

LAP: A Linearize and Project Method for Solving Inverse Problems with Coupled Variables

James L. Herring

Department of Mathematics and Computer Science, Emory University
Atlanta, Georgia 30322, USA
herrinj@gmail.com

James G. Nagy

Department of Mathematics and Computer Science, Emory University
Atlanta, Georgia 30322, USA
jnagy@emory.edu

Lars Ruthotto

Department of Mathematics and Computer Science, Emory University
Atlanta, Georgia 30322, USA
lruthotto@emory.edu

Abstract. Many inverse problems involve two or more sets of variables that represent different physical quantities but are tightly coupled with each other. For example, image super-resolution requires joint estimation of the image and motion parameters from noisy measurements. Exploiting this structure is key for efficiently solving these large-scale optimization problems, which are often ill-conditioned.

In this paper, we present a new method called Linearize And Project (LAP) that offers a flexible framework for solving inverse problems with coupled variables. LAP is most promising for cases when the subproblem corresponding to one of the variables is considerably easier to solve than the other. LAP is based on a Gauss–Newton method, and thus after linearizing the residual, it eliminates one block of variables through projection. Due to the linearization, this block can be chosen freely. Further, LAP supports direct, iterative, and hybrid regularization as well as constraints. Therefore LAP is attractive, e.g., for ill-posed imaging problems. These traits differentiate LAP from common alternatives for this type of problem such as variable projection (VarPro) and block coordinate descent (BCD). Our numerical experiments compare the performance of LAP to BCD and VarPro using three coupled problems whose forward operators are linear with respect to one block and nonlinear for the other set of variables.

Key words and phrases : Nonlinear Least-Squares, Gauss–Newton Method, Inverse Problems, Regularization, Image Processing, Variable Projection

2010 AMS Mathematics Subject Classification — 65F10, 65F22, 65M32

1. Introduction

We present an efficient Gauss–Newton method called Linearize And Project (LAP) for solving large-scale optimization problems whose variables consist of two or more blocks representing, e.g., different physics. Problems with these characteristics arise, e.g., when jointly reconstructing image and motion parameters from a series of noisy, indirect, and transformed measurements. LAP is motivated by problems in which the blocks of variables are nontrivially coupled with each other, but some of the blocks lead to well-conditioned and easy-to-solve subproblems. As two examples of such problems arising in imaging, we consider super resolution [29, 10, 4] and motion corrected Magnetic Resonance Imaging (MRI) [1, 6].

A general approach to solving coupled optimization problems is to use alternating minimization strategies such as Block Coordinate Descent (BCD) [23, 26]. These straightforward approaches can be applied to most objective functions and constraints and also provide flexibility for using various regularization strategies. However, alternating schemes have been shown to converge slowly for problems with tightly coupled blocks [26, 4].

One specific class of coupled optimization problems, which has received much attention, is separable nonlinear least-squares problems; see, e.g., [18, 17, 27]. Here, the variables can be partitioned such that the residual function is linear in one block and nonlinear in the other. For brevity, we will refer to the sets of variables as *linear* and *nonlinear block of variables*, respectively. One common method for solving such problems is Variable Projection (VarPro) [18, 17, 27]. The idea is to derive a nonlinear least-squares problem of reduced size by eliminating the the linear block of variables through projections. VarPro is most effective when the projection, which entails solving a linear least-squares problem, can be computed cheaply and accurately. When the number of linear variables in the problem is large, iterative methods can be used to compute the projection [11], but iterative methods can become inefficient when the least-squares problem is ill-posed. Further, standard iterative methods do not provide bounds on the optimality, which are needed to ensure well-defined gradients; see our discussion in Sec. 2.2. We note that recently proposed methods such as in [9] can provide that information and are, thus, attractive options in these cases. Another limitation of VarPro is that it is not straightforward to incorporate inequality constraints on the linear variables. Some progress has been made for box-constraints using a pseudo-derivative approach [34] leading to approximate gradients for the nonlinear problem. Finally, VarPro limits the options for adaptive regularization parameter selection strategies for the least-squares problem.

As an alternative to the methods above, we propose the LAP method for solving coupled optimization problems. Our contributions can be summarized as follows:

- We propose an efficient iterative method called LAP that computes the Gauss–Newton step at each iteration by eliminating one block of variables

through projection and solving the reduced problem iteratively. Since projection is performed after linearization, any block can be eliminated. Hence the LAP framework offers superior flexibility compared to existing projection-based approaches, e.g., by supporting various types of regularization strategies and the option to impose constraints for all variables.

- We demonstrate LAP’s flexibility for different regularization strategies including Tikhonov regularization using the discrete gradient operator with a fixed regularization parameter and a hybrid regularization approach [5], which simultaneously computes the search direction and selects an appropriate regularization parameter at each iteration.
- We use projected Gauss–Newton to implement element-wise lower and upper bound constraints with LAP on the optimization variables. This is a distinct advantage over VarPro, where previously proposed methods for incorporating inequality constraints require using approximate gradients.
- We present numerical experiments for several separable nonlinear least-squares problems. The problems are characterized by linear imaging models and nonlinear motion models with applications including 2D and 3D super-resolution and motion correction for magnetic resonance imaging. We compare the performance of LAP with block coordinate descent (BCD) and variable projection (VarPro) by analyzing convergence, CPU timings, number of matrix-vector multiplications, and the solution images and motion parameters.
- We provide our MATLAB implementation of LAP including the examples used in this paper freely at

<https://github.com/herrinj/LAP>

Our paper is organized as follows: Sec. 2 introduces a general formulation of the motion-corrected imaging problem, which we use to motivate and illustrate LAP, and briefly reviews BCD and VarPro; Sec. 3 explains our proposed scheme, LAP, for the coupled Gauss–Newton iteration along with a discussion of regularization options and implementation of bound constraints using projected Gauss–Newton; and Sec. 4 provides experimental results for several examples using LAP and compares it with BCD and VarPro. We end with a brief summary and some concluding remarks.

2. Motion-Corrected Imaging Problem

In this section, we give a general description of coupled optimization problems arising in image reconstruction from motion affected measurements and briefly review Block Coordinate Descent (BCD) and Variable Projection (VarPro).

We follow the guidelines in [25], and consider images as continuously differentiable and compactly supported functions on a domain of interest, $\Omega \subset \mathbb{R}^d$ (typically, $d = 2$ or 3). We assume that the image attains values in a field \mathbb{F} where $\mathbb{F} = \mathbb{R}$ corresponds to real-valued and $\mathbb{F} = \mathbb{C}$ to complex-valued images. We denote by $\mathbf{x} \in \mathbb{F}^n$ a discrete image obtained by evaluating a continuous image at the cell-centers of a rectangular grid with n cells.

The discrete transformation $\mathbf{y} \in \mathbb{R}^{d \cdot n}$ is obtained by evaluating a function $y : \Omega \rightarrow \mathbb{R}^d$ at the cell-centers and can be visualized as a transformed grid. For the rigid, affine transformations of primary interest in this paper, transformations are comprised of shifts and rotations and can be defined by a small set of parameters denoted by the variable \mathbf{w} , but in general, the number of parameters defining a transformation may be large. We observe that under a general transformation $\mathbf{y}(\mathbf{w})$, the cell-centers of a transformed grid do not align to the cell-centers of the original grid, so to evaluate a discretized image under a transformation \mathbf{y} , we must interpolate using the known image coefficients \mathbf{x} . This interpolation can be represented via a sparse matrix $T(\mathbf{y}(\mathbf{w})) \in \mathbb{R}^{n \times n}$ determined by the transformation. For the examples in this paper, we use bilinear or trilinear interpolation corresponding to the dimension of the problem, but other alternatives are possible; see, e.g., [25] for alternatives. The transformed version of the discrete image \mathbf{x} is then expressed as a matrix-vector product $T(\mathbf{y}(\mathbf{w}))\mathbf{x}$.

Using the above definitions of discrete images and their transformations, we then consider the discrete, forward problem for N distinct data observations,

$$\mathbf{d}_k = K_k T(\mathbf{y}(\mathbf{w}_k)) \mathbf{x} + \epsilon_k, \quad \text{for all } k = 1, 2, \dots, N, \quad (1)$$

where $\mathbf{d}_k \in \mathbb{F}^{m_k}$ is the measured data, $K_k \in \mathbb{F}^{m_k \times n}$ is a matrix corresponding to the problem-specific image operator, and ϵ_k is image noise, which we assume to be independently and identically distributed Gaussian noise.

In this paper, we focus on the case in which the motion can be modeled by a small number of parameters. In our case, $\mathbf{w}_k \in \mathbb{R}^3$ or \mathbb{R}^6 models rigid transformations in 2D and 3D, respectively. The total dimension of the motion parameters across all data measurements is then given by $p = 3N$ or $p = 6N$ for $d = 2$ and $d = 3$, respectively. In the application at hand, we note that $p \ll n$. To simplify our notation, we use the column vectors $\mathbf{d} \in \mathbb{F}^m$ where $m = m_k \cdot N$ and $\mathbf{w} \in \mathbb{R}^p$ to denote the data and motion parameters for all N measurements, respectively.

