ROBUST NON-NEGATIVE LEAST SQUARES USING SPARSITY

Filip Elvander, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson

Centre for Mathematical Sciences, Lund University, Sweden.
email: {filipelv, sia, aj}@maths.lth.se

ABSTRACT

Sparse, non-negative signals occur in many applications. To recover such signals, estimation posed as non-negative least squares problems have proven to be fruitful. Efficient algorithms with high accuracy have been proposed, but many of them assume either perfect knowledge of the dictionary generating the signal, or attempts to explain deviations from this dictionary by attributing them to components that for some reason is missing from the dictionary. In this work, we propose a robust non-negative least squares algorithm that allows the generating dictionary to differ from the assumed dictionary, introducing uncertainty in the setup. The proposed algorithm enables an improved modeling of the measurements, and may be efficiently implemented using a proposed ADMM implementation. Numerical examples illustrate the improved performance as compared to the standard non-negative LASSO estimator.

***Index Terms*—** robust non-negative least squares, ADMM

1. INTRODUCTION

Non-negative least squares problems occur in a wide variety of fields, such as hyperspectral imaging [1, 2], DNA microarray analysis [3, 4], deconvolution [5], audio processing [6], and spectroscopy [7], and the general problem has as a result attracted notable attention in the recent literature (see, e.g., [8–10]). Much of this work has focused on finding computationally efficient and statistically reliable non-negative least squares (NN-LS) estimators, as well as examining their convergence properties. In this work, we build upon these earlier efforts, striving to formulate a robust NN-LS estimator, allowing the measured signal to be constructed from vectors deviating from those occurring in the dictionary. This would, for instance, be the case when the dictionary elements are formed from some kind of clean reference signals that are then not occurring exactly in the same form in the actual measurements, such as would commonly occur if the reference signals are measured in a laboratory environment, whereas the measurements of interest are formed using, e.g., fewer samples, cheaper and/or less precise instruments (see also, e.g., [7, 11–13]).

To elaborate on this notion, consider the measured non-negative signal

$$\mathbf{y} = \begin{bmatrix} y(1) & \dots & y(N) \end{bmatrix}^T \tag{1}$$

where $(\cdot)^T$ denotes the transpose, and

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{e} \tag{2}$$

where $\mathbf{X}$ is an $N \times m$ dictionary matrix with non-negative entries, $\mathbf{a}$ is a sparse non-negative vector of amplitudes, with unknown sparsity, K, and $\mathbf{e}$ an additive non-negative noise. If the dictionary matrix $\mathbf{X}$ is known, one may thus form the non-negative least squares estimate as

$$\begin{aligned} &\underset{\mathbf{a}}{\text{minimize}} && \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 \\ &\text{s.t} && \mathbf{a} \geq \mathbf{0} \end{aligned} \tag{3}$$

where the inequality constraint should be interpreted elementwise. This problem may be solved efficiently using, for instance, a non-negativity constrained LASSO, i.e.,

$$\begin{aligned} &\underset{\mathbf{a}}{\text{minimize}} && \|\mathbf{y} - \mathbf{X}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1 \\ &\text{s.t} && \mathbf{a} \geq \mathbf{0} \end{aligned} \tag{4}$$

or via computationally more efficient solutions such as the one presented in, e.g., [10]. However, the accuracy of the estimate of $\mathbf{a}$ obtained using such methods depends strongly on the assumption that the dictionary $\mathbf{X}$ is perfectly known. If the assumed dictionary deviates significantly from the $\mathbf{X}$ generating the signal, such approaches will not allow $\mathbf{a}$ to be accurately reconstructed. Similarly, the accuracy of the estimates will degrade in cases when some columns of $\mathbf{X}$ are unknown, thus modeling the scenario of unknown substances being present in the signal (as examined in [7]). In applications where the reference signals used to construct the dictionary may differ from those actually measured, such as would commonly occur when different measurement equipments are used to form the different measurements, it is reasonable to assume that the elements of the dictionary $\mathbf{X}$ are only approximately the same as those occurring in the measurements, allowing $\mathbf{X}$ to be decomposed as $\mathbf{X} = \mathbf{\Phi} + \mathbf{\Delta}$,

This work was supported in part by the Swedish Research Council, Carl Trygger's foundation, and the Royal Physiographic Society in Lund.

