

Hybrid Projection Methods for Solution Decomposition in Large-Scale Bayesian Inverse Problems

Chung, Julianne; Jiang, Jiahua; Miller, Scot ; Saibaba, Arvind

DOI:
[10.1137/22M1502197](https://doi.org/10.1137/22M1502197)

License:
None: All rights reserved

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Chung, J, Jiang, J, Miller, S & Saibaba, A 2023, 'Hybrid Projection Methods for Solution Decomposition in Large-Scale Bayesian Inverse Problems', *SIAM Journal on Scientific Computing*, pp. S97-S119.
<https://doi.org/10.1137/22M1502197>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

First Published in *SIAM Journal on Scientific Computing*, published by the Society for Industrial and Applied Mathematics (SIAM). © by SIAM. Unauthorized reproduction of this article is prohibited.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

HYBRID PROJECTION METHODS FOR SOLUTION DECOMPOSITION IN LARGE-SCALE BAYESIAN INVERSE PROBLEMS*

JULIANNE CHUNG[†], JIAHUA JIANG[‡], SCOT M. MILLER[§], AND ARVIND K. SAIBABA[¶]

Abstract. We develop hybrid projection methods for computing solutions to large-scale inverse problems, where the solution represents a sum of different stochastic components. Such scenarios arise in many imaging applications (e.g., anomaly detection in atmospheric emissions tomography) where the reconstructed solution can be represented as a combination of two or more components and each component contains different smoothness or stochastic properties. In a deterministic inversion or inverse modeling framework, these assumptions correspond to different regularization terms for each solution in the sum. Although various prior assumptions can be included in our framework, we focus on the scenario where the solution is a sum of a sparse solution and a smooth solution. For computing solution estimates, we develop hybrid projection methods for solution decomposition that are based on a combined flexible and generalized Golub–Kahan process. This approach integrates techniques from the generalized Golub–Kahan bidiagonalization and the flexible Krylov methods. The benefits of the proposed methods are that the decomposition of the solution can be done iteratively, and the regularization terms and regularization parameters are adaptively chosen at each iteration. Numerical results from photoacoustic tomography and atmospheric inverse modeling demonstrate the potential for these methods to be used for anomaly detection.

Key words. inverse problems, hybrid methods, generalized Golub–Kahan, flexible methods, Tikhonov regularization, Bayesian inverse problems

MSC codes. 65F22, 65F10

DOI. 10.1137/22M1502197

1. Introduction. In many inverse problems, the ability to efficiently and accurately detect anomalies from observed data can have significant benefits. For example, in atmospheric inverse modeling, large-scale anomalous emissions of greenhouse gases and air pollution pose threats to human health, state emissions targets, and energy security. We need inverse models that can identify anomalous emissions events quickly—so the leak or malfunction in question can be fixed. However, inverse models that can identify anomalous emissions events require *more complicated prior models*, in particular, models that can incorporate multiple complex sources with different

*Received by the editors June 9, 2022; accepted for publication (in revised form) February 10, 2023; published electronically July 18, 2023.

<https://doi.org/10.1137/22M1502197>

Funding: This work was partially supported by the National Science Foundation ATD program under grants DMS-2026841, 2026830, and 2026835. This work was partially supported by National Natural Science Foundation of China under grant 12101406, and Shanghai Science and Technology Innovation Program under grant 21YF1429100. This material was also based upon work partially supported by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[†]Department of Mathematics, Emory University, Atlanta, GA, USA (jmchung@emory.edu).

[‡]Department of Mathematics, University of Birmingham, UK (j.jiang.3@bham.ac.uk).

[§]Department of Environmental Health and Engineering, Johns Hopkins University, Baltimore MD, USA (smill191@jhu.edu).

[¶]Department of Mathematics, North Carolina State University, Raleigh, NC, USA (asaibab@ncsu.edu).

smoothness properties. It is desirable to capture both anomalies (e.g., sparsely distributed events representing anomalous emissions like natural gas blowouts or point sources like large power plants) and smooth regions (e.g., representing broad scale emissions patterns from area sources). Moreover, the quality of the reconstruction depends crucially on the *choice of appropriate hyperparameters* that govern the prior and noise distributions, and estimating these parameters prior to inversion can be prohibitively expensive. For these and other inverse problems where the solution must capture different stochastic properties, we describe efficient and flexible iterative methods for reconstructing solutions that have a combination of smooth regions with sparse anomalies (e.g., for anomaly detection).

More specifically, we consider linear inverse problems of the form

$$(1.1) \quad \mathbf{d} = \mathbf{A}\mathbf{s} + \boldsymbol{\delta} \quad \text{with} \quad \mathbf{s} = \mathbf{s}_1 + \mathbf{s}_2,$$

where the goal is to reconstruct the desired parameters $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$, given forward operator (or parameter-to-observable map) $\mathbf{A} \in \mathbb{R}^{m \times n}$ and the observed data $\mathbf{d} \in \mathbb{R}^m$. We assume that the measurement errors $\boldsymbol{\delta}$ are realizations of Gaussian random variables, i.e., $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, where \mathbf{R} is a symmetric positive definite (SPD) matrix, and that \mathbf{s}_1 and \mathbf{s}_2 are mutually independent and are realizations from *different* distributions [29, 10]. Contrary to most inverse problems that involve estimating the unknown parameters \mathbf{s} , a distinguishing feature of the solution decomposed approach is that both sets of parameters \mathbf{s}_1 and \mathbf{s}_2 are estimated from the data \mathbf{d} , even if \mathbf{s} is desired in the end.

Contributions and overview. In this paper we propose a new computational framework for solving inverse problems in which the solution is assumed to be the sum of two components: a “smooth” background and a sparse term that represents anomalies. Following the Bayesian approach, we derive a posterior distribution for the unknown solution components and focus on efficiently computing the maximum a posteriori (MAP) estimate. Our approach has three main components:

1. Use of a majorization-minimization (MM) scheme to solve the optimization problem for the MAP estimate as a sequence of iteratively reweighted least-squares problems that carefully reweight only the sparse term;
2. a novel flexible, generalized Golub–Kahan (FGGK) iterative method for efficiently generating a single basis for approximately solving the reweighted least-squares subproblems; and
3. robust methods for automatically selecting the regularization parameters within the projected solution space at each iteration.

The main novelty of this paper is that in contrast to inner-outer methods or alternating approaches that solve a sequence of least-squares/optimization problems from “scratch,” our approach successively builds a single basis with which we seek approximate solutions for the components. To accomplish this task, we develop the FGGK approach, a new Krylov subspace solver, by paying close attention to the computational cost; like the generalized Golub–Kahan (genGK) process [17], it avoids forming inverses of covariance matrices and uses the same number of matrix-vector products with the forward operator and covariance matrices, but has the ability to incorporate information about both of the solution components. By automatically selecting the regularization parameter at each iteration, we avoid the need to solve repeated optimization problems to estimate the parameters.

We demonstrate the performance of our approach on a series of large-scale test problems that represent anomaly detection in dynamic tomography and atmospheric

inverse modeling, where for the last example that contains over a million unknowns, our method is capable of achieving satisfactory results in fewer than 50 iterations.

An outline of the paper is as follows. In section 2 we describe a Bayesian approach for (1.1) and provide an overview of related works that incorporate multiple stochastic components for inverse problems. The proposed hybrid projection method for solution decomposition is described in section 3, along with a description of some algorithmic considerations. Heuristic methods to choose regularization parameters are provided in Appendix A. Numerical results are provided in section 4, and conclusions are provided in section 5.

2. Bayesian inverse problems with solution decomposition. In this section, we describe a Bayesian approach for solving (1.1). We assume that the errors δ and the unknowns \mathbf{s} are mutually independent. For anomaly detection and more generally for problems where the solution can be decomposed, we consider the case where \mathbf{s}_1 follows a multivariate Gaussian distribution, and the components of \mathbf{s}_2 are independent and follow the univariate Laplace distribution [3, Chapter 4.3]. The Laplace prior promotes sparsity in \mathbf{s}_2 (e.g., the solution itself is sparse and contains many zeros or a representation in some frequency domain is sparse) [28].

In the geostatistical framework [37, 32, 30, 31], we model the unknown function $s(\zeta)$, where $\zeta \in \mathbb{R}^d$ represents the coordinates in space, as a realization of a random field. We express this realization as a sum of two terms,

$$s(\zeta) = s_1(\zeta) + s_2(\zeta), \quad s_2(\zeta) := \sum_{k=1}^p \beta_k \psi_k(\zeta),$$

where s_1 is the realization of a random field that captures the smooth features, ψ_k are deterministic basis functions, and β_k are coefficients to be determined. As is the prevalent approach, we take the random field s_1 to be Gaussian, which is characterized by a mean function $\mu_1(\zeta)$ and a covariance function $\lambda^{-2}\kappa(\zeta, \zeta')$, where λ is a parameter that controls the precision and is a hyperparameter that must be determined. For short, we write $s_1 \sim \text{GP}(\mu_1(\zeta), \lambda^{-2}\kappa(\zeta, \zeta'))$, where GP denotes a Gaussian process. We consider a set of grid points $\{\zeta_j\}_{j=1}^n$ on which we represent the unknown random field. Define the vector $\mathbf{s}_1 = [s_1(\zeta_1) \ \dots \ s_1(\zeta_n)]^\top$, then it follows that

$$\mathbf{s}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \lambda^{-2}\mathbf{Q}),$$

where $\boldsymbol{\mu}_1 = [\mu_1(\zeta_1) \ \dots \ \mu_1(\zeta_n)]^\top$ and $\mathbf{Q}_{ij} = \kappa(\zeta_i, \zeta_j)$ for $1 \leq i, j \leq n$.

Since anomalies are localized, we take $p = n$ and basis functions $\psi_k(\zeta) = \delta(\zeta - \zeta_k)$ as Dirac delta functions.¹ We promote sparsity in the coefficients β_k as follows: we assume that the coefficients β_k are independent of one another and the random field s_1 , and impose the univariate Laplace distribution with mean $[\boldsymbol{\mu}_2]_j$ and scale parameter $2\alpha^{-2}$, that is,

$$\beta_j \sim \mathcal{L}([\boldsymbol{\mu}_2]_j, 2\alpha^{-2}), \quad 1 \leq j \leq n.$$

Similarly to \mathbf{s}_1 , we define the vector \mathbf{s}_2 with components $[\mathbf{s}_2]_j = \beta_j$ for $1 \leq j \leq n$.

In summary, we have the prior model

$$(2.1) \quad \mathbf{s}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \lambda^{-2}\mathbf{Q}) \quad \text{and} \quad [\mathbf{s}_2]_j \sim \mathcal{L}([\boldsymbol{\mu}_2]_j, 2\alpha^{-2}), \quad 1 \leq j \leq n,$$

¹Although this model corresponds to the assumption that point source locations coincide with grid points, this assumption is not strictly needed for the algorithm. Other basis functions could be considered.

where $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^n$, \mathbf{Q} is SPD, and $\lambda \neq 0, \alpha \neq 0$ are scaling parameters. For problems of interest, computing the inverse and square root of \mathbf{R} is inexpensive (e.g., \mathbf{R} is often a diagonal matrix), but explicit computation of \mathbf{Q} (or its inverse or square root) may not be possible. However, we assume that matrix-vector multiplications involving \mathbf{A} , \mathbf{A}^\top , and \mathbf{Q} can be done efficiently (e.g., in $\mathcal{O}(n \log n)$ operations rather than $\mathcal{O}(n^2)$ operations for an $n \times n$ matrix); for details, see [2]. This framework can be extended to spatiotemporal models, but we do not provide the details here.

