CROSS-PRODUCTS LASSO

David Luengo , Javier Vía , Sandra Monzón , Tom Trigano and Antonio Artés-Rodríguez

Dep. of Circuits and Systems Engineering, Universidad Politécnica de Madrid, 28031 Madrid (Spain).
Dep. of Communications Engineering, Universidad de Cantabria, 39005 Santander (Spain).
Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, 28911 Leganés (Spain).
Dep. of Electrical Engineering, Shamoon College of Engineering, Ashdod (Israel).

ABSTRACT

Negative co-occurrence is a common phenomenon in many signal processing applications. In some cases the signals involved are sparse, and this information can be exploited to recover them. In this paper, we present a sparse learning approach that explicitly takes into account negative co-occurrence. This is achieved by adding a novel penalty term to the LASSO cost function based on the cross-products between the reconstruction coefficients. Although the resulting optimization problem is non-convex, we develop a new and efficient method for solving it based on successive convex approximations. Results on synthetic data, for both complete and overcomplete dictionaries, are provided to validate the proposed approach.

Index Terms— negative co-occurrence, sparsity-aware learning, LASSO, sparse coding

1. INTRODUCTION

Co-ocurrence has been extensively exploited during the last forty years in areas such as computer vision at both low [1] and high [2] processing levels, but it has received much less attention by the signal processing community. However, co-occurrence is relatively common in signal processing applications, especially negative co-occurrence. Focusing on the unidimensional case, a typical example can be found in the biomedical signal processing of intracardiac electrograms: after a cardiac cell activation there exists a so called "refractory period" where the cell cannot be excited [3]. A second example can be found in the analysis of spectrometric data stemming from Type I counters, where a detector (e.g., Geiger) which records incoming particles has a specific associated dead time during which the system is not able to record another particle interaction [4].

Both of the above mentioned applications share another common feature: the signals involved admit a sparse representation using overcomplete dictionaries, i.e., they can be analyzed using sparse coding techniques [5], [6]. In sparse coding or compressed sensing techniques, the use of additional information to obtain a sparsity pattern in accordance with some physical/biological knowledge has already been investigated in [7]. Two additional examples are the fused LASSO [8], which imposes both sparsity and flatness of the obtained coefficients profile (making it valuable to describe mass spectroscopy data), and the elastic net, which favors sparsity obtained by correlated variables [9]. Recent contributions have also investigated, both theoretically and practically, the use of algorithms which encourage sparsity by clusters of coefficients [10, 11]. However, to the best of our knowledge, encouraging sparsity by taking into account negative co-occurrence has not been considered yet.

In this paper we introduce a novel sparse learning algorithm that explicitly enforces negative co-occurrence by incorporating a new penalty term, based on the cross-products of the reconstruction coefficients, to the LASSO cost function. Hence, we call our approach cross-products LASSO (CP-LASSO). Unfortunately, this leads to a non-convex optimization problem, but we show that it can be solved efficiently in an approximate way using successive convex approximations (SCA). Results on synthetic data, both for complete and overcomplete dictionaries, are provided to validate the proposed approach.

The paper is organized as follows. Section 2 shows the problem formulation. The novel regularization function and its efficient minimization using SCA are described in detail in Section 3. Section 4 shows numerical results on several synthetic data sets. Finally, Section 5 concludes the paper.

2. PROBLEM FORMULATION

2.1. Signal Model

In this paper we focus on discrete-time signals generated by an unknown latent sparse activation signal going through an

LTI system, and contaminated by noise,[1]

$$y[n] = p[n] * h[n] + w[n], \tag{1}$$

where $w[n]$ is the discrete-time noise process, $h[n]$ is the channel's impulse response (we assume a causal FIR channel of length L), $*$ denotes the standard linear convolution operator, and $p[n]$ is the sparse latent signal (also known as spike or activations train),

$$p[n] = \sum_{k=0}^{K-1} A_k \delta[n - N_k], \tag{2}$$

where $\delta[n]$ denotes Kronecker's delta, K is the total number of spikes in $p[n]$, A_k is the amplitude of the k-th spike and N_k its arrival time. We also assume that after each activation there is a *negative co-occurrence period* during which new activations cannot occur, i.e., $N_k - N_{k-1} > N_{\min}$. Substituting (2) into (1), the signal model finally becomes

