

# Adaptive A-optimal experimental design for linear dynamical systems

J. Fohring and E. Haber

July 27, 2015

## Abstract

While optimal experimental design for static linear inverse problems has been well studied, there is little in the way of experimental design methods for dynamical systems. In particular, problems where historic data is included in the optimal design problem, such that the experiment tracks the motion of the model. Thus, we propose an adaptive design method which minimizes the adapted mean square error (amse) of the regularized model estimate, defined by introducing a monitor function which scales the mean squared error according to historic model estimates.

We present the adaptive design method for two model estimation formulations based on the classification of the error in the dynamic process. First, we consider the case where there is no error in the dynamical system and formulate the inverse problem as a PDE constrained optimization problem to recover the initial condition. And secondly, we include the error in the dynamics and formulate the inverse problem as a Kalman smoothing system to recover models for all times.

We demonstrate our approach with A-optimal design criteria for a borehole seismic tomography experiment, where the motion of the model is governed by advection in a constant velocity field, and show that we are able to recover optimal designs which track the motion of the model.

## 1 Introduction

While there exists a vast area of research in optimal experimental design for inverse problems of static problems [? ? ? ? ? ? ? ? ? ?], there is little in the way of optimal design algorithms for ill-posed inverse problems where the model is governed by a dynamical system. Although static design methods can be applied to time dependent partial differential equations (PDE's) [? ?], 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 at 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  $t_1, \dots, t_n$ , 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 dynamical system, and 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 versus 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 [? ]. 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.

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 builds on A-optimal design algorithms presented in [? ] for large-scale ill-posed inverse problems, and on work presented in [? ] 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 [? ]). 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 [? ? ] 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-optimal 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 apply the new adaptive optimal design criteria developed in Section 2 to the case where the model parameters are governed by a dynamical system. We investigate the cases where the dynamic process is either noiseless, or noisy. Section 4 outlines the numerical optimization method for solving the design problems, and finally in section 5 we present a small model problem used to test the adaptive design method. In sections 6 and 7 the results are presented for the two cases of the dynamics.

## 2 Adaptive experimental design

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

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

where  $\mathbf{F}_1$  is the forward operator at time  $t_1$  and  $\mathbf{F}_2$  is the forward operator at time  $t_2$ . The measurement error  $\boldsymbol{\epsilon}$  is assumed to be normally distributed uncorrelated noise,  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{W}^{-1})$ , with zero mean and covariance matrix  $\mathbf{W}^{-1}$ .

At this point we do not estimate any dynamics, our goal is simply to design the experiments at times  $t_1$  and  $t_2$  such that we obtain the “best” recovery of  $\mathbf{m}$  by some criteria. As we see next, this scenario fits the problem where the dynamical system is “exact” (containing no noise) and  $\mathbf{m}$  represents the dynamical system parameters that are time invariant, such as the initial condition.

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

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

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

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

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

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

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

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

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

$$\operatorname{mse}(\mathbf{w}, \mathbf{m}) = \mathbf{E} \|\hat{\mathbf{m}}_1 - \mathbf{m}\|_2^2 = \operatorname{trace}[\mathbf{F}_1 \mathbf{C}^{-2} \mathbf{F}_1^\top \mathbf{W}_1] + \alpha^2 \mathbf{E} \|\mathbf{C}^{-1} \mathbf{L}^\top \mathbf{L} (\hat{\mathbf{m}}_0 - \mathbf{m})\|_2^2. \quad (6)$$

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

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

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

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

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 all elements are larger than 0. For example, in our previous work we proposed to minimize the function

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

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

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

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

which leads to the estimate

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

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

$$\phi_2^\beta(\mathbf{w}_2) = \text{trace} \left[ (\mathbf{F}_1^\top \mathbf{W}_1 \mathbf{F}_1 + \mathbf{F}_2^\top \mathbf{W}_2 \mathbf{F}_2 + \alpha \mathbf{L}^\top \mathbf{L})^{-1} \right] + \beta \sum \mathbf{w}_2 \quad 0 \leq \mathbf{w}_i. \quad (12)$$

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,  $\hat{\mathbf{m}}_1$  has some “interesting” features and some “boring” features. To be more specific, assume that the difference

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

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

$$\tau = \boldsymbol{\delta} \odot \boldsymbol{\delta}, \quad (13)$$

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

$$\text{amse}(\mathbf{w}, \mathbf{m}) = \mathbf{E} \|\hat{\mathbf{m}}_2 - \mathbf{m}\|_{\tau_1}^2 = \mathbf{E}[(\hat{\mathbf{m}}_2 - \mathbf{m})^\top \text{diag}(\tau_1)(\hat{\mathbf{m}}_2 - \mathbf{m})]. \quad (14)$$

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

$$\phi_2(\mathbf{w}_2) = \text{trace} \left[ \text{diag}(\tau_1)(\mathbf{F}_1^\top \mathbf{W}_1 \mathbf{F}_1 + \mathbf{F}_2^\top \mathbf{W}_2 \mathbf{F}_2 + \alpha \mathbf{L}^\top \mathbf{L})^{-1} \right]. \quad (15)$$

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  $t_1, \dots, t_{k-1}$  using the forward operators  $\mathbf{F}_1, \dots, \mathbf{F}_{k-1}$  and assume that we want to design an experiment for the problem at time  $t_k$ . The design criteria in this case reads

$$\phi_k^\beta(\mathbf{w}_k) = \text{trace} \left[ \text{diag}(\tau_{k-1}) \left( \sum_{j=1}^k \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top \mathbf{L} \right)^{-1} \right] + \beta \sum \mathbf{w}_k \quad 0 \leq \mathbf{w}_i, \quad (16)$$

where  $\tau_{k-1}$  is the monitor function that measures 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 assumes the model,  $\mathbf{m}$  is static. In this section we address the case where the model is governed by a (potentially noisy) dynamical system.

