

Optimal experimental design for reservoir monitoring

J. Fohring and E. Haber

November 7, 2014

Abstract

Efficiently monitoring subsurface flow using geophysical imaging techniques is important to several applications such as aquifer characterization or enhanced oil recovery projects. To increase efficiency, the optimal number of measurements and their locations required for a geophysical survey are computed by an A-optimal experimental design method for ill-posed problems. A-optimal design minimizes the prior covariance matrix associated with target subsurface model. To obtain a prior covariance matrix for the future unknown model, the posterior covariance matrix associated with the current MAP model estimate is propagated forward in time. The MAP estimate is computed from historic geophysical data. The method is demonstrated for a two dimensional borehole seismic tomography experiment and a three dimensional DC resistivity survey. An advection model was applied to describe subsurface flow and propagate the error covariance matrices forward in time.

1 Introduction

- Monitoring; Many problems in geophysics address the issue of monitoring subsurface flow, such as aquifer characterization, enhanced oil recovery, or saltwater intrusion. In general one would like to image the motion of certain subsurface parameters over a period of time. One method to do so is to collect measurements dependent on these parameters and invert through a mathematical relationship to recover parameter distributions. When monitoring for these dynamic parameters, the choice of where the data are collected becomes important as resolution is both temporally and spatially dependent.
- much has been done in experimental designs for linear and nonlinear problems...A optimal, etc.. appropriate references. inverting for fluid parameters given geophysical imaging...refs...talk about seismic, and helmoltz and simultaneous sources
- Determining an the design of the data collection experiment requires knowledge of the target parameter distribution. Typically optimal experimental design methods rely on minimizing the posterior error associated with inverting and thus imaging the

parameter distribution. This poses a problem, particularly for static cases, since the point of the data collection is to image the unknown target. However, if the dynamics governing the motion of the unknown target are known, even to a small degree in accuracy, then it is possible to design the data collection experiment to be conducted in the future as the target moves. In order to design for a future experiment it is necessary to know the initial distribution of target parameters as well as the dynamics. If the initial condition is provided with the dynamics, then either an iterative approach or and all at once approach can be taken to determine data collection designs for a set of time points.

- An all at once method would seek to recover all of the designs for a set of time points given the dynamics and the initial condition. This method requires an 'all at once' prior covariance matrix. The generation of such a matrix is difficult... This method also puts a great deal of trust into the dynamic model. If for example, the velocity field in which a tracer is advected is known to very low accuracy then the designs and location of the tracer target far into the future would have significant error.
- An iterative or adaptive design method would make use of data collected for each experiment to update both the current covariance matrix and can be used to update the dynamic model parameters as well. This way, after each data collection experiment is conducted a design for the next time is computed with higher accuracy information.
- computational limitations: covariance matrices are large and dense, not necessarily spd, making factorization/ reduction techniques difficult. geophysical experiments are computationally expensive. Helmholtz- simultaneous source...

2 A-Optimal experimental design for linear ill-posed problems

To begin, a brief review of A-optimal experimental design for ill-posed problems is given as presented in full detail in [?].

Consider the discrete static ill-posed inverse problem of estimating model parameters $\mathbf{m} \in \mathbb{R}^N$ from noisy observation data $\mathbf{d} \in \mathbb{R}^M$,

$$\mathbf{d} - \mathbf{F}\mathbf{m} = \epsilon,$$

where \mathbf{F} is an $M \times N$ matrix with each row representing an experiment.

Assuming a Bayesian approach is taken with the posterior estimate $\hat{\mathbf{m}}$ conditioned on a known prior distribution $\pi_m = N(\mu, \Sigma_m)$, and that the error ϵ , is Gaussian $N(0, \mathbf{W}^{-1})$, the Bayesian posterior estimate is obtained by minimizing the negative log-likelihood of Bayes rule,

$$\begin{aligned} \hat{\mathbf{m}}(\mathbf{w}) &= \operatorname{argmin} (\mathbf{F}\mathbf{m} - \mathbf{d})^\top \mathbf{W}(\mathbf{F}\mathbf{m} - \mathbf{d}) + (\mathbf{m} - \mu)^\top \Sigma_m^{-1}(\mathbf{m} - \mu) \\ &= (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_m^{-1})^{-1}(\mathbf{F}^\top \mathbf{W} \mathbf{d}_w + \Sigma_m^{-1} \mu). \end{aligned}$$

The weighting matrix $\mathbf{W} = \text{diag}[\mathbf{w}]$ and $\mathbf{w} = (\omega_1, \dots, \omega_M)^\top$, relates to the data variance as $\text{Var}(d_i) = 1/\omega_i$.

