

A Truncated Levenberg-Marquardt Algorithm for the Calibration of Highly Parameterized Nonlinear Models

Stefan Finsterle¹ and Michael B. Kowalsky

Lawrence Berkeley National Laboratory, Earth Sciences Division

One Cyclotron Road, MS 90-1116, Berkeley, California, SAFinsterle@lbl.gov

Abstract

We propose a modification to the Levenberg-Marquardt minimization algorithm for a more robust and more efficient calibration of highly parameterized, strongly nonlinear models of multiphase flow through porous media. The new method combines the advantages of truncated singular value decomposition with those of the classical Levenberg-Marquardt algorithm, thus enabling a more robust solution of underdetermined inverse problems with complex relations between the parameters to be estimated and the observable state variables used for calibration. The truncation limit separating the solution space from the calibration null space is re-evaluated during the iterative calibration process. In between these re-evaluations, fewer forward simulations are required, compared to the standard approach, to calculate the approximate sensitivity matrix. Truncated singular values are used to calculate the Levenberg-Marquardt parameter updates, ensuring that safe small steps along the steepest-descent direction are taken for highly correlated parameters of low sensitivity, whereas efficient quasi-Gauss-Newton steps are taken for independent parameters with high impact. The performance of

¹ Corresponding author; SAFinsterle@lbl.gov, Earth Sciences Division, 1 Cyclotron Road, MS 90-1116, Berkeley, CA 94720; phone: (510) 486-5205; fax: (510) 486-5686

the proposed scheme is demonstrated for a synthetic data set representing infiltration into a partially saturated, heterogeneous soil, where hydrogeological, petrophysical, and geostatistical parameters are estimated based on the joint inversion of hydrological and geophysical data.

Keywords

multiphase flow; inverse modeling; Levenberg-Marquardt algorithm; parameterization; iTOUGH2

1. Introduction

Calibrating a hydrologic model against selected observations of the system state is a standard method to estimate model parameters, which are then assumed to reflect either site-specific material properties or physical processes that are represented by effective coefficients. The calibration process is an integral part of model development (Carrera et al., 2005), and involves the following steps: (1) selection of calibration data and assessment of their error structure; (2) development of a conceptual, mathematical, and numerical model capable of simulating the system behavior of interest under measurement conditions; (3) definition of parameters to be estimated (parameterization); (4) formulation of an objective function as a measure of misfit between model prediction and measured data; (5) selection of an optimization algorithm to identify the minimum of the objective function; and (6) assessment of model appropriateness, parameter identifiability, estimation error, and prediction uncertainty. All these steps are interlinked. For example, the experimental design and the type of observed data available for

calibration affect the model to be developed. The data and model determine the criteria for a suitable parameterization. The expected error structure of both the data and model affect the objective function, the topology of which then influences the choice of an effective minimization algorithm.

Despite its universal appeal and wide use in science and engineering, the formulation and solution of the inverse problem is inherently difficult. Underlying assumptions are not always made consistent, and related decisions are seldom explicitly discussed. As a result, claims about the nature of the estimates and associated uncertainties often remain without a theoretical basis or practical justification. This is especially true for strongly nonlinear models, models that are either over- or underparameterized, or data that exhibit non-normal, correlated, or systematic errors.

Minimization algorithms for the solution of nonlinear inverse problems have received considerable attention in the scientific community for a long time, starting with Carl Friedrich Gauss, who introduced the method of least squares for the analysis of astronomical and geodetic data during the last decade of the eighteenth century (Gauss, 1821). Gauss made significant contributions to all aspects of parameter estimation, providing a detailed discussion of measurement errors, a probabilistic justification of the sum of squares as the objective function, a method to minimize this function, and an analysis of estimation uncertainty. While Gauss considered his method suitable also for non-normal errors, and his algorithm can be used (iteratively) for weakly nonlinear inverse problems, it has been continuously refined to address strongly nonlinear models. For example, the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) combines the robustness of the steepest-descent method with the quadratic convergence

rate of the Gauss-Newton method to effectively identify the minimum of a convex objective function. More recently, a great number of mostly heuristic minimization algorithms have been proposed to handle high-dimensional objective functions with multiple local minima. While computationally very demanding, they have also been applied to inverse problems in hydrology (for summary discussions, see, for example, Regis and Shoemaker, 2004; Vrugt et al., 2009).

The sole purpose of all these methods is to find the (preferably global) minimum of the objective function given a forward model and a finite, predefined set of adjustable, potentially constrained parameters. It is important to realize that the topology of the objective function to be minimized is fixed. It is predetermined by (1) the measured data values, (2) the functional form of the forward model, (3) the chosen parameterization, and (4) the functional form of the objective function itself. The most often chosen objective function is the sum of squared weighted residuals. The objective functions of so-called ill-posed inverse problems have poorly defined minima as a result of insufficient parameter sensitivity or strong parameter correlations, which are often an indication that the problem is over-parameterized.

The convexity of the objective function and thus the well-posedness and stability of an inverse problem is often improved by regularization, either by including prior information about the parameters to be estimated (Carrera and Neuman, 1986), or by smoothness or homogeneity conditions—often referred to as Tikhonov regularization (Tikhonov and Arsenin, 1977)—which is standard practice in geophysical imaging. By adjusting the trade-off parameter that assigns relative weights to data matching and regularization, any inverse problem can be stabilized and a unique solution found.

However, obtaining this solution comes at the expense of losing information contained in the observed data, and may lead to biased estimates. For example, if there is a conceptual difference between the data used as prior information and the corresponding model-related parameter to be estimated by inverse modeling, the latter will be biased. Such conceptual differences include scale effects (e.g., if core-scale permeability measurements are used as prior information in a regional-scale groundwater model), or any discrepancy in the features or processes that need to be captured in an effective model parameter. Similarly, smoothness constraints may inappropriately homogenize stratigraphic discontinuities in the field (Moore and Doherty, 2006), which is then compensated for by adjusting hydrogeologic parameter values. It is therefore essential to remove inherent inconsistencies between the data matching and regularization terms and to appropriately weight them, before a formal inversion is attempted. In most cases, this precludes the dynamic adjustment of the trade-off parameter as part of the optimization procedure to increase the stability of the inversion. Nevertheless, if the regularization scheme and trade-off parameters are carefully selected (Alcolea et al., 2006), regularization fulfills its purpose of introducing additional information and of stabilizing the inversion.

Monte-Carlo based methods are alternative approaches for the calibration of highly parameterized models; an overview is given in Hendricks Franssen et al. (2009). The main advantage of these methods is that they examine the topology of the objective function over an extended region in the parameter space, and thus provide insights into the uniqueness and accuracy of the solution. Many of these sampling-based methods, however, require a substantial amount of forward simulations, which is often impractical

if the model incorporates coupled physical behavior and complex features, as does the one presented here. We therefore focus on a derivative-based method and examine its suitability for solving a highly parameterized inverse problem when a forward simulation is computationally expensive.

In what follows, we attempt to solve potentially an ill-posed inverse problem without changing the topology of the objective function. Nevertheless, in certain applications, the objective function may contain well-justified contributions from regularization, which results in a formulation similar to the hybrid regularization methodology described by Tonkin and Doherty (2005). The problem we address is simpler in that it is only concerned with the minimization algorithm rather than the formulation of the inverse problem itself.

We first present the standard Levenberg-Marquardt minimization algorithm, before modifying it to become a more robust and more efficient subspace method. The advantages of the proposed approach will be demonstrated using a synthetic hydrogeophysical inverse problem.

