A differential equations approach to $l_1$-minimization with applications to array imaging

Miguel Moscoso$^1$, Alexei Novikov$^2$, George Papanicolaou$^3$ and Lenya Ryzhik$^3$

$^1$ Escuela Politécnica Superior, Universidad Carlos III de Madrid, Madrid, Spain
$^2$ Department of Mathematics, Pennsylvania State University, PA, USA
$^3$ Department of Mathematics, Stanford University, CA, USA

E-mail: moscoso@math.uc3m.es, anovikov@math.psu.edu, papanicolaou@stanford.edu and ryzhik@math.stanford.edu

Received 9 March 2012, in final form 31 July 2012
Published 6 September 2012
Online at stacks.iop.org/IP/28/105001

Abstract
We present an ordinary differential equation approach to the analysis of algorithms for constructing $l_1$ minimizing solutions to underdetermined linear systems of full rank. It involves a relaxed minimization problem whose minimum is independent of the relaxation parameter. An advantage of using the ordinary differential equations is that energy methods can be used to prove convergence. The connection to the discrete algorithms is provided by the Crandall–Liggett theory of monotone nonlinear semigroups. We illustrate the effectiveness of the discrete optimization algorithm in some sparse array imaging problems.

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

1. Introduction
We consider the solution of large underdetermined linear systems of equations $Ax = y$ where $A \in \mathbb{R}^{m \times n}$ is a given matrix, $y \in \mathbb{R}^m$ is a known vector of $m \ll n$ measurements and $x \in \mathbb{R}^n$ is the unknown signal or image to be estimated. We assume that $A$ has full rank equal to $m$. We want to find the solutions of this system with minimal $l_1$ norm $\|x\|_{l_1}$,

$$\min \|x\|_{l_1}, \text{ subject to } y = Ax,$$

(1.1)

Our motivation is array imaging problems, which is an application discussed in this paper, but such sparsity inducing constrained minimization problems, where the $l_1$ norm of the solution vector is used, arise in many other applications in signal and image processing [18, 23, 33, 34, 43, 44].

A lot of research has been devoted to developing algorithms for solving efficiently (1.1) and its relaxed form

$$\min \left\{ \tau \|x\|_{l_1} + \frac{1}{2} \|y - Ax\|^2 \right\}.$$ 

(1.2)
Here, and throughout the paper, $\|q\|$ denotes the $l_2$ norm of a vector $q$. In (1.2), the exact constraint $y = A\mathbf{x}$ is relaxed so as to take into account possible measurement noise and $\tau$ is a positive real parameter that promotes sparse solutions when it is large enough.

The iterative shrinkage-thresholding algorithm (ISTA) is the usual gradient descent method applied to (1.2). It has been used in many different applications with great success, such as [16, 20–22, 28, 48], just to mention a few. The ISTA algorithm generates a sequence of iterates $\{x_k\}$ of the form

$$x_{k+1} = \eta_h(x_k - h\nabla f(x_k)).$$

Here, $h$ is the step size,

$$\eta_h(x) = \begin{cases} 
    x - a, & \text{if } x > a, \\
    0, & \text{if } -a < x < a, \\
    x + a, & \text{if } x < -a
\end{cases}$$

is the shrinkage-thresholding operator and $\nabla f(x_k)$ denotes the gradient of $f(x) = \frac{1}{2}\|y - A\mathbf{x}\|^2$ at the current iterate $x_k$. Thus, $\nabla f(x_k) = A^*(A\mathbf{x}_k - y)$, where $A^*$ denotes the complex conjugate transpose of $A$. The algorithm (1.3) involves only simple matrix–vector multiplications followed by a shrinkage-thresholding step.

For a fixed value of $\tau$, the solution to (1.2) differs in general from the solution of (1.1). In other words, exact recovery from noiseless data is not achieved by solving (1.2), unless the regularization parameter $\tau$ is sent to zero. However, it is well known that the convergence of (1.3) is slow for small values of the parameter $\tau$. This issue is considered in detail in [8].

Variants of (1.3) have been proposed to speed up its convergence rate. In [3], for example, a fast version of ISTA is proposed (FISTA, described in more detail below in section 3) that has as easy an implementation as (1.3) but has a much better convergence rate.

In this paper, we present an ordinary differential equation (ODE) approach to an ISTA for solving $l_1$-minimization problems independent of the regularization parameter $\tau$. We use a generalized Lagrange multiplier, or augmented Lagrangian, approach [4, 30, 32, 40, 42] to the relaxed problem (1.2) to impose exact recovery of the solution to (1.1). The exact solution is sought through an efficient algorithm obtained from a min–max variational principle, which is a special case of the Arrow–Hurwitz–Uzawa algorithm [1]. We prove that this algorithm yields the exact solution for all values of the parameter $\tau$. Our only assumption is that the matrix $A$ has full rank. The performance of the algorithm, with and without noise in the data, is explored through several numerical simulations of array imaging problems.

The connection of the ODE method to the iterative shrinkage algorithm is provided by the Crandall–Liggett theory [19], which analyzes the convergence of an implicit finite difference discretization of the ODE. The theory works for infinite-dimensional, monotone nonlinear problems as well. A detailed discussion and analysis of monotone operators and splitting methods can be found in [5, 2, 41].

The min–max variational principle used here is also behind the Bregman and linearized Bregman iterative algorithms [31, 39, 48, 49]. The fully implicit version of the algorithm is also analyzed in detail in [17, 27] using different techniques. Many other methods have been proposed in the literature to solve (1.1) and (1.2). We just mention here some of them: orthogonal matching pursuit [10], homotopy [26, 38, 45], interior-point methods [47], gradient projection [29], sub-gradient descent methods in primal and dual spaces [7, 35] and proximal gradient in combination with iterative shrinkage thresholding [3, 36, 37].

Finding the constrained, minimal $l_1$-norm solution in (1.1) does not imply that this solution vector has minimal support, even though the $l_1$ norm is sparsity promoting. Nevertheless, in many applications, in imaging in particular, this optimization method does produce the minimal
support, or the minimal $l_0$ norm solution. The theory of compressed sensing \cite{11-14, 24, 25, 46} gives conditions under which constrained $l_1$- and $l_0$-minimizations are equivalent. We do not address this issue here.

The paper is organized as follows. In section 2, we motivate our approach, summarize our main results and describe the numerical algorithm. Theorems 2.6 and 2.4 are the main results of this paper. A key ingredient in the proof of these theorems is theorem 2.7 proved in section 4. The proof of the variational principle of theorem 2.2 is presented in section 6. This result is originally due to \cite{42} but we present it here for the convenience of the reader. In section 3, we show the performance of the algorithm with and without noise in the data using some numerical experiments of array imaging. Finally, section 7 contains conclusions.

2. Formulation and main results

We consider the constrained optimization problem (1.1) under the assumptions that (1.1) has a unique minimizer $\hat{x}$, and that $A$ has full rank, so the matrix $AA^*$ is invertible.

2.1. The min–max variational principle

In order to find the minimizer $\hat{x}$, we recall the variational formulation of the $l_1$-minimization problem \cite{4, 32, 40, 42}. Define the function

$$F(x, z) = \tau \|x\|_1 + \frac{1}{2} \|Ax - y\|^2 + \langle z, y - Ax \rangle,$$

for $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, and set

$$\bar{F} = \max_x \min_z \{F(x, z)\}. \tag{2.1}$$

**Proposition 2.1.** Problem (2.1) has a solution, that is $-\infty < \bar{F} < +\infty$, and the max–min is attained.

**Proof.** The function $F(x, z)$ is convex in $x$, and $\lim_{x \to +\infty} F(x, z) = +\infty$, for any fixed $z$. Thus, $F(x, z)$ attains its minimum for a fixed $z$. Let us denote

$$l(x) = \tau \|x\|_1 + \frac{1}{2} \|Ax - y\|^2, \tag{2.2}$$

and

$$h(z) = \min_x F(x, z) = \min_{x} \{l(x) + \langle z, y - Ax \rangle\}. \tag{2.3}$$

As the function $l(x)$ is convex, and $l(x) \to +\infty$, as $|x| \to \infty$, it follows that $h$ is concave, as a minimum of affine functions, and $h(z) \to -\infty$, as $|z| \to \infty$. Thus, it attains its maximum $\max_z h(z)$. \hfill \Box

In order to motivate the functional (2.1) we look at another natural way to impose the constraint in (1.1) by using a Lagrange multiplier. If we consider a functional

$$\tau \|x\|_1 + \langle z, y - Ax \rangle, \tag{2.4}$$

then (at least, formally) its Euler–Lagrange equations for the extremum give us the sub-differential optimality condition

$$[A^* z] = \begin{cases} \tau, & \text{if } \hat{x}_i > 0, \\ -\tau, & \text{if } \hat{x}_i < 0, \end{cases} \text{ and } |[A^* z]| \leq \tau \text{ if } \hat{x}_i = 0. \tag{2.5}$$

It is, however, difficult to work with (2.4), because if some of the entries of $A^* z$ are larger than $\tau$ in absolute value, then (2.4) is not bounded from below as a function of $x$. Furthermore,
even if $z$ is chosen according to the sub-differential condition (2.5), then the minimum may not be unique, even if $A$ is invertible. Indeed, consider a simple example: minimize $|x|$ subject to $x = 1$. Suppose $\tau = 1$, then (2.4) is $|x| + z(1 - x)$. Then $z = 1$ satisfies the sub-differential condition, and (2.4) becomes

$$|x| + (1 - x) = \begin{cases} 1, & \text{if } x > 0, \\ 1 - 2x, & \text{if } x < 0, \end{cases}$$

which has no unique minimum. The addition of a quadratic term to (2.4) regularizes this degeneracy. Since the function $I(x)$ in (2.2) is convex, (2.3) may be interpreted (up to a sign) as a generalized Legendre transform of $I(x)$.

