One-Dimensional Probability Density Function Estimation by Positive Quartic C^2-Spline Interpolation and Smoothing

Daichi KITAHARA† Isao YAMADA†

†Department of Communications and Computer Engineering, Tokyo Institute of Technology

Abstract Spline is a continuous function piecewise-defined by polynomials and is widely used for interpolation and smoothing of observed data. In 1994, Heß and Schmidt proposed a positive quartic C^2-spline interpolation for estimation of a non-negative and twice continuously differentiable function. In this paper, first we generalize the positive quartic C^2-spline interpolation to the positive quartic C^2-spline smoothing. Then we propose two estimation methods of a probability density function from its histogram by extending the ideas of the positive quartic C^2-spline interpolation and smoothing. Finally numerical experiments show the effectiveness of the proposed methods.

1 Introduction

The probability density function (PDF) estimation [1]–[3] is a classical but has been central problem of significant impact to many branches of mathematical sciences and engineerings including the Bayesian signal processing [4], pattern recognition [5], [6], quantization [7], filtering [8], data analysis [9], and combustion science [10], etc. Inherently, the PDF estimation requires construction of a positive continuous function from given finite data. The estimation of continuous functions has been a common task in many areas of signal processing where the spline function, a smooth piecewise polynomial, has often been used [11], e.g., in super-resolution of images or videos [12], [13], computer aided design [14], [15], and regression analysis of data [16], [17], due to its flexibility and optimality (see, e.g., Fact 1) in many senses.

However the spline interpolation has been hardly applicable to the PDF estimation because (i) the data point to be interpolated is not available, and (ii) the positivity of the interpolating spline function is not guaranteed in general. On guaranteeing the positivity of the spline function, Heß and Schmidt showed a way to construct a positive quartic C^2-spline function which interpolates given non-negative data samples [18]. These situations suggest that the remaining issues for sound applications of the spline function to the PDF estimation is to establish a more flexible construction of a positive spline function than just interpolating data points.

In this paper, we propose two novel PDF estimations by using positive quartic C^2-splines. For this purpose, in Section 2, as a preliminary, we first extend the standard spline smoothing to the positive quartic C^2-spline smoothing where the obtained spline function is guaranteed to satisfy an optimality in the sense of a balance between the data fidelity and the smoothness. The proposed PDF estimations are given in Section 3, the first method, for a use in the situation where a relatively rough histogram is available, is realized by modifying the positive quartic C^2-spline smoothing shown in Section 2. The second one, for a use in the situation where an almost ideal histogram of which each bin approximates sufficiently corresponding area of the PDF to be estimated is available, is realized by modifying the idea found in [18]. In Section 4, by numerical experiments, we compare the performances of the proposed methods with the kernel density estimation [1], [2] which has been used widely for the PDF estimation. The experiments for a Gaussian mixture show the effectiveness of the proposed methods.

2 Preliminaries

Let $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{Z}_+$ denote respectively the set of all real numbers, non-negative real numbers and non-negative integers. A boldface letter denotes a vector or a matrix depending on the situation. For any vector $\boldsymbol{x} \in \mathbb{R}^n$, $[\boldsymbol{x}]_i$ denotes the ith component of $\boldsymbol{x}$, and $\|\boldsymbol{x}\| := \sqrt{\sum_{i=1}^n [\boldsymbol{x}]_i^2}$.

2.1 Spline Function

Let $\Delta_n := \{x_i\}_{i=0}^n$ be a grid on an area $\Omega := [x_0, x_n] \subset \mathbb{R}$ s.t. $x_0 < x_1 < \cdots < x_n$, and let $d, \rho \in \mathbb{Z}_+$ s.t. $0 \leq \rho < d$. Define

$$\mathcal{S}_d^\rho(\Delta_n) := \{f \in C^\rho(\Omega) \,|\, \forall i \; f = f_i \in \mathbb{P}_d \text{ over } [x_i, x_{i+1}]\}$$

as the set of all spline functions of degree d and smoothness ρ on Δ_n, where $C^\rho(\Omega)$ stands for the set of all ρ-times continuously differentiable functions over the interior of Ω, and $\mathbb{P}_d$ denotes the set of all polynomials whose degree is d at most.

Assume that we observe samples of a twice continuously differentiable function $f^* : \Omega \to \mathbb{R}$ with additive noise $\epsilon_i \in \mathbb{R}$ on Δ_n, i.e., we observe $\zeta_i := f^*(x_i) + \epsilon_i$ at x_i ($i = 0, 1, \ldots, n$). In this situation, the spline function is often used as an approximation of f^* because it guarantees the optimality shown in the following fact [19].

