Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Shifted linear systems in electromagnetics. Part I: Systems with identical right-hand sides.

Rainer Schlundt[1],

Franz-Josef Schmückle[2], Wolfgang Heinrich[2]

submitted: 8th June 2009

[1] Weierstrass Institute for Applied
Analysis and Stochastics
Mohrenstraße 39
10117 Berlin, Germany
E-Mail: schlundt@wias-berlin.de

[2] Ferdinand-Braun-Institut
für Höchstfrequenztechnik
Gustav-Kirchhoff-Str. 4
12489 Berlin, Germany
E-Mail: fj.Schmueckle@fbh-berlin.de
w.heinrich@ieee.org

No. 1420
Berlin 2009

2000 *Mathematics Subject Classification.* 35Q60, 65F10, 65F15, 65N22, 78M25.

Key words and phrases. Microwave device, Maxwell's equations, Scattering matrix, Boundary value problem, PML boundary condition, Eigenvalue problem, Linear algebraic equations, Multiple shifts, Krylov subspace method, Polynomial preconditioning.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We consider the solution of multiply shifted linear systems for a single right-hand side. The coefficient matrix is symmetric, complex, and indefinite. The matrix is shifted by different multiples of the identity. Such problems arise in a number of applications, including the electromagnetic simulation in the development of microwave and mm-wave circuits and modules.
The properties of microwave circuits can be described in terms of their scattering matrix which is extracted from the orthogonal decomposition of the electric field. We discretize the Maxwell's equations with orthogonal grids using the Finite Integration Technique (FIT).
Some Krylov subspace methods have been used to solve multiply shifted systems for about the cost of solving just one system. We use the QMR method based on coupled two-term recurrences with polynomial preconditioning.

Contents

List of Figures

1 Introduction

Today, electromagnetic simulation forms an indispensable tool in the development of microwave circuits. The description of the boundary of the computational domain has always been a key issue in bringing up efficiency of electromagnetic simulation. The Perfectly Matched Layer (PML) concept provides an excellent solution to this issue. However, the benefits of PML do not come for free. In the frequency-domain case, the material tensors worsen the numerical properties of the system of equations to be solved, which results in increased CPU time [20].
The subject under investigation are three-dimensional structures of arbitrary geometry which are connected to the remaining circuit by transmission lines. Ports are defined at the outer terminations of the transmission lines (see Fig. 1).

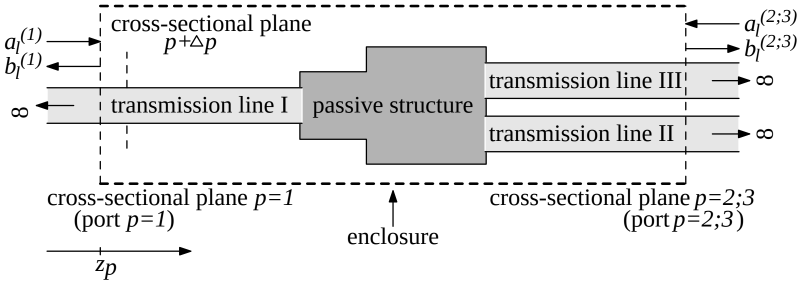

Figure 1: The basic structure under investigation

Calculating the excitations at the ports, one obtains eigenvalue problems and then large-scale systems of linear algebraic equations. In general, the computation of the eigenvalue problem and of the system of linear algebraic equations have to be done for several frequencies. Moreover, these linear equation problems have to be solved repeatedly for different right-hand sides. The number of right-hand sides depends on the number of ports and modes.

2 Scattering Matrix

The scattering matrix describes the structure in terms of the wave modes on the transmission line sections at the ports. We consider all exciting modes with amplitudes a_l towards the discontinuity and all amplitudes b_l outwards from the discontinuity (see Fig. 1). As example for the waves at the left port of Fig. 1 the transverse mode field at a cross-sectional plane z is given by

$$\vec{E}_t(z) = \sum_{l=1}^{m^{(p)}} a_l \vec{E}_{t,l} e^{-\jmath k_{z_l} z} + \sum_{l=1}^{m^{(p)}} b_l \vec{E}_{t,l} e^{+\jmath k_{z_l} z} = \sum_{l=1}^{m^{(p)}} w_l(z) \vec{E}_{t,l} \tag{1}$$

with

$$w_l(z) = a_l e^{-\jmath k_{z_l} z} + b_l e^{+\jmath k_{z_l} z} = \tilde{a}_l(z) + \tilde{b}_l(z), \tag{2}$$

where k_{z_l} is the propagation constant. We consider the application of (1) with (2) at a pair of neighboring cross-sectional planes z_p and $z_{p+\Delta p}$. That means, we have to solve m_s boundary value problems (see [9, 10]) with the boundary conditions

$$\vec{E}_{t,\nu} = \sum_{\rho=1}^{m_s} \bar{w}_{\rho,\nu} \vec{E}_{t,l}(z_p), \quad \rho = l + \sum_{q=1}^{p-1} m^{(q)}, \quad p = 1(1)\overline{p}, \quad \nu = 1(1)m_s, \tag{3}$$

in order to compute $w_m^{(p+\Delta p)}$ where the weighted amplitude sums $w_m^{(p)}$ are given. $m^{(p)}$ denotes the number of modes which have to be taken into account at the port p. $\overline{p}$ is the number of ports. The modes on a port p are numbered with l, $l = 1(1)m^{(p)}$. That means, the dimension m_s of this matrix is determined by the total number of modes at all ports. We get $\vec{E}_{t,l}(z_p)$ solving eigenvalue problems for the transmission lines.

The scattering matrix S (see [10]) is defined by

$$\vec{\bar{b}}_\nu = S \vec{\bar{a}}_\nu, \quad \nu = 1(1)m_s \ , \tag{4}$$

or

$$\bar{b}_{\rho,\nu} = \sum_{\sigma=1}^{m_s} S_{\rho,\sigma} \cdot \bar{a}_{\sigma,\nu}, \quad \rho, \nu = 1(1)m_s \ . \tag{5}$$

3 Boundary Value Problem

A three-dimensional boundary value problem can be formulated using the integral form of Maxwell's equations in the frequency domain [1] in order to compute the electromagnetic field:

$$\begin{array}{rclcrcl} \oint_{\partial\Omega} \vec{H} \cdot d\vec{s} & = & \jmath\omega \int_{\Omega} [\epsilon] \vec{E} \cdot d\vec{\Omega} & , & \oint_{\Omega} [\epsilon] \vec{E} \cdot d\vec{\Omega} & = & 0 \ , \\ \oint_{\partial\Omega} \vec{E} \cdot d\vec{s} & = & -\jmath\omega \int_{\Omega} [\mu] \vec{H} \cdot d\vec{\Omega} & , & \oint_{\Omega} [\mu] \vec{H} \cdot d\vec{\Omega} & = & 0 \ , \end{array} \tag{6}$$

$$\vec{D} = [\epsilon]\vec{E}, \quad \vec{B} = [\mu]\vec{H}. \tag{7}$$

The electric and magnetic flux densities $\vec{D}$ and $\vec{B}$ are complex functions of the spatial coordinates. $\omega = 2\pi f$ is the angular frequency of the sinusoidal excitation. f denotes the frequency.

At the ports p the transverse electric field $\vec{E}_t(z_p)$ is given by superposing weighted transmission line modes $\vec{E}_{t,l}(z_p)$ (see (1)):

$$\vec{E}_t(z_p) = \sum_{l=1}^{m^{(p)}} w_l(z_p)\vec{E}_{t,l}(z_p)\,. \tag{8}$$

All other parts of the surface of the computation domain are assumed to be an electric or a magnetic wall:

$$\vec{E} \times \vec{n} = 0\,, \quad \vec{H} \times \vec{n} = 0\,. \tag{9}$$

We introduce a complex permittivity $[\epsilon]$ and a complex permeability $[\mu]$ diagonal tensor to obtain a reflection-free interface between the computational area and the lossy PML region:

$$[\epsilon] = (\epsilon)[\Lambda^{(\epsilon)}]\,, \quad [\mu] = (\mu)[\Lambda^{(\mu)}] \tag{10}$$

with

$$(\epsilon) = \mathrm{diag}(\epsilon_x, \epsilon_y, \epsilon_z)\,, \quad (\mu) = \mathrm{diag}(\mu_x, \mu_y, \mu_z)\,. \tag{11}$$

$[\Lambda^{(\epsilon)}]$ and $[\Lambda^{(\mu)}]$ are defined for a PML in x-, y-, or z-direction as follows ($\nu \in \{\epsilon, \mu\}$):

$$[\Lambda^{(\nu)}] = \left\{ \begin{array}{l} [\Lambda^{(\nu)}]_x = \mathrm{diag}(\frac{1}{\lambda_\nu}, \lambda_\nu, \lambda_\nu) \\ [\Lambda^{(\nu)}]_y = \mathrm{diag}(\lambda_\nu, \frac{1}{\lambda_\nu}, \lambda_\nu) \\ [\Lambda^{(\nu)}]_z = \mathrm{diag}(\lambda_\nu, \lambda_\nu, \frac{1}{\lambda_\nu}) \end{array} \right\} \quad \text{with} \tag{12}$$

$$\lambda_\nu = 1 - \jmath\frac{\kappa_\nu}{\nu_0\omega} \quad \text{and} \quad \frac{\kappa_\epsilon}{\epsilon_0} = \frac{\kappa_\mu}{\mu_0}\,. \tag{13}$$

In case of overlapping at edges and corners the resulting PML tensor is the product of the PML tensors of the individual PML walls that form the edges and corners, respectively.

