In this paper, we consider efficient methods for computing solutions to and estimating uncertainties in dynamic inverse problems, where the parameters of interest may change during the measurement procedure. Compared to static inverse problems, incorporating prior information in both space and time in a Bayesian framework can become computationally intensive, in part, due to the large number of unknown parameters. In these problems, explicit computation of the square root and/or inverse of the prior covariance matrix is not possible, so we consider efficient, iterative, matrix-free methods based on the generalized Golub–Kahan bidiagonalization that allow automatic regularization parameter and variance estimation. We demonstrate that these methods for dynamic inversion can be more flexible than standard methods and develop efficient implementations that can exploit structure in the prior, as well as possible structure in the forward model. Numerical examples from photoacoustic tomography, space-time deblurring, and passive seismic tomography demonstrate the range of applicability and effectiveness of the described approaches. Specifically, in passive seismic tomography, we demonstrate our approach on both synthetic and real data. To demonstrate the scalability of our algorithm, we solve a dynamic inverse problem with approximately 43,000 measurements and 7.8 million unknowns in under 40 s on a standard desktop.
Keywords: dynamic inversion, Bayesian methods, Tikhonov regularization, generalized Golub–Kahan, Matern covariance kernels, tomographic reconstruction

Supplementary material for this article is available online

(Some figures may appear in colour only in the online journal)

1. Introduction

The goal of an inverse problem is to use data, that is collected or measured, to estimate unknown parameters given some assumptions about the forward model [25, 59]. In many applications, the problem is assumed to be static, in the sense that the underlying parameters do not change during the measurement process. However, in many realistic scenarios such as in passive seismic tomography [62, 63] or dynamic electrical impedance tomography [52, 53], the underlying parameters of interest change dynamically. Incorporating prior information regarding temporal smoothness in reconstruction algorithms can lead to better reconstructions. However, this presents a significant computational challenge since many large spatial reconstructions may need to be computed, e.g. at each time step and for many time points. For example, in passive seismic tomography, geophones are used to collect measurements from seismic events (e.g. earthquakes) occurring in 1–2 week intervals over 3–4 months, and the goal is to obtain 3D spatial reconstructions of the elastic properties of the sub-surface for each time interval (e.g. to monitor changing stress conditions). As another application, in medical imaging, during the data acquisition process the reconstruction algorithms need to account for patient motion or non-affine temporal changes, and this can be modeled as a dynamic inverse problem.

Here we consider a discrete dynamic inverse problem with unknowns in space and time where the goal is to reconstruct parameters \( s_i \in \mathbb{R}^{n_s} \) from observations \( d_i \in \mathbb{R}^{m_i} \) for \( i = 1, \ldots, n_t \). Here \( n_s \) refers to the number of spatial grid points, \( n_t \) the number of time points, and \( m = \sum_i m_i \) is the total number of measurements over all time points. For some problems, the number of measurements may correspond to the number of sensors or spatial measurement locations and thus may be the same for all time points. Let

\[
    s = \begin{bmatrix} s_1 \\ \vdots \\ s_{n_t} \end{bmatrix}, \quad \text{and} \quad d = \begin{bmatrix} d_1 \\ \vdots \\ d_{n_t} \end{bmatrix},
\]

then we are interested in the following dynamic problem,

\[
    d = As + \epsilon,
\]

where \( A \in \mathbb{R}^{m \times n_s n_t} \) models the forward process which is assumed linear and \( \epsilon \) represents noise or measurement errors in the data. We assume that \( \epsilon \sim \mathcal{N}(0, R) \), where \( R \) is a positive definite matrix whose inverse and square root are inexpensive (e.g. a diagonal matrix with positive diagonal entries). Given \( A \) and \( d \), the goal of the inverse problem is to reconstruct \( s \). Since these problems are typically ill-posed, regularization is often required to compute a reasonable solution.

To solve the dynamic inverse problem, we adopt a Bayesian approach where the measured data and the parameters to be recovered (here, the space-time unknowns) are modeled as random variables. Additionally we assume that the prior distribution for \( s \) is modeled as
a Gaussian distribution with mean \( \mu = [\mu_1^\top \ldots \mu_n^\top]^\top \) and positive-definite covariance matrix \( Q \). That is, \( s \in \mathcal{N}(\mu, \lambda^{-2} Q) \), where \( \lambda \) is a (yet to be determined) scaling parameter for the precision matrix. For dynamic problems, covariance matrices \( R \) and \( Q \) contain information in both space and time. Then Bayes’ rule is used to combine the likelihood, the prior distribution, and the posterior distribution,

\[
\pi(s \mid d) \propto \pi(d \mid s)\pi(s)
\]

\[
\propto \exp \left( -\frac{1}{2} \| As - d \|^2_{R^{-1}} - \frac{\lambda^2}{2} \| s - \mu \|^2_{Q^{-1}} \right).
\]

where \( \| x \|_M = \sqrt{x^\top M x} \) for any symmetric positive definite matrix \( M \). The maximum a posteriori (MAP) estimate, which is the peak of the posterior distribution, can be obtained by minimizing the negative log likelihood of (4), i.e.

\[
s(\lambda) = \arg \min_{s \in \mathbb{R}^{nsnt}} \frac{1}{2} \| As - d \|^2_{R^{-1}} + \frac{\lambda^2}{2} \| s - \mu \|^2_{Q^{-1}}.
\]

Notice that for dynamic inverse problems, computing the MAP estimate requires solving for \( nsnt \) unknowns. The main challenge here is that for the applications under consideration, \( n_1 \) is typically \( O(10^5) - O(10^6) \) and \( n_t \) is typically \( O(10^2) - O(10^3) \). The resulting prior covariance matrices have \((n_1n_t)^2\) or \(O(10^{18})\) entries. Storing such covariance matrices is infeasible, much less performing computations with them. Clearly there is a need for developing specialized numerical methods for tackling the immense computational challenges arising from dynamic inverse problems. Our strategy is to use a combination of highly structured representations of prior information along with efficient numerical methods that can exploit these representations.

1.1. Overview of main contributions

In this work, we adopt a Bayesian framework for solving dynamic inverse problems. We derive two efficient methods for computing MAP estimates, where the distinguishing features of our approach compared to previous methods are that we can incorporate a wide class of spatiotemporal priors, include time-dependent observation operators, and enable automatic regularization parameter selection. The resulting solvers are highly efficient and scalable to large problem sizes. In addition to the MAP estimate, we develop an efficient representation of the posterior covariance matrix using the generalized Golub–Kahan (gen-GK) bidiagonalization. This low-rank approximation can be used for uncertainty quantification, by estimating the variance of the distribution. Using several real-world imaging applications (including both simulated and real data), we show that our methods are well suited for a wide class of dynamic inverse problems, and we demonstrate scalability of our algorithms.

1.2. Related work

The literature on dynamic inverse problems is large, and it is not our intention to provide a detailed overview. We mention a few related approaches that are relevant to our work.

A popular approach for space-time reconstructions uses Kalman filters and smoothers. However, textbook implementations of these methods can be prohibitively expensive. This is
because they require the storage and computation of covariance matrices that scale as $O(n^2)$. One approach is to use an efficient representation of the state covariance matrix, as a low-rank perturbation of an appropriately chosen matrix \cite{39, 41}. Efficient computational techniques for the Kalman filter especially tailored to the random-walk forecast model were proposed by \cite{32, 50}. More details on how this forecast model fits within our framework can be found in appendix B.

Significant simplifications can be made if we assume that measurement errors are independent in time and reconstruct the parameters of interest only using the data available from the current time step. However, several other authors, see for example Schmitt and collaborators \cite{52, 53}, have emphasized the importance of including temporal priors in many practical applications. They considered a total-variation type temporal smoothness prior with a simple spatial prior (the identity matrix) and showed that their approach achieved superior results in faster computational time than other statistical approaches such as Kalman smoothers. Furthermore, various works on dynamic evolutionary algorithms have been developed for dynamic electrical impedance tomography \cite{42, 51}.

Our approach is more general in that we allow for a variety of spatial priors where the resulting covariance matrices are dense, unwieldy, and only available via matrix-vector multiplication. Furthermore, we consider more general forward models and consider hybrid iterative approaches so that the regularization parameter $\lambda$ can be automatically estimated. Selecting regularization parameters can be a delicate (yet very important) task in both static and dynamic settings. In \cite{52}, the regularization (or precision) parameter decoupled in space and time, and the resulting two parameters were required algorithmic inputs.

In section 2, we describe the problem set-up and address various changes of variables that can be used. We also provide a brief overview of generalized hybrid methods to efficiently solve static inverse problems. Efficient methods for approximating the MAP estimate, i.e. solving (5), in the space-time formulation will be described in section 3, where special cases of problem structure will be considered for efficiency. Efficient variance estimation methods based on the gen-GK bidiagonalization are described in section 4. Numerical results are presented in section 5 for various simulated imaging problems and for real data from passive seismic tomography. Conclusions and discussions are provided in section 6.

2. Problem set up and background

One main goal in the Bayesian framework is to efficiently compute the MAP estimate, and in this paper, we are mainly interested in cases where $Q$ is a very large, dense matrix, so computing $Q$ and $Q^{-1}$ or their factorizations is not feasible. Such scenarios arise, for example, when working with Gaussian random fields, in which case forming $Q$ explicitly may not be feasible, but computing matrix-vector products (MVPs) with $Q$ can be done efficiently \cite{49}.

First, we describe various problem formulations for computing the MAP estimate and describe a change of variables so that gen-GK methods can be used. Then, a brief overview of generalized hybrid methods is provided in section 2.2 for completeness, and a discussion on various choices for modeling temporal priors is provided in section 2.3.