Given a set of measurements $\{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N\}$, the motion-corrected imaging problem consists of jointly estimating the underlying image parameters and the motion parameters in (1). We formulate this as the coupled optimization problem

$$\min_{\mathbf{x} \in \mathcal{C}_x, \mathbf{w} \in \mathcal{C}_w} \Phi(\mathbf{x}, \mathbf{w}) = \frac{1}{2} \|\mathbf{KT}(\mathbf{w})\mathbf{x} - \mathbf{d}\|^2 + \frac{\alpha}{2} \|\mathbf{Lx}\|_2^2, \quad (2)$$

where the matrices \mathbf{K} and \mathbf{T} have the following block structure

$$\mathbf{K} = \begin{bmatrix} K_1 & & & \\ & K_2 & & \\ & & \ddots & \\ & & & K_N \end{bmatrix} \quad \text{and} \quad \mathbf{T}(\mathbf{w}) = \begin{bmatrix} T(\mathbf{y}(\mathbf{w}_1)) \\ T(\mathbf{y}(\mathbf{w}_2)) \\ \vdots \\ T(\mathbf{y}(\mathbf{w}_N)) \end{bmatrix}.$$

We denote by $\mathcal{C}_x \subset \mathbb{F}^n$ and $\mathcal{C}_w \subset \mathbb{R}^p$ rectangular sets used to impose bound constraints on the image and motion parameters. Lastly, we regularize the problem by adding a chosen regularization operator, \mathbf{L} (e.g., a discrete image gradient

or the identity matrix) and a regularization parameter, $\alpha > 0$, that balances minimizing the data misfit and the regularity of the reconstructed image. We note that finding a good regularization parameter α is a separate, challenging problem which has been widely researched [5, 7, 20, 35]. One strength of LAP is that, for $\mathbf{L} = \mathbf{I}$, it allows for regularization methods that automatically select the α parameter [5, 12, 14, 15]. In our numerical experiments, we investigate one such hybrid method for automatic regularization parameter selection [5] as well as direct regularization using a fixed α value.

Problems of the form (2) are often referred to as separable nonlinear least-squares problems. They are a specific class of coupled optimization problems characterized by being nonlinear in one block of variables, \mathbf{w} , and linear in the other, \mathbf{x} . Several optimization approaches exist for solving such problems. One option is a fully coupled Gauss–Newton approach to optimize over both sets of variables simultaneously by solving a single linear system. However, this approach does not exploit the convexity of the problem in the image variables, resulting in small Gauss–Newton steps due to the nonlinearity of the motion parameters [4]. Two methods, which have been shown to be preferable to the fully coupled optimization for solving separable nonlinear least-squares problems, are Block Coordinate Descent, which represents a fully decoupled optimization approach, and Variable Projection, which represents a partially coupled, optimization approach. We now provide a brief review of those two methods before introducing LAP.

2.1. Block Coordinate Descent (BCD). BCD represents a fully decoupled approach to solving coupled optimization problems such as (2); see, e.g., [26]. In BCD, the optimization variables are partitioned into a number of blocks. The method then sequentially optimizes over one block of variables while holding all the others fixed. After one cycle in which all subsets of variables have been optimized, one iteration is completed. The process is then iterated until convergence. For this paper, we separate the variables into the two subsets of variables suggested by the structure of the problem, one for the image variables and another for the motion variables. At the k th iteration, we fix \mathbf{w}_k and obtain the updated image \mathbf{x}_{k+1} by solving

$$\mathbf{x}_{k+1} = \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}_x} \Phi(\mathbf{x}, \mathbf{w}_k). \quad (3)$$

Afterwards, we fix our new guess for \mathbf{x}_{k+1} and optimize over the motion \mathbf{w} ,

$$\mathbf{w}_{k+1} = \operatorname{argmin}_{\mathbf{w} \in \mathcal{C}_w} \Phi(\mathbf{x}_{k+1}, \mathbf{w}). \quad (4)$$

These two steps constitute a single iteration of the method. We note that BCD is decoupled in the sense that while optimizing over one set of variables, we neglect optimization over the other. This degrades convergence for tightly coupled problems [26]. However, BCD has many advantages. It is applicable to general coupled problems including ones that are nonlinear in all blocks of variables. Also, it allows for straightforward implementation of bound constraints

and supports various types of regularization. For our numerical experiments, we solve the BCD imaging problem (3) inexactly using a single step of projected Gauss–Newton with bound constraints and various regularizers, which we introduce in Section 3.3. The optimization problem in the second step (4) is small-dimensional and separable, and we perform a single step of Gauss–Newton with a direct solver to compute the search direction.

2.2. Variable Projection (VarPro). VarPro is frequently used to solve separable nonlinear least-squares problems such as the one in (2); see, e.g., [17, 27]. The key idea in VarPro is to eliminate the linear variables (here, the image) by projecting the problem onto a reduced subspace associated with the nonlinear variables (here, the motion) and then solving the resulting reduced, nonlinear optimization problem. In our problem (2), eliminating the image variables requires solving a linear least-squares problem involving the matrix $T(\mathbf{w})$ that depends on the current motion parameters. We express the projection by

$$\mathbf{x}(\mathbf{w}) = \operatorname{argmin}_{\mathbf{x} \in \mathcal{C}_x} \Phi(\mathbf{x}, \mathbf{w}). \quad (5)$$

Substituting this expression in for \mathbf{x} , we then obtain a reduced dimensional problem in terms of the nonlinear variable \mathbf{w} ,

$$\min_{\mathbf{w} \in \mathcal{C}_w} \Phi(\mathbf{x}(\mathbf{w}), \mathbf{w}). \quad (6)$$

The reduced problem is solved to recover the motion parameters, noting that by solving (5) at each iteration, we simultaneously recover iterates for the image. Assuming $\mathcal{C}_w = \mathbb{R}^p$ (unconstrained case), we see that the first-order necessary optimality condition in (6) is

$$0 = \nabla_{\mathbf{w}} \Phi(\mathbf{x}(\mathbf{w}), \mathbf{w}) + \nabla_{\mathbf{w}} \mathbf{x}(\mathbf{w}) \nabla_{\mathbf{x}} \Phi(\mathbf{x}(\mathbf{w}), \mathbf{w}). \quad (7)$$

Note that in the absence of constraints on \mathbf{x} (i.e., $\mathcal{C}_x = \mathbb{F}^n$) the second term on the right hand side of (7) vanishes due to the first-order optimality condition of (6). However, this is not necessarily the case when $\mathcal{C}_x \neq \mathbb{F}^n$ or when (5) is solved with low accuracy. In those cases $\nabla_{\mathbf{x}} \Phi(\mathbf{x}(\mathbf{w}), \mathbf{w})$ does not equal 0 and computing $\nabla_{\mathbf{w}} \mathbf{x}(\mathbf{w})$, which can be as hard as solving the original optimization problem (2), is inevitable. In such situations, neglecting the second term in (7) may considerably degrade the performance of VarPro.

3. Linearize and Project (LAP)

We now introduce the LAP method for solving the coupled optimization problem (2). We begin by linearizing the residual in (2) following a standard Gauss–Newton framework. In each iteration, computing the search direction then requires solving a linear system that couples the image and motion parameters. We do this by projecting the coupled problem onto one block of variables. This offers flexibility in terms of regularization and can handle bound constraints on both the motion and image variables. Furthermore, it also allows the user to freely choose which block of variables is eliminated via projection.

We introduce our approach to solving the linear coupled problem for the Gauss–Newton step by breaking the discussion into several subsections. We start with a subsection introducing our strategy of projection onto the image space for an unconstrained problem where $\mathcal{C}_x = \mathbb{F}^n$ and $\mathcal{C}_w = \mathbb{R}^p$. This is followed by a subsection on the various options for image regularization that our projection approach offers. Lastly, we extend LAP to a projected Gauss–Newton framework to allow for element-wise bound constraints on the solution, i.e., when \mathcal{C}_x and \mathcal{C}_w are proper subsets of \mathbb{F}^n and \mathbb{R}^p , respectively.

3.1. Linearizing the Problem. We begin considering a Gauss–Newton framework to solve the coupled problem (2) in the unconstrained case, i.e., $\mathcal{C}_x = \mathbb{F}^n$ and $\mathcal{C}_w = \mathbb{R}^p$. To solve for the Gauss–Newton step at each iteration, we first reformulate the problem by linearizing the residual $\mathbf{r}(\mathbf{x}, \mathbf{w}) = \mathbf{K}\mathbf{T}(\mathbf{w})\mathbf{x} - \mathbf{d}$ around the current iterate, $(\mathbf{x}_0, \mathbf{w}_0)$. Denoting this residual as $\mathbf{r}_0 = \mathbf{r}(\mathbf{x}_0, \mathbf{w}_0)$, we can write its first-order Taylor approximation as

$$\mathbf{r}(\mathbf{x}_0 + \delta\mathbf{x}, \mathbf{w}_0 + \delta\mathbf{w}) \approx \mathbf{r}_0 + \begin{bmatrix} \mathbf{J}_x & \mathbf{J}_w \end{bmatrix} \begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{w} \end{bmatrix}, \quad (8)$$

where $\mathbf{J}_x = \nabla_x \mathbf{r}_0^\top$ and $\mathbf{J}_w = \nabla_w \mathbf{r}_0^\top$ are the Jacobian operators with respect to the image and motion parameters, respectively. They can be expressed as

$$\mathbf{J}_x = \mathbf{T}^\top \mathbf{K}^\top, \quad \text{and} \quad \mathbf{J}_w = \text{diag}(\nabla_{\mathbf{w}_1} (T(\mathbf{y}(\mathbf{w}_1))\mathbf{x}), \dots, \nabla_{\mathbf{w}_N} (T(\mathbf{y}(\mathbf{w}_N))\mathbf{x}))^\top \mathbf{K}^\top,$$

where each term $\nabla_{\mathbf{w}_k} (T(\mathbf{y}(\mathbf{w}_k))\mathbf{x})$ is the gradient of the transformed image at the transformation $\mathbf{y}(\mathbf{w}_k)$; see [4] for a detailed derivation. Both $\mathbf{J}_x \in \mathbb{F}^{m \times n}$ and $\mathbf{J}_w \in \mathbb{F}^{m \times p}$ are sparse and we recall that in the application at hand $p \ll n$.

