Abstract. In a typical inverse problem, a spatially distributed parameter in a physical model is estimated from indirect measurements of model output. Since measurements are stochastic in nature, so is any parameter estimate. Moreover, in the Bayesian setting, the choice of regularization corresponds to the definition of the prior probability density function, which in turn is an uncertainty model for the unknown parameters. For both of these reasons, significant uncertainties exist in the solution of an inverse problem. Thus to fully understand the solution, quantifying these uncertainties is important. When the physical model is linear and the error model and prior are Gaussian, the posterior density function is Gaussian with a known mean and covariance matrix. However, the electrical impedance tomography inverse problem is nonlinear, and hence, no closed form expression exists for the posterior density. The typical approach for such problems is to sample from the posterior, and then use the samples to compute statistics (such as the mean and variance) of the unknown parameters. Sampling methods for electrical impedance tomography have been studied by various authors in the inverse problems community. However, up to this point the focus has been on the development of increasingly sophisticated implementations of the Gibbs sampler, whose samples are known to converge very slowly to the correct density for large-scale problems. In this paper, we implement a recently developed sampling method called randomize-then-optimize (RTO), which provides nearly independent samples for each application of an appropriate numerical optimization algorithm. The sample density for RTO is not the posterior density, but RTO can be used as a very effective proposal within a Metropolis-Hastings algorithm to obtain samples from the posterior. Here our focus is on implementing the method on synthetic examples from electrical impedance tomography, and we show that it is both computationally efficient and provides good results. We also compare RTO performance with the Metropolis Adjusted Langevin Algorithm and found RTO to be much more efficient.

Key words. inverse problems, Bayesian methods, Markov Chain Monte Carlo, numerical optimization, electrical impedance tomography

AMS Subject Classifications: 15A29, 65F22, 65C05, 65C60, 94A08

1. Introduction. Solving an inverse problem requires estimating unknown parameters in a physical model from indirect measurements of model output. The unknown parameter vector typically corresponds to a numerical discretization of a continuously defined, spatially distributed parameter in the physical model, and hence is high-dimensional. This leads to an over-parameterized statistical model, and hence, estimates have large variance, or in other words, are unstable with respect to perturbations in the observations. Regularization methods, which have been studied for decades in the inverse problems literature, are the typical fix for such instability.

Over the past ten years or so, an increasing number of researchers in inverse problems have drawn the connection between regularized solutions of inverse problems...
and Bayesian statistics [13]. Since the measured data in an inverse problem will always contain noise, a statistical approach is desirable and is even necessary if one would like to quantify uncertainty in the unknown parameters.

Given a posterior density function from Bayes’ Law for a given inverse problem, uncertainty quantification often requires a sampling method, particularly if the dependence of the parameters on the data is nonlinear. Perhaps the classic nonlinear inverse problem is electrical impedance tomography (EIT), and for this application a number of Markov chain Monte Carlo (MCMC) sampling techniques have been implemented, all based on the Gibbs sampler [14, 15, 19, 29]. However, these techniques typically suffer from slow convergence and high CPU demand in large-scale cases. Another class of MCMC methods used for large-scale inverse problems – which construct Gaussian proposals using the Hessian of the likelihood function – are presented in [18, 22].

In this paper, we implement the randomize-then-optimize (RTO) algorithm introduced in [2] for sampling from the posterior in EIT. RTO employs the optimization algorithm used for computing the maximum a posterior (MAP) estimator, with only minor modification, to obtain posterior samples. In [2], the RTO probability density function is derived and then is used to sample from small- and medium-scale posterior density functions arising in parameter estimation for nonlinear models. Here, our focus is on implementing RTO for the problem of sampling from the posterior in the large-scale EIT case.

To motivate RTO, it is helpful to first consider the linear Gaussian case

\[ \mathbf{v} = \mathbf{U}\gamma + \mathbf{\epsilon}, \]  

where \( \mathbf{U} \in \mathbb{R}^{m \times n} \) is the ill-conditioned forward model matrix, \( \mathbf{v} \in \mathbb{R}^m \) is the measured data, \( \gamma \in \mathbb{R}^n \) is the unknown, and \( \mathbf{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m) \). The likelihood function for statistical model (1.1) then has the form

\[ p(\mathbf{v}|\gamma, \lambda) \propto \lambda^{m/2} \exp \left( -\frac{\lambda}{2} \|\mathbf{U}\gamma - \mathbf{v}\|^2 \right), \]  

where \( \lambda = 1/\sigma^2 \) is the noise precision (inverse-variance). If we additionally assume a Gaussian prior of the form

\[ p(\gamma|\delta) \propto \delta^{n/2} \exp \left( -\frac{\delta}{2} (\gamma - \bar{\gamma})^T \mathbf{L} (\gamma - \bar{\gamma}) \right), \]  

where \( \delta \mathbf{L} \) is the prior precision (inverse-covariance) matrix, then the conditional probability density for \( \gamma \) is given by Bayes’ Law:

\[ p(\gamma|\mathbf{v}, \lambda, \delta) \propto p(\mathbf{v}|\gamma, \lambda)p(\gamma|\delta) \]

\[ \propto \exp \left( -\frac{\lambda}{2} \|\mathbf{U}\gamma - \mathbf{v}\|^2 - \frac{\delta}{2} (\gamma - \bar{\gamma})^T \mathbf{L} (\gamma - \bar{\gamma}) \right) \]

\[ = \exp \left( -\frac{1}{2} \|\hat{\mathbf{U}}\gamma - \hat{\mathbf{v}}\|^2 \right), \]  

where

\[ \hat{\mathbf{U}} = \begin{bmatrix} \lambda^{1/2}\mathbf{U} \\ \delta^{1/2}\mathbf{L}^{1/2} \end{bmatrix}, \quad \hat{\mathbf{v}} = \begin{bmatrix} \lambda^{1/2}\mathbf{v} \\ \delta^{1/2}\mathbf{L}^{1/2}\bar{\gamma} \end{bmatrix}. \]  

\[ (1.5) \]
A key point is that samples from $p(\gamma|v,\lambda,\delta)$ can be computed by repeatedly solving the optimization problem
\begin{equation}
\gamma = \arg \min_{\gamma} ||\hat{U}\gamma - (\hat{v} + \eta)||^2, \quad \eta \sim \mathcal{N}(0, I). \tag{1.6}
\end{equation}

To see this, note that the solution of the normal equations for this least squares problem is given by
\begin{equation}
\gamma = (\hat{U}^T \hat{U})^{-1} \hat{U}^T (\hat{v} + \eta) = (\lambda U^T U + \delta L)^{-1} (\lambda U^T v + \delta L \bar{\gamma}) + w, \quad w \sim \mathcal{N}(0, (\lambda U^T U + \delta L)^{-1}),
\end{equation}
which has probability density function $p(\gamma|v,\lambda,\delta)$. This optimization-based approach for sampling from the posterior density function is implemented in the case of large-scale linear inverse problems in [1].

