Adaptive A-optimal experimental design for dynamical systems

J. Fohring and E. Haber

June 22, 2015

Abstract

The optimal collection of data required to infer dynamic model parameters governing a time dependent partial differential equation (PDE) through the inversion process is important to may applications, particularly as it can reduce subsequent costs by reducing total measurements. Popular methods which reduce the number of data required seek a set of measurements which minimize the mean square error (mse) of the estimated model parameters. For linear ill-posed model estimation problems, these methods depend only on the physics of the problem and information provided through regularization. Thus historic data are not included, nor does the design necessarily track the dynamic target.

To include prior data in the design process, we reformulate the model estimation inverse problem to include both the dynamic PDE and historic data. We present two model estimation formulations based on the classification of the error in the dynamic process, and then define adaptive optimal design, which minimizes the adapted mean square error (amse) of the estimated models. The amse is defined such that historic model estimates are included in the model optimization problem through a monitor function. This monitor function depends on prior model estimators and scales the mse such that areas of interest contribute to the design. In this way, minimizing the amse in the design optimization problem results in designs where historic data and the dynamics contribute to the design of the future experiment.

We demonstrate the adaptive optimal design method for a two dimensional target whose motion is governed by a tracer advection model, and where a seismic borehole tomography survey was used to image the target in the subsurface. Using this model we show that the adaptive design method produces designs which track the motion of the target with a significantly reduced set of data.

1 Introduction

While there exists a vast area of research in optimal experimental design for inverse problems of static problems [14, 12, 13, 6, 8, 10, 5, 18, 1], there is little in the way of optimal design algorithms for ill-posed inverse problems where the model is governed by a dynamic system. Although static design methods can be applied to time dependent partial differential equations (PDE's) [2, 14], this approach results in an a-priori estimate of the designs for "all times" which depend only on prior information provided in the form of a regularization (or prior) to the ill-posed problem. Current design techniques do not utilize information that is collected in early times in order to design experiments at later times.

In many realistic scenarios, the dynamics are sufficiently slow to allow for improved designs after a first experiment has been conducted. The available experimental data contains valuable information about the target to be recovered. This information should thus be included in the

design algorithm for future experiments. Similarly, if we collect data at times $t_1, \ldots, t_n$, it should be used for the design of the experiment at time t_{n+1}. In this paper we present an adaptive method to determine an optimal experimental design for an experiment where the underlying model is a dynamic system where the governing dynamic PDE is known. Our method can be seen as a post-priori design, which gradually improves the design given the recovery at previous time steps.

The idea of a-priori estimates verses post-priori estimates is not new. For example, in adaptive mesh refinement, when numerically solving partial differential equations one often uses a-priori estimates in order to solve the problem on one mesh and then refines the mesh in order to obtain a more accurate solution [7]. The posterior estimate thus contributes to the decision as to where to refine the mesh. In our case we apply a similar logic to influence the design algorithm in such a way that we use solutions from earlier times to obtain a better optimal design at later times.

More multigrid refs?

There are several considerations to be made when developing an optimal experimental design method. One such item is what exactly is meant by an "optimal experiment"? In our context we refer to the problem of determining the optimal placement of sensors to infer model parameters through the inversion of data. Although there are several design criteria for such inverse problems, we focus on A-optimal design in our work. Generalization to other design criteria is straight forward. The A-optimal design method minimizes the mean squared error (MSE) of the model estimator, which, for linear inverse problems, is equivalent to the minimization of the trace of the posterior covariance matrix (from the Bayesian perspective). Our work here builds on A-optimal design algorithms presented in [14] for large-scale ill-posed inverse problems, and on work presented in [11] where historical data was used to recover the properties of a dynamic system.

There are two main paradigms when considering dynamic systems in the context of an inverse problem. In the first, the dynamics are "exact", that is, we assume that the dynamic system does not contain errors and that we can make some indirect measurements of the dynamic process. In this case, the goal is to evaluate the parameters of the dynamic system, and the initial conditions (see [11]). A second approach is to consider errors in the dynamic system that are modeled as random errors. This leads to Kalman filtering or smoothing (see [16, 3] and references therein) depending on the application. We address the experimental design for both cases in this work.

The paper is outlined as follows. In Section 2 we review the A-experimental design criteria and show its shortcomings when dealing with adaptive design. We then present a new design criteria that enables us to adapt the experiment based on the collected experimental data. In Section 3 we adapt the design criteria developed in Section 2 to the case of dynamic systems.

needs work

2 Adaptive experimental design

To present the adaptive experimental design method we begin by reviewing the derivation of the A-optimal design criteria from first principals. We start by considering only two measurement vectors of the form

$$\mathbf{F}_1\mathbf{m} + \boldsymbol{\epsilon}_1 = \mathbf{d}_1 \quad \text{and} \quad \mathbf{F}_2\mathbf{m} + \boldsymbol{\epsilon}_2 = \mathbf{d}_2 \tag{1}$$

where $\mathbf{F}_1$ is the forward operator at time t_1 and $\mathbf{F}_2$ is the forward operator at time t_2. At this point we do not estimate any dynamics, our goal is to design the experiments at times t_1 and t_2

such that we obtain the "best" recovery by some criteria. As we see next, this scenario fits the problem where the dynamical system is "exact" (that is, contain no noise) and $\mathbf{m}$ represents the dynamical system parameters that are time invariant such as initial conditions.

Two different designs can be considered. First, it is possible to perform *a-priori* design, that is, to design the experiments $\mathbf{F}_1$ and $\mathbf{F}_2$ prior to the data collection. This is the case when the time between t_1 and t_2 is much shorter than the time for the processing of the data. A second approach is to use *post-priori* design. The idea here is to use the results obtained from the first experiment in order to design a "better" second experiment. Our goal is to first obtain an a-priori design for the first experiment and then, after the data is collected and processed, use the estimator obtained at t_1 in order to design the data collection at time t_2.

