Sparse Modeling for Astronomical Data Analysis

Shiro Ikeda\textsuperscript{1}, Hirokazu Odaka\textsuperscript{2} and Makoto Uemura\textsuperscript{3}

\textsuperscript{1} The Institute of Statistical Mathematics, 10-3, Midoricho, Tachikawa 190-8562, Japan
\textsuperscript{2} Kavli Institute for Particle Astrophysics and Cosmology (KIPAC), Menlo Park, CA 94025, U.S.A.
\textsuperscript{3} Hiroshima Astrophysical Science Center, Hiroshima University, Kagamiyama 1-3-1, Higashi-Hiroshima 739-8526, Japan
E-mail: shiro@ism.ac.jp

Abstract. For astronomical data analysis, there have been proposed multiple methods based on sparse modeling. We have proposed a method for Compton camera imaging. The proposed approach is a sparse modeling method, but the derived algorithm is different from LASSO. We explain the problem and how we derived the method.

1. Introduction
In astronomy, different imaging methods are used depending on the wavelength. Compton camera imaging is one of the imaging method to visualize gamma-ray sources from 100keV to several MeV. It is an important technology and high-resolution Compton cameras will be installed in the next-generation X-ray observatory ASTRO-H \cite{1, 2}, which is scheduled for launch soon.

Imaging of Compton camera is not straightforward because only a small portion of photons are absorbed after Compton scattering and the direction of arrival of each photon is not known directly. Since photon detection is a stochastic process, the image reconstruction problem can be formulated as a statistical estimation problem \cite{3, 4, 5}. We follow the framework developed for COMPTEL \cite{6}, and employ the bin-mode estimation (BME) method.

Since the gamma-ray source is sparsely distributed in the image, it is natural to apply sparse modeling method. However, LASSO type estimator does not work in the original form. We had to develop a new method for the Compton camera imaging. In this paper, we explain the method we have developed \cite{7, 8} from a Bayesian viewpoint.

We have developed an estimation method for Compton camera imaging \cite{7, 8}.

2. Compton camera imaging
2.1. Compton camera system
Compton camera system is schematically shown in figure 1. Most of the photons will travel through the camera without any interaction, but some of the photons will be scattered by a scatterer and absorbed by an absorber.

When a photon from $\mathbf{u} = (\alpha, \delta)$ (in equatorial coordinates) is detected, it is scattered at $\mathbf{x} = (x_1, x_2)$ of the scatterer and absorbed at $\mathbf{y} = (y_1, y_2)$ of the absorber. Let $E_1$ and $E_2$ be the energies of the recoil electron and the scattered photon, respectively. At each event, $\mathbf{x}$, $\mathbf{y}$, $E_1$ and $E_2$ are measured.
This is the so called Compton scattering and the relation between the scattering angle $\theta$ and the energies is known. It is

$$\cos \theta = 1 - m_e c^2 \left( \frac{1}{E_2} - \frac{1}{E_1 + E_2} \right),$$

where $m_e$ and $c$ are the mass of an electron and the speed of light, respectively. Thus, from a single event, we know the photon was from a point on a cone (figure 1). The goal of the Compton camera imaging is to reconstruct the gamma-ray intensity map $\lambda(\mathbf{u})$ on the celestial sphere from collected information. This is not straightforward since a single scattering event does not provide direct information of the direction of arrival.

From the measurement of a single scattering event, we receive the scattering angle $\theta$ and the relative position of the event on the scatterer and absorber, i.e. $(\mathbf{x} - \mathbf{y})$ (because the distance from the astronomical objects are large). Let us define $\mathbf{v} = ((\mathbf{x} - \mathbf{y}), \cos \theta)$ as the information collected from a single event. In the following of the paper, the vector $\mathbf{u}$ is discretized into pixels, and $\mathbf{v}$ are discretized into bins, therefore they are finite discrete variables.

We would like to formulate the estimation problem with a probabilistic framework. Let us start by defining distributions. Let $s(\mathbf{u})$ be the probability that a photon from pixel $\mathbf{u}$ is absorbed. We define $\rho(\mathbf{u})$ as the probability that an absorbed photon arrives from $\mathbf{u}$, more precisely,

$$\rho(\mathbf{u}) = \frac{\lambda(\mathbf{u})s(\mathbf{u})}{\sum_{\mathbf{u}'} \lambda(\mathbf{u}')s(\mathbf{u}')}.$$  

Next, we define $p(\mathbf{v} \mid \mathbf{u})$ as the probability that an absorbed photon from $\mathbf{u}$ is detected at $\mathbf{v}$, and $q(\mathbf{v})$ as the probability that an absorbed photon is detected at $\mathbf{v}$. The relation between these probability distributions are summarized in the following equation

$$q(\mathbf{v}) = \sum_{\mathbf{u}} p(\mathbf{v} \mid \mathbf{u}) \rho(\mathbf{u}).$$

Figure 1. Compton camera.
All of $\rho(u)$, $p(v \mid u)$, and $q(v)$ are multinomial distributions conditional to the events that a photon is absorbed by the camera. After observing many photons, the empirical distribution of $q(v)$ is observed. If $p(v \mid u)$ and $s(u)$ is known, it is possible to estimate $\rho(u)$ and $\lambda(u)$.

The problem is how to prepare $p(v \mid u)$ and $s(u)$. Here we utilize a numerical method. A Compton camera system was simulated and a lot of photons were randomly drawn numerically using a software [9]. The results were accumulated to compute $p(v \mid u)$ and $s(u)$. This method is general and easily implemented.