2. Truncated Levenberg-Marquardt Algorithm

Parameter estimation by automatic calibration of a numerical model is an iterative process in which a global measure of misfit between observed data and the corresponding model output is successively reduced by updating parameters based on the value, gradient, or curvature of the objective function. We are interested in inverse problems where the relationship between the n model parameters of interest (vector \mathbf{p}) and the m

model outputs (vector \mathbf{z}) is highly nonlinear and evaluated by a computationally expensive forward model. We limit the discussion here to the minimization of the standard, weighted least-squares objective function

$$S = (\mathbf{z}^* - \mathbf{z}(\mathbf{p}))^T \mathbf{C}_{zz}^{-1} (\mathbf{z}^* - \mathbf{z}(\mathbf{p})), \quad (1)$$

even though the Levenberg-Marquardt algorithm is flexible enough to also minimize other objective functions, such as those presented in Finsterle and Najita (1998). In (1), \mathbf{z}^* is the vector holding the measured data corresponding to model output $\mathbf{z}(\mathbf{p})$; the difference between the measured data and the model output is the residual vector \mathbf{r} . The $m \times m$ positive-definite matrix $\mathbf{C}_{zz} = s_0^2 \mathbf{V}_{zz}$ is referred to as the observation covariance matrix. It is often taken as a diagonal matrix with its elements σ_i set to the variance of the measurement error. However, this interpretation assumes that the underlying conceptual model is error free, and that errors are independent. Since the residuals also include modeling errors, the validity of these assumptions is questionable, as demonstrated in Lehtikoinen et al. (2010) and Finsterle and Zhang (2010). From a practical perspective, the purpose of \mathbf{C}_{zz} is to weight measurements of different quality, to scale observations of different types, to make the objective function dimensionless, and to weight the data fitting error relative to regularization terms. These purposes can also be fulfilled by a cofactor matrix \mathbf{V}_{zz} . However, it is desirable to interpret the weighting matrix as the inverse of a covariance matrix of the residuals; the dimensionless scalar s_0^2 can then be used to test whether our prior stochastic assumptions about the residuals are consistent with the actual residuals obtained after model calibration (Finsterle and Pruess,

1995). For normally distributed residuals, \mathbf{C}_{zz} then becomes the stochastic model of the maximum-likelihood estimation framework.

The Levenberg-Marquardt algorithm belongs to a class of methods that are based on a quadratic approximation of the objective function S evaluated at a given point \mathbf{p}_k in the parameter space. The parameter set will be iteratively updated by $\Delta\mathbf{p}_k$ such the objective function $S(\mathbf{p}_{k+1}) < S(\mathbf{p}_k)$. If first and second derivatives of S are available, a quadratic model of the objective function can be obtained from the first three terms of the Taylor-series expansion:

$$S(\mathbf{p}_{k+1}) \approx S(\mathbf{p}_k) + \mathbf{g}_k^T \Delta\mathbf{p}_k + \frac{1}{2} \Delta\mathbf{p}_k^T \mathbf{H}_k \Delta\mathbf{p}_k. \quad (2)$$

Here, \mathbf{g}_k is the gradient vector, and \mathbf{H}_k is the Hessian matrix at iteration k . The minimum of the right-hand side of (2) is achieved if $\Delta\mathbf{p}_k$ minimizes the quadratic function

$$\Phi(\Delta\mathbf{p}) = \mathbf{g}_k^T \Delta\mathbf{p} + \frac{1}{2} \Delta\mathbf{p}^T \mathbf{H}_k \Delta\mathbf{p}. \quad (3)$$

Defining the $m \times n$ Jacobian matrix as

$$\mathbf{J} = -\frac{\partial \mathbf{r}}{\partial \mathbf{p}} = \frac{\partial \mathbf{z}}{\partial \mathbf{p}} = \begin{bmatrix} \frac{\partial z_1}{\partial p_1} & \dots & \frac{\partial z_1}{\partial p_n} \\ \vdots & & \vdots \\ \frac{\partial z_m}{\partial p_1} & \dots & \frac{\partial z_m}{\partial p_n} \end{bmatrix}, \quad (4)$$

the gradient vector at iteration k can be written for the weighted least-squares objective function as

$$\mathbf{g}_k = -2\mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{r}_k. \quad (5)$$

The Hessian is an $n \times n$ matrix with the second partial derivatives of the objective function. For least-squares problems, the Hessian can be written as

$$\mathbf{H}_k = 2 \left(\mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{J}_k + \sum_{i=1}^m (\mathbf{C}_{zz}^{-1/2} \mathbf{r})_i \mathbf{G}_i \right), \quad (6)$$

where $\mathbf{C}_{zz}^{1/2} = \mathbf{V} \mathbf{D}^{1/2} \mathbf{V}^{-1}$, with \mathbf{V} and \mathbf{D} holding, respectively, the eigenvectors and eigenvalues of the weighting matrix \mathbf{C}_{zz}^{-1} , and $\mathbf{G}_i = \left[\frac{\partial^2 r_i}{\partial \mathbf{p} \partial \mathbf{p}^T} \right]$ is the Hessian of the residuals. Denoting the first term in (6) by \mathbf{F} (also referred to as the Fisher information matrix) and second term by \mathbf{B} , it becomes obvious that the Hessian of a least-squares objective function consists of a combination of first- and second-order information:

$$\mathbf{H}_k = 2(\mathbf{F}_k + \mathbf{B}_k). \quad (7)$$

Note that \mathbf{B} is zero if the model is linear. For highly nonlinear models, \mathbf{B} becomes significant if the residuals are large, such as when parameter combinations are far away from the minimum or when data are very noisy. Also note that the positive and negative residuals in \mathbf{B} do not cancel one another, i.e., the Hessian is not necessarily a positive-definite matrix.

At the minimum of (3), $\Delta \mathbf{p}_k$ satisfies the linear system

$$\mathbf{H}_k \Delta \mathbf{p}_k = -\mathbf{g}_k. \quad (8)$$

Combining (5), (7) and (8) yields Newton's method:

$$\Delta \mathbf{p}_k = (\mathbf{F}_k + \mathbf{B}_k)^{-1} \mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{r}_k. \quad (9)$$

The various iterative solutions to the nonlinear least-squares problem are based on different approximations to the Hessian. In particular, the Gauss-Newton method is based on the premise that the first-order term $(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J}) \equiv \mathbf{F}$ of the Hessian dominates the second-order term \mathbf{B} . This assumption is justified for linear and mildly nonlinear problems and for nonlinear problems near the solution, where the residuals are expected to be small, i.e., when the objective function is sufficiently smaller than the eigenvalues of \mathbf{F} . The Gauss-Newton method simply sets \mathbf{B} to zero, ensuring that the Hessian is positive-definite and approximating the actual objective function by a quadratic function. The Gauss-Newton update is given by

$$\Delta \mathbf{p}_k = \left(\mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{J}_k \right)^{-1} \mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{r}_k, \quad (10)$$

which is the solution of the linear least-squares problem. For nonlinear models, the parameter vector is iteratively updated, $\mathbf{p}_{k+1} = \mathbf{p}_k + \Delta \mathbf{p}_k$, where $\Delta \mathbf{p}_k$ is the new Gauss-Newton parameter update vector calculated for each iteration.

