On the problem of granulometry for a degraded Boolean image model

Alfred O. Hero III
Dept. of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48109-2122, USA
email: hero@eecs.umich.edu

Abstract

We consider a geometric coverage process consisting of a random number of disks, or grains, having random radii and positions in the plane. Our objective is granulometry: estimation of a parameter of the disk radius distribution, which is important in diverse applications such bio-assay, balistics, and numerical taxonomy. These disks are only incompletely observed due to mutual occlusion, spatial blurring and additive noise. We use a measurement channel paradigm to derive an expectation-maximization (EM) type estimation algorithm and a distortion-rate lower bound on estimation error.

I. Introduction

Here we treat a problem of parametric estimation from an image consisting of a Boolean process with spatial blurring and additive Gaussian noise. This type of geometric model is pertinent to many applications. The first such application was described in 1955 by Picinbono [10] for modeling the transparency of a photographic film composed of silver grains of random diameter for which the number and spatial positions of these grains are given by a homogeneous Poisson process. The model introduced here extends that of [10] by incorporation of a spatial point spread function and an additive noise into the measurements.

II. The Boolean Model

Let $\Theta = [\Theta_1, \ldots, \Theta_p]^T$ be a vector of random variables taking values $\theta = [\theta_1, \ldots, \theta_p]^T$ in $\mathbb{R}^p$ and having a joint density $f_\Theta(\theta)$. Our goal is to develop a MAP estimator of Θ and to specify lower bounds on the mean square estimation error (MSE). Estimation of Θ is based on an observed image $Y = \{Y(u) : u \in I\}$ composed of an signal image S and a noise image W. Here $I = [-a, a] \times [-a, a]$ denotes the support of the image and $|I| = 4a^2$ denotes its area. The signal S is generated by a marked point process $dM = \{dM(u) : u \in I\}$ whose distribution depends on Θ. The process dM creates N disks centered at positions $\{t_i\}^N$ in I and with radii $\{R_i\}_{i=1}^N$, $R_i \in (0, \infty)$. Conditioned on $\Theta = \theta$ and N, $\{U_i\}_{i=1}^N$ and $\{R_i\}_{i=1}^N$ are assumed mutually independent and i.i.d. with marginal densities $f_{U|\Theta}(u|\theta) = 1/|I|$ and $f_{R|\Theta}(r|\theta)$, respectively. Here N is a Poisson r.v. with conditional rate $E[N|\Theta] = E[N] = \Lambda > 0$ and intensity $\lambda = \Lambda/|I|$ which are indpendent of Θ. Under these assumptions the joint distribution of dM is closed form and estimation of Θ from dM is easily studied [8], [12]. This is no longer true when additive noise and blurring are introduced into the observations giving rise to a model:

$$Y(u) = S(u) + W(u), \quad u \in I \tag{1}$$

where W is a spatially white zero mean Gaussian noise with spectral power level $N_o/2$, and $S(u)$ is the blurred Boolean superposition

$$S(u) = h(u) \star g(u; dM). \tag{2}$$

where $h(u)$ is a spherically symmetric point spread function and $g(u, dM) = 1 - T(u)$ is the opacity function introduced by Picinbono

$$g(u; dM) = \prod_{i=1}^{N} D\left(\frac{u - U_i}{R_i}\right), \tag{3}$$

where $D(u)$ is the indicator function of a disc of radius 1 centered at the origin. Note that $g(u; dM)$ is a binary function which is non-zero only if there exists at least one disc covering the point u. In Figure 5 three realisations of the images S and Y are shown for a linear radial density $f_{R|\Theta}(r|\theta)$ whose slope is controlled by $\Theta \in [-1, 1]$

The Boolean model (3) is also called a "coverage process" model [5] and has been used for many applications in the life sicences, stereology, balistics [3]. The problem studied in this paper is known as granulometry [11] and consists of estimating an attribute of the density $f_{R|\Theta}$, e.g. the mean surface area $\pi E[R_i^2|\Theta]$ of the disks. The additive noise and blurring model (1), account for the effects of physical transcription of the image, e.g. electronic, mechanical, or chemical recording processes. Unfortunately, there is no analytical representation of the joint distribution of Y, Θ and thus optimal estimation Θ is much more difficult than for the case of direct observation of dM.