A-optimal experimental designs for well-posed inverse problems are defined by minimizing the average variance of the estimate $\hat{\mathbf{m}}$, with covariance matrix $\Sigma_{\mathbf{m}}$ [?]. To this end, minimizing the average variance is equivalent to minimizing the trace of the posterior covariance matrix,

$$\phi_{AB}(\mathbf{w}) = \text{tr}[(\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{\mathbf{m}}^{-1})^{-1}]. \quad (1)$$

Given ϕ_{AB} , a design can be computed by solving a sparsity controlled optimal design optimization problem [?]. The design weights ω_i are determined by minimizing $\phi(\mathbf{w})$ with an l_1 penalty on \mathbf{w} to promote sparsity,

$$\hat{\mathbf{w}} = \arg \min \phi(\mathbf{w}) + \beta \|\mathbf{w}\|_1 \quad (2a)$$

$$\text{s.t. } 0 \leq \omega_i. \quad (2b)$$

It is important to note that in this formulation $\hat{\mathbf{m}}$ is chosen to be a linear (or affine) function of the data, although other estimators could be used, such as a Tikhonov estimator. Since the idea is to design experiments at future time steps for dynamic problems where the estimators covariance matrix can be estimated, the simple Bayesian estimator is sufficient.

2.1 Designing experiments for dynamic model parameters

To illustrate the method of estimating the covariance matrix associated with a future model, we first consider the dynamic tracer advection problem [?], which describes the transport of a solute in a fully saturated porous media

$$\begin{aligned} \frac{\partial(c\rho)}{\partial t} + \nabla \cdot c\rho \vec{\mathbf{u}} &= 0, \\ \text{subject to } c\rho(0, \vec{\mathbf{x}}) &= c_0\rho(\vec{\mathbf{x}}), \end{aligned} \quad (3a)$$

with hydraulic conductivity \mathbf{K} , and tracer concentration c in a fluid of density ρ . The fluid velocity field $\vec{\mathbf{u}}$ satisfies a simple source-sink (q) flow model

$$\nabla \cdot \vec{\mathbf{u}} = q \quad (3b)$$

$$\vec{\mathbf{u}} = \mathbf{K}(\vec{\mathbf{x}}) \nabla p \quad (3c)$$

$$p(0, \vec{\mathbf{x}}) = p_0(\vec{\mathbf{x}}), \quad (3d)$$

with either Dirichlet or Neumann boundary conditions on the pressure p .

With the goal being to image the motion of the tracer, $c(t, \vec{\mathbf{x}})$, we assume that changes in c amount to changes in geophysical properties. This implies that $m(c)$ exists, and that the motion of m is governed by Equation (3a).

Since Equation (3a) is linear with respect to c , a particle in cell discretization was chosen. Applying this discretization technique on a uniform grid with c at the cell centers and $\vec{\mathbf{u}}$ on the edges, results in the following linear transformation:

$$\mathbf{m}_{k+1} = \mathbf{T}(\mathbf{u})\mathbf{m}_k, \quad (4)$$

where the transport matrix \mathbf{T} is an interpolation matrix dependent only on the velocity field \mathbf{u} , see [?] for more details.

Given an imaging data set for time k , an estimate $\hat{\mathbf{m}}_k$ can be computed in the usual way such that $\hat{\mathbf{m}}_k$ is a random variable with a Gaussian distribution. Since the linear transformation of a Gaussian random variable is also a random variable, then

$$\begin{aligned} \hat{\mathbf{m}}_{k+1} &= \mathbf{T}(\mathbf{u})\hat{\mathbf{m}}_k, \\ \Sigma_{k+1} &= \mathbf{T}\Sigma_{\hat{\mathbf{m}}_k}\mathbf{T}^\top, \end{aligned}$$

(See [?] for a complete proof). That is, the posterior covariance matrix $\Sigma_{\hat{\mathbf{m}}_k}$ associated with $\hat{\mathbf{m}}_k$ is propagated forward in time with the known dynamics of the problem. This provides the prior distribution required to compute an A-optimal experimental design for the next survey at the time point $k + 1$.