For strongly nonlinear models, if the parameter vector \mathbf{p}_k is far away from the optimum parameter set, the Hessian is not necessarily a positive-definite matrix, and the approximation $(\mathbf{J}^T \mathbf{C}_{zz}^{-1} \mathbf{J})$ used by the Gauss-Newton method may not lead to an efficient or successful step. In the Levenberg-Marquardt method, the approximation to the Hessian is made positive definite by replacing \mathbf{B} in (7) with a $n \times n$ diagonal matrix $\lambda_k \mathbf{D}_k$:

$$\Delta \mathbf{p}_k = \left(\mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{J}_k + \lambda_k \mathbf{D}_k \right)^{-1} \mathbf{J}_k^T \mathbf{C}_{zz}^{-1} \mathbf{r}_k \quad (11)$$

The scalar λ is the so-called damping or Levenberg parameter (Levenberg, 1944). If λ_k is zero, $\Delta \mathbf{p}_k$ is identical to the Gauss-Newton step; as $\lambda_k \rightarrow \infty$, the approximation of the

Hessian becomes diagonally dominant. Consequently, $\Delta \mathbf{p}_k$ becomes parallel to the steepest-descent direction, and the step length approaches zero. After each iteration, the Levenberg parameter is either increased or decreased following a scheme proposed by Marquardt (1963). Far away from the minimum, i.e., during the first few iterations, a relatively large value of λ_k is chosen, leading to small steps along the gradient of the objective function. Stepping along the steepest-descent direction is a robust strategy, ensuring that $S(\mathbf{p}_{k+1}) < S(\mathbf{p}_k)$ for sufficiently large λ_k . However, the step length $\|\Delta \mathbf{p}_k\|_2$ may be very small and minimization becomes inefficient. Therefore, λ_k is decreased by a factor of $1/\nu$ after each successful step, where $\nu > 1$ is the so-called Marquardt parameter. With decreasing λ_k , $\Delta \mathbf{p}_k$ approaches the Gauss-Newton step with its quadratic convergence rate. If an unsuccessful step is proposed, i.e., the objective function is increased, λ_k is increased by ν .

Matrix \mathbf{D} in (11) is sometimes referred to as the Tikhonov matrix because its function is to prevent the approximation to the Hessian to become singular, which is similar to the purpose of the regularization matrix discussed above. However, it is important to realize that we are not changing the inverse problem to be solved (as regularization does) by invoking (11); the Tikhonov matrix just changes the trust region and thus step direction and step length within the Levenberg-Marquardt method. We will consider four choices for \mathbf{D} and examine their performance for highly parameterized inverse problems.

The simplest and most often cited choice for \mathbf{D} is the identity matrix, applying uniform damping to each parameter increment. In fact, Levenberg (1944) demonstrated that this choice is optimal at the solution. However, without providing a discussion, he also

suggested a scheme where the diagonal elements of \mathbf{D} are given by the reciprocals of the diagonal elements of \mathbf{F} , improving the stability of \mathbf{H} by assigning the highest damping to the diagonal element of \mathbf{F} with the lowest value. Next, we propose two additional schemes where the diagonal elements of \mathbf{D} are further scaled by the inverse of the eigenvalues of \mathbf{F} , and, as a further modification, we use the eigenvalues of \mathbf{F} to truncate the parameter space, i.e., only parameters associated with a sufficiently large eigenvalue of the Hessian will be updated. This last scheme is motivated by the following considerations.

- (1) The eigenvalues of the Fisher information matrix \mathbf{F} indicate the amount of information parameters contribute to the solution of the inverse problem. Parameter combinations along the eigenvector with the largest eigenvalue change the calculated system behavior $\mathbf{z}(\mathbf{p})$ most dramatically, and can thus most accurately be identified from the measured data.
- (2) Conversely, small eigenvalues result from a lack of sensitivity. The corresponding rows of \mathbf{F} likely lead to instability or even singularity of \mathbf{H} ; they need to be dampened more strongly by adding a larger value to the corresponding diagonal element of \mathbf{B} .
- (3) In addition to parameter sensitivity, the eigenvalues also reflect statistical correlations among the parameters. Two parameters that are strongly correlated (and thus cannot be independently determined from the data) have similar columns in the Jacobian matrix, resulting in a Fisher information matrix with small eigenvalues. Using the proposed scheme, such parameters will be either

- (4) In a strongly nonlinear optimization problem (such as the calibration of a multiphase flow model considered here), parameter sensitivities may vary substantially as the minimization algorithm proceeds through the parameter space. Far away from the minimum, certain parameters may have very small or even zero sensitivity coefficients, whereas the same parameters may turn out to be the most significant ones near the solution. For example, a poor initial guess of the parameter vector \mathbf{p} may lead to fully water saturated conditions at a sensor location where water content or capillary pressure is measured, resulting in zero sensitivity coefficients for all two-phase flow parameters. As other parameters are updated, this region of the flow domain may become unsaturated, i.e., the calculated water content and capillary pressures are now sensitive to the two-phase flow parameters of interest, allowing them to be estimated by inverse modeling. It is therefore essential that parameters can be temporarily removed from the minimization to avoid singularity of \mathbf{H} , which would cause premature termination of the calibration process. The procedure also must allow these parameters to be reactivated as soon as they become sensitive.
- (5) Truncating the parameter space has the additional advantage that fewer columns of the Jacobian matrix need to be evaluated by the perturbation method. Since sensitivity coefficients change for nonlinear problems, however, it is essential to re-evaluate the full Jacobian matrix after every few iterations. This dynamic determination of the solution subspace is efficient. Important parameters are

The truncated Levenberg-Marquardt algorithm proposed here is summarized in Fig. 1. The truncation level k can either be predetermined and fixed throughout the inversion, updated during the inversion as a function of the Fisher matrix (which changes due to the nonlinearity of the model), or it can be relaxed according to an empirical schedule. If truncation is merely employed to obtain a more efficient solution to an otherwise robust, overdetermined inverse problem, k can be adjusted heuristically as the minimization proceeds, so that more and more parameters may enter the calibration solution space. Tonkin and Doherty (2005) propose setting k as the number of eigenvalues that exhibit a ratio to the largest eigenvalue that is greater than 10^{-6} . As an alternative to this empirical threshold value, the dimension of the calibration solution space can be chosen such that the final value of the objective function (excluding contributions from Tikhonov regularization) is commensurate with the expected measurement noise level (Finsterle and Pruess, 1995; Tonkin and Doherty, 2005; Moore and Doherty, 2005), or that a predefined maximum prediction uncertainty is not exceeded (in case the inversion is part of an estimation-prediction framework). Evidently, choosing the appropriate truncation level is not a straightforward task, may be problem dependent, and thus needs some further experimentation.

Recall that while only k columns of the Jacobian matrix are needed for the truncated Levenberg-Marquardt update, the full Jacobian needs to be evaluated every few iterations as the sensitivity coefficients change along the solution path. After a step in the parameter space has been taken, it may not be necessary to accurately recalculate the sensitivity

coefficients. Instead, the Jacobian may be efficiently corrected using the Broyden rank-one update (Broyden, 1965), which takes the objective function evaluated at the new parameter set to estimate its curvature and to update the Jacobian matrix. The Broyden update may be acceptable if the determinant of the resulting Fisher information matrix does not change dramatically. This scheme has been implemented into the iTOUGH2 optimization code (Finsterle, 2004), which provides inverse modeling capabilities for the TOUGH2 suite of simulators (Pruess et al., 1999) for nonisothermal multiphase, multicomponent flow in fractured-porous media.

