Transient Analysis of High-Speed Channels via Newton-GMRES Waveform Relaxation

Salvatore Bernardo Olivadese, Stefano Grivet-Talocia
Dip. Elettronica e Telecomunicazioni, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
Ph +39 011 0904104, Fax +39 011 0904099
e-mail stefano.grivet@polito.it

***Abstract*—This paper presents a technique for the numerical simulation of coupled high-speed channels terminated by arbitrary nonlinear drivers and receivers. The method builds on a number of existing techniques. A Delayed-Rational Macromodel is used to describe the channel in compact form, and a general Waveform Relaxation framework is used to cast the solution as an iterative process that refines initial estimates of transient scattering waves at the channel ports. Since a plain Waveform Relaxation approach is not able to guarantee convergence, we turn to a more general class of nonlinear algebraic solvers based on a combination of the Newton method with a Generalized Minimal Residual iteration, where the Waveform Relaxation equations act as a preconditioner. The convergence of this scheme can be proved in the general case. Numerical examples show that very few iterations are indeed required even for strongly nonlinear terminations.**

I. Introduction and Motivation

This paper discusses a numerical technique for the simulation of high-speed channels with arbitrary terminations. This type of simulation is nowadays essential to characterize the signal degradation effects due to the frequency-dependent channel dispersion and losses, combined with the spurious effects possibly introduced by the nonlinearities of drivers and receivers. The most common metric to assess the quality of a high-speed link is the eye diagram opening.

Eye diagram simulations are performed in time-domain. In order to obtain a rich statistical information, long bit sequences must be simulated, imposing stringent requirements on the efficiency of the adopted solver. Most recent approaches based on inverse Fourier transform of the channel responses and fast convolutions [1] sacrifice accuracy for speed by approximating driver and receiver characteristics with linear models. Although very fast, these approaches may not capture important effects on the resulting waveforms due to the presence of nonlinearities. In this paper, we aim at a general scheme that will not impose such brute approximation.

Standard SPICE simulations [2] provide an alternative and accurate approach. SPICE solvers are excellent to handle nonlinear circuits but generally fail in providing an efficient solution of large distributed circuits like the fully coupled high-speed channels considered in this work, even if state-of-the-art behavioral macromodels are used to represent the channel [3].

Recently, the class of Waveform Relaxation (WR) schemes [4] has been revitalized for high-speed channel simulation. The basic idea of WR is to partition a large system or circuit into subparts that are weakly interacting. Each part is solved independently by neglecting the inter-part interactions, which are used as correction terms within an iterative loop. This approach is very appealing, since it enables naive parallelization on modern multicore computing architectures. Recent contributions on WR simulation of transmission lines and high-speed channels are available in [5], [6] and [7], [8].

Despite its promise for fast solution of mixed nonlinear-distributed circuits, the WR approach may fail to converge in some cases [9]. It is clear that any numerical scheme, as fast as it can be, cannot be employed for systematic and routine channel analysis when the occurrence of instability due to lack of convergence cannot be controlled a priori. This consideration motivated our approach in [10], where the transient simulation approach was cast as the solution of a large-scale and very sparse linear system, whose matrices represent the discretized operators of channel and terminations. In particular, the Krylov subspace iteration denoted Generalized Minimal RESidual (GMRES) scheme [11] was used in [10]. Fast convergence was obtained by using the WR equations as a preconditioner.

By its very nature, this WR-GMRES scheme is only applicable to linear terminations. Here, we extend this approach to the general case of nonlinear terminations. The nonlinear algebraic system obtained by discretizing the dynamic operators of channel and terminations is solved by a combination of Newton method with the GMRES solver, known as Newton-GMRES scheme [12]. Again, the WR part is used as a preconditioner. We apply this scheme to industrial channel models terminated by strongly nonlinear terminations, designed on purpose to stress the solver. The numerical results show that convergence occurs, as expected, in very few iterations, which are further reduced by the WR preconditioner. SPICE validations confirm the correctedness of the waveforms.

