Adaptive A-optimal experimental design for imaging dynamic targets

J. Fohring and E. Haber

March 12, 2015

Abstract

Efficiently monitoring subsurface flow using geophysical imaging techniques is important to several applications such as aquifer characterization or enhanced oil recovery projects. One method of increasing monitoring efficiency is to control the number of measurements required for a geophysical survey in an optimal way. To optimally design future geophysical experiments, we use A-optimal design methods while incorporating the flow dynamics.

A-optimal design methods minimize the posterior covariance matrix associated with estimating the target subsurface model. We reformulate the MAP estimate for the prior target in order to design for the future target, such that the problem is constrained by the flow dynamics. In this way, the prior experiments collected data and flow dynamics contribute to the design of the next experiment, thus incorporating updated information to the design algorithm.

The method is demonstrated for a two dimensional borehole seismic tomography experiment and a three dimensional seismic survey. An advection model is applied to describe subsurface flow and propagate the error covariance matricies 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. As well as for large scale ill posed, refs. inverting

1

for fluid parameters given geophysical imaging...refs...talk about seismic, and helmolz and simultaneous sources

- Determining 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 covariance matrix, associated with estimating the target through the inversion of geophysical data. 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 date collection designs for a set of time points.

- ideally, incorporation of real data in the estimation of historic, but also future experimental designs would lead to the best estimates. Adaptively, chase the target, by recomputing the design after data is collected.

- We also consider what happens if we wish to recover/update the parameters governing the fluid flow. How does this change designs for the future time step?

- computational limitations: (constrained all at once, no error) no historic info included, (Kalman formulation, error) covariance matrices are large and dense, making factorization/ reduction techniques difficult. geophysical experiments are computationally expensive. (adaptive, constrained) no covariance matrices to worry about, includes last data set/design, can easily tack on regularization.

2   Mathematical formulation

This section introduces the notation for the governing equations of the subsurface dynamics and borehole tomography geophysics experiment, presents a brief overview of A-Optimal experimental design, and finally discussion of our approach to designing experiments for the flow simulation while including historic experimental designs.

2.1   Tracer dynamics

The subsurface dynamics are described by the tracer advection equations as presented by [**?**], which characterize the transport of a solute in a fully saturated porous media,

$$\frac{\partial(c\rho)}{\partial t} + \nabla \cdot c\rho\vec{\mathbf{u}} = 0, \tag{1a}$$
$$\text{subject to } c\rho(0, \vec{\mathbf{x}}) = c_0\rho(\vec{\mathbf{x}}),$$

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

$$\nabla \cdot \vec{\mathbf{u}} = q \tag{1b}$$

$$\vec{\mathbf{u}} = \mathbf{K}(\vec{\mathbf{x}})\nabla p \tag{1c}$$

$$p(0, \vec{\mathbf{x}}) = p_0(\vec{\mathbf{x}}), \tag{1d}$$

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 (1a). From now on were refer only to m.

To discretize equations (1), particle in cell discretization was chosen such that equation (1a) is given by,

$$\mathbf{m}_k = \mathbf{T}(\mathbf{u})\mathbf{m}_{k-1}$$

and the fluid velocity field $\mathbf{u}$ was computed by solving equations (??) with hydraulic conductivities for $\mathbf{K}$ ranging from
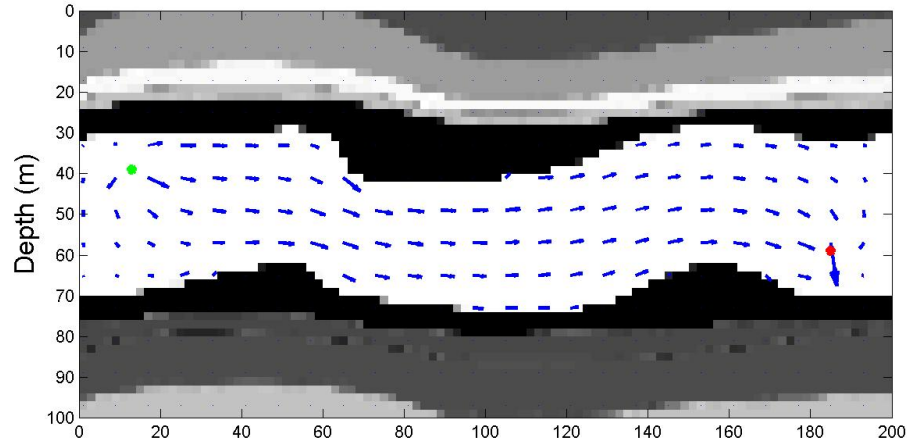