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

$$\mathbf{m}_k = \mathbf{T} \mathbf{m}_{k-1} + \boldsymbol{\eta}_k, \quad (17a)$$

$$\mathbf{d}_k = \mathbf{F} \mathbf{m}_k + \boldsymbol{\epsilon}_k, \quad (17b)$$

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

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  $\boldsymbol{\eta}_k = \mathbf{0}$ ,  $\forall k$ . Such a scenario is often assumed in the case of history matching [? ]. 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  $\boldsymbol{\eta}_k \neq \mathbf{0}$  and therefore, the models  $\mathbf{m}_1, \dots, \mathbf{m}_k$  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

$$\mathbf{d}_k = \mathbf{F}_k \mathbf{m}_0 + \boldsymbol{\epsilon}_k \quad (18)$$

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.

Given the assumption that there is no error, the estimation of the initial model  $\mathbf{m}_0$  given  $k$  data sets is

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

and therefore the design function is identical to 16

$$\phi_k^\beta(\mathbf{w}_k) = \operatorname{trace} \left[ \operatorname{diag}(\boldsymbol{\tau}_{k-1}) \left( \sum_{j=1}^k \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top \mathbf{L} \right)^{-1} \right] + \beta \sum \mathbf{w}_k, \quad 0 \leq \mathbf{w}_i. \quad (20)$$

### 3.2 Noisy model dynamics

To address the case where the dynamics are inexact we begin by rewriting the system of equations for all  $\mathbf{m}$ 's,

$$\begin{array}{ll} \mathbf{F}\mathbf{m}_0 - \mathbf{d}_0 = \boldsymbol{\epsilon}_0 & \\ \mathbf{F}\mathbf{m}_1 - \mathbf{d}_1 = \boldsymbol{\epsilon}_1 & \mathbf{m}_1 - \mathbf{T}\mathbf{m}_0 = \boldsymbol{\eta}_1 \\ \vdots & \vdots \\ \mathbf{F}\mathbf{m}_k - \mathbf{d}_k = \boldsymbol{\epsilon}_k & \mathbf{m}_k - \mathbf{T}\mathbf{m}_{k-1} = \boldsymbol{\eta}_k. \end{array}$$

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  $\mathbf{W}_0 = \operatorname{diag}(\mathbf{w}_0)$  is known or can be designed for as was suggested in previous work [? ]. Also note that if  $\boldsymbol{\eta} = \mathbf{0}$  we end up with the previous linear propagation of  $\mathbf{m}_0$ .

In order to apply the amse method we formulate the problem as in Kalman smoothing [? ? ]. Recalling that  $\boldsymbol{\eta}_k \sim N(0, \mathbf{Q}_k)$ , and assuming that all  $\mathbf{Q}_k$  are known, we estimate all models

$\mathbf{m}_0, \dots, \mathbf{m}_k$  by minimizing the following system

$$\begin{pmatrix} \hat{\mathbf{m}}_0 \\ \vdots \\ \hat{\mathbf{m}}_k \end{pmatrix} = \operatorname{argmin} \frac{1}{2} \left\| \begin{pmatrix} \mathbf{W}_0 & & \\ & \ddots & \\ & & \mathbf{W}_k \end{pmatrix}^{\frac{1}{2}} \left( \begin{pmatrix} \mathbf{F} & & \\ & \ddots & \\ & & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{m}_0 \\ \vdots \\ \mathbf{m}_k \end{pmatrix} - \begin{pmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_k \end{pmatrix} \right) \right\|_2^2 \quad (21)$$

$$\begin{aligned} & + \frac{1}{2} \left\| \begin{pmatrix} \mathbf{Q}_1^{-1} & & \\ & \ddots & \\ & & \mathbf{Q}_k^{-1} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} -\mathbf{T} & \mathbf{I} & & \\ & \ddots & \ddots & \\ & & -\mathbf{T} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{m}_0 \\ \vdots \\ \mathbf{m}_k \end{pmatrix} \right\|_2^2 \\ & + \frac{\alpha}{2} \left\| \begin{pmatrix} \mathbf{L} & & \\ & \ddots & \\ & & \mathbf{L} \end{pmatrix} \begin{pmatrix} \mathbf{m}_0 \\ \vdots \\ \mathbf{m}_k \end{pmatrix} \right\|_2^2. \end{aligned} \quad (22)$$

Defining the transport matrix  $\bar{\mathbf{T}}$  for all times as

$$\bar{\mathbf{T}} = \begin{pmatrix} -\mathbf{T} & \mathbf{I} & & \\ & \ddots & \ddots & \\ & & -\mathbf{T} & \mathbf{I} \end{pmatrix},$$

and  $\widehat{\mathbf{m}}_k = (\widehat{\mathbf{m}}_0^\top, \dots, \widehat{\mathbf{m}}_k^\top)^\top$ ,  $\bar{\mathbf{m}}_k = (\mathbf{m}_0^\top, \dots, \mathbf{m}_k^\top)^\top$ ,  $\bar{\mathbf{W}}_k = \operatorname{diag}(\mathbf{w}_0, \dots, \mathbf{w}_k)$ , and  $\bar{\mathbf{Q}}_k^{-1} = \operatorname{blkdiag}(\mathbf{Q}_1^{-1}, \dots, \mathbf{Q}_k^{-1})$  allows us to write the estimation of  $\mathbf{m}$ ,  $\forall k$  in compact form as

$$\widehat{\mathbf{m}}_k = \operatorname{argmin} \frac{1}{2} \left\| (\mathbf{I} \otimes \mathbf{F}) \bar{\mathbf{m}}_k - \bar{\mathbf{d}}_k \right\|_{\bar{\mathbf{W}}_k}^2 + \frac{1}{2} \left\| \bar{\mathbf{T}} \bar{\mathbf{m}}_k \right\|_{\bar{\mathbf{Q}}_k^{-1}}^2 + \frac{\alpha}{2} \left\| (\mathbf{I} \otimes \mathbf{L}) \bar{\mathbf{m}}_k \right\|^2. \quad (23)$$