In this paper, we extend (1.6) to the case of a nonlinear function $U$; specifically, we replace (1.1) by
\begin{equation}
v = U(\gamma) + \epsilon, \tag{1.7}
\end{equation}
where $U : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is nonlinear, $v \in \mathbb{R}^m$ is the measured data, $\gamma \in \mathbb{R}^n$ is the unknown, and $\epsilon \sim \mathcal{N}(0, \sigma^2 I_m)$. As in the linear case, the likelihood function for statistical model (1.7) takes the form
\begin{equation}
p(v|\gamma, \lambda) \propto \lambda^{m/2} \exp\left(-\frac{\lambda}{2} \|U(\gamma) - v\|^2\right), \tag{1.8}
\end{equation}
while the prior is assumed to have the same form (1.3). For us, the prior covariance matrix $L^{-1}$ will be defined by a Gaussian covariance function on the computational domain used for the numerical solution of the elliptic partial differential equation that defines EIT [15]. Finally, by Bayes' Law, the posterior density function is given by
\begin{equation}
p(\gamma|v, \lambda, \delta) \propto \exp\left(-\frac{\lambda}{2} \|U(\gamma) - v\|^2 - \frac{\delta}{2} (\gamma - \bar{\gamma})^T L (\gamma - \bar{\gamma})\right) \\
= \exp\left(-\frac{1}{2} \|\hat{U}(\gamma) - \hat{v}\|^2\right), \tag{1.9}
\end{equation}
where
\begin{equation}
\hat{U}(\gamma) = \begin{bmatrix} \lambda^{1/2} U(\gamma) \\ \delta^{1/2} L^{1/2} \gamma \end{bmatrix}, \quad \hat{v} = \begin{bmatrix} \lambda^{1/2} v \\ \delta^{1/2} L^{1/2} \bar{\gamma} \end{bmatrix}. \tag{1.10}
\end{equation}

Here, we extend the idea of (1.6) to the case where $U$ is nonlinear. The obvious extension is to replace $U\gamma$ by $U(\gamma)$ in (1.6), yielding
\begin{equation}
\gamma = \arg \min_{\psi} ||\hat{U}(\psi) - (\hat{v} + \eta)||^2, \quad \eta \sim \mathcal{N}(0, I). \tag{1.11}
\end{equation}

In a variety of cases, we have found that the samples defined by (1.11) are quite accurate approximations of posterior samples [2]; indeed, the randomized maximum likelihood (RML) method of [4, Section 2.1] advocates using samples from (1.11) as approximate samples from $p(\gamma|v, \lambda, \delta)$.
This motivates using (1.11) as a proposal mechanism within a Metropolis-Hastings or importance sampling framework to obtain theoretically correct posterior samples. However, this would require knowledge of the probability density function for $\gamma$ defined by (1.11). We have not been able to derive this probability density, but in [2] we derive the probability density for the related optimization problem

$$\gamma = \arg \min_{\psi} \| \bar{Q}^T (\bar{U}(\psi) - (\bar{v} + \eta)) \|^2, \quad \eta \sim \mathcal{N}(0, I),$$

(1.12)

where $\bar{Q}$ is the $m \times n$ matrix obtained from the thin QR-factorization of the Jacobian $J(\bar{\gamma})$ of $\bar{U}$ evaluated at the MAP estimator $\bar{\gamma} = \arg \max_{\gamma} p(\gamma|v, \lambda, \delta)$. We note that the samples computed from (1.12) are still exact in the linear case, and in the nonlinear case, we have found that the samples are often very close to the posterior, as in RML.

We call (1.12) randomize-then-optimize (RTO), and we will sketch the derivation of the probability density function for $\gamma$ defined by (1.12) below; it has the form

$$p_{\gamma}(\gamma) \propto c(\gamma)p(\gamma|v, \lambda, \delta),$$

(1.13)

where $p(\gamma|v, \lambda, \delta)$ is the posterior density function (1.9), and $c(\gamma)$ is a scaling term. Finally, we use (1.12, 1.13) as an independence proposal within a Metropolis-Hastings (MH) algorithm for sampling from $p(\gamma|v, \lambda, \delta)$, and we denote the resulting method by RTO-MH.

The remainder of the paper is organized as follows. In Section 2, we derive the RTO probability density function and then describe the RTO-MH algorithm for sampling from $p(\gamma|v, \lambda, \delta)$. We also briefly describe a state of the art method for sampling high-dimensional posteriors, the Hessian-preconditioned Metropolis Adjusted Langevin Algorithm (MALA) [23], which resembles the recently proposed stochastic Newton MCMC sampler [18]. In Section 3, we implement RTO on synthetic EIT test cases and compare its performance with MALA.

2. The RTO Probability Density and Metropolis-Hastings Sampler. In this section, we give a brief overview of the derivation of the RTO probability density function $p_{\gamma}(\gamma)$. A more detailed derivation is given in [2]. First, for notational simplicity, we drop the ′′′ notation in (1.12), and we assume that the MAP estimation problem,

$$\bar{\gamma} = \arg \min_{\psi} \frac{1}{2} \| \bar{U}(\psi) - v \|^2,$$

has a unique solution $\bar{\gamma}$, and that $U$ is continuously differentiable with rank $n$ Jacobian $J(\gamma)$ for all $\gamma$ in the domain of $U$.

