

Sparse regression algorithm for activity estimation in γ spectrometry

Y. Sepulcre, T. Trigano and Y. Ritov

January 31, 2014

Abstract

We consider the counting rate estimation of an unknown radioactive source, which emits photons at times modeled by an homogeneous Poisson process. A spectrometer converts the energy of incoming photons into electrical pulses, whose number provides a rough estimate of the intensity of the Poisson process. When the activity of the source is high, a physical phenomenon known as pileup effect distorts direct measurements, resulting in a significant bias to the standard estimators of the source activities used so far in the field. We show in this paper that the problem of counting rate estimation can be interpreted as a sparse regression problem. We suggest a post-processed, non-negative, version of the Least Absolute Shrinkage and Selection Operator (LASSO) to estimate the photon arrival times. The main difficulty in this problem is that no theoretical conditions can guarantee consistency in sparsity of LASSO, because the dictionary is not ideal and the signal is sampled. We therefore derive theoretical conditions and bounds which illustrate that the proposed method can none the less provide a good, close to the best attainable, estimate of the counting rate activity. The good performances of the proposed approach are studied on simulations and real datasets.

1 Introduction

Rate estimation of a point process is an important problem in nuclear spectroscopy. An unknown radioactive source emits photons at random times, which are modeled by an homogeneous Poisson process. Each photon which interacts with a semiconductor detector produces electron-hole pairs, whose migration generates an electrical pulse of finite duration. We can therefore estimate the activity of the source by counting the number of activity periods of the detector. We refer the reader to [12] and [13] for further insights on the physical aspects in this framework. However, when the source is highly radioactive, the durations of the electrical pulses may be longer than their interarrival times, thus the pulses can overlap. In gamma spectrometry, this phenomenon is referred to as pileup. Such a distortion

induces an underestimation of the activity, which become more severe as the counting rate increases. This issue is illustrated in Figure 1.

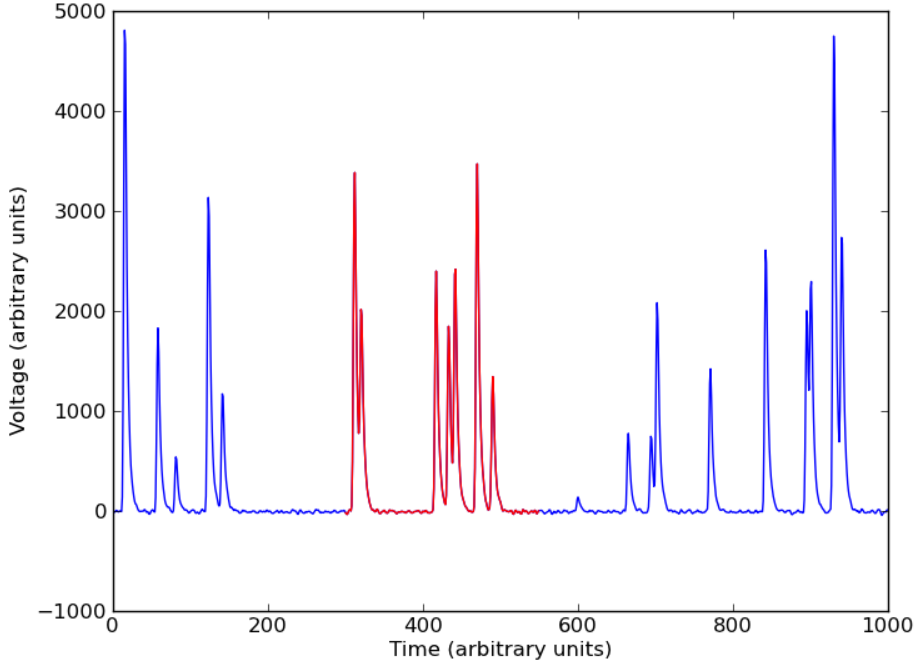


Figure 1: Example of a spectrometric signal. The red part is an example of piled up electrical pulses.

In its mathematical form, the current intensity as a function of time can be modeled as a general shot-noise process

$$y(t) \triangleq \sum_{k \geq 1} E_k \Phi_k(t - T_k), \quad (1)$$

where $\{E_k, k \geq 1\}$ and $\{\Phi_k(s), k \geq 1\}$ are respectively the energy and the shape of the electrical pulse associated to the k -th photon, and $y(t)$ defines the continuous time recorded signal. The pulse shapes $\{\Phi_k(s), k \geq 1\}$ are assumed to belong to a parametric family of functions Γ_Θ , $\Theta \subset \mathbb{R}^n$. The restriction of the signal to a maximal segment where it is strictly positive is referred to as a *busy period*, and where it is 0 as *idle period*. In practice, we observe of sampled version of (1) with additional noise, and wish to estimate from this recorded digital signal the counting rate activity.

The problem of activity estimation has been extensively studied in the field of nuclear instrumentation since the 1960's (see [7] or [17] for a detailed review of these early contributions; classical pileup rejection techniques are detailed in [1]). Early papers on pileup correction focus specifically on activity correction methods, such as the VPG (Virtual Pulse Generator) method described in [25, 26]. Moreover, it must be stressed that these techniques are strongly related to the instrumentation used for the experiments. Recent offline methods

are based on direct inversion techniques [20] or computationally intensive methods [4], and are usually not fitted for very high counting rates. It is of interest to consider fast, event-by-event pile-up correctors for real-time applications, as proposed in [22] for calorimetry and in [18] for scintillators. One of the main advantages of the methods developed in [20] is that they do not rely on any shape information of the time signal, but rather on the alternance of the idle and busy periods of the detector. However, when the activity of the radioactive source is too high, we observe very few transitions from busy to idle periods, thus making this information statistically irrelevant.

In the latter case, it is therefore necessary to introduce additional assumptions on the pulse shapes (e.g. to specify Γ_{Θ}), and to estimate both the signal sample path on a relevant basis. This can be formally viewed as a regression problem. However, due to the nature of the physical phenomenon, and since Poisson processes usually represent occurrences of rare events, the regressor chosen to estimate the signal must be sparse as well. Since the seminal papers [19] and [9], representation of sparse signals has received a considerable attention, and significant advances have been made both from the theoretical and applied point of view. Several recent contributions [11] suggest efficient algorithms yielding estimators with good statistical properties, thus making sparse regression estimators a possible option for real-time processing. In this paper, we chose to use a modification of LASSO with a positivity constraint [11]. Indeed, LASSO provides a sparse solution close to the real signal for the ℓ_2 -norm. However, since we are not interested in the reconstruction of the signal for activity estimation, but rather in the Poissonian arrival times, it is of interest to investigate the consistency in selection of the sparsity pattern. Numerous recent works have been devoted to this general question about LASSO, the first ones being [31] and [15]. Both papers introduced independently the so-called *irrepresentability* condition as a necessary condition for selection consistency. More recently, [24] developed the conditions under which the irrepresentability condition is also a sufficient one. We also refer to [30], [16], [21] and for recent results on consistency in the ℓ_2 -sense for the signal estimation; note however that the estimation of the activity of the source is related to the selection consistency issue, whereas the consistency in the ℓ_2 sense should be used for energy spectrum reconstruction. The problem we address in this paper shares also similarities with the reconstruction of sampled signals with finite rate of innovation [23]. In the latter, the authors present a method based on the use of the annihilator filter used in error-correction coding, which allows to reconstruct perfectly a Poisson driven signal made of splines of piecewise polynomials, even when it is not bandwidth limited. This leads to a purely algebraic reconstruction of the signal when it can be decomposed on a known functional base. However, the cornerstone for algebraic reconstruction is the full knowledge of this base, which is not the case in our framework.

The paper is organized as follows. Section 2 presents the model and the derivation of the estimator of the counting rate. This estimation can be roughly seen as a post-processed version of the non-negative LASSO. Though (1) is rather close to a standard linear regression one, the presented problem is difficult to address, since the discrete signal stemming from $y(t)$ is not generated from a specific, known dictionary. Moreover, it is impossible to infer

the exact number of Poissonian arrivals between two sampling points. Both considerations imply that theoretical conditions (e.g. derived in [24]) which ensure consistency in sparsity are not met in this case.. We therefore present in Section 3 theoretical results showing that the activity of the source can be recovered almost as well as the best estimator we could build from a full knowledge of the Poisson process and discrete observations with a high probability. Finally, section 4 illustrates on some applications the effectiveness of the proposed approach, both on simulations and real data, with comments. Details of the calculations and proofs of the presented results are detailed in the appendix.

2 Sparse regression based method for activity estimation

2.1 Model and assumptions

We observe a signal uniformly sampled on some subdivision $\mathcal{T} \triangleq \{0 = t_0, t_1, t_2, \dots, t_{N-1}\}$ with sampling period Δt , stemming from a discrete version of (1):

$$y_i = \sum_{n=1}^M E_n \Phi_n(t_i - T_n) + \varepsilon_i, \quad 0 \leq i \leq N - 1, \quad (2)$$

where $\{T_n, 1 \leq n \leq M\}$ is the sample path of an homogeneous Poisson process with constant unknown intensity λ , $\{E_n, 1 \leq n \leq M\}$ is a sequence of independent and identically distributed (iid) random variables representing the photons energies, with unknown probability density function f , $\{\Phi_n, 1 \leq n \leq M\}$ is a sequence of functions to be defined later which characterize the electric pulse shapes generated by the photons, and $\{\varepsilon_i, 0 \leq i \leq N - 1\}$ is a sequence of iid Gaussian random variables with zero mean and variance σ^2 representing the additional noise of the input signal. Alternatively, when defining the matrix $\mathbf{\Phi} \triangleq [\Phi_n(t_i - T_n)]_{0 \leq i \leq N-1, 1 \leq n \leq M}$ and the vectors $\mathbf{y} \triangleq [y_0, y_1, \dots, y_{N-1}]^T$, $\mathbf{E} \triangleq [E_1, \dots, E_M]^T$ and $\boldsymbol{\varepsilon} \triangleq [\varepsilon_0, \dots, \varepsilon_{N-1}]^T$, (2) can be rewritten in a matricial form:

$$\mathbf{y} = \mathbf{\Phi} \mathbf{E} + \boldsymbol{\varepsilon}. \quad (3)$$

All along the paper, it is assumed for convenience that N is an even number. The problem to address is the estimation of λ given \mathbf{y} . However, no T_n belongs to \mathcal{T} with probability 1. We thus introduce the following integer subset related to the closest sample times from the Poisson arrivals T_n :

$$P_0 \triangleq \{\lfloor T_n / \Delta t \rfloor; n = 1, \dots, M\}, \quad (4)$$

