A New Parameter Choice Method for Inverse Problems with Poisson Noise

Elias Salomão Helou and Lucas Eduardo Azevedo Simões
ICMC/USP – Campus de São Carlos
Caixa Postal 668
13.560-970 São Carlos, SP, Brazil
http://www.icmc.usp.br/~elias

***Abstract*—On image recovery and reconstruction, signal detection is often performed by some photon-counting process, thereby leading to data that can be modeled as Poisson variables. With such problems in mind, we provide a new parameter methodology choice for the regularization of linear inverse problems where the data are affected by Poisson noise.**

***Keywords*-parameter choice; Poisson noise; inverse problems; emission tomography.**

I. Introduction

Imaging technologies often detect the signal to be later recovered or reconstructed as an image by counting the arrival of photons emitted by or transmitted through the object. This is the case, to mention a few examples, of CCD (Charge-Coupled Device) sensors [1], and emission [2], [3] and transmission tomography [4]. Since counting processes are usually well modeled by a Poisson distribution, it is natural to consider these data as Poisson distributed random variables. Furthermore, image reconstruction from tomographic data or image recovery from blurred or otherwise distorted images are well known to be ill-posed problems, requiring some sort of regularization to offer meaningful solutions from noisy data.

Contributions: The present paper introduces a new parameter choice method for the regularization of linear ill-posed problems with Poisson data, which is a modification of the approach by Santos and De Pierro [5].

A. Tykhonov-Like Regularization

One of the most popular regularization methods is Tykhonov regularization [6] and its variants. The basic form of Tykhonov regularization for the linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$ with $A \in \mathbb{R}^{m\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$ is:

$$\min \quad \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma\|\boldsymbol{x}\|_2^2,$$

where $\gamma \geq 0$ is the so-called *regularization parameter*, which control the smoothness imposed to the solution. For larger values of γ, the solution becomes smoother at the cost of being less correlated to the data. This bias is not only acceptable, but desired: it is best to have a biased estimator with reasonable variance than an unbiased one whose variability is too large.

More sofisticated models inspired on Tykhonov's method exist, as measures other than the squared 2-norm are useful and some constraints may convey important information about the problem. In general, one will try to solve problems such as

$$\min_{\boldsymbol{x}\in C} \quad f(A\boldsymbol{x}, \boldsymbol{b}) + \gamma r(\boldsymbol{x}).$$

Such optimization-based regularization procedures are very important and widely used in practice, but require careful choice of the regularization parameter γ. Choosing γ in order to obtain a good trade-off between smoothness and data-fitness is the problem we tackle in the present paper. When the error is gaussian, a large amount of methods exists for the choice of γ [7]. However, when it comes to Poisson data, there are much fewer choices, among which we mention Bardsley and Goldes' approach [8], Bertero *et al.*'s method [9] and Santos and De Pierro's technique [5]. We briefly review these methods and introduce our own in the next section.

II. Parameter Choice Rules for Linear Ill-Posed Problems with Poisson Data

This section describes existing methods for choosing an adequate regularization parameter for the solution of a system $A\boldsymbol{x} = \boldsymbol{b}$.

A. Bardsley and Goldes' Discrepancy Principle

The classical Morozov's Discrepancy Principle [6] is intended to be used with gaussian data. Based on the fact that, in such a case, $E\left[\|(A\boldsymbol{x})_i - b_i\|_2^2\right] = \sigma^2$, where σ is the standard deviation of the b_i, it selects γ such that

$$\|A\boldsymbol{x}^\gamma - \boldsymbol{b}\|_2^2 = m\sigma^2.$$

However, for Poisson noise, we actually have the different relation $E\left[\|(A\boldsymbol{x})_i - b_i\|_2^2\right] = E[b_i]$. Because of that, it is justifiable to use the following equality as a criterion:

$$\|\operatorname{diag}(A\boldsymbol{x}^\gamma)^{-1/2}(A\boldsymbol{x}^\gamma - \boldsymbol{b})\|_2^2 = m, \tag{1}$$

where $\operatorname{diag}(\boldsymbol{x})$ is the diagonal matrix with non-zero entries corresponding to the components of $\boldsymbol{x}$.

B. Bertero et al.'s Discrepancy Principle

Another approach is given in [9]. The key idea is to observe the following result, proven in [10]:

Lemma 1. *Let*

$$F(x,\lambda) = 2\left\{x\log\frac{x}{\lambda} - x + \lambda\right\},$$

then, if x_λ is a random variable which is Poisson distributed with parameter λ, we have

$$E[F(x_\lambda, \lambda)] = 1 + O\left(\frac{1}{\lambda}\right).$$

Therefore, if the b_i are large enough we have

$$KL(\boldsymbol{b}, A\boldsymbol{x}^\dagger) \approx \frac{m}{2},$$

where KL is the Kullback-Leibler cross-entropy [11]. Thus it makes sense to consider choosing a value of γ for which we have

$$KL(\boldsymbol{b}, A\boldsymbol{x}^\gamma) = \frac{m}{2}. \tag{2}$$

C. Santos and De Pierro's Approach

A completely different route was taken by Santos and De Pierro in [5]. Their approach consists of trying to approximate

$$E[KL(A\boldsymbol{x}^\dagger, A\boldsymbol{x}^\gamma)]$$

in order to minimize this value, the procedure proposed by the authors is to first generate $\boldsymbol{\omega} \in \mathbb{R}^m$ as a sequence of independently distributed standard gaussian variables and then to find γ such that it minimizes

$$\sum_{i=1}^{m} (A\boldsymbol{x}^\gamma)_i - \boldsymbol{b}^T \mathbf{log}(A\boldsymbol{x}^\gamma) - \frac{m\boldsymbol{\omega}^T \operatorname{diag}(\boldsymbol{b})\{\mathbf{log}(A\boldsymbol{x}_+^\gamma) - \mathbf{log}(A\boldsymbol{x}_-^\gamma)\}}{\delta\|\boldsymbol{\omega}\|_2^2}, \tag{3}$$

since this minimizer will, in average and approximately, also minimize $E\left[KL(A\boldsymbol{x}^\dagger, A\boldsymbol{x}^\gamma)\right]$.

