Sparse regression algorithm for activity estimation in $\gamma$ spectrometry

Y. Sepulcre, T. Trigano and Y. Ritov

February 28, 2012

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 the measurements and introduces a significant bias to the standard estimators of the source activities. 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 version of the Least Absolute Shrinkage and Selection Operator (LASSO) to estimate the photon arrival times, and derive necessary conditions which guarantee that the arrival times will be selected. The 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 [11] and [12] 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]

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

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

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

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. By analogy with queuing models, we call \( \{W(t), t \geq 0\} \) the workload process. The pulse shapes \( \{\Phi_k(s), k \geq 1\} \) are assumed to belong to a parametric family of functions \( \Gamma_\Theta, \Theta \subset \mathbb{R}^n \). The restriction of the workload process 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 [16] 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 [24, 25]. 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 [17] 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 states (where no photon impinges to the detector) and busy states (when we observe
electrical pulses) of the detector. However, when the activity of the radioactive source is too high, we observe very few transitions from busy to idle states, 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 \( \Gamma_\Theta \)), and to estimate both the workload 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 workload must be sparse as well. Since the seminal papers [19] and [8], 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 [10] suggest efficient algorithms yielding estimators with good statistical properties, thus making sparse regression estimators a possible option for real-time processing. Indeed, LASSO provides a sparse solution close to the real signal for the \( \ell_2 \)-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 [28] and [14]. Both papers introduced independently the so-called \emph{irrepresentability} condition as a necessary condition for selection consistency. More recently, [23] developed the conditions under which the irrepresentability condition is also a sufficient one. We also refer to [27], [15], [21] and for recent results on consistency in the \( \ell_2 \)-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 \( \ell_2 \) sense should be used for energy spectrum reconstruction.

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 LASSO. This sparse regressor is briefly recalled, as well as its main properties regarding consistency in selection. We 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. Section 4 illustrates on some applications the effectiveness of the proposed approach, both on simulations and real data. 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

Assume that we observe a signal uniformly sampled on \( \mathcal{T} \triangleq \{0 = t_0, t_1, t_2, \ldots, t_{N-1} = T\} \). We further on denote the subdivision step \( T/N \) by \( \delta t \), and by \( y \triangleq [y_0, y_1, \ldots, y_{N-1}]^T \) the
resulting observations. The observations stem from a discrete version of the generalized shot-noise process [18]:

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

where \{T_n, 1 \leq n \leq M\} is the sample path of an homogeneous Poisson process with constant intensity \( \lambda \), \{E_n, 1 \leq n \leq M\} is a sequence of independent and identically distributed (iid) random variables representing the photon 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 \( \sigma^2 \) representing the additional noise of the input signal. We further denote by \( \overline{y} = (\overline{y}_1, \cdots, \overline{y}_{N-1}) \) the noise-free part of \( y \), in other words \( \overline{y}_i = \sum_{n=1}^{M} E_n \Phi_n(t_i - T_n) \) for all \( i = 0, \cdots, N - 1 \).

For most spectrometers, an electrical pulse created by a single photon has a characteristic shape, usually in most detectors a rapid growth created by the charge collection and an exponential decay as the charges migrate to the detector. We therefore make the following assumptions:

**Assumption 1** the functions \( \Phi_n \) belong to a fixed positive span of truncated Gamma functions

\[ \Gamma_{\theta}(t) = c_{\theta} \cdot t^{\theta_1} \cdot e^{-\theta_2 t} \cdot 1(0 \leq t \leq \tau), \tag{3} \]

where \( 0 < \tau < T \), \( \theta = (\theta_1, \theta_2) \in \mathbb{R}_+^2 \) and \( c_{\theta} \) is a normalizing constant so that \( \frac{1}{N} \sum_{i=0}^{N-1} \Gamma_{\theta}(t_i)^2 = 1 \) for all \( n \).

**Assumption 2** The random variables \{\( E_n, 1 \leq n \leq M \)\} are bounded by positive and known constants:

\[ 0 < E_{\text{min}} \leq E_n \leq E_{\text{max}}, \text{ for all } n. \tag{4} \]

### 2.2 Overview of the estimation procedure

In the following, the cardinality of a set \( A \) is denoted by \( |A| \). Our objective is to estimate \( \lambda \) given \( y \). 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 \( \lambda \). Therefore, \( \lambda \) can be consistently estimated by

\[ \lambda_e \triangleq \frac{M}{T_M}. \tag{5} \]

However, \( y \) is a discrete-time signal, therefore \( \lambda_e \) cannot be attained since we are restricted to use only times in \( \mathcal{T} \). Define \( \overline{T}_n \) as the closest point to \( T_n \) in \( \mathcal{T} \):

\[ T_n \triangleq \overline{T}_n + d_n, \quad 0 \leq |d_n| < \delta t \tag{6} \]
and $\widetilde{M}$ as the number of distinct values of $\widetilde{T}_n$, the best estimate of $\lambda$ is defined as

$$\lambda_{\text{opt}} \triangleq \frac{\widetilde{M}}{T_M}. \quad (7)$$

It is likely that $\lambda_{\text{opt}} < \lambda_c$, since $\widetilde{M} < M$; however, provided $\lambda \delta t$ is small, $\lambda_c$ and $\lambda_{\text{opt}}$ should remain close. The main idea of this paper is to plug in (7) estimates of $M$ and $T_M$, as now explained.

