Chapter 8
Numerical Stability Properties

Phillip Regalia and Richard Le Borne

Abstract Designers of algorithms must not only solve the problem of interest, but do so using methods which are robust under perturbations in the data as well as the intermediate parameters of the method. More generally, it is often the case that the actual problem of interest is too complicated to solve directly; simplifying assumptions are necessary. At each stage, from problem identification, to the setup of the problem to be solved using some method, to the ultimate algorithm to be implemented in code, perturbations and their effects must be anticipated and analyzed. Stability is the property that assesses the level of robustness to perturbations that is required before the computed solution given by an algorithm can be used with confidence. The origin of the perturbation can vary, as pointed out above. What is important, however, is to have analysis that supports the premise that a small change in the problem results in a small change to the solution.

8.1 Introduction

The use of the QR-decomposition (vs. QR-algorithm) depends greatly on its reputation for providing consistently usable results. When an algorithm’s reputation suffers, whether deserved or not, users often seek an alternative. It is therefore important, if not essential, to first establish the criteria in which a reliable solution can be guaranteed. For example, the reputation of Gaussian elimination suffered greatly in the 1940s because of its inability to always provide a usable solution. It was not until an analysis performed by J. H. Wilkinson [1], that introduced the
8.2 Preliminaries

The usefulness of a computed solution is a statement assessing its numerical accuracy. When the estimate of the desired solution is a vector, required is a means for assessing its accuracy. The vector norm assigns a single number to a vector and this property is very useful for assessing the error in a given quantity. There are different vector norms, but the most often used are considered equivalent in $\mathbb{R}^N$ in that one norm is within a constant factor of another. For example, the Euclidean norm, or two-norm, is defined as the square-root of the sum of squares of vector elements. For the vector of filter weights, $w = [w_1, w_2, \ldots, w_N]^T$, its Euclidean norm, $\|w\|_2$ is given by

$$\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \cdots + w_N^2} \quad (8.1)$$

$$= \sqrt{\sum_{i=1}^{N} w_i^2} \quad (8.2)$$

For matrices, the norm again is used to associate a single number to it but its definition is chosen to be compatible with vector norms. For example, the matrix norm associated with the Euclidean norm is called the spectral norm. For $X \in \mathbb{R}^{N \times M}$, its spectral norm, $\|X\|_2$, is defined as

$$\|X\|_2 = \max_{\|w\|_2 = 1} \|Xw\|_2 \quad (8.3)$$

$$= \sqrt{\lambda_{\text{max}}(X^T X)} \quad (8.4)$$

$$= \sigma_{\text{max}}(X), \quad (8.5)$$

where $\lambda_{\text{max}}(X^T X)$ is the maximum eigenvalue of $X^T X$ and $\sigma_{\text{max}}(X)$ is the largest singular value of the data matrix $X$. The singular value of a matrix will next be defined. For a full discussion on vector and matrix norms a good source is [2, Chapter 2].

Suppose at time index $k$, we are given the data matrix $X(k) \in \mathbb{R}^{(k+1) \times (N+1)}$ of rank $r$. Then there are unitary matrices $U(k) \in \mathbb{R}^{(k+1) \times (k+1)}$ and $V(k) \in \mathbb{R}^{(N+1) \times (N+1)}$ where $U^T(k) U(k) = I \in \mathbb{R}^{(k+1) \times (k+1)}$ and $V^T(k) V(k) = I \in \mathbb{R}^{(N+1) \times (N+1)}$ and a diagonal matrix $\Sigma(k) \in \mathbb{R}^{(k+1) \times (N+1)}$ where $\Sigma(k) = \text{diag}(\sigma_1(k), \sigma_2(k), \ldots, \sigma_r(k), 0, \ldots, 0) = \text{diag}(\Sigma_+(k), 0, \ldots, 0)$ with $\sigma_1(k) \geq \sigma_2(k) \geq \cdots \geq \sigma_r > 0$. Then the singular value decomposition is given by: