CrossMark

ORIGINAL PAPER

Element-wise uniqueness, prior knowledge, and data-dependent resolution

Keith Dillon[1] · Yeshaiahu Fainman[2]

Abstract Techniques for finding regularized solutions to underdetermined linear systems can be viewed as imposing prior knowledge on the unknown vector. The success of modern techniques, which can impose priors such as sparsity and non-negativity, is the result of advances in optimization algorithms to solve problems which lack closed-form solutions. Techniques for characterization and analysis of the system to determine when information is recoverable, however, still typically rely on closed-form solution techniques such as singular value decomposition or a filter cutoff estimate. In this letter we propose optimization approaches to broaden the approach to system characterization. We start by deriving conditions for when each unknown element of a system admits a unique solution, subject to a broad class of types of prior knowledge. With this approach we can pose a convex optimization problem to find "how unique" each element of the solution is, which may be viewed as a generalization of resolution to incorporate prior knowledge. We find that the result varies with the unknown vector itself, i.e., it is data-dependent, such as when the sparsity of the solution improves the chance it can be uniquely reconstructed. The approach can be used to analyze systems on a case-by-case basis, estimate the amount of important information present in the data, and quantitatively understand the degree to which the regularized solution may be trusted.

✉ Keith Dillon
  kdillon1@tulane.edu

[1] Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA

[2] Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA

1 Introduction

We focus on techniques that use norms such as the ℓ_1-norm (sum of absolute elements) or the ℓ_∞-norm (maximum absolute element) for regularization and/or denoising of an underdetermined linear system, $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a known $m \times n$ matrix with $m < n$, $\mathbf{b}$ is a known measurement vector, and $\mathbf{x}$ is the unknown vector we seek. These techniques are generally not solvable in closed form (unlike, e.g., regularization with the ℓ_2-norm). However modern optimization methods can incorporate such information without difficulty using linear inequalities or convex conic constraints [2]. In this paper we will develop a framework for analyzing the results from such approaches, with specific focus on those which may be specifically formulated as linear inequality constraints (see Appendix for some examples). In broad terms, instead of considering the restrictions on $\mathbf{x}$ in the set $\{\mathbf{x} \mid \mathbf{Ax} = \mathbf{b}\}$, we focus on the (hopefully smaller and hence more informative) set $\{\mathbf{x} \mid \mathbf{Ax} = \mathbf{b}, \ \mathbf{Dx} \geq \mathbf{d}\}$. We first ask the question: Subject to this new constraint, has $\mathbf{x}$ become unique? We then extend this to the question: How well does this new information improve our ability to resolve $\mathbf{x}$?

Uniqueness has been extensively studied for the case of ℓ_1-regularization, where we are concerned, for example, with whether the solution found via Basis Pursuit [8] is unique. This is especially interesting because under the right conditions this solution is the optimally sparse solution (e.g., [12]). Published conditions for uniqueness come in several forms, such as the restricted isometry property [6], the null-space property [12], and neighborliness properties [11,13]. A significant limitation of these methods is their computational intractability for realistic system sizes [22]. This renders them unusable for analysis of systems, except for those with special structure that may be addressed theoretically (e.g., the random matrix designs used in compressed

sensing [23]). Additionally, non-negativity constraints have received increased interest recently due to their relationship to the ℓ_1-regularized case [5,14]. In this application, if the true solution is sparse enough (and a necessary condition for the matrix holds), then the system has a unique non-negative solution. There is no regularization in this case, and the non-negativity is directly applied as deterministic constraints on the solution. Box constraints on $\mathbf{x}$ are a related case which has received some interest as well [15,19]. In all these approaches, however, the goal is a single cutoff which may be determined for the system itself, whereas as we show in this letter, the answer actually varies, generally depending on the data and even between elements of the unknown vector.

Uniqueness can be directly related to system resolution, as suggested in Backus–Gilbert theory [1,3], though the approach is limited to ℓ_2-based penalties. Stark [21] proposed an extension of this approach to incorporate arbitrary forms of prior knowledge using optimization. However it is not clear whether the optimization problem is tractable for particular implementations, and a formulation for discrete systems is not provided. A different direction is introduced by Candès [7], where gains due to sparsity of the unknown vector are described in terms of a super-resolution factor, essentially a higher-resolution cutoff. However this method requires the unknown to have a very particular structure, such as an impulse train.

In this paper we will formulate a novel approach to uniqueness by providing conditions on an element-wise (i.e., coordinate-wise) basis. This approach allows us to directly use convex optimization theory and makes the relationship to the classical case (i.e., with no prior knowledge) clear. Further, we may relax the conditions with a test for uniqueness that, when it fails, can provide a resolution estimate for the system. The estimates can be formulated as linear programs which can be efficiently solved using off-the-shelf software [4]. Finally, we provide simulations for different super-resolution scenarios, demonstrating how achievable resolution varies depending on both prior knowledge used and with the object itself, and how we are able to extract additional high-resolution information which would otherwise be lost if we used a single global resolution cutoff.

2 Methods

In our analysis we will neglect noise and model errors, presuming they are addressed by a prior denoising step, and so assume our underdetermined system $\mathbf{Ax} = \mathbf{b}$ has infinite solutions, which form the set,

$$F_{\text{EC}} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{Ax} = \mathbf{b}\}. \tag{1}$$

The subscript "EC" implies the solutions are equality-constrained. In this paper we will consider the following set

which has an added restriction representing our prior knowledge about the solution,

$$F_{\text{M}} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b}, \mathbf{Dx} \geq \mathbf{d}\}, \tag{2}$$

where the subscript "M" implies mixed constraints. By defining $\mathbf{A}$, $\mathbf{D}$, $\mathbf{b}$, and $\mathbf{d}$ in Eq. (2) appropriately we may represent a variety of cases (see Appendix). For example, we can consider the incorporation of non-negativity, as well as forms of regularization and denoising, and combinations of these.

