H^2/H^∞ Approach to the Histogram Method for Density Estimation

M. Nagahara[†1], K. I. Sato[‡2], and Y. Yamamoto[†3]

[†]Graduate School of Informatics, Kyoto University, [‡]Graduate School of Economics, Kyoto University
Kyoto, 606-8501, Japan
[1]nagahara@ieee.org, [2]KenjiSato@e-mail.jp, [3]yy@i.kyoto-u.ac.jp

Abstract: In this paper, we study nonparametric density estimation by the histogram method. Histogram is interpreted as quantization, which decreases the amount of information. Then interpolation (or estimation) of the missing information is needed. To achieve this, we introduce sampled-data H^2/H^∞ optimization. We design the reconstruction system which optimizes the worst case error between the original PDF and the estimation. The optimization is formulated by linear matrix inequalities and equalities. Numerical examples are illustrated to show effectiveness of our method.

Keywords: Density estimation, sampled-data control, histogram method.

1. INTRODUCTION

Probability density functions (PDF's, henceforth) or probability distributions play a fundamental role in both analysis and design of stochastic systems. Theoretically, it is often assumed that such information about randomness is given a priori. However, in practice, we usually have only limited information about that. On one hand, the random nature of such systems hinders us to obtain the exact distribution. On the other hand, because the analysis and design are usually executed on a digital processor, acquired data must be discretized in both time and value and this severely reduces the information. In this paper, we propose a method that weakens the latter difficulty. That is, we are concerned about estimating a PDF from quantized data. Our analysis assumes no specific function form for the PDF, and hence our method is supposed to be classified into nonparametric estimation. At the same time, it might be worth mentioning that we exploit a couple of ideas which were given rise in the signal processing and control literatures.

The first idea that we rely on owes Widrow [14]. These papers study the estimation problem from quantized samples by applying Shannon's sampling theorem [10], [11]. We refer to the Widrow's result as Widrow's quantization theorem. This theorem assumes that the original characteristic function is band-limited up to the frequency π/Δ, where Δ is the step size of the uniform quantizer. This frequency corresponds to the Nyquist frequency in Shannon's sampling theorem. The assumption which is called *band-limiting condition* is quite restrictive because many random variables such as uniform, exponential, Erlangian, or gamma ones do not satisfy the assumption [8].

The second helpful idea has come from the control theory. In order to avoid the stringent assumption of Shannon's theorem, sampled-data control theory proved useful in [6]. The paper shows that the sampled-data H^2/H^∞ optimization provides a way to obtain a filter without the band-limiting assumption. By altering the method of [6] into that suitable for Widrow's quantization theorem, we will be able to obtain a better estimation.

In this paper, this alternative way to estimating a PDF is proposed. Unlike Widrow's method, we are unable to obtain the exact PDF. Such a restriction is inevitable because we dismiss the band-limiting condition. Instead, we apply the sampled-data control theory [2] that enables us to minimize the undesirable error in the analog domain. A designed filter, which generates an estimated PDF, will be combined with an up-sampler to compose an interpolator [12]. Like the method of [6], our estimation can be accomplished by efficient numerical optimization.

The organization of this paper is as follows. Section 2 defines nonparametric density estimation problem which we discuss in this paper. In section 3, we introduce Widrow's quantization theorem. In section 4, we first point out problems in Widrow's estimation. Then to solve these problems, we propose a new estimation method by sampled-data H^∞ optimization. We discuss a multi-dimensional case in section 5. Numerical examples are given in section 6. Section 7 concludes our results.