After linearizing the residual around the current iterate, we substitute the approximation (8) for the residual term in (2) to get

$$\min_{\delta\mathbf{x}, \delta\mathbf{w}} \hat{\Phi}(\delta\mathbf{x}, \delta\mathbf{w}) = \frac{1}{2} \|\mathbf{J}_x \delta\mathbf{x} + \mathbf{J}_w \delta\mathbf{w} + \mathbf{r}_0\|^2 + \frac{\alpha}{2} \|\mathbf{L}(\mathbf{x}_0 + \delta\mathbf{x})\|^2 \quad (9)$$

By solving this problem we obtain the updates for the image and motion parameters, denoted by $\delta\mathbf{x}$ and $\delta\mathbf{w}$, respectively. As (9) is based on a linearization it is common practice to solve it only to a low accuracy; see, e.g., [26]. Note that solving (9) directly using an iterative method equates to the fully coupled Gauss–Newton approach mentioned in Sec. 2, which has been observed to converge slowly during optimization. This motivates LAP’s projection-based strategy to solving the linearized problem for the Gauss–Newton step, which we now present.

3.2. Projecting the problem onto the image space. Recall that VarPro is restricted to projecting onto the nonlinear variables (in our case, this would be the motion parameters, which requires us to solve one large-scale image reconstruction problem per iteration). With our framework, there is no such restriction; the coupled linear problem (9) can be projected onto either set of variables. Because there are a small number of nonlinear variables, to solve the coupled linear problem in (9), we propose projecting the problem onto the image

space and solving for $\delta\mathbf{x}$. To project, we first note that the first-order optimality condition for the linearized problem with respect to $\delta\mathbf{w}$ is

$$0 = \nabla_{\delta\mathbf{w}}(\mathbf{J}_x\delta\mathbf{x} + \mathbf{J}_w\delta\mathbf{w} + \mathbf{r}_0)^\top (\mathbf{J}_x\delta\mathbf{x} + \mathbf{J}_w\delta\mathbf{w} + \mathbf{r}_0)$$

or equivalently,

$$0 = 2(\mathbf{J}_w^\top\mathbf{J}_w\delta\mathbf{w} + \mathbf{J}_w^\top\mathbf{J}_x\delta\mathbf{x} + \mathbf{J}_w^\top\mathbf{r}_0).$$

Solving this condition for $\delta\mathbf{w}$, we get

$$\delta\mathbf{w} = -(\mathbf{J}_w^\top\mathbf{J}_w)^{-1}(\mathbf{J}_w^\top\mathbf{J}_x\delta\mathbf{x} + \mathbf{J}_w^\top\mathbf{r}_0). \quad (10)$$

We can then substitute this expression for $\delta\mathbf{w}$ into (9) and group terms. This projects the problem onto the image space and gives a new problem in terms of $\delta\mathbf{x}$, the Jacobians \mathbf{J}_x and \mathbf{J}_w , and the residual \mathbf{r}_0 given by

$$\min_{\delta\mathbf{x}} \frac{1}{2} \left\| \left(\mathbf{I} - \mathbf{J}_w(\mathbf{J}_w^\top\mathbf{J}_w)^{-1}\mathbf{J}_w^\top \right) \mathbf{J}_x\delta\mathbf{x} + \left(\mathbf{I} - \mathbf{J}_w(\mathbf{J}_w^\top\mathbf{J}_w)^{-1}\mathbf{J}_w^\top \right) \mathbf{r}_0 \right\|^2 + \frac{\alpha}{2} \|\mathbf{L}(\mathbf{x}_0 + \delta\mathbf{x})\|^2$$

or more succinctly,

$$\min_{\delta\mathbf{x}} \frac{1}{2} \left\| \mathbf{P}_{\mathbf{J}_w}^\perp (\mathbf{J}_x\delta\mathbf{x} + \mathbf{r}_0) \right\|^2 + \frac{\alpha}{2} \|\mathbf{L}(\mathbf{x}_0 + \delta\mathbf{x})\|^2 \quad (11)$$

where $\mathbf{P}_{\mathbf{J}_w}^\perp = \mathbf{I} - \mathbf{J}_w(\mathbf{J}_w^\top\mathbf{J}_w)^{-1}\mathbf{J}_w^\top$ is a projection onto the orthogonal complement of the column space of \mathbf{J}_w . This least-squares problem can be solved using an iterative method with an appropriate right preconditioner [3, 24, 33]. In particular, we observe that

$$\mathbf{P}_{\mathbf{J}_w}^\perp \mathbf{J}_x = \mathbf{J}_x - \mathbf{J}_w(\mathbf{J}_w^\top\mathbf{J}_w)^{-1}\mathbf{J}_w^\top\mathbf{J}_x$$

is a low rank perturbation of the operator \mathbf{J}_x since $\text{rank}(\mathbf{P}_{\mathbf{J}_w}^\perp) \leq p \ll n$. Hence, a good problem-specific preconditioner for \mathbf{J}_x should be a suitable preconditioner for the projected operator.

It is important to emphasize that in our approach, the matrix $\mathbf{J}_w^\top\mathbf{J}_w \in \mathbb{R}^{p \times p}$ is moderately sized and symmetric positive-definite if \mathbf{J}_w is full rank. Therefore, it is computationally efficient to compute its Cholesky factors once per outer Gauss–Newton iteration and reuse them to invert the matrix when needed. Furthermore, one can use a thin QR factorization of \mathbf{J}_w to compute the Cholesky factors and avoid forming $\mathbf{J}_w^\top\mathbf{J}_w$ explicitly. We use this strategy to increase the efficiency of iteratively solving (11) when using matrix-free implementations, and in practice have not seen rank-deficiency. After solving the preconditioned projected problem for $\delta\mathbf{x}$, one obtains the accompanying motion step $\delta\mathbf{w}$ via (10).

3.3. Regularization. After linearizing and projecting onto the space of the linear, image variables, we solve the regularized least-squares problem (11). For our applications, this problem is high-dimensional, and we use an iterative method to approximately solve it with low accuracy. Also, this problem is ill-posed for our numerical tests and thus requires regularization. Here we discuss the types of direct, iterative, and hybrid methods for regularization that can be used in LAP.

We first note that (11) is a Tikhonov least-squares problem. Problems of this form are well-studied in the literature; see, e.g., [8, 21, 22, 35]. The quality of an iterative solution for this type of problem depends on the selection of an appropriate regularization parameter α and a regularization operator \mathbf{L} . Common choices for \mathbf{L} include the discretized gradient operator $\mathbf{L} = \nabla_h$ using forward differences and the identity $\mathbf{L} = \mathbf{I}$. Additionally, problems including non-quadratic regularizers, e.g., total variation [32] or p -norm based regularizers can be addressed by solving a sequence of least-squares problems that include weighted ℓ_2 regularizers [30, 31, 36], or more efficiently via a hybrid Krylov subspace approach [12]. Efficient solvers for optimization problems with quadratic regularizers are also a key ingredient of splitting-based methods, e.g., the Split Bregman method for ℓ_1 regularized problems [16]. Thus, while we restrict ourselves to quadratic regularization terms for the scope of this paper, LAP is suitable for a broader class of regularization options. For more information on regularization parameter selection, see [21, 35].

Another approach is iterative regularization [8, 21], which aims to stop an iterative method at the iteration that minimizes the reconstruction error. It exploits the fact that the early iterations of iterative methods like Landweber and Krylov subspace methods contain the most important information about the solution and are less affected by noise in the data. In contrast, later iterations of these methods contain less information about the solution and are more noise-affected. This results in a reduction of the reconstruction error for the computed solution during the early iterations of the iterative method followed by its increase in later iterations, a phenomenon known as semi-convergence. However, in practice iterative regularization can be difficult as the reconstruction error at each iterate is unknown and the quality of the solution is sensitive to an accurate choice of stopping iterate.

Hybrid regularization methods represent further alternatives that seek to combine the advantages of direct and iterative regularization, see e.g., [5, 12, 14, 15] and the references therein. Hybrid methods use Tikhonov regularization with a new regularization parameter α_k at each step of an iterative Krylov subspace method such as LSQR [28]. Direct regularization for the early iterates of a Krylov subspace method is possible due to the small dimensionality of the Krylov subspace. This makes SVD-based parameter selection methods for choosing α_k computationally feasible at each iteration. The variability of α_k at each iteration helps to stabilize the semi-convergence behavior of the iterative regularization from the Krylov subspace method, making it less sensitive to the choice of stopping iteration. Thus, hybrid methods combine the advantages of both direct and iterative regularization while avoiding the cost of SVD-based parameter selection for the full-dimensional problem and alleviating the sensitivity to stopping criteria which complicates iterative regularization.

Each of the methods in this paper necessitates solving a regularized system in the image variable given by (11) for LAP, (5) for VarPro, and (3) for BCD. For these problems, we use both direct and hybrid regularization when possible

to demonstrate the flexibility of the LAP approach. We begin by running LAP with the discrete gradient regularizer, $\mathbf{L} = \nabla_h$ and a fixed α . This approach is also feasible within the VarPro framework and is straightforward to implement using BCD. To test hybrid regularization, we use the HyBR method [5] (via the interface given in the IR Tools package [13]) for LAP and BCD to automatically choose α using a weighted GCV method. We note that because such hybrid regularization methods are tailored to linear inverse problems, they cannot be used to directly solve the nonlinear VarPro optimization problem (5).

