JOSA COMMUNICATIONS

Stable estimation of the probability density function of intensity from photon frequency counts

Charles L. Byrne*

Code 5120, U.S. Naval Research Laboratories, Washington, D.C. 20375

Bruce Martin Levine and J. C. Dainty†

Institute of Optics, University of Rochester, Rochester, New York 14627

Received March 23, 1984; accepted July 6, 1984

A numerical solution to the problem of estimating the probability density of integrated intensity, $P(W)$, given a measured histogram of photon counts is described. The solution has a minimum norm in a Hilbert space that is weighted according to a prior estimate of $P(W)$ and is stable by virtue of a simple but powerful method of regularization. Additional stabilization is achieved by restricting the number of input photon-count frequencies.

1. INTRODUCTION

A common problem in the measurement of photon-count statistics is to infer the probability density function (PDF) of intensity from those of the photon counts. Let $P(W)$, where $W \geq 0$, be the PDF of integrated intensities, and let $p(n)$, where $n = 0, 1, 2, \ldots$, be the PDF of photon counts resulting from light distributed by $P(W)$. In the semiclassical approximation, the two PDF's are related through the Poisson transform[1] by the relation

$$p(n) = \int_0^\infty (W^n/n!)e^{-W}P(W)\mathrm{d}W. \qquad (1.1)$$

A histogram of measured photon-count probabilities is modeled as

$$y(n) = p(n) + \eta, \qquad (1.2)$$

where the term η is a zero mean and arbitrary variance–covariance structured random error. Saleh[2] showed that an analytic solution for $P(W)$ given $p(n)$ is possible in principle. Solutions based on analytic methods have been proposed[3,4] and have been demonstrated to work when $p(n)$ is known exactly. When $p(n)$ is estimated from a measured histogram and substituted into these formulas, even small sampling errors cause wild fluctuations in the resulting estimates for $P(W)$. It will be shown that a more stable numerical estimate of $P(W)$ can be formed given the vector of $y(n)$, a prior estimate of the basis PDF $P_0(W)$, and its corresponding photon-count PDF $p_0(n)$.

2. HILBERT-SPACE FORMULATION

In order to consider $P(W)$ to be a member of a suitably defined Hilbert space, rewrite Eq. (1.1) as

$$p(n) = \int_0^\infty P(W)[P_0(W)W^n/n!]e^{-W}P_0^{-1}(W)\mathrm{d}W. \qquad (2.1)$$

Here, $P(W)$ is a member of the Hilbert space $G = L^2(0, \infty, Q_0^{-1})$, where $Q_0(W) = e^W P_0(W)$. The inner product of the two functions $f(W)$ and $g(W)$ in G is defined as

$$\langle f, g\rangle_G = \int_0^\infty f(W)g(W)Q_0^{-1}(W)\mathrm{d}W, \qquad (2.2)$$

so that, with

$$g_n(W) = P_0(W)W^n/n!,$$

Eq. (2.1) is written as

$$p(n) = \langle P, g_n\rangle_G, \qquad n = 0, 1, \ldots.$$

Let the transformation $\mathbf{A}$: $G \rightarrow H = R^{N+1}$ (the Euclidean $N+1$ space) be that which associates with any $P(W)$ in G the vector $\mathbf{p} = [p(0), p(1), \ldots, p(N)]^T$ (T denotes the transpose of the vector). The adjoint mapping is $\mathbf{A}^T$: $H \rightarrow G$ and assigns to each $\mathbf{h} = [h_0, h_1, \ldots, h_N]^T$ in R^{N+1} the function

$$\mathbf{A}^T\mathbf{h} = P_0(W)\sum_{n=0}^{N} h_n W^n/n!. \qquad (2.3)$$

The data vector $\mathbf{y}$ is in H, and there will be exact solutions to the equation

$$\mathbf{y} = \mathbf{A}\mathbf{g}. \qquad (2.4)$$

In particular, the minimum norm solution g_{mn} given by Eq. (A3) is

$$g_{mn} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}. \qquad (2.5)$$

The transformation $\mathbf{A}\mathbf{A}^T$ associates with each $\mathbf{h}$ in H the vector of data values $\mathbf{y}$, and, in particular, the mth value is

$$y(m) = \int_0^\infty \left[P_0(W)\sum_{n=0}^{N} h_n W^n/n!\right]e^{-W}(W^m/m!)\mathrm{d}W$$

$$= \sum_{n=0}^{N} h_n \left[\int_0^\infty P_0(W)(W^{m+n}/m!n!)e^{-W}\mathrm{d}W\right]$$