The basic algorithm illustrating the design process for a linear observation method is outlined in Algorithm 1. In this algorithm $\hat{\mathbf{m}}_k$ is estimated using a Tikhonov method with a

Algorithm 1 Iterative Optimal ϕ_{AB} Design

- 1: Initialize: Σ_0 ▷ estimate from prior information before going to the field!
 - 2: **for** $k = 1$ to n **do** ▷ n is the maximum number of experiments, or time points
 - 3: $\hat{\mathbf{w}}_k(\Sigma_k) \rightarrow \text{diag}(\hat{\mathbf{w}}_k)\mathbf{F} = \mathbf{F}_k$ ▷ compute the optimal design by solving 2
 - 4: $\mathbf{d}_k = \mathbf{F}_k\mathbf{m}_k + \epsilon_k$ ▷ collect data
 - 5: $\hat{\mathbf{m}}_k = (\mathbf{F}_k^\top\mathbf{F}_k + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{F}_k^\top\mathbf{d}_k$ ▷ estimate the current model
 - 6: $\Sigma_k \leftarrow \mathbf{T}(\mathbf{F}_k^\top\mathbf{F}_k + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\mathbf{T}^\top$ ▷ propagate the error
 - 7: **end for**
-

regularization operator \mathbf{L} . To design the experiment, only data from the current time step is used, and the velocity field \mathbf{u} is given.

2.2 Numerical optimization for dynamic A-optimal design

Finding an optimal design for the time step $k + 1$ involves solving the optimization problem 2 by minimizing the objective functional,

$$\mathcal{J}_{k+1}(\mathbf{w}) = \text{tr}[(\mathbf{F}^\top \text{diag}(\mathbf{w})\mathbf{F} + \Sigma_{k+1}^{-1})^{-1}] + \beta \mathbf{e}^\top \mathbf{w}, \quad (5)$$

where \mathbf{e} is a vector of ones.

This is a difficult task for large problems, particularly due to the trace of a large dense matrix, the computation of the derivative of the trace, and the requirement of the inverse of

the large, dense propagated posterior covariance matrix $\Sigma_{k+1} = \mathbf{T}(\mathbf{F}_k^\top \mathbf{F}_k + \alpha \mathbf{L}^\top \mathbf{L})^{-1} \mathbf{T}^\top$. To this end, several actions can be taken to reduce the computational costs.

First, a stochastic Hutchinson trace estimator is used to approximate the trace [?, ?] such that Equation (5) can be written as

$$\mathcal{J}_{k+1}(\mathbf{w}) = \mathbf{v}^\top (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{v} + \beta \mathbf{e}^\top \mathbf{w}, \quad (6)$$

where \mathbf{v} is a random vector with values 1 and -1 with equal distribution. This approximation makes the computation of the derivative of \mathcal{J} tractable for large problems.

A second issue arises when computing the derivative of \mathcal{J} , as this involves computing the derivative of an inverse matrix,

$$\nabla_{\mathbf{w}} \mathcal{J}_{k+1}(\mathbf{w}) = \mathbf{v}^\top \nabla_{\mathbf{w}} \left((\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{v} \right) + \beta \mathbf{e}.$$

To compute the derivative of \mathcal{J} we define

$$\mathbf{z} = (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{v} \quad \Leftrightarrow \quad (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1}) \mathbf{z} = \mathbf{v}, \quad (7)$$

and differentiate implicitly to obtain

$$\mathbf{F}^\top \text{diag}(\mathbf{F} \mathbf{z}) + (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1}) \nabla_{\mathbf{w}} \mathbf{z} = \mathbf{0}.$$

The gradient of \mathbf{z} is then given by

$$\nabla_{\mathbf{w}} \mathbf{z} = -(\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{F}^\top \text{diag}(\mathbf{F} \mathbf{z}).$$

Thus, the first term in $\nabla_{\mathbf{w}} \mathcal{J}_{k+1}(\mathbf{w})$ is

$$-\mathbf{v}^\top (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{F}^\top \text{diag}(\mathbf{F} \mathbf{z}).$$

Taking the transpose

$$-\text{diag}(\mathbf{F} \mathbf{z}) \mathbf{F} (\mathbf{F}^\top \mathbf{W} \mathbf{F} + \Sigma_{k+1}^{-1})^{-1} \mathbf{v} \quad \Leftrightarrow \quad -\text{diag}(\mathbf{F} \mathbf{z}) (\mathbf{F} \mathbf{z}) = -\mathbf{F} \mathbf{z} \odot \mathbf{F} \mathbf{z},$$

so that the derivative of the objective function is then

$$\nabla_{\mathbf{w}} \mathcal{J}_{k+1}(\mathbf{w}) = -\mathbf{F} \mathbf{z} \odot \mathbf{F} \mathbf{z} + \beta \mathbf{e}. \quad (8)$$

Since the design problem (2) is nonlinear with respect to \mathbf{w} , a Newton type iterative method is used with a backtracking line search. This solution method requires both the evaluation of \mathcal{J} and its gradient, which are conveniently obtained by solving for \mathbf{z} only once.

