

SOLUTION OF THE TIME-DOMAIN INVERSE RESISTIVITY PROBLEM IN THE MODEL REDUCTION FRAMEWORK PART I. ONE-DIMENSIONAL PROBLEM WITH SISO DATA*

V. DRUSKIN[†], V. SIMONCINI[‡], AND M. ZASLAVSKY[§]

Abstract. Many time-domain problems in engineering applications can be described by means of a parameter dependent time-invariant dynamic systems. We are interested in parameter estimation, by fitting available transient measurements using the nonlinear least square method. As the main application, we consider the control source electromagnetic method (CSEM) of geophysical exploration governed by the diffusion Maxwell system, where the unknown parameters describe the spatial distribution of electrical resistivity. We propose a novel model reduction approach for constructing an efficient approximation of the Jacobian. The reduction is based on projection of the state variable onto a Rational Krylov subspace (RKS), and it allows us to split the time and space dependence of the derivative. We examine several popular RKSs and single out the H_2 -optimal subspace that not only minimizes the approximation error but completely annuls its influence on the inversion result. Preliminary numerical experiments with a simplified one-dimensional, single-input/single-output CSEM setting are reported to validate our strategy.

Key words. Gauss-Newton, Jacobian, rational Krylov subspace, model reduction, H_2 optimality, time-domain Maxwell's system, CSEM, inverse problem.

1. Introduction. Many time-domain problems in remote-sensing, such as radar imaging, seismic and electromagnetic geophysical exploration, can be described by time-invariant evolutionary partial differential equations (PDEs) or systems on $\mathbb{R}^3 \times [0, \infty]$, that can be written in the evolutionary form

$$A(r)u + u_t = 0, \quad u|_{t=0} = b. \quad (1.1)$$

Here $A(r)$ is a (first or second order) PDE operator in \mathbb{R}^3 , r is a PDE coefficient dependent on spatial coordinates and independent on time, u is the state variable dependent on both the space and time coordinates, and b is a compactly supported source distribution in \mathbb{R}^3 (see [20], [1]). The measurements are usually performed by weighting the state variable with respect to a receiver distribution q compactly supported in \mathbb{R}^3 , i.e., for given source the measured data (time-domain transfer functions) can be presented using three-dimensional integral such as $y(t) = \int_{\mathbb{R}^3} qu dV$. The inverse problem can be loosely formulated as an estimation of r (under some assumptions) via a number of such transfer functions measured with the help of several source and receiver distributions. To fix the ideas, we will assume that (1.1) comes from the diffusion Maxwell system

$$\nabla \times \rho \nabla \times u + u_t = 0, \quad u|_{t=0} = b, \quad (1.2)$$

where the state variable u is the magnetic vector field, $\rho \in L_\infty[\mathbb{R}^3]$ is the uniformly positive electrical resistivity distribution and b and q are respectively the magnetic source and receiver distributions on \mathbb{R}^3 .

The inverse problem is normally solved by minimization of the regularized data misfit functional with the help of the Gauss-Newton (GN) or nonlinear Conjugate

*Version of Sept 30, 2011.

[†]Schlumberger-Doll Research, 1 Hampshire st. Cambridge, MA 02139 USA (druskin1@slb.com)

[‡]Dipartimento di Matematica, Università di Bologna, Piazza di Porta S. Donato 5, I-40127 Bologna, Italy and CIRSA, Ravenna, Italy (valeria.simoncini@unibo.it).

[§]Schlumberger-Doll Research, 1 Hampshire st. Cambridge, MA 02139 USA (mzaslavsky@slb.com)

Gradient (a.k.a. electromagnetic migration) methods; see, e.g., [17], [14] and references therein. In these approaches, the Frechet derivatives of the transfer functions with respect to r are computed using the adjoint formulation. The regularized GN approach supplied with a line search is known as the most powerful optimization algorithm in principle, however its application to the large-scale time domain problems is hindered by the high dimensionality of the time-domain data sets (and, consequently, the sizes of the corresponding Jacobians). For example, a typical measurement set in the so-called time-domain control sources electromagnetic (tCSEM) problems may contain from 10^2 to 10^4 transfer functions and each of them can be discretized on the temporal grids spanning from 10^2 to 10^4 nodes. Therefore straightforward application of the GN algorithm would result in the data sets with up to 10^8 entries.

A look at the problem setting readily reveals that the problem corresponds to a standard multi-input/multi-output (MIMO) control theory formulation, with b and q being respectively the input and output weight vectors. It is well known that the complexity of such problems can be greatly reduced using well developed tools of the model order reduction (MOR) theory. MOR approaches were successfully applied to the forward modeling in [5], [8]; Obviously, the next logical step is to apply the model reduction approach to the Jacobian as well.

In this work we consider the model reduction approach based on the Rational Krylov subspace (RKS) projection method. We derive a representation for the reduced order Jacobian as the product of a time-dependent and a stationary part. This formula allows us to limit the size of the Jacobian per output (per given source-receiver pair) per model parameter, to twice the order of the reduced transfer function and, therefore, to simplify the computation of the Hessian.

A proper choice of the RKS is indeed critical for the success of the inversion. It is intuitive, that the latter must be greatly affected by the accuracy of the reduced order transfer function. So, it is not surprising that the best results can be obtained with the RKS satisfying the Meier-Luenberger necessary H_2 (Hardy space) optimality condition. However we found that such subspaces not only minimize the approximation error but completely annul its influence on the inversion result (even if the subspace is not optimal globally). In fact, we show that the approximation error belongs to the (left) null-space of the reduced Jacobian. In many cases the subspaces satisfying the Meier-Luenberger conditions can be efficiently computed using the so-called iterative rational Krylov algorithm (IRKA), [13]. We compare the inversion on such subspaces with those using other nearly optimal RKS's based on Zolotarev problem and adaptive pole selection algorithm [8], [9].

In this paper (Part I) we limit our exposition to the main concepts of our approach when applied to an example of the single-input/single-output (SISO) problem, and we consider a simplified parametric inversion. We derive an expression for an approximate Jacobian of the reduced order model (ROM) and consider special properties of the H_2 -optimal ROMs. We illustrate our analysis by some numerical examples on a 1D inverse problem corresponding to a simplified CSEM setting. The numerical results shows efficiency of the model reduction approaches with properly selected rational Krylov subspaces, in particular, with the H_2 -optimal ones.

In the second part (Part II) we plan to extend our approach to the solution of the full scale regularized MIMO multidimensional inverse problems.

