



Fast communication

Imposing uniqueness to achieve sparsity

Keith Dillon*, Yu-Ping Wang

Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA



ARTICLE INFO

Article history:

Received 17 July 2015

Received in revised form

11 November 2015

Accepted 11 December 2015

Available online 29 December 2015

Keywords:

Sparsity

Regularization

Uniqueness

Non-negativity

Underdetermined linear systems

Convex optimization

ABSTRACT

In this paper we take a novel approach to the regularization of underdetermined linear systems. Typically, a prior distribution is imposed on the unknown to hopefully force a sparse solution, which often relies on uniqueness of the regularized solution (something which is typically beyond our control) to work as desired. But here we take a direct approach, by imposing the requirement that the system takes on a unique solution. Then we seek a minimal residual for which this uniqueness requirement holds. When applied to systems with non-negativity constraints or forms of regularization for which sufficient sparsity is a requirement for uniqueness, this approach necessarily gives a sparse result. The approach is based on defining a metric of distance to uniqueness for the system, and optimizing an adjustment that drives this distance to zero. We demonstrate the performance of the approach with numerical experiments.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Modern approaches to sparse solutions in linear regression or inverse problems are often viewed as MAP estimation [11] techniques. For example, Basis Pursuit [5] and LASSO [20], employing the ℓ_1 -norm, can be formulated with a Gaussian likelihood for additive noise and a Laplace prior [13]. However, the choice of prior distribution itself is only imposed because it often achieves a desirable result, namely a result that, in the noise-free case, can be shown to be equal to the minimum ℓ_0 -norm solution. Hence the approach generally amounts to a heuristic technique. Indeed a great deal of research in compressed sensing [24] has focused on theoretical guarantees for when the desired sparse result will be achieved for an underdetermined linear system. For example, the major conditions for uniqueness, such as the restricted isometry property [4], the nullspace property [8], or the k -neighborliness property [9], provide guarantees for when the minimum ℓ_1 -norm solution for the noise-free

underdetermined system equals the minimal ℓ_0 -norm solution. The key to this relationship is uniqueness of the minimizer (i.e., the situation where there is only solution which achieves the minimum).

This question of a unique minimizer is mathematically equivalent to the question of whether a related linear system has a unique non-negative solution. Based on this relationship, [10,1–3,22,23] utilize results regarding the ℓ_1 norm or derive similar results to develop uniqueness conditions for non-negative systems. Conceptually this situation is much easier to envision; an underdetermined system has m equations and n unknowns with $m < n$, and non-negativity provides n inequalities to further restrict solutions. For the solution to be unique, we need at least $n - m$ of the inequalities to be active and acting as equalities somehow due to the structure of the problem. This, in turn, means the corresponding elements of the unknown must be zero and hence the unknown must be sparse. Note that the non-negative system that is related to the set of minimum ℓ_1 -norm minimizers is a case with a very specific structure. For non-negative systems with more general structure, a new and easily verified row-space condition [3] must also be tested in order to determine if

* Corresponding author. Tel.: +1 504 988 1547.

E-mail address: kdillon1@tulane.edu (K. Dillon).

the system can have a unique non-negative solution. Overall, the implication of uniqueness is the same that we can use a more computationally tractable norm to calculate the ℓ_0 -norm result.

When non-negativity constraints are imposed as true prior knowledge in an inverse problem, such as to impose known physical properties, for example, the perspective based on establishing uniqueness guarantees fits quite well. But for applications such as variable or basis selection, the approach again amounts to a heuristic with a true goal of achieving a solution with desirable properties, i.e., one that is sparse. Slawski and co-authors have investigated the theory and applications of non-negative least squares (NNLS) as a competing technique versus the popular ℓ_1 -norm models for such applications [17–19]. Other researchers have extended non-negativity-based techniques to include additional means to enforce sparsity. In [12] non-negativity constraints are combined with ℓ_1 -norm regularization. In [15] non-negativity constraints are incorporated into an orthogonal matching pursuit algorithm. However in all these techniques, uniqueness of the solution is important to the quality of the results, yet is a property which depends on both the matrix and the solution, hence cannot be guaranteed. And so the authors investigate the incorporation of additional heuristics, based on conventional sparse regularization techniques, to increase the likelihood of a unique solution.

The technique presented here differs in that we propose an “additional ingredient” that is a requirement for uniqueness itself, hence we guarantee a unique solution. We will focus on the non-negative system as a general case, and start in the next section by reviewing the relationship to the ℓ_1 -regularized and non-negative least-squares techniques. Then we will derive uniqueness conditions for the solution set and provide an algorithm to enforce them on the system. Finally we demonstrate the performance of the algorithm with simulated examples where we will demonstrate the ability of the method to enforce unique solutions for a variety of models.

2. Theory

We will address the linear system $\mathbf{Ax} = \hat{\mathbf{b}} + \mathbf{n} = \mathbf{b}$, where \mathbf{A} is a known $m \times n$ matrix with $m < n$; \mathbf{b} is a known vector we wish to approximate with few columns of \mathbf{A} , and \mathbf{x} is an unknown vector we would like to estimate; \mathbf{n} is a “noise” vector about which we only have statistical information. The NNLS technique [14] solves $\min_{\mathbf{x} \geq 0} \|\mathbf{Ax} - \mathbf{b}\|_2^2$, or equivalently,

$$\mathbf{x}^* = \arg \min_{\substack{\mathbf{x} \geq 0, \Delta \mathbf{b} \\ \mathbf{Ax} = \mathbf{b} + \Delta \mathbf{b}}} \|\Delta \mathbf{b}\|_2^2 \quad (1)$$

