x}_n+\frac{\tau^2}{2}\vec{v}, \tag{2.44}$$

and we see that this overdamped method is just a particular case of the damped one with an effective time step $\tau'=\tau^2/2$. Since the value of α is fixed with respect to the mesh-size τ, the optimal values obtained in the previous section cannot be applied here in general, so we expect this method to be less efficient than the damped one. Nevertheless, we will perform its analysis: here we do not need an extra variable and the method is convergent simply if the matrix $I-\tau M$ has all its eigenvalues of modulus less than one.

$$\lambda \text{ is eigenvalue of } I-\tau M \iff |I-\tau M-\lambda I|=0 \tag{2.45}$$

$$\iff \left|M-\frac{1-\lambda}{\tau}I\right|=0 \iff \frac{1-\lambda}{\tau} \text{ is eigenvalue of } M\,. \tag{2.46}$$

Again, letting μ be any eigenvalue of M, we have

$$\mu=\frac{1-\lambda}{\tau} \iff \lambda=1-\tau\mu\,. \tag{2.47}$$

Now all λ's are real and the condition for all of them to be of modulus less than one corresponds to

$$0<\tau<\frac{2}{\mu_+}, \tag{2.48}$$

where μ_+ is the largest eigenvalue of M. This condition by itself is not sufficient to determine the optimal value of τ. In fact, what would be required is to minimize all the possible values of λ. This implies, in turn, knowing, beforehand, all the eigenvalues of M, which is not realistic in the general case. On the other hand, we may estimate the error decay to go as $|\lambda|^n$, n being the number of iterations, which

implies that we should try to minimize this value as much as possible. We, thus, may consider as a reference the two following values:

$$\tau_+ = \frac{1}{\mu_+}, \quad \tau_- = \frac{1}{\mu_-}. \tag{2.49}$$

We expect the optimal value of τ, that is, the one that minimizes the number of iterations to achieve a given precision, to lie between these two.

2.5 Singular Matrix

Let be the linear dissipative ($\alpha > 0$) mechanical system:

$$\frac{\mathrm{d}^2\vec{x}}{\mathrm{d}t^2} + \alpha\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} + A^{\mathrm{T}}A\vec{x} = \vec{0}. \tag{2.50}$$

If the matrix A is singular, the equation $A\vec{x} = \vec{0}$ has an infinite number of solutions, otherwise only the trivial solution $\vec{x} = \vec{0}$ exists. As a consequence, when numerically computing the solution of the equation above with arbitrary initial conditions, we have two possible behaviours:

- Matrix A is *non-singular*: in this case all numerical solutions converge to $\vec{0}$ as time goes to infinity.
- Matrix A is *singular*: in this case some numerical solution converges to a vector different from $\vec{0}$ as time goes to infinity.

It is clear that the fixed points of the dynamical system correspond to vectors belonging to the kernel of A. The problem is to determine whether some are non-null and use this as a criterion to determine whether a matrix is singular or not.

We will see in the implementations in next chapter how to proceed in a practical way but here we illustrate the two behaviours through the following example.

Example 2.5.1. Let be

$$A = \begin{pmatrix} -1 & -2 & -8 & -12 & 4 \\ -3 & 4 & 8 & 15 & -7 \\ -1 & 3 & 11 & 17 & -7 \\ 1 & -1 & -4 & -6 & 3 \\ 1 & 3 & 12 & 19 & -6 \end{pmatrix}.$$

This is a non-singular matrix, and the iterative process tends to $\vec{0}$, independently of the initial vector chosen.

On the other hand,

$$A = \begin{pmatrix} 2 & -2 & -8 & -9 & 7 \\ -6 & 4 & 8 & 12 & -10 \\ -5 & 3 & 11 & 13 & -11 \\ 2 & -1 & -4 & -5 & 4 \\ -4 & 3 & 12 & 14 & -11 \end{pmatrix}$$

is singular. Starting, for instance, with initial vector $\vec{x}_0 = (1,0,0,0,1)^{\mathrm{T}}$, we obtain numerically convergence towards vector:

$$(-0.243500, 0.243500, 0.324667, -0.0811667, 0.405833)^{\mathrm{T}}.$$

Our result is valid for a general matrix. Thus, we may determine the singular character of a matrix computing numerically some solutions of an associated linear dissipative mechanical system.

A different possibility is to check whether matrix A has a zero eigenvalue. We will address this case in Chap. 5.

2.6 Exercises

2.1 Show that no positive values μ_+ and μ_- exist such that the overdamped method is just the damped one with optimal values for the parameters α and τ.

2.2 (a) Compute M, $\mu_\pm$, τ, α and determine the value of $|\lambda|$ for matrix

$$A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}.$$

(b) Compute the estimated number of iterations needed to obtain an absolute error $\|A\vec{x}_n - \vec{b}\|$ less that 10^{-12}.

2.3 Matrix A, defined as

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

is well conditioned, since the rows correspond to an orthogonal set of vectors.

(a) Compute M, $\mu_\pm$, τ, α and determine the value of $|\lambda|$.

(b) Compute the estimated number of iterations needed to obtain an absolute error $\|A\vec{x}_n - \vec{b}\|$ less that 10^{-12}. Is there a significant difference with the result of the previous exercise?