III. Composite Channel Representation

The measurements Y are related to the parameters Θ through the conditional density $f_{Y|\Theta} = \{f_{Y|\Theta}(y|\theta)\}_{y,\theta}$ or equivalently through the log-likelihood function $l(\theta) = \ln f_{Y|\Theta}(y|\theta)$. Since Θ is a random vector of parameters we can associate $f_{Y|\Theta}$ with transition probabilities of a measurement channel C. Let X be an arbitrary random variable. Then from the Bayes identity: $f_{Y|\Theta}(y|\theta) = \int_X f_{Y|X,\Theta}(y|x,\theta) f_{X|\Theta}(x|\theta)dx$. When X satisfies $f_{Y|X,\Theta}(y|x,\theta) =$ independent of θ, the Bayes identity affirms that C is decomposable into a cascade of two channels C_1 and C_2 whose transition probabilities are, respectively, $f_{X|\Theta}$ et $f_{Y|X}$. In the language of the EM algorithm, discussed below, X is a complete data set that carries more information about Θ than does Y [9]. Now, in the context of the model (1) a natural choice for X is the marked point process dM which gives the decomposition illustrated in Figure 1.

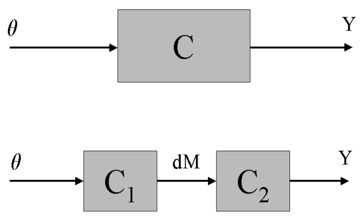

Fig. 1. *(a) Statistical representation of Y as the output of C with input Θ. (b) Decomposition of C into C_1 and C_2.*

IV. An EM-type MAP Estimator

The EM algorithm takes the form

Initialization: θ^0, $k = 0$

For $k = 1, \ldots$

- [E Step] `Estimate`

$$Q(\theta, \theta^k) = E[\ln f_{X|\Theta}(\underline{X}|\theta)|Y, \Theta = \theta^k] \qquad (4)$$

- [M Step] `Maximize`

$$\theta^{k+1} = \text{argmax}_\theta \ \left\{Q(\theta|\theta^k) + \ln f_\Theta(\theta)\right\} \qquad (5)$$

The exact EM algorithm is impossible to implement since the expectation $Q(\theta|\overline{\theta}) = E[\ln f_{X|\Theta}(\underline{X}|\theta)|Y, \Theta = \overline{\theta}]$ is not computable in closed form. We propose a linear approximation which was first introduced in [1] and consists of making a first order approximation to $Q(\theta|\overline{\theta})$, i.e. we replace the non-linear conditional mean estimator $E[\ln f_{X|\Theta}(\underline{X}|\theta)|Y, \Theta = \overline{\theta}]$ by the linear least mean square error estimator of $\ln f_{X|\Theta}(\underline{X}|\theta)$ given $Y, \Theta = \overline{\theta}$.

Under the assumption of large I this approximation takes the form [6]:

$$\hat{Q}^{(1)}(\theta|\overline{\theta}) = \Lambda \int_0^\infty (1 + r^2 q(Y, \overline{\theta})) f_{R|\Theta}(r|\overline{\theta}) \ln f_{R|\Theta}(r|\theta) dr$$

$$q(Y, \overline{\theta}) =$$

$$H(0)\pi\lambda e^{-\pi\lambda m_2(\overline{\theta})} \ \frac{1}{|I|} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\text{FT}_\text{e}(\omega)}{|H(\omega)|^2 \Phi_T(\omega|\overline{\theta}) + N_o/2} d\omega$$

where $\text{FT}_\text{e}(\omega)$ is the 2D Fourier transform of the residual error $e(u) = Y(u) - E[Y(u)|\overline{\theta}]$ over $u \in I$, and $\Phi_T(\omega|\overline{\theta})$ is the 2D Fourier transform of $\text{cov}(T(u), T(0)|\overline{\theta})$ over $u \in I$.

