Adaptive A-optimal experimental design for dynamical systems

J. Fohring and E. Haber

June 13, 2015

Abstract

The optimal collection of data required to infer the dynamic model parameters governing a partial differential equation (PDE) through the inversion process is important to may applications. One method of optimally collecting data is to control the number of measurements required for a given experiment. Popular methods of optimal data collection, or optimal experimental design, seek a set of measurements which minimize the mean square error (mse) of the estimated model parameters. However, these methods do not include any historic data collected at earlier times.

To include historic 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 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. The 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 our 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.

1   Introduction

While there exists a vast area of research in optimal experimental design for inverse problems for static problems [? ? ? ? ? ? ? ? 1], 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) [2?], this approach results in an a priori estimate of the design for "all times" which depend only on prior information provided in the form of 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 is sufficiently slow to allow for improved design after a first experiment is conducted. The available 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 underline model is a

dynamical system with some known dynamics. Our method can be seen as a post-priori design, which gradually improve the design given the recovery at previous time steps.

The idea of a-priori estimates verses post-priori estimators 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 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 here, builds on A-optimal design algorithms presented in [?] for large-scale ill-posed inverse problems, and on work presented in [9] where historical data was used to recover the properties of a dynamical system.

There are two main paradigms when considering dynamical systems in the context of an inverse problem. In the first, the dynamics is "exact", that is, we assume that the dynamical system does not contain errors and that we can make some indirect measurements of the dynamics. In this case, the goal is to evaluate the parameters of the dynamical system and the initial conditions (see [9]). A second approach is to consider errors in the dynamical systems that are modeled as random errors. This leads to Kalman filtering or smoothing (see [? 3] and reference within) 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 derive from first principals the mse of a linear estimator, then present the adapted mse and define its monitor functions. In section 3 we briefly outline the two methods for combining two different PDEs governing the imaging technique and the dynamic motion, and present the adaptive design algorithms. In section 4 we outline the numerical optimization method for solving the design problem, and finally in sections 5 and 6 we present numerical examples demonstrating the adaptive A-optimal design method.

2   Adaptive experimental design

To present the adaptive experimental design method we begin by reviewing the derivation of A-optimal design from first principals. We start by considering only two measurements 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. 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 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 $\mathbf{d}_1$, which is accomplished by solving the ill-posed optimization problem

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

where $\mathbf{W}_1 = \text{diag}(\mathbf{w}_1)$ is a matrix of inverse standard deviations, that is, the inverse covariance of the noise. Minimizing (2) yields the estimator

$$\widehat{\mathbf{m}}_1 = (\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{d}_1 + \alpha\mathbf{L}^\top\mathbf{L}\widehat{\mathbf{m}}_0). \tag{2}$$

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

$$\begin{aligned}\widehat{\mathbf{m}}_1 - \mathbf{m} \ &= \ \mathbf{C}^{-1}\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1\mathbf{m} + \mathbf{C}^{-1}\mathbf{F}_1^\top\mathbf{W}_1\boldsymbol{\epsilon}_1 + \alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\widehat{\mathbf{m}}_0 \\ &+ \ (\alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\mathbf{m} - \alpha\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}\mathbf{m}) - \mathbf{m}.\end{aligned} \tag{3}$$

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

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

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

$$\text{mse}(\mathbf{w}, \mathbf{m}) = \mathbf{E}\|\widehat{\mathbf{m}}_1 - \mathbf{m}\|^2 = \text{trace}[\mathbf{F}_1\mathbf{C}^{-2}\mathbf{F}_1^\top\mathbf{W}_1] + \alpha^2\mathbf{E}\|\mathbf{C}^{-1}\mathbf{L}^\top\mathbf{L}(\widehat{\mathbf{m}}_0 - \mathbf{m})\|^2. \tag{5}$$

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

$$\phi(\mathbf{w}_1) = \text{mse}(\mathbf{w}_1) = \text{trace}[\mathbf{C}^{-2}\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1] + \alpha\text{trace}[\mathbf{C}^{-2}\mathbf{L}^\top\mathbf{L}]. \tag{6}$$

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

$$\phi_1(\mathbf{w}_1) = \text{trace}\left[(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\right]. \tag{7}$$

In the design problem we seek to obtain a better estimate for $\mathbf{m}$ and therefore we minimize $\phi_1(\mathbf{w}_1)$ with an additional cost on $\mathbf{w}_1$, for example, in our previous work we proposed to minimize the function

$$\phi_1^\beta(\mathbf{w}_1) = \text{trace}\left[(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\right] + \beta\sum\mathbf{w}_1 \qquad 0 \leq \mathbf{w} \tag{8}$$

3

which balances the minimization of the least square error and the cost of the experiment [10].

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

$$\frac{1}{2}\|\mathbf{F}_1\mathbf{m} - \mathbf{d}_1\|_{\mathbf{W}_1}^2 + \frac{1}{2}\|\mathbf{F}_2\mathbf{m} - \mathbf{d}_2\|_{\mathbf{W}_2}^2 + \frac{\alpha}{2}\|\mathbf{L}(\mathbf{m} - \widehat{\mathbf{m}}_0)\|_2^2. \tag{9}$$

which leads to the estimate

$$\widehat{\mathbf{m}}_2(\mathbf{w}_2) = (\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \mathbf{F}_2^\top\mathbf{W}_2\mathbf{F}_2 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{d}_1 + \mathbf{F}_2^\top\mathbf{W}_2\mathbf{d}_2 + \alpha\mathbf{L}^\top\mathbf{L}\widehat{\mathbf{m}}_0) \tag{10}$$

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

$$\phi_2(\mathbf{w}_2) = \mathsf{trace}\left[(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \mathbf{F}_2^\top\mathbf{W}_2\mathbf{F}_2 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\right]. \tag{11}$$

While this design optimization problem can be solved, it does not take into consideration the fact that we already have an estimate of the model, $\widehat{\mathbf{m}}_1$ and therefore, does not include the information we have already obtained at time t_1 in order to plan a better design. It is easy to verify that any linear estimator of the data yields a covariance matrix that is independent of $\mathbf{d}_1$, and therefore the approach above does not yield a post-priori estimator.

