Journal of Computational Mathematics, Vol.19, No.1, 2001, 9–14.

ANALYSIS OF A MECHANICAL SOLVER FOR LINEAR SYSTEMS OF EQUATIONS*[1)]

Luis Vázquez
(*Dept. Matemática Aplicada, Facultad de Informática, Universidad Complutense, 28040-Madrid, Spain*)

Salvador Jiménez
(*Dept. Matemática y Física Aplicadas, Universidad Alfonso X El Sabio, Avda. Universidad 1, 28691-Villanueva de la Cañada, Madrid, Spain*)

Dedicated to the 80th birthday of Professor Feng Kang

Abstract

In this contribution we analyse some fundamental features of an iterative method to solve systems of linear equations, following the approach introduced in a previous work[1]. Such questions range from optimal parameters and initial conditions to comparison with other methods. An interesting result is that *a priori* we can give an estimation of the number of iterations to get a given accuracy.

Key words: Iterative method, Linear systems, Classical dynamics.

1. Introduction

A new approach to solve systems of linear equations, equivalent to solve the motion of a damped harmonic oscillator, has been proposed in a previous paper[1]. Due to this parallelism, we call such methods *Mechanical Solvers* for systems of linear equations. The present study is devoted to the analysis of these methods.

Let be the linear system

$$A\vec{x} = \vec{b} \tag{1}$$

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

$$\vec{x}_{tt} + \alpha\vec{x}_t + A\vec{x} = \vec{b}. \tag{2}$$

If A has a positive real spectrum, we have

$$\lim_{t\to\infty} \vec{x}(t) = A^{-1}\vec{b} \tag{3}$$

which is the solution of the linear system (1). Different equations of motion can be proposed for the system above, of the form

$$\vec{x}_{tt} + \alpha\vec{x}_t + M\vec{x} = \vec{v} \tag{4}$$

* Received November 8, 2000.

1) This work has been partially supported by the Comisión Interministerial de Ciencia y Tecnología of Spain under grant PB98-0850.

such that:

$$M\vec{x} = \vec{v} \iff A\vec{x} = \vec{b} \tag{5}$$

In order to avoid problems with the spectrum of A, we may choose

$$M = A^{\mathrm{T}}A,\ \vec{v} = A^{\mathrm{T}}\vec{b}. \tag{6}$$

Although this may not be a good idea if A is ill conditionned [2], we ensure that M is symmetric and positive definite by construction and thus has a real, positive definite spectrum. This will be used in what follows.

The next step is to solve the differential equation with a simple finite-difference scheme, such as:

$$\frac{\vec{x}_{n+1} - 2\vec{x}_n + \vec{x}_{n-1}}{\tau^2} + \alpha\frac{\vec{x}_{n+1} - \vec{x}_{n-1}}{\tau} + M\vec{x}_n = \vec{v} \tag{7}$$

Every finite-difference method associated to (4) will define an iterative process to solve the system (5).

2. Analysis of the Numerical Scheme

Although a single equation is more accurate, for the sake of the analysis we translate (7) into a system of two equations. Keeping in mind the Mechanical analogy we define:

$$\vec{p}_n = \frac{\vec{x}_{n+1} - \vec{x}_n}{\tau} \tag{8}$$

with this and (7) the scheme becomes

$$\begin{cases} \vec{x}_{n+1} = \vec{x}_n + \tau\vec{p}_n \\ \\ \left(\dfrac{\alpha}{2}I + \tau M\right)\vec{x}_{n+1} + \left(1 + \dfrac{\tau\alpha}{2}\right)\vec{p}_{n+1} = \dfrac{\alpha}{2}\vec{x}_n + \vec{p}_n + \tau\vec{v} \end{cases} \tag{9}$$

where I is the $m \times m$ identity matrix. Let us write this in block-matrix form as:

$$\underbrace{\left(\begin{array}{c|c} \frac{\alpha}{2}I + \tau M & \left(1 + \frac{\tau\alpha}{2}\right)I \\ \hline I & \mathcal{O} \end{array}\right)}_{N_+} \underbrace{\left(\begin{array}{c} \vec{x}_{n+1} \\ \hline \vec{p}_{n+1} \end{array}\right)}_{\vec{Y}_{n+1}} = \underbrace{\left(\begin{array}{c|c} \frac{\alpha}{2}I & I \\ \hline I & \tau I \end{array}\right)}_{N_-} \underbrace{\left(\begin{array}{c} \vec{x}_n \\ \hline \vec{p}_n \end{array}\right)}_{\vec{Y}_n} + \underbrace{\left(\begin{array}{c} \tau\vec{v} \\ \hline \vec{0} \end{array}\right)}_{\vec{W}} \tag{10}$$

and define N_+, N_-, $\vec{Y}_{n+1}$, $\vec{Y}_n$ and $\vec{W}$ as indicated in the previous formula. We have thus an iterative process that we may write formally as

$$\vec{Y}_{n+1} = (N_+)^{-1}N_-\vec{Y}_n + (N_+)^{-1}\vec{W} \tag{11}$$

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

$$N \equiv (N_+)^{-1}N_- \tag{12}$$

of modulus strictly less than 1. Let us compute those eigenvalues:

$$\lambda \text{ is eigenvalue of } N \iff \left|\begin{array}{c|c} (1-\lambda)\frac{\alpha}{2}I - \lambda\tau M & \left[1 - \lambda\left(1 + \frac{\tau\alpha}{2}\right)\right]I \\ \hline (1-\lambda)I & \tau I \end{array}\right| = 0 \tag{13}$$