2.1. Problem formulations

Notice that by setting the derivative of the objective function in (5) to zero, the desired MAP estimate $s(\lambda)$ is the solution to the system of equations.
\( (A^T R^{-1} A + \lambda^2 Q^{-1}) s = A^T R^{-1} d + \lambda^2 Q^{-1} \mu. \)  

For problems where matrix factorizations \( Q^{-1} = L_Q^T L_Q \) and \( R^{-1} = L_R^T L_R \) are possible, a common approach to compute the MAP estimate is to solve the following *general-form Tikhonov problem*,

\[
\min_{x \in \mathbb{R}^{nsnt}} \frac{1}{2} \| \text{L}_R (A s - d) \|_2^2 + \frac{\lambda^2}{2} \| \text{L}_Q (s - \mu) \|_2^2.
\]  

In this case, the solution can be computed as \( s = \mu + L_Q^{-1} x \) where \( x \) solves the *priorconditioned problem* [6, 7],

\[
\min_{x \in \mathbb{R}^{nsnt}} \frac{1}{2} \| \text{L}_R (A L_Q^{-1} x - b) \|_2^2 + \frac{\lambda^2}{2} \| x \|_2^2.
\]

In the applications that we consider, the covariance matrices can be very large and dense, so the storage and computational costs to obtain factorizations and/or inverses of \( Q \) can be prohibitive. In order to avoid matrix factorizations of \( Q \) and/or expensive linear solves with \( Q \), a different change of variables was proposed in [10], where

\[
x \leftarrow Q^{-1}(s - \mu), \quad b \leftarrow d - A \mu,
\]

so that (6) reduces to the modified system of equations

\[
(A^T R^{-1} A Q + \lambda^2 I) x = A^T R^{-1} b.
\]

In summary, with this change of variables, the MAP estimate is given by \( s(\lambda) = \mu + Q x(\lambda) \), where \( x(\lambda) \) is the solution to the following optimization problem

\[
\min_{x \in \mathbb{R}^{nsnt}} \frac{1}{2} \| A Q x - b \|_2^2 + \frac{\lambda^2}{2} \| x \|_2^2.
\]

Hybrid iterative methods for approximating \( x(\lambda) \) were described in [10] for generic \( A, R, \) and \( Q \), but here we focus on their applications and extensions to dynamic problems. For completeness, we give a brief description of these methods in section 2.2 and refer the interested reader to [10] or appendix A for further details.

In this paper, we choose to focus on iterative methods; however, we briefly mention an alternative formulation. For relatively small problems where the number of overall measurements \( m \) is small, i.e. \( O(10^3) \sim O(10^4) \), a simple change of variables along with the Sherman–Morrison formula can be used to avoid \( Q^{-1} \). The MAP estimate can be computed as \( s(\lambda) = \mu + Q A^T \xi(\lambda) \) where

\[
(A Q A^T + \lambda^2 R) \xi(\lambda) = d - A \mu.
\]

Notice that in terms of solving linear systems, the number of unknowns has reduced from \( n_s n_t \) to \( m \). A direct solver could be used to solve (12), but forming \( QA^T \) may be computationally prohibitive, costing \( O(n_s n_t m^2) \). Iterative methods could be used, but in these cases, it may be difficult to know a good regularization parameter *a priori*. Further simplifications would be possible for problems where \( m_i \) is constant for all \( i = 1, \ldots, n_t \) and \( A \) is also a Kronecker product, i.e. \( A = A_t \otimes A_s \), so that

\[
Q A^T = (Q_t \otimes Q_s) (A^T_t \otimes A^T_s) = Q_s A^T_t \otimes Q_t A^T_s
\]

and

\[
Q A^T = (A_t, Q_s, A^T_t) \otimes (A_s, Q_t, A^T_s).
\]
As mentioned earlier, we do not pursue this approach since it is computationally expensive when the number of measurements is large.

2.2. Generalized hybrid methods

The basic idea behind the generalized hybrid methods is first to generate a basis \( \mathbf{V}_k \) for the Krylov subspace

\[
\mathcal{S}_k \equiv \text{Span}\{\mathbf{V}_k\} = \mathcal{K}_k(\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}, \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b})
\]

(15)

where \( \mathcal{K}_k(\mathbf{M}, \mathbf{d}) = \text{Span}\{\mathbf{d}, \mathbf{M}\mathbf{d}, \ldots, \mathbf{M}^{k-1}\mathbf{d}\} \), and second to solve (11) in this subspace. A basis for \( \mathcal{S}_k \) can be generated using the gen-GK bidiagonalization process\(^5\) summarized in algorithm 1. We initialize \( \beta_1 = \| \mathbf{b} \|_{\mathbf{R}^{-1}} \) and \( \mathbf{u}_1 = \mathbf{b} / \beta_1 \), and at the end of \( k \) steps, we have the matrices

\[
\begin{align*}
\mathbf{U}_{k+1} & \equiv [\mathbf{u}_1, \ldots, \mathbf{u}_{k+1}], \\
\mathbf{V}_k & \equiv [\mathbf{v}_1, \ldots, \mathbf{v}_k], \quad \text{and} \quad \mathbf{B}_k \equiv \\
& \begin{bmatrix}
\alpha_1 \\
\beta_2 \\
\vdots \\
\alpha_k \\
\beta_{k+1}
\end{bmatrix}
\end{align*}
\]

(16)

that in exact arithmetic satisfy

\[
\mathbf{A}\mathbf{Q}\mathbf{V}_k = \mathbf{U}_{k+1}\mathbf{B}_k, \quad \mathbf{U}_{k+1}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{I}_{k+1}, \quad \text{and} \quad \mathbf{V}_k^\top \mathbf{Q}\mathbf{V}_k = \mathbf{I}_k.
\]

(17)

Then, we seek approximate solutions to (11) of the form \( \mathbf{x}_k = \mathbf{V}_k \mathbf{z}_k \), so that \( \mathbf{x}_k \in \mathcal{S}_k \), where the coefficients \( \mathbf{z}_k \) can be determined by solving the following problem

\[
\min_{\mathbf{z}_k \in \mathcal{S}_k} \frac{1}{2} \| \mathbf{A}\mathbf{Q}\mathbf{x}_k - \mathbf{b} \|_{\mathbf{R}^{-1}}^2 + \frac{\lambda^2}{2} \| \mathbf{x}_k \|_{\mathbf{Q}}^2 \\
\quad \Leftrightarrow \min_{\mathbf{z}_k \in \mathbb{R}^L} \frac{1}{2} \| \mathbf{B}_k \mathbf{z}_k - \beta_1 \mathbf{e}_1 \|_2^2 + \frac{\lambda^2}{2} \| \mathbf{z}_k \|_2^2,
\]

(18)

where the equivalency uses the relations in (17). For fixed \( \lambda \), an approximate MAP estimate can be recovered by undoing the change of variables,

\[
\mathbf{s}_k = \mu + \mathbf{Q}\mathbf{x}_k = \mu + \mathbf{Q}\mathbf{V}_k(\mathbf{B}_k^\top \mathbf{B}_k + \lambda^2 \mathbf{I})^{-1} \mathbf{B}_k^\top \beta_1 \mathbf{e}_1.
\]

(19)

\(^5\)Generalized Golub–Kahan methods were first proposed by Benbow [3] for generalized least squares problems, and used in several applications, see e.g. [1, 2, 38]. However, the specific form of the bidiagonalization was developed in [10].
where, now, $\mathbf{s}_k \in \mu + \mathbf{Q}\mathbf{S}_k$. For $\lambda = 0$, we call this approach genLSQR. The choice of a good regularization parameter $\lambda$ is important for the quality of the solution. If $\lambda$ is not known a priori, hybrid methods can take advantage of the shift-invariance property of Krylov subspaces and select $\lambda$ adaptively and automatically by utilizing well-known SVD-based regularization parameter selection schemes [23, 25, 59] for the projected problem (18), since $\mathbf{B}_k$ is only of size $(k + 1) \times k$. In particular, we choose the regularization parameter based on weighted generalized cross validation approach (see appendix A for details). Henceforth, we refer to this approach as genHyBR. As a benchmark for simulated experiments, we also consider the optimal regularization parameter $\lambda_{\text{opt}}$, which minimizes the 2-norm of the error between the reconstruction and the truth.

We comment on the computational cost and convergence of the genLSQR and genHyBR algorithms. In addition to MVPs with $\mathbf{A}$ and $\mathbf{A}^\top$ that are required for the standard GK bidiagonalization [19], each iteration of gen-GK bidiagonalization requires two MVPs with $\mathbf{Q}$ and two solves with $\mathbf{R}$ (which are assumed to be cheap); in particular, we emphasize that algorithm 1 avoids $\mathbf{Q}^{-1}$ and $\mathbf{L}_\mathbf{Q}$, due to the change of variables in (9). An additional computational cost of $\mathcal{O}(k^3)$ is required at each iteration $k$ for computing the coefficients $\mathbf{z}_k$ and the regularization parameter. For fixed $\lambda$ and in exact arithmetic, iterates $\mathbf{s}_k$ are mathematically equivalent to both conjugate gradient with weighted inner products and LSQR on a priori-conditioned problem, see [10].

**Algorithm 1.** Generalized Golub–Kahan (gen-GK) bidiagonalization.

| Require: Matrices $\mathbf{A}$, $\mathbf{R}$ and $\mathbf{Q}$, and vector $\mathbf{b}$. |
|---|
| 1: $\beta_1 \mathbf{u}_1 = \mathbf{b}$, where $\beta_1 = \|\mathbf{b}\|_{\mathbf{R}^{-1}}$. |
| 2: $\alpha_1 \mathbf{v}_1 = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_1$. |
| 3: for $i = 1, \ldots, k$ do |
| 4: $\beta_{i+1} \mathbf{u}_{i+1} = \mathbf{A} \mathbf{Q} \mathbf{v}_i - \alpha_i \mathbf{u}_i$, where $\beta_{i+1} = \|\mathbf{A} \mathbf{Q} \mathbf{v}_i - \alpha_i \mathbf{u}_i\|_{\mathbf{R}^{-1}}$. |
| 5: $\alpha_{i+1} \mathbf{v}_{i+1} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_{i+1} - \beta_{i+1} \mathbf{v}_i$, where $\alpha_{i+1} = \|\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_{i+1} - \beta_{i+1} \mathbf{v}_i\|_{\mathbf{Q}}$ |
| 6: end for |
| 7: return $\mathbf{U}_{k+1}, \mathbf{V}_{k+1} = [\mathbf{Y}_k, \mathbf{v}_{k+1}], \text{ and } \mathbf{B}_k$ |

2.3. **Modeling prior covariances**

Following the geostatistical approach, we model the unknown field $s(\mathbf{p}, t)$ as a realization of a spatio-temporal random function $Z(\mathbf{p}, t)$ for $\mathbf{p} \in \mathbb{R}^d$ and $t \in \mathbb{R}$. We assume that the covariance function is stationary in space and stationary in time; for simplicity here, we also assume that the mean is zero. In other words, we assume that the covariance function satisfies

$$\text{cov}\{Z(\mathbf{p}_1, t_1), Z(\mathbf{p}_2, t_2)\} = C(\mathbf{p}_1 - \mathbf{p}_2, t_1 - t_2),$$

(20)

where $C : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_+$ is a positive definite covariance function.