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. For example, 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. It may also be the case that we wish to design only for the new best estimate of $\mathbf{m}$, $\widehat{\mathbf{m}}_1$ itself. This implies that $\mathbf{m}_0 = 0$.

To this end, we introduce a **monitor function**, $\boldsymbol{\tau}(\boldsymbol{\delta})$ that measures the change in the model. Examples for this function are

$$\begin{aligned}
\boldsymbol{\tau}_s &= \boldsymbol{\delta} \odot \boldsymbol{\delta} \\
\boldsymbol{\tau}_B &= \boldsymbol{\delta} > 0 = 1
\end{aligned}$$

but other functions can be used. The first monitor function, τ_s, simply measures the change in the estimator compared to the a-priori information in a 2-norm, while the second, τ_N, is a binary function which evenly measures those changes.

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

$$\mathsf{amse}(\mathbf{w}, \mathbf{m}) = \|\widehat{\mathbf{m}}_2 - \mathbf{m}\|_{\boldsymbol{\tau}_1}^2 = (\widehat{\mathbf{m}}_2 - \mathbf{m})^\top\mathsf{diag}(\boldsymbol{\tau}_1)(\widehat{\mathbf{m}}_2 - \mathbf{m}). \tag{13}$$

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

$$\phi_2(\mathbf{w}_2) = \mathsf{trace}\left[\mathsf{diag}(\boldsymbol{\tau})(\mathbf{F}_1^\top\mathbf{W}_1\mathbf{F}_1 + \mathbf{F}_2^\top\mathbf{W}_2\mathbf{F}_2 + \alpha\mathbf{L}^\top\mathbf{L})^{-1}\right]. \tag{14}$$

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

In general, the design for the kth experiment given data collected for times $t_1, ..t_{k-1}$ is

$$\phi_k(\mathbf{w}_k) = \mathsf{trace}\left[\mathsf{diag}(\boldsymbol{\tau}_{k-1})\left(\sum_{j=1}^{k}\mathbf{F}_j^\top\mathbf{W}_j\mathbf{F}_j + \alpha\mathbf{L}^\top\mathbf{L}\right)^{-1}\right] \tag{15}$$

$$\boldsymbol{\tau}_{k-1}(\widehat{\mathbf{m}}_{k-1} - \widehat{\mathbf{m}}_{k-2}) \tag{16}$$

$$\widehat{\mathbf{m}}_{k-1} = \mathsf{argmin}\ \frac{1}{2}\sum_{j=1}^{k-1}\|\mathbf{F}_j\mathbf{m} - \mathbf{d}_j\|_2^2 + R(\mathbf{m}). \tag{17}$$

If the model $\mathbf{m}$ is a dynamic target, then some consideration must be made as to how the dynamics are incorporated.

3 Dynamic target

In the previous section we presented a formulation for adaptive A-optimal experimental design that incorporates historic data given a set of measurements. However, we did not distinguish between a static model $\mathbf{m}$ or a model that is changing over time. To design for a moving target, we must address how the dynamics describing its motion are defined.

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

$$\begin{aligned}\mathbf{m}_k &=& \mathbf{T}\mathbf{m}_{k-1} + \boldsymbol{\eta}_k, && \text{(18a)}\\ \mathbf{d}_k &=& \mathbf{F}\mathbf{m}_k + \boldsymbol{\epsilon}_k, && \text{(18b)}\end{aligned}$$

where the transport matrix $\mathbf{T} : \mathbb{R}^N \to \mathbb{R}^N$ is a discretization of the PDE governing the dynamics in the system. The noise vectors $\boldsymbol{\eta}_k$ and $\boldsymbol{\epsilon}_k$ are assumed to be Gaussian normal, $\boldsymbol{\eta}_k \sim N(\boldsymbol{\eta}_0, \mathbf{Q}_k)$ and $\boldsymbol{\epsilon}_k \sim N(\boldsymbol{\epsilon}_0, \mathbf{W}_k^{-1})$.

In combining equations (18a) and (18b) there are two instances of the dynamics that result in different formulations. These are whether there is noise in the equations describing the motion or not, $\boldsymbol{\eta}_k \overset{?}{=} \mathbf{0}$. We present differing formulations for the so called noiseless case $\boldsymbol{\eta} = \mathbf{0}$, and the noisy case $\boldsymbol{\eta} \neq \mathbf{0}$ in the following sections.

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

$$\begin{aligned}\mathbf{m}_k &=& \mathbf{T}\mathbf{m}_{k-1}, &\text{ or}\\ \mathbf{m}_k &=& \mathbf{T}^{k-1}\mathbf{m}_0.\end{aligned}$$

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

$$\mathbf{d}_k = \mathbf{F}_k \mathbf{m}_0 + \boldsymbol{\epsilon}_k \tag{19}$$

where $\mathbf{F}_k = \mathbf{F}\mathbf{T}^{k-1}$, and the estimation of the initial model $\mathbf{m}_0$ given k data sets is

$$\widehat{\mathbf{m}}_{0,k} = \mathsf{argmin} \ \frac{1}{2} \sum_{j=1}^{k} \|\mathbf{F}\mathbf{T}^{j-1}\mathbf{m}_0 - \mathbf{d}_j\|_{\mathbf{W}_j}^2 + \frac{\alpha}{2}\|\mathbf{L}\mathbf{m}_0\|_2^2. \tag{20}$$

Substituting the definition of $\mathbf{F}_k$ into the amse leads to the amse design functional,

$$\phi_k(\mathbf{w}_k) = \mathsf{trace} \left[\mathsf{diag}(\boldsymbol{\tau}) \Big(\sum_{j=1}^{k} (\mathbf{F}\mathbf{T}^{j-1})^\top \mathbf{W}_j \mathbf{F}\mathbf{T}^{j-1} + \alpha \mathbf{L}^\top \mathbf{L} \Big)^{-1} \right]. \tag{21}$$