2.1 Uniqueness conditions

Our first goal is to derive conditions for uniqueness of the kth element x_k in F_M, for any selected $k \in \{1, \ldots, n\}$. To do this we will use optimization problems to solve for bounds on each element of $\mathbf{x}$. When we refer to bounds on an element, we imply the maximum and minimum values that unknown element may take which are consistent with the information we have, as we investigated in [10]. The bounds of the kth element (for any $k \in \{1, \ldots, n\}$) of a solution to a system are the scalar values given by

$$x_k^{(max)} = \max\{x_k \in \mathbb{R} \mid \mathbf{x} \in F\}, \tag{3}$$

$$x_k^{(min)} = \min\{x_k \in \mathbb{R} \mid \mathbf{x} \in F\}. \tag{4}$$

An element x_k is uniquely determined if $x_k^{(max)} = x_k^{(min)}$. We can test whether this is the case with the optimization problem,

$$\begin{aligned} \delta_k = \max_{\mathbf{x}} x_k \ - \ &\min_{\mathbf{x}} x_k \ = \ \max_{\mathbf{x},\mathbf{x}'} (x_k - x_k') \\ \mathbf{Ax} = \mathbf{b} \quad &\mathbf{Ax} = \mathbf{b} \quad \mathbf{Ax} = \mathbf{b} \\ \mathbf{Dx} \geq \mathbf{d} \quad &\mathbf{Dx} \geq \mathbf{d} \quad \mathbf{Ax}' = \mathbf{b} \\ & \mathbf{Dx} \geq \mathbf{d} \\ & \mathbf{Dx}' \geq \mathbf{d}. \end{aligned} \tag{5}$$

If the optimal value is $\delta_k = 0$, then x_k must be uniquely determined. Equation (5) forms a linear program, and we can use duality theory for linear programming [9] to find an upper bound on δ_k. The dual can be written as

$$\begin{aligned} \tilde{\delta}_k = \min_{\mathbf{y},\mathbf{y}',\mathbf{z},\mathbf{z}'} &\mathbf{b}^T (\mathbf{y} - \mathbf{y}') + \mathbf{d}^T (\mathbf{z} - \mathbf{z}') \\ &\mathbf{A}^T \mathbf{y} + \mathbf{D}^T \mathbf{z} = \mathbf{e}_k \\ &\mathbf{A}^T \mathbf{y}' + \mathbf{D}^T \mathbf{z}' = -\mathbf{e}_k \\ &\mathbf{z} \leq \mathbf{0} \\ &\mathbf{z}' \geq \mathbf{0}. \end{aligned} \tag{6}$$

We form uniqueness conditions by requiring a feasible point exists such that the objective equals zero, giving the conditions,

$$\mathbf{b}^T \left(\mathbf{y} - \mathbf{y}' \right) + \mathbf{d}^T \left(\mathbf{z} - \mathbf{z}' \right) = 0$$
$$\mathbf{A}^T \mathbf{y} + \mathbf{D}^T \mathbf{z} = \mathbf{e}_k$$
$$\mathbf{A}^T \mathbf{y}' + \mathbf{D}^T \mathbf{z}' = -\mathbf{e}_k \qquad (7)$$
$$\mathbf{z} \leq \mathbf{0}$$
$$\mathbf{z}' \geq \mathbf{0}.$$

As Eq. (5) calculates the difference between a maximum and minimum over the same set, $\delta_k \geq 0$. Further, if a solution exists to Eq. (7), then $\tilde{\delta}_k = 0$, since $\delta_k \leq \tilde{\delta}_k$, and by duality theory we have $0 \leq \delta_k \leq \tilde{\delta}_k = 0$. Finally, strong duality holds for linear programs under very general conditions (which we presume to hold), which requires $\delta_k = \tilde{\delta}_k$.

To understand the conditions of Eq. (7), note that if $\mathbf{D}$ and $\mathbf{d}$ are set to zeros [and hence we are back to the classical case of Eq. (1)], then the conditions can be met for any $\mathbf{y}$ such that $\mathbf{A}^T \mathbf{y} = \mathbf{e}_k$. Note that $\mathbf{e}_k$ is a column of the identity matrix, and so for the classical case $\mathbf{y}$ is simply a (transposed) row of the left inverse of $\mathbf{A}$. This condition can therefore be viewed as an element-wise version of the condition that $\mathbf{A}$ is non-singular. Note that this classical condition does not depend on $\mathbf{b}$, while the conditions of Eq. (7) do. Since $\mathbf{b} = \mathbf{Ax}$, uniqueness when there is prior knowledge included will (in general) depend on the particular value of $\mathbf{x}$ in each case. Further, for the case where there is no solution to $\mathbf{A}^T \mathbf{y} = \mathbf{e}_k$, we may still able to solve the equation $\mathbf{A}^T \mathbf{y} + \mathbf{D}^T \mathbf{z} = \mathbf{e}_k$, if we can find an appropriate choice of $\mathbf{z}$. So the prior knowledge represented by $\mathbf{Dx} \geq \mathbf{d}$ results in a restriction on the possible $\mathbf{x}$, but a relaxation of the uniqueness conditions. As a simple example, an underdetermined linear system cannot have a unique solution, but it may have a unique non-negative solution.

