Robust sparse recovery with non-negativity constraints

Martin Slawski and Matthias Hein
Department of Computer Science, Saarland University
Email: {ms, hein}@cs.uni-saarland.de

***Abstract*—It has been established recently that sparse non-negative signals can be recovered using non-negativity constraints only. This result is obtained within an idealized setting of exact sparsity and absence of noise. We propose non-negative least squares – without any regularization – followed by thresholding for the noisy case. We develop conditions under which one can prove a finite sample result for support recovery and tackle the case of an approximately sparse target. Under weaker conditions, we show that non-negative least squares is consistent for prediction. As illustration, we present a feature extraction problem from Proteomics.**

I. Introduction

In various applications, the sparse target $\beta^* \in \mathbb{R}^p$ to be recovered is known to be non-negative. Several recent papers discuss to what extent this additional prior knowledge may simplify the problem of recovering β^* from n, $n < p$, uncorrupted linear measurements $y = X\beta^*$. In [1], [2], [3], it is pointed out that ℓ_1-minimization is no longer needed if the set $A = \{\beta : y = X\beta, \ \beta \succeq 0\}$ is a singleton. Donoho and Tanner [2] study the faces of the cone $X\mathbb{R}_+^p$ generated by the columns of X, showing that for random matrices with entries from a symmetric distribution, A fails to be a singleton with high probability if $n < 2p$ already for $s = 0$, where $s = |S|, \ S = \{j : \beta_j^* > 0\}$. On the other hand, they show that with X as the concatenation of a row of ones and a random Gaussian matrix $\widetilde{X}$, the faces of $X\mathbb{R}_+^p$ are in a one-to-one relation with those of $\widetilde{X}T^{p-1}$, where T^{p-1} is the standard simplex in $\mathbb{R}^p$, i.e. A is a singleton if and only if $\operatorname{argmin}_{\beta \in \widetilde{A}} \mathbf{1}^\top \beta, \ \widetilde{A} = \{\beta : \widetilde{X}\beta^* = \widetilde{X}\beta, \ \beta \succeq 0\}$ is. A similar result is shown in [3] with $\widetilde{X}$ replaced by a random binary matrix. In [4], we have generalized these two positive results to concatenations of random isotropic sub-Gaussian matrices and a row of ones as well as to random matrices with entries from a sub-Gaussian distribution on $\mathbb{R}_+$. A major shortcoming of these results is that they are derived within a little realistic noise-free setting, and it is unclear how they can be transferred to the noisy case. Contradicting the well-established paradigm in statistics suggesting that a regularizer is necessary to prevent over-adaptation to noise, we show that such a transfer is indeed possible.

II. Sparse recovery for the noisy case

A. Approach

In [4], we assume that $y = X\beta^* + \varepsilon$, where ε is zero-mean sub-Gaussian noise with parameter σ. We suggest to find a minimizer $\widehat{\beta}$ of the non-negative least squares (NNLS) criterion $\min_{\beta \succeq 0} \|y - X\beta\|_2^2$ first, and to estimate the support S of β^* by $\widehat{S}(\lambda) = \{j : \widehat{\beta}_j(\lambda) > 0\}$, where $\widehat{\beta}(\lambda)$ is obtained by hard thresholding $\widehat{\beta}$ with threshold $\lambda \geq 0$, i.e. all components of $\widehat{\beta}$ smaller than λ are set to zero.

B. Key condition and main result

In the noiseless case, S can be recovered if $X_S\mathbb{R}_+^s$ is a face of $X\mathbb{R}_+^p$, i.e. there exists a hyperplane separating the cone generated by the columns of the support $\{X_j\}_{j \in S}$ from the cone generated by the columns of the off-support $\{X_j\}_{j \in S^c}$. For the noisy case, we employ a quantitative notion of separation captured by the constant

$$\widehat{\tau}(S) = \max_{\tau, \ w : \|w\|_2 \leq 1} \tau \ \text{ sb.t. } X_S^\top w = 0, \ \ n^{-1/2} X_{S^c}^\top w \succeq \tau \mathbf{1}.$$

From convex duality, it is easy to see that $\widehat{\tau}(S)$ equals the distance of the subspace spanned by X_S and the simplex generated by X_{S^c}. Based on this relation, we investigate how $\widehat{\tau}(S)$ scales in dependency of n, p, s. We find that $\widehat{\tau}^2(S)$ is of the order s^{-1} minus a random deviation term for the random designs well-suitable for sparse recovery in the noiseless case as mentioned in Section 1.
A brief, qualitative version of our main result is as follows.

Theorem. *Set* $\lambda > \frac{2\sigma}{\widehat{\tau}^2(S)} \sqrt{\frac{2 \log p}{n}}$. *If* $\min_{j \in S} \beta_j^* > \widetilde{\lambda}, \ \widetilde{\lambda} = \lambda C(S)$, *for a constant* $C(S)$, $\widehat{\beta}(\lambda)$ *satisfies* $\|\widehat{\beta}(\lambda) - \beta^*\|_\infty \leq \widetilde{\lambda}$, *and* $\widehat{S}(\lambda) = S$, *with high probability*.

III. Approximately sparse targets

Using a lower bound on $\widehat{\tau}(S)$ again, we can bound the reconstruction error as long as β^* is concentrated on components in S.

IV. Prediction consistency

We show that for a broad classes of non-negative designs, NNLS posses a 'self-regularizing property' which prevents over-adaption to noise. For these designs, the mean square prediction error $n^{-1}\|X\widehat{\beta} - X\beta^*\|_2^2$ is upper bounded by a term of order $O(\|\beta^*\|_1 \sqrt{\log p/n})$, a result resembling that obtained in [5] for ℓ_1-regularized least squares.

V. Application

An important challenge in the analysis of protein mass spectrometry data is to extract peptide masses from a raw spectrum. In [6], this is formulated as a sparse recovery problem with non-negativity constraints in the presence of heteroscedastic noise. It is demonstrated that NNLS plus thresholding with a locally adaptive threshold out-performs standard sparse recovery methods.

References

[1] A. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4813–4820, 2008.
[2] D. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Disc. Comp. Geometry*, vol. 43, pp. 522–541, 2010.
[3] M. Wang, W. Xu, and A. Tang, "A unique nonnegative solution to an undetermined system: from vectors to matrices," *IEEE Trans. Signal Proc.*, vol. 59, pp. 1007–1016, 2011.
[4] M. Slawski and M. Hein, "Non-negative least squares for sparse recovery in the presence of noise," In preparation, 2011.
[5] E. Greenshtein and Y. Ritov, "Persistence in high-dimensional linear predictor selection and the virtue of overparametrization," *Bernoulli*, vol. 6, pp. 971–988., 2004.
[6] M. Slawski, R. Hussong, A. Tholey, T. Jakoby, B. Gregorius, A. Hildebrandt, and M. Hein, "Peak pattern deconvolution for Protein Mass Spectrometry by Non-Negative Least Squares/Least Absolute Deviation template matching," *submitted*, 2011.