4 Maxwellian Grid Equations

Maxwellian grid equations are formulated for staggered nonequidistant rectangular grids (see Fig. 2) using the Finite Integration Technique with lowest order integration formulae [1, 14, 23]:

$$\oint_{\partial\Omega} \vec{f} \cdot d\vec{s} \rightarrow \sum(\pm f_i s_i)\,, \quad \int_{\Omega} \vec{f} \cdot d\vec{\Omega} \rightarrow f\Omega\,, \quad \oint_{\Omega} \vec{f} \cdot d\vec{\Omega} \rightarrow \sum(\pm f_i \Omega_i)\,. \tag{14}$$

The discretized form of (6) results in an equation for each field component. Pre-

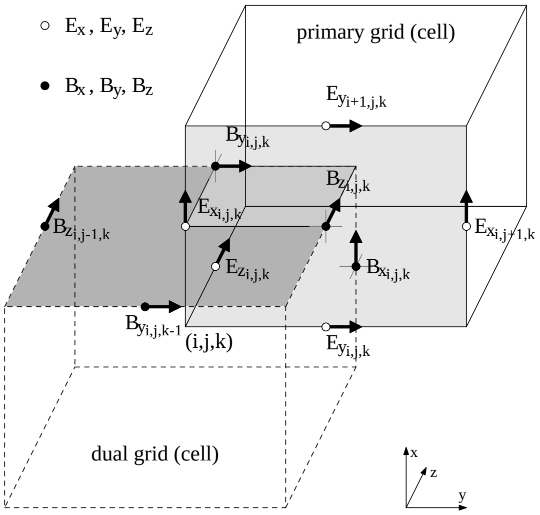

Figure 2: Primary and dual grid

senting each equation using matrices provides a compact form:

$$\begin{array}{rclcrcl} \tilde{C} D_{\tilde{s}/\tilde{\mu}} \vec{b} & = & \jmath\omega\epsilon_0\mu_0 D_{\tilde{A}\tilde{\epsilon}} \vec{e} & , & \tilde{S} D_{\tilde{A}\tilde{\epsilon}} \vec{e} & = & 0\,, \\ C D_s \vec{e} & = & -\jmath\omega D_A \vec{b} & , & S D_A \vec{b} & = & 0\,. \end{array} \tag{15}$$

The diagonal matrices $D_{\tilde{s}/\tilde{\mu}}$, $D_{\tilde{A}\tilde{\epsilon}}$, D_s, and D_A represent all cell quantities. The so-called curl $(C, \tilde{C})$ and source matrices $(S, \tilde{S})$ describe the topology of the two grids with the following properties (see [24]):

$$SC = 0\,, \quad \tilde{S}\tilde{C} = 0\,, \quad C = \tilde{C}^T\,. \tag{16}$$

4.1 System of Linear Algebraic Equations

Using (16), eliminating the components of the magnetic flux density $(\vec{b})$ in (15), and multiplying by $D_s^{1/2}$ yields a symmetric form of linear algebraic equations:

$$(D_s^{1/2} C^T D_{\tilde{s}/\tilde{\mu}} D_A^{-1} C D_s^{1/2} - k_0^2 D_{\tilde{A}\tilde{\epsilon}}) D_s^{1/2} \vec{e} = 0\,, \tag{17}$$

where $k_0 = \omega\sqrt{\epsilon_0\mu_0}$ denotes the wavenumber in vacuum. Moreover, the gradient of the electric field divergence

$$[\epsilon]\nabla([\epsilon]^{-2}\nabla \cdot [\epsilon]\vec{E}) = 0 \tag{18}$$

is equivalent to the matrix equation

$$(D_s^{-1/2} D_{\tilde{A}\tilde{\epsilon}} \tilde{S}^T D_{\tilde{V}\tilde{\epsilon}\tilde{\epsilon}}^{-1} \tilde{S} D_{\tilde{A}\tilde{\epsilon}} D_s^{-1/2}) D_s^{1/2} \vec{e} = 0\,. \tag{19}$$

The diagonal matrix $D_{\tilde{V}\tilde{\epsilon}\tilde{\epsilon}}$ is a volume matrix for the 8 partial volumes of the dual elementary cell.

The addition of Eqs. (17) and (19) yields the form

$$(A^{(G)} - k_0^2 D^{(G)})x^{(G)} = 0\,. \tag{20}$$

Taking into account the constitutive relations (7), the boundary conditions (9), and the transmission line modes $\vec{E}_{t,l}(z_p)$ (see (1)) we transform Eq. (20) into an inhomogeneous linear system of equations where its right-hand side depends on (8). For it, we use the notations given in Table 1. Thus, we get from Eq. (20):

Table 1: Notations

$x^{(G)} = (x_E^{(G)}, x_I^{(G)})^T$, vector of the unknown electric flux densities			
$x_E^{(G)} = (x_{2d}^{(G)}, x_{3d}^{(G)})^T$, given components of the solution vector $x^{(G)}$			
$x_E^{(G)}$	external points, $\dim(x_E^{(G)}) = n_E$	$x_{2d}^{(G)}$	solution of the $2d$ eigenvalue problem
$x_I^{(G)}$	internal points, $\dim(x_I^{(G)}) = n_I, n = n_E + n_I$	$x_{3d}^{(G)}$	given boundary points of the $3d$ problem
$A^{(G)}$	$= (A_E^{(G)}, 0) + (0, A_I^{(G)})$	$\dim(A^{(G)}) = (n_I, n)\,,$ $\dim(A_E^{(G)}) = (n_I, n_E)\,,$ $\dim(A_I^{(G)}) = (n_I, n_I)$	
$D^{(G)}$	$= (0, D_I^{(G)})$	$\dim(D^{(G)}) = (n_I, n)\,,$ $\dim(D_I^{(G)}) = (n_I, n_I)$	
I_E	identity	$\dim(I_E) = (n_E, n_E)$	

$$\begin{pmatrix} 0 \\ A^{(G)} \end{pmatrix} \begin{pmatrix} x_E^{(G)} \\ x_I^{(G)} \end{pmatrix} = \begin{pmatrix} I_E - I_E & 0 \\ A_E^{(G)} & A_I^{(G)} - k_0^2 D_I^{(G)} \end{pmatrix} \begin{pmatrix} x_E^{(G)} \\ x_I^{(G)} \end{pmatrix} = 0\,,$$

$$\begin{pmatrix} I_E & 0 \\ 0 & A_I^{(G)} - k_0^2 D_I^{(G)} \end{pmatrix} \begin{pmatrix} x_E^{(G)} \\ x_I^{(G)} \end{pmatrix} + \begin{pmatrix} -I_E & 0 \\ A_E^{(G)} & 0 \end{pmatrix} \begin{pmatrix} x_E^{(G)} \\ x_I^{(G)} \end{pmatrix} = 0\,,$$

$$\begin{pmatrix} I_E & 0 \\ 0 & A_I^{(G)} - k_0^2 D_I^{(G)} \end{pmatrix} \begin{pmatrix} x_E^{(G)} \\ x_I^{(G)} \end{pmatrix} = \begin{pmatrix} x_E^{(G)} \\ -A_E^{(G)} x_E^{(G)} \end{pmatrix} = \begin{pmatrix} x_E^{(G)} \\ b_I^{(G)} \end{pmatrix} = b^{(G)}\,, \tag{21}$$

$$(A_I^{(G)} - k_0^2 D_I^{(G)})x_I^{(G)} = b_I^{(G)}\,. \tag{22}$$

Therefore, the systems of linear algebraic equations (21) and (22) are to be solved, respectively.