2.2 Resolution

Now we will relax the uniqueness conditions to provide a metric which we can then use to compare the improvement due to various cases of prior knowledge. To motivate the approach, consider the classical case again. If a $\mathbf{y}$ can be found such that $\mathbf{A}^T \mathbf{y} = \mathbf{e}_k$, then we can compute $\mathbf{y}^T \mathbf{b} = \mathbf{y}^T \mathbf{Ax} = \mathbf{e}_k^T \mathbf{x} = x_k$. So $\mathbf{y}$ is a linear functional that computes x_k from the data. In the event that finding such a functional is not possible, our goal is to find one that gets as close as possible. As depicted in Fig. 1, we replace $\mathbf{e}_k$ with a vector $\mathbf{c}$ that has some spread over multiple elements. To find the $\mathbf{c}$ closest to $\mathbf{e}_k$ we use an optimization problem such as the following,

$$d_k^{(EC)} = \min_{\mathbf{c}, \mathbf{y}} \|\mathbf{c}\|$$
$$\mathbf{A}^T \mathbf{y} = \mathbf{c}$$
$$\mathbf{c} \geq 0 \qquad (8)$$
$$c_k = 1.$$

In the case where $\mathbf{A}^T \mathbf{y} = \mathbf{e}_k$ has a solution, Eq. (8) will achieve $\mathbf{c} = \mathbf{e}_k$. Otherwise, the result will be a metric of how
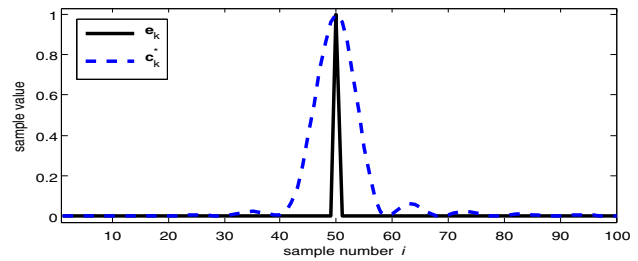

Fig. 1 $\mathbf{e}_k$ versus relaxed result for $k = 50$ with $n = 100$

similar $\mathbf{c}$ could be made to $\mathbf{e}_k$. To provide a intuitively meaningful metric, we included the constraint $\mathbf{c} \geq \mathbf{0}$ and for the norm use a ℓ_2-norm weighted with distance (in terms of spatial or temporal location of the samples) from the kth element. If the weighting increases quadratically, $\mathbf{c}$ can be viewed as a distribution over space, and the metric can be viewed as its variance. So the optimization seeks the distribution $\mathbf{c}$ about the element of interest x_k with the minimum spatial variance, such that $\mathbf{c}^T \mathbf{x}$, the local average over the spatial region, may be uniquely determined.

Similarly, the conditions of Eq. (7) can be used to form the analogous optimization problem subject to prior knowledge,

$$d_k^{(M)} = \min_{\mathbf{c}, \mathbf{y}, \mathbf{y}', \mathbf{z}, \mathbf{z}'} \|\mathbf{c}\|$$
$$\mathbf{b}^T \left(\mathbf{y} - \mathbf{y}' \right) + \mathbf{d}^T \left(\mathbf{z} - \mathbf{z}' \right) = 0$$
$$\mathbf{A}^T \mathbf{y} + \mathbf{D}^T \mathbf{z} = \mathbf{c}$$
$$\mathbf{A}^T \mathbf{y}' + \mathbf{D}^T \mathbf{z}' = -\mathbf{c} \qquad (9)$$
$$\mathbf{z} \leq \mathbf{0}$$
$$\mathbf{z}' \geq \mathbf{0}$$
$$\mathbf{c} \geq 0$$
$$c_k = 1.$$

The constraints are linear, so this is a convex optimization problem.

3 Simulations

To demonstrate the approach, we formed three different simulations. We used CVX [17,18] to solve the optimization problems, with the problems of Eqs. (8) and (9); the matrices were formed as described in the Appendix. We also used other published methods for comparison where possible.

3.1 Example 1: Structured one-dimensional system

First we simulated a one-dimensional system which performs a low-pass filtering and downsamples the result by a factor of two. The true vector $\mathbf{x}$, the convolution kernel, and the low-pass-filtered result $\mathbf{b}$ are shown in Fig. 2. In Fig. 3 we compare $\mathbf{x}^{(true)}$ to some regularized estimates, including Basis Pursuit and non-negative least-squares (NNLS) reconstructions, as
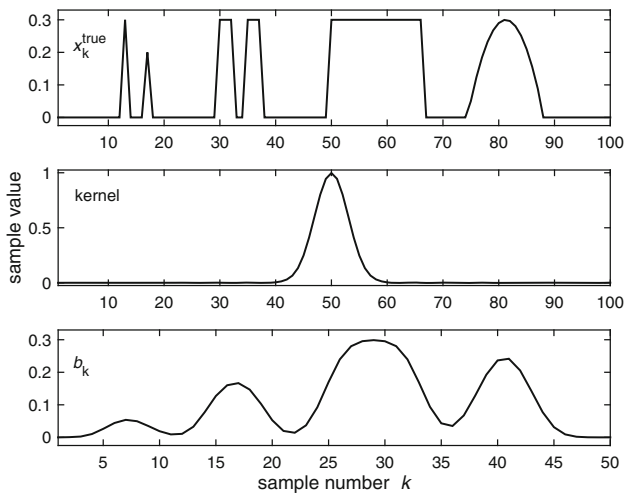
Fig. 2 Test input $\mathbf{x}^{(true)}$, the true values of the unknown vector, kernel convolved with $\mathbf{x}$ prior to downsampling, and measured data $\mathbf{b} = \mathbf{A}\mathbf{x}^{(true)}$; $\mathbf{A}$ is $m \times n$ with $n = 100$ and $m = 50$
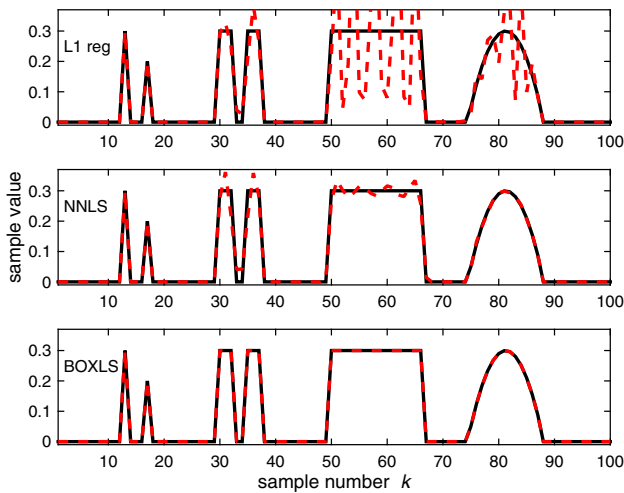