(dealing with columns to get an uppertriangular block matrix:)

$$\iff \left| M - \frac{1-\lambda}{\lambda\tau}\left[\frac{\alpha}{2} - \frac{1}{\tau} + \frac{\lambda}{\tau}\left(1 + \frac{\alpha\tau}{2}\right)\right] I \right| = 0 \tag{14}$$

$$\iff \left(1 + \frac{\alpha\tau}{2}\right)\lambda^2 + (\mu\tau^2 - 2)\lambda + \left(1 - \frac{\alpha\tau}{2}\right) = 0 \tag{15}$$

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

$$\lambda_\pm(\mu,\tau,\alpha) = \frac{2 - \mu\tau^2 \pm \tau\sqrt{\mu^2\tau^2 - 4\mu + \alpha^2}}{2 + \alpha\tau} \tag{16}$$

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 smaller than 1).

A fundamental property is that for any eigenvalue μ of M, we have

$$\lambda_+(\mu,\tau,\alpha)\,\lambda_-(\mu,\tau,\alpha) = \frac{2 - \tau\alpha}{2 + \tau\alpha} \tag{17}$$

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 $\lambda_\pm$ imaginary, it means that both would have a modulus less than 1, and 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: $\lambda_+ = \bar{\lambda}_-$ imaginary (not real)

$$\iff \mu^2\tau^2 - 4\mu + \alpha^2 \leq 0 \iff \mu \in [\mu_-, \mu_+] \tag{18}$$

where:

$$\mu_- = \frac{2 - \sqrt{4 - \tau^2\alpha^2}}{\tau^2}, \quad \mu_+ = \frac{2 + \sqrt{4 - \tau^2\alpha^2}}{\tau^2}. \tag{19}$$

This can be inverted to give:

$$\tau = \frac{2}{\sqrt{\mu_+ + \mu_-}}, \quad \alpha = 2\sqrt{\frac{\mu_+\,\mu_-}{\mu_+ + \mu_-}}. \tag{20}$$

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 to be symmetric and positive definite.

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

$$\mu = \sigma^2. \tag{21}$$

We may thus define σ_+ and σ_-. Once these values are known (or equivalently μ_+ and μ_-, and we will see later how they can be estimated) we compute τ and α and get an *a priori* estimate of the rate of convergence. From (16) we have

$$|\lambda| = \sqrt{\frac{2 - \tau\alpha}{2 + \tau\alpha}} = \frac{\sigma_+ - \sigma_-}{\sigma_+ + \sigma_-} \tag{22}$$

and **we may estimate the error at iteration step n with $|\lambda|^n$.**

3. Numerical Examples

3.1 Dimension 2:

We consider a very simple case:

$$A = \begin{pmatrix} 4 & 2 \\ -1 & 3 \end{pmatrix}, \vec{b} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} M = \begin{pmatrix} 17 & 5 \\ 5 & 13 \end{pmatrix}, \tag{23}$$

In this 2×2 case, M has just two eigenvalues, μ_1 and μ_2:

$$\begin{cases} \mu_+ + \mu_- = \mu_1 + \mu_2 = \text{Trace}(M) = 30 \\ \mu_+ \, \mu_- = \mu_1 \, \mu_2 = \det(M) = \det(A)^2 = 14^2. \end{cases} \tag{24}$$

$$\begin{cases} \tau = \frac{2}{\sqrt{30}} \approx 0.365148372 \\ \alpha = \frac{28}{\sqrt{30}} \approx 5.112077203 \end{cases} \tag{25}$$

and we have $|\lambda| \approx 0.186$. If we want a precision less than 10^{-12} the estimated number of iteration is 17. In figure 1, we compare this method with Jacobi, Gauss-Seidel and the Method of the Steepest Descent.