The first order necessary conditions for optimality are given by

$$J(\bar{\gamma})^T (\bar{U}(\bar{\gamma}) - v) = 0.$$

(2.1)

Let $J(\bar{\gamma}) = \bar{Q}\bar{R}$ be the thin QR-factorization of $J(\bar{\gamma})$, then $\bar{Q} \in \mathbb{R}^{m \times n}$ with orthonormal columns, and since $\bar{R}$ is invertible, (2.1) is equivalent to $\bar{Q}^T (\bar{U}(\bar{\gamma}) - v) = 0$, which we express as

$$F_{\bar{\gamma}}(\bar{\gamma}) = \bar{Q}^T v, \quad \text{where} \quad F_{\bar{\gamma}}(\bar{\gamma}) \overset{def}{=} \bar{Q}^T \bar{U}(\bar{\gamma}).$$

(2.2)

Now, we perturb the right-hand side of (2.2) with iid Gaussian noise and use $F_{\bar{\gamma}} : \mathbb{R}^n \to \mathbb{R}^n$ to define the random variable $\gamma$ via the inverse mapping

$$\gamma = F_{\bar{\gamma}}^{-1}(\bar{Q}^T v + \epsilon), \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, I).$$

(2.3)

4
The nonlinear transformation (2.3) is well-defined if: (i) samples of \( y \equiv Q^T v + \epsilon \) lie in the range of \( F_y \); and (ii) the Jacobian of \( F_y \), given by \( J_y \equiv Q^T J(\gamma) \), is invertible for all \( \gamma \) in the domain of \( U \). We make the later assumption and can overcome (i) by restricting \( y \) to the range of \( F_y \) (see [2] for arguments making this rigorous).

Finally, we derive the probability density for \( \gamma \) defined by (2.3). First, the probability density function for \( y \) is given by \( p_y(y) \propto \exp \left( -\frac{1}{2} \| y - Q^T v \|^2 \right) \). Then, by the theory of transformations of a random vector [8], we obtain

\[
p_{\gamma}(\gamma) = \det (J_{\gamma}(\gamma)) \ p_y(F_{\gamma}(\gamma)) = \det (Q^T J(\gamma)) \exp \left( -\frac{1}{2} \| Q^T (U(\gamma) - v) \|^2 \right) = c(\gamma)p(\gamma|v), \tag{2.4}
\]

where \( p(\gamma|v, \lambda, \delta) \propto \exp \left( -\frac{1}{2} \| U(\gamma) - v \|^2 \right) \) is the posterior density function (1.9), and

\[
c(\gamma) = \det (Q^T J(\gamma)) \exp \left( \frac{1}{2} \| r(\gamma) \|^2 - \frac{1}{2} \| Q^T r(\gamma) \|^2 \right), \tag{2.5}
\]

with \( r(\gamma) \equiv U(\gamma) - v \).

In practice, we sample from \( p_{\gamma}(\gamma) \) by repeatedly solving the following optimization problem

\[
\gamma = \arg \min_{\psi} \left\{ \ell(\psi, \epsilon) \equiv \| Q^T (U(\psi) - (v + \epsilon)) \|^2 \right\}, \quad \text{where} \quad \epsilon \sim \mathcal{N}(0, I), \tag{2.6}
\]

which we assume has a unique solution for all \( \epsilon \sim \mathcal{N}(0, I) \). As mentioned above, we must reject those \( Q^T(v + \epsilon) \) not in the range of \( F_{\gamma}(\psi) = Q^T U(\psi) \), or equivalently, those \( (\gamma, \epsilon) \) for which \( \ell(\gamma, \epsilon) > 0 \). In our numerical implementation, we reject those \( (\gamma, \epsilon) \) for which \( \ell(\gamma, \epsilon) > \eta \), where \( 1 >> \eta > 0 \). This procedure for computing samples from \( p_{\gamma}(\gamma) \) motivates our name for the method, randomize-then-optimize.

Finally, we made the assumption above that \( Q^T J(\gamma) \) is invertible, which must be checked. We do this by computing the QR-factorization of \( Q^T J(\gamma) \) for each sample \( \gamma \): let \( Q^T J(\gamma) = Q \gamma R_{\gamma} \), then \( Q^T J(\gamma) \) is invertible if the diagonal elements \( r_{ii}^\gamma = |R_{\gamma}|_{ii} \) are nonzero for \( i = 1, \ldots, n \). Fortunately, we can also use this QR-factorization to compute (2.5); in practice, we compute

\[
\ln c(\gamma) = \sum_{i=1}^{n} \ln |r_{ii}^\gamma| + \frac{1}{2} \left( \| r(\gamma) \|^2 - \| Q^T r(\gamma) \|^2 \right), \tag{2.7}
\]

which we will use for the Metropolis-Hastings acceptance ratio below.

### 2.1. The RTO-Metropolis-Hastings Algorithm

Now, we show how to use the RTO density (2.4), (2.5) as an independence proposal within a Metropolis-Hastings (MH) algorithm. Given a previous sample \( \gamma^{k-1} \) and a proposed RTO sample \( \gamma^* \), the acceptance ratio for the MH method is given by [9]

\[
r = \min \left( 1, \frac{p(\gamma^*|v, \lambda, \delta)p(\gamma^{k-1})}{p(\gamma^{k-1}|v, \lambda, \delta)p(\gamma^*)} \right) = \min \left( 1, \frac{p(\gamma^*|v, \lambda, \delta)c(\gamma^{k-1})p(\gamma^{k-1}|v, \lambda, \delta)}{p(\gamma^{k-1}|v, \lambda, \delta)c(\gamma^*)p(\gamma^*|v, \lambda, \delta)} \right) = \min \left( 1, \frac{c(\gamma^{k-1})/c(\gamma^*)}{c(\gamma^*)} \right). \tag{2.8}
\]
Thus the acceptance ratio depends only upon the values of the scaling function \( c \) defined in (2.5). For large-scale problems, such as EIT, the computation of the ratio \( c(\gamma_k - 1)/c(\gamma^*) \) can be numerically unstable. Hence we advocate using

\[
c(\gamma_k - 1)/c(\gamma^*) = \exp\left(\ln c(\gamma_k - 1) - \ln c(\gamma^*)\right),
\]

(2.9)

where \( \ln c(\gamma) \) is defined in (2.7).