Notations

We use the following notations. $\mathbb{R}$ and $\mathbb{Z}$ are the sets of the real numbers and the integers, respectively. L^2 is the Lebesgue spaces consisting of the square integrable real functions on $\mathbb{R}$. For $w \in L^2$, $\mathcal{F}w$ denotes the Fourier transform of w. We denote by $\chi_{_I}$ the indicator function on a set $I \subset \mathbb{R}$; $\chi_{_I}(x) = 1$ if $x \in I$ and $\chi_{_I}(x) = 0$ if $x \notin I$. For a matrix M, $M^\top$ and tr$M$ are respectively the transpose and the trace of M. For a symmetric matrix P, $P \succ 0$, $P \succeq 0$ and $P \prec 0$ denote positive, nonnegative, and respectively negative definite matrices. For a linear time-invariant (continuous-time or discrete-time) system with state space matrices $\{A, B, C, D\}$,

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right](s) \ \text{ or } \ \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right](z)$$

denotes the transfer function.

2. DENSITY ESTIMATION PROBLEM

In this section, we define our problem. Assume that there are samples drawn from a PDF $p(x)$, which is unknown. Our concern in the present paper is how to estimate $p(x)$ from the samples. In ordinary practice the acquired samples are stored and processed in digital systems, and hence they must be converted to digital data, that is quantized. We assume that this quantization is a uniform one with the step size Δ and the no-overload input range $((-M-1)\Delta, (M+1)\Delta)$ (see [9]), which is defined by

$$Q : \mathbb{R} \to \mathcal{A} := \{-M\Delta, (-M+1)\Delta, \dots, M\Delta\},$$

$$Qx := \begin{cases} k\Delta, & \left(k - \frac{1}{2}\right)\Delta \le x < \left(k + \frac{1}{2}\right)\Delta, \\ & \quad k = -M+1, \dots, M-1, \\ M\Delta, & x \ge \left(M - \frac{1}{2}\right)\Delta, \\ -M\Delta, & x < \left(-M - \frac{1}{2}\right)\Delta. \end{cases} \tag{1}$$

Then, our problem is formulated as follows:

Problem 1: Given samples $\widetilde{x}_1, \widetilde{x}_2, \dots, \widetilde{x}_N$ which are i.i.d. (independent, identically distributed) and quantized by the uniform quantizer defined by (1), estimate the original PDF $p(x)$ from which the samples are drawn.

This problem is ill-posed because there are infinitely many solutions. To avoid ill-posedness, we have to restrict the function space to which the original PDF belongs. A unique solution to the problem is obtained by Widrow et al. by restricting the function space to that of band-limited functions [14]. We introduce this solution in the next section.

3. WIDROW'S QUANTIZATION THEOREM

In this section, we introduce Widrow's quantization theorem.

First, we analyze the uniform quantization Q (see (1)) of a PDF $p(x)$. The samples drawn from $p(x)$ and quantized by Q can be interpreted as samples of a discrete PDF $c[k]$. This discrete PDF is obtained by *area-sampling* [14], denoted by $\widetilde{\mathcal{S}}_\Delta p$, that is,

$$c[k] = \int_{(k-1/2)\Delta}^{(k+1/2)\Delta} p(x)dx =: (\widetilde{\mathcal{S}}_\Delta p)[k]. \tag{2}$$

This can be rewritten by

$$\begin{aligned} c[k] &= \int_{-\infty}^{\infty} \chi_{[-\Delta/2, \Delta/2)}(k\Delta - x)p(x)dx \\ &= (q_\Delta * p)(k\Delta) =: \mathcal{S}_\Delta(q_\Delta * p)[k], \\ q_\Delta(x) &:= \chi_{[-\Delta/2,\Delta/2)}(x). \end{aligned}$$

Thus, quantization can be interpreted as convolution by q_Δ and ideal sampling $\mathcal{S}_\Delta$. Note that the Fourier transform (frequency response) of q_Δ is given by

$$Q_\Delta(j\omega) = \frac{1}{j\omega}\left(e^{j\omega\frac{2}{\Delta}} - e^{-j\omega\frac{2}{\Delta}}\right). \tag{3}$$