2.2. Estimation

Equation (3) shows the relation between $q(v)$ and $\rho(u)$. Since $u$ and $v$ are discrete, $q(v)$ and $\rho(u)$ are positive vectors and $p(v \mid u)$ is a positive matrix. We assume $p(v \mid u)$ is known and $q(v)$ is observed empirically. Our goal is to estimate $\rho(u)$ under this condition. We further know $\rho(u)$ is sparse.

Although this is an estimation of a sparse vector, we cannot use LASSO type estimator [10]. We explain the reason from a Bayesian viewpoint.

The LASSO estimation can be viewed as the MAP (maximum a posteriori) estimation. The cost function is equivalent to the negative of the log posterior distribution. The posterior distribution consists of two terms: the log likelihood and the log of sparse prior. Let us consider the posterior distribution of the present problem.

Likelihood term Suppose $N$ photons are detected independently and $t$-th photon is absorbed at $v_t$, $(t = 1, \cdots, N)$. The log likelihood function is defined as follows

$$L(\rho) = \sum_{t=1}^{N} \log q(v_t) = \sum_{t=1}^{N} \log \sum_{u} p(v_t \mid u) \rho(u).$$

(4)

This is different from the squared loss term of LASSO.

Prior term The widely used prior term of LASSO is proportional to $\| \cdot \|_{\ell_1}$. However the vector $\rho(u)$ is a multinomial distribution, therefore $\|\rho(u)\|_{\ell_1} = \sum_u \rho(u) = 1$ and the L1 norm of the vector is constant. Instead of using L1 norm, we employ $\alpha \sum_u \log(1/\rho(u))$. Since the term becomes large for a small $\rho(u)$, it encourages the solution to be sparse. The induced prior distribution is the Dirichlet distribution, which is the conjugate prior of the multinomial distribution.

Now, the log of the posterior distribution is

$$\sum_{t=1}^{N} \log \sum_{u} p(v_t \mid u) \rho(u) - \alpha \sum_u \log \rho(u).$$

cost function to minimise is the negative of the loss posterior, which becomes

$$C = -\sum_{t=1}^{N} \log \sum_{u} p(v_t \mid u) \rho(u) + \alpha \sum_u \log \rho(u).$$

Algorithm

The algorithm for computing the optimal $\rho(u)$ can be easily implemented with a modified EM (Expectation-Maximization) algorithm [11]. It looks like the Iterative Shrinkage
Thresholding algorithm for LASSO.

\[(E\text{-step})\quad q^{(l)}(v) = \sum_u p(v | u) \rho^{(l)}(u), \quad (5)\]

\[(M\text{-step})\quad \rho^{(l+1)}(u) \propto \max \left[ 0, \left( \sum_{t=1}^N \frac{p(v_t | u)}{q^{(l)}(v_t)} \rho^{(l)}(u) - \alpha \right) \right]. \quad (6)\]

where \(l\) starts from 0 and increases by 1 at each iteration. By choosing \(\alpha\) between 0 and 1, the optimal vector becomes sparse. When the difference between \(q^{(l+1)}(v)\) and \(q^{(l)}(v)\) becomes sufficiently small, the update is terminated. For more details of the algorithm, see [8].

3. Conclusion
Since the spatial distribution of astronomical objects is sparse in many cases, the idea of sparse modeling is often effective for astronomical data analysis. We have utilized the idea for Compton camera imaging and explained the method in this paper.

The popular LASSO type estimation employs L1-norm regularization term. However, it does not work in the current situation since the target vector for estimation is a multinomial distribution whose L1-norm is constant. We employed the logarithm to obtain the sparse solution. In the Bayesian framework, our choice corresponds to setting the prior as the Dirichlet distribution, which is conjugate to the multinomial distribution.

Acknowledgments
The present study was financially supported by JSPS KAKENHI Grant Number 25120007 and 25120008.

References
[1] Takahashi T et al 2010 The ASTRO-H mission Proc. SPIE 7732 77320Z
[2] Tajima H and et al 2010 Soft gamma-ray detector for the ASTRO-H mission Proc. SPIE 7732 773216
[3] Tomitani T and Hirasawa M 2002 Image reconstruction from limited angle Compton camera data Phys. Med. Biol. 47 (12) 2129–45
[4] Parra L C 2000 Reconstruction of cone-beam projections from Compton scattered data IEEE trans. Nucl. Sci. 47 (4) 1543–50
[5] Xu D and He Z 2006 Filtered back-projection in 4\(\pi\) Compton imaging with a single 3D position sensitive CdZnTe detector IEEE trans. Nucl. Sci. 53 (5) 2787–96
[6] Bloemen H et al 1994 COMPTEL imaging of the Galactic disk and the separation of diffuse emission and point sources Astrophys. J. Suppl. S. 92 (2) 412–23
[7] Ikeda S, Odaka H, Uemura M, Takahashi T, Watanabe S and Takada S 2013 Compton camera imaging Proc. 2013 7th Int. Conf. on Sensing Technology (Wellington) 674–7
[8] Ikeda S, Odaka H, Uemura M, Takahashi T, Watanabe S and Takada S 2014 Bin mode estimation methods for Compton camera imaging Nucl. Instrum. Methods Phys. Res. A 760 46–56
[9] Odaka H et al Development of an integrated response generator for Si/CdTe semiconductor Compton cameras Nucl. Instrum. Methods Phys. Res. A 624 (2) 303–9
[10] Tibshirani R 1996 Regression shrinkage and selection via the lasso J. R. Stat. Soc. Series B Stat. Methodol. 267–88
[11] Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm J. R. Stat. Soc. Series B Stat. Methodol. 39 1–38