ITERATED SPARSE RECONSTRUCTION FOR ACTIVITY ESTIMATION IN NUCLEAR SPECTROSCOPY

Y. Sepulcre

Jerusalem College of Engineering
Department of Computer Science
Jerusalem, Israel

T. Trigano

Shamoon College of Engineering
Department of Electrical Engineering
Ashdod, Israel

ABSTRACT

The aim of nuclear spectroscopy is to provide as many information as possible regarding the activity and the content of an unknown radioactive source. Due to some random perturbations called pileup phenomenon, electrical pulses recorded by the spectrometric apparatus may overlap. Recent developments in compressive sensing and sparse signal reconstruction allow us to build efficient algorithms for the estimation of the activity. We suggest in this paper an iterative algorithm which improves this estimation. This algorithm is based on successive sparse approximations of the residual signal with a slowly decreasing sparsity parameter, so that at each step pulses with smaller amplitude are detected. The algorithm is detailed and its performances compared to other methods based on sparse reconstruction. Results show that the proposed approach perform better than the state-of-the-art methods, and emphasize the usefulness of the iterative scheme.

***Index Terms*—** Sparse reconstruction, nuclear spectroscopy, counting rate estimation, LASSO

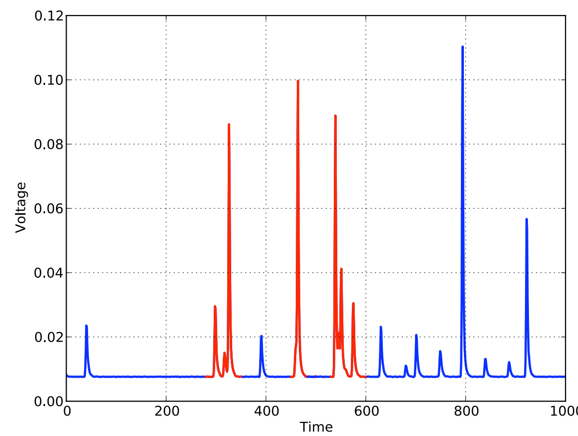


Fig. 1. Example of spectrometric signal from an High Purity Germanium (HPGe) detector (arbitrary units). The parts displayed in red are pileups.

1. INTRODUCTION

In nuclear spectrometry experiments, particles interact with a detector at random times, creating electrical pulses which are afterwards analyzed [1]. The practitioner is usually interested in measuring the activity of radioactive sources and identifying them. However, when the activity of the radioactive source is high, generated pulses may overlap. This phenomenon is known in the literature as pileup effect. An example of real signal with pileups is displayed in Figure 1.

The pileup phenomenon motivates the search for algorithms which allow to separate clusters of electrical pulses for a better identification activity estimation[2]. However, most methods are not fitted to high counting rates, since they do not rely on any shape information of the time signal. When taking into account the typical shape of individual pulses, the problem of pileup correction can be viewed as a sparse regression problem. In the recent years [3], representation of sparse signals has received a considerable attention, and significant advances have been made both from the theoretical and applied point of view. More recently, previous contributions suggested post-processed version of the LASSO [4] in order to estimate the activity of the source, both in homogeneous [5] and inhomogeneous [6] cases. These post-processing steps are used to compensate the incompleteness of the dictionary used, and usually provide very good results close to the optimum. However, it is observed in practice that these methods may be inefficient for very high counting rates. It is therefore necessary to improve the sparse estimation of the signal, whether by adapting the dictionary to the data at hand [7, 8] or by means of iterated sparse reconstructions [9]. We present in this contribution an iterative adaptation of the algorithm described in [5]. The main advantage of the proposed approach is that it does not depend on any *ad hoc* threshold, and is therefore suitable for higher counting rates. The paper is organized as follows: we recall in section 2 the model used and describe how to estimate the activity of the source in an homogeneous setting, and present the iterative algorithm to adapt the sparsity parameter in order to obtain a better estimation of the activity. Results are presented in Section 3, showing that the proposed approach outperforms the previous counting rates estimation techniques in the field.