3.4. Optimization. In Sec. 3.1 and 3.2, we introduced LAP for (2) for the case of an unconstrained problem. In practice for imaging problems such as (2), bounds on the image and/or motion variables are often known, in which case imposing such prior knowledge into the inverse problem is desirable. This section details how the LAP strategy can be coupled with projected Gauss–Newton to impose element-wise bound constraints on the solutions for \mathbf{x} and \mathbf{w} . We introduce projected Gauss–Newton following the description in [19]. The method represents a compromise between a full Gauss–Newton, which converges quickly when applied to the full problem and projected gradient descent, which allows for the straightforward implementation of bound constraints. For projected Gauss–Newton, we separate the step updates $\delta\mathbf{x}$ and $\delta\mathbf{w}$ into two sets: the set of variables for which the bound constraints are inactive (the inactive set) and the set of variables for which the bound constraints are active (the active set). We denote these subsets for the image by $\delta\mathbf{x}_{\mathcal{I}} \subset \delta\mathbf{x}$ and $\delta\mathbf{x}_{\mathcal{A}} \subset \delta\mathbf{x}$ where the subscript \mathcal{I} and \mathcal{A} denote the inactive and active sets, respectively. Identical notation is used for the inactive and active sets for the motion.

On the inactive set, we take the standard Gauss–Newton step at each iteration using the LAP strategy described in Sec. 3.2. This step is computed by solving (9) restricted to the inactive set. Thus, (11) becomes

$$\min_{\delta\mathbf{x}} \frac{1}{2} \left\| \hat{\mathbf{P}}_{\mathbf{J}_w}^\perp (\hat{\mathbf{J}}_x \delta\mathbf{x}_{\mathcal{I}} + \mathbf{r}_0) \right\|^2 + \frac{\alpha}{2} \|\hat{\mathbf{L}}(\mathbf{x}_{0,\mathcal{I}} + \delta\mathbf{x}_{\mathcal{I}})\|^2, \quad (12)$$

where $\hat{\mathbf{J}}_x$, $\hat{\mathbf{J}}_w$, $\hat{\mathbf{P}}_{\mathbf{J}_w}^\perp$, and $\hat{\mathbf{L}}$ are \mathbf{J}_x , \mathbf{J}_w , $\mathbf{P}_{\mathbf{J}_w}$, and \mathbf{L} restricted to the inactive set via projection. We then obtain the corresponding motion step on the inactive set by

$$\delta\mathbf{w}_{\mathcal{I}} = -(\hat{\mathbf{J}}_w^\top \hat{\mathbf{J}}_w)^{-1} (\hat{\mathbf{J}}_w^\top \hat{\mathbf{J}}_x \delta\mathbf{x}_{\mathcal{I}} + \hat{\mathbf{J}}_w^\top \mathbf{r}_0). \quad (13)$$

This equation is analogous to (10) for the unconstrained problem. Note that LAP’s projection strategy is only used on the inactive set. Thus, the constraints do not affect the optimality condition for the projection to eliminate the $\delta\mathbf{w}_{\mathcal{I}}$ block of variables. The projected least-squares problem (12) is also unaffected. Also, for the special case when the upper and lower bounds on all variables are $-\infty$ and ∞ and all variables belong to the inactive set at each iteration, the method reduces to standard Gauss–Newton as presented in Sec. 3.2.

For the active set, we perform a scaled, projected gradient descent step given by

$$\begin{bmatrix} \delta \mathbf{x}_{\mathcal{A}} \\ \delta \mathbf{w}_{\mathcal{A}} \end{bmatrix} = - \begin{bmatrix} \tilde{\mathbf{J}}_x^\top \mathbf{r}_0 \\ \tilde{\mathbf{J}}_w^\top \mathbf{r}_0 \end{bmatrix} - \alpha \begin{bmatrix} \tilde{\mathbf{L}}^\top \tilde{\mathbf{L}}(\mathbf{x}_{0,\mathcal{A}} + \delta \mathbf{x}_{\mathcal{A}}) \\ \mathbf{0} \end{bmatrix}. \quad (14)$$

where again $\tilde{\mathbf{J}}_x$, $\tilde{\mathbf{J}}_w$, and $\tilde{\mathbf{L}}$ represent the projection of \mathbf{J}_x , \mathbf{J}_w , and \mathbf{L} onto the active set. We note that the regularization parameter α should be consistent for both (12) and (14). When using direct regularization with a fixed α , this is obvious. However, for the hybrid regularization approach discussed in Sec. 3.3, a choice is required. We set α on the active set at each iteration to be the same as the α adaptively chosen by the hybrid regularization on the inactive set at the same iterate.

The full step for a projected Gauss–Newton iteration is then given by a scaled combination of steps on the inactive and active sets

$$\begin{bmatrix} \delta \mathbf{x} \\ \delta \mathbf{w} \end{bmatrix} = \begin{bmatrix} \delta \mathbf{x}_{\mathcal{I}} \\ \delta \mathbf{w}_{\mathcal{I}} \end{bmatrix} + \gamma \begin{bmatrix} \delta \mathbf{x}_{\mathcal{A}} \\ \delta \mathbf{w}_{\mathcal{A}} \end{bmatrix}. \quad (15)$$

Here, the parameter $\gamma > 0$ is a weighting parameter to reconcile the difference in scales between the Gauss–Newton and gradient descent steps. To select this parameter, we follow the recommendation of [19] and use

$$\gamma = \frac{\max(\|\delta \mathbf{x}_{\mathcal{I}}\|_\infty, \|\delta \mathbf{w}_{\mathcal{I}}\|_\infty)}{\max(\|\delta \mathbf{x}_{\mathcal{A}}\|_\infty, \|\delta \mathbf{w}_{\mathcal{A}}\|_\infty)}. \quad (16)$$

This choice of γ ensures that the projected gradient descent step taken on the active set is no larger than the Gauss–Newton step taken on the inactive set, and we have used it with no ill effects in practice.

After combining the steps for both the inactive and active sets using (15), we use a projected Armijo line search for the combined step to ensure the next iterate obeys the problem’s constraints. A standard Armijo line search chooses a step size $0 < \eta \leq 1$ by backtracking from the full Gauss–Newton step ($\eta = 1$) to ensure a reduction of the objective function [26]. The projected Armijo line search satisfies a modified Armijo condition given by

$$\Phi(\mathbf{P}_{\mathcal{C}_x}(\mathbf{x} + \eta \delta \mathbf{x}), \mathbf{P}_{\mathcal{C}_w}(\mathbf{w} + \eta \delta \mathbf{w})) \leq \Phi(\mathbf{x}, \mathbf{w}) + c\eta \mathbf{Q}(\nabla \Phi(\mathbf{x}, \mathbf{w}))^\top \begin{bmatrix} \delta \mathbf{x} \\ \delta \mathbf{w} \end{bmatrix}. \quad (17)$$

Here, $\mathbf{P}_{\mathcal{C}_x}$ and $\mathbf{P}_{\mathcal{C}_w}$ are projections onto the feasible set for the image and motion variables, respectively, and $\mathbf{Q}(\nabla \Phi(\mathbf{x}, \mathbf{w}))$ is the projected gradient. The constant $c \in (0, 1)$ determines the necessary reduction for the line search; we set $c = 10^{-4}$ as suggested in [26]. Under the projections $\mathbf{P}_{\mathcal{C}_x}$ and $\mathbf{P}_{\mathcal{C}_w}$, variables that would leave the feasible region are projected onto the boundary and join the active set for the next iteration. However, the projection does not prevent variables from leaving the active set and joining the inactive set. This necessitates updating the inactive and active sets after the line search at each next iteration. Also note that for the special, unconstrained case, the line search reverts to the standard Armijo line search with no need for projection. However, for

this case, the problem reverts to the standard Gauss–Newton framework which should remove the necessity for a line search in most cases.

Lastly, we discuss the choice of stopping criteria for the projected Gauss–Newton method, which again depends on the type of regularization used. For a fixed regularization parameter α , we monitor the relative change of the objective function and the norm of the projected gradient including the regularizer term. The projected gradient is the first-order optimality condition of the constrained problem in (2) and can be computed via a projection [2]. The projection is necessary because the bound constraints prevent gradient entries corresponding to variables with minima outside the feasible region from converging to zero. When using hybrid regularization, we must consider the variability of α_k at each Gauss–Newton iteration. Selecting a different α_k parameter at each iteration changes the weight of the regularization term in the objective function (2) and its projected gradient at each iteration. This makes the full objective function value and projected gradient for these methods unreliable stopping criteria. Instead, we monitor both the norm of the difference between the current iterate and the previous iterate and the difference between the previous and current objective function values for the data misfit term of the objective function. We stop the Gauss–Newton method when either of those values drops below a certain threshold, indicating a stagnation of the method. This allows us to monitor the behavior of the Gauss–Newton iteration without being subject to the fluctuations associated with the varying α_k .

3.5. Summary of the Method. We summarize the discussion of LAP by presenting the entire projected Gauss–Newton algorithm using the LAP framework. This provides a framework to efficiently solve the coupled imaging problems of interest while including the flexible options for regularization and bound constraints which helped motivate our approach. The complete algorithm is given in Algorithm 1.

4. Numerical Experiments

We now test LAP for three coupled imaging problems. These are a two-dimensional super-resolution problem, a three-dimensional super-resolution problem, and a linear MRI motion correction problem. For all examples, we compare the results using LAP to those of VarPro and BCD. We look at the quality of the resultant image, the ability of the methods to correctly determine the motion parameters across all data frames, the number of iterations required during optimization, the cost of optimization in terms matrix-vector multiplications, and the CPU time to reach a solution. To compare the quality of the resultant image and motion, we use relative errors with respect to the known, true solutions. The matrix-vector multiplications of interest are those by the Jacobian operator associated with the linear, imaging variable \mathbf{J}_x . Matrix-vector multiplications by this operator are required in the linear least-squares systems for all three methods and are the most computationally expensive operation within the optimization.