Let $\{\theta^{(k)} = (\theta_1^{(k)} , \theta_2^{(k)}); k = 1, 2, \ldots, p\}$ be a discrete domain of gamma functions parameters. We consider the span of all these gamma shapes $\Gamma_{\theta^{(k)}}$, sampled on $\mathcal{T}$. For any $\tau$ in $[0, T]$, we can define the following $N \times p$ matrix whose columns are the samplings of the previous gamma functions shifted by $\tau$:

$$A(\tau) \triangleq [\Gamma_{\theta^{(k)}}(t_i - \tau)]_{0 \leq i \leq N - 1, 1 \leq k \leq p} \cdot \quad (8)$$

When $\tau = t_j$ we denote $A(\tau)$ by $A_j$, and shall refer to this matrix as the time block associated to the $j$-th point. We can now define a global dictionary $A$ by concatenating the time blocks $A_j$:

$$A = [A_0 \ A_1 \ \cdots \ A_{N-1}] \quad (9)$$

Should the $T_n$ belong to $\mathcal{T}$ and the $\Phi_n$ to $\Gamma_{\theta^{(k)}}$, then the vector $\beta$ such that $\tilde{y} = A\beta$ would be naturally sparse. In a realistic situation the sampled shapes $\Phi_n(t_i - T_n)$ do not belong to the span of $A$, because $T_n \notin \mathcal{T}$ almost surely and $\Phi_n$ is random. However, provided $\lambda \delta t$ remains sufficiently small, the projection of individual pulses on the span of $A$ yields only few consecutive non-zero coefficients. If the signal is estimated as $\sum_{m=0}^{N-1} A_m \beta_m$, the set $J(\beta) = \{0 \leq m \leq N - 1; \beta_m \neq 0\}$ still contains much fewer elements than $N$. We would like to recover this sparse block pattern $J(\beta)$, so we make use of the LASSO estimator [19]:

$$\hat{\beta}(r) = \arg \min_{\{\beta \in \mathbb{R}^{Np}\}} \left\{ \frac{1}{2N} \left\| y - \sum_{m=0}^{N-1} A_m \beta_m \right\|_2^2 + r \sum_{m=0}^{N-1} |\beta_m|_{t_i} \right\}, \quad (10)$$

where the tuning parameter $r$ quantifies the tradeoff between sparsity and estimation precision. LASSO provides a sparse estimator $\hat{\beta} = [\hat{\beta}_0, \cdots, \hat{\beta}_{N-1}]^T$ such that the linear model $\tilde{y} = A\hat{\beta}$ approximates accurately the signal $y$. In practice, (10) can be efficiently computed by the LARS algorithm [10]. Note that the group-LASSO [26] also exploits the time blocks decomposition of $\beta$ and provide a block-sparse regressor. However, in this paper, we cannot assume the groups to be fully known, due to the incompleteness of $A$, and present only the results achieved by the classical LASSO in our problem.

Assuming the solution (10) is known, the estimation of $\lambda$ should be carefully done. It is tempting to estimate the arrival times with the set $J(\hat{\beta})$ and the total number of occurrences by $\hat{M} = |J(\hat{\beta})|$, then plug this data into (7). However $J(\hat{\beta})$ may contain consecutive active time blocks which do not all correspond to real arrival times. This is not surprising: since $A$ is incomplete, $J(\beta)$ may itself be distinct from $\{\tilde{T}_n, n \geq 0\}$. Another explanation is that
columns in close time blocks are too correlated to ensure block sparsity coherence of the LASSO \[28\]. In this paper we suggest an additional thresholding step to the estimation of \( \lambda \) to circumvent this issue, that is

1. solve \((10)\) to obtain \( \hat{\beta}(r) \).

2. set all the \( \hat{\beta}_m \) such that \( \|\hat{\beta}_m\|_1 < \eta \) to zero, where \( \eta \) is a user defined threshold;

3. estimate recursively \( \hat{T}_n \overset{\Delta}{=} \min_t \{ t > \hat{T}_{n-1} ; \hat{\beta}_{t-1} = 0, \hat{\beta}_t \neq 0 \} \), and \( \hat{M} \overset{\Delta}{=} | \{ t \in \mathcal{T} ; \hat{\beta}_{t-1} = 0, \hat{\beta}_t \neq 0 \} | \).

4. Estimate the activity as

\[
\hat{\lambda}(r, \eta) \overset{\Delta}{=} \frac{\hat{M}}{\hat{T}_\hat{M}}
\]

We refer to steps 2 and 3 as ”post processing” steps. Both steps can be heuristically understood as follows: step 2 in 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.

3 Theoretical results

In order to guarantee some consistency in estimation as well as in selection, previous works imposed conditions on the dictionary \( A \), for instance low correlations between columns \[9, 6, 13, 5\] or positivity of minors of specific sizes \[15, 6, 27\]. The estimation procedure described in this paper is close to \[15\], which suggest improvements of LASSO by hard-thresholding coefficients, so that only representative variables are selected. In \[28, 23\], the irrepresentability condition is presented, and is proved to be necessary if we wish selection consistency. Roughly speaking if \( A_1 \) and \( A_2 \) denote respectively the active columns and their complement, one can define the two Gram matrices \( A_{11} = \frac{1}{N} A_1^T A_1 \), \( A_{21} = \frac{1}{N} A_2^T A_1 \). The irrepresentability condition in its less general form stresses that \( \| A_{21} A_{11}^{-1} \|_\infty < 1 - \eta \) for \( \eta > 0 \), where the infinite (triple) norm \( \| B \|_\infty \) of an operator \( B \) is defined as \( \| B \|_\infty = \max \| Bz \|_\infty ; \) moreover to ensure sign consistency the parameter \( r \) should be chosen between two bounds depending on the least singular value of \( A_{11} \). We adapt in this section these conditions to our framework, and derive bounds of our proposed estimator in a general way.

3.1 Characterization of block sparsity patterns

Given \( \mathbf{y} \) and \( A \), we are interested in subsets of \( \mathcal{T} \) which support vectors \( \beta \) approximating well \( \mathbf{y} \), whereas satisfying an assumption similar to \[\mathbf{4}\]. For \( P \subset \mathcal{T} \) any subset of subdivision
points, let \( A_P \) be the concatenation of time blocks \( A_j \) for \( j \in P \). For any vector \( x \in \mathbb{R}^K \), \( x \geq 0 \) shall mean that every component of \( x \) is nonnegative. Following (8), for any vector \( \beta \in \mathbb{R}^{Np} \) we shall denote by \( \beta = [\beta_0^T, \beta_1^T, \ldots, \beta_{N-1}^T]^T \) its natural decomposition along time blocks. Recall that \( J(\beta) \) designates the block sparsity pattern of \( \beta \), that is \( J(\beta) \triangleq \{ 0 \leq m \leq N - 1, \beta_m \neq 0 \} \).