2. Problem formulation. We would like to adhere to the conventional MOR setting, therefore we will view (1.1) as a semidiscrete parabolic problem where $b, u(t) \in \mathbb{R}^n$, $0 < A(r) = A^T(r) \in \mathbb{R}^{n \times n}$ is a matrix valued function of the vector parameter

$r \in \mathcal{S}$, where \mathcal{S} is a compact set in \mathbb{R}^n with all positive components, and we assume that n is very large. The solution to (1.1) can be written as

$$u(t, r) = e^{-tA(r)}b. \quad (2.1)$$

In general, the approach considered here only requires the uniqueness of the inverse problem (with respect to r) and the smoothness of $A(r)$, on \mathcal{S} . In this framework,

$$A \approx \nabla \times \rho \nabla \times. \quad (2.2)$$

Alternatively, A may stem from the FD discretization of the self-adjoint elliptic operator, namely $A \approx \nabla \cdot \rho \nabla$. In both cases, we consider

$$A = D^T \text{diag}(r)D,$$

where $r \approx \rho$ and $D \in \mathbb{R}^{n \times n}$ approximates the corresponding first order PDE operator. Here and below, if r is a vector, then $\text{diag}(r)$ is a diagonal matrix with the components of r as diagonal elements. For a given r , let

$$y(t, r) = q^T e^{-tA(r)}b \in L_2[0, \infty] \quad (2.3)$$

be the operator for the forward problem, for some $b, q \in \mathbb{R}^n$, where we restrict our analysis to single vectors b and q . The vector-function of the observed measurements $d(t)$ is given by

$$d(t) = y(t, r^{true}) + \Delta(t),$$

where $\Delta(t)$ is the measurement error, and r^{true} is the sought after resistivity. Recovering r^{true} from the data $d(t)$ may be performed by minimizing the misfit $y(t, r^{true}) - d(t)$ in an appropriate norm. In this case, r^{approx} can be estimated as

$$r^{approx} = \arg \min_{r \in \mathcal{S}} \Phi(r), \quad \Phi(r) = \frac{1}{2} \|y(\cdot, r) - d\|_{L_2}^2, \quad (2.4)$$

where $\|f\|_{L_2}^2 = \int_0^\infty f(t)^2 dt$; see, e.g., [18].

We assume that this minimization problem has unique solution. However, this is known to be an ill-posed problem. To circumvent this obstacle one must restrict the size of the compact set of admissible solutions \mathcal{S} . To make the minimization problem above numerically tractable, a Tikhonov regularization penalty term is usually added to the functional, see e.g., [2]. However, to significantly simplify the derivation of our new idea, we just consider formulation (2.4) and reserve the Tikhonov formulation to part II. We assume that \mathcal{S} is a compact convex k -dimensional linear manifold in \mathbb{R}^n , i.e., it is a compact convex subset in a k -dimensional affine subspace. For k sufficiently small, no regularization is needed.

We end this section with some notation. We shall denote with $f \diamond g$ the standard inner product in $L_2[0, \infty]$, namely $f \diamond g = \int_0^\infty f(t)g(t)dt$, so that the norm above is $\|f\|_{L_2}^2 = f \diamond f$. In particular, this inner product will also be used component-wise for vectors of functions:

$$(f_1, \dots, f_k)^T \diamond (g_1, \dots, g_j) = \begin{bmatrix} f_1 \diamond g_1 & \cdots & f_1 \diamond g_j \\ f_2 \diamond g_1 & \cdots & f_2 \diamond g_j \\ \vdots & \ddots & \vdots \\ f_k \diamond g_1 & \cdots & f_k \diamond g_j \end{bmatrix},$$

giving rise to matrices of corresponding size.

3. Subspace approximation of the Jacobian. The solutions of (2.4) lying in the interior of \mathcal{S} are zeros of the gradient of $\Phi(r)$ projected on the tangent space of \mathcal{S} , i.e., they satisfy the equation

$$(J(\cdot, \delta r)S)^T \diamond [y(\cdot, r) - d(\cdot)] = 0, \quad (3.1)$$

where J is the Jacobian operator

$$J : \delta r \in \mathbb{R}^n \mapsto \delta y \in L_2[0, \infty],$$

with $\delta y = \frac{d}{ds}y(t, r + s\delta r)|_{s=0}$, while the columns of $S \in \mathbb{R}^{n \times k}$ are formed by an (orthogonal) basis of the tangent space of \mathcal{S} . The projected Jacobian drives the approximation to the resistivity function by means of the following projected Gauss-Newton iteration:

$$r_{j+1} = r_j - S(S^T J^T J S)^{-1} S^T J^T \diamond (y(\cdot, r_j) - d(\cdot)), \quad (3.2)$$

that differs from the general Gauss-Newton iteration (cf., e.g., [16], [19]) by projection on \mathcal{S} .

One way to compute J would be to use the following well known adjoint formula (see, e.g., [17])

$$\delta y(t, r) = - \int_0^t [Du(t - \tau)]^T \text{diag}(r) Du(\tau) d\tau. \quad (3.3)$$

Here we take a different approach, and in the following we propose an approximation to δy , and thus to the action of the Jacobian J , by exploiting a subspace approximation of the forward problem solution.

Let \mathcal{U}_m be an m -dimensional subspace of \mathbb{R}^n , $m \ll n$ and for a given r , let $z_i \in \mathbb{R}^n$, $\|z_i\| = 1$, $\theta_i \in \mathbb{R}$, $i = 1, \dots, m$ be the Ritz pairs of $A(r)$ on that subspace. Then the Galerkin approximation of u in (2.1) in that subspace can be computed as

$$u_m(t, r) = \sum_{i=1}^m e^{-\theta_i t} z_i z_i^T b.$$

With this approximation, and for $q = b$, we can obtain an approximate solution to y in (2.3) as

$$y_m(t, r) = b^T u_m(t, r) = \sum_{i=1}^m e^{-\theta_i t} c_i, \quad (3.4)$$

where $c_i = (z_i^T b)^2$. We thus propose to modify the original inverse problem in (2.4) by using the approximation y_m in the optimization process, so that the projected problem can be written as

$$r^{approx} = \arg \min_{r \in \mathcal{S}} \Phi_m(r), \quad \Phi_m(r) = \frac{1}{2} \|y_m(\cdot, r) - d\|_{L_2}^2. \quad (3.5)$$

Using the chain rule we obtain

$$\delta y_m(t, \delta r) = \frac{d}{ds} y_m(r + s\delta r)|_{s=0} = \sum_{i=1}^m e^{-\theta_i t} \delta c_i - \sum_{i=1}^m t e^{-\theta_i t} c_i \delta \theta_i,$$

where

$$\delta c_i = \frac{d}{ds} c_i, \quad \delta \theta_i = \frac{d}{ds} \theta_i, \quad i = 1, \dots, m.$$

We have thus obtained $\delta y_m(t, \delta r)$ as the following inner product:

$$\delta y_m(t, \delta r) = \mathbf{e}_m(t)^T \begin{bmatrix} \delta c_1 \\ \vdots \\ \delta c_m \\ -c_1 \delta \theta_1 \\ \vdots \\ -c_m \delta \theta_m \end{bmatrix}, \quad (3.6)$$

with $\mathbf{e}_m(t)^T = [e^{-t\theta_1}, \dots, e^{-t\theta_m}, te^{-t\theta_1}, \dots, te^{-t\theta_m}]$, showing that the dependence on t only appears in $\mathbf{e}_m(t)$. We shall show that in fact, the inner product (3.6) implicitly defines a matrix $B \in \mathbb{R}^{2m \times n}$, independent of t , namely

$$B : \delta r \in \mathbb{R}^n \mapsto (\delta c_1, \dots, \delta c_m, \delta \theta_1, \dots, \delta \theta_m)^T \in \mathbb{R}^{2m}, \quad (3.7)$$

that is the Jacobian of the spectral problem, so as to give

$$\delta y_m(t, \delta r) = \mathbf{e}_m(t)^T B \delta r.$$

Formula (3.6) thus defines the Ritz approximation based Jacobian

$$J_m(t) = \mathbf{e}_m(t)^T B, \quad (3.8)$$

where $B = B(r)$, and

$$J_m : \delta r \in \mathbb{R}^n \mapsto \delta y \in L_2[0, \infty].$$

With this construction, the projected version of (3.1) for the new problem (3.5) can be written as

$$S^T B^T \mathbf{e}_m \diamond [y_m(\cdot, r) - d] = 0. \quad (3.9)$$

As already mentioned, solutions to (3.5) are also solutions to (3.9). We next show that a proper selection of the approximation space allows us to give sufficient, as well as necessary conditions, for a solution to the original problem to be also a solution of (3.9).