Figure 1: Flow field

2.2 Seismic borehole 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 $m(t, \vec{x})$, or inverse acoustic velocity of the media, along a ray path Γ,

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

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

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

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 $\mathbf{F}\mathbf{m}_k = \mathbf{d}_k$, for $k = 1, \ldots, n$. The tomography experiment was setup with 3 boreholes with 5 sources in the east borehole, 10 surface receivers, 10 receiver in the center borehole, and 10 receivers in the west borehole. The boreholes are 200m apart and 100m deep, as shown in figure
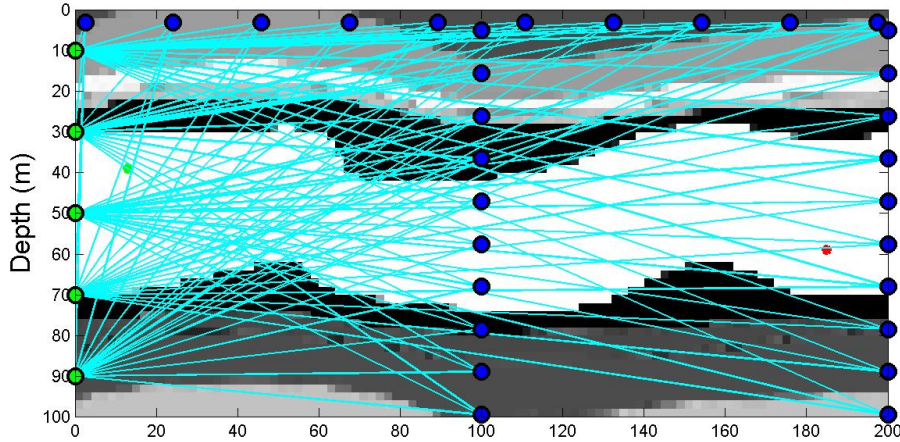

Figure 2: Initial experiment setup. Sources are pictured in green on the left and receivers in blue on the surface, center, and right. The initial setup covers the entire flow domain.

4

2.3 A-Optimal experimental design

To begin, a brief review of Bayesian A-optimal experimental design 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 $\mathbf{d} \in \mathbb{R}^M$,

$$\mathbf{d} - \mathbf{Fm} = \epsilon,$$

where $M = ns \times nr$ is the number of sources times the number of receivers, and $\mathbf{F}$ is an $M \times N$ matrix with each row representing an experiment.

Assuming a Bayesian approach is taken with the posterior estimate $\widehat{\mathbf{m}}$ conditioned on a known prior distribution $\pi_m = N(\mu, \boldsymbol{\Sigma})$, 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 resulting in,

$$\begin{aligned}
\widehat{\mathbf{m}}(\mathbf{w}) &= \operatorname{argmin} (\mathbf{Fm} - \mathbf{d})^\top \mathbf{W}(\mathbf{Fm} - \mathbf{d}) + (\mathbf{m} - \mu)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{m} - \mu) \\
&= (\mathbf{F}^\top \mathbf{W}\mathbf{F} + \boldsymbol{\Sigma}^{-1})^{-1}(\mathbf{F}^\top \mathbf{W}\mathbf{d} + \boldsymbol{\Sigma}^{-1}\mu).
\end{aligned}$$

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

A-optimal experimental designs for well-posed inverse problems are defined by minimizing the average variance of the estimate $\widehat{\mathbf{m}}$, thus minimizing the posterior covariance matrix of $\widehat{\mathbf{m}}$ [**?**],

$$\phi_{A_B}(\mathbf{w}) = \operatorname{tr}[(\mathbf{F}^\top \mathbf{W}\mathbf{F} + \boldsymbol{\Sigma}^{-1})^{-1}]. \tag{2}$$

Given ϕ_{A_B}, 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,

$$\begin{aligned}
\widehat{\mathbf{w}} &= \arg \min \phi(\mathbf{w}) + \beta \|\mathbf{w}\|_1 \\
&\text{s.t.} \quad 0 \geq \omega_i.
\end{aligned} \tag{3}$$

It is important to note that in this formulation $\widehat{\mathbf{m}}$ 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 linear dynamic problems where the estimators covariance matrix can be estimated, the simple Bayesian estimator is sufficient.