II. Problem statement

Figure 1 provides a graphical illustration of the stucture under investigation. A P-port (P even) fully-coupled high-speed channel $\mathcal{H}$ is loaded by terminations $\mathcal{F}_i, i = 1, \ldots, P$. The channel is possibly characterized by a complex topology, including routing through package, connectors, vias, and board, and is formed by $P/2$ individual point to point interconnects $\mathcal{H}_q,\ q = 1, \ldots, P/2$ that are coupled through

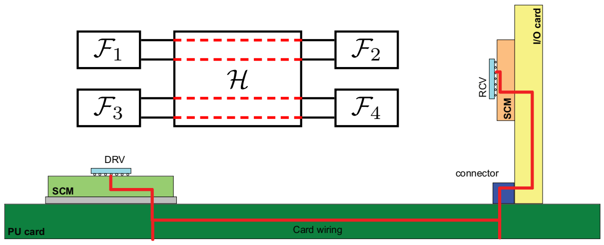

Fig. 1. A high-speed channel and its schematic representation.

near end electromagnetic interaction. The channel is described by its $P \times P$ scattering matrix, available at a finite set of frequencies through direct measurement or cascade connection of electromagnetic models of its constitutive parts. The channel examples that are considered in this work (courtesy of IBM) have $P = 18$ ports. The terminations $\mathcal{F}_i$ are supposed to be independent and decoupled from each other. Such terminations may be available as transistor-level circuits or behavioral models.

A. Discrete-time channel macromodels

Since the numerical simulation will have to be performed in time domain, we start by casting the channel characteristics through a time-domain macromodel. Due to its distributed nature, with possibly large end-to-end propagation delays, we choose to describe the channel through a Delayed-Rational Macromodel (DRM), whose closed form expression is [3]

$$H^{i,j}(s) = \sum_{m=0}^{M^{i,j}} \sum_{n=1}^{N_m^{i,j}} \frac{R_{mn}^{i,j}}{s - p_{mn}^{i,j}} e^{-s\tau_m^{i,j}} + D^{i,j} \tag{1}$$

in the Laplace s-domain and

$$h^{i,j}(t) = \sum_{m=0}^{M^{i,j}} \sum_{n=1}^{N_m^{i,j}} R_{mn}^{i,j} e^{p_{mn}^{i,j}(t-\tau_m^{i,j})} u(t-\tau_m^{i,j}) + D^{i,j}\delta(t) \tag{2}$$

in time domain, where i, j denote output and input port, respectively, and $\tau_m^{i,j}$ are delays corresponding to the various arrival times of the signal reflections caused by internal and port discontinuities. These delays and the coefficients $p_{mn}^{i,j}$, $R_{mn}^{i,j}$ are identified as described in [2], [3] directly from the scattering matrix samples.

An advantage of the DRM form is evident after time discretization of (2) with a fixed time step δt over the interval $t \in [0, T]$. Denoting $t_k = k\delta t$ and extracting a single impulse response term $h(t) = e^{p(t-\tau)} u(t - \tau)$ from (2), we see that the discretized convolution with an arbitrary signal $a(t)$ leads to an output signal $b(t)$ whose samples are obtained via

$$b(t_k) = (h * a)(t_k) \simeq \alpha_0 b(t_{k-1}) + \sum_{\nu=0}^{2} \beta_\nu a(t_{k-\nu-\bar{k}}) \tag{3}$$

where $\bar{k} = \left\lfloor \frac{\tau}{\delta t} \right\rfloor$. This corresponds to a recursive convolution with suitable precomputed coefficients [7], which results particularly efficient if compared to a standard convolution.

We collect all incident and reflected scattering waves at the channel ports into time-dependent vectors $\boldsymbol{a}(t)$ and $\boldsymbol{b}(t)$. After time discretization, all time samples $\boldsymbol{a}(t_k), \boldsymbol{b}(t_k)$ for $k = 0, \ldots, K$ with $K = T/\delta t$ are collected into block vectors $\mathbf{a}$ and $\mathbf{b}$. The channel operator that returns $\mathbf{b}$ in terms of $\mathbf{a}$ reads

$$\mathbf{b} = \mathcal{H}\,\mathbf{a}\,, \tag{4}$$

where $\mathcal{H}$ can be interpreted as a very large size matrix, which is never computed and stored, but whose application to vector $\mathbf{a}$ requires a linear superposition of recursive convolution terms (3) in only $O(KP)$ operations.