From this perspective we can view NNLS as seeking a minimal system adjustment to get a feasible \mathbf{x}^* in the set

$$S_{NN} = \{\mathbf{x} | \mathbf{Ax} = \mathbf{b}', \mathbf{x} \geq \mathbf{0}\}, \quad (2)$$

where $\mathbf{b}' = \mathbf{b} + \Delta \mathbf{b}$. It can be shown that for all optimal solutions $(\Delta \mathbf{b}^*, \mathbf{x}^*)$ to Eq. (1), the component $\Delta \mathbf{b}^*$ will be unique. Hence we only need to consider the set S_{NN} given

this $\Delta \mathbf{b}^*$. In other words, we can use results from the well-known noise-free case. A necessary condition for uniqueness is the requirement that the row space of \mathbf{A} intersects the positive orthant [3]. Mathematically this means the system $\mathbf{A}^T \mathbf{y} = \boldsymbol{\beta}$ has some solution \mathbf{y} for which $\boldsymbol{\beta}$ has all positive elements. Geometrically it means that the polytope [25] formed by S_{NN} must be finite in size [7]. Note that if a general system was converted into an equivalent non-negative one by replacing the general signal with the difference of two non-negative signals representing positive and negative channels, the resulting system matrix would violate the positive orthant condition. We will presume throughout this paper that the row space of \mathbf{A} intersects the positive orthant for all matrices.

Of course, the NNLS technique need not result in a sparse solution at all. Hence some techniques also include ℓ_1 -regularization or other ingredients in addition to non-negativity. On the other hand, many popular techniques, such as LASSO, impose ℓ_1 -regularization alone. LASSO (in the form of Basis Pursuit denoising) can be viewed as a heuristic technique where λ is chosen to trade off sparsity of the solution with minimal adjustment to the model, which can be posed in a form similar to Eq. (1), as follows:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}, \Delta \mathbf{b}} \{ \|\Delta \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1 \} \\ \mathbf{Ax} = \mathbf{b} + \Delta \mathbf{b}. \quad (3)$$

When \mathbf{A} is underdetermined (i.e., $n > m$), the situation we are interested in here, it is known that the LASSO solution may not be unique [16]. However, as with NNLS, the residual $\Delta \mathbf{b}$ is always unique [21]. So, again, we can address the question of uniqueness by focusing on a noise-free case, here the Basis Pursuit problem, $\alpha = \min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\mathbf{Ax} = \mathbf{b}'$. In this case the question of uniqueness applies to the solutions in the set, $F = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{Ax} = \mathbf{b}', \|\mathbf{x}\|_1 \leq \alpha\}$, which can be posed as a set of the form S_{NN} . However, for intuition consider F as depicted in Fig. 1, the intersection of ℓ_1 -norm ball of radius α , and the affine set of solutions to the linear system, $\{\mathbf{x} | \mathbf{Ax} = \mathbf{b}'\}$. For the example in Fig. 1, the solution is non-unique, as F contains an interval of points on the nearest face of the ball. This set F is the set of all possible minimizers which achieve an equal minimum ℓ_1 norm. It is not clear what is the best way to handle this situation. From the MAP estimation perspective, any point on the intersection is equally likely; the distribution for $P(\mathbf{x} | \mathbf{b}')$ will be uniform over F . A typical algorithm may yield a very-undesirable dense solution in the interior of F .

In all the above cases, the most fundamental question we would like to be able to answer is whether an underdetermined system has a unique non-negative solution. That will be the first goal of this paper. If desired, we can then easily apply our results to other types of problems such as NNLS, Basis Pursuit, and LASSO via a mathematically equivalent system as noted above.

2.1. Uniqueness conditions

To start, we will presume that we have a compatible non-negatively constrained system determined by \mathbf{A} and \mathbf{b} , arising either directly from our application, or, for

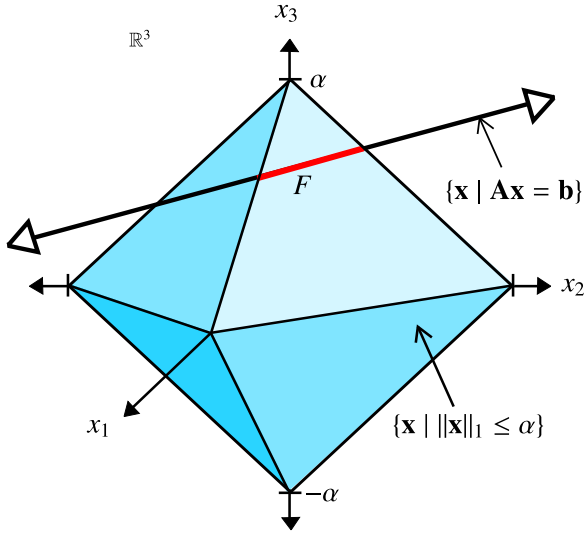


Fig. 1. Nonunique intersection on ℓ_1 -norm ball and line containing solutions, demonstrating potential failure of ℓ_1 -norm based methods. The intersection is the set F , depicted by the red segment, which is contained in the nearest face of the ball. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

example, from the NNLS or LASSO technique (from which we will henceforth replace \mathbf{b}' with \mathbf{b}), and wish to investigate the uniqueness of the solutions in the set S_{NN} . By compatible non-negatively constrained system here we mean at least one non-negative solution must exist. In [7] we investigated bounds for the solution set for non-negative systems by using linear programming to find the maximum and minimum of each element of the unknown. A similar technique was employed in [21] to test for uniqueness in LASSO. We can extend this idea to form conditions for uniqueness of x_k , the k th element of \mathbf{x} , over S_{NN} . First we test for uniqueness as follows:

$$\delta_k = \max_{\mathbf{x} \geq \mathbf{0}} x_k - \min_{\mathbf{x} \geq \mathbf{0}} x_k = \max_{\mathbf{x} \geq \mathbf{0}, \mathbf{x}' \geq \mathbf{0}} (x_k - x'_k)$$

$$\begin{matrix} \mathbf{Ax} = \mathbf{b} & \mathbf{Ax} = \mathbf{b} & \mathbf{Ax} = \mathbf{b} \\ & & \mathbf{Ax}' = \mathbf{b}. \end{matrix} \quad (4)$$