Based on this idea, Widrow et al. proposed a reconstruction scheme based on Shannon's sampling theorem. To see this, we define the space BL^2 of band-limited functions:

$$BL^2 := \left\{p \in L^2 : (\mathcal{F}p)(\omega) = 0, |\omega| > \pi/\Delta\right\}.$$

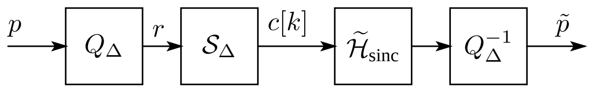

Fig. 1 Density estimation by Widrow's method

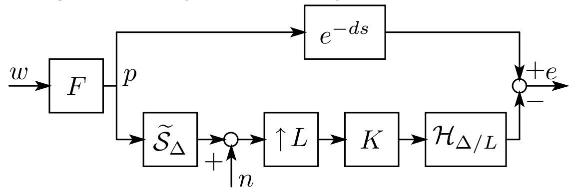

Fig. 2 Block diagram for the optimal filter K.

Then, Widrow et al. have shown a solution to Problem 1 as follows [14].

Theorem 1 (Widrow et al.) Assume that $p \in BL^2$. Then $p(x)$ can be perfectly reconstructed from the discrete PDF of the quantized samples $\{\widetilde{x}_n\}$.

The reconstruction procedure by this theorem is as follows. First, construct a histogram from the samples $\{\widetilde{x}_n\}$, with its intervals $(-\infty, -(M+1/2)\Delta)$, $[(k-1/2)\Delta, (k+1/2)\Delta)$, $k = -M+1, \dots, M-1$, and $[(M-1/2)\Delta, \infty)$. If we have sufficiently many samples, $\{c_k\}$ can be approximated from the frequency of each interval divided by the number of the samples. Let $r = q_\Delta * p$. Then, if the band-limiting assumption holds, r can be perfectly reconstructed from the discrete data $\{c[k]\}$ by

$$r(x) = \sum_{k=-\infty}^{\infty} c[k]\mathrm{sinc}(x/\Delta - k) =: (\widetilde{\mathcal{H}}_{\text{sinc}}c)(x),$$

where $\mathrm{sinc}(x) := \sin(\pi x)/(\pi x)$. Then, the original PDF $p(x)$ can be obtained by the convolution $p = r * q_\Delta^{-1}$, where q_Δ^{-1} is the inverse system of q_Δ. This procedure is shown in Fig. 1.

4. H^2/H^∞ OPTIMAL ESTIMATION

In practice, the band-limiting assumption $p \in BL^2$ can be too restrictive because many important probability density functions such as uniform, exponential, Erlangian, or gamma ones do *not* satisfy the assumption (see [8]). We therefore introduce another signal space FL^2 defined by

$$FL^2 := \{p \in L^2 : p = Fw, w \in L^2\},$$

where F is a stable and strictly causal system with a rational transfer function $F(s)$. Although this space does not include band-limited functions (since the sinc function does not have a rational transfer function), this space includes many functions such as the distributions mentioned above [8]. The function $F(s)$ is treated as a *smoothing parameter* [1] in our density estimation.

Assuming that the original PDF $p(x)$ is in FL^2 defined by we consider the reconstruction problem by the error system shown in Fig. 2. In this figure, an L^2 signal w is filtered by $F(s)$. This $F(s)$ is a presumed frequency domain model for the original PDF p. Then, this p is sampled by the generalized sampler $\widetilde{\mathcal{S}}_\Delta$ defined by

$$(\widetilde{\mathcal{S}}_\Delta p)[k] := \int_{(k-1/2)\Delta}^{(k+1/2)\Delta} p(x)dx$$

and a discrete PDF $\{c[k]\}$ is obtained. To reconstruct p from the discrete data $\{c[k]\}$, we adopt the interpolation technique in multirate signal processing. First, the discrete data $\{c[k]\}$ is upsampled by $\uparrow L$, which is defined by

$$((\uparrow L)c)[k] = \begin{cases} c[l], & k = Ll,\ l = 0, 1, \dots \\ 0, & \text{otherwise} \end{cases}$$

