

Statistical method for resolving the photon–photoelectron-counting inversion problem

Jinlong Wu^a, Tiejun Li^{a,*}, Xiang Peng^b, Hong Guo^b

^a LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, PR China

^b CREAM Group, State Key Laboratory of Advanced Optical Communication Systems and Networks (Peking University) and Institute of Quantum Electronics, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, PR China

ARTICLE INFO

Article history:

Received 10 December 2009

Received in revised form 12 October 2010

Accepted 13 October 2010

Available online 21 October 2010

Keywords:

Quantum key distribution

Photon statistics

Infinite binomial mixture model

Coarse-graining

Bayesian methods

Ill-posed problem

ABSTRACT

A statistical inversion method is proposed for the photon–photoelectron-counting statistics in quantum key distribution experiment. With the statistical viewpoint, this problem is equivalent to the parameter estimation for an infinite binomial mixture model. The coarse-graining idea and Bayesian methods are applied to deal with this ill-posed problem, which is a good simple example to show the successful application of the statistical methods to the inverse problem. Numerical results show the applicability of the proposed strategy. The coarse-graining idea for the infinite mixture models should be general to be used in the future.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Quantum key distribution (QKD) creates a secret key between authorized partners connected by a quantum channel and a classical authenticated channel [1]. The photon statistics of a QKD source is crucial for security analysis and needs to be characterized [2]. In doing so, experimentally, a passive scheme with a beam splitter (BS) and an inefficient detector (see Fig. 1) is used to monitor the photon statistics [2]. The inefficient detector can be treated as a virtual beam splitter placed in front of an ideal detector. Suppose that the probability of inputting n photons is p_n and the probability of detecting m photoelectrons by the detector is q_m , then q_m is the binomial transformation of p_n and p_n can be theoretically recovered by the inverse binomial transformation of q_m [3]

$$q_m = \sum_{n=m}^{\infty} p_n C_n^m \xi^m (1-\xi)^{n-m}, \quad \xi \in (0, 1), \quad (1)$$

$$p_n = \sum_{m=n}^{\infty} q_m C_m^n \xi^{-n} (1-\xi^{-1})^{m-n}, \quad \xi \in (0.5, 1). \quad (2)$$

where C_n^m is the combinatorial number of picking m unordered outcomes from n possibilities, and $\xi = t_B t_D$. Formally, the direct and inverse transformations satisfy the duality relation if we interchange the role of m and n in Eq. (1), and replace ξ by ξ^{-1} , we obtain the inverse transformation in Eq. (2).

* Corresponding author. Tel.: +86 10 62757592; fax: +86 10 62751801.

E-mail addresses: tieli@pku.edu.cn, tieli@math.pku.edu.cn (T. Li), xiangpeng@pku.edu.cn (X. Peng), hongguo@pku.edu.cn (H. Guo).

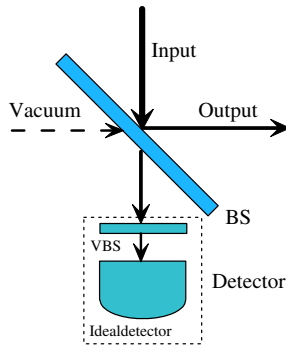


Fig. 1. The simplified passive scheme to monitor the photon statistics. BS: beam splitter (transmittance: t_B), VBS: virtual beam splitter (transmittance: t_D). The inefficient detector (efficiency: t_D) is modeled by VBS and an ideal detector (efficiency: 100%).

However, when a large number of photons are monitored with a finite-resolution and noisy detector, it is challenging to implement the direct inversion of p_n based on Eq. (2) on computer. For confirming this, in what follows, we prove that the inversion of photon–photoelectron-counting problem is ill-posed. To resolve this practical issue, a new statistical inversion method, i.e., Bayesian method, is proposed in our paper, in which the coarse-graining strategies for treating realistic detection resolution and estimating parameters are applied. Using this new method, the reconstruction of p_n with the data set from an actual QKD experiment [2] is carried out successfully. Since Bayesian method is based on the parameter estimation for the infinite binomial mixture model, naturally, it is compared with the classical EM algorithm in our paper. It is shown that the new method is more robust to deal with the photon–photoelectron-counting inversion problem than EM even though the effect from noisy detection is considered. Furthermore, to handle the infinite binomial mixture model numerically, we propose the coarse-graining idea to reduce it to a finite mixture model. We believe that this simple idea of piecewise approximation from traditional numerical analysis can be generalized and applied to other infinite mixture models provided that the mixture weight has suitable regularity.

The rest of the paper is organized as follows. In Section 2, we present some basic mathematical studies for the binomial decaying transformation Eqs. (1) and (2). In Section 3, the ill-posedness is mainly discussed. The statistical inversion method is proposed and implemented to the experimental data in Section 4. Finally we discuss some implementation issues and the possible extensions in the future.

2. Preliminary mathematical study

We will state the following basic properties for the direct and inverse binomial transformations without proof.

Lemma 1. Under the conditions that $p_n \geq 0$ and $\sum_{n=0}^{\infty} p_n = 1$, we have

$$\sum_{m=0}^{\infty} q_m = 1 \quad \text{and} \quad q_m \geq 0, \quad m = 0, 1, \dots \tag{3}$$

For the notational simplicity, we define the space of probability sequences $\{p_n\}_{n=0}^{\infty}$ as ℓ^1_+ , which is a convex set in ℓ^1 space. Suppose the random variable $X \sim \{p_n\}_{n=0}^{\infty}$ and $Y \sim \{q_m\}_{m=0}^{\infty}$, then

Lemma 2. If $\mathbb{E}X = \mu$, $\text{Var}(X) = \sigma^2$, then

$$\mathbb{E}Y = \xi\mu, \quad \text{Var}(Y) = \xi^2\sigma^2 + \xi(1 - \xi)\mu. \tag{4}$$

Lemma 3. If $p_n = C_N^n \eta^n (1 - \eta)^{N-n}$ ($n \leq N$), that is, X follows the binomial distribution $B(N, \eta)$ with parameters N and η , then

$$Y \sim B(N, \xi\eta). \tag{5}$$

Lemma 4. If X follows the Poisson distribution $\mathcal{P}(\lambda)$, then $Y \sim \mathcal{P}(\xi\lambda)$.

All the proof of the above lemmas can be done directly, which can be referred to [4].

Denote the linear mapping from $\{p_n\}$ to $\{q_m\}$ defined in Eq. (1) as

$$\mathbb{K} : \{p_n\} \mapsto \{q_m\} \tag{6}$$

and the corresponding inverse mapping as \mathbb{K}^{-1} . We have

Lemma 5. The mapping

$$\mathbb{K} : \ell_+^1 \rightarrow \ell_+^1, \quad \mathbb{K} : \ell^1 \rightarrow \ell^1. \quad (7)$$

Lemma 6. If $\zeta \in (0.5, 1)$, then the series on the right hand side of the inverse binomial transformation Eq. (2) absolutely converge to p_n if $\{q_m\}$ is given as in Eq. (1). Furthermore, the inverse transformation holds for any $\{p_n\} \in \ell^1$.

Proof. For the inverse binomial transformation

$$p_n = \zeta^{-n} (1 - \zeta^{-1})^{-n} \sum_{m=n}^{\infty} q_m C_m^n (1 - \zeta^{-1})^m, \quad (8)$$

define $b_m = |(1 - \zeta^{-1})^m C_m^n|$, then we have

$$\frac{b_{m+1}}{b_m} = |1 - \zeta^{-1}| \frac{m+1}{m+1-n}, \quad (9)$$

which tends to $|1 - \zeta^{-1}|$ when m goes to ∞ . If $\zeta \in (0.5, 1)$, then $|1 - \zeta^{-1}| \in (0, 1)$. We have the absolute convergence with the D'Alemberts ratio test and the comparison strategy.

Now suppose $\{q_m\}$ is given as in Eq. (1), then we may rearrange the terms by the absolute convergence

$$\begin{aligned} \sum_{m=n}^{\infty} q_m C_m^n \zeta^{-n} (1 - \zeta^{-1})^{m-n} &= \sum_{m=n}^{\infty} \left(\sum_{k=m}^{\infty} p_k C_k^m \zeta^m (1 - \zeta)^{k-m} \right) C_m^n \zeta^{-n} (1 - \zeta^{-1})^{m-n} \\ &= \sum_{k=n}^{\infty} C_k^n p_k \sum_{m=n}^k C_{k-n}^{m-n} (1 - \zeta)^{k-m} (\zeta - 1)^{m-n} = p_n + \sum_{k=n+1}^{\infty} C_k^n p_k \sum_{m=n}^k C_{k-n}^{m-n} (1 - \zeta)^{k-m} (\zeta - 1)^{m-n} = p_n. \end{aligned} \quad (10)$$

The last equality holds because of the Pascal's rule.

For the general $\{p_n\} \in \ell^1$ case, we have $\{q_m\} \in \ell^1$, thus the absolute convergence and the argument in Eq. (10) also holds. The proof is completed. \square

Proposition 1. For $\zeta \in (0.5, 1)$, \mathbb{K} is injective but not surjective from ℓ_+^1 to ℓ_+^1 .

Proof. Denote $\mathbf{p} = \{p_n\}_{n=0}^{\infty}$. From Lemma 6, if $\mathbb{K}\mathbf{p} = \mathbf{0}$ and $p \in \ell^1$, we have $p = \mathbf{0}$ by the inverse formula. \mathbb{K} is injective.