$$= \sum_{n=0}^{N} h_n \{[(m+n)!/m!n!]p_0(m+n)\}.$$

In matrix notation, this becomes

0740-3232/84/111132-04$02.00

$$\mathbf{y} = \mathbf{P}_0\mathbf{h}, \tag{2.6}$$

where the mapping $\mathbf{A}\mathbf{A}^T$ is represented by the matrix $\mathbf{P}_0$ in Eq. (2.6). Numerical estimates of $P(W)$ are obtained by taking

$$\mathbf{h} = \mathbf{P}_0^{-1}\mathbf{y} \tag{2.7}$$

and

$$g_{mn}(W) = P_0(W) \sum_{n=0}^{N} h_n W^n/n!. \tag{2.8}$$

The matrix $\mathbf{P}_0$ is frequently ill conditioned in practice, so that small changes in the data y produce large changes in the numerical values of g_{mn}. Stabilization of the estimates requires use of regularization methods.

3. REGULARIZED SOLUTION

As is discussed in Appendix A, the regularized solution of Eq. (2.6) compromises between an exact solution and one that controls $\|\mathbf{g}\|$. This solution, given by Eq. (A4), is

$$\mathbf{g}_r = (\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y}. \tag{3.1}$$

One cannot obtain numerical values from Eq. (3.1) since $\mathbf{A}^T\mathbf{A}$ is an infinite dimensional operator. To obtain a solution in terms of a possible matrix inversion, clear the inverse from Eq. (3.1) and rearrange the terms to get

$$\mathbf{g}_r = (1/\epsilon)\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{g}_r)$$
$$= \mathbf{A}^T\mathbf{h}_r. \tag{3.2}$$

To find $\mathbf{h}_r$, again multiply Eq. (3.1) by $(\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I})$ and then by $\mathbf{A}$. Because $\mathbf{P}_0 = \mathbf{A}\mathbf{A}^T$ is invertible, it follows that

$$\mathbf{h}_r = (\mathbf{P}_0 + \epsilon\mathbf{I})^{-1}\mathbf{y}. \tag{3.3}$$

One obtains numerical estimates by substituting Eq. (3.3) into Eq. (3.2) and, as in Eq. (2.8),

$$g_r(W) = P_0(W) \sum_{n=0}^{N} h_{rn} W^n/n!. \tag{3.4}$$

One sees that the regularization involves only the adding of $\epsilon > 0$ to the main diagonal of the matrix $\mathbf{P}_0$. Even a small value of ϵ ($\approx 10^{-4}$) is sufficient in some cases to stabilize $\|\mathbf{g}\|$. So long as ϵ is greater than the computing machine's zero and not so large as to dominate the other terms in the matrix ($\mathbf{P}_0 + \epsilon\mathbf{I}$), a stable solution should be obtained. Generally, the regularized solution $\mathbf{g}_r$ is not consistent with the data [i.e., it is not a true solution of the matrix equation (2.4)]. By using Eqs. (3.2) and (3.3), one sees that, for

$$\mathbf{A}\mathbf{g}_r = (\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T + \epsilon\mathbf{I})^{-1}\mathbf{y},$$

only as $\epsilon \rightarrow 0$ does $\mathbf{g}_r$ approach data consistency.

The regularization presented here is sometimes referred to as Tychonov regularization[5] or as Miller regularization.[6] In fact, the basic idea of involving a measure of the norm of the solution in the function being minimized occurs in a variety of areas. This approach can also be viewed as an extension of the Hoerl–Kennard ridge regression estimator[7] to the case of infinite dimensional regression. The use of the prior basis PDF $P_0(W)$ in the construction of G also has application in problems of spectral estimation and signal processing.[8,9]

4. NUMERICAL EXAMPLE

We used a computer simulation to determine if a negative exponential PDF can be estimated from its Poisson transform pair, the Bose–Einstein PDF; i.e., for

$$p(n) = \mu^n/(1 + \mu)^{n+1} \tag{4.1}$$

Eq. (1.1) can be solved to obtain

$$P(W) = 1/\mu \exp(-W/\mu). \tag{4.2}$$

A sample of 10,000 counting intervals was simulated in a computer by using Eq. (4.1) and the parameter $\mu = 4$. The number of photon counts n was accumulated in a frequency histogram shown in Table 1. The data fit was accomplished by using a Bose–Einstein prior PDF with $\mu = 1$. It follows that the following matrices and functions are:

$\mathbf{y} = [p(n) + \eta]$, i.e., the vector of simulated photon-count probabilities,

$$\mathbf{P}_0 = \left[\frac{(m+n)!}{m!n!} p_0(m+n)\right] = \left[\frac{(m+n)!}{m!n!}\left(\frac{1}{2}\right)^{m+n}\right],$$

$P_0(W) = e^{-W}$.