Finally, for completeness, we present the full RTO-MH algorithm.

**The RTO Metropolis-Hastings (RTO-MH) Algorithm**

1. Choose initial vector \( \gamma^0 \), \( 1 >> \eta > 0 \), and total samples \( N \). Set \( k = 1 \).
2. Compute an RTO sample \( \gamma^* \) using (2.6) with corresponding \( \epsilon^* \).
3. If \( \ell(\gamma^*, \epsilon^*) > \eta \), return to Step 2, else go to Step 4.
4. Define the acceptance probability \( r \) by (2.8) with ratio computed using (2.9).
5. Simulate \( u \sim \mathcal{U}(0, 1) \). If \( u < r \), set \( \gamma_k = \gamma^* \), else set \( \gamma_k = \gamma_k - 1 \).
6. If \( k < N \), set \( k = k + 1 \) and return to Step 2.

As mentioned above, RTO can also be used within an importance sampling framework, as in [2]. However, for EIT we found that the importance sampling method for correcting the RTO samples had numerical issues.

2.2. Metropolis Adjusted Langevin Algorithm. In nonlinear inverse problems, standard random walk Metropolis samplers and even adaptive methods quickly become infeasible as the dimension of the problem grows. For instance, in the synthetic EIT test case in Section 3.2.3, the Delayed Rejection Adaptive Metropolis algorithm (DRAM, [10]) fails to comprehensively explore the target within a reasonable number of iterations. In high-dimensional nonlinear problems, it is crucial that the sampler is guided efficiently using additional information about the shape of the target distribution, such as gradients. One such MCMC method is the Metropolis Adjusted Langevin algorithm (MALA, [23]). In Section 3, we compare RTO to MALA in terms of efficiency. Here, we briefly present a preconditioned version of the MALA algorithm, following the notation of [18].

MALA is based on the Langevin stochastic differential equation (SDE)

\[
d\gamma = -A \nabla \gamma (\log p(\gamma | v, \lambda, \delta)) dt + \sqrt{2} A^{1/2} dW,
\]

(2.10)

where \( W \) is a vector of Brownian motions, \( \nabla \gamma (\log p(\gamma | v, \lambda, \delta)) \) is the gradient of the negative log-posterior and \( A = A^{1/2}(A^{1/2})^T \) is a positive definite preconditioning matrix. This SDE has the posterior \( p(\gamma | v, \lambda, \delta) \) as its stationary distribution [26], which means that the trajectories of the solution sample the posterior. Discretizing the above SDE with the Euler-Maruyama method yields the following update, which we will use as a proposal in the Metropolis-Hastings sampler:

\[
\gamma^* = \gamma_k - A \nabla \gamma (\log p(\gamma_k | v, \lambda, \delta)) \Delta t + \sqrt{2} \Delta t A^{1/2} z,
\]

(2.11)

where \( z \sim N(0, I) \). The preconditioning matrix \( A \) is needed to scale the proposals efficiently. A good choice is to set \( A = H(\gamma_k)^{-1} \), where \( H(\gamma_k) \) is the Hessian of the negative log-posterior. This choice yields the stochastic Newton sampler [18], which resembles the Newton’s method for optimization. Setting \( A = I \) gives the standard Langevin proposal, which resembles the steepest descent optimizer.

Computing local Hessians at each point can be computationally very challenging. In practice, good sampling performance is often obtained using a fixed Hessian.
computed at the MAP estimate, $A = H(\gamma_{\text{MAP}})^{-1}$. In our applications, we use the standard Gauss-Newton approximation for the likelihood part of the Hessian, which, in our notation, gives $H(\gamma_{\text{MAP}}) \approx \lambda J(\gamma_{\text{MAP}})^T J(\gamma_{\text{MAP}}) + \delta L$.

The step size $\Delta t$ in MALA is a tuning parameter that is similar to the scale of the proposal covariance matrix in standard random walk samplers. In our implementation, we use an adaptive method that attempts to tune $\Delta t$ during the sampling so that a desired acceptance probability $\alpha$ is achieved. We target $\alpha = 0.58$, following the optimal scaling results for MALA [21].

Computationally, the Hessian preconditioned MALA is obviously lighter than RTO, since for each sample a single linear system of equations with coefficient matrix $\lambda J(\gamma_{\text{MAP}})^T J(\gamma_{\text{MAP}}) + \delta L$ must be solved. However, the quality of the RTO samples is typically better. RTO and MALA are compared for the EIT problem in Section 3.

3. RTO-MH applied to the synthetic EIT data. In this section, we test the RTO-MH sampling method on the EIT inverse problem. We begin with a brief description of the EIT model in Section 3.1.

3.1. The EIT Model and its Numerical Solution. Electrical impedance tomography (EIT) is a diffusive imaging modality which is based on electrical boundary measurements. In EIT, a set of alternating currents are injected into the object through an array of boundary electrodes, and the resulting potentials on the electrodes are measured. Based on these measurements, the internal conductivity distribution within the object is reconstructed. The applications of EIT include, e.g., industrial process monitoring [30], medical imaging [5], geophysical exploration [3] and non-destructive testing [16].

The best available model for physically realizable quasi-stationary measurements is referred to as the complete electrode model (CEM), which is of the form [6]

$$\nabla \cdot (\gamma \nabla \varphi) = 0, \quad \vec{r} \in \Omega \quad (3.1)$$
$$\varphi + z_\ell \gamma \frac{\partial \varphi}{\partial \vec{n}} = \Phi_\ell, \quad \vec{r} \in e_\ell, \quad \ell = 1, \ldots, L \quad (3.2)$$
$$\int_{e_\ell} \gamma \frac{\partial \varphi}{\partial \vec{n}} dS = I_\ell, \quad \ell = 1, \ldots, L \quad (3.3)$$
$$\gamma \frac{\partial \varphi}{\partial \vec{n}} = 0, \quad \vec{r} \in \partial \Omega \cup \bigcup_{\ell=1}^L e_\ell \quad (3.4)$$

where $\gamma = \gamma(\vec{r})$ and $\varphi = \varphi(\vec{r})$ are the electrical conductivity and potential distribution, respectively, and $\vec{r} \in \Omega$ denotes the spatial coordinate. Further, $e_\ell$ is the area under the $\ell$th electrode, $z_\ell$ is the contact impedance between the $\ell$th electrode and the object, $\Phi_\ell$ and $I_\ell$ are the amplitudes of the electrode potential and current, respectively, $\vec{n}$ is the outward unit normal and $L$ is the number of electrodes. In addition, the charge conservation law must be satisfied and a reference level of the potential $\varphi$ needs to be fixed:

$$\sum_{\ell=1}^L I_\ell = 0, \quad \sum_{\ell=1}^L \Phi_\ell = 0. \quad (3.5)$$

It has been shown that conditions (3.5) ensure the uniqueness and existence of the (weak) solution of the CEM [25].