where $\lfloor x \rfloor$ denotes the closest integer to x . Note that provided $\lambda \Delta t \ll 1$, P_0 is a sparse subset of $\{0, 1, \dots, N - 1\}$. We further on denote by $\bar{\mathbf{y}}$ the noise-free part of the signal (3), that is $\bar{\mathbf{y}} \triangleq \mathbf{\Phi} \mathbf{E}$. All along the paper, it is assumed that the random variables $\{E_n, 1 \leq n \leq M\}$

are bounded by positive and known constants E_{\min}, E_{\max} :

$$0 < E_{\min} \leq E_n \leq E_{\max}, \quad n = 1, \dots, M. \quad (5)$$

In practice, neither \mathbf{E} nor Φ are known, so the problem cannot be seen as a standard regression problem. Nevertheless, a single electrical pulse Φ_n has a specific shape, characterized in most detectors by a rapid growth created by the charge collection followed by an exponential decay as the charges migrate to the detector electrodes. Thus, a natural idea is to rely on some user predefined dictionary of truncated gamma shapes in order to obtain a modelization of (3) we can work with. Since a gamma shape is parametrized by two scale and shape parameters, we define a set of p pairs of such parameters by $\boldsymbol{\theta} \triangleq \{(\theta_1^{(s)}, \theta_2^{(s)}); s = 1, 2, \dots, p\}$. For all $s = 1, \dots, p$, we define the following pulse shape:

$$\Gamma_s(t) \triangleq c_s t^{\theta_1^{(s)}} \exp(-\theta_2^{(s)} t) \mathbf{1}(0 < t \leq \tau \Delta t),$$

where τ is a positive constant integer defining the common support of the pulse shapes, and c_s is a normalizing constant chosen so that $\frac{1}{N} \sum_{i=0}^{N-1} \Gamma_s(t_i)^2 = 1$. Accordingly, we define the following $N \times p$ matrix \mathbf{A}_k whose columns are sampled versions of the previously defined pulse shapes, translated by t_k :

$$\mathbf{A}_k \triangleq [\Gamma_s(t_i - t_k)]_{0 \leq i \leq N-1, 1 \leq s \leq p}. \quad (6)$$

We further on refer to \mathbf{A}_k in (6) as the *time block* associated to the k -th point. We now define a global dictionary \mathbf{A} by concatenating these time blocks:

$$\mathbf{A} = [\mathbf{A}_0 \quad \mathbf{A}_1 \quad \dots \quad \mathbf{A}_{N-1}]. \quad (7)$$

Note that (7) defines a $N \times Np$ matrix with full rank N . Therefore, an equivalent version of (3) consists in some linear decomposition of $\bar{\mathbf{y}}$ along the columns of \mathbf{A} , for some unknown regressor $\boldsymbol{\beta}$. Recall that with probability 1, no T_n belongs to \mathcal{T} neither Φ_n is a column of \mathbf{A} . In that sense, we shall say that our dictionary \mathbf{A} is *incomplete*. Therefore, the model investigated throughout the rest of the paper is

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad (8)$$

in which $\boldsymbol{\delta} \triangleq \bar{\mathbf{y}} - \mathbf{A}\boldsymbol{\beta}$ denotes the discrepancy between the decomposition on Φ and \mathbf{A} , and which $\mathbf{A}\boldsymbol{\beta}$ belongs to the closed positive span \mathcal{C} defined as

$$\mathcal{C} \triangleq \left\{ \sum_{i \in P_0} \mathbf{A}_i x_i ; x_i \in \mathbb{R}_+^p \text{ such that } E_{\min} \leq \frac{1}{\sqrt{N}} \|\mathbf{A}_i x_i\|_2 \leq E_{\max} \right\}. \quad (9)$$

Note that \mathcal{C} parametrizes models supported on timeblocks indexed by P_0 only. The reason for bounding $\|\mathbf{A}_i x_i\|_2$ in (9) can be understood in light of (5). Since P_0 contains essentially

all the information one could ever retrieve from T_n , the set of all the T_n and $\Delta t P_0$ tend to be identical as Δt tends to 0. Since we expect the decomposition $\mathbf{A}\boldsymbol{\beta}$ to be quite close to the decomposition $\boldsymbol{\Phi}\mathbf{E}$, it is rather natural to focus on the best discrepancy measure $\boldsymbol{\delta}$ one could ever get when imposing similar constraints on the energies encoded in the model $\mathbf{A}\boldsymbol{\beta}$. Following (8), and since we want to focus on model with small discrepancy, we use in the rest of the paper $\alpha \triangleq \|\boldsymbol{\delta}\|_2/\sqrt{N}$ as a standard discrepancy measure.

2.2 Additional notations

We introduce here for convenience the notations used in the following sections. Given any finite discrete set I , we denote by $|I|$ its cardinality. We denote by $\mathbf{1}_I$, the column vector of length $|I|$ whose all coefficients are equal to 1. If \mathbf{u}, \mathbf{v} are two vectors of identical size, we shall write $\mathbf{u} < \mathbf{v}$ (respectively $\mathbf{u} \leq \mathbf{v}$) when all entries of $\mathbf{v} - \mathbf{u}$ are positive (nonnegative).

If n is any integer in $\{1, \dots, Np\}$, thus indexing a column of \mathbf{A} , we shall refer to this column by A_n ; similarly for any regressor $\boldsymbol{\beta} \in \mathbb{R}^{Np}$ the n th entry is denoted by β_n . Given I a subset of $\{0, 1, \dots, N-1\}$, we denote by \mathbf{A}_I the submatrix obtained by concatenation of times blocks whose index belong to I , namely

$$\mathbf{A}_I \triangleq [\dots, \mathbf{A}_k, \dots]_{k \in I} . \quad (10)$$

Given two subsets I, J of $\{0, 1, \dots, N-1\}$, we define the Gram matrix of size $p|I| \times p|J|$ associated to A_I and A_J as

$$\mathbf{G}_{I,J} \triangleq \frac{1}{N} \mathbf{A}_I^T \mathbf{A}_J, \quad (11)$$

and whenever $I = J$ and reduces to one singleton, we shall drop the dependency in I, J and write more simply \mathbf{G} , since by the very construction of \mathbf{A} , the Gram matrix of one timeblock is independent of the block index.

Given $\boldsymbol{\beta}$ a vector of size pN , it will be naturally decomposed along the timeblocks: $\boldsymbol{\beta} = [\boldsymbol{\beta}_0^T; \dots; \boldsymbol{\beta}_{N-1}^T]^T$, where for all i , $\boldsymbol{\beta}_i \in \mathbb{R}^p$. We define the block pattern of $\boldsymbol{\beta}$ as

$$J(\boldsymbol{\beta}) \triangleq \{i; \boldsymbol{\beta}_i \neq 0\} . \quad (12)$$

Given some integer k in $\{0, 1, \dots, N-1\}$, we define for all positive real α the α -neighborhood of k as:

$$V_\alpha(k) \triangleq [k - \alpha; k + \alpha] \cap \{0, 1, \dots, N-1\}, \quad (13)$$

and denotes his complement by $\overline{V_\alpha(k)} = \{0, 1, \dots, N-1\} \setminus V_\alpha(k)$. Alternatively, (13) can be reformulated accordingly to the correlations between blocks, since the discrete correlation between two shapes is a decreasing function of the distance between their time shifts, and is zero whenever this distance is greater than τ . Therefore, given a integer k in $\{0, 1, \dots, N-1\}$, we define some neighbourhood of k accordingly to some specified correlation level $0 \leq \rho \leq 1$:

$$\mathcal{T}_\rho(k) \triangleq \{0, 1, \dots, N-1\} \setminus \{j; \max |\mathbf{G}_{\{k\}, \{j\}}| < \rho\}. \quad (14)$$

In other words, $j \in \overline{\mathcal{T}_\rho(k)}$ if and only if every column in the j -th timeblock has a correlation lower than ρ with every column in \mathbf{A}_k , expressing the fact that t_j is somehow 'distant' from t_k . Due to these considerations, we can associate to $\rho \in [0, 1]$ a real α such that $V_\alpha(k) = \mathcal{T}_\rho(k)$. Obviously, when $\rho' \leq \rho$ one has $\mathcal{T}_\rho(k) \subset \mathcal{T}_{\rho'}(k)$, that is $\alpha \leq \alpha'$.

We eventually define two quantities which will appear in the theoretical bounds obtained. If k is an integer such that $\tau \leq k \leq N - \tau - 1$, we shall define

$$\mathcal{G} \triangleq \frac{1}{N} \sum_{l \in V_\tau(k)} \max_{i,j} \mathbf{A}_k^T \mathbf{A}_l(i, j) \quad (15)$$

as the sum of all maximal correlations per block with respect to the k -th timeblock (note that \mathcal{G} is independent of k), and

$$\mathbf{t}(x) \triangleq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$$

2.3 Overview of the estimation procedure

Recall that our objective is to estimate λ given \mathbf{y} . It is well known that if $\{T_n, 1 \leq n \leq M\}$ are the points of an homogeneous Poisson process, the inter-arrival times are iid random variables with common exponential distribution with parameter λ . Therefore, λ can be consistently estimated by

$$\lambda_c \triangleq \frac{M}{T_M}. \quad (16)$$

However, \mathbf{y} is a discrete-time signal, therefore (16) cannot be attained since we are restricted to use only times in \mathcal{T} . Therefore, the best estimate of λ attainable in practice is defined as

$$\lambda_{\text{opt}} \triangleq \frac{|P_0|}{\Delta t \max P_0}. \quad (17)$$

It is likely that $\lambda_{\text{opt}} < \lambda_c$, since $|P_0| < M$; however, provided $\lambda \Delta t$ is small, λ_c and λ_{opt} should remain close. The main idea of this paper is therefore to plug in (17) estimates of M and T_M , as now explained. If the signal is modelled by (8), the set $J(\boldsymbol{\beta})$ still contains much fewer elements than N . Thus, we would like to recover first $J(\boldsymbol{\beta})$, and make use of a non-negative LASSO estimator (NNLASSO) [19, 11]:

$$\widehat{\boldsymbol{\beta}}(r) = \arg \min_{\{\boldsymbol{\beta} \in \mathbb{R}^{N_p}\}} \left\{ \frac{1}{2N} \left\| \mathbf{y} - \sum_{m=0}^{N-1} \mathbf{A}_m \boldsymbol{\beta}_m \right\|_2^2 + r \sum_{m=0}^{N-1} |\boldsymbol{\beta}_m|_{\ell_1} \right\} \quad (18)$$

such that $\boldsymbol{\beta} \geq 0$,

where the tuning parameter r quantifies the tradeoff between sparsity and estimation precision. NNLASSO provides a sparse estimator $[\widehat{\boldsymbol{\beta}}_0^T(r), \dots, \widehat{\boldsymbol{\beta}}_{N-1}^T(r)]^T$ such that the linear model $\widehat{\mathbf{y}} = \mathbf{A} \widehat{\boldsymbol{\beta}}(r)$ approximates accurately the signal \mathbf{y} . In practice, (18) can be efficiently computed by a modification of the LARS algorithm [11]. Note that the group-LASSO [29]

also exploits the time blocks decomposition of $\boldsymbol{\beta}$ and provide a block-sparse regressor. However, in this paper, we cannot assume the groups to be fully known, due to the incompleteness of \mathbf{A} .