Before we discuss the design of the data acquisition for time t_2 we review the a-priori design of the experiment at time t_1. Consider the estimation of $\mathbf{m}$ given $\mathbf{d}_1$, accomplished by solving the ill-posed optimization problem

$$\widehat{\mathbf{m}}_1 = \text{argmin } \frac{1}{2}\|\mathbf{F}_1\mathbf{m} - \mathbf{d}_1\|^2_{\mathbf{W}_1} + \frac{\alpha}{2}\|\mathbf{L}(\mathbf{m} - \widehat{\mathbf{m}}_0)\|^2_2. \tag{2}$$

where $\mathbf{W}_1 = \mathsf{diag}(\mathbf{w}_1)$ is a matrix of inverse standard deviations, that is, the inverse covariance of the noise, $\mathbf{L}$ is some regularization operator, $\widehat{\mathbf{m}}_0$ is the current estimate of the model prior to having data, and α is a regularization parameter. Minimizing (2) yields the estimator

$$\widehat{\mathbf{m}}_1 = (\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{d}_1 + \alpha\mathbf{L}^\top\mathbf{L}\widehat{\mathbf{m}}_0). \tag{3}$$

Defining the matrix $\mathbf{C} = (\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})$, and recalling that $\mathbf{F}_1\mathbf{m} + \boldsymbol{\epsilon}_1 = \mathbf{d}_1$ we can write the error in the recovery as

$$\begin{aligned} \widehat{\mathbf{m}}_1 - \mathbf{m} &= \mathbf{C}^{-1}\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1\mathbf{m} + \mathbf{C}^{-1}\mathbf{F}_1^\top\mathbf{W}_1\boldsymbol{\epsilon}_1 + \alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\widehat{\mathbf{m}}_0 \\ &+ (\alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\mathbf{m} - \alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\mathbf{m}) - \mathbf{m}. \end{aligned} \tag{4}$$

Collecting terms and using the definition of $\mathbf{C}$ we obtain

$$\widehat{\mathbf{m}}_1 - \mathbf{m} = \mathbf{C}^{-1}\mathbf{F}_1^\top\mathbf{W}_1\boldsymbol{\epsilon}_1 + \alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}(\widehat{\mathbf{m}}_0 - \mathbf{m}). \tag{5}$$

Squaring and taking the expectation over $\boldsymbol{\epsilon}_1$ and recalling that $\boldsymbol{\epsilon}_1 \sim N(\mathbf{0}, \mathbf{W}_1^{-1})$, we obtain that the mean square error is

$$\mathsf{mse}(\mathbf{w}, \mathbf{m}) = \mathbf{E}\|\widehat{\mathbf{m}}_1 - \mathbf{m}\|^2 = \mathsf{trace}[\mathbf{F}_1\mathbf{C}^{-2}\mathbf{F}_1^\top\mathbf{W}_1] + \alpha^2\mathbf{E}\|\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}(\widehat{\mathbf{m}}_0 - \mathbf{m})\|^2. \tag{6}$$

If we further assume that $\mathbf{m} - \widehat{\mathbf{m}}_0$ is Gaussian with a zero mean and a covariance $(\alpha\mathbf{L}^\top\mathbf{L})^{-1}$ then the mean squared error is

$$\phi(\mathbf{w}_1) = \mathsf{mse}(\mathbf{w}_1) = \mathsf{trace}[\mathbf{C}^{-2}\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1] + \alpha\mathsf{trace}[\mathbf{C}^{-2}\mathbf{L}^\top\mathbf{L}]. \tag{7}$$

Finally, using the linearity of the trace and the definition of $\mathbf{C}$ we obtain that

$$\phi_1(\mathbf{w}_1) = \mathsf{trace}\left[(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\right]. \tag{8}$$

In the design problem we seek to obtain a better estimate for $\mathbf{m}$ where we assume that we can control the (inverse) standard deviation, $\mathbf{w}_1$, of the collected data. Data that are assigned with infinite standard deviation ($w_i = 0$) are not collected.

This leads to the minimization of the penalized mse, where we minimize $\phi_1(\mathbf{w}_1)$ with an additional cost on $\mathbf{w}_1$ that are larger that 0. For example, in our previous work we proposed to minimize the function

$$\phi_1^{\beta}(\mathbf{w}_1) = \mathsf{trace}\left[(\mathbf{F}_1^{\top}\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^{\top}\mathbf{L})^{-1}\right] + \beta\sum \mathbf{w}_1 \qquad 0 \leq \mathbf{w} \tag{9}$$

which balances the minimization of the mse and the cost of the experiment [12].

Consider now using the design problem of estimating $\mathbf{m}$ given the estimated model $\widehat{\mathbf{m}}_1$. One could rewrite the recovery problem in a similar way, that is

$$\widehat{\mathbf{m}}_2 = \arg\min \ \frac{1}{2}\|\mathbf{F}_1\mathbf{m} - \mathbf{d}_1\|^2_{\mathbf{W}_1} + \frac{1}{2}\|\mathbf{F}_2\mathbf{m} - \mathbf{d}_2\|^2_{\mathbf{W}_2} + \frac{\alpha}{2}\|\mathbf{L}(\mathbf{m} - \widehat{\mathbf{m}}_0)\|_2^2. \tag{10}$$

which leads to the estimate

$$\widehat{\mathbf{m}}_2(\mathbf{w}_2) = (\mathbf{F}_1^{\top}\mathbf{W}_1\mathbf{F}_1 + \mathbf{F}_2^{\top}\mathbf{W}_2\mathbf{F}_2 + \alpha\mathbf{L}^{\top}\mathbf{L})^{-1}(\mathbf{F}_1^{\top}\mathbf{W}_1\mathbf{d}_1 + \mathbf{F}_2^{\top}\mathbf{W}_2\mathbf{d}_2 + \alpha\mathbf{L}^{\top}\mathbf{L}\widehat{\mathbf{m}}_0) \tag{11}$$

Note that we assume that $\mathbf{w}_1$ is fixed and therefore the new estimate is a function of $\mathbf{w}_2$ alone. If we repeat the steps above then we obtain that the A-optimal design for time t_2 yields the minimization of the function

$$\phi_2^{\beta}(\mathbf{w}_2) = \mathsf{trace}\left[(\mathbf{F}_1^{\top}\mathbf{W}_1\mathbf{F}_1 + \mathbf{F}_2^{\top}\mathbf{W}_2\mathbf{F}_2 + \alpha\mathbf{L}^{\top}\mathbf{L})^{-1}\right] + \beta\sum \mathbf{w}_2. \tag{12}$$

At this point it is worth pausing for a moment and noting an important feature of the experimental design criteria. The design criteria does not depend on the data. This observation is true for any of the design criteria (that is, C,D and E designs). Furthermore, it is easy to verify that **any** linear estimator of the data yields a covariance matrix that is independent of $\mathbf{d}_1$. This implies that using current design criteria we are unable to use the estimated model obtained at time t_1 to obtain a better estimate at time t_2.

In order to use the information obtained at time t_1 for the design of the experiment at time t_2 we now present the concept of *adaptive design*.

Assume that the model, $\widehat{\mathbf{m}}_1$ has some "interesting" features and some "boring" features. To be more specific, assume that the difference

$$\boldsymbol{\delta}_1 = |\widehat{\mathbf{m}}_1 - \widehat{\mathbf{m}}_0|$$

is small in some norm over some region and large in others. The goal then is to better estimate the new features that appear in the model.

To this end, we introduce the **monitor function**, $\boldsymbol{\tau}(\boldsymbol{\delta})$ that measures the change in the model. For example, to start, consider the function

$$\boldsymbol{\tau}_s = \boldsymbol{\delta} \odot \boldsymbol{\delta}, \tag{13}$$

that simply measures the change in the estimator compared to the previously known estimator.