To prove that \mathbb{K} is not surjective in ℓ_+^1 , we construct a probability sequence $\{q_m\}$ such that $q_m = 0$ when m is odd. In this case, for all p_n when n is odd, we have $p_n < 0$ from the inverse transformation (8) since $(1 - \zeta^{-1}) \in (-1, 0)$. The proof is completed. \square

Proposition 2 (Continuity of \mathbb{K}). For any $\{p_n\}, \{\hat{p}_n\} \in \ell^1$, suppose $\mathbf{q} = \mathbb{K}\mathbf{p}$ and $\hat{\mathbf{q}} = \mathbb{K}\hat{\mathbf{p}}$, then

$$\|\hat{\mathbf{q}} - \mathbf{q}\|_{\ell^1} \leq \|\hat{\mathbf{p}} - \mathbf{p}\|_{\ell^1}. \quad (11)$$

The equality can be achieved in ℓ^1 but not in ℓ_+^1 . This shows the operator \mathbb{K} is a bounded linear operator in ℓ^1 with norm 1.

Proof. We have

$$\|\hat{\mathbf{q}} - \mathbf{q}\|_{\ell^1} = \sum_{m=0}^{\infty} \left| \sum_{n=m}^{\infty} (\hat{p}_n - p_n) C_n^m \zeta^m (1 - \zeta)^{n-m} \right| \leq \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} |\hat{p}_n - p_n| C_n^m \zeta^m (1 - \zeta)^{n-m} = \sum_{n=0}^{\infty} |\hat{p}_n - p_n| = \|\hat{\mathbf{p}} - \mathbf{p}\|_{\ell^1}.$$

From the above procedure it is easy to observe that the equality holds only when either $\hat{p}_n \geq p_n$ or $\hat{p}_n \leq p_n$ for all n . The proof is completed. \square

3. Ill-posedness of the inversion problem

Proposition 3 (Non-compactness of \mathbb{K}). As an operator from ℓ^1 to ℓ^1 , \mathbb{K} is not compact.

Proof. Take the unit vectors $(e_i)_n = \delta_{in}$, $n = 0, 1, 2, \dots$. We have

$$(\mathbf{q}_i := \mathbb{K}e_i)_m = C_i^m \zeta^m (1 - \zeta)^{i-m}, \quad m = 0, 1, \dots, i. \quad (12)$$

To show \mathbb{K} is not compact, we will prove for any sufficiently large M , there exist $N > M$ such that

$$\|\mathbf{q}_M - \mathbf{q}_N\|_{\ell^1} \geq C \sim \mathcal{O}(1), \tag{13}$$

where C is a constant which is independent of M and N .

For any M , define

$$\delta_M = \min\{q_{Mj}, j = 0, 1, \dots, M\} = \min\{\xi^M, (1 - \xi)^M\}. \tag{14}$$

From the de Moivre–Laplace theorem [5], the asymptotics of $B(n, \xi)$ can be approximated with the Gaussian $N(n\xi, n\xi(1 - \xi))$ perfectly when n is sufficiently large. We have

$$\Delta_N = \max\{q_{Nj}, j = 0, 1, \dots, N\} \sim \frac{1}{\sqrt{2\pi N\xi(1 - \xi)}} \rightarrow 0 \text{ as } N \rightarrow \infty. \tag{15}$$

We can take N large enough such that $\Delta_N < \delta_M$. With these choices, we obtain

$$\begin{aligned} \|\mathbf{q}_M - \mathbf{q}_N\|_{\ell^1} &= \sum_{j=0}^N |q_{Mj} - q_{Nj}| = \sum_{j=0}^M (q_{Mj} - q_{Nj}) + \sum_{j=M+1}^N q_{Nj} \geq 1 + \sum_{j=M+1}^N q_{Nj} - (M + 1) \min\{\xi^M, (1 - \xi)^M\} \\ &\geq 1 - (M + 1)\xi^M. \end{aligned} \tag{16}$$

When M goes to infinity, $(M + 1)\xi^M \rightarrow 0$, which completes the proof. \square

Though \mathbb{K} is not compact, the ill-posedness of the inverse problem can be easily seen from Eq. (2). Suppose we have a small perturbation $\mathbf{q} = \varepsilon_0 \mathbf{e}_k$, where $\varepsilon_0 \ll 1$ and k is large. From Eq. (2), we obtain the k th component of \mathbf{p} as

$$p_k = \varepsilon_0 \xi^{-k}. \tag{17}$$

Since $\xi \in (0.5, 1)$, ξ^{-k} will be extremely large when $k \gg 1$, which is about 10^7 in the experimental data. This ill-posedness drives us to take some other approach to handle this problem.

Another aspect of the ill-posedness of Eq. (2) is that it is the sum of an alternating series, which brings huge round-off errors and thus achieves no accuracy for the final result though it is convergent, theoretically. This can be explicitly shown in Fig. 2. We set $\xi = 0.7$ in this numerical example. We first choose a Gaussian-like discrete distribution \mathbf{p} which has support at $n = 0, 1, \dots, 160$. Then we apply the direct binomial transformation Eq. (1) to obtain \mathbf{q} which has the same support. The direct transformation is well-conditioned. Now we try to recover \mathbf{p} from the inverse binomial transformation Eq. (2). The computed \mathbf{p} oscillates at the scale $\mathcal{O}(10^8)$ with the traditional floating point computation as can be shown in Fig. 2(b). If we track the converging history for some fixed n , say $n = 50$, we find that with oscillations at the scale $\mathcal{O}(10^{16})$ at the beginning, it converges to 1.5715 finally. But the exact value $p_{50} = 0.0043$. This is exactly due to the huge round-off error and the ill-posedness of this inversion formula. In our problem setting, the support indices of \mathbf{p} are in $\mathcal{O}(10^7)$, which is not an applicable range of the inverse formula. Similar situation holds for different choices of the parameter ξ .

The ill-posedness of \mathbb{K}^{-1} makes that Eq. (2) is not useful for the problems with large photon number. We will explore this issue from a new viewpoint in the next section.

4. Statistical inversion

4.1. Coarse-graining of the model

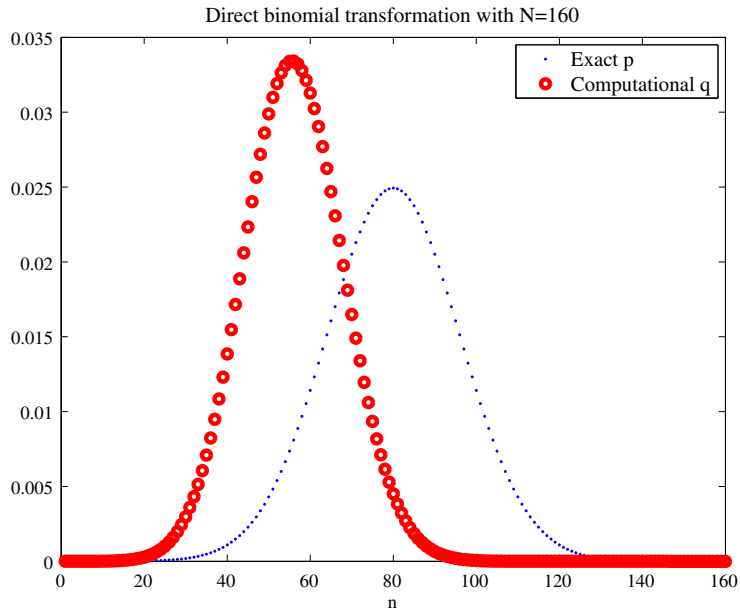
As the deterministic inversion formula does not work for the large number of photons, we propose a statistical inversion method to resolve this inverse problem for the following three reasons:

- (1) The probability distribution q_m cannot be obtained exactly in the realistic measurements. Instead, we only observe some realizations of it. This means that the statistical fluctuation of q_m cannot be neglected, which is amplified by the direct inversion with Eq. (2) due to the ill-posedness. What is more, we only know the interval that each observation of the number of photoelectrons fall in from each measurement, which corresponds to the sum of q_m in different intervals.
- (2) The detection noise exists in the measurements. We should take into account the effect in the inversion methods among which the statistical inversion method is naturally chosen.
- (3) There is a perfect statistical interpretation for the direct binomial transformation Eq. (1), which makes the statistical inversion strategy more natural than the other possible deterministic methods.

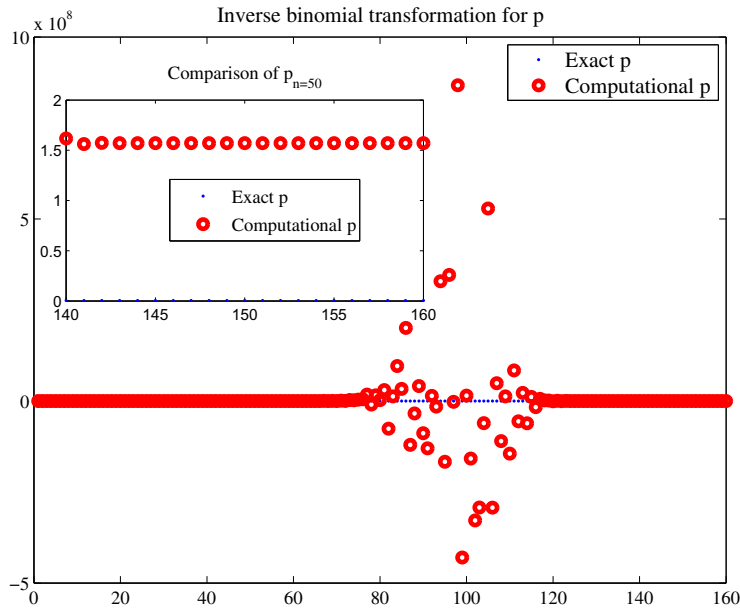
The statistical approach is based on the following basic observations.

Proposition 4. *The transformation Eq. (1) can be interpreted as an infinite binomial mixture model*

$$q_m = \sum_{n=0}^{\infty} p_n B(m; n, \xi), \tag{18}$$



(a) Direct binomial transformation from a known \mathbf{p}



(b) Inverse binomial transformation from the obtained \mathbf{q}

Fig. 2. The behavior of the direct and inverse binomial transformations. In (a), the pre-set Gaussian like discrete distribution \mathbf{p} is shown with dots. The \mathbf{q} obtained by direct transformation Eq. (2) is shown with circles. In (b), the reconstructed \mathbf{p} from Eq. (2) is shown with circles compared with the exact \mathbf{p} shown with dots. The reconstructed \mathbf{p} oscillates at the scale $\mathcal{O}(10^8)$, which reflects the ill-posedness of the inversion formula. The inset in (b) shows the converging history of p_{50} compared with the exact value 0.0043. The computed p_{50} by Eq. (2) is 1.5715, which has no probabilistic meaning at all! The parameter $\xi = 0.7$ in this example.

where $B(m; n, \xi)$ is the binomial distribution with parameters n and $\xi \in (0, 1)$. We take the definition $B(m; n, \xi) = 0$ if $m > n$.