Under assumptions (1.1) and (2.1) and using Bayes' theorem, the posterior probability density function is given by

$$(2.2) \quad \begin{aligned} \pi_{\text{post}}(\mathbf{s}_1, \mathbf{s}_2 \mid \mathbf{d}) &= \frac{\pi(\mathbf{d} \mid \mathbf{s}_1, \mathbf{s}_2)\pi(\mathbf{s}_1)\pi(\mathbf{s}_2)}{\pi(\mathbf{d})} \\ &\propto \exp\left(-\frac{1}{2}\|\mathbf{A}\mathbf{s} - \mathbf{d}\|_{\mathbf{R}^{-1}}^2 - \frac{\lambda^2}{2}\|\mathbf{s}_1 - \boldsymbol{\mu}_1\|_{\mathbf{Q}^{-1}}^2 - \frac{\alpha^2}{2}\|\mathbf{s}_2 - \boldsymbol{\mu}_2\|_1\right), \end{aligned}$$

where $\|\cdot\|_1$ denotes the 1-norm of a vector, $\|\mathbf{x}\|_{\mathbf{M}^2} = \mathbf{x}^\top \mathbf{M} \mathbf{x}$ for any SPD matrix \mathbf{M} , and \propto means “proportional to.” In the Bayesian framework, the solution is the posterior distribution. Notice that the posterior is not Gaussian.

In this manuscript, we describe new hybrid projection methods to efficiently approximate the MAP estimate, which corresponds to the mode of the posterior distribution and is the solution to the following optimization problem:

$$(2.3) \quad \min_{\mathbf{s}_1 \in \mathbb{R}^n, \mathbf{s}_2 \in \mathbb{R}^n} \|\mathbf{A}(\mathbf{s}_1 + \mathbf{s}_2) - \mathbf{d}\|_{\mathbf{R}^{-1}}^2 + \lambda^2\|\mathbf{s}_1 - \boldsymbol{\mu}_1\|_{\mathbf{Q}^{-1}}^2 + \alpha^2\|\mathbf{s}_2 - \boldsymbol{\mu}_2\|_1.$$

Computational challenges. Optimization problems such as (2.3) can be computationally challenging to solve, and there are three main computational concerns. First, an accurate reconstruction of \mathbf{s} will rely heavily on being able to obtain good estimates of regularization parameters λ and α , which can be very difficult to estimate prior to solution computation. Second, for many problems with nonstandard priors (e.g., priors defined on nonstructured grids), explicit computation of \mathbf{Q} (or its inverse or square root) may not be possible. Generalized hybrid iterative methods which are based on the genGK bidiagonalization can be used to solve problems of the form

$$(2.4) \quad \min_{\mathbf{s}_1 \in \mathbb{R}^n} \|\mathbf{A}\mathbf{s}_1 - \mathbf{d}\|_{\mathbf{R}^{-1}}^2 + \lambda^2\|\mathbf{s}_1 - \boldsymbol{\mu}_1\|_{\mathbf{Q}^{-1}}^2$$

and are described in [17]. Third, it is well known that solving the ℓ_1 regularized problem can be computationally difficult, due to nondifferentiability at the origin as well as the need to use expensive nonlinear or iteratively reweighted optimization schemes. These inner-outer approaches can get very costly [20, 44], which has led to the development of accelerated alternative methods such as split Bregman methods [51] and iterative shrinkage thresholding algorithms [4], where an iterative two-step process is used. In [19], a weighted sum of mixed regularizing penalties were proposed to promote sparsity. However, these methods require a priori selection of various parameters (including the regularization and shrinkage parameter), which can be cumbersome [38]. Iterative alternating sequential algorithms that promote sparsity in solutions via a hierarchical Bayesian framework that postulate a conditionally Gaussian prior model and gamma hyperpriors are described in [8, 7, 11]. Connections between priorconditioning in the context of Bayesian hyperpriors and flexible preconditioning for iterative methods were made in [9]. Flexible Krylov methods have been

Algorithm 2.1 Solving (2.3) using alternating optimization

- 1: Initialize $\mathbf{s}_2^{(0)}, k = 1$
 - 2: **while** not converged **do**
 - 3: Solve $\mathbf{s}_1^{(k)} = \arg \min_{\mathbf{s}_1} \|\mathbf{A}\mathbf{s}_1 - (\mathbf{d} - \mathbf{A}\mathbf{s}_2^{(k-1)})\|_{\mathbf{R}^{-1}}^2 + \lambda^2 \|\mathbf{s}_1 - \boldsymbol{\mu}_1\|_{\mathbf{Q}^{-1}}^2$
 - 4: Solve $\mathbf{s}_2^{(k)} = \arg \min_{\mathbf{s}_2} \|\mathbf{A}\mathbf{s}_2 - (\mathbf{d} - \mathbf{A}\mathbf{s}_1^{(k)})\|_2^2 + \alpha^2 \|\mathbf{s}_2 - \boldsymbol{\mu}_2\|_1$
 - 5: Set $k = k + 1$
 - 6: **end while**
-

proposed [22, 14] as a means to both avoid inner-outer schemes and allow automatic regularization parameter selection. Such methods can be used to solve problems of the form

$$(2.5) \quad \min_{\mathbf{s}_2 \in \mathbb{R}^n} \|\mathbf{A}\mathbf{s}_2 - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{s}_2 - \boldsymbol{\mu}_2\|_1$$

and are described in [14].

A natural approach to solving (2.3) would utilize an alternating optimization scheme (see Algorithm 2.1), but this approach may be very slow and requires unrealistic initialization vectors. Instead, we describe in section 3 an iterative FGK approach that combines the flexible and generalized Golub–Kahan projection methods to efficiently generate a basis for solving inverse problems with mixed prior models.

Related approaches and ideas. The idea to incorporate multiple stochastic components for inverse problems is not necessarily new (see, e.g., earlier works [29, 10]), nor is it restricted to atmospheric imaging. For example, different texture models have been investigated for improved breast cancer imaging [34]. However, previous methods to handle multiple stochastic components are quite costly and often rely on simplifying assumptions. For example, in [52], the authors disaggregate the unknown fluxes to account for the biospheric and fossil fuel components separately, but simple Gaussian priors were used (a stationary, separable exponential model for the biospheric fluxes and a diagonal covariance matrix for the fossil fuel fluxes). In the special case where both priors are Gaussian, i.e., $\mathbf{s}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{Q}_1)$ and $\mathbf{s}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{Q}_2)$ with $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^n$ and SPD matrices $\mathbf{Q}_1, \mathbf{Q}_2$, it can be shown that $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \mathbf{Q}_1 + \mathbf{Q}_2)$, and methods for mixed Gaussian priors can be used [13]. Furthermore, if sparsity or a sparsity decomposition is desired, methods based on robust principal component analysis have been developed, e.g., in dynamic magnetic resonance imaging, to separate the solution (reshaped into a matrix) into a low-rank plus a sparse matrix [48], but such methods are too restrictive for the problems of interest.

We remark that although optimization problem (2.3) has a similar flavor to elastic net regularization [54, 27, 12] and other ℓ_1 - ℓ_2 problems [53], our approach is fundamentally different. First, these approaches linearly combine ℓ_1 and ℓ_2 regularization for \mathbf{s} and do not split the solution into two stochastic components. That is, they assume that the entire solution vector is included in both regularization terms, e.g., for elastic net,

$$\min_{\mathbf{s} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{s}\|_2^2 + \alpha^2 \|\mathbf{s}\|_1.$$

Second, typically iterative numerical methods of active set type are employed to solve elastic net regularized problems, but the inclusion of nontrivial regularizers makes these approaches infeasible.

3. Hybrid projection methods for solution decomposition. In this section, we describe an efficient computational method to approximate the MAP estimate given in (2.3). We develop a *combined* hybrid projection method that builds on the generalized and flexible Golub–Kahan processes and inherits many of the main computational benefits from previously developed hybrid approaches.

We begin in subsection 3.1 by describing an MM approach to handle the 1-norm regularization term in (2.3). Although this inner-outer optimization approach is computationally infeasible, it motivates the use of flexible preconditioning techniques that, extended and combined with genGK methods, are described in subsection 3.2. In particular, we describe an FGGK projection method, with some discussion on algorithmic considerations (e.g., breakdown) and computational considerations, and then we describe a hybrid projection approach based on the FGGK projection. We pay special attention to methods to choose regularization parameters for the projected problem in Appendix A.

3.1. MM approach. Various methods have been developed for approximating the solution of an ℓ_1 -regularized problem (2.5), ranging from iterative shrinkage algorithms to iterative reweighted norms [4, 44, 23]. In this subsection, we provide an overview of the MM approach for approximating the solution of (2.3), which requires solving a sequence of optimization problems. For this discussion, we assume that λ and α are fixed.

We begin with the following change of variables,

$$(3.1) \quad \mathbf{s}_1 = \boldsymbol{\mu}_1 + \mathbf{Q}\mathbf{x}, \quad \mathbf{s}_2 = \boldsymbol{\mu}_2 + \boldsymbol{\xi}, \quad \text{and} \quad \mathbf{c} = \mathbf{d} - \mathbf{A}\boldsymbol{\mu}_1 - \mathbf{A}\boldsymbol{\mu}_2,$$

and get the optimization problem

$$(3.2) \quad \min_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}^n} f(\mathbf{x}, \boldsymbol{\xi}) = \|\mathbf{A}\mathbf{Q}\mathbf{x} + \mathbf{A}\boldsymbol{\xi} - \mathbf{c}\|_{\mathbb{R}^{-1}}^2 + \lambda^2 \|\mathbf{x}\|_{\mathbf{Q}}^2 + \alpha^2 \|\boldsymbol{\xi}\|_1.$$

Notice that we have removed all instances of \mathbf{Q}^{-1} . Next, to handle the ℓ_1 term, we use the MM approach to convert optimization problem (3.2) into a sequence of reweighted least-squares problems. For some $\epsilon > 0$, we consider $|\xi| \approx \varphi_\epsilon(\xi) = \sqrt{\xi^2 + \epsilon}$ and approximate $\|\boldsymbol{\xi}\|_1 \approx \sum_{j=1}^n \varphi_\epsilon(\xi_j)$. The corresponding objective function is

$$(3.3) \quad f_\epsilon(\mathbf{x}, \boldsymbol{\xi}) = \|\mathbf{A}\mathbf{Q}\mathbf{x} + \mathbf{A}\boldsymbol{\xi} - \mathbf{c}\|_{\mathbb{R}^{-1}}^2 + \lambda^2 \|\mathbf{x}\|_{\mathbf{Q}}^2 + \alpha^2 \sum_{j=1}^n \varphi_\epsilon(\xi_j).$$

Let $\xi^{(k)}$ be the iterate at the k th step of the MM approach, then from (1.5) in [33], we have the majorization relationship

$$\varphi_\epsilon(\xi) = \sqrt{\xi^2 + \epsilon} \leq \sqrt{(\xi^{(k)})^2 + \epsilon} + \frac{1}{2\sqrt{(\xi^{(k)})^2 + \epsilon}} (\xi^2 - (\xi^{(k)})^2) =: \psi_\epsilon(\xi | \xi^{(k)}),$$

i.e., the function $\psi_\epsilon(\xi | \xi^{(k)})$ is said to majorize the function $\varphi_\epsilon(\xi)$ at $\xi^{(k)}$. Applying the majorization relationship to (3.3), we can define the surrogate function

$$g_\epsilon(\mathbf{x}, \boldsymbol{\xi} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) = \|\mathbf{A}\mathbf{Q}\mathbf{x} + \mathbf{A}\boldsymbol{\xi} - \mathbf{c}\|_{\mathbb{R}^{-1}}^2 + \lambda^2 \|\mathbf{x}\|_{\mathbf{Q}}^2 + \alpha^2 \sum_{j=1}^n \psi_\epsilon(\xi_j | \xi_j^{(k)}),$$

where the current iterate is given by $(\mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)})$.

It is easily verified that $f_\epsilon(\mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) = g_\epsilon(\mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)})$ and

$$f_\epsilon(\mathbf{x}, \boldsymbol{\xi}) \leq g_\epsilon(\mathbf{x}, \boldsymbol{\xi} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) \quad \forall \mathbf{x}, \boldsymbol{\xi} \in \mathbb{R}^n.$$

These two conditions mean that the surrogate function $g_\epsilon(\mathbf{x}, \boldsymbol{\xi} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)})$ matches the objective function $f_\epsilon(\mathbf{x}, \boldsymbol{\xi})$ at the current iterate and majorizes the surrogate function for every point, respectively. This means that as long as we choose the next iterate $(\mathbf{x}^{(k+1)}, \boldsymbol{\xi}^{(k+1)})$ such that the surrogate g_ϵ decreases, then we ensure that this decreases the objective f_ϵ since

$$\begin{aligned} f_\epsilon(\mathbf{x}^{(k+1)}, \boldsymbol{\xi}^{(k+1)}) &\leq g_\epsilon(\mathbf{x}^{(k+1)}, \boldsymbol{\xi}^{(k+1)} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) \\ &\leq g_\epsilon(\mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) = f_\epsilon(\mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}). \end{aligned}$$

The first inequality and the final equality are due to the majorization properties of the surrogate function g_ϵ . The second inequality is satisfied by choosing the next iterates in a manner to ensure that the surrogate is decreased; it is important to note that it is not necessary to minimize the surrogate at each iteration.