B. Termination equations

Using the above notation, the block scattering vector $\mathbf{b}$ collects the time samples of the incident scattering wave into the terminations. The reflected scattering wave samples can be formally described as

$$\mathbf{a} = \mathcal{F}(\mathbf{b})\,, \tag{5}$$

where $\mathcal{F}$ is a block-diagonal nonlinear operator. In the linear case, as assumed in [10], this expression becomes

$$\mathbf{a} = \mathcal{G}\mathbf{b} + \mathcal{Q}\mathbf{u}\,, \tag{6}$$

where $\mathcal{G}, \mathcal{Q}$ are large and very sparse matrices which can be automatically obtained from a Modified Nodal Analysis (MNA) of the termination circuits, and $\mathbf{u}$ collects all time samples of the independent sources within the terminations.

C. Two-level Waveform Relaxation

A two-level WR scheme was presented in [7] to iteratively solve the nonlinear system of coupled channel and termination equations (4)-(5). The channel is hierarchically partitioned by first splitting inter-channel couplings $\mathcal{C}$ from the individual transmission and reflection coefficients $\mathcal{D}$ as $\mathcal{H} = \mathcal{D} + \mathcal{C}$, and then splitting each decoupled channel from its terminations. This gives rise to the following iterative scheme

$$\begin{aligned} &\text{for} \quad \mu = 1, \ldots, \mu_{\max} \\ &\quad \text{for} \quad \nu = 1, \ldots, \bar{\nu} \\ &\qquad \mathbf{b}_{\mu,\nu} = \mathcal{D}\,\mathbf{a}_{\mu,\nu-1} + \boldsymbol{\theta}_{\mu-1}\,, \\ &\qquad \mathbf{a}_{\mu,\nu} = \mathcal{F}\left(\mathbf{b}_{\mu,\nu}\right)\,, \\ &\quad \boldsymbol{\theta}_\mu = \mathcal{C}\,\mathbf{a}_{\mu,\bar{\nu}}\,, \end{aligned} \tag{7}$$

denoted as WR-LPTP (from "Longitudinal Partitioning-Transverse Partitioning") in [7]. Note that channel and termination equations are never coupled and solved concurrently, but they are just evaluated alternatively during LPTP iterations. This scheme, which corresponds to a fixed point iteration, is only characterized by conditional convergence and is not generally applicable, as discussed in [9]. However, when convergence holds, the solution is attained very fast, as far as a compact form of the nonlinear termination operator $\mathcal{F}$ is available, e.g., as a behavioral macromodel.

D. The linear case

By restricting the analysis to linear terminations only (6), in [10] we formulated the channel simulation as a solution of a large-scale sparse linear system

$$(\mathcal{I}-\mathcal{G}\mathcal{H})\mathbf{a}=\mathcal{Q}\mathbf{u}\,, \tag{8}$$

where application of individual operators $\mathcal{G}$, $\mathcal{H}$ and $\mathcal{Q}$ can be performed very fast. Due to this fact, in [10] we proposed to use an iterative scheme based on Krylov-subspace projection in order to obtain the solution without inverting or factoring the system matrix $(\mathcal{I}-\mathcal{G}\mathcal{H})$. The method of choice was the well-known GMRES scheme, applied to compute successive estimates of the solution $\mathbf{a}_n=\mathbf{a}_0+\boldsymbol{\delta}_n$, where $\boldsymbol{\delta}_n$ are elements of the Krylov subspace

$$\mathcal{K}_n=\operatorname{span}\left\{\mathbf{r}_0,(\mathcal{I}-\mathcal{G}\mathcal{H})\mathbf{r}_0,\ldots,(\mathcal{I}-\mathcal{G}\mathcal{H})^{n-1}\mathbf{r}_0\right\} \tag{9}$$

and $\mathbf{r}_0=\mathcal{Q}\mathbf{u}-(\mathcal{I}-\mathcal{G}\mathcal{H})\mathbf{a}_0$. In order to further speedup convergence (guaranteed by the GMRES scheme) the system was preconditioned by an approximate inverse obtained from the channel operator by neglecting the coupling terms $\mathcal{C}$, as

$$(\mathcal{I}-\mathcal{G}\mathcal{D})^{-1}(\mathcal{I}-\mathcal{G}\mathcal{H})\mathbf{a}=(\mathcal{I}-\mathcal{G}\mathcal{D})^{-1}\mathcal{Q}\mathbf{u}\,. \tag{10}$$

Practical application of this preconditioner in the construction of the Krylov subspace (9) is possible [10] with a slight modification of the inner WR-LPTP loop (7).