Fig. 3 Regularized estimates with different techniques: ℓ_1-regularized, a.k.a Basis Pursuit (L1); non-negative least squares (NNLS); box-constrained least squares (BOXLS). *Dashed trace* is a estimate, and *solid trace* is true $\mathbf{x}$ for comparison

well as "BOXLS," a result analogous to NNLS but with box constraints (both a lower and upper constraints) on $\mathbf{x}$, where we use the constraint $0 \le x_i \le 0.3$ for each element of $\mathbf{x}$.

We see that ℓ_1-regularization did not yield a very accurate result; on the left side of $\mathbf{x}$, where the signal is locally sparse we have a correct estimate, but on the right side of the plot where $\mathbf{x}$ is denser, the estimate is incorrect. NNLS gave a better result, but still was incorrect in the densest region in the center right of the plot. BOXLS produced an apparently perfect result, as we used both the true upper and lower limits as prior knowledge.

Figure 4 gives element-wise resolution estimates computed via multiple different methods. We calculated a discrete implementation of the Backus–Gilbert method [1,20], which we see performs similarly to the equality-constrained method
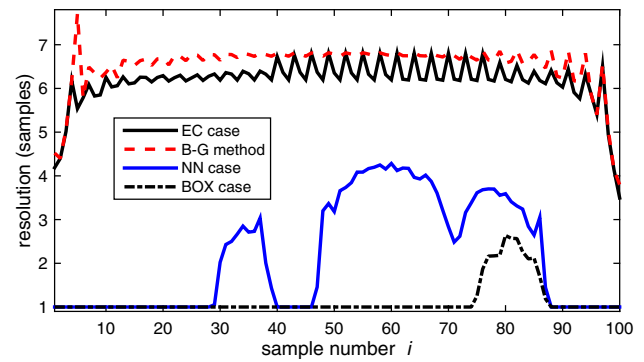
Fig. 4 Resolution estimate for each sample for different cases: EC case computed using Eq. (8), discrete implementation of the Backus–Gilbert method (B–G), and NN and BOX cases based on Eq. (9) utilizing non-negativity and box constraints, respectively
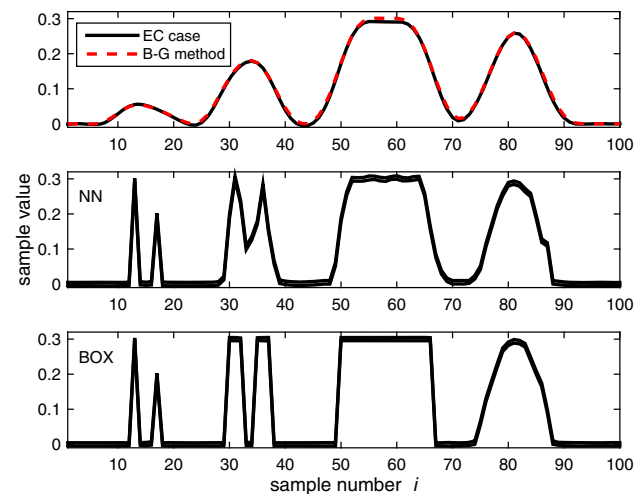
Fig. 5 Low-resolution estimates of $\mathbf{x}$ for different cases; essentially an adaptive estimate that varies in resolution depending on the best resolution achievable at each sample

of Eq. (8). The resolution is also given using Eq. (9) for non-negativity and box constraints. We also provide element-wise "low-resolution" estimates using the optimal resolution cells, i.e., $\mathbf{c}^T \mathbf{b}$, analogous to $\mathbf{e}_k^T \mathbf{x}$, in Fig. 5. The equality-constrained and Backus–Gilbert methods return essentially constant resolutions (except for edge effects) which quantify the amount of low-pass filtering performed by the kernel. The box-constrained case achieves best resolution (resolution = 1 sample implies $\mathbf{c} = \mathbf{e}_k$) for most of the elements, as we might have guessed given the accurate reconstruction, except in a small interval around sample 80. This poorer-resolution region underlines the fact that an apparently accurate regularized reconstruction does not necessarily imply a unique solution and hence a sufficient system resolution. The non-negative case achieved results in between. These results demonstrate the key determinant of uniqueness and resolution improvement with prior knowledge, which is active constraints, be they active non-negativity constraints (mean-

ing zeros in the signal) for a sparse signal, or a signal reaching both min and max values for a box constraints.

