HIGH RESOLUTION SPARSE ESTIMATION OF EXPONENTIALLY DECAYING SIGNALS

Johan Swärd, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson

Dept. of Mathematical Statistics, Lund University, Sweden

ABSTRACT

We consider the problem of sparse modeling of a signal consisting of an unknown number of exponentially decaying sinusoids. Since such signals are not sparse in an oversampled Fourier matrix, earlier approaches typically exploit large dictionary matrices that include not only a finely spaced frequency grid but also a grid over the considered damping factors. The resulting dictionary is often very large, resulting in a computationally cumbersome optimization problem. Here, we instead introduce a novel dictionary learning approach that iteratively refines the estimate of the candidate damping factor for each sinusoid, thus allowing for both a quite small dictionary and for arbitrary damping factors, not being restricted to a grid. The performance of the proposed method is illustrated using simulated data, clearly showing the improved performance as compared to previous techniques.

Index Terms— Sparse signal modeling, Spectral analysis, Sparse reconstruction, Parameter estimation.

1. INTRODUCTION

Exponentially decaying sinusoidal signals occur naturally in a wide range of fields, such as, for instance, radio frequency spectroscopy, wireless communications, sonar and radar (see, e.g., [1, 2] and the references therein). Commonly, the measurements suffer from various kinds of interference signals or corrupting additive colored noise. Furthermore, the number of modes present in the signal are generally unknown or may vary over time, typically necessitating some form of model order selection decision. Given such difficulties, it is often of interest to formulate non-parametric or semi-parametric modeling techniques, imposing only mild assumptions of the *a priori* knowledge of the signal structure. Popular solutions include the so-called dCapon, dAPES, and dIAA spectral estimators which form a generalized spectral estimate of the signal, constructing their spectral representation over both the frequency and damping dimensions [1, 2] (see also [3, 4]). Although this form of techniques are robust to the model assumptions, they suffer difficulties in separating closely spaced frequency and damping modes from each other, and typically require notable computational efforts if not implemented carefully [4]. As an alternative, one may use sparse modeling of the signal, forming a large dictionary of all potential frequencies and damping candidates, thus generally having vastly more columns than rows. For a given signal and the resulting dictionary matrix, one thus wishes to find the sparsest solution to the resulting linear set of equations mapping the signal to a linear combination of a few of the columns of the dictionary. Such techniques have successfully been applied to line spectral data, and the topic has attracted notable attention in the recent literature (see, e.g., [5–11]). Although these algorithms appear quite different from each other, they share the property that the considered dictionary grid should be selected sufficiently fine to allow for a sparse signal representation (see also [12, 13]), which, if extended to also consider damped modes, necessitates a large dictionary matrix containing elements with a sufficiently fine grid over the range of both the potential frequencies and damping candidates (see, e.g., [2, 14, 15]); this will be particularly noticeable if treating large data sets, or data sets with multiple measurement dimensions, such as in NMR measurements. In order to mitigate this problem, we here propose a novel dictionary learning approach wherein we iteratively decompose the signal with a fixed small dictionary, adaptively learning the dictionary elements best suited to enhance sparsity. To this effect, we initially form a coarsely spaced dictionary with undamped modes over the range of considered frequency candidates, iteratively adapting both the frequency and damping settings for the dictionary elements, thereby also allowing for both a reduction and an expansion of the number of dictionary elements considered in the optimization.

2. SPARSE DICTIONARY LEARNING

Let $y(n)$ denote a signal of interest that may be well modelled as a sum of K exponentially decaying sinusoids, i.e.,

$$y(n) = \sum_{k=1}^{K} \alpha_k e^{-2\pi j f_k n - \beta_k n} + e(n) \qquad (1)$$

for $n = 1, \ldots, N$, where f_k, β_k, and α_k are the frequency, damping factor, and complex amplitude of the k:th component, respectively, whereas $e(n)$ denotes an additive circularly