Thus, the MM algorithm for solving (3.2) is as follows: Given initial guesses $(\mathbf{x}^{(0)}, \boldsymbol{\xi}^{(0)})$, solve the following sequence of reweighted least-squares problems,

$$\begin{aligned} (\mathbf{x}^{(k+1)}, \boldsymbol{\xi}^{(k+1)}) &= \arg \min_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}^n} g_\epsilon(\mathbf{x}, \boldsymbol{\xi} | \mathbf{x}^{(k)}, \boldsymbol{\xi}^{(k)}) \\ (3.4) \quad &= \arg \min_{\mathbf{x} \in \mathbb{R}^n, \boldsymbol{\xi} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{Q}\mathbf{x} + \mathbf{A}\boldsymbol{\xi} - \mathbf{c}\|_{\mathbb{R}^{-1}}^2 + \lambda^2 \|\mathbf{x}\|_{\mathbf{Q}}^2 + \alpha^2 \|\mathbf{D}(\boldsymbol{\xi}^{(k)})\boldsymbol{\xi}\|_2^2, \end{aligned}$$

where terms from g_ϵ that do not depend on \mathbf{x} and $\boldsymbol{\xi}$ have been dropped and the diagonal matrix $\mathbf{D}(\boldsymbol{\xi}) = \text{diag}([2\sqrt{\xi_i^2 + \epsilon}]^{-1/2})_{i=1}^n$. To get the solution after removing the change of variables, we get $\mathbf{s}^{(k+1)} = \boldsymbol{\mu}_1 + \mathbf{Q}\mathbf{x}^{(k+1)} + \boldsymbol{\mu}_2 + \boldsymbol{\xi}^{(k+1)}$.

The convergence of the MM scheme has been established; see, e.g., [26]. However, notice that minimizing the surrogate requires solving a large optimization problem (3.4) with $2n$ unknowns at each iteration. For small problems, one could solve the corresponding normal equations. For larger problems, an iterative method could be used to solve the reweighted least-squares problems, but this often leads to expensive inner-outer solves [44]. Instead, we describe in the next section an approach that avoids inner-outer schemes by exploiting flexible preconditioning techniques, following recent works, e.g., [22, 14]. That is, we solve optimization problem (3.4) approximately at each step.

3.2. FGGK iterative method. In this section, we describe iterative projection methods that can be used for approximating the solution for inverse problems with solution decomposition (e.g., for anomaly detection). We exploit aspects of both the flexible and generalized Golub–Kahan projection methods and develop a solution decomposition hybrid projection approach, henceforth dubbed **sdHybr**, to approximate the MAP estimate (2.3). Similarly to all hybrid projection methods that combine iterative projection methods with variational regularization techniques, there are two main components. First, we generate a single basis for the solution (which in this case includes two sets of solution vectors) by exploiting a flexible preconditioning framework integrated with a genGK bidiagonalization. Second, we compute an approximate solution to the inverse problem by solving an optimization problem in the projected subspace where regularization parameters can be estimated automatically.

FGGK process. We consider solving problems of the form (3.4) by incorporating a changing diagonal matrix, which rescales the norms, to generate a basis for the solution. Suppose we are given \mathbf{A} , \mathbf{Q} , \mathbf{R} , \mathbf{c} , and a sequence of invertible matrices $\{\mathbf{D}_j\}_{j=1}^k$. We initialize the iterations with $m_{1,1} = \|\mathbf{c}\|_{\mathbf{R}^{-1}}$ and $\mathbf{u}_1 = \mathbf{c}/m_{1,1}$; furthermore, take $\mathbf{v}_1 = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_1$ and $t_{1,1} = \|\mathbf{v}_1\|_{\mathbf{Q}}$. The FGGK iterative process generates vectors \mathbf{z}_k , \mathbf{v}_k , and \mathbf{u}_{k+1} such that at the k th iteration, we have

$$(3.5) \quad m_{k+1,k} \mathbf{u}_{k+1} = \mathbf{A} \mathbf{Q} \mathbf{v}_k + \mathbf{A} \mathbf{D}_k^{-1} \mathbf{v}_k - \sum_{j=1}^k m_{j,k} \mathbf{u}_j,$$

$$(3.6) \quad t_{k+1,k} \mathbf{v}_{k+1} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_{k+1} - \sum_{j=1}^k t_{j,k} \mathbf{v}_j.$$

In the first step, we expand the basis \mathbf{u}_{k+1} by including the vectors $\mathbf{A} \mathbf{Q} \mathbf{v}_k$ and $\mathbf{A} \mathbf{D}_k^{-1} \mathbf{v}_k$, and we orthogonalize against the previous basis vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ using Gram–Schmidt with the inner product $\langle \cdot, \cdot \rangle_{\mathbf{Q}}$. Similarly, we expand the basis vectors \mathbf{v}_{k+1} with the vector $\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_{k+1}$ and orthogonalize against the previous vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ using the inner product $\langle \cdot, \cdot \rangle_{\mathbf{R}^{-1}}$. Finally we ensure that both $\mathbf{u}_j, \mathbf{v}_j$ are normalized so that $\|\mathbf{u}_j\|_{\mathbf{R}^{-1}} = \|\mathbf{v}_j\|_{\mathbf{Q}} = 1$, and let $\mathbf{w}_j = \mathbf{D}_j^{-1} \mathbf{v}_j$ for $1 \leq j \leq k+1$.

For notational convenience, consider the augmented matrices

$$(3.7) \quad \widehat{\mathbf{A}} = [\mathbf{A} \quad \mathbf{A}] \in \mathbb{R}^{m \times 2n} \quad \text{and} \quad \widehat{\mathbf{Q}} = \begin{bmatrix} \mathbf{Q} & \\ & \mathbf{I} \end{bmatrix} \in \mathbb{R}^{(2n) \times (2n)}.$$

Equations (3.5) and (3.6) can be summarized in matrix form as

$$(3.8) \quad \widehat{\mathbf{A}} \widehat{\mathbf{Q}} \mathbf{Z}_k = \mathbf{U}_{k+1} \mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{V}_{k+1} \mathbf{T}_{k+1},$$

where the search basis \mathbf{Z}_k takes the form

$$(3.9) \quad \mathbf{Z}_k = [\mathbf{z}_1 \quad \dots \quad \mathbf{z}_k] = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \\ \mathbf{w}_1 & \dots & \mathbf{w}_k \end{bmatrix} = \begin{bmatrix} \mathbf{V}_k \\ \mathbf{W}_k \end{bmatrix} \in \mathbb{R}^{2n \times k}.$$

Furthermore, we have two matrices $\mathbf{M}_k \in \mathbb{R}^{(k+1) \times k}$ and $\mathbf{T}_{k+1} \in \mathbb{R}^{(k+1) \times (k+1)}$ that are upper Hessenberg and upper triangular, respectively. The basis vectors are collected in the matrices $\mathbf{U}_{k+1} = [\mathbf{u}_1 \quad \dots \quad \mathbf{u}_{k+1}] \in \mathbb{R}^{m \times (k+1)}$ and $\mathbf{V}_{k+1} = [\mathbf{v}_1 \quad \dots \quad \mathbf{v}_{k+1}] \in \mathbb{R}^{n \times (k+1)}$ which satisfy the orthogonality conditions (in exact arithmetic)

$$(3.10) \quad \mathbf{U}_{k+1}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{I}_{k+1}, \quad \mathbf{V}_{k+1}^\top \mathbf{Q} \mathbf{V}_{k+1} = \mathbf{I}_{k+1}.$$

Solving the least-squares problem. Thus far, we have described the FGGK process as an iterative method to generate a basis for the solution. Next, we seek approximate solutions to the least-squares problem (3.4) by using the FGGK relations above to obtain a sequence of projected least-squares problems. For clarity of presentation, we assume that the parameters λ and α are fixed.

To determine the optimal coefficients \mathbf{f} in (3.12), we plug the FGGK relations (3.8) into the objective function to get

$$(3.11) \quad \mathbf{f}_k = \arg \min_{\mathbf{f} \in \mathbb{R}^k} \|\mathbf{M}_k \mathbf{f} - m_{1,1} \mathbf{e}_1\|_2^2 + \lambda^2 \|\mathbf{f}\|_2^2 + \alpha^2 \|\mathbf{W}_k \mathbf{f}\|_2^2,$$

which is equivalent to

$$(3.12) \quad \min_{\mathbf{x} = \mathbf{V}_k \mathbf{f}, \mathbf{y} = \mathbf{W}_k \mathbf{f}} \|\mathbf{A} \mathbf{Q} \mathbf{x} + \mathbf{A} \mathbf{y} - \mathbf{c}\|_{\mathbf{R}^{-1}}^2 + \lambda^2 \|\mathbf{x}\|_{\mathbf{Q}}^2 + \alpha^2 \|\mathbf{y}\|_2^2.$$

Our ultimate goal is to obtain the decomposed solution $\mathbf{s} = \mathbf{s}_1 + \mathbf{s}_2$. Based on the discussion above and using the change of variables (3.1), we get

$$\mathbf{s}_1^{(k)} := \boldsymbol{\mu}_1 + \mathbf{Q}\mathbf{V}_k\mathbf{f}_k, \quad \mathbf{s}_2^{(k)} := \boldsymbol{\mu}_2 + \mathbf{W}_k\mathbf{f}_k,$$

where $\mathbf{f}_k \in \mathbb{R}^k$ are the coefficients obtained by solving (3.11). In summary, an approximation to the MAP estimate at the k th iteration of the **sdHybr** method would be $\mathbf{s}^{(k)} = \boldsymbol{\mu}_1 + \mathbf{Q}\mathbf{V}_k\mathbf{f}_k + \boldsymbol{\mu}_2 + \mathbf{W}_k\mathbf{f}_k$, where \mathbf{f}_k solves the optimization problem (3.11).

Efficient QR update. Similar to what is done in [14], we can take the thin QR factorization of $\mathbf{W}_k = \mathbf{Q}_{W,k}\mathbf{R}_{W,k}$, then (3.11) becomes

$$(3.13) \quad \mathbf{f}_k = \arg \min_{\mathbf{f} \in \mathbb{R}^k} \|\mathbf{M}_k\mathbf{f} - m_{1,1}\mathbf{e}_1\|_2^2 + \lambda^2\|\mathbf{f}\|_2^2 + \alpha^2\|\mathbf{R}_{W,k}\mathbf{f}\|_2^2.$$

An efficient update of the QR factorization of \mathbf{W}_k is as follows. Suppose we have computed the thin-QR factorization

$$(3.14) \quad \mathbf{W}_k = \mathbf{Q}_{W,k}\mathbf{R}_{W,k},$$

where $\mathbf{Q}_{W,k}^\top\mathbf{Q}_{W,k} = \mathbf{I}_k$ and $\mathbf{R}_{W,k}$ is an upper triangular matrix. Then we can efficiently update the QR factorization of \mathbf{W}_k in $\mathcal{O}(nk)$ flops using Gram–Schmidt as

$$(3.15) \quad \begin{aligned} \mathbf{W}_{k+1} &= \begin{bmatrix} \mathbf{W}_k & \mathbf{w}_{k+1} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Q}_{W,k}\mathbf{R}_{W,k} & \mathbf{w}_{k+1} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \mathbf{Q}_{W,k} & (\mathbf{I} - \mathbf{Q}_{W,k}\mathbf{Q}_{W,k}^\top)\mathbf{w}_{k+1}/\beta_{k+1} \end{bmatrix}}_{\mathbf{Q}_{W,k+1}} \underbrace{\begin{bmatrix} \mathbf{R}_{W,k} & \mathbf{Q}_{W,k}^\top\mathbf{w}_{k+1} \\ \mathbf{0} & \beta_{k+1} \end{bmatrix}}_{\mathbf{R}_{W,k+1}}, \end{aligned}$$

where $\beta_{k+1} = \|(\mathbf{I} - \mathbf{Q}_{W,k}\mathbf{Q}_{W,k}^\top)\mathbf{w}_{k+1}\|_2$. For additional numerical stability, one can use another round of Gram–Schmidt or instead use Householder QR updates.