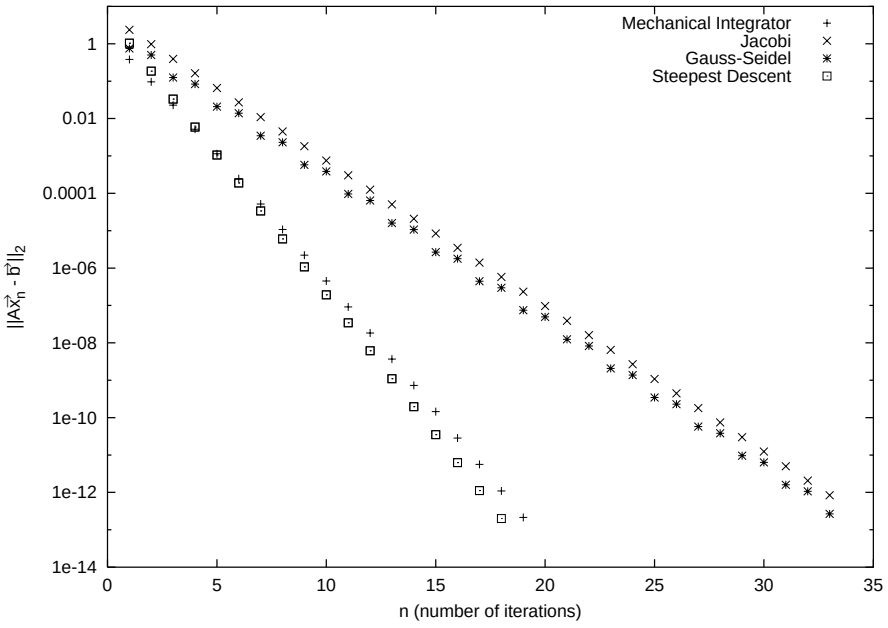

Figure 1: Comparison with other iterative methods: dimension 2

3.2 Dimension 5:

We consider here a randomly generated matrix but somewhat diagonal dominant, so Jacobi and Gauss-Seidel methods can converge.

$$A = \begin{pmatrix} -33 & -9 & 8 & -5 & -10 \\ 2 & 23 & -5 & 6 & 10 \\ 9 & -12 & 35 & -7 & 3 \\ 14 & 9 & -8 & 33 & 10 \\ -5 & -15 & -7 & 3 & 30 \end{pmatrix}, \vec{b} = \begin{pmatrix} 1 \\ -5 \\ 7 \\ 11 \\ -3 \end{pmatrix} \tag{26}$$

The eigenvalue μ_+ can be estimated using the Power Method on $A^T A$ (in general, since a great precision is not necessary, only a few iterations are needed). In order to estimate μ_-, we apply the power method to the matrix $\mu_+ I - M$ of which the greatest eigenvalue is $\mu_+ - \mu_-$. In this case we obtain $\tau \approx 0.03537269250$, $\alpha \approx 34.90239342$, and $|\lambda| \approx 0.486$. To obtain a precision better than 10^{-12}, we estimate 39 iterations. The comparison with other methods is plotted in figure 2

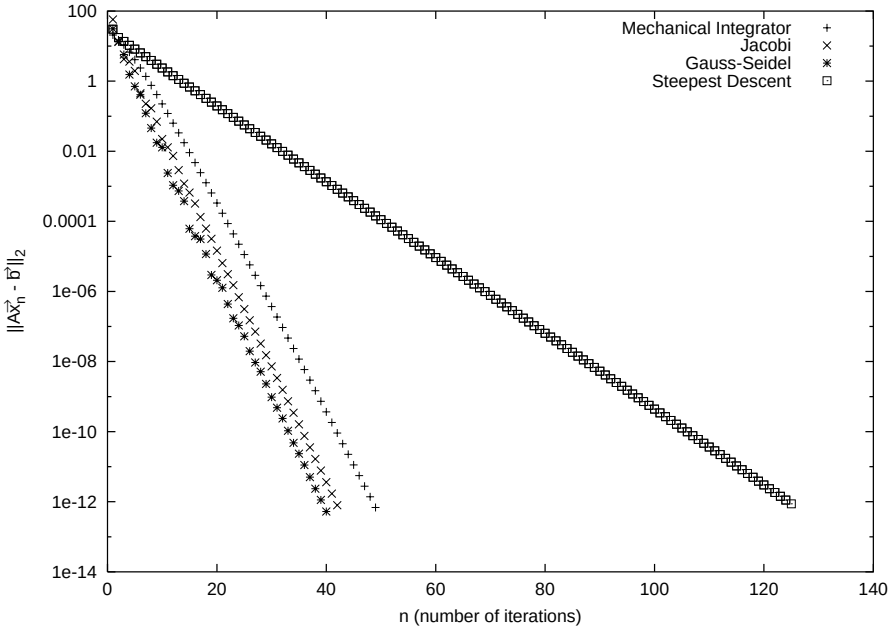

Figure 2: Comparison with other iterative methods: dimension 5

4. Conclusion

The method is a general one: it does not depend on a specific form or properties of the original matrix. It is a simple method and can be as fast as the most common ones. Furthermore it has an *a priori* estimate of the number of iteration needed to achieve a given precision, which

may decide whether or not the method is useful for a given specific problem. To optimize the convergence, it is necesary to compute μ_- and μ_+, and this can be done in a simple way by means of a few iterations of the Power Method.

References

[1] L. Vázquez, J.L. Vázquez-Poletti, *Journal of Computational Mathematics,* (2000) in press.

[2] See the discussion of formula (2.7.40), page 85 of: W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery and J.G.P. Barnes, *Numerical Recipes in C*, second edition, Cambridge University Press, Cambridge 1995.