3.2 Example 2: Chirped impulse train

Next we formed a model consisting of impulses with varying amplitudes and intervals, so we could compare the method to the estimates of [7], which require such structure. The pulses are monotonically decreasing at a linear rate to discern the cutoff where the pulse repetition rate becomes too high for different methods. In this example, in addition to a low-pass filtering kernel as in the first example, we imposed a hard low-pass cutoff at a frequency of 75 cycles, corresponding to a wavelength of $\lambda_c = 13.3$ samples. Figure 6 gives the true signal, the filtered version **b**, and the ℓ_1-regularized reconstruction via Basis Pursuit.

Figure 7 gives element-wise resolution estimates using the discrete Backus–Gilbert method, the estimate of Eq. (9) utilizing non-negativity, and a cutoff estimated using the principles of [7] which is labeled SRF (denoting a super-resolution factor) limit. For the Backus–Gilbert method we again see the essentially constant behavior, independent of signal structure. For Eq. (9) optimization we see an estimate of high resolution (i.e., a single pixel) for the left half of the signal, where the impulses are widely spaced; as the pulse intervals become shorter, the resolution transitions to the spectral cutoff of approximately 13 samples. This roughly agrees with our ability to discern individual pulses in **b** and in the accuracy of the ℓ_1 solution in Fig. 6. The resolution estimate is more conservative, as it determines when samples are able to be uniquely determined at the given resolution, while a probability maximization approach such as ℓ_1-regularization may still serendipitously achieve a correct estimate. However
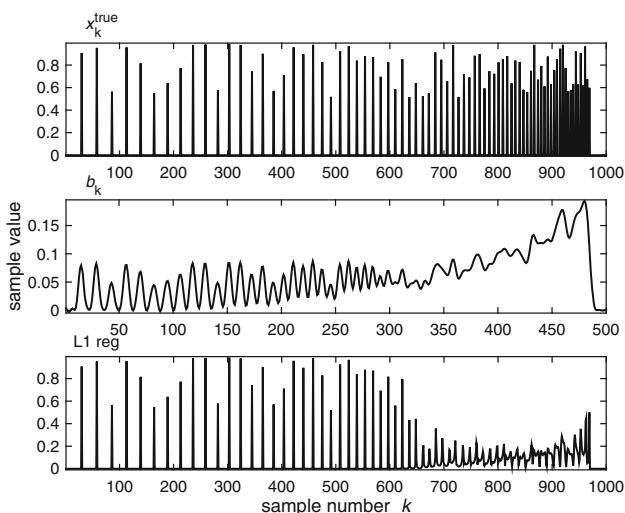
Fig. 6 Chirped impulse train, low-pass-filtered version, and ℓ_1-regularized estimate
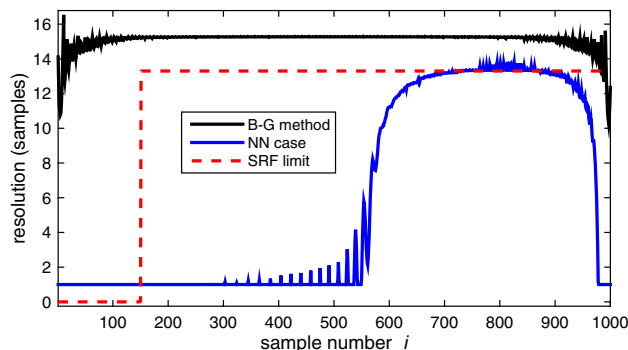
Fig. 7 Resolution estimate for Backus–Gilbert (B–G) method, the method of Eq. (9) utilizing non-negativity (NN case), and an analytical estimate based on [7] (SRF limit)

our result tells us that the ℓ_1-regularized result is not reliable for these shorter pulse intervals.

The SRF limit was determined according to [7], where unlimited super-resolution of the impulse is possible as long as the spacing is at least $1.87\lambda_c$, for real signals. Note that this result is significantly more conservative as it does not take advantage of non-negativity, and it requires a cutoff where the result (as long as it is composed of impulses) may be infinitely resolved; hence, the resolution estimate is zero (meaning zero-width resolution cells, and perfect resolvability), while our estimate is "1," to the far left of the signal, where the pulse intervals are greater than approximately 29. For intervals shorter than that of cutoff, we set the estimate equal to the filter cutoff for the system. Note that here we also presumed the cutoff could be applied to a signal on a partial basis, rather than discarding the high-resolution signal completely due to the less-resolvable region on the right.

### 3.3 Example 3: Two-dimensional image

In the final simulation, we analyze the resolution for a noisy blurred image, again taking into account non-negativity as our prior knowledge. We used the non-negative denoising (NNDN) formulation given in the Appendix. The true image is given in Fig. 8, a blurred and downsampled version with 1 % noise is given in Fig. 9, and a NNLS estimate is given in Fig. 10.