2.4 Repeat the previous exercise using now the matrix

$$A = \begin{pmatrix} 1\,0\,1 \\ 0\,1\,0 \\ -1\,0\,1 \end{pmatrix}.$$

2.5 Hilbert matrices are known to be ill conditioned [30]. Let be the Hilbert matrix of order two:

$$H = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix},$$

and a linear system: $H\vec{x} = \vec{b}$. Since Hilbert matrices are symmetric and positive definite, we do not need to transform them, building $M = H^{\mathrm{T}}H$, but we may use directly $M = H$ in our mechanical method.

(a) Find the corresponding values of $\mu_{\pm}$, τ, α and determine the value of $|\lambda|$.
(b) Compute the estimated number of iterations needed to obtain an absolute error $\|H\vec{x}_n - \vec{b}\|$ less than 10^{-12}. Compare with Exercise 2.3.

2.6 Repeat the previous exercise, using now the Hilbert matrix of order three:

$$H = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix},$$

and compare with Exercise 2.4.

2.7 Compute, for the overdamped method, the estimates $\tau_{\pm}$ for the 2×2 matrices of Exercises 2.3 and 2.5.

2.8 Repeat the previous exercise, now with the 3×3 matrices of Exercises 2.4 and 2.6.

2.9 In the study of wave propagation in one-dimensional media, for instance for chains of coupled oscillators or for partial differential equations simulated in finite differences, linear systems occur with three-diagonal symmetric matrices of the form:

$$A = \begin{pmatrix} a & b & 0 & \cdots & 0 & 0 & 0 \\ b & a & b & 0 & \cdots & 0 & 0 \\ 0 & b & a & b & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & b & a & b & 0 \\ 0 & 0 & \cdots & 0 & b & a & b \\ 0 & 0 & 0 & \cdots & 0 & b & a \end{pmatrix},$$

with $a, b \in \mathbb{R}$. The eigenvalues of these matrices are:

$$a + 2|b| \cos\left(\frac{k\pi}{q+1}\right), \quad k = 1, \ldots, q, \tag{2.51}$$

q being the number of rows and columns of the matrix. The condition number of a matrix A in the matrix 2-norm $\|\cdot\|_2$ is (see Eq. (2.7.5), page 80, of [14]):

$$\kappa_2(A) = \|A\|_2 \, \|A^{-1}\|_2 = \frac{\sigma_{\max}}{\sigma_{\min}}.$$

(a) Find this condition number for A, as a function of q. Would be such a system well or ill conditioned?
(b) Find the optimal values of parameters α and τ and determine the value of $|\lambda|$. Keep in mind that, since A is symmetric we may use it as matrix M.
(c) Typical values are $a = -2$ and $b = 1$. For these values, compute the estimated number of iterations needed to obtain an absolute error $\|A\vec{x}_n - \vec{b}\|$ less than 10^{-12} as a function of q.

2.10 (a) Show that if A is a 2×2 singular matrix, the value of α given by (2.36) is zero. What value for α can you suggest in such a case?
(b) Is α zero for any singular matrix A regardless of its dimensions?

2.11 Companion matrices appear in eigenvalue problems, in relation with characteristic polynomials

$$\lambda^q + a_{q-1}\lambda^{q-1} + \cdots + a_1\lambda + a_0 \, .$$

They are sparse matrices with the coefficients a_k in the last column (with opposite signs) and the elements of the second diagonal, just below the main one, all equal to 1:

$$C_q = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdots & -a_0 \\ 1 & 0 & 0 & 0 & \cdots & -a_1 \\ 0 & 1 & 0 & 0 & \cdots & -a_2 \\ 0 & 0 & 1 & 0 & \cdots & -a_3 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{q-1} \end{pmatrix}.$$

By direct computations it can be established that $C^{\mathrm{T}}C$ has a first $(q-1) \times (q-1)$ block which is just the identity matrix, has values $-a_1$ to $-a_{q-1}$ as first elements of the last column and row, while the (q, q) element is $\sum_{k=0}^{q-1} a_k^2$:

$$C_q^{\mathrm{T}} C_q = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ 0 & 0 & 1 & \cdots & 0 & -a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{q-1} \\ -a_1 & -a_2 & -a_3 & \cdots & -a_{q-1} & \sum_{k=0}^{q-1} a_k^2 \end{pmatrix}.$$

It is easily seen that this matrix has eigenvalue $\lambda = 1$ with multiplicity $q-2$, the two other being the roots of the second order equation:

$$\lambda^2 - \left(1 + \sum_{k=0}^{q-1} a_k^2\right)\lambda + a_0^2 = 0\,.$$

(a) As a preliminary step, determine C_q the expression of all the eigenvalues of $C_q^{\mathrm{T}} C_q$.

(b) Consider now $q = 5$, and the polynomial

$$\lambda^5 - 6\lambda^4 + 10\lambda^3 - 11\lambda + 6\,.$$

Compute $\mu_\pm$ and the optimal values of the parameters τ and α.

(c) Consider now $q = 4$, with arbitrary coefficients a_k. Using the previous result, determine for C_4 the values of $\mu_\pm$ and the optimal values of the parameters τ and α.

http://www.springer.com/978-1-4614-5911-8

Newtonian Nonlinear Dynamics for Complex Linear and Optimization Problems

Vazquez, L.; Jimenez, S.

2013, XII, 140 p., Hardcover
ISBN: 978-1-4614-5911-8