We briefly review various formulations for space-time covariance kernels, before delving into our specific choices of kernels. Perhaps the most convenient representation can be obtained if we assume that the covariance function is separable in space and time, and isotropic in these variables. Then $C$ takes the form

$$C(\mathbf{p}, t) = C_S(\|\mathbf{p}\|) C_T(|t|) \quad \forall (\mathbf{p}, t) \in \mathbb{R}^d \times \mathbb{R},$$

(21)

where $C_S(\|\mathbf{p}\|)$ and $C_T(|t|)$ are isotropic, purely spatial and purely temporal covariance functions, respectively, and the resulting matrices $\mathbf{Q}$ have the Kronecker product structure. The separability assumption is common in the statistics literature [17, 30], and tests for separability
can be found in [15]. The Kronecker product structure has computational advantages which we will exploit in section 3 to develop efficient algorithms. While mathematically and computationally convenient, it is important to recognize the potential shortcomings of the separability assumption. The key issue is that the prior models do not allow for interactions in variability of space and time; for a detailed discussion, see [30]. However, even if the covariance kernel is not separable, one may approximate it using a separable covariance kernel [17], where the resulting covariance matrix approximation can be represented as a Kronecker product or a sum of Kronecker products. Such approximations were studied in [57] and have been shown to be successful in the context of image deblurring, e.g. [11, 26, 36].

Many nonseparable covariance kernels have been proposed that model space-time interactions of variability. One such approach uses $C(p, t) = \varphi(\sqrt{c_1 |p|^2 + c_2 |t|^2})$, where $c_1, c_2$ are weights that control the correlation of the space and time variables, and $\varphi(\cdot)$ is an appropriate covariance function. Another approach is to use a product-sum model

$$C(p, t) = a_0 C_0(p) C_T(t) + a_1 C_1(p) + a_2 C_T(t) \quad \forall (p, t) \in \mathbb{R}^d \times \mathbb{R}$$

(22)

where $a_0, a_1$ and $a_2$ are non-negative coefficients and $C_0, C_1$ and $C_T$ are isotropic, purely spatial and purely temporal covariance functions, respectively. A review of these covariance kernels is provided in [18].

A wide variety of choices for $Q$ can be included in our framework and thus incorporated in the methods described below. Next we give a few examples of temporal and spatial priors that are well-suited for our problems.

2.3.1 Specific choices of covariance kernels. A common approach is to use Gaussian random fields where the entries of the covariance matrix are computed directly as $\langle Q_i \rangle_{ij} = \kappa(|t_i - t_j|)$, where $\{t_i\}_{i=1}^n$ are the time points. A popular choice for $\kappa(\cdot)$ is from the Matérn family of covariance kernels [43], which form an isotropic, stationary, positive-definite class of covariance kernels. We define the covariance kernel in the Matérn class as

$$C_{\nu, \ell}(r) = \frac{1}{2^{\nu-1} \Gamma(\nu)} \left( \frac{r \sqrt{2 \nu}}{\ell} \right)^\nu K_\nu \left( \frac{r \sqrt{2 \nu}}{\ell} \right)$$

(23)

where $\Gamma$ is the Gamma function, $K_\nu(\cdot)$ is the modified Bessel function of the second kind of order $\nu$, and $\ell$ is a scaling factor. The choice of parameter $\nu$ in equation (23) defines a special form for the covariance. For example, when $\nu = 1/2$, $C_{\nu, \ell}$ corresponds to the exponential covariance function, and if $\nu = 1/2 + p$ where $p$ is a non-negative integer, $C_{\nu, \ell}$ is the product of an exponential covariance and a polynomial of order $p$. Also, in the limit as $\nu \to \infty$, $C_{\nu, \ell}$ converges to the Gaussian covariance kernel, for an appropriate scaling of $\ell$. Another related family of covariance kernels is the $\gamma$-exponential function [43],

$$\kappa(r, \gamma, \ell) = \exp \left(-\frac{r}{\ell \gamma}\right) \quad 0 < \gamma \leq 2.$$  

(24)

Just as $C_T(\cdot)$ can be a $\gamma$-exponential function, or chosen from the Matérn class, $C_S(\cdot)$ can be chosen in the same way. Therefore, $Q_s$ has entries $\langle Q_s \rangle_{ij} = \kappa(|p_i - p_j|)$, where $\{p_i\}_{i=1}^n$ are the spatial locations. In the applications of interest, the number of spatial locations $n_s$ is much larger than the number of time points $n_t$; thus the storage of $Q_s$ is challenging, as is employing it in iterative methods, since the cost of an MVP is $O(n_s^2)$. Both the storage and computational cost can be reduced to $O(n_s \log n_s)$ using the FFT based approach or $H$-matrix approach. This has been reviewed in [47, 48].
3. Generalized hybrid methods for dynamic inverse problems

In this section, we describe various approaches based on the gen-GK bidiagonalization for computing MAP estimates for dynamic inverse problems. These iterative approaches are desirable for problems where both $A$ and $Q$ may be extremely large, or for problems where these matrices are not explicitly stored but can be accessed via function calls to compute MVPs with $A, A^\top$, and $Q$ efficiently. In section 3.1, we describe an ‘all-at-once’ generalized hybrid method that requires MVPs with covariance matrix $Q$. This requirement is quite general and includes matrices such as $Q$ being a Kronecker product, a sum of Kronecker products, or a convolution operator. Then for problems where the number of time points is small and $A$ and $R$ are also Kronecker products, we describe an efficient decoupled approach in Section 3.2.

3.1. Simultaneous generalized hybrid approach

The first approach uses genHyBR as summarized in section 2.2 to solve for all unknown variables (e.g. in space and time) simultaneously. Since the number of unknowns can be quite large in the ‘all-at-once’ approach and the gen-GK vectors $V_k$ must be stored for hybrid methods, we assume that solutions can be captured in relatively few iterations or that appropriate preconditioning can be used so that $k$ remains small. Next we describe efficiencies that can be gained for problems where $Q$ is a Kronecker product, but we reiterate that the simultaneous approach has applicability beyond the cases presented here.

For problems where $Q$ is a Kronecker product [31], MVPs with $Q$ can be computed efficiently as

$$Qx = (Q_s \otimes Q_t)x = \text{vec}(Q_sXX^\top),$$

where vec and mat are operations such that vec unfolds a matrix $X = R_n \times n_t$ into a vector by stacking column-wise and mat folds it back, i.e.

$$\text{vec}(X) \in R^{m \times 1} \quad \text{mat}\left[\text{vec}(X)\right] = X.$$

Assuming the cost of an MVP with $Q_s$ is $O(n_s \log n_s)$, the cost of $Qx$ is $O(n_t n_s \log n_s + n_sn_t^2)$, which is significantly smaller than the naive cost of $O(n_t^2n_s^2)$. Further reductions in computational cost can be achieved by parallelizing the matrix-matrix multiplications, e.g. MVPs of $Q_s$ with the columns of $X$.

3.1.1. Separable forward operator. If, additionally, the number of measurements at each time step is the same, which we denote by $\bar{m}$, and $A = A_s \otimes A_t$, and $R = R_t \otimes R_s$, where $R_t \in R_{n_t \times n_t}$ and $R_s \in R_{\bar{m} \times \bar{m}}$ are the temporal and spatial noise covariance matrices respectively, then MVPs required for the gen-GK bidiagonalization algorithm can be computed efficiently. That is, for vectors $x \in R^{n_t \times n_s}$ and $y \in R^{\bar{m} \times n_t}$,

$$AQx = \text{vec}(A_sQ_sXX^\top A_t^\top), \quad \text{and} \quad A^\top R^{-1}y = \text{vec}(A_s^\top R_t^{-1}YR_s^{-\top}A_t),$$

where $X \equiv \text{mat}(x) \in R^{n_t \times n_s}$ and $Y \equiv \text{mat}(y) \in R^{\bar{m} \times n_t}$.

3.2. Decoupled generalized hybrid approach

For problems where $A, R$, and $Q$ are all Kronecker products and $n_t$ is relatively small such that $Q_t$ and its factor $L_t$ are feasible, we describe a decoupled genHyBR approach. The normal equations corresponding to the generalized least-squares problem in (11) can be written as
\[
(Q^\top A^\top R^{-1}AQ + \lambda^2 Q)x = Q^\top A^\top R^{-1}b.
\]  
We can exploit the fact that \( Q = L_l^\top L_l \otimes Q_r \), and let \( y = (L_r \otimes I)x \), alternatively \( Y = XL_l^\top \). We introduce the variable \( H = A^\top R^{-1}A \); similarly, we also define \( H_r = A_r^\top R_r^{-1}A_r \) and \( H_l = A_l^\top R_l^{-1}A_l \). The normal equations simplify, when we left-multiply by \( 8.5838 \). By expanding the above expression columnwise, the key observation is that all the equations decouple, which can be done using the genHyBR method. Note that a transformation back of variables is needed, and \( \text{opt} \) be its singular value decomposition, then

\[
[Q_r \otimes Q_r] H (Q_l^\top \otimes Q_l) + \lambda^2 [I \otimes Q_r] y = (L_r \otimes Q_r) A^\top R^{-1} b.
\]

Using the properties of Kronecker products, this equation can alternatively be written as the following generalized-Sylvester equation

\[
Q_l H_l Q_r Y_l H_r L_l^\top + \lambda^2 Q_r Y_r = C \hat{A}_r,
\]

where, for simplicity, we introduce \( C \equiv Q_l A_l^\top R_l^{-1} B R_l^{-1/2} \) and \( \hat{A}_r \equiv R_r^{-1/2} A_r L_r^\top \).