4.2 Eigenvalue Problem

The vector $x_{2d}^{(G)}$ is the solution of the $2d$ eigenvalue problem. In the following, we consider a longitudinally homogeneous transmission line. Thus, any field can be expanded into a sum of so-called modal fields which vary exponentially in the logitudinal direction:

$$\vec{E}(x, y, z \pm 2h) = \vec{E}(x, y, z)e^{\mp \jmath k_z 2h} \ , \tag{23}$$

where k_z is the propagation constant, and $2h$ is the length of an elementary cell in z-direction (see Fig. 3). Thus, we get a two-dimensional eigenvalue problem for the transverse electric fields $\vec{y} = \vec{E}_{t,l}(z_p)$, $l = 1(1)m^{(p)}$, (see (8)) on the transmission line region:

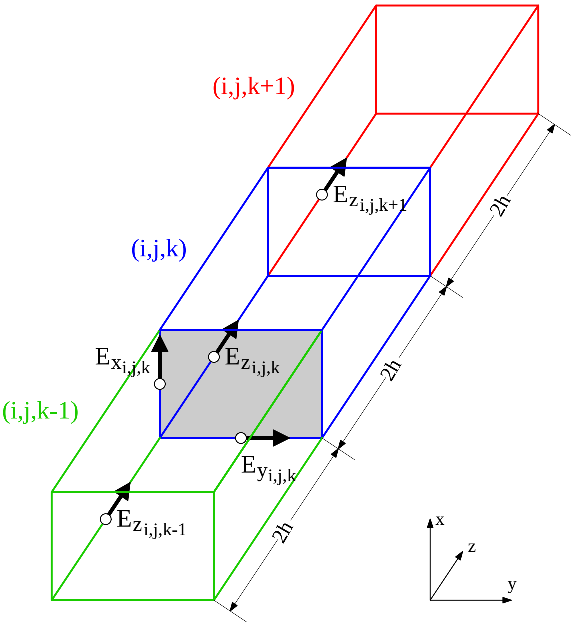

Figure 3: Transmission line

$$A\vec{y} = \gamma\vec{y}, \ \ \gamma = e^{-\jmath k_z 2h} + e^{+\jmath k_z 2h} - 2 = -4\sin^2(hk_z). \tag{24}$$

A detailed derivation of the eigenvalue problem can be found in [11, 12, 13, 15].

5 QMR Algorithm for Shifted Matrices

We consider the iterative solution of large systems of linear algebraic equations which not only have multiple right-hand sides, but also have multiple shifts for each right-hand side. The generalized form of Eq. (22) is the problem

$$(\alpha_j A + \beta_j D)x^{(j,k)} = b^{(k)} \ , \quad \alpha_j, \beta_j \in \mathbb{C} \ , \tag{25}$$

with $j = 1, \ldots, n_s$ and $k = 1, \ldots, n_b$. Let n_b be the number of right-hand sides and n_s be the number of shifts. The Matrix A is complex, symmetric, and indefinite. A standard way to solve systems with multiple right-hand sides is to use a block approach [19, 6].

Eq. (25) can be transformed into

$$(D^{-1/2}AD^{-1/2} + \alpha_j^{-1}\beta_j I)(\alpha_j D^{1/2}x^{(j,k)}) = D^{-1/2}b^{(k)}\ . \tag{26}$$

Thus, we get then the common equation

$$A^{(j)}x^{(j,k)} = (A + \sigma_j I)x^{(j,k)} = b^{(k)}\ , \quad \sigma_j \in \mathbb{C}\,. \tag{27}$$

For the special case that all right-hand sides in (27) are identical, i.e.,

$$x^{(j)} = x^{(j,k)}\ ,\ b^{(k)} = b \quad \text{for all} \quad k = 1, \ldots, n_b\,, \tag{28}$$

it is straightforward to exploit the shift structure when solving the n_s systems by Krylov subspace methods. We use the initial guess $x_0^{(j)} = 0$ for all j. In this case, the Krylov subspaces for all n_s systems are identical:

$$\mathcal{K}_m(A + \sigma_j I, b) = \mathcal{K}_m(A, b) \quad \text{for all} \quad j = 1, \ldots, n_s \quad \text{and} \quad m \geq 1\,. \tag{29}$$

This means that the computation of suitable basis vectors for the underlying Krylov subspaces has to be performed only once.

A lot of Krylov subspace methods have been developed for shifted matrix problems. We present a shifted coupled two-term algorithm without look-ahead for

$$A^{(j)}x^{(j)} = (A + \sigma_j I)x^{(j)} = b\ , \quad \sigma_j \in \mathbb{C}\ , \quad j = 1, \ldots, n_s\,. \tag{30}$$

Unfortunately, standard preconditioning techniques with a preconditioner

$$M = M_1M_2 = (M_1M_2)^T = M^T, \tag{31}$$

such as SSOR preconditioning, destroy the special structure when they are applied to shifted linear systems. The only technique we are aware of that allows to preserve the shifted structure is polynomial preconditioning (see [5]).

Using standard preconditioning techniques, we apply the coupled two-term QMR algorithm to the shifted linear systems [5, 7]

$$\tilde{A}^{(j)}\tilde{x}^{(j)} = \tilde{b}\ , \quad j = 1, \ldots, n_s\ , \tag{32}$$

with

$$\begin{gathered} \tilde{A}^{(j)} = M_1^{-1}(A + \sigma_j M_1M_2)M_2^{-1} = M_1^{-1}AM_2^{-1} + \sigma_j I\ , \\ \tilde{b} = M_1^{-1}b \quad \text{and} \quad \tilde{x}^{(j)} = M_2x^{(j)}\,. \end{gathered} \tag{33}$$

It is easy to see that the linear systems (30) and (32) with (33) are not equivalent. Only, for $M = I$ the systems (30) and (32) are equivalent.

It is possible to write the resulting algorithm in terms of quantities corresponding to the system (30). This is what we have done below.
We have the following analogies:

$$\begin{array}{llllll} v_n & \longrightarrow & \tilde{v}_n = M_1^{-1} v_n & , & w_n & \longrightarrow & \tilde{w}_n = M_2^{-T} w_n \,, \\ p_n & \longrightarrow & \tilde{p}_n = M_2 p_n & , & q_n & \longrightarrow & \tilde{q}_n = M_1^T q_n \,, \end{array}$$

$$\tilde{w}_n^T \tilde{v}_n = w_n^T M_2^{-1} M_1^{-1} v_n = w_n^T M^{-1} v_n \,,$$
$$\tilde{q}_n^T \tilde{p}_n = q_n^T M_1 M_2 p_n = q_n^T M p_n \,,$$
$$\tilde{q}_n^T \tilde{A} \tilde{p}_n = q_n^T M_1 M_1^{-1} A M_2^{-1} M_2 p_n = q_n^T A p_n \,.$$

The resulting coupled two-term QMR algorithm is as follows.

0. **Input**: A, $\{\sigma_1, \ldots, \sigma_{n_s}\}$, b.
   For $j = 1, \ldots, n_s$, set $x_0^{(j)} = 0$ and $r_0 = b$.
   Compute $\rho_1 = \|M_1^{-1} r_0\|$ and set $v_1 = r_0/\rho_1$.
   For $j = 1, \ldots, n_s$, set $p_0^{(j)} = d_0^{(j)} = 0$, $c_0^{(j)} = \epsilon_0^{(j)} = 1$, $\vartheta_0^{(j)} = 0$, $\eta_0^{(j)} = -1$.
   Choose $j_1 \in \{1, \ldots, n_s\}$: $\tilde{A}^{(j_1)} \tilde{x}^{(j_1)} = \tilde{b}$ is the seed system.

For $n = 1, 2, \ldots,$ **do**:

1. Compute $\delta_n = v_n^T M^{-1} v_n$.
   If $\delta_n = 0$, then stop.

2. **For all** $j = 1, \ldots, n_s$ for which $x_n^{(j)}$ has not converged yet:

   - If $\epsilon_{n-1}^{(j)} = 0$, then stop.
   - Compute
     $$p_n^{(j)} = M^{-1} v_n - p_{n-1}^{(j)} (\rho_n \delta_n / \epsilon_{n-1}^{(j)}) \,.$$
   - Compute
     $$\begin{aligned} \epsilon_n^{(j)} &= p_n^{(j)T} A^{(j)} p_n^{(j)} = p_n^{(j)T} (A + \sigma_j M) p_n^{(j)} \,, \\ \beta_n^{(j)} &= \epsilon_n^{(j)} / \delta_n \,, \\ \hat{v}_{n+1} &= A^{(j_1)} p_n^{(j_1)} - v_n \beta_n^{(j_1)} = (A + \sigma_{j_1} M) p_n^{(j_1)} - v_n \beta_n^{(j_1)} \,, \\ \rho_{n+1} &= \|M_1^{-1} \hat{v}_{n+1}\| \,. \end{aligned}$$
   - Compute
     $$\begin{aligned} \vartheta_n^{(j)} &= \frac{\rho_{n+1}}{c_{n-1}^{(j)} |\beta_n^{(j)}|} \,, \\ c_n^{(j)} &= \frac{1}{\sqrt{1 + \vartheta_n^{(j)2}}} \,, \\ \eta_n^{(j)} &= -\eta_{n-1}^{(j)} \frac{\rho_n c_n^{(j)2}}{\beta_n^{(j)} c_{n-1}^{(j)\,2}} \,, \\ d_n^{(j)} &= p_n^{(j)} \eta_n^{(j)} + d_{n-1}^{(j)} \vartheta_{n-1}^{(j)\,2} c_n^{(j)2} \,, \\ x_n^{(j)} &= x_{n-1}^{(j)} + d_n^{(j)} \,. \end{aligned}$$

```
        • If $\rho_{n+1} = 0$, then stop.
          Otherwise, set
                    $v_{n+1} = \hat{v}_{n+1}/\rho_{n+1}$.
    End for $(j)$.
  3. If all $x_n^{(j)}$ have converged, then stop.
End for $(n)$.
```

5.1 Implementation Details

In this section, we present a detailed description of the implementation of the coupled two-term Lanczos algorithm for shifted linear systems [7].
First, we consider the non-shifted system $Ax = b$. The construction for the basis vectors p_k and v_k can be written compactly in matrix form:

$$V_n = P_n U_n \ , \ AP_n = V_{n+1} L_n \, . \tag{34}$$

Here, U_n is an upper triangular matrix and L_n is an upper Hessenberg matrix given by

$$U_n = \begin{pmatrix} 1 & u_{12} & \dots & u_{1n} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & u_{n-1n} \\ 0 & \dots & 0 & 1 \end{pmatrix} \quad \text{and} \quad L_n = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ \rho_2 & l_{22} & \ddots & \vdots \\ 0 & \rho_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & l_{nn} \\ 0 & \dots & \dots & \rho_{n+1} \end{pmatrix} .$$

Furthermore, it follows that

$$\begin{array}{llllll} l_{in} = 0 & , & i = 1, 2, \dots, n-1 & , & l_{nn} = \beta_n = \epsilon_n/\delta_n \, , \\ u_{in} = 0 & , & i = 1, 2, \dots, n-2 & , & u_{n-1n} = \rho_n \delta_n/\epsilon_{n-1} \, . \end{array}$$

Therefore, the matrix U_n is upper bidiagonal and L_n is a lower bidiagonal matrix.