In a straightforward extension of the previous section, the amse design optimization function for a design at time  $t_k$  is then

$$\begin{aligned} \Phi_k^\beta(\mathbf{w}_k) &= \operatorname{trace} \left[ \operatorname{diag}(\bar{\boldsymbol{\tau}}_k) \left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right)^{-1} \right] + \beta \sum \mathbf{w}_i, \\ 0 &\leq \mathbf{w}_i. \end{aligned} \quad (24)$$

where  $\bar{\boldsymbol{\tau}}_k = (\boldsymbol{\tau}_0(\widehat{\mathbf{m}}_0), \dots, \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}))$ .

Note that we assume that we are at time  $t_{k-1}$  and that experiments  $1, \dots, k-1$  have already been conducted. Therefore, we are only solving for  $\mathbf{w}_k$ . At this point,  $\widehat{\mathbf{m}}_k$  is not known. To predict  $\widehat{\mathbf{m}}_k$  at time  $t_k$  we simply propagate the previous estimator ahead in time,  $\widehat{\mathbf{m}}_k \approx \mathbf{T}\widehat{\mathbf{m}}_{k-1}$ .

The two formulations of the adaptive design method for a dynamical system, yield two different numerical optimization problems. We discuss numerical optimization techniques for the solution of these problems in the following section.

## 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.

## 4.1 Noiseless dynamics

Finding an optimal design for time step  $k$  requires the minimization of the objective functional (16), which involves computing the trace of a large dense matrix. To avoid such an expensive calculation, we use the stochastic Hutchinson trace estimator to approximate the trace. Applying the trace estimator, Equation (16) becomes

$$\phi_k^\beta(\mathbf{w}_k) = \mathbf{v}^\top \text{diag}(\boldsymbol{\tau}) \left( \mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G} \right)^{-1} \mathbf{v} + \beta \mathbf{e}^\top \mathbf{w}_k,$$

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

Note that computing the the gradient of  $\phi_k^\beta(\mathbf{w}_k)$  involves computing the gradient of an inverse matrix. This computation is easily carried out by defining  $\mathbf{z}$  such that

$$\mathbf{z} = (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G})^{-1} \mathbf{v} \quad \Leftrightarrow \quad (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G}) \mathbf{z} = \mathbf{v}, \quad (25)$$

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

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

Solving for the gradient of  $\mathbf{z}$  we have,

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

Substituting  $\nabla_{\mathbf{w}_k} \mathbf{z}$  back into  $\nabla \phi_k^\beta$ , transposing, defining the matrix  $\mathbf{C} = \mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G}$ , and the vectors,  $\mathbf{y}$  and  $\mathbf{z}$  by,

$$\mathbf{C} \mathbf{z} = \mathbf{v} \quad \mathbf{C} \mathbf{y} = \boldsymbol{\tau} \odot \mathbf{v},$$

the gradient of the noiseless design function is

$$\nabla \phi_k^\beta(\mathbf{w}_k) = -(\mathbf{F}_k \mathbf{z}) \odot (\mathbf{F}_k \mathbf{y}) + \beta \mathbf{e}. \quad (26)$$

Unfortunately due to the inclusion of the monitor function  $\boldsymbol{\tau}$ , the computation of the gradient is slightly more complicated compared with the classic A-optimal design case. Here we must solve two linear systems,  $\mathbf{C} \mathbf{z} = \mathbf{v}$  and  $\mathbf{C} \mathbf{y} = \boldsymbol{\tau} \odot \mathbf{v}$  compared with a single linear system for the classic A design. However, since only matrix vector products are required, we can make use of the conjugate gradient method to solve these systems avoiding ever forming  $\mathbf{C}$ .

Due to the non-linearity of the design problem, a gradient descent method is used for the solution of the problem with a backtracking line search to solve for  $\mathbf{w}_k$ . We have also experimented with nonlinear conjugate gradient and limited memory BFGS for the solution of the problem.

## 4.2 Noisy dynamics

To minimize the noisy design function  $\Phi_k^\beta(\mathbf{w}_k)$ , we again first apply the Hutchinson trace estimator to approximate the trace operation in equation (24),

$$\Phi_k^\beta(\mathbf{w}_k) = \bar{\mathbf{v}}^\top \text{diag}(\bar{\boldsymbol{\tau}}_k) \left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right)^{-1} \bar{\mathbf{v}} \quad (27)$$



where  $\bar{\boldsymbol{\tau}} = (\boldsymbol{\tau}_0, \boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_k)^\top$  and  $\bar{\mathbf{v}}$  is a  $kN \times 1$  vector of evenly distributed values of  $[-1, 1]$ . As in the noiseless case, we again note that the computation of the gradient of  $\Phi_k^\beta$  involves the computation of the gradient of a large dense inverse matrix. Therefore we again define the vector  $\bar{\mathbf{z}} = (\mathbf{z}_0, \dots, \mathbf{z}_k)^\top$  such that

$$\left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right) \bar{\mathbf{z}} = \bar{\mathbf{v}}, \quad (28)$$

differentiate with respect to  $\mathbf{w}_k$ , and solve for  $\nabla_{\mathbf{w}_k} \bar{\mathbf{z}}$  to obtain

$$\nabla \bar{\mathbf{z}} = - \left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right)^{-1} \mathbf{B}. \quad (29)$$

where

$$\mathbf{B} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{F}^\top \text{diag}(\mathbf{F} \mathbf{z}_k) \end{pmatrix}. \quad (30)$$