When the radial density is exponential, $f_{R|\Theta}(r|\theta) = \theta e^{-\theta r}$, $r > 0$, $\theta \in (0, \infty)$ we find an analytical form for the M-step of the approximate EM algorithm obtained by replacing Q in (5) by $\hat{Q}^{(1)}$:

$$\theta^{k+1} = \max\left\{0, \ \theta^k \ \frac{(\theta^k)^2 + 2! \ q(Y, \theta^k)}{(\theta^k)^2 + 3! \ q(Y, \theta^k)}\right\}.$$

The above equation is obtained under the assumption of a (improper) diffuse prior density of Θ. When the radial density is linear, as in Fig. 5, the M-step is not explcit and must be found numerically. In Figure 2 the likelihood function trajectory $\ln f(Y|\theta^k)$ is illustrated for several realizations of Y, Θ, a linear radial density, and Θ uniform over $[-1,1]$. Note that convergence is quite rapid in each case. The bias (0.03) and variance $(0.1)^2$ of the EM algorithm correspond to approximately 10% improvement over a standard non-linear least squares fit of the model covariance function to the sample covariance function.

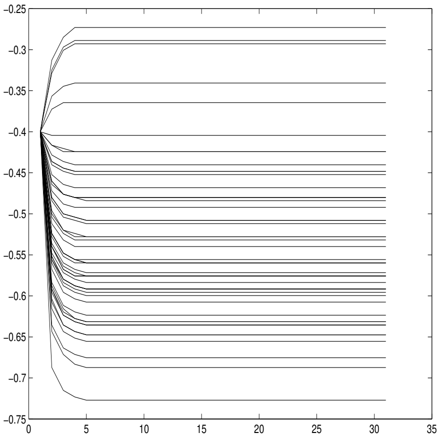

Fig. 2. *Comparison of the likelihood trajectories for several realizations of Y, Θ for the same blur function and noise power as in Fig. 5 and $\theta = 0.5$.*

V. A Shannon Lower Bound on MSE

For an estimator $\hat{\Theta} = [\hat{\Theta}_1, \ldots, \hat{\Theta}_p]^T$ define the total mean square error $\mathrm{MSE} = \sum_{j=1}^{p} \mathrm{E}[(\Theta_i - \hat{\Theta}_i)^2]$. Let V and Z be two random variables with mutual information $I(V;Z) = E[\ln P_{Z|V}(Z|V)/P_Z(Z)]$. Let $\rho(V,Z)$ be the squared distance (distortion) between the source V and an estimate $\hat{V}(Z)$ based on Z. Shannon theory asserts that for any upper bound d on $\overline{\rho} = E[\rho(V,Z)]$ the capacity $C = \sup_{P_V} I(V,Z)$ of the channel with input V and output Z must be at least as large as $R_\rho(d) = \inf_{P_{Z|V}:\overline{\rho}\leq d} I(V,Z)$, which is called the rate-distortion function. $R_\rho(d)$ is strictly decreasing over $d < d_{max}$ where d_{max} is the sum of the *a priori* variances of the components of V. Thus, defining the inverse $R_\rho^{-1}(\bullet)$ we have the lower bound

$$d = \mathrm{MSE} \geq \min\{\mathrm{d_{max}},\ \mathrm{R}_\rho^{-1}(\mathrm{C})\} \qquad (6)$$

To this lower bound the Shannon bound [2] can be applied $R_\rho(d) \leq H(V) - \frac{1}{2}\ln(2\pi d e)$, where $H(V) = E[-\ln P_V(V)]$. Furthermore, Shannon's data processing theorem asserts that if C is the capacity of a channel composed of a cascade of two channels with capacities C_1 and C_2 [4], then

$$C \leq \min\{C_1, C_2\}.$$

Upon application of these two Shannon bounds to (6) we obtain the following lower bound

$$\mathrm{MSE} \geq \frac{1}{2\pi \mathrm{e}}\ \mathrm{e}^{2\mathrm{H(V)}}\ \mathrm{e}^{-2\min\{\mathrm{C_1,C_2}\}}. \qquad (7)$$

Now identifying the data $Y = Z$ and the parameters $\theta = V$ in (7) and using the decomposition $C = C_1 \bullet C_2$ illustrated in Fig. 1 we can evaluate (7) once C_1 and C_2 are available.