A summary of the **sdHybr** method is provided in Algorithm 3.1. Notice that in a fully Bayesian approach, the regularization parameters λ and α are part of the prior. Thus, they should be set in advance. Any uncertainty in the prior can be expressed, e.g., in terms of hyperpriors [9], but estimation of these parameters requires expensive sampling schemes [3, 1] that can quickly prove to be computationally infeasible. Furthermore, if poor hyperparameters are selected, a fully Bayesian approach may lead to poor solutions. To circumvent such challenges, we provide in Appendix A some heuristic approaches that can be used to automatically estimate the regularization parameters at each iteration, i.e., λ_k and α_k . We remark that various existing stopping criteria can be used. A systematic study of computational methods for hyperparameter selection is outside the scope of this article.

Computational Cost. Each iteration of the **sdHybr** method requires one matrix-vector product with \mathbf{A} and one with its adjoint (let $T_{\mathbf{A}}$ denote the cost of one matrix-vector product with \mathbf{A} or its adjoint, two matrix-vector products with \mathbf{Q} (similarly, denoted $T_{\mathbf{Q}}$), one matrix-vector product with \mathbf{R}^{-1} (denoted $T_{\mathbf{R}^{-1}}$), one matrix-vector product with \mathbf{D}_k^{-1} (denoted $T_{\mathbf{D}_k^{-1}}$), the inversion of diagonal matrix \mathbf{D}_k that is $\mathcal{O}(n)$ floating point operations (flops), and additional $\mathcal{O}(k(m+n))$ flops for the summation calculation in (3.5) and (3.6). To compute the solution of the projected problem (3.13), the cost is $\mathcal{O}(k^3)$ flops, since \mathbf{M}_k is an upper Hessenberg matrix. And the cost of forming \mathbf{x} and \mathbf{y} to get \mathbf{s} is $\mathcal{O}(k^2(m+n))$ flops. Thus, the overall cost of the algorithm is

$$(3.16) \quad T_{\text{sdHybr}} = 2kT_{\mathbf{A}} + 2kT_{\mathbf{Q}} + kT_{\mathbf{R}^{-1}} + \mathcal{O}(k^2(m+n)) + \mathcal{O}(k^4) \text{ flops.}$$

Algorithm 3.1 sdHybr

Require: Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, positive definite matrices $\mathbf{Q} \in \mathbb{R}^{n \times n}$ and $\mathbf{R} \in \mathbb{R}^{m \times m}$, vector $\mathbf{c} \in \mathbb{R}^m$, $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathbb{R}^n$. Invertible matrix $\mathbf{D}_1 = \mathbf{I}_n \in \mathbb{R}^{n \times n}$.

- 1: Initialize $\mathbf{u}_1 = \mathbf{c}/m_{1,1}$, where $m_{1,1} = \|\mathbf{c}\|_{\mathbf{R}^{-1}}$ and $\mathbf{v}_1 = \mathbf{0}, k = 1$.
- 2: **while** stopping criteria are not satisfied **do**
- 3: $\mathbf{h} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{u}_k, t_{j,k} = \mathbf{h}^\top \mathbf{Q} \mathbf{v}_j$ for $j = 1, \dots, k-1$
- 4: $\mathbf{h} = \mathbf{h} - \sum_{j=1}^{k-1} t_{j,k} \mathbf{v}_j, t_{k,k} = \|\mathbf{h}\|_{\mathbf{Q}}, \mathbf{v}_k = \mathbf{h}/t_{k,k}$
- 5: $\mathbf{z}_k = \begin{bmatrix} \mathbf{v}_k \\ \mathbf{w}_k \end{bmatrix}, \mathbf{V}_k = [\mathbf{v}_1 \ \dots \ \mathbf{v}_k], \mathbf{W}_k = [\mathbf{w}_1 \ \dots \ \mathbf{w}_k]$, where $\mathbf{w}_k = \mathbf{D}_k^{-1} \mathbf{v}_k$.
- 6: $\mathbf{h} = \mathbf{A}(\mathbf{Q} \mathbf{v}_k + \mathbf{w}_k), m_{j,k} = \mathbf{h}^\top \mathbf{R}^{-1} \mathbf{u}_j$ for $j = 1, \dots, k$
- 7: $\mathbf{h} = \mathbf{h} - \sum_{j=1}^k m_{j,k} \mathbf{u}_j, m_{k+1,k} = \|\mathbf{h}\|_{\mathbf{R}^{-1}}, \mathbf{u}_{k+1} = \mathbf{h}/m_{k+1,k}$
- 8: Update QR factorization using (3.15) to obtain $\mathbf{W}_{k+1} = \mathbf{Q}_{k+1} \mathbf{R}_{k+1}$
- 9: Solve (3.13) to get $\mathbf{f}_k(\lambda_k, \alpha_k)$ with selected regularization parameters λ_k, α_k .
- 10: $\mathbf{s}_1^{(k)} = \boldsymbol{\mu}_1 + \mathbf{Q} \mathbf{V}_k \mathbf{f}_k, \mathbf{s}_2^{(k)} = \boldsymbol{\mu}_2 + \mathbf{W}_k \mathbf{f}_k$.
- 11: $\mathbf{D}_{k+1} = \mathbf{D}(\mathbf{W}_k \mathbf{f}_k)$
- 12: $k = k + 1$
- 13: **end while**
- 14: **return** Approximations $\mathbf{s}_1^{(k)}$ and $\mathbf{s}_2^{(k)}$ that define the sum $\mathbf{s}^{(k)} = \mathbf{s}_1^{(k)} + \mathbf{s}_2^{(k)}$.

Notice that compared to the MM method, the projected problem (3.11) for **sdHybr** is cheaper to solve than (3.4) at each MM iteration, since it is an optimization problem over a smaller dimensional space.

Variations of the proposed approach. We remark that various projection methods that combine flexible and generalized GK methods could be considered besides the **sdHybr** approach; however, for stability and for proper selection of regularization parameters, we found this projection approach (Algorithm 3.1) to be the most computationally appealing. For example, a naïve first approach would be to use a genGK approach to handle the \mathbf{Q} -norm regularizer and use the flexible Golub–Kahan approach to handle the 1-norm separately in (3.2). This would generate two solution subspaces that each contain orthonormal columns but are not orthogonal to each other, and the number of unknowns for the projected problem would be $2k$. Although an efficient QR update could be used, there are potential issues with breakdown (e.g., due to linear dependence of subspace vectors). Furthermore, we found that this approach can be sensitive to initializations. The FGGK process described above avoids this by working with stacked solution vectors and a projected problem of order k .

Another natural approach to combine flexible and genGK methods would be to reformulate the problem such that the genGK vectors include the flexible preconditioner. This approach is described in Appendix B and provides a nice alternative, but the main caveat is that selecting regularization parameters in a hybrid framework becomes more challenging. In particular, a hybrid framework based on the projection method described in Appendix B requires formulating one regularization parameter as a fixed scalar multiple of the other or utilizing more expensive optimization procedures. This is due to the inability to simplify the norm for the regularizer when considering different regularization parameters. In contrast, the FGGK procedure leads to a projected problem (3.11) for which existing parameter selection approaches can be naturally applied, as described in the next section.

Remarks on the solution decomposition method. The FGGK process was designed to solve the sequence of optimization problems (3.4) involving a diagonal

weighting matrix that changes at each iteration. We make a few comments when the weighting matrix is fixed at each iteration. Recall that the Krylov subspace associated with the matrix \mathbf{E} and vector \mathbf{g} is defined as

$$\mathcal{K}_k(\mathbf{E}, \mathbf{g}) \equiv \text{Span}\{\mathbf{g}, \mathbf{E}\mathbf{g}, \dots, \mathbf{E}^{k-1}\mathbf{g}\}.$$

Assume that \mathbf{D}_j with $j = 1, \dots, k$ are fixed, that is $\mathbf{D}_1 = \dots = \mathbf{D}_k = \widehat{\mathbf{D}}$ and, furthermore, assume $\widehat{\mathbf{D}}$ is invertible. For the FGGK process, based on the relations in (3.5) and (3.6), it can be shown that the columns of \mathbf{U}_k and \mathbf{V}_k form \mathbf{R}^{-1} -orthogonal and \mathbf{Q} -orthogonal (c.f., (3.10)) bases for the Krylov subspaces,

$$\begin{aligned} \text{Span}\{\mathbf{U}_k\} &= \mathcal{K}_k(\mathbf{A}(\mathbf{Q} + \widehat{\mathbf{D}}^{-1})\mathbf{A}^\top \mathbf{R}^{-1}, \mathbf{c}), \\ \text{Span}\{\mathbf{V}_k\} &= \mathcal{K}_k(\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}(\mathbf{Q} + \widehat{\mathbf{D}}^{-1}), \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{c}), \end{aligned}$$

respectively. This means that when the diagonal weighting matrices are fixed, the subspaces generated using the FGGK approach are Krylov subspaces.

4. Numerical results. In this section, we investigate the performance of the proposed `sdHybr` method using various examples from image processing. In subsection 4.1, we consider a hypothetical atmospheric transport problem where the goal is to recover emissions maps that cover North America for detecting anomalies. Then, in subsection 4.2 we consider dynamic spherical means tomography reconstruction, where the true image combines moving smooth components and sparsely positioned dots. Finally, in subsection 4.3 we consider a more challenging case study from NASA's Orbiting Carbon Observatory 2 (OCO-2) satellite that includes an atmospheric transport model and dynamic data. For the last example, the true solution is not synthetically generated as a sum of random fields. For each of the case studies, we show that `sdHybr` methods are able to capture both smooth and sparse components.

We compare the performance of `sdHybr` methods to that of generalized hybrid methods [17] and flexible hybrid algorithms [14], which we denote by `genHyBR` and `fHybr`, respectively. For the hybrid methods, we consider regularization parameters selected using the unbiased predictive risk estimation (UPRE) method, discrepancy principle (DP), and the weighted generalized cross validation (WGCV) method. For `sdHybr`, this corresponds to solving nonlinear constrained optimization problems (A.2), (A.3), and (A.4), respectively. For this task, we use a quasi-Newton method as implemented in the MATLAB `fminunc` function with an initial guess of $\lambda = -0.5$ and $\gamma = -0.5$. For the stopping criteria for `sdHybr`, the iterative method is terminated if either of the following two criteria is satisfied: (i) a maximum number of iterations is reached; (ii) the GCV stopping function \widehat{G} defined in (A.5) reaches the minimum or flattens out. The following experiments ran on a laptop computer with Intel i5 CPU 2 GHz and 16G memory.