This work was supported in part by the Swedish Research Council, the Swedish Foundation for Cooperation in Research and Higher Education (STINT), the European Research Council (ERC, Grant agreement n. 261670) and the Crafoord's and Carl Trygger's foundations.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

symmetric Gaussian noise. Here, both the number of modes, K, and the parameters detailing each such mode, i.e., f_k, β_k, and α_k, for $k = 1, \ldots, K$, are considered unknown, as is the variance of the additive noise, σ_e^2. Stacking the observations of the signal into a vector

$$\mathbf{y} = \begin{bmatrix} y(1) & \ldots & y(N) \end{bmatrix}^T \tag{2}$$

where $(\cdot)^T$ denotes the transpose, allows the signal to be expressed using a sparse signal representation, such that it may be formulated as a linear combination of only a few out of a range of potential candidate modes, i.e.,

$$\mathbf{y} = \sum_{k=1}^{P} \alpha_k \mathbf{w}_{f_k}(\beta_k) + \mathbf{e} = \mathbf{W}(\boldsymbol{\beta})\mathbf{a} + \mathbf{e} \tag{3}$$

where the noise vector $\mathbf{e}$ is constructed similarly to $\mathbf{y}$, and

$$\mathbf{W}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{w}_1(\beta_1) & \ldots & \mathbf{w}_P(\beta_P) \end{bmatrix} \tag{4}$$

$$\mathbf{w}_{f_k}(\beta_k) = \begin{bmatrix} e^{j2\pi f_k - \beta_k} & \ldots & e^{(j2\pi f_k - \beta_k)N} \end{bmatrix}^T \tag{5}$$

$$\mathbf{a} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_P \end{bmatrix}^T \tag{6}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \ldots & \beta_P \end{bmatrix}^T \tag{7}$$

with P denoting the total number of dictionary elements, typically formed such that $\{f_k\}_{k=1}^P$ are chosen uniformly over the frequency range of interest, e.g., $[0, 1]$, for a given set of damping factors, $\{\beta_\ell\}$, for $\ell = 1, \ldots, P$. To allow for high-resolution estimates without imposing any prior knowledge of the frequency locality of the modes, we here assumed that $P \gg N$, although to reduce the computational complexity in our implementation, we actually form the dictionary using only about N elements, which combined with an iterative zooming, as detailed below, is more than sufficient to yield the desired resolution. As the damping factors are unknown, the dictionary elements need to be initiated to some reasonable estimate, for instance, setting all damping constants equal to zero. The unknown amplitudes and damping constants may then be found as the solution to the optimization problem

$$\min_{\mathbf{a}, \boldsymbol{\beta}} ||\mathbf{y} - \mathbf{W}(\boldsymbol{\beta})\mathbf{a}||_2^2 + \boldsymbol{\rho}(\mathbf{a}) \tag{8}$$

where $\boldsymbol{\rho}(\mathbf{a})$ is a sparsity promoting penalty function, on $\mathbf{a}$, often selected as the ℓ_1 norm [6, 16], leading to a convex optimization problem, for a fixed $\boldsymbol{\beta}$, which will thus reliably converge to a global minimum without the need of a good initialization (see, e.g., [17]). We note that in many cases, even further sparsity is desirable, which may then be obtained by using other heuristics for approximating the function that counts the number of non-zero elements, e.g., by using a concave penalty function [18], so that larger non-zero elements are penalized (relatively) less than small non-zero elements, or by iteratively forming improved ℓ_0 approximations [19], although it should be stressed that such approaches lead to more difficult, non-convex, optimization problems, commonly suffering from numerous local minima. Without loss of generality, we are in this work limiting our attention to using only an ℓ_1 penalty, as the found solutions are deemed to be sufficiently sparse, but the algorithm could easily be modified to using other penalties as well. After forming the initial estimates of $\mathbf{a}$ and $\boldsymbol{\beta}$ using (8), we proceed to the dictionary learning step wherein both the frequency and damping for each candidate mode in the dictionary are updated appropriately, and where, if desired, the number of dictionary elements may be reduced or expanded as deemed appropriate. The Fourier transform of an exponentially damped sinusoid reveals that the power spectral density decays as