2. ITERATIVE SPARSE REGRESSION FOR COUNTING RATE ESTIMATION

2.1. Problem formulation

We observe a spectroscopic signal on a finite time interval stemming from particles impinging on a detector, and assume that the signal is sampled on $\mathcal{T} \triangleq \{t_i,\ 1 \leq i \leq N\}$. This signal is modeled classically by a generalized sampled shot noise process [10]:

$$y_i = \sum_{n=1}^{M} E_n h_n(t_i - T_n) + \varepsilon_i,\ 1 \leq i \leq N\,, \tag{1}$$

where $\{T_n,\ 1 \leq n \leq M\}$ are the observed points of the sample path and represent the arrival times of the particles, $\{E_n,\ 1 \leq n \leq M\}$ their energies and $\{h_n,\ 1 \leq n \leq M\}$ the associated pulses shapes. A physically plausible assumption is that the $\{T_n,\ 1 \leq n \leq M\}$ defines a sample path of an homogeneous Poisson process with unknown counting rate λ. The signal contains additional Gaussian noise ε_i with variance σ^2. We furthermore assume that the pulse shapes h_n are causal and normalized, that is $h_n(t) = 0$ for all $t < 0$ and integrates to 1. In this paper we aim to estimate λ given the sole signal $\mathbf{y} \triangleq [y_1 \cdots y_N]^T$. The standard Maximum Likelihood estimate of λ is

$$\hat{\lambda}_{\text{opt}} = \frac{M}{T_M}\,, \tag{2}$$

however M and T_M are unknown in practice. In [5] a method was proposed which is based on sparse linear regression and estimate the arrival times T_n and their number M; it was also proved to reach much better results than traditional thresholding techniques, especially in high intensity regimes. We recall here the detail of this approach.

On HPGe detectors used for example in Gamma spectrometry, an electrical pulse created by a single photon has a characteristic shape created by the charge collection and migration in the detector. In this framework, it is assumed that the h_n's belong to a parametric family of gamma functions

$$\Gamma_{\boldsymbol{\theta}}(t) = c_{\boldsymbol{\theta}}\, t^{\theta_1} \cdot e^{-\theta_2\, t}\, 1(t \geq 0)\,; \tag{3}$$

where the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ belong to a discrete subset of $\mathbb{R}_+^2$ of cardinal p, and $c_{\boldsymbol{\theta}}$ is a normalizing constant. Equation (1) can be related to a sparse regression problem as follows. For each $j < N$ we define a *time block* A_j whose columns are the shapes (3) sampled and translated at t_j:

$$A_j \triangleq \begin{bmatrix} \Gamma_{\theta_1}(t_1 - t_j) & \cdots & \Gamma_{\theta_p}(t_1 - t_j) \\ \vdots & \vdots & \vdots \\ \Gamma_{\theta_1}(t_N - t_j) & \cdots & \Gamma_{\theta_p}(t_N - t_j) \end{bmatrix};$$

A global dictionary $\mathbf{A}$ is then defined as

$$\mathbf{A} = [A_1\ A_2\ \cdots\ A_{N-1}]\,. \tag{4}$$

In this paper we furthermore normalize $\mathbf{A}$ so that the Gram matrix $\frac{1}{N}\mathbf{A}^T\mathbf{A}$ has ones all along the main diagonal. Should the T_n belong to $\mathcal{T}$, the model (1) could be expressed as a standard regression problem $\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \triangleq [\varepsilon_1^T\ \varepsilon_2^T\ \ldots\ \varepsilon_N^T]^T$ models the random noise, and $\boldsymbol{\beta} \triangleq [\boldsymbol{\beta}_1^T\ \boldsymbol{\beta}_2^T\ \ldots\ \boldsymbol{\beta}_N^T]^T$ is a vector supported in the time blocks indexed by the Poissonian sequence $\{T_n,\ n \geq 0\}$, thus making it a sparse regressor. Note that in the latter ε_j and $\boldsymbol{\beta}_j$ are vectors of size p.