The first observation is that if (1.1) has a unique minimum $\hat{x}$, then the variational principle (2.1) finds $\hat{x}$ exactly.

**Theorem 2.2.** Assume that (1.1) has a unique minimum $\hat{x}$. Then we have

$$\tau \| \hat{x} \|_1 = \max_{z} \min_{x} F(x, z). \quad (2.6)$$

Moreover, we have $\tau \| \hat{x} \|_1 = F(\hat{x}, z)$ for any $z$, and if $\min_{z} F(x, z) = \tau \| \hat{x} \|_1$ for some fixed $z$, then Argmin $F(x, z) = \hat{x}$.

This result can be found in [42] in a much greater generality. We present its proof below in the particular case we are interested in for the convenience of the reader.

It is remarkable that (2.6) holds for any value of $\tau > 0$—this gives us a freedom to choose $\tau$ large or small, depending on a particular application. We also have the following well-known result [42], which follows from the proof of theorem 2.4 below.

**Theorem 2.3.** Assume that (1.1) has a unique minimizer $\hat{x}$. Then, there exists a vector $z$ such that $[A^* z]_i = \text{sgn}(\hat{x})$ if $\hat{x}_i \neq 0$, and $\|A^* z\| \leq 1$ if $\hat{x}_i = 0$.

We say that $z$ satisfies the sub-differential condition if there exists a minimizer of (1.1) such that

$$[A^* z]_i = \tau \text{sgn}(\hat{x}) \text{ if } \hat{x}_i \neq 0, \text{ and } \|A^* z\| \leq \tau \text{ if } \hat{x}_i = 0. \quad (2.7)$$

We note that (2.7) is weaker than the sub-differential condition of [12]—there it is required that $\|A^* z\| < \tau$ if $\hat{x}_i = 0$, while we do not require the strict inequality. It follows from the proof of theorem 2.3 that the exact extremum of $F(x, z)$ is achieved for any $z$ that satisfies the sub-differential condition (2.7). Such $z$ is not unique but, of course, our interest is not in finding $z$ but in finding the minimizer of (1.1).

### 2.2. The ordinary differential equations method

In order to find $\hat{x}$, ideally, we would like to take the ODE point of view and generate a trajectory $(x(t), z(t))$ of the following system:

$$\frac{dx}{dt} = -\nabla_x F(x, z), \quad \frac{dz}{dt} = \nabla_z F(x, z). \quad (2.8)$$

with the hope that $x(t) \to \hat{x}$ as $t \to +\infty$. There is an obvious degeneracy in the problem, namely, $F(\hat{x}, z) = \tau \| \hat{x} \|_1$ for all $z \in \mathbb{R}^n$. Hence, we can only hope to recover $\hat{x}$ as there is no ‘optimal’ $z$.

The obvious technical difficulty is that the function $F(x, z)$ is not differentiable in $x$ at the points where $x_j = 0$ for some $j = 1, \ldots, n$. Following [19], we interpret solutions of (2.8) as follows. Given $x \in \mathbb{R}^n$, let the sub-differential \( \partial \|x\|_1 \) be a subset of $\mathbb{R}^n$:

$$\partial \|x\|_1 = \text{sgn}(x_1) \times \cdots \times \text{sgn}(x_n).$$
Here, \( \text{sgn}(s) \), for \( s \in \mathbb{R} \), is understood to be a subset of \( \mathbb{R} \); \( \text{sgn}(s) = \{1\} \) if \( s > 0 \), \( \text{sgn}(s) = \{-1\} \) if \( s < 0 \) and \( \text{sgn}(s) = [-1, 1] \) if \( s = 0 \). Then, instead of treating the system of ODEs (2.8) with a discontinuous right-hand side, we consider
\[
\frac{dx}{dt} - A^*(z - Ax + y) \in -\tau \hat{\partial} \|x\|_l,
\]
\[
\frac{dz}{dt} = y - Ax,
\]
(2.9)
supplemented by the initial data \( x(0) = x_0 \) and \( z(0) = 0 \). We say that \((x(t), z(t))\) is a strong solution to (2.9) on a time interval \( 0 \leq t \leq T \) if \( x(t) \) and \( z(t) \) are continuous, differentiable for almost all \( t \in [0, T] \), \( x(0) = x_0 \), \( z(0) = 0 \), and (2.9) holds for almost all \( t \in [0, T] \).

An important observation is that (2.9) is contractive, or, accretive in the sense of Crandall and Liggett [19]. That is, the following property holds: given any pair \((x_1, z_1)\), \((x_2, z_2)\) and any \( \xi_1 \in \hat{\partial} \|x_1\|_l \), \( \xi_2 \in \hat{\partial} \|x_2\|_l \), we have
\[
\langle A^*(z_1 - Ax_1) - \tau \xi_1 - A^*(z_2 - Ax_2), \xi_2 \rangle \geq (\langle z_1 - z_2, \xi_2 \rangle - \langle z_1 - z_2, \xi_2 \rangle)^2 \leq 0.
\]
(2.10)
The last inequality above follows from the component-wise monotonicity of the subdifferential \( \hat{\partial} \|x\|_l \). It follows from (2.10) and theorems I and II of [19] that (2.9) has a unique strong solution. Our first result shows that this solution converges as \( t \to +\infty \) to \( \bar{x} \), the minimizer of (1.1).

**Theorem 2.4.** Let (1.1) have a unique minimizer \( \bar{x} \). Then, for any \( \delta > 0 \) there exists \( T = T(\delta) \) such that the solution of (2.9) satisfies
\[
\|x(t) - \bar{x}\| < \delta, \text{ for all } t > T.
\]
(2.11)
The time \( T(\delta) \) depends only on \( \delta \), the initial data \( x_0 \) and \( \|AA^*\| \) but not on the dimension \( n \).

We note that theorem 2.4 follows from the existence of \( z \) (theorem 2.3 or [42]) and the fact that (2.9) is contractive, see e.g. [9]. Our proof of theorem 2.4, however, does not rely on the existence of \( z \).

### 2.3. The discrete algorithm

Crandall–Liggett theory [19] implies that (2.9) should not be discretized explicitly. Hence, a possible strategy to solve (2.9) numerically is to consider the following implicit scheme:
\[
\frac{x_{k+1} - x_k}{\Delta t} = -\tau \xi_{k+1} + A^*(z_{k+1} + y - Ax_{k+1}),
\]
\[
\frac{z_{k+1} - z_k}{\Delta t} = y - Ax_{k+1},
\]
(2.12)
with the initial data \( x_0 = x \) and \( z_0 = 0 \). Here, \( \xi_{k+1} \) is a vector in the set \( \hat{\partial} \|x_{k+1}\|_l \).

A simple way to understand how (2.12) works is to consider the toy problem
\[
\dot{r} = -\text{sgn}(r).
\]
(2.13)
An explicit discretization
\[
\frac{r_{k+1} - r_k}{\Delta t} = -\xi_k,
\]
with \( \xi_k \in \text{sgn}(r_k) \), will start oscillating around \( r = 0 \) as soon as \( r_k \in [-\Delta t, \Delta t] \), and will never converge to \( x = 0 \) for \( \Delta t > 0 \). On the other hand, the implicit discretization
\[
\frac{r_{k+1} - r_k}{\Delta t} = -\xi_{k+1},
\]
(2.14)
with $\xi_{k+1} \in \text{sgn}(r_{k+1})$ behaves differently. If $r_k \in [-\Delta t, \Delta t]$, then the implicit nature of this scheme shows that it is impossible to have $\xi_{k+1} = \pm 1$, which forces $\xi_{k+1} = r_k/\Delta t$ and $r_{k+1} = 0$. The implicit scheme is actually equivalent to soft thresholding:

$$r_{k+1} = \eta_{\Delta t}(r_k).$$

The function $\eta_{\Delta t}$ here is defined by (1.4). This simple example already shows both the importance of using an implicit discretization and that the implicit scheme has a simple explicit realization (2.15).