Then the upsampled signal $(\uparrow L)c$ is filtered by $K(z)$ which we design. The output of this filter is converted to a continuous function by the zero-order hold $\mathcal{H}_{\Delta/L}$ with period Δ/L, and we obtain an estimation $\tilde{p}$ of the original PDF p. In the error system Fig. 2, the estimation $\tilde{p}$ is compared with d-shifted version $p(x-d)$ of the original PDF $p(x)$. The objective is to attenuate the *continuous* reconstruction error e with respect to the H^∞ norm of the error system

$$J_1(K) := \|\{e^{-ds} - \mathcal{H}_{\Delta/L}K(\uparrow L)\widetilde{\mathcal{S}}_\Delta\}F(s)\|\|_\infty. \quad (4)$$

At the same time, to reduce the quantization noise n from the output, we limit the H^2 norm given by

$$J_2(K) := \|\mathcal{H}_{\Delta/L}K(\uparrow L)\|_2. \quad (5)$$

Our objective is to find K which minimizes $J_1(K)$ subject to $J_2(K) < \lambda$ for given $\lambda > 0$.

The sampled-data systems in (4) and (5) can be reduced to discrete-time ones via H^∞ discretization or fast-sample fast-hold (FSFH) method [2], [3], [6]. Assuming that the filter $K(z)$ is an FIR filter

$$K(z) = \sum_{l=0}^{N-1} a_l z^{-l},$$

we have the following discrete-time error systems $E_1(z)$ and $E_2(z)$ for the sampled-data systems in (4) and (5) [3],

$$E_1(z) = T_{11}(z) + W(\boldsymbol{a})T_{12}(z),$$
$$E_2(z) = W(\boldsymbol{a})T_{22}(z),$$

where $\boldsymbol{a} := [a_0, a_1, \dots, a_{N-1}]$ is the vector of the FIR coefficients, and $W(\boldsymbol{a}) := \begin{bmatrix} a_0 I & a_1 I & \dots & a_{N-1}I \end{bmatrix}$. The systems T_{11}, T_{12}, and T_{22} are all finite-dimensional discrete-time LTI systems. Thanks to the $\boldsymbol{a}$-affine structure of $E_1(z)$ and $E_2(z)$, their state-space representations are given by

$$E_i(z) = \left[\begin{array}{c|c} A_i & B_i \\ \hline C_i(\boldsymbol{a}) & D_i(\boldsymbol{a}) \end{array}\right](z), \quad i = 1, 2,$$

where A_i and B_i are constant matrices and $C_i(\boldsymbol{a})$ and $D_i(\boldsymbol{a})$ affinely depend on the design parameter $\boldsymbol{a}$. By using the Kalman-Yakubovic-Popov lemma [4], our optimization can be described as follows: given $\lambda > 0$, find the optimal FIR coefficient $\boldsymbol{a}$ which minimizes $\gamma > 0$ subject to

$$P_1 = P_1^\top \succ 0, \quad P_2 = P_2^\top \succ 0$$
$$\begin{bmatrix} A_1^\top P_1 A_1 - P_1 & A_1^\top P_1 B_1 & C_1(\boldsymbol{a})^\top \\ B_1^\top P_1 A_1 & -\gamma I + B_1^\top P_1 B_1 & D_1(\boldsymbol{a})^\top \\ C_1(\boldsymbol{a}) & D_1(\boldsymbol{a}) & -\gamma I \end{bmatrix} \prec 0,$$
$$\begin{bmatrix} A_2^\top P_2 A_2 - P_2 & C_2(\boldsymbol{a})^\top \\ C_2(\boldsymbol{a}) & -I \end{bmatrix} \prec 0,$$
$$\begin{bmatrix} B_2^\top P_2 B_2 - \lambda^2 & D_2(\boldsymbol{a})^\top \\ D_2(\boldsymbol{a}) & -I \end{bmatrix} \prec 0,$$
$$V(\boldsymbol{a}) \succeq 0, \quad \mathrm{tr}V(\boldsymbol{a}) = L,$$