For a given r , let y_m^o be a so-called H_2 -optimal m -th order approximation of y , i.e.,

$$\|y_m^o - y\|_{L_2} = \min_{c_1, \dots, c_m, \theta_1, \dots, \theta_m} \|y_m - y\|_{L_2}.$$

The necessary condition of H_2 optimality is zero gradient

$$\frac{\partial}{\partial c_i} (y_m - y) \diamond (y_m - y) = 0, \quad \frac{\partial}{\partial \theta_i} (y_m - y) \diamond (y_m - y) = 0, \quad i = 1, \dots, m. \quad (3.10)$$

Using (3.4) we obtain $\frac{\partial}{\partial c_i}(y_m - y) \diamond (y_m - y) = 2 \exp(-\theta_i t) \diamond (y_m - y)$ and $\frac{\partial}{\partial \theta_i}(y_m - y) \diamond (y_m - y) = 2t \exp(-\theta_i t) \diamond (y_m - y)$, so condition (3.10) can be rewritten as

$$\mathbf{e}_m(t) \diamond [\tilde{y}_m - y] = 0. \quad (3.11)$$

For shorthand we shall call H_2 -extremal the subspaces satisfying the above necessary conditions of H_2 optimality. We distinguish it from the optimal space, because it is well known that the H_2 functional is generally nonconvex and may contain multiple extremal points [3].

THEOREM 3.1. [Meier and Luenberger, 1967] *Let us consider y_m obtained using the rational Krylov subspace (RKS)¹:*

$$\mathcal{U}_m = \text{span}\{\varphi, A\varphi, \dots, A^{m-1}\varphi\}, \quad \varphi = \prod_{j=1}^m (A + s_j I)^{-1} b. \quad (3.12)$$

Then condition

$$s_i = \theta_i, \quad i = 1, \dots, m \quad (3.13)$$

is necessary and sufficient for (3.11) to hold (H_2 extremality condition).

For completeness, we shall give a slightly simplified proof of this classical result.

Proof. Let us denote with f the Laplace-transforms $\tilde{f}(s) = \int_0^\infty e^{-st} f(t) dt$ for any $f \in L_2[0, \infty]$. Then

$$\begin{aligned} \mathbf{e}_m(t) \diamond f &= \left(\int_0^\infty e^{-t\theta_1} f(t) dt, \int_0^\infty e^{-t\theta_m} f(t) dt, \dots, \int_0^\infty t e^{-t\theta_1} f(t) dt, \int_0^\infty t e^{-t\theta_m} f(t) dt \right)^T \\ &= \left(\tilde{f}(\theta_1) \dots, \tilde{f}(\theta_m), \frac{d}{ds} \tilde{f}|_{s=\theta_1} \dots, \frac{d}{ds} \tilde{f}|_{s=\theta_m} \right)^T. \end{aligned}$$

By means of the chain rule, we see that condition (3.11) is equivalent to the interpolation conditions

$$(\tilde{y} - \tilde{y}_m^o)|_{s=\theta_i} = 0, \quad \frac{d}{ds}(\tilde{y} - \tilde{y}_m^o)|_{s=\theta_i} = 0, \quad i = 1, \dots, m. \quad (3.14)$$

It is known (see [12]) that (3.13) is the necessary and sufficient condition for the transfer-function of the Galerkin RKS approximation to satisfy (3.14). \square

We are thus ready to state our conditions on the exactness of the solution.

THEOREM 3.2. *Let us consider (3.9) with $d = y(r^{true})$ ($\Delta = 0$), y_m obtained using the Rational Krylov subspace defined in (3.12). Then the condition (3.13) is sufficient for r^{true} to be a solution of (3.9). Moreover, if rank of BS is $2m$, then this condition is also necessary.*

Proof. Assume that (3.13) holds. Then according to Theorem 3.1, the optimal solution $y_m^o(r^{true})$ obtained in the rational Krylov subspace (3.12) satisfies (3.11). By multiplying it from the left by $S^T B^T$ we obtain (3.9).

If rank of BS is $2m$, then $(BS)^T$ is full column rank, so that (3.9) simplifies to (3.11). Applying Theorem 3.1 in the reverse order we obtain the necessity. \square

This key result requires some further explanation to be explored in an appropriate manner. Intuitively we expect the error of the inverse problem to be at least of the order of $\|y_m - y\|_{L_2}$, i.e., for the H_2 -optimal subspace it would be at least of the order of the optimal approximation error $\|y_m^o - y\|_{L_2}$. However our result is stronger than

¹The space \mathcal{U}_m may equivalently be defined as $\mathcal{U}_m = \text{span}\{(A + s_1 I)^{-1} b, (A + s_2 I)^{-1} (A + s_1 I)^{-1} b, \dots, \prod_{j=1}^m (A + s_j I)^{-1} b\}$.

that: the error is not just minimized, but its impact on the solution of the inverse problem is also completely canceled out.

If the reduced order model is not exact, then

$$\Phi_m(r^{true}) > \Phi(r^{true}) = 0.$$

However, r^{true} is still an extremal point of Φ_m , if we use an H_2 -extremal subspace (and provided the measurement error is negligible). Obviously, the extremal set of Φ_m can be significantly larger than the one of Φ , since using reduced order transfer function in the inverse problem may increase the ambiguity of the solution. However, the minimization of Φ_m can become unique after restricting r to a proper admissible set \mathcal{S} , in which case the approximation error does not affect the solution of the inverse problem if we use an H_2 -extremal subspace. As a consequence, we see that H_2 -extremal approximations not only may minimize the approximation error, but they can completely eliminate the error of the solution of the inverse problem. Moreover, the subspace does not have to be *globally optimal* for the error cancellation. We should point out however, that the global optimality is desirable for more accurate approximation and better conditioning of the reduced order Jacobian. The complete result is stated in the following corollary.

COROLLARY 3.3. *Let (3.9) (obtained using RKS) with $d = y(r^{true})$ have a unique solution $r^{approx} \in \mathcal{S}$. Then condition (3.11) is sufficient for $r^{true} = r^{approx}$. If rank of BS is $2m$ then this condition is necessary as well.*

We should point out that ROM Jacobians of parameter-dependent dynamic systems were recently studied in [4]. They proved the exactness of the parametric Jacobian of a ROM at the *interpolation points*. Their result is valid for an arbitrary RKS. Our results are different in the sense that we study a Jacobian property that is global in time and specific for H_2 -extremal subspaces. We showed that if the approximate transfer function y_m^o is computed via projection onto an H_2 -extremal subspace, then the error $y_m^o - y$ is in the null-space of J_m . Therefore, r^{true} remains a solution of the extremal equation with the exact data (though enlargement of J_m 's null-space can be a source of additional nonuniqueness indeed).