Given the monitor function, rather than minimizing the mean square error, we propose to minimize the adaptive mean square error, amse defined by

$$\mathsf{amse}(\mathbf{w}, \mathbf{m}) = \|\widehat{\mathbf{m}}_2 - \mathbf{m}\|^2_{\boldsymbol{\tau}_1} = (\widehat{\mathbf{m}}_2 - \mathbf{m})^{\top}\mathsf{diag}(\boldsymbol{\tau}_1)(\widehat{\mathbf{m}}_2 - \mathbf{m}). \tag{14}$$

The idea behind the amse is to obtain a tighter bound on the model where the estimator exhibits large changes compared to the known a-priori estimator.

Starting from equation (5), modifying it to deal with the amse, and repeating the calculation above we obtain that the expectation over the amse is

$$\phi_2(\mathbf{w}_2) = \mathsf{trace}\left[\mathsf{diag}(\boldsymbol{\tau})(\mathbf{F}_1^\top \mathbf{W}_1 \mathbf{F}_1 + \mathbf{F}_2^\top \mathbf{W}_2 \mathbf{F}_2 + \alpha \mathbf{L}^\top \mathbf{L})^{-1}\right]. \tag{15}$$

Similar to the non-adaptive case, the adaptive design is obtained by minimizing the (penalized) function $\phi_2(\mathbf{w}_2)$.

The above concept can easily be extended to any number of time steps. Consider $k-1$ experiments that are conducted at times $t_1, ..t_{k-1}$ using the forward operators $\mathbf{F}_1, \ldots, \mathbf{F}_{k-1}$ and assume that we want to design an experiment for the problem at time t_k. The design criteria in this case reads

$$\phi_k(\mathbf{w}_k) = \mathsf{trace}\left[\mathsf{diag}(\boldsymbol{\tau}_{k-1})\Big(\sum_{j=1}^{k} \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top \mathbf{L}\Big)^{-1}\right] \tag{16}$$

where $\boldsymbol{\tau}_{k-1}$ is the monitor function that measure the difference between the model estimated at time t_{k-1} and the model estimated at time t_{k-2}.

3 Design for dynamical systems

In the previous section we presented a formulation for adaptive A-optimal experimental design that incorporates historic data given a set of measurements assuming that the model, $\mathbf{m}$ is static. In this section we consider the case that the model is governed by a dynamical system.

Consider the discrete dynamic process and imaging described by the following set of linear equations

$$\mathbf{m}_k = \mathbf{T}\mathbf{m}_{k-1} + \boldsymbol{\eta}_k, \tag{17a}$$
$$\mathbf{d}_k = \mathbf{F}\mathbf{m}_k + \boldsymbol{\epsilon}_k, \tag{17b}$$

where the transport matrix $\mathbf{T} : \mathbb{R}^N \to \mathbb{R}^N$ is a discretization of a PDE governing the dynamics in the system. The noise vectors $\boldsymbol{\eta}_k$ and $\boldsymbol{\epsilon}_k$ are assumed to be Gaussian normal, $\boldsymbol{\eta}_k \sim N(0, \mathbf{Q}_k)$ and $\boldsymbol{\epsilon}_k \sim N(0, \mathbf{W}_k^{-1})$.

In combining equations (17a) and (17b) there are two instances of the dynamics that result in different formulations. Assume first that the dynamics are exact, that is $\boldsymbol{\eta}_k = 0,\ \forall k$. Such a scenario is often assumed in the case of history matching [17]. In this case the dynamics are determined by the initial model, $\mathbf{m}_0$ that is to be evaluated using the measurements at later times. The second instance refers to the case that $\boldsymbol{\eta}_k \neq 0$ and therefore, the models $\mathbf{m}_1, \ldots, \mathbf{m}_k$ need to be estimated from the data. We now discuss both cases in detail.

3.1 Noiseless model dynamics

Setting $\boldsymbol{\eta}_k = \mathbf{0}$ allows us to write the model at time t_k as a linear mapping of the initial model $\mathbf{m}_0$, such that

$$\mathbf{m}_k = \mathbf{T}\mathbf{m}_{k-1}, \quad \text{or} \quad \mathbf{m}_k = \mathbf{T}^k \mathbf{m}_0.$$

To this end the data for time step t_k can be written in terms of the initial model

$$\mathbf{d}_k = \mathbf{F}_k \mathbf{m}_0 + \boldsymbol{\epsilon}_k \tag{18}$$

where $\mathbf{F}_k = \mathbf{F}\mathbf{T}^k$. Note that this case is exactly the case described in the previous section and therefore we can directly use the amse criteria.

The estimation of the initial model $\mathbf{m}_0$ given k data sets is

$$\widehat{\mathbf{m}}_{0,k} = \text{argmin} \ \frac{1}{2}\sum_{j=1}^{k} \|\mathbf{F}\mathbf{T}^j\mathbf{m}_0 - \mathbf{d}_j\|^2_{\mathbf{W}_j} + \frac{\alpha}{2}\|\mathbf{L}(\mathbf{m}_0 - \widehat{\mathbf{m}}_0)\|_2^2. \tag{19}$$

Substituting the definition of $\mathbf{F}_k$ into Equation (16) leads to the optimization problem

$$\begin{aligned} \min_{\mathbf{w}_k} \ & \left\{ \text{tr}\Big[\mathsf{diag}(\boldsymbol{\tau}_k)\Big(\sum_{j=1}^{k} \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top\mathbf{L}\Big)^{-1}\Big] + \beta \sum \mathbf{w}_k \right\} \\ \text{s.t.} \ \ & 0 \geq \mathbf{w}_i, \end{aligned} \tag{20}$$

3.2 Noisy model dynamics

I'm assuming you want to redo this whole section?

To include the noise inherent in the dynamic model we begin by considering three measurements of the dynamic target $\mathbf{m}_k$,

$$\begin{aligned} \mathbf{F}\mathbf{m}_0 - \mathbf{d}_0 &= \boldsymbol{\epsilon}_0, \\ \mathbf{F}\mathbf{m}_1 - \mathbf{d}_1 &= \boldsymbol{\epsilon}_1, \qquad \mathbf{m}_1 - \mathbf{T}\mathbf{m}_0 = \boldsymbol{\eta}_1 \\ \mathbf{F}\mathbf{m}_2 - \mathbf{d}_2 &= \boldsymbol{\epsilon}_2, \qquad \mathbf{m}_2 - \mathbf{T}\mathbf{m}_1 = \boldsymbol{\eta}_2. \end{aligned}$$

It is assumed in this case that the first experiment $\mathbf{d}_0$, will either be conducted such that all data are collected, or will be conducted in a naive way. In either case it is assumed that $\mathbf{W}_0 = \mathsf{diag}(\mathbf{w}_0)$ is known. Also note that if $\boldsymbol{\eta} = \mathbf{0}$ we end up with the previous linear propagation of $\mathbf{m}$.