Equations (3.3) and (3.4) were solved to obtain regularized estimates of the intensity PDF $g_r(W)$. Figure 1 shows these

Table 1. Summary of Photon-Count Simulation[a]

Count	Frequency	Count	Frequency	Count	Frequency
0	1984	12	140	24	13
1	1602	13	98	25	6
2	1280	14	100	26	5
3	1012	15	64	27	6
4	832	16	52	28	2
5	647	17	44	29	0
6	501	18	43	30	5
7	452	19	37	31	0
8	338	20	24	32	0
9	236	21	18	33	3
10	223	22	16	34	2
11	199	23	13	–	–

[a] Overflow counts, 3; total sample size, 10,000.

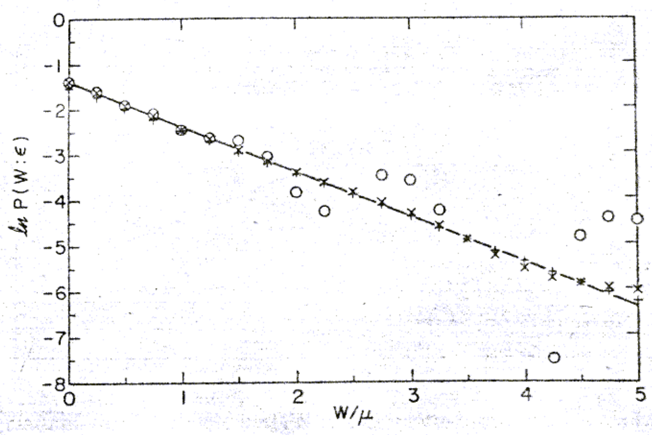

Fig. 1. Regularized estimates of $P(W)$ from simulated Bose–Einstein photon counts. Simulation parameters: sample size, 10,000; $\mu = 4$. Regularization parameters: $\mu_0 = 1$, ϵ; O, 0.0001; ×, 0.01; +, 0.1. The first 29 photon-count frequencies were used to compute $\mathbf{P}_0$.

estimates for differing values of ϵ on a logarithmic scale for the intensity PDF and on a normalized scale for W. Estimates are defined as admissable if they are positive and less than one. For $\epsilon = 0$, there are no admissable estimates; however, one can see that an admissible estimation is obtained for all values of W by using $\epsilon = 10^{-4}$. The exponentially increasing values of ϵ generally improve the estimation until too large a value biases the estimate of $P(W)$ at the origin and at its tails.

5. DISCUSSION

Other approaches to the estimation of $P(W)$, such as those of Simar[10] and of Tucker,[11] focus on the asymptotic properties of the probability density estimators. For finite data these estimators are discrete probabilities supported on a finite number of points (about $N/2$). These estimators are good (as Simar shows) for cases in which the compounding PDF is essentially a sum of point masses, i.e., a finite Poisson mixture. For the problem considered in this paper, such estimators are inappropriate. One advantage of Eq. (3.4) is that it is a continuous function of intensity.

Finite sampling errors in estimating the photon-count frequencies by $y(n)$ are also an important consideration in regularizing the $\mathbf{P}_0$ matrix. Matrix condition can be improved by restricting the input photon counts only to reliable estimates. Reducing the dimension of $\mathbf{P}_0$, and hence also the basis vector $\mathbf{h}$, is not severe for smooth functions, such as the exponential given in Section 4. This truncation, however, may reduce the range of W/μ over which accurate estimates of intensity may be obtained.

The importance of matrix conditioning of $\mathbf{P}_0$ is illustrated in Figs. 2 and 3. When the correct prior parameter is used, the resulting estimates are accurate only through a small range of intensities for all values of ϵ, as shown by Fig. 2. Further conditioning of $\mathbf{P}_0$ is still needed to obtain a satisfactory estimation. This is accomplished in Fig. 3, in which only relative photon-count frequencies greater than 0.1 are used. This suggests consideration of a measure for the condition of $\mathbf{P}_0$, namely, the ratio of the smallest to the largest eigenvalue. The conditioning measure would also aid in making the choice for a minimum acceptable value for ϵ.