The direct incorporation of the probabilistic interpretation Eq. (18) leads to the maximum likelihood estimate (MLE) of \mathbf{p} associated with the likelihood function through T number of measurements for $\{m_t\}$, that is

$$L(m_{1:T}|\mathbf{p}) = \prod_{t=1}^T \left(\sum_{n=0}^{\infty} p_n B(m_t; n, \xi) \right). \tag{19}$$

But determining infinite parameters p_n from finite number of measurements is not a feasible task. And the measurement resolution of m is finite due to the accuracy of measuring instrument. Thus, we take the following two coarse-graining strategies to make our statistical model more applicable:

- (1) *Coarse-graining of the measurements.* In realistic measurements, we only know the measured number of photoelectrons falls in K different disjoint intervals

$$I_1 \cup I_2 \cup \dots \cup I_K = [0, M_1] \cup [M_1 + 1, M_2] \cup \dots \cup [M_{K-1} + 1, M_K], \tag{20}$$

where $M_K = \infty$. We will take the convention that we only consider the integer points in the interval here and in what follows. With this partition, it naturally leads to the infinite mixture of K -multinomial distributions

$$\tilde{q}_k = \sum_{n=0}^{\infty} p_n B_k(n, \xi), \quad k = 1, 2, \dots, K, \tag{21}$$

where $B_k(n, \xi) = \sum_{m \in I_k} B(m; n, \xi)$.

- (2) *Coarse-graining of the parameters.* We should also truncate the infinite parameter inversion problem to a finite parameter problem. We take the following piecewise constant approximation to the probability weights $\{p_n\}$. We approximately need

$$p_n = \text{Constant} = \hat{p}_l, \quad n \in J_l = [N_{l-1} + 1, N_l], \quad l = 1, 2, \dots, L + 1, \tag{22}$$

where $N_0 = -1, N_{L+1} = \infty$. And we define $\Delta N_l = N_l - N_{l-1}$ ($l = 1, 2, \dots, L + 1$). To ensure the normalization, we obtain $\hat{p}_{L+1} = 0$ and

$$\sum_{l=1}^L \Delta N_l \hat{p}_l = 1. \tag{23}$$

With this piecewise constant approximation, we have

$$\tilde{q}_k = \sum_{n=0}^{\infty} p_n B_k(n, \xi) \approx \sum_{l=1}^L \tilde{p}_l B_{kl}(\xi). \tag{24}$$

Here, we take the definition

$$\tilde{p}_l = \hat{p}_l \Delta N_l, \quad B_{kl}(\xi) = \frac{1}{\Delta N_l} \sum_{n \in J_l} B_k(n, \xi). \tag{25}$$

It is easy to verify that

$$\sum_{l=1}^L \tilde{p}_l = 1, \quad \sum_{k=1}^K B_{kl}(\xi) = 1 \tag{26}$$

from Eqs. (23) and (25). The final formulation Eq. (24) reduces the infinite binomial mixture model to a finite mixture model.

We should emphasize that the reduction procedure presented above is quite general for any infinite mixture models. With this piecewise constant approximation, we take the basic assumption about the smoothness of the mixture weights, which is reasonable for many physical models. What is more, we can obtain more accurate \mathbf{p} with a sequential refinement step by step. At the very beginning, we can take a very coarse division with which we get a macroscopic view about the result. Then we refine the grid and the results at the coarser scale supply a very good initial to speed up the convergence. More degrees of freedom, better accuracy. In the finest scale, we recover the original mixture model with the only approximation that we truncate the components of \mathbf{p} to a finite size. We believe this would be a general procedure to deal with the future infinite mixture models.

The MLE for this reduced model will be

$$\max_{\tilde{p}_l \geq 0, \sum_l \tilde{p}_l = 1} L(Z_{1:T} | \tilde{\mathbf{p}}) = \prod_{t=1}^T \left(\sum_{l=1}^L \tilde{p}_l B_{Z_t, l}(\xi) \right) = \prod_{k=1}^K \left(\sum_{l=1}^L \tilde{p}_l B_{kl}(\xi) \right)^{\#(k)} \tag{27}$$

where $Z_t \in \{1, 2, \dots, K\}$ is the coarse-grained measurement and $\#(k)$ is the number of appearance of k in $Z_{1:t}$. The natural candidate to obtain $\tilde{\mathbf{p}}$ is the classical EM algorithm.

4.2. EM algorithm

The real photoelectron data from QKD experiment [2] is shown in Fig. 3. There are totally 35,658 measurements. Due to the accuracy of the measuring instrument, the number of the observed photoelectrons falls into 37 intervals which have the range from 1.35×10^7 to 1.55×10^7 . In the experiment, the parameter $\xi = 0.76$.

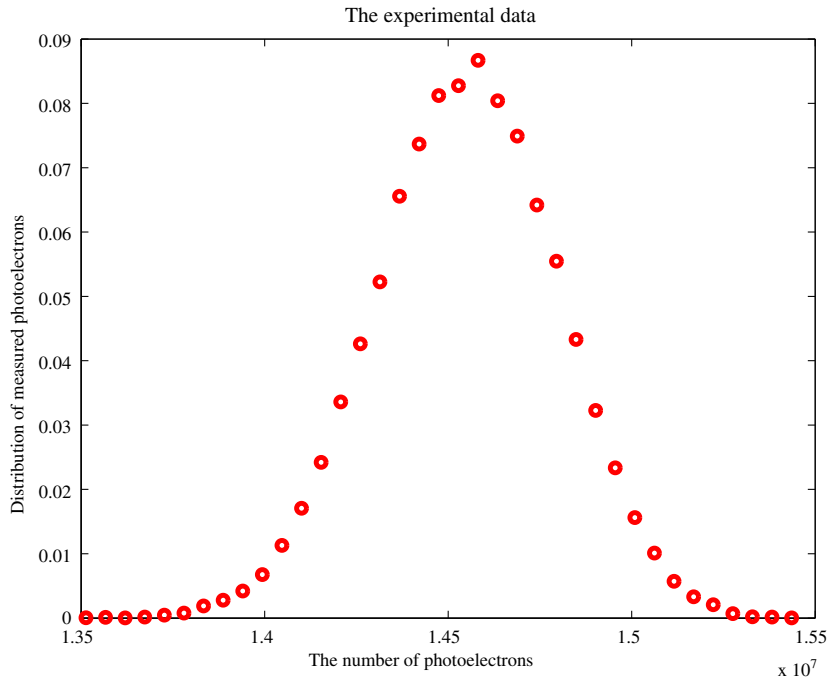


Fig. 3. The empirical distribution of the measured photoelectrons. Due to the accuracy of the measuring instrument, the number of the observed photoelectrons falls into 37 intervals which have the range from about 1.35×10^7 to 1.55×10^7 .

To apply the EM algorithm, we reformulate this problem with hidden variables $C_{1:T}$, which represent the class label that the t th sample belongs to. In this case the full likelihood function including the hidden variables will be

$$L(Z_{1:T}, C_{1:T} | \tilde{\mathbf{p}}) = \prod_{t=1}^T \tilde{p}_{C_t} B_{Z_t, C_t}(\xi) \quad (28)$$

The EM algorithm for this finite mixture model is composed of two steps.

Algorithm 1 (EM algorithm). The mixture weight $\tilde{\mathbf{p}}$ is obtained by the iterations of the following two steps.

- Step 1. E-step to compute the expectation values for the membership variables of each class k :

$$y_{kl} = \frac{\tilde{p}_l B_{kl}(\xi)}{\sum_{l=1}^L \tilde{p}_l B_{kl}(\xi)}, \quad k = 1, 2, \dots, K; \quad l = 1, 2, \dots, L. \quad (29)$$

- Step 2. M-step to compute the mixture weight

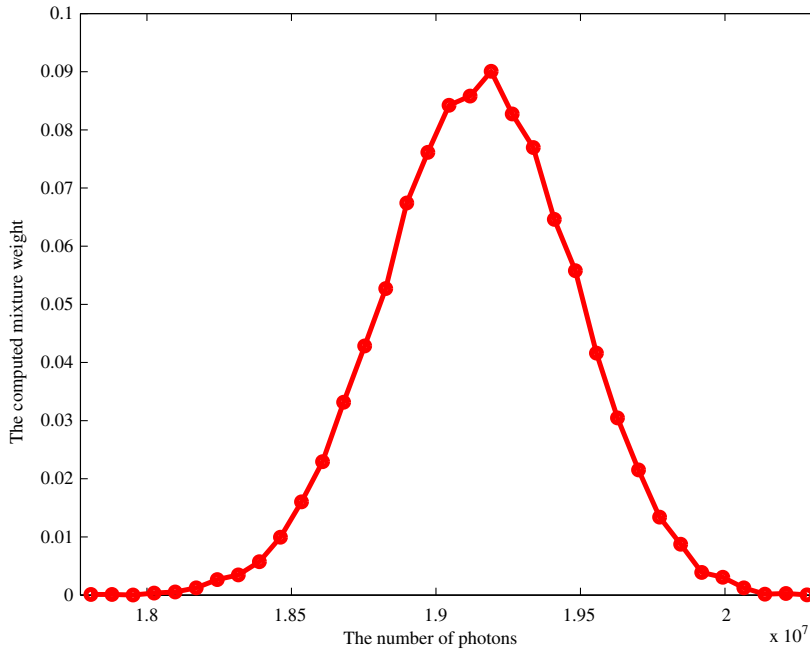
$$\tilde{p}_l = \frac{1}{T} \sum_{k=1}^K \#(k) y_{kl}, \quad l = 1, 2, \dots, L. \quad (30)$$