Algorithm 1 Linearize and Project (LAP)

- 1: Given \mathbf{x}_0 and \mathbf{w}_0
 - 2: Compute active and inactive sets, \mathcal{A} and \mathcal{I}
 - 3: Evaluate $\Phi(\mathbf{x}_0, \mathbf{w}_0)$, \mathbf{r}_0 , $\mathbf{J}_x^{(0)}$ and $\mathbf{J}_w^{(0)}$
 - 4: **for** $k = 1, 2, 3, \dots$ **do**
 - 5: Compute the step the inactive set with LAP using (12) and (13)
 - 6: Compute the step on the active set using projected gradient descent (14)
 - 7: Combine the steps using (15) and (16)
 - 8: Perform the projected Armijo line search satisfying (17), update $\mathbf{x}_k, \mathbf{w}_k$
 - 9: Update active and inactive sets, \mathcal{A} and \mathcal{I}
 - 10: Evaluate $\Phi(\mathbf{x}_k, \mathbf{w}_k)$, \mathbf{r}_k , and $\mathbf{J}_x^{(k)}$ and $\mathbf{J}_w^{(k)}$
 - 11: **if** Stopping criteria satisfied **then**
 - 12: **return** $\mathbf{x}_k, \mathbf{w}_k$
 - 13: **end if**
 - 14: **end for**
-

4.1. **Two-Dimensional Super Resolution.** Next, we run numerical experiments using a relatively small two-dimensional super resolution problem. To construct a super resolution problem with known ground truth image and motion parameters, we use the 2D MRI dataset provided in FAIR [25] (original resolution 128×128) to generate 32 frames of low-resolution test data (resolution 32×32) after applying 2D rigid body transformations with randomly chosen parameters. Gaussian white noise is added using the formula

$$\mathbf{d}_k = \bar{\mathbf{d}}_k + \epsilon_k \quad \text{where} \quad \epsilon_k = \mu \frac{\|\bar{\mathbf{d}}_k\|_2}{\|\mathbf{n}_k\|_2} \mathbf{n}_k$$

where $\bar{\mathbf{d}}_k$ denotes a noise free low-resolution data frame, μ is the percentage of noise and \mathbf{n}_k is a vector of normally distributed random values with mean 0 and standard deviation 1. Our experiments show results with $\mu = 1\%$, 2% , and 3% noise added to the low-resolution data frames. The resulting super-resolution problem is an optimization problem of the form (2). Here, the imaging operator \mathbf{K} is a block diagonal matrix composed of 32 identical down-sampling matrices $K = K_k$ along the diagonal, which relate the high-resolution image to the lower resolution one via block averaging. The total number of parameters in the optimization is 16,528 corresponding to 16,384 for the image and 96 for the motion.

As mentioned in [4], the choice of initial guess is crucial in super-resolution problems, so we perform rigid registration of the low-resolution data onto the first volume (resulting in 31 rigid registration problems) to obtain a starting guess for the motion parameters. The resulting relative error of the parameters is around 2%. Using these parameters, we solve the linear image reconstruction problem to obtain a starting guess for the image with around 5% relative error.

We then solve the super-resolution problem using LAP, VarPro, and BCD. For all three approaches, we compare two regularization strategies. All three

methods are run using the gradient operator, $\mathbf{L} = \nabla_h$ with a fixed regularization parameter $\alpha = 0.01$. LAP and BCD are then run with the Golub-Kahan hybrid regularization approach detailed in Section 3.3 (denoted as HyBR in tables and figures). As previously mentioned, HyBR cannot be applied directly to the VarPro optimization problem (5), so for VarPro we use a fixed $\alpha = 0.01$ and the identity as our second regularizer, $\mathbf{L} = \mathbf{I}$. To allow for comparison, we use the same regularization parameter as in [4], which may not be optimal for any of the methods presented. For more rigorous selection criteria, we refer the reader to the methods mentioned in Sec. 3.3. For LAP and BCD, we add element-wise bound constraints on the image space in the range $[0, 1]$ for both choices of regularizer, with both bounds active in practice. The number of active bounds varies for different noise levels and realizations of the problem but can include as many as 30 – 35% of the image variables for both LAP and BCD. VarPro is run without bound constraints. Neither constraints nor regularization are imposed on the motion parameters for the three methods.

All three methods require solving two different types of linear systems, one associated with the image variables and another with the motion variables. LAP solves these systems to determine the Gauss–Newton step. We use LSQR [28] with a stopping tolerance of 10^{-2} to solve (11) for both regularization approaches. The motion step (10) is computed using the Cholesky factors of $\mathbf{J}_w^\top \mathbf{J}_w$. VarPro requires solving the linear system (5) within each function call. For both choices of regularization, we solved this system by running LSQR for a fixed number of 20 iterations. Recall that this system must be solved to a higher accuracy than the similarly sized systems in LAP and BCD to maintain the required accuracy in the gradient; see Sec. 2.2. Gauss–Newton on the reduced dimensional VarPro function (6) then requires solving the reduced linear system in the motion parameters, which we solve using Cholesky factorization on the normal equations. For BCD, coordinate descent requires alternating solutions for the linear system in the image, (3), and a nonlinear system in the motion parameters, (4). For both of these, we take a single projected Gauss–Newton step. For the image, this is solved using LSQR with a stopping tolerance of 10^{-2} using MATLAB’s `lsqr` function for direct regularization and HyBR’s LSQR for hybrid regularization. LSQR with this tolerance is also used for the corresponding problems in the 3D super-resolution and MRI motion correction examples in Sec. 4.2 and 4.3. For the motion, we use Cholesky factorization on the normal equations. During these solves, we track computational costs for the number of matrix-vector multiplications by the Jacobian operator associated with the image, \mathbf{J}_x . For LAP and BCD, these multiplications are required when solving the system for the Gauss–Newton step for the image step, while for VarPro, they are necessary for the least-squares solve within the objective function.

We compare results for the methods using the relative errors of the resultant image and motion parameters. We separate relative errors for the image and motion. Plots of these errors against iteration can be seen in Fig. 2 for the problem with 2% added noise. The corresponding resulting image for LAP for

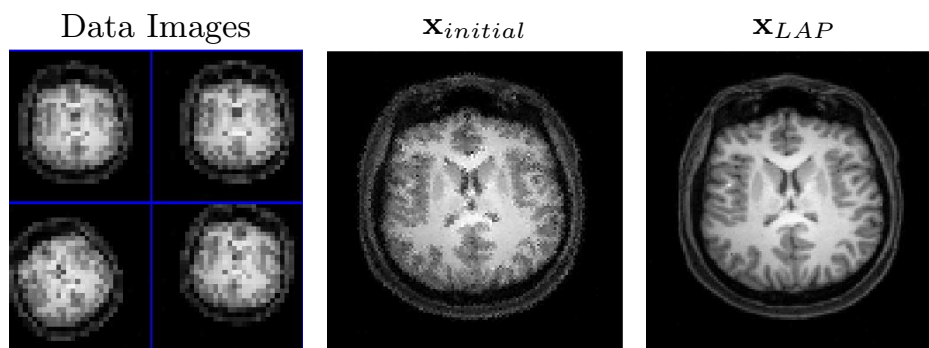


Figure 1. The images from left to right show a montage of data frames, the initial guess $\mathbf{x}_{initial}$, and the reconstructed image \mathbf{x}_{LAP} for LAP using the discrete gradient regularizer for the 2D super resolution problem with 2% noise.

the 2% error case can be seen in Fig. 1. Lastly, a table of relevant values including the average number of iterations, minimum errors for image and motion, matrix-vector multiplications, and CPU timings for the methods taken over 10 different realizations of the problem for all three noise levels is in Table 1.

For direct regularization using the discrete gradient operator, the solutions for all three methods are comparable in terms of the relative error for the motion, with LAP and BCD slightly outperforming VarPro for the relative error of the recovered images. This is a direct result of the element-wise bound constraints on the resultant images using these methods. Furthermore, these solutions are superior to those for all three methods using hybrid regularization or the identity operator, suggesting that this is a more appropriate regularizer for this problem. LAP with the discrete gradient operator recovers the most accurate reconstructed image of the three methods and achieves better or comparable recovery of the motion parameters. This is observable in the relative error plots for the 2% added noise case in Fig. 2, and for the problem with all three noise levels in Table 1. We can also see from the relative error plots that LAP tends to recover the correct motion parameters earlier in the Gauss–Newton iterations than either BCD or VarPro. In terms of method cost, both the LAP and BCD iterations cost significantly less in terms of time and matrix-vector multiplications than those of VarPro, resulting in faster CPU times and fewer matrix-vector multiplies for the entire optimization. However, while BCD is also relatively cheap in terms of matrix-vector multiplies and CPU time, LAP outperforms it in terms of solution quality. Overall, the LAP approach compares favorably to VarPro and BCD for this example in terms of both the resulting solutions and cost, outperforming both methods.