To find an upper bound on δ_k , we use duality theory for linear programs [6]. The dual of Eq. (4) is the linear program,

$$\tilde{\delta}_k = \min_{\mathbf{y}, \mathbf{y}'} \mathbf{b}^T (\mathbf{y} - \mathbf{y}')$$

$$\begin{matrix} \mathbf{A}^T \mathbf{y} \geq \mathbf{e}_k \\ \mathbf{A}^T \mathbf{y}' \leq \mathbf{e}_k, \end{matrix} \quad (5)$$

where \mathbf{e}_k is a vector of zeros except a one in the k th element. For intuition, consider what we get without non-negativity constraints, where the possible solutions are the affine set $\{\mathbf{x} | \mathbf{Ax} = \mathbf{b}\}$. It is straightforward to show that x_k takes on a unique value if and only if there is a solution $\hat{\mathbf{y}}$ to $\mathbf{A}^T \hat{\mathbf{y}} = \mathbf{e}_k$. If we write these conditions as $\hat{\mathbf{y}}_{(k)}^T \mathbf{A} = \mathbf{e}_k^T$ for $k = 1, \dots, n$, and combine them into a matrix system, we find that these versions of the dual variables $\hat{\mathbf{y}}$ are simply columns of the left inverse $\hat{\mathbf{Y}}^T \mathbf{A} = \mathbf{I}$, where \mathbf{I} is the identity

matrix. If we similarly combine the n dual optimization problems of Eq. (5), we get the following combined problem for all elements (note that inequalities operate element-wise for matrices throughout this paper):

$$\alpha = \min_{\mathbf{Y}, \mathbf{d}} \|\mathbf{d}\|$$

$$\begin{matrix} (\mathbf{Y} - \mathbf{Y}')^T \mathbf{b} = \mathbf{d} \\ \mathbf{Y}^T \mathbf{A} \geq \mathbf{I} \\ \mathbf{Y}'^T \mathbf{A} \leq \mathbf{I}, \end{matrix} \quad (6)$$

where $\|\cdot\|$ can be any norm. This vector \mathbf{d} has elements which are the lengths of sides of the smallest box containing S_{NN} , the set of non-negative solutions. Its magnitude ($\|\mathbf{d}\|$, the “distance from uniqueness”) gives us a metric for how tightly the system restricts the possible values of the unknown. Its length is the worst-case distance between any two members of the solution set.

Theorem 1. *The compatible system $\mathbf{Ax} = \mathbf{b}$ has a unique non-negative solution if and only if $\mathbf{d} = \mathbf{0}$.*

Proof. If $\mathbf{d} = \mathbf{0}$, then all $\tilde{\delta}_k$ are zero. As $\tilde{\delta}_k$ are upper bounds for δ_k , each x_k can only take on a unique scalar value and hence \mathbf{x} is unique. For the reverse direction, since we assume that the system is compatible, Eq. (4) has an optimal solution. Further note that the optimal \mathbf{d} will be the same regardless of norm, hence we can presume that an ℓ_1 -norm is used and use the strong duality conditions for linear programs to require that $\tilde{\delta}_k = \delta_k$. If \mathbf{x} is unique then we must have $\tilde{\delta}_k = 0$ for all k , and by strong duality this means $\tilde{\delta}_k = 0$ for all k , i.e., $\mathbf{d} = \mathbf{0}$. \square

2.2. Algorithm

Next we assume that we are given a system which does not have a unique solution, and our goal is to find a “nearby” system for which $\mathbf{d} = \mathbf{0}$, by adjusting \mathbf{b} . Specifically we desire a residual $\Delta \mathbf{b}$ such that $(\mathbf{Y} - \mathbf{Y}')^T (\mathbf{b} + \Delta \mathbf{b}) = \mathbf{0}$ in Eq. (6). However, a variable $\Delta \mathbf{b}$ here produces a non-convex quadratic constraint, as we will have products between $\Delta \mathbf{b}$ and both \mathbf{Y} and \mathbf{Y}' . We will address this with an iterative algorithm based on successive linear approximations of this constraint. Given a tolerance μ , we solve the following algorithm for successive updates to \mathbf{b} :

- Step 1: Solve Eq. (6) for \mathbf{d}^* , \mathbf{Y}^* and \mathbf{Y}'^* .
- Step 2: If $\|\mathbf{d}\| \leq \mu$, stop, otherwise choose ϵ where $0 \leq \epsilon < \|\mathbf{d}\|$, and solve Eq. (7) for $\Delta \mathbf{b}$ and \mathbf{x} using the \mathbf{d}^* , \mathbf{Y}^* and \mathbf{Y}'^* computed in Step 1:

$$\beta = \min_{\mathbf{x} \geq \mathbf{0}, \Delta \mathbf{b}} \|\Delta \mathbf{b}\|$$

$$\begin{matrix} \|(\mathbf{Y}^* - \mathbf{Y}'^*)^T \Delta \mathbf{b} + \mathbf{d}^*\| \leq \epsilon \\ \mathbf{Ax} = \mathbf{b} + \Delta \mathbf{b} \end{matrix} \quad (7)$$

The constraints here are linear, so the problem is convex. Note that Eq. (7) can be viewed as a NNLS problem (compare to Eq. (1)) with an added constraint to impose uniqueness.

- Step 3: Update \mathbf{b} by accumulating new residual adjustment from Step 2 (i.e., $\mathbf{b} = \mathbf{b} + \Delta \mathbf{b}$) and goto Step 1.

The following theorem proves that a solution found via the above algorithm produces a system with a unique solution following an iteration where we picked $\epsilon = 0$, which we may choose to do in a single iteration or after multiple iterations (the only difference will be how globally optimal the resulting $\Delta \mathbf{b}$ might be). In practice we should use a small positive value instead of zero for ϵ , due to numerical precision issues.