3   Designing experiments for dynamic targets

There are several considerations to be made when addressing the problem of designing experiments for a dynamic system, such as whether or not to account for error in the dynamics? Should historic data be used to update the dynamic model parameters? Does doing so change the optimal design? Does including all historic designs to get an optimal design for the future time step differ from a design computed using only the last design? Here we will consider these questions.

To begin, consider the dynamic process and imaging described by the following set of linear equations

$$\mathbf{m}_k = \mathbf{T}(\mathbf{u})\mathbf{m}_{k-1} + \eta_{\mathbf{k}}, \tag{4a}$$

$$\mathbf{d}_k = \mathbf{F}\mathbf{m}_k + \epsilon_{\mathbf{k}}, \tag{4b}$$

where the transport matrix $\mathbf{T} : \mathbb{R}^N \to \mathbb{R}^N$ is dependent on the parameter $\mathbf{u}(\mathbf{x})$, and the noise vectors η and ϵ are assumed to be Gaussian normal, $\eta \sim N(\eta_0, \mathbf{\Sigma}_\eta)$ and $\epsilon \sim N(\epsilon_0, \mathbf{W}^{-1})$.

There are several ways to look at this problem when considering how to estimate and design for $\mathbf{m}_k$. We could argue that the noise in the dynamics is insignificant, set $\eta_k = 0$, and write the system as one equation. This leads to the A-optimal designs presented in [**?**], where given a Bayesian estimation for $\mathbf{m}_0$, designs are recovered for all subsequent time points presented in Section 3.1. Alternately, we could design for a linear Kalman filter formulation. In Bayesian terms, the dynamics provide all the information for the prior distribution. Estimations for $\mathbf{m}_k$ can either be written in an iterative manner, or as an all at once estimation (Kalman smoothing). Designing for the Kalman estimator results in designs for a set of times, without the inclusion of historic data, and where the error associated with the dynamics is propagated in time, present in Section 3.2.

However, in reality estimating an optimal design should depend on available data from the previous experiment. To include this data we reformulate the estimation of the previous model $\mathbf{m}_{k-1}$ by including the constraint $\mathbf{m}_k = \mathbf{T}\mathbf{m}_{k-1}$. In Section 3.3 write the problem as an iterative method that computes an optimal design for a future time point based on either all historic data, or one the last data set collection.

3.1   Design, no noise in dynamics

Beginning with the simplest case, where $\eta_k = 0$ and $\mathbf{u}$ is known in equations (4a) such that $\mathbf{T} \leftarrow \mathbf{T}(\mathbf{u})$, we will build an objective function in terms of an initial condition $\mathbf{m}_0$ for a set of time points $k = 1, ..., n$ for all models $\overline{\mathbf{m}} = [\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_n]^\top$ given a set of data $\bar{\mathbf{d}} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_n]^\top$,

Writing the dynamic equation in block form gives

$$
\underbrace{\begin{pmatrix} -\mathbf{I} & & & & \\ \mathbf{T} & -\mathbf{I} & & & \\ & \mathbf{T} & -\mathbf{I} & & \\ & & \ddots & \ddots & \\ & & & \mathbf{T} & -\mathbf{I} \end{pmatrix}}_{=:\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_n \end{pmatrix}}_{=:\mathbf{m}} + \underbrace{\begin{pmatrix} \mathbf{T} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{=:-\mathbf{B}} \mathbf{m}_0 = \mathbf{0}. \tag{5}
$$

Thus, the set of models can then be written as the solution to the system

$$
\overline{\mathbf{m}} = \mathbf{A}(\mathbf{u})^{-1}\mathbf{B}(\mathbf{u})\mathbf{m}_0. \tag{6}
$$