Let \( \hat{A}_r = U_r \Sigma_r V_r^\top \) be its singular value decomposition, then \( L_r H_r L_l^\top = V_r \Sigma_r^2 V_l^\top \). We make another change of variables \( Z \equiv YV_r = X L_l^\top V_r \), and we get

\[
Q_l H_l Q_r \Sigma_r^2 + \lambda^2 Q_r Z = C U_r \Sigma_r,
\]

where we have multiplied on the right by \( V_r \). By expanding the above expression columnwise, the key observation is that all the equations decouple,

\[
(\sigma_i^2 Q_l H_l Q_r + \lambda^2 Q_r) z_i = \sigma_i C u_{i,j} \quad i = 1, \ldots, n_l.
\]

Notice that for \( \sigma_i = 0 \), \( z_i = 0 \), whereas for \( \sigma_i > 0 \), the solution can be obtained by solving the least-squares problem,

\[
\min_{\kappa \in \mathbb{R}^n} \frac{1}{2} \| \sigma_i A_i Q_r z_i - B R_l^{-1/2} u_{i,j} \|^2_{R_l^{-1}} + \frac{\lambda^2}{2} \| z_i \|^2_{Q_r},
\]

which can be done using the genHyBR method. Note that a transformation back of variables must be made. This is summarized in algorithm 2.

\begin{algorithm}
\caption{Decoupled genHyBR.}
\begin{algorithmic}[1]
\Require \( A, A, R, R, L_r, B \)
1: Compute the SVD of \( \hat{A}_r = R_r^{-1/2} A_r L_r^\top = U_r \Sigma_r V_r^\top \)
2: for \( i = 1, \ldots, n \) do
3: \quad Apply genHyBR to (33) to obtain \( z_i \).
4: end for
5: Form \( Z = [z_1 \ldots z_n] \). \( X = Z V_l^\top L_l^\top \) and \( S = Q_r X Q_l^\top \).
\end{algorithmic}
\end{algorithm
This approach is embarrassingly parallel, which may lead to enormous computational savings. Additionally, solving the sequence of decoupled systems may be done efficiently either by recycling of Krylov subspaces [40] or by effective preconditioning [46].

4. Estimation of posterior variances using gen-GK

For the problem considered here, the posterior distribution $s|d$ is Gaussian with

$$s|d \sim \mathcal{N}(\Gamma_{\text{post}} A^T R^{-1} b, \Gamma_{\text{post}})$$

where $\Gamma_{\text{post}} \equiv (\lambda^2 Q^{-1} + H)^{-1}$ (34)

and $H = A^T R^{-1} A$. The MAP estimate corresponds to the mode of the posterior distribution and provides information regarding the ‘most likely’ estimate. On the other hand, the posterior variances defined as the diagonals of the posterior covariance $\Gamma_{\text{post}}$ provide a measure of the spread of the posterior distribution around the posterior mean. For static inverse problems, estimating the posterior covariance matrix is known to be computationally challenging [48, 49]. For dynamic problems, the problem is further exacerbated since the posterior covariance matrix is of size $n_s n_t \times n_s n_t$. Moreover, this matrix is dense and forming it explicitly to obtain the diagonal entries is computationally infeasible. In this work, we exploit intermediate information from genHyBR, which is used to compute a MAP estimate, to estimate the posterior variances. Following section 3, we consider a simultaneous and a decoupled approach.

Before we explain how to compute variances, we make the following remark. In the sequel, we will use algorithm 1 with one minor modification, namely, at each step we explicitly re-orthogonalize the vectors in $U_k$ and $V_k$. Numerical experience suggests that this marginally increases the computational cost by $O(k^2 (m + n))$ but considerably improves the accuracy of the variance computations.

4.1. Simultaneous approach

Recall that after $k$ iterations of the gen-GK bidiagonalization process, we have matrices $B_k, U_{k+1}$ and $V_k$ satisfying relations in (17). Let $B_k^T B_k = W_k \Theta_k W_k^T$ be the eigenvalue decomposition with eigenvalues $\theta_1, \ldots, \theta_k$ and let $Z_k = Q V_k W_k$, then we get the following low-rank approximation

$$QHQ \approx Q V_k B_k^T B_k V_k^T Q = Z_k \Theta_k Z_k^T.$$  (35)

Using (35) and the Woodbury formula, we obtain the approximation

$$\Gamma_{\text{post}} \approx Q (\lambda^2 Q + Z_k \Theta_k Z_k^T)^{-1} Q$$

$$= \lambda^{-2} Q - \lambda^{-2} Z_k (I_k + \lambda^2 \Theta_k^{-1})^{-1} Z_k^T$$

$$= \lambda^{-2} Q - Z_k \Delta_k Z_k^T \equiv \hat{\Gamma}_{\text{post}}$$  (36)

where

$$\Delta_k \equiv \lambda^{-2} \begin{bmatrix} \frac{\theta_1}{\theta_1 + \lambda^2} & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & \frac{\theta_k}{\theta_k + \lambda^2} \end{bmatrix} \in \mathbb{R}^{k \times k}.$$  (37)

Notice that we have an efficient representation of $\hat{\Gamma}_{\text{post}}$ as a low-rank perturbation of the prior $\lambda^{-2} Q$. In summary, diagonal entries of $\hat{\Gamma}_{\text{post}}$ can provide estimates of diagonal entries of $\Gamma_{\text{post}}$. 


where the main computational requirement is to obtain the diagonals of \( \mathbf{Q} \) and the diagonals of the rank-\( k \) perturbation. Therefore, the only additional computational cost for estimating the posterior variance is \( O(k^3 + k^2 n_s n_t) \).

An approximation of this kind was previously explored in [4, 5, 14, 49]; however, the error estimates developed in the above references assume that the exact eigenpairs are available. If the Ritz pairs converge to the exact eigenpairs of the matrix \( \mathbf{Q} \mathbf{H} \), then furthermore, the optimality result in [56, theorem 2.3] applies here as well.

4.2. Decoupled approach

For cases where \( \mathbf{A} = \mathbf{A}_t \otimes \mathbf{A}_s \), we develop a decoupled strategy for estimating the posterior variances by exploiting the structure described in section 3.2. In contrast to the simultaneous approach, multiple Krylov subspaces are constructed and used in the decoupled approach. Using notation defined in section 3.2, the posterior covariance matrix is given by

\[
\Gamma_{\text{post}} = (\mathbf{Q}_t^{-1} \otimes \lambda^2 \mathbf{Q}_s^{-1} + \mathbf{H}_t \otimes \mathbf{H}_s)^{-1}
\]

\[= (\mathbf{L}_t^\top \mathbf{V}_t \otimes \mathbf{I})(\mathbf{I} \otimes \lambda^2 \mathbf{Q}_s^{-1} + \mathbf{\Sigma}_s^2 \otimes \mathbf{H}_s)^{-1}(\mathbf{V}_t^\top \mathbf{L}_t \otimes \mathbf{I}),
\]

where the matrix in the center is a block-diagonal matrix whose diagonal blocks are \((\lambda^2 \mathbf{Q}_s^{-1} + \sigma_j^2 \mathbf{H}_s)^{-1}\) for \( i = 1, \ldots, n_t \). Analogous to the simultaneous approach, gen-GK approximations for each \( i \) denoted by \( \mathbf{B}_{ij} \) and \( \mathbf{V}_{ij} \) can be used to get low-rank approximations

\[
(\lambda^2 \mathbf{Q}_s^{-1} + \sigma_j^2 \mathbf{H}_s)^{-1} \approx \lambda^{-2} \mathbf{Q}_s - \mathbf{Z}_{k,i} \mathbf{\Delta}_{k,i} \mathbf{Z}_{k,i}^\top, \quad i = 1, \ldots, n_t
\]

where \( \mathbf{B}_{ij} \mathbf{V}_{ij} = \mathbf{W}_{ij} \mathbf{\Theta}_{ij} \mathbf{W}_{ij}^\top \) is an eigenvalue decomposition,

\[
\mathbf{Z}_{k,i} = \mathbf{Q}_s \mathbf{V}_{ij} \mathbf{W}_{ij}, \quad \text{and} \quad \mathbf{\Delta}_{k,i} = \lambda^{-2}(\mathbf{I}_k + \lambda^2 \mathbf{\Theta}_{k,i}^{-1})^{-1}.
\]

Notice that depending on the stopping criteria for the gen-GK process, each time point may have a different rank \( k \). We omit this dependence for clarity of presentation, and denote the low-rank blocks \( \mathbf{D}_i = \mathbf{Z}_{k,i} \mathbf{\Delta}_{k,i} \mathbf{Z}_{k,i}^\top \).

In summary, an approximation to the posterior covariance matrix is given by

\[
\Gamma_{\text{post}} \approx \lambda^{-2} \mathbf{Q} - (\mathbf{L}_t^\top \mathbf{V}_t \otimes \mathbf{I}) \begin{bmatrix} \mathbf{D}_1 & \cdots & \mathbf{D}_n \end{bmatrix} \begin{bmatrix} \mathbf{V}_t^\top \mathbf{L}_t \otimes \mathbf{I} \end{bmatrix},
\]

where the diagonals can be computed for timestep \( i = 1, \ldots, n_t \) as

\[
\text{diag} \left[ (\mathbf{e}_i^\top \mathbf{I}) \Gamma_{\text{post}} (\mathbf{e}_i \otimes \mathbf{I}) \right] \approx \lambda^{-2} \left( [\mathbf{Q}_s]_{ii} \otimes \text{diag}(\mathbf{Q}_s) \right) - \sum_{j=1}^{m_i} (\mathbf{e}_j^\top \mathbf{L}_t \mathbf{e}_i)^2 \text{diag}(\mathbf{D}_i),
\]

where \([\mathbf{Q}_s]_{ij}\) is the \( ij \)-th diagonal entry of \( \mathbf{Q}_s \) and the operation \( \text{diag}(\cdot) \) returns a vector containing the diagonals of a matrix.

