Learning to unmix from Poisson measurements with application to $\gamma$-spectroscopy

Jérôme Bobin  
CEA Saclay  
DRF/Irfu/SEDI/LCS  
91191 Gif-sur-Yvette, France  
Email: jerome.bobin@cea.fr

Jiaxin Xu  
IRSN  
PSE-ENV/SAME/LMRE  
91400 Orsay, France  
Email: jiaxin.xu@irsn.fr

Anne de Vismes Ott  
IRSN  
PSE-ENV/SAME/LMRE  
91400 Orsay, France  
Email: anne.de-vismes@irsn.fr

Christophe Bobin  
CEA Saclay  
DRT/LIST/DM2I/LNHB  
91191 Gif-sur-Yvette, France  
Email: christophe.bobin@cea.fr

Abstract—Spectral unmixing is a key linear inverse problem to analyze radioactivity in the environment (aerosol samples) with $\gamma$-ray spectroscopy. Current approaches are agnostic to the available archive of past measurements, which bring highly valuable information to perform accurate radionuclide activity estimation. For that purpose, we propose an unrolled version of the ADMM (alternating directions method of multipliers) algorithm to tackle the underlying maximum Poisson likelihood regression problem along with a learning of a data-driven regularization. Preliminary results show the efficiency of the proposed approach, with improved estimation accuracy both in estimation bias and variance.

**Problem Formulation:** A single observed $\gamma$-ray spectrum $b$ is composed of $t$ samples that measure the number of photons in different channels of energy as illustrated in Fig[1]. It is modeled as the linear combination of $n$ individual radionuclides $\{a_i\}_{i=1,\ldots,n}$, to which a background signal $m$ is added. Furthermore, an observation $b$ follows a Poisson distribution: $b \sim \text{Poisson}(\Phi a + m)$, where $\Phi = [\phi_1, \cdots, \phi_n]$ and the background $m$ are generally known accurately. For the purpose of environmental radioactivity applications, the challenge is estimate the activity with very high accuracy, with low estimation bias and variance. A standard approach consists in estimating $a$ by solving a least-square regression problem with positivity constraint [1], which does not precisely account for the Poisson nature of the noise. Recently, we showed that replacing the Euclidean metric with a Poisson neg-loglikelihood dramatically improves the bias as well as the variance of the underlying estimator [1]. The problem to be tackled takes the generic form:

$$
\min_a \lambda \mathcal{J}(a) + \Phi a + m - b \odot \log(\Phi a + m),
$$

(1)

where $\odot$ is the Hadamard product. The term $\mathcal{J}$ stands for a penalization term based on some prior knowledge about $a$. The second term is the neg-loglikelihood of the Poisson distribution. If common methods in the field only impose the non-negativity of $a$, ones could benefit from the available archive of hundreds to thousands of routinely taken measurements to further learn a data-driven prior $\mathcal{J}(\cdot)$, which should improve the both the estimation bias/variance.

For that purpose, algorithm unrolling methods have recently emerged in machine learning. Such methods consist in building a recurrent network that reproduces that inversion process of common optimization algorithms (e.g. Forward-Backward Splitting [2-4] and proximal primal-dual [5] algorithms or Neumann Network -NN- [6]) while allowing to learn a data-driven regularization from some training set. To our knowledge, an unrolled version of the ADMM algorithm [7] has been introduced in [8] to specifically tackle deconvolution problems in MRI.

**Proposed method:** In this work, we propose an unrolled version of the ADMM algorithm to tackle the following equivalent of the problem in Eq.(1):

$$
\min_{u,a,v} \max_{\rho} \mathcal{J}_\Phi(a) + u - b \odot \log(u) + v^T(u - \Phi a - m) + \frac{\rho}{2} \|u - \Phi a - m\|^2,
$$

where $u$ is an extra variable, $v$ is the dual variable related to the constraint $u = \Phi a + m$ and $\rho$ is a positive scalar. The regularization term $\mathcal{J}_\Phi$ now depends on some parameters $\Theta$ to be learnt from the available training set.