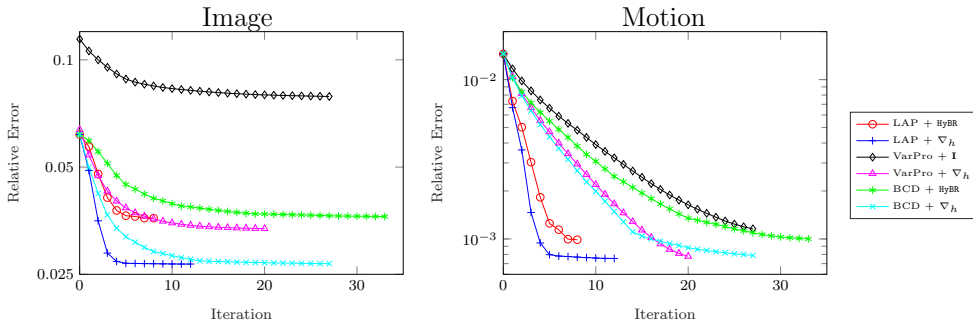


Figure 2. These plots show the relative errors for both the reconstructed image and the motion parameters for the 2D super resolution problem with 2% added noise. We note that methods using the discrete gradient regularizer reach lower minima for the image in this problem, and LAP with the discrete gradient regularizer outperforms both VarPro and BCD in recovering the motion parameters in the early iterations of the method for all noise levels tested.

4.2. Three-Dimensional Super Resolution. The next problem is a larger three-dimensional super-resolution problem. Again, we use a 3D MRI dataset provided in FAIR [25] to construct a super-resolution problem with a known ground truth image and motion parameters. The ground truth image (resolution $160 \times 96 \times 144$) is used to generate 128 frames of low-resolution test data (resolution $40 \times 24 \times 32$). Each frame of data is shifted and rotated by a random 3D rigid body transformation, after which it is downsampled using block averaging. Lastly, Gaussian white noise is added to each low-resolution data frame in the same way as for the two-dimensional super-resolution problem, and we run the problem for $\mu = 1\%$, 2% , and 3% added noise per data frame. The resulting optimization problem has 2,212,608 unknowns, 2,211,840 for the image and 768 for the motion parameters. The data has dimension 5,898,240. The formulation of the problem is identical to that of the two-dimensional super-resolution problem with appropriate corrections for the change in dimension. The imaging operator \mathbf{K} in the three-dimensional example is block diagonal with 128 identical down-sampling matrices K which relate the high-resolution image to the low-resolution data by block averaging.

The initial guess for the three-dimensional problem is generated using the same strategy as in the two-dimensional case. To generate a guess for the motion parameters, we register all frames onto the first frame (thus solving 127 rigid registration problems.) This gives an initial guess for the motion with approximately 3% relative error. Using this initial guess, we then solve a linear least-squares problem to obtain an initial guess for the image with a relative error of around 13%. We note that this is a poorer initial guess when compared to the true solution, than the one obtained for the two-dimensional super-resolution

		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
1% Noise	LAP + HyBR	11.1	5.34e-2	1.82e-2	93.4	14.3
	LAP + ∇_h	13.1	3.76e-2	1.78e-2	62.2	15.4
	VarPro + \mathbf{I}	25.7	8.60e-2	1.84e-2	554.0	60.1
	VarPro + ∇_h	21.1	4.12e-2	1.81e-2	462.0	56.0
	BCD + HyBR	12.0	5.93e-2	2.03e-2	68.1	22.9
	BCD + ∇_h	29.7	3.95e-2	1.79e-2	89.6	55.4
2% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + HyBR	12.0	6.38e-2	1.86e-2	77.0	12.5
	LAP + ∇_h	15.9	4.38e-2	1.81e-2	66.4	15.9
	VarPro + \mathbf{I}	29.6	1.03e-1	1.86e-2	632.0	64.7
	VarPro + ∇_h	21.4	5.05e-2	1.85e-2	468.0	54.9
	BCD + HyBR	13.4	7.10e-2	2.15e-2	53.1	25.0
BCD + ∇_h	29.9	4.55e-2	1.82e-2	91.7	57.0	
3% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + HyBR	12.3	7.54e-2	2.28e-2	68.6	16.3
	LAP + ∇_h	18.5	5.34e-2	2.22e-2	74.2	22.0
	VarPro + \mathbf{I}	34.4	1.27e-1	2.32e-2	728.0	86.7
	VarPro + ∇_h	23.9	6.17e-2	2.13e-2	518.0	72.5
	BCD + HyBR	15.0	7.86e-2	2.32e-2	52.9	31.9
BCD + ∇_h	29.7	5.42e-2	2.24e-2	83.2	63.9	

Table 1. This table shows data for the 2D super-resolution for multiple values of added Gaussian noise. The columns from left to right give the stopping iteration, the relative error of the solution image, the relative error of the solution motion, number of matrix-vector multiplications during optimization, and time in seconds using `tic` and `toc` in `MATLAB`. All values are averages taken from 10 instances with different motion parameters, initial guesses, and noise realizations. The best results for each column is bold-faced.

example. This may impact the quality of the solution obtainable for examples with large amounts of noise.

For this example, we test LAP, VarPro, and BCD using only the discrete gradient regularizer due to its better performance in the 2D example. We use identical parameters to the 2D problem for the regularization parameter, bound constraints on the image variables, and accuracy of the iterative LSQR. The lone difference in the problem setup from the 2D case is in solving the linear system (5) for VarPro. Instead of the 20 fixed LSQR iterations from the 2D case, we run a fixed number of 50 iterations to achieve the required accuracy for the larger system. For all three methods, the reduced system in the motion is solved using Cholesky factorization on the normal equations.

The resulting solution image for LAP and the relative error plots for all three methods for the images and motion parameters can be found in Figs. 3 and 4, respectively. Like the 2D super-resolution example, LAP converges faster to the motion parameters in the early iterations than BCD and VarPro and succeeds in

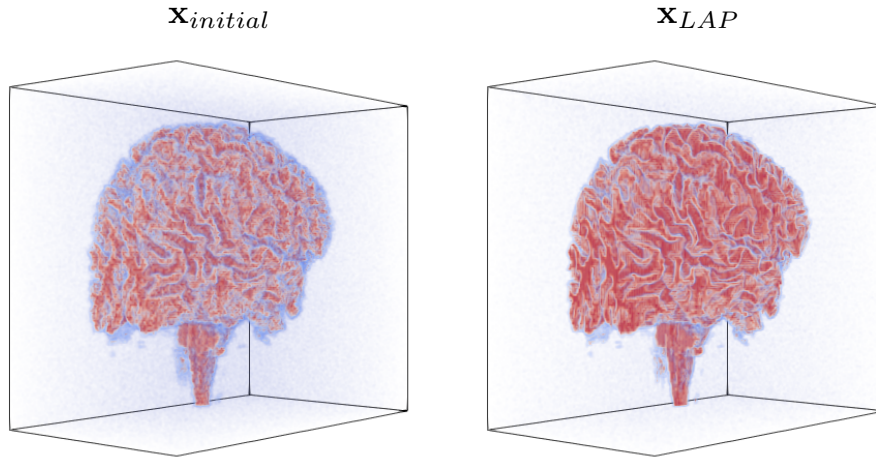


Figure 3. This figure shows 3D volume renderings of the reconstructed images using LAP for the 3D super resolution problem with 2% noise. The volume on the left shows the initial guess, $\mathbf{x}_{initial}$. On the right is the reconstructed solution using LAP, \mathbf{x}_{LAP} with the discrete gradient regularizer.

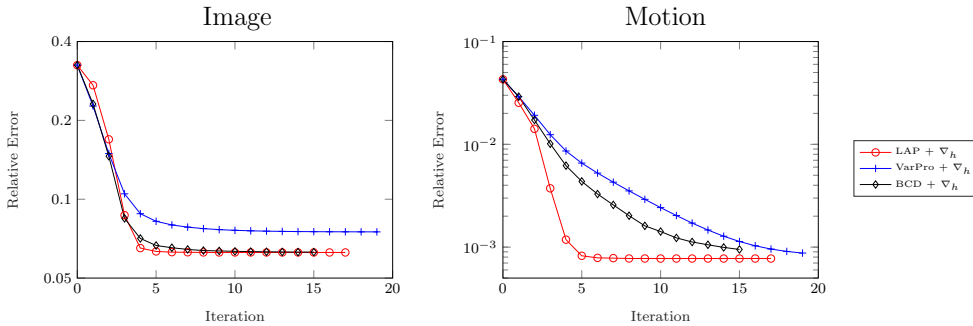


Figure 4. This figure plots the relative errors for both the reconstructed image and the motion parameters for the 3D super resolution for 2% added noise. LAP succeeds in capturing the correct motion parameters in fewer iterations than VarPro and BCD and in recovering images of comparable quality.

reaching lower relative errors for both the recovered image and motion parameters. Again, VarPro's iterations are far more expensive in terms of matrix-vector multiplications and CPU time as seen in Table 2. BCD performs similarly with LAP in terms of the reconstructed image, but it does less well at recovering the motion parameters and is slightly more expensive in terms of CPU time due to a higher number of function calls.

1% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + ∇_h	14.1	6.23e-2	8.65e-4	22.1	2.30e3
	VarPro + ∇_h	16.4	7.21e-2	9.55e-4	1.79e3	1.23e4
	BCD + ∇_h	11.3	6.25e-2	1.40e-3	25.7	4.26e3
2% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + ∇_h	15.3	6.35e-2	9.27e-4	21.0	2.15e3
	VarPro + ∇_h	18.0	7.57e-2	1.00e-3	1.95e3	1.22e4
	BCD + ∇_h	12.2	6.37e-2	1.59e-3	26.2	3.75e3
3% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + ∇_h	16.2	6.48e-2	8.97e-4	20.3	2.94e3
	VarPro + ∇_h	17.2	8.08e-2	9.52e-4	1.88e3	1.07e4
	BCD + ∇_h	11.7	6.50e-2	1.53e-3	25.2	4.21e3

Table 2. This table presents data for the solution of the 3D super resolution for multiple values of added Gaussian noise. The columns from left to right give the stopping iteration, relative error of the solution image, relative error of the solution motion, number of matrix-vector multiplies during optimization, and time in seconds using `tic` and `toc` in `MATLAB`. All values are averages taken from 10 separate problems with different motion parameters, initial guesses, and noise realizations.