5. Numerical results

In this section, we provide three examples from image reconstruction. The first is a model problem from dynamic photoacoustic tomography (PAT) reconstruction that illustrates the
significant improvements that can be obtained by including a temporal prior. Then we consider an image deblurring problem that can exploit the decoupled methods, and we finish with a severely ill-posed passive seismic tomography (PST) problem, where we include results from real field measurements. In all of the results, genHyBR solutions correspond to (19).

5.1. Dynamic photoacoustic tomography (PAT)

PAT is a hybrid imaging modality that combines the rich contrast of optical imaging with the high resolution of ultrasound imaging, thereby producing higher resolution in vivo images with lower patient risk (e.g. requiring no ionizing radiation and no contrast agents) and lower cost and inconvenience than other imaging modalities. In modern PAT systems, transducers are rotated around an object and data is acquired in time. Most current methods for PAT image reconstruction such as explicit inversion formulas, time reversal, or series solution are not suited for such systems. A significant limitation in extending these methods to dynamic PAT is that accurate tomographic reconstruction relies on expensive motion estimation and artifact removal software, since the object being imaged may move during data acquisition. Chung and Nguyen [8] recently studied a continuous model of PAT reconstruction under motion for specific parameterized motion models and developed specialized algorithms for these scenarios. However, here we seek a sequence of reconstructions, rather than just one reconstruction, and we accomplish this in a Bayesian setting by incorporating Matérn type priors in both space and time. Previous works on dynamic tomography include [21, 22, 27], but many of these rely on parameterized models, whereas our framework can incorporate other realistic scenarios (other than motion) such as non-stationary optical illumination and inaccurate transducer responses [33, 54, 60, 61, 64].

We consider the discrete problem where $s_i \in \mathbb{R}^n$ is the discretized desired solution at time point $i$ and let $\{z_i\}, i = 1, \ldots, n_t$ denote the locations of the transducers. At each transducer location $z_i$, assume there are $r$ radii and let $A_i \in \mathbb{R}^{r \times ns}$ be the corresponding projection matrix for that location (i.e. $A_i s_i$ is the discrete circular Radon transform of $s$ on circles centered at $z_i$). Thus, the observed spherical projection measurements for all $r$ radii are contained in vector

$$d_i = A_i s_i + e_i,$$

where $e_i \in \mathbb{R}^r$ is additive Gaussian noise that is independent and identically distributed. Then the forward model has the form

$$
\begin{bmatrix}
d_1 \\
\vdots \\
d_{n_t}
\end{bmatrix} =
\begin{bmatrix}
A_1 \\
\vdots \\
A_{n_t}
\end{bmatrix}
\begin{bmatrix}
s_1 \\
\vdots \\
s_{n_t}
\end{bmatrix} +
\begin{bmatrix}
e_1 \\
\vdots \\
e_{n_t}
\end{bmatrix}
$$

where the goal is to estimate the desired images $s_i$, given the observations $d_i$.

For this example, we consider a numerical phantom consisting of 120 time frames, where each image is of size $256 \times 256$ and represents a superposition (i.e. a linear combination) of six circular absorbers. A similar phantom was used in [61]. See figure 1(a) for sample true images and the supplementary material (stacks.iop.org/IP/34/024005/mmedia) for a video. Measurements were taken at 120 equidistant angles between 0 and 357 at 3 degree intervals, and each projection consists of 363 radii. White noise was added to the observations with $R = .0017^2 I$ (this corresponds to a noise level of 0.02). Results presented here use $R$ as above.

6 Here we assume the 2D image is vectorized column-wise.
although noise estimation algorithms could be used [13]. The sinogram of size $363 \times 120$ is shown in figure 1(b), providing a total number of $43\,560$ observations.

For the dynamic inverse problem, the total number of unknowns is $256 \times 256 \times 120 = 7864\,320$. Note that our problem is severely underdetermined. This is contrary to other dynamic PAT setups, e.g. [61], where a complete sinogram image is obtained at each time point. In those cases, our decoupled genHyBR approach could be used.

Given the observed sinogram in figure 1, suppose we ignore the underlying dynamics and compute a static reconstruction. That is, we solve the following (inaccurate) model problem,

$$\min_{s \in \mathbb{R}^{n_s}} \| d - A s \|^2_{R^{-1}} + \lambda^2 \| s \|^2_{Q^{-1}}$$  (48)

with $d = [d_1^\top \cdots d_{n_t}^\top]^\top$ and $A = [A_1^\top \cdots A_{n_t}^\top]^\top$. To solve the static problem, we used genHyBR to obtain one reconstructed image of size $256 \times 256$, where the regularization parameter was picked using the WGCV criterion. Here we set $Q_s$ to be a covariance matrix that is determined from the Matérn covariance function $C_3 = C_{1.001}(\cdot)$. The static reconstruction is provided in figure 1(c), where it is evident that the reconstruction is able to locate the objects of interest but can not provide dynamic information.

Next we consider three cases that use simultaneous genHyBR to solve the dynamic problem (47):

- genHyBR where $Q$ is generated from a Matérn kernel $C_{1.01}(\sqrt{c_1} |p|^2 + c_2 |t|^2)$ where $c_1 = 1$ and $c_2 = 4 \times 10^{-4}$. Here, $Q$ can not be represented as a Kronecker product, but MVPs can still be done efficiently.
- genHyBR with $Q = Q_t \otimes Q_s$ where $Q_t = I$ and $Q_s$ corresponds to $C_3(\cdot) = C_{1.001}(\cdot)$.
- genHyBR with $Q = Q_t \otimes Q_s$ where $Q_t$ and $Q_s$ correspond to $C_{1.01}(\cdot) = C_{2.5,3}(\cdot)$ and $C_{1.01}(\cdot) = C_{1.001}(\cdot)$ respectively.

Image reconstructions at various time points, along with the corresponding true images, are provided in figure 2, where all of the results use the WGCV parameter after $10$ iterations.
Compared to the static reconstruction in figure 1(c), the Matérn reconstructions in the second row reveal changes over time. However, the more striking comparison occurs when using the separable covariance functions and comparing the results with $Q_t = I$ to those with $Q_t \neq I$ (see rows 3 and 4 in figure 2). As noted earlier, our dynamic PAT problem is severely underdetermined, and this example illustrates that including a temporal prior can be crucial to revealing dynamics of the imaged object. In table 1 we provide the computed regularization parameters for each approach, along with the relative errors computed as $\|s_t - s_{true}\|_2/\|s_{true}\|_2$. These values are consistent with the quality of the reconstructions in figure 2 and are not significantly improved with reorthogonalization of gen-GK vectors. For reconstructions that assume $Q$ is a Kronecker product, genHyBR took around 37s and, with reorthogonalization, around 106s\(^7\). Partial reorthogonalization could be used but was not investigated here.

Next, we show that variance estimates (i.e. approximations to diagonals of the posterior covariance matrix) can be obtained with minimal additional costs (here, in 15s). In figure 3

\(^7\) All timings were recorded on a MacPro, OSX Yosemite, 2.7 GHz 12-Core Intel Xeon E5, 64G memory in Matlab 2014b using default computing options.
we provide results for Matérn and $Q_t \neq I$ (corresponding to the MAP estimates in the 2nd and 4th rows of figure 2), where we note that overall variances are on the order of $10^{-5}$ and $10^{-3}$ respectively. We observe that solutions corresponding to earlier and later time points (e.g. $i = 1$ and $i = 120$) contain higher variances (i.e. greater uncertainty), with smaller variances in the center regions of the images, especially for $Q_t = I$. Variance images for $Q_t = I$ were essentially constant with mean value 0.0257 and standard deviation $2.3 \times 10^{-6}$ and thus are omitted.

In summary, we have shown that the gen-GK bidiagonalization can be used for the efficient computation of MAP estimates and variance estimates for dynamic PAT problems where the underlying object is changing slowly relative to the rate of image acquisition. Various choices for the prior covariance matrices could be included in this framework.

5.2. Space-time image deblurring

In dynamic image deblurring, the goal is to reconstruct a sequence of clear images from a sequence of blurred, noisy images. We consider a simulated problem. In figure 4, we provide 9 true images and the corresponding observed images, all of size $50 \times 50$ pixels. The blur matrix was taken to be $A = A_t \otimes A_s$ where $A_s$ represents a 2D Gaussian point spread function with spread parameter $\sigma = 0.07$ and bandwidth 3 and $A_t$ represents a 1D Gaussian blur with spread parameter $\sigma = 1$ and bandwidth 3. The noise level is set to be 0.02 such that $R = 0.04372I$. The problem set-up is a modification of the ‘blur’ example from the Regularization Tools toolbox [24].

We compare the following methods: LSQR ($Q = I$ and $\lambda = 0$), HyBR opt ($Q = I$ and optimal $\lambda$), genLSQR, and genHyBR with the optimal regularization parameter $\lambda_{opt}$ and the WGCV regularization parameter $\lambda_{wgcv}$. For genLSQR and genHyBR, we used $Q = Q_0 \otimes Q_t$ where $Q_0$ and $Q_t$ correspond to $C_T(\cdot) = C_{1.5,.3}(\cdot)$ and $C_S(\cdot) = C_{5,.007}(\cdot)$ respectively. Relative errors per iteration provided in figure 5 reveal similar behavior as that described in [10]. In particular, LSQR and genLSQR are plagued by semiconvergence (i.e. the ‘U’-shaped error curve that results from noise contamination during inversion), which can be avoided in the hybrid variants with the selection of the optimal regularization parameter. WGCV is able to provide a fairly good regularization parameter, but as indicated by the black diamond, the automatic stopping iteration (based on a GCV method) terminated the process at iteration 22. Relative errors for later iterations are provided to illustrate stabilization of the method.
Since \( A \) also has Kronecker product structure, the decoupled genHyBR approach can be used to compute MAP approximations (see algorithm 2). We denote ‘decoupled \( \lambda_{wgcv} \)’ to be the solution using the decoupled approach with a fixed regularization parameter \( \lambda_{wgcv} \), and ‘decoupled \( \lambda_{wgcv}^{(i)} \)’ refers to using a different regularization parameter for each subproblem. WGCV-selected regularization parameters \( \lambda_{wgcv}^{(i)} \) and corresponding stopping iterations \( k^{(i)}_{\text{stop}} \) are provided in table 2, along with regularization parameters \( \lambda_{wgcv} \) and \( \lambda_{opt} \). We remark that the regularization parameters \( \lambda_{wgcv}^{(i)} \) in decoupled approach decrease with increasing index \( i \); this can be attributed to the scaling factor from the singular values and the changing right hand sides.