In the present context, the advantage of ADMM is that it allows to split the inversion of spectral signature dictionary and the application of the regularization in two distinct steps. A single iteration $k$ of the proposed Learned-ADMM, and subsequently each layer of the resulting recurrent network, reads as:

- **Update of $u$:** minimizing Eq.(1) with respect to $u$ leads to:

$$
u(k+1) = \text{prox}_{1/\rho} \Phi \Phi^T a(k) + m - 1/\rho \nu(k),$$

where $\text{prox}_{1/\rho}$ is proximal operator of the Poisson neg-likelihood with scaling parameter $1/\rho$ [9].

- **Update of $a$:** updating $a$ for fixed parameters $\Theta$ is done as follows:

$$a(k+1) = \text{argmin}_a \mathcal{J}_\Phi(a) + \frac{\rho}{2} \|\Phi a - m + 1/\rho \nu(k) - u(k+1) - a(k)\|^2.$$  

Since the dictionary of spectral signatures $\Phi$ is not orthogonal, this problem does not admit a closed-form solution. Instead of resorting to a numerical evaluation with an extra iterative procedure, it is rather approximated with a projected least-square estimate of the form: $a(k+1) = R_\Theta(\Phi^T(u(k+1) - m + 1/\rho \nu(k)))$, where $\Phi^T$ is the pseudo-inverse of $\Phi$.

The operator $R_\Theta$ stands for a shrinkage operator that depends on parameters $\Theta$ that are updated during training procedure so as to minimize the mean estimation bias $\sum_{p} ||a_p - a_p^*||$, from a training set of $T$ samples $\{\{b_p,a_p^*\}\}_{p=1,\ldots,T}$. The processing of routine aerosol measurements generally leads to activities that do not vary to a large extent for natural radionuclides. Therefore, the goal of learning $R_\Theta$ is to capture this standard regime as well as the intricate correlation between the radionuclides’ estimated activities. For that purpose, $R_\Theta$ is encoded with a MLP (multilayer perceptron) with $\ell$ layers and $n$ hidden units per layer; activation functions are rectified linear units (ReLU). Trainable variables are therefore the weight matrices and the bias for all layers.

- **Gradient ascent on $v$:** the dual variable is updated as: $\nu(k+1) = \nu(k) + \rho (u(k+1) - \Phi a(k+1) - m)$. The parameter $\rho$ is also trained along with the inversion procedure.

Preliminary results are based on experiments, which have been performed on realistic simulations of $\gamma$-spectra. These results are described and commented in Figures [5] and [3]. More extensive results, especially on real data, will be presented at the conference. This work is partly funded by the European Community through the grant LENA (ERC StG - contract no. 678282).
The data are composed of the 4 radionuclides that are routinely measured in the environment (\(^{7}\text{Be}\), \(^{22}\text{Na}\), \(^{30}\text{K}\), \(^{210}\text{Pb}\)) and \(^{137}\text{Cs}\), a commonly sought-after artificial radionuclide. The training and testing sets are composed of \(T = 1000\) samples unless stated otherwise; validation is carried out on 10% of the training samples. In these sets, the variability of the simulated activities is based on the levels that are measured on routine measurements.

In these experiments, \(L = 5\) provide the best results; more layers did not provide significant improvements. The data-driven shrinkage operator \(\mathcal{R}_{\text{Be}}\) is implemented multilayer-perceptron with \(\ell\) layers and \(n\) hidden units per layer; activation functions are rectified linear units (ReLU). In this application, \(\ell = 2\) and \(n = 5\) (i.e., one per radionuclide) provided a simple, robust and effective regularization. More complex models with more layers or undercomplete weight matrices did not lead to improved results. However, and in contrast to standard practice in unrolling approaches, learning individual parameters (weights and biases) for each recurrent block yielded significantly better estimation results without increasing the robustness of the training. The proposed network has been implemented with Keras with Tensorflow backend [12], the ADAM optimizer is used with a learning rate of \(10^{-4}\). The initial first guess for \(u^{(0)}\) and \(v^{(0)}\) set to 0 and \(a^{(0)}\) is set as the PoissonML estimate.