4.3. MRI Motion Correction. The final test problem is a two-dimensional MRI motion correction problem. The goal in this MRI application is to reconstruct a complex-valued MRI image from its Fourier coefficients that are acquired block-wise in a sequence of measurements. Since the measurement process typically requires several seconds or minutes, in some cases the object being imaged moves substantially. Motion renders the Fourier samples inconsistent and — without correction — results in artifacts and blurring in the reconstructed MRI image. To correct for this, one can instead view the collected MRI data as a set of distinct, complex-valued Fourier samplings, each measuring some portion of the Fourier domain and subject to some unknown motion parameters. The resulting problem of recovering the unknown motion parameters for each Fourier sampling and combining them to obtain a single motion-corrected MRI image fits into the coupled imaging framework presented in this paper.

The forward model for this problem was presented by Batchelor *et al.* [1]. In their formulation, our imaging operator \mathbf{K} is again block diagonal with diagonal blocks K_k for $k = 1, 2, \dots, N$ given by

$$K_k = \mathbf{A}_k \mathcal{F} \mathbf{S}.$$

Here, \mathbf{S} is a complex-valued block rectangular matrix containing the given coil sensitivities of the MRI machine, \mathcal{F} is block diagonal with each block a two-dimensional Fourier transform (2D FFT), and \mathbf{A}_k is a block diagonal matrix with rectangular blocks containing selected rows of the identity corresponding to the Fourier sampling for the k th data observation. As with the other examples, the imaging operator \mathbf{K} is multiplied on the right by the block rectangular matrix \mathbf{T}

with $T(\mathbf{y}(\mathbf{w}_k))$ blocks modeling the motion parameters of each Fourier sampling. We note that the cost of matrix-vector multiplications by this imaging operator is dominated by the 2D FFTs in block \mathcal{F} , and we note that for 32 receiver coils, the cost of a single matrix-vector multiplication will require $32N$ 2D FFTs of size 128×128 where N is the number of Fourier samplings in the data set. Additionally, we note that the presence of these FFT matrices prevents us from explicitly storing the matrix and necessitates passing it as a function call for all of the methods. This also applies to the Jacobian with respect to the image, \mathbf{J}_x . However, because it is relatively small in size, \mathbf{J}_w can still be computed and stored explicitly.

We use the dataset provided in the `alignedSENSE` package [6] to set up an MRI motion correction with a known true image and known motion parameters. To this end, we generate noisy data by using the forward problem (1). The ground truth image with resolution 128×128 is rotated and shifted by a random 2D rigid body transformation. The motion affected image is then observed on 32 sensors with known sensitivities. Each of these 32 observations is then sampled in Fourier space. For our problem, each sampling corresponds to $1/16$ of the Fourier domain, meaning that $N = 16$ samplings (each with unknown motion parameters) are needed to have a full sampling of the whole space. We sample using a Cartesian parallel two-dimensional sampling pattern [6]. Noise is added to the data using the formula

$$\mathbf{d} = \bar{\mathbf{d}} + \epsilon \quad \text{where} \quad \epsilon = \mu \frac{\|\bar{\mathbf{d}}\|_\infty}{\|\mathbf{n}\|_2} \mathbf{n}.$$

Here, $\bar{\mathbf{d}}$ is the noise free data, μ is the percentage of noise added, and \mathbf{n} is a complex-valued vector where the entries of $\text{Re}(\mathbf{n})$ and $\text{Im}(\mathbf{n})$ are normally-distributed random numbers with mean 0 and standard deviation 1. We run the problem for $\mu = 5\%$, 10% , and 15% added noise. The resulting data has dimension $\frac{128 \times 128 \times 32}{16} \times 16 = 524,288$. The MRI motion correction optimization problem fits within the framework in Eq. 2 and has 16,432 unknowns corresponding to 16,384 for the image and 48 for the motion.

As with previous examples, we solve the MRI motion correction problem using LAP, VarPro, and BCD. The setup and parameters are similar to those of the 2D super-resolution. We use the discrete gradient and HyBR as regularization options for LAP and BCD, and for VarPro we regularize using the discrete gradient operator and the identity operator. For the non-hybrid regularizers, we fix $\alpha = 0.01$. For the least-squares problems in the image variables for LAP and BCD, we solve using LSQR with an identical tolerance to the super-resolution examples. For VarPro, we use LSQR with a tolerance of 10^{-8} or a maximum of 100 iterations to solve (5) to maintain accuracy in the gradient. Cholesky factorization on the normal equations is used for the lower-dimensional solves in the motion. No bound constraints are applied to any of the methods for this example because element-wise bound constraints on the real or imaginary parts of the complex-valued image variables will affect the angle (or phase) of the solution, which is undesirable.

For an initial guess for the motion parameters, we start with $\mathbf{w} = 0$ (corresponding to a relative error of 100%). Using this initialization, we solve a linear least-squares problem to get an initial guess for the image. For 10% added Gaussian noise in the data, the initial guess for the image has a relative error of around 35%. We show the initial guess in Fig. 5.

LAP, VarPro, and BCD provide fairly accurate reconstructions of both the image and motion parameters for quite large values of noise using either HyBR or the identity as a regularizer; see Figs. 5 and 6. This is likely due to the fact that the problem is not severely ill-posed and is highly over-determined (32 sensor readings for each point in Fourier space.) For this example, the hybrid regularization approach for LAP and BCD produces the best results, with LAP requiring considerably fewer iterations. We remark that the best regularization from this problem differs from the super-resolution problems, which shows the importance of the flexibility that LAP offers for regularizing the image. The comparative speed of LAP is observable for the relative error plots for the problem with 10% noise and further evidenced in Table 3 for all noise levels over 10 separate realizations of the problem. For the gradient-based regularizer, all three methods do not recover the motion parameters accurately. We also note that the number of iterations and their cost is an important consideration for this problem. Because of the distance of the initial guess from the solution, this problem requires more iterations than the super-resolution examples. Additionally, the high number of 2D FFTs required for a single matrix-vector multiplication makes multiplications by the Jacobian \mathbf{J}_x and \mathbf{J}_w expensive. Table 3 shows that LAP outperforms VarPro and BCD for both choices of regularizer by requiring fewer, cheaper iterations in terms of both time and matrix-vector multiplications. The difference in cost is most dramatic when compared with VarPro again due to the large number of FFTs required for a single matrix-vector multiplication and the large number of such multiplications required within each VarPro function call. For BCD and LAP, the number of matrix-vector multiplications is similar, but BCD requires more iterations for convergence. Overall, we see that LAP is a better choice for this problem and that it provides better reconstructions of both the image and motion in fewer, cheaper iterations.

5. Summary and Conclusion

We introduce a new method, called Linearize And Project (LAP), for solving large-scale inverse problems with coupled blocks of variables in a projected Gauss-Newton framework. Problems with these characteristics arise frequently in applications, and we exemplify and motivate LAP using joint reconstruction problems in imaging that aim at estimating image and motion parameters from a number of noisy, indirect, and motion-affected measurements. By design, LAP is most attractive when the optimization problem with respect to one block of variables is comparably easy to solve. LAP is very flexible in the sense that it supports different regularization strategies, simplifies imposing equality and inequality constraints on both blocks of variables, and does not require (as in

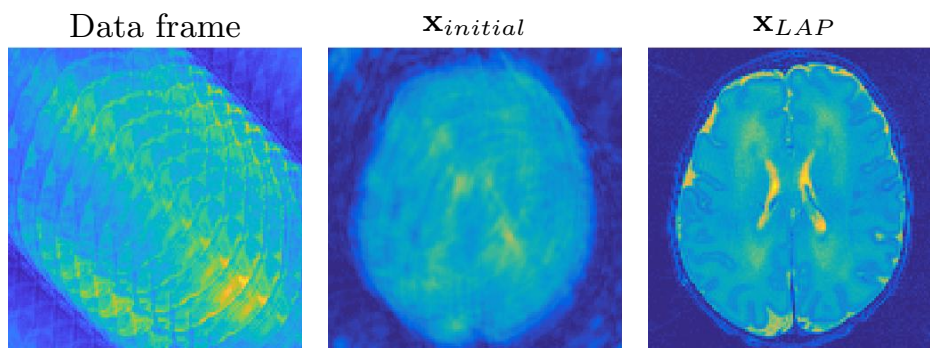


Figure 5. This figure shows a single inverted data sampling (*left*), the initial guess for the image (*center*), and reconstructed image (*right*) for the MRI motion correction problem with 10% noise. The solution image shown is for LAP using the HyBR regularizer. Note that the MRI images are the modulus of the complex-valued images recovered.