To estimate $(\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2)^\top$ given $(\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2)^\top$ we minimize the following system with respect to $(\mathbf{m}_0, \mathbf{m}_1, \mathbf{m}_2)^\top$

$$\begin{aligned} &\frac{1}{2}\left\| \begin{pmatrix} \mathbf{W}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_2 \end{pmatrix}^{\frac{1}{2}} \left(\begin{pmatrix} \mathbf{F} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{m}_0 \\ \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} - \begin{pmatrix} \mathbf{d}_0 \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{pmatrix} \right) \right\|_2^2 \\ &+ \frac{1}{2}\left\| \begin{pmatrix} \mathbf{Q}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2^{-1} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} -\mathbf{T} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{T} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{m}_0 \\ \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix} \right\|_2^2 \end{aligned} \tag{21}$$

Taking the gradient of (21), setting it equal to zero, and solving for $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)^\top$ results in the following estimator

$$\begin{pmatrix} \widehat{\mathbf{m}}_0 \\ \widehat{\mathbf{m}}_1 \\ \widehat{\mathbf{m}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{F}^\top\mathbf{W}_0\mathbf{F} + \mathbf{T}^\top\mathbf{Q}_0^{-1}\mathbf{T} & -\mathbf{T}^\top\mathbf{Q}_0^{-1} & \mathbf{0} \\ -\mathbf{Q}_0^{-1}\mathbf{T} & \mathbf{F}^\top\mathbf{W}_1\mathbf{F} + \mathbf{Q}_0^{-1} + \mathbf{T}^\top\mathbf{Q}_1^{-1}\mathbf{T} & -\mathbf{T}^\top\mathbf{Q}_1^{-1} \\ \mathbf{0} & -\mathbf{Q}_1^{-1}\mathbf{T} & \mathbf{F}^\top\mathbf{W}_2\mathbf{F} + \mathbf{Q}_1^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\mathbf{d}}_0 \\ \widehat{\mathbf{d}}_1 \\ \widehat{\mathbf{d}}_2 \end{pmatrix}$$

where $\widehat{\mathbf{d}}_k = \mathbf{F}^\top \mathbf{W}_k^{\frac{1}{2}} \mathbf{d}_k$.

The mse of $(\widehat{\mathbf{m}}_0, \widehat{\mathbf{m}}_1, \widehat{\mathbf{m}}_2)$ is then

$$\mathsf{tr}\left[\begin{pmatrix} \mathbf{F}^\top\mathbf{W}_0\mathbf{F} + \mathbf{T}^\top\mathbf{Q}_0^{-1}\mathbf{T} & -\mathbf{T}^\top\mathbf{Q}_0^{-1} & \mathbf{0} \\ -\mathbf{Q}_0^{-1}\mathbf{T} & \mathbf{F}^\top\mathbf{W}_1\mathbf{F} + \mathbf{Q}_0^{-1} + \mathbf{T}^\top\mathbf{Q}_1^{-1}\mathbf{T} & -\mathbf{T}^\top\mathbf{Q}_1^{-1} \\ \mathbf{0} & -\mathbf{Q}_1^{-1}\mathbf{T} & \mathbf{F}^\top\mathbf{W}_2\mathbf{F} + \mathbf{Q}_1^{-1} \end{pmatrix}^{-1}\right].$$

Now, consider that we wish to find the design $\mathbf{w}_2$ after having already carried out experiments $\mathbf{w}_0$ and $\mathbf{w}_1$. In the tradition of A-optimal design we would again minimize the mse. However, as is evident the mse does not depend on the data $\mathbf{d}_0, \mathbf{d}_1$, thus we will again introduce the amse and scale the mse by the available estimates of $\mathbf{m}_k$,

$$\begin{pmatrix} \mathsf{diag}(\boldsymbol{\tau}_0(\widehat{\mathbf{m}}_0)) & & \\ & \mathsf{diag}(\boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1)) & \\ & & \mathsf{diag}(\boldsymbol{\tau}_2(\mathbf{T}\widehat{\mathbf{m}}_1)) \end{pmatrix}.$$

Note that since the objective is to design for the third experiment, an estimate for $\mathbf{m}_2$ does not yet exist, thus to approximate $\widehat{\mathbf{m}}_2$ we use the dynamics to propagate $\widehat{\mathbf{m}}_1$ forward one time step.

Defining $\mathbf{S}_k = \mathbf{F}^\top\mathbf{W}_k\mathbf{F} + \mathbf{Q}_{k-1}^{-1} + \mathbf{T}^\top\mathbf{Q}_k^{-1}\mathbf{T}$ and $\mathbf{A}_k = -\mathbf{Q}_k^{-1}\mathbf{T}$ for $k = 0, 1...$, the design for the k^{th} experiment minimizes the general amse

$$\boldsymbol{\Phi}_k(\mathbf{w}_k) = \tag{22}$$
$$\mathsf{tr}\left[\mathsf{diag}\begin{pmatrix} \boldsymbol{\tau}_0(\widehat{\mathbf{m}}_0) \\ \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix}\begin{pmatrix} \mathbf{S}_0 & \mathbf{A}_0^\top & 0 & 0 \\ \mathbf{A}_0 & \mathbf{S}_1 & \mathbf{A}_1^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \mathbf{A}_{k-1} & \mathbf{F}^\top\mathbf{W}_k\mathbf{F} + \mathbf{Q}_{k-1}^{-1} \end{pmatrix}^{-1}\right]$$

Given the amse, the noisy A-optimal design optimization problem is constructed in exactly the same fashion as the noiseless case, by minimizing the amse for $\mathbf{w}_k$ with a sparsity penalty on $\mathbf{w}_k$.

$$\begin{aligned} \min_{\mathbf{w}_k} \quad & \left\{ \mathsf{tr}\left[\mathsf{diag}\begin{pmatrix} \boldsymbol{\tau}_0(\widehat{\mathbf{m}}_0) \\ \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix}\begin{pmatrix} \mathbf{S}_0 & \mathbf{A}_0^\top & 0 & \\ \mathbf{A}_0 & \mathbf{S}_1 & \mathbf{A}_1^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{F}^\top\mathbf{W}_k\mathbf{F} + \mathbf{Q}_{k-1}^{-1} \end{pmatrix}^{-1}\right] + \beta\mathbf{e}^\top\mathbf{w}_k \right\} \\ \text{s.t.} \quad & 0 \geq \omega_i. \end{aligned}$$

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

This formulation provides an alternate method of designing for a future experiment given historic data, where the error in the dynamics is now propagated in time.

Given the adaptive optimal design methods, we now discuss the necessary numerical optimization.