Lemma 1. *The optimal \mathbf{x}^* found by Eq. (7) with $\epsilon = 0$, using the \mathbf{d}^* , \mathbf{Y}^* and \mathbf{Y}^{*T} computed in Eq. (6), will be the unique non-negative solution in the new system modified by the optimal $\Delta \mathbf{b}^*$, i.e., $\{\mathbf{x}^*\} = \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b} + \Delta \mathbf{b}^*, \mathbf{x} \geq \mathbf{0}\}$.*

Proof. This holds for any feasible point for the problem of Eq. (7), the optimal is simply a choice with minimal residual. Given a feasible point $(\hat{\mathbf{x}}, \Delta \hat{\mathbf{b}})$, $\hat{\mathbf{x}}$ is a solution in the set $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b} + \Delta \hat{\mathbf{b}}, \mathbf{x} \geq \mathbf{0}\}$, which means the system is compatible. Further, $\Delta \hat{\mathbf{b}}$ solves the system $(\mathbf{Y}^* - \mathbf{Y}^{*T})^T \Delta \hat{\mathbf{b}} = -\mathbf{d}^*$. By combining the constraints from Eq. (6) with our equation for $\Delta \hat{\mathbf{b}}$, we have

$$\begin{aligned} (\mathbf{Y}^* - \mathbf{Y}^{*T})^T (\mathbf{b} + \Delta \hat{\mathbf{b}}) &= \mathbf{0} \\ \mathbf{Y}^{*T} \mathbf{A} &\geq \mathbf{I} \\ \mathbf{Y}^{*T} \mathbf{A} &\leq \mathbf{I}. \end{aligned} \quad (8)$$

These equations are the constraints of Eq. (6) with \mathbf{b} replaced by $\mathbf{b} + \Delta \hat{\mathbf{b}}$. This means we can perform the optimization problem of Eq. (6) with \mathbf{b} also replaced by $\mathbf{b} + \Delta \hat{\mathbf{b}}$ and get the minimizer $\mathbf{d} = \mathbf{0}$. By Theorem 1, this means the non-negative solution to $\mathbf{A}\mathbf{x} = \mathbf{b} + \Delta \hat{\mathbf{b}}$ must be unique. \square

It can further be shown that it is possible to achieve any desired tolerance μ on the size of the solution set by choosing $\epsilon = \mu$, using the fact that $\mu = \|\mathbf{d}_{\text{final}}\| = \|(\mathbf{Y}^* - \mathbf{Y}^{*T})^T (\mathbf{b} + \Delta \mathbf{b})\| = \|(\mathbf{Y}^* - \mathbf{Y}^{*T})^T \Delta \mathbf{b} + \mathbf{d}^*\| \leq \epsilon$. Finally we argue that the closer we make ϵ to $\|\mathbf{d}^*\|$ in each iteration, the more accurately our linearized constraint fits the true constraint at each iteration, as it is quadratic. This is not necessary for Theorem 1, which only deals with feasible points, but in practice it suggests that we can achieve a smaller net $\|\Delta \mathbf{b}\|$ by performing more iterations. We will demonstrate the performance of single versus multiple iteration settings in the next section.

3. Simulations

We performed a variety of numerical simulations to demonstrate the technique. For each simulation, given a particular n and m , we formed a matrix using a choice of model to generate elements. For most cases we generated a true \mathbf{x} with K nonzero values at uniformly random locations, and generated each of those K values using a uniform distribution over the interval $[0, 1]$ (unless stated otherwise). We added Gaussian noise to the resulting measurement vector \mathbf{b} to achieve an SNR of 10:1. We then performed a denoising step using NNLS to make a compatible system as a starting point. Finally we estimated \mathbf{d} for the system, if it was unbounded (meaning our system did not fulfill the positive orthant condition), or if it was very small (meaning our system essentially had a unique

non-negative solution already), the system was discarded and a new one was randomly generated with the same parameters. For the algorithm we used a tolerance of $\mu = 0.01$.

For example, Fig. 2 gives the true solution and the unique solution estimated via our technique for a matrix generated using uniformly distributed random elements with values in $[0,1]$, with $n=100$, $m=30$, and $K=15$. For this example we used $\epsilon=0.01$, requiring only a single iteration to achieve the final tolerance. Fig. 2 also gives the non-negative ℓ_1 (min ℓ_1 -norm over non-negative \mathbf{x}). In Fig. 3 we give \mathbf{d} for the initial system as (denoted as \mathbf{d}_0), which gives the range each element of the solution could initially take. After running our algorithm we computed \mathbf{d} again ($\mathbf{d}_{\text{unique}}$), measuring the range taken by each element of $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b} + \Delta \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$. As we can see from Fig. 3, $\mathbf{d}_{\text{unique}}$ is essentially zero meaning our system now admits a unique solution. Fig. 3 also gives the corresponding initial (\mathbf{b}_0) and adjusted ($\mathbf{b}_0 + \Delta \mathbf{b} = \mathbf{b}_{\text{unique}}$) versions of \mathbf{b} . We can see that \mathbf{b}_0 required only a very-small adjustment to form $\mathbf{b}_{\text{unique}}$ in this case. Figs. 4 and 5 provide similar results for

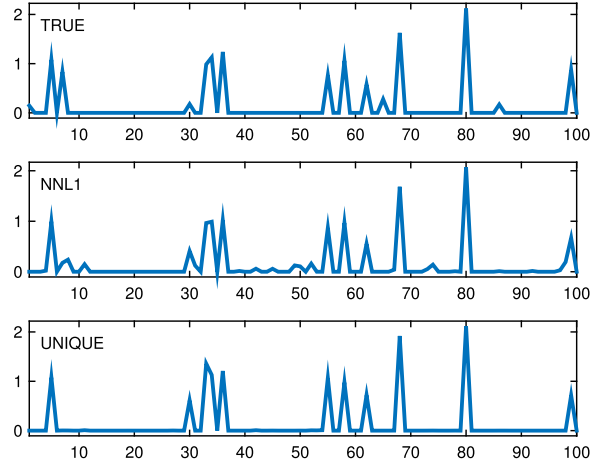


Fig. 2. True, Non-negative- ℓ_1 (NNL1), and unique solutions for uniform random matrix example, $n = 100$, $m = 30$, $K = 15 = \|\mathbf{x}_{\text{true}}\|_0$.