In the QMR method, the vectors v_k and p_k generated by the Lanczos algorithm are used as a basis for the Krylov subspace $\mathcal{K}_n(A, r_0)$. The nth QMR iterate is then defined by

$$x_n = x_0 + V_n z_n \, . \tag{35}$$

Setting $y_n = U_n z_n$, we can rewrite x_n as follows:

$$x_n = x_0 + V_n U_n^{-1} y_n \, , \tag{36}$$

where y_n is the unique solution of the least squares problem

$$y_n = \arg\min_{y \in \mathbb{C}^n} \| \, \|r_0\| e_1 - L_n y \, \| \, . \tag{37}$$

We now return to shifted linear systems of the form (32).

The construction of the basis vectors $p_k^{(j)}$ and v_k for the Krylov subspace can be written as follows:

$$V_n = P_n^{(j)} U_n^{(j)} \,, \quad A^{(j)} P_n^{(j)} = (A + \sigma_j M) P_n^{(j)} = V_{n+1} L_n^{(j)} \,. \tag{38}$$

$U_n^{(j)}$ is an upper bidiagonal matrix with $u_{n-1n}^{(j)} = \rho_n \delta_n / \epsilon_{n-1}^{(j)}$ and $L_n^{(j)}$ is a lower bidiagonal matrix with $l_{nn}^{(j)} = \beta_n^{(j)} = \epsilon_n^{(j)} / \delta_n$.
We set $x_0^{(j)} = 0$, $j = 1, \ldots, n_s$. Then, the nth QMR iterates are defined by

$$x_n^{(j)} = V_n z_n^{(j)} = V_n {U_n^{(j)}}^{-1} y_n^{(j)} \,, \tag{39}$$

where $y_n^{(j)}$ is unique solution of the jth least squares problem

$$y_n^{(j)} = \arg \min_{y \in \mathbb{C}^n} \| \, \|r_0\| e_1 - L_n^{(j)} y \, \| \,. \tag{40}$$

Using

$$A^{(j)} = A + \sigma_j M = A + \sigma_{j_1} M + (\sigma_j - \sigma_{j_1}) M = A^{(j_1)} + (\sigma_j - \sigma_{j_1}) M$$

the term $\epsilon^{(j)}$ can be expressed as $\epsilon^{(j_1)}$:

$$\begin{aligned} \epsilon_n^{(j)} &= {p_n^{(j)}}^T A^{(j)} p_n^{(j)} \\ &= {p_n^{(j)}}^T (A^{(j_1)} + (\sigma_j - \sigma_{j_1}) M) p_n^{(j)} \\ &= {p_n^{(j)}}^T A^{(j_1)} p_n^{(j)} + (\sigma_j - \sigma_{j_1}) {p_n^{(j)}}^T M p_n^{(j)} \,. \end{aligned} \tag{41}$$

The resulting algorithm is as follows.

```
0. For $j = 1, \ldots, n_s$, set $\mu_0^{(j)} = 0$ and $\gamma_0^{(j)} = 0$.

For $n = 1, 2, \ldots,$ do:

  1. For $j = 1, \ldots, n_s$, do:

       • $\mu_n^{(j)} = \mu_{n-1}^{(j)} (\rho_n \delta_n / \epsilon_{n-1}^{(j)})^2 + \epsilon_{n-1}^{(j)} (\rho_n \delta_n (1/\epsilon_{n-1}^{(j_1)} - 1/\epsilon_{n-1}^{(j)}))^2$
       • $\gamma_n^{(j)} = \delta_n + \gamma_{n-1}^{(j)} (\rho_n \delta_n / \epsilon_{n-1}^{(j)})^2$
       • $\epsilon_n^{(j)} = \epsilon_n^{(j_1)} + \mu_n^{(j)} + (\sigma_j - \sigma_{j_1}) \gamma_n^{(j)}$

     End for $(j)$.

End for $(n)$.
```

The vector $p_n^{(j)}$ can be expressed recursively as

$$\begin{array}{rclcl} p_n^{(j)} & = & M^{-1}v_n & - & p_{n-1}^{(j)}(\rho_n\delta_n/\epsilon_{n-1}^{(j)}) \\ & = & M^{-1}v_n & - & p_{n-1}^{(j_1)}(\rho_n\delta_n/\epsilon_{n-1}^{(j_1)})+ \\ & & & & p_{n-1}^{(j_1)}(\rho_n\delta_n/\epsilon_{n-1}^{(j_1)}) - p_{n-1}^{(j)}(\rho_n\delta_n/\epsilon_{n-1}^{(j)}) \\ & = & p_n^{(j_1)} & + & p_{n-1}^{(j_1)}(\rho_n\delta_n(1/\epsilon_{n-1}^{(j_1)} - 1/\epsilon_{n-1}^{(j)}))+ \\ & & & & (p_{n-1}^{(j_1)} - p_{n-1}^{(j)})(\rho_n\delta_n/\epsilon_{n-1}^{(j)}) \end{array}$$

and $p_0^{(j_1)} = p_0^{(j)} = 0$.
Also, the scalar product $p_n^{(j)T} M p_n^{(j)}$ can be expressed recursively as

$$\begin{array}{rclcl} p_n^{(j)T} M p_n^{(j)} & = & v_n^T M v_n & + & p_{n-1}^{(j)}{}^T M p_{n-1}^{(j)}(\rho_n\delta_n/\epsilon_{n-1}^{(j)})^2 \\ & = & \delta_n & + & p_{n-1}^{(j)}{}^T M p_{n-1}^{(j)}(\rho_n\delta_n/\epsilon_{n-1}^{(j)})^2 \end{array}$$

and $p_0^{(j)} = 0$.

6 Polynomial Preconditioning

We consider the shifted linear system

$$A^\sigma x = (A + \sigma I)x = b\,. \tag{42}$$

We use polynomial preconditioning to speed up the convergence of the iterative methods for the solution of (42), i.e.,

$$s^\sigma(A^\sigma)A^\sigma x = s^\sigma(A^\sigma)b \tag{43}$$

for left preconditioning and

$$A^\sigma s^\sigma(A^\sigma)y = s^\sigma(A^\sigma)A^\sigma y = b\,, \quad x = s^\sigma(A^\sigma)y\,, \tag{44}$$

for right preconditioning, respectively. Here, s^σ is a suitable chosen polynomial of a small degree. Both linear systems (43) and (44) are equivalent. We seek a polynomial s^σ with the following two properties [4, 3]:

- The coefficient matrix $s^\sigma(A^\sigma)A^\sigma$ is again a shifted matrix.
- The convergence of the iterative method, applied to the preconditioned system, is speed up optimally.

First, for any polynomial, we can represent $A^\sigma s^\sigma(A^\sigma)$ in the form

$$A^\sigma s^\sigma(A^\sigma) = (A + \sigma I)s^\sigma(A + \sigma I) = As(A) + \tau I \tag{45}$$

with $\tau \in \mathbb{C}$. Note that s^σ, s, and τ are related by

$$(z+\sigma)s^\sigma(z+\sigma) = zs(z) + \tau \quad \text{and} \quad \tau = \sigma s(-\sigma)\,. \tag{46}$$

We note that the coefficient matrix $As(A)$ of the preconditioned system (45) is Hermitian if, and only if, s is a real polynomial. In order to guarantee that $As(A)$ is nonsingular, we require that $s(z) \neq 0$ for all $z \in S$ with

$$\varrho(A) \subseteq S = [a,b] \cup [c,d]\,, \quad c < d < 0 < a < b\,,$$

where $\varrho(A)$ is the spectrum of A.