4 Numerical optimization

In this section we present the calculation of the gradients required to solve the design optimization problems for both the noiseless and noisy dynamics. We chose the steepest descent method to solve the non-linear design optimization problem.

4.1 Noiseless dynamics

Finding an optimal design for time step k involves minimizing the objective functional (**??**). This is a difficult task for large problems, particularly due to the presence of 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.

To avoid computing the trace explicitly, a stochastic Hutchinson trace estimator is used such that Equation (**??**) can be written as

$$\phi_k(\mathbf{w}_k) = \mathbf{v}^\top \mathsf{diag}(\boldsymbol{\tau}) \Big(\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G} \Big)^{-1} \mathbf{v},$$

where $\mathbf{G} = \sum_{j=1}^{k-1} \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top \mathbf{L}$ and $\mathbf{v}$ is a random vector with equally distributed values of 1 and -1 [15].

The calculation of the gradient of $\phi_k(\mathbf{w}_k)$ involves computing the derivative of an inverse matrix times a vector,

$$\nabla_{\mathbf{w}_k} \phi_k(\mathbf{w}_k) = \mathbf{v}^\top \mathsf{diag}(\boldsymbol{\tau}) \; \nabla_{\mathbf{w}_k} \left(\Big(\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G} \Big)^{-1} \mathbf{v} \right). \tag{23}$$

This calculation can easily be carried out by defining $\mathbf{z}$ such that

$$\mathbf{z} = (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G})^{-1} \mathbf{v} \quad \Leftrightarrow \quad (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G}) \mathbf{z} = \mathbf{v}, \tag{24}$$

recalling that $\mathbf{W}_k = \mathsf{diag}(\mathbf{w}_k)$, and differentiating implicitly to obtain

$$\mathbf{F}_k^\top \mathsf{diag}(\mathbf{F}_k \mathbf{z}) + (\mathbf{F}_k^\top \mathsf{diag}(\mathbf{w}_k) \mathbf{F}_k + \mathbf{G}) \nabla_{\mathbf{w}_k} \mathbf{z} = \mathbf{0}.$$

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

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

Transposing, defining $\mathbf{C} = \mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G}$, and $\mathbf{C}\mathbf{y} = \mathsf{diag}(\boldsymbol{\tau})\mathbf{v}$, the gradient of the design objective function, equation (**??**), is then

$$-\mathsf{diag}(\mathbf{F}_k \mathbf{z}) \mathbf{F}_k \mathbf{y} + \beta \mathbf{e}, \tag{25}$$

where $\mathbf{F}_k = \mathbf{F}\mathbf{T}^k$. Unfortunately due to the inclusion of the monitor function $\boldsymbol{\tau}$, the gradient is not as a "nice" as in the classic A-optimal design since we must now solve two linear systems, $\mathbf{C}\mathbf{z} = \mathbf{v}$ and $\mathbf{C}\mathbf{y} = \mathsf{diag}(\boldsymbol{\tau})\mathbf{v}$. However, we can make use of the conjugate gradient method to solve the systems to avoid ever forming $\mathbf{C}$ since only matrix vector products are required.

Since the design problem is nonlinear with respect to $\mathbf{w}_k$, a Newton type iterative method (steepest descent) is used with a backtracking line search.

4.2 Noisy dynamics

Minimizing the objective function for noisy dynamics is somewhat more involved then for the noiseless case. First, to approximate the trace operation and reduce computational cost we will

again apply a Hutchinson trace estimator to equation (22),

$$\mathbf{\Phi}(\mathbf{w}_k) = \begin{pmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}^\top \mathsf{diag}\begin{pmatrix} \boldsymbol{\tau}_0(\widehat{\mathbf{m}}_0) \\ \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix} \begin{pmatrix} \mathbf{S}_0 & \mathbf{A}_0^\top & 0 & \\ \mathbf{A}_0 & \mathbf{S}_1 & \mathbf{A}_1^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}, \tag{26}$$

where each vector $\mathbf{v}_k$ is a random vector of values $+1, -1$ with equal distribution.

To compute the gradient of $\mathbf{\Phi}$ we define the vector $(\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_k)^\top$ such that

$$\begin{pmatrix} \mathbf{S}_0 & \mathbf{A}_0^\top & 0 & \\ \mathbf{A}_0 & \mathbf{S}_1 & \mathbf{A}_1^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix} \begin{pmatrix} \mathbf{z}_0 \\ \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \end{pmatrix} = \begin{pmatrix} \mathbf{v}_0 \\ \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{pmatrix} \tag{27}$$

and note that the computation of the gradient of equation (26) requires calculating the gradient of $(\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_k)^\top$. We then implicitly differentiate to obtain $\nabla_{\mathbf{w}_k}(\mathbf{z}_0, \mathbf{z}_1, ..., \mathbf{z}_k)^\top$. Recalling that $\mathbf{S}_k = \mathbf{F}^\top \mathbf{W}_k \mathbf{F} + \mathbf{Q}_{k-1}^{-1} + \mathbf{T}^\top \mathbf{Q}_k^{-1} \mathbf{T}$ and $\mathbf{A}_k = -\mathbf{Q}_k^{-1}\mathbf{T}$, and that only $\mathbf{S}_k(\mathbf{w}_k))$ depends on $\mathbf{w}_k$ it is easy to show that the gradient is given by

$$\underbrace{\begin{pmatrix} \mathbf{S}_0 & \mathbf{A}_0^\top & 0 & \\ \mathbf{A}_0 & \mathbf{S}_1 & \mathbf{A}_1^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix}}_{\mathbf{E}} \begin{pmatrix} \nabla \mathbf{z}_0 \\ \nabla \mathbf{z}_1 \\ \vdots \\ \nabla \mathbf{z}_k \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ -\mathbf{F}^\top \mathsf{diag}(\mathbf{F}\mathbf{z}_k) \end{pmatrix}}_{\mathbf{B}} \tag{28}$$

Note that the gradient of $\mathbf{z}$ is a matrix with dimensions $kN \times N$.

Defining the large matrix as $\mathbf{E}$ and the right hand side as $\mathbf{B}$, we can write the gradient of $\mathbf{z}$ as

$$\begin{pmatrix} \nabla \mathbf{z}_0 \\ \nabla \mathbf{z}_1 \\ \vdots \\ \nabla \mathbf{z}_k \end{pmatrix} = \mathbf{E}^{-1}\mathbf{B}, \tag{29}$$

and thus, after transposing, the gradient of $\mathbf{\Phi}$ is

$$\nabla \mathbf{\Phi}(\mathbf{w}_k) = \mathbf{B}^\top \mathbf{E}^{-\top}(\boldsymbol{\upsilon} \odot \bar{\mathbf{v}}) \tag{30}$$

where $\boldsymbol{\upsilon}$ is the vector of all $\boldsymbol{\tau}_k$'s and $\bar{\mathbf{v}}$ is the vector of all $\mathbf{v}_k$'s. Because the matrix $\mathbf{E}$ is not symmetric, GMRES [19] (generalized minimum residual norm) was used to solve the system $\mathbf{E}^{-\top}(\boldsymbol{\upsilon} \odot \bar{\mathbf{v}})$ without having to explicitly form the big matrix $\mathbf{E}$.