where $V(\boldsymbol{a}) = \mathrm{diag}\{a_1, a_2, \dots, a_{N-1}\}$. The inequality $V(\boldsymbol{a}) \succeq 0$ is for the positivity of the estimated PDF $\tilde{p}(x)$, that is, $\tilde{p}(x) \geq 0$. The equality $\mathrm{tr}V(\boldsymbol{a}) = L$ is for the unity of the whole integral of $\tilde{p}(x)$, that is,

$$\int_{-\infty}^{\infty} \tilde{p}(x)dx = 1.$$

For these conditions on $V(\boldsymbol{a})$, see [7]. These are LMI's and an LME, and the optimal coefficient $\boldsymbol{a}$ can be obtained effectively by using an optimization softwares such as MATLAB. If the up-sampling ratio L is large (e.g., $L = 32$ or 64) and the number N of $K(z)$ is also large, the computation will be hard. In such a case, a cutting plane method can be applied. See [13].

5. MULTI-DIMENSIONAL PDF ESTIMATION

By using the one-dimensional filter $K(z)$ given in the previous section, we can construct a higher dimensional PDF estimation system. In the case of M-dimensional PDF's, we adopt *separable* filter [5], that is,

$$K(z_1, z_2, \dots, z_M) = K_1(z_1)K_2(z_2)\cdots K_M(z_M),$$

where each $K_i(z_i)$ is a one-dimensional filter. By this approach, the computational complexity increases exponentially with the dimension M. In real system, this can be applied when the dimension is at most 2 or 3.

6. NUMERICAL EXAMPLES

We here show an example of density estimation. We set the original PDF $p(x)$ to be the uniform distribution in the interval $[-5, 5]$. Note that since the support of $p(x)$ is $[-5, 5]$ (compact), $p(x)$ does not satisfy the band-limiting assumption of Widrow's theorem. The number of samples drawn from the PDF is 10^5. Then the samples are uniformly quantized with step size $\Delta = 1$. For our design, we set the design parameters as follows. The transfer function (smoothing parameter) $F(s)$ is set as

$$F(s) = \frac{\omega_1\omega_2}{(s+\omega_1)(s+\omega_2)},$$

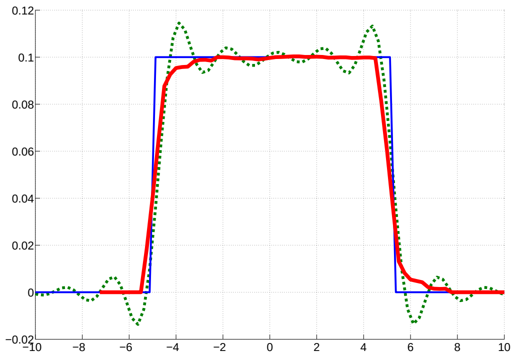

Fig. 3 Density estimation: original PDF (thin line), Widrow's estimation (dots), and proposed estimation (thick line)

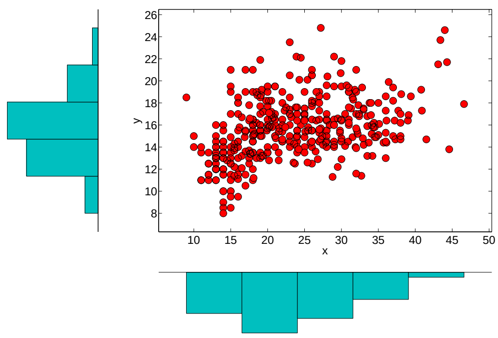

Fig. 4 2-dimensional samples and histograms

where $\omega_1 := \pi/10$ and $\omega_2 := \pi/20$. The up-sampling ratio L is 4. The number of taps is $N = 16$. Fig. 3 shows the estimated PDF's by Widrow's method and the proposed one. It can be seen that the PDF of Widrow's estimation has negative values. On the other hand, our PDF is always positive. Moreover, Widrow's estimation shows large ripples, while our method well reconstructs the original PDF. The ripples are caused by the high frequency components in $\Psi(\omega)$ which is not considered in Widrow's theorem. By this example, we can see the effectiveness of our method.