Substituting the gradient of  $\bar{\mathbf{z}}$  back into the gradient of the noisy design objective function and transposing gives,

$$\nabla \Phi_k^\beta(\mathbf{w}_k) = -\mathbf{B}^\top \left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right)^{-\top} (\bar{\boldsymbol{\tau}} \odot \bar{\mathbf{v}}) + \beta \mathbf{e}. \quad (31)$$

To solve the system for  $\bar{\mathbf{z}}$  and for

$$\mathbf{y} = \left( (\mathbf{I} \otimes \mathbf{F})^\top \bar{\mathbf{W}}_k (\mathbf{I} \otimes \mathbf{F}) + \bar{\mathbf{T}}^\top \bar{\mathbf{Q}}_k^{-1} \bar{\mathbf{T}} + \alpha \mathbf{I} \otimes (\mathbf{L}^\top \mathbf{L}) \right)^{-\top} (\bar{\boldsymbol{\tau}} \odot \bar{\mathbf{v}}) \quad (32)$$

the preconditioned conjugate gradient method was used to avoid forming a the large matrix. We again used a steepest descent method for the non-linear design optimization.

## 5 Numerical Examples

We illustrate the adaptive design method with a toy problem designed to fully demonstrate the ability of our method to track motion and areas of interest in the model. The toy example consists of a circular volume of tracer moving vertically in a whole space. The dynamics governing the motion of the tracer are described by a simple tracer advection model. To image the tracer we chose a borehole seismic tomography survey.

The partial differential equations governing both the seismic tomography experiment and the dynamics 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  $\Gamma$ ,

$$\mathbf{d}_{j,k}(t) = \int_{\Gamma_{j,k}} m(t, \vec{\mathbf{x}}) d\ell. \quad (33)$$

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

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

Equation (33) is discretized on a standard cell-centered 2D grid using a finite volume scheme, with  $m$  located at the cell centers.

The discrete tomography experiment is then,

$$\mathbf{F}\mathbf{m}_k = \mathbf{d}_k, \text{ for } k = 0, 1, \dots, n. \quad (34)$$

Here  $\mathbf{F}$  is a discretization of the ray paths. Note that we assume many ray paths and the goal of the experimental design is to choose a subset of rays that “best” recover the model,  $\mathbf{m}$ . The initial survey setup is pictured in figure 1, 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 model.

## 5.2 Dynamics: Tracer advection

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

$$\begin{aligned} \frac{\partial(c\rho)}{\partial t} + \nabla \cdot c\rho\vec{\mathbf{u}} &= 0, \\ \text{subject to } c\rho(0, \vec{\mathbf{x}}) &= c_0\rho(\vec{\mathbf{x}}), \end{aligned} \quad (35a)$$

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

$$\nabla \cdot \vec{\mathbf{u}} = q \quad (35b)$$

$$\vec{\mathbf{u}} = \mathbf{K}(\vec{\mathbf{x}})\nabla p \quad (35c)$$

$$p(0, \vec{\mathbf{x}}) = p_0(\vec{\mathbf{x}}), \quad (35d)$$

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 (35a) [? ]. From now on we will refer only to  $m$  assuming that there is a direct relation between  $m$  and  $c$ .

A particle in cell discretization [? ] was chosen to discretize the advection of the tracer, such that equation (35a) is

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

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 four neighboring cell centers accordingly (see [? ] for a detailed description of the discretization).

Equations (35b) and (35c) were discretized on a staggered grid using the finite volume method with the components of the velocity field located on the cell faces, and the tracer concentration at the cell centers.

The fluid velocity field  $\mathbf{u}$  was computed by solving equations (35b) and (35c) with a constant hydraulic conductivity for  $10\text{kg/s}^2\text{m}^2$ , and pumping rate of  $150\text{m}^3/\text{s}$ .