4.1. Case study 1: A hypothetical atmospheric transport problem. We investigate a synthetic atmospheric transport problem, where observations $\mathbf{d} \in \mathbb{R}^{98880}$ are generated as in (1.1) with $\mathbf{A} \in \mathbb{R}^{98880 \times 3222}$ representing a forward atmospheric model and \mathbf{s} representing the true emissions map which is a synthetically generated summation of a randomly generated smooth image and an image with sparse anomalies. Henceforth, we denote the true emissions vector as $\mathbf{s}_{\text{true}} \in \mathbb{R}^{3222}$. See Figure 1. The smooth image is generated by a Matérn kernel [50, equation (4.14)] with parameters $\nu = 2.5$ and $\ell = 0.05$. For the image with sparse speckles, a few of the speckles acquire the maximum value, whereas the remaining have varying values. The goal is

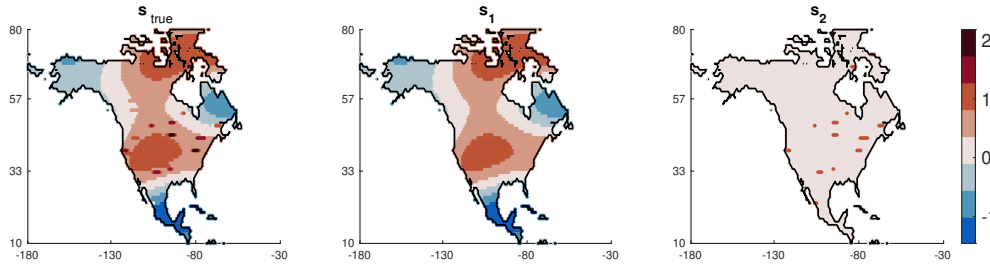


FIG. 1. Atmospheric transport example, case study 1. The true emissions image \mathbf{s}_{true} provided on the left is the sum of smooth image \mathbf{s}_1 and sparse image \mathbf{s}_2 , i.e., $\mathbf{s}_{\text{true}} = \mathbf{s}_1 + \mathbf{s}_2$. Colormaps for all images are the same.

to reconstruct the unknown set of states or fluxes in space, where the spatial resolution is $1^\circ \times 1^\circ$. This resolution is coarser than ideal for detecting superemitters but provides a nice testbed example. In this case study, the observations \mathbf{d} are sampled at the locations and times of OCO-2 observations during July through mid-August 2015, and the atmospheric model \mathbf{A} is from NOAA's CarbonTracker-Lagrange project [41, 36]. Specifically, these atmospheric modeling simulations are from the weather research and forecasting stochastic time-inverted Lagrangian transport model modeling system [35, 42]. Note that we do not use realistic CO_2 emissions in this case study (c.f., case study 3). Instead, we use randomly generated emissions to create a relatively simple, initial test case for the algorithms proposed here. We add Gaussian white noise corresponding to a 4% noise level to the observations, i.e., $\frac{\|\delta\|_2}{\|\mathbf{A}\mathbf{s}_{\text{true}}\|_2} = 0.04$.

We obtain reconstructions using the proposed **sdHybr** method and provide relative reconstruction error norms per iteration. These are computed as

$$\frac{\|\mathbf{s}_k - \mathbf{s}_{\text{true}}\|_2}{\|\mathbf{s}_{\text{true}}\|_2},$$

where \mathbf{s}_k is the reconstruction at the k th iteration. In the left plot of Figure 2, we provide relative reconstruction error norms per iteration using the optimal regularization parameter at each iteration, which is not available in practice. In the right plot, we see that similar results are obtained using the DP-selected regularization parameters. Results for **sdHybr** with WGCV and UPRE are very similar, so we do not provide them here. For comparison, we provide the relative reconstruction error norms per iteration of **genHyBR** and **fHybr** for both the optimal and DP-selected regularization parameters. For **genHyBR** and **sdHybr**, we let \mathbf{Q} represent a Matérn kernel with $\nu = 0.5$ and $\ell = 0.5$. All considered hybrid methods include a variety of stopping criteria. The tolerance for the GCV function was set to $\delta_{\text{GCV}} = 10^{-6}$ and the maximum number of iterations is 50. The diamonds denote the (automatically selected) stopping iterations. **sdHybr** with different parameter selection resulted in similar stopping points.

We observe that **sdHybr** reconstructions produce smaller relative reconstruction errors than **genHyBR** and **fHybr**, demonstrating that our solution decomposition can be beneficial. This result is also evident in the image reconstructions displayed in Figure 3. In the top row of Figure 3, we provide the overall **sdHybr-dp** reconstruction, along with the computed estimates of \mathbf{s}_1 and \mathbf{s}_2 . In the bottom row, we provide the reconstructions obtained using **genHyBR-dp** and **fHybr-dp**. We observe that the **genHyBR** reconstruction captures the smooth regions but fails to reconstruct the anomalies, while the **fHybr** reconstruction captures the anomalies but lacks smoothness in the background.

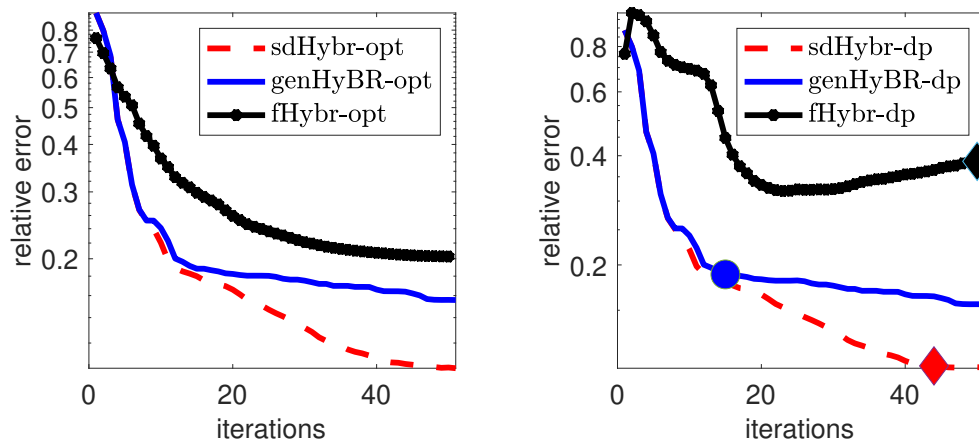


FIG. 2. Atmospheric transport, case study 1: Relative reconstruction error norms per iteration of *sdHybr*, *genHyBR*, and *fHybr*. Results in the left plot correspond to selecting the optimal regularization parameters at each iteration, and results in the right plot correspond to the DP-selected regularization parameters.

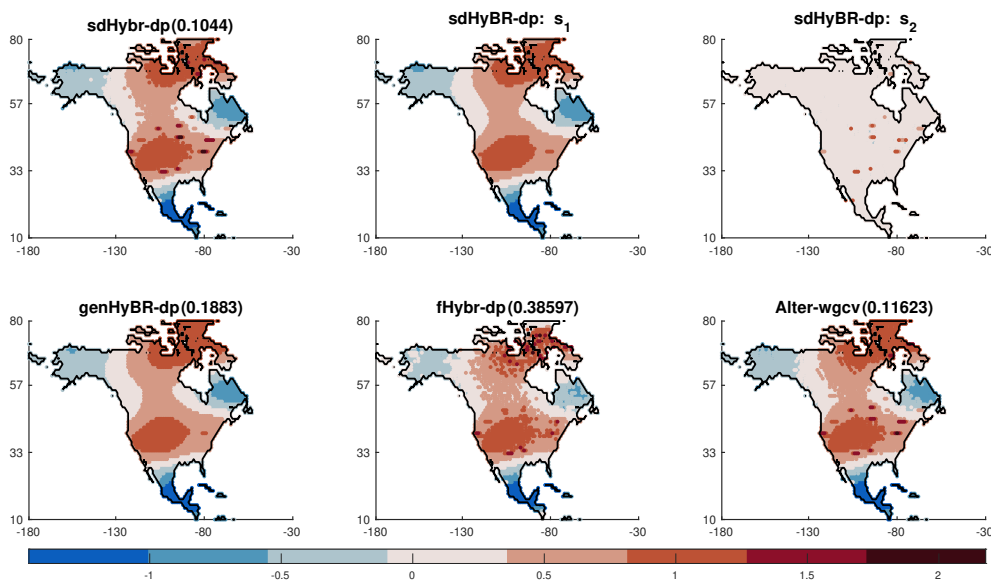


FIG. 3. Atmospheric transport, case study 1: Reconstructions with relative reconstruction error norms provided in the titles. Decomposed solutions computed using *sdHybr-dp*.

In the bottom row of Figure 3, we also provide a reconstruction using *Alter-wgcv*, which is an alternating optimization approach Algorithm 2.1. Both *sdHybr* and *Alter-wgcv* capture the sparse anomalies as well as the smooth background. The maximum number of iterations for *Alter-wgcv* was 200. We found that the *Alter* approach requires high accuracy of the algorithms used to compute solutions in the alternating framework, hence the larger number of iterations for each solve within *Alter* and the overall longer CPU times: *sdHybr* took 19.15 seconds and *Alter* took 253.97 seconds. Furthermore, we observed that *Alter* is very sensitive to the accuracy of the initial guess. Moreover, only WGCV was able to provide *Alter* reconstructions that could distinguish anomalies.

4.2. Case study 2: Dynamic spherical means tomography. In this experiment, we consider a dynamic tomography setup where the goal is to reconstruct a sequence of images from a sequence of projection datasets. Such scenarios are common in dynamic photoacoustic or dynamic electrical impedance tomography, where the underlying parameters change during the data acquisition process [49, 47, 24]. Reconstruction is particularly challenging for nonlinear or nonparametric deformations and often requires including a spatiotemporal prior [18, 46]. In spatiotemporal inversions, classic approaches (e.g., those based on parametric covariance families or separable covariance functions) may not be rich enough to capture the phenomena of interest alone, and multiple priors may be required to promote different spatiotemporal properties.

In this example, we consider a sequence of 20 true images (e.g., time points), where each image is 128×128 and represents a sum of a smooth image and a sparse image. That is, the true image at the t th time point can be represented as the sum of two images: the smooth image \mathbf{s}_1 was generated using a truncated Karhunen–Loève expansion using 30 basis vectors, with a Matérn covariance kernel defined with two spatial and one temporal dimensions (see, for example, [18]). We will refer to this as a three-dimensional (3D) Matérn kernel. We also take $\nu = 0.2$ and $\ell = 0.2$. The sparse image \mathbf{s}_2 was generated using a star cluster example, and the two images were summed together. In Figure 4, we provide three of the true image decompositions (i.e., for time points $t = 1, 10, 20$). Notice that although sparsely distributed, the spots in \mathbf{s}_2 take pixel values in a larger range than pixel values in \mathbf{s}_1 .

We consider a linear problem of the form (1.1), where

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}_1 & & \\ & \ddots & \\ & & \mathbf{F}_{20} \end{bmatrix} \in \mathbb{R}^{20 \times 181 \times 128^2},$$

where \mathbf{F}_t represents a spherical projection matrix corresponding to 20 equally spaced angles between t and $340+t$ for $t = 1, \dots, 20$, and $\mathbf{d} \in \mathbb{R}^{20 \times 181}$ contains projection data. For this, we use the `PRspherical` test problem from the `IRTools` toolbox [21, 25],

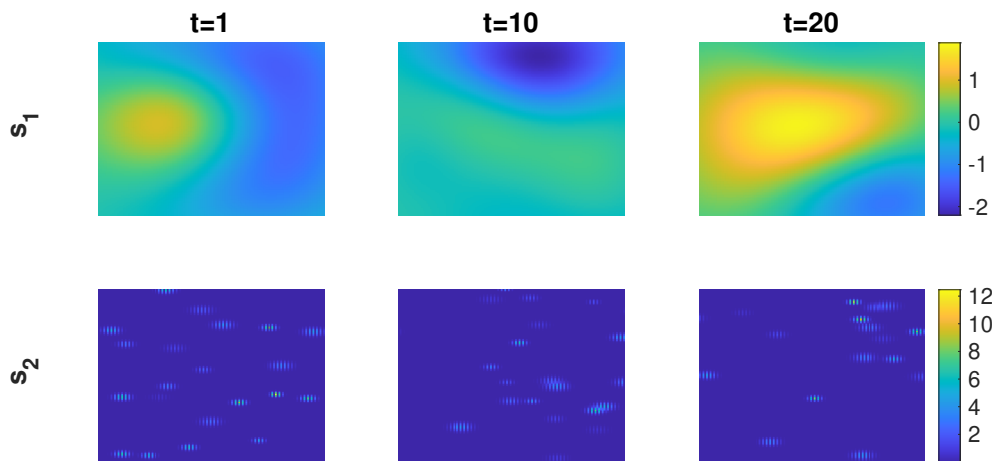


FIG. 4. *Dynamic spherical tomography example, case study 2. We provide three true images corresponding to time points $t = 1, 10, 20$. For each time point the true image is a sum of a smooth image and a sparse image.*

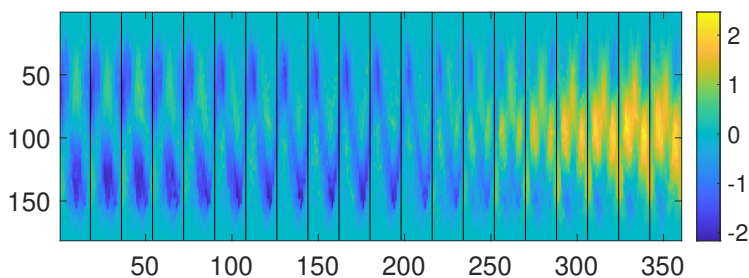


FIG. 5. *Dynamic spherical tomography example, case study 2. There are 20 observed sinograms corresponding to 20 time points.*