**Definition 1** Let \( \alpha \) and \( \beta_{\min} < \beta_{\max} \) be positive numbers, and \( P \subset \mathcal{T} \). We shall say that \( P \in \mathcal{P}_{\beta_{\min}, \beta_{\max}, \alpha} \) iff:

\[
\min_{\mathcal{C}} \frac{1}{\sqrt{N}} \| y - A_P \beta \|_2 < \alpha,
\]

where \( \mathcal{C} \) denotes the set of all nonnegative vectors \( \beta = [\beta_0^T, \beta_1^T, \ldots, \beta_{N-1}^T]^T \) such that \( J(\beta) = P \) and that \( \beta_{\min} \leq \| \beta_j \|_1 \leq \beta_{\max} \) for all \( 0 \leq j \leq N - 1 \). As a limit case of the previous definition, we shall denote by \( \mathcal{P}_{\alpha} \) the set of supports defined as in (12) with \( \beta_{\min} = 0, \beta_{\max} = +\infty \).

Equivalently (12) expresses a condition on the solution of a constrained least squares problem associated to \( y \), hence we can assume that the residual \( \delta_P = y - A_P \beta \) and \( \epsilon \) are independent random vectors for \( \beta \) in \( \mathcal{C} \). Note that by definition the set \( \mathcal{P}_{\beta_{\min}, \beta_{\max}, \alpha} \) could be empty. In addition of the previous definition, for any subset \( P \), one can define an associated potential intensity estimate as:

\[
\lambda_P = \frac{|P|}{\max P}.
\]

Note that setting \( P = \{ \tilde{T}_n, \ n \geq 0 \} \) in the latter yields \( \lambda_P = \lambda_{\text{opt}} \) as defined in (6). Therefore, we shall provide in the next section numerical conditions involving \( P, \beta_{\min}, \beta_{\max}, \alpha \), the noise level \( \sigma \) and \( A \), so that the obtained estimator is close to \( \lambda_{\text{opt}} \) with a high probability.

### 3.2 Main results

We introduce hereafter further conventions: the complementary subset of \( P \) is denoted by \( \overline{P} \); for \( P, Q \) two such subsets, we define the matrix \( A_{PQ} \) as

\[
A_{PQ} \triangleq \frac{1}{N} A_P^T A_Q.
\]

Given \( P \subset \mathcal{T} \), for any value of the parameter \( r > 0 \) define \( \hat{\beta}_P(r) = [\hat{\beta}_{P,0}^T; \ldots; \hat{\beta}_{P,N-1}^T]^T \) as the LASSO minimizer (10) under the additional constraint \( J(\hat{\beta}_P) \subseteq P \). Let \( P \in \mathcal{P}_{\alpha} \) for some \( \alpha > 0 \), we can write \( y = A_P \beta_P + \delta_P + \epsilon \), where \( \beta_P \) is a (non negative) vector, and \( \delta_P \) is the residual vector \( y - A_P \beta_P \). We also define the function \( t(x) = \frac{1}{\sqrt{2\pi}} x^{-1} e^{-\frac{x^2}{2}} \), which classically provides an upper bound of the tail of the Gaussian distribution. The next proposition describes for which values of \( r \) the regressor \( \hat{\beta}_P \) is also a global minimizer of the LASSO.
Proposition 1  Denote by $\mu_{\text{min}} > 0$ the least eigenvalue of $A_{PP}$. Suppose the matrix infinite norm $\|A_{PP}\|_\infty$ satisfies for some $\eta > 0$:

$$\|A_{PP}\|_\infty < \frac{\mu_{\text{min}}}{\sqrt{|P|} \cdot p} (1 - \eta)$$

then the following condition holds on the operator infinite norm of $A_{PP}^{-1}$:

$$\|A_{PP}A_{PP}^{-1}\|_\infty < 1 - \eta$$

Moreover, if $r$ is chosen such that

$$r > \max \left\{ \frac{2 \|\delta_P\|_2}{\eta \sqrt{N}}, \frac{2 \sqrt{2} \sigma}{\eta} \sqrt{\frac{\log(N - |P|) p}{N}} \right\},$$

then $\hat{\beta}_P$ is a minimizer of the LASSO functional \(10\) with probability tending to 1 as $N \to \infty$.

**Proof:**  See Appendix B.

In the later results, we assume for convenience that $N$ is an even number, and we define

$$Q_{N/2}(x) \Delta = \frac{\Gamma(N/2, \sqrt{2x})}{\Gamma(N/2)} = e^{-x} \sum_{k=0}^{N/2-1} \frac{x^k}{k!}.$$  

(18)

The asymptotic behavior of \(18\) are detailed in Lemma 1. We also define $\mathcal{G} = \frac{1}{N} A_j^T A_j$ as the Gram matrix of any single time block (obviously independent of $j$ in our setting detailed at 2.2), which has positive entries less or equal than 1. The following result will be the main technical tool for the two main theorems.

Proposition 2  Suppose we have $P \in \mathcal{P}_\alpha$ for $\alpha > 0$, that is $y = A_P \tilde{\beta}_P + \delta_P + \epsilon$ for $\tilde{\beta}_P \geq 0$ and $\|\delta_P\|_2 < \alpha \sqrt{N}$. Let now $r$ be a LASSO parameter value, and $\hat{\beta}(r)$ be the LASSO regressor. Define

$$P_{\alpha+r} \Delta = \{ k \in P; \|\mathcal{G} \tilde{\beta}_{P,k}\|_\infty > \alpha + r \}$$

(19)

as the subset of the most representative blocks of $\tilde{\beta}_P$; Then

$$\max P_{\alpha+r} - \tau \leq \max J(\hat{\beta}(r))$$

(20)

and denoting by $N_{\alpha+r}$ the number of connected components of $\bigcup_{k \in P_{\alpha+r}} (k - \tau, k + \tau)$:

$$N_{\alpha+r} \leq |J(\hat{\beta}(r))|$$

(21)

with probability greater than

$$1 - p \sum_{k \in P_{\alpha+r}} \left( \frac{\sqrt{N} \left[ \|\mathcal{G} \tilde{\beta}_{P,k}\|_\infty - \alpha - r \right]}{2 \sigma} \right)^2 - \frac{4 \alpha^2}{\sigma^2 N} - Q_{N/2} \left( \frac{N}{2} \sigma^2 + \alpha^2 \right).$$