Theorems I and II of [19] not only provide existence of a strong solution to (2.9) but also show that it can be found by the implicit scheme (2.12).

**Proposition 2.5.** Solution of (2.12) converges as $\Delta t \to 0$, uniformly on finite time intervals, to the unique strong solution of (2.9).

Theorem 2.4 and proposition 2.5 together imply immediately the following theorem.

**Theorem 2.6.** Let the sequence $x_n, z_n$ solve (2.12) with the initial data $x_0 = x, z_0 = 0$. Given any $\delta > 0$ there exists $h > 0$ and $T > 0$, so that for all $0 < \Delta t < h$ and all $k > [T/\Delta t]$ we have $|x_k - x| < \delta$. The time $T$ depends on $\delta$, the initial data $x \in \mathbb{R}^n$, and the norm $\|A^*x\|$.

Note that the scheme (2.12) is expensive to implement because one needs to solve a linear system of equations at each iteration. If, however, one examines the proof of theorems I and II in [19], it is clear that the only term that should be discretized implicitly is $\text{sgn}$—the other terms can be discretized explicitly, keeping the statement of proposition 2.5 intact. Hence, the result of theorem 2.6 applies equally well to an Euler quasi-explicit modification of (2.12):

$$x_{k+1} = x_k - \xi_{k+1} + \Delta t A^*(z_k + y - Ax_k),$$

$$z_{k+1} = z_k + \Delta t (y - Ax_k),$$

where $\xi_{k+1} \in \tau \Delta t \partial ||x_{k+1}||_l$ is a vector in the subdifferential of $\tau \Delta t ||x_{k+1}||_l$. As in the toy problem (2.13)–(2.15), it is equivalent to soft thresholding:

$$x_{k+1} = \eta_{\tau \Delta t}(x_k + \Delta t A^*(z_k + y - Ax_k)),$$

$$z_{k+1} = z_k + \Delta t (y - Ax_k).$$

We call (2.17) the generalized Lagrangian multiplier algorithm (GeLMA). This scheme is simple to implement numerically. The key observation to achieve this simplicity is to use implicit discretization for the sub-differential, and explicit discretization for everything else. Thus (2.17) only requires matrix–vector multiplication followed by soft thresholding. The GeLMA algorithm converges if $\Delta t < 1/||A||$—that condition simply comes from the usual constraint for an explicit scheme for a linear system.

We also note that one can mimic the ODE proof of theorem 2.4 directly on the numerical scheme, eliminating, in particular, the dependence of $h$ on $\delta$. Our objective, however, in part, is to explain the effectiveness of shrinking-thresholding algorithms in the language of differential equations, potentially opening the way for the application of other continuous techniques in such problems. Therefore, we have chosen to concentrate on the ODE proof.

### 2.3.1. The regularized ordinary differential equations.

Since the system (2.9) has a ‘bad’ right-hand side, working with it directly is technically inconvenient. Hence, in order to prove theorem 2.4, from which theorem 2.6 follows, we consider a regularized system, introducing a single-valued approximation of $\text{sgn}$:

$$G_\varepsilon(s) = \begin{cases} 
1, & \text{if } s > \varepsilon, \\
\frac{s}{\varepsilon}, & \text{if } |s| < \varepsilon, \\
-1, & \text{if } s < -\varepsilon.
\end{cases}$$

6
Here $\varepsilon > 0$ is a small regularization parameter that will be sent to zero at the end. With a slight abuse of notation, here, and in other instances when this should cause no confusion, we will also denote by $G_\varepsilon(x)$ a vector-valued function with components $G_\varepsilon(x) = (G_\varepsilon(x_1), G_\varepsilon(x_2), \ldots, G_\varepsilon(x_n))$. The regularized version of (2.9) is

$$\frac{dz_\varepsilon}{dt} = -\tau G_\varepsilon(x_\varepsilon) + A^*(z_\varepsilon + y - Ax_\varepsilon), \quad \frac{dx_\varepsilon}{dt} = y - Ax_\varepsilon.$$  \hspace{1cm} (2.18)

It has the same form as (2.8), with $F(x, z)$ being replaced by a differentiable approximation

$$F_\varepsilon(x, z) = \tau \sum_{j=1}^n r_\varepsilon(x_j) + f(x) + \langle z, y - Ax \rangle.$$  \hspace{1cm} (2.19)

Here,

$$r_\varepsilon(s) = \begin{cases} |s|, & \text{if } s > \varepsilon, \\ s^2/(2\varepsilon) + \varepsilon/2, & \text{if } |s| < \varepsilon, \\ |s|, & \text{if } s < -\varepsilon \end{cases}$$

is an approximation of $|s|$ known as the Huber function. We will denote below

$$\|x\|_\varepsilon = \sum_{j=1}^n r_\varepsilon(x_j).$$  \hspace{1cm} (2.20)

though, of course, this is not a norm as it does not vanish at $x = 0$.

**Theorem 2.7.** Let (1.1) have a unique minimizing $\tilde{x}$. Then, for any $\delta > 0$ there exists $\varepsilon_0 = \varepsilon_0(\delta, n)$ and $T = T(\delta)$ such that for any $\varepsilon, 0 < \varepsilon < \varepsilon_0$, the solution of (2.18) satisfies

$$\|x_\varepsilon(t) - \tilde{x}\| < \delta, \text{ for all } t > T.$$  \hspace{1cm} (2.21)

The time $T(\delta)$ depends only on $\delta$, the initial data $x_0$ and $\|AA^*\|$ but not on the dimension $n$.

When the minimizer of (1.1) is not unique, the proof of theorem 2.7 can be easily adapted to show that for any $\delta > 0$ there exists $\varepsilon_0(\delta)$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and any limit point (as $t \to +\infty$) $\tilde{x}_\varepsilon$ of the trajectory $x_\varepsilon(t)$, we have $\|\tilde{x}_\varepsilon - \tilde{x}\| < \delta$ for some minimizer $\tilde{x}$ of (1.1).

Theorem 2.7 is the key ingredient in the proof of theorem 2.6: together with *a priori* bounds on $x_\varepsilon(t)$ obtained in the course of its proof, they show that solution $x(t)$ of (2.9) is the limit of $x_\varepsilon(t)$ as $\varepsilon \to 0$, and thus it obeys the same bounds as $x_\varepsilon(t)$, finishing the proof.

### 3. Application to array imaging

In this section, we illustrate the performance of our algorithm for array imaging of localized scatterers. The problem is to determine the location and reflectivities of small scatterers by sending a narrow band (single frequency) probing signal of wavelength $\lambda$ from an active array and recording the backscattered field on this array [6]. In this paper, we consider only single illumination by the central element of the array.

#### 3.1. Array imaging in homogeneous media

The array has $N$ transducers located at positions $x_p$ ($p = 1, \ldots, N$) separated from each other by a given distance. In each numerical experiment, there are $M$ point-like scatterers of unknown reflectivities $\rho_j > 0$ located at unknown positions $y_j$ ($j = 1, \ldots, M$). The scatterers are assumed to be within a bounded region at a distance $L$ from the array, called the image window.
(IW). We discretize this IW with a uniform mesh of $K$ points $y_j$ ($j = 1, \ldots, K$), and assume that each scatterer is located at one of these $K$ grid points, so $\{y_{n_1}, \ldots, y_{n_K}\} \subset \{y_1, \ldots, y_K\}$.

Furthermore, we assume that the medium between the array and the scatterers is homogeneous so wave propagation between any two points $x$ and $y$ is modeled by the free space Green function

$$\hat{G}_0(x, y, \omega) = \frac{\exp(-i\kappa |x - y|)}{4\pi |x - y|},$$

(3.1)

where $\kappa = \omega/c = 2\pi/\lambda$ and $c$ is the reference wave speed in the medium. We also assume that the scatterers are well separated or are weak, so multiple scattering among them is negligible (this is the Born approximation). Under these conditions, the backscattered field measured at $x_i$ due to a pulse sent from $x$, and reflected by the $M$ scatterers in the IW, is given by

$$b_i(\omega) = \sum_{j=1}^{M} \rho_j \hat{G}_0(x_i, y_j, \omega) \hat{G}_0(y_j, x_i, \omega).$$

(3.2)

Next, we write the linear system that relates the reflectivity $\rho_j$ at each grid point $y_j$ of the IW ($j = 1, \ldots, K$) and the data $b_i(\omega)$ measured at the array ($i = 1, \ldots, N$). To this end, we introduce the reflectivity vector $\rho_0 = (\rho_0, \rho_2, \ldots, \rho_K)^T \in \mathbb{R}^K$ and the data vector $b(\omega) = (b_1, b_2, \ldots, b_N)^T \in \mathbb{R}^N$, where the superscript $T$ means transpose. Thus, the image is a gridded array of $K$ pixels, and the data are stacked into a vector of $N \ll K$ components. Furthermore, there are only a few scatterers in the IW so the vector $\rho_0$ is sparse.