In two dimensions, an element-wise estimate performed for every pixel becomes challenging due to the large number of pixels. However we may make use of several different tactics to reduce computational time. First note that estimates for different pixels may be calculated completely independently, allowing parallelization up to the number of available processors. In our case we utilized a quad core processor achieving a reduction in time by a factor of four. Further, cases where the pixel values are uniquely determined (resolution achieves unity) can be screened using an efficient feasibility check of the uniqueness equations of Eq. (7).
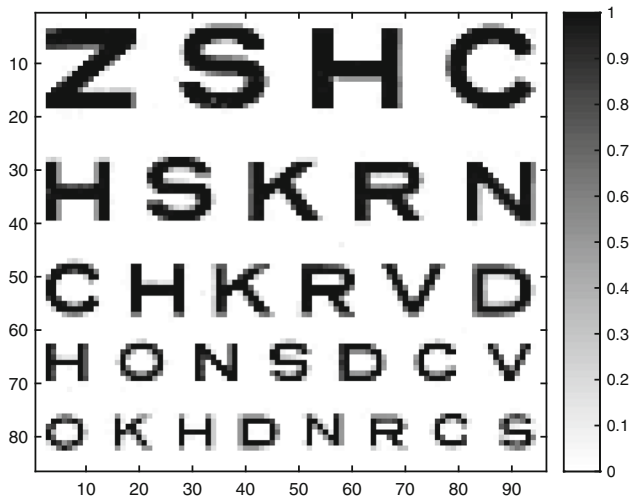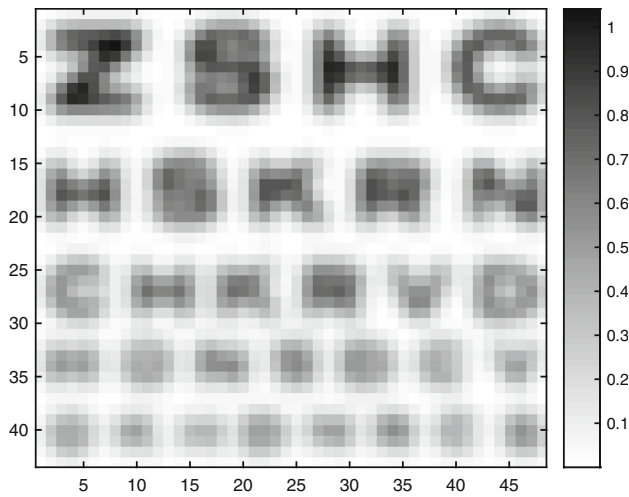
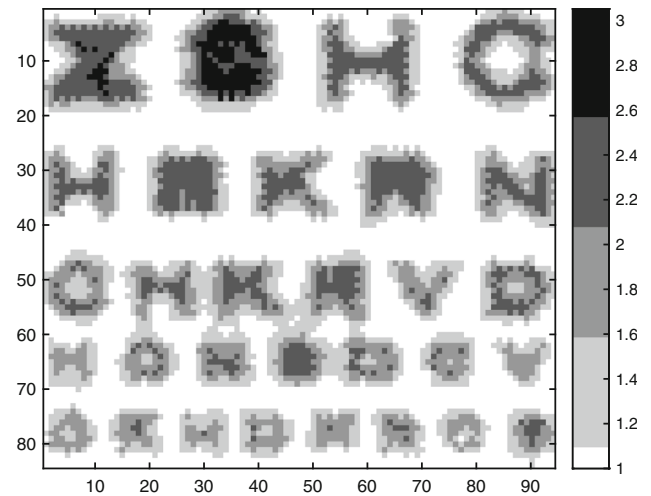**Fig. 8** True image of an eyechart, prior to filtering
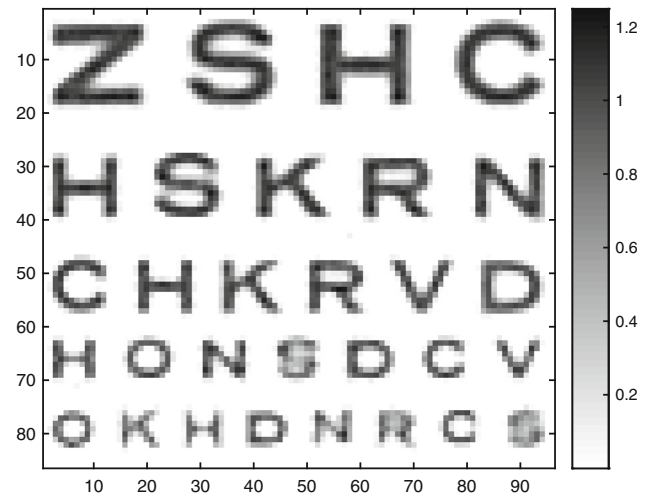
Fig. 10 Non-negative least-squares estimate of image

Fig. 9 Low-resolution image; result of blurring, downsampling by a factor of two, and addition of 1 % noise

Fig. 11 Element-wise resolution estimates; note that the pixels for smallest characters achieve roughly double the resolution of the largest characters; also note that pixels consistently achieve poorer resolution for the characters which were most poorly reconstructed in Fig. 10, such as the letter "s"

Resolutions for approximately 40 % of the pixels could be determined this way for our example. Finally, for larger signals or images one may truncate a local region for each estimate with a sliding window, to provide a problem small enough to be tractable but large enough to include sufficient neighboring pixels for a given location. In all, we computed the resolution estimate of Fig. 11 in approximately 4 h on a 3.2 GHz desktop processor with general optimization software. For comparison, the 1000-sample estimate of Example 2 took approximately 5 min.

In this example, the smallest characters had features on the order of a pixel across, while a resolution of 1.5–2.0 pixels was mostly achieved for them. The largest characters, conversely, had features four pixels in size, while a resolution of two to three pixels was achieved. As a result, we are able to better discern the larger characters despite the limit of a lower resolution. The Backus–Gilbert resolution for this problem

was computed using a version of the algorithm which can accommodate noise [20], yielding a uniform resolution estimate of 3.5 pixels. As before, this is roughly the worst case compared with the estimates found with Eq. (9). Hence, even with a more sophisticated resolution estimates which incorporate prior knowledge, if only a single resolution cutoff is sought it often will not show improvement over conventional resolution estimates.