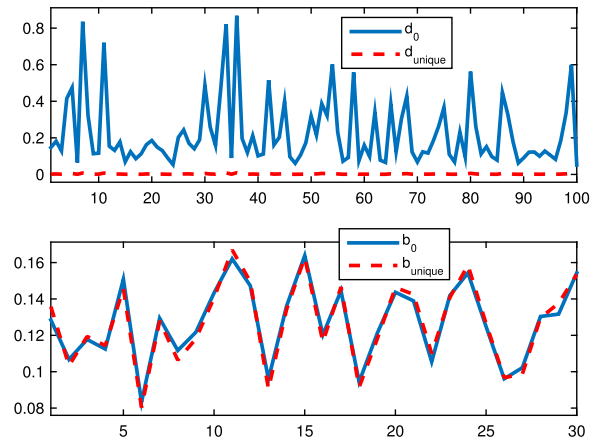


Fig. 3. Displacement (\mathbf{d}) from uniqueness for original and unique system, and output vector (\mathbf{b}) for original and unique system.

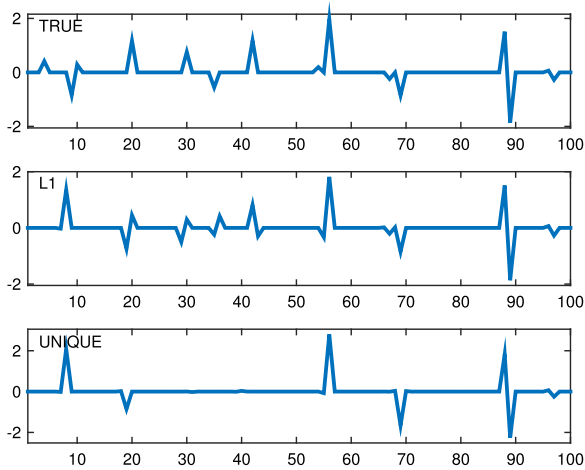


Fig. 4. True, minimum ℓ_1 (L1), and unique solutions for non-bounded example, $n=100$, $m=30$, $K=15$.

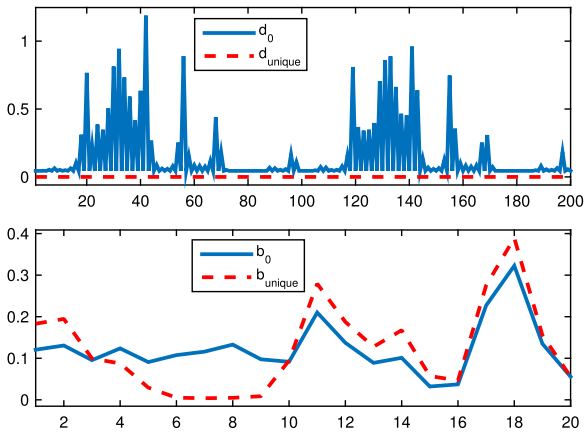


Fig. 5. Displacement (\mathbf{d}) from uniqueness for original and unique system, and output vector (\mathbf{b}) for original and unique system. Non-bounded example.

a non-bounded example for which the minimum ℓ_1 solution was not unique, using a model that performs a low-pass filtering operation. The system was converted to a non-negative system in $2n$ dimensions, then the algorithm was applied to enforce a unique solution, and finally the resulting system was converted back to n dimensions, yielding a system with a unique minimum ℓ_1 norm.

We investigated the algorithm's performance by generating 100 realizations using models with uniform random elements (Fig. 6) and binary random elements (Fig. 7), where for each realization we used a new random \mathbf{A} , new true \mathbf{x} (changing both the values and locations), and new noise. We compared the cardinality of the result computed with our algorithm both with single-iteration settings, and with a multiple-iteration approach where ϵ was chosen to be 90 percent of the maximum element of \mathbf{d} computed in each previous iteration. For the single-iteration unique solutions in Fig. 6, we find a clear trend towards sparser solutions as $\|\mathbf{d}_0\|$ (the initial distance from uniqueness) increases, while in Fig. 7 we see that

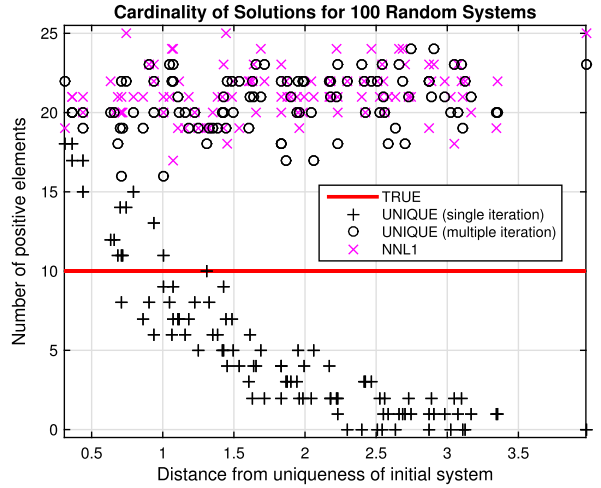


Fig. 6. Number of positive elements ($x_i \geq 0.02$) for true, Non-negative ℓ_1 denoised (NNL1), and unique solutions for 100 different noisy realizations of uniform model, SNR=10, $n=100$, $m=25$, $K=10 = \|\mathbf{x}_{true}\|_0$.