Figure 4. Dynamic image deblurring problem. (a) True images. (b) Observed, blurred images.

Figure 5. Relative reconstruction errors for the dynamic image deblurring problem. LSQR and HyBR correspond to \( Q = I \), and genLSQR and genHyBR correspond to \( Q = Q_r \otimes Q_c \).

Since \( A \) also has Kronecker product structure, the decoupled genHyBR approach can be used to compute MAP approximations (see algorithm 2). We denote ‘decoupled \( \lambda_{wgcv} \)’ to be the solution using the decoupled approach with a fixed regularization parameter \( \lambda_{wgcv} \), and ‘decoupled \( \lambda_{wgcv}^{(i)} \)’ refers to using a different regularization parameter for each subproblem. WGCV-selected regularization parameters \( \lambda_{wgcv}^{(i)} \) and corresponding stopping iterations \( k^{(i)}_{\text{stop}} \) are provided in table 2, along with regularization parameters \( \lambda_{wgcv} \) and \( \lambda_{opt} \). We remark that the regularization parameters \( \lambda_{wgcv}^{(i)} \) in decoupled approach decrease with increasing index \( i \); this can be attributed to the scaling factor from the singular values and the changing right hand sides.
The relative reconstruction error for the decoupled $\lambda_{\text{wgcv}}$ reconstruction was 0.2461, which is slightly smaller than that of (simultaneous) genHyBR WGCV which was 0.2507 at termination. Furthermore, we observed that allowing different regularization parameters can result in smaller reconstruction error. The relative error for the decoupled $\lambda_{\text{wgcv}}^{(i)}$ reconstruction was 0.2198. Also, allowing a different Krylov subspace for each reconstruction can be beneficial in reducing 'ghosting' errors from neighboring images, as evident in the absolute error images provided in figure 6, where the same color axis is used per $i$ and black corresponds to larger errors. HyBR opt and genHyBR opt are only provided for reference, since they require the optimal regularization parameter that is not available in practice.

Variance estimates for genHyBR WGCV are provided in the first row of figure 7, where a similar observation to the dynamic PAT example can be made. Specifically, variances for solutions at earlier ($i = 1$) and later ($i = 9$) time points contain overall higher variances, possibly due to one-sided temporal information. Furthermore, the low variance estimates in the lower right corner (around the plus sign) imply little variation in solutions in this sharp region. The variance estimates in the second row illustrate that the decoupled variance estimate approach with fixed regularization parameter (as explained in section 4.2) can provide a good approximation.

We note that decoupled $\lambda_{\text{wgcv}}^{(i)}$ does not directly fit our framework; however, a modification of $Q_{i}^{-1}$ may be used to incorporate the changing regularization parameters. In summary, the
A decoupled approach can be used for both MAP and variance estimation if $A$, $Q$, and $R$ are all Kronecker products.

As a final remark, in section 3.2 we assume that the regularization parameter is fixed; however, in this section, we present results for which the regularization parameter is allowed to be

![Figure 7](image-url) Variance estimates for genHyBR WGCV and decoupled $\lambda_{wgcv}$ for time points 1, 3, 6, and 9. The top row refers to the simultaneous approach, whereas the bottom row corresponds to the decoupled approach.

![Table 3](image-url) Dates and number of rays for 8 time periods studied.

| Begin date | End date | Number of rays |
|------------|----------|----------------|
| Jan 14     | Jan 24   | $m_1 = 13,604$ |
| Jan 25     | Jan 31   | $m_2 = 8512$   |
| Feb 1      | Feb 17   | $m_3 = 9255$   |
| Feb 8      | Feb 14   | $m_4 = 27121$  |
| Feb 15     | Feb 21   | $m_5 = 41282$  |
| Feb 22     | Feb 28   | $m_6 = 13755$  |
| Mar 1      | Mar 17   | $m_7 = 11774$  |
| Mar 8      | Mar 16   | $m_8 = 66,553$ |

![Figure 8](image-url) Relative reconstruction errors in the observable regions per iteration for the PST checkerboard example.
This certainly has benefits as demonstrated; however, its statistical meaning is not fully clear and is worth exploring in future work.

5.3. Passive seismic tomography (PST)

Recent advances in PST have enabled the monitoring of mining-induced stress redistribution in coal and hardrock mines [34, 35, 62]. The basic goal of microseismic tomography is to image subsurface properties by using the many low-magnitude seismic events (e.g. micro-earthquakes) that are recorded by a microseismic monitoring system in a deep mine. Using time-lapse PST tomogram reconstructions, we can better understand the stress redistribution within the rock mass so that trends preceding and following significant seismic events can be analyzed. However, obtaining these 3D spatial reconstructions in real-time is a computationally challenging task that consists of solving a sequence of very large, often nonlinear, inverse problems.

Figure 9. Slices from the 4th volume of the simulated PST example. The top row contains slices of the true checkerboard volume, the second row contains the HyBR reconstruction (with $Q_t = I$ and $Q_s = I$), and the bottom two rows contain genHyBR reconstructions for $Q_t = I$ and $Q_s \neq I$ respectively. In these images, only the regions with ray coverage have been highlighted.
In this work, we consider a simplified, linear PST problem in a dynamic framework and investigate gen-GK methods for computing reconstructions. The basic formulation of the problem is the same as (2) where $s$ is a discretization of the velocity model, $d$ contains the observed travel times or recorded sinogram, and $A$ simulates a ray trace operation. We consider the situation in which measurements are taken in periodic time intervals, and the goal is to generate velocity models over time, from which the changing conditions within the rock mass, inferred to be changing due to induced seismicity, can be obtained.

5.3.1. PST simulated data. As is commonly done in practice, we begin with a simulated problem where the goal is to reconstruct a ‘checkerboard’ image [62]. We create eight checkerboard volumes of size $50 \times 66 \times 61$ voxels to represent true images, where the values of the checkerboard structure are generated to be reciprocals of $20\,000 \pm 10\%$ (i.e. values are $4.545 \times 10^{-5}$ and $5.555 \times 10^{-5}$) in the region of the volume which is seismically ‘observable’. Cross-sections from the 4th such generated structures are provided in the top row of figure 9. Then we used straight-path ray trace matrices $A_i$ for $i = 1, \ldots, 8$ from real mine data to generate sinograms. Each matrix corresponds to seismic events that occurred and

Figure 10. Isosurfaces and contours for real PST data at various time points. The first column corresponds to the time independent genHyBR approach, and the second and third columns correspond to the simultaneous genHyBR approach with $Q_t = I$ and $Q_t \neq I$ respectively.

In this work, we consider a simplified, linear PST problem in a dynamic framework and investigate gen-GK methods for computing reconstructions. The basic formulation of the problem is the same as (2) where $s$ is a discretization of the velocity model, $d$ contains the observed travel times or recorded sinogram, and $A$ simulates a ray trace operation. We consider the situation in which measurements are taken in periodic time intervals, and the goal is to generate velocity models over time, from which the changing conditions within the rock mass, inferred to be changing due to induced seismicity, can be obtained.

5.3.1. PST simulated data. As is commonly done in practice, we begin with a simulated problem where the goal is to reconstruct a ‘checkerboard’ image [62]. We create eight checkerboard volumes of size $50 \times 66 \times 61$ voxels to represent true images, where the values of the checkerboard structure are generated to be reciprocals of $20\,000 \pm 10\%$ (i.e. values are $4.545 \times 10^{-5}$ and $5.555 \times 10^{-5}$) in the region of the volume which is seismically ‘observable’. Cross-sections from the 4th such generated structures are provided in the top row of figure 9. Then we used straight-path ray trace matrices $A_i$ for $i = 1, \ldots, 8$ from real mine data to generate sinograms. Each matrix corresponds to seismic events that occurred and
were detected in a given time period; see table 3. Observed data were constructed using (2) where \( \varepsilon \) represented Gaussian noise with \( R = 0.0015^2 I \). In summary, the number of unknowns for this problem is \( 50 \times 66 \times 61 \times 8 = 1610400 \) and the total number of observations is \( m = 191,856 \) where the number of observations per time point is provided in table 3.

In figure 8, we provide relative reconstruction errors in the observable regions per iteration for various methods:

- \( \text{genLSQR} \) corresponds to \( C_T(\cdot) = C_{3.5,0.09}(\cdot) \), \( C_S = C_{2.5,0.025}(\cdot) \) and \( \lambda = 0 \).
- \( \text{genHyBR} Q_t \neq I \) corresponds to \( C_T(\cdot) = C_{3.5,0.09}(\cdot) \) and \( C_S = C_{2.5,0.025}(\cdot) \).
- \( \text{genHyBR} Q_t = I \) corresponds to \( C_S(\cdot) = C_{2,5,0.025}(\cdot) \).

Here \( C_{\nu,\ell}(\cdot) \) corresponds to a covariance matrix determined from a Matern kernel with parameters \( \nu \) and \( \ell \) as defined in section 2.3, and we assume, for simplicity, that \( p \) and \( t \) are equally spaced on a grid that is normalized to \([0,1]\). Regularization parameter \( \lambda_{\text{opt}} \) was used for \( \text{genHyBR} \).