Let us consider the vector

$$\hat{\mathbf{g}}_0(y_j, \omega) = (\hat{G}_0(x_1, y_j, \omega), \hat{G}_0(x_2, y_j, \omega), \ldots, \hat{G}_0(x_N, y_j, \omega))^T$$

that represents the signal at the array due to a point source at $y_j$ in the IW. Due to the spatial reciprocity $\hat{G}_0(x_i, y_j, \omega) = \hat{G}_0(y_j, x_i, \omega)$, it can also be interpreted as the illumination vector of the array at position $y_j$. With this notation, we can write the linear system

$$A_w \rho_0 = b(\omega).$$

(3.3)

where $A_w$ is an $N \times K$ matrix whose $j$th column is given by $\hat{G}_0(y_j, x_i, \omega) \hat{\mathbf{g}}_0(y_j, \omega)$. Since $N \ll K$, (3.3) is an underdetermined linear system, and hence there can be many configurations of scatterers that match the data vector $b(\omega)$. Array imaging is to solve (3.3) for $\rho_0$.

A related problem to (3.3) has been studied in [15] in array imaging of localized scatterers from intensity-only measurements. Intensity measurements are interpreted as linear measurements of a rank 1 matrix associated with the unknown reflectivities. Since the rank minimization problem is NP hard, it is replaced by the minimization of the nuclear norm of the decision matrix. This makes the problem convex and solvable in polynomial time. It is shown that exact recovery can be achieved by minimizing this problem.

### 3.2. Numerical simulations

We consider here numerical experiments in 2D. Our linear array consists of 100 transducers that are one wavelength apart. Hence, the aperture of the array is $a = 100$. In each numerical experiment there are a few point-like scatterers of different reflectivities at a distance 120 from the array. The IW is discretized with 41 $\times$ 41 grid points. Hence, we have 1681 unknowns and 100 measurements. All the spatial units are expressed in units of the wavelength $\lambda$ of the illuminating source.

Figure 1 shows results from various scatterer’s configurations with no noise in the data. In the top row, we display the original scatterer’s configurations and in the bottom row the corresponding images obtained by the $\ell_1$-minimization GeLMA algorithm (2.17). These
results show that this algorithm recovers the positions and reflectivities of the scatterers exactly when there is no noise in the data. To examine this issue more clearly we plot in figure 2 the vector solutions $\rho$ (small red circles) and the exact vectors $\rho_0$ (large blue circles) for these three scatterer’s configurations. There is no apparent difference between the exact and recovered solutions. Both, localization (support recovery) and strength estimation (reflectivities) are solved exactly in all the cases.

An interesting feature of the GeLMA algorithm (2.17) is that it attains the exact solution of the basis pursuit problem for large values of the regularization parameter $\tau$. This speeds up the convergence rate. Informally, this speed-up of convergence can be seen from the coercivity estimate (2.10) and the error estimate (4.7). Note that for other popular gradient-based algorithms, such as ISTA or FISTA [3], $\tau$ has to be smaller than $\|A^T\omega b\|_{\ell_\infty}$. Otherwise, they converge to the (maximally sparse) zero solution $\rho = 0$. To examine this property in more detail, we show in figure 3 (left panel) the plots of the $\ell_2$ distance to the exact solution $\|\rho - \rho_0\|$ as a function of the iteration number for various values of $\tau = \alpha \|A^T\omega b\|_{\ell_\infty}$: $\alpha = 2$ (solid line), $\alpha = 5$ (dashed line), $\alpha = 10$ (dot-dashed line) and $\alpha = 20$ (dotted line). We observe
that the larger the value of $\tau$, the faster the convergence rate. Furthermore, for all the values of $\tau$, the algorithm achieves the exact solution $\rho_0$.

In figure 3 (right panel) we compare the convergence rates of the GeLMA algorithm and the FISTA algorithm

$$\rho^{(k)} = \eta \tau \xi^{(k)} - \alpha \nabla f(\xi^{(k)}), \quad (3.4)$$

$$\alpha_{k+1} = \frac{1 + \sqrt{1 + 4\alpha^2}}{2}, \quad (3.5)$$

$$\xi^{(k+1)} = \rho^{(k)} + \frac{\alpha - 1}{\alpha_{k+1}}(\rho^{(k)} - \rho^{(k-1)}), \quad (3.6)$$

for $\tau = 0.01\|A^T_b(\omega)\|_{l_2}$. We choose a small value of $\tau$ because we are considering noiseless data in these examples. In (3.4)-(3.6), $\rho_i$ and $\xi_i = \rho_i$ are given, and $\alpha_0 < 2/L$. We observe that the convergence rate of the FISTA algorithm (solid line) for $\tau = 0.01\|A^T_b(\omega)\|_{l_2}$ is much slower than the convergence rate of the GeLMA algorithm for $\tau = 20\|A^T_b(\omega)\|_{l_2}$. Even more, the FISTA algorithm with $\tau = 0.01\|A^T_b(\omega)\|_{l_2}$ does not obtain the exact solution. To achieve the exact solution, we would have to let $\tau \to 0$.

Next, we examine the performance of the GeLMA algorithm under noise contaminated data $b(\omega) + e(\omega)$. The noise vector $e(\omega)$ is generated by independent Gaussian random variables with zero mean and standard deviation $\beta \|b(\omega)\|_{l_2}/\sqrt{N}$. Here, $\beta$ is a parameter that measures the noise strength. In figure 4, we show the results for $\beta = 0.05$ (left column), $\beta = 0.1$ (middle column) and $\beta = 0.3$ (right column). For a fixed step size $\Delta t$, the regularization parameter $\tau = \alpha \|A^T_b(\omega)\|_{l_2}$ controls the sparsity of the solution. Hence, one expects the algorithm to be more stable with respect to additive noise when $\tau$ is large. We plot in figure 4 the recovered images using different values of $\tau$: $\alpha = 2$ (top row), $\alpha = 20$ (middle row) and $\alpha = 200$ (bottom row). We observe in the top row that the location of the scatterers is recovered exactly when there is 5% noise in the data (left plot). The recovered reflectivities are also quite
Figure 4. Impact of the regularization parameter $\tau = \alpha \| A^T b(\omega) \|_{l_\infty}$ on the reconstructions for different amounts of noise in the data. Top row: recovered images with $\alpha = 2$ and 5% noise (left), 10% noise (middle) and 30% noise (right). Middle row: the same as the top row but for $\alpha = 20$. Bottom row: the same as the top row but for $\alpha = 200$.

close to the real ones. However, when the noise increases to 10% (middle plot), one scatterer is missing in the recovered image that also shows some ghost scatterers. As expected, the image gets worse when the noise is 30%, as can be seen in the right plot. The results are much better when we increase the value of $\alpha$ to 20 (middle row). With 5% noise in the data (left plot) both the location and reflectivities of the scatterers are very close to the real ones. Even with 10% noise in the data (middle plot), we can determine the location of the four scatterers. However, with 30% noise we miss the fourth scatterer. Finally, we show the recovered images using $\alpha = 200$ in the bottom row. For 5% and 10% noise (left and middle images, respectively), the location of the scatterers is exact. Furthermore, the recovered reflectivities are very sharp. However, we still miss the location of one scatterer when there is 30% noise in the data, as can be seen in the right image of the bottom row of this figure. We plan to investigate in detail the robustness of the algorithm with respect to noise in a future publication.