$$\frac{2\beta}{(2\pi\Delta f)^2 + \beta^2} \tag{9}$$

where Δf is the distance from the frequency from frequency of the actual spectral peak and β the damping factor. The initial estimates will thus not be sparse, although the found spectral peak will still be at the correct frequency. Similarly, for a coarse initial grid spacing, closely spaced components or off-grid frequencies will still result in peaks in neighboring grid-points. As a result, one would obtain a sparser solution if one estimated a suitable damping factor for each of the peaks found in the initial estimate, as well as refine the used frequency grid for the peak location to allow for off-grid frequencies fitting the data better. This is done by estimating the frequency and damping factor for each of the modes found in (8), restricting the frequencies to move only within the range of the neighboring frequency grid points, such that a dictionary element may not move away further in frequency than a few grid points, updating the dictionary element accordingly. The dictionary is then updated using these refined dictionary elements, whereafter the entire process is iterated. To clarify, in order to form the refined dictionary elements, consider the residual from the solution of (8), i.e.,

$$\boldsymbol{\varepsilon} = y - W(\boldsymbol{\beta})\hat{\mathbf{a}} \tag{10}$$

Algorithm 1 Seperating frequency and damping algorithm

Initialize $\boldsymbol{\beta}^0 = \mathbf{0}$, $f_k^0 = k/N$ for $k = 1, \ldots, N$
for $m = 1, \ldots$ **do**
  $\mathbf{a}^m = \underset{\mathbf{a}}{\operatorname{argmin}} ||y - W(\boldsymbol{\beta}^{m-1})\mathbf{a}||_2^2 + \boldsymbol{\rho}(\mathbf{a})$
  $\boldsymbol{\varepsilon}^m = y - W(\boldsymbol{\beta}^{m-1})\mathbf{a}^m$
  $P^m =$number of peaks of $\mathbf{a}^m$
  for $j = 1, \ldots P^m$ **do**
    $I_j =$ index of peak j
    $\Gamma_j =$ indexes of peak j and the adjacent elements.
    $\left\{\hat{\beta}_j\,, \hat{f}_j\right\} = \arg\max_{f,\beta} \left\{\boldsymbol{\varepsilon}_\Gamma^*(\boldsymbol{\beta})\, \Pi_{\mathbf{w}}\, \boldsymbol{\varepsilon}_\Gamma(\boldsymbol{\beta})\right\}$
  end for
  Update W with $\boldsymbol{\beta}^m$ and a finer grid around $\mathbf{f}^m$.
end for

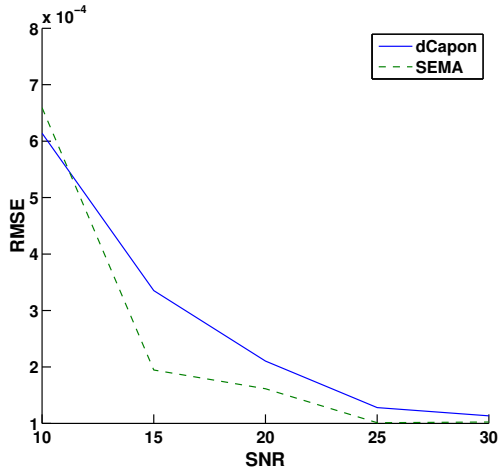

Fig. 1. The RMSE of the frequency estimation as a function of SNR. The signal contains three modes.

