Signal Processing 123 (2016) 1–8

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

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.

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 rowspace 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).

http://dx.doi.org/10.1016/j.sigpro.2015.12.009
0165-1684/

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,

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

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

$$S_{NN}=\left\{\mathbf{x}|\mathbf{Ax}=\mathbf{b}',\mathbf{x}\geq\mathbf{0}\right\}, \tag{2}$$

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 rowspace 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 rowspace 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:

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

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