Next, we turn to the question of optimal choice of polynomial s. We have two different cases:

- $zs(z) > 0 \quad \forall z \in S$
- $zs(z) > 0 \quad \forall z \in [a,b] \quad \text{and} \quad zs(z) < 0 \quad \forall z \in [c,d]\,.$

If the last case holds, then the preconditioned system remains indefinite. We can now state the main result in the following form [4]:

> Let $S = [a,b] \cup [c,d]$ be the union of a positive and negative interval with $c < d < 0 < a < b$ and $\Gamma = \{(\gamma,\delta) \in \mathbb{R} \times \mathbb{R} \,:\, \delta > 0\}$ a parameter set. The optimal polynomial $s^*(z)$ of
>
> $$w(\gamma,\delta) = \min_s \|f - zs\|_g\,, \quad \|f - zs\|_g = \max_{z \in S} |g(z)(f(z) - zs(z))|\,, \tag{47}$$
>
> where
>
> $$g(z) = \begin{cases} 1 & \text{if} \quad z > 0 \\ \delta & \text{if} \quad z < 0 \end{cases}\,, \quad f(z) = \begin{cases} 1 & \text{if} \quad z > 0 \\ \gamma & \text{if} \quad z < 0 \end{cases}$$
>
> is an indefinite polynomial preconditioner with
>
> $$\gamma = \frac{\bar{d} + \bar{c}}{\bar{b} + \bar{a}} \quad \text{and} \quad \delta = \frac{\bar{b} - \bar{a}}{\bar{d} - \bar{c}}\,.$$
>
> The numbers $\bar{a}$, $\bar{b}$, $\bar{c}$, and $\bar{d}$ are defined by
>
> $$\bar{a} = \min_{z \in [a,b]} zs(z)\,, \quad \bar{b} = \max_{z \in [a,b]} zs(z)\,, \quad \bar{c} = \min_{z \in [c,d]} zs(z)\,, \quad \text{and} \quad \bar{d} = \max_{z \in [c,d]} zs(z)\,.$$
>
> Moreover, there exist parameters γ_0 and δ_0, $(\gamma_0, \delta_0) \in \Gamma$, such that $s^*(z, \gamma_0, \delta_0)$ is an optimal indefinite polynomial preconditioner.

(47) is a linear Chebyshev approximation problem depending on the two parameters $(\gamma, \delta) \in \Gamma$. We seek to approximate $f(z)$ by polynomials of the form $zs(z)$ in the weighted uniform norm $\|.\|_g$. The standard tool for the numerical solution of such general real Chebyshev approximation problems is the method of Remez. The Remez type procedure is based on the equioscillation property [4].

6.1 Remez Algorithm

Let $l \in \mathbb{N}$, $\gamma \in \mathbb{R}$, and $\delta > 0$ be given. In the following, let s be any candidate for the optimal polynomial with degree $l-1$ and real coefficients. We introduce the so-called residual polynomial

$$p(z) = p(z,s) = 1 - zs(z) \tag{48}$$

corresponding to s and therefore

$$g(z)(f(z) - zs(z)) = \begin{cases} p(z) & \text{if} \quad z > 0 \\ \delta(\gamma - 1 + p(z)) & \text{if} \quad z < 0 \end{cases} . \tag{49}$$

We seek a polynomial s with $l+1$ extremal points

$$c \le z_1 < z_2 < \cdots < z_k \le d \,, \quad a \le z_{k+1} < \cdots < z_l < z_{l+1} \le b$$

and a number $y \in \mathbb{R}$ such that

$$g(z_j)(f(z_j) - z_j s(z_j)) = \begin{cases} (-1)^{j-1} y & \text{for} \quad j = 1, \ldots, k \\ (-1)^j y & \text{for} \quad j = k+1, \ldots, l+1 \end{cases} . \tag{50}$$

Moreover, if s is optimal, then $w(\gamma, \delta) = |y|$ (see (47)). For any $k \in \{1, \ldots, n+1\}$, denote by

$$Z_k = \{(z_1, \ldots, z_{n+1}) \,:\, c \le z_1 < \cdots < z_k \le d, \; a \le z_{k+1} < \cdots < z_{l+1} \le b\}$$

the set of all possible z_j for which (50) holds. To each $Z \in Z_k$, there is a unique polynomial $s(z) = s(z, Z)$ and a unique number $y = y(Z) \in \mathbb{R}$ such that

$$\begin{aligned} p(z) &= \sum_{j=1}^{k} (1 - \gamma + (-1)^{j-1} y/\delta) L_j(z) + \sum_{j=k+1}^{l+1} (-1)^j y L_j(z) \,, \\ L_j(z) &= \prod_{\substack{i=1 \\ i \ne j}}^{l+1} \frac{z - z_i}{z_j - z_i} \,, \quad L_j(z_i) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases} \,, \end{aligned} \tag{51}$$

and

$$y = \frac{1 + (\gamma - 1) \sum_{j=1}^{k} L_j(0)}{(1/\delta) \sum_{j=1}^{k} (-1)^{j-1} L_j(0) + \sum_{j=k+1}^{l+1} (-1)^j L_j(0)} \,. \tag{52}$$

The Lagrange interpolation formula $L_j(z)$ can be rewritten in such a way that it can evaluated and updated in $\mathcal{O}(l)$ operations. The numerator of L_j can be written as the quantity

$$L(z) = (z - z_1)(z - z_2) \cdots (z - z_{l+1}) = \prod_{j=1}^{l+1} (z - z_j) \tag{53}$$

divided by $(z - z_j)$. We define the barycentric weights by

$$w_j = \frac{1}{\prod\limits_{\substack{k=1 \\ k \neq j}}^{l+1} (z_j - z_k)} = \frac{1}{L'(z_j)} \,, \quad j = 1, \dots, l+1 \,, \tag{54}$$

and thus we can write L_j as

$$L_j(z) = L(z) \frac{w_j}{z - z_j} \quad \text{and} \quad L_j(0) = -L(0) \frac{w_j}{z_j} \,. \tag{55}$$

Using (53) – (55), we transform the Eqs. (51) and (52) into

$$p(z) = L(z) \left(\sum_{j=1}^{k} (1 - \gamma + (-1)^{j-1} y/\delta) \frac{w_j}{z - z_j} + \sum_{j=k+1}^{l+1} (-1)^j y \frac{w_j}{z - z_j} \right) \tag{56}$$

and

$$y = - \frac{1 - (\gamma - 1) L(0) \sum_{j=1}^{k} w_j / z_j}{L(0) \left((1/\delta) \sum_{j=1}^{k} (-1)^{j-1} w_j / z_j + \sum_{j=k+1}^{l+1} (-1)^j w_j / z_j \right)} \,. \tag{57}$$

The resulting Remez algorithm is as follows.

1. Choose $l \in \mathbb{N}$, $k \in \{1, \dots, l+1\}$, and $l+1$ alternation points z_j such that $Z = (z_1, \dots, z_{l+1}) \in Z_k$, $c \leq z_1 < \cdots < z_k \leq d < 0 < a \leq z_{k+1}, \cdots < z_{l+1} \leq b$.
2. Evaluate y and the residual polynomial $p(z)$ by means Lagrange interpolation polynomials $L_j(z)$:

$$p(z_j) = \left\{ \begin{array}{ll} 1 - \gamma + (-1)^{j-1} y/\delta & \text{for} \quad j = 1, \dots, k \\ (-1)^j y & \text{for} \quad j = k+1, \dots, l+1 \end{array} \right. .$$

3. Find local extrema of $g(z)(f(z) - zs(z))$ on mesh S, and form a new reference point set $Z' = (z'_1, \dots, z'_{l+1}) \in Z_k$.
4. If algorithm not converged go to step 2.

$$\begin{array}{rcl} |g(z)(f(z) - zs(z))| & \leq & |y| \quad \text{for any} \;\; z \in S \\ |g(z_j)(f(z_j) - z_j s(z_j))| & = & |y| \quad \text{for} \;\; j = 1, \dots, l+1 \end{array}$$

A practical procedure for computing the approximate local extrema in step 3 can be found in [8]. The initial choice of reference point set Z is completely arbitrary. The Remez algorithm tends to self-correct a bad choice of reference points by the exchange process. A possible choice of the initial reference point set Z is:

- Equidistant nodes with spacing $h = (d - c)/(k - 1)$ on the interval $[c, d]$ and with spacing $h = (b - a)/(l - k)$ on the interval $[a, b]$.

- Chebyshev points of the first kind. Zeros of the Chebyshev polynomial on $[-1,1]$ are

$$\cos\frac{(2j-1)\pi}{2l+2}\,, \quad j=1,\dots,l+1\,.$$

- Chebyshev points of the second kind. Extrema of the Chebyshev polynomial on $[-1,1]$ are

$$\cos\frac{(j-1)\pi}{l}\,, \quad j=1,\dots,l+1\,.$$

Now, we consider the problem of how to obtain the bounds $a,b,c,d \in \mathbb{R}$ in the set $S=[a,b]\cup[c,d]$ where $c<d<0<a<b$. Ideally, a, b, c, and d are the four extreme eigenvalues of A. Some results of this problem can be found in [22, 2, 21].