In minimizing equation (21) a sparsity penalty, $\beta\|\mathbf{w}_k\|_1$, is imposed on the weight vector to guarantee a design with a minimal number of measurements. Additionally we require that the weights all be positive, as it does not make sense to have negative variances. The adaptive design optimization problem is then

$$\min_{\mathbf{w}_k} \ \left\{ \mathsf{trace}\Big[\mathsf{diag}(\boldsymbol{\tau})\Big(\sum_{j=1}^{k} \mathbf{F}_j^\top \mathbf{W}_j \mathbf{F}_j + \alpha \mathbf{L}^\top \mathbf{L} \Big)^{-1} \Big] + \beta \mathbf{e}^\top \mathbf{w}_k \right\}$$
$$\text{s.t.} \ \ 0 \geq \omega_i, \tag{22}$$

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

The adaptive design method is outlined in algorithm 1 below. Note that in step 5 of the

---

Algorithm 1 Adaptive Optimal Design: Noiseless dynamics

1: Given a prior $\widehat{\mathbf{m}}_0$ for $\mathbf{m}$, set $\boldsymbol{\tau}(\widehat{\mathbf{m}}_0)$ ▷ if there is no prior, $\boldsymbol{\tau} = \mathbf{I}$
2: **for** $k = 1$ to n **do** ▷ n time points
3:     $\mathbf{w}_k = \mathsf{argmin} \ \ \phi_k(\mathbf{w}_k) + \beta \mathbf{e}^\top \mathbf{w}_k$ ▷ design using equation (21)
4:     Given $\mathbf{w}_k$ collect data $\mathbf{d}_k$
5:     $\widehat{\mathbf{m}}_{0,k} = \mathsf{argmin} \ \frac{1}{2}\sum_{j=1}^{k}\|\mathbf{F}\mathbf{T}^{j-1}\mathbf{m}_0 - \mathbf{d}_j\|_2^2 + R(\mathbf{m}_0)$ ▷ re-estimate $\mathbf{m}_0$
6:     $\boldsymbol{\tau}(\widehat{\mathbf{m}}_{0,k})$ ▷ Recompute the monitor function $\boldsymbol{\tau}$
7: **end for**

---

algorithm we have left the regularization as an arbitrary function even though the design functional was derived for a linear regularization. The regularization adjusts the ill-posed inverse problem such that the solutions best estimate the true model parameter distributions, so when calculating the monitor function it is imperative to get a best estimate since this directly effects the experimental design. For example, one could choose total variation as a regularization to promote sharper edges then say the smoothing gradient operator if this is expected of the model.

3.2 Noisy model dynamics

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

$$\mathbf{F}\mathbf{m}_1 - \mathbf{d}_1 = \boldsymbol{\epsilon}_1,$$
$$\mathbf{F}\mathbf{m}_2 - \mathbf{d}_2 = \boldsymbol{\epsilon}_2, \qquad \mathbf{m}_2 - \mathbf{T}\mathbf{m}_1 = \boldsymbol{\eta}_2$$
$$\mathbf{F}\mathbf{m}_2 - \mathbf{d}_2 = \boldsymbol{\epsilon}_2, \qquad \mathbf{m}_3 - \mathbf{T}\mathbf{m}_2 = \boldsymbol{\eta}_3.$$

It is assumed in this case that the first experiment will either be estimated in the regular way with no prior estimate or will be conducted in a naive way. In either case it is assumed that $\mathbf{W}_1 = \mathsf{diag}(\mathbf{w}_1)$ is known. Also note that if $\boldsymbol{\eta} = \mathbf{0}$ we end up with the previous linear propagation of $\mathbf{m}$.

To estimate $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)^\top$ given $(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)^\top$ we minimize the following system with respect to $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)^\top$

$$\frac{1}{2} \left\| \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W}_3 \end{pmatrix}^{\frac{1}{2}} \left(\begin{pmatrix} \mathbf{F} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F} \end{pmatrix} \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{pmatrix} - \begin{pmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \end{pmatrix} \right) \right\|_2^2$$
$$+ \frac{1}{2} \left\| \begin{pmatrix} \mathbf{Q}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_2^{-1} \end{pmatrix}^{\frac{1}{2}} \begin{pmatrix} -\mathbf{T} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\mathbf{T} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{pmatrix} \right\|_2^2 \tag{23}$$

Taking the gradient of (23), setting it equal to zero, and solving for $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3)^\top$ results in the following estimator

$$\begin{pmatrix} \widehat{\mathbf{m}}_1 \\ \widehat{\mathbf{m}}_2 \\ \widehat{\mathbf{m}}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{F}^\top \mathbf{W}_1 \mathbf{F} + \mathbf{T}^\top \mathbf{Q}_1^{-1} \mathbf{T} & -\mathbf{T}^\top \mathbf{Q}_1^{-1} & \mathbf{0} \\ -\mathbf{Q}_1^{-1} \mathbf{T} & \mathbf{F}^\top \mathbf{W}_2 \mathbf{F} + \mathbf{Q}_1^{-1} + \mathbf{T}^\top \mathbf{Q}_2^{-1} \mathbf{T} & -\mathbf{T}^\top \mathbf{Q}_2^{-1} \\ \mathbf{0} & -\mathbf{Q}_2^{-1} \mathbf{T} & \mathbf{F}^\top \mathbf{W}_3 \mathbf{F} + \mathbf{Q}_2^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{\mathbf{d}}_1 \\ \widehat{\mathbf{d}}_2 \\ \widehat{\mathbf{d}}_3 \end{pmatrix}$$

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

The mse of $(\widehat{\mathbf{m}}_1, \widehat{\mathbf{m}}_2, \widehat{\mathbf{m}}_3)$ is then

$$\mathsf{tr} \left[\begin{pmatrix} \mathbf{F}^\top \mathbf{W}_1 \mathbf{F} + \mathbf{T}^\top \mathbf{Q}_1^{-1} \mathbf{T} & -\mathbf{T}^\top \mathbf{Q}_1^{-1} & \mathbf{0} \\ -\mathbf{Q}_1^{-1} \mathbf{T} & \mathbf{F}^\top \mathbf{W}_2 \mathbf{F} + \mathbf{Q}_1^{-1} + \mathbf{T}^\top \mathbf{Q}_2^{-1} \mathbf{T} & -\mathbf{T}^\top \mathbf{Q}_2^{-1} \\ \mathbf{0} & -\mathbf{Q}_2^{-1} \mathbf{T} & \mathbf{F}^\top \mathbf{W}_3 \mathbf{F} + \mathbf{Q}_2^{-1} \end{pmatrix}^{-1} \right].$$