A. Point Process Channel C_1

Using the fact that among all point processes dM with the same intensity the Poisson process has highest entropy we obtain a bound on C_1 similar to the expression obtained in [8, Lemma 4]

$$C_1 \leq C_1^* = \Lambda \sup_{f_\Theta} \int f_\Theta(\theta) \int dr\ f_{R|\Theta}(r|\theta) \ln \frac{f_{R|\Theta}(r|\theta)}{f_R(r)} d\theta$$

$$f_R(r) \stackrel{\mathrm{def}}{=} \int f_\Theta(\theta) f_{R|\Theta}(r|\theta) d\theta,$$

C_1^* is simply the capacity of a purely Poisson channel which is equal to the maximum mean Kullback distance between the conditional density $f_{R|\Theta}(r|\theta)$ and the marginal $f_R(r)$. Thus $C_1 = 0$ when $f_{R|\Theta}(r|\theta)$ is constant in θ and thus identical to $f_R(r)$. In this case neither dM nor Y carry any information about Θ. For the case of a linear radial density $f_{R|\Theta}(r|\theta)$, $\theta \in [-1,1]$, the source density f_θ which attains capacity C_1^* is easily determined and has an approximately quadratic form, as indicated in Figure 3. The resulting capacity is the linearly increasing function of Λ: $C_1^* = \Lambda a$ where $a \approx 0.0698$.

B. Continuous Process Channel C_2

Using the fact that among all continuous processes Y with fixed covariance function the Gaussian process has highest entropy we obtain the following bound [6]

$$C_2^* = \frac{|I|}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ln\left(1 + \frac{\Phi_S(\omega)}{N_o/2}\right) d\omega \qquad (8)$$

where $\Phi_S(\omega)$ is the power spectral density of the signal component S.

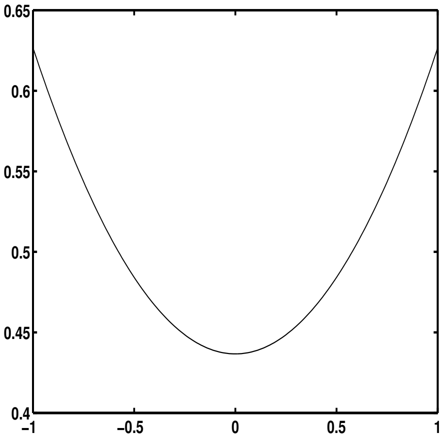

Fig. 3. *The density f_{Θ}^{*} that maximises the mutual information $I(\Theta, dM)$ and attains capacity for the case of linear $f_{R|\Theta}$ as shown in Figure 5.*

Define the fonction $p(u) = (1 - \|u\|/2)^{2+\lambda}$ and its Fourier transform $P(\|\omega\|)$. By making a rectangular to polar coordinate transformation in (8) we obtain the simplification

$$C_2^* = \pi|I| \int_0^\infty \rho \ln\left(1 + \kappa\, \frac{|H(\rho)|^2 P(\rho)}{N_o/2}\right) d\rho \qquad (9)$$

where, $M_\theta(t) = E[e^{t\theta}]$ is the characteristic function of f_Θ and

$$\kappa = e^{-\pi\lambda/3}\left[M_\theta(-\pi\lambda/3) - M_\theta^2(-\pi\lambda/6)e^{-\pi\lambda/3}\right].$$