4 Discussion

In this paper we gave uniqueness conditions for each element of a system of equations and inequality constraints. This

element-wise approach allowed our conditions to be tested using convex optimization, which, in turn, allowed us to estimate resolution on an element-wise basis and incorporating prior knowledge. As we saw with the simulated examples, regularization techniques such as NNLS and Basis Pursuit can achieve higher-resolution results than a conventional resolution estimate suggests. Indeed, this is the very important reason for using such methods. The additional information our resolution estimates provide allows us to better understand such regularized results. For example, in the simulation, reconstruction of the letter "s" consistently achieved lower resolution (slightly apparent in the poorer reconstruction of this letter in the NNLS result). Knowing this, we might ascribe lower confidence to such letters in a subsequent classification stage. Further, we saw that while the resolution cutoff remained fixed (i.e., it was spatially invariant across the image), the resolution improved as the character size got smaller; hence, the smallest characters could actually be reconstructed surprisingly well using NNLS for this example, due to the low noise level and non-negativity prior.

Generally, the resolution cell estimate is very interesting when inequality constraints are included, as it yields a data-dependent result. In the case of non-negativity, this result depends on the sparsity of the elements which are mixed with our element of interest. In the case of more general inequality constraints, the sparsity condition would be replaced with a metric of the number of active constraints. For the simulated cases, essentially super-resolution problems, this mixing is localized so we see the effect due to the active constraints in local regions. For such a system, a concentrated resolution estimate makes sense. For more arbitrary systems a concentrated resolution cell may not be achievable. This would imply that the ambiguity between high-resolution elements cannot be explained with any locally concentrated combination. Our method could easily be extended to such problems, to find the best resolution cell via some other desirable property, such as the smallest number of combined pixels without regard for localization. There are also a variety of ways one could estimate the most compact resolution cell for each pixel. The ℓ_2-norm was used here as it yielded an intuitive interpretation in terms of the variance of a distribution over space or time.

The computational complexity of the technique requires one optimization problem per estimate, which poses a challenge for larger problems. In the simulations we described a number of ways to alleviate this, including windowing the problem, screening of unique samples, and parallelization. A variety of other strategies may be helpful as well. When the low-resolution distributions are large, the estimates at neighboring elements are mostly redundant. We can therefore choose to increase the spacing between estimates such that we still achieve a covering of all elements. Further, while we used an off-the-shelf solver, one can typically achieve sig-

nificant improvements with a customized algorithm which takes advantage of the structure of the problem.

Appendix

In this appendix we will describe how a selection of variations on prior knowledge can be formulated as linear inequality constraints. Again the classical case with no prior knowledge is based on the solution set F_{EC}, with $\mathbf{D} = \mathbf{0}$ and $\mathbf{d} = \mathbf{0}$.

$$F_{\text{EC}} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b}\}. \tag{10}$$

Application of our bounds testing problem with the feasible set $\mathbf{x} \in F_{EC}$ forms an equality-constrained linear program [16], for which optimality conditions give the row space condition $\mathbf{A}^T \mathbf{y} = \mathbf{e}_k$.

Non-negativity results in the solution set

$$F_{NN} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}. \tag{11}$$

This can be implemented in our system with the simple definitions, $\mathbf{D} = \mathbf{I}$, $\mathbf{d} = \mathbf{0}$, using the identity matrix and a vector of zeros.

$\ell_1 - regularization$ can be formulated as a case of non-negativity, which can be used to determine whether we have a unique optimal solution to the Basis Pursuit problem [8],

$$\alpha = \min_{\mathbf{x}} \|\mathbf{x}\|_1$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{12}$$

This can be tested by analyzing in the uniqueness of the solutions in the following set,

$$F_{BP} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b}, \|\mathbf{x}\|_1 = \alpha\}$$
$$= \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} = \mathbf{b}, \|\mathbf{x}\|_1 \leq \alpha\}. \tag{13}$$

This is equivalent to the following non-negative system,

$$F_{NN} = \{\hat{\mathbf{x}} \in \mathbb{R}^{2n} | \hat{\mathbf{A}}\hat{\mathbf{x}} = \hat{\mathbf{b}}, \hat{\mathbf{x}} \geq \mathbf{0}\}. \tag{14}$$

With the definitions

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{A}, -\mathbf{A} \\ \mathbf{1}^T \end{pmatrix}, \quad \hat{\mathbf{b}} = \begin{pmatrix} \mathbf{b} \\ \alpha \end{pmatrix}. \tag{15}$$

This can be seen by defining $\mathbf{x} = \hat{\mathbf{x}}_{(1)} - \hat{\mathbf{x}}_{(2)}$, where $\hat{\mathbf{x}}^T = \left(\hat{\mathbf{x}}_{(1)}^T, \hat{\mathbf{x}}_{(2)}^T\right)$ and $\hat{\mathbf{x}}_{(1)} \geq \mathbf{0}, \hat{\mathbf{x}}_{(2)} \geq \mathbf{0}$. We relate bounds found using the feasible set of Eq. (14) to the bounds for the set of Eq. (13) by noting that at the minimum, where we get α as the optimal for Eq. (12), $\hat{\mathbf{x}}_{(1)}$ and $\hat{\mathbf{x}}_{(1)}$ are complementary. If

they were not, we could take advantage of this fact to reduce the minimum of $\|\mathbf{x}\|_1 = \hat{\mathbf{x}}_{(1)} + \hat{\mathbf{x}}_{(2)}$ further.