5 Numerical Examples

To demonstrate the adaptive design method we begin with a toy problem designed to fully demonstrate the ability of our method to track the motion of a target. The toy model problem consists

of a circular target of constant concentration which is moved vertically in a whole space of constant background hydraulic conductivity. The motion of the target is governed by a time invariant velocity field generated by a simple source-sink system with no diffusion. To image the target we chose a geophysical borehole seismic tomography method since it is linear with respect to $\mathbf{m}$, tends to cover a large area, and provides an effective picture when tracking a target.

The partial differential equations governing both the imaging and the transport of the target are discretized on a 2D computational domain $\Omega = [0, 400] \times [0, 100]$ meters, divided into $[100, 50]$ cells of width $h = [4, 2]$ meters.

5.1 Imaging: seismic borehole tomography

Borehole seismic tomography measures the travel time of an acoustic wave traveling between sources n_s and receivers n_r. Travel times are given by integrating the slowness $m(t, \vec{\mathbf{x}})$ or inverse acoustic velocity of the media along a linear ray path Γ,

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

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

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

Equation (5.1) is discretized on a standard staggered 2D grid using a finite volume scheme placing m and $\mathbf{d}$ at the cell centers, see figure 1 below.

The discrete tomography experiment in matrix form is then,

$$\mathbf{F}\mathbf{m}_k = \mathbf{d}_k, \text{ for } k = 1, \ldots, n. \tag{31}$$

The initial survey setup is pictured in figure 2, and includes 2 boreholes with 20 sources in the east borehole and 30 receivers in the west borehole. The initial design was chosen to overly cover all the area in the flow domain, with the goal of significantly reducing the number of data required to image the moving target.

5.2 Dynamics: Tracer advection

The subsurface dynamics are described by the tracer advection equations as presented in [9], which characterize the transport of a solute in a fully saturated porous media,

$$\frac{\partial(c\rho)}{\partial t} + \nabla \cdot\, c\rho\vec{\mathbf{u}} = 0, \tag{32a}$$
$$\text{subject to } c\rho(0, \vec{\mathbf{x}}) = c_0\rho(\vec{\mathbf{x}}),$$

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

$$\nabla \cdot\, \vec{\mathbf{u}} = q \tag{32b}$$
$$\vec{\mathbf{u}} = \mathbf{K}(\vec{\mathbf{x}})\nabla p \tag{32c}$$
$$p(0, \vec{\mathbf{x}}) = p_0(\vec{\mathbf{x}}), \tag{32d}$$

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 is linear, and that the motion of m is governed by Equation (32a). From now on we will refer only to $\mathbf{m}$. For more information see [11].

To discretize the advection of the tracer, a particle in cell discretization was chosen such that equation (32a) is then given by,

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

The matrix $\mathbf{T}(\mathbf{u})$ is the transpose of a bi-linear interpolation matrix which pushes mass from a cell center along the path $\mathbf{u}\Delta t$ and spreads it to the 4 neighboring cell centers accordingly.

The remainder of the flow equations were discretized on a staggered grid using the finite volume method with the components of the velocity field located on the cell edges, and the tracer concentration at the cell centers.

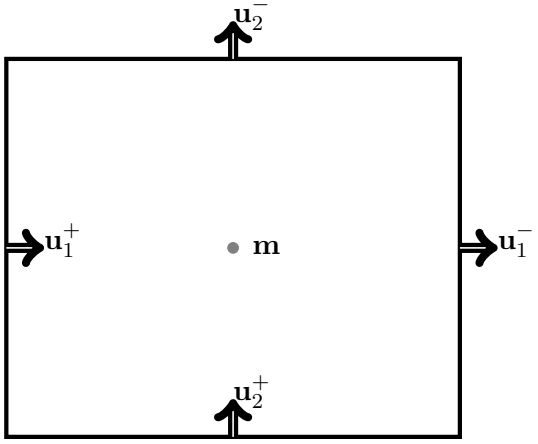

Figure 1: An example of a stagered grid cell with model parameters at the cell center and fluid velocity field on the edges.

The fluid velocity field $\mathbf{u}$ was computed by solving equations (32b) and (32c) with a constant hydraulic conductivity for 1e$-$03 kg/s^{2}m^2.

Tomography rays	Flow field

Figure 2: Initial experiment setup. Sources are pictured in green on the left and receivers in blue on. The initial setup covers the entire flow domain with 20 sources and 30 receivers.

5.3 Regularization

We chose a gradient operator for the linear regularization operator $\mathbf{L} = \mathsf{GRAD}$ to promote smoothness in the recovered models. The gradient maps from cell centers to edges with Dirichlet boundary conditions.

Two different inversions were carried out to recover models $\mathbf{m}_k$. First a linear inversion was performed with the gradient operator, and second a non-linear inversion was performed with smoothed total variation (TV) as the regularization [4]. The total variation regularization promotes discontinuous boundaries, or sharp edges. This is particularly valid for our experiment since we have chosen advection without a diffusion term to transport a concentration of tracer. The tracer model is either 100% concentrated in the blob and zero everywhere else. In this ideal case there should be no diffusion of the tracer, and thus the boundaries should remain sharp as it is advected along in the fluid velocity field.

In continuous space smoothed total variation is described by the following relation,

$$R(m) = \alpha \int \phi(|\nabla|)\ d\vec{\mathbf{x}}. \tag{33}$$

A standard discrete approximation for the smoothed total variation regularization found in [4] is,

$$R(\mathbf{m}) = \alpha h^2 \mathbf{e}^\top \sqrt{\mathbf{A}_f^c \left((\mathsf{GRAD}\ \mathbf{m}) \odot (\mathsf{GRAD}\ \mathbf{m})\right) + \epsilon}, \tag{34}$$

where the gradient operator, GRAD, maps from cell centers to cell faces, and the averaging matrix $\mathbf{A}_f^c$ maps from faces to cell centers.

Although we have formulated the optimal design based on a linear estimation of $\mathbf{m}$, it is not unreasonable to estimate $\mathbf{m}_k$ or $\mathbf{m}_0$ with a non-linear regularization, provided that this will contribute to recovering the best estimate of $\mathbf{m}_k$. As you can see in figure 3 below, the overall relative model error for the total variation regularization is significantly lower than that of the models recovered using the linear regularization. Thus the estimates obtained using the TV regularization are more accurate with respect to the true model for this problem.