Now, consider that we wish to find the design $\mathbf{w}_3$ after having already carried out experiments $\mathbf{w}_1$ and $\mathbf{w}_2$. 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 $\mathbf{d}_1, \mathbf{d}_2$, thus we will again introduce the amse and scale the mse by the available estimates of $\mathbf{m}_k$,

$$\begin{pmatrix} \mathsf{diag}(\boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1)) & & \\ & \mathsf{diag}(\boldsymbol{\tau}_2(\widehat{\mathbf{m}}_2)) & \\ & & \mathsf{diag}(\boldsymbol{\tau}_3(\mathbf{T}\widehat{\mathbf{m}}_2)) \end{pmatrix}.$$

Note that since the objective is to design for experiment 3, an estimate for $\mathbf{m}_3$ does not yet exist, thus to approximate $\widehat{\mathbf{m}}_3$ we use the dynamics to propagate $\widehat{\mathbf{m}}_2$ forward one time step.

7

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 $\mathbf{A}_k = -\mathbf{Q}_k^{-1}\mathbf{T}$, the design for the kth experiment minimizes the general amse

$$\mathbf{\Phi}_k(\mathbf{w}_k) = \tag{24}$$

$$\mathrm{tr}\left[\mathrm{diag}\begin{pmatrix} \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \boldsymbol{\tau}_2(\widehat{\mathbf{m}}_2) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix} \begin{pmatrix} \mathbf{F}^\top \mathbf{W}_1 \mathbf{F} + \mathbf{T}^\top \mathbf{Q}_1^{-1}\mathbf{T} & \mathbf{A}_1^\top & 0 & \\ \mathbf{A}_1 & \mathbf{S}_2 & \mathbf{A}_2^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{F}^\top \mathbf{W}_k \mathbf{F} + \mathbf{Q}_{k-1}^{-1} \end{pmatrix}^{-1} \right] \tag{25}$$

The noisy 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 $\mathbf{w}_k$.

$$\min_{\mathbf{w}_k}\left\{ \mathrm{tr}\left[\mathrm{diag}\begin{pmatrix} \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \boldsymbol{\tau}_2(\widehat{\mathbf{m}}_2) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{A}_1^\top & 0 & \\ \mathbf{A}_1 & \mathbf{S}_2 & \mathbf{A}_2^\top & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k(\mathbf{w}_k) \end{pmatrix}^{-1} \right] + \beta \mathbf{e}^\top \mathbf{w}_k \right\}$$

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

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

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.

> probably should discuss that we are not considering $\mathbf{m}_k$ as random variables, but as the mean. then don't have to propagate all the covariance matrices.. Just need to estimate $\mathbf{Q}$'s, or maybe at the beginning for the section. ie. it's not Kalman filtering

The adaptive design method for noisy dynamics is outlined in algorithm 2 below.

Algorithm 2 Adaptive Optimal Design: Noisy dynamics

1: Given a design for experiment 1, $\mathbf{w}_1$ for $\mathbf{m}_1$ ▷ A naive design with no dynamics included
2: **for** $k = 2$ to n **do**   ▷ n time points
3:      $\mathbf{w}_k = \mathsf{argmin} \quad \mathbf{\Phi}_k(\mathbf{w}_k) + \beta \mathbf{e}^\top \mathbf{w}_k$ ▷ design using equation (24)
4:      Given $\mathbf{w}_k$ collect data $\mathbf{d}_k$
5:      $\widehat{\mathbf{m}}_k = \mathsf{argmin} \quad \frac{1}{2}\|\mathbf{F}\mathbf{m}_k - \mathbf{d}_k\|_2^2 + R(\mathbf{m}_k)$ ▷ estimate $\mathbf{m}_k$
6:      $\boldsymbol{\tau}_{k+1}(\mathbf{T}\widehat{\mathbf{m}}_k)$ ▷ Propagate the estimator to compute the monitor function $\boldsymbol{\tau}_{k+1}$
7: **end for**

---

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 steepest descent method to solve the non-linear design optimization problem, with either

4.1 Noiseless dynamics

Finding an optimal design for time step k involves minimizing the objective functional (21). 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 (21) 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 [?].

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

$$\nabla_{\mathbf{w}_k}\phi_k(\mathbf{w}_k) = \mathbf{v}^\top \mathsf{diag}(\boldsymbol{\tau})\,\nabla_{\mathbf{w}_k}\left(\Big(\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G}\Big)^{-1}\mathbf{v}\right). \tag{26}$$

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

$$\mathbf{z} = (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G})^{-1}\mathbf{v} \quad \Leftrightarrow \quad (\mathbf{F}_k^\top \mathbf{W}_k \mathbf{F}_k + \mathbf{G})\mathbf{z} = \mathbf{v}, \tag{27}$$

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 (21), is then

$$- \mathsf{diag}(\mathbf{F}_k \mathbf{z})\mathbf{F}_k \mathbf{y} + \beta \mathbf{e}, \tag{28}$$

where $\mathbf{F}_k = \mathbf{F}\mathbf{T}^{k-1}$. The computation of the Hessian of ϕ is fairly straight forward if one notes that the gradient $\nabla \mathbf{y} = \nabla \mathbf{z}$,

$$\mathbf{H}(\mathbf{w}_k) = -\mathsf{diag}\big(\mathbf{F}_k(\mathbf{z} - \mathbf{y})\big)\mathbf{F}_k \nabla \mathbf{z}. \tag{29}$$

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

Computing the gradient of equation (24) 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,

$$\boldsymbol{\Phi}(\mathbf{w}_k) = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}^{\top} \mathsf{diag} \begin{pmatrix} \boldsymbol{\tau}_1(\widehat{\mathbf{m}}_1) \\ \boldsymbol{\tau}_2(\widehat{\mathbf{m}}_2) \\ \vdots \\ \boldsymbol{\tau}_k(\mathbf{T}\widehat{\mathbf{m}}_{k-1}) \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \mathbf{A}_1^{\top} & 0 & \\ \mathbf{A}_1 & \mathbf{S}_2 & \mathbf{A}_2^{\top} & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{pmatrix}. \qquad (30)$$