Assuming that the solution (18) is known, the estimation of λ should be carefully done. It is tempting to estimate the arrival times with the set $J(\widehat{\boldsymbol{\beta}}(r))$ and the total number of occurrences by $|J(\widehat{\boldsymbol{\beta}}(r))|$, then plug this data into (17). However $J(\widehat{\boldsymbol{\beta}}(r))$ may contain consecutive active time blocks which do not all correspond to real arrival times. This is not surprising: since \mathbf{A} is incomplete, $J(\boldsymbol{\beta})$ may itself be distinct from P_0 . In this paper we suggest an additional thresholding step to the estimation of λ to circumvent this issue, that is

1. solve (18) to obtain $\widehat{\boldsymbol{\beta}}(r)$.
2. set all the $\widehat{\boldsymbol{\beta}}_m(r)$ such that $\|\widehat{\boldsymbol{\beta}}_m(r)\|_1 < \eta$ to zero, where η is a user defined threshold to be precised later;
3. estimate the arrival times recursively $\widehat{T}_n \triangleq \min_{k=0, \dots, N-1} \{k\Delta t > \widehat{T}_{n-1}; \widehat{\boldsymbol{\beta}}_{k-1}(r) = 0, \widehat{\boldsymbol{\beta}}_k(r) \neq 0\}$, and $\widehat{M} \triangleq |\{k; \widehat{\boldsymbol{\beta}}_{k-1}(r) = 0, \widehat{\boldsymbol{\beta}}_k(r) \neq 0\}|$.
4. Estimate the activity as

$$\widehat{\lambda}(r, \eta) \triangleq \frac{\widehat{M}}{\widehat{T}_M} \quad (19)$$

We refer to steps 2 and 3 as "post processing" steps. Both steps can be heuristically understood as follows: step 2 is introduced since time blocks containing negligible weights are probably selected to improve slightly the estimation, but are not related to pulses start; indeed in realistic situations all the pulses considered, including the real ones, start with similar sharp slopes, but decrease differently, which makes these "negligible" time blocks appear behind the pulse start. In step 3 we merge consecutive selected time blocks due to high correlations between blocks and incompleteness of the dictionary, as mentioned above. Clearly a good estimation of λ is conditioned by a careful choice of both sparsity and thresholding parameters r and η . A reasonable a practical choice is to set them accordingly to the noise variance, as seen in the applications section. Note also that the cornerstone of Step 2 is that λ is small enough with respect to the signal sampling frequency, so that consecutive active blocks would unlikely correspond to two distinct events. Even if this thresholding fails in case of extremely high counting rates, as seen in the application sections, we emphasize that it covers most spectrometric applications, making it very relevant in practice.

Note also that alternative methods to (18), which are based on iterative and reweighting procedures, exist [8]. These methods seem appealing for higher activities, since they are known to provide sparser solutions than NN-LASSO. Similarly, sparse Bayesian learning techniques [27, 28] are known to provide sparser results in the case of very correlated dictionaries. Nevertheless, in practice, the fact that the spectrometric signal \mathbf{y} does not stem

directly from \mathbf{A} cripples their performances, and the post-processing introduced in the latter remains necessary even in this case, as seen later in the applications section.

3 Theoretical results

In order to guarantee some consistency in estimation as well as in selection, previous works imposed conditions on the dictionary \mathbf{A} : for instance low correlations between columns [10, 6, 14, 5] or positivity of minors of specific sizes [16, 6, 30]. The estimation procedure described in this paper is close to [16], which suggest improvements of LASSO by hard-thresholding coefficients, so that only representative variables are selected. In [31, 24], the so called *irrepresentability* condition is introduced, and is proved to be necessary if we wish selection consistency with high confidence. In our case remember we wish to recover P_0 , thus the next subsection details an irrepresentability condition on $\max_{\|z\|_\infty \leq 1} \|\mathbf{G}_{P_0, P_0}^{-1} \mathbf{G}_{P_0, P_0} z\|_\infty$ so that>NNLASSO could theoretically select variables belonging only to P_0 . Nevertheless, this insight is not relevant practically. Therefore, further theorems 1 and 2 do not use this assumption, and rather compare the actual timeblock pattern and the one obtained with>NNLASSO in terms of intersecting neighborhoods.

3.1 Exact timeblock recovery and bounds for sparsity pattern approximation

For any value of the parameter $r > 0$, recall that we defined $\widehat{\beta}(r)$ as the>NNLASSO minimizer (18). Next proposition is closely adapted from [24], and provides, under very specific conditions on the dictionary \mathbf{A} , some range of values of r such that the>NNLASSO minimizer selects only time blocks from P_0 .

Proposition 1 *Assume that for all vectors \mathbf{z} of length $|P_0|$ such that $\mathbf{z} \leq \mathbf{1}_{P_0}$ the following assumption holds:*

$$\mathbf{G}_{P_0, P_0}^{-1} \mathbf{G}_{P_0, P_0} \mathbf{z} < (1 - \eta_0) \mathbf{1}_{P_0} \quad (20)$$

for some $0 < \eta_0 < 1$. If the parameter r is chosen such that

$$r > \max \left\{ \frac{2\alpha}{\eta_0}; \frac{2\sqrt{2}\sigma}{\eta_0} \sqrt{\frac{\log(N - |P_0|)p}{N}} \right\}, \quad (21)$$

then $\widehat{\beta}(r)$ is supported by P_0 with probability tending to 1 as N tends to infinity.

Proof: See Appendix B. ■

Though similar to standard conditions which ensure the consistency of LASSO appearing e.g. in [24], Proposition 1 is of little practical use, since the dictionary \mathbf{A} does not satisfy the latter conditions. In light of (21), we can also remark that a good choice of the sparsity parameter depends on many terms (α, P_0) unknown in practice. This illustrates the need of

further results, since the question arose in this paper is whether and in which measure standard sparse methods could provide sufficiently good results even when theoretical conditions are not met. The two main theorems are based on the following proposition², whose aim is to compare the true sparsity pattern with the NNLASSO one, for convenient choice of r , and adequate block thresholding. The short technical lemma (1), which is a direct consequence of the definition of \mathcal{C} , see (9), will be used in the proof.

Proposition 2 *Define the following threshold value*

$$\eta \triangleq \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{4(2\tau + 1)E_{\max}}, \quad (22)$$

and assume that r satisfies

$$r + \alpha < \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{2 E_{\max}}. \quad (23)$$

Then, there exists $0 \leq \rho \leq 1$ (dependent on r and η , but independent of $\boldsymbol{\beta}$) such that, for all k in P_0 , there is an integer m in $\mathcal{T}_\rho(k)$ so that $\|\widehat{\boldsymbol{\beta}}_m\|_1 \geq \eta$ with probability greater than

$$1 - p \mathfrak{t} \left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{4 E_{\max} \sigma} \right) - p(2\tau + 1) \mathfrak{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right). \quad (24)$$

Conversely, there exists $0 \leq \mu \leq 1$ (dependent on η , but independent from $\boldsymbol{\beta}$ and r) such that, for any k satisfying $\|\widehat{\boldsymbol{\beta}}_k\|_1 \geq \eta$ we have:

$$\Pr(P_0 \cap \mathcal{T}_\mu(k) \neq \emptyset) > 1 - \|\widehat{\boldsymbol{\beta}}_k\|_0 \mathfrak{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right). \quad (25)$$

Proof: See Appendix C. ■

Proposition 2 is of practical interest. Roughly speaking, it states that provided the threshold parameter of the post-processing steps and the sparsity parameter are set accordingly to (22) and (23), then any element of the close to optimal sparsity pattern P_0 has in his close vicinity an integer k so that $\|\widehat{\boldsymbol{\beta}}_k\|_1$ is selected in step 2 of our algorithm, and conversely. Therefore, the latter result closely relates $J(\widehat{\boldsymbol{\beta}})$ after post-processing to P_0 , thus connecting $\widehat{\lambda}(r, \eta)$ to λ_{opt} . It is clear that in practice, the value of α is unknown. However, the results still stands for smaller values of η , such as the one selected in the application section.

3.2 Confidence bounds for counting rate estimators

The two next theorems are strongly based on Proposition 2. They provide computable bounds of confidence intervals for $\widehat{\lambda}(r, \eta) - \lambda_{opt}$, where $\widehat{\lambda}(r, \eta)$ and λ_{opt} were respectively

defined in (19) and (17). Recall that only the blocks of $\widehat{\boldsymbol{\beta}}(r)$ selected by the criterion $\|\widehat{\boldsymbol{\beta}}_k(r)\|_1 > \eta$ are used to estimate the arrival times. In the next theorems, we define for brevity $\widehat{\boldsymbol{\beta}}(r, \eta)$ as the obtained vector after post-processing, that is

$$\widehat{\boldsymbol{\beta}}(r, \eta) \triangleq [\widehat{\boldsymbol{\beta}}_k^T(r) \mathbf{1}(\|\widehat{\boldsymbol{\beta}}_k(r)\|_1 > \eta)]_{0 \leq k \leq N-1}^T$$

and for any subset X of $\{0, 1, \dots, N-1\}$ defined as a union of discrete intervals, we denote by $I(X)$ the number of these intervals.