$$y[n] = \sum_{k=0}^{K-1} A_k h[n - N_k] + w[n] \tag{3}$$

This discrete-time model can be formulated more compactly in matrix form as

$$\mathbf{y} = \mathbf{H}\mathbf{a} + \mathbf{w}, \tag{4}$$

where $\mathbf{w} = [w[0], \ldots, w[N-1]]^\top$ is the $N \times 1$ noise vector; $\mathbf{a} = [a[0], \ldots, a[N-1]]^\top$ is the $N \times 1$ sparse amplitudes vector ($a[n] \neq 0 \Leftrightarrow n = N_k$ for some k); and $\mathbf{H}$ is the $N \times N$ channel matrix.

2.2. Reconstruction Model

If the channel's impulse response is known, then we can use (4) to recover the activations from the observations by imposing some of the sparse restrictions described in Section 3. Unfortunately, $h[n]$ is often unknown, and estimating it directly from the observations can be problematic for low signal to noise ratios (SNRs). Besides, in many applications we are interested in the arrival times of the activations rather than in $h[n]$. In these cases, a common alternative is performing sparse learning on the following reconstruction model

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{\Phi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, \ldots, \mathbf{\Phi}_M]$ is the $N \times MN$ dictionary matrix, with $M \geq 1$ indicating the number of basis signals in the dictionary and $\mathbf{\Phi}_m$ being the m-th $N \times N$ dictionary matrix ($1 \leq m \leq M$), constructed from the m-th discrete-time basis waveform, $\phi_m[n] \neq 0 \Leftrightarrow 0 \leq n \leq N_m - 1$; $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_M^\top]^\top$ is the $MN \times 1$ sparse coefficients vector,

[1]This formulation can be easily extended to Q activation signals [12, 13]. However, here we focus on a single activation for the sake of simplicity.

with $\boldsymbol{\beta}_m = [\beta_m[0], \ldots, \beta_m[N-1]]^\top$ for $1 \leq m \leq M$; and $\boldsymbol{\varepsilon} = [\varepsilon[0], \ldots, \varepsilon[N-1]]^\top$ is the $N \times 1$ excess noise vector. On the one hand, when $M = 1$ we have a complete dictionary, and (5) has the same structure as (4), although the basis functions can be different. On the other hand, $\mathbf{\Phi}$ is an overcomplete dictionary when $M > 1$.

3. RESTRICTED SPARSE LEARNING

3.1. Prior Work: LASSO plus post-processing

A first approach for minimizing the reconstruction error of the model, subject to a sparsity contraint and respecting the negative co-ocurrence period, was recently proposed in [12]. First of all, an initial estimate of $\boldsymbol{\beta}$ is obtained, using LASSO [14],

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{MN}} \left\{ \|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{6}$$

where $\|\mathbf{x}\|_2$ denotes the L_2 norm of $\mathbf{x}$, $\|\boldsymbol{\beta}\|_1$ is the L_1 norm of $\boldsymbol{\beta}$, and λ is the regularization parameter. Unfortunately, the reconstruction obtained using (6) is unlikely to respect the restriction between activation times imposed by the negative co-occurrence period, especially for unknown channels and low SNRs. Hence, after the computation of $\hat{\boldsymbol{\beta}}$, [12] introduces a second step that estimates the samples associated to the arrival times of the spikes recursively as follows:

$$\hat{N}_k = \arg\max_{1 \leq n \leq N} \left\{ \|\hat{\boldsymbol{\beta}}[n]\|_1 \mathbb{I}(\eta < \|\hat{\boldsymbol{\beta}}[n]\|_1 < \|\hat{\boldsymbol{\beta}}[\hat{N}_{k-1}]\|_1) \right\}$$

$$\text{s.t.} \quad |\hat{N}_k - \hat{N}_\ell| > N_{\min}, \quad \text{for} \quad 1 \leq \ell \leq k-1, \tag{7}$$

where $\hat{\boldsymbol{\beta}}[n] = [\hat{\beta}_1[n], \ldots, \hat{\beta}_M[n]]^\top$, $\mathbb{I}(\cdot)$ is an indicator function, i.e., a function that takes a value equal to one if the logical condition is fulfilled, and zero otherwise; and η is a user-defined threshold, used to discard the $\hat{\boldsymbol{\beta}}[n]$ with a small L_1 norm, which provide no information about spike localization.