To compute the gradient of $\boldsymbol{\Phi}$ we define the vector $(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k)^{\top}$ such that

$$\begin{pmatrix} \mathbf{S}_1 & \mathbf{A}_1^{\top} & 0 & \\ \mathbf{A}_1 & \mathbf{S}_2 & \mathbf{A}_2^{\top} & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix} \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_k \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_k \end{pmatrix} \qquad (31)$$

and note that the gradient of equation (24) requires the gradient of $(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k)^{\top}$. We then implicitly differentiate to obtain $\nabla_{\mathbf{w}_k}(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k)^{\top}$. Recalling that $\mathbf{S}_k = \mathbf{F}^{\top}\mathbf{W}_k\mathbf{F} + \mathbf{Q}_{k-1}^{-1} + \mathbf{T}^{\top}\mathbf{Q}_k^{-1}\mathbf{T} + \alpha\mathbf{L}^{\top}\mathbf{L}$ 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

$$\underbrace{\begin{pmatrix} \mathbf{S}_1 & \mathbf{A}_1^{\top} & 0 & \\ \mathbf{A}_1 & \mathbf{S}_2 & \mathbf{A}_2^{\top} & 0 \\ 0 & \ddots & \ddots & \ddots \\ & 0 & \mathbf{A}_{k-1} & \mathbf{S}_k \end{pmatrix}}_{\mathbf{E}} \begin{pmatrix} \nabla\mathbf{z}_1 \\ \nabla\mathbf{z}_2 \\ \vdots \\ \nabla\mathbf{z}_k \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ -\mathbf{F}^{\top}\mathsf{diag}(\mathbf{F}\mathbf{z}_k) \end{pmatrix}}_{\mathbf{B}} \qquad (32)$$

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 $\mathbf{z}$ is

$$\begin{pmatrix} \nabla\mathbf{z}_1 \\ \nabla\mathbf{z}_2 \\ \vdots \\ \nabla\mathbf{z}_k \end{pmatrix} = \mathbf{E}^{-1}\mathbf{B}, \qquad (33)$$

and thus the gradient of $\boldsymbol{\Phi}$ as

$$\nabla\boldsymbol{\Phi}(\mathbf{w}_k) = \mathbf{B}^{\top}\mathbf{E}^{-\top}(\boldsymbol{v} \odot \bar{\mathbf{v}}) \qquad (34)$$

where $\boldsymbol{v}$ 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 was used to solve the system $\mathbf{E}^{-\top}(\boldsymbol{v} \odot \bar{\mathbf{v}})$ without having to explicitly form the big matrix $\mathbf{E}$.

5   Numerical Examples

To demonstrate the adaptive dynamic design we begin with a toy problem designed to fully demonstrate the ability of the methods 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. 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 $\mathbf{m}$ and tends to cover a large area.

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 $m = [100, 50]$ 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,

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

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 s and $\mathbf{d}$ at the cell centers, see figure 1 below.

The discrete tomography experiment in matrix form is then,

$$\mathbf{F}\mathbf{m}_k = \mathbf{d}_k, \text{ for } k = 1, \ldots, n. \tag{35}$$

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 [8], which characterize the transport of a solute in a fully saturated porous media,

$$\frac{\partial (c\rho)}{\partial t} + \nabla \cdot c\rho\vec{\mathbf{u}} = 0, \tag{36a}$$
$$\text{subject to } c\rho(0, \vec{\mathbf{x}}) = c_0\rho(\vec{\mathbf{x}}),$$

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

$$\nabla \cdot \vec{\mathbf{u}} = q \tag{36b}$$
$$\vec{\mathbf{u}} = \mathbf{K}(\vec{\mathbf{x}})\nabla p \tag{36c}$$
$$p(0, \vec{\mathbf{x}}) = p_0(\vec{\mathbf{x}}), \tag{36d}$$

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 (36a). From now on we will refer only to $\mathbf{m}$.

To discetize the advection of the tracer, a particle in cell discretization was chosen such that equation (36a) 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 is to the 4 neighboring cells 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.
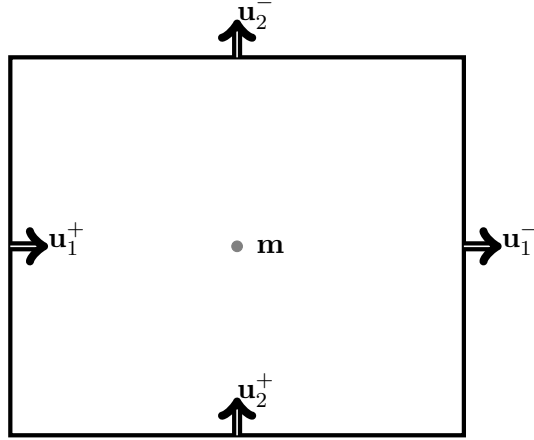

Figure 1: An example of a stagered grid cell with model parameters at the cell center and fluid velocity field on the edges.

The fluid velocity field $\mathbf{u}$ was computed by solving equations (36b) and (36c) with a constant hydraulic conductivity for $1e-03$ kg/s^{2}m^2.

5.3 Regularization

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

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 as the regularization [4] because it is possible that the models can have discontinuities or sharp boundaries. This is particularly valid for our experiment since we have chosen advection without any diffusion term to transport a concentration of tracer. In this ideal case there should be no diffusion of the tracer.