where $\mathbf{\Phi}$ is the known part of the dictionary and $\mathbf{\Delta}$ is a matrix of perturbations. In this work, we assume that $\mathbf{\Delta}$ is small relative to $\mathbf{\Phi}$, such that the dictionary elements are similar, although not precisely the same, as the signatures occurring in the measurements. Without loss of generality, we further assume that $\mathbf{\Phi}$ has been normalized such that $\|\mathbf{\Phi}_i\|_2 = 1$, for each dictionary column i, and that $\|\mathbf{\Delta}_i\|_2 \leq \epsilon$, for some $\epsilon > 0$. Given these assumptions, we propose a computationally efficient and robust NN-LS estimator allowing for some flexibility in the assumed dictionary.

2. ROBUST NON-NEGATIVE LEAST SQUARES

In order to achieve the desired robustness, we form the estimates of $\mathbf{a}$ by allowing the columns of the dictionary used in the estimation to vary within balls of radii ϵ, centered on the columns of $\mathbf{\Phi}$. This can be accomplished by introducing a new variable, $\mathbf{W}$, defined such that every column of $\mathbf{W}$ equals the corresponding column of $\mathbf{X}$, scaled with the corresponding amplitude of $\mathbf{a}$. Thus, $\mathbf{W1} = \mathbf{Xa}$, where $\mathbf{1}$ is a vector of ones. As a result, the optimization problem can be expressed as

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & \|\mathbf{y} - \mathbf{W1}\|_2^2 + \lambda \sum_{i=1}^{m} \|\mathbf{W}_i\|_2 \\ \text{s.t} \quad & \mathbf{W} \geq \mathbf{0} \\ & \left\| \frac{\mathbf{W}_i}{\|\mathbf{W}_i\|_2} - \mathbf{\Phi}_i \right\|_2 \leq \epsilon, \ i = 1, \ldots, m \end{aligned} \tag{5}$$

with the inequalities interpreted elementwise. Here, the second term of the objective function penalizes the norm of the columns of $\mathbf{W}$, thus having a sparsifying effect such that only a few of the columns will be non-zero. The second constraint in (5) ensures that the dictionary vectors of $\mathbf{W}$ do not deviate too much from the assumed dictionary, i.e., $\mathbf{\Phi}$. This may equivalently be expressed as

$$\|\mathbf{W}_i\|_2 (2 - \epsilon^2) - 2\mathbf{W}_i^T \mathbf{\Phi}_i \leq 0, \ \forall i \tag{6}$$

which is a convex constraint for $\epsilon^2 < 2$. Thus, (5) is a convex optimization problem, which may be solved using standard convex optimization software such as, for instance, CVX [14,15]. However, such estimators often scale poorly with increasing data lengths, typically ensuring an unnecessary high computational complexity. To alleviate this, we here proceed to introduce an alternating direction method of multipliers (ADMM) framework (see, e.g., [16,17]), which decomposes the optimization into a series of simpler problems, which each can be solved more efficiently than the original problem. After obtaining the $\hat{\mathbf{W}}$ minimizing (5), an estimate of the component a_i of $\mathbf{a}$ may be formed as

$$\hat{a}_i = \left\| \hat{\mathbf{W}}_i \right\|_2 \tag{7}$$

Algorithm 1 The proposed RONNIE algorithm

initialize $k = 0$, $\mathbf{Z}_0 = \mathbf{0}^{N \times m}$ and
    $\mathbf{U}_0^{(\ell)} = \mathbf{D}_0^{(\ell)} = \mathbf{0}^{N \times m}$, $\ell = 1, 2, 3$
repeat {ADMM scheme}
    $\mathbf{Z}_k = \frac{1}{3} \sum_{j=1}^{3} \left(\mathbf{U}_k^{(j)} + \mathbf{D}_k^{(j)} \right)$
    $\mathbf{U}_{k+1}^{(1)} = \mathbf{T}\left(\boldsymbol{\zeta}_{k,1}, \frac{m}{\mu} \right)$
    $\mathbf{U}_{k+1}^{(2)} = \mathbf{S}_{\text{block}}\left(\mathbf{S}_{+}\left(\boldsymbol{\zeta}_{k,2} \right), \frac{\lambda}{\mu} \right)$
    $\mathbf{U}_{k+1}^{(3)} = \mathbf{\Gamma}\left(\boldsymbol{\zeta}_{k,3}, \epsilon \right)$
    $\mathbf{D}_{k+1} = \mathbf{D}_k - \left(\mathbf{Z}_{k+1} - \mathbf{U}_{k+1} \right)$
    $k \leftarrow k + 1$
until convergence
$\hat{\mathbf{W}} = \mathbf{Z}_k$