For \( \text{LSQR} \) and \( \text{HyBR} \) (defined in section 5.2), we follow current practice and take an initial guess \( \mathbf{s}_0 \) to be a constant image with all entries equal to \( 5 \times 10^{-5} \). For the \( \text{genHyBR} \) reconstructions, we use a physically-informed prior mean and take \( \mathbf{\mu} \) to be a constant image with
all entries equal to $5 \times 10^{-5}$. We found these inclusions to be critical for obtaining physically meaningful results.

Oftentimes in PST, the temporal prior is ignored (e.g. $Q_t = I$) and reconstructions for each time point are done independently (with $Q_s = I$). Comparing HyBR and genHyBR for $Q_t = I$, we observe that better reconstructions can be obtained by including a spatial prior $Q_s$. Furthermore, these results show that incorporating a temporal prior may lead to additional improvements. Cross-sections of the 4th volume are provided in figure 9 for HyBR, and genHyBR for $Q_t = I$ and $Q_t \neq I$.

5.3.2. PST real data. As demonstrated in the simulated problem, the small ray path coverage makes dynamic PST a highly ill-posed problem. Next, we consider the real field data measurements and reconstruct a time-lapse of eight volumes using genHyBR. The true volumes are unknown; here we only show isosurfaces and comparisons to currently used algorithms. Along with expert field knowledge, this information can aid in evaluating the reconstructions and the potential for future improvements.

We present three reconstructions using genHyBR. The first approach essentially mimics what is done in practice, which is to compute reconstructions independent of time. Here we used genHyBR for each time period with $Q_s$ corresponding to Matérn covariance function $C_5(\cdot) = C_{10.5, 0.006}(\cdot)$. In the time independent approach, different regularization parameters and stopping iterations were selected for each reconstruction. Then we used simultaneous genHyBR with $Q_t = I$ and $Q_t$ corresponding to $C_7(\cdot) = C_{3.3}(\cdot)$. In both cases, we used $Q_s$ as defined above. For all of these experiments, WGCV was used to select the regularization parameter and automatic stopping criteria was used as described in [12] with a maximum of 10 iterations. Obtaining one dynamic PST reconstruction after 10 simultaneous genHyBR iterations took approximately 11 s.

High velocity iso-values (corresponding to a value of 20,025 or a slowness of $4.9938 \times 10^{-5}$) and contours for different time intervals are shown in figures 10 and 11.
the temporal prior can result in better reconstructions, especially for time intervals with very few observations (e.g. Jan 25–31 and Feb 1–17). In addition, genHyBR with \( Q \neq I \) gives more detailed information and locality of high stresses than the time independent reconstructions, where the isosurfaces are more dispersed.

We compare our reconstructions to a typical field reconstruction using the simultaneous iterative reconstruction technique (SIRT) for the time period February 22–28. In figure 12 we display cross-sections of genHyBR reconstructions, along with the SIRT reconstruction. These correspond to the 24th, 40th, and 27th slices in the \( x \), \( y \), and \( z \) dimensions respectively. We remark that the genHyBR reconstructions all provide smoother, more localized reconstructions of high-velocity zones. It is worth mentioning that these reconstructions have assumed that the forward model has ‘straight-ray’ paths and a typical approach in mining would be to use this reconstruction as an initial guess for obtaining reconstructions with a more sophisticated, nonlinear ‘curved-ray’ forward model. This is another topic of future work.

6. Conclusions

We consider a Bayesian framework for solving dynamic inverse problems, and we developed efficient, iterative, matrix-free methods based on the gen-GK bidiagonalization. A wide range of priors can be incorporated in our framework. We focused on priors that are modeled as Gaussian random fields with special attention to space-time covariance kernels for which MVPs can be computed efficiently. We first focus on computing the MAP estimate. In the simultaneous approach, a solution for the entire unknown in space-time is solved in an ‘all-at-once’ manner. When the observation operator also has Kronecker product structure, a series of variable transformations enables the problem to decouple in time. Both the simultaneous and decoupled approaches leverage the efficient iterative solvers developed in our previous work [10], and has the added benefit that the simultaneous approach allows for automatic selection of the regularization parameter. In addition to the MAP estimate, we describe methods that reuse intermediate information contained in the iterative solvers to estimate the variance of the posterior distribution. Several examples from image processing, including new applications to PST, demonstrate scalability of our algorithms and illustrate the broad applicability of our work.

Acknowledgments

We would like to acknowledge Creighton mines for providing the raw data for the PST application, and we thank the referees for their suggestions to improve the manuscript. Furthermore, this material is based upon work supported by the National Science Foundation under Grant DMS-1654175, and some of this work was conducted as a part of the SAMSI Program on Optimization 2016-2017, supported by the National Science Foundation under Grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute.

Appendix A. Additional details on generalized hybrid methods

In this section, we provide more details on the derivation of generalized hybrid methods, which is based on the gen-GK bidiagonalization in algorithm 1. Given matrices \( A, R, Q \), and vector \( b \), with initializations \( \beta_1 = \| b \|_R^{-1}, u_1 = b/\beta_1 \) and \( \alpha_1 v_1 = A^T R^{-1} u_1 \), the \( k \)th iteration of the gen-GK bidiagonalization procedure generates vectors \( u_{k+1} \) and \( v_{k+1} \) such that
\( \beta_{k+1} u_{k+1} = A Q v_k - \alpha_k u_k \quad (A.1) \)

\( \alpha_{k+1} v_{k+1} = A^\top R^{-1} u_{k+1} - \beta_{k+1} v_k \quad (A.2) \)

where scalars \( \alpha_i, \beta_i \geq 0 \) are chosen such that \( \|u_i\|_{R^{-1}} = \|v_i\|_Q = 1 \). At the end of \( k \) steps, we have

\[
B_k \equiv \begin{bmatrix}
\alpha_1 \\
\beta_2 \\
\alpha_2 \\
\beta_3 \\
\vdots \\
\alpha_k \\
\beta_{k+1}
\end{bmatrix}, \quad U_{k+1} \equiv [u_1, \ldots, u_{k+1}], \quad \text{and} \quad V_k \equiv [v_1, \ldots, v_k],
\]

(A.3)

where the following relations hold up to machine precision,

\( U_{k+1} \beta_1 e_1 = b \quad (A.4) \)

\( AQV_k = U_{k+1} B_k \quad (A.5) \)

\( A^\top R^{-1} U_{k+1} = V_k B_k^\top + \alpha_{k+1} v_{k+1} e_{k+1}^\top \quad (A.6) \)

Furthermore, in exact arithmetic, matrices \( U_{k+1} \) and \( V_k \) satisfy the following orthogonality conditions

\( U_{k+1}^\top R^{-1} U_{k+1} = I_{k+1} \quad \text{and} \quad V_k^\top Q V_k = I_k \quad (A.7) \)

We can define the residual at step \( k \) to be \( r_k = AQx_k - b \). It follows from Equations (A.4)–(A.6) that

\( r_k \equiv AQx_k - b = U_{k+1} (B_k z_k - \beta_1 e_1) \quad (A.8) \)

\( A^\top R^{-1} r_k = V_{k+1} (B_k z_k - \beta_1 e_1) \quad (A.9) \)

To obtain coefficients \( z_k \), we take \( z_k \) that minimizes the genLSQR problem,

\[
\min_{z_k \in \mathbb{R}^{k}} \frac{1}{2} \| r_k \|_{R^{-1}}^2 + \frac{\lambda^2}{2} \| x_k \|_Q^2 \Leftrightarrow \min_{z_k \in \mathbb{R}^{k}} \frac{1}{2} \| B_k z_k - \beta_1 e_1 \|_{Q}^2 + \frac{\lambda^2}{2} \| z_k \|_{2}^2 .
\]

(A.10)

Note that \( B_k \) is relatively small (of size \( (k+1) \times k \)), so its SVD can be computed efficiently, \( B_k = U_k \Sigma_k V_k^\top \). Then for fixed \( \lambda \), the solution to the projected problem is given by

\( z_k = (B_k^\top B_k + \lambda^2 I)^{-1} B_k^\top \beta_1 e_1 = V_k \Sigma_k \tilde{U}_k \beta_1 e_1 \quad (A.11) \)

where \( \Sigma_k \tilde{U}_k \) is the Moore–Penrose pseudoinverse.

Although a wide range of regularization parameter selection methods can be used in our framework, in this paper we consider a variant of the generalized cross validation (GCV) approach. The GCV parameter is selected to minimize the GCV function \([20]\) corresponding to the general-form Tikhonov problem (7),

\[
G(\lambda) = \frac{n \| A s(\lambda) - d \|_{R^{-1}}^2}{\text{trace} (I_n - L_R A A^\top A)} .
\]

(A.12)
where $A^\dagger = (A^\top R^{-1}A + \lambda^2 Q^{-1})^{-1}A^\top L R$. We have assumed $\mu = 0$ for simplicity. At the $k$th iteration, the GCV parameter corresponding to the projected problem (18) minimizes,

$$G_{\text{proj}}(\lambda) \equiv k \left\| (I - B_k B_k^\dagger) \beta_k e_i \right\|^2_2 + \text{trace} \left( (I_{k+1} - B_k B_k^\dagger) e_1^\top e_1 \right),$$

(A.13)

where $B_k^\dagger = (B_k^\top B_k + \lambda^2 I)^{-1}B_k^\top$. A weighted-GCV (WGCV) approach [12] has been suggested for use within hybrid methods, where a weighting parameter is introduced in the denominator of (A.13). We denote $\lambda_{\text{wgcv}}$ to be the regularization parameter computed using WGCV. Previous work on parameter selection within hybrid methods include [9, 12, 16, 28, 44, 45].