4. Proofs of theorems 2.4, 2.3 and 2.7

Theorems 2.4 and 2.3 are easy consequences of theorem 2.7 and its proof.
4.1. Outline of the proof of theorem 2.7

Let \(\tilde{x}\) be the unique minimizer of (1.1) We write \(x_\varepsilon = \tilde{x} + q_\varepsilon\) and obtain

\[
\frac{d q_\varepsilon}{dt} = -\tau G_\varepsilon (\tilde{x} + q_\varepsilon) + A^* (z_\varepsilon - A q_\varepsilon), \quad \frac{dz_\varepsilon}{dt} = -A q_\varepsilon.
\]  

(4.1)

Our goal is now to show that \(q_\varepsilon(t) \to 0\) as \(t \to +\infty\). If we take the time-derivative of the first equation in (4.1), and use the second equation, we obtain

\[
\dot{q}_\varepsilon + A^* A (q_\varepsilon + q_\varepsilon) = -\tau g^\varepsilon (\tilde{x} + q_\varepsilon) \dot{q}_\varepsilon.
\]  

(4.2)

Here, \(g^\varepsilon(x)\) is a diagonal matrix with the entries on the main diagonal given by

\[
g^\varepsilon_i(x) = \begin{cases} 0, & \text{if } |x_i| > \varepsilon, \\ 1/\varepsilon, & \text{if } |x_i| < \varepsilon. \end{cases}
\]  

(4.3)

Note that (4.2) is simply an equation for an oscillator with friction and a forcing term on the right-hand side. As the matrix \(A^* A\) is singular, the oscillator is degenerate. Therefore, it is reasonable to expect that the friction term \(A^* A \dot{q}_\varepsilon\) in (4.2) by itself would ensure that \(A q_\varepsilon(t) \to 0\) as \(t \to +\infty\), provided that the forcing does not interfere. However, the friction alone cannot send \(q_\varepsilon(t)\) to zero since it is degenerate. Moreover, in showing that \(q_\varepsilon(t)\) becomes small as \(t \to +\infty\), one has to use the fact that \(\tilde{x}\) is the minimizer of (1.1) and not just any solution of \(Ax = y\). The strategy of the proof is (i) to establish uniform bounds on \(q_\varepsilon(t)\) and \(z_\varepsilon(t)\), and (ii) show that any limit point of \(q_\varepsilon(t)\) as \(t \to +\infty\) is close to zero.

The \textit{a priori} bounds are obtained in several steps. We first describe the required intermediate lemmas, and present their proofs later. The first step in the proof is the following lemma that provides a Lyapunov function for (1.1), and establishes a bound on \(\|A q_\varepsilon(t)\|\).

**Lemma 4.1.** There exists a constant \(C_0 > 0\) that is independent of \(\varepsilon\) (and depends only on the initial data \(x_0\)) so that

\[
\|\dot{q}_\varepsilon(t)\|^2 + \|A q_\varepsilon(t)\|^2 + \int_0^t \|A q_\varepsilon(s)\|^2 \, ds < C_0,
\]  

(4.4)

for all \(\varepsilon < \varepsilon_0\) and all \(t > 0\).

The bound on \(\|A q_\varepsilon\|\) in lemma 4.1 leads to a uniform bound on \(z_\varepsilon(t)\).

**Lemma 4.2.** There exists a constant \(C > 0\) that is independent of \(\varepsilon > 0\) so that \(\|z_\varepsilon(t)\| < C\) for all \(t > 0\).

The next step is to show that \(A q_\varepsilon(t)\) is small for large times. Since \(\dot{z}_\varepsilon = -A q_\varepsilon\), it follows from lemma 4.2 that

\[
\int_{t_1}^{t_2} A q_\varepsilon(s) \, ds
\]  

is uniformly bounded for all \(t_{1,2} > 0\). Together with the integral bound on \(A q_\varepsilon(t)\) in lemma 4.1, this shows that \(A q_\varepsilon(t)\) becomes small at some ‘not too large’ time.

**Lemma 4.3.** There exist two constants \(C_{1,2} > 0\) that are independent of \(\varepsilon \in (0, \varepsilon_0)\) so that for any \(k \in \mathbb{N}\) there exists a time \(t_k < C_1 k^3\) such that for all \(t \in (t_k, t_k + C_2 k)\) we have \(\|A q_\varepsilon(t)\| < 1/k\) for all \(\varepsilon < \varepsilon_0\).

Next, using the bounds in lemmas 4.1 and 4.2, as well as the precise form of the forcing term in (4.2), we obtain a uniform bound for \(\|q_\varepsilon(t)\|\).

**Lemma 4.4.** There exists a constant \(C > 0\) so that we have

\[
\|\tilde{x} + q_\varepsilon(t)\| \leq C,
\]  

(4.5)

for all \(t > 0\) and all \(\varepsilon > 0\).
The bound on $\|q_\varepsilon(t)\|$ allows us to strengthen lemma 4.3 to include a bound on $\dot{q}_\varepsilon(t)$ ‘at some times’ as well.

**Lemma 4.5.** There exists a constant $C > 0$ that is independent of $\varepsilon \in (0, \varepsilon_0)$ so that for any $k \in \mathbb{N}$ there exists a time $s_k < Ck^3$ such that $\|Aq_\varepsilon(s_k)\| + \|\dot{q}_\varepsilon(s_k)\|^2 \leq C/k$ for all $\varepsilon < \varepsilon_0$.

The Lyapunov function in lemmas 4.1 and 4.5 together imply that $\dot{q}_\varepsilon(t)$ and $\ddot{q}_\varepsilon(t)$ are not only ‘small sometimes’ but rather tend to zero as $t \to +\infty$.

**Corollary 4.6.** There exists a constant $C > 0$ that is independent of $\varepsilon \in (0, \varepsilon_0)$ so that for any $n \in \mathbb{N}$ there exists a time $s_n = s_n(\varepsilon) < Cn^3$ such that $\|Aq_\varepsilon(s_n)\|^2 + \|\dot{q}_\varepsilon(s_n)\|^2 \leq C/n$ for all $\varepsilon < \varepsilon_0$ and all $s > s_n$.

Corollary 4.6 shows that the right-hand side of the ODE system (4.1) is small as $t \to +\infty$. The final step in the proof is to show that this implies that $q_\varepsilon(t)$ is small, and it is here that the condition that $\bar{x}$ is the minimizer of (1.1) comes into play.

### 4.2. The end of the proof of theorem 2.7

It follows from corollary 4.6 that for any $\delta_0 > 0$ there exist $T = T(\delta_0)$ and $\varepsilon_0 = \varepsilon_0(\delta_0)$,

$$
\|A^\varepsilon z_\varepsilon(t) - \tau G_\varepsilon(\bar{x} + q_\varepsilon(t))\| \leq \delta_0, \quad \|Aq_\varepsilon(t)\| \leq \delta_0,
$$

for all $\varepsilon \leq \varepsilon_0$ and $t > T$. The first inequality in (4.6) implies

$$
\|A^\varepsilon z_\varepsilon(t) - \tau G_\varepsilon(\bar{x} + \bar{q}_\varepsilon(t)), \bar{x} + q_\varepsilon(t)\| \leq \delta_0 \|\bar{x} + q_\varepsilon(t)\|.
$$

Using the second inequality from (4.6) in

$$
\|\langle A^\varepsilon z_\varepsilon(t), \bar{x} + q_\varepsilon(t) \rangle - \langle A^\varepsilon z_\varepsilon(t), \bar{x} \rangle\| \leq \delta_0 \|z_\varepsilon(t)\|
$$

and denoting

$$
\mathcal{N}_\varepsilon(x) = \sum_i x_i G_\varepsilon(x_i),
$$

we obtain

$$
\|\tau \mathcal{N}(\bar{x} + q_\varepsilon(t)) - \langle A^\varepsilon z_\varepsilon(t), \bar{x} \rangle\| \leq \delta_0 (\|z_\varepsilon(t)\| + \|\bar{x} + q_\varepsilon(t)\|).
$$

It also follows from the first inequality in (4.6) that

$$
\|A^\varepsilon z_\varepsilon(t)\|_{\mathcal{H}} \leq \tau + \delta_0,
$$

and thus

$$
\|A^\varepsilon z_\varepsilon(t), \bar{x}\| \leq (\tau + \delta_0) \|ar{x}\|_{\mathcal{H}}.
$$

As a consequence,

$$
\mathcal{N}(\bar{x} + q_\varepsilon(t)) - \|ar{x}\|_{\mathcal{H}} \leq \frac{\delta_0}{\tau} (\|z_\varepsilon(t)\| + \|\bar{x} + q_\varepsilon(t)\|).
$$

and therefore

$$
\|\bar{x} + q_\varepsilon(t)\|_{\mathcal{H}} - \|ar{x}\|_{\mathcal{H}} \leq \frac{\delta_0}{\tau} (\|z_\varepsilon(t)\| + \|\bar{x} + q_\varepsilon(t)\|) + \varepsilon_0 n. \quad (4.7)
$$

Here, $n$ is the dimension of $q_\varepsilon$. As $\bar{x}$ is the unique minimizer, for any $\delta$ we can choose $\alpha$ and $\delta_0$ sufficiently small so that estimates

$$
\|\bar{x} + q_\varepsilon(t)\|_{\mathcal{H}} - \|ar{x}\|_{\mathcal{H}} \leq \alpha, \quad \|Aq_\varepsilon(t)\| \leq \delta_0
$$

$^4$ The quantity $\mathcal{N}_\varepsilon(x)$ plays essentially the same role as $\|x\|_{\mathcal{H}}$ defined in (2.20). They are, however, quantitatively slightly different.
imply that \( \|q_\varepsilon\| < \delta \). Hence it remains to use uniform boundedness of \( \bar{x} + q_\varepsilon(t) \) and \( z_\varepsilon(t) \) and choose \( \delta_0 \) and \( \varepsilon_0 \) so that
\[
\frac{\delta_0}{\tau} (\|\bar{x}\|_\ell + \|z_\varepsilon(t)\| + \|\bar{x} + q_\varepsilon(t)\|) + \varepsilon_0 \leq \alpha.
\]
This finishes the proof of theorem 2.7 except for the proof of lemmas 4.1–4.5 and corollary 4.6.