3. EFFICIENT ADMM IMPLEMENTATION

To efficiently implement (5), we exploit the ideas in [18] on how to extend the ADMM framework to the sum of more than two convex functions. In order to do so, let

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \mathbf{I} \end{bmatrix}^T \tag{8}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. Furthermore, define the sets $\mathbb{B}_j$ as

$$\mathbb{B}_j = \left\{ \mathbf{U}_j \in \mathbb{R}^N \mid \left\| \frac{\mathbf{U}_j}{\|\mathbf{U}_j\|} - \mathbf{\Phi}_j \right\|_2 \leq \epsilon \right\} \tag{9}$$

Then, the minimization in (5) may be rewritten as

$$\underset{\mathbf{W}}{\text{minimize}} \sum_{i=1}^{3} g_i(\mathbf{W}) \tag{10}$$

where

$$g_1(\mathbf{U}) = \frac{1}{2} \|\mathbf{y} - \mathbf{U1}\|_2^2 \tag{11}$$

$$g_2(\mathbf{U}) = \lambda \sum_{j=1}^{m} \|\mathbf{U}_j\|_2 + \mathbf{\Omega}_{\{\mathbb{R}_+^{N \times m}\}}(\mathbf{U}) \tag{12}$$

$$g_3(\mathbf{U}) = \sum_{j=1}^{m} \mathbf{\Omega}_{\{\mathbb{B}_j\}}(\mathbf{U}_j) \tag{13}$$

with

$$\mathbf{\Omega}_{\{\mathbb{B}\}}(\mathbf{U}) = \begin{cases} 0 & \text{if } \mathbf{U} \in \mathbb{B} \\ +\infty & \text{if } \mathbf{U} \notin \mathbb{B} \end{cases} \tag{14}$$

for any set $\mathbb{B}$. Let $\mathbf{Z} \in \mathbb{R}^{N \times m}$ be the primal optimization variable, and introduce the auxiliary variable $\mathbf{U}$, and dual variable $\mathbf{D}$, defined as

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}^{(1)T} & \mathbf{U}^{(2)T} & \mathbf{U}^{(3)T} \end{bmatrix}^T \tag{15}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}^{(1)T} & \mathbf{D}^{(2)T} & \mathbf{D}^{(3)T} \end{bmatrix}^T \tag{16}$$

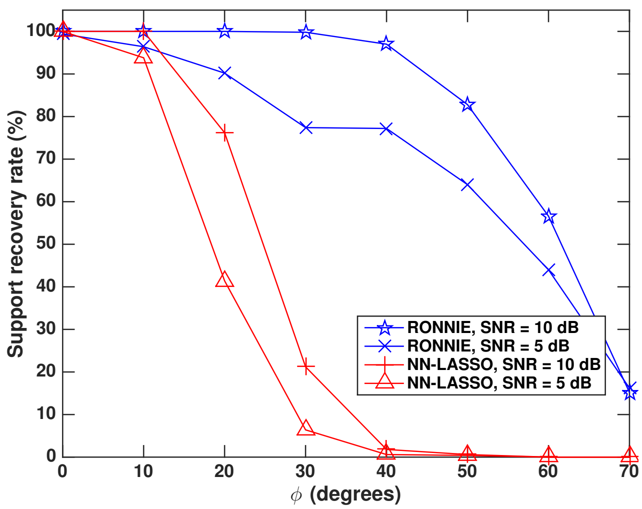

Fig. 1. Percentage of estimates where the support of the signal is recovered, for signals with three present signatures, as a function of the angle between each signature and its perturbed counterpart, ϕ.