Assuming that the imaging technique does not change in time, the negative log data likelihood is given by

$$
\frac{1}{2}\Big((\mathbf{I} \otimes \mathbf{F})\mathbf{A}(\mathbf{u})^{-1}\mathbf{B}(\mathbf{u})\mathbf{m}_0 - \bar{\mathbf{d}}\Big)^{\top} \overline{\mathbf{W}} \left((\mathbf{I} \otimes \mathbf{F})\mathbf{A}(\mathbf{u})^{-1}\mathbf{B}(\mathbf{u})\mathbf{m}_0 - \bar{\mathbf{d}}\right) \tag{7}
$$

where, $\overline{\mathbf{W}}$ is a block diagonal matrix with diagonal weight matrices $\mathbf{W}_1, \mathbf{W}_2, ..., \mathbf{W}_n$ on its main diagonal for time points $k = 1, 2, ..., n$.

Given a prior distribution for $\overline{\mathbf{m}} \sim N(\overline{\mathbf{m}}_0, \Sigma_{\overline{\mathbf{m}}})$, with $\Sigma_{\overline{\mathbf{m}}} = \text{blkdiag}(\Sigma_{\eta_1}, .., \Sigma_{\eta_n})$, the posterior covariance for the Bayes estimate $\widehat{\mathbf{m}}_0$ is

$$
\Sigma_{\widehat{\mathbf{m}}_0} = \left(((\mathbf{I} \otimes \mathbf{F})\mathbf{A}^{-1}\mathbf{B})^{\top} \, \overline{\mathbf{W}} \, (\mathbf{I} \otimes \mathbf{F})\mathbf{A}^{-1}\mathbf{B} + \Sigma_{\overline{\mathbf{m}}}^{-1}\right)^{-1} \tag{8}
$$

Thus the Bayesian A-optimal design functional becomes

$$
\phi_{A_B}(\overline{\mathbf{w}}) = \text{tr}\left[\left(((\mathbf{I} \otimes \mathbf{F})\mathbf{A}^{-1}\mathbf{B})^{\top} \, \overline{\mathbf{W}} \, (\mathbf{I} \otimes \mathbf{F})\mathbf{A}^{-1}\mathbf{B} + \Sigma_{\overline{\mathbf{m}}}^{-1}\right)^{-1}\right], \tag{9}
$$

where the designs recovered are for all times points $k = 1, ..., n$. This method was demonstrated in [?] for and advection diffusion dynamic flow problem. Estimating designs in this way requires that we fully accept the accuracy of the dynamics, and that we know the prior distribution covariance at each time point. We could choose a linear regularization operator $\mathbf{L}$ such that $\mathbf{L}^{\top}\mathbf{L} = \Sigma_{\mathbf{k}}^{-1}$, as in a Tikhonov formulation to deal with the priors, but it is apparent that in reality the parameters governing the dynamics are not specifically known, or are known to low accuracy, thus reducing confidence in designs for later time points. In the case of the dynamics, the dependence on their own subsurface parameters should be considered. Ideally we would like to be able to iteratively update $\mathbf{u}$ once imaging data is collected, such that the designs are recomputed given this new and more accurate information.

3.2 Designing with noise in the dynamics

Again we assume that $\mathbf{u}$ is known in equations (4a) so that $\mathbf{T} \leftarrow \mathbf{T}(\mathbf{u})$, and that no initial condition is needed since there is error in the estimates of $\mathbf{m}$. Denoting $\overline{\mathbf{m}} = [\mathbf{m}_1, \mathbf{m}_2, ..., \mathbf{m}_n]^{\top}$

and a set of data $\bar{\mathbf{d}} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_n]^\top$ and assuming the initial condition is given by the noise $\eta_0 = \mathbf{m}_0$ and $\mathbf{T}_0 = \mathbf{I}$ we can write

$$\underbrace{\begin{pmatrix} \mathbf{I} & 0 & & & \\ -\mathbf{T} & \mathbf{I} & 0 & & \\ & -\mathbf{T} & \mathbf{I} & \ddots & \\ & & \ddots & \ddots & 0 \\ & & & -\mathbf{T} & \mathbf{I} \end{pmatrix}}_{=:\mathbf{A}} \underbrace{\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \\ \vdots \\ \mathbf{m}_n \end{pmatrix}}_{=:\overline{\mathbf{m}}} = \mathbf{0}. \tag{10}$$