In a realistic setting T_n does not belongs to $\mathcal{T}$, and the actual pulse shapes are only close to the positive span of (3) in the ℓ_2 sense. A sparse regressor $\widehat{\beta}$ can still be computed from $\mathbf{y}, \mathbf{A}$ by means of LASSO [4], basis pursuit [3] and so forth; provided the sampling rate is small enough, the time support provided by such a regressor is expected to share similarities with the true beginnings of pulses. In [5], the algorithm for counting rate estimation is mostly based on a LASSO regressor $\widehat{\beta}$ obtained as follows:

$$\widehat{\beta}(r) \triangleq \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y} - \mathbf{A}\boldsymbol{\beta}\|^2 + r\,|\boldsymbol{\beta}|_1 \right\}, \tag{5}$$

where r is a parameter which controls the sparsity of $\widehat{\beta}$. In order to compensate false time detection, we needed to threshold active blocks giving small contribution and merge consecutive active blocks into one single event. This method performs much better than state of the art algorithms. One main drawback is that close arrival times can be merged into one single event, and that it requires *ad hoc* thresholding of active time blocks. In case of very high activities, clusters of active blocks can hide distinct events, thus yielding an underestimation of λ. We seek here for a better estimate, and we show that an improvement can be obtained by iteratively decreasing the sparsity parameter in the LASSO, in other words by checking the "history" provided by the LARS algorithm [11], as now explained.

2.2. Iterative sparse regression for intensity estimation

We consider a sampled signal $\mathbf{y}$ and a dictionary $\mathbf{A}$ built as previously. The LARS algorithm introduced in [11] provides an iterative method to compute the solution $\widehat{\beta}(r)$ in (5) for a given r. Based on the constraints imposed by the KKT conditions of (5), it builds a continuous path $\widehat{\beta}(s)$ from $s = \infty$ to $s = r$. Whenever s crosses critical values $\{s_k\}_{k \geq 0}$, new variables are added or withdrawn to the support of $\widehat{\beta}(s)$, that is the set of its non-zero coefficients. Inside each interval (s_{k+1}, s_k) this support does not vary and the coefficient values vary linearly with the parameter r. The counting rate estimator suggested in this paper is based on the same principle. It differs from the LARS in two aspects: at each stage, we also update a set of active arrival times, and we keep track of all the history of these sets to build estimates of M and T_M. The proposed algorithm is summarized in Algorithm 1, and is detailed in the next paragraph.

Input :
$\mathbf{y}$: input signal;
$\mathbf{A}$: dictionary built in (4);
$\{r_n\}_{0\le n\le n_{\max}}$: a decreasing sequence of positive sparstiy parameters.
Output: $\widehat{\lambda}$ estimator of the counting rate λ.

Initialize $\mathbf{T} = \emptyset$;
for $j = 0, \cdots, n_{\max}$ **do**
  Compute $\widehat{\beta}(r_j)$ the solution of (5) ;
  Select the active blocks
  $\mathcal{J}(r_j) = \left\{ i\,;\ \sum_{\beta_k \in \widehat{\boldsymbol{\beta}}_i(r_j)} \beta_k > 0 \right\}$;
  Decompose $\mathcal{J}(r_j)$ into a union of consecutive active blocks:
  $\mathcal{J}(r_j) = \bigcup_{k=1}^{s} \{i_k, i_k+1, \cdots, i_k+l_k\},$;
  Update the set of estimated arrival times:
  $\mathbf{T} \leftarrow \mathbf{T} \cup \{i_1, \cdots, i_s\}$;
end
Compute the plug-in counting rate estimate:
$\widehat{\lambda} = \dfrac{|\mathbf{T}|}{\max \mathbf{T}}$.