III. A New Rule for Parameter Choice

Now we present a new approach for the problem of selecting an appropriate regularization parameter. Our rationale is basically the same of the previous one, except that we try to minimize

$$E\left[\|A\boldsymbol{x}^\dagger - A\boldsymbol{x}^\gamma\|_2^2\right].$$

Our main tool is the following result, the proof of which we omit due to space constraints:

Theorem 1. *Assume $F_\lambda := A\boldsymbol{x}^\gamma$ is twice continuously differentiable and let $\boldsymbol{\omega} \sim \mathcal{N}(\boldsymbol{0}, I)$, then we have*

$$E\left[\|A\boldsymbol{x}^\dagger - A\boldsymbol{x}^\gamma\|_2^2\right] = K + E\left[\|A\boldsymbol{x}^\gamma\|_2^2\right] - 2E[\boldsymbol{b}^T A\boldsymbol{x}^\gamma] + E\left[\frac{\boldsymbol{\omega}^T \operatorname{diag}(\boldsymbol{b})\big(F_\gamma(\boldsymbol{b}+\delta\boldsymbol{\omega}) - F_\gamma(\boldsymbol{b}-\delta\boldsymbol{\omega})\big)}{\delta}\right] + E\left[O\left(\|\boldsymbol{\epsilon}\|_2^3 + \frac{\|\boldsymbol{\epsilon}\|_2^2}{\delta} + \delta\right)\right],$$

where K is a constant independent of γ.

We remark that the fourth term on the right hand side above stems from a different estimator for the trace of a matrix than the one used by Santos and De Pierro in [5] and by Helou in [12], therefore arising to a different approximation than what would be obtained by straightforward application of [12, Theorem 1]. We have also been able to improve the error estimate of [12] because our particular case was simpler than the more general results presented there.

Theorem 1 right above immediately provides us with a practical recipe for finding a suitable regularization parameter: first generate a vector of m independent standard normal variables $\boldsymbol{\omega}$, then compute, for the selectable values of the regularization parameter $\{\gamma_1, \gamma_2, \dots, \gamma_r\}$, the reconstructions $\boldsymbol{x}^{\gamma_i}$, $\boldsymbol{x}^{\gamma_i}(\boldsymbol{b}+\delta\boldsymbol{\omega})$ and $\boldsymbol{x}^{\gamma_i}(\boldsymbol{b}-\delta\boldsymbol{\omega})$. Finally, find which γ_i minimizes

$$\|A\boldsymbol{x}^{\gamma_i}\|_2^2 - 2\boldsymbol{b}^T A\boldsymbol{x}^{\gamma_i} + \frac{\boldsymbol{\omega}^T \operatorname{diag}(\boldsymbol{b}) A\big(\boldsymbol{x}^{\gamma_i}(\boldsymbol{b}+\delta\boldsymbol{\omega}) - \boldsymbol{x}^{\gamma_i}(\boldsymbol{b}-\delta\boldsymbol{\omega})\big)}{\delta}. \tag{4}$$

Since the above quantity is, in average, a good approximation to $E\left[\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2\right]$, this parameter choice should lead to good results most of the time.

IV. Conclusions and Future Work

We have already performed experiments regarding tomographic image reconstruction which show that our approach is competitive and, in fact, performs better than existing methods most of the time, however lack of space makes it impossible to present them here. We plan to extend these experiments and generalize our method within a short period of time in order to publish the results as a full paper.

References

[1] D. L. Snyder and A. M. Hammoud, "Image recovery from data acquired with a charge-coupled-device camera," *Journal of the Optical Society of America A*, vol. 10, no. 5, pp. 1014–1023, 1993.

[2] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

[3] Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *Journal of the American Statistical Association*, vol. 80, no. 389, pp. 8–20, 1985.

[4] K. Lange and J. A. Fessler, "Globally convergent algorithms for maximum *a Posteriori* transmission tomography," *IEEE Transactions on Image Processing*, vol. 4, no. 10, pp. 1430–1438, 1995.

[5] R. J. Santos and Á. R. De Pierro, "A new parameters choice method for ill-posed problems with Poisson data and its application to emission tomographic imaging," *International Journal of Tomography & Statistics*, vol. 11, no. s09, pp. 33–52, 2009.

[6] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Kluwer Academic Publishers, 2000.

[7] F. Bauer and M. A. Lukas, "Comparing parameter choice methods for regularization of ill-posed problems," *Mathematics and Computers in Simulation*, vol. 81, no. 9, pp. 1795–1841, 2011.

[8] J. M. Bardsley and J. Goldes, "Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation," *Inverse Problems*, vol. 25, no. 9, p. 095005, 2009.

[9] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, "A discrepancy principle for Poisson data," *Inverse Problems*, vol. 26, no. 10, pp. 105 004–105 023, 2010.

[10] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero, "Efficient gradient projection methods for edge-preserving removal of poisson noise," *Inverse Problems*, vol. 25, no. 4, p. 045010, 2009.

[11] I. Csiszár, "Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems," *The Annals of Statistics*, vol. 19, no. 4, pp. 2032–2066, 1991.

[12] E. S. Helou, "A new parameter choice rule for optimization models for inverse problems with poisson noise," in *Anais do I Congresso de Matemática Aplicada e computacional da Região Sudeste - I CMAC Sudeste*. SBMAC, 2011.