6.2 Nelder-Mead Method

The Nelder-Mead method [17] or downhill simplex method is a commonly used nonlinear optimization algorithm for unconstrained real functions. It is a numerical method for multidimensional minimization, that is, finding the minimum of an objective function of more than one independent variable. The method requires only function evaluations, not derivates. It uses the concept of a simplex, which is a polyhedron of $m+1$ vertices in m dimensions. Examples of simplexes include a line segment on a line, a triangle on a plane, a tetrahedron in three-dimensional space and so forth. The problem can be written as

$$\min_x F(x)\,, \quad x\in\mathbb{R}^m\,, \quad F\in\mathcal{C}(\mathbb{R}^m,\mathbb{R})\,. \tag{58}$$

The Nelder-Mead algorithm is a simple, intuitive, and relatively robust method that approaches the optimum in great steps in the beginning of the search. It must be started not just with a single point, but with $m+1$ points, defining an initial simplex:

$$\Sigma=\{x^1,\dots,x^{m+1}\}\,.$$

Let s^j be the center of gravity with respect to x^j:

$$s^j=\frac{1}{m}\sum_{\substack{i=1\\ i\neq j}}^{m+1} x^i\,.$$

There are three construction (basic) principles to determine a new point of the simplex Σ.

1. Reflexion of the corner x^j at the center s^j. The new point is determined from

$$x^r=s^j+\alpha(s^j-x^j)\,, \quad 0<\alpha\le 1\,.$$

2. Expansion of the corner x^j in the direction $(s^j - x^j)$. The new point is determined from
$$x^e = s^j + \beta(s^j - x^j) = s^j + \frac{\beta}{\alpha}(x^r - s^j)\,, \quad \beta > \alpha\,.$$
3. Contraction with three different types. $0 < \gamma < \alpha$ denotes a contraction constant.
   (a) Partial interior contraction of x^j in the direction $(s^j - x^j)$. The new point is determined from
   $$x^c = s^j + \gamma(x^j - s^j)\,.$$
   (b) Partial exterior contraction of x^j in the direction $(s^j - x^j)$. The new point is determined from
   $$x^c = s^j + \gamma(s^j - x^j) = s^j + \frac{\gamma}{\alpha}(x^r - s^j)\,.$$
   (c) Total contraction (shrink step) to x^j. All points will replaced by
   $$x^i = (x^i + x^j)/2\,, \quad \forall\, i \in \{1, \ldots, m+1\}\,.$$

Note that standard values are $\alpha = 1$, $\beta = 2$, and $\gamma = 1/2$.

It is possible to extend the Nelder-Mead method to simple bounds and nonlinear inequality and equality constraints [16]:

$$\begin{gathered}\min_x F(x)\,, \quad x \in \mathbb{R}^m\,, \quad F \in \mathcal{C}(\mathbb{R}^m, \mathbb{R})\,, \\ x_i^{min} \le x_i \le x_i^{max},\ i = 1, \ldots, m, \\ g_i(x) \le 0,\ i = 1, \ldots, n_i, \quad h_i(x) = 0,\ i = n_i + 1, \ldots, n_c\,. \end{gathered} \tag{59}$$

An adaptive linear penalty function is used to handle general inequality and equality constraints. The problem (59) is rewritten in an unconstrained penalized form,

$$\begin{gathered}\min_x L(x, \lambda)\,, \quad x \in \mathbb{R}^m\,, \quad \lambda \in \mathbb{R}^{n_c}\,, \quad L \in \mathcal{C}(\mathbb{R}^m \times \mathbb{R}^{n_c}, \mathbb{R})\,, \\ L(x, \lambda) = F(x) + \sum_{i=1}^{n_i} \lambda_i \max(0, g_i(x)) + \sum_{i=n_i+1}^{n_c} \lambda_i \max(0, |h_i(x)|)\,. \end{gathered} \tag{60}$$

The penalty parameters $\lambda_i, i = 1, \ldots, n_c$, are updated after each determination of a new simplex point by the Nelder-Mead algorithm. The updating scheme consists of increasing penalty parameters of violated constraints. They are initialized as 0.

If $(L(x^{new}, \lambda^k) \le L(x^{best}, \lambda^k)$ **then**

- $s > 0$, positive step size
- $\lambda_i^{k+1} = \lambda_i^k + s \max(0, g_i(x^{new}))\,, \quad i = 1, \ldots, n_i\,,$
- $\lambda_i^{k+1} = \lambda_i^k + s \max(0, |h_i(x^{new})|)\,, \quad i = n_i + 1, \ldots, n_c\,,$

- $x^{best} = \arg \min_{x \in \{x^{new}, x^{best}, \Sigma\}} L(x, \lambda^{k+1})$,

End if.

The original Nelder-Mead algorithm was developed for unbounded domain problems. The new points can leave the domain after either the reflexion or the expansion operation. Its coordinates are projected on the bounds:

$$\begin{aligned} &\text{if } (x_i < x_i^{min}) \text{ set } x_i = x_i^{min}\,, \\ &\text{if } (x_i > x_i^{max}) \text{ set } x_i = x_i^{max}\,. \end{aligned}$$

6.3 Asymptotic Convergence Factor

Using the Chebyshev approximation problem (47), we now construct an objective function $F(x)$ (see (58)) for the Nelder-Mead method. For the choice of a suitable preconditioner, it is cruical to have error bounds for the iterates. We express the Krylov subspace $\mathcal{K}_m(A, r_0)$ in terms of polynomials $q(z)$:

$$\mathcal{K}_m(A, r_0) = \{q(A)r_0 \,:\, q \in \Pi_{m-1}\}\,,$$

where the notation Π_{m-1} will be used for the set of all complex polynomials of degree at most $m-1$. Thus, we have the following minimal residual property

$$\|b - Ax_m\| = \min_{x \in x_0 + \mathcal{K}_m} \|b - Ax\|\,, \quad x_m \in x_0 + \mathcal{K}_m\,. \tag{61}$$

Using (61), we can deduce the following result:

$$\frac{\|b - Ax_m\|}{\|b - Ax_0\|} \le \mathcal{E}_m(a, b, c, d)\,, \quad m = 1, 2, \ldots\,, \tag{62}$$

with

$$\mathcal{E}_m(a, b, c, d) = \min_{p \in \Pi_m^{(r)}, p(0)=1} \; \max_{z \in [a,b] \cup [c,d]} |p(z)|\,, \tag{63}$$

where $\Pi_m^{(r)}$ denotes the set of all real polynomials of degree at most m. Unfortunately, the solution of (63) is explicitly know only for special cases. For the general case holds

$$\lim_{m \to \infty} (\mathcal{E}_m(a, b, c, d))^{1/m} = \kappa(a, b, c, d)\,, \quad 0 < \kappa(a, b, c, d) < 1\,. \tag{64}$$

$\kappa(a, b, c, d)$ is usually called the asymptotic convergence factor. An explicit formula for κ for two intervals in terms of elliptic integrals is derived in [4].

$$R_F(x, y, z) = \frac{1}{2} \int_0^\infty \frac{dt}{\sqrt{(t+x)(t+y)(t+z)}}\,, \quad x, y, z \ge 0\,, \tag{65}$$

is the standard form of the elliptic integral of the first kind.

Let $c < d < 0 < a < b$. Then

$$\begin{aligned} \kappa(a,b,c,d) &= \frac{\vartheta_4(\pi(v_0 - M)/(2K), q)}{\vartheta_4(\pi(v_0 + M)/(2K), q)}\,, \\ \vartheta_4(\psi, q) &= 1 + 2\sum_{j=1}^{\infty}(-1)^j q^{j^2}\cos(2\psi j)\,, \end{aligned} \qquad (66)$$

where

$$\begin{aligned} q &= e^{-\dfrac{\pi K'}{K}}\,, \quad k = \sqrt{\frac{(a-d)(b-c)}{(a-c)(b-d)}}\,, \quad K = R_F(1, 0, 1-k^2)\,, \\ K' &= R_F(1,0,k^2)\,, \quad M = -\sqrt{\frac{a-c}{b-c}}\, R_F\left(1, \frac{b-a}{b-c}, \frac{b-a}{b-d}\right)\,, \\ v_0 &= -\sqrt{\frac{a(b-d)}{b(a-d)}}\, R_F\left(1, \frac{d(a-b)}{b(a-d)}, \frac{c(a-b)}{b(a-c)}\right)\,. \end{aligned} \qquad (67)$$

For $J \in \mathbb{N}$, we set

$$\vartheta_4^{(J)}(\psi, q) = 1 + 2\sum_{j=1}^{J}(-1)^j q^{j^2}\cos(2\psi j)\,.$$

If J is choosen large enough, the finite series $\vartheta_4^{(J)}(\psi, q)$ will yield a sufficiently accurate approximation to $\vartheta_4(\psi, q)$. For the calculation of the integral R_F, we use a procedure due to Carlson [18].

We now return to the right polynomial preconditioned system (44). Next, we state error bounds for this system. Setting

$$\bar{a} = \min_{z\in[a,b]} zs(z)\,, \quad \bar{b} = \max_{z\in[a,b]} zs(z)\,, \quad \bar{c} = \min_{z\in[c,d]} zs(z)\,, \text{ and } \bar{d} = \max_{z\in[c,d]} zs(z)\,,$$

it follows that

$$\varrho(As(A)) \subset \bar{S} = [\bar{a}, \bar{b}] \cup [\bar{c}, \bar{d}] \quad \text{and} \quad \bar{c} < \bar{d} < 0 < \bar{a} < \bar{b}.$$

Analogous to (62), we get the estimates

$$\frac{\|b - As(A)y_m\|}{\|b - As(A)y_0\|} = \frac{\|b - Ax_m\|}{\|b - Ax_0\|} \le \mathcal{E}_m(\bar{a}, \bar{b}, \bar{c}, \bar{d})\,, \quad m = 1, 2, \ldots\,.$$

Furthermore, the error bound behaves like

$$\mathcal{E}_m(\bar{a}, \bar{b}, \bar{c}, \bar{d}) \approx (\kappa(\bar{a}, \bar{b}, \bar{c}, \bar{d}))^m$$

for large m.