4. Approximate Jacobian and derivation of the matrix B . In the numerical solution of (3.9) by means of the Gauss-Newton method, each iteration realizes (3.2), where the Ritz approximations y_m and J_m are used instead of y and J , respectively. In this section we give more details on the actual implementation of (3.2), which requires the explicit application of J_m , and thus the derivation of B . We also recall that J_m depends on the current projection space, therefore the (possibly optimal) set $\{s_1, \dots, s_m\}$ of rational Krylov subspace poles has to be updated at each iteration.

For a given m , consider the Jacobian as defined in (3.8), namely $J_m = \mathbf{e}_m(t)^T B$. With this notation,

$$J_m^T J_m = B^T \mathbf{e}_m(t) \diamond \mathbf{e}_m(t)^T B = B^T M B,$$

where, for all $i, j = 1, \dots, m$,

$$\begin{aligned} M &= \begin{bmatrix} (e^{-t\theta_i} \diamond e^{-t\theta_j})_{i,j} & (e^{-t\theta_i} \diamond te^{-t\theta_j})_{i,j} \\ (te^{-t\theta_i} \diamond e^{-t\theta_j})_{i,j} & (te^{-t\theta_i} \diamond te^{-t\theta_j})_{i,j} \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{1}{\theta_i + \theta_j}\right)_{i,j} & \left(\frac{1}{(\theta_i + \theta_j)^2}\right)_{i,j} \\ \left(\frac{1}{(\theta_i + \theta_j)^2}\right)_{i,j} & \left(\frac{1}{(\theta_i + \theta_j)^3}\right)_{i,j} \end{bmatrix} \in \mathbb{R}^{2m \times 2m}. \end{aligned}$$

Moreover, for any sufficiently regular function f , we also have

$$J_m^T f = B^T (\mathbf{e}_m(t) \diamond f) = B^T [(e^{-t\theta_i} \diamond f)_i, (te^{-t\theta_i} \diamond f)_i]^T \in \mathbb{R}^{2m}.$$

With this notation, the projected Gauss-Newton iteration (3.2) for our problem is given by

$$r_{j+1} = r_j - S(S^T J_m^T J_m S)^{-1} S^T J_m^T (y(\cdot, r_j) - d), \quad (4.1)$$

where we are assuming here that $J_m^T (y(\cdot, r_j) - d)$ is well defined, and that $(J_m^T J_m)^{-1}$ exists.

Let m_* be the degree of the minimal polynomial of A with respect to b , i.e., the monic polynomial p of lowest degree such that $p(A)b = 0$.

PROPOSITION 4.1. *For $i = 1, \dots, m_*$, let (θ_i, z_i) be the Ritz pairs of A in \mathcal{U}_{m_*} , and let $c_i = (b^T z_i)^2$. Moreover, let $\delta A = D^T \text{diag}(\delta r) D$. Then*

$$\delta c_i = -2z_i^T b b^T (A - \theta_i I)^\dagger \delta A z_i, \quad \delta \theta_i = (D z_i)^T \text{diag}(\delta r) D z_i, \quad i = 1, \dots, m_*, \quad (4.2)$$

where $(A - \theta_i I)^\dagger$ denotes the pseudo-inverse of $(A - \theta_i I)$.

Proof. First order perturbation for the eigenvalues of the symmetric matrix A gives $\delta \theta_i = z_i^T \delta A z_i$ (cf. [11, sec.7.2.2]), from which the expression for $\delta \theta_i$ follows. For the first formula, we recall (cf. [11, sec.7.2.2]) that $(A - \theta_i I) \delta z_i = \delta \theta_i z_i - \delta A z_i$, so that

$$\delta z_i = \delta \theta_i (A - \theta_i I)^\dagger z_i - (A - \theta_i I)^\dagger \delta A z_i = -(A - \theta_i I)^\dagger \delta A z_i.$$

Since $\delta c_i = \delta (z_i^T b)^2 = 2z_i^T b \delta z_i^T b$, the result follows. \square

With the explicit form of the two derivatives, we can derive an expression for the matrix B implicitly introduced in (3.7).

THEOREM 4.2. *With the previous notation, let $Q_{m_*} = [D z_1 z_1^T b, \dots, D z_m z_m^T b]$, and $W_{m_*} = D[(A - \theta_1 I)^\dagger b, \dots, (A - \theta_{m_*} I)^\dagger b]$. Then*

$$B_{m_*} = - \begin{bmatrix} 2(D z_1 z_1^T b \circ D(A - \theta_1 I)^\dagger b)^T \\ \vdots \\ 2(D z_{m_*} z_{m_*}^T b \circ D(A - \theta_{m_*} I)^\dagger b)^T \\ (D z_1 z_1^T b \circ D z_1 z_1^T b)^T \\ \vdots \\ (D z_m z_m^T b \circ D z_m z_m^T b)^T \end{bmatrix} = -[2Q_{m_*} \circ W_{m_*}, Q_{m_*} \circ Q_{m_*}]^T,$$

where \circ denotes the (element-wise) Hadamard product.

Proof. We recall that for any conforming vectors x, y , it holds that $x^T \text{diag}(\delta r) y = [x \circ y]^T \delta r$. Since B_{m_*} must satisfy $[\delta c_{1:m_*}; -c_i \delta \theta_{1:m_*}] = B_{m_*} \delta r$, the expression for each row of B_{m_*} follows using Proposition 4.1. \square

REMARK 4.3. *The simple structure for B_{m_*} highlighted in Theorem 4.2 makes us wonder, if at least for some cases of square B_{m_*} its inverse can be also written in explicit form. For more general tridiagonal matrices A , with more than n parameters, this question was answered positively in [6].*

Since m_* may be large, the computation of the true Jacobian becomes expensive. Assume next that the approximation space \mathcal{U}_m has size $m < m_*$, and let (θ_i, y_i) , $i = 1, \dots, m$ be the eigenpairs of $H_m = U_m^T A U_m$, so that $z_i = V_m y_i$ are the corresponding Ritz vectors; let also $\Theta_m = \text{diag}(\theta_1, \dots, \theta_m)$. The derivative of y_m is

$$\delta y_m = -b^T U_m t(\delta \Theta_m) \exp(-t\Theta_m) U_m^T b + 2b^T U_m Y_m \exp(-t\Theta_m) \delta(U_m Y_m)^T b.$$

We then approximate $\delta(U_m Y_m) \approx U_m \delta Y_m$, that is, we consider the perturbation in the basis negligible. This is clearly a strong assumption, however it allows us to keep the matrix simple and cheap to compute. Under these conditions, and following the derivation of Theorem 4.2, we obtain

$$\delta y_m \approx -b^T U_m t(\delta \Theta_m) \exp(-t \Theta_m) U_m^T b + 2b^T U_m Y_m \exp(-t \Theta_m) (\delta Y_m)^T U_m^T b,$$

that is,

$$\delta y_m \approx -\mathbf{e}_m(t)^T [2Q_m \circ W_m, Q_m \circ Q_m]^T \delta r,$$

where now, $Q_m = [Dz_1(z_1^T b), \dots, Dz_m(z_m^T b)]$, and

$$W_m = DU_m[(H_m - \theta_1 I)^\dagger U_m^T b, \dots, (H_m - \theta_m I)^\dagger U_m^T b].$$

Note that for $m \ll m_*$, the matrix B_m may have much fewer rows than columns, as $2m$ may be smaller than the dimension of the space of admissible solutions. In the experiments of section 6, however, the space \mathcal{S} is very small, and $2m$ is usually larger than the dimension of \mathcal{S} .