### Appendix B. Other examples that fit our framework

As mentioned in the introduction, the random-walk forecast model [29, 37, 55, 58] was previously considered for its computational advantages [32, 50]. We show how this model also fits within our framework. Assume that the state $s_i$ undergoes the following dynamics for $i = 1, \ldots, n_t - 1$

$$s_{i+1} = s_i + \epsilon_i \quad \epsilon_i \sim \mathcal{N}(0, Q_s),$$

(B.1)

with initial conditions $s_1 \sim \mathcal{N}(0, Q_s)$. We can then express the distribution of the state $s$ as

$$\pi(s) \propto \exp \left( -\frac{1}{2} \sum_{i=1}^{n_t-1} (s_{i+1} - s_i)^\top Q_s^{-1}(s_{i+1} - s_i) - \frac{1}{2} s_1^\top Q_s^{-1} s_1 \right).$$

(B.2)

Thus, $s$ is a Gaussian distribution with zero mean and covariance matrix $Q_s \otimes Q_s$, where

$$Q_s^{-1} = \begin{bmatrix}
2 & -1 & & \\
-1 & 2 & -1 & \\
& \ddots & \ddots & \ddots \\
& & -1 & 2 & -1 \\
& & & -1 & 1
\end{bmatrix}.$$  

(B.3)

The matrix $Q_s$ has an explicit representation and is the so-called vert minijvert matrix with $(i,j)$th entry of $Q_s$ equal to $\min\{i, j\}$. A similar representation is also available for the forecast model $s_{ti+1} = \alpha_i s_i + \beta_i \epsilon_i$, but will not be considered here.

Another approach assumes that $Q_s = (L_t^\top L_t + \gamma I)^{-1}$ where $L_t$ is a sparse discretization of a differential operator and $\gamma$ is a small positive parameter to ensure $Q_s$ is positive definite. For example, a common choice is to enforce smoothness in time by selecting

$$L_t = \begin{bmatrix}
\frac{1}{t_2-t_1} & \frac{1}{t_3-t_2} & \cdots & \\
\frac{1}{t_3-t_2} & \frac{1}{t_4-t_3} & \cdots & \\
& \ddots & \ddots & \ddots \\
& & \frac{1}{t_{n_t-1}-t_{n_t-2}} & \cdots & \\
& & & \frac{1}{t_{n_t-1}} & \frac{1}{t_{n_t-1}}
\end{bmatrix} \in \mathbb{R}^{(n_t-1)\times n_t}.$$  

(B.4)
Although not derived within a Bayesian framework, Schmitt and collaborators [52, 53] considered such temporal priors along with a standard Tikhonov term to enforce spatial smoothness. In fact, it is possible to show that their algorithm approximates the MAP estimate, which in our framework corresponds to $R = I, \mu = 0,$ and

$$Q = \left( \lambda_s^2 I + \lambda_t^2 B^T B \right)^{-1} = \left( \frac{\lambda_s^2}{\lambda_t^2} L^T L + \frac{\lambda_t^2}{\lambda_s^2} I \right)^{-1} \otimes \lambda_t^{-2} I \otimes \lambda_s^{-2} I,$$

where $B = L \otimes I$. Here $\lambda_s$ and $\lambda_t$ correspond to regularization parameters in space and time respectively, and the MAP estimate minimizes the function,

$$\Phi(s) = \|A s - d\|_2^2 + \lambda_s^2 \|s\|_2^2 + \lambda_t^2 \|B s\|_2^2.$$

In this paper, we focus on Matérn kernels and their covariance matrices for both the spatial and temporal priors.

**ORCID iDs**

Julianne Chung @ https://orcid.org/0000-0002-6760-4736

**References**

[1] Arioli M 2013 Generalized Golub–Kahan bidiagonalization and stopping criteria SIAM J. Matrix Anal. Appl. 34 571–92
[2] Arioli M and Orban D 2013 Iterative methods for symmetric quasi-definite linear systems—part I: Theory Cahier du GERAD G-2013-32 (Montréal, QC, Canada)
[3] Benbow S J 1999 Solving generalized least-squares problems with LSQR SIAM J. Matrix Anal. Appl. 21 166–77
[4] Bui-Thanh T, Burstedde C, Ghattas O, Martin J, Stadler G and Wilcox L C 2012 Extreme-scale UQ for Bayesian inverse problems governed by PDEs Proc. of the Int. Conf. on High Performance Computing, Networking, Storage and Analysis (IEEE Computer Society Press)
[5] Bui-Thanh T, Ghattas O, Martin J and Stadler G 2013 A computational framework for infinite-dimensional Bayesian inverse problems part I: the linearized case, with application to global seismic inversion SIAM J. Sci. Comput. 35 A2494–523
[6] Calvetti D and Somersalo E 2005 Priorconditioners for linear systems Inverse problems 21 1397–418
[7] Calvetti D 2007 Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective J. Comput. Appl. Math 198 378–95
[8] Chung J and Nguyen L 2017 Motion estimation and correction in photoacoustic tomographic reconstruction SIAM J. Imaging Sci. 10 216–42
[9] Chung J and Palmer K 2015 A hybrid LSQR algorithm for large-scale Tikhonov regularization SIAM J. Sci. Comput. 37 S562–80
[10] Chung J and Saibaba A 2017 Generalized hybrid iterative methods for large-scale Bayesian inverse problems SIAM J. Sci. Comput. 39 S24–46
[11] Chung J M, Kilmer M E and O’Leary D P 2015 A framework for regularization via operator approximation SIAM J. Sci. Comput. 37 B332–59
[12] Chung J, Nagy J G and O’Leary D P 2008 A weighted GCV method for Lanczos hybrid regularized Electron. Trans. Numer. Anal. 28 149–67
[13] Domoho D L 1995 De-noising by soft-thresholding IEEE Trans. Inf. Theory 41 613–27
[14] Flath H P, Wilcox L C, Akcèlik V, Hill J, van Bloemen Waanders B and Ghattas O 2011 Fast algorithms for Bayesian uncertainty quantification in large–scale linear inverse problems based on low-rank partial Hessian approximations SIAM J. Sci. Comput. 33 407–32
[15] Fuentes M 2006 Testing for separability of spatial—temporal covariance functions J. Stat. Plan. Inference 136 447–66
[16] Gazzola S, Novati P and Russo M R 2015 On Krylov projection methods and Tikhonov regularization Electron. Trans. Numer. Anal. 44 83–123
[17] Genton M G 2007 Separable approximations of space-time covariance matrices Environmetrics 18 681–95
[18] Gneiting T, Genton M G and Guttorp P 2006 Geostatistical space-time models, stationarity, separability, and full symmetry Monogr. Stat. Appl. Probab. 107 151
[19] Golub G H and Kahan W 1965 Calculating the singular values and pseudoinverse of a matrix SIAM J. Numer. Anal. 2 205–24
[20] Golub G H, Heath M and Wahba G 1979 Generalized cross-validation as a method for choosing a good ridge parameter Technometrics 21 215–23
[21] Hahn B N 2014 Efficient algorithms for linear dynamic inverse problems with known motion Inverse Problems 30 035008
[22] Hahn B N 2015 Dynamic linear inverse problems with moderate movements of the object: Ill-posedness and regularization Inverse Problems Imaging 9 395–413
[23] Hanke M and Hansen P C 1993 Regularization methods for large-scale problems Surv. Math. Ind. 3 253–315
[24] Hansen P C 1994 Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems Numer. Algorithms 6 1–35
[25] Hansen P C 2010 Discrete Inverse Problems: Insight and Algorithms (Philadelphia: SIAM) (https://doi.org/10.1137/1.9780898718836)
[26] Kamm J and Nagy J G 1998 Kronecker product and SVD approximations in image restoration Linear Algebra Appl. 284 177–92
[27] Katsievich A, Silver M and Zamyatin A 2011 Local tomography and the motion estimation problem SIAM J. Imaging Sci. 4 200–19
[28] Kilmer M E and O’Leary D P 2001 Choosing regularization parameters in iterative methods for ill-posed problems SIAM J. Matrix Anal. Appl. 22 1204–21
[29] Kim K, Kim B, Kim M, Lee Y and Vauhkonen M 2001 Image reconstruction in time-varying electrical impedance tomography based on the extended Kalman filter Meas. Sci. Technol. 12 1032
[30] Kyriakidis P C and Journel A G 1999 Geostatistical space—time models: a review Math. Geol. 31 651–84
[31] Laub A J 2005 Matrix Analysis for Scientists & Engineers (Philadelphia, PA: SIAM)
[32] Li J Y, Ambikasaran S, Darve E F and Kitanidis P K 2014 A Kalman filter powered by H2-matrices for quasi-continuous data assimilation problems Water Resour. Res. (https://doi.org/10.1002/2013WR014607)
[33] Lou Y, Wang K, Oraevsky A A and Anastasio M A 2016 Impact of nonstationary optical illumination on image reconstruction in optoacoustic tomography JOSA A 33 2333–47
[34] Luxbacher K, Westman E, Swanson P and Karfakis M 2008 Three-dimensional time-lapse velocity tomography of an underground longwall panel Int. J. Rock Mech. Min. Sci. 45 478–85
[35] Ma X, Westman E C, Fahrman B P and Thibodeau D 2016 Imaging of temporal stress redistribution due to triggered seismicity at a deep nickel mine Geomech. Energy Environ. 5 55–64
[36] Nagy J G, Ng M K and Perrone L 2004 Kronecker product approximation for image restoration with reflexive boundary conditions SIAM J. Matrix Anal. Appl. 25 829–41
[37] Nenna V, Pidlisecky A and Knight R 2011 Application of an extended Kalman filter approach to inversion of time-lapse electrical resistivity imaging data for monitoring recharge Water Resour. Res. 47 W10525
[38] Orban D and Arioli M 2017 Iterative Solution of Symmetric Quasi-Definite Linear Systems (Philadelphia: SIAM) (https://doi.org/10.1137/1.9781611974737)
[39] Paninski L 2010 Fast Kalman filtering on quasilinear dendritic trees J. Comput. Neurosci. 28 211–28
[40] Parks M L, De Sturler E, Mackey G, Johnson D D and Maiti S 2006 Recycling Krylov subspaces for sequences of linear systems SIAM J. Sci. Comput. 28 1651–74
[41] Pnevmatikakis E A, Rad K R, Huggins J and Paninski L 2014 Fast Kalman filtering and forward–backward smoothing via a low-rank perturbative approach J. Comput. Graph. Stat. 23 316–39
[42] Rashid A, Kim S, Liu D and Kim K 2016 A dynamic oppositional biogeography-based optimization approach for time-varying electrical impedance tomography Physiol. Meas. 37 820