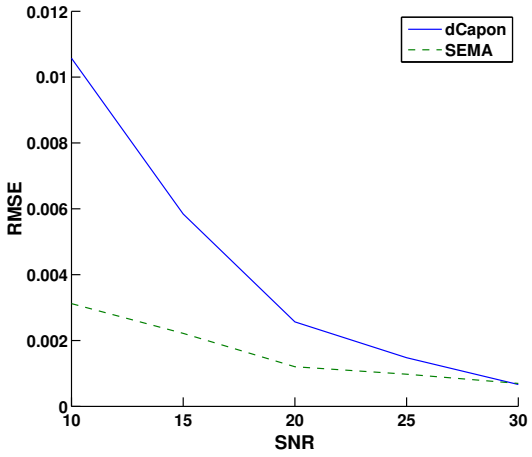

Fig. 2. The RMSE of the damping estimation as a function of SNR. The signal contains three modes.

where $\hat{\mathbf{a}}$ denotes the vector minimizing (8) with $\boldsymbol{\beta} = \mathbf{0}$. Let I_j denote the index of the jth peak in $\hat{\mathbf{a}}$, and form Γ_j from the adjacent indices on each side of I_j, i.e.,

$$\Gamma_j = \begin{bmatrix} I_{j-1} & I_j & I_{j+1} \end{bmatrix}^T \tag{11}$$

The estimates of the corresponding dictionary elements can thus be refined using the minimization

$$\mathcal{C} = \min_{f,\beta,a} \left\| \boldsymbol{\varepsilon} + W_\Gamma(\boldsymbol{\beta}_\Gamma)\hat{\mathbf{a}}_\Gamma - \mathbf{w}_f(\beta)a \right\|_2 \tag{12}$$

$$= \min_{f,\beta,a} \left\| \boldsymbol{\varepsilon}_\Gamma(\boldsymbol{\beta}) - \mathbf{w}_f(\beta)a \right\|_2 \tag{13}$$

with

$$\boldsymbol{\varepsilon}_\Gamma(\boldsymbol{\beta}) \triangleq \boldsymbol{\varepsilon} + W_\Gamma(\boldsymbol{\beta}_\Gamma)\hat{\mathbf{a}}_\Gamma \tag{14}$$

where by subscribing a set of indexes the corresponding elements or columns of a vector or matrix are inferred, respectively. Solving for the unknown amplitude using least squares, for a fixed frequency and damping constant, yields

$$\hat{a} = \frac{\mathbf{w}_f^*(\beta)\boldsymbol{\varepsilon}_\Gamma}{\mathbf{w}_f^*(\beta)\mathbf{w}_f(\beta)} \tag{15}$$

with $(\cdot)^*$ denoting the conjugate transpose, which, if inserted in (13), implies

$$\left\{\hat{\beta}_j\,,\,\hat{f}_j\right\} = \arg\max_{f,\beta} \left\{ \boldsymbol{\varepsilon}_\Gamma^*(\boldsymbol{\beta})\,\Pi_{\mathbf{w}}\,\boldsymbol{\varepsilon}_\Gamma(\boldsymbol{\beta}) \right\} \tag{16}$$

where

$$\Pi_{\mathbf{w}} = \frac{\mathbf{w}_f(\beta)\mathbf{w}_f^*(\beta)}{\mathbf{w}_f^*(\beta)\mathbf{w}_f(\beta)} = \psi_{\mathbf{w}}\mathbf{w}_f(\beta)\mathbf{w}_f^*(\beta) \tag{17}$$

with

$$\psi_{\mathbf{w}}^{-1} = \sum_{k=1}^{N} e^{-2\beta} = \frac{1-e^{-2\beta N}}{1-e^{-2\beta}} \tag{18}$$

It is worth noting that (16) may be solved computationally efficiently using, for instance, a golden section search (see, e.g., [20]). Using the found $\hat{\beta}_j$ and $\hat{f}_j$ estimates, for each peak in $\hat{\mathbf{a}}$, we form new refined dictionary elements using these estimates, as well as create a finer frequency grid around the so-obtained new dictionary element by also including new dictionary elements with frequencies close to the refined peak frequency, while simultaneously omitting dictionary elements with zero amplitude response. This creates an effective zooming procedure over frequency, which when iterated allows for high-resolution parameter estimates, without the need of *a priori* knowledge of the model orders, or a large dictionary matrix causing a computationally cumbersome solution. The resulting algorithm, here termed the Sparse Exponential Mode Analysis (SEMA) algorithm, is summarized in Algorithm 1. The computational complexity of the algorithm is dominated by solving (8), the cost of which scales comparably to calculating the least norm solution to an underdetermined system of equations, i.e., requiring about $\mathcal{O}(NP\min(N,P))$ operations [21], which should be compared to the approximately $\mathcal{O}(NPN_\beta\min(N,PN_\beta))$ operations required for algorithms that consider N_β grid points for the damping parameter.