5. The algorithm. We are ready to summarize the core steps of our implementation of the Gauss-Newton iteration (4.1):

Algorithm: Given r_0, A, b, d, S, m

for $j = 0, 1, \dots$

1. Compute orthonormal basis U_m of approximation space for $A(r_j)$
2. Compute $H_m = U_m^T A(r_j) U_m$ and Ritz pairs (θ_i, z_i)
3. Compute matrix M_m , and matrix B_m
4. Compute projected matrix $G_m = S^T B_m^T M_m B_m S$
5. Update $r_{j+1} = r_j - S G_m^{-1} (S B_m^T \mathbf{e}_m \diamond (y_m(\cdot, r_j) - d))$
6. If converged, then stop

Since y_m is known, the operation $\mathbf{e}_m \diamond y_m(\cdot, r_j)$ could be performed exactly as follows:

$$\mathbf{e}_m \diamond y_m(\cdot, r_j) = \begin{bmatrix} b^T U_m (H_m + \theta_1 I)^{-1} U_m^T b \\ \vdots \\ b^T U_m (H_m + \theta_m I)^{-1} U_m^T b \\ b^T U_m (H_m + \theta_1 I)^{-2} U_m^T b \\ \vdots \\ b^T U_m (H_m + \theta_m I)^{-2} U_m^T b \end{bmatrix}.$$

However, since similar computation cannot be performed with the observed data d , the whole operation $(\mathbf{e}_m \diamond y_m(\cdot, r_j) - \mathbf{e}_m \diamond d)$ is computed by means of a quadrature formula, at the same time nodes considered for the other numerical quadratures.

If m is sufficiently small, then G_m is nonsingular and sufficiently well-conditioned, hence no regularization is required to perform step 5.

The first step in the algorithm above is the most crucial one within the projected Gauss-Newton iteration. If an H_2 -optimal space can be determined, this provides a sufficient condition for determining the sought after resistivity (cf. Theorem 3.2), at least in the ideal case when B_m is computed exactly. It turns out that condition (3.13) required by Theorem 3.2 is satisfied if the projection space is obtained as the

m	GN it	$\ y - d\ _{L_2}$	$\frac{\ r - r^{true}\ }{\ r^{true}\ }$	$\ S^T J_m^T(y - d)\ $	IRKA its
2	1	3.4579e-02	8.9086e-01	2.2647e-05	19
	2	1.3661e-02	5.6792e-01	3.3775e-06	17
	3	3.8083e-03	2.2173e-01	4.6851e-07	18
	4	5.7437e-04	2.9047e-02	3.9068e-08	17
	5	3.8361e-04	3.4636e-04	4.3727e-10	16
	6	3.8358e-04	1.1656e-05	1.4727e-11	13
	7	3.8358e-04	3.8918e-07	4.9170e-13	10
	8	3.8358e-04	1.2994e-08	1.6421e-14	7
	9	3.8358e-04	4.3747e-10	5.4858e-16	5
8	1	3.4546e-02	8.9086e-01	2.2782e-05	55
	2	1.3794e-02	5.7161e-01	3.5125e-06	47
	3	4.0075e-03	2.3256e-01	5.2030e-07	43
	4	5.6615e-04	3.8238e-02	5.4269e-08	41
	5	1.4880e-05	1.0319e-03	1.3522e-09	37
	6	2.9944e-07	6.5004e-06	1.0172e-12	30
	7	2.9923e-07	1.7509e-11	2.5424e-17	15

TABLE 6.1

Experiments for Example 6.1. Typical convergence history of Gauss-Newton iteration. Dimension of admissible solutions space: $\dim(\mathcal{S})=k=1$.

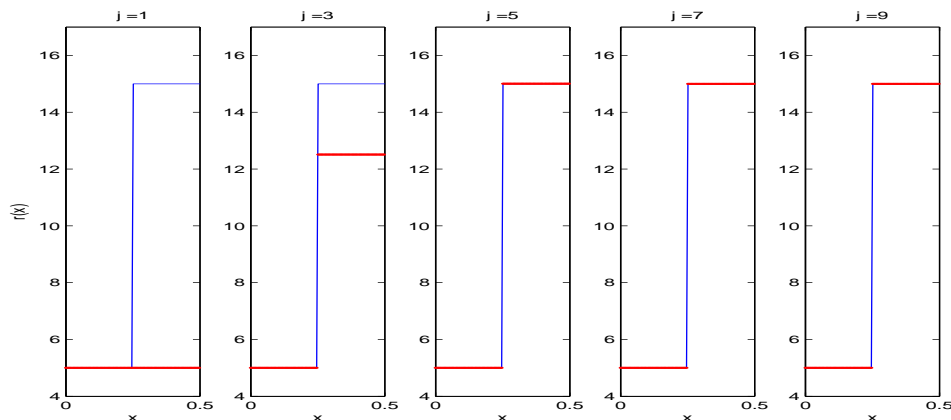


FIG. 6.1. Experiments for Example 6.1. Typical convergence history of $r_j(x)$, $x \in [0, \frac{1}{2}]$ during Gauss-Newton iteration. Dimension of admissible solutions space $\dim(\mathcal{S})=k=1$. r^{true} is piecewise constant, with value $r^{true}(x) = 15$ for $x \in [\frac{1}{4}, \frac{1}{2}]$ and five elsewhere. Projection space dimension $m=2$.

EXAMPLE 6.1. In this example we consider $k = \dim(\mathcal{S}) = 1$, corresponding to standard parametric inversion. Assuming that the problem is defined in the spatial interval $[0, \frac{1}{2}]$, we consider the case where \mathcal{S} represents piecewise constant functions, with *known* (fixed) $\rho = 5$ for $x \in [0, \frac{1}{4}]$ and unknown constant resistivity for $x \in]\frac{1}{4}, \frac{1}{2}]$ set to be 15 (cf. the continuous curve in the plots of Figure 6.1).