Fact 1 (Spline function as a solution of a variational problem) *In the set of all twice continuously differentiable functions over the interior of Ω, i.e., $C^2(\Omega)$, there is a unique function $s^* \in C^2(\Omega)$ minimizing*

$$\sum_{i=0}^n |\zeta_i - s(x_i)|^2 + \lambda \int_{x_0}^{x_n} |s''(x)|^2 \, dx,$$

and it is a natural cubic spline $s^* \in \mathcal{S}_3^2(\Delta_n)$, where $\lambda > 0$ is a parameter controlling the trade-off between the data fidelity and the smoothness of the solution.

2.2 Positive Quartic C^2-Spline

The problem of our interest is to approximate a non-negative function $f^* : \Omega \to \mathbb{R}_+$ with use of its non-negative data $\zeta_i := f^*(x_i) + \epsilon_i \geq 0$. In [18], the positive C^2-spline interpolation is considered, and it is shown that the lowest degree is $d = 4$ for guaranteeing the existence of a spline function $s \in \mathcal{S}_d^2(\Delta_n)$ which satisfies $s(x) \geq 0$ for all $x \in \Omega$ and $s(x_i) = \zeta_i$ $(i = 0, 1, \ldots, n)$. Therefore in this paper, we use $\mathcal{S}_4^2(\Delta_n)$ as the set of all candidates of estimate of f^* and introduce the following problem as a natural extension of the standard spline smoothing.

Problem 1 (One-dimensional positive spline smoothing) *Find $s^* \in \mathcal{S}_4^2(\Delta_n)$ minimizing*

$$\sum_{i=0}^{n} |\zeta_i - s(x_i)|^2 + \lambda \int_{x_0}^{x_n} |s''(x)|^2 \, dx$$

subject to

$$s(x) \geq 0 \text{ for all } x \in \Omega.$$

In order to solve the above problem, we employ the quadratic expression of the integral in the cost function and a sufficient condition [18] for the positivity of $s \in \mathcal{S}_4^2(\Delta_n)$. Define $z_i := s(x_i)$, $p_i := s'(x_i)$ and $P_i := s''(x_i)$ for $i = 0, 1, \ldots, n$, and define $h_i := x_{i+1} - x_i$ for $i = 0, 1, \ldots, n-1$. Then for ensuring the existence of $s \in \mathcal{S}_4^2(\Delta_n)$, z_i, p_i, and P_i have to satisfy

$$12(z_i - z_{i+1}) + 6h_i(p_i + p_{i+1}) + h_i^2(P_i - P_{i+1}) = 0 \quad (1)$$

for $i = 0, 1, \ldots, n-1$. The integration in each interval $[x_i, x_{i+1}]$ is expressed as

$$\int_{x_i}^{x_{i+1}} |s''(x)|^2 \, dx = \boldsymbol{b}_i^T \boldsymbol{A}_i \boldsymbol{b}_i,$$

where $\boldsymbol{b}_i := (p_i, p_{i+1}, P_i, P_{i+1})^T \in \mathbb{R}^4$ and

$$\boldsymbol{A}_i := \frac{1}{30h_i} \begin{pmatrix} 36 & -36 & 3h_i & 3h_i \\ -36 & 36 & -3h_i & -3h_i \\ 3h_i & -3h_i & 4h_i^2 & -h_i^2 \\ 3h_i & -3h_i & -h_i^2 & 4h_i^2 \end{pmatrix}.$$

A sufficient condition in [18] for the positivity of $s \in \mathcal{S}_4^2(\Delta_n)$ over Ω is expressed as

$$\begin{cases} z_0 \geq 0, \\ 4z_i + h_i p_i \geq 0 \quad (i = 0, 1, \ldots, n-1), \\ 12z_i + 6h_i p_i + h_i^2 P_i \geq 0 \quad (i = 0, 1, \ldots, n-1), \\ 4z_i - h_{i-1} p_i \geq 0 \quad (i = 1, 2 \ldots, n), \\ z_n \geq 0. \end{cases} \quad (2)$$