Algorithm 1: Counting rate estimation by iterative LASSO

Define a decreasing sequence $\{r_n\}_{0\le n\le n_{\max}}$ of positive sparsity parameters in order to solve (5). This corresponds to reconstruct the same signal $\mathbf{y}$ with decreasing ℓ_1 penalizations. We build iteratively the set of estimated arrival times $\mathbf{T}$ as follows: at step n, the LASSO regressor (5) $\widehat{\beta}(r_n)$ is computed. Then we choose the time blocks indexed by each $i \le N$ such that the sum of the coefficients of $\widehat{\boldsymbol{\beta}}_i(r_n)$ is positive; in other words we select the time-blocks pattern

$$\mathcal{J}(r_n) = \left\{ i\,;\ \sum_{\beta_k \in \widehat{\boldsymbol{\beta}}_i(r_n)} \beta_k > 0 \right\}. \qquad (6)$$

Note, however, that consecutive active time blocks in (6) likely correspond to the same electrical pulse, since the signal is not created from $\mathbf{A}$ in the strict sense and since the *irrepresentability* condition does not hold [12]: indeed consecutive time blocks are usually highly correlated. Therefore we decompose (6) into a disjoint union of clusters of consecutive active times blocks

$$\mathcal{J}(r_n) = \bigcup_{k=1}^{s} \{i_k, i_k+1, \cdots, i_k+l_k\}, \qquad (7)$$

where for all $k = 1 \dots s$: $i_k \le i_k + l_k < i_{k+1}$ and $i_{s+1} = N$. The set $\mathbf{T}$ is then updated by including $\{i_1, \cdots, i_s\}$. It is noteworthy that the size of $\mathbf{T}$ can only grow as iterations are performed. By doing so, we keep track of possible clusters of active time blocks which merged into a single cluster as r_j decreased, and attenuate the underestimation of λ.

Eventually, the counting rate is estimated by plugging the number of elements in $\mathbf{T}$ (denoted by $|\mathbf{T}|$) and its maximum value in (2), namely

$$\hat{\lambda} \triangleq \frac{|\mathbf{T}|}{\max \mathbf{T}}. \qquad (8)$$

The performances of the proposed approach are strongly related to the choice of the sequence $\{r_n\}_{0\le n\le n_{\max}}$. Intuitively, this sequence should decrease slowly enough in order to catch all the changes in (7) one after the other and to detect all disjoint intervals where the LASSO pattern stays constant. The initial value r_0 must be taken so that only one time block is active. From [11], a sensible choice is to set

$$r_0 = \left\| \frac{1}{N} \mathbf{A}^T \mathbf{y} \right\|_\infty,$$

and setting $r_{n_{\max}} = \sigma$ prevents to overfit the input signal. We emphasize that the choice of $\{r_n\}_{0\le n}$ is not unique. A practical setting criterion is that $r_n - r_{n+1}$ should be smaller than the minimum distance between two consecutive critical values in the *LARS* iterations. An optimal (in the sense of computational cost) sequence so that the latter criterion depends on $\mathbf{A}$ as well as on the type of signals $\mathbf{y}$ we have to analyze, and finding the optimum is beyond the scope of the present paper.

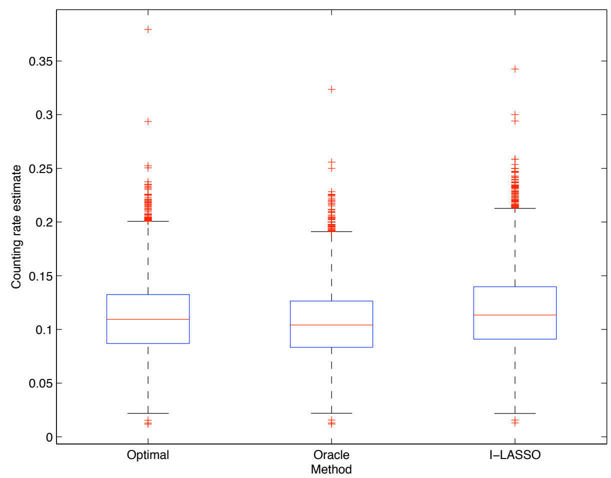
There are two main differences between the approach presented in this paper and the one in [5]: in the latter, a single sparsity parameter r is chosen from the very beginning, whereas we keep trace here of all beginnings of clusters created in the process; moreover the only threshold we apply is on the correlation r_j (which stays above known σ to prevent overfitting), whereas time blocks are not thresholded by some ad hoc method as before. These advantages make the proposed method more suitable for the practitioner.