Figure 3: Plot of the relative error ($\|\mathbf{m}_1^{true} - \widehat{\mathbf{m}}_1\|/(\|\mathbf{m}_1^{true}\|)$) versus the regularization parameter α for both the linear regularization and smoothed total variation.

I'm not super happy with this figure...it's just to illustrate the hand wavy argument about using different regularizations to reconstruct m's.

5.4 Monitor function

We chose the binary monitor function

$$\boldsymbol{\tau}_B = \begin{cases} 0, & \text{if } \boldsymbol{\delta} \leq 0 \\ 1, & \text{if } \boldsymbol{\delta} > 0 \end{cases}$$

where $\boldsymbol{\delta} = \mathbf{m}_k - \mathbf{m}_{\text{background}}$ to provide historic data to the adaptive A-optimal design problem. The background model $\mathbf{m}_{\text{background}}$, is the same model used to compute the velocity field in equation (32c). In other words, we are designing only to track the target instead of fully imaging the entire domain.

6 Design results: Noiseless dynamics

To generate data, the initial tracer concentration was marched along in time for time steps of 25 days and measured by conducting a tomography survey at each time with Gaussian noise added to each data set. In total 9 experiments were conducted for times $t_0, t_1, ..., t_8$.

The initial design for time point t_0 did not include the dynamics. This amounts to setting $\boldsymbol{\tau}_0 = 1$ in algorithm **??**. The design algorithm is therefore identical to the static case and produces an optimal design based on the physics and that of the linear regularization operator $\mathbf{L}$, without any historic data. This is apparent in both the plot of the weights for experiment 1 at time t_0 in figure 4 and the image of the rays in figure 5, where the design specifies fewer rays that cover the entire domain. In other words, the design algorithm has no information about the location of the target at this point.

For each experiment the amse was plotted versus the number of nonzero weights for a set of regularization parameters β, see figure 4 for examples for times t_0, t_2, t_4. From these curves, the best set of weights were chosen such that the weights were sparse, but also so the amse was kept reasonably small. The experiment weights are also plotted.

Exp	Sparsity	Weights
t_0		
t_2		
t_4		

Figure 4: Noiseless dynamics: Plots of the amse vs. the number of nonzero weights (left column), and the weights used to conduct the experiment for experiments 1,3,and 5, at times t_0, t_2, t_4.

After each each experimental design was determined, both $\mathbf{m}_0$ and the current model $\mathbf{m}_k$ were reconstructed from the reduced set of data corresponding to the optimal experimental design. The new estimate for $\mathbf{m}_0$ was then used to compute $\boldsymbol{\tau}_k$. In figure 5 the rays corresponding to the optimal experimental design data used to estimate the current models are pictured in blue over an image of the model. Note that the rays tend to follow the target as it moves through the domain, and rays which do not pass through the target are not included in the design. The total number of data required to image the models is significantly reduced. In most cases the number of data are approximately 40 of a possible 600. There do appear to be some rays that do not pass through the target that are included in the optimal design, and thus would not contribute to the image reconstruction. However, since the reconstruction of $\mathbf{m}_0$ is never perfect and there is noise in the data, these spurious rays are expected.

It is also apparent that the number of spurious rays increases further in time. In particular, at time t_6 it appears that the design algorithm has a harder time generating a design that captures the target well. There are many more rays which do not pass through the target. This is likely due to the increasing number of multiplications by the transport matrix $\mathbf{T}$ as time progresses. This is by no means a perfect process.

It is also apparent in the reconstructions of $\mathbf{m}_k$ with smoothed total variation regularization does a better job, even when the optimal design was computed for the linear regularization model

estimation problem. Again, this is to be expected since we have chosen a model with discontinuous edges, and are transporting the target only by advection.

exp t_0	exp t_1	exp t_2	exp t_3	exp t_4	exp t_5	exp t_6	exp t_7	exp t_8
$\#d = 350$	$\#d = 43$	$\#d = 46$	$\#d = 57$	$\#d = 35$	$\#d = 35$	$\#d = 30$	$\#d = 25$	$\#d = 19$
$\mathbf{m}_{t_0}^{\mathsf{True}}$	$\mathbf{m}_{t_1}^{\mathsf{True}}$	$\mathbf{m}_{t_2}^{\mathsf{True}}$	$\mathbf{m}_{t_3}^{\mathsf{True}}$	$\mathbf{m}_{t_4}^{\mathsf{True}}$	$\mathbf{m}_{t_5}^{\mathsf{True}}$	$\mathbf{m}_{t_6}^{\mathsf{True}}$	$\mathbf{m}_{t_7}^{\mathsf{True}}$	$\mathbf{m}_{t_8}^{\mathsf{True}}$
$\mathbf{m}_{t_0}^{\mathsf{TV}}$	$\mathbf{m}_{t_1}^{\mathsf{TV}}$	$\mathbf{m}_{t_2}^{\mathsf{TV}}$	$\mathbf{m}_{t_3}^{\mathsf{TV}}$	$\mathbf{m}_{t_4}^{\mathsf{TV}}$	$\mathbf{m}_{t_5}^{\mathsf{TV}}$	$\mathbf{m}_{t_6}^{\mathsf{TV}}$	$\mathbf{m}_{t_7}^{\mathsf{TV}}$	$\mathbf{m}_{t_8}^{\mathsf{TV}}$
$\mathbf{m}_{t_0}^{\mathsf{Lin}}$	$\mathbf{m}_{t_1}^{\mathsf{Lin}}$	$\mathbf{m}_{t_2}^{\mathsf{Lin}}$	$\mathbf{m}_{t_3}^{\mathsf{Lin}}$	$\mathbf{m}_{t_4}^{\mathsf{Lin}}$	$\mathbf{m}_{t_5}^{\mathsf{Lin}}$	$\mathbf{m}_{t_6}^{\mathsf{Lin}}$	$\mathbf{m}_{t_7}^{\mathsf{Lin}}$	$\mathbf{m}_{t_8}^{\mathsf{Lin}}$

Figure 5: Noiseless dynamics: Optimal designs and recovered models for 9 experiments. The top row shows the rays and the recovered model, with the number of rays (#d = 350) given below, followed by the true models, the models recovered using total variation, and finally the models recovered using the linear gradient regularization.

7 Results: Noisy dynamics

To demonstrate the adaptive method for the noisy formulation presented in section 3.2, the tracer target was marched along in time while determining an optimal survey design for the future experiment given all historic data. For the first experiment $\mathbf{w}_0$, the initial naive experiment computed in the noiseless example was used. Each model was reconstructed at each time step using the reduced data sets determined by the optimal design. Models were reconstructed using the non-linear smoothed total variation regularization, and also with the linear regularization. The covariance matrices $\mathbf{Q}_k$ were set as a scale $\alpha_k \mathbf{I}$ for the initial experiment, where $\alpha_k = 10^{-8}$.