The forward problem of EIT is computing the potential distribution $\varphi = \varphi(\vec{r})$ and the electrode potentials $\Phi_\ell$, given the conductivity distribution $\gamma(\vec{r})$, the contact
impedances $z_\ell$ and the current pattern $I = (I_1, \ldots, I_L)$. The solution of the forward problem can be approximated using the finite element method (FEM), based on the weak form of (3.1)–(3.5) written in [25]. For the finite element (FE) approximation of the CEM, see [27]. The resulting observation model is of the non-linear form (1.7) where $v$ is a vector consisting of electrode potentials corresponding to a set of different current patterns $I$. Vector $\gamma$ is a finite dimensional representation of the conductivity distribution $\gamma(\vec{r})$. In this paper we represent $\gamma$ using a piecewise linear basis, while the potential distribution $\varphi$ is approximated using a piecewise second order polynomial basis. The mapping $U$ is linear with respect to electrode currents $I$ and non-linear with respect to conductivity $\gamma$ and the contact impedances $z_\ell$. In this paper, we assume that the contact impedances are known. The observation noise $\epsilon$ is assumed to be Gaussian $\epsilon \sim N(0, \sigma^2 I_n)$. The variance of the noise, $\sigma^2$, can usually be determined experimentally, see [11]. In the following numerical studies we assume that $\lambda^{-1} = \sigma^2 = (2 \times 10^{-4} \cdot \max_i(v_i))^2 I_n$. This is known to be an accurate assumption for a certain commercial EIT device built by Numcore Ltd, Finland.

The statistical inverse problem of EIT is to construct the posterior distribution of the conductivity $\gamma$ given the noisy observations $v$. In the next sections, we apply RTO-MH to compute the approximate statistics of the posterior distribution.

3.2. Numerical Experiments. We apply the proposed sampling method to numerical experiments with four different target conductivities. In Section 3.2.1, we consider a case where the smoothness prior model is ‘correct’ in the sense that the true conductivity is a realization from the prior probability distribution. In Section 3.2.2, two different targets with annular internal structures are considered. Here, a structural smoothness prior model is used. Finally, in Section 3.2.3, the true target conductivity consists of a sharp conductive inclusion on a homogeneous background, while the smoothness prior is used as the prior probability model. In the last section, we also compare the results of the RTO based sampling method to results given by the delta method (sometimes referred to as the Laplace approximation) and to MALA.

3.2.1. Target 1: smooth conductivity distribution. In the first test case, the target conductivity is selected to be a sample drawn from a probability distribution promoting spatial smoothness. The same probability distribution is used as the prior model in the inversion. The model is written as follows. As in [17], the covariance matrix $C = \frac{1}{5} L^{-1}$ is constructed by writing

$$C(i,j) = a \exp \left\{ -\frac{||r_i - r_j||^2}{2b^2} \right\} + c \delta_{ij} \quad (3.6)$$

where $r_i = (x_i, y_i)$ is the spatial coordinate corresponding to a discrete conductivity value $\gamma_i$, and $a$, $b$ and $c$ are parameters that control the variance and the spatial smoothness of the conductivity. Clearly, the variance of the conductivity at point $r_i$ is $\text{var}(\gamma_i) = C(i,i) = a + c$. Here, we define the correlation length $\ell$ as the distance where the cross-covariance $C(i,j)$ drops to 1% of the $\text{var}(\gamma_i)$. It is easy to show that in the case of model (3.6), the connection between the parameter $b$ and $\ell$ is

$$b = \frac{\ell}{\sqrt{2 \ln(100) - 2 \ln(1 + c/a)}} \quad (3.7)$$

In the first test case, the correlation length is selected to be $\ell = 5$ cm and the variance is set to $\text{var}(\gamma_i) = 0.05 \text{ (mS/cm)}^2 \forall i$. In (3.6), the parameter $c$ is selected to be
relatively small; \( c = 10^{-3} \text{var}(\gamma_i) \). The expectation of the conductivity \( \bar{\gamma} \) is set to 1 mS/cm.

The synthetic measurement data is generated by solving numerically the forward problem of EIT (see Section 3.1) corresponding to a circular 2D target, and adding noise to the computed electrode potentials. For simulating EIT measurements, 16 equidistantly located boundary electrodes are modeled. The radius of the circular target is 10 cm, and the width of each electrode is 2 cm. Pairwise current injection patterns are selected, such that currents are injected between electrodes \( i \) and \( j \), where \( i = 1, 5, 9, 13 \), and \( 1 < j < 16, j \neq i \), resulting a total of 54 different current patterns. Here, the electrodes are numbered in counterclockwise order, beginning from the rightmost electrode. Corresponding to each current injection, the electrode potentials are measured between all adjacent electrodes. In the forward simulations, the number of elements in the FE mesh is 4172, and the numbers of basis functions for the conductivity 2183 and for the electrode potential 8537.

In RTO-MH sampling, 1000 samples from the posterior distribution are computed. To avoid an inverse crime, the FE mesh used in inverse computations is selected to be sparser than that used for generating the synthetic data; the number of elements in the inverse mesh is 1360. The number of basis functions for the conductivity – and hence, the dimension of the unknown \( \gamma \) – is 745. The number of piecewise second order polynomial basis functions for the electric potential is 2849.