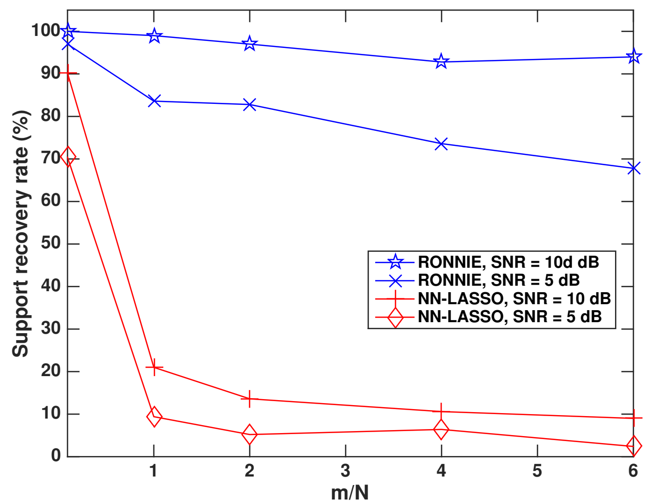

Fig. 2. Percentage of estimates where the support of the signal is recovered, for signals with three present signatures, when varying the relative sizes of the number of columns of $\mathbf{\Phi}$ (m) and the number of data points (N).

where $\mathbf{U}^{(j)}$ and $\mathbf{D}^{(j)} \in \mathbb{R}^{N\times m}$.

With these definitions, the $k+1$ updating step of the ADMM may be expressed as

$$\mathbf{Z}_{k+1} = \arg\min_{\mathbf{Z}} \left\| \mathbf{G}\mathbf{Z} - \mathbf{U}_k - \mathbf{D}_k \right\|_F^2 \tag{17}$$

where

$$\mathbf{U}_{k+1}^{(\ell)} = \arg\min_{\mathbf{U}^{(\ell)}} g_\ell\left(\mathbf{U}^{(\ell)}\right) + \frac{\mu}{2}\left\| \mathbf{U}^{(\ell)} - \boldsymbol{\zeta}_{k,\ell} \right\|_F^2 \tag{18}$$

and

$$\boldsymbol{\zeta}_{k,\ell} = \mathbf{Z}_{k+1} - \mathbf{D}_k^{(\ell)} \tag{19}$$

for $\ell = 1, 2, 3$.
Simplifying the minimization further, the explicit updating may be expressed as

$$\mathbf{Z}_{k+1} = \frac{1}{3}\sum_{j=1}^{3}(\mathbf{U}_k^{(j)} + \mathbf{D}_k^{(j)}) \tag{20}$$

$$\mathbf{U}_{k+1}^{(1)} = \mathbf{T}\left(\boldsymbol{\zeta}_{k,1}, \frac{m}{\mu}\right) \tag{21}$$

$$\mathbf{U}_{k+1}^{(2)} = \mathbf{S}_{\text{block}}\left(\mathbf{S}_+\left(\boldsymbol{\zeta}_{k,2}\right), \frac{\lambda}{\mu}\right) \tag{22}$$

$$\mathbf{U}_{k+1}^{(3)} = \mathbf{\Gamma}\left(\boldsymbol{\zeta}_{k,3}, \epsilon\right) \tag{23}$$

where μ is an inner convergence variable and $[\,\cdot\,]_j$ denotes column j of the enclosed matrix. The function $\mathbf{T}\big(\boldsymbol{\zeta}_{k,1}, \frac{m}{\mu}\big)$ in (21) returns a matrix whose jth column is

$$\frac{1}{\mu}\left(\mathbf{y} - \left(1 + \frac{m}{\mu}\right)^{-1}\left(\frac{m}{\mu}\mathbf{y} + \sum_{i=1}^{m}[\boldsymbol{\zeta}_{k,1}]_i\right)\right) + [\boldsymbol{\zeta}_{k,1}]_j \tag{24}$$

For the update in (22), the operator $\mathbf{S}_+$ operates elementwise on a matrix, setting negative elements to zero, while leaving positive ones unchanged. Similarly, the operator $\mathbf{S}_{\text{block}}$ operates columnwise, shrinking each non-zero column as

$$\mathbf{U}_j \to \frac{\|\mathbf{U}_j\|_2 - \lambda/\mu}{\|\mathbf{U}_j\|_2}\mathbf{U}_j \tag{25}$$