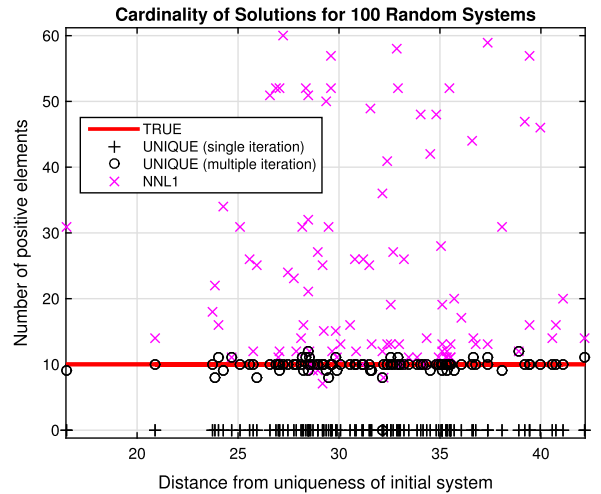


Fig. 7. Number of positive elements ($x_i \geq 0.02$) for true, Non-negative ℓ_1 denoised (NNL1), and unique solutions for 100 different noisy realizations of convolution model, SNR=10, $n=100$, $m=11$, $K=10 = \|\mathbf{x}_{true}\|_0$. In this case, the NNL1 technique often failed to yield a sparse solution while the proposed technique consistently succeeded using the multiple-iteration algorithm.

only the multiple-iteration approach produced non-trivial results. We also computed the minimum non-negative ℓ_1 solutions using the same denoised starting point as the proposed algorithm, and plotted the cardinalities of all of these solutions versus $\|\mathbf{d}_0\|$. We see that for the uniform system, we found comparable performance with both the proposed multiple-iteration algorithm and the NNL1 solution, while for the binary system, the proposed algorithm demonstrated superior performance, consistently returning a sparse solution.

Next we repeated similar 100-realization simulations for a range of different model sizes using matrices formed by three different random distributions. The results are summarized in Figs. 8–10, where the range of resulting cardinalities achieved at each m is shown by the solid

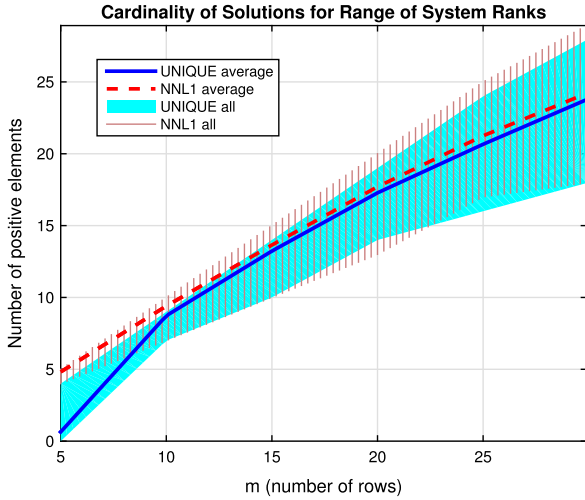


Fig. 8. Range of cardinality of solutions for Non-negative ℓ_1 denoised (NNL1), and unique solutions using $\epsilon = 0.9 \max(\mathbf{d})$, for 100 different noisy realizations of uniform random system, for each value of m ranging in steps of 5 between 5 and 30 (600 total realizations), $SNR=10$, $n=100$, $K=10 = \|\mathbf{x}_{true}\|_0$.

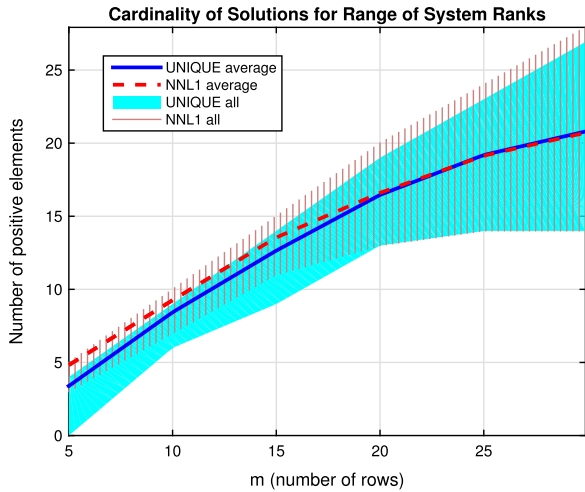


Fig. 9. Range of cardinality of solutions for Non-negative ℓ_1 denoised (NNL1), and unique solutions using $\epsilon = 0.9 \max(\mathbf{d})$, for 100 different noisy realizations of Gaussian random system, again for each m between 5 and 30, $SNR=10$, $n=100$, $K=10 = \|\mathbf{x}_{true}\|_0$.

regions. For each model we varied m between a minimum of 5 and a maximum of the highest that was reasonably possible to still find non-unique systems (systems for which the solution was not already unique without adjustment) with that approach, generally around 25–30. At around $m=5$ we see that the proposed algorithm began to return all-zero results as the single-iteration case did, suggesting that a more sophisticated linearization is needed (e.g., more iterations). At higher values of m , we see comparable results for the proposed and NNL1 algorithms for the uniform and Gaussian models, generally achieving a cardinality comparable to m . For the binary model, we see that the NNL1 algorithm consistently returns many non-sparse results as seen in Fig. 7. In Fig. 11 we give the average (Pearson) correlation for each m for the methods

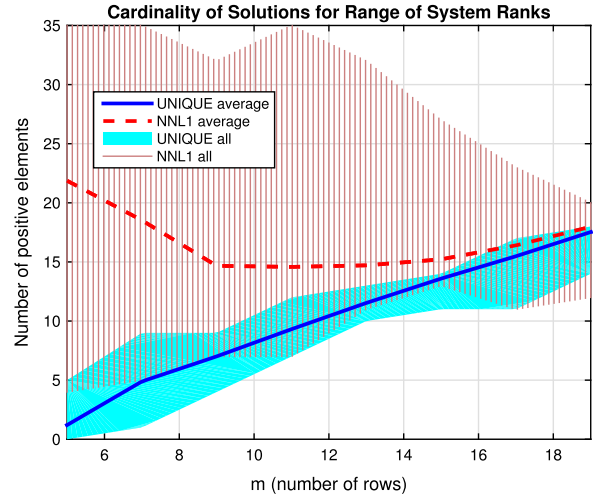


Fig. 10. Range of cardinality of solutions for Non-negative ℓ_1 denoised (NNL1), and unique solutions using $\epsilon = 0.9 \max(\mathbf{d})$, for 100 different noisy realizations of binary random system for each m between 5 and 20, $SNR=10$, $n=50$, $K=10 = \|\mathbf{x}_{true}\|_0$. Again we see the failure to achieve a sparse solution for many cases with the NNL1 technique.