In this test case, the acceptance probabilities in RTO-MH are high – 851 RTO samples out of 1000 are accepted in step 4 of the RTO-MH algorithm. This indicates that the RTO probability density is a good approximation for the posterior density. We also note that for all of the examples that we consider, it is very rarely the case that a sample \( \gamma \) is rejected because it is not in the range of \( F\bar{\gamma} \). Moreover, the condition that the Jacobian of \( F\bar{\gamma} \) is invertible has never failed in the examples we have considered.

The target conductivity distribution of the first test case (Target 1) and the corresponding results are shown in Figure 3.1. Both the MAP estimate and the CM estimate are depicted. Both estimates are represented in the same scale as the true distribution. Main features of the conductivity are tracked relatively well by both estimators, especially near the boundaries. In the center of the circular domain, the contrast is slightly lower. This is an expected result, because in general the EIT measurements are more sensitive to the conductivity in the neighborhood of the boundary than in the middle of the target. The difference between the MAP and CM estimate is very small.

We also compute the 99 % credibility limits for the conductivity. That is, for each conductivity parameter \( \gamma_i \), \( i = 1, \ldots, m \), we compute the estimates \( \gamma_i^{\text{min}} \) and \( \gamma_i^{\text{max}} \) such that \( p(\gamma_i < \gamma_i^{\text{min}}) = p(\gamma_i > \gamma_i^{\text{max}}) = 0.005 \) where \( p \) denotes the sample based estimate for the probability. We denote the width of the 99 % credibility interval, i.e. the spatial distribution of the difference \( \gamma_i^{\text{max}} - \gamma_i^{\text{min}} \), by \( \mathcal{W}_{99\%}^{\text{est}} \). The width of the 99 % credibility interval is illustrated in Figure 3.1 (bottom right). Clearly, the uncertainty is highest in the middle of the target and smallest near the boundary. This is again a result of EIT measurements being more sensitive to the conductivity near the boundary than the conductivity in the center. Note that since the prior model is Gaussian, the 99 % credibility limits for the conductivity before the EIT measurements are \( E(\gamma) \pm 2.58 \sqrt{\text{var}(\gamma_i)} \). Hence \( \mathcal{W}_{99\%}^{\text{pr}} \), the width of the 99 % credibility interval before the measurements, is \( \mathcal{W}_{99\%}^{\text{pr}} = 2 \cdot 2.58 \sqrt{\text{var}(\gamma_i)} \equiv 1.1538 \text{ mS/cm} \), which is much larger than the posterior credibility interval seen on the bottom right in Figure 3.1.
Fig. 3.1. Target 1: the true conductivity distribution (top left), the MAP estimate (top right), the CM estimate (bottom left) and the estimated width of the 99% credibility interval $W_{99\%}^{\text{post}}$ (bottom right).

The profile of the target conductivity along line $x = y$ is drawn in Figure 3.2. Also the profiles of the CM estimate and the sample based 99% posterior credibility limits are plotted. The true conductivity profile lies clearly between the 99% credibility limits.

3.2.2. Targets 2 and 3: internal structures. Next, we test the RTO-MH sampling with two targets (Targets 2 and 3), which both include an annular internal structure. The two target conductivities are shown in Figures 3.3 and 3.5 (top left). In both targets, the annular structure is centered at (2 cm, 2 cm), and the inner and outer radii of the annulus are 3.5 cm and 5 cm, respectively. We define the subdomains $\Omega_1$, $\Omega_2$ and $\Omega_3$ of the target $\Omega = \{ \| \mathbf{r} \|_2^2 < (10\, \text{cm})^2 \}$ as

\[
\begin{align*}
\Omega_1 &= \{ \| \mathbf{r} - (2\, \text{cm}, 2\, \text{cm}) \|_2^2 < (3.5\, \text{cm})^2 \} \\
\Omega_2 &= \{ (3.5\, \text{cm})^2 \leq \| \mathbf{r} - (2\, \text{cm}, 2\, \text{cm}) \|_2^2 \leq (5\, \text{cm})^2 \} \\
\Omega_3 &= \{ \mathbf{r} \in \Omega, \| \mathbf{r} - (2\, \text{cm}, 2\, \text{cm}) \|_2^2 > (5\, \text{cm})^2 \}
\end{align*}
\]

and the two target conductivity distributions as

\[
\gamma(\mathbf{r}) = \begin{cases} 
1.5 \, \text{mS/cm}, & \mathbf{r} \in \Omega_1 \\
\gamma_{\text{ann}}, & \mathbf{r} \in \Omega_2 \\
1 \, \text{mS/cm}, & \mathbf{r} \in \Omega_3.
\end{cases}
\]

In both targets, the conductivity of the annular structure $\gamma_{\text{ann}}$ is lower than the background conductivity: in Target 2, $\gamma_{\text{ann}} = 0.5 \, \text{mS/cm}$, and in Target 3, $\gamma_{\text{ann}} = 1 \, \text{mS/cm}$. The profile of the target conductivity along line $x = y$ is drawn in Figure 3.2. Also the profiles of the CM estimate and the sample based 99% posterior credibility limits are plotted. The true conductivity profile lies clearly between the 99% credibility limits.
0.05 mS/cm. Hence, the only difference between the two test cases is that in Target 3, the conductivity contrast between the annulus and the background is higher than in the case of Target 2.

The EIT measurements are generated as in the case of Target 1. For the forward simulations, the FE mesh is constructed so that the mesh element boundaries follow the edges of the annulus; this allows for a relatively smooth approximation of the annulus boundary shape. The number of elements in the FE mesh is 6834, and the numbers of basis functions for the conductivity are 3530 and for the electrode potential 13893.