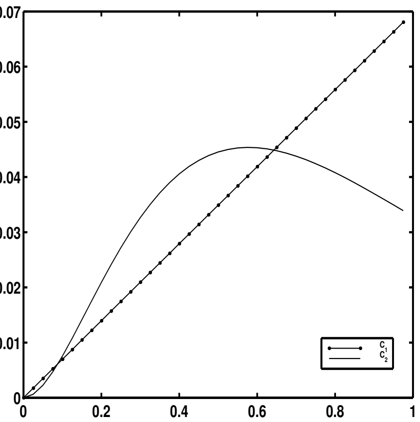

Fig. 4. *C_1^* and C_2^* as a function of the intensity λ for linear $f_{R|\Theta}$ shown in Figure 5 Here $f_\Theta(\theta)$ is uniform on $[-1,1]$.*

In Figure 4, C_1^* and C_2^* are plotted as a function of the intensity $\lambda = \Lambda/|I|$ for uniform density $f_\Theta(\theta)$ over $[-1,1]$ and the same values for I, N_o, and σ as used in Figure 5. Recall that it is the minimum of C_1^* and C_2^* that determine the Shannon bound (7). Notice that C_1^* increases in λ: estimation Θ from direct measurements dM always benefits from an increase in the number of points N. On the other hand, C_2^* takes a maximum value, decaying to zero as λ becomes large: estimates of Θ based on degraded measurements Y suffer from an increasing number of occlusions that must occur as the number of disks become large. This degradation for large λ is to be contrasted with the case of a linear superposition model studied in [7].

The Shannon bound as illustrated in Figure 4 separates estimator performance into two λ operating regions, one for $C_1^* < C_2^*$, the Poisson noise limited region, the other for $C_2^* < C_1^*$, Gaussian noise limited region. There are thus three regions $\lambda \in [0, 0.1]$, $\lambda \in (0.1, 0.65]$ and $\lambda > 0.65$: the only region where the Poisson limited region is attainable from measurements Y is for values $\lambda \in (0.1, 0.65]$. The boundaries of the λ regions depend on $|I|$, N_o, σ and f_Θ.

References

[1] N. Antoniadis and A. O. Hero, "Time delay estimation for filtered Poisson processes using an EM-type algorithm," *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 2112–2123, 1994.

[2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs NJ, 1971.

[3] H. Elias, *Stereology*, Springer, Berlin, 1967.

[4] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.

[5] P. Hall, *Introduction to the theory of coverage processes*, Wiley, New York, 1988.

[6] A. O. Hero, "Sur un problème d'estimation pour des processus de poisson composeés et filtrés," *Traitement du Signal*, vol. 15, no. 6, pp. 493–502, 1999.

[7] A. O. Hero, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Trans. on Inform. Theory*, vol. 37, no. 1, pp. 92–106, Jan. 1991.

[8] A. O. Hero, "Lower bounds on estimator performance for energy invariant parameters of multi-dimensional Poisson processes," *IEEE Trans. on Inform. Theory*, vol. 35, pp. 843–858, July 1989.

[9] A. O. Hero and J. A. Fessler, "A recursive algorithm for computing CR-type bounds on estimator covariance," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1205–1210, July 1994.

[10] B. Picinbono, "Modèle statistique suggéré par la distribution de grains d'argent dans les films photographiques," *Comptes Rendus de l'Académie des Sciences*, vol. Séance du 6 Juin, pp. 2206–2208, 1955.

[11] B. D. Ripley, *Spatial statistics*, Wiley, New York, 1981.

[12] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer-Verlag, New York, 1991.