The results are pictured in figure 6, below. It is apparent that we were again able to track the motion of the target with a significantly reduced set of measurements. However, in this case the overall number of rays passing through the target were higher than that of the noiseless formulation. There also appear to be less spurious rays. The recoveries for $\mathbf{m}_k$ using the linear regularization term approximate the true models to a better extent that those of the noiseless formulation.

Unfortunately although the designs appear to be more accurate in this case, the computation time is significantly longer than the noiseless formulation. This is particularly true if the full system is formed instead of using GMRES to solve the big systems for all of the $\mathbf{z}$'s.

Things to quantify - relative data fit error. Also should comment on computation time? The noiseless method is faster. For the noisy case I solved it two ways, building the whole system and using backslash, or not forming anything and using gmres. Both are slow. Debating on leaving this as a qualitative explanation vs. actually quoting some numbers...Thoughts Eldad?

exp t_0	exp t_1	exp t_2	exp t_3	exp t_4	exp t_5	exp t_6	exp t_7	exp t_8
$\#d = 297$	$\#d = 58$	$\#d = 62$	$\#d = 77$	$\#d = 65$	$\#d = 65$	$\#d = 81$	$\#d = 54$	$\#d = 14$
$\mathbf{m}_{t_0}^{\mathsf{True}}$	$\mathbf{m}_{t_1}^{\mathsf{True}}$	$\mathbf{m}_{t_2}^{\mathsf{True}}$	$\mathbf{m}_{t_3}^{\mathsf{True}}$	$\mathbf{m}_{t_4}^{\mathsf{True}}$	$\mathbf{m}_{t_5}^{\mathsf{True}}$	$\mathbf{m}_{t_6}^{\mathsf{True}}$	$\mathbf{m}_{t_7}^{\mathsf{True}}$	$\mathbf{m}_{t_8}^{\mathsf{True}}$
$\mathbf{m}_{t_0}^{\mathsf{TV}}$	$\mathbf{m}_{t_1}^{\mathsf{TV}}$	$\mathbf{m}_{t_2}^{\mathsf{TV}}$	$\mathbf{m}_{t_3}^{\mathsf{TV}}$	$\mathbf{m}_{t_4}^{\mathsf{TV}}$	$\mathbf{m}_{t_5}^{\mathsf{TV}}$	$\mathbf{m}_{t_6}^{\mathsf{TV}}$	$\mathbf{m}_{t_7}^{\mathsf{TV}}$	$\mathbf{m}_{t_8}^{\mathsf{TV}}$
$\mathbf{m}_{t_0}^{\mathsf{Lin}}$	$\mathbf{m}_{t_1}^{\mathsf{Lin}}$	$\mathbf{m}_{t_2}^{\mathsf{Lin}}$	$\mathbf{m}_{t_3}^{\mathsf{Lin}}$	$\mathbf{m}_{t_4}^{\mathsf{Lin}}$	$\mathbf{m}_{t_5}^{\mathsf{Lin}}$	$\mathbf{m}_{t_6}^{\mathsf{Lin}}$	$\mathbf{m}_{t_7}^{\mathsf{Lin}}$	$\mathbf{m}_{t_8}^{\mathsf{Lin}}$

Figure 6: Noisey dynamics: Optimal designs and recovered models for 9 experiments. The top row shows the rays and the recovered model, with the number of rays (ex. #d = 297) given below. The second row pictures true models, the third shows the models recovered using total variation, and finally the bottom shows models recovered using the linear gradient regularization.

8 Concluding remarks

In this paper we have presented a new method for the design of experiments for a dynamic target which incorporates historic data. Our approach is based on coupling the dynamic governing partial differential equation with the geophysical imaging partial differential equation to estimate model parameters, and the introduction of a monitor function to scale the posterior mean squared error (amse) according to historic model estimators. Our optimal design approach minimizes the amse to determine an optimal experimental design for a future experiment. As a result of this method, optimal designs are conditioned on the both the physics of the problem and historic data. This leads to designs which track the motion of the target, and estimation of the tagret that require a minimum number of data points.

9 References

[1] Jonathan B. Ajo-Franklin. Optimal experiment design for time-lapse traveltime tomography. *Geophysics*, 74(4):Q27–Q40, July 2009.

[2] Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A-Optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized l_0 - sparsification. *SIAM Journal on Scientific Computing*, 36(5):2122–2148, 2014.

[3] Aleksandr Aravkin. *Robust Methods for Kalman Filtering / Smoothing and Bundle Adjustment.* PhD thesis, 2010.

[4] U. M. Ascher, E. Haber, and H. Huang. On Effective Methods for Implicit Piecewise Smooth Surface Recovery. *SIAM Journal on Scientific Computing*, 28(1):339–358, January 2006.

[5] Anthony Curtis Atkinson and Alexander N. Donev. *Optimum Experimental Designs*. Clarendon Press, 1992.

[6] André Bardow. Optimal experimental design of ill-posed problems: The METER approach. *Computers & Chemical Engineering*, 32(1-2):115–124, January 2008.

[7] W. Briggs, V. Henson, and S. McCormick. *A Multigrid Tutorial, Second Edition*. Society for Industrial and Applied Mathematics, second edition, 2000.

[8] Kathryn Chaloner and Isabella Verdinelli. Bayesian Experimental Design : A Review. *Statistical Science*, 10(3):273–304, 1995.

[9] Zhangxin Chen, Guanren Huan, and Yuanle Ma. *Computational Methods for Multiphase Flows in Porous Media*. SIAM, 2006.

[10] Andrew Curtis. Optimal experiment design: cross-borehole tomographic examples. *Geophysical Journal International*, 136(3):637–650, March 1999.

[11] J Fohring, E Haber, and L Ruthotto. Geophysical imaging of fluid flow in porous media. *SIAM Journal on Scientific Computing*, 36(5):218–236, 2014.

[12] E Haber, L Horesh, and L Tenorio. Numerical methods for experimental design of large-scale linear ill-posed inverse problems. *Inverse Problems*, 24(5):055012, October 2008.

[13] E Haber, L Horesh, and L Tenorio. Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems. *Inverse Problems*, 26(2):025002, February 2010.

[14] Eldad Haber, Zhuojun Magnant, Christian Lucero, and Luis Tenorio. Numerical methods for A-optimal designs with a sparsity constraint for ill-posed inverse problems. *Computational Optimization and Applications*, 52(1):293–314, April 2011.

[15] M.F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2):433–450, January 1990.

[16] R E Kalman. A New Approach to Linear Filtering and Prediction Problems 1. *Transaction of the ASME - Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[17] Dean S. Oliver and Yan Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, July 2010.

[18] Friedrich Pukelsheim and William J. Studden. E-Optimal Designs for Polynomial Regression. *The Annals of Statistics*, 21(1):402–415, March 1993.

[19] Youcef Saad and Martin H. Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7(3):856–869, 1986.