(22)
Assume now that the LASSO solution $\hat{\beta}(r)$ has only nonnegative coefficients. For all $l \in \mathcal{T}$ define $\hat{C}_l = \| G \hat{\beta}_l(r) \|_{\infty} - r$. Then we have

$$\max\{k, \hat{C}_k > 2\alpha\} - \tau \leq \max P$$

(23)

and denoting by $\hat{N}_{2\alpha + r}$ the number of connected components of $\bigcup_{\{l: \hat{C}_l > 2\alpha\}} (l - \tau, l + \tau)$ then:

$$\hat{N}_{2\alpha + r} \leq |P|$$

(24)

both with probability greater than $1 - |\{k, \hat{C}_k > 2\alpha\}| \cdot t\left(\frac{\sqrt{N} \alpha}{2\sigma}\right)$.

**Proof:** See Appendix C.

Proposition 2 is of practical interest. Roughly speaking, its first statement says that if $P$ is a sparsity pattern which allows a good representation of $y$, then the maximal value of the LASSO sparsity pattern $J(\hat{\beta}(r))$ and its cardinality are lower bounded by terms depending on the most representative elements of $P$ only with a high probability. Conversely, the second part of Proposition 2 states that the maximum and cardinality of the set of the most significant active blocks of $\hat{\beta}(r)$ can be upper bounded by terms depending on $P$ only. Therefore, the latter result relates $\hat{\beta}_P$ (which is most likely connected to the true sparsity pattern) to the solution of LASSO $\hat{\beta}(r)$ after thresholding.

**Remark** Note that the probability upper bounds presented in Proposition 2 may be negative if $\sqrt{N} \alpha$ tends to 0. This illustrates the necessity of a more selective thresholding in that case, since in the latter we tend to accept all active blocks as arrival times since $\alpha$ tends to 0. In other words, one should threshold by numbers greater than the LASSO parameter to get reliable comparisons in (21) and (24).

**Remark** Regarding the last result in Proposition 2, up to our knowledge no constraint ensuring the positivity of the LASSO regressor $\hat{\beta}$ appears in previous works. Nevertheless the results obtained on real data show it is a case worth considering: the true signal is not so far from being built out from $A$, and the energies are positive by nature. Actually, it would have been more convenient to consider the LASSO under the additional constraint that all blocks contain nonnegative coefficients only, for this is a crucial point in many proofs. Note that the optimization problem remains convex, thus this issue can be addressed, at least theoretically.

Proposition 2 is used to prove next theorems, where we compare an “oracle” intensity $\lambda_P$ to the proposed intensity estimator (11) computed for a threshold $\eta > 0$. We recall that the blocks of $\hat{\beta}(r)$ are selected by thresholding $\| \hat{\beta}_l(r) \|_1 > \eta$ to obtain selected times $\hat{T}_1, \ldots, \hat{T}_M$ as the minima of each open component formed by consecutive active blocks. We eventually compute $\hat{\lambda}(r, \eta) = \hat{M}/\hat{T}_M$. 

9
Theorem 1 Assume that all the notations and assumptions of Proposition 2 hold, and that \( P \in \mathcal{P}_{\eta,\mu,\alpha} \) for \( 0 < \eta < \mu \) (the value \( \mu \) is not important in this result). Define \( a \triangleq \min_{i,j} g(i,j) \), and let \( 0 < r < a \eta - \alpha \) be a LASSO parameter, provided the right member is indeed positive; denote by \( \hat{\lambda}(r,2\eta) = \frac{\hat{M}}{\hat{T}_M} \) be the estimator computed by [11]. Then one has the following inequality controlling possible overestimation of \( \lambda_P \):

\[
\hat{\lambda}(r,2\eta) - \lambda_P \leq \frac{\hat{M}}{\hat{N}_{2a+r}} \frac{\max J(\hat{\beta}(r))}{\hat{T}_M} \left( \max\{k, \|\hat{\beta}_k\|_1 > 2\eta - \tau \} - 1 \right) \lambda_P; \quad (25)
\]

and this holds with probability greater than

\[
1 - \frac{4 \alpha^2}{\sigma^2 N} - |P| \Pr \left( \frac{\sqrt{N} \left| a \eta - \alpha - r \right|}{2 \sigma} \right) - Q_{N/2} \left( \frac{N}{2} \frac{\sigma^2 + \alpha^2}{\sigma^2} \right) - |J(\hat{\beta})|t \left( \frac{\sqrt{N} \alpha}{2 \sigma} \right)
\]

Proof: See appendix 12.

Note that in (25) the only unknown term besides \( \lambda_P \) is \( \hat{N}_{2a+r} \), which depends on the unknown error bound \( \alpha \); nevertheless the quotient \( \hat{M}/\hat{N}_{2a+r} \) appearing in (25) justifies this third step of the proposed algorithm. Indeed, this is the best strategy to expect \( \hat{M}/\hat{N}_{2a+r} \) to be as close to 1 as possible, thus minimizing the risk to overestimate the intensity.

The last result relies on a “block mutual coherence” bound introduced in (26). It says roughly that if the correlations between active blocks are appropriately bounded, the underestimation of the activity is also controlled by a lower bound.