An example of 2-dimensional density estimation is also shown. Fig. 4 shows given 2-dimensional data. From the data, we estimate the original PDF. The obtained estimation is shown in Fig. 5.

7. CONCLUSION

In this paper, we have proposed a new nonparametric density estimation from quantized samples via sampled-data H^∞ optimization. Our estimation is formulated by signal reconstruction with an up-sampler and a digital filter, which can be effectively used in digital systems such as DSP's (Digital Signal Processors). The optimization of the filter is formulated by an H^2/H^∞ one, and the optimal filter can be computed by LMI's and an LME.

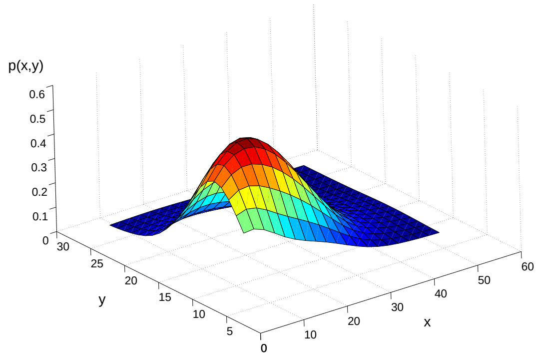

Fig. 5 2-dimensional density estimation

Higher dimensional case have been also considered. Design examples have been illustrated to show effectiveness of the proposed method.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[2] T. Chen and B. A. Francis, *Optimal Sampled-data Control Systems*, Springer, 1995.

[3] H. Ishii and Y. Yamamoto, Sampled-data H^∞ and H^2/H^∞ design of multirate D/A converter, Trans. of the ISCIE, Vol. 11, No. 10, pp. 585–592, 1998 (in Japanese).

[4] T. Iwasaki, *LMI and Control*, Shoko-do, 1997 (in Japanese).

[5] W.-S. Lu and A. Andoniou, *Two-Dimensional Digital Filters*, Marcel Dekker, 1992.

[6] M. Nagahara and Y. Yamamoto, A new design for sample-rate converters, *Proc. of 39th IEEE CDC*, pp. 4296–4301, 2000.

[7] M. Nagahara, K. I. Sato, and Y. Yamamoto, H^∞ optimal nonparametric density estimation from quantized samples, *Proc. of the 40th ISCIE SSS*, 2008.

[8] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, Fourth Ed., Mc Graw Hill, 2002.

[9] R. Schreier and G. C. Temes, *Understanding Delta-Sigma Data Converters*, Wiley Interscience, 2005.

[10] C. E. Shannon, Communication in the presence of noise, *Proc. IRE*, Vol. 37, No. 1, pp. 10–21, 1949.

[11] M. Unser, Sampling — 50 years after Shannon, *Proc. IEEE*, Vol. 88, No. 4, pp. 569–587, 2000.

[12] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, 1993.

[13] Y. Wakasa, D. Yasufuku, M. Nagahara, and Y. Yamamoto, Sampled-data design of interpolators using the cutting-plane method, *Proc. of 41st IEEE CDC*, pp. 2564–2569, 2002.

[14] B. Widrow, I. Kollar, and M.-C. Liu, Statistical theory of quantization, *IEEE Trans. on Instrumentation and Measurement*, Vol. 45, No. 2, pp. 353 – 361, 1996.