3.2. Cross-Products LASSO (CP-LASSO)

A novel approach, based on introducing the restriction imposed by the negative co-occurrence period into the cost function, is proposed here. This can be done by incorporating an additional penalty term to the LASSO cost function which discourages the presence of non-null coefficients associated to nearby basis functions. The new cost function proposed is

$$\begin{aligned} J_{\text{one-step}} = & \|\mathbf{y} - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \\ & + \rho \sum_{n=0}^{N-1} \sum_{m=1}^{M} \sum_{\substack{k=-N_{\min} \\ k \neq 0}}^{N_{\min}} \|\beta_m[n]\boldsymbol{\beta}[n+k]\|_0, \end{aligned} \tag{8}$$

where $\|\mathbf{x}\|_0$ denotes the L_0 "norm" of $\mathbf{x}$, ρ is an additional regularization parameter, and $\boldsymbol{\beta}[n + k] = [\beta_1[n + k], \ldots, \beta_M[n + k]]^\top$ is an $M \times 1$ vector containing all the coefficients associated to the $(n + k)$-th sample.

Since the L_0 "norm" is generally intractable, the usual approach taken is substituting it by the more tractable L_1 norm, which provides an equivalent solution under certain conditions. Performing this standard relaxation, and after some algebra [13], the modified cost function given by (8) turns into

$$J_{\text{CP-LASSO}} = \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 + \rho\|(\mathbf{B}^\top \otimes \mathbf{I}_M)\boldsymbol{\beta}\|_1, \quad (9)$$

where $\otimes$ denotes the Kronecker product of two matrices [15]; $\mathbf{I}_M$ is the $M \times M$ identity matrix; and $\mathbf{B}^\top$ is a $2MN_{\min} \times N$ matrix whose n-th column, following the notation of [16], is given by $\mathbf{b}_n = \mathbf{B}^\top(:,n) = [\boldsymbol{\beta}^\top[n - N_{\min}], \ldots, \boldsymbol{\beta}^\top[n - 1], \boldsymbol{\beta}^\top[n+1], \ldots, \boldsymbol{\beta}^\top[n - N_{\min}]]^\top$, with $\boldsymbol{\beta} = 0$ for $k < 0$. Since the novel cost function incorporates the cross-products between reconstruction coefficients to the LASSO, we call the novel approach cross-products LASSO (CP-LASSO).

3.3. Successive Convex Approximations (SCA)

Unfortunately, the new penalty term introduced in (9) leads to a complicated non-convex optimization problem. Therefore, in this subsection we present an algorithm, based on successive convex approximations (SCA) [17, 18, 19], for finding a (local) solution of the constrained version of the Cross-Products LASSO. In particular, the problem to be solved can be formulated as

$$\begin{aligned}
&\underset{\boldsymbol{\beta}, \mathbf{c}}{\text{minimize}} && \|\boldsymbol{\beta}\|_1 + \mathbf{c}^\top \boldsymbol{\Gamma} \mathbf{c} \\
&\text{subject to} && |\beta_k| = c_k, \qquad k = 1, \ldots, MN \\
& && \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2 \leq \xi,
\end{aligned} \quad (10)$$

where β_k (resp. c_k) is the k-th entry of $\boldsymbol{\beta}$ (resp $\mathbf{c}$), ξ is some user-defined tolerable residual error, and the symmetric matrix $\boldsymbol{\Gamma}$, with zeros along its main diagonal, penalizes the cross products of the absolute values of $\boldsymbol{\beta}$. That is, the entry $\gamma_{k,\ell}$ in the k-th row and ℓ-th column of $\boldsymbol{\Gamma}$, induces a penalization $\gamma_{k,\ell} c_k c_\ell = \gamma_{k,\ell}|\beta_k||\beta_\ell|$. Here we must point out that the proposed algorithm considers the general case in which some $\gamma_{k,\ell}$ can take negative values, then rewarding co-ocurrence.