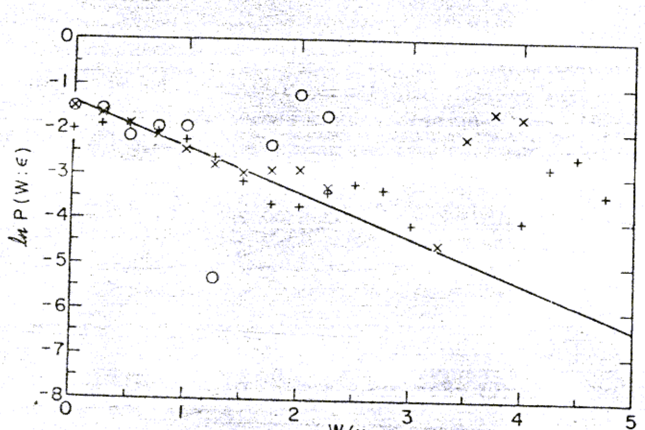

Fig. 2. Regularized estimates of $P(W)$ from simulated Bose–Einstein photon counts. Simulation parameters: sample size, 10,000; $\mu = 4$. Regularization parameters: $\mu_0 = 4$, ϵ; O, 0.0001; ×, 0.01; +, 0.1. The first 29 photon-count frequencies were used to compute $\mathbf{P}_0$.

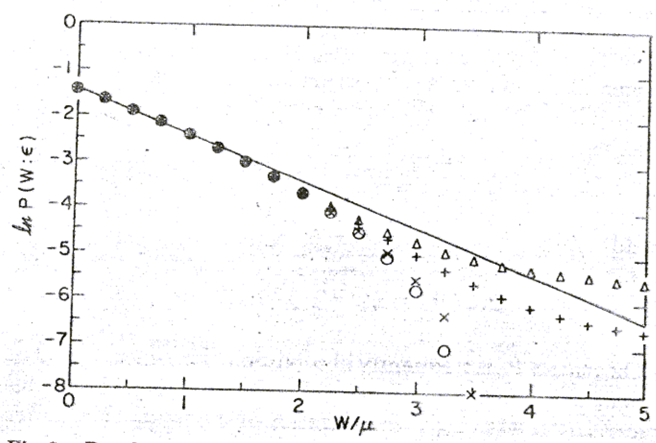

Fig. 3. Regularized estimates of $P(W)$ from simulated Bose–Einstein photon counts. Simulation parameters: sample size, 10,000; $\mu = 4$. Regularization parameters: $\mu_0 = 4$, ϵ; O, 0.0; ×, 0.0001; +, 0.0005; △, 0.001. The first four photon-count frequencies were used to compute $\mathbf{P}_0$.

No positivity constraints have been applied, nor has the integral of the estimated PDF been forced to equal one. This leaves open the possibility of using Eq. (3.4) iteratively with a succession of prior estimates $P_0(W)$, each obtained by clipping the previous regularized estimate $P_r(W)$ to make it positive and calling that a new prior estimate. Note that this procedure should iteratively discover a better basis for which the resulting regularized estimate is more positive.

APPENDIX A

Let G and H be arbitrary (possibly infinite dimensional) Hilbert spaces and the operator $\mathbf{A}$: $G \rightarrow H$ be any continuous linear transformation. For the vector $\mathbf{h}_0$, a fixed but arbitrary member of H, the equation

$$\mathbf{A}\mathbf{g} = \mathbf{h}_0 \tag{A1}$$

may have no solutions, one unique solution, or infinitely many solutions. In general, one considers the minimum-norm least-squares solution, defined as the unique $\mathbf{g}$ in G of minimum norm among all $\mathbf{g}$ for all $\|\mathbf{A}\mathbf{g} - \mathbf{h}_0\|$, as minimum. Denote this solution by $\mathbf{g}_{mnls}$. Two special cases are important.

Case 1. **A** *is one to one.* There is a unique **g** minimizing Eq. (A1) so that the minimum-norm least-squares solution is the least-squares solution; i.e.,

$$\mathbf{g}_{mnls} = \mathbf{g}_{ls} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{h}_0, \tag{A2}$$

where $\mathbf{A}^T$ is the adjoint transformation mapping H into G.

Case 2. **A** *is onto* H. The transformation $\mathbf{A}^T$ is one to one, and there is at least one solution to Eq. (A1). The minimum-norm least-squares solution is now the minimum-norm solution, or

$$\mathbf{g}_{mnls} = \mathbf{g}_{mn} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{h}_0. \tag{A3}$$

If small changes in $\mathbf{h}_0$ produce large changes in the minimum-norm solution $\mathbf{g}_{mn}$, then one can regularize by replacing the minimum-norm solution with a least-squares solution to a slightly different problem, which is formulated below.