Theorem 2 Assume that conditions of Theorem 1 hold, and let \( C \) be the maximal correlation between two columns belonging to two distinct time blocks in \( P \), that is

\[
C \triangleq \max_{(k,m) \in P^2 \atop k \neq m} \left\{ \max_{i,j} A_{k,m}(i,j) \right\};
\]

assume that the following condition holds:

\[
a \eta - \alpha - 2C |P| \mu - 2\sqrt{2C (\sigma^2 + \alpha^2)} > 0, \quad (26)
\]

and that \( r \) has been chosen so that

\[
2C \frac{\sigma^2 + \alpha^2}{r} + r < a \eta - \alpha - 2C |P| \mu. \quad (27)
\]
Then the following inequality controlling possible underestimation of $\lambda_P$ does hold under probability bounded from below as in Theorem 1:

$$
\lambda_P \left[ \frac{\hat{M}}{|J(\hat{\beta}(r))|} \left( 1 - \frac{\tau}{T_M} \right) - 1 \right] \leq \hat{\lambda}(r; 2\eta) - \lambda_P; \quad (28)
$$

**Proof:** See appendix D.

Note that the quantity $\frac{\hat{M}}{|J(\hat{\beta}(r))|} \left( 1 - \frac{\tau}{T_M} \right)$ is less than 1, thus the LHS of (28) controls indeed some underestimation. The two latter theorems are written as a function of a general subset $P$, though in practice the main case of interest is when $P$ is the support generated by the $\tilde{T}_n$, thus providing $\lambda_{opt}$. Not surprisingly, the obtained results depend on unknowns quantities such as $|P|$; indeed the inequalities depend heavily on the ratio $\lambda\delta t$, which controls both the sparsity of the optimal $P$ and the accuracy of the dictionary.

One could interpret $\frac{|J(\hat{\beta}(r))|}{\hat{M}}$ in (28) as some average number of active consecutive blocks (which stay so after thresholding by $2\eta$) per estimated arrival time; this emphasizes that the sparsity parameter $r$ should be chosen large enough in order to get a significant upper bound on the activity. More interestingly, (26) shares strong similarity with traditional bounds on the so called “mutual coherence” of the dictionary enforcing selection consistency [5]; indeed the term “block mutual coherence” condition is adequate here, and it is satisfying for our purpose since we do not aim for precise selection consistency. Since $C$ depends on the minimal distance between points in $P$, (26) is satisfied as soon as this distance is greater than some fixed positive number which depends on the shapes in the dictionary (i.e. their parameters) and the sampling width $\delta t$. The dictionary is set up from the beginning, independently from the observed Poissonian sample, therefore this last question is equivalent to the one solved in Lemma 2 in appendix A.

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 $\lambda$ 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 $\lambda$ varies from 0.05 to 0.4, which corresponds to physical activities from $5.10^5$ 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 \( \{\gamma_n, n \geq 0\} \) 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.

In this paper we chose a dictionary made of Gamma functions, in order to represent a charge collection increase and an exponential decrease for one single photon’s pulse; more specifically, assuming that we observe \( N \) points of the sample signal, the \( j \)-th column of the dictionary \( A \) is build accordingly to (8) and (9). 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 \( \theta_1, \theta_2 \). 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 post-processing step described in Proposition 2 and Theorem 1. This allows to use the proposed approach on real-world experiments where fast algorithms and small dictionaries for real-time implementations are in order. For one given activity, the estimator is computed 1000 times by means of the proposed method, and by means of the standard method aforementioned, both in (I) and (II) cases. In all cases, the parameter \( r_N \) was chosen equal to 3\( \sigma \).

### 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 \( \lambda \) (the actual value of \( \lambda \) 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 \( \lambda \) 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 \( \lambda \). Regarding the estimator obtained through the Lasso reconstruction of the signal, the results obtained in
cases where $\lambda$ 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 0.05 to 0.2 the obtained results are very close to the actual counting rate.

When $\lambda$ 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 $\lambda$ is underestimated. However the data is obtained from a sampled signal, therefore the actual $\lambda$ 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 $\lambda_{opt}$ instead of $\lambda$. It is is made in Figures 4(a) and 4(b). We observe an almost linear fit between both estimators, thus showing numerically that the proposed approach provides values similar to $\lambda_{opt}$, which is the best estimate we could build from a full knowledge of $T_n$ and of the sampled signal, but is in practice unattainable.

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 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 $3.99e4$, which conforms both to the simulations and the knowledge of the dataset.

5 Conclusion

We presented in this paper a method based on sparse representation of a sampled spectroscopic signal to estimate the activity of an unknown radioactive source. Based on a crude dictionary, we suggest a post-processed version of the LASSO to estimate the number of
Figure 4: Comparison of $\hat{\lambda}$ with $\lambda_{opt}$.

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

individual photonic 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 the LASSO to estimate the input signal in a sparse manner, we can derive some conditions which guarantee that the number of individual pulses and arrival times are well estimated. 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 $\lambda$ 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 For each positive integer $k$, let $Q_k(x)$ be the incomplete Gamma function defined by [18], and let $x_k$ be any sequence of real numbers such that for some $\gamma > 1$ we have $x_k > \gamma k$ for all $k$. Then for all $m \in \mathbb{R}$:

$$k^m Q_k(x_k) = O \left( \frac{k^{m+1/2} \exp((1 - \gamma + \log \gamma)k)}{k \to \infty} \right)$$

Proof: Since $x_k > k$, we have for each $0 \leq i \leq k$ $x_k^i / i! < x_k^k / k!$. Thus

$$Q_k(x_k) < e^{-x_k} k \frac{x_k^{k-1}}{(k-1)!} = \frac{k^2}{k!} e^{-x_k} x_k^{k-1}$$

As $k$ tends to infinity, Stirling’s formula gives $k! \sim \sqrt{2\pi k} \left( \frac{k}{e} \right)^k$, hence the right hand side (RHS) of (29) is equivalent to the left hand side (LHS) of the following inequality

$$\frac{1}{\sqrt{2\pi}} k^{-k+3/2} x_k^{k-1} \exp(k-x_k) \leq \frac{1}{\sqrt{2\pi}} k^{1/2} \exp((k-1) \log \gamma + (1-\gamma)k)$$

When multiplied by any power of $k$, the RHS of the latter inequality tends to 0 as $k$ goes to infinity, since $1 - \gamma + \log \gamma < 0$, which completes the proof. ■

Lemma 2 Suppose a homogeneous Poisson process of intensity $\lambda$ 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 $\delta$ is bounded from below by $1 - \lambda^2 T \delta$.