The optimization problem in (10) is difficult to solve, since the cost function is not convex whenever $\boldsymbol{\Gamma} \neq 0$. Moreover, the first set of constraints is not convex. However, noting that $\|\boldsymbol{\beta}\|_1 = \mathbf{1}^\top \mathbf{c}$, and introducing the constraint[2] $1 + 2\boldsymbol{\Gamma}\mathbf{c} \geq 0$, we ensure that the cost function increases with c_k. Thus, the optimization problem is

$$\begin{aligned}
&\underset{\boldsymbol{\beta}, \mathbf{c}}{\text{minimize}} && \mathbf{1}^\top \mathbf{c} + \mathbf{c}^\top \boldsymbol{\Gamma} \mathbf{c} \\
&\text{subject to} && |\beta_k| \leq c_k, \qquad k = 1, \ldots, MN \\
& && \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2 \leq \xi \\
& && 1 + 2\boldsymbol{\Gamma}\mathbf{c} \geq 0,
\end{aligned} \quad (11)$$

Algorithm 1 SCA for Cross-Products LASSO
___
Input: $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$, ξ and $\mathbf{y}$.
Output: Recovered signal $\boldsymbol{\beta}$.
Initialize $\mathbf{c}_0 = 0$.
Obtain the Matrices $\boldsymbol{\Gamma}_+$ and $\boldsymbol{\Gamma}_-$ from the EV of $\boldsymbol{\Gamma}$
repeat
    Solve the convex optimization problem in (12)
    Update $\mathbf{c}_0 = \mathbf{c}$
until Convergence
___

where the main difficulty resides in the non-convex cost function. In order to deal with this difficulty and find a solution of the original Karush-Kuhn-Tucker (KKT) conditions [20], we apply the SCA methodology [17, 18, 19]. The main idea is replacing the non-convex functions by a sequence of local convex approximations, which must satisfy three conditions:

- The value of the original function, $f(\cdot)$, and its convex approximation, $\tilde{f}(\cdot)$, at the reference point $\mathbf{x}_0$ should be the same, i.e., $f(\mathbf{x}_0) = \tilde{f}(\mathbf{x}_0)$.

- The gradients at the reference point should coincide, i.e., $\nabla_f(\mathbf{x}_0) = \nabla_{\tilde{f}}(\mathbf{x}_0)$.

- The convex approximation must be an over-estimator of $f(\cdot)$, i.e., $\tilde{f}(\mathbf{x}) \geq f(\mathbf{x}), \forall \mathbf{x}$.

In our particular case, given a reference value $\mathbf{c}_0$ for the vector $\mathbf{c}$, the cost function can be approximated by $\mathbf{1}^\top \mathbf{c} + \mathbf{c}^\top \boldsymbol{\Gamma}_+ \mathbf{c} + 2\mathbf{c}_0^\top \boldsymbol{\Gamma}_-(\mathbf{c} - \mathbf{c}_0)$, where $\boldsymbol{\Gamma}_+$ and $\boldsymbol{\Gamma}_-$ are the positive semidefinite and negative semidefinite parts of $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_+ + \boldsymbol{\Gamma}_-$. It is easy to check that this approximation satisfies the previous conditions, and therefore, the convex problem to be solved in each iteration of the proposed algorithm is finally,[3]

$$\begin{aligned}
&\underset{\boldsymbol{\beta}, \mathbf{c}}{\text{minimize}} && \mathbf{1}^\top \mathbf{c} + \mathbf{c}^\top \boldsymbol{\Gamma}_+ \mathbf{c} + 2\mathbf{c}_0^\top \boldsymbol{\Gamma}_-(\mathbf{c} - \mathbf{c}_0) \\
&\text{subject to} && |\beta_k| \leq c_k, \qquad k = 1, \ldots, MN \\
& && \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\beta}\| \leq \xi \\
& && 1 + 2\boldsymbol{\Gamma}\mathbf{c} \geq 0.
\end{aligned} \quad (12)$$

The overall procedure is summarized in Algorithm 1, where the initial value for $\mathbf{c}$ ($\mathbf{c}_0 = 0$), reduces the cost function to the convex envelope of the original non-convex cost function.

4. NUMERICAL RESULTS

4.1. Known Channel Matrix

For the first experiment we randomly generate channels of length $L = 20$ first, with $\tilde{h}[n]$ following a zero-mean and unit-variance Gaussian distribution, i.e., $\tilde{h}[n] \sim \mathcal{N}(0, 1)$, and then obtain $h[n]$ through an energy normalization: $h[n] =$

---

[2]Although this constraint is redundant at this point, and it can be completely avoided if all $\gamma_{k,\ell} \geq 0$, it will become relevant soon.