We now return to our optimization problem (47). For the optimal polynomial $s^*(z)$ of $w(\gamma, \delta)$ follows

$$\begin{aligned} \bar{a}(s^*) &= 1 - w(\gamma,\delta)\,, \quad \bar{b}(s^*) = 1 + w(\gamma,\delta)\,, \\ \bar{c}(s^*) &= \gamma - \frac{w(\gamma,\delta)}{\delta}\,, \quad \bar{d}(s^*) = \gamma + \frac{w(\gamma,\delta)}{\delta}\,. \end{aligned}$$

s^* is an indefinite polynomial preconditioner if, and only if, $(\gamma, \delta) \in \Gamma_w$:

$$\Gamma_w = \{(\gamma,\delta) \in \mathbb{R} \times \mathbb{R} \,:\, \delta > 0,\, w(\gamma,\delta) < 1, \text{ and } \gamma < -w(\gamma,\delta)/\delta\}\,.$$

The objective function $F(x)$ of (58) is replaced by the asymptotic convergence factor

$$\kappa(1 - w(\gamma,\delta),\, 1 + w(\gamma,\delta),\, \gamma - \frac{w(\gamma,\delta)}{\delta},\, \gamma + \frac{w(\gamma,\delta)}{\delta})$$

with $x = (\gamma, \delta)$ and $m = 2$. Thus, we have to solve a nonlinear optimization problem with simple constraints. The function $F(x)$ is continous, but only piecewise differentiable. There exist parameters γ_0 and δ_0 such that

$$F(\gamma_0, \delta_0) = \min_{(\gamma,\delta)\in\Gamma_w} F(\gamma,\delta)\,, \quad (\gamma_0,\delta_0) \in \Gamma_w\,. \tag{68}$$

6.4 Implementation Details

The polynomials $s(z)$ and $s^\sigma(z + \sigma)$ (see (46)) are given by

$$s(z) = \sum_{i=0}^{l} c_i z^i \quad \text{and} \quad z^\sigma(z+\sigma) = \sum_{i=0}^{l} c_i^\sigma (z^i + \sigma)\,,$$

respectively. The coefficients c_i, $i = 0, \dots, l$, and the shift σ are known. The goal is to evaluate the parameter τ and the coefficients c_i^σ, $i = 0, \dots, l$.

$$\begin{aligned} (z+\sigma)s^\sigma(z+\sigma) &= zs(z) + \tau \\ (z+\sigma)\sum_{i=0}^{l} c_i^\sigma (z+\sigma)^i &= z\sum_{i=0}^{l} c_i z^i + \tau \\ (z+\sigma)\sum_{i=0}^{l} c_i^\sigma \sum_{j=0}^{i} \binom{i}{j} z^{i-j}\sigma^j &= \sum_{i=0}^{l} c_i z^{i+1} + \tau \\ \sum_{i=0}^{l} c_i^\sigma \sum_{j=0}^{i} \binom{i}{j} z^{i+1-j}\sigma^j + \sigma \sum_{i=0}^{l} c_i^\sigma \sum_{j=0}^{i} \binom{i}{j} z^{i-j}\sigma^j &= \sum_{i=0}^{l} c_i z^{i+1} + \tau \end{aligned} \tag{69}$$

Equating the coefficients of z^0 $(j = i)$ in Eq. (69) and from (46) we get

$$\tau = \sigma \sum_{i=0}^{l} c_i^\sigma \sigma^i = \sigma \sum_{i=0}^{l} c_i(-\sigma)^i = \sigma \sum_{i=0}^{l} (-1)^i c_i \sigma^i\,.$$

The coefficients c_i^σ, $i = 0, \dots, l$, are solutions of the following linear system of equations. Equating the coefficients of z^{k+1}, $k = 0, \dots, l$, results in:

1. right-hand side of (69): $i+1 = k+1$, c_k

2. left-hand side of (69):

   (a) $i+1-j = k+1$, $j = i-k \geq 0$, $\sum\limits_{i=k}^{l} c_i^\sigma \binom{i}{i-k} \sigma^{i-k}$

   (b) $i-j = k+1$, $j = i-k-1 \geq 0$, $\sum\limits_{i=k+1}^{l} c_i^\sigma \binom{i}{i-k-1} \sigma^{i-k}$

3. $\sum\limits_{i=k}^{l} c_i^\sigma \binom{i}{i-k} \sigma^{i-k} + \sum\limits_{i=k+1}^{l} c_i^\sigma \binom{i}{i-k-1} \sigma^{i-k} = c_k\,.$

After some transformations we get the coefficients c_i^σ as the solution of an upper triangular system of equations.

```
0. Input: $\{c_0, \ldots, c_l\}$.
   $c_l^\sigma = c_l$.

For $k = l-1, \ldots, 0$ do:

    1. $c_k^\sigma = c_k - \sum_{i=k+1}^{l} \binom{i+1}{k+1} \sigma^{i-k} c_i^\sigma\,.$

End for.
```

We now consider the computation of the coefficients c_k, $k = 0, \ldots, l$, from the Lagrange polynomial $L_j(z)$ (see (51)). Using ansatz

$$L_j(z) = \prod_{\substack{i=1\\ i\neq j}}^{l+1} \frac{z - z_i}{z_j - z_i} = \sum_{k=1}^{l+1} a_{jk} z^{k-1}$$

we get

$$L_j(z_i) = \delta_{ji} = \sum_{k=1}^{l+1} a_{jk} z_i^{k-1}$$

and therefore

$$\begin{pmatrix} \delta_{j1} \\ \vdots \\ \delta_{ji} \\ \vdots \\ \delta_{jl+1} \end{pmatrix}^T = \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jk} \\ \vdots \\ a_{jl+1} \end{pmatrix}^T \begin{pmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \vdots & & \vdots & & \vdots \\ z_1^{k-1} & \cdots & z_i^{k-1} & \cdots & z_{l+1}^{k-1} \\ \vdots & & \vdots & & \vdots \\ z_1^l & \cdots & z_i^l & \cdots & z_{l+1}^l \end{pmatrix}. \tag{70}$$

Consequently, we have to solve (70) for $j = 1, \ldots, l+1$ with the transpose of the matrix. This is a Vandermonde matrix.

$$\begin{pmatrix} 1 & \cdots & z_1^{k-1} & \cdots & z_1^l \\ \vdots & & \vdots & & \vdots \\ 1 & \cdots & z_i^{k-1} & \cdots & z_i^{k-1} \\ \vdots & & \vdots & & \vdots \\ 1 & \cdots & z_{l+1}^{k-1} & \cdots & z_{l+1}^l \end{pmatrix} \begin{pmatrix} a_{j1} \\ \vdots \\ a_{jk} \\ \vdots \\ a_{jl+1} \end{pmatrix} = \begin{pmatrix} \delta_{j1} \\ \vdots \\ \delta_{ji} \\ \vdots \\ \delta_{jl+1} \end{pmatrix} \tag{71}$$

One could in principle solve Eq. (71) by standard techniques for linear equations. A more efficient method is derived in [18].

Using (48), (51), (52), and (70) we can compute the coefficients c_k of the polynomial $s(z)$ (see (46)).

$$\begin{aligned} p(z_i) = \sum_{j=1}^{l+1} y_j L_j(z_i) &= \sum_{j=1}^{l+1} y_j \delta_{ji}\,, \quad y_j = p(z_j) \\ &= \sum_{j=1}^{l+1} y_j \sum_{k=1}^{l+1} a_{jk} z_i^{k-1} = \sum_{k=1}^{l+1} \sum_{j=1}^{l+1} y_j a_{jk} z_i^{k-1} \\ &= \sum_{k=1}^{l+1} b_{k-1} z_i^{k-1}\,, \quad b_{k-1} = \sum_{j=1}^{l+1} y_j a_{jk}\,, \quad b_0 = 1 \end{aligned}$$

From (48) we get $s(z) = z^{-1}(1 - p(z))$. This results in

$$c_k = -b_{k+1} \quad \text{for} \quad k = 0, \ldots, l-1\,.$$

7 Numerical Results

A nonequidistant mesh of 57 664 elementary cells including graded PML regions is used for the discretization of (6), that means the order of the system of linear algebraic equations is 172 992 (see (21)). The number of internal points $x_I^{(G)}$ (see Table 1 and Eq. (22)) is 152 608. The stopping criterion was a reduction of the norm of the residual for the preconditioned system (45) by 10^{-8}.

Based on the family of approximation problems (47), we have computed indefinite polynomial preconditioners. For this purpose, the Remez algorithm in Section 6.1 was used. Using the Nelder-Mead method (see Section 6.2), optimal indefinite polynomial preconditioners were computed by solving the constrained optimization problem (68) numerically.