Finally, for the updating step in (23), one has to solve

$$\begin{aligned} &\underset{[\mathbf{U}^{(3)}]_i}{\text{minimize}} && \left\| [\mathbf{U}^{(3)}]_i - [\mathbf{Z}_{k+1}]_i + [\mathbf{D}_k^{(3)}]_i \right\|_2^2 \\ &\text{s.t} && \left\| [\mathbf{U}^{(3)}]_i \right\|_2 (2-\epsilon^2) - 2[\mathbf{U}^{(3)}]_i^T \mathbf{\Phi}_i \leq 0 \end{aligned} \tag{26}$$

If one considers the Karush-Kuhn-Tucker (KKT) conditions for the optimum of this problem (see, e.g., [19]), the solutions, as a function of the Lagrange multiplier γ, are given by

$$[\mathbf{U}^{(3)}]_i(\gamma) = \frac{\tilde{\mathbf{Z}}_i + 2\gamma\mathbf{\Phi}_i}{\left\|\tilde{\mathbf{Z}}_i + 2\gamma\mathbf{\Phi}_i\right\|_2}\left(\left\|\tilde{\mathbf{Z}}_i + 2\gamma\mathbf{\Phi}_i\right\|_2 - \gamma(2-\epsilon^2)\right) \tag{27}$$

where

$$\tilde{\mathbf{Z}}_i = [\mathbf{Z}_{k+1}]_i - [\mathbf{D}_k^{(3)}]_i \tag{28}$$

Thus, each instance of (26) may be solved efficiently by solving a 1-dimensional optimization problem in γ; in fact, the optimal γ is the positive root of a quadratic equation. The function $\mathbf{\Gamma}$ returns the matrix whose ith column is the solution $[\mathbf{U}^{(3)}]_i(\gamma^*)$, where γ^* is the corresponding solution of (26) using the parametrization (27). This yields the update (23). The estimate of $\mathbf{W}$ is formed as $\hat{\mathbf{W}} = \mathbf{Z}_k$ in the last

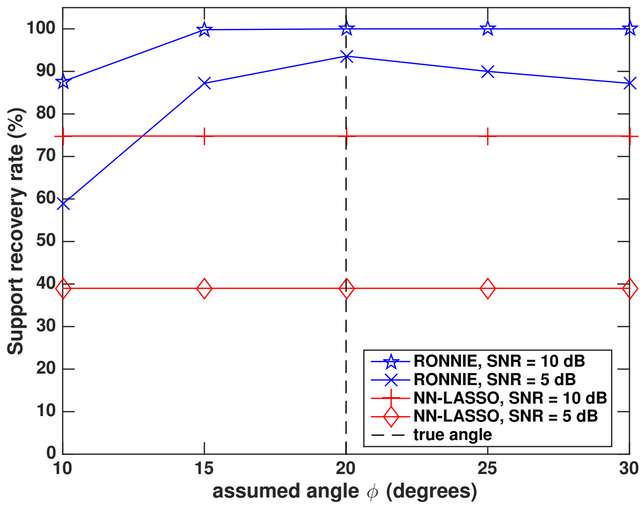

Fig. 3. Percentage of estimates where the support of the signal is recovered, for signals with three present signatures, when varying the assumed perturbation angle ϕ, with the true angle being 20 degrees.

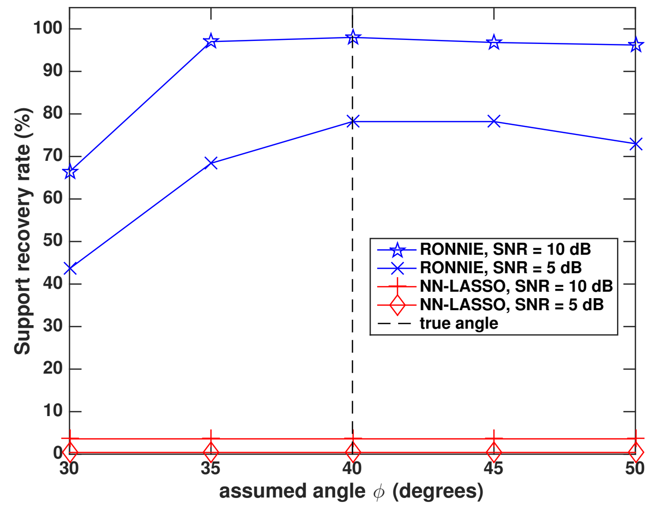

Fig. 4. Percentage of estimates where the support of the signal is recovered, for signals with three present signatures, when varying the assumed perturbation angle ϕ, with the true angle being 40 degrees.

iteration k. The full ADMM implementation is presented in Algorithm 1. We term the resulting method the RObust Non-Negative constraIned Estimator (RONNIE).