The results of [10] indicate that even with a serial (non parallelized) implementation, the obtained preconditioned WR-GMRES scheme is able to obtain the solution in a runtime that is between 2–3 times larger than the corresponding WR-LPTP direct solution, but still about 10 times faster than a direct SPICE solution. This motivates our interest to extend this framework to the general nonlinear case.

III. The WR-NGMRES solver

The WR-LPTP approach (7) is able to solve high-speed channels with nonlinear terminations, but is not guaranteed to converge. Conversely, the GMRES approach to solve (8) is guaranteed to converge, but is not able to deal with nonlinearities. Hence the motivation for this work, which extends the GMRES approach by integrating it with the Newton scheme, in order to include general nonlinear terminations with guaranteed convergence.

A. The NGMRES scheme

Equations (4) and (5) can be cast as a nonlinear system

$$\mathbf{b}-\mathcal{H}\mathcal{F}(\mathbf{b})=\mathcal{N}(\mathbf{b})=\mathbf{0}. \tag{11}$$

The well known basic Newton sequence for the iterative solution of (11) is

$$\mathbf{b}_{n+1}=\mathbf{b}_n-\left[\mathcal{J}(\mathbf{b}_n)\right]^{-1}\mathcal{N}(\mathbf{b}_n), \tag{12}$$

with n iteration number and $\mathcal{J}(\mathbf{b}_n)$ the Jacobian matrix of $\mathcal{N}$ evaluated at $\mathbf{b}_n$. Equation (12) can be considered as the solution of a local linear model of $\mathcal{N}(\mathbf{b})$, providing an improved approximate solution of the nonlinear system (11). Practical application of the Newton scheme involves

1) an estimation of $\mathcal{N}(\mathbf{b}_n)$,
2) the numerical solution of linear system
$$\mathcal{J}(\mathbf{b}_n)\mathbf{s}=-\mathcal{N}(\mathbf{b}_n), \tag{13}$$
3) and the construction of $\mathbf{b}_{n+1}=\mathbf{b}_n+\lambda\mathbf{s}$, where the step length λ is selected to guarantee a decrease in the residual error $\|\mathcal{N}(\mathbf{b}_{n+1})\|<\|\mathcal{N}(\mathbf{b}_n)\|$.

The second item is clearly the most time consuming, since the evaluation of the Jacobian matrix and its inversion is prohibitive for large-scale systems as in our application. However, an approximate solution for the descent step $\mathbf{s}$ can be computed to satisfy the inequality

$$\|\mathcal{J}(\mathbf{b}_n)\mathbf{s}+\mathcal{N}(\mathbf{b}_n)\|\le\eta_n\|\mathcal{N}(\mathbf{b}_n)\| \tag{14}$$

usually denoted as the inexact Newton condition [13], through an iterative scheme that does not even require to compute, store and factor $\mathcal{J}(\mathbf{b}_n)$. The GMRES method is used here to compute the step $\mathbf{s}$ with (14) as a stop condition. The parameter η_n is determined as discussed in [12].

The starting point $\mathbf{b}_0$ for the above NGMRES scheme can be evaluated with one or two iterations of the LPTP scheme (7). This choice guarantees that $\mathbf{b}_0$ is already quite close to the true solution, so that standard results on the local convergence theory of Newton method [14] apply. To achieve global convergence, we use the linear search method denoted as the Armijo rule [15], which guarantees a good estimate of the optimal step length λ towards the solution, avoiding oscillations in the convergence process.

B. Preconditioning

According to the general theory of iterative Krylov methods [16], the convergence of the GMRES scheme strongly depends on the clustering of the system matrix eigenvalues. To improve this clustering and obtain a fast convergence, a suitable preconditioning scheme can be used.