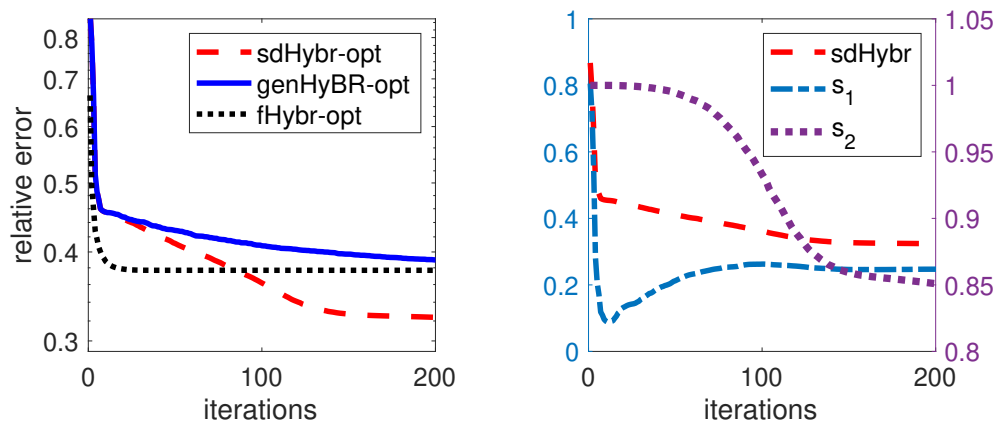


FIG. 6. *Dynamic spherical tomography example, case study 2. Relative reconstruction error norms per iteration are provided in the left plot for sdHybr-opt , genHyBR-opt , and fHybr-opt . In the right plot, we provide the relative reconstruction errors for \mathbf{s}_1 (left axis) and \mathbf{s}_2 (right axis) separately. The relative reconstruction errors for sdHybr-opt from the left plot are provided again for reference.*

and to simulate measurement error we add 2% Gaussian noise. The collection of 20 observed sinograms are concatenated and provided in Figure 5.

For the reconstructions, we used prior covariance matrix \mathbf{Q} representing a 3D Matérn kernel with $\nu = 0.5$ and $\ell = 0.4$. We focus on a comparison of methods using optimal regularization parameters, and we provide relative reconstruction error norms computed per iteration for sdHybr-opt , genHyBR-opt , and fHybr-opt in the left plot of Figure 6. We observe that sdHybr-opt can achieve smaller overall reconstruction error norms compared to genHyBR-opt and fHybr-opt . In the right plot of Figure 6, we provide relative reconstruction errors for \mathbf{s}_1 and \mathbf{s}_2 separately. An interesting observation is that in early iterations, the sdHybr-opt method seems to reconstruct better approximations of \mathbf{s}_1 , and in later iterations, reconstructions seem to capture features in \mathbf{s}_2 .

The main benefit of the solution decomposition approach can be seen in the reconstructions. For time point $t = 1$, we provide image reconstructions in Figure 7, along with the corresponding true image. We observe that sdHybr-opt can reconstruct better solutions than genHyBR-opt and fHybr-opt . Moreover, the solution decomposition approach can simultaneously solve for both components, which means that we have two separate image reconstructions \mathbf{s}_1 and \mathbf{s}_2 . Notice that the sdHybr-opt

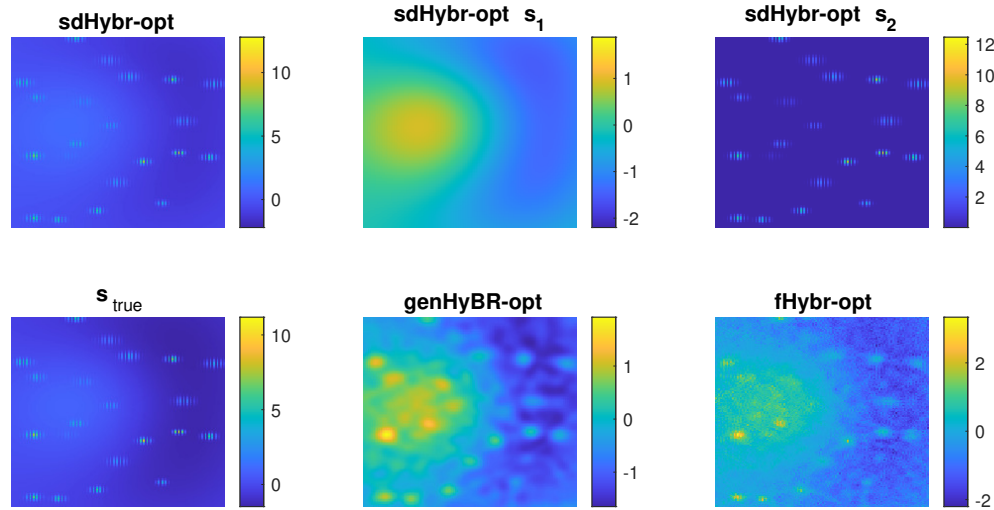


FIG. 7. *Dynamic spherical tomography example, case study 2. For time point $t = 1$, the $sdHybr-opt$ solution is provided along with reconstructed solutions s_1 and s_2 in the top row. In the bottom row, the true image along with $genHyBR$ and $fHybr$ reconstructions with the optimal regularization parameter are provided.*

solution can better capture the larger values in s_2 while simultaneously capturing the smooth features in s_1 .

We obtained similar results for the automatic parameter selection techniques, so we omit those here. Also, we acknowledge that the choice of hyperparameters will be important for the overall reconstruction. Additional investigation is necessary to determine appropriate Matérn hyperparameters for the prior covariance matrix, and this is a topic of future work.

4.3. Case study 3: Atmospheric inverse modeling based on NASA's OCO-2 satellite. For this case study, we consider a more realistic atmospheric inverse model, where the goal is to estimate CO_2 fluxes across North America using observations from NASA's OCO-2 satellite. The setup parallels that in [41, 36], so we just provide an overview here.

We consider a linear model of the form (1.1), where the aim is to estimate CO_2 fluxes at 3-hourly temporal resolution over 41 days (approximately 6 weeks from late June through July 2015) and at $1^\circ \times 1^\circ$ latitude-longitude spatial resolution. This setup corresponds to 3,222 unknowns per 3-hour time interval; hence, $\mathbf{s} \in \mathbb{R}^{328 \cdot 3222}$. For \mathbf{s}_{true} , we use CO_2 fluxes from NOAA's CarbonTracker product (version 2019b). Although a decomposition of $\mathbf{s}_{true} = \mathbf{s}_1 + \mathbf{s}_2$ is not available, we observe that, similarly to actual atmospheric models, the true fluxes contain a combination of large, sparsely distributed values which correspond to anomalies (e.g., fires, anthropogenic emissions, or anomalies in biospheric fluxes) and smooth, broad regions of surface fluxes with small-scale variability. Synthetic satellite observations given in $\mathbf{d} \in \mathbb{R}^{19,156}$ are generated as in (1.1), where \mathbf{A} represents the atmospheric transport simulation described in subsection 4.1 and $\boldsymbol{\delta}$ is added noise to represent measurement errors. The noise covariance matrix \mathbf{R} is $\sigma^2 \mathbf{I}$, where $\sigma = 0.5648$, which corresponds to a noise level of 50%. More specifically, for $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ the noise level of the observation corresponds to adding $\boldsymbol{\epsilon} = \sigma \mathbf{n}$ where $\sigma = \text{nlevel} \cdot \frac{\|\mathbf{A} \mathbf{s}_{true}\|_2}{\|\mathbf{n}\|_2}$. Notice that this inverse problem is significantly underdetermined, and thus, appropriate prior information plays a key role.

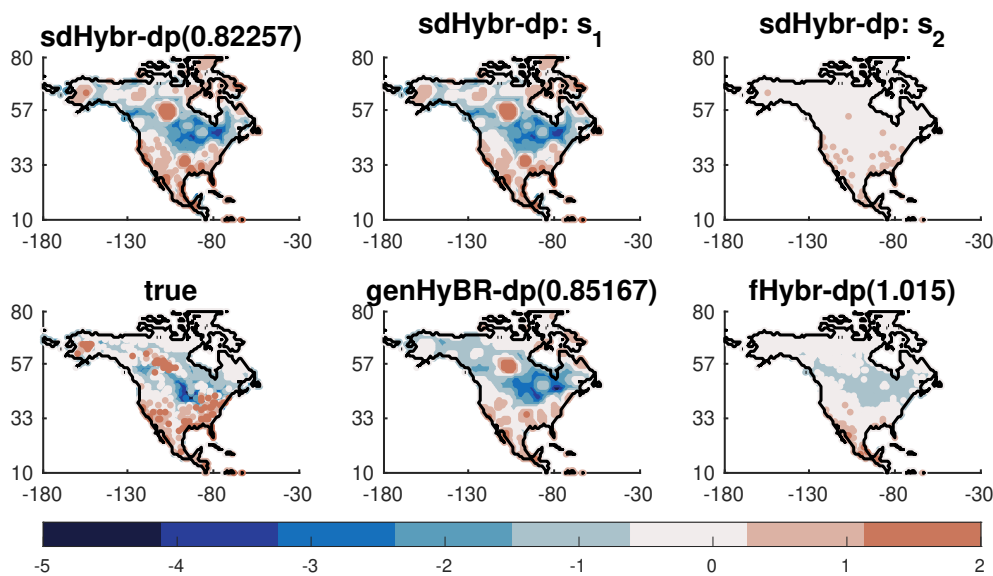


FIG. 8. *OCO-2* example, case study 3. In the top row, we provide the averaged computed reconstruction of the fluxes (in $\mu\text{mol m}^{-2}\text{s}^{-1}$) for *sdHybr*, along with the average reconstructions for s_1 and s_2 . In the bottom row, we provide the averaged true fluxes for reference, along with the reconstructions obtained using *genHyBR* and *fHybr*. Relative reconstruction error norms for the spatiotemporal fluxes are provided in the titles, and all results correspond to using the DP selected regularization parameters.

Similarly to previous approaches [52, 41], we consider a prior covariance matrix, $\mathbf{Q} = \lambda^{-2}\mathbf{Q}_t \otimes \mathbf{Q}_s$, where \mathbf{Q}_t represents the temporal covariance and \mathbf{Q}_s represents the spatial covariance in the fluxes. These covariance matrices are defined by kernel functions

$$(4.1) \quad k_t(d_t; \theta_t) = \begin{cases} 1 - \frac{3}{2} \left(\frac{d_t}{\theta_t} \right) + \frac{1}{2} \left(\frac{d_t}{\theta_t} \right)^3 & \text{if } d_t \leq \theta_t, \\ 0 & \text{if } d_t > \theta_t, \end{cases}$$

$$(4.2) \quad k_s(d_s; \theta_s) = \begin{cases} 1 - \frac{3}{2} \left(\frac{d_s}{\theta_s} \right) + \frac{1}{2} \left(\frac{d_s}{\theta_s} \right)^3 & \text{if } d_s \leq \theta_s, \\ 0 & \text{if } d_s > \theta_s, \end{cases}$$

where d_t is the day difference between two unknowns, d_s is the spherical distance between two unknowns, and θ_t, θ_s are kernel parameters. In this setting, we set $\theta_t = 9.854$ and $\theta_s = 555.42$, as in [41].

We compute spatiotemporal reconstructions using *sdHybr*, *genHyBR*, and *fHybr*, all using the automatically selected regularization parameters using the discrepancy principle. We provide in Figure 8 the overall average image of flux reconstructions, along with the average true image. We observe that *genHyBR* does fairly well at estimating the broad regions of flux estimates, and *fHybr* is not able to capture a good reconstruction. The average reconstruction of our proposed *sdHybr* method best captures both sources and sinks present in the true average image. Furthermore, a significant benefit of *sdHybr* is the ability to obtain a solution decomposition. The reconstructions of s_1 and s_2 from *sdHybr* are provided in the top row of Figure 8.