3. Example

An illustrative example is presented here to demonstrate the four schemes for specifying the Tikhonov matrix, i.e., where the diagonal matrix \mathbf{D} is (1) the identity matrix, (2) the reciprocals of the diagonal elements of the normal matrix \mathbf{F} , (3) the same as (2) but divided by the eigenvalues of \mathbf{F} , and (4) the same as (3), but truncated at a user-specified threshold. We evaluate the performance of the various forms of the Levenberg-Marquardt algorithm using an example that involves the estimation of hydrogeological, petrophysical, and geostatistical parameters using synthetically generated hydrological and geophysical data from an infiltration experiment conducted in the vadose zone. This synthetic example considers infiltration of water from a pond into variably saturated, heterogeneous but structured soil with a shallow water table at a depth of 3 m. During the two-day ponded infiltration experiment and subsequent three-day redistribution period, daily averages of the infiltration rate are measured with an accuracy of 0.1 L/day. Two monitoring boreholes are equipped with 40 neutron probes for daily water content

measurements with a standard deviation of 0.02. The same boreholes are also used for six ground penetrating radar (GPR) surveys, one prior to infiltration, and then one every day for five days. In each of these surveys, arrival times are recorded for 133 antenna pairs with a measurement noise of 0.5 ns. The locations of these observation points are visualized in Fig. 2b.

The soil is heterogeneous with a standard deviation for the \log_{10} -permeability of 1.0. The field is spatially correlated as described by a spherical, anisotropic semivariogram, with a correlation length of 2.0 m along the main direction of the stratigraphic layers (which dip at an angle of 20° from the horizontal) and 0.4 m perpendicular to it. Sequential Gaussian simulation (Deutsch and Journel, 1992) is used to generate realizations of this spatially correlated permeability field (see Fig. 2a); the capillary strength parameter of the van Genuchten model (van Genuchten, 1980) is deterministically correlated to the local permeability value according to Leverett (1941).

The model consists of a two-dimensional model domain with heterogeneous permeabilities (Fig. 2a), a constant pressure boundary condition at the bottom (representing the water table), an atmospheric boundary condition at the surface, a prescribed pressure in the pond (representing water depth during the infiltration period), and no-flow boundaries on the side. The simulated saturation distribution after one day of infiltration (Fig. 2b) reflects the intricacy of the heterogeneous soil, making it obvious that point measurements and a homogeneous evolution model would be inadequate to infer the system state and soil structure. GPR travel times are simulated using the straight-ray method, which is based on a high-frequency approximation that calculates the arrival time of the first amplitude departure of the transmitted wave, ignoring the

remainder of the waveform. A volumetric mixing formula (Roth et al., 1990) is used as the petrophysical function that calculates the effective dielectric constant from the local saturation, porosity, and the known dielectric constants of the individual phases. This synthetic model has been described in detail in Kowalsky et al. (2004) and Finsterle and Kowalsky (2008).

To test the performance of the truncated Levenberg-Marquardt algorithm, we deliberately formulated an ill-posed inverse problem in which a total of 30 parameters are simultaneously estimated based on the noisy, synthetically generated data of flow rates, water contents, and GPR arrival times. Of the 30 unknowns, ten are hydrological parameters (absolute permeability in horizontal and vertical directions, porosity, and seven parameters of the van Genuchten relative permeability and capillary pressure functions), two are petrophysical properties (the dielectric constant of the solids and the mixture exponent), four are geostatistical parameters (variance, correlation length, anisotropy factor, and orientation), and 14 are permeability modifiers at 14 so-called pilot points (deMarsily et al., 1984; RamaRao et al., 1995; Gómez-Hernández et al., 1997), which are used as conditioning points during the geostatistical simulation of the heterogeneous permeability field. Table 1 summarizes the parameters.

This overparameterized, highly nonlinear inverse problem is solved using the Levenberg-Marquardt algorithm with different strategies regarding the Tikhonov matrix. Fig. 3 shows the reduction of the objective function versus the number of forward simulations (which is proportional to the CPU time required to solve the inverse problem). The small fluctuations in the objective function represent the changes in the objective function as each parameter is perturbed by 1% of its value to numerically evaluate derivatives. Large

spikes are trial evaluations that led to an unsuccessful uphill step, triggering an increase in the Levenberg parameter, whereas step reductions indicate the successful completion of an iteration.

If using the diagonal elements of the normal matrix as the dampening coefficients, the algorithm encounters a local minimum after a single iteration and terminates. Using the identity matrix leads to steep reduction of the objective function at early iterations, but becomes less efficient as the minimum is approached. If the Tikhonov matrix is scaled by the inverse of the eigenvalues, the objective function is further reduced and approaches the theoretical minimum achievable for the given data noise after approximately 400 forward simulations. The truncated Levenberg-Marquardt algorithm outperforms the other approaches during the first few iterations when it reaches a relatively low objective function value. Further fine-tuning, however, is prevented by the fact that more than a third of the adjustable parameters are below the truncation threshold of 10^{-4} and are thus assigned to the calibration null space. This is further illustrated in Fig. 4, which shows the ranking of the eigenvalues (scaled by the maximum eigenvalue) at the last iteration. The parameter associated with the largest eigenvalue is the reference permeability in the vertical direction, reflecting the fact that gravity-driven flow determines the system state during the infiltration experiment. Porosity is the second highest ranked parameter, as it not only directly affects water content measurements, but also—by means of the petrophysical model—the dielectric constant and thus the velocity with which the electromagnetic wave travels from the transmitting to the receiving GPR antennas. Geostatistical, other hydrological, and the two petrophysical parameters have higher eigenvalues than the first permeability modified at the pilot point closest to the

infiltration pond, which is ranked number 11. Ten of the 13 parameters below the truncation threshold of 10^{-4} are pilot points. This does not imply that soil heterogeneity cannot be estimated in general. In this example, the truncation threshold of 10^{-4} was chosen relatively high to better demonstrate the mechanics of the algorithm. Moreover, by curtailing the parameter space and thus the algorithm's flexibility to adjust the heterogeneity of the soil, the goodness-of-fit may be considered poor according to the criteria discussed in Finsterle and Zhang (2010), calling for further model refinement.

The data are also insufficient to identify three hydrological parameters, specifically the cut-off value for extrapolating the capillary pressure function near residual saturation, which is a parameter of low sensitivity during an infiltration experiment (but may be essential if soil drying is considered). Note that scaling the Tikhonov matrix by the inverse of these small eigenvalues prevents the system of normal equations from becoming numerically singular.

Fig. 5 shows a scatter plot of the final GPR-time residuals. The cloud of residuals is widest if the identity matrix is used (blue symbols), and tightest for the inversion result obtained with the eigenvalue-based Tikhonov matrix without truncation of the parameter space (green symbols). (The residuals after the attempted inversion with the normal matrix are not shown, as the objective function itself indicates that the model is an unlikely fit to the data.) The relative quality of the matches revealed in the scatter plot is consistent with the final values of the objective function shown in Fig. 3. The plot also shows a random pattern about the diagonal line, indicating that the calibrated model is devoid of obvious systematic errors. Similar conclusions can be drawn for the analysis of the water-content and infiltration-rate residuals. While a more detailed residual analysis is

essential for the assessment of the models ability to reproduce the data, such an analysis is outside the scope of this paper; a detailed discussion can be found in Finsterle and Zhang (2010). The reconstructed permeability field is very close to the true field shown in Fig. 2a. However, this outcome is too optimistic and a result of the fact that a synthetic example was chosen with a known seed value used by the random number generator; this issue is further discussed in Finsterle and Kowalsky (2008).