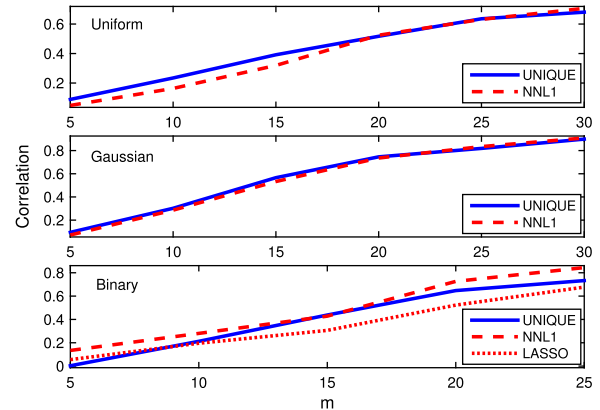


Fig. 11. Average Pearson correlation between estimates and true solutions for each m for the simulations of Figs. 8–10. Since the NNL1 solutions often failed to be sparse for the Binary model, and hence the correlation may demonstrate over-fitting, a non-negative LASSO estimate was calculated using increased regularization to achieve comparable sparsity to the UNIQUE solution.

between the estimate and the true solution. For the uniform and Gaussian models we see that the proposed algorithm generally achieved a slightly higher correlation with the true solution for smaller values of m , presumably due to the bias of the ℓ_1 estimate. At higher m the NNL1 correlation was slightly higher, probably due to the 0.01 gap the proposed algorithm allows in uniqueness for numerical reasons. For the binary model, the NNL1 achieved a higher correlation, but recall that this was the case where NNL1 often returned dense solutions. To make a fairer comparison, we imposed ℓ_1 regularization via LASSO to achieve a comparable sparsity for the failure cases (the regularization parameter was increased until a solution with cardinality of at most m was achieved), which led to a lower correlation overall.

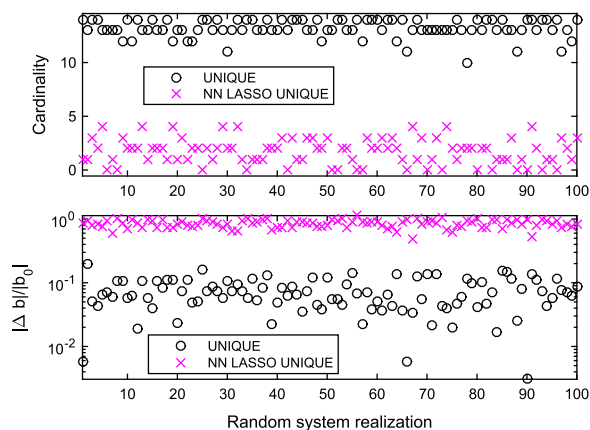


Fig. 12. Cardinality and fractional change in $\Delta \mathbf{b}$ for unique solutions using proposed algorithm versus LASSO with $\epsilon = 0.9 \max(\mathbf{d})$, for 100 different noisy realizations of uniform random system, for $\text{SNR} = 10$, $m = 15$, $n = 100$, $K = 10 = \|\mathbf{x}_{\text{true}}\|_0$. We see that the proposed algorithm generally achieved a sparse solution with cardinality comparable to m , and with a change in b comparable to the SNR. Attempting to manufacture a system that has a unique solution using LASSO was far less successful, yielding results that were overly sparse with an excessively large fractional change in b .

Finally we compared the proposed approach to a LASSO-based technique in achieving a system with a unique solution. We compared the proposed algorithm with a non-negative LASSO algorithm, wherein we increased the regularization parameter λ until the adjusted system $\{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{b} + \Delta \mathbf{b}_{\text{LASSO}}, \mathbf{x} \geq \mathbf{0}\}$ has a unique solution (the smallest λ where \mathbf{d} is below our threshold of 0.01). Note that this is stricter than a requirement that only the set of LASSO minimizers be unique. In Fig. 12 we give the cardinality and fractional residual (computed as $\|\Delta \mathbf{b}\| / \|\mathbf{b}\|$) for each of the 100 realizations. We see that generally the LASSO approach used a much larger adjustment to the system to achieve a unique system solution, and for many cases was unable to produce a non-trivial unique solution.

4. Discussion

In this paper we derived uniqueness conditions for the non-negative solution to an underdetermined linear system using optimization theory. The conditions provide a kind of distance from uniqueness for when the system is non-unique, which we used in an algorithm to adjust the system and make a similar system which has a unique solution. We provided simulated examples to demonstrate the qualitative behavior of the technique where we demonstrated a broad ability to find non-trivial unique solutions over a range of systems. For systems with elements drawn from a continuous distribution, performance appears comparable to ℓ_1 -based methods, while for more structured systems such as the binary model, the performance of the proposed algorithm appears to be superior, presumably due to difficulties in achieving a unique optimal with the ℓ_1 -norm in such cases. The question of reconstruction error is beyond the scope of this paper as we do not make assumptions about matrix properties

beyond the positive orthant condition, but the correlations in the simulations proved comparable to and often better than achieved with ℓ_1 -based techniques, suggesting that this would be a fruitful area for further research.