For a time invariant imaging technique, the Bayesian negative log likelihood is given by

$$\mathcal{J}(\overline{\mathbf{m}}) = \frac{1}{2}\left((\mathbf{I} \otimes \mathbf{F})\overline{\mathbf{m}} - \bar{\mathbf{d}}\right)^\top \overline{\mathbf{W}} \left((\mathbf{I} \otimes \mathbf{F})\overline{\mathbf{m}} - \bar{\mathbf{d}}\right) + \frac{1}{2}\left(\mathbf{A}\overline{\mathbf{m}}\right)^\top \mathbf{\Sigma}_{\overline{\mathbf{m}}}^{-1}\left(\mathbf{A}\overline{\mathbf{m}}\right) \tag{11}$$

where, $\overline{\mathbf{W}}$ is a block diagonal matrix with diagonal weight matrices $\mathbf{W}_0, \mathbf{W}_1, ..., \mathbf{W}_n$ on its main diagonal, and $\mathbf{\Sigma}_{\overline{\mathbf{m}}} = \mathrm{blkdiag}(\mathbf{\Sigma}_{\eta_1}, .., \mathbf{\Sigma}_{\eta_n})$.

The posterior covariance matrix obtained after minimizing $\mathcal{J}(\overline{\mathbf{m}})$ for the Bayes estimate $\overline{\mathbf{m}}$ is given by

$$\mathbf{H} = \left((\mathbf{I} \otimes \mathbf{F})^\top \overline{\mathbf{W}}(\mathbf{I} \otimes \mathbf{F}) + \mathbf{A}^\top \mathbf{\Sigma}_{\overline{\mathbf{m}}}^{-1}\mathbf{A}\right)^{-1} \tag{12}$$

Thus, computing an optimal design entails minimizing the trace of the inverse of the matrix $\mathbf{H}$. Again,

$$\phi_{A_B}(\overline{\mathbf{w}}) = \mathrm{tr}\left[\left((\mathbf{I} \otimes \mathbf{F})^\top \overline{\mathbf{W}}(\mathbf{I} \otimes \mathbf{F}) + \mathbf{A}^\top \mathbf{\Sigma}_{\overline{\mathbf{m}}}^{-1}\mathbf{A}\right)^{-1}\right], \tag{13}$$

This formulation provides another method of designing all of the experiments given the covariance matrices for each time point. The main difference in this formulation than the previous section is that the error is now propagated in time. However, this method still does not incorporate historic designs or collected data, and it requires storing and manipulating large dense covariance matrices.

3.3 Adaptive historic design

In this section we present two reformulations of the model estimation problems such that historic data/designs are included. In this way the future design will depend information from the previous experiments and the dynamics of the problem. We will consider the formulation with a linear and a non-linear regularization term.

3.3.1 Linear regularization

In order to estimate the design for an experiment at time k given a design $\mathbf{w}_{k-1}$, and thus collected data set $\mathbf{d}_{k-1}^c$, from the previous time point $k-1$ the model objective function is constructed in order to re-estimate $\mathbf{m}_{k-1}$ while including historic data,

$$\mathcal{G}(\mathbf{m}, \mathbf{w}) = \sum_{k=1}^{n} \left(\frac{1}{2} \|\mathbf{F}_{k-1}\mathbf{m}_{k-1} - \mathbf{d}_{k-1}^{c}\|_{2}^{2} + \frac{1}{2} \|\mathbf{FTm}_{k-1} - \mathbf{d}_{k}\|_{\mathbf{W}_{k}}^{2} + \frac{1}{2} \|\mathbf{Lm}_{k-1}\|_{2}^{2} \right) \quad (14)$$

where $\mathbf{F}_{k-1} = \mathbf{W}_{k-1}\mathbf{F}$, and $\mathbf{m}_k = \mathbf{Tm}_{k-1}$. The operator $\mathbf{L}$ is a smoothing regularization operator. One could argue in this instance that the first term including the historic data can be interpreted as a prior such that the regularization term is not necessarily required. In this way, while it appears that this is a frequentist approach, it can still be interpreted as a Bayesian one if both the collected data misfit term and the regularization combine as a prior.