The performance of our method is reported in Table 6.1 for $m = 2$ and $m = 8$ (shown in the first column). The subsequent columns show: the Gauss-Newton iteration number, the L_2 norm of the misfit ($\|y - d\|_{L_2}$), the error in the resistivity, the norm of the misfit Jacobian, and the number of IRKA iterations to obtain an

TABLE 6.2

Experiments for Example 6.1. Comparison of Gauss-Newton performance using different projection spaces. Dimension of admissible solutions space $k = 1$.

m	G-N its	$\ y_j - d\ _{L_2}$	$\frac{\ r_j - r^{true}\ }{\ r^{true}\ }$	$\ S^T J_m^T(y_j - d)\ $	method
2	10	3.8358e-04	1.1236e-11	1.7560e-17	IRKA
	12	2.3256e-03	2.0976e-02	1.9325e-17	RKSM
	12	2.3256e-03	2.0975e-02	1.9082e-17	Adapt RKSM
4	8	2.2334e-05	4.1868e-10	5.5433e-16	IRKA
	8	2.1254e-04	2.0927e-03	9.2517e-16	RKSM
	8	3.2202e-04	3.0019e-03	1.4546e-16	Adapt RKSM
6	8	2.4928e-06	6.4802e-12	1.2520e-17	IRKA
	8	1.2513e-05	1.5614e-05	1.3636e-17	RKSM
	8	2.8916e-05	5.6807e-06	7.2642e-18	Adapt RKSM
8	7	2.9923e-07	1.7509e-11	2.5424e-17	IRKA
	7	8.8486e-07	1.7976e-08	1.1180e-17	RKSM
	8	3.9417e-06	1.1939e-06	5.7947e-18	Adapt RKSM

H_2 -optimal space at each Gauss-Newton iteration (a stopping tolerance of 10^{-12} was used in IRKA). The digits show that even for an extremely tiny reduced order model, a good accuracy of r^{true} is obtained after few Gauss-Newton iterations. A larger dimension m of the projection space clearly provides faster convergence of the Gauss-Newton iteration, with a final more accurate reproduction of the data, corresponding to a smaller minimum. On the other hand, we observe that IRKA becomes more expensive, as the total number of iterations is significantly higher, and the cost of each IRKA iteration is correspondingly larger. We mention that the previous optimal set of poles was chosen to initialize the current IRKA iteration; this explains the lower number of IRKA iterations as the Gauss-Newton recurrence proceeds (last column of Table 6.1).

Note that after $\|S^T J_m^T(y - d)\|$ has reached machine precision, further Gauss-Newton iterations will no longer improve the current approximate solution. In all experiments we have thus decided to stop the iteration when this quantity becomes sufficiently small. This strategy may also be adopted as a stopping criterion.

In Table 6.2 we report comparisons when the method employs an H_2 non-optimal rational Krylov subspace for the projection phase. In particular, we considered the standard Rational Krylov subspace with Zolotarev nodes², so as to ensure no adaptation to the spectrum [8], and the adaptive Rational Krylov subspace, recently introduced in [9]. In both cases, a rough estimate of the spectral interval of $A(r_0)$ was used to provide initial spectral information. In this table, only the total number of Gauss-Newton iterations is reported for each method and for each choice of m . We observe that especially for small m , only the optimal projection space is capable of achieving a good approximation to the sought after resistivity. On the other hand, the cost of IRKA is far higher than that of the other methods, and thus the approaches are not comparable in terms of computational costs.

EXAMPLE 6.2. With the same data as in the previous example, we assume here that the observed data is perturbed by noise, namely $d(t) = y(r^{true}, t) + \Delta$, $t \geq 0$. The components of the vector Δ were selected as a random normally distributed

²A different sequence of nodes was precomputed for each value of m .

TABLE 6.3

Experiments for Example 6.2. Dimension of admissible solutions space $k = 1$. Noisy data: $d(t) = y(r^{true}, t) + \Delta(t) \quad t \geq 0, \|y(r^{true}, t)\|_{L_2} \approx 0.15$.

m	$\frac{\ \Delta\ _{L_2}}{\ y^{true}\ _{L_2}}$	GN it	$\ y_m - d\ _{L_2}$	$\frac{\ r_j - r^{true}\ }{\ r^{true}\ }$	$\ (J_m S)^T(y_m - d)\ $	IRKA its
2	6e-3	1	3.4587e-02	8.9086e-01	2.2643e-05	19
		2	1.3695e-02	5.6798e-01	3.3774e-06	17
		3	3.9375e-03	2.2191e-01	4.6862e-07	18
		4	1.1555e-03	2.9305e-02	3.9113e-08	17
		5	1.0737e-03	1.0043e-04	4.3192e-10	16
		6	1.0737e-03	2.5296e-04	1.4482e-11	12
		7	1.0737e-03	2.4112e-04	4.8139e-13	10
2	6e-5	1	3.4579e-02	8.9086e-01	2.2647e-05	19
		2	1.3661e-02	5.6792e-01	3.3775e-06	17
		3	3.8083e-03	2.2173e-01	4.6851e-07	18
		4	5.7450e-04	2.9050e-02	3.9068e-08	17
		5	3.8381e-04	3.4390e-04	4.3722e-10	16
		6	3.8378e-04	1.4069e-05	1.4724e-11	12
		7	3.8378e-04	2.0260e-06	4.9159e-13	10
8	6e-3	1	3.4555e-02	8.9086e-01	2.2778e-05	55
		2	1.3828e-02	5.7168e-01	3.5122e-06	47
		3	4.1300e-03	2.3275e-01	5.2038e-07	43
		4	1.1492e-03	3.8520e-02	5.4309e-08	41
		5	1.0001e-03	1.3164e-03	1.3600e-09	37
		6	9.9999e-04	2.8009e-04	1.2094e-12	30
		7	9.9999e-04	2.7916e-04	1.9344e-16	16
8	6e-5	1	3.4546e-02	8.9086e-01	2.2782e-05	55
		2	1.3794e-02	5.7161e-01	3.5125e-06	47
		3	4.0075e-03	2.3256e-01	5.2031e-07	43
		4	5.6625e-04	3.8241e-02	5.4269e-08	41
		5	1.7927e-05	1.0347e-03	1.3523e-09	37
		6	1.0001e-05	3.5713e-06	1.0191e-12	30
		7	1.0001e-05	2.7919e-06	3.0108e-17	15

perturbation of the corresponding components of $y(r^{true}, t)$, so that $\|\Delta\|_{L_2} = \varepsilon$, and we chose $\varepsilon = 10^{-5}, 10^{-3}$; Note that $\|y(r^{true}, t)\|_{L_2} \approx 0.15$. The performance of the new method for $m = 2, 8$ is reported in Table 6.3. The iteration progresses up to the final attainable accuracy of the minimized function, which corresponds to $\|\Delta\|_{L_2}$, as expected. We also observe that the method is able to obtain an estimate of the true resistivity which is correct at least in the first two digits.

EXAMPLE 6.3. For the same original problem data as in Example 6.1, we next consider a space of admissible solutions of size $k = 5$, corresponding to the subspace of dimension k associated with piecewise constant functions (cf. the continuous curve in the plot of Figure 6.2). This is a somewhat contrived example, as the contrast is rather limited; larger contrasts would require the use of a more general optimization procedure with regularization and line search.