When solving the inverse problem, we assume that the position and shape of the annular structure are known, and construct a structured mesh also for the inverse computations. The number of elements in the inverse mesh is 5168, and the numbers of basis functions for the conductivity are 2681 and 10529, respectively. We construct the prior model taking into account the structure as follows; for an alternative construction of a structural smoothness prior, see [13]. Define index sets $\mathcal{I}_1, \mathcal{I}_2$ and $\mathcal{I}_3$, such that $i \in \mathcal{I}_k$, when $r_i \in \Omega_k$, $k = 1, 2, 3$. We write the covariance $C(i,j)$ between conductivities corresponding to spatial coordinates $r_i, i \in \mathcal{I}_k$ and $r_j, j \in \mathcal{I}_t$ in the form of Equation (3.6) when $k = t$ and set $L^{-1}(i,j) = C(i,j) = 0$, when $k \neq t$. This means that the conductivities in the three different subdomains – inside the annulus ($\Omega_1$), annular structure ($\Omega_2$) and outside the annulus ($\Omega_3$) – are mutually uncorrelated, while the conductivity within each subdomain is correlated. The correlation length $\ell$ corresponding to conductivity within the annular structure $\Omega_2$ is set very large, $\ell = 10^8$ cm. This practically implies that the conductivity of the annulus is constant. In $\Omega_1$ and $\Omega_3$, by contrast, the correlation length is only $\ell = 4$ cm.

Both the expectation and the variance of the conductivity are set to be constant.
in the whole domain: $\bar{\gamma} = 1$ mS/cm and $\text{var}(\gamma_i) = 0.0278$ (mS/cm)$^2$. The selected prior variance corresponds to 99% credibility limits $(1 \pm 0.43)$ mS/cm.

In RTO-MH sampling, 500 samples from the RTO probability density are drawn in both test cases. The acceptance probabilities are again relatively high: 416 and 438 RTO samples are accepted in cases 2 and 3, respectively.

The results corresponding to Target 2 are illustrated in Figure 3.3. The MAP and the CM estimates and the width of the 99% credibility interval $W_{\text{post}}^{99\%}$ are depicted. Again, the conductivity distribution is tracked relatively well by both MAP and CM estimators. The image of $W_{\text{post}}^{99\%}$ reveals that in the subdomains $\Omega_1$ and $\Omega_3$ the posterior uncertainty increases towards the center of the target. In the subdomain $\Omega_2$, however, $W_{\text{post}}^{99\%}$ is almost constant. This is a result of the very strong smoothness assumption corresponding to subdomain $\Omega_2$ included in the prior model. This also explains why $W_{\text{post}}^{99\%}$ is significantly smaller in $\Omega_2$ than in $\Omega_1$ and $\Omega_3$, even though $W_{\text{pr}}^{99\%}$ is constant in the whole domain.

In Figure 3.4 the profiles along line $x = y$ are drawn. In subdomains $\Omega_1$ and $\Omega_3$, the true conductivity is clearly between the 99% posterior credibility limits. Note that the conductivity is estimated relatively well also in subdomain $\Omega_1$, where the true conductivity is higher than the upper 99% prior credibility limit 1.43 mS/cm. This result demonstrates the robustness of the method with respect to the prior model: the true conductivity in $\Omega_1$ lays inside the 99% posterior credibility interval, even though it does not lay inside the 99% prior credibility interval. Also in $\Omega_2$, the true conductivity which is lower than the lower 99% prior credibility limit, is only slightly
outside the narrow 99 % posterior credibility interval.

The results corresponding to Target 3 are shown in Figures 3.5 and 3.6. In subdomains $\Omega_2$ and $\Omega_3$, the reconstructed conductivities are again feasible. Especially in subdomain $\Omega_2$, the reconstructed conductivity is very close to the true value $\gamma_{\text{ann}} = 0.05$ mS/cm, and the 99 % posterior credibility interval is narrow in this region. The posterior estimate of $\Omega_2$ conductivity is very good, taking into account that in this test case $\gamma_{\text{ann}}$ is relatively far from the 99 % prior credibility interval. Further, since the conductivity of the annular structure $\Omega_2$ is very low, it practically acts like an electrical insulator. For this reason, the electric current density through the annulus is small, and the sensitivity of the EIT measurements to the conductivity in $\Omega_1$ (inside the annular structure) is low. Since the EIT measurements carry practically no information on the conductivity inside the annulus, the posterior probability distribution of the conductivity in $\Omega_1$ is almost equal to the prior distribution. This hypothesized effect is clearly visible in Figures 3.5 and 3.6: 1) both in MAP and CM estimates, the reconstructed conductivity in $\Omega_1$ is close to the prior expectation of the conductivity, 1 mS/cm; and 2) the width of the 99 % posterior credibility interval $W_{\text{post}}^{99\%}$ is very high in $\Omega_1$. The profiles plotted in the Figure 3.6 reveal that in $\Omega_1$ ($r \in [-0.5\text{cm}, 4.5\text{cm}]$), the 99 % posterior credibility limits are close to 99 % prior credibility limits ($1 \pm 0.43$) mS/cm, as predicted. The posterior estimate of the conductivity in subdomain $\Omega_1$ in the case of Target 3 differs remarkably from the that in case of Target 2, in which the conductivity contrast between subdomains $\Omega_2$ and $\Omega_3$ was lower. These results support the feasibility of the RTO-MH method for estimating the posterior distributions in non-linear inverse problems.

3.2.3. Target 4: Sharp conductive inclusion. As the last numerical example, we consider a case where the true conductivity consists of a sharp conductive inclusion on a homogeneous background. The conductivity of the inclusion is 1 mS/cm
Fig. 3.5. Target 3: the true conductivity distribution (top left), the MAP estimate (top right), the CM estimate (bottom left) and the estimated width of the 99% credibility interval $W_{99\%}^{post}$ (bottom right).

Fig. 3.6. Target 3: Profiles of the true conductivity (solid gray line), the CM estimate (solid black line), the 99% posterior credibility limits of the conductivity (dashed black lines) and the 99% prior credibility limits of the conductivity (dotted gray lines).
and the background conductivity is 0.2 mS/cm. The true conductivity distribution is shown in Figure 3.7 (top left).

As the prior probability, we use the Gaussian smoothness prior model introduced in Section 3.2.1, with covariance matrix defined by (3.6) and (3.7), with $\ell = 8$ cm, $\text{var}(\gamma_i) = 1 (\text{mS/cm})^2 \forall i$, $c = 10^{-3}\text{var}(\gamma_i)$, and $a = 0.999$. Thus $b = 8/\sqrt{2\ln(0.01)}$, which defines (3.6) completely.