For this experiment, we remark that the 50% error level corresponds to $\sigma = 0.56$ ppm, which is aspirational for real data inverse modeling studies using *OCO-2* data

but not too far from the error levels recommended in the recent OCO-2 model intercomparison project. These values average $\sigma = 0.92$ ppm for land nadir/land glint observations and $\sigma = 0.62$ ppm for ocean glint observations. Similarly, recommended error levels for OCO-3 are $\sigma = 0.77$ ppm (averaged across all observation types) [6]. Furthermore, these noise levels may become more realistic in the future as observational and atmospheric modeling errors decline [40, 39]. Thus, these case studies allow us to explore the capabilities and behavior of the proposed inverse modeling algorithms at many different error levels.

In particular, the results of this case study demonstrate that **sdHybr** can yield accurate results for complex, spatiotemporal atmospheric inverse modeling. The proposed method enables anomaly detection, due to the simultaneous reconstruction of two separate reconstructions, one containing smooth, broad regions and another capturing sparsely distributed anomalies. Moreover, the use of efficient hybrid projection methods means that these methods can be paired with the adjoint of an atmospheric model (where explicit construction of \mathbf{A} is replaced with efficient matrix-vector products with \mathbf{A} and \mathbf{A}^\top) and with automatic selection of regularization parameters.

5. Conclusions. We have described hybrid projection methods for efficiently computing solutions to large-scale inverse problems, where the solution consists of two characteristically different solutions. We focus on the scenario where the desired solution is a sum of a sparse solution and a smooth solution (e.g., a framework that can be used for anomaly detection) and describe an FGGK hybrid iterative approach that confers several advantages. The approaches are *efficient*, in part because they converge quickly, they exploit efficient matrix-vector multiplications, and they avoid expensive inner-outer optimization schemes by leveraging recent work on flexible preconditioning techniques. These methods are also *automatic* since hyperparameters and stopping criteria can be determined as part of the iterative algorithm. We describe various problems and alternative formulations that also fit within our framework, so these methods can be utilized for a wide range of problems. Numerical results from various applications, including dynamic inverse problems and atmospheric inverse modeling, demonstrate the benefits and potential for our approach.

Future work includes extending and applying the proposed method for real satellite data, that when coupled with the GEOS-Chem chemical transport model, requires a reformulation to handle the adjoint. In addition, techniques for uncertainty quantification can be extended to this framework by utilizing a linearization approach to approximate the posterior at the MAP estimate with a Gaussian distribution [3, 5] or by exploiting previously computed bases for the solution subspaces as described in [45].

Appendix A. Heuristics for selecting regularization parameters. In this section, we describe various methods for selecting regularization parameters λ and α for the projected problem (3.13). Notice that the solution to this problem is given by

$$\mathbf{f}_k(\lambda, \alpha) = (\mathbf{M}_k^\top \mathbf{M}_k + \lambda^2 \mathbf{I} + \alpha^2 \mathbf{R}_{W,k}^\top \mathbf{R}_{W,k})^{-1} \mathbf{M}_k^\top m_{1,1} \mathbf{e}_1.$$

We write $\mathbf{f}_k(\lambda, \alpha)$ to denote the explicit dependence of \mathbf{f}_k on the regularization parameters λ and α .

Let $\mathbf{C}_k(\lambda, \alpha) = (\mathbf{M}_k^\top \mathbf{M}_k + \lambda^2 \mathbf{I} + \alpha^2 \mathbf{R}_{W,k}^\top \mathbf{R}_{W,k})^{-1} \mathbf{M}_k^\top$, and denote the projected residual as

$$\mathbf{r}_k^{\text{proj}}(\lambda, \alpha) = \mathbf{M}_k \mathbf{f}_k(\lambda, \alpha) - m_{1,1} \mathbf{e}_1.$$

We use hybrid regularization techniques to estimate the regularization parameters in the projected space similarly to [13]. However, the key difference is that (2.3) is no longer a Tikhonov type problem, so the techniques do not apply directly. Therefore, the parameter selection techniques used on the projected problem we present below are heuristics inspired by [13].

In order to provide a benchmark for comparing the parameter selection methods, we define the optimal parameters as

$$(A.1) \quad (\lambda^{\text{proj}}, \alpha^{\text{proj}}) = \arg \min_{\lambda, \alpha} \|\mathbf{s}_k(\lambda, \alpha) - \mathbf{s}_{\text{true}}\|_2^2,$$

where \mathbf{s}_{true} is the true solution that is not available in practice. As was mentioned, we use it merely to test the performance of the parameter selection methods. We can select parameters λ, α using the *UPRE method* for the projection problem, where

$$(A.2) \quad (\lambda^{\text{proj}}, \alpha^{\text{proj}}) = \arg \min_{\lambda, \alpha} \frac{1}{k} \left\| \mathbf{r}_k^{\text{proj}}(\lambda, \alpha) \right\|_2^2 + \frac{2}{k} \text{tr}(\mathbf{M}_k \mathbf{C}_k(\lambda, \alpha)) - 1.$$

Notice that the noise level of the problem should be included in the definition of \mathbf{R} . Another common approach is to use the *DP*, where parameters λ, α are selected such that

$$(A.3) \quad (\lambda^{\text{proj}}, \alpha^{\text{proj}}) = \arg \min_{\lambda, \alpha} \left| \left\| \mathbf{r}_k^{\text{proj}}(\lambda, \alpha) \right\|_2^2 - m\tau \right|,$$

where $\tau \geq 1$ is a safety factor. Without a priori knowledge of the noise level, another option is to use an extension of the *WGCV* method. The basic idea is to select parameters

$$(A.4) \quad (\lambda^{\text{proj}}, \alpha^{\text{proj}}) = \arg \min_{\lambda, \alpha} \frac{\left\| \mathbf{r}_k^{\text{proj}}(\lambda, \alpha) \right\|_2^2}{(\text{tr}(\mathbf{I}_k - \omega \mathbf{M}_k \mathbf{C}_k(\lambda, \alpha)))^2},$$

where $\omega = k/m$ [43]. Various parameter choice methods can be used here since the projected problem is small. We point the interested reader to other works on regularization parameter selection in the context of hybrid methods [14].

Motivated by the approaches described in [15, 16], we introduce three stopping criteria for the FGGK process in the solution decomposition hybrid approach. The iterative process is terminated if either of these conditions is satisfied: (i) a maximum number of iterations is attained; (ii) the GCV function defined in terms of the iteration,

$$(A.5) \quad \widehat{G}(k) = \frac{k \left\| \mathbf{r}_k^{\text{proj}}(\lambda, \alpha) \right\|_2^2}{(\text{tr}(\mathbf{I}_k - \mathbf{M}_k \mathbf{C}_k(\lambda, \alpha)))^2}$$

reaches the minimum or flattens out. In addition to these criteria, one could also consider stopping if the gradient of the objective function (3.3) is sufficiently small; however, we did not need this in our implementation.

Appendix B. Alternative to FGGK. Here we briefly derive an alternative to FGGK. Recall that we have to solve the sequence of systems (3.4) where now $\mathbf{D}_k = \mathbf{D}(\boldsymbol{\xi}^{(k)})$. Let $\mathbf{y} = \mathbf{D}_k \boldsymbol{\xi}$, so that the optimization problem can be recast as

$$(B.1) \quad \min_{\mathbf{x}, \mathbf{y}} \left\| \widehat{\mathbf{A}} \widehat{\mathbf{Q}} \begin{bmatrix} \mathbf{I} & \\ & \mathbf{D}_k^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \mathbf{c} \right\|_2^2 + \left\| \begin{bmatrix} \lambda \mathbf{I} & \\ & \alpha \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \right\|_{\widehat{\mathbf{Q}}}^2,$$

Algorithm B.1 Alternative flexible, generalized Golub–Kahan process

```

1: Initialize  $\mathbf{u}_1 = \mathbf{c}/\|\mathbf{c}\|_{\mathbf{R}^{-1}}$ 
2: for  $i = 1, \dots, k$  do
3:    $\mathbf{w} = \widehat{\mathbf{A}}^\top \mathbf{R}^{-1} \mathbf{u}_i$ ,  $h_{j,i} = \mathbf{w}^\top \widehat{\mathbf{Q}} \mathbf{v}_j$  for  $j = 1, \dots, i-1$ 
4:    $\mathbf{w} = \mathbf{w} - \sum_{j=1}^{i-1} h_{j,i} \mathbf{v}_j$ ,  $h_{i,i} = \|\mathbf{w}\|_{\widehat{\mathbf{Q}}}$ ,  $\mathbf{v}_i = \mathbf{w}/h_{i,i}$ 
5:    $\mathbf{z}_i = \begin{bmatrix} \mathbf{I} \\ \mathbf{D}_i^{-1} \end{bmatrix} \mathbf{v}_i$ 
6:    $\mathbf{w} = \widehat{\mathbf{A}} \widehat{\mathbf{Q}} \mathbf{z}_i$ ,  $g_{j,i} = \mathbf{w}^\top \mathbf{R}^{-1} \mathbf{u}_j$  for  $j = 1, \dots, i$ 
7:    $\mathbf{w} = \mathbf{w} - \sum_{j=1}^i g_{j,i} \mathbf{y}_j$ ,  $g_{i+1,i} = \|\mathbf{w}\|_{\mathbf{R}^{-1}}$ ,  $\mathbf{u}_{i+1} = \mathbf{w}/g_{i+1,i}$ 
8: end for

```

where $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{Q}}$ were defined in (3.7). We can derive a similar process as before, where after k iterations of this process, we obtain

$$(B.2) \quad \widehat{\mathbf{A}} \widehat{\mathbf{Q}} \mathbf{Z}_k = \mathbf{U}_{k+1} \mathbf{G}_k \quad \text{and} \quad \widehat{\mathbf{A}}^\top \mathbf{U}_{k+1} = \mathbf{V}_{k+1} \mathbf{H}_{k+1},$$

where

- $\mathbf{Z}_k = [\mathbf{z}_1 \ \dots \ \mathbf{z}_k] = [\mathbf{L}_1 \mathbf{v}_1 \ \dots \ \mathbf{L}_k \mathbf{v}_k] \in \mathbb{R}^{2n \times k}$ with $\mathbf{L}_i = \begin{bmatrix} \mathbf{I} \\ \mathbf{D}_i^{-1} \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$ contains solution basis vectors;
- $\mathbf{G}_k = [g_{i,j}]_{i=1, \dots, k+1; j=1, \dots, k} \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg;
- $\mathbf{H}_{k+1} = [h_{i,j}]_{i,j=1, \dots, k+1} \in \mathbb{R}^{(k+1) \times (k+1)}$ is upper triangular;
- $\mathbf{U}_{k+1} = [\mathbf{u}_1 \ \dots \ \mathbf{u}_{k+1}] \in \mathbb{R}^{m \times (k+1)}$ with $\mathbf{u}_1 = \mathbf{c}/\|\mathbf{c}\|_{\mathbf{R}^{-1}}$; and
- $\mathbf{V}_{k+1} = [\mathbf{v}_1 \ \dots \ \mathbf{v}_{k+1}] \in \mathbb{R}^{2n \times (k+1)}$ such that

$$\mathbf{U}_{k+1}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} = \mathbf{I}_{k+1}, \quad \mathbf{V}_{k+1}^\top \widehat{\mathbf{Q}} \mathbf{V}_{k+1} = \mathbf{I}_{k+1}.$$

The specific algorithm is given in Algorithm B.1.

Next, assume that we seek solution $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathcal{R}(\mathbf{Z}_k)$, i.e., $\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \mathbf{Z}_k \mathbf{f}$ for some $\mathbf{f} \in \mathbb{R}^k$.