4. NUMERICAL RESULTS

We proceed to examine the performance of the proposed estimator using synthetic signals. As colinear perturbations of normalized vectors will have no detrimental effect on the performance on neither the NN-LASSO nor RONNIE, we here examine the effects of orthogonal perturbations. Let $N = 60$, $m = 90$, and choose the elements of $\mathbf{\Phi}$ as the absolute values of independent identically distributed (i.i.d.) samples from a Student's t distribution with 4 degrees of freedom. In order to examine the method's robustness to perturbations not supported by the dictionary elements, each measured perturbed signature j was formed as

$$\mathbf{X}_j = \frac{(1-\alpha)\mathbf{\Phi}_j + \alpha\mathbf{\Delta}_j}{\sqrt{(1-\alpha)^2 + \alpha^2}} \tag{29}$$

where $\mathbf{\Phi}_j$ is the jth column of $\mathbf{\Phi}$, and $\mathbf{\Delta}_j$ a perturbation orthogonal to $\mathbf{\Phi}_j$, for $\alpha \in (0,1)$, ensuring that the angle between $\mathbf{\Phi}_j$ and $\mathbf{X}_j$ is ϕ. The connection between the angle ϕ, in radians, the size ϵ of the uncertainty sphere, and α may be expressed as

$$\epsilon = \sqrt{2\left(1 - \cos\left(\phi\right)\right)} \tag{30}$$

$$\phi = \arccos\left(\frac{1-\alpha}{\sqrt{(1-\alpha)^2 + \alpha^2}}\right) \tag{31}$$

It is assured that $\mathbf{X}_j$ is a non-negative vector. The measured signal is composed of three such perturbed signatures and additive noise according to $\mathbf{y} = \sum_{j=1}^{3} \mathbf{X}_j + \mathbf{e}$, where the elements of $\mathbf{e}$ are chosen from an exponential distribution.

In the simulation study, the assumed size ϵ of the uncertainty sphere, or equivalently the angle ϕ, as well as the regularization parameter λ for RONNIE and the NN-LASSO are set using hand-tuning. Figure 1 shows the probability of correctly recovering the support of the signal for the proposed method as compared to the NN-LASSO in (4), as a function of the angle between the dictionary signature and the perturbed signature, ϕ, for varying signal-to-noise ratios (SNRs), here defined as

$$\text{SNR} = 10\log_{10}\frac{\left\|\sum_{j=1}^{3}\mathbf{X}_j\right\|_2}{\|\mathbf{e}\|_2} \tag{32}$$

The presented results are obtained using 500 Monte Carlo simulations. As can be seen from the figure, the performance of the NN-LASSO suffers greatly when ϕ increases, as large perturbations causes it to distribute the power of the three present signatures over a larger part of the dictionary. In contrast, RONNIE displays the expected robustness to the orthogonal perturbation, exhibiting almost flawless performance for SNR 10 dB and perturbation angles below and including 40 degrees. The performance of the NN-LASSO and RONNIE are both worse for SNR = 5 dB, but for RONNIE, the support recovery rate is still over 70% even for a perturbation angle of 40 degrees.

Proceeding to examine the sensitivity to the relative sizes of N and m, we fix $N = 40$ and vary m so that $m/N = [0.1, 0.5, 1, 2, 4, 6]$. This is done for $\phi = 30$ degrees, with the rest of the experiment setup being identical to the previous one. The results are shown in Figure 2. As can be seen from the figure, the performance of the NN-LASSO quickly deteriorates when the number of basis functions increase relative to the number of signal samples. This is caused by the increasing probability that one of the non-present columns of $\mathbf{X}$ will be more correlated to the signal than the orthogonally perturbed true columns $\mathbf{X}_j$, when m increases, causing the NN-LASSO to erroneously pick columns $\mathbf{X}_i$ with the largest inner products $\mathbf{X}_i^T\mathbf{y}$. It should further be noted that the performance of RONNIE does not seem to suffer significantly from increasing the ratio m/N; the support recovery rate is still over 90% when we have 6 times more columns in $\mathbf{X}$ than signal samples for SNR 10 dB. For SNR 5 dB, the support recovery rate is lower, but it is still almost 70% for RONNIE when we have 6 times more columns in $\mathbf{X}$ than signal samples.