Note that the Target 4 does not feature smoothness in the whole domain, because the edges of the inclusion are sharp, i.e., the smoothness prior is not an optimal prior model in the selected test case. In this example, the FE mesh used in the inverse computations is lower than in the previous sections; here, the dimension of the unknown $\gamma$ is 745. We sample from the posterior density by computing an MCMC chain of length 1000.

The MAP and CM estimates are shown in Figure 3.7. The position of the conductive inclusion is tracked relatively well. However, the conductivity of the inclusion is underestimated, and the inclusion is significantly smoother than that in the true conductivity. These are clearly effects of the smoothness prior.

As above, we also plot the element-wise 99% credibility interval $W_{99\%}^\text{post}$ in the lower-right in Figure 3.7. As in Case 1, the posterior credibility interval profile is lowest near the boundary where the sensors are located. However, unlike in Case 1, the profile is not
Fig. 3.8. Target 4: element-wise sample variance from the RTO samples (left) and using the delta method (right).

(even approximatively) symmetric or highest in the middle of the domain. In contrast, the posterior samples have highest variance in the position of the conductive inclusion. This is again an expected result, because the contrast between the conductive inclusion and the background is high; as pointed out in [16], after a certain limit, the increase of the conductivity of the inclusion has only a very small effect on EIT measurements, if the background conductivity remains unchanged. In other words, in such a case the inference of the conductivity of the inclusion based on EIT data solely is subject to high uncertainty. Of course, the complementary information carried by the prior model also affects the posterior uncertainty; nevertheless, when the conductivity of the inclusion is high, the posterior uncertainty at the location of the inclusion is generally higher than in the surrounding background.

For comparison, in Figure 3.8 we plot the element-wise sample variance of $\gamma$ together with the approximation of the variance of $\gamma$ given by the diagonal values of the matrix $\lambda^{-1}(\lambda J(\gamma_{\text{MAP}})T J(\gamma_{\text{MAP}}) + \delta L)^{-1}$; note that the delta method in statistics uses the Gaussian with this covariance matrix and mean $\gamma_{\text{MAP}}$ to compute approximate confidence intervals for parameters estimated from nonlinear models. We see in Figure 3.8 that the delta method yields an element-wise variance approximation that is noticeably (though not significantly) larger than the element-wise sample variance. Thus, although sampling from the posterior using RTO is more computationally demanding, for this example it yields tighter confidence bounds than the delta method.

**Comparison to MALA.** We compare RTO to MALA in the 745 dimensional synthetic EIT problem discussed above. We sample 1000 samples using RTO and 20000 samples using MALA. Our goal is to compare the relative effective sample size (ESS) for each method, taking into account the computational effort needed to produce the samples.

The ESS for the $i$th variable $\gamma_i$ is commonly defined as $\text{ESS}_i = N/\tau_i$, where $\tau_i$ is the integrated autocorrelation time, defined as $\tau_i = 1 + 2 \sum_{k=1}^\infty c_i(k)$, where $c_i(k)$ is the lag-$k$ autocorrelation of variable $i$. To estimate $\tau_i$, we use Sokal’s adaptive periodogram estimator [24]. The ESS gives an estimate of how many independent samples the MCMC chain of length $N$ corresponds to.

The quality of RTO samples is often much better than that of MALA in terms of
ESS alone. However, the computational cost for producing one sample is higher. To take this into account, we compute a relative effective sample size \( \text{RESS}_i = \frac{\text{ESS}_i}{M} \), where \( M \) is the total number of model/Jacobian evaluations used. In MALA, only one new evaluation is needed per sample \( (M = N) \). In RTO, the number \( M \) depends on how fast the optimizations converge. The relative ESS gives an estimate of how many independent samples per evaluation the samplers produce. In the synthetic EIT case, the extra effort for computing the correction factor (2.5) amounts to approximately one evaluation, which is taken into account in the comparison.

The results are summarized in Table 3.1. In this example, the optimizations within RTO converged fast; around 7-10 iterations were needed in average. In addition, the autocorrelations for RTO were remarkably smaller than for MALA. Taking the computational effort into account, RTO yields about 0.05 independent samples per model/Jacobian evaluation, around 36 times more than MALA. That is, in this particular case, RTO was more than an order of magnitude faster than MALA.

Table 3.1

<table>
<thead>
<tr>
<th></th>
<th>( N ) and ( M )</th>
<th>( \tau ): mean [min-max]</th>
<th>( \text{RESS} ): mean [min-max]</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALA</td>
<td>( N=M=20000 )</td>
<td>891.4 [223.5-2358.7]</td>
<td>0.0014 [0.0004-0.0045]</td>
</tr>
<tr>
<td>RTO</td>
<td>( N=1000, M=9297 )</td>
<td>2.2 [1.2-3.9]</td>
<td>0.051 [0.028-0.087]</td>
</tr>
</tbody>
</table>

4. Conclusions. We have shown that the randomize-then-optimize method (RTO) presented in [2] can be used to sample from the posterior density function that arises in the solution of the large-scale, nonlinear EIT inverse problem. RTO is implemented as follows. First, the data and the prior are resampled (randomize), and then a modified MAP estimation problem is solved (optimize) yielding an RTO sample. The form of the probability density function for the RTO samples is given, and RTO is used as an independence proposal within the Metropolis-Hastings algorithm, yielding the RTO Metropolis-Hastings (RTO-MH) method.

We test RTO-MH on four synthetic EIT examples, and in all cases, the sample mean and true conductivity have good agreement. Moreover, we use the samples to compute 99% credibility intervals for the conductivity, which in all cases make sense intuitively. In the last example, we compare the RTO sample mean and sample variance with those obtained using the MAP estimate and the approximate variance computed using the delta method. The sample mean and MAP estimates are very similar, while the sample variance is slightly smaller than the approximation computed using the delta method.

Finally, we compare RTO-MH with the state of the art Metropolis Adjusted Langevin Algorithm (MALA) sampling scheme. Both were applied to a synthetic EIT example, and per independent sample, MALA was found to require approximately 36 times more model/Jacobian evaluations than RTO-MH, making RTO-MH significantly more computationally efficient.

Acknowledgments. The second author would like to acknowledge the support provided by Academy of Finland (projects 270174 and 273536).

REFERENCES

17