With the given data, $K = 37$ is chosen. Now we first take $L = 35$ to get a macroscopic view about the mixture weights \tilde{p}_l at the coarsest scale. To determine the approximate support of p_n , we take advantage of Eq. (4) which gives

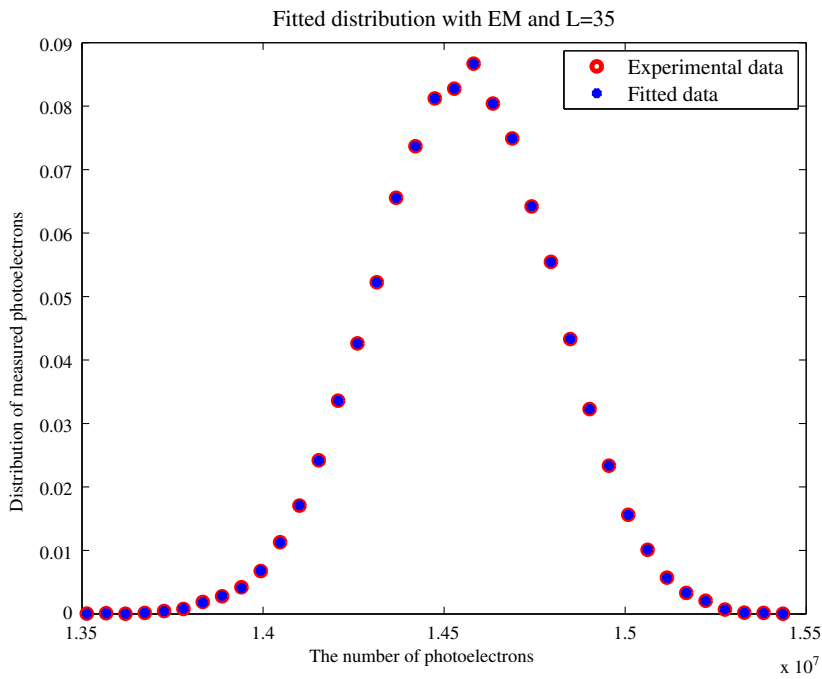
$$[1.35, 1.55] \times 10^7 / 0.76 \approx [1.777, 2.032] \times 10^7.$$

We subdivide the interval $[1.777 \times 10^7, 2.032 \times 10^7]$ into L segments and apply the classical EM algorithm. The results are shown in Fig. 4. With the computed \tilde{p}_l in Fig. 4(a), we obtain the fitted distribution shown with solid dots in Fig. 4(b). Compared with the experimental distribution, it works pretty well.

To increase the accuracy and resolution, we take larger $L = 50$. In this case, we hope to get better result. But the computed \tilde{p}_l shows wild oscillations even though the fitted distribution looks perfect (Fig. 5). The further increase of L will cause more disastrous oscillations but with better fitting. This fact is due to the ill-posedness of the inversion problem, which has been discussed in Section 3. From the physical point of view, these oscillations are not what we expected and it is believed that the mixture weights should have better regularity among all the possible candidates. We should select the smoother one with reasonable assumptions. In the inverse problem community, one introduces the additional deterministic regularization



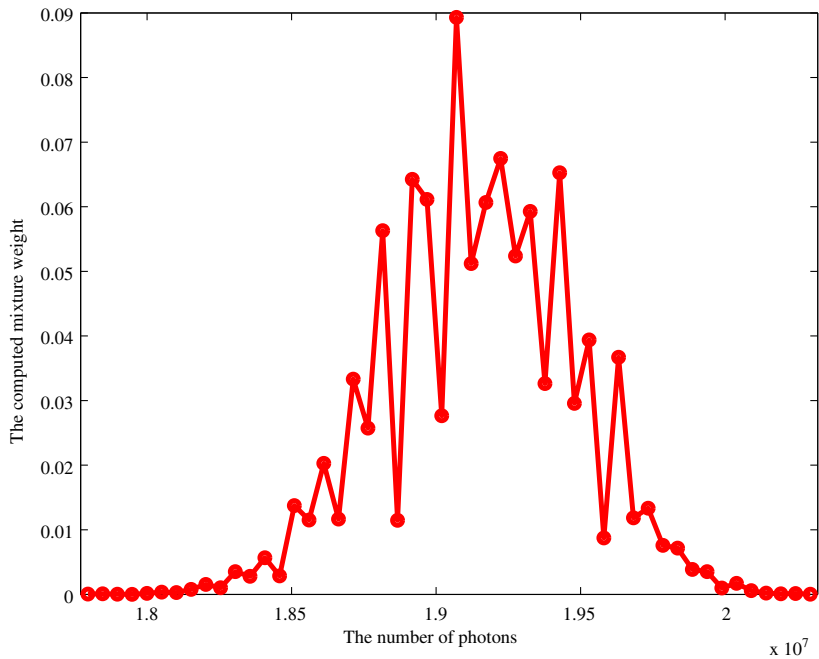
(a) The computed mixture weight with EM algorithm and $L = 35$



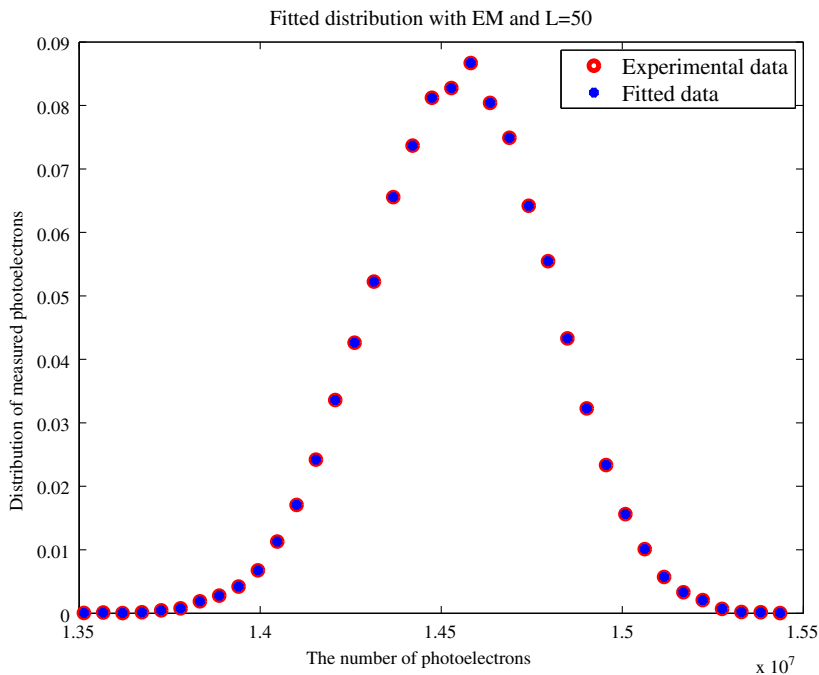
(b) The comparison between the real data and the fitted distribution

Fig. 4. The results obtained by EM algorithm and $L = 35$. (a) shows the computed mixture weights \bar{p}_i , in which only the piecewise constant values are drawn. With this coarse-grained \bar{p}_i , the fitted distribution shown with solid dots in (b). Compared with the experimental distribution, it fits quite well.

terms to deal with this ill-posedness, such as the Tikhonov regularization or Landwieber iteration, etc. [6]. Here we take another regularization strategy – the Bayesian framework – since it is more naturally embedded in our statistical inversion approach [7].



(a) The computed mixture weight with EM algorithm and $L = 50$



(b) The comparison between the real data and the fitted distribution when $L = 50$

Fig. 5. The results obtained by EM algorithm and $L = 50$. From (a) we observe the wild oscillations in the computed mixture weights \bar{p}_i . Even with this oscillating \bar{p}_i , we still have the perfect match between the experimental distribution and the fitted distribution.

4.3. Bayesian methods

The most important message in [7] to deal with the statistical inverse problem is to apply the prior distribution π_{pr} , which plays the role of regularization in the deterministic inversion. Their relation will be identified in what follows.

With this general framework, once the priori distribution $\pi_{pr}(\tilde{\mathbf{p}})$ is given, the posterior distribution $\pi_{post}(\tilde{\mathbf{p}})$ has the form

$$\pi_{post}(\tilde{\mathbf{p}}|Z_{1:t}) \propto L(Z_{1:T}|\tilde{\mathbf{p}})\pi_{pr}(\tilde{\mathbf{p}}). \tag{31}$$

For the finite mixture model with unknown mixture weights, the classical conjugate prior distribution will be the Dirichlet distribution $Dir(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_L)$ and $\sum_{l=1}^L \alpha_l = 1$. The sampling of the posterior distribution can be done by Gibbs sampling method by introducing the hidden variable C_t as before. But this prior distribution does not help for the current statistical problem since it does not take into account the special feature of this problem. We propose the smooth prior [7] to ensure the regularity of the mixture weights under the basic assumption that the distribution p_n is smooth enough. That is,

$$\pi_{pr}(\tilde{\mathbf{p}}) \propto \exp \left\{ -\gamma \sum_{l=1}^{L+1} (\tilde{p}_l - \tilde{p}_{l-1})^2 \right\} \cdot \chi_A(\tilde{\mathbf{p}}), \tag{32}$$

where $\chi_A(\cdot)$ is the indicator function with support A . Here, A is defined as

$$A = \left\{ \tilde{p}_l \geq 0, \sum_{l=1}^L \tilde{p}_l = 1 \right\}.$$

We take the definition $\tilde{p}_0 = \tilde{p}_{L+1} = 0$ for the simplification of the notations, and γ is the regularization parameter. Note that the function in the exponent is the discrete counterpart of H^1 norm (the L^2 norm is automatically bounded because $\tilde{\mathbf{p}}$ is a probability vector).