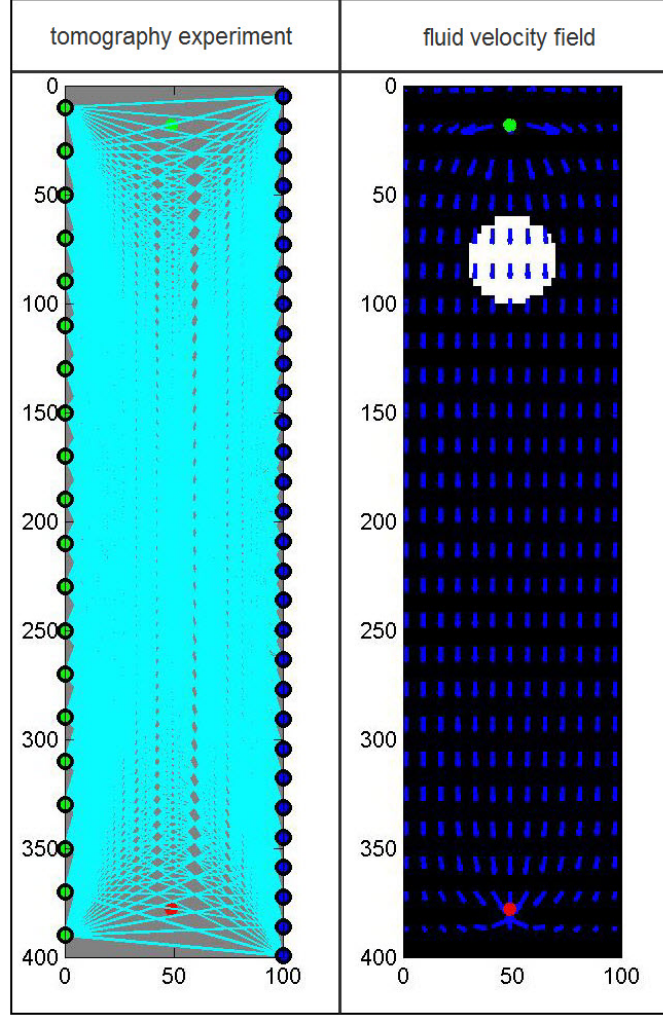


Figure 1: Initial experiment setup. Tomography sources are pictured in green on the left and receivers in blue on the right. The initial setup covers the entire flow domain with 20 sources and 30 receivers. There is one source of fluid, pumped at a constant rate from the top green point, and a sink at the bottom (red) of the domain. Flow is in the downward direction.

### 5.3 Regularization

Two different inversions were carried out to recover models  $\mathbf{m}_k$ , by apply two different regularization functions. First a linear inversion was performed where the operator  $\mathbf{L}$  was chosen to be a gradient operator,  $\mathbf{L} = \text{GRAD}$ , to promote smoothness in the recovered models.

Second, a non-linear inversion was performed with smoothed total variation (TV) as the regularization [? ]. 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 concentrated in the circle 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,

$$R(m) = \alpha \int \phi(|\nabla m|) d\vec{x}. \quad (36)$$

where the convex function  $\phi(c) = \sqrt{c^2 + \epsilon}$ , is an approximation to the  $\ell_1$  norm.

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

$$R(\mathbf{m}) = \alpha h^2 \mathbf{e}^\top \sqrt{\mathbf{A}_f^c ((\text{GRAD } \mathbf{m}) \odot (\text{GRAD } \mathbf{m}))} + \epsilon, \quad (37)$$

where the gradient operator,  $\text{GRAD}$ , 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  $\mathbf{m}_k$ , it is not unreasonable to estimate  $\mathbf{m}_k$  with a non-linear regularization, provided that this will contribute to recovering the best estimates of  $\mathbf{m}_k$ . As can be seen in figure 2 below, the overall relative model error for estimates recovered with the total variation regularization is significantly lower than that of the models estimated using the linear regularization. Thus the estimates obtained using the TV regularization are more accurate with respect to the true model for this particular problem. Since the monitor function has a very large impact on the optimal design, it is desirable to get the best possible estimate of  $\mathbf{m}_k$ .

In all cases, the regularization parameter had the value  $\alpha_k = 100$ . This value was chosen following a cooling strategy which starts with a large value for  $\alpha$ . Alpha is then decreased incrementally after each inversion, until the data misfit is approximately equal to noise.

### 5.4 Monitor function

To design only for the tracer motion, we define delta as  $\delta = |\mathbf{m}_k - \mathbf{m}_{\text{background}}|$ . Because neither the estimated model nor the background model are perfect, we chose a threshold value  $s$  to remove unwanted or erroneous information from delta where the background model  $\mathbf{m}_{\text{background}}$ , is the model used to compute the velocity field in equation (35c).

## 6 Design results: Noiseless dynamics

For the noiseless case, data were simulated by marching the initial tracer concentration along in time, for time steps of 25 days and measured by conducting a tomography survey at each time