4.3. Proof of theorem 2.4

Fix \( T_\delta \) such that \( |q_\varepsilon(t)| < \delta \) for all \( T > T_\delta \). We know from the Arzela–Ascoli theorem that \( q_\varepsilon(t) \rightarrow q(t) \) and \( z_\varepsilon \rightarrow z(t) \) uniformly on \([0, T_\delta]\), after extracting a subsequence, and the functions \( q(t) \) and \( z(t) \) are Lipschitz on \([0, T]\), with the Lipschitz constant independent of \( \delta > 0 \). The second equation in (4.1) and the dominated convergence theorem imply that
\[
z(t) = -\int_0^t Aq(s) \, ds,
\]
whence
\[
\dot{z} = -Aq, \quad z(0) = 0. \tag{4.9}
\]
The family \( f_\varepsilon(t) = G_\varepsilon(\bar{x} + q_\varepsilon(t)) \) is uniformly bounded in \( L^2[0, T_\delta] \). Hence, after possibly extracting a subsequence, it converges weakly in \( L^2[0, T_\delta] \) to a limit \( f(s) \). The (vector-valued) function \( f(s) \) satisfies the following properties: (i) \(-1 \leq f_j(t) \leq 1\), for all \( 0 \leq t \leq T_\delta \), \( 1 \leq j \leq N \), and (ii) if \( q_\varepsilon(t) \neq -\bar{x}_j \), then \( f_j(t) = \text{sgn}(\bar{x}_j + q_j) \). It follows that for any \( 0 \leq t_1 < t_2 \leq T_\delta \) we have
\[
q(t_2) - q(t_1) = -\tau \int_{t_1}^{t_2} f(s) \, ds + \int_{t_1}^{t_2} A^*(z(s) - Aq(s)) \, ds. \tag{4.10}
\]
The aforementioned properties of \( f(t) \) imply that \( x(t) = \bar{x} + q(t) \) is a strong solution of (2.9). Uniqueness of the strong solution \([19]\) implies that the whole family \( x_\varepsilon(t) = \bar{x} + q_\varepsilon(t) \), \( z_\varepsilon(t) \) converges to the solution of (2.9). The conclusion of theorem 2.4 now follows from theorem 2.7.

4.4. Proof of theorem 2.3

Theorem 2.7 implies that as \( \varepsilon \rightarrow 0 \) and \( t \rightarrow \infty \), along a subsequence, we have \( \bar{z}_\varepsilon \rightarrow \bar{z} \) and \( q_\varepsilon \rightarrow 0 \). Then the first estimate in (4.6) implies that \( \lambda = \bar{z}/\tau \) satisfies
\[
[\bar{A}^* \bar{x}_j] = \text{sgn} \bar{x}_j, \quad \text{if } \bar{x}_j \neq 0,
\]
\[
-1 \leq [\bar{A}^* \bar{x}_j] \leq 1, \quad \text{if } \bar{x}_j = 0. \tag{4.11}
\]
This completes the proof of theorem 2.3.

5. Proofs of auxiliary lemmas for the proof of theorem 2.7

5.1. Proof of lemma 4.1

Multiplying (4.2) by \( \dot{q}_\varepsilon(t) \) gives
\[
\frac{1}{2} \frac{d}{dt} (\|\dot{q}_\varepsilon(t)\|^2 + \|Aq_\varepsilon(t)\|^2) = -\|A\dot{q}_\varepsilon(t)\|^2 - \tau \langle g' (\bar{x} + q_\varepsilon(t)) \dot{q}_\varepsilon(t), \dot{q}_\varepsilon(t) \rangle. \tag{5.1}
\]
Let
\[ N^e_T = \int_0^T (g^e(\tilde{x} + q_\epsilon(s))\dot{q}_\epsilon(s), \dot{q}_\epsilon(s)) \, ds = \sum_{j=1}^n \int_0^T g^e_j(\tilde{x}_j + q_\epsilon_j(s))|\dot{q}_\epsilon_j(s)|^2 \, ds \geq 0; \] (5.2)
then integrating (5.1) in time, we obtain
\[ \frac{1}{2}(\|\dot{q}_\epsilon(0)\|^2 + \|Aq_\epsilon(0)\|^2) = \frac{1}{2}(\|\dot{q}_\epsilon(T)\|^2 + \|Aq_\epsilon(T)\|^2) + \tau N^e_T + \int_0^T \|A\dot{q}_\epsilon(t)\|^2 \, dt \] (5.3)
and (4.4) follows. Note that \(\|\dot{q}_\epsilon(0)\|\) is uniformly bounded in \(\epsilon > 0\) since the function \(G_\epsilon(s)\) takes values in the interval \([-1, 1]\).

\[ \Box \]

5.2. Proof of lemma 4.2

Differentiating the second equation in (4.1), we obtain
\[ \ddot{z}_\epsilon(t) + AA^* (\ddot{z}_\epsilon(t) + z_\epsilon(t)) = \tau AG_\epsilon (\ddot{x} + q_\epsilon(t)). \] (5.4)
Let us multiply this equation by \(e^t\) and integrate to obtain
\[ \int_0^t e^t \ddot{z}_\epsilon(s) \, ds + e^t AA^* z_\epsilon(t) = \tau A \int_0^t e^s G_\epsilon (\ddot{x} + q_\epsilon(s)) \, ds, \] (5.5)
since \(z(0) = 0\). We estimate, using (4.4),
\[ \left\| \int_0^t e^t \ddot{z}_\epsilon(s) \, ds \right\| \leq \left( \int_0^t e^{2s} \int_0^s \|\ddot{z}_\epsilon(s)\|^2 \, ds \right)^{1/2} \leq Ce^t \left( \int_0^t \|A\dot{q}_\epsilon(s)\|^2 \, ds \right)^{1/2} \leq Ce^t. \]
As \(|G_\epsilon, j| \leq 1\) for all \(1 \leq j \leq n\), we also have
\[ \left\| A \int_0^t e^t G_\epsilon (\ddot{x} + q_\epsilon(s)) \, ds \right\| \leq Ce^t. \]
Since the matrix \(AA^*\) is invertible, we obtain from (5.5) that \(\|z_\epsilon(t)\| \leq C\).

\[ \Box \]

5.3. Proof of lemma 4.3

Let us set \(y_\epsilon(t) = Aq_\epsilon(t)\). As \(z_\epsilon(t)\) is uniformly bounded, there exists a constant \(C_2 > 0\) that is independent of \(\epsilon\) so that
\[ \left\| \int_{t_1}^{t_2} y_\epsilon(s) \, ds \right\| < C_2/2, \] (5.6)
for all \(0 < t_1 < t_2\). If we take an integer \(n_0 = C_2k\), we have
\[ \left\| \frac{1}{n_0} \int_{t}^{t+n_0} y_\epsilon(s) \, ds \right\| < \frac{1}{2k}, \] (5.7)
for all \(t > 0\). Applying the mean-value theorem to each coordinate, we have
\[ \|y_\epsilon(t) - \frac{1}{n_0} \int_{t}^{t+n_0} y_\epsilon(s) \, ds \| \leq \int_{t}^{t+n_0} \|\dot{y}_\epsilon(\xi)\| \, d\xi. \]
Using the Cauchy–Schwartz inequality, we obtain
\[ \|y_\epsilon(t)\| \leq \frac{1}{2k} + \sqrt{n_0} \left( \int_{t}^{t+n_0} \|\dot{y}_\epsilon(s)\|^2 \, ds \right)^{1/2}. \] (5.8)
Lemma 4.1 implies that given \(n_0\) there exist less than \(4C_2k^2n_0\) integers \(l\) such that
\[ \int_{t}^{t+2n_0} \|\dot{y}_\epsilon(s)\|^2 \, ds \geq \frac{1}{4k^2n_0}. \]
It follows that there exists \( t_k < Ck^3 = 4Ck^2n_0 \) such that
\[
\int_{t_k}^{t_k+2n_0} \|\dot{y}_x(s)\|^2 \, ds < \frac{1}{4k^2n_0}.
\]
Then, for all \( t \in (t_k, t_k + n_0) \) we have
\[
\int_{t}^{t+n_0} \|\dot{y}_x(s)\|^2 \, ds < \frac{1}{4k^2n_0},
\]
while
\[
\|y_x(t)\| \leq \frac{1}{2k} + \sqrt{n_0} \sqrt{\frac{1}{4k^2n_0}} = \frac{1}{k}.
\]
(5.9)
for all \( t \in (t_k, t_k + n_0) \). \( \square \)