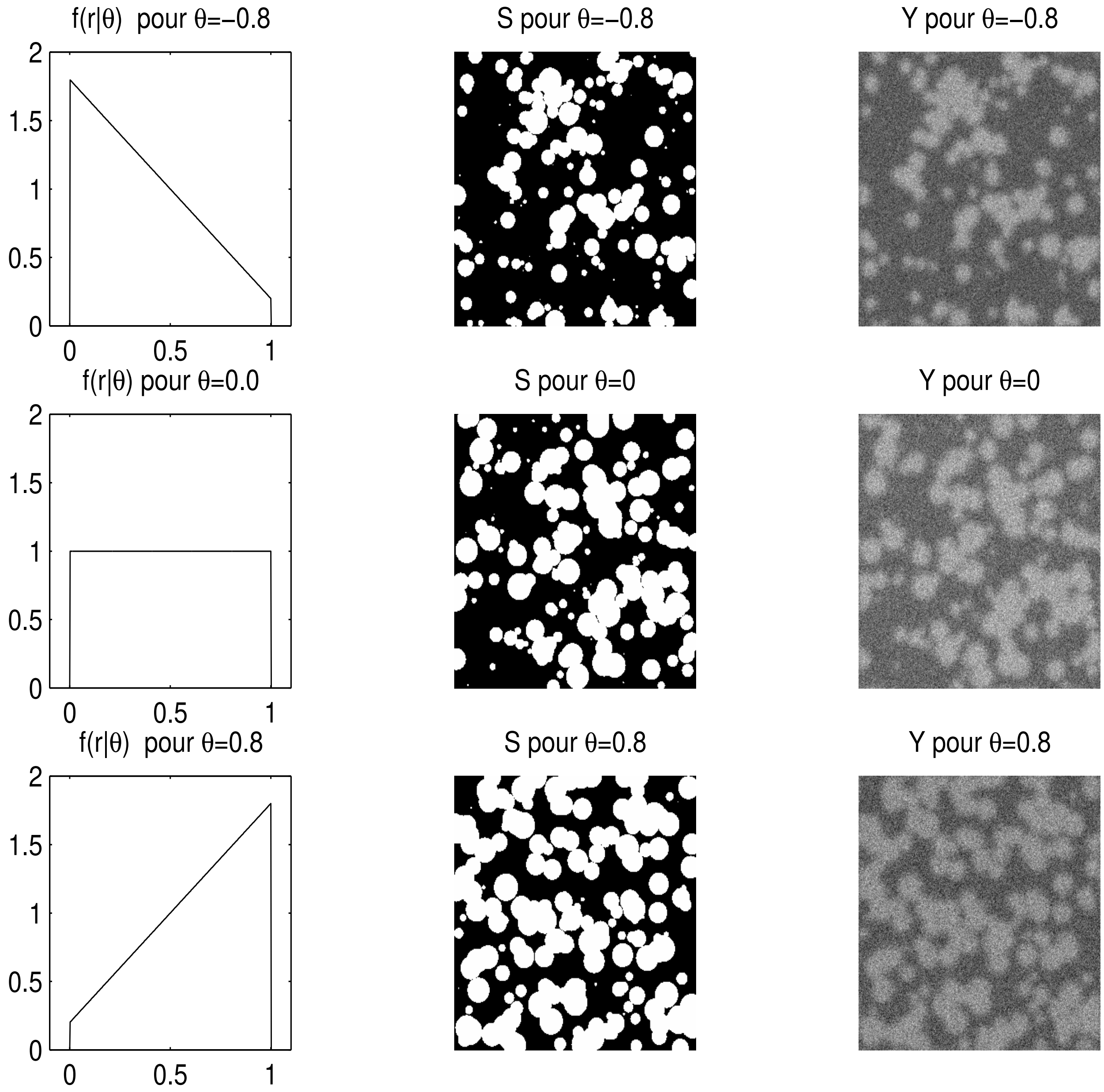

Fig. 5. *Three realisations of the images S et Y for linear radial densities (in mm) shown in the first column. Image size is 20×20 mm and intensity increases from black to white, The intensity $\lambda = 0.5$ corresponds to an average of $\Lambda = 200$ discs in each image. In the third column the SNR is 3dB and the PSF $h(u)$ is a symmetric Gaussienne of standard width $2 * \sigma = 0.66mm$.*