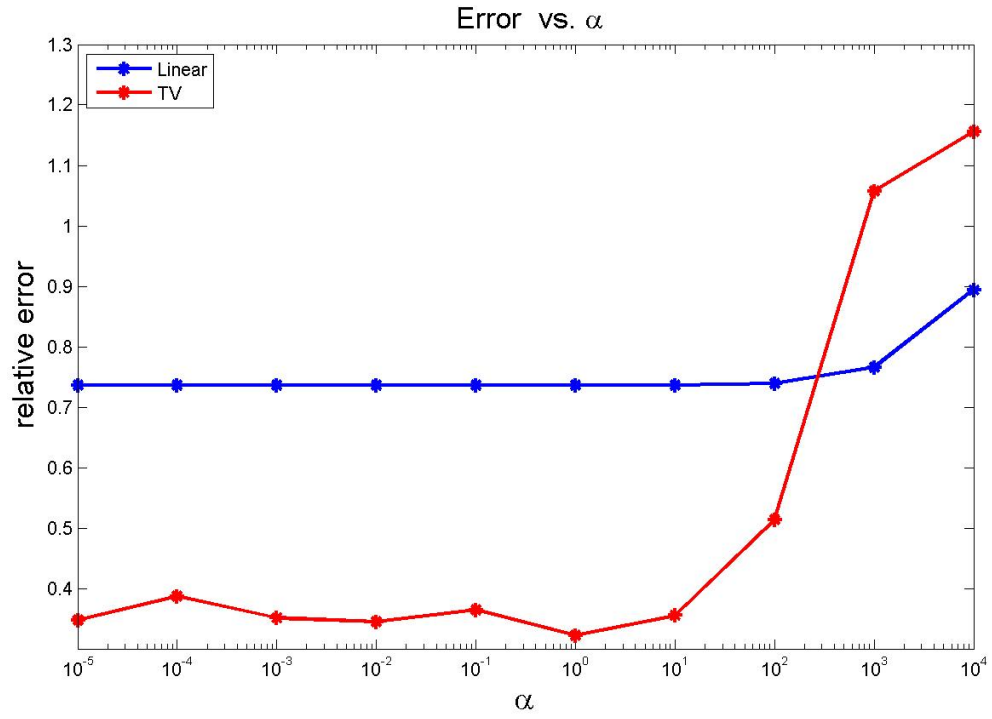


Figure 2: Plot of the relative error ( $\|\mathbf{m}_1^{true} - \hat{\mathbf{m}}_1\|/(\|\mathbf{m}_1^{true}\|)$ ) versus the regularization parameter  $\alpha$  for both the linear regularization and smoothed total variation.

with 4% Gaussian noise added to each data set. In total 9 experiments were conducted for times  $t_0, t_1, \dots, t_8$ .

The initial design for time point  $t_0$  did not include the dynamics. This amounts to setting  $\tau_0 = 1$  everywhere. The design optimization problem in this case is identical to the static case and produces an optimal design based only on the physics and information from the linear regularization operator  $\mathbf{L}$ . This is apparent in both the plot of the weights for experiment 1 at time  $t_0$  in figure 3 and the image of the rays in figure 4, where the design specifies fewer rays that cover the entire domain.

For each experiment the amse was plotted versus the number of nonzero weights for a set of regularization parameters  $\beta$ , see figure 3 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.

After a set of weights were chosen, both  $\mathbf{m}_0$  and the current model  $\mathbf{m}_k$  were reconstructed from the reduced set of data  $\mathbf{d}_k$ . The values of  $\mathbf{d}$  which contribute to the reduced data set correspond to the non-zero weights of  $\mathbf{w}_k$ . The total number of data required to image the models is significantly reduced. In most cases the number of data are on average 40 of a possible 600. In figure 4 the rays corresponding to  $\mathbf{d}_k$  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 pass through the tracer. There are some spurious rays that do not pass through the tracer which are included in the optimal design. However, since the design depends on the monitor function  $\tau(\hat{\mathbf{m}}_0)$ , and the reconstructions of  $\hat{\mathbf{m}}_0$  are never exact, these spurious rays are expected.

It is also apparent that the number of spurious rays increases further in time even though estimates of  $\mathbf{m}_0$  improve as more data are collected. 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 partially due to the increasing number of multiplications by the transport matrix  $\mathbf{T}$  as time progresses, compounding errors, but also the loss of mass as the tracer exits the domain out of the sink. However, the designs still provide enough information to recover models which indicate the location of the tracer.

The reconstructions of  $\mathbf{m}_k$  with smoothed total variation as the regularization are much closer to the true models, 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.

## 7 Results: Noisy dynamics

For the first experiment  $\mathbf{w}_0$ , the initial naive experiment computed in the noiseless example was used in the subsequent design estimations. Each model was reconstructed at each time step after 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 assigned a scalar value for the variance for each time step to represent iid noise in the dynamics, such that  $\mathbf{Q}_k = \sigma_k \mathbf{I}$ . We assigned a constant  $\sigma_k = \pm 0.2$  for all experiments.

The results are pictured in figure 5, below. It is apparent that adapted A-optimal designs were again able to track the motion of the tracer with a significantly reduced set of measurements. The recoveries for  $\mathbf{m}_k$  using the linear regularization term approximate the true models to a better