Box constraints define the following set,

$$F_{BOX} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b}, \ \mathbf{d}_{min} \le \mathbf{x} \le \mathbf{d}_{max}\}. \quad (16)$$

Here $\mathbf{d}_{min}$ and $\mathbf{d}_{max}$ are vectors defining the box. We can formulate this as Eq. (2) with the definitions

$$\mathbf{D} = \begin{pmatrix} +\mathbf{I} \\ -\mathbf{I} \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} +\mathbf{d}_{min} \\ -\mathbf{d}_{max} \end{pmatrix}. \quad (17)$$

We can view this as a more general version of regularization with the infinity norm, e.g.,

$$F_{BOX} = \{\mathbf{x} \in \mathbb{R}^n | \mathbf{Ax} = \mathbf{b}, \ \|\mathbf{x}\|_\infty \le d\}. \quad (18)$$

Denoising can be viewed as a dual to regularization, where rather than requiring the solution set be regular, we require the error in the linear system to be regular, as in the following,

$$F_{DN} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{Ax} - \mathbf{b}\| \le \sigma, \}. \quad (19)$$

We can also form a denoised version of the non-negativity case using the infinity norm as follows,

$$F_{NND} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{Ax} - \mathbf{b}\|_\infty \le \sigma, \ \mathbf{x} \ge \mathbf{0}\}. \quad (20)$$

This can be formulated as mixed constraints with no linear constraint term (i.e., "$\mathbf{A}$" and "$\mathbf{b}$" in the original linear system are all zeros), and with

$$\mathbf{D} = \begin{pmatrix} -\mathbf{A} \\ \mathbf{A} \\ \mathbf{I} \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} -\mathbf{b} - \sigma\mathbf{1} \\ \mathbf{b} - \sigma\mathbf{1} \\ \mathbf{0} \end{pmatrix}. \quad (21)$$

References

1. Backus, G., Gilbert, F.: The resolving power of gross earth data. Geophys. J. R. Astron. Soc. **16**(2), 169–205 (1968)
2. Becker, S.R., Cands, E.J., Grant, M.C.: Templates for convex cone problems with applications to sparse signal recovery. Math. Program. Comput. **3**(3), 165–218 (2011)
3. Bertero, M., Mol, C.D., Pike, E.R.: Linear inverse problems with discrete data. I. General formulation and singular system analysis. Inverse Probl. **1**(4), 301 (1985)
4. Boyd, S.P., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
5. Bruckstein, A.M., Elad, M., Zibulevsky, M.: On the uniqueness of non-negative sparse and redundant representations. In: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, pp. 5145–5148 (2008)
6. Candes, E.J.: The restricted isometry property and its implications for compressed sensing. Comptes Rendus Mathematique **346**(910), 589–592 (2008)
7. Candes, E.J., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. Commun. Pure Appl. Math. **67**, 906–956 (2014)
8. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)
9. Dantzig, G.: Linear Programming and Extensions. Princeton University Press, Princeton (1998)
10. Dillon, K., Fainman, Y.: Bounding pixels in computational imaging. Appl. Opt. **52**(10), D55–D63 (2013)
11. Donoho, D.L.: Neighborly Polytopes and Sparse Solutions of Underdetermined Linear Equations. In: Technical Report, Stanford University (2005)
12. Donoho, D.L., Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. Proc. Natl. Acad. Sci. **100**(5), 2197–2202 (2003)
13. Donoho, D.L., Tanner, J.: Neighborliness of randomly projected simplices in high dimensions. Proc. Natl. Acad. Sci. USA **102**(27), 9452–9457 (2005)
14. Donoho, D.L., Tanner, J.: Sparse nonnegative solution of underdetermined linear equations by linear programming. Proc. Natl. Acad. Sci. USA **102**(27), 9446–9451 (2005)
15. Donoho, D.L., Tanner, J.: Counting the faces of randomly-projected hypercubes and orthants, with applications. Discrete Comput. Geom. **43**(3), 522–541 (2010)
16. Gill, P.E., Murray, W., Wright, M.H.: Numerical Linear Algebra and Optimization. Advanced Book Program. Addison-Wesley Pub. Co., Boston (1991)
17. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control. Lecture Notes in Control and Information Sciences, vol. 371, pp. 95–110. Springer, Berlin/Heidelberg (2008)
18. Grant, M. and Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.0 beta. http://cvxr.com/cvx (2013)
19. Petra, S., Schrader, A., Schnrr, C.: 3D tomography from few projections in experimental fluid dynamics. In: Nitsche, W., Dobriloff, C. (eds.) Imaging Measurement Methods for Flow Analysis, No. 106 in Notes on Numerical Fluid Mechanics and Multidisciplinary Design, pp. 63–72. Springer, Berlin, Heidelberg (2009)
20. Press, W.H.: Numerical Recipes. The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge (2007)
21. Stark, P.B.: Generalizing resolution. Inverse Probl. **24**(3), 034014 (2008)
22. Tillmann, A., Pfetsch, M.: The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Trans. Inf. Theory **60**(2), 1248–1259 (2014)
23. Willett, R.M., Marcia, R.F., Nichols, J.M.: Compressed sensing for practical optical imaging systems: a tutorial. Opt. Eng. **50**(7), 072 (2011). 601-072, 601-13