Then the projected problem

$$(B.3) \quad \min_{\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \in \mathcal{R}(\mathbf{Z}_k)} \left\| \widehat{\mathbf{A}} \widehat{\mathbf{Q}} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} - \mathbf{c} \right\|_2^2 + \left\| \begin{bmatrix} \lambda \mathbf{I} & \\ & \alpha \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \right\|_{\widehat{\mathbf{Q}}}^2$$

becomes

$$\min_{\mathbf{f} \in \mathbb{R}^k} \left\| \mathbf{G}_k \mathbf{f} - \|\mathbf{c}\|_{\mathbf{R}^{-1}} \mathbf{e}_1 \right\|_2^2 + \left\| \begin{bmatrix} \lambda \mathbf{I} & \\ & \alpha \mathbf{I} \end{bmatrix} \mathbf{Z}_k \mathbf{f} \right\|_{\widehat{\mathbf{Q}}}^2.$$

The important point here is that, while the projected problem is in a smaller dimensional space, it is not obvious how to simultaneously estimate the parameters λ and α . One choice would be to assume $\alpha = \lambda \lambda_\alpha$, where λ_α is fixed and estimate λ using techniques similar to Appendix A. However, the choice we made in subsection 3.2 allows for the estimation of both parameters within the projected space, while maintaining the same computational cost. This highlights the novelty of the proposed approach.

REFERENCES

- [1] S. AGAPIOU, J. M. BARDSLEY, O. PAPASPILIOPOULOS, AND A. M. STUART, *Analysis of the Gibbs sampler for hierarchical inverse problems*, SIAM/ASA J. Uncertain. Quantif., 2 (2014), pp. 511–544.
- [2] S. AMBIKASARAN, J. LI, P. KITANIDIS, AND E. DARVE, *Large-scale stochastic linear inversion using hierarchical matrices*, Comput. Geosci., 17 (2012), pp. 913–927.
- [3] J. M. BARDSLEY, *Computational Uncertainty Quantification for Inverse Problems*, Comput. Sci. Eng. 19, SIAM, Philadelphia, 2018.
- [4] A. BECK AND M. TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [5] T. BUI-THANH, O. GHATTAS, J. MARTIN, AND G. STADLER, *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 (2013), pp. A2494–A2523.
- [6] B. BYRNE, ET AL., *National CO2 budgets (2015-2020) inferred from atmospheric CO2 observations in support of the global stocktake*, Earth Syst. Sci. Data Discuss., 15 (2023), pp. 963–1004.
- [7] D. CALVETTI, M. PRAGLIOLA, AND E. SOMERSALO, *Sparsity promoting hybrid solvers for hierarchical Bayesian inverse problems*, SIAM J. Sci. Comput., 42 (2020), pp. A3761–A3784.
- [8] D. CALVETTI, M. PRAGLIOLA, E. SOMERSALO, AND A. STRANG, *Sparse reconstructions from few noisy data: Analysis of hierarchical Bayesian models with generalized gamma hyperpriors*, Inverse Problems, 36 (2020), 025010.
- [9] D. CALVETTI AND E. SOMERSALO, *An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Vol. 2, Springer, New York, 2007.
- [10] D. CALVETTI AND E. SOMERSALO, *Subjective knowledge or objective belief? An oblique look to Bayesian methods*, in Large-Scale Inverse Problems and Quantification of Uncertainty, Wiley, Chichester, UK, 2010, pp. 33–70.
- [11] D. CALVETTI, E. SOMERSALO, AND A. STRANG, *Hierarchical Bayesian models and sparsity: ℓ_2 -magic*, Inverse Problems, 35 (2019), 035003.
- [12] D.-H. CHEN, B. HOFMANN, AND J. ZOU, *Elastic-net regularization versus ℓ_1 -regularization for linear inverse problems with quasi-sparse solutions*, Inverse Problems, 33 (2016), 015004.
- [13] T. CHO, J. CHUNG, AND J. JIANG, *Hybrid projection methods for large-scale inverse problems with mixed Gaussian priors*, Inverse Problems, 37 (2021), 044002.
- [14] J. CHUNG AND S. GAZZOLA, *Flexible Krylov methods for ℓ_p regularization*, SIAM J. Sci. Comput., 41 (2019), pp. S149–S171.
- [15] J. CHUNG, J. G. NAGY, AND D. P. O’LEARY, *A weighted GCV method for Lanczos hybrid regularization*, Electron. Trans. Numer. Anal., 28 (2008), pp. 149–167.
- [16] J. CHUNG AND K. PALMER, *A hybrid LSMR algorithm for large-scale Tikhonov regularization*, SIAM J. Sci. Comput., 37 (2015), pp. S562–S580.
- [17] J. CHUNG AND A. K. SAIBABA, *Generalized hybrid iterative methods for large-scale Bayesian inverse problems*, SIAM J. Sci. Comput., 39 (2017), pp. S24–S46.
- [18] J. CHUNG, A. K. SAIBABA, M. BROWN, AND E. WESTMAN, *Efficient generalized Golub-Kahan based methods for dynamic inverse problems*, Inverse Problems, 34 (2018), 024005.
- [19] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pure Appl. Math., 57 (2004), pp. 1413–1457.
- [20] M. FORNASIER, S. PETER, H. RAUHUT, AND S. WORM, *Conjugate gradient acceleration of iteratively re-weighted least squares methods*, Comput. Optim. Appl., 65 (2016), pp. 205–259.
- [21] S. GAZZOLA, P. C. HANSEN, AND J. G. NAGY, *IR Tools: A MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 81 (2019), pp. 773–811.
- [22] S. GAZZOLA AND J. G. NAGY, *Generalized Arnoldi-Tikhonov method for sparse reconstruction*, SIAM J. Sci. Comput., 36 (2014), pp. B225–B247.
- [23] I. F. GORODNITSKY AND B. D. RAO, *A new iterative weighted norm minimization algorithm and its applications*, in IEEE Sixth SP Workshop on Statistical Signal and Array Processing, IEEE, Piscataway, NJ, 1992, pp. 412–415.
- [24] B. N. HAHN, *Efficient algorithms for linear dynamic inverse problems with known motion*, Inverse Problems, 30 (2014), 035008.
- [25] P. C. HANSEN AND J. S. JØRGENSEN, *AIR tools II: Algebraic iterative reconstruction methods, improved implementation*, Numer. Algorithms, 79 (2018), pp. 107–137.
- [26] Y. HUANG AND Z. JIA, *Some results on the regularization of LSQR for large-scale discrete ill-posed problems*, Sci. China Math., 60 (2017), pp. 701–718.

- [27] B. JIN, D. A. LORENZ, AND S. SCHIFFLER, *Elastic-net regularization: Error estimates and active set methods*, *Inverse Problems*, 25 (2009), 115022.
- [28] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160, Springer, New York, 2006.
- [29] J. P. KAIPIO AND E. SOMERSALO, *Estimating anomalies from indirect observations*, *J. Comput. Phys.*, 181 (2002), pp. 398–406.
- [30] P. K. KITANIDIS, *Parameter uncertainty in estimation of spatial functions: Bayesian analysis*, *Water Resour. Res.*, 22 (1986), pp. 499–507.
- [31] P. K. KITANIDIS, *Quasi-linear geostatistical theory for inverting*, *Water Resour. Res.*, 31 (1995), pp. 2411–2419.
- [32] P. K. KITANIDIS AND E. G. VOUMVORIS, *A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations*, *Water Resour. Res.*, 19 (1983), pp. 677–690.
- [33] K. LANGE, *MM Optimization Algorithms*, SIAM, Philadelphia, 2016.
- [34] Z. LI, A. DESOLNEUX, S. MULLER, AND A.-K. CARTON, *A novel 3D stochastic solid breast texture model for x-ray breast imaging*, in *International Workshop on Breast Imaging*, Springer, Cham, Switzerland, 2016, pp. 660–667.
- [35] J. C. LIN, C. GERBIG, S. C. WOFYSY, A. E. ANDREWS, B. C. DAUBE, K. J. DAVIS, AND C. A. GRAINGER, *A near-field tool for simulating the upstream influence of atmospheric observations: The stochastic time-inverted Lagrangian transport (STILT) model*, *J. Geophys. Res. Atmos.*, 108 (2003).
- [36] X. LIU, A. L. WEINBREN, H. CHANG, J. TADIĆ, M. E. MOUNTAIN, M. E. TRUDEAU, A. E. ANDREWS, Z. CHEN, AND S. M. MILLER, *Data reduction for inverse modeling: An adaptive approach v1.0*, *Geosci. Model Dev.*, 14 (2021), pp. 1–21.
- [37] G. MATHERON, *The intrinsic random functions and their applications*, *Adv. Appl. Probab.*, 5 (1973), pp. 439–468.
- [38] H. MATSUI AND S. KONISHI, *Variable selection for functional regression models via the l1 regularization*, *Comput. Statist. Data Anal.*, 55 (2011), pp. 3304–3310.
- [39] S. M. MILLER AND A. M. MICHALAK, *The impact of improved satellite retrievals on estimates of biospheric carbon balance*, *Atmos. Chem. Phys.*, 20 (2020), pp. 323–331.
- [40] S. M. MILLER, A. M. MICHALAK, V. YADAV, AND J. M. TADIĆ, *Characterizing biospheric carbon balance using CO₂ observations from the OCO-2 satellite*, *Atmos. Chem. Phys.*, 18 (2018), pp. 6785–6799.
- [41] S. M. MILLER, A. K. SAIBABA, M. E. TRUDEAU, M. E. MOUNTAIN, AND A. E. ANDREWS, *Geostatistical inverse modeling with very large datasets: An example from the orbiting carbon observatory 2 (OCO-2) satellite*, *Geosci. Model Dev.*, 13 (2020), pp. 1771–1785.
- [42] T. NEHRKORN, J. ELUSZKIEWICZ, S. WOFYSY, J. LIN, C. GERBIG, M. LONGO, AND S. FREITAS, *Coupled weather research and forecasting-stochastic time-inverted Lagrangian transport (WRF-STILT) model*, *Meteorol. Atmos. Phys.*, 107 (2010), pp. 51–64.
- [43] R. A. RENAUT, S. VATANKHAH, AND V. E. ARDESTANI, *Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems*, *SIAM J. Sci. Comput.*, 39 (2017), pp. B221–B243.
- [44] P. RODRIGUEZ AND B. WOHLBERG, *An efficient algorithm for sparse representations with ℓ^p data fidelity term*, in *Proceedings of 4th IEEE Andean Technical Conference (ANDESCON)*, IEEE, Piscataway, NJ, 2008.
- [45] A. K. SAIBABA, J. CHUNG, AND K. PETROSKE, *Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems*, *Numer. Linear Algebra Appl.*, 27 (2020), e2325.
- [46] U. SCHMITT AND A. K. LOUIS, *Efficient algorithms for the regularization of dynamic inverse problems: I. Theory*, *Inverse Problems*, 18 (2002), p. 645–658.
- [47] U. SCHMITT, A. K. LOUIS, C. H. WOLTERS, AND M. VAUHKONEN, *Efficient algorithms for the regularization of dynamic inverse problems: II. Applications*, *Inverse Problems*, 18 (2002), p. 659–676.
- [48] B. TRÉMOULHÉAC, N. DIKAIOS, D. ATKINSON, AND S. R. ARRIDGE, *Dynamic MR image reconstruction-separation from undersampled (k,t) -space via low-rank plus sparse prior*, *IEEE Trans. Med. Imaging*, 33 (2014), pp. 1689–1701.
- [49] K. WANG, J. XIA, C. LI, L. V. WANG, AND M. A. ANASTASIO, *Fast spatiotemporal image reconstruction based on low-rank matrix estimation for dynamic photoacoustic computed tomography*, *J. Biomed. Opt.*, 19 (2014), 056007.
- [50] C. K. WILLIAMS AND C. E. RASMUSSEN, *Gaussian Processes for Machine Learning*, Vol. 2, MIT Press, Cambridge, MA, 2006.

- [51] K. XIONG, G. ZHAO, G. SHI, AND Y. WANG, *A convex optimization algorithm for compressed sensing in a complex domain: The complex-valued split Bregman method*, *Sensors*, 19 (2019), 4540.
- [52] V. YADAV, A. M. MICHALAK, J. RAY, AND Y. P. SHIGA, *A statistical approach for isolating fossil fuel emissions in atmospheric inverse problems*, *J. Geophys. Res. Atmosph.*, 121 (2016), pp. 12490–12505.
- [53] P. YIN, Y. LOU, Q. HE, AND J. XIN, *Minimization of ℓ_{1-2} for compressed sensing*, *SIAM J. Sci. Comput.*, 37 (2015), pp. A536–A563.
- [54] H. ZOU AND T. HASTIE, *Regularization and variable selection via the elastic net*, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 67 (2005), pp. 301–320.