Theorem 1 *Under the same assumptions and settings as in Proposition 2, and define a_ρ, a_μ as the integers satisfying $V_{a_\rho}(k) = \mathcal{T}_\rho(k)$, $V_{a_\mu}(k) = \mathcal{T}_\mu(k)$, for any integer k in $[\tau, N-1-\tau]$ (to avoid interval truncature). We get that*

$$\widehat{\lambda}(r, \eta) - \lambda_{opt} \leq \lambda(r, \eta) \times \left[1 - \frac{\widehat{T}_{\widehat{M}}/\Delta t}{a_\rho + \max J(\widehat{\boldsymbol{\beta}}(r, \eta))} \frac{I\left(\bigcup_{1 \leq j \leq \widehat{M}} V_{a_\mu}(\widehat{T}_j)\right)}{\widehat{M}} \right] \quad (26)$$

with probability greater than

$$1 - p\mathbf{t} \left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i, j)^{1/2}}{4 E_{\max} \sigma} \right) - p(2\tau + 1)\mathbf{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right) - \|\widehat{\boldsymbol{\beta}}(r)\|_0 \mathbf{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right)$$

Proof: See appendix D. ■

Roughly speaking, Theorem 1 gives a lower bound to $\widehat{\lambda}(r, \eta) - \lambda_{opt}$ in terms of the>NNLASSO sparsity block pattern, making it numerically computable. It is straightforward to see that the term between brackets in the lower bound of (26) is positive. The integers $\widehat{T}_{\widehat{M}}/\Delta t, \max J(\widehat{\boldsymbol{\beta}}(r, \eta))$ are the extremities of the last interval in $J(\widehat{\boldsymbol{\beta}}(r, \eta))$, thus the shorter this interval is the less underestimated λ_{opt} is. Moreover, the term $\widehat{M}^{-1} I\left(\bigcup_{1 \leq j \leq \widehat{M}} V_{a_\mu}(\widehat{T}_j)\right)$ is equal to one as soon as $\widehat{T}_i - \widehat{T}_{i-1} > a_\mu$ for all $1 < i \leq \widehat{M}$, that is when the estimated times are sufficiently spaced. Note that this justifies our choice to define the \widehat{T}_i 's as the beginnings of the open components, since times belonging to such an open component are not likely to have separated a_μ -neighbourhoods. Thus, in the most favorable case, $\widehat{T}_{\widehat{M}} = \Delta t \max J(\widehat{\boldsymbol{\beta}}(r, \eta))$ and the number of components is equal to \widehat{M} , in other words

$$\widehat{\lambda}(r, \eta) - \lambda_{opt} \leq \lambda(r, \eta) \left(1 + \frac{\max J(\widehat{\boldsymbol{\beta}}(r, \eta))}{a_\rho} \right)^{-1},$$

thus showing that $\widehat{\lambda}(r, \eta)$ is close to λ_{opt} with a high probability. The asymptotic study as well as theoretical insights on the lengths of intervals inside the>NNLASSO timeblock pattern, in terms of r and other quantities involved in the problem, are beyond the scope of the present paper. The next theorem provides a computable lower bound for $\widehat{\lambda}(r, \eta) - \lambda_{opt}$.

Theorem 2 *Under the same assumptions and conventions used in Theorem 1, suppose moreover that*

$$(\lambda\Delta t)^2 N a_\rho < 1. \quad (27)$$

Then

$$\widehat{\lambda}(r, \eta) - \lambda_{opt} \geq \widehat{\lambda}(r, \eta) \times \left[1 - \frac{|J(\widehat{\boldsymbol{\beta}}(r, \eta))|}{\widehat{M}} \cdot \frac{\widehat{T}_{\widehat{M}}/\Delta t}{\max J(\widehat{\boldsymbol{\beta}}(r, \eta)) - a_\mu} \right] \quad (28)$$

with probability greater than

$$1 - (\lambda\Delta t)^2 N a_\rho - |P_0| \left[\text{pt} \left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i, j)^{1/2}}{4 E_{\max} \sigma} \right) - (p(2\tau + 1) + \|\widehat{\boldsymbol{\beta}}(r)\|_0) \text{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right) \right].$$

Proof: See appendix D. ■

In the latter, the quotient $\frac{|J(\widehat{\boldsymbol{\beta}}(r, \eta))|}{\widehat{M}}$ in (28) can be interpreted as some average number of active consecutive blocks after thresholding. Note that the probability given here is lower than the one in Theorem 1.

4 Applications

We present in this section results on realistic simulations, which emphasize the effectiveness of the proposed approach when compared to a standard method (comparison to a fixed threshold and estimation of λ by means of the idle times of the detector, see [20]). The proposed algorithm for counting rate estimation is then studied on a real dataset.

4.1 Results on simulations

4.1.1 Experimental settings

The performances of the proposed approach are investigated for 50 points of an homogeneous Poisson process whose intensity λ varies from 0.1 to 0.4, which corresponds to physical activities from 1.10^6 and up to 4.10^6 photons per second when the signal is sampled to 10 MHz. Those numbers are related to high or very high radioactive activities, as mentioned for example in [1]. The energies $\{E_n, n \geq 0\}$ are drawn accordingly to a Gaussian density truncated at 0, with mean 50 and variance 5. We present both results in the case of a good Signal Noise Ratio ($\sigma = 1$), as can be found in Gamma spectrometry applications.

Assuming that we observe N points of the sample signal, the j -th column of the dictionary \mathbf{A} is build accordingly to (7). The grid $\boldsymbol{\theta}$ is taken uniform on $(0, 10]^2$, with subdivision step

0.1. In order to check the robustness of the approach and its practical implementation for real-time instrumentation, the signals are simulated in two different settings:

- for each point of the Poisson process, a shape is taken randomly from the dictionary A ; this case will later on be denoted by (I).
- for each point of the Poisson process, a shifted Gamma is created with randomly chosen parameters θ_1, θ_2 . In our experiments, both parameters are drawn uniformly accordingly to $\theta_1 \sim U([0; 10])$ and $\theta_2 \sim U([0; 2])$ (case denoted by (II))

It is obvious that the standard framework for regression is (I); however, as mentioned earlier, we also want to investigate how the algorithm behaves when the dictionary is not rich enough to cover all the possible shapes, and check the effectiveness of the additional post-processing step introduced in the latter sections. This allows to use the proposed approach on real-world experiments where fast algorithms and small dictionaries for real-time implementations must be used. For one given activity, the estimator is computed 10000 times by means of the proposed method, and by means of the standard method aforementioned, both in (I) and (II) cases. Ideally, the parameters η and r should be chosen accordingly to (22) and (23); however, these bounds are unknown in a real-life experiment, for the radioactive source is generally unknown (and, consequently, so are E_{min} and E_{max}). Both on simulations and real data validations, we found out that taking the parameter $\eta = 3\sigma$, and setting the parameter r so that $\|\mathbf{y} - \mathbf{A}\hat{\boldsymbol{\beta}}(r)\|_2 \leq \sigma\sqrt{N}$ provided a good compromise between sparsity and good ℓ_2 precision.

4.1.2 Simulation results and discussion

Figure 2 represents a portion of the simulated signal in case (II) for $\lambda = 1$, as well as the provided estimation and estimated time arrivals. We can observe that the obtained regressor fits well the incoming signal, and that a careful choice of r_N allows to find most of the arrival times. The boxplots displayed in Figures 3(a) and 3(b) represent the distribution of the estimators of λ (the actual value of λ is displayed in the x -axis) when using the standard method counting rate estimation, and the results obtained by our method are given in Figures 3(c) and 3(d). It can be seen from these results that the proposed algorithm provides an estimator with smaller variance, thus making it more appropriate for counting rate estimation.

The high variance in the standard thresholding method can be easily explained. As λ increases, so does the number of pileups, hence the number of individual pulses and arrival times are underestimated. Both phenomena yield a poor estimate of λ . Regarding the estimator obtained with the proposed algorithm, the results obtained in cases where λ is high (e.g. greater than 0.15) are much better than those of the standard method: we observe a much smaller variance, and for the intensities 0.1 to 0.2 the obtained results are very close to the actual counting rate.

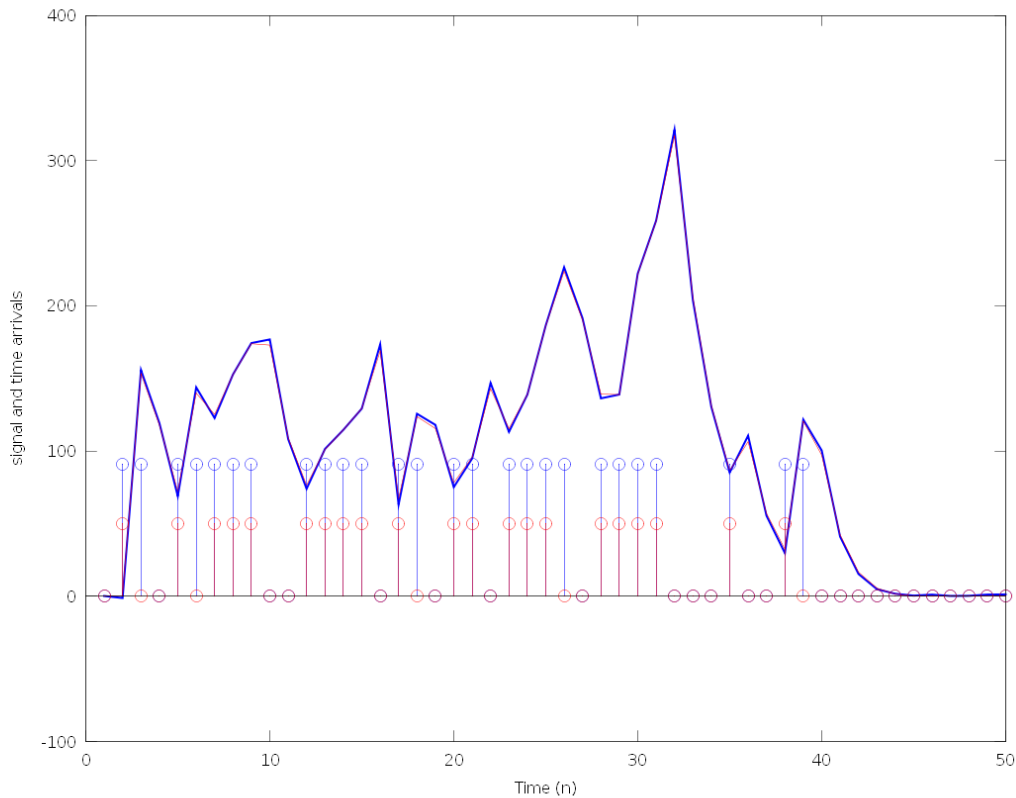
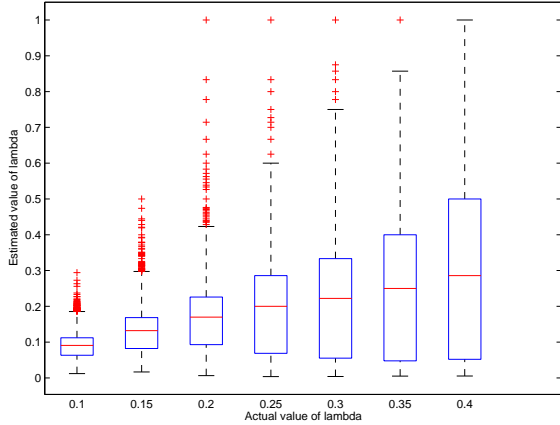


Figure 2: Simulated signal (blue) and NNLASSO regressor (red), with associated time arrivals.