[3]Note that the constraint $1 + 2\boldsymbol{\Gamma}\mathbf{c} \geq 0$ plays a crucial role in (12), ensuring that the first set of constraints is satisfied with equality $|\beta_k| = c_k$.

SNR (dB)	LASSO					CP-LASSO				
	$N_d(0)$	$N_d(1)$	$N_d(2)$	N_v	N_{fa}	$N_d(0)$	$N_d(1)$	$N_d(2)$	N_v	N_{fa}
0	18.94	32.10	37.96	95.15	47.63	16.22	31.32	36.58	61.52	40.87
5	13.51	16.07	16.24	8.13	19.60	18.16	27.81	29.42	1.04	7.56
10	9.84	10.10	10.11	2.05	28.61	19.13	26.59	26.95	0	8.39
15	10.04	10.17	10.17	0.94	28.25	19.67	25.78	25.94	0	9.55
20	9.50	9.58	9.58	0.70	28.99	20.06	26.46	26.54	0	8.45

Table 1. Results (averaged over 100 experiments) for unknown channel matrix using a Hanning window as basis element in the reconstruction dictionary when $L = N_{\min} = 10$.

$N_{\min}$	$\bar{K}$	LASSO			CP-LASSO		
		N_r	N_d	N_{fa}	N_r	N_d	N_{fa}
5	73.32	34.41	34.12	0.29	60.79	52.81	7.97
10	42.02	11.66	11.65	0.01	31.38	29.31	2.07
15	29.54	5.73	5.73	0.00	19.24	18.85	0.39
20	22.86	3.33	3.33	0.00	13.33	13.14	0.18
25	18.62	3.38	3.38	0.00	12.15	12.11	0.04

Table 2. Results (averaged over 100 experiments) for random known channel matrices when $L = 20$ and SNR = 10 dB.

$\tilde{h}[n]/\sqrt{\frac{1}{L}\sum_{n=0}^{L-1}|\tilde{h}[n]|^2}$. The amplitudes of the arrivals also follow a Gaussian distribution, $A_k \sim \mathcal{N}(0,1)$, and the number of zero samples between two consecutive spikes is equal to $N_{\min}$ plus a discretized Poisson process with rate κ (i.e., the expected inter-arrival time is $N_{\min} + (\kappa + 1)/\kappa$). The channel matrix is assumed to be known, so we set $\boldsymbol{\Phi} = \mathbf{H}$.

Table 2 shows the results when $\kappa = 1$ and $N = 500$ for LASSO (which corresponds to using (10) with $\boldsymbol{\Gamma} = 0$) and CP-LASSO ($\rho = 100$). Since we are interested in recovering the latent spikes, this table shows the number of activations recovered (N_r), the number of correctly located detections (N_d), and the number of false alarms (N_{fa}), altogether with the average number of spikes ($\bar{K}$). CP-LASSO outperforms LASSO regarding N_d, although at the expense of a higher value of N_{fa} in this case.

4.2. Unknown Channel Matrix

For the second example we use activations that follow the shape given in [21] for electrograms with $L = 11$. The amplitude of the activations is again normally distributed, i.e., $A_k \sim \mathcal{N}(0,1)$, and the negative co-occurrence period is set to $N_{\min} = 10$. The activations arrive periodically now, with a period equal to $N_{\min} + 1$. We assume an unknown channel with a known length, and use a Hanning window of length $L = 10$ to generate the reconstruction dictionary. Table 1 shows the results for $L = 10$ in terms of the number of correct detections ($N_d(k)$ is the number of detections within distance k of a true activation, i.e., $N_d(0) = N_d$ in Table 2), number of violations of the negative co-occurrence period (N_v) and number of false alarms (N_{fa}). In this case CP-LASSO outperforms LASSO w.r.t. all performance measures.
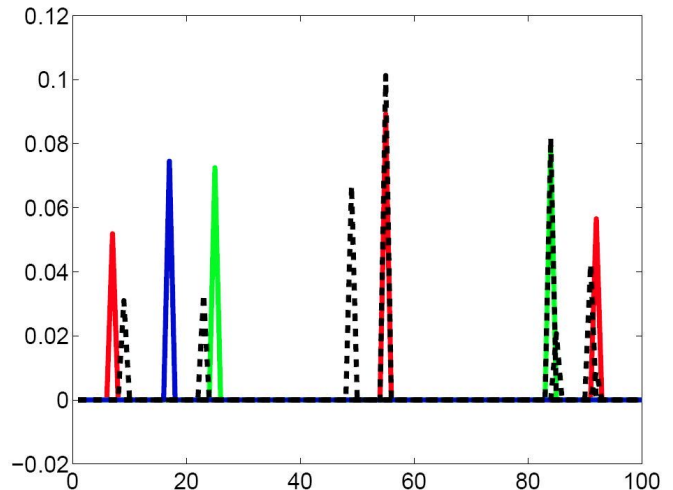