We consider the situation in which the terminations are non-homogeneous, as in the case of drivers, whose characteristic equations depend also on the discrete time samples of internal sources $\mathbf{u}$, which can be considered as a constant vector in the solution process. In this case (5) becomes

$$\mathbf{a}=\mathcal{F}(\mathbf{b},\mathbf{u})=\underbrace{\mathcal{F}(\mathbf{b},\mathbf{u})-\mathcal{F}(\mathbf{b}_0,\mathbf{u})}_{\mathcal{G}_{nl}(\mathbf{b},\mathbf{u})}+\underbrace{\mathcal{F}(\mathbf{b}_0,\mathbf{u})}_{\mathcal{Q}_{nl}(\mathbf{u})}\,. \tag{15}$$

Considering that the channel operator can be approximated by neglecting inter-channel couplings as $\mathcal{H}\approx\mathcal{D}$ as in Sec. II-C, we can obtain an approximate solution $\hat{\mathbf{b}}$ by solving

$$\hat{\mathcal{N}}(\hat{\mathbf{b}})=\hat{\mathbf{b}}-\mathcal{D}\left[\mathcal{G}_{nl}(\hat{\mathbf{b}},\mathbf{u})+\mathcal{Q}_{nl}(\mathbf{u})\right]=0 \tag{16}$$

through few WR (fixed-point) iterations, as in the inner loop of (7). We apply this iteration, as for the linear case (10), as a means of accelerating convergence in the iterative GMRES solution of (13). This plays the role of a preconditioner and allows to significantly reduce the number of iterations in the determination of the Newton step.

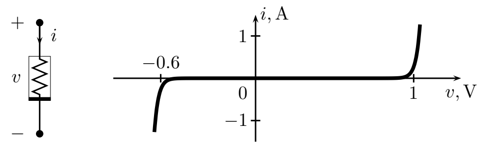

Fig. 2. Nonlinear protection circuit and its DC characteristic.

IV. Results

We apply the proposed WR-NGMRES scheme to the transient simulation of a 18-port high-speed channel, see Fig. 1, with nonlinear terminations. A victim net (port 9) is driven by a pseudorandom bit sequence through a 10Ω driver, whereas all other nets (odd-numbered ports) are driven by clock signals (1Ω internal impedance). All nets are terminated (even-numbered ports) by 1 pF capacitances. Protection circuits with DC characteristic depicted in Fig. 2 are further connected in parallel at ports 2 and 10. The strong nonlinearities induced by these elements combined by the terminations mismatch provide an excellent benchmark for the proposed solver.

The voltages at ports 2 and 10 comptued with and without the protection circuits are depicted in Fig. 3. The figure clearly shows the effect of the nonlinearities of the protection circuits, which when removed lead to a completely different solution. The waveforms obtained by SPICE demonstrate the accuracy of the proposed solution, which was obtained by the preconditioned WR-NGMRES scheme in 6 outer iterations with a relative accuracy of 10^{-6}. This performance was confirmed by many other test cases, not displayed here due to lack of space. We only remark that the preconditioning scheme applied to 20 different benchmarks was able to reduce by 50% on average the total number of iterations, with a peak reduction of 75%.

V. Conclusions

We presented an algebraic approach to the transient simulation of high-speed channels with nonlinear terminations, which combines a Waveform Relaxation framework with a Krylov-subspace iterative scheme and an inexact Newton method. The main advantage is generality, since the limitations of previous WR implementations in terms of convergence [9] or restriction to the linear case [10] are eliminated. As in previous WR approaches involving longitudinal partitioning, channel and termination operators are not solved concurrently but only forward evaluated during the iterations. This leaves the flexibility of solving each decoupled termination independently, with the preferred solver. Since all such simulations are independent, the proposed scheme is an excellent candidate for parallelization, which is under way. A future report will document detailed runtime and speedup factors of both serial and parallel implementations of the proposed WR-NGMRES scheme.