5.4. Proof of lemma 4.4

Let us recall (4.2):
\[
\ddot{q}_t + A^*A(\dot{q}_t + q_t) = -\tau g^f(\bar{x} + q_t)\dot{q}_t.
\]
(5.10)

Multiply this equation by \( q_t \) and integrate:
\[
\langle q_t, q_t(t) \rangle - (q_t(0), q_t(0)) + \frac{1}{2} \|Aq_t(t)\|^2 - \frac{1}{2} \|Aq_t(0)\|^2 + \int_t^r \|Aq_t(s)\|^2 \, ds = \int_0^s \|\dot{q}_t(s)\|^2 \, ds - \tau \int_0^s \langle g^f(\bar{x} + q_t)\dot{q}_t, q_t \rangle \, ds.
\]
(5.11)

Next, set
\[
v_t(t) = -\int_0^t q_t(s) \, ds,
\]
so that \( z_t(t) = Av_t(t) \) and \( v_t(0) = 0 \). We rewrite (4.1) as
\[
\frac{d}{dt} = -\tau G_t(\bar{x} + q_t) + A^*A(v_t - q_t), \frac{dv_t}{dt} = -A q_t.
\]
(5.12)

Consider the function
\[
Q(t) = \frac{1}{2} \|A(v_t(t) - q_t(t))\|^2 + \tau \|\bar{x} + q_t(t)\|_l^2.
\]
(5.13)

Then we have
\[
\frac{dQ}{dt} = \tau (G_t(\bar{x} + q_t, \dot{q}_t) - \langle A^*A(v_t - q_t), \dot{q}_t \rangle + \langle A^*A(v_t - q_t), \dot{v}_t \rangle)
= -\|\dot{q}_t\|^2 + \frac{1}{2} \frac{d}{dt} \|Av_t\|^2 - \langle A\dot{v}_t, Aq_t \rangle = -\|\dot{q}_t\|^2 + \frac{1}{2} \frac{d}{dt} \|z_t\|^2 + \|\dot{z}_t\|^2.
\]
(14.1)

As \( z_t(0) = 0 \), it follows that
\[
\frac{1}{2} \|A(v_t(t) - q_t(t))\|^2 + \tau \|\bar{x} + q_t(t)\|_l^2 - \frac{1}{2} \|A(v_t(0) - q_t(0))\|^2 - \tau \|\bar{x} + q_t(0)\|_l^2
= \frac{\|z_t(t)\|^2}{2} - \int_0^t \|\dot{q}_t(s)\|^2 \, ds + \int_0^t \|\dot{z}_t(s)\|^2 \, ds.
\]
This can be re-written as
\[
\frac{1}{2} \|z_t(t) - Aq_t(t)\|^2 + \tau \|\bar{x} + q_t(t)\|_l^2 + \int_0^t \|\dot{q}_t(s)\|^2 \, ds = \frac{\|z_t(t)\|^2}{2} + \int_0^t \|Aq_t(s)\|^2 \, ds + C_0.
\]
(5.15)
Adding (5.11) and (5.15) gives

\[ (q_\epsilon(t), \dot{q}_\epsilon(t)) + \frac{1}{2} \| A q_\epsilon(t) \|^2 + \frac{1}{2} \| z_\epsilon(t) - A q_\epsilon(t) \|^2 + \tau \| \bar{x} + q_\epsilon(t) \|_l^2 \]

\[ = -\tau \int_0^t \langle g^\prime (\bar{x} + q_\epsilon) \dot{q}_\epsilon, q_\epsilon \rangle \, ds + \frac{\| z_\epsilon(t) \|^2}{2} + C_0, \tag{5.16} \]

with the constant \( C_0 \) that only depends on the initial data. Then lemmas 4.1 and 4.2 imply

\[ (q_\epsilon(t), \dot{q}_\epsilon(t)) + \tau \| \bar{x} + q_\epsilon(t) \|_l^2 = -\tau \int_0^t \langle g^\prime (\bar{x} + q_\epsilon) \dot{q}_\epsilon, q_\epsilon \rangle \, ds + r(t), \tag{5.17} \]

with a uniformly bounded function \( r(t) \): \( |r(t)| \leq C \). We claim that there exists \( C > 0 \) that is independent of \( \epsilon \) and \( t \) so that

\[ \left| \int_0^t \langle g^\prime (\bar{x} + q_\epsilon) \dot{q}_\epsilon, q_\epsilon \rangle \, ds \right| \leq C. \tag{5.18} \]

Indeed, let us fix some \( 1 \leq j \leq n \) and look at

\[ I = \int_0^t \langle g^\prime_j (\bar{x} + q_\epsilon(s)) q_{\epsilon,j}(s) \dot{q}_{\epsilon,j}(s) \rangle \, ds = \frac{1}{\epsilon} \sum_{k=1}^Q \int_{\bar{s}_k}^{\bar{s}_{k+1}} q_{\epsilon,j}(s) \dot{q}_{\epsilon,j}(s) \, ds \]

\[ = \frac{1}{2\epsilon} \sum_{k=1}^Q (\| q_{\epsilon,j}(s_k) \|^2 - \| q_{\epsilon,j}(s_{k+1}) \|^2). \tag{5.19} \]

Here, \( (s_k, s_{k+1}) \), \( k = 1, \ldots, Q \), are the time intervals that \( q_{\epsilon,j}(s) \) spends in the interval \( (-\bar{x}_j - \epsilon, -\bar{x}_j + \epsilon) \) and \( q_{\epsilon,j}(s_k) = \bar{x}_j \pm \epsilon \), depending on whether \( q_j \) enters this interval from above or below, and similarly for \( q_{\epsilon,j}(s_{k+1}) \). Since \( q_{\epsilon,j}(s) \) is continuous, we see that \( q_{\epsilon,j}(s_k) = q_j(s_{k+1}) \). Hence, (5.19) is a telescoping sum giving

\[ I = \frac{1}{2\epsilon} (\| q_{\epsilon,j}(s_Q) \|^2 - \| q_{\epsilon,j}(s_1) \|^2). \]

As both terms on the right-hand side above can take only the values \( -\bar{x}_j \pm \epsilon \), we conclude that \( |I| \leq C \), so that (5.18) holds. Now, (5.17) becomes

\[ (q_\epsilon(t), \dot{q}_\epsilon(t)) + \tau \| \bar{x} + q_\epsilon(t) \|_l^2 \leq C. \tag{5.20} \]

As \( \|x\| \leq \|x\|_l \), using the triangle inequality, we obtain the following inequality for \( m_\epsilon(t) = \| q_\epsilon(t) \| \):

\[ m_\epsilon(t) \dot{m}_\epsilon(t) + C_1 m_\epsilon(t) \leq C_2. \]

Now, the comparison principle implies that \( m_\epsilon(t) \leq C \) for all \( t > 0 \), and the proof of lemma 4.4 is complete. \( \square \)

5.5. Proof of lemma 4.5

Let us choose \( t_k \) and \( t_k' \) as in the proof of lemma 4.3. The estimate for \( \| A q_\epsilon(t_k) \| \) is exactly as in that lemma.

Next, dividing (5.15) by \( C_k k = t_k' - t_k \) we obtain, due to the boundedness of \( z_\epsilon(t) \) and \( q_\epsilon(t) \),

\[ \frac{1}{t_k' - t_k} \int_{t_k}^{t_k'} \| \dot{q}_\epsilon(s) \|^2 \, ds \leq \frac{C}{k} + \frac{C}{k} \int_{t_k}^{t_k'} \| A q_\epsilon(s) \|^2 \, ds \leq \frac{C}{k} + \frac{C}{k^2}. \tag{5.21} \]

It follows that there exists a time \( s_k \in (t_k, t_k') \) such that \( \| \dot{q}_\epsilon(s_k) \| \leq C/\sqrt{k}. \) \( \square \)
5.6. Proof of corollary 4.6

This follows immediately from lemma 4.5 and (5.1), as the latter implies that
\[
\|\tilde{y}_i(t)\|^2 + \|Aq_i(t)\|^2 \leq \|\tilde{y}_i(s_0)\|^2 + \|Aq_i(s_0)\|^2 \leq \frac{C}{n},
\]
for all \( t > s_0 \).