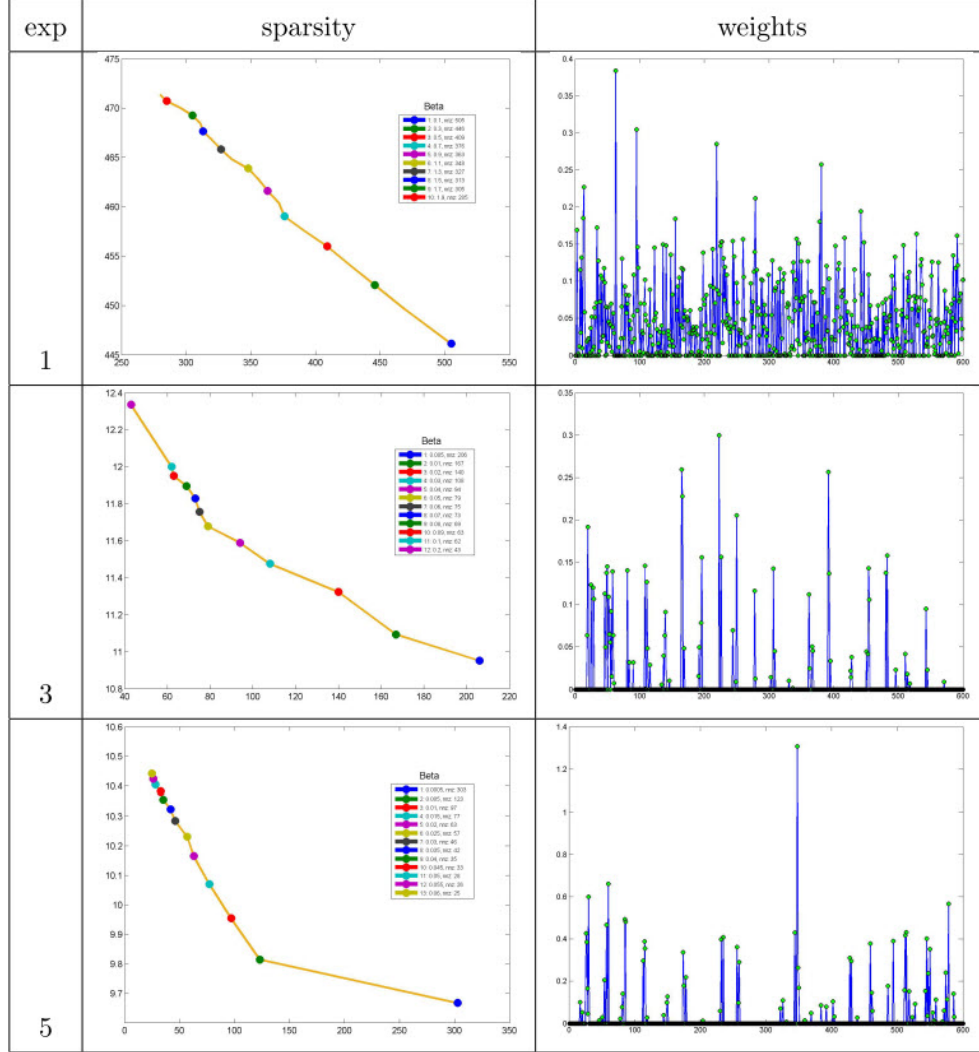


Figure 3: Noiseless dynamics: Plots of the amse vs. the number of nonzero weights (left column), and the weights used to conduct the experiment for experiments 1, 3, and 5, at times  $t_0$ ,  $t_2$ ,  $t_4$ .

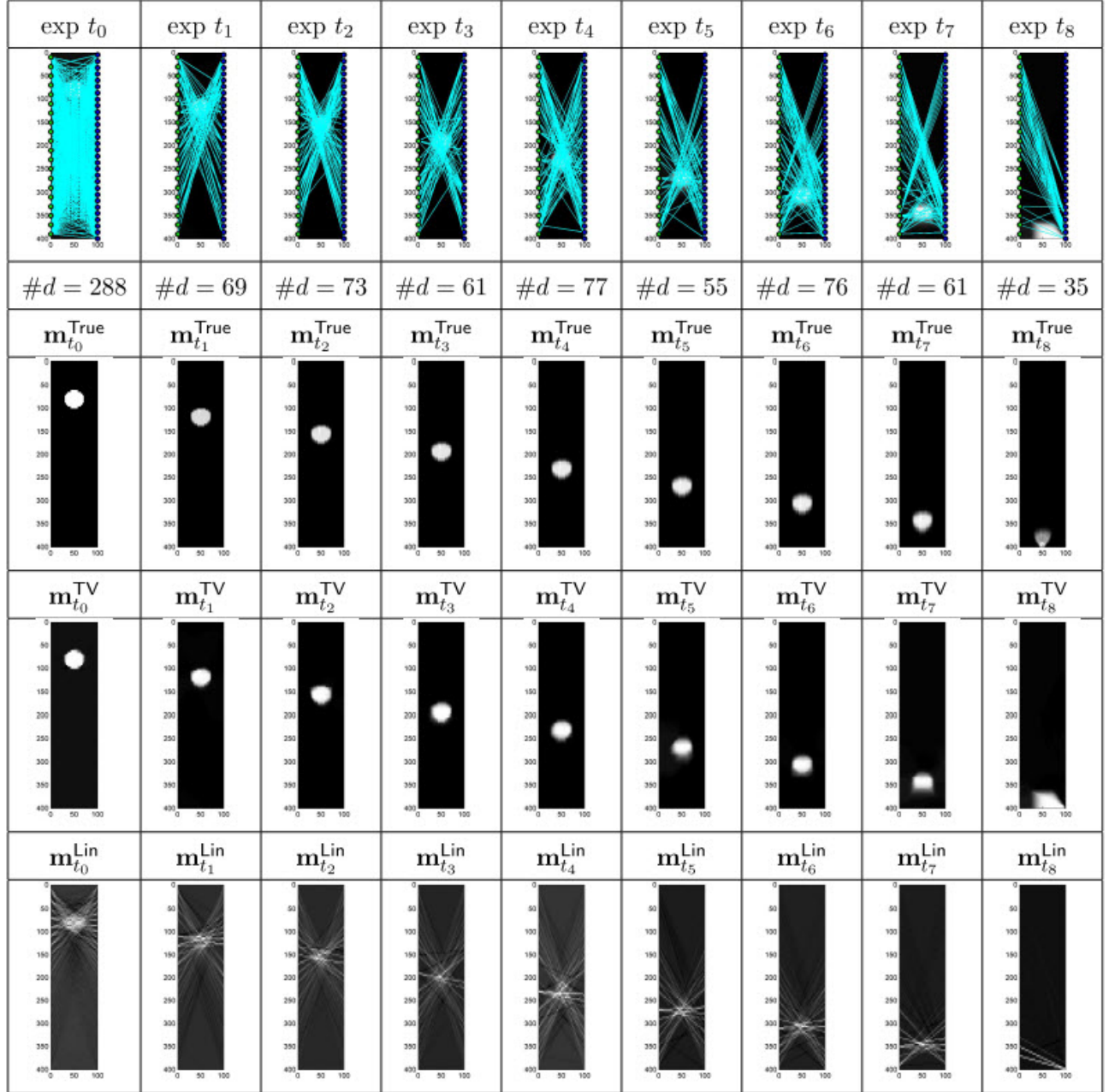


Figure 4: Noiseless dynamics: Optimal designs and recovered models for 9 experiments. The top row shows the rays and the recovered model, with the number of rays ( $\#d = 350$ ) given below, followed by the true models, the models recovered using total variation, and finally the models recovered using the linear gradient regularization.



extent that those of the noiseless formulation. However, those recovered with the smoothed TV regularization were again used in the monitor function.