3. RESULTS AND DISCUSSION

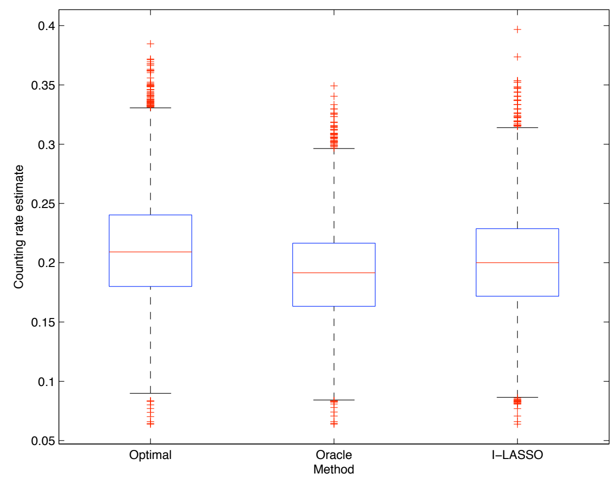
We present in this section the performances of the proposed adaptation procedure on simulations similar to [5].

3.1. Experimental settings

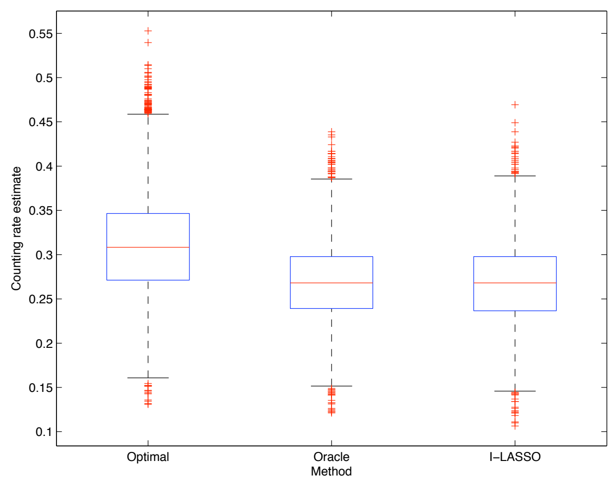
We investigate several values of λ ranging from 0.1 to 0.4, while the sampling rate stays equal to 1. Compared to real-life experiments this corresponds to counting rates from 1.10^6 to 4.10^6 photons per second with a digital apparatus with sampling rate 10 MHz. For each intensity value, we draw 10^4 signals. Pulse shapes are created by drawing randomly two



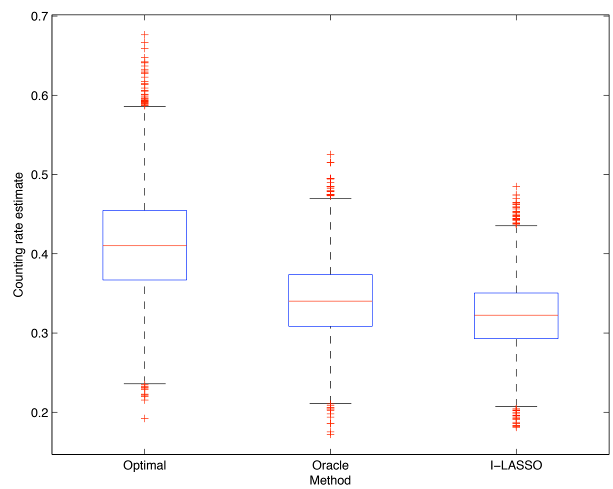
(a) $\lambda = 0.1$.



(b) $\lambda = 0.2$.