Multiple improvements or variations are possible with this algorithm. A more intelligent choice of iteration settings would likely address the failure cases, such as with small m . More generally, note that Lemma 1 applies to any feasible point, meaning we can choose an optimal however we wish. We chose the minimum residual on $\Delta \mathbf{b}$ defined using the ℓ_2 -norm, but we might employ a different norm such as the ℓ_1 -norm, for example, to get a sparse residual. We might also impose other desired properties such as additional solution sparsity via a ℓ_1 -norm regularization term on the unique solution itself, yielding a hybrid approach. We might also utilize other means to convert a general linear system into a non-negative one which satisfies the positive orthant condition, for example by augmenting the matrix with an additional row that is optimized somehow for the particular scenario. Finally, it would be straightforward to replace the non-negativity constraint with other constraints, such as a conic constraint or other convex inequalities.

The approach of finding the nearest unique solution is interesting from multiple perspectives. First is simply its ability to indirectly provide a sparse solution, as sparsity is the basis for uniqueness in a non-negative system and the related ℓ_1 -norm based techniques. But another interesting perspective is the idea of uniqueness itself as a form of prior knowledge or model design requirement. If our goal is to choose $\Delta \mathbf{b}$ to minimize the uncertainty in \mathbf{x} , then a choice for which \mathbf{x} must be unique is the best possible in a sense. The unique solution might also be utilized as a component of the solution set, as in dimensionality reduction, rather than a model in itself.

Acknowledgments

The authors wish to thank the NIH (R01 GM109068, R01 MH104680) and NSF (1539067) for support.

References

- [1] A.M. Bruckstein, M. Elad, M. Zibulevsky, On the uniqueness of non-negative sparse & redundant representations, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, ICASSP 2008, IEEE, Las Vegas, NV, USA, March 2008, pp. 5145–5148.
- [2] A.M. Bruckstein, M. Elad, M. Zibulevsky, On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations, IEEE Trans. Inf. Theory 54 (November (11)) (2008) 4813–4820.
- [3] A.M. Bruckstein, M. Elad, M. Zibulevsky, Sparse non-negative solution of a linear system of equations is unique, in: The Third International Symposium on Communications, Control and Signal Processing, 2008, ISCCSP 2008, IEEE, St. Julians, Malta, March 2008, pp. 762–767.
- [4] Emmanuel J. Candes, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (May (9–10)) (2008) 589–592.
- [5] Scott Shaobing Chen, David L. Donoho, Michael A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (January (1)) (2001) 129–159.
- [6] George Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, NJ, USA, 1998 August.

- [7] Keith Dillon, Yeshaiahu Fainman, Bounding pixels in computational imaging, *Appl. Opt.* 52 (April (10)) (2013) D55–D63.
- [8] David L. Donoho, Michael Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization, *Proc. Natl. Acad. Sci. USA* 100 (March (5)) (2003) 2197–2202.
- [9] David L. Donoho, Jared Tanner, Neighborliness of randomly projected simplices in high dimensions, *Proc. Natl. Acad. Sci. USA* 102 (July (27)) (2005) 9452–9457.
- [10] David L. Donoho, Jared Tanner, Sparse nonnegative solution of underdetermined linear equations by linear programming, *Proc. Natl. Acad. Sci. USA* 102 (July (27)) (2005) 9446–9451.
- [11] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall PTR, Upper Saddle River, NJ, USA, 1998.
- [12] M.A. Khajehnejad, A.G. Dimakis, Xu Weiyu, B. Hassibi, Sparse recovery of nonnegative signals with minimal expansion, *IEEE Trans. Signal Process.* 59 (January (1)) (2011) 196–208.
- [13] Samuel Kotz, Tomasz Kozubowski, Krzysztof Podgorski, *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Economics, Engineering, and Finance*, Springer, New York, NY, USA, 2001 June.
- [14] C. Lawson, R. Hanson, *Solving least squares problems. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics*, Philadelphia, PA, USA, January 1995.
- [15] Tsung-han Lin, H.T. Kung, *Stable and Efficient Representation Learning with Nonnegativity Constraints*, 2014, pp. 1323–1331.
- [16] Michael R. Osborne, Brett Presnell, Berwin A. Turlach, On the LASSO and its dual, *J. Comput. Graph. Stat.* 9 (June (2)) (2000) 319–337.
- [17] Martin Slawski, Matthias Hein, Sparse recovery by thresholded non-negative least squares, in: J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, vol. 24, Curran Associates, Inc., Granada, Spain, 2011, pp. 1926–1934.
- [18] Martin Slawski, Matthias Hein, Non-negative least squares for high-dimensional linear models: consistency and sparse recovery without regularization, *Electron. J. Stat.* 7 (2013) 3004–3056.
- [19] Martin Slawski, Matthias Hein, E. Campus, Sparse recovery for protein mass spectrometry data, in: *Practical Applications of Sparse Modeling*, 2014, p. 79.
- [20] Robert Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. Ser. B* 58 (1994) 267–288.
- [21] Ryan J. Tibshirani, The Lasso problem and uniqueness, *Electron. J. Stat.* 7 (2013) 1456–1490.
- [22] Meng Wang, Ao Tang, Conditions for a unique non-negative solution to an underdetermined system, in: *The 47th Annual Allerton Conference on Communication, Control, and Computing*, 2009, Allerton 2009, IEEE, Monticello, IL, USA, September 2009, pp. 301–307.
- [23] Meng Wang, Weiyu Xu, Ao Tang, Unique “nonnegative” solution to an underdetermined system: from vectors to matrices, *IEEE Trans. Signal Process.* 59 (March (3)) (2011) 1007–1016.
- [24] Rebecca M. Willett, Roummel F. Marcia, Jonathan M. Nichols, Compressed sensing for practical optical imaging systems: a tutorial, *Opt. Eng.* 50 (7) (2011) 072601–072613.
- [25] Günter M. Ziegler, *Lectures on Polytopes*, Springer, New York, NY, USA, 1995.