Fig. 1. Activations recovered using LASSO (dashed black line) and CP-LASSO (each colour corresponds to a different basis function).

Finally, we show an example of the activations recovered using an overcomplete dictionary with $M = 3$ in Figure 1. The synthetic data are generated using a random channel with $L = 20$, $N = 100$ and $N_{\min} = 10$. We build the reconstruction dictionary using the mexican hat wavelet with three different variances: $\sigma_1^2 = 0.1$, $\sigma_2^2 = 1$ and $\sigma_3^2 = 10$. The signal contains 8 activations. Using LASSO we recover 8 activations, not always significant, and we have two violations of the negative co-occurrence period. Using CP-LASSO we only detect 6 activations, but there are no violations of the co-occurrence period and all of them are relevant.

5. DISCUSSION

In this paper we have shown how to incorporate a negative co-occurrence restriction into a sparse learning problem. The proposed approach builds on LASSO, adding a novel penalty term. Although some authors have derived approaches to obtain sparse signals that respect some physical/biological restrictions in the past (see e.g. [7, 8, 9]), no approach based on the cross-products has been developed in the literature as far as we know. Finally, we also make use of the successive convex approximations (SCA) idea in order to optimize the resulting non-convex problem.

6. REFERENCES

[1] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-3, no. 6, pp. 610 –621, nov. 1973.

[2] Myung Jin Choi, J.J. Lim, A. Torralba, and A.S. Willsky, "Exploiting hierarchical context on a large database of object categories," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, june 2010, pp. 129 –136.

[3] J. Jalife, M. Delmar, J. Anumonwo, O. Berenfeld, and J. Kalifa, *Basic Cardiac Electrophysiology for the Clinician*, Wiley, 2 edition, 2011.

[4] G. F. Knoll, *Radiation Detection and Measurement*, Wiley, 3nd edition, 2000.

[5] Bruno A. Olshausen and David J. Fieldt, "Sparse coding with an overcomplete basis set: a strategy employed by v1," *Vision Research*, vol. 37, pp. 3311–3325, 1997.

[6] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.

[7] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982 –2001, Apr. 2010.

[8] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society Series B*, vol. 67, no. 1, pp. 91–108, 2005.

[9] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B*, vol. 67, no. 2, pp. 301–320, Jan. 2005.

[10] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, Jan. 2012.

[11] Y.C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042 –3054, June 2010.

[12] Sandra Monzón, Tom Trigano, David Luengo, and Antonio Artés-Rodríguez, "Sparse spectral analysis of atrial fibrillaion electrograms," in *2012 IEEE Machine Learning for Signal Processing Workshop (MLSP)*, Santander (Spain), 23–26 Sep. 2012.

[13] David Luengo, Javier Vía, Sandra Monzón, Tom Trigano, and Antonio Artés-Rodríguez, "Blind analysis of EGM signals: Sparsity-aware formulation," Tech. Rep., Dic. 2012, arXiv.

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

[15] Charles F. Van Loan, "The ubiquitous Kronecker product," *Journal of Computational and Applied Mathematics*, vol. 123, pp. 85–100, 2000.

[16] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, MA (USA), 3rd edition, 1996.

[17] M. Chiang, Chee Wei Tan, D.P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[18] Barry R. Marks and Gordon P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[19] M. Avriel, *Advances in geometric programming*, Plenum Press, New York, 1980.

[20] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, March 2004.

[21] Gerald Fischer, Markus Ch. Stühlinger, Claudia-N. Nowak, Leonard Wieser, Bernhard Tilg, and Florian Hintringer, "On computing dominant frequency from bipolar intracardiac electrograms," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 165–169, January 2007.