4.3.1. Maximum a posteriori (MAP) estimate

With the posteriori distribution, one approach to obtain the mixture weights is the *maximum a posteriori (MAP) estimate* through which $\tilde{\mathbf{p}}$ is found

$$\tilde{\mathbf{p}} = \arg \max_{\tilde{\mathbf{p}}} L(Z_{1:T}|\tilde{\mathbf{p}})\pi_{pr}(\tilde{\mathbf{p}}). \tag{33}$$

This approach is essentially equivalent to a deterministic regularization method. Define the function $H(\tilde{\mathbf{p}}|Z_{1:T})$ as

$$H(\tilde{\mathbf{p}}|Z_{1:T}) = -\ln L(Z_{1:T}|\tilde{\mathbf{p}}), \tag{34}$$

then solving Eq. (33) is equivalent to

$$\min \left(H(\tilde{\mathbf{p}}|Z_{1:T}) + \gamma \sum_{l=1}^{L+1} (\tilde{p}_l - \tilde{p}_{l-1})^2 \right) \tag{35}$$

with constraints $\tilde{\mathbf{p}} \in A$.

No matter whether Eqs. (33) or (35) is used, it is too complicated to be solved because of the multi-variate polynomial form Eq. (27) with high degree. Now we introduce the hidden variable C_t and then take the same idea to obtain $\tilde{\mathbf{p}}$ by alternating iterations as EM algorithm. To be specific, we aim at maximizing the extended posterior

$$\max_{\tilde{\mathbf{p}}} L(Z_{1:T}, C_{1:T}|\tilde{\mathbf{p}})\pi_{pr}(\tilde{\mathbf{p}}) \tag{36}$$

with the EM algorithm, which is called variational EM in some literatures.

Algorithm 2 (EM algorithm for MAP). The MAP of the mixture weight $\tilde{\mathbf{p}}$ can be obtained by the iterations of the following two steps.

- Step 1. E-step to compute the expectation values for the membership variables of each class k :

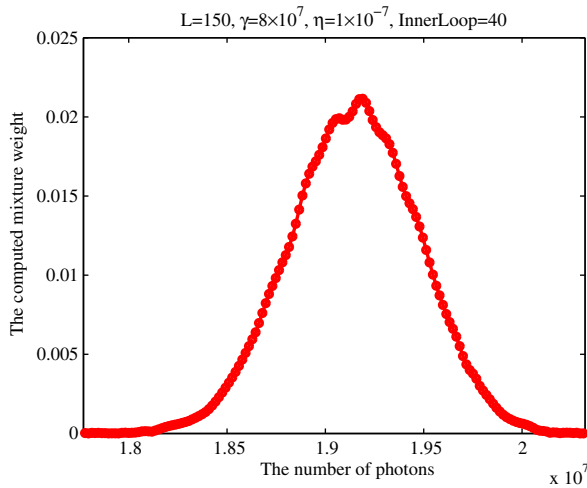
$$y_{kl} = \frac{\tilde{p}_l B_{kl}(\xi)}{\sum_{l=1}^L \tilde{p}_l B_{kl}(\xi)}, \quad k = 1, 2, \dots, K; \quad l = 1, 2, \dots, L. \tag{37}$$

- Step 2. M-step to compute the mixture weight

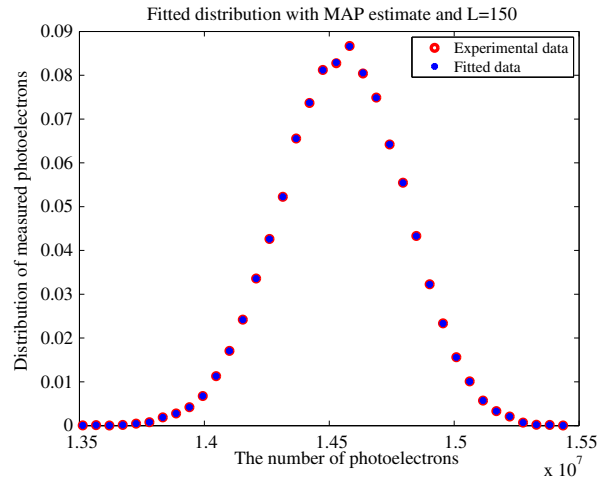
$$\max_{\tilde{\mathbf{p}}} \left(\prod_{l=1}^L \tilde{p}_l^{\sum_{k=1}^K z^{(k)} y_{kl}} \right) \cdot \exp \left\{ -\gamma \sum_{l=1}^{L+1} (\tilde{p}_l - \tilde{p}_{l-1})^2 \right\} \cdot \chi_A(\tilde{\mathbf{p}}). \tag{38}$$

In the M-step above, to satisfy the constraints that $\tilde{\mathbf{p}} \in A$, we take the following procedure. We introduce the generalized variables $y_l \in \mathbb{R}$ such that

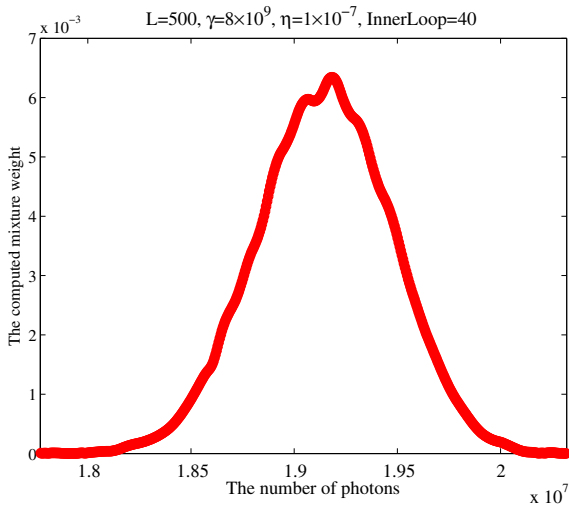
$$\tilde{p}_l = \frac{\exp(y_l)}{\sum_{l=1}^L \exp(y_l)}. \tag{39}$$



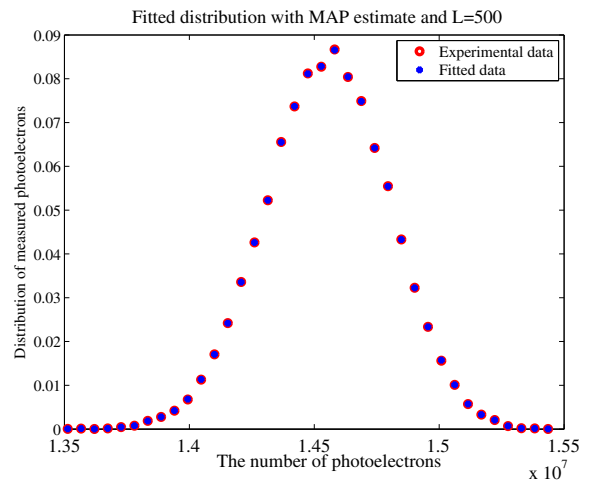
(a) The computed mixture weight with MAP estimate and $L = 150$



(b) The comparison between the real data and the fitted distribution when $L = 150$ by MAP estimate.



(c) The computed mixture weight with MAP estimate and $L = 500$



(d) The comparison between the real data and the fitted distribution when $L = 500$ by MAP estimate.

Fig. 6. The results obtained by MAP estimate and $L = 150$ and 500 . The wild oscillations are removed by the introduction of the smooth prior. From (a) and (c) we can obtain a relatively smooth profile for the computed mixture weights \hat{p}_i . With this \hat{p}_i , we have the perfect match between the experimental distribution and the fitted distribution ((b) and (d)).

With this transformation, the obtained $\hat{\mathbf{p}}$ will satisfy the probabilistic constraints automatically, and the optimization with respect to y_l becomes an un-constrained problem. Then we can choose all the possible methods for un-constrained optimizations. Since the accurate y is only an intermediate step in the iterations and is not necessary to be known, simply, we just use the steepest descent method. Here, we remark that Eq. (39) is not a one-to-one mapping, but a simple and reasonable one. The computed mixture weights through MAP estimation are shown in Fig. 6 when $L = 150$ and 500 . With the smooth priors, the new profile of $\hat{\mathbf{p}}$ is quite smooth even with so larger L . To suppress the oscillations, we take the parameters $\gamma = 8 \times 10^7$ and the time step size $\eta = 1 \times 10^{-7}$ in the steepest descent method in the M-step when $L = 150$. In each M-step, we take 40 inner iterations to find an improved $\hat{\mathbf{p}}$. When $L = 500$, we should take $\gamma = 8 \times 10^9$ to suppress the wilder oscillations. The other parameters are the same as $L = 150$. Whether $L = 150$ or 500 , we get the perfect match between the fitted distribution and the experimentally obtained histogram.

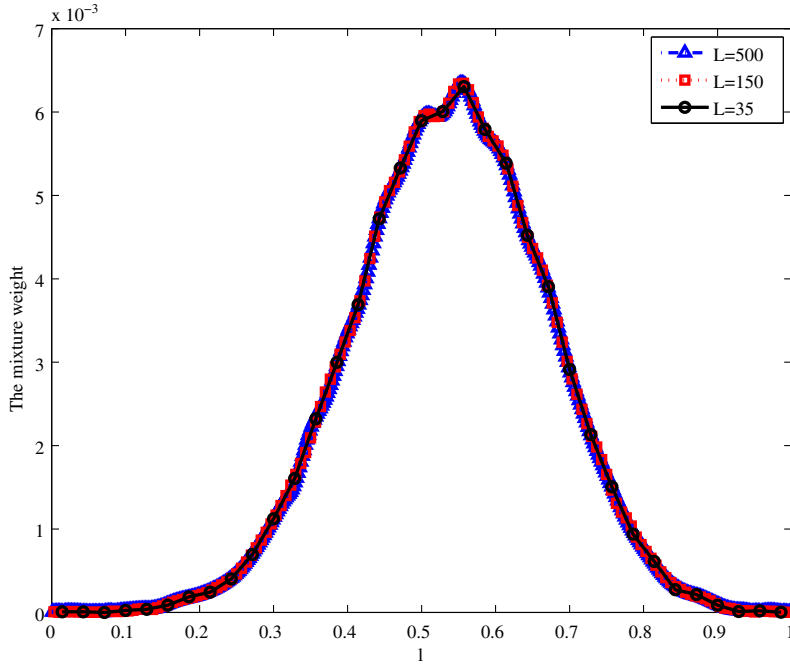


Fig. 7. The comparison of three obtained mixture weights \tilde{p}_l for $L = 35, 150, 500$ when the domain is normalized to the common interval $[0, 1]$. The perfect coincidence of the three curves validates the convergence of the results.