At this point it is important to note that computing \mathbf{z} requires the explicit inverse of the prior covariance matrix Σ_{k+1} . For small scale problems it is possible to compute and

store the inverse of this matrix and to use direct solvers to compute \mathbf{z} . However, there is no guarantee that $\Sigma_{\mathbf{k}+1}$ will be positive definite, nor in our case invertable.

For case where $\Sigma_{\mathbf{k}+1}$ is SPD, Lanczos bidiagonalization can be applied by first reducing the system (7) to standard form. This process also requires forming $\Sigma_{\mathbf{k}+1}$ and computing its Cholesky factorization. However, the Cholesky factorization is evaluated outside of the design optimization, so depending on the problem size, it is somewhat computationally cheap.

For large scale problems forming $\Sigma_{\mathbf{k}+1}$, computing its inverse explicitly, and storing the resulting large dense matrix becomes intractable. To avoid forming and computing such an expensive matrix, Krylov subspace methods can be applied, such as the Conjugate Gradient method. This allows us to take advantage of matrix vector products and avoid forming $\Sigma_{\mathbf{k}+1}$. If, as above, a Tikhonov estimate is used to estimate $\hat{\mathbf{m}}_k$ and thus Σ_k , then

$$\Sigma_{\mathbf{k}+1}^{-1} = \mathbf{T}^{-\top} (\mathbf{F}_k^\top \mathbf{F}_k + \alpha \mathbf{L}^\top \mathbf{L}) \mathbf{T}^{-1}. \quad (9)$$

In this way, solving for \mathbf{z} with an iterative method requires the solution to the systems:

$$\mathbf{T}(\mathbf{u})\mathbf{y} = \mathbf{z} \quad (10)$$

$$\mathbf{T}(\mathbf{u})^\top \mathbf{x} = (\mathbf{F}_k^\top \mathbf{F}_k + \alpha \mathbf{L}^\top \mathbf{L})\mathbf{y}. \quad (11)$$

Given that \mathbf{T} is sparse, and provided it is non-singular, then \mathbf{z} can be computed fairly efficiently without ever forming the posterior covariance at time k or its inverse, and thus never forming or storing $\Sigma_{\mathbf{k}+1}^{-1}$.

2.3 Understanding the covariance matrix

It is important to note that there is no guarantee that $\Sigma_{\mathbf{k}+1}$ will be positive definite. In fact it is possible that the transport matrix \mathbf{T} can be singular. This generally occurs for zero flux boundary conditions; the assumption being that there is no tracer coming in from outside of the domain.

The zeros in the transport matrix translate to entries in $\Sigma_{\mathbf{k}+1}$ with zero variances, or infinite certainty that there is no mass entering a cell from outside the domain. One way to rectify the singularity is to remove this information from the covariance matrix.

By introducing the $l \times N$ projection matrix \mathbf{P} , which maps from \mathbb{R}^N to the reduced space \mathbb{R}^l , the reduced model and reduced covariance matrix are given by

$$\hat{\mathbf{m}}_{\mathbf{k}+1}^r = \mathbf{P}\mathbf{T}\hat{\mathbf{m}}_k, \quad (12)$$

$$\Sigma_{\mathbf{k}+1}^r = \mathbf{P}\Sigma_{\mathbf{k}+1}\mathbf{P}^\top. \quad (13)$$

This is a reasonable approach since the experiment we are designing for depends on where the tracer is currently. Therefore, the cells associated with zero variances will not contribute to the optimal design. This method not only alleviates the singularities, but also reduces the overall size of the problem. [SHOW this here, what T looks like, what Sigma looks like](#)

A second approach requires the inclusion of boundary conditions. In particular, we redefine the estimated model at time $k + 1$,

$$\hat{\mathbf{m}}_{k+1} = \mathbf{T}\hat{\mathbf{m}}_k + \xi_k, \quad (14)$$

so that it is composed of the transported estimate plus a random model error which lives just outside the domain. See figure [need a figure to illustrate this](#).