3. NUMERICAL EXAMPLES

We proceed to examine the performance of the proposed method using simulated data. We initially consider $N = 128$ samples of a signal containing three modes, where the frequency and damping constants are chosen uniformly over

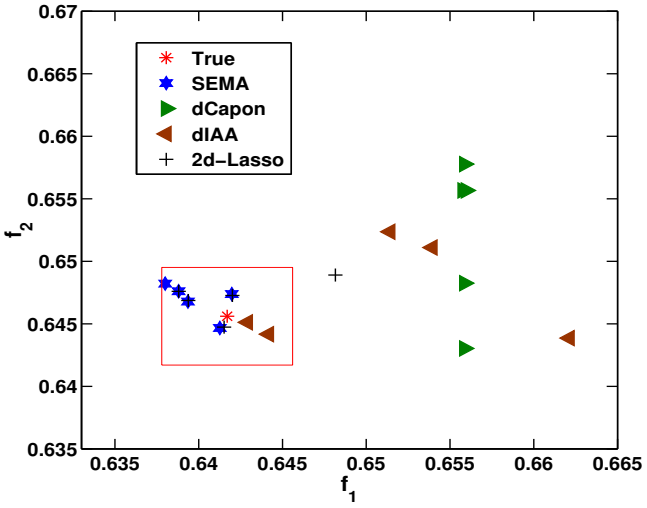

Fig. 3. The result of resolving two closely spaced spectral peaks. The red square indicates the distance $1/(2N)$ from the true frequencies.

$[0,\ 0.025]$ and $[0,\ 1]$, respectively. As we will examine the issue of resolution separately, we have here initially ensured that no modes are closer in frequency than $1/N$. Figures 1 and 2 show the resulting performance of the proposed SEMA algorithm as compared to the dCapon estimate [1], as a function of the signal-to-noise-ratio (SNR), defined as $\log_{10}(||\mathbf{y}||_2^2/N\sigma^2)$, where σ^2 denotes the variance of the noise. The figures show the root mean squared error (RMSE) of the frequency and damping estimates, here defined as

$$\text{RMSE} = \sqrt{\frac{1}{MK}\sum_{m=1}^{M}\sum_{k=1}^{K}\left(\theta_{m,k} - \hat{\theta}_{m,k}\right)^2} \qquad (19)$$

where θ_m, k denotes the estimate of either the frequency or the damping of mode k and Monte-Carlo simulation m, M is the total number of Monte-Carlo simulations, and K is the number of modes. These results have been obtained using 175 Monte-Carlo simulations. In this example, the size of the frequency and damping grid for dCapon was selected to be 6000×6000, uniformly covering frequencies and damping factors in $[0, 1]$ and $[0, 0.025]$, respectively, and using the recommended filter length of $N/4$. The SEMA algorithm on the other hand use a dictionary containing only 128 elements, with 40 elements being inserted in the dictionary refinement step in the iteration. As is clear from the figures, the proposed SEMA algorithm yields notably better estimates, without requiring a large dictionary grid over both dimensions, thereby allow for a substantially faster implementation. It is also worth noting that the dCapon estimation errors are here larger than the smallest possible error that is attainable given the current grid size, implying that the grid size does not in itself limit the quality of the estimates.