As discussed above, the Levenberg-Marquardt algorithm with an eigenvalue-based Tikhonov matrix performs best for this particular example and other inversion problems examined by the authors. However, as with any local method used for solving highly nonlinear inverse problems, we cannot guarantee that the proposed approach is superior to others in all cases. In fact, it is difficult to provide conclusive arguments why the algorithm converged better than its alternatives. Truncating the parameter space has the obvious advantage of reducing the computational burden of evaluating the Jacobian matrix. In addition, it appears that the large perturbations taken by insensitive parameters at a given iteration may lead the inversion astray in subsequent iterations, when these parameters may become more sensitive. The dynamic regularization and truncation method keeps these parameters contained when needed, but allows them to adjust should they become relevant. Moreover, the length and orientation of the parameter update is different in the case where step size limitations are invoked, or parameter bounds are reached. The overall impact, however, is difficult to assess as the solution path is tortuous and the behavior depends on the details of the implementation of the algorithm.

4. Concluding Remarks

The calibration of multiphase flow models is challenging because (1) a large number of parameters is needed to capture a variety of coupled processes and to adequately describe the heterogeneous structure of the subsurface, (2) the governing equations are highly nonlinear, and (3) the solution of the forward problem is computationally very expensive. This challenge calls for an inversion approach that allows for a flexible parameterization and robust minimization of the objective function, whose topology is complicated and may change significantly along the solution path. While regularization methods may improve the robustness of the solution, they fundamentally change the inverse problem that is solved. Instead of changing the objective function to be minimized, we focused on improving the minimization algorithm itself. To handle highly nonlinear, potentially overparameterized inverse problems in a robust and potentially more efficient manner, a modified, truncated Levenberg-Marquardt algorithm was proposed and examined using an example that involves the joint inversion of hydrogeological and geophysical data.

It was demonstrated that using the inverse of the eigenvalues of the Fisher information matrix as damping coefficients yielded better estimates with a significantly smaller number of forward evaluations. Truncating the parameter space based on these eigenvalues led to improved performance during the initial iterations, but may result in suboptimal parameter sets if the truncation limit is set too high. This difficulty may be addressed by dynamically lowering the truncation limit, or by optimizing for superparameters (i.e., parameters that are aligned with the eigenvectors) as proposed by Tonkin and Doherty (2005).

Solving multiphase inverse problems will remain difficult. It foremost requires a solid understanding of the involved physical processes and their representation in a numerical

model. A careful design of characterization experiments or monitoring systems is equally important to ensure that the measured data contain information about the relevant parameters. Finally, robust and efficient minimization algorithms are needed to solve the related nonlinear inverse problem. The enhancements discussed here are intended to make the calibration of highly parameterized multiphase flow models more feasible, thus reducing the impact of systematic errors on estimated parameters and subsequent model predictions.

Acknowledgment

We would like to thank Andrés Alcolea and two anonymous reviewers, as well as Michael Commer (LBNL) for their thoughtful and constructive review comments. Discussions with Knútur Árnason, ÍSOR, Iceland, are also greatly appreciated. This work was supported by Laboratory Directed Research and Development (LDRD) funding from Berkeley Lab, provided by the Director, Office of Science, of the U.S. Department of Energy, and as part of the Subsurface Science Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Resources under Award Number DE-AC02-05CH11231.

References

- Alcolea, A., J. Carrera, and Medina A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources* 29(11), 1678–1689, doi:10.1016/j.advwatres.2005.12.009.
- Broyden, C.G., 1965. Class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation* 19(92), 577–293, doi:10.2307/2003941.
- Carrera, C., Neuman, S.P., 1986. Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information. *Water Resources Research* 22(2), 199–210.
- Carrera J., A. Alcolea, A. Medina, J. Hidalgo, and L.J. Slooten L.J., 2005. Inverse problem in hydrology, *Hydrogeology Journal* 13, 206–222, doi: 10.1007/s10040-004-0404-7.

- deMarsily, G.H., G. Lavedan, M. Boucher, and G. Fasanino (1984). Interpretation of inference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. In: Verly G. et al. (editors). Geostatistics for natural resources characterization. Part 2. D. Reidel Pub. Co., 831–849.
- Deutsch, C.V., Journel, A.G., 1992. GSLIB-Geostatistical Software Library and User's Guide. Oxford University Press, New York.
- Finsterle, S., 2004. Multiphase inverse modeling: Review and iTOUGH2 applications. *Vadose Zone Journal* 3, 747–762.
- Finsterle, S., Kowalsky, M.B., 2008. Joint hydrological-geophysical inversion for soil structure identification. *Vadose Zone Journal* 7, 287–293, doi:10.2136/vzj2006.0078.
- Finsterle, S., Pruess, K., 1995. Solving the estimation-identification problem in two-phase flow modeling. *Water Resources Research* 31(4), 913–924.
- Finsterle, S., Najita, J., 1998. Robust estimation of hydrogeologic model parameters. *Water Resources Research* 34(11), 2939–2947.
- Finsterle, S., Zhang, Y., 2010. Error handling strategies in multiphase inverse modeling. *Computers and Geosciences* (this issue).
- Gauss, C.F., 1821. Theory of the Combination of Observations Least Subject to Errors. (in: Stewart (trans.), SIAM, Philadelphia, PA, 1995).
- Gómez-Hernández, J.J., A. Sahuquillo, and J.E. Capilla (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1. Theory. *Journal of Hydrology* 204(1–4), 162–174.

- Hendricks Franssen, H.J., A. Alcolea, M. Riva, M. Bakr, N. van de Wiel, F. Stauffer and A. Guadagnini, 2009. A comparison of seven methods for the inverse modelling of groundwater flow. Application to the characterisation of well catchments. *Advances in Water Resources*, 32(6), 851–872, doi:10.1016/j.advwatres.2009.02.011.
- Kowalsky, M.B., Finsterle, S., Rubin, Y., 2004. Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone. *Advances in Water Resources* 27(6), 583–599.
- Lehikoinen, A., Huttunen, J.M.J., Finsterle, S., Kowalsky, M.B. Kaipio, J.P., 2010. Dynamic inversion for hydrological process monitoring under model uncertainty. *Water Resources Research* 46, W04513, doi:10.1029/2009WR008470.
- Levenberg, K., 1944. A method for the solution of certain nonlinear problems in least squares. *The Quarterly of Applied Mathematics* 2, 164–168.
- Leverett, M. C., 1941. Capillary behavior in porous solids. *Transactions of the Society of Petroleum Engineering AIME* 142, 152–169.
- Marquardt, D.W., 1963. An algorithm for least squares estimation of nonlinear parameters, *SIAM Journal of Applied Mathematics* 11, 431–441.
- Moore, C., Doherty, J., 2005. Role of the calibration process in reducing model predictive error. *Water Resources Research* 41, W05020, doi:10.1029/2004WR003501.
- Moore, C., Doherty, J., 2006. The cost of uniqueness in groundwater model calibration. *Advances in Water Resources* 29, 605–623.