## 8 Concluding remarks

In this paper we have presented a new method for the design of experiments for dynamical systems which tracks the motion of specific areas of the model by incorporating historic data. The motivation for not simply applying static optimal design methods to the dynamic problem is that the mean squared error of the current estimated model never contains information from previous data. To this end, the design optimization problem “knows” nothing about the changes in the model due to the dynamics and thus designs only reflect information from the physics and the regularization of the ill-posed model estimation problem.

Our adaptive optimal experimental design approach is based on the introduction of a monitor function to scale the mean squared error (amse) of an estimator whose motion is governed by a dynamical system, according to historic data. To determine a design for a future experiment, we minimize the amse with the added constraint that the design is sparse. We consider two model estimation problems whose construction depends on the characterization of the error in the dynamical system. This leads to experimental designs which track the changes in the model with a reduced set of measurements. We have tested our methodology for a geophysical imaging inverse problem where the dynamics of the model parameters are governed by advection in porous media.

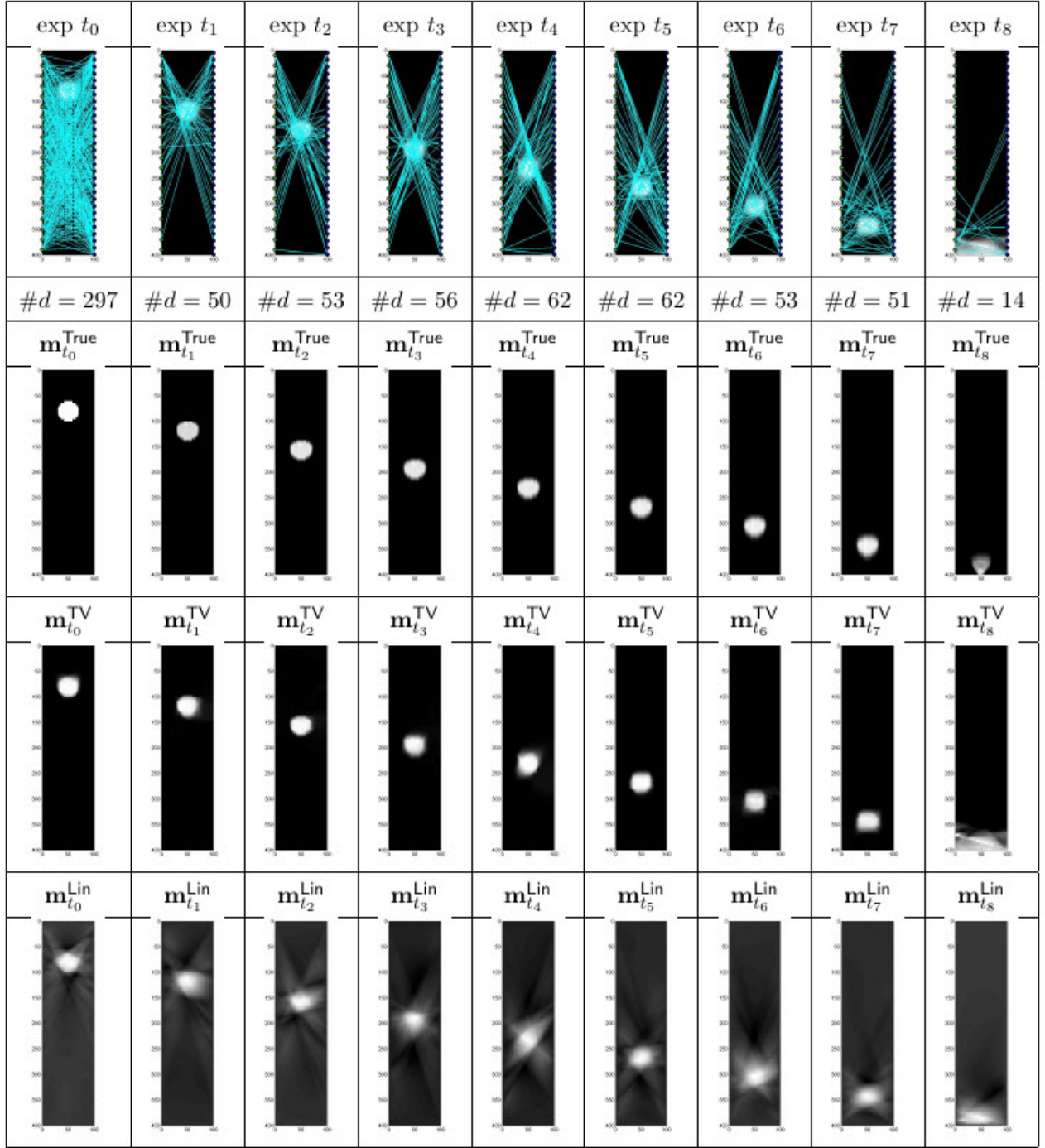


Figure 5: Noisy dynamics: Optimal designs and recovered models for 9 experiments. The top row shows the rays and the recovered model, with the number of rays (ex.  $\#d = 297$ ) given below. The second row pictures true models, the third shows the models recovered using total variation, and finally the bottom shows models recovered using the linear gradient regularization.