We compare cumulative iteration counts required to individually solve each of the n_s linear systems (see (30)) using coupled two-term QMR algorithm with the number of iterations required to solve all the n_s systems simultaneously with the QMR algorithm for shifted matrices. Table 2 shows the number of iteration required to individually solve without polynomial preconditioning. Table 3 shows the numbers

Table 2: Number of iterations for each shifted linear system

Number of shifted system	1	2	3	4	5
QMR iterations	97	872	2 843	137	508
Accumulated QMR iterations	97	969	3 812	3 949	4 457
Number of shifted system	6	7	8	9	10
QMR iterations	1 558	718	732	1 177	1 958
Accumulated QMR iterations	6 015	6 733	7 465	8 642	10 600

Table 3: Number of iterations for shifted matrices

Number of shifted linear systems	$n_s = 4$	$n_s = 10$
No preconditioning	2 348	2 348
Preconditioning: (γ, δ)	1 534	1 534
Preconditioning: (γ_0, δ_0)	1 390	1 390

of iterations for shifted matrices with polynomial preconditioning for $n_s = 4$ and $n_s = 10$, respectively. We choose the parameters $(\gamma, \delta) \in \Gamma_w$ such that the two intervals of $\bar{S} = [\bar{a}, \bar{b}] \cup [\bar{c}, \bar{d}]$ containing the eigenvalues of the preconditioned matrix $As(A)$ have the same length and position as the original intervals of $S = [a, b] \cup [c, d]$, i.e.,

$$\frac{b+a}{d+c} = \frac{\bar{b}+\bar{a}}{\bar{d}+\bar{c}} \quad \text{and} \quad \frac{b-a}{d-c} = \frac{\bar{b}-\bar{a}}{\bar{d}-\bar{c}}\,.$$

Thus, we have

$$\gamma = \frac{d+c}{b+a} \quad \text{and} \quad \delta = \frac{b-a}{d-c}\,.$$

Note that for the example considered here

$$\gamma \approx -0.2241 \quad \text{and} \quad \delta \approx 4.4610\,.$$

We have also computed the 'optimal' parameters $(\gamma_0, \delta_0) \in \Gamma_w$:

$$\gamma_0 \approx -0.7339 \quad \text{and} \quad \delta_0 \approx 4.3407\,.$$

The degree of the Lagrange polynomial $L_j(z)$ (see (51)) is $l = 9$. We choose for the preconditioner $s(A)$ (see (45)) the linear case. This choice decreases the numerical effort and is more stable. Our seed linear system is the system with the index $j_1 = 3$. The comparison of the iteration numbers of the coupled two-term QMR algorithm for shifted matrices with the individual solution of each linear system shows the important advantage. Once more the linear polynomial preconditioner increases the benefit. The computation of the 'optimal' parameters (γ_0, δ_0) leads to better results.

8 Conclusions

We have derived polynomial preconditioners for indefinite linear systems which lead to indefinite preconditioned coefficient matrices. Such polynomials can be obtained via the solution of linear Chebyshev approximation problems depending on two parameters. The concept of the asymptotic convergence factors leads to an optimal preconditioner. The Nelder-Mead method or downhill simplex method is a commonly used algorithm for the multidimensional minimization. A Remez type procedure for the numerical solution of the linear Chebyshev approximation problem was outlined.

One problem is to find suitable informations on the location of the eigenvalues of A, i.e., the bounds of the two intervals $[a, b]$ and $[c, d]$. Another problem is the choice of the seed linear system. For restarted methods, the crucial question is: Are the residuals collinear? For minimal residual methods the residuals are in general not collinear.

References

[1] Klaus Beilenhoff, Wolfgang Heinrich, and Hans L. Hartnagel. Improved finite-difference formulation in frequency domain for three-dimensional scattering problems. *IEEE Transactions on Microwave Theory and Techniques*, 40, No. 3:540–546, 1992.

[2] Ljiljana Cvetkovic, Vladimir Kostic, and Richard S. Varga. A new Geršgorin-type eigenvalue inclusion set. *Electronic Transactions on Numerical Analysis*, 18:73–80, 2004.

[3] R.W. Freund. *Krylov subspace methods for complex non-Hermitian linear systems*. Habilitation thesis, Universität Würzburg, 1991. Also available as RIACS Technical Report 91.11.

[4] R.W. Freund. On polynomial preconditioning and asymptotic convergence factors for indefinite Hermitian matrices. *Linear Algebra and Its Applications*, 154–156:259–288, 1991. An extended version of this paper is available as RIACS Technical Report 89.32.

[5] R.W. Freund. Solution of shifted linear systems by Quasi-Minimal Residual iterations. In L. Reichel, A. Ruttan, and R.S. Varga, editors, *Numerical Linear Algebra*, pages 101–121. W. de Gruyter, 1993.

[6] R.W. Freund and W. Malhotra. A Block-QMR algorithm for non-Hermitian linear systems with multiple right-hand sides. *Linear Algebra and Its Applications*, 254:119–157, 1997.

[7] R.W. Freund and N.M. Nachtigal. An implementation of the QMR method based on coupled two-term recurrences. *SIAM J. Sci. Comput.*, 15:313–337, 1994.

[8] G.H. Golub and L.B. Smith. Algorithm 414: Chebyshev approximation of continuous functions by a Chebyshev system of functions. *Communications of the ACM*, 14, No. 11:737–746, 1971.

[9] Georg Hebermehl, Friedrich-Karl Hübner, Rainer Schlundt, Thorsten Tischler, Horst Zscheile, and Wolfgang Heinrich. Simulation of microwave and semiconductor laser structures including absorbing boundary conditions. In Eberhard Bänsch, editor, *Challenges in Scientific Computing - CISC2002*, volume 35 of *Lecture Notes in Computational Science and Engineering, Springer Verlag*, pages 131–159, 2003.

[10] Georg Hebermehl, Jürgen Schefter, Rainer Schlundt, Thorsten Tischler, Horst Zscheile, and Wolfgang Heinrich. Simulation of microwave and semiconductor laser structures including PML: Computation of the eigen mode problem, the boundary value problem, and the scattering matrix. In A. Anile, G. Ali, and G. Mascali, editors, *Proc. 5th International Workshop Scientific Computing in Electrical Engineering (SCEE), Capo D'Orlando, Italy, September 5–9, 2004, Springer Verlag*, pages 203–214, 2006.

[11] Georg Hebermehl, Rainer Schlundt, Horst Zscheile, and Wolfgang Heinrich. Improved numerical solutions for the simulation of monolithic microwave integrated circuits. WIAS Preprint no. 236, Weierstraß-Institut für Angewandte Analysis und Stochastik, http://www.wias-berlin.de/publications/preprints/236/, 1996.

[12] Georg Hebermehl, Rainer Schlundt, Horst Zscheile, and Wolfgang Heinrich. Simulation of monolithic microwave integrated circuits. WIAS Preprint no. 235, Weierstraß-Institut für Angewandte Analysis und Stochastik, http://www.wias-berlin.de/publications/preprints/235/, 1996.

[13] Georg Hebermehl, Rainer Schlundt, Horst Zscheile, and Wolfgang Heinrich. Eigen mode solver for microwave transmission lines. *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 16:108–122, 1997.

[14] Georg Hebermehl, Rainer Schlundt, Horst Zscheile, and Wolfgang Heinrich. Improved numerical methods for the simulation of microwave circuits. *Surveys on Mathematics for Industry*, 9, No. 2:117–129, 1999.

[15] Georg Hebermehl, Rainer Schlundt, Horst Zscheile, and Wolfgang Heinrich. On the computation of eigen modes for lossy microwave transmission lines including perfectly matched layer boundary conditions. *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 20:948–964, 2001.

[16] M.A. Luersen, R. Le Riche, and F. Guyon. A constrained, globalized, and bounded Nelder-Mead method for engineering optimization. *Structural and Multidisciplinary Optimization*, 27:43–54, 2004.

[17] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[18] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in Fortran 77: the art of scientific computing.* Cambridge University Press, 2nd ed., 1992.

[19] Y. Saad. *Iterative Methods for Sparse Linear Systems.* PWS Publishing Company, 1996.

[20] Prodyut K. Talukder, Franz-Josef Schmückle, Rainer Schlundt, and Wolfgang Heinrich. Optimizing the FDFD method in order to minimize PML-related numerical problems. In *IEEE MTT-S Int. Microwave Symp. Dig.*, pages 293–296, 2007.

[21] Richard S. Varga. *Geršgorin and his circles.* Springer Verlag, Berlin, Germany, 2004.

[22] Richard S. Varga and Alan Krautstengl. On Geršgorin-type problems and ovals of Cassini. *Electronic Transactions on Numerical Analysis*, 8:15–20, 1999.

[23] T. Weiland. A discretization method for the solution of Maxwell's equations for six-component fields. *Electronics and Communication (AEÜ)*, 31:116–120, 1977.

[24] T. Weiland. On the unique numerical solution of Maxwellian eigenvalue problems in three dimensions. *Particle Accelerators (PAC)*, 17:277–242, 1985.