\( \square \)

6. The proof of theorem 2.2

We will use theorem 2.3 in order to prove theorem 2.2. The role of the vector \( z \) that satisfies the sub-differential condition can be seen from the following lemma.

Lemma 6.1. Suppose the sub-differential condition does not hold for a particular \( z \). Then for this \( z \), we have a strict inequality
\[
h(z) = \min_{\tilde{x}} F(x, z) < \tau \|\tilde{x}\|_l.
\]

Proof. Assume that \( z \) does not satisfy the sub-differential condition, that is, either

- \(|A^*z_i| > \tau\) for some \( i \), or
- \(|A^*z_i| \leq \tau\), but \(|A^*z_i| \neq \tau \text{sign}(\tilde{x}_i)\) for some \( i \) such that \( \tilde{x}_i \neq 0 \).

We will show that
\[
F(\tilde{x} + q, z) < F(\tilde{x}, z) = \tau \|\tilde{x}\|_l,
\]
for some (sufficiently small) \( q \), which implies (6.1). We will now construct \( q \) explicitly.

Assume first that (i) holds: \(|A^*z_i| > \tau\) for some \( i \). Then, set \( r = |A^*z_i| \), and choose \( q \) so that
\[
q_k = \begin{cases} 
\varepsilon \text{ sign}(r), & \text{if } k = i, \\
0, & \text{otherwise}.
\end{cases}
\]

We have
\[
F(\tilde{x} + q, z) = \tau \|\tilde{x} + q\|_l + \frac{1}{2} \|Aq\|^2 - (A^*z, q)
= \tau \|\tilde{x}\|_l + \tau (|\tilde{x}_i + \varepsilon \text{ sign}(r)| - |\tilde{x}_i|) + \frac{1}{2} \|Aq\|^2 \varepsilon |r|
\leq \tau \|\tilde{x}\|_l + \varepsilon \tau - \varepsilon |r| + \frac{1}{2} \|Aq\|^2 \leq \tau \|\tilde{x}\|_l + \varepsilon \tau - \varepsilon |r| + Ce^2 < \tau \|\tilde{x}\|_l,
\]
provided that we choose \( \varepsilon \) sufficiently small.

Similarly, if (ii) holds, pick some \( i \) such that \( \tilde{x}_i \neq 0 \) but \(|A^*z_i| = \tau \text{sign}(\tilde{x}_i)\). Assume first that \(|A^*z_i| = r \text{sign}(\tilde{x}_i)\) with \( 0 < |r| < \tau \). Pick \( \varepsilon \in (0, |\tilde{x}_i|) \) and choose \( q \) with the components
\[
q_k = \begin{cases} 
-\varepsilon \text{ sign}(\tilde{x}_i), & \text{if } k = i, \\
0, & \text{otherwise}.
\end{cases}
\]

The computation is similar:
\[
F(\tilde{x} + q, z) = \tau \|\tilde{x} + q\|_l + \frac{1}{2} \|Aq\|^2 - (A^*\tilde{x}, q) = \tau \|\tilde{x}\|_l + \tau (|\tilde{x}_i - \varepsilon \text{sign} (\tilde{x}_i)| - |\tilde{x}_i|)
+ \frac{1}{2} \|Aq\|^2 + \varepsilon r \leq \tau \|\tilde{x}\|_l - \varepsilon \tau + \varepsilon r
+ \frac{1}{2} \|Aq\|^2 \leq \tau \|\tilde{x}\|_l - \varepsilon \tau + \varepsilon r + Ce^2 < \tau \|\tilde{x}\|_l,
\]
provided that \( \varepsilon \) is sufficiently small. The last case to consider is when (ii) holds, but \(|A^*z_i| = -r \text{sign}(\tilde{x}_i)\). We still choose \( q \) as in (6.3), and the computation is identical to (6.4), with \( r = -\tau \). This completes the proof of lemma 6.1. \( \square \)
**Proof of theorem 2.2.** We trivially have
\[ h(z) = \min F(x, z) \leq F(\hat{x}, z) = \tau \|\bar{z}\|_{l_1}, \]
for all \( z \). Thus, the conclusion of theorem 2.2 would follow if we show that there exists \( \bar{z} \) such that \( h(\bar{z}) = \tau \|\bar{z}\|_{l_1} \). That is, we need to show that for any \( q \neq 0 \) and some \( \bar{z} \), we have
\[ F(\bar{x} + q, \bar{z}) = \tau \|\bar{x} + q\|_{l_1} + \frac{1}{2}\|Aq\|^2 - (A^*\bar{z}, q) > F(\bar{x}, \bar{z}) = \tau \|\bar{x}\|_{l_1}. \]  
(6.5)

We claim that (6.5) is true for any \( \bar{z} \) that satisfies the sub-differential condition (2.7)—recall that theorem 2.3 implies that such \( \bar{z} \) exists. Let \( \bar{z} \) satisfy the sub-differential condition (2.7):
\[ |A^*\bar{z}|_i = \tau \text{ sign } \bar{x}_i, \quad \text{if } i \in S_1, \]
\[ ||A^*\bar{z}||_l \leq \tau, \quad \text{if } i \in S_0. \]
(6.6)
(6.7)

We denoted here by \( S_1 \) the set of indices \( i \) such that \( \bar{x}_i \neq 0 \), and by \( S_0 \) the set of indices \( i \) such that \( \bar{x}_i = 0 \).

The function \( F(\bar{x} + q, \bar{z}) \) is convex in \( q \). Hence, it suffices to show that \( q = 0 \) is a strict local minimum, that is, show that (6.5) holds for \( q \) small enough. In particular, we may assume that
\[ \text{sign}(\bar{x}_i + q_i) = \text{sign}(\bar{x}_i), \quad \text{if } i \in S_1. \]
(6.8)

Now, we obtain from (6.7):
\[ \tau |q_i| - |A^*\bar{z}|q_i \geq 0, \quad i \in S_0, \]
(6.9)
while for \( i \in S_1 \), we use (6.8) and (6.6) to obtain
\[ \tau |\bar{x}_i + q_i| - |A^*\bar{z}|q_i = \tau (\text{sgn}\bar{x}_i)(\bar{x}_i + q_i) - \tau (\text{sgn}\bar{x}_i)\bar{x}_i = \tau (\text{sgn}\bar{x}_i)\bar{x}_i = \tau |\bar{x}_i|, \quad i \in S_1. \]
(6.10)

We deduce from (6.9) and (6.10) that
\[ F(\bar{x} + q, \bar{z}) = \tau \|\bar{x} + q\|_{l_1} + \frac{1}{2}\|Aq\|^2 - (A^*\bar{z}, q) \]
\[ = \sum_{i \in S_1} (\tau |\bar{x}_i + q_i| - |A^*\bar{z}|q_i) + \sum_{i \in S_0} (\tau |q_i| - |A^*\bar{z}|q_i) \]
\[ + \frac{1}{2}\|Aq\|^2 \geq \sum_{i \in S_1} \tau |\bar{x}_i| + \frac{1}{2}\|Aq\|^2 = \tau \|\bar{x}\|_{l_1} + \frac{1}{2}\|Aq\|^2. \]
(6.11)

Therefore, we have \( F(\bar{x} + q, \bar{z}) > \tau \|x\|_{l_1} \) unless \( Aq = 0 \). However, if \( Aq = 0 \), then
\[ F(\bar{x} + q, z) = \tau \|\bar{x} + q\|_{l_1} > \tau \|\bar{x}\|_{l_1}, \]
because \( \bar{x} \) is the unique minimizer of (1.1). Therefore, (6.5) holds for all \( q \).  
\[ \square \]

**7. Conclusions**

We have shown using ordinary differential equation methods that the relaxed \( l_1 \)-minimization algorithm for problems with underdetermined linear constraints converges independent of the regularization parameter. In the examples in array imaging, the observed convergence rates are faster than the theory implies, which means that more analysis is needed. The algorithm is robust to noise although we have not shown this theoretically. Finally, as the convergence rates are independent of dimension, generalization to the infinite-dimensional case is straightforward.
Acknowledgments

MM was supported by the Spanish Ministry of Education and Science grant FIS2010-18 473, AN was supported by NSF grant DMS-0908011, LR was supported by AFOSR NSSEFF fellowship, and NSF grant DMS-0908 507, and GP by AFOSR grant FA9550-11-1-0266. We thank Laurent Demanet for an interesting discussion, Alberto Bressan for bringing reference [19] to our attention, and Jalal Fadili for taking time to explain the literature on Arrow–Hurwicz–Uzawa algorithm, and pointing out the paper [17] to us.

References


[27] Esser E, Zhang X and Chan T F 2010 A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science SIAM J. Imaging Sci. 3 1015–46