(c) $\lambda = 0.3$.



(d) $\lambda = 0.4$.

Fig. 2. Boxplots of the counting rates estimates, optimal (left), oracle (middle) and I-LASSO (right). We observe that even in the case of very high activity, the obtained values remain close to the oracle

parameters θ_1, θ_2 inside $[0; 2]$. The energies E_n are drawn accordingly to a truncated Gaussian density with mean 20 and variance 9. We assume a good Signal Noise Ratio (SNR) ($\sigma = 0.1$), which is the case for HPGe detectors. We compare the estimations of λ obtained by the proposed estimate (8) (denoted by I-LASSO) with the optimal estimate (2) and with the best estimate attainable with sampled data, denoted by Oracle, and defined as

$$\lambda_{\text{oracle}} = \frac{M'}{\max_{1 \leq n \leq M} \lfloor T_n \rfloor},$$

where $\lfloor x \rfloor$ is the integer part of x and M' is the number of distinct values of $\lfloor T_n \rfloor, n = 1 \ldots M$. In our simulations we chose an arithmetically decreasing sequence, that is we defined a small parameter $\rho > 0$ and set $r_j = r_0 - j\rho$ for all $j > 0$, as long as r_j remains greater than σ.

3.2. Commented results

Boxplots illustrating the obtained results are displayed in Figure 2. When compared to the results of [5], we observe that I-LASSO provides slightly better biases and variances, especially for high counting rates where we actually detect more pulses.

Since there are relatively high correlations between variables belonging to close time blocks LASSO is bound to estimate some single pulses by using consecutive time blocks, even for high parameter values. Note that running LASSO once with an inappropriate r increases the risk of misdetection, because the pulses could not be well "separated" by the correlation level fixed by r. This justifies the use of an iterative approach.

Secondly, as r decreases blocks whose purpose is to improve the signal estimation can be activated, whereas they do not indicate actual beginnings of pulses. They tend to further merge together clusters of important time blocks (which have higher coefficients). However, since these lasts were separated when r was higher, we avoid misdetections.

Note that we considered a time block as active if the sum of its coefficients are positive: indeed it may happen that blocks contain negative coefficients, and among them some could not be related to a pulse start; simulations showed that selecting all time block containing one active variable at least usually leads to underperformances. Another way could be to select variables having only positive correlation with the residual.

4. CONCLUSION

In the particular problem of sparse regression for counting rate estimation, we defined a simple iterative procedure exploiting all the model selections obtainable by LASSO progression. This method makes no use of any thresholding step, and performs quite well for intensities range of real practical interest. In future works we aim to use this procedure to address other problems involving pulses separations, for instance estimating the energy spectrum of the signal.

5. REFERENCES

[1] G. F Knoll, *Radiation Detection and Measurement*, Wiley, 2nd edition, 1989.

[2] C. Michotte and M. Nonis, "Experimental comparison of different dead-time correction techniques in single-channel counting experiments," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 608, no. 1, pp. 163–168, Sept. 2009.

[3] S. Chen, D. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33—61, 1998.

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

[5] T. Trigano, Y. Sepulcre, M. Tal, and Y. Mashiach, "On a sparse regression algorithm for activity estimation in gamma spectrometry," in *2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI)*, 2010, pp. 268–272.

[6] T. Trigano, Y. Sepulcre, M. Roitman, and U. Aferiat, "On nonhomogeneous activity estimation in gamma spectrometry using sparse signal representation," in *2011 IEEE Statistical Signal Processing Workshop (SSP)*. June 2011, pp. 649–652, IEEE.

[7] T. Zhang, "Adaptive Forward-Backward greedy algorithm for learning sparse representations," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4689–4708, 2011.

[8] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, 2010.

[9] N. Meinshausen, "Relaxed LASSO," *Computational Statistics and Data Analysis*, pp. 374–393, 2007.

[10] D. L Snyder, *Random Point Processes*, Wiley, 1975.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[12] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (LASSO)," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.