**Fig. 1.** Description of the data: Example of a single spectrum with its individual components. These simulations have been obtained using the MCNP simulation code [https://laws.lanl.gov/hosts/mcnp.lanl.gov/index.shtml]. The data are composed of the 4 radionuclides that are routinely measured in the environment (\(^{7}\text{Be}\), \(^{22}\text{Na}\), \(^{30}\text{K}\), \(^{210}\text{Pb}\)) and \(^{137}\text{Cs}\), a commonly sought-after artificial radionuclide. The training and testing sets are composed of \(T = 1000\) samples unless stated otherwise; validation is carried out on 10% of the training samples. In these sets, the variability of the simulated activities is based on the levels that are measured on routine measurements.

**Fig. 2.** L-ADMM implementation: each recurrent block is composed of a main block where \(u\) and \(a\) are updated sequentially. The update of the dual variable \(v\) appears externally and is the only variable that explicitly cumulates information in the recurrence; this is reminiscent of skip connections. Several values for the number of recurrent layers \(L\) have been tested. In these experiments \(L = 5\) provides the best results; more layers did not provide significant improvements. The data-driven shrinkage operator \(\mathcal{R}_{\text{Be}}\) is implemented multilayer-perceptron with \(\ell\) layers and \(n\) hidden units per layer; activation functions are rectified linear units (ReLU). In this application, \(\ell = 2\) and \(n = 5\) (i.e., one per radionuclide) provided a simple, robust and effective regularization. More complex models with more layers or undercomplete weight matrices did not lead to improved results. However, and in contrast to standard practice in unrolling approaches, learning individual parameters (weights and biases) for each recurrent block yielded significantly better estimation results without increasing the robustness of the training. The proposed network has been implemented with Keras with Tensorflow backend [12], the ADAM optimizer is used with a learning rate of \(10^{-4}\). The initial first guess for \(u^{(0)}\) and \(v^{(0)}\) set to 0 and \(a^{(0)}\) is set as the PoissonML estimate.

**Fig. 3.** Robustness with respect to the sample size: in these experiments, 4 different algorithms are compared for tackling spectral unmixing problems: i) least-squares with non-negativity constraint (LS - see [1]), ii) a Neumann Network (NN) with a trainable prior that is similar to the one we implemented in L-ADMM, iii) non-negative Poisson log-likelihood minimization (PoissonML - [10]), and iv) the proposed L-ADMM approach. The robustness of the learning-based approaches with respect to the training set size is first evaluated. For that purpose, the relative bias \(|\hat{a} - a^\ast|/a^\ast\) is computed for each radionuclide. The present figure displays the results when the total number of counts (total number of measured photons) if 15.10^3 (the results are for for \(^{7}\text{Be}\) but are similar for the 4 other radionuclides); the median value as well as the 25% and 75% percentiles across the test set are features. This shows that for a number of samples larger than 1000, the L-ADMM algorithm provide a significantly smaller bias as well as narrower variance, which is expected with a properly learnt prior regularization.

**Fig. 4.** Evolution as a function of the total number of counts: This figure displays the evolution of the relative bias for the estimated \(^{210}\text{Pb}\) activity when the total number of counts evolves from 15.10^3 to 240.10^3. Unsurprisingly, the bias decreases when the number of counts increases since the noise level decreases. However, the proposed L-ADMM algorithm provides a significantly lower bias, especially for lower number of counts. A similar phenomenon is observed for the NN algorithm when the number of counts is equal to 15.10^3.

**REFERENCES**