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

$$R(m) = \alpha \int \phi(|\nabla|) \, d\vec{\mathbf{x}}. \tag{37}$$
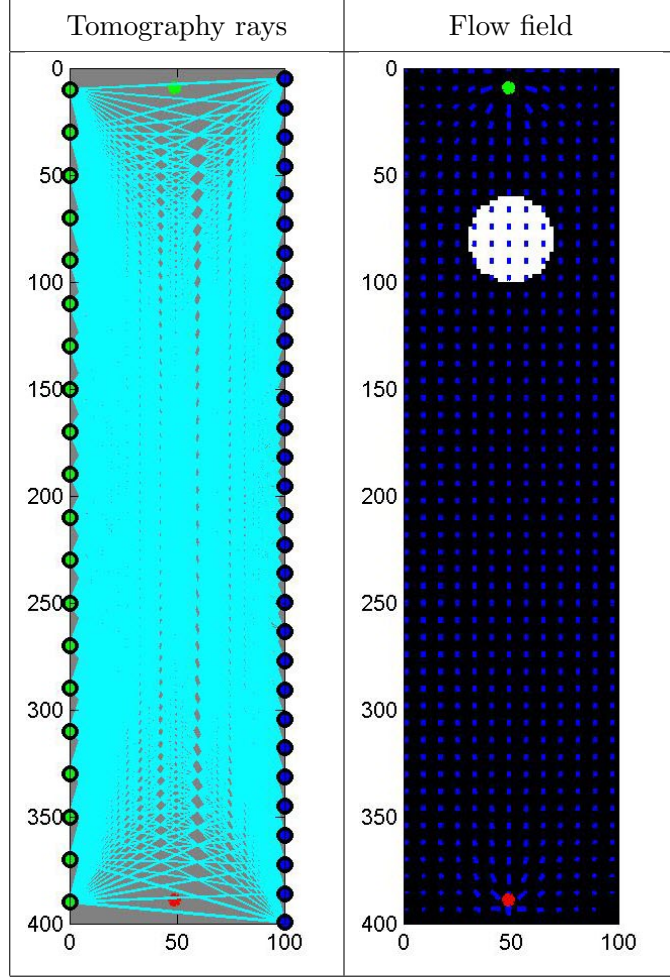
Figure 2: Initial experiment setup. Sources are pictured in green on the left and receivers in blue on. The initial setup covers the entire flow domain with 20 sources and 30 receivers.

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

$$R(\mathbf{m}) = \alpha h^2 \mathbf{e}^\top \sqrt{\mathbf{A}_f^c \left((\mathsf{GRAD}\ \mathbf{m}) \odot (\mathsf{GRAD}\ \mathbf{m}) \right) + \epsilon}, \tag{38}$$

where the GRAD gradient operator again 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 is significantly lower than that of the models recovered using the linear regularization.

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

6   Results: Noiseless dynamics

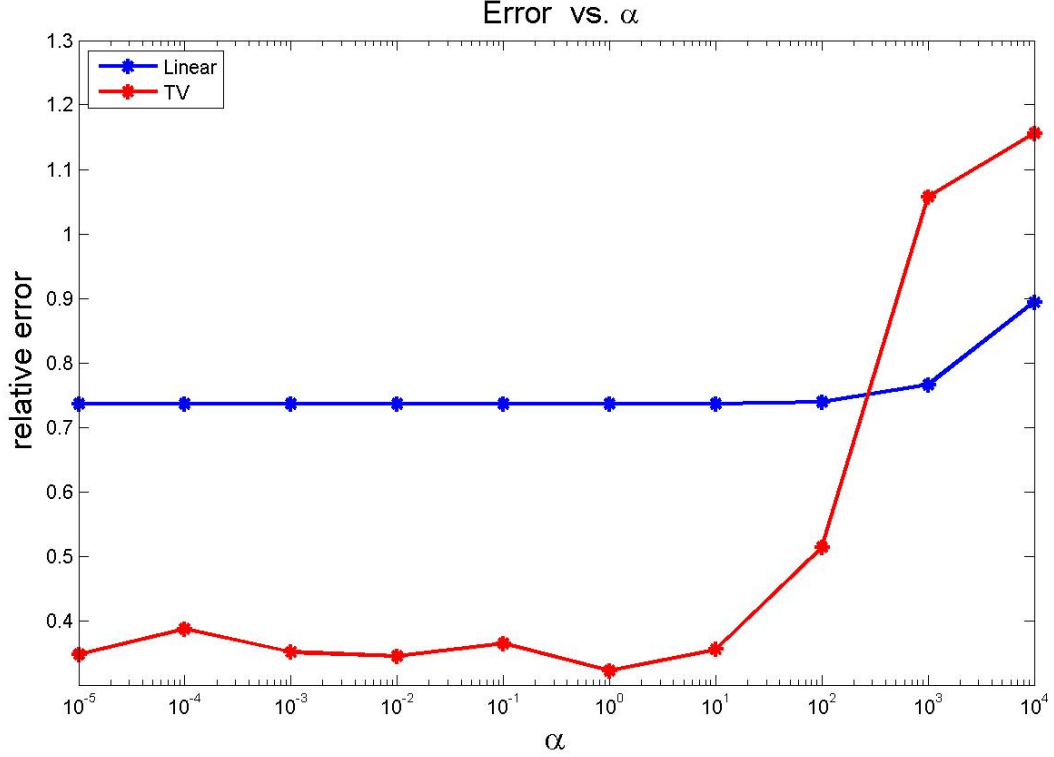To generate data, the initial tracer concentration was marched along in time for time steps of 25 days, and measured by conducting a tomography survey. Gaussian noise was added to each data set. In total 9 experiments were conducted.

The initial design for time point t_0 did not include the dynamics, which amounts to setting $\tau = 1$ in algorithm 1. 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 6.

For each experiment the mse was plotted versus the number of nonzero weights for a set of regularization parameters β, see figure 4 for images 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 mse was kept reasonably small. The experiment weights are also plotted.