Assume that **A** is onto H so that Eq. (A1) has exact solutions. Let K be the Hilbert space of ordered pairs $\mathbf{k} = (\mathbf{g}, \mathbf{h})$, that is, $K = G \oplus H$. The inner product is

$$\langle \mathbf{k}_1, \mathbf{k}_2 \rangle = \langle (\mathbf{g}_1, \mathbf{h}_1), (\mathbf{g}_2\, \mathbf{h}_2) \rangle$$
$$= \alpha \langle \mathbf{g}_1, \mathbf{g}_2 \rangle + (1 - \alpha) \langle \mathbf{h}_1, \mathbf{h}_2 \rangle,$$

so

$$\|\mathbf{k}\|^2 = \|(\mathbf{g}, \mathbf{h})\|^2 = \alpha \|\mathbf{g}\|^2 + (1 - \alpha) \|\mathbf{h}\|^2.$$

Define the transformation $\mathbf{B}$: $G \rightarrow K$ by the equation

$$\mathbf{Bg} = (\mathbf{g}, \mathbf{Ag}).$$

Then the adjoint transformation $\mathbf{B}^T$: $K \rightarrow G$ satisfies the relation

$$\mathbf{B}^T(\mathbf{g}, \mathbf{h}) = \alpha \mathbf{g} + (1 - \alpha) \mathbf{A}^T \mathbf{h}.$$

Given a fixed $\mathbf{h}_0$ in H, let $\mathbf{k}_0 = (\mathbf{0}, \mathbf{h}_0)$. Unless $\mathbf{h}_0 = \mathbf{0}$, $\mathbf{k}_0$ is not in the range of **B**. Then find a least-squares solution of the equation

$$\mathbf{Bg} = \mathbf{k}_0.$$

From the above, it is

$$\mathbf{g}_{ls} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{k}_0$$
$$= [(1 - \alpha) \mathbf{A}^T \mathbf{A} + \alpha \mathbf{I}]^{-1} (1 - \alpha) \mathbf{A}^T \mathbf{h}_0$$
$$= (\mathbf{A}^T \mathbf{A} + \epsilon \mathbf{I})^{-1} \mathbf{A}^T \mathbf{h}_0, \qquad \epsilon = \alpha/(1 - \alpha). \qquad \text{(A4)}$$

The $\mathbf{g}_{ls}$ minimizes $\|\mathbf{Bg} - \mathbf{k}_0\|$, so $\mathbf{g}_{ls}$ also minimizes the quantity

$$(1 - \alpha) \|\mathbf{Ag} - \mathbf{h}_0\|^2 + \alpha \|\mathbf{g}\|^2.$$

The least-squares solution of Eq. (A4) is called the regularized solution in that a strict solution to Eq. (A1) is relaxed in order to gain control over the quantity $\|\mathbf{g}\|$.

ACKNOWLEDGMENTS

Bruce Martin Levine and J. C. Dainty wish to thank the U.S. Air Force Office of Scientific Research for partial support of this research under grant AFOSR-81-0003.

* Permanent address, Department of Mathematics, The Catholic University of America, Washington, D.C. 20064.

† Permanent address, Blackett Laboratory, Imperial College, London SW7 2BZ, England.

REFERENCES

1. L. Mandel, "Fluctuations of photon beams: the distribution of the photo-electrons," Proc. Phys. Soc. London **74**, 233 (1959).
2. B. Saleh, *Photoelectron Statistics* (Springer-Verlag, Berlin, 1978), p. 83.
3. E. Wolf and C. L. Mehta, "Determination of statistical properties of light from photoelectric measurements," Phys. Rev. Lett. **13**, 705 (1964).
4. G. Bedard, "Analysis of light fluctuations from photon counting statistics," J. Opt. Soc. Am. **57**, 1201 (1967).
5. A. N. Tychonov, "On the solution of incorrectly put problems and the regularization method," Dokl. Akad. Nauk SSSR **151**, 501 (1963).
6. K. Miller, "Least squares methods for ill-posed problems with a prescribed bound," SIAM J. Math. Anal. **1**, 52 (1970).
7. A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation of nonorthogonal problems," Technometrics **12**, 55 (1970).
8. C. L. Byrne and R. M. Fitzgerald, "An approximation theoretic approach to maximum entropy spectral analysis," IEEE Trans. Acoust. Speech Signal Process. **ASSP-31**, 734 (1983).
9. C. L. Byrne and R. M. Fitzgerald, "Spectral estimators that extend the maximum entropy and maximum likelihood methods," SIAM J. Appl. Math. **44**, 425 (1984).
10. L. Simar, "Maximum likelihood estimation of a compound Poisson process," Ann. Stat. **4**, 1200 (1976).
11. H. G. Tucker, "An estimate of the compounding distribution of a compound Poisson distribution," Theory Probab. Its Appl. (USSR) **8**, 195 (1963).