Proof: Let us compute the probability that one interarrival time is smaller than $\delta$. 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

$$f((T_1, \ldots, T_n) | N_T=n)(u_1, \ldots, u_n) = f((u_1, \ldots, u_n)) \frac{n!}{T^n} \mathbf{1}_{0=u_0 \leq u_1 \leq \cdots \leq u_n \leq T}.$$
where \( \{U(i), i = 1 \ldots n\} \) are the order statistics of \( n \) independent random variables uniformly distributed on \([0, T]\). When \( n \leq 1 \) whether the Poissonian instants are separated by a distance greater than \( \delta \) is an empty question; when \( n \geq 2 \) and \( 2 \leq i \leq n \) we have:

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

where \( \Omega_i = \{0 \leq u_1 \leq \cdots \leq u_n \leq T; u_i - u_{i-1} \leq \delta\} \). For all \( 1 \leq k \leq n \) we set \( \text{incr}_k = u_k - u_{k-1} \) (we set \( u_0 = 0 \) just above), so it is equivalent to write

\[
\Omega_i = \{\text{incr}_k \geq 0, 1 \leq i \leq n; \quad \text{incr}_i \leq \delta \text{ and } \sum_{k=1}^n \text{incr}_k \leq T\}
\]

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

\[
\Omega_i = \bigcup_{0 \leq t \leq \delta} \hat{\Omega}_i(t);
\]

\[
\hat{\Omega}_i(t) \overset{\Delta}{=} \{\text{incr}_j \geq 0, \text{ for all } j \neq i, \text{incr}_i = t; \sum_{j \neq i} \text{incr}_j \leq T - t\}
\]

 Integrating now along the variable \( t \) we obtain:

\[
\text{Vol}(\Omega_i) = \int_0^\delta \text{Vol}(\hat{\Omega}_i(t)) \, dt
\]

\[
= \int_0^\delta \frac{(T - t)^{n-1}}{(n-1)!} \, dt = \frac{1}{n!} [T^n - (T - \delta)^n]
\]

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

\[
P(T_i - T_{i-1} \leq \delta \text{ for some } 2 \leq i \leq n \mid N_T = n)
\]

\[
\leq (n - 1) \left[1 - \left(1 - \frac{\delta}{T}\right)^n\right]
\]

We can of course consider 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 \( \delta \) is not greater than

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

\[
= \lambda T - [\lambda(T - \delta) - 1] \exp(-\lambda \delta) - 1
\]

The lemma follows from Taylor inequality. ■
B Proof of Proposition 1

All along the proof we will write \( \hat{\beta}_P = \hat{\beta}_P(r) \) for simplicity. Let \( z \) such that \( \|z\|_\infty \leq 1 \), since the \( \ell_\infty \)-norm is dominated by the \( \ell_2 \)-norm we get

\[
\| A_{P^T} A_{P^{1/2}} \cdot z \|_\infty \leq \| A_{P^T} \|_\infty \cdot \| A_{P^{1/2}} \cdot z \|_2 \leq \frac{1}{\mu_{\min}} \| A_{P^T} \|_\infty \sqrt{|P| p}
\]

By (15) we obtain (16). The rest of the proof follows [23] with mild modifications. KKT conditions for \( \hat{\beta}_P \) imply therefore the existence of a vector \( z_P \in [-1; 1]|P|p \) such that:

\[
\beta_P - \hat{\beta}_P = A_{P^T}^{-1} \begin{bmatrix} r z_P - \frac{1}{N} A^T_P (\delta_P + \varepsilon) \end{bmatrix} \tag{30}
\]

Still by KKT conditions \( \hat{\beta}_P \) is a global minimum as soon as:

\[
\frac{1}{N} A^T_P \begin{bmatrix} \delta_P + \varepsilon + A_P (\hat{\beta}_P - \hat{\beta}_P) \end{bmatrix} = r z_P \tag{31}
\]

with \( \|z_P\|_\infty < 1 \). Plugging in (31) the expression obtained at (30), and introducing the matrix

\[
H_P = \frac{1}{\sqrt{N}} (A_{P^T} - A_P A_{P^{1/2}} A_{P^{1/2}}^T),
\]

we obtain:

\[
\frac{1}{\sqrt{N}} H_P^T (\delta_P + \varepsilon) + r A_{P^T} A_{P^{1/2}}^{-1} z_P = r z_P
\]

By assumption (16) it is sufficient that the following condition holds:

\[
\left\| \frac{1}{\sqrt{N}} H_P^T (\delta + \varepsilon) \right\|_\infty < r \eta \tag{32}
\]

Note that \( H_P \) can be rewritten as \( H_P = \left( I - \left( \frac{A_{P^T}}{\sqrt{N}} \right) A_{P^{1/2}}^{-1} \left( \frac{A_P}{\sqrt{N}} \right)^T \right) \left( \frac{1}{\sqrt{N}} A_{P^T} \right) \), showing that the columns of \( H_P \) are the projections of the normalized columns of \( A_{P^T} \) onto the orthogonal complement of the columns of \( A_P \). It thus follows that all columns of \( H_P \) have normalized \( \ell_2 \)-norm bounded by 1 since this is true for \( A_{P^T} \). Denoting by \( H_i \) any column of \( H_P \), then \( H_i^T \varepsilon \) is a Gaussian random variable of variance less than \( \sigma^2 \). Consequently:

\[
P \left( \left\| \frac{1}{\sqrt{N}} H_P^T \varepsilon \right\|_\infty \geq \frac{r \eta}{2} \right) \leq \sum_i P \left( |H_i^T \varepsilon| \geq \frac{\sqrt{N} r \eta}{2} \right) \leq 2 t \left( \frac{\sqrt{N} r \eta}{2 \sigma} \right) (N - |P|)p . \tag{33}
\]
In order to make \((33)\) tend to 0, we need that
\[
r \geq 2 \sqrt{2} \sigma \sqrt{\frac{\log(N - |P|)}{N}} / \eta.
\]
Now we have also
\[
\left| \frac{1}{\sqrt{N}} H^T_i \delta \right| \leq \frac{1}{\sqrt{N}} \| \delta_P \|_2 < \alpha,
\]
hence \(\left\| \frac{1}{\sqrt{N}} H^T_P \delta \right\|_\infty < \frac{r}{2} \) as soon as \(r > \frac{2 \| \delta_P \|_2}{\eta \sqrt{N}}\).