- RamaRao, B. S., de Marsily, G., Marietta, M.G., 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. *Water Resources Research*, 31(3), 475-493
- Regis, R.G., Shoemaker, C.A., 2004. Local function approximation in evolutionary algorithms for the optimization of costly functions. *IEEE Transactions on Evolutionary Computation* 8(5), 490–505.
- Roth, K.R., Schulin, R., Flühler, H., Attinger, W., 1990. Calibration of time domain reflectometry for water content measurement using a composite dielectric approach. *Water Resources Research*, 26(10), 2267–2273.
- Tikhonov, A.N., Arsenin, V.Y., 1977. *Solutions to Ill-Posed Problems*. John Wiley, New York, pp. 272.
- Tonkin, M., Doherty, J. 2005. A hybrid regularized inversion methodology for highly parameterized models. *Water Resources Research* 41, W10412, doi:10.1029/2005WR003995.
- van Genuchten, M. T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Journal of the American Soil Sciences Society* 44, 892–898.
- Vrugt, J., Robinson, B.A. Hyman, J.M., 2009. Self-adaptive multimethod search for global optimization in real-parameter spaces. *IEEE Transactions on Evolutionary Computation* 13(2), 234–259.

Table 1.

Hydrological, petrophysical, and geostatistical parameters fixed and estimated by joint inversion of hydrological and geophysical data

Parameter	Value
<i>Hydrogeological Parameters</i>	
Reference horizontal permeability, $\log(k_{ref,h} [\text{m}^2])$	-12.0
Reference vertical permeability, $\log(k_{ref,v} [\text{m}^2])$	-13.0
Porosity, ϕ	0.25
Reference capillary-strength parameter, $\log(1/\alpha_{ref} [\text{Pa}])$	3.0
Residual liquid saturation, S_{lr}	0.1
Irreducible liquid saturation, S_{li}	0.1
Saturation exponent, η	0.5
van Genuchten parameter, n	2.7
van Genuchten parameter, m	0.7
Cutoff for linear extrapolation of capillary pressure, ε	0.05
<i>Petrophysical Parameters</i>	
Mixture exponent, n	0.5
Dielectric constant, solid, κ_s	4.0
<i>Geostatistical Parameters</i>	
Permeability modifiers at 14 pilot points, $\log(k/k_{ref})$	0.0
Correlation length, L [m]	0.4
Geometric anisotropy ratio, ω	5.0
Orientation, β [°]	20.0
Sill value, $\log(c)$	1.0

Figure Captions

Figure 1. Truncated Levenberg-Marquardt algorithm.

Figure 2. (a) Spatially correlated permeability field, and (b) liquid saturation distribution after one day of water release, locations of neutron probes in boreholes (squares), and GPR straight-ray paths.

Figure 3. Levenberg-Marquardt minimization of the objective function with different damping matrices.

Figure 4. Ranked scaled eigenvalues of Fisher information matrix.

Figure 5. Scatter plot of final GPR-time residuals after calibration with different Tikhonov matrices.

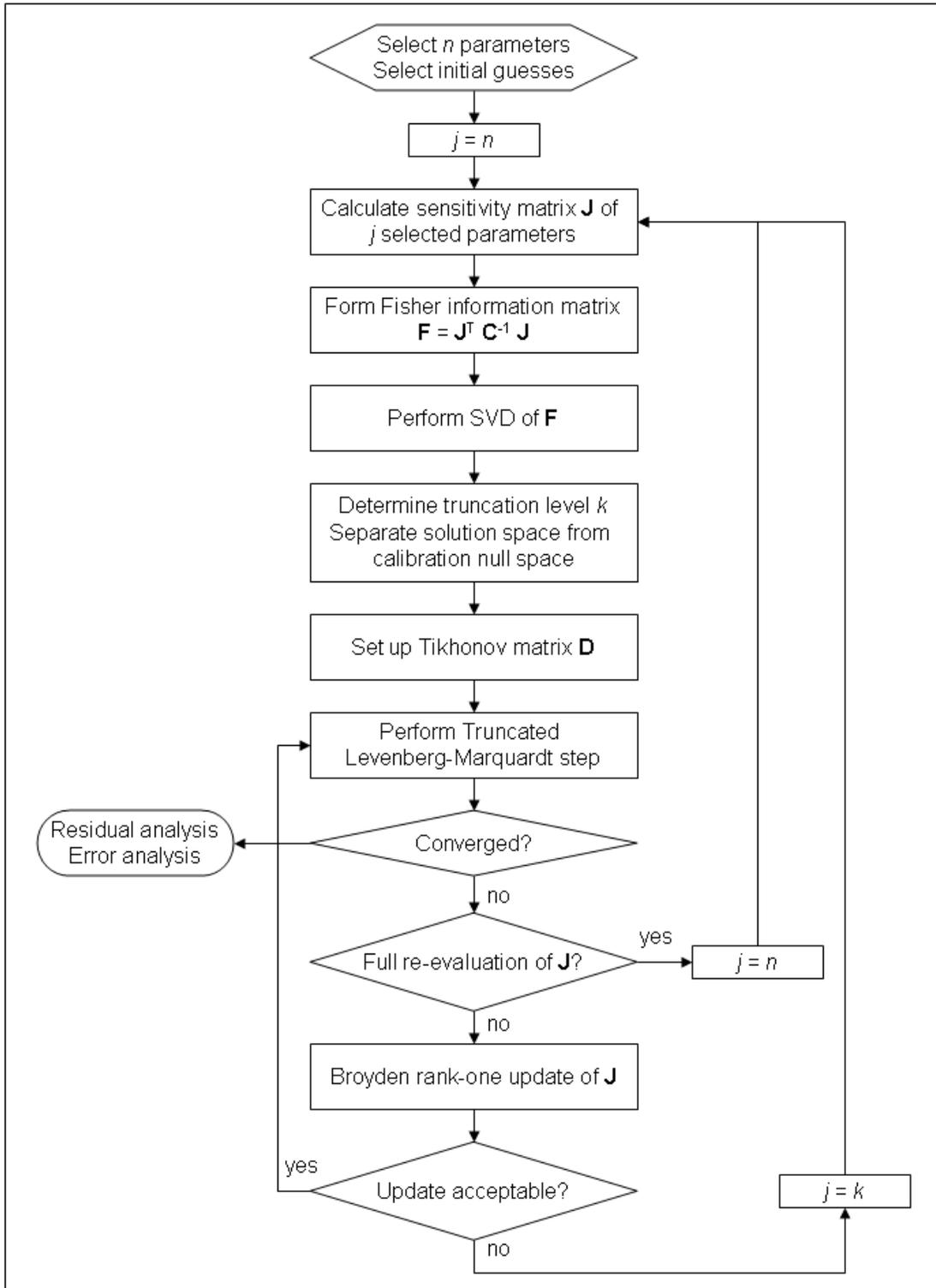
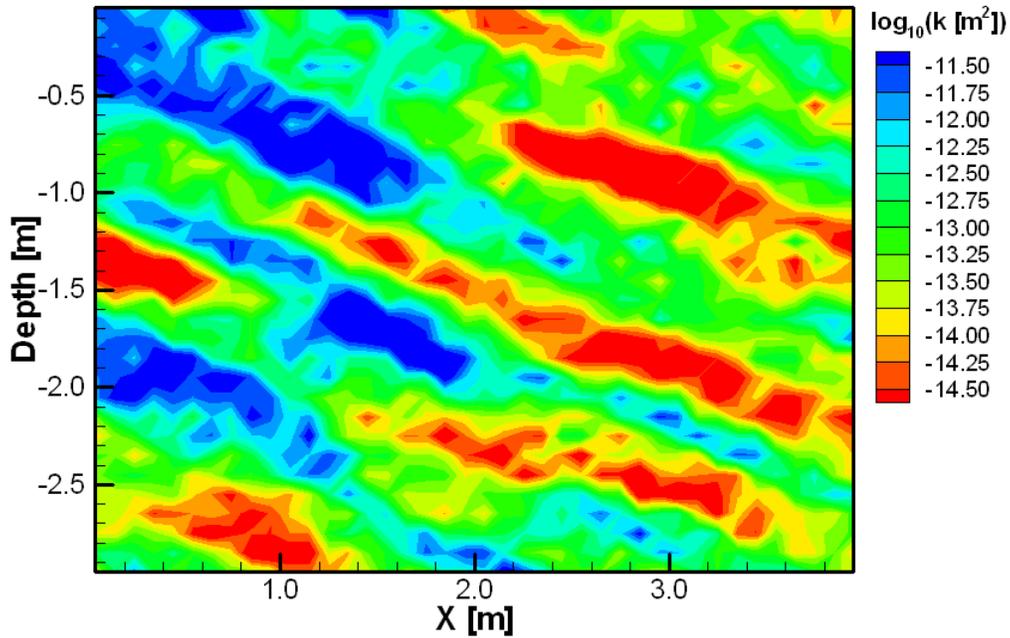
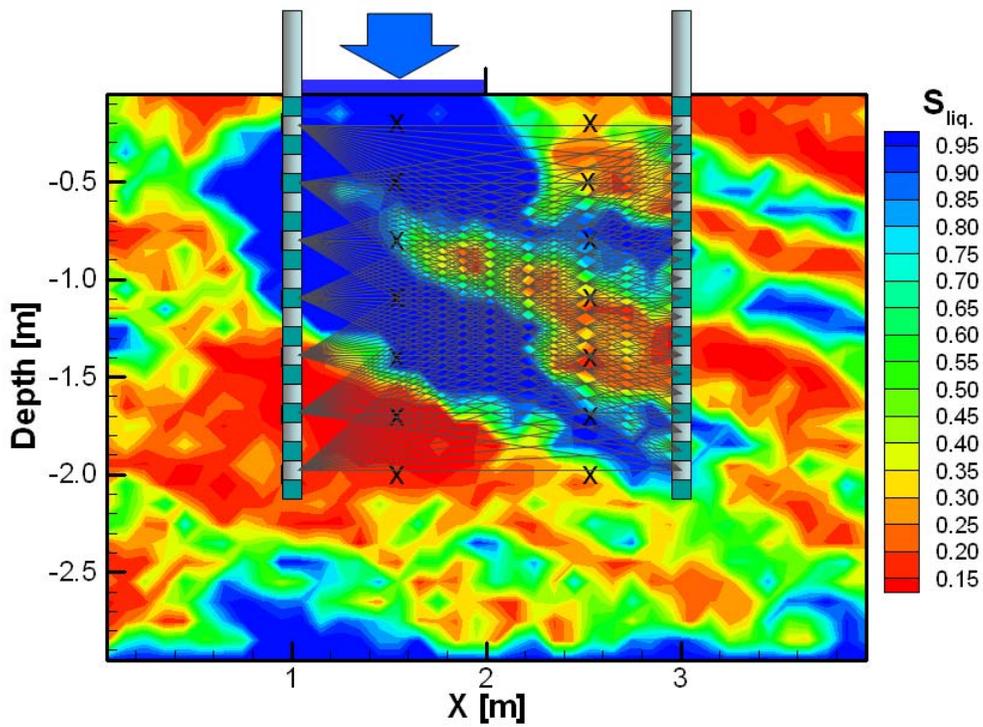