Minimizing equation (14) with respect to $\mathbf{m}_{k-1}$ yields the posterior covariance matrix

$$\mathbf{\Sigma}_{\mathbf{k-1}}(\mathbf{w_k}) = \left(\mathbf{F}_{\mathbf{k-1}}^{\top}\mathbf{F}_{\mathbf{k-1}} + \mathbf{T}^{\top}\mathbf{F}^{\top}\mathbf{W_k}\mathbf{FT} + \mathbf{L}^{\top}\mathbf{L} \right)^{-\mathbf{1}}. \quad (15)$$

By formulating the problem in this manner, one can adaptively update optimal designs after each experiment has been carried out.

The design functional to then be minimized is given by

$$\phi_k(\mathbf{w}_k) = \text{tr}\left[\left(\mathbf{F}_{k-1}^{\top}\mathbf{F}_{k-1} + \mathbf{T}^{\top}\mathbf{F}^{\top}\mathbf{W}_k\mathbf{FT} + \mathbf{L}^{\top}\mathbf{L} \right)^{-1} \right]. \quad (16)$$

A second benefit of building the design objective function to depend on an historic design and data set is that the design problem does not grow in size as time progresses. If we were to incorporate all historic data and designs, then we would end up with a rather large system. However, if all historic information is included, one could argue that this would provide a more accurate estimate and design.

The basic algorithm illustrating the adaptive design process for a linear observation method is outlined in Algorithm 1.

Algorithm 1 Iterative Optimal ϕ_{A_B} Design

1: Initialize: $\mathbf{F}_0$ ▷ Estimate a naive design
2: **for** $k = 1$ to n **do**   ▷ n time points
3:     $\widehat{\mathbf{w}}_k(\mathbf{F}_{k-1}) \rightarrow \text{diag}(\widehat{\mathbf{w}}_k)\mathbf{F} = \mathbf{F}_k$ ▷ compute the optimal design by solving 3 with 16
4:     $\mathbf{d}_k = \mathbf{F}_k\mathbf{m}_k + \epsilon_k$ ▷ collect data
5:     $\mathbf{F}_{k-1} \leftarrow \mathbf{F}_k$ ▷ update F
6: **end for**

3.3.2 Non-linear regularization

At this point it is important to consider what information is available when estimating the optimal design. In equation (14), we added collected data misfit term, a linear regularization

term promotes smoothness in the model $\mathbf{m}_k$, which of course appears in equation (15) as $\mathbf{L}^\top \mathbf{L}$. However, although the design $\mathbf{F}_k$ in the design objective function, the collected data does not. To include the collected data, we can instead impose a non-linear regularization function,

$$\mathcal{G}(\mathbf{m}, \mathbf{w}) = \sum_{k=1}^{n} \left(+ \frac{1}{2} \|\mathbf{FTm}_{k-1} - \mathbf{d}_k\|_{\mathbf{W}_k}^2 + R(\mathbf{m}_{k-1})\|_2^2 \right). \tag{17}$$

Since there is no closed form for the posterior covariance matrix for a non-linear inverse problem, we must apply an approximation as in [**?**]. Assigning the non-linear regularization functional results in an approximation to the inverse Hessian (or posterior covariance matrix) that depends on $\mathbf{m}$. Choosing a Gauss-Newton approximation of the Hessian, the optimal design optimization problem (3) becomes a bi-level optimization with an inequality constraint,

$$\min_{\mathbf{w}_k} \quad \mathrm{tr}\left[\left(\mathbf{T}^\top \mathbf{F}^\top \mathbf{W}_k \mathbf{FT} + \mathbf{S}^\top(\widehat{\mathbf{m}}_{k-1}) \mathbf{S}(\widehat{\mathbf{m}}_{k-1}) \right)^{-1} \right] \tag{18a}$$