To show the convergence of the computed mixture weights after adding the smooth prior, we normalize the domain of \tilde{p}_l to the common real interval $[0, 1]$, and plot these \tilde{p}_l for $L = 35, 150, 500$ in the same figure (Fig. 7). The perfect coincidence of the three curves validates the convergence of the results.

4.3.2. Bayesian sampling

The virtue of the Bayesian approach is that it cannot only get an estimate about the parameters such as the MAP, but also the whole posterior distribution. It will be easier to extract different estimates from the whole distribution. To sample the posterior distribution, we take the Gibbs sampling technique [8] which can be summarized as follows.

Algorithm 3 (Gibbs sampling for the posterior distribution). The posterior distribution Eq. (31) of the mixture weight $\tilde{\mathbf{p}}$ can be sampled by the repeated iterations of the following two steps:

- Step 1. Sample the hidden variable Z_l conditioned on the known $\tilde{\mathbf{p}}$. This can be done with the following substeps:
 - Step 1.1. Computing the expectation values for the membership variables of each class. That is

$$y_{kl} = \frac{\tilde{p}_l B_{kl}(\xi)}{\sum_{l=1}^L \tilde{p}_l B_{kl}(\xi)}, \quad k = 1, 2, \dots, K; \quad l = 1, 2, \dots, L, \tag{40}$$

which is the same as the EM algorithm.

- Step 1.2. Suppose the number of observed measurements falls in class k is $\sharp(k)$, then sample the L -vector $\mathbf{n}_k \sim \text{Multi}(\sharp(k), (y_{kl})_{l=1}^L)$, the multinomial distribution with parameter $(y_{k1}, y_{k2}, \dots, y_{kL})$.
- Step 1.3. Compute the sum of vectors

$$\mathbf{n} = \sum_{k=1}^K \mathbf{n}_k, \tag{41}$$

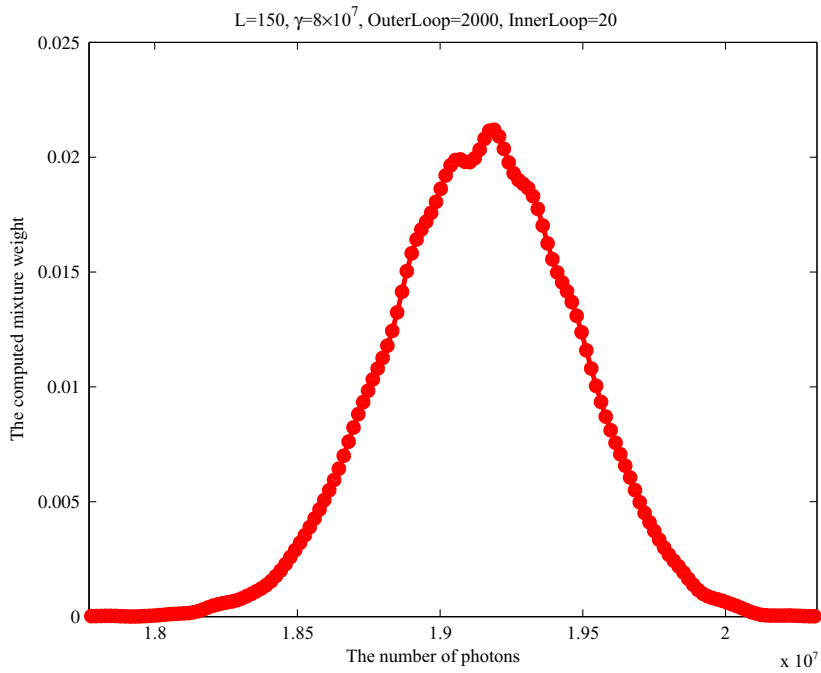
which satisfies $\sum_{l=1}^L n_l = T$.

- Step 2. Sample the mixture weight $\tilde{p}_l \sim f(\tilde{\mathbf{p}})$ conditioned on the known Z_l :

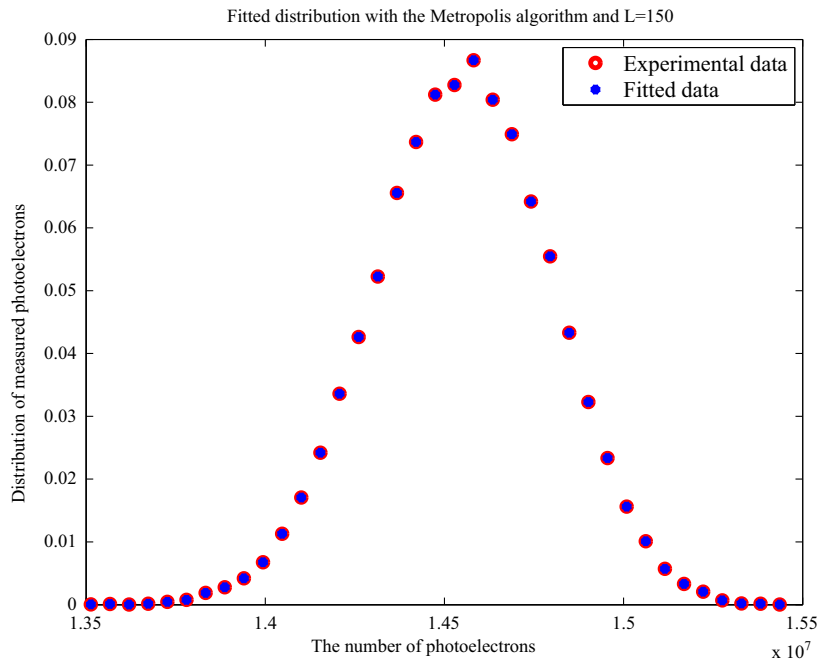
$$f(\tilde{\mathbf{p}}) \propto \left(\prod_{l=1}^L \tilde{p}_l^{n_l} \right) \cdot \exp \left\{ -\gamma \sum_{l=1}^{L+1} (\tilde{p}_l - \tilde{p}_{l-1})^2 \right\} \cdot \chi_A(\tilde{\mathbf{p}}). \tag{42}$$

To sample $f(\tilde{\mathbf{p}})$, we take the following proposal to ensure the probabilistic constraints $\tilde{\mathbf{p}} \in A$.

- Step 2.1. Generate an $(L - 1)$ -point uniformly distributed random variable $H \in \{1, 2, \dots, L - 1\}$.
- Step 2.2. For $\tilde{p}_H, \tilde{p}_{H+1}$, generate a uniformly distributed random variable $R \sim \mathcal{U}[0, 1 - \sum_{l \neq H, H+1} \tilde{p}_l]$.



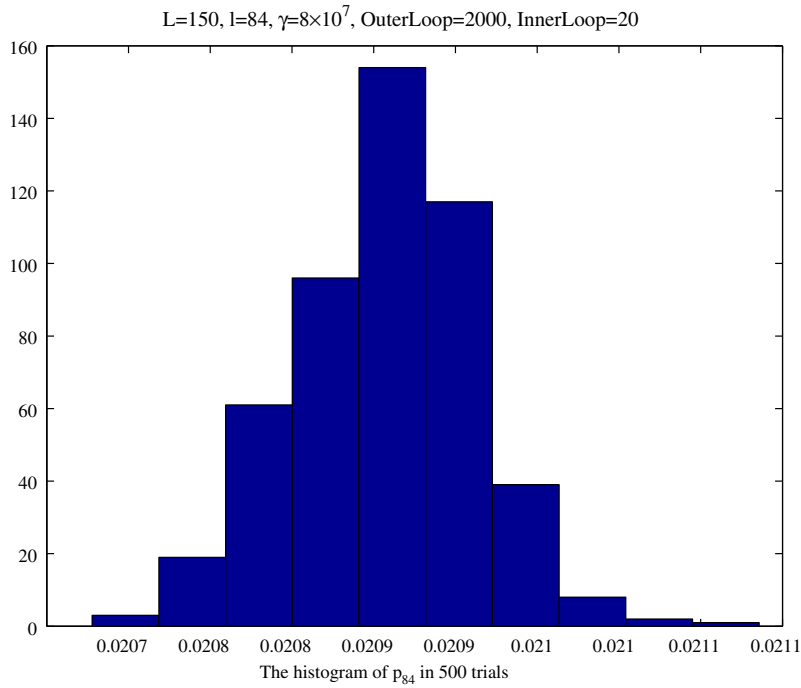
(a) The computed posterior mean for the mixture weight with Gibbs sampling strategy and $L = 150$



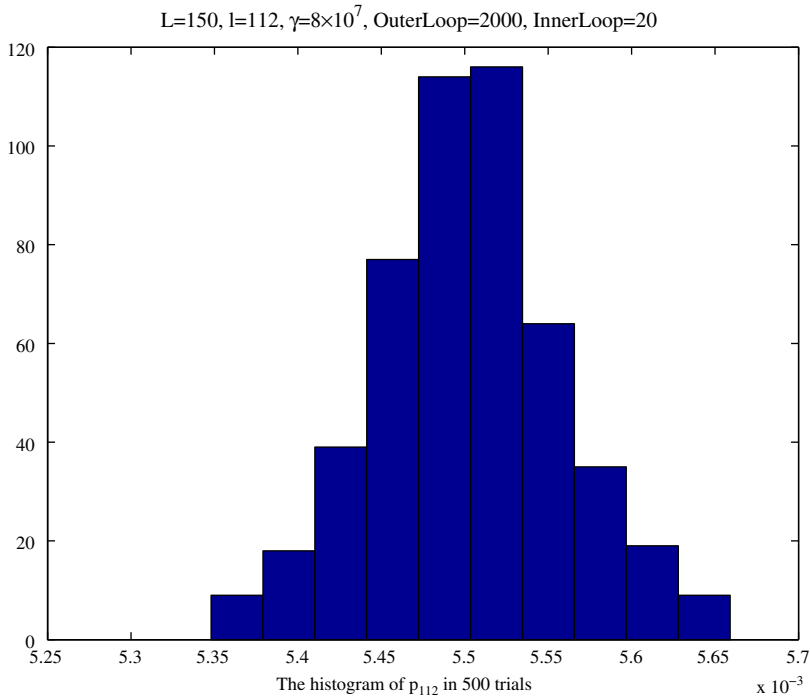
(b) The comparison between the real data and the fitted distribution when $L = 150$ by taking the posterior mean.