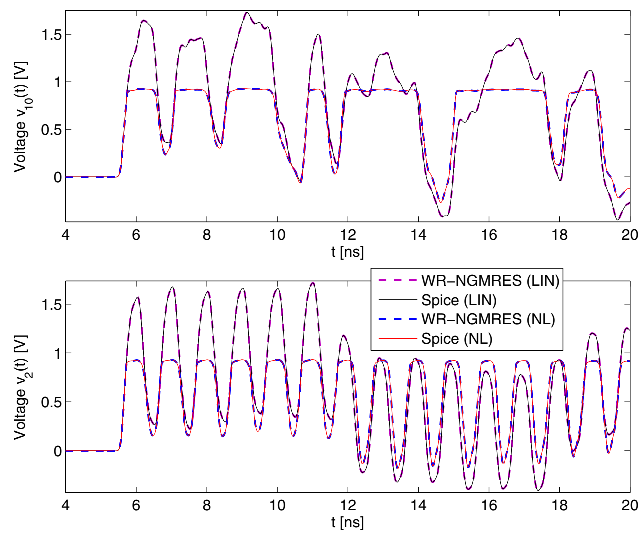

Fig. 3. Received voltage on victim (top) and aggressor (bottom) channels. SPICE and WR-NGMRES results are compared with (NL) and without (LIN) nonlinear protection circuits.

References

[1] G.Balamurugan *et al.*, “Modeling and Analysis of High-Speed I/O Links,” *IEEE Trans. on Advanced Packaging*, vol. 32, no. 2, pp. 237–247, May 2009.

[2] A. Chinea *et al.*, “Signal Integrity Verification of Multi-Chip Links using Passive Channel Macromodels,” *IEEE Trans. on Components, Packaging and Manufacturing Technology*, vol. 1, pp. 920–933, June 2011.

[3] A. Chinea, P. Triverio, S. Grivet-Talocia, “Delay-Based Macromodeling of Long Interconnects from Frequency-Domain Terminal Responses,” *IEEE Transactions on Advanced Packaging*, Vol. 33, No. 1, pp. 246–256, Feb. 2010.

[4] E.Lelarasmee, A.E.Ruehli, A.L.Sangiovanni-Vincentelli, “The Waveform Relaxation Method for Time-Domain Analysis of Large Scale Integrated Circuits”, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 1, N. 3, July 1982, pp. 131–145.

[5] N.M.Nakhla, A.E.Ruehli, M.S.Nakhla, R.Achar, “Simulation of coupled interconnects using waveform relaxation and transverse partitioning,” *IEEE Trans. on Advanced Packaging*, vol. 29, no. 1, pp. 78–87, 2006

[6] N.Nakhla, A.E.Ruehli, M.S.Nakhla, R.Achar, C.Chen, “Waveform Relaxation Techniques for Simulation of Coupled Interconnects With Frequency-Dependent Parameters”, *IEEE Transactions on Advanced Packaging*, Vol.30, N.2, 2007, pp. 257–269.

[7] V.Loggia, S.Grivet-Talocia, H.Hu, “Transient simulation of complex high-speed channels via Waveform Relaxation”, *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 1, pp. 1823–1838, November 2011.

[8] W.T.Beyene, “Applications of Multilinear and Waveform Relaxation Methods for Efficient Simulation of Interconnect-Dominated Nonlinear Networks,” *IEEE Trans. on Advanced Packaging*, vol. 31, no. 3, pp. 637–648, Aug. 2008.

[9] S. Grivet-Talocia, and V. Loggia, “Fast Channel Simulation via Waveform Over-Relaxation,” in *15th IEEE Workshop on Signal Propagation on Interconnects, Napoli, Italy*, May 8–11, 2011.

[10] H. Hu, S. Grivet-Talocia, “A Preconditioned Waveform Relaxation Solver for Signal Integrity Analysis of High-Speed Channels,” *EMC Europe 2012*, Rome, 17–21 Sept. 2012.

[11] Y.Saad, M.H.Schultz, “GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems,” *SIAM J. Sci. Stat. Comput.*, Vol. 7, No. 3, July 1986

[12] C.T. Kelley, “Iterative Methods for Linear and Nonlinear Equations,” *SIAM J. Sci. Stat. Comput.*, Vol. 16 in Frontiers in Applied Mathematics.

[13] R. Dembo, S. Eisenstat, and T. Steihaug, “Inexact Newton methods,” *SIAM J. Numer. Anal.*, 19 (1982), pp. 400–408.

[14] J.M. Oretga, W.C.Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables,* Accademic Press, New York, 1970.

[15] L. Armijo, “ Minimization of functions having Lipschitz-continuous first partial derivatives,” *Pacific J. Math.*, 16 (1966), pp 1–3.

[16] L.N. Trefethen, D. Bau III, *Numerical linear algebra,* SIAM, Philadelphia, 1997.