C Proof of Proposition 2

For any \(k \in \mathcal{T}\) and any positive number \(\rho\), we first define \(\mathcal{T}_{\rho,k}\) as the following subset of \(\mathcal{T}\):
\[
\mathcal{T}_{\rho,k} \triangleq \mathcal{T} \setminus \{ t_j \in \mathcal{T}, A_{kj}(i,j) \leq \rho, 1 \leq i, j \leq p \} \quad (34)
\]
In other words, the set \(\mathcal{T}_{\rho,k}\) gathers sampling times \(t_j\) “close” to \(t_k\) in the sense that some correlations between columns of the \(j\)-th time block and columns of the \(k\)-th time block contains columns are above the threshold \(\rho\). Obviously for \(\rho' \geq \rho\) we have \(\mathcal{T}_{\rho',k} \subseteq \mathcal{T}_{\rho,k}\).

We keep the same notations as in Appendix B. Let \(\alpha > 0\) and \(P\) be the support of a nonnegative regressor \(\widehat{\beta}_P\) satisfying
\[
\left\| y - A_P \widehat{\beta}_P \right\|_2 = \| \delta_P \|_2 < \alpha \sqrt{N}.
\]
Let \(\widehat{\beta}\) be the LASSO regressor. For all \(k\) in \(P\), using the same notations as in Proposition 2 we have by the KKT conditions:
\[
\mathcal{G} \widetilde{\beta}_{P,k} + \sum_{k \neq m \in P} A_{k,m} \widetilde{\beta}_{P,m} - \sum_l A_{k,l} \widehat{\beta}_l
\]
\[
= rz_k - \frac{1}{N} A^T_k (\delta_P + \epsilon), \quad (35)
\]
for some vector \(z_k \in [-1, 1]^p\). Let \(0 \leq \gamma \leq 1\); since all elements of \(A_{k,l}\) are bounded by 1, we can write \(\left\| \sum_l A_{k,l} \widehat{\beta}_l \right\|_\infty \leq (1 - \gamma) \sum_{l \in \mathcal{T}_{\gamma,k}} \| \widehat{\beta}_l \|_1 + \gamma \| \widehat{\beta} \|_1\); since the vector \(\sum_{k \neq m \in P} A_{k,m} \widetilde{\beta}_{P,m}\) is nonnegative, it follows immediately from (35) that:
\[
\left\| \mathcal{G} \widetilde{\beta}_{P,k} \right\|_\infty - \alpha - \frac{1}{N} \left\| A^T_k \epsilon \right\|_\infty - r
\]
\[
< (1 - \gamma) \sum_{l \in \mathcal{T}_{\gamma,k}} \| \widehat{\beta}_l \|_1 + \gamma \| \widehat{\beta} \|_1. \quad (36)
\]
We can bound roughly the norm \(\| \widehat{\beta} \|_1\), since \(\widehat{\beta}\) minimizes the Lasso functional:
\[
\frac{1}{2N} \| y - A \widehat{\beta} \|_2^2 + r \| \widehat{\beta} \|_{\ell_1} \leq \frac{1}{2N} \| \delta_P + \epsilon \|_2^2 + r \| \widehat{\beta}_P \|_{\ell_1}, \quad (37)
\]
and as a result:
\[ \| \hat{\beta} \|_{\ell_1} \leq \| \tilde{\beta}_P \|_{\ell_1} + \frac{1}{2Nr} \| \delta_P + \epsilon \|_2^2 \]  
(38)

(36) and (38) imply the weaker inequality
\[ \| G \tilde{\beta}_{P,k} \|_\infty - \alpha - \frac{1}{N} \| A^T_k \epsilon \|_\infty - r < (1 - \gamma) \sum_{l \in I_{\gamma,k}} \| \hat{\beta}_l \|_1 + \gamma(\| \tilde{\beta}_P \|_1 + \frac{1}{2Nr} \| \delta_P + \epsilon \|_2^2) . \]  
(39)

The three terms in the RHS of (39) are bounded as follows. First expand the quadratic expression
\[ \frac{1}{N} \| \delta_P + \epsilon \|_2^2 \leq \alpha^2 + 2 \frac{\epsilon^T \delta_P}{N} + \frac{1}{N} \| \epsilon \|_2^2 ; \]  
(40)

since \( \epsilon \) and \( \delta_P \) are independent, the variable \( Z = \frac{\epsilon^T \delta_P}{\sqrt{N} \sqrt{\sigma^2}} \) is centered with variance smaller than \( \frac{\sigma^2}{N} \). On the other hand, \( \frac{1}{\sigma^2} \| \epsilon \|_2^2 \) is distributed according to a \( \chi^2(N) \) distribution, we thus get when \( N \) is even
\[ P \left( \frac{1}{N} \| \delta_P + \epsilon \|_2^2 > 2\sigma^2 + \alpha^2 \right) \leq Q_{N/2} \left( \frac{N \sigma^2 + \alpha^2}{\sigma^2} \right) + 4 \frac{\alpha^2}{\sigma^2 N} . \]

Denote by \( \tilde{C}_k \triangleq \| G \tilde{\beta}_{P,k} \|_\infty - \alpha \); for all \( k \) in \( P_{a+r} \) we can write
\[ P \left( \frac{1}{N} \| A^T_k \epsilon \|_\infty > \frac{\tilde{C}_k - r}{2} \right) \leq p t \left( \frac{\sqrt{N} (\tilde{C}_k - r)}{2 \sigma} \right) . \]

For all \( k \) in \( P_{a+r} \), the real number \( \gamma_k \) defined by \( \gamma_k = \frac{\| G \tilde{\beta}_{P,k} \|_\infty - \alpha - r}{2 (\| \tilde{\beta}_P \|_1 + \frac{\sigma^2 + \alpha^2}{\tau})} \) is between 0 and 1/2, so (39) yields
\[ P \left( \sum_{l \in I_{\gamma_k,k}} \| \hat{\beta}_l \|_1 > 0 \right) \geq 1 - p t \left( \frac{\sqrt{N} (\tilde{C}_k - r)}{2 \sigma} \right) - \frac{4 \alpha^2}{\sigma^2 N} - Q_{N/2} \left( \frac{N \sigma^2 + \alpha^2}{2 \sigma^2} \right) . \]