Figure 2. Truncated Levenberg-Marquardt algorithm.



(a)



(b)

Figure 2. (a) Spatially correlated permeability field, and (b) liquid saturation distribution after one day of water release, locations of neutron probes in boreholes (squares), and GPR straight-ray paths.

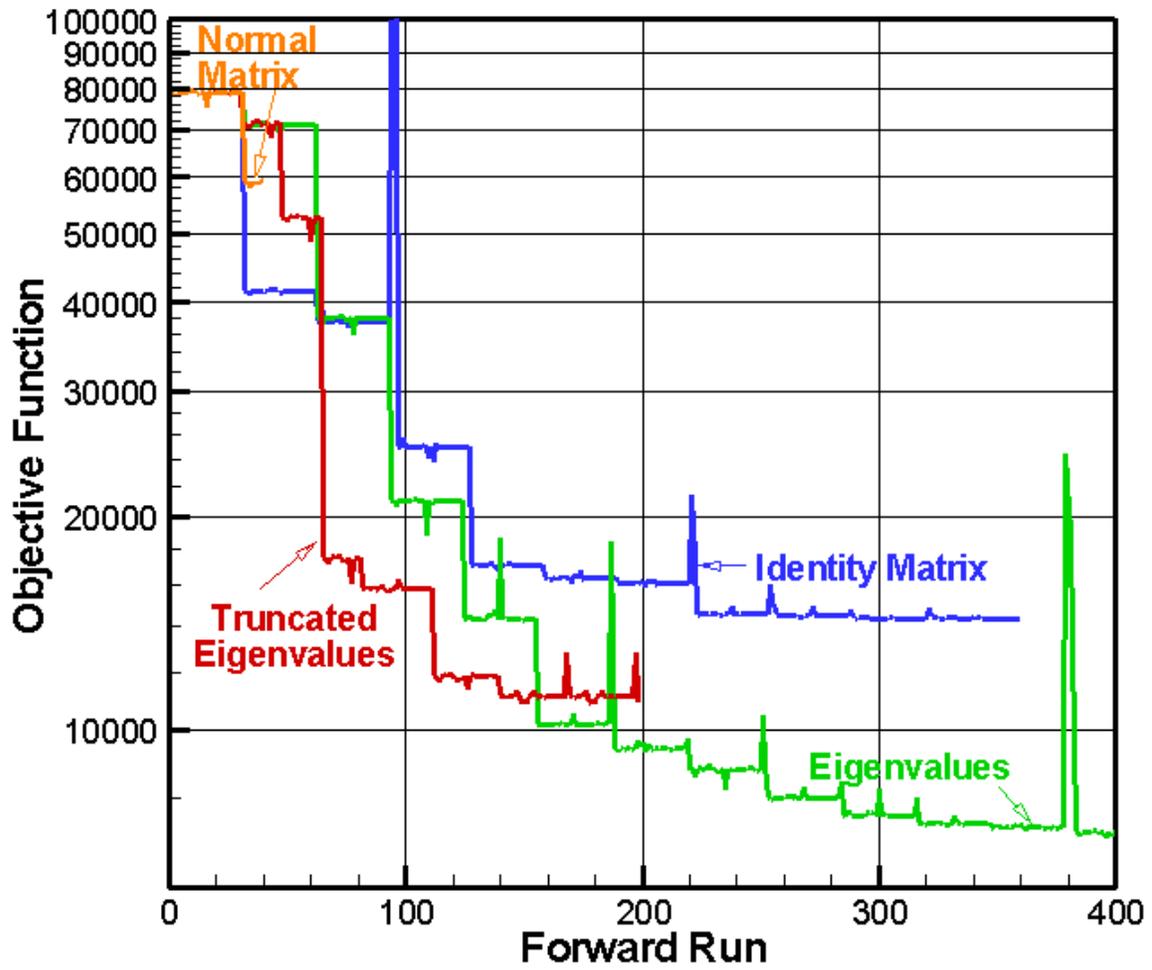


Figure 3. Levenberg-Marquardt minimization of the objective function with different damping matrices.