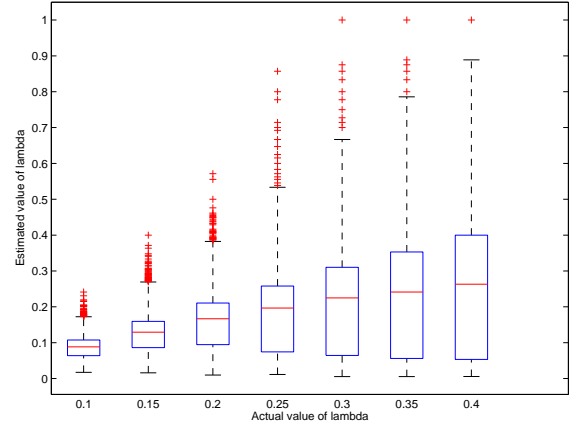
When λ becomes higher, several pulses are likely to start between two consecutive sampling points. Thus, the suggested algorithm may be misled in treating both as one single impulse, which explains why λ is underestimated. However the data is obtained from a sampled signal, therefore the actual λ cannot be well estimated when $\lambda\Delta t$ is too high. Indeed, a better insight is obtained when comparing the values of our estimate with λ_{opt} instead of λ . This is done in Figures 4(a) and 4(b). We observe an almost linear fit between both estimators, in accordance with Theorems 1 and 2, thus showing numerically that the proposed approach provides values similar to λ_{opt} , which is the best estimate we could build from a full knowledge of T_n and of the sampled signal.

4.2 Applications on real data

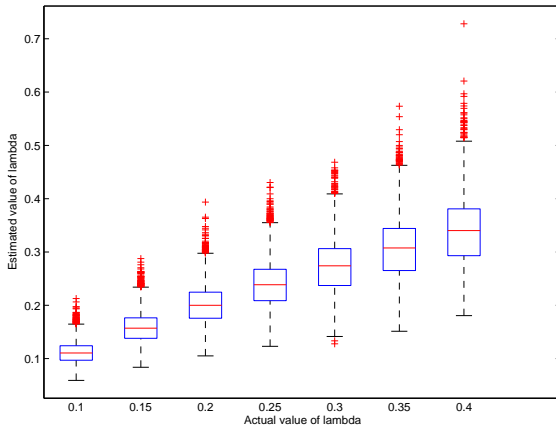
We applied the proposed method for counting rate estimation on real spectrometric data from the ADONIS system described in [3], which is sampled to 10 MHz. The actual activity is 400000 photons per second, which corresponds to an intermediate activity. Figure 5 shows the use of the proposed algorithm on a real dataset. It can be observed from the latter figures



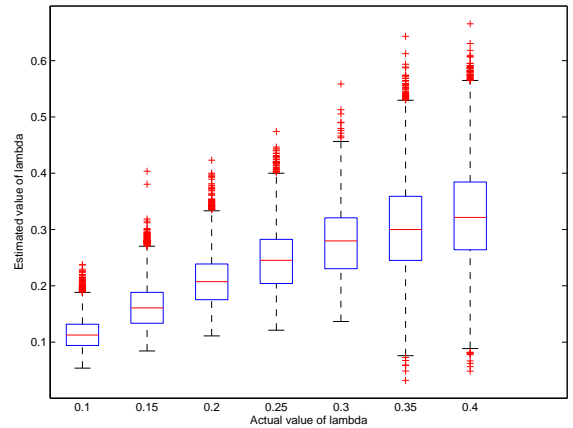
(a) Estimated versus actual values of λ for the standard method - case (I)



(b) Estimated versus actual values of λ for the standard method - case (II)



(c) Estimated versus actual values of λ for the proposed method - case (I)

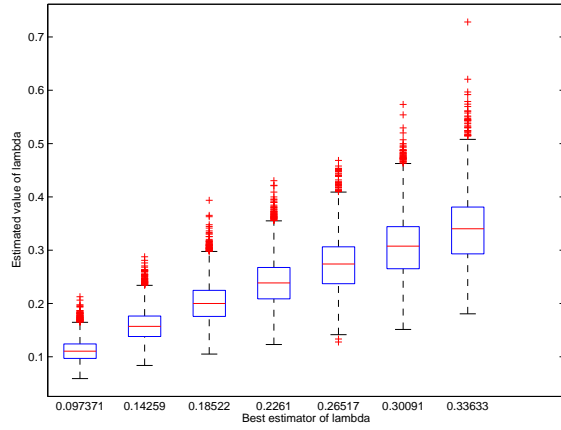


(d) Estimated versus actual values of λ for the proposed method - case (II)

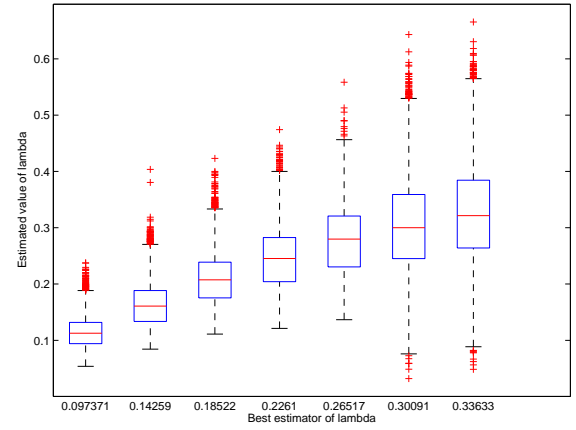
Figure 3: Distribution of the obtained counting rate estimators.

that a very incomplete dictionary is more than sufficient to retrieve the starting points of each individual pulses. However, the post-processing step we suggest in this paper is required to estimate the activity of the radioactive source. The obtained estimated activity is $3.99 \cdot 10^4$, which conforms both to the simulations and the knowledge of the dataset.

We illustrate the importance of the post-processing steps for real data in Figure 6, and compare the performances of NNLASSO, the sparse Bayesian learning (SBL) of [27] and the reweighted ℓ_1 procedure described in [8]. We observe that SBL seems to provide a better “inner-block sparsity”, in the sense that SBL provides active blocks with fewer active coefficients. This is due to the fact that SBL performs usually better than NNLASSO when the columns of \mathbf{A} are highly correlated. However, as it can be seen from the repartition of the coefficients, particularly for the coefficients indexes ranging from 1000 to 3000, singles pulses are always reconstructed by means of several active consecutive blocks of \mathbf{A} , even



(a) Proposed estimate of λ versus λ_{opt} - case (I)



(b) Proposed estimate of λ versus λ_{opt} - case (II)

Figure 4: Comparison of $\hat{\lambda}$ with λ_{opt} .

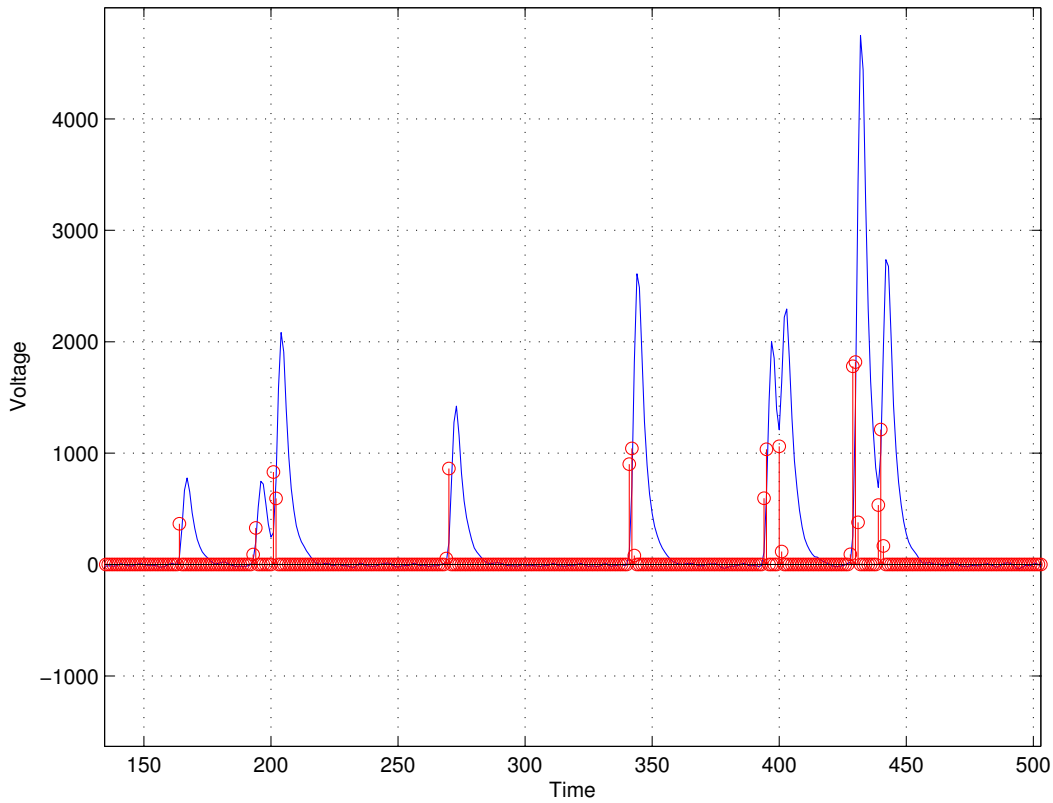
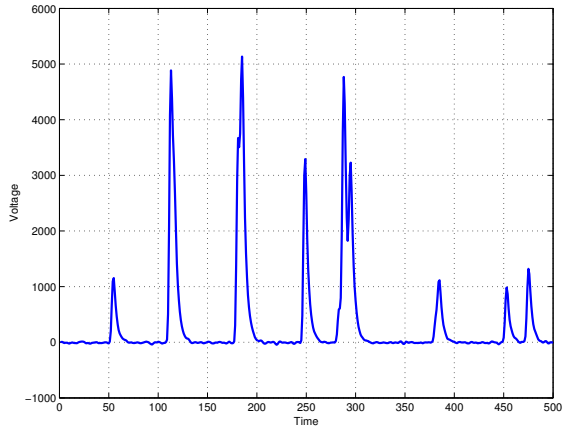
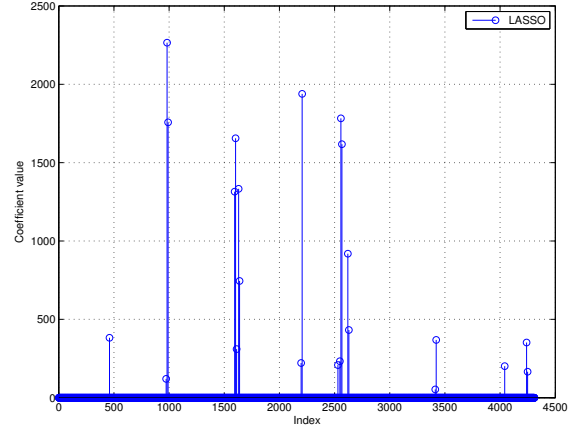


Figure 5: Results on real data: input discrete signal (blue), and active/inactive blocks (red). We observe several well-separated pileups.