$$\text{s.t.} \quad \widehat{\mathbf{m}}_{k-1} = \operatorname{argmin} \frac{1}{2} \|\mathbf{F}_{k-1}\mathbf{m}_{k-1} - \mathbf{d}_{k-1}^c\|_2^2 + R(\mathbf{m}_{k-1}) \tag{18b}$$

$$0 \geq \omega_i, \tag{18c}$$

where $\mathbf{S}(\mathbf{m})$ is the gradient of the regularization functional $R(\mathbf{m})$. Typically solving the lower level optimization would require a stochastic approach of averaging over a set of possible models. However, in this case the historic data is available so that the best estimator $\widehat{\mathbf{m}}_{k-1}$ can be recovered so that it is not necessary to sample average over a set of possibilities.

4   Numerical optimization

In this section we present the computation of the gradients of the linear and non-linear design functionals.

4.1   Linear adaptive design

Finding an optimal design for the time step k involves solving the optimization problem 3 by minimizing the objective functional (16),

$$\widehat{\mathbf{w}_k} \quad = \quad \operatorname{argmin} \mathcal{J}_k := \left\{ \mathrm{tr}\left[\left(\mathbf{T}^\top \mathbf{F}^\top \mathbf{W}_k \mathbf{FT} + \mathbf{F}_{k-1}^\top \mathbf{F}_{k-1} + \mathbf{L}^\top \mathbf{L} \right)^{-1} \right] + \beta \mathbf{e}^\top \mathbf{w}_k \right\}$$

$$\text{s.t.} \quad 0 \geq \omega_i.$$

where $\mathbf{e}$ is a vector of ones (the constraint allows us to write the 1-norm as a sum).

This is a difficult task for large problems, particularly due to the trace of a large dense matrix and the computation of the derivative of the trace. 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 (??) can be written as

$$\mathcal{J}_k(\mathbf{w}_k) = \mathbf{v}^\top (\mathbf{T}^\top \mathbf{F}^\top \mathbf{W}_k \mathbf{F}\mathbf{T} + \mathbf{G})^{-1}\mathbf{v} + \beta \mathbf{e}^\top \mathbf{w}_k, \tag{19}$$

where $\mathbf{G} = \mathbf{F}_{k-1}^\top \mathbf{F}_{k-1} + \mathbf{L}^\top \mathbf{L}$ and $\mathbf{v}$ 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. From now on we drop the subscript k for clarity,

$$\nabla_{\mathbf{w}}\mathcal{J}(\mathbf{w}) = \mathbf{v}^\top \nabla_{\mathbf{w}}\Big(((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})^{-1}\mathbf{v}\Big) + \beta \mathbf{e}.$$

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

$$\mathbf{z} = ((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})^{-1}\mathbf{v} \quad \Leftrightarrow \quad ((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})\mathbf{z} = \mathbf{v}, \tag{20}$$

and differentiate implicitly to obtain

$$(\mathbf{F}\mathbf{T})^\top \mathrm{diag}(\mathbf{F}\mathbf{T}\mathbf{z}) + ((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})\nabla_{\mathbf{w}}\mathbf{z} = 0.$$

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

$$\nabla_{\mathbf{w}}\mathbf{z} = -((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})^{-1}(\mathbf{F}\mathbf{T})^\top \mathrm{diag}(\mathbf{F}\mathbf{T}\mathbf{z}).$$

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

$$-\mathbf{v}^\top ((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})^{-1}(\mathbf{F}\mathbf{T})^\top \mathrm{diag}(\mathbf{F}\mathbf{T}\mathbf{z}).$$

Taking the transpose

$$-\mathrm{diag}(\mathbf{F}\mathbf{T}\mathbf{z})\mathbf{F}\mathbf{T}((\mathbf{F}\mathbf{T})^\top \mathbf{W}\mathbf{F}\mathbf{T} + \mathbf{G})^{-1}\mathbf{v} \quad \Leftrightarrow \quad -\mathrm{diag}(\mathbf{F}\mathbf{T}\mathbf{z})(\mathbf{F}\mathbf{T}\mathbf{z}) = -\mathbf{F}\mathbf{T}\mathbf{z} \odot \mathbf{F}\mathbf{T}\mathbf{z},$$

so that the derivative of the objective function is then

$$\nabla_{\mathbf{w}}\mathcal{J}_k(\mathbf{w}) = -\mathbf{F}\mathbf{T}\mathbf{z} \odot \mathbf{F}\mathbf{T}\mathbf{z} + \beta \mathbf{e}. \tag{21}$$

Since the design problem (3) 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. However, for large scale problems, the evaluation of $\mathbf{z}$ is not cheap.