In the following section, a simple example is presented to illustrate the different solution methods for a small problem with a given velocity field and small displacement so that \mathbf{T} is positive definite.

3 Numerical example 1: Iterative design, known flow model

To demonstrate iterative optimal experimental design, two geophysical imaging techniques are considered. First, a travel time seismic tomography geophysical survey is applied to image the advection of a tracer in the subsurface. An initial spatial distribution of the tracer is provided, and the tracer is advected with the flow of water injected from one well and pumped out at another. Seismic tomography data are measured at each time step as the initial tracer model is advected using the transport matrix \mathbf{T} . Secondly, a direct current (DC) resistivity survey applied to the same flow model. Since the DC problem results in a non-linear system of equations, we will actually consider the linearized case and design for the Jacobian, outlined further in section...

Three solution methods are examined when designing for these two geophysical surveys; solving the optimization problem given the full inverse of the prior covariance matrix, solving the using Lanczos bidiagonalization, and solving the optimization problem without forming the prior covariance matrix. Computational time will be compared for the three methods.

Since computational expense increases with the refinement of the spatial discretization, in section ... designs are compared for different mesh sizes with identical time steps to identify the possibility of designing experiments on a coarse mesh.

Later we will investigate the case where the velocity and initial tracer concentration are unknown.

3.1 Travel time tomography

A typical tomography experiment measures the travel times between sources n_s and receivers n_r . Travel times are given by integrating the slowness $s(t, \vec{\mathbf{x}})$, or inverse acoustic velocity of the media, along a ray path Γ ,

$$\mathbf{d}_{j,k}(t) = \int_{\Gamma_{j,k}} s(t, \vec{\mathbf{x}}) d\ell.$$

Assuming the data are noisy and measured at times (t_0, \dots, t_n) the tomography problem is given by

$$\mathcal{F}s(t_k, \vec{x}) + \epsilon_k = \mathbf{d}(t_k).$$

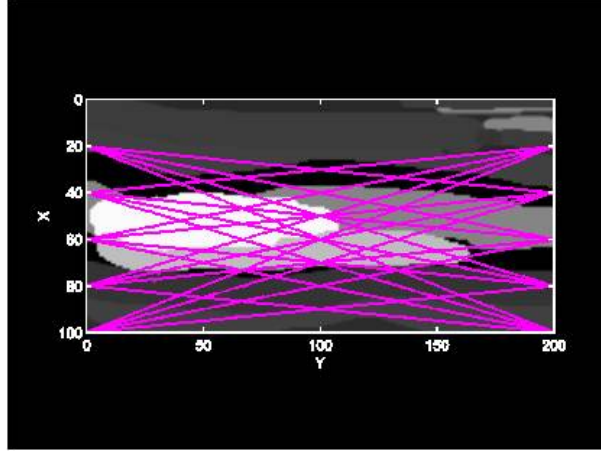


Figure 1: *Visualizations of hydraulic conductivity with tomography sources on the left and receivers on the right. The ray paths $\Gamma_{j,k}$ are shown in pink.*

Discretizing on a staggered 2D grid using a finite volume scheme places s and \mathbf{d} at the cell centers. Therefore the discrete tomography experiment at each time step is then given by $\mathbf{F}\mathbf{s}_k = \mathbf{d}_k$, for $k = 1, \dots, n$. The tomography experiment was setup with 20 sources in the east borehole and 20 receivers in the west borehole. The boreholes are 200m apart and 50m deep, as shown in figure....**better describe the experimental setup**

3.1.1 Comparing solution methods

To compare solutions methods, an experiment was conducted where data were collected from all source receiver combinations. Initial estimates of Σ_0 and the best Tikhonov regularization parameter α_0 were obtained by a standard cooling algorithm and Pareto curve .

3.1.2 Mesh refinement

compare for identical (small) time step, the designs. would be nice to get the same design results! Try to estimate the sensitivity of the design to the mesh size. how to compare? what metric?

3.1.3 Time step restriction

talk about the singularity of \mathbf{T} for time steps greater than h . What to do about this...

3.2 DC Resistivity

3.2.1 Comparing solution methods

The big difference is whether or not we have sigma...pcg is slow

3.2.2 Mesh refinement

mesh refinement. If designs can be done on a coarse mesh, then it's fine to form sigma inverse. This is powerful

3.2.3 Time step restriction

hopefully can figure out what to do here.

4 Numerical example 2: Iterative design, estimating the velocity as well

5 Concluding Remarks

6 References