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 τ. In figure 6 the rays corresponding to the optimal experimental design data used to estimate the current models are pictured in blue over and image
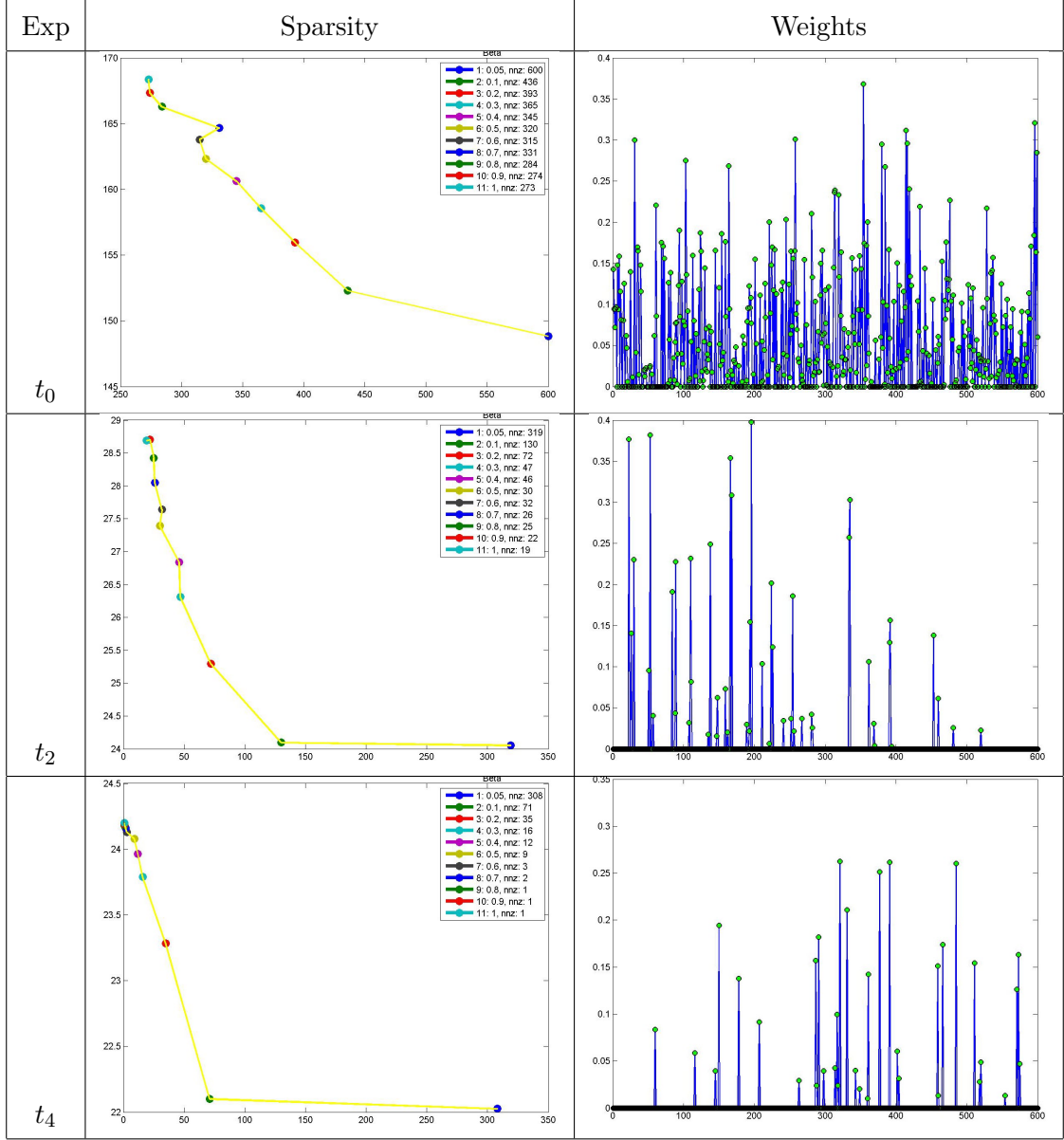
Figure 4: Noiseless dynamics: Plots of the amse vs. the number of nonzero weights (left column), and the weights used to conduct the experiment for experiments 1,3,and 5, at times t_0, t_2, t_4.

of the model. Note that the rays ten to follow the model as it moves through the domain. 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 rays that do not pass through the target, 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 somewhat expected.

It is also apparent in the reconstructions of $\mathbf{m}_k$ that the smoothed total variation does a better

job, even when the optimal design was computed for the linear regularization when estimating the target. Again, this is to be expected since we have chose a model with discontinuous edges, and are transporting the target only by advection.

7 Results: Noisy dynamics

To demonstrate the formulation presented in section 3.2, we again march a the tracer along in time while determining an optimal survey design for the future experiment, given a current data. Models $\mathbf{m}_k$ are reconstructed at each time step again using the reduced data sets, for the non-linear regularization, and again for the linear regularization.

8 Concluding remarks

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 l0 - 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] Zhangxin Chen, Guanren Huan, and Yuanle Ma. *Computational Methods for Multiphase Flows in Porous Media*. SIAM, 2006.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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.