Fig. 8. The results obtained by Gibbs sampling and $L = 150$. The wild oscillations are removed by the introduction of the smooth prior. The posterior mean for the mixture weight is very similar to the MAP estimate. The perfect match between the experimental and the fitted distribution is also ensured.

- Step 2.3. Take $\tilde{p}_H = R, \tilde{p}_{H+1} = (1 - \sum_{l \neq H, H+1} \tilde{p}_l) - R$ and keep the other components of $\tilde{\mathbf{p}}$. Then we apply the Metropolis–Hastings strategy to make the decision to accept or reject the proposal.



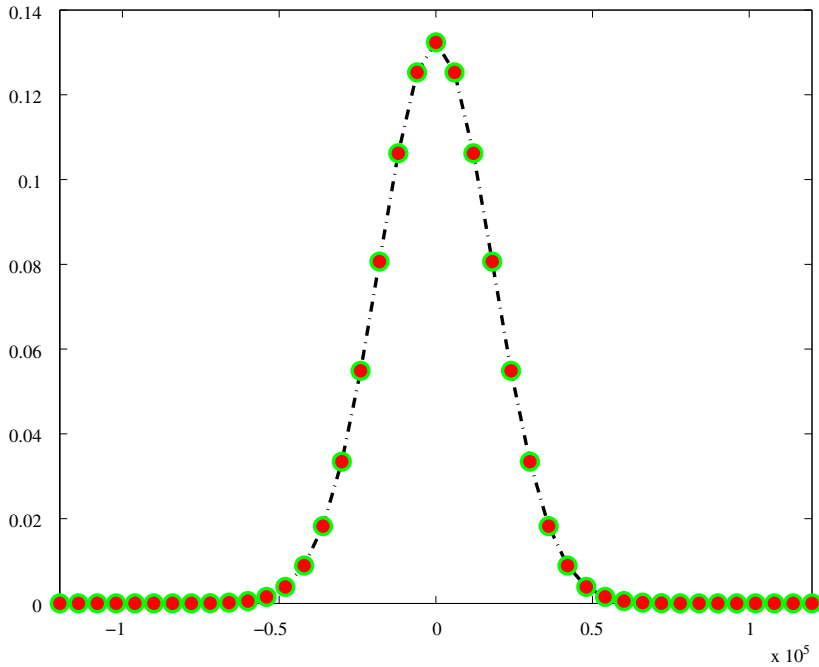
(a) The obtained histogram for \tilde{p}_l with Gibbs sampling strategy when $L = 150$ and $l = 84$.



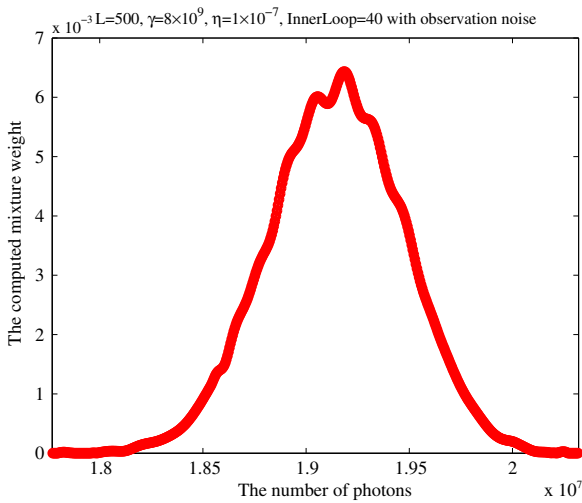
(b) The obtained histogram for \tilde{p}_l with Gibbs sampling strategy when $L = 150$ and $l = 112$.

Fig. 9. The histograms for \tilde{p}_l (left panel $l = 84$, right panel $l = 112$) from 500 samples with Gibbs sampling for the posterior distribution when $L = 150$. From this histogram, one can compute the posterior mean or some other statistical quantities.

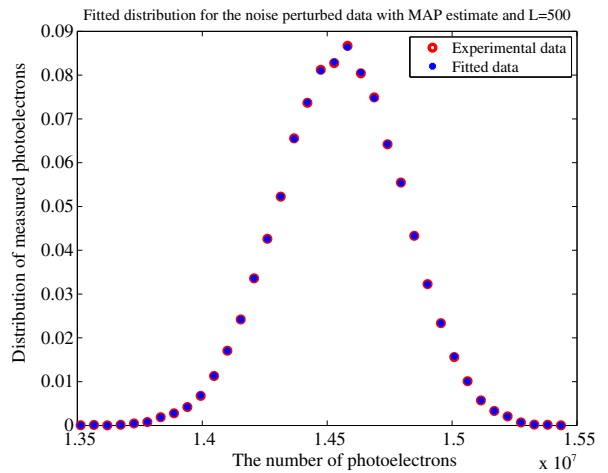
probability mass function of discrete **noise** distribution when noise exists (K=20,span=6000)



(a) The assumed detection noise.



(b) The computed mixture weight for the assumed noise perturbed data with MAP estimate and $L = 500$.

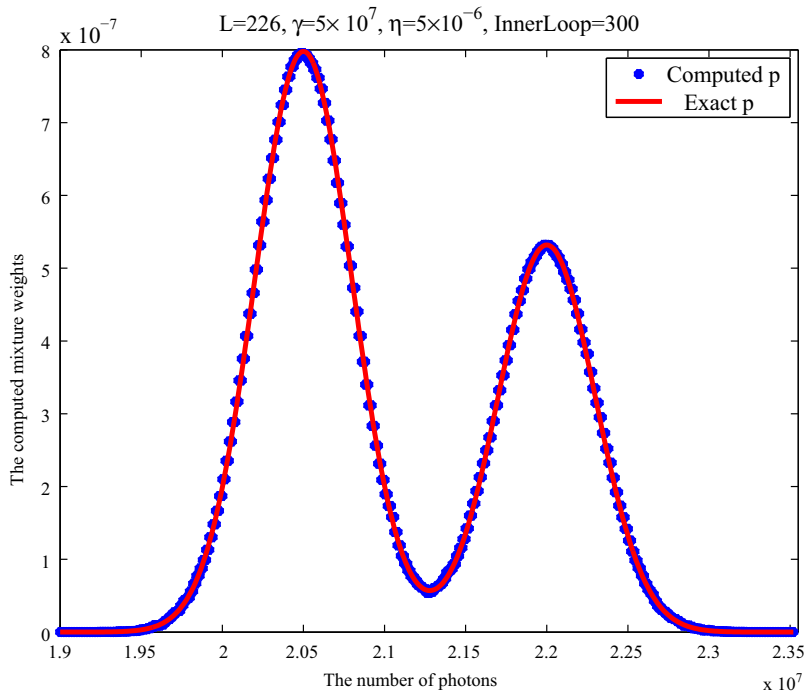


(c) The comparison between the real data and the fitted distribution when $L = 500$ by MAP estimate.

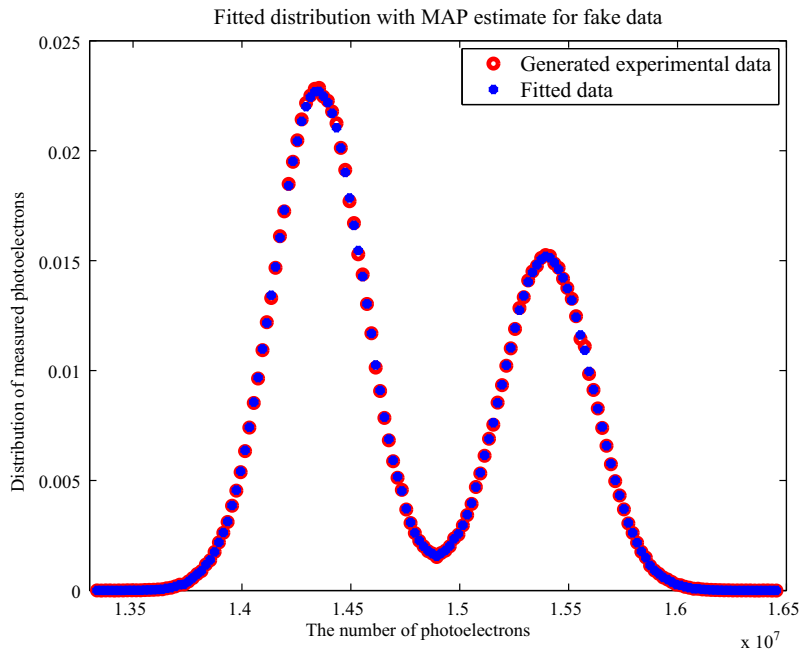
Fig. 10. The results obtained by MAP estimate for the data perturbed by the assumed noise and $L = 500$. The detection noise is assumed in $O(10^5)$ scale and with finite possible values. The proposed method before can be applied here without any difficulty.