The H_2 -optimal space of size m was generated using IRKA with a stopping tolerance of 10^{-12} , as in the previous examples. Table 6.4 reports the performance of the proposed method for $k = 5$ and various value of m . In all cases except for $m = 2$, the method achieved a rather accurate value of the resistivity, while the optimization function is increasingly better minimized as the projection space increases,

TABLE 6.4

Experiments for Example 6.3. Accuracy of Gauss-Newton iteration as m (projection space dimension) varies. Dimension of admissible solutions space: $k = 5$.

m	GN its	$\ y - d\ _{L_2}$	$\frac{\ r - r^{true}\ }{\ r^{true}\ }$	$\ S^T J_m^T(y - d)\ $
2	15	7.4295e-04	3.7106e-02	1.2027e-10
4	15	3.9362e-05	2.1294e-06	3.0095e-14
6	10	4.1931e-06	7.1387e-08	2.0746e-16
8	10	5.6193e-07	5.5721e-09	6.7678e-16
10	7	7.9802e-08	1.0529e-09	1.4269e-17
12	7	1.0891e-08	2.0363e-09	1.5794e-16
14	10	1.4165e-09	4.5174e-10	4.8247e-16

as expected. For $m = 2$, the Jacobian is not full rank, since $2m < k$, therefore we do not expect our method to converge. A full history of convergence for $m = 6, 12$ is reported in Figure 6.3; the plots show that a larger m seems to allow for a faster convergence of the procedure, as a better approximation to the Jacobian is obtained. Finally, in Figure 6.4 we report the behavior of the GN method as a function of the projection space dimension m , in terms of final attained error in the misfit and in the resistivity function. Results with both IRKA and RKSM are displayed, confirming the better behavior of IRKA for small m in the resistivity function approximation.

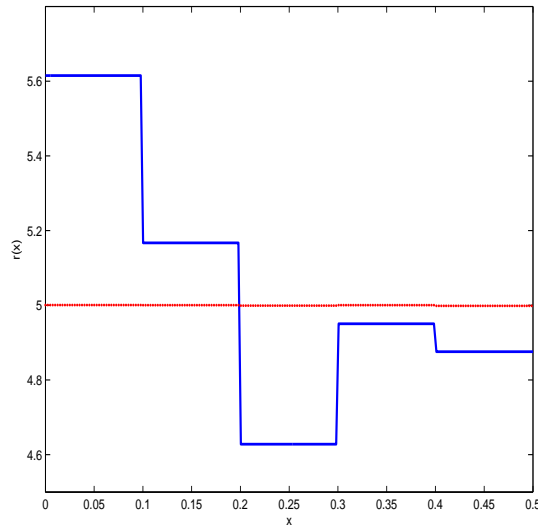


FIG. 6.2. Experiments for Example 6.3. Exact function r^{true} (piecewise constant, continuous function in the plot), and starting approximation equal to five, for $x \in [0, \frac{1}{2}]$. Dimension of admissible solutions space $\dim(S)=5$.

EXAMPLE 6.4. Next we consider a more general example, where the data vector d is not generated as a perturbation of the true forward problem solution, but rather as the true solution of the same model problem with a finer discretization. We do this to test the robustness of our approach, so as to ensure that no *inverse crime* is performed in our algorithm; see, e.g., [7], [15] for a discussion on the inverse crime concept. For an admissible solutions space of size $k = 1$ of piecewise constants, we collect $d_*(t_i)$ as

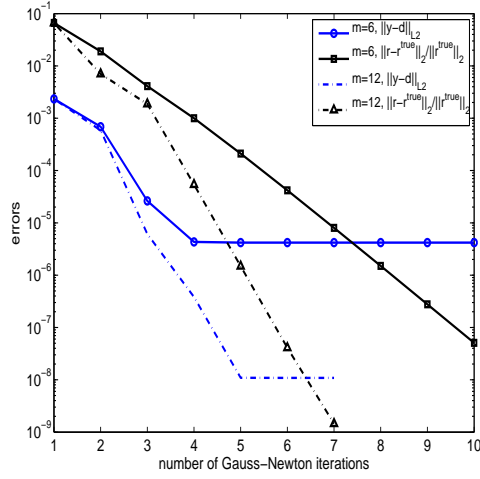


FIG. 6.3. Experiments for Example 6.3. Behavior of the Gauss-Newton iteration for $m = 6, 12$. Dimension of admissible solutions set $\dim(S)=5$.

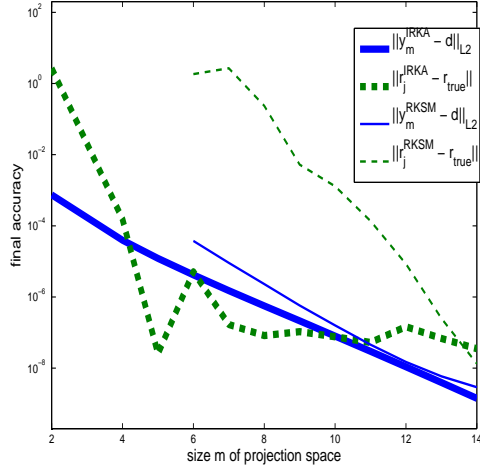


FIG. 6.4. Experiments for Example 6.3. Behavior of the Gauss-Newton iteration as a function of the projection space dimension. Dimension of admissible solutions set $\dim(S)=5$. Results with projection spaces from IRKA and RSKM are shown.

$d_*(t_i) = \exp(t_i A(r^{true}))b$, where $t_i \in [10^{-8}, 10^8]$ as before, and A is the 1600×1600 matrix obtained with a fine discretization in (2.2). We then approximate r^{true} as in the data of Example 6.1, with a space discretization of 200 nodes, by minimizing $\|y_m - d_*\|_{L_2}$. We observe that for these data and $n = 200$,

$$\|y^{true} - d_*\|_{L_2} = 6.6407e - 04, \quad \frac{\|y^{true} - d_*\|_{L_2}}{\|y^{true}\|_{L_2}} = 1.1852e - 03,$$

where y^{true} is determined as $y^{true} = b^T e^{-t_i A(r^{true})} b$ on the 200 nodes grid.

The convergence history of the Gauss-Newton iteration for $m = 2$ is reported in Table 6.5, showing that our approach is sufficiently robust, and it achieves the misfit

TABLE 6.5

Experiments for Example 6.4. Convergence history of Gauss-Newton iteration with data d from a finer discretization ($n = 1600$). Dimension of admissible solutions space: $\dim(\mathcal{S})=k = 1$. Projection space dimension $m = 2$.

GN its	$\ y - d\ _{L_2}$	$\ S^T J_m^T(y - d)\ $	IRKA its
1	3.5178e-02	2.3016e-05	19
2	1.4071e-02	3.4331e-06	17
3	4.0415e-03	4.8069e-07	18
4	6.5445e-04	4.2949e-08	17
5	4.3263e-04	2.6176e-10	16
6	4.3257e-04	8.4806e-12	12
7	4.3257e-04	2.7314e-13	10
8	4.3257e-04	8.7976e-15	7

$\|y^{true} - d_*\|_{L_2}$ in few iterations. Figure 6.5 reports the approximation to r^{true} as the iteration proceeds, with a final relative error of 3%.