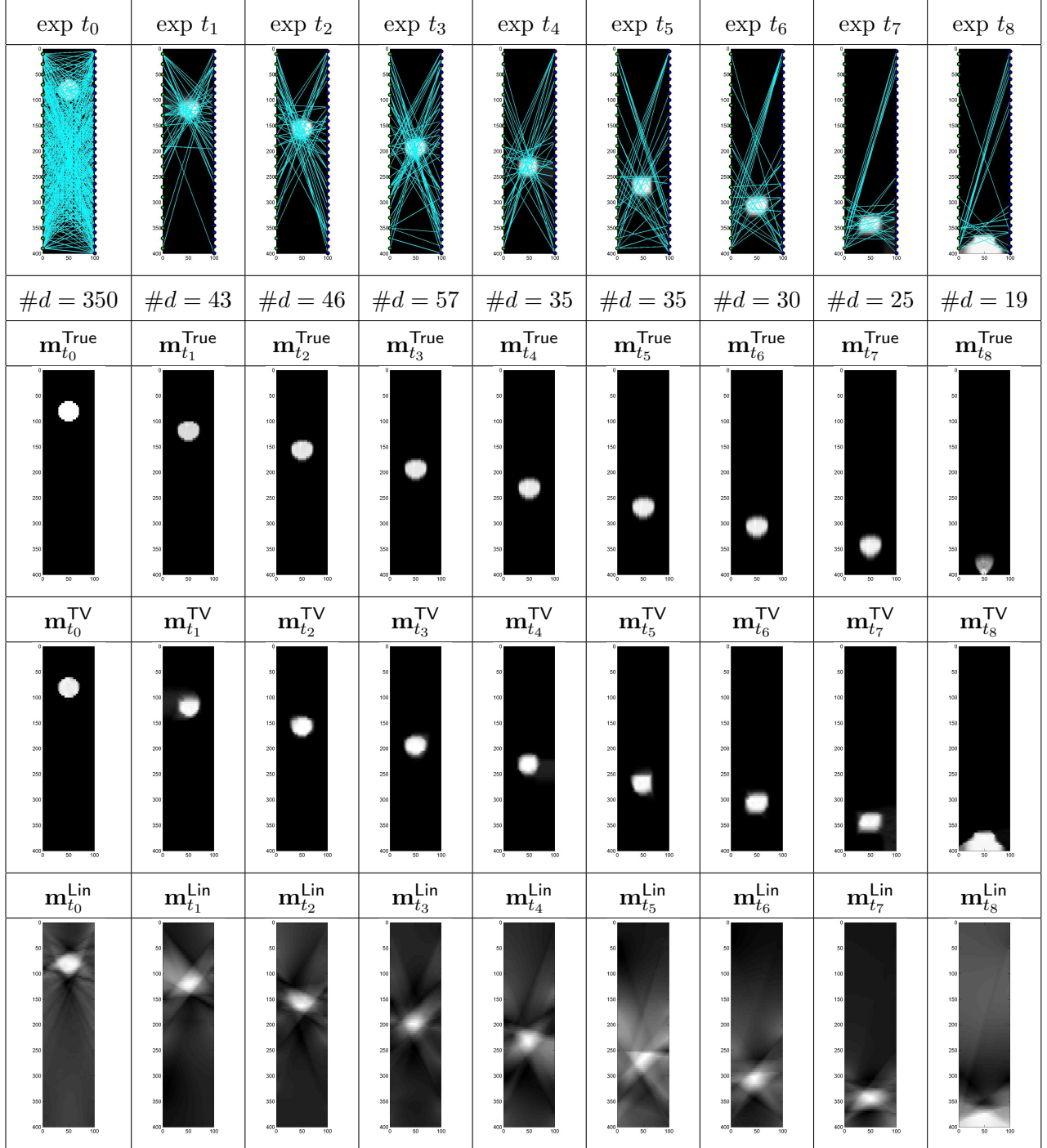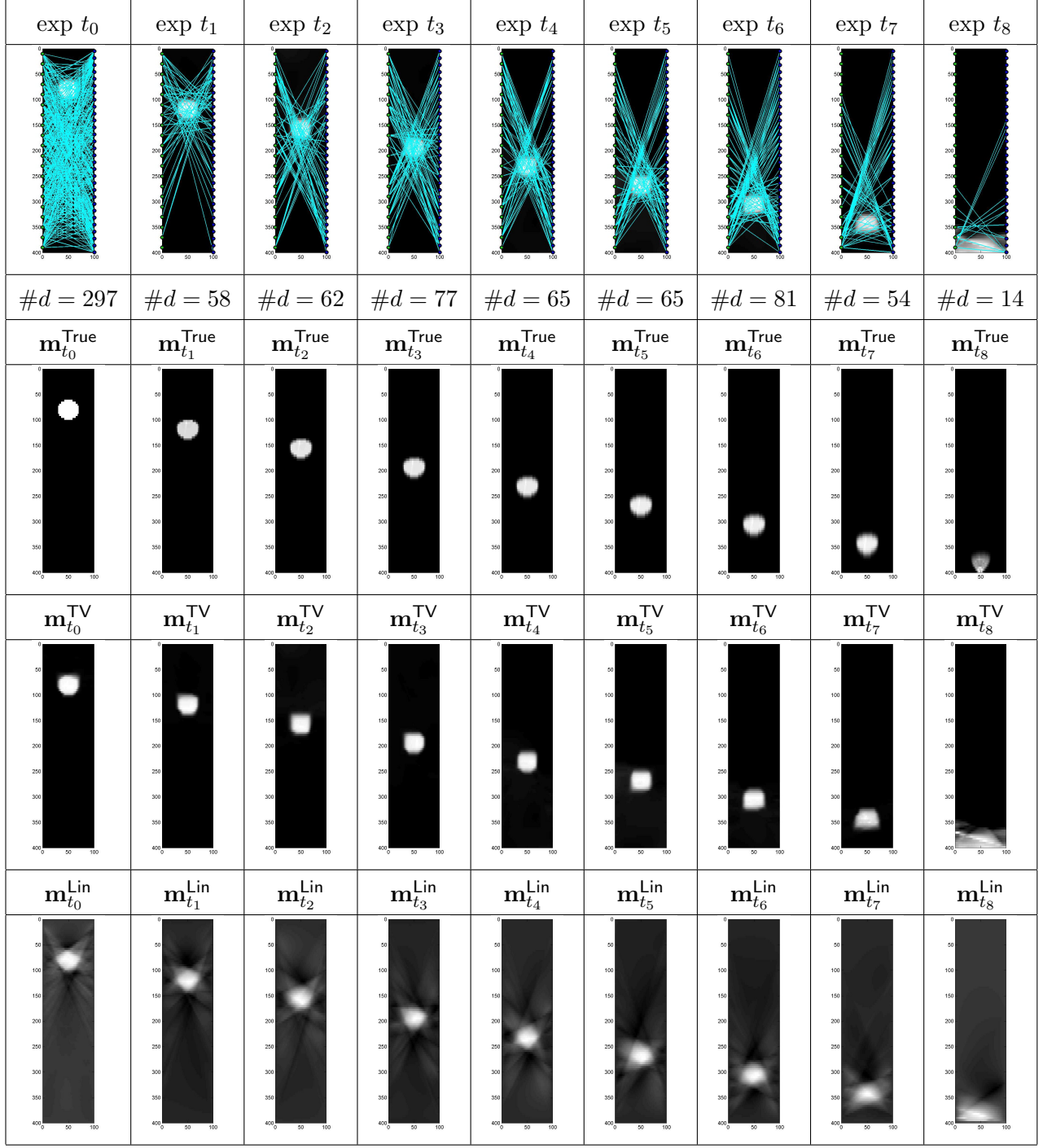| exp t_0	exp t_1	exp t_2	exp t_3	exp t_4	exp t_5	exp t_6	exp t_7	exp t_8
$\#d = 297$	$\#d = 58$	$\#d = 62$	$\#d = 77$	$\#d = 65$	$\#d = 65$	$\#d = 81$	$\#d = 54$	$\#d = 14$
$\mathbf{m}_{t_0}^{\text{True}}$	$\mathbf{m}_{t_1}^{\text{True}}$	$\mathbf{m}_{t_2}^{\text{True}}$	$\mathbf{m}_{t_3}^{\text{True}}$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

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.