One important feature of the current proposal is that it is symmetric and irreducible, which makes the decision step more easily to be implemented. This idea is drawn from the Gibbs sampling for the Ising model [8] and is improved by taking two components as a whole to make the change each time. Though other asymmetric proposal, such as the general Metropolis–Hastings algorithm, may be considered to deal with the asymmetry, the current one is simpler to reduce the computational effort and is general to deal with any parameters with probabilistic constraints.

With the Bayesian sampling technique, the computed posterior mean of $\hat{\mathbf{p}}$ and the fitted distribution versus the experimental data when $L = 150$ are shown in Fig. 8. The regularization parameter $\gamma = 8 \times 10^7$. We do the outer sampling loop for



(a) The comparison between the exact mixture weight and the computed posterior mean with MAP estimation and $L = 226$



(b) The comparison between the real data and the fitted distribution when $L = 226$ by MAP estimation.

Fig. 11. The results obtained by MAP estimation and $L = 226$. The wild oscillations are removed by the introduction of the smooth prior. The posterior mean for the mixture weight is very similar to the MAP estimate. The perfect match between the experimental and the fitted distribution is also ensured.

2000 times in Algorithm 3 and in Step 2 we do the inner Gibbs sampling for 20 times. The whole processes are repeated for 500 times and we take the numerical mean from the final results.

The computational effort for the Bayesian sampling of the posterior distribution is much larger than the MAP estimate procedure. But compared with the MAP estimate, the Bayesian sampling is more powerful and gives more information since it cannot only produce the posterior mean (one estimate for the parameters), but also the whole distribution. When $L = 150$, the posterior distributions for some selected components of \tilde{p}_l are shown in Fig. 9.

4.4. Detection noise concerned

Commonly, it is difficult to consider the detection noise in the photon–photoelectron-counting inversion problem due to its ill-posedness. Using our method, it can be incorporated into the current framework very easily.

The detection noise R , the real photoelectron number Y and the experimental observation of photoelectron number \bar{Y} satisfy

$$\bar{Y} = Y + R, \quad (43)$$

where the statistics of R is assumed discrete-valued and the distribution vector is known as $\mathbf{r} = (r_0, r_1, \dots, r_N)$. Experimentally, Y and R are independent of each other since they origin from the totally independent sources. For example, dark counts are a kind of noise commonly existed in imperfect detector and independent of photon detection, and emerges as Poissonian events [9]. Then the distribution of \bar{Y} can be given as

$$\bar{q}_m = \sum_{j=0}^N \sum_{n=0}^{\infty} p_n B(m-j; n, \xi) r_j, \quad (44)$$

With the same procedure as in Section 4.1, we can reduce the infinite mixture model Eq. (44) to the finite mixture model

$$\bar{q}_k \approx \sum_{l=1}^L \tilde{p}_l B_{kl}(\xi), \quad (45)$$

where \tilde{p}_l is defined the same as before except that

$$B_{kl}(\xi) = \frac{1}{\Delta N_l} \sum_{n \in J_l} \sum_{m \in I_k} \sum_{j=0}^N B(m-j; n, \xi) r_j. \quad (46)$$

For this reduced model, all of the optimization and sampling procedures can be applied and we only list the results with MAP estimate in Fig. 10.

In the computations, we assume the detection noise is of $\mathcal{O}(10^5)$ scale which is far less than the received number of photoelectrons. The statistics of R , the computed mixture weight for the noise perturbed data with MAP estimate and the comparison between the real data and the fitted distribution are shown in Fig. 10(a)–(c), respectively. Since the magnitude of the perturbation noise is small, the final mixture weights look similar with the unperturbed case. Generally, the proposed method can be applied without any difficulty in our cases.

4.5. Applications to the fake data

To further verify the validity of the proposed strategy, we apply it to the fake data which are generated by a known model. In doing so, we choose the parameter $\xi = 0.7$, and then set up the mixture weight \mathbf{p} from the discretization of a two-component Gaussian mixture model, whose distribution is shown in Fig. 11(a) with solid curve. The support of the exact \mathbf{p} is from $n = 1.9 \times 10^7$ to $n = 2.352 \times 10^7$. With this known \mathbf{p} , we generate 10^6 observation data from the direct binomial transformation and assume the accuracy of the measuring instrument is of $\mathcal{O}(10^4)$ scale and thus the observation data are split into $K = 157$ intervals (see Fig. 11(b)). For these fake data, we apply the proposed MAP estimate to the case $L = 226$ in which each interval contains 2×10^4 indices. In the set up of the prior distribution, the regularization parameter γ should be taken as large as 5×10^7 to suppress the oscillations. The time step size in the steepest descent method is taken as 5×10^{-6} and it is repeated for 300 times. The obtained MAP estimate for the piecewise constant mixture weight coincides with the exact \mathbf{p} perfectly, which confirms the effectiveness of our method. With this obtained \mathbf{p} , the fitted distribution perfectly matches the fake experimental data.

5. Discussion

In this paper, we present a statistical approach to perform the photon–photoelectron-counting inversion problem in quantum key distribution experiment. The proposed approach naturally takes into account the special structure of the binomial decaying transformation and applies the statistical inversion framework to this infinite binomial mixture model. We further proposed the piecewise constant coarse graining procedure to reduce the infinite mixture model to a finite K -multinomial mixture model. This idea is general and can be applied to the other infinite mixture model reductions with the basic assumption that the mixture weights have suitable spatial regularity. To overcome the ill-posedness issue of inversion problem, we further incorporate the smooth priors within the Bayesian framework. Both the MAP estimate and the Gibbs sam-

pling method are proposed to obtain the parameter estimation of the mixture weights. This recipe is proved to be successful to the real and fake data.

There are still some issues which need to be emphasized. For all of the numerical inverse problems, one should apply the regularization technique [6], in which the choice of the regularization parameter γ is a key point. In our method, this regularization is brought from the priori distribution. But there is no unique practical strategy to choose γ since the different magnitude of regularization parameters will give different solution and it is difficult to know which is the best. This is a common issue for all inverse problems. For our model, from the viewpoint of the continuum limit of the regularization term we have

$$-\mu \int_{l_{\min}}^{l_{\max}} \dot{p}^2 dl$$

and the discretization of the above integral with different mesh size gives

$$-\frac{\mu}{\Delta_L} \sum_{l=1}^{L+1} (\tilde{p}_l - \tilde{p}_{l-1})^2.$$

We obtain the consistency condition for different γ_L which corresponds to the same continuum limit

$$\gamma_L = \frac{\mu}{\Delta_L} \propto L. \tag{47}$$

The relation Eq. (47) gives a rationale for choosing γ_L in the code. While in our realistic simulations, we just increase γ_L only large enough to suppress the oscillations in Fig. 5. This strategy avoids introducing additional biasing smoothing. In fact the obtained mixture weight is fairly robust to the choice of γ_L in our numerical simulations.

It is also interesting to note one limit regime when $\gamma \rightarrow \infty$. In this non-interesting regime, the obtained mixture weights will resemble the parabola shape

$$p_l = -\frac{1}{\mathcal{Z}}(l - l_{\min})(l - l_{\max}), \quad l = l_{\min}, \dots, l_{\max}, \tag{48}$$

where l_{\min}, l_{\max} are the left-most and right-most indices of \mathbf{p} taking the value 0, respectively. \mathcal{Z} is a normalization constant.

For the choice of the priors, it is worth noting that the smooth prior is not the only choice. One may also take the other smoothers such as the maximum entropy prior, etc. For the other problems with the features [7], one can choose the total variation prior. But this should depend on the specific problems. In our problem, we believe the smoothness is more reasonable and physically relevant.

Another interesting topic related to this problem is to investigate its continuous approximation

$$q(y) = \int_{\mathbb{R}} p(x) \frac{1}{\sqrt{2\pi x \zeta (1 - \zeta)}} \exp\left(-\frac{(y - x\zeta)^2}{2x\zeta(1 - \zeta)}\right) dx \tag{49}$$

when the observed number of photoelectrons is very large. This continuous approximation preserves the correct relation between the first-order moments. To further study the property of this continuous transformation, designing the numerical methods based on this form and exploring the accuracy will be an interesting future topic.

It is challenging to rigorously prove the error estimate of the proposed algorithm, which will be a future research topic.

Acknowledgments

We thank Rensheng Zhu for some useful discussion. T. Li and J. Wu are supported by the National Science Foundation of China under Grant No. 10871010 and the National Basic Research Program under Grant No. 2005CB321704. X. Peng and H. Guo acknowledge financial support from the Key Project of the National Natural Science Foundation of China (Grant No. 60837004), and the National Hi-Tech Program of China (863 Program).

References

- [1] V. Scarani, H. Bechmann-Pasquinucci, N.J. Cerf, M. Dušek, N. Lütkenhaus, M. Peev, The security of practical quantum key distribution, *Rev. Mod. Phys.* 81 (2009) 1301–1350.
- [2] X. Peng, H. Jiang, B.J. Xu, X. Ma, H. Guo, Experimental quantum key distribution with an untrusted source, *Opt. Lett.* 33 (2008) 2077–2079.
- [3] T. Kiss, U. Herzog, U. Leonhardt, Compensation of losses in photodetection and in quantum-state measurements, *Phys. Rev. A* 52 (1995) 2433–2435.
- [4] Y. Hu, X. Peng, T. Li, H. Guo, On the Poisson approximation to photon distribution for faint lasers, *Phys. Lett. A* 367 (2007) 173–176.
- [5] A.N. Shiryaev, *Probability*, Springer-Verlag, Berlin and Heidelberg, 1996.
- [6] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, New York, 1996.
- [7] K. Kaipio, *E. Somersalo, Statistical and Computational Inverse Problem*, Springer-Verlag, New York, 2004.
- [8] J.S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag, New York, 2001.
- [9] H. Lee, U. Yurtsever, P. Kok, G. M Hockney, C. Adami, S. L Braunstein, J.P. Dowling, Towards photostatic from photon-number discriminating detectors, *J. Mod. Opt.* 51 (2004) 1517–1528.