Next, we evaluate the sensitivity of RONNIE to different choices of the assumed perturbation angle, ϕ. Here, we set $N = 60$, $m = 90$, and consider SNRs 10 and 5 dB. The angle ϕ of the perturbation is then fixed, and we apply RONNIE using different assumed values of ϕ. The other aspects of the signal construction is identical to the earlier experiments. The results are shown in Figures 3 and 4, for true angles $\phi = 20$ and 40 degrees, respectively. The support recovery rate for the NN-LASSO is also included as reference, although its performance of course does not depend on the assumed angle ϕ. As can be seen in the figures, RONNIE displays considerable robustness to the choice of ϕ, yielding support recovery rates higher than the NN-LASSO for all considered angles. As can be seen from the figures, the support recovery rate of RONNIE is only lowered considerably when the assumed angle ϕ is set well below the true angle, though the method does not suffer from setting the assumed angle significantly higher than the true angle.

5. REFERENCES

[1] M.D. Iordache, J.M. Biourcas-Dias, and A. Plaza, "Sparse Unmixing of Hyperspectral Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, June 2011.

[2] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral Unmixing via L1-2 Sparsity-Constrained Nonnegative Matrix Factorization," *IEEE Transactions on Geoscience a*, vol. 49, no. 6, pp. 4282–4297, June 2011.

[3] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 275–285, June 2008.

[4] H. Kim and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for micromicro data analysis," *Bioinformatics*, vol. 23, no. 12, pp. 1495–1502, May 2007.

[5] Y. Lin and D.D. Lee, "Bayesian Regularization and Nonnegative Deconvolution for Room Impulse Response Estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 839–847, March 2006.

[6] A. Cont, S. Dubnov, and D. Wessel, "Realtime Multiple-Pitch and Multiple-Instrument Recognition for Music Signals Using Sparse Non-Negative Constraints," in *10th International Conference on Digital Audio Effects*, France, September 2007, vol. 57.

[7] D. Wu, M. Yaghoobi, M. Davies, and R. Clewes, "A Sparse Regularized Model for Raman Spectral Analysis," *Signal Processing for Defence*, 2014.

[8] A.M. Bruckstein, M. Elad, and M. Zibulevsky, "Sparse Non-Negative Solution of a Linear System of Equations is Unique," in *ISCCSP*, Malta, 12-14 March 2008.

[9] J. Chen, C. Richard, J-C.M. Bermudez, and P. Honeine, "Variants of Non-Negative Least-Mean-Square Algorithm and Convergence Analysis," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3990–4005, August 2014.

[10] M. Yaghoobi, D. Wu, and M.E. Davies, "Fast Non-Negative Orthogonal Matching Pursuit," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1229–1233, September 2015.

[11] S. D. Somasundaram, A. Jakobsson, and E. Gudmundson, "Robust Nuclear Quadrupole Resonance Signal Detection Allowing for Amplitude Uncertainties," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 887–894, March 2008.

[12] M. Rubsamen and M. Pesavento, "Maximally Robust Capon Beamformer," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2030–2041, 2013.

[13] N. R. Butt, S. I. Adalbjörnsson, S. D. Somasundaram, and A. Jakobsson, "Robust Fundamental Frequency Estimation in the Presence of Inharmonicities," in *38th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, May 26–31, 2013.

[14] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[15] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008, http://stanford.edu/∼boyd/graph_dcp.html.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[17] N. Parikh and S. Boyd, "Proximal Algorithms," *Found. Trends Optim.*, vol. 1, pp. 127–239, 2014.

[18] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Algorithms for imaging inverse problems under sparsity regularization," in *Proc. 3rd Int. Workshop on Cognitive Information Processing*, May 2012, pp. 1–6.

[19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.