To conclude note that \( \sum_{l \in I_{\gamma_k,k}} \| \hat{\beta}_l \|_1 > 0 \) implies the weaker condition that \( \hat{\beta}_l \neq 0 \) for some \( l \in (k - \tau, k + \tau) \); this proves (21) and (20).
The second part of the proposition is proved as follows. Assume that \( \hat{\beta} \) has nonnegative components. If \( \hat{C}_k > 2\alpha \), using a similar argument as in the previous proof with the KKT conditions expressed at the time block \( k \), we get with probability greater than \( 1 - t \left( \frac{\sqrt{N} \alpha}{2\sigma} \right) \) that for any \( \gamma \in [0, 1] \):

\[
\frac{\hat{C}_k - \alpha}{2} < (1 - \gamma) \sum_{j \in \mathcal{T}_{\gamma,k} \cap P} \| \tilde{\beta}_{P,j} \|_1 + \gamma \| \tilde{\beta}_P \|_1
\]

(41)

Setting \( 0 < \gamma_k = \frac{\hat{C}_k - \alpha}{2 \| \beta_P \|_1} \) we have \( \sum_{j \in \mathcal{T}_{\gamma_k,k} \cap P} \| \tilde{\beta}_{P,j} \|_1 > 0 \), i.e. \( \mathcal{T}_{\gamma_k,k} \cap P \neq \emptyset \). Hence (24) and (23) follow with probability bounded as written.

D Proofs of Theorem 1 and Theorem 2

In order to prove Theorem 1, we combine the results of (20) and (24): whenever \( a \eta - \alpha > r \) note that \( P_{\alpha + r} = P \) (19), so one has

\[
\max P - \tau \leq \max J(\hat{\beta})
\]

(42)

with probability at least \( 1 - |P| p t \left( \frac{\sqrt{N} \left[ a \eta - \alpha - r \right]}{2\sigma} \right) - \frac{4 \alpha^2}{\sigma^2 N} - Q_{N/2} \left( \frac{N}{2} \frac{\sigma^2 + \alpha^2}{\sigma^2} \right) \). Now by Proposition 2 and the trivial inequality \( |\{ k, \hat{C}_k > 2\alpha \}| \leq |J(\hat{\beta})| \) the following inequalities occur with the probability announced in the statement:

\[
\frac{\hat{N}_{2\alpha + r}}{\max J(\hat{\beta}(r))} \frac{\max \{ k, \hat{C}_k > 2\alpha \} - 2\tau}{\max \{ k, \hat{C}_k > 2\alpha \} - \tau} \leq \frac{\hat{N}_{2\alpha + r}}{\max J(\hat{\beta})} \frac{\max P - \tau}{\max P} \leq \frac{|P|}{\max P} = \lambda_P
\]

(43)

(25) is straightforward from (13), since any \( k \) in \( \mathcal{T} \) such that \( \| \tilde{\beta}_k \|_1 > 2\eta \) satisfies in particular \( \hat{C}_k > 2\alpha \), since the stronger \( r + 2\alpha < 2a\eta \) does hold, thus Theorem 1 holds.

The proof of Theorem 2 relies on elements of the proof of Proposition 2. In the latter, we found that whenever \( 0 < \gamma \) satisfies

\[
\gamma < \min_{k \in P_{\alpha + r}} \left\{ \frac{\| S_{\tilde{\beta}_{P,k}} \|_{\infty} - \alpha - r}{2 \left( \| \tilde{\beta}_P \|_1 + \frac{\sigma^2 + \alpha^2}{r} \right)} \right\},
\]

then the number of connected components of the union \( \bigcup_{k \in P_{\alpha + r}} \mathcal{T}_{\gamma,k} \) is lesser or equal than \( |J(\hat{\beta}(r))| \). If \( r \) satisfies (27), we get that \( P_{\alpha + r} = P \) because in particular \( a\eta > \alpha + r \).
Moreover, when
\[ C < \frac{a\eta - \alpha - r}{2 \left( |P|\mu + \frac{\sigma^2 + \alpha^2}{r} \right)}, \]
the previous number of connected components is simply equal to $|P|$. This gives precisely condition (27), which can be realized when (26) is satisfied: indeed the LHS of (27) achieves its minimum at $r = \sqrt{2 a (\sigma^2 + \alpha^2)}$, which is then plugged back. These considerations together with (21), (23) imply the inequality:
\[ \lambda_P \leq \frac{|J(\tilde{\beta}(r))|}{\max\{k, \tilde{C}_k > 2\alpha\} - \tau}; \quad (44) \]
Since $\tilde{T}_M \leq \max\{k, \tilde{C}_k > 2\alpha\}$, the result follows from (11) and (44). There is therefore an interval of values of $r$ for which (44) occurs under the estimated probability, which proves Theorem 2.

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] S. Chen, D. Donoho, , and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.
[9] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. Information Theory, IEEE Transactions on, 47(7):2845–2862, 2001.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. Annals of Statistics, 32(2):407–499, 2004.

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

[12] W. R. Leo. Techniques for Nuclear and Particle Physics Experiments: A How-To Approach. Springer, 2nd rev edition, January 1994.

[13] K. Lounici. Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. Electronic Journal of Statistics, 2:90–102, 2008.

[14] N. Meinshausen and P. Buhlmann. High-dimensional graphs and variable selection with the lasso. The Annals of Statistics, 34(3):1436–1462, 2006.

[15] N. Meinshausen and B. Yu. Lasso-type recovery of sparse representations for high-dimensional data. The Annals of Statistics, 37(1):246–270, 2009.

[16] 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.

[17] 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.

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

[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. Buhlmann. 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. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using l1-constrained quadratic programming (Lasso). IEEE Trans. Inf. Theor., 55(5):2183–2202, 2009.
[24] G. P Westphal. Real-Time correction for counting losses in nuclear pulse spectroscopy. *Journal of Radioanalytical and Nuclear Chemistry*, 70:387–397, 1982.

[25] 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.

[26] 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.

[27] 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.

[28] P. Zhao and B. Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541—2563, November 2006.