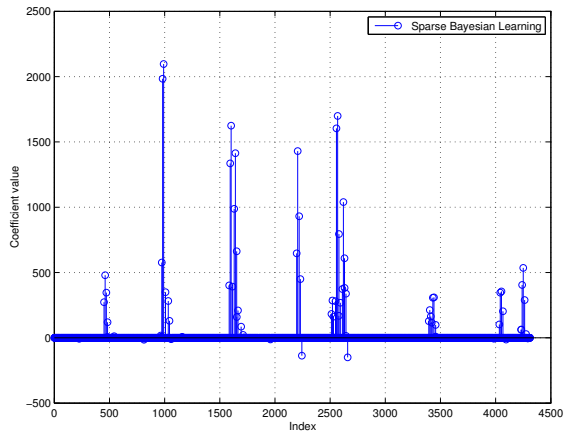
with methods providing sparser solutions than NN-LASSO. This illustrates that in practice, the post-processing steps cannot be circumvented by improving sparsity.



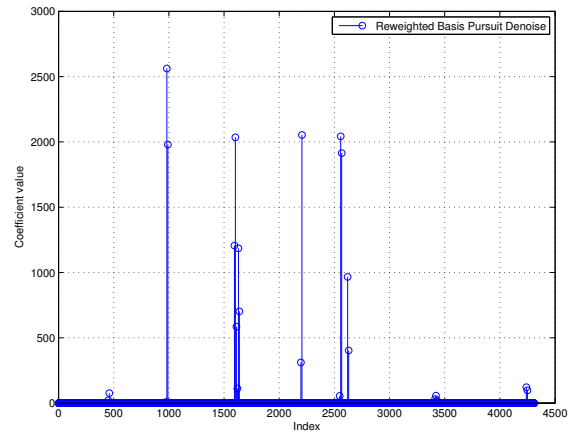
(a) Spectrometric signal



(b) $\hat{\beta}(r)$ obtained with NNLASSO



(c) $\hat{\beta}(r)$ obtained with SBL [27]



(d) $\hat{\beta}(r)$ obtained with Reweighted Basis Pursuit Denoising [8]

Figure 6: Comparison of different sparse regression algorithms on real data

5 Conclusion

We presented in this paper a method based on sparse representation of a sampled spectrometric signal to estimate the activity of an unknown radioactive source. Based on a crude dictionary, we suggest a post-processed variation of a non-negative LASSO to estimate the number of individual electrical pulses and their arrival times. Results on simulations and real data both emphasize the efficiency of the method, and the small size of the dictionary make the implement for real-time applications accessible. It was theoretically shown that although the standard conditions are not met *per se* for NNLASSO to estimate the actual P_0 , we can derive some conditions which guarantee that the number of individual pulses and arrival times are well estimated nevertheless. This is made possible by the fact that we do not wish to reconstruct the input signal, but rather find some partial information. Further aspects should focus on the joint estimation of λ and the energy distribution, as

well as the estimation of the activity in a nonhomogeneous case, and should appear in future contributions.

A Technical lemmas

Lemma 1 *If $\mathbf{A}\boldsymbol{\beta} \in \mathcal{C}$ then for all index $m \in J(\boldsymbol{\beta})$ we have:*

$$\|\boldsymbol{\beta}_m\|_1 \leq \frac{E_{\max}}{\min_{i,j} \mathbf{G}(i,j)^{1/2}}$$

and

$$\min_{i,j} \mathbf{G}(i,j)^{1/2} \frac{E_{\min}^2}{E_{\max}} \leq \|\mathbf{G}\boldsymbol{\beta}_m\|_{\infty}$$

Proof: For all $m \in J(\boldsymbol{\beta})$ we have $E_{\min}^2 \leq \boldsymbol{\beta}_m^T \mathbf{G} \boldsymbol{\beta}_m \leq E_{\max}^2$, and on the other hand

$$\min_{i,j} \mathbf{G}(i,j) \cdot \|\boldsymbol{\beta}_m\|_1^2 \leq \boldsymbol{\beta}_m^T \mathbf{G} \boldsymbol{\beta}_m \leq E_{\max}^2$$

so the first assertion follows. The second is proved by using the previous result in the following way

$$\begin{aligned} E_{\min}^2 &\leq \boldsymbol{\beta}_m^T \mathbf{G} \boldsymbol{\beta}_m \leq \|\boldsymbol{\beta}_m\|_1 \|\mathbf{G}\boldsymbol{\beta}_m\|_{\infty} \\ &\leq \frac{E_{\max}}{\min_{i,j} \mathbf{G}(i,j)^{1/2}} \|\mathbf{G}\boldsymbol{\beta}_m\|_{\infty}, \end{aligned}$$

This concludes the proof. ■

Lemma 2 *Suppose a homogeneous Poisson process of intensity λ is observed in the interval $[0, T]$, and let $\delta > 0$ such that $\lambda^2 T \delta < 1$. The probability that all interarrival times are greater than δ is bounded from below by $1 - \lambda^2 T \delta$.*

Proof: We compute the probability that one interarrival time is smaller than δ . Denote by T_n the n -th point of the process sample path, and by N_t the number of points on $[0, t]$. It is known (see e.g. [2]) that

$$\begin{aligned} f_{((T_1, \dots, T_n) | N_T = n)}(u_1, \dots, u_n) \\ = f_{(U_{(1)}, U_{(2)}, \dots, U_{(n)})}(u_1, u_2, \dots, u_n) = \frac{n!}{T^n} \mathbf{1}_{0 = u_0 \leq u_1 \leq \dots \leq u_n \leq T}, \end{aligned}$$

where $\{U_{(i)}, i = 1 \dots n\}$ are the order statistics of n independent random variables uniformly distributed on $[0, T]$. We get for $n \geq 2$ and $2 \leq i \leq n$ that

$$P(T_i - T_{i-1} \leq \delta | N_T = n) = \frac{n!}{T^n} \text{Vol}(\Omega_i)$$

where $\Omega_i \triangleq \{0 \leq u_1 \leq \dots \leq u_n \leq T; u_i - u_{i-1} \leq \delta\}$ and Vol denotes the volume of the latter space. For all $1 \leq k \leq n$ we set $\text{incr}_k = u_k - u_{k-1}$ (and $u_0 = 0$), so it is equivalent to write

$$\Omega_i = \{\text{incr}_k \geq 0, 1 \leq k \leq n;$$

$$\text{incr}_i \leq \delta \text{ and } \sum_{k=1}^n \text{incr}_k \leq T\}$$

We have now the decomposition of Ω_i along the (disjoint) slices defined by $\text{incr}_i = t$, $0 \leq t \leq \delta$:

$$\Omega_i = \bigcup_{0 \leq t \leq \delta} \tilde{\Omega}_i(t);$$

$$\tilde{\Omega}_i(t) \triangleq \{\text{incr}_j \geq 0, \text{ for all } j \neq i, \text{incr}_i = t; \sum_{j \neq i} \text{incr}_j \leq T - t\}$$

Integrating now along the variable t we obtain:

$$\begin{aligned} \text{Vol}(\Omega_i) &= \int_0^\delta \text{Vol}(\tilde{\Omega}_i(t)) dt \\ &= \int_0^\delta \frac{(T-t)^{n-1}}{(n-1)!} dt = \frac{1}{n!} [T^n - (T-\delta)^n] \end{aligned}$$

Hence we have $P(\{T_i - T_{i-1} \leq \delta\} | N_t = n) = 1 - (1 - \frac{\delta}{T})^n$ therefore we get for all $n \geq 2$ that

$$\begin{aligned} P(T_i - T_{i-1} \leq \delta \text{ for some } 2 \leq i \leq n | N_t = n) \\ \leq (n-1) \left[1 - \left(1 - \frac{\delta}{T}\right)^n \right] \end{aligned}$$

We can of course write that the same probability is equal to 0 as $n = 0, 1$. Due to the equality $\sum_{n \geq 2} (n-1) \frac{x^n}{n!} = (x-1) \exp(x) + 1$, we get by conditioning that the probability that one interarrival time is smaller than δ is not greater than

$$\begin{aligned} \exp(-\lambda T) \sum_{n \geq 2} \frac{\lambda^n T^n}{n!} (n-1) \left[1 - \left(1 - \frac{\delta}{T}\right)^n \right] \\ = \lambda T - [\lambda(T-\delta) - 1] \exp(-\lambda \delta) - 1 \end{aligned}$$

The lemma follows from Taylor inequality. ■

B Proof of Proposition 1

In its very essence, the proof follows [24] with mild modifications. We introduce $\hat{\beta}_0(r)$ as the minimizer of the NNLASSO problem with sparsity parameter $r > 0$ under the additional

constraint $J(\widehat{\boldsymbol{\beta}}_0(r)) \subseteq P_0$. The aim is to prove that, under the conditions stated in the result, this vector is a global minimizer of the unconstrained problem. Due to (18), the KKT conditions with the additional constraint are:

$$\begin{aligned} \frac{\mathbf{A}_{P_0}^T}{N} \left[\mathbf{A}_{P_0} \boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon} - \mathbf{A}_{P_0} \widehat{\boldsymbol{\beta}}_0(r) \right] &= r \mathbf{z}_{P_0} \\ \mathbf{z}_{P_0} &\in (-\infty, 1]^{p|P_0|} \end{aligned} \quad (29)$$