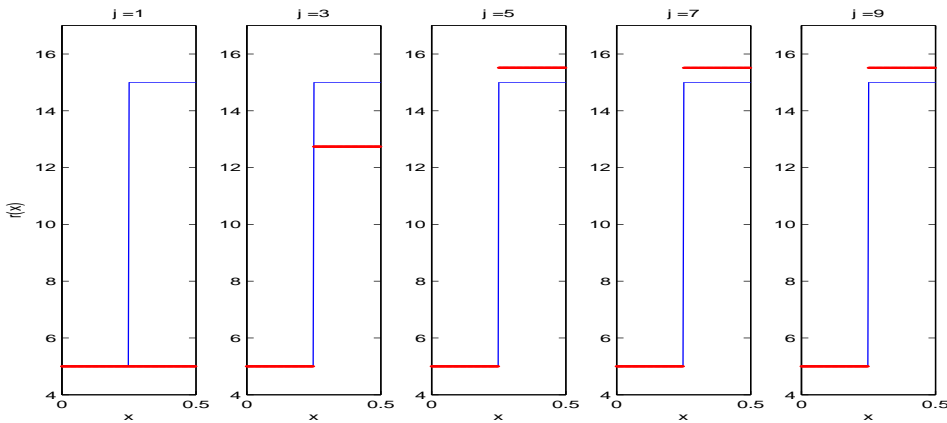


FIG. 6.5. Experiments for Example 6.4. Convergence history of r_j , for $x \in [0, \frac{1}{2}]$ during Gauss-Newton iteration. The final relative error in r is 3%. Dimension of admissible solutions space $\dim(\mathcal{S})=1$. r^{true} is piecewise constant (continuous function in the plot). Projection space dimension $m = 2$.

7. Conclusions. In this paper we have outlined the path towards the implementation of a model reduction approach for evolutionary problems such as the diffusion Maxwell equation, of great relevance in geophysical exploration. Our numerical experiments show that the idea of approximating the Jacobian of the forward problem solution in a rich though very small approximation space is very promising, coping with many of the difficulties typical of inverse problems. Although this paper has necessarily focused on the 1D SISO case, we believe that substantially the same approximation strategy, with the Gauss-Newton iteration enriched by regularization and line search, can be employed to attack the far more challenging 3D MIMO case. Very early numerical experiments not reported here seem to confirm our expectations; a full report will be the topic of Part II of this project.

Our experiments show significant dimensionality reduction of the Jacobian using cheap *non-optimal* (more precisely, approximately optimized using Zolotarev or

adaptive shift selection strategies) subspaces, however, the most drastic reduction is achieved using the H_2 optimal subspaces. A crucial element for the overall computational efficiency of the latter approach is the IRKA algorithm, iteratively generating the H_2 optimal subspaces. Unfortunately, it presently requires the solution of a large number of auxiliary linear systems with the operator $A(r)$ for a given r . Albeit limited, our numerical experience seems to show that if these solves become overwhelming, non-optimal rational Krylov subspaces may be more advantageous. However, we believe that a fine tuning of the method, one can significantly improve IRKA's performance in the framework of the Gauss-Newton algorithm, and using H_2 optimal subspaces will eventually result in a winning strategy.

Finally, we should point out that the approach of this paper is not limited to the diffusion Maxwell system, and can be extended to other inverse problems for time-invariant dynamic system, e.g., the inverse hyperbolic (seismic) problem.

Acknowledgments. The work of the second author was partially funded by the Project PRIN 2008 N. 20083KLJEZ. We thank Liliana Borcea, Garret Flagg, Tarek Habashy, Aria Abubakar and Alexander Mamonov for very valuable discussions.

REFERENCES

- [1] A.L. Levshin A.A. Kaufman. Acoustic and elastic wave fields in geophysics, part i. In *Methods in Geochemistry and Geophysics*, volume 52. Elsevier, 2000.
- [2] A. Abubakar, T. M. Habashy, V. L. Druskin, L. Knizhnerman, and D. Alumbaugh. 2.5d forward and inverse modeling for interpreting low-frequency electromagnetic measurements. *Geophysics*, 73:F165, 2008.
- [3] A.C. Antoulas, C.A. Beattie, and S. Gugercin. Interpolatory model reduction of large-scale dynamical systems. In J. Mohammadpour and K. Grigoriadis, editors, *Efficient Modeling and Control of Large-Scale Systems*, J. Mohammadpour and K. Grigoriadis editors, Springer-Verlag, IS BN 978-1-4419-5756-6, Publication date: Feb. 2010. Springer-Verlag, February 2010. ISBN 978-1-4419-5756-6.
- [4] Ulrike Baur, Christopher Beattie, Peter Benner, and Serkan Gugercin. Interpolatory Projection Methods for Parameterized Model Reduction. Chemnitz scientific computing preprints, Technische Universität Chemnitz, 2009. To appear in SIAM J. Scient. Comput.
- [5] Ralph-Uwe Boerner, Oliver G. Ernst, and Klaus Spitzer. Fast 3-D simulation of transient electromagnetic fields by model reduction in the frequency domain using Krylov subspace projection. *Geophys. J. Int.*, 173:766–780, 2008.
- [6] L. Borcea, V. Druskin, and L. Knizhnerman. On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids. *Comm. Pure Appl. Math.*, 58(9):12–31, 2005.
- [7] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, Berlin, 1992.
- [8] V. Druskin, L. Knizhnerman, and M. Zaslavsky. Solution of large scale evolutionary problems using rational Krylov subspaces with optimized shifts. *SIAM J. Sci. Comput.*, 31(5):3760–3780, 2009.
- [9] V. Druskin, C. Lieberman, and M. Zaslavsky. On adaptive choice of shifts in rational krylov subspace reduction of evolutionary problems. *SIAM J. Sci. Comput.*, 32(5):2485–2496, 2010.
- [10] G. Flagg, C.A. Beattie, and S. Gugercin. Convergence of the Iterative Rational Krylov Algorithm. Technical report, Virginia Tech, 2011. Available as arXiv:1107.5363v1.
- [11] G. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.
- [12] E. Grimme. *Krylov projection methods for model reduction*. PhD thesis, The University of Illinois at Urbana-Champaign, 1997.
- [13] S. Gugercin, A. C. Antoulas, and C. Beattie. H_2 model reduction for large-scale linear dynamical systems. *SIAM J. Matrix Anal. Appl.*, 30:609–638, 2008.
- [14] E. Haber, D. Oldenburg, and R. Shekhtman. Inversion of 3D time domain electromagnetic data. *Geophys. J. Int.*, 171:550–564, 2004.

- [15] Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.
- [16] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer, New York, 1999.
- [17] C. R. Vogel and J. G. Wades. Analysis of costate discretizations in parameter estimation for linear evolution equations. *SIAM J. Control and Optimization*, 33(1):227–254, January 1995.
- [18] Curtis R. Vogel. Sparse matrix computations arising in distributed parameter identification. *SIAM J. Matrix Anal. Appl.*, 20:1027–1037, 1999.
- [19] Curtis R. Vogel. *Computational Methods for Inverse Problems*. Frontiers in Applied Mathematics. SIAM, Philadelphia, 2002.
- [20] M.S. Zhdanov. Geophysical electromagnetic theory and methods. In *Methods in Geochemistry and Geophysics*, volume 43. Elsevier, 2009.