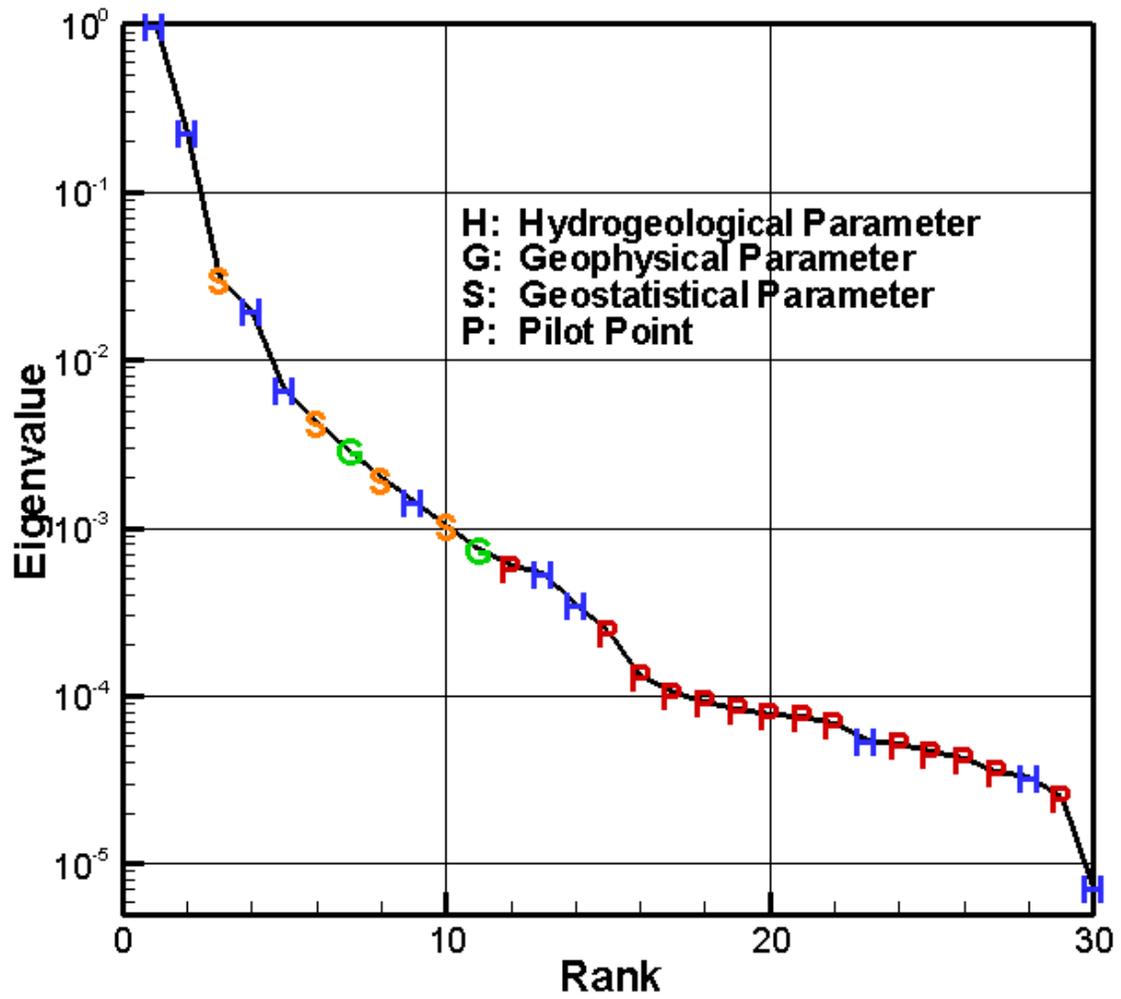


Figure 4. Ranked scaled eigenvalues of Fisher information matrix.

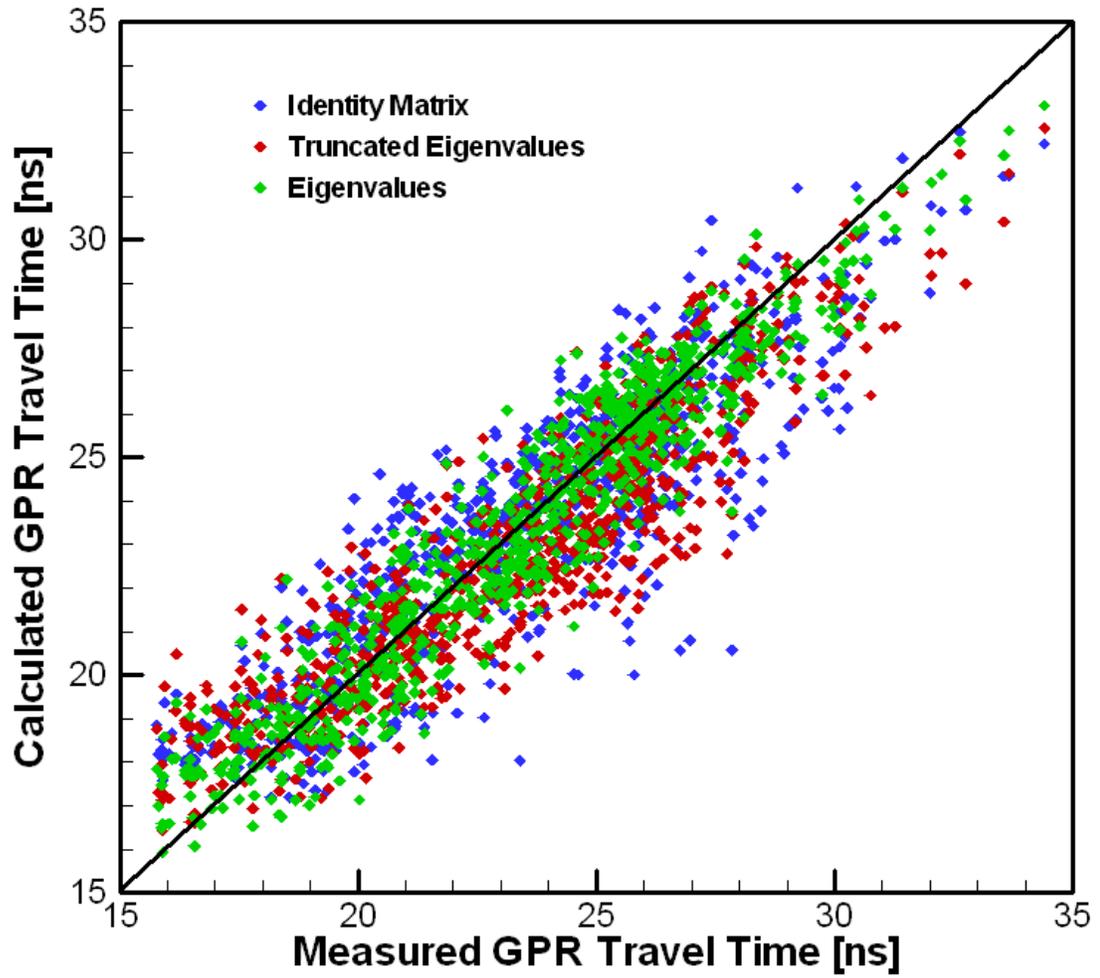


Figure 5. Scatter plot of final GPR-time residuals after calibration with different Tikhonov matrices.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.