Next, we examine the ability of the methods to resolve two closely spaced spectral lines. In this case, we consider a signal containing two sinusoidal components with frequencies, $f_1 = 0.6417$ and $f_2 = 0.6456$, with random damping constants being drawn uniformly from $[0, 0.025]$. Figure 3 illustrates the resulting frequency estimates as obtained from 5 Monte-Carlo simulations, using SNR = 20 dB, and SNR $= 20$. The figure shows the estimates obtained using SEMA, dCapon, dIAA, and for a Lasso method with a dictionary containing both frequencies and damping factors and exploiting a zooming similar to the one used in SEMA. Here, the frequency grid for dCapon and dIAA has been selected to only be formed on $[0.63, 0.67]$, allowing the methods notable *a priori* information on the frequency region of interest. The damping grid ranges over $[0, 0.025]$ and has size 500 for all methods, except for the used Lasso method, where, due to complexity reasons, it is set to 10. As seen in the figure, both the proposed method and the Lasso method clearly manage to resolve the two peaks, whereas dCapon and dIAA, in most cases, are unable to find the correct peaks. In the figure, the red square indicates the region $1/(2N)$ around the true frequencies. Throughout these simulations, we have used $\boldsymbol{\rho}(\mathbf{a}) = \lambda||\mathbf{a}||_1$, there λ has been selected such as being the average of the dominant amplitudes and that of the remaining spectrum, where the former is found as the mean of the amplitude of the k_0 largest peaks of the periodgram, whereas the latter is computed as the mean of the remaining periodogram estimate. Here, k_0 denotes the number of considered peaks, being equivalent to the model order given to the competitive algorithms. It should be stressed that selecting λ in this way does not imply using k_0 as the assumed model order, only acting as a guideline for how this weighting might be selected.

4. REFERENCES

[1] P. Stoica and T. Sundin, "Nonparametric NMR Spectroscopy," *J. Magn. Reson.*, vol. 152, pp. 57–69, 2001.

[2] E. Gudmundson, P. Stoica, J. Li, A. Jakobsson, M. D. Rowe, J. A. S. Smith, and J. Ling, "Spectral Estimation of Irregularly Sampled Exponentially Decaying Signals with Applications to RF Spectroscopy," *J. Magn. Reson.*, vol. 203, no. 1, pp. 167–176, March 2010.

[3] F. J. Frigo, J. A. Heinen, J. A. Hopkins, T. Niendorf, and B. J. Mock, "Using Peak-Enhanced 2D-Capon Analysis with Single Voxel Proton Magnetic Resonance Spectroscopy to Estimate T2* for Metabolites," in *Proc. of ISMRM*, 2004, vol. 12, p. 2437.

[4] G. O. Glentis and A. Jakobsson, "Computationally efficient damped Capon and APES spectral estimation," in *21st European Signal Processing Conference*, Marrakech, Morocco, Sept. 9-13 2013.

[5] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Reweighted Minimum Norm Algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, March 1997.

[6] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul,, jul 2008, pp. 10225–10229.

[7] P. Stoica, Jian Li, and Hao He, "Spectral Analysis of Nonuniformly Sampled Data: A New Approach Versus the Periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858, March 2009.

[8] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source Localization and Sensing: A Nonparametric Iterative Approach Based on Weighted Least Squares," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 425–443, January 2010.

[9] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, March 2011.

[10] P. Stoica, P. Babu, and J. Li, "SPICE : a novel covariance-based sparse estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629 –638, Feb. 2011.

[11] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 15801590, July 2012.

[12] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182 – 2195, May 2011.

[13] P. Stoica and P. Babu, "Sparse Estimation of Spectral Lines: Grid Selection Problems and Their Solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.

[14] S. I. Adalbjörnsson and A. Jakobsson, "Sparse Estimation of Spectroscopic Signals," in *19th European Signal Processing Conference, EUSIPCO 2011*, Barcelona, Spain, 2011.

[15] S. Sahnoun, E. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Applied SP*, vol. 60, pp. –, 2012.

[16] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

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

[18] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[19] M. M. Hyder and K. Mahata, "An Improved Smoothed ℓ^0 Approximation Algorithm for Sparse Representation," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2194–2205, April 2010.

[20] L. C. Böiers, *Mathematical methods of optimization*, Studentlitteratur AB, 2010.

[21] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, April 2004.