To reduce computational costsit is possible to first preform a Cholesky factorization on $\mathbf{G} = \mathbf{R}^\top \mathbf{R}$, reduce equation (20) to standard form, and then compute the Lanczos bidiagonalization of $\mathbf{F}\mathbf{T}\mathbf{R}^{-1}$. This process requires calculating the Cholesky factorization and calculating $\mathbf{R}^{-1}$. However, this is evaluated only once outside of the design optimization, so depending on the problem size, it is computationally cheap.

Alternately, instead of the Lanczos approach Krylov subspace methods can be applied, such as Conjugate Gradients, allowing us to take advantage of matrix vector products and avoid ever forming $\mathbf{G}$.

4.2   Non-Linear adaptive design

5   Numerical example: Adaptive design

To demonstrate iterative optimal experimental design, a seismic borehole tomography survey is chosen to image tracer advection in 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 $\mathbf{T}$.

5.1   Experimental setup

The computational domain, $\Omega = [0, 100] \times [0, 200]$ meters, is divided into $m = [50, 100]$ cells of width $h = [1, 1]$ meters and consists of a layered model of five rock types.
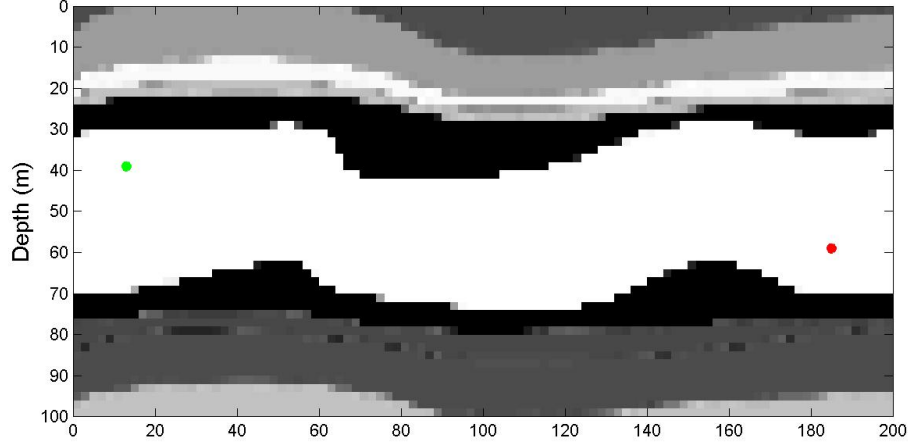

Figure 3: Initial layered Earth model.

5.2   Regularization

linear regularization gradient operator nonlinear regularization Next, the regularization functional for tracer is discretized. Noting that $\mathbf{m}$ is discretized on a cell-centered grid, a standard discrete approximation for the smoothed total variation regularization is applied [**?**]

$$R^m(\mathbf{m}) = h^2 \mathbf{e}^\top \sqrt{\mathbf{A}_f^c \left((\mathsf{GRAD}\ \mathbf{m}) \odot (\mathsf{GRAD}\ \mathbf{m}) \right) + \epsilon}, \tag{22}$$

13

where GRAD is a standard 2-point discretization of the gradient of a cell-centered variable, which maps from cell-centers to faces, as described in [**?**, **?**]. $\mathbf{A}_f^c$ is an averaging matrix from cell-faces to cell-centers. Note that we somewhat abuse the notation, and that the square root of a vector is the point-wise square root.

6 Numerical example 2: Helmhotz maybe

This depends on how far ahead in time we design...The all at once approach assumes the velocity field is accurate(ish)...could still solve the all at once problem for updated velocity field too.

7 Concluding Remarks