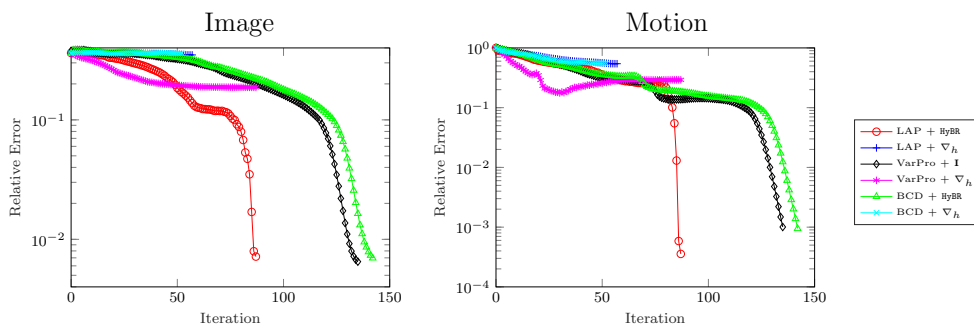


Figure 6. This figure shows the relative errors for both the reconstructed image and the motion parameters for the MRI motion correction problem for three levels of noise. LAP with hybrid regularization achieves better reconstructions of the image and motion parameters for fewer iterations than either VarPro and BCD.

the case of VarPro) that the forward problem depends linearly on one set of variables. In our numerical experiments using four separable nonlinear least-squares problems, we showed that LAP is competitive and often superior to VarPro and BCD with respect to accuracy and efficiency.

LAP is as general as alternating minimization methods such as Block Coordinate Descent. However while BCD ignores the coupling between the variable blocks when computing updates, LAP takes it into consideration. Thus, in our

		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
5% Noise	LAP + HyBR	76.6	3.55e-3	1.59e-4	3.18e2	5.86e2
	LAP + ∇_h	76.2	3.74e-3	1.67e-4	3.62e2	4.32e2
	VarPro + \mathbf{I}	116.0	7.93e-2	4.32e-2	2.19e4	1.20e4
	VarPro + ∇_h	115.9	7.92e-2	4.33e-2	2.19e4	1.52e4
	BCD + HyBR	128.6	4.35e-2	2.27e-2	4.03e2	1.01e3
	BCD + ∇_h	116.6	7.83e-2	4.20e-2	4.27e2	8.59e2
10% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + HyBR	76.8	6.56e-3	3.25e-4	3.11e2	6.52e2
	LAP + ∇_h	76.1	6.88e-3	4.15e-4	3.55e2	5.19e2
	VarPro + \mathbf{I}	116.0	8.08e-2	4.33e-2	2.19e4	1.12e4
	VarPro + ∇_h	116.0	8.08e-2	4.32e-2	2.19e4	1.27e4
	BCD + HyBR	128.2	4.53e-2	2.21e-2	3.89e2	1.24e3
BCD + ∇_h	116.0	8.02e-2	4.21e-2	4.12e2	1.05e3	
15% Noise		Iter.	Rel. Err. \mathbf{x}	Rel. Err. \mathbf{w}	MatVecs.	Time(s)
	LAP + HyBR	77.3	9.49e-3	4.69e-2	3.07e2	5.02e2
	LAP + ∇_h	75.0	1.61e-2	2.92e-2	3.40e2	3.61e2
	VarPro + \mathbf{I}	115.8	8.29e-2	4.36e-2	2.18e4	9.98e3
	VarPro + ∇_h	115.7	8.28e-2	4.35e-2	2.18e4	1.27e4
	BCD + HyBR	127.0	4.80e-2	2.21e-2	3.47e2	1.06e3
BCD + ∇_h	129.0	8.20e-2	4.17e-2	3.99e2	8.17e2	

Table 3. This table shows the results of LAP, VarPro, and BCD for solving the MRI motion correction example for multiple regularizers and varying levels of added noise. Averaged over 10 realizations of the problem, the columns are stopping iteration, the relative error of the solution image, the relative error of the solution motion, number of matrix-vector multiplies during optimization, and time in seconds using `tic` and `toc` in `MATLAB`. LAP outperforms the other methods in terms of solution quality, computational cost, and CPU time.

experiments LAP requires a considerably smaller number of iterations, matrix-vector multiplications, and CPU time than BCD to achieve similar accuracy.

LAP is not limited to separable nonlinear least-squares problems and thus more broadly applicable than VarPro. Since LAP projects after linearization, it provides the opportunity to freely choose which block of variables gets eliminated. For example, in our numerical examples in Sec. 4.1–4.3, LAP eliminates the parameters associated with the motion (which are of comparably small dimension) when computing the search direction in the projected Gauss–Newton scheme. Due to the robustness of Gauss–Newton methods, it suffices to solve the imaging problem iteratively to low accuracy. By contrast, VarPro eliminates the image variables that enter the residual in a linear way. While this leads to a small-dimensional nonlinear optimization problem for the motion, each iteration requires solving the imaging problem to relatively high accuracy to obtain reliable gradient information; see also Sec 2.2. This can be problematic for large and

ill-posed imaging problems, and thus LAP can in some cases reduce runtimes by an order of magnitude; see Tables 2 and 3.

It is worth noting that the key step of LAP, which is the projection onto one variable block when solving the approximated Newton system, is a block elimination, and the reduced system corresponds to the Schur-complement.

To allow for comparison with VarPro we focused on separable nonlinear least-squares problems. In future work, we will study the performance of LAP to solve general coupled nonlinear optimization problems.

ACKNOWLEDGEMENT

This work is supported by Emory’s University Research Committee and National Science Foundation (NSF) awards DMS 1522760 and DMS 1522599.

References

- [1] P. Batchelor, D. Atkinson, P. Irarrazaval, D. Hill, J. Hajnal, and D. Larkman. Matrix Description of General Motion Correction Applied to Multishot Images. *Magnetic resonance in medicine*, 54(5):1273–1280, 2005.
- [2] A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, 2014.
- [3] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996.
- [4] J. Chung, E. Haber, and J. Nagy. Numerical methods for coupled super-resolution. *Inverse Problems*, 22(4):1261–1272, 2006.
- [5] J. Chung, J. Nagy, and D. O’Leary. A weighted GCV method for Lanczos hybrid regularization. *Electronic Transactions on Numerical Analysis*, 2008.
- [6] L. Cordero-Grande, R. Teixeira, E. Hughes, J. Hutter, A. Price, and J. Hajnal. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Transactions on Computational Imaging*, 2(3):266–280, 2016.
- [7] E. de Sturler and M. Kilmer. A Regularized Gauss–Newton Trust Region Approach to Imaging in Diffuse Optical Tomography. *SIAM Journal on Scientific Computing*, 33(5):3057–3086, 2011.
- [8] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 2000.
- [9] R. Estrin, D. Orban, and M. Saunders. LSLQ: an Iterative Method for Linear Least-Squares with an Error Minimization Property. Technical Report G-2017-05, Les Cahiers du GERAD, 2017.
- [10] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.
- [11] J. Fohring, E. Haber, and L. Ruthotto. Geophysical Imaging of Fluid Flow in Porous Media. *SIAM Journal on Scientific Computing*, 36(5):S218–S236, 2014.
- [12] S. Gazzola and J. G. Nagy. Generalized Arnoldi–Tikhonov method for sparse reconstruction. *SIAM J. Sci. Comput.*, 36(2):B225–B247, 2014.
- [13] S. Gazzola, J. G. Nagy, and P. C. Hansen. IR Tools: A MATLAB package of iterative regularization methods and large-scale test problems. arXiv:1712.05602, to appear in *Numer. Algor.*, 2018.
- [14] S. Gazzola and P. Novati. Automatic parameter setting for Arnoldi–Tikhonov methods. *J. Comp. Appl. Math.*, 256(15):180–195, 2014.
- [15] S. Gazzola, P. Novati, and M. R. Russo. Embedded techniques for choosing the parameter in Tikhonov regularization. *Num. Lin. Alg. Appl.*, 21(6):796–812, 2014.

- [16] T. Goldstein and S. Osher. The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, Jan. 2009.
- [17] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, 19(2):R1, 2003.
- [18] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, 10(2):413–432, 1973.
- [19] E. Haber. *Computational Methods in Geophysical Electromagnetics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.
- [20] E. Haber and D. Oldenburg. A GCV based method for nonlinear ill-posed problems. *Computational Geosciences*, 4:41–63, 2000.
- [21] P. C. Hansen. *Rank-deficient and discrete ill-posed problems*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.
- [22] P. C. Hansen. *Discrete Inverse Problems: Insight and Algorithms*. SIAM, Philadelphia, PA, 2010.
- [23] R. Hardie, K. Barnard, and E. Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12):1621–1633, 1997.
- [24] M. Hestenes and E. Stiefel. *Methods of conjugate gradients for solving linear systems*, volume 49. NBS, 1952.
- [25] J. Modersitzki. *FAIR: flexible algorithms for image registration*, volume 6 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
- [26] J. Nocedal and S. Wright. Numerical optimization. *Springer Science*, 35:67–68, 1999.
- [27] D. O’Leary and B. Rust. Variable projection for nonlinear least squares problems. *Computational Optimization and Applications*. *An International Journal*, 54(3):579–593, 2013.
- [28] C. C. Paige and M. A. Saunders. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, 1982.
- [29] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.
- [30] P. Rodríguez and B. Wohlberg. An iteratively reweighted norm algorithm for total variation regularization. In *Proceedings of the 40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2006.
- [31] P. Rodríguez and B. Wohlberg. An efficient algorithm for sparse representations with ℓ^p data fidelity term. In *Proceedings of 4th IEEE Andean Technical Conference (ANDESCON)*, 2008.
- [32] L. I. RUDIN, S. Osher, and E. FATEMI. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60(1-4):259–268, 1992.
- [33] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.
- [34] D. Sima and S. V. Huffel. Separable nonlinear least squares fitting with linear bound constraints and its application in magnetic resonance spectroscopy data quantification. *Journal of computational and applied mathematics*, 203(1):264–278, 2007.
- [35] C. Vogel. *Computational methods for inverse problems*, volume 23. SIAM, 2002.
- [36] B. Wohlberg and P. Rodríguez. An iteratively reweighted norm algorithm for minimization of total variation functionals. *IEEE Signal Processing Letters*, 14:948–951, 2007.