We deduce from (29) the equality $\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_0(r) = \mathbf{G}_{P_0, P_0}^{-1} \left[r \mathbf{z}_{P_0} - \frac{1}{N} \mathbf{A}_{P_0}^T (\boldsymbol{\delta} + \boldsymbol{\varepsilon}) \right]$. Using this equality, $\widehat{\boldsymbol{\beta}}_0(r)$ will be a global minimizer of NNlasso as soon as

$$\frac{\mathbf{A}_{P_0}^T}{N} \left[\mathbf{A}_{P_0} \mathbf{G}_{P_0, P_0}^{-1} \left(r \mathbf{z}_{P_0} - \frac{\mathbf{A}_{P_0}^T (\boldsymbol{\delta} + \boldsymbol{\varepsilon})}{N} \right) + \boldsymbol{\delta} + \boldsymbol{\varepsilon} \right] < r \mathbf{1}_{\overline{P_0}},$$

or, equivalently, when

$$\begin{aligned} \left[\frac{\mathbf{A}_{\overline{P_0}}}{\sqrt{N}} - \frac{\mathbf{A}_{P_0}}{\sqrt{N}} \mathbf{G}_{P_0, P_0}^{-1} \mathbf{G}_{P_0, \overline{P_0}} \right]^T \frac{(\boldsymbol{\delta} + \boldsymbol{\varepsilon})}{\sqrt{N}} \\ < r \left[\mathbf{1}_{\overline{P_0}} - \mathbf{G}_{\overline{P_0}, P_0} \mathbf{G}_{P_0, P_0}^{-1} \mathbf{z}_{P_0} \right]. \end{aligned} \quad (30)$$

For convenience, we now define

$$\mathbf{H} \triangleq \frac{\mathbf{A}_{\overline{P_0}}}{\sqrt{N}} - \frac{\mathbf{A}_{P_0}}{\sqrt{N}} \mathbf{G}_{P_0, P_0}^{-1} \mathbf{G}_{P_0, \overline{P_0}},$$

and denote by H_i the i -th column of \mathbf{H} . Note that \mathbf{H} can be rewritten as

$$\mathbf{H} = \left(I - \left(\frac{\mathbf{A}_{P_0}}{\sqrt{N}} \right) \mathbf{G}_{P_0, P_0}^{-1} \left(\frac{\mathbf{A}_{P_0}}{\sqrt{N}} \right)^T \right)^T \frac{1}{\sqrt{N}} \mathbf{A}_{\overline{P_0}},$$

showing that the columns of \mathbf{H} are the projections of the normalized columns of $\mathbf{A}_{\overline{P_0}}$ onto the orthogonal complement of the columns of \mathbf{A}_{P_0} . It thus follows that all columns the H_i 's have normalized ℓ_2 -norm bounded by 1 since this is true for $\mathbf{A}_{\overline{P_0}}$. Due to assumption (20), it is sufficient that the following condition holds to get (30):

$$\max_i \frac{1}{\sqrt{N}} H_i^T (\boldsymbol{\delta} + \boldsymbol{\varepsilon}) < r \eta_0 ; \quad (31)$$

we now use the fact that the random variable $H_i^T \boldsymbol{\varepsilon}$ is Gaussian of variance less than σ^2 , consequently:

$$\begin{aligned} &P \left(\max_i \frac{1}{\sqrt{N}} H_i^T \boldsymbol{\varepsilon} \geq \frac{r \eta_0}{2} \right) \\ &\leq \sum_i P \left(H_i^T \boldsymbol{\varepsilon} \geq \frac{\sqrt{N} r \eta_0}{2} \right) \\ &\leq p(N - |P|) \mathfrak{t} \left(\frac{\sqrt{N} r \eta_0}{2 \sigma} \right) \end{aligned} \quad (32)$$

In order to make (32) tend to 0, we need that

$$r \geq \frac{2\sqrt{2}\sigma}{\eta_0} \sqrt{\frac{\log(N - |P|)p}{N}}.$$

Now we have also $\frac{1}{\sqrt{N}}H_i^T \boldsymbol{\delta} \leq \alpha$ by Cauchy-Schwarz inequality, and we have $\alpha < r\eta_0/2$ as soon as $r > 2\alpha/\eta_0$. The result follows.

C Proof of Proposition 2

We keep the same notations as in Appendix B. Let $r > 0$ be any sparsity parameter, and $\widehat{\boldsymbol{\beta}}(r)$ be the corresponding NNLASSO regressor, denoted shortly by $\widehat{\boldsymbol{\beta}}$ in the rest of this proof. The KKT conditions, combined with the inequality derived from Cauchy-Schwarz $\frac{1}{N}\mathbf{A}^T\boldsymbol{\delta} \leq \alpha\mathbf{1}_{pN}$, yields

$$\frac{\mathbf{A}^T}{N} [\mathbf{A}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \boldsymbol{\delta} + \boldsymbol{\varepsilon}] \leq r\mathbf{1}_{pN},$$

or, equivalently,

$$\frac{\mathbf{A}^T\mathbf{A}}{N}\boldsymbol{\beta} - (r + \alpha)\mathbf{1}_{pN} + \frac{\mathbf{A}^T\boldsymbol{\varepsilon}}{N} \leq \frac{\mathbf{A}^T\mathbf{A}}{N}\widehat{\boldsymbol{\beta}}. \quad (33)$$

For some k in P_0 , it follows from (33) that

$$\|\mathbf{G}\boldsymbol{\beta}_k\|_\infty - (r + \alpha) - \frac{1}{N}\|\mathbf{A}_k^T\boldsymbol{\varepsilon}\|_\infty \leq \|\mathbf{G}_{k,V_\tau(k)}\widehat{\boldsymbol{\beta}}_{V_\tau(k)}\|_\infty.$$

Define $0 \leq \rho \leq 1$; since all the coefficients of the Gram matrices considered are bounded by 1, we have

$$\begin{aligned} \|\mathbf{G}\boldsymbol{\beta}_k\|_\infty - (r + \alpha) - \frac{1}{N}\|\mathbf{A}_k^T\boldsymbol{\varepsilon}\|_\infty &\leq (1 - \rho) \sum_{l \in \mathcal{J}_\rho(k)} \|\widehat{\boldsymbol{\beta}}_l\|_1 + \rho\|\widehat{\boldsymbol{\beta}}_{V_\tau(k)}\|_1. \end{aligned} \quad (34)$$

We can say now two things: first, when $r + \alpha < \|\mathbf{G}\boldsymbol{\beta}_k\|_\infty$, we have

$$\begin{aligned} \Pr\left(\frac{1}{N}\|\mathbf{A}_k^T\boldsymbol{\varepsilon}\|_\infty > \frac{\|\mathbf{G}\boldsymbol{\beta}_k\|_\infty - (r + \alpha)}{2}\right) &< p\mathbf{t}\left(\sqrt{N} \frac{\|\mathbf{G}\boldsymbol{\beta}_k\|_\infty - (r + \alpha)}{2\sigma}\right); \end{aligned}$$

secondly, for any column A_n of \mathbf{A} which is active in $\widehat{\boldsymbol{\beta}}$, (33) reduces to the equality $\frac{A_n^T}{N}(\mathbf{A}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}) - r = \frac{A_n^T}{N}\mathbf{A}\widehat{\boldsymbol{\beta}}$, and since $\widehat{\boldsymbol{\beta}}$ has non negative entries it follows that $\widehat{\boldsymbol{\beta}}_n$ is smaller than $\frac{1}{N}A_n^T(\mathbf{A}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}) - r$; now summing up these inequalities overall such $n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_{V_\tau(k)})$ yields

$$\begin{aligned} \|\widehat{\boldsymbol{\beta}}_{V_\tau(k)}\|_1 &\leq \frac{1}{N}\|\mathbf{A}_{V_\tau(k)}^T\mathbf{A}\boldsymbol{\beta}\|_1 + \\ &\sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_{V_\tau(k)})} \left(\frac{1}{N}A_n^T\boldsymbol{\varepsilon} - r + \alpha\right). \end{aligned} \quad (35)$$

Since every $n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_{V_\tau(k)})$ can be expressed as $n = pl + s$ with l in $V_\tau(k) \cap J(\widehat{\boldsymbol{\beta}})$ and $s \leq p$, we have

$$\Pr\left(\sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_{V_\tau(k)})} \left(\frac{1}{N} A_n^T \boldsymbol{\varepsilon} - r + \alpha\right) > 0\right) \leq p(2\tau + 1) \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right),$$

thus (35) yields

$$\Pr\left(\|\widehat{\boldsymbol{\beta}}_{V_\tau(k)}\|_1 \leq \frac{1}{N} \|\mathbf{A}_{V_\tau(k)}^T \mathbf{A} \boldsymbol{\beta}\|_1\right) > 1 - p(2\tau + 1) \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right). \quad (36)$$

So far, we showed that

$$\frac{1}{(1 - \rho) |\mathcal{T}_\rho(k)|} \left[\frac{\|\mathbf{G} \boldsymbol{\beta}_k\|_\infty - (r + \alpha)}{2} - \rho \frac{1}{N} \|\mathbf{A}_{V_\tau(k)}^T \mathbf{A} \boldsymbol{\beta}\|_1 \right] \leq \max_{l \in \mathcal{T}_\rho(k)} \|\widehat{\boldsymbol{\beta}}_l\|_1 \quad (37)$$

with probability greater than

$$1 - p \mathfrak{t}\left(\sqrt{N} \frac{\|\mathbf{G} \boldsymbol{\beta}_k\|_\infty - (r + \alpha)}{2\sigma}\right) - p(2\tau + 1) \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right).$$

Now, choosing r accordingly to (23) and using Lemma 1 yields

$$\frac{\|\mathbf{G} \boldsymbol{\beta}_k\|_\infty - (r + \alpha)}{2} > \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{4 E_{\max}},$$

and if ρ is taken equal to 0 the LHS of (37) is greater than η as chosen in (22). On the other hand, the term $\frac{1}{N} \|\mathbf{A}_{V_\tau(k)}^T \mathbf{A} \boldsymbol{\beta}\|_1$ can be bounded by means of Lemma 1 as follows:

$$\begin{aligned} \frac{1}{N} \|\mathbf{A}_{V_\tau(k)}^T \mathbf{A} \boldsymbol{\beta}\|_1 &\leq \sum_{l \in V_\tau(k)} \sum_{m \in V_\tau(\{l\})} \|\mathbf{G}_{\{l\}, \{m\}} \boldsymbol{\beta}_m\|_1 \\ &\leq p \sum_{l \in V_\tau(k)} \sum_{m \in V_\tau(\{l\})} (\max_{i,j} \mathbf{G}_{\{l\}, \{m\}}(i,j)) \|\boldsymbol{\beta}_m\|_1 \\ &\leq \frac{(2\tau + 1) \mathfrak{G} p E_{\max}}{\min_{i,j} \mathbf{G}(i,j)^{1/2}}. \end{aligned}$$

Therefore, if (23) holds, we introduce

$$C_{r,\rho} \triangleq \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{E_{\max}} - \alpha - r - \frac{(2\tau + 1) \mathfrak{G} p E_{\max}}{\min_{i,j} \mathbf{G}(i,j)^{1/2}} \rho$$

and

$$\rho_r \triangleq \sup \left\{ \rho \in [0, 1] ; \eta \leq \frac{C_{r,\rho}}{(1-\rho)|\mathcal{T}_\rho(k)|} \right\}, \quad (38)$$

thus inequality (37) shows that if $k \in P_0$ then there exists $m \in \mathcal{T}_{\rho_r}(k)$ such that

$$\Pr(\|\widehat{\boldsymbol{\beta}}_m\|_1 \geq \eta) > 1 - pt \left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{4 E_{\max} \sigma} \right) - (2\tau + 1)pt \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right),$$

which completes the proof of the first part of Proposition 2. The converse part is proved using a similar argument as follows. For any k , we denote by $\text{Supp}(\widehat{\boldsymbol{\beta}}_k)$ the set of indices of the non-zero entries in the vector $\widehat{\boldsymbol{\beta}}_k$. For k chosen such that $\|\widehat{\boldsymbol{\beta}}_k\|_1 \geq \eta$, then for all n in $\text{Supp}(\widehat{\boldsymbol{\beta}}_k)$ we have $\frac{1}{N} A_n^T \mathbf{A} \boldsymbol{\beta} = r + \frac{1}{N} A_n^T \mathbf{A} \widehat{\boldsymbol{\beta}} - \frac{1}{N} A_n^T (\boldsymbol{\delta} + \boldsymbol{\varepsilon})$. Considerations analog to those yielding (35),(36) allow to obtain:

$$\begin{aligned} \sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \frac{1}{N} A_n^T \mathbf{A} \boldsymbol{\beta} &\geq \sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \left[r - \alpha - \frac{1}{N} A_n^T \boldsymbol{\varepsilon} \right] \\ &+ \sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \frac{1}{N} A_n^T \mathbf{A} \widehat{\boldsymbol{\beta}}_k \end{aligned}$$

hence

$$\Pr \left(\sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \frac{1}{N} A_n^T \mathbf{A} \boldsymbol{\beta} > \min_{i,j} \mathbf{G}(i,j) \|\widehat{\boldsymbol{\beta}}_k\|_0 \eta \right) > 1 - \|\widehat{\boldsymbol{\beta}}_k\|_0 t \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right). \quad (39)$$

Now let ρ in $[0, 1]$, we can write similarly to (34)

$$\begin{aligned} &\sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \frac{1}{N} A_n^T \mathbf{A} \boldsymbol{\beta} \\ &= \sum_{n \in \text{Supp}(\widehat{\boldsymbol{\beta}}_k)} \frac{1}{N} A_n^T \mathbf{A} (\boldsymbol{\beta}_{\mathcal{T}_\rho(k)} + \boldsymbol{\beta}_{\overline{\mathcal{T}_\rho(k)}}) \\ &< \|\widehat{\boldsymbol{\beta}}_k\|_0 (\|\boldsymbol{\beta}_{\mathcal{T}_\rho(k)}\|_1 + \rho \|\boldsymbol{\beta}_{\overline{\mathcal{T}_\rho(k)} \cap V_\tau(k)}\|_1). \end{aligned} \quad (40)$$

Since by Lemma 1 we have

$$\|\boldsymbol{\beta}_{\overline{\mathcal{T}_\rho(k)} \cap V_\tau(k)}\|_1 \leq \frac{|\overline{\mathcal{T}_\rho(k)} \cap V_\tau(k)| E_{\max}}{\min_{i,j} \mathbf{G}(i,j)^{1/2}},$$

(39) leads to introduce the next correlation level, similarly to (38):

$$\begin{aligned} \mu \triangleq \sup \left\{ \rho \in [0, 1] ; \rho |\overline{\mathcal{T}_\rho(k)} \cap V_\tau(k)| \right. \\ \left. \leq \frac{\eta \cdot \min_{i,j} \mathbf{G}(i,j)^{3/2}}{E_{\max}} \right\}. \end{aligned}$$

Thus, (39) implies that whenever $\|\widehat{\boldsymbol{\beta}}_k\|_1 \geq \eta$, one has

$$\Pr(P_0 \cap \mathcal{T}_\mu(k) \neq \emptyset) > 1 - \|\widehat{\boldsymbol{\beta}}_k\|_0 \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right),$$

which completes the proof.

D Proofs of Theorem 1 and Theorem 2

We first prove Theorem 1 : by Proposition 2, we have

$$\begin{aligned} \Pr\left(\max P_0 - a_\rho \leq \max J(\widehat{\boldsymbol{\beta}}(r, \eta))\right) &> 1 \\ &- p \mathfrak{t}\left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i, j)^{1/2}}{4 E_{\max} \sigma}\right) \\ &- p(2\tau + 1) \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right) \end{aligned}$$

and moreover each of the maximal distinct intervals contained in $\bigcup_{1 \leq j \leq \widehat{M}} V_{a_\mu}(\widehat{T}_j)$ intersects P_0 under probability greater than

$$\begin{aligned} 1 - \sum_{k \in J(\widehat{\boldsymbol{\beta}}(r, \eta))} \|\widehat{\boldsymbol{\beta}}_k(r)\|_0 \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right) \\ \geq 1 - \|\widehat{\boldsymbol{\beta}}(r)\|_0 \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right) \end{aligned}$$

thus under the same probability we have $I(\bigcup_{1 \leq j \leq \widehat{M}} V_{a_\mu}(\widehat{T}_j)) \leq |P_0|$, and combining these two results we obtain the lower bound

$$\lambda_{opt} \geq \frac{1}{\Delta t} \frac{I\left(\bigcup_{1 \leq j \leq \widehat{M}} \mathcal{T}_\mu(\widehat{T}_j)\right)}{a_\rho + \max J(\widehat{\boldsymbol{\beta}}(r, \eta))},$$

which is equivalent to (26) after factorization by $\widehat{\lambda}(r, \eta)$.

Theorem 2 is proved in a similar manner. Using again Proposition. (2), we have

$$\begin{aligned} \Pr(\max J(\widehat{\boldsymbol{\beta}}(r, \eta)) - a_\mu \leq \max P_0) \\ > 1 - \|\widehat{\boldsymbol{\beta}}_{\max J(\widehat{\boldsymbol{\beta}}(r, \eta))}\|_0 \mathfrak{t}\left(\sqrt{N} \frac{r - \alpha}{\sigma}\right). \end{aligned}$$

Using the same argument as above, we also get that

$$|J(\widehat{\boldsymbol{\beta}}(r, \eta))| \geq I\left(\bigcup_{k \in P_0} V_{a_\rho}(k)\right)$$

with probability greater than

$$1 - |P_0| \left[pt \left(\sqrt{N} \frac{E_{\min}^2 \min_{i,j} \mathbf{G}(i,j)^{1/2}}{4 E_{\max} \sigma} \right) - p(2\tau + 1) \mathbf{t} \left(\sqrt{N} \frac{r - \alpha}{\sigma} \right) \right]$$

Now the a_ρ -neighbourhoods $V_{a_\rho}(k)$ for $k \in P_0$ are all disjoint with probability bounded as in Lemma 2 when taking $T = N\Delta t$; the result follows

References

- [1] ANSI. *American National Standard for Calibration and Use of Germanium Spectrometers for the Measurement of Gamma-Ray Emission Rates of Radionuclides*. American National Standards Institute, 1999.
- [2] F. Baccelli and P. Brémaud. *Elements of Queueing Theory*. Springer, 2002.
- [3] E. Barat, T. Dautremer, L. Laribiere, J. Lefevre, T. Montagu, and J.-C. Trama. Adonis : a new system for high count rate hpge gamma spectrometry. In *Nuclear Science Symposium Conference Record, 2006. IEEE*, volume 2, pages 955–958, 2006.
- [4] F. Belli, B. Esposito, D. Marocco, M. Rivaa, Y. Kaschuck, and G. Bonheure. A method for digital processing of pile-up events in organic scintillators. *Nuclear Instruments and Methods, Phys. Res. A*, 595(2):512–519, 2008.
- [5] Z. Ben-Haim, Y. C Eldar, and M. Elad. Coherence-Based performance guarantees for estimating a sparse vector under random noise. *IEEE Transactions on Signal Processing*, 58(10):5030–5043, October 2010.
- [6] P. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.
- [7] Q. Bristow. *Theoretical and experimental investigations of coincidences in Poisson distributed pulse trains and spectral distortion caused by pulse pileup*. PhD thesis, Carleton University, Ottawa, Canada, 1990.
- [8] E. Candes, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, December 2008.
- [9] S. Chen, D. Donoho, , and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.

- [10] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *Information Theory, IEEE Transactions on*, 47(7):2845–2862, 2001.
- [11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.
- [12] G. F Knoll. *Radiation Detection and Measurement*. Wiley, 2nd edition, 1989.
- [13] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-To Approach*. Springer, 2nd rev edition, January 1994.
- [14] K. Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of Statistics*, 2:90–102, 2008.
- [15] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.
- [16] N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics*, 37(1):246–270, 2009.
- [17] 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: Accelerators, Spectrometers, Detectors and Associated Equipment*, 608(1):163–168, September 2009.
- [18] R. Novak and M. Vencelj. Gauss-Seidel iterative method as a Real-Time pileup solver of scintillator pulses. *IEEE Transactions in Nuclear Science*, 56(6):3680–3687, 2009.
- [19] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.
- [20] T. Trigano, A. Souloumiac, T. Montagu, F. Roueff, and E. Moulines. Statistical pileup correction method for HPGe detectors. *IEEE Transactions on Signal Processing*, 55(10):4871–4881, oct. 2007.
- [21] S. Van de Geer and P. Bühlmann. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.
- [22] M. Vencelj, K. Bucar, R. Novak, and H. J Wortche. Event by event pile-up compensation in digital timestamped calorimetry. *Nuclear Instruments and Methods, Phys. Res. A*, 607(3):581–586, 2009.
- [23] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, June 2002.

- [24] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theor.*, 55(5):2183–2202, 2009.
- [25] G. P Westphal. Real-Time correction for counting losses in nuclear pulse spectroscopy. *Journal of Radioanalytical and Nuclear Chemistry*, 70:387–397, 1982.
- [26] G.P. Westphal. Instrumental correction of counting losses in nuclear pulse spectroscopy. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 10-11(Part 2):1047–1050, May 1985.
- [27] D. Wipf and S. Nagarajan. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. *Journal of Selected Topics in Signal Processing*, 4(2):317–329, 2010.
- [28] David P. Wipf. Sparse estimation with structured dictionaries. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems*, pages 2016–2024, 2011.
- [29] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, B*, 68(1):49–67, 2006.
- [30] C. H Zhang and J. Huang. The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, 36(4):1567–1594, 2008.
- [31] P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541—2563, November 2006.