2.3 Problem 1 under Sufficient Condition

Define $\boldsymbol{\zeta}, \boldsymbol{z}, \boldsymbol{p}, \boldsymbol{P} \in \mathbb{R}^{n+1}$, $\boldsymbol{b} \in \mathbb{R}^{2n+2}$ and $\boldsymbol{s} \in \mathbb{R}^{3n+3}$ as

$$\left.\begin{aligned} \boldsymbol{\zeta} &:= (\zeta_0, \zeta_1, \ldots, \zeta_n)^T \\ \boldsymbol{z} &:= (z_0, z_1, \ldots, z_n)^T \\ \boldsymbol{p} &:= (p_0, p_1, \ldots, p_n)^T \\ \boldsymbol{P} &:= (P_0, P_1, \ldots, P_n)^T \end{aligned}\right\},$$

$\boldsymbol{b} := (\boldsymbol{p}^T, \boldsymbol{P}^T)^T$ and $\boldsymbol{s} := (\boldsymbol{z}^T, \boldsymbol{p}^T, \boldsymbol{P}^T)^T = (\boldsymbol{z}^T, \boldsymbol{b}^T)^T$. We solve Problem 1 under (2) by reducing to the following convex optimization problem for $\boldsymbol{s}$.

Problem 1.A (Problem 1 under sufficient condition (2)) *Find $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ minimizing*

$$\|\boldsymbol{\zeta} - \boldsymbol{z}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b}$$

subject to

$$\boldsymbol{E}_1 \boldsymbol{s} = \boldsymbol{0} \text{ and } \boldsymbol{G}\boldsymbol{s} \geq \boldsymbol{0},$$

where $\boldsymbol{A} \in \mathbb{R}^{(2n+2)\times(2n+2)}$ is a symmetric positive semi-definite matrix constructed by using the components of $\boldsymbol{A}_i$ the two constrains $\boldsymbol{E}_1\boldsymbol{s} = \boldsymbol{0}$ (s.t. $\boldsymbol{E}_1 \in \mathbb{R}^{n\times(3n+3)}$) and $\boldsymbol{G}\boldsymbol{s} \geq \boldsymbol{0}$ (s.t. $\boldsymbol{G} \in \mathbb{R}^{(3n+2)\times(3n+3)}$) are respectively equivalent to (1) and (2).

Problem 1.A can be solved with use of the alternating direction method of multipliers (ADMM) [20] by reformulation to the ADMM-form (see Appendix) as follows.

Problem 1.B (Reformulation of Problem 1.A for ADMM) *Find $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ minimizing*

$$\|\boldsymbol{\zeta} - \boldsymbol{z}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b} + \iota_{C_1}(\boldsymbol{s}) + \iota_{C_2}(\boldsymbol{G}\boldsymbol{s}),$$

where ι_C denotes the indicator function of a nonempty closed convex set C, i.e.,

$$\iota_C(\boldsymbol{x}) := \begin{cases} 0 & \text{if } \boldsymbol{x} \in C, \\ \infty & \text{if } \boldsymbol{x} \notin C, \end{cases}$$

and nonempty closed convex sets C_1 and C_2 are defined as

$$C_1 := \left\{ \boldsymbol{s} \in \mathbb{R}^{3n+3} \,\middle|\, \boldsymbol{E}_1 \boldsymbol{s} = \boldsymbol{0} \right\}$$

and

$$C_2 := \left\{ \boldsymbol{\nu} \in \mathbb{R}^{3n+2} \,\middle|\, \boldsymbol{\nu} \geq \boldsymbol{0} \right\} = \mathbb{R}_+^{3n+2}.$$

In Appendix, by considering $\boldsymbol{x}$, $\boldsymbol{L}$, $f(\boldsymbol{x})$, and $g(\boldsymbol{L}\boldsymbol{x})$ respectively as $\boldsymbol{s}$, $\boldsymbol{G}$, $\|\boldsymbol{\zeta} - \boldsymbol{z}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b} + \iota_{C_1}(\boldsymbol{s})$, and $\iota_{C_2}(\boldsymbol{G}\boldsymbol{s})$, the first step of (4) is expressed as

$$\boldsymbol{s}_{k+1} = \operatorname*{argmin}_{\boldsymbol{E}_1 \boldsymbol{s} = \boldsymbol{0}} \|\boldsymbol{\zeta} - \boldsymbol{z}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b} + \frac{1}{2\gamma} \|\boldsymbol{\nu}_k - \boldsymbol{G}\boldsymbol{b} - \boldsymbol{\xi}_k\|^2,$$

which is solved by the method of Lagrange multipliers, and the second step of (4) is computed by the metric projection P_{C_2} onto C_2.[1] As a consequence, Problem 1.B is solved by the following iteration:

$$\left\lfloor\begin{aligned} \boldsymbol{s}_{k+1} &= \boldsymbol{\Lambda}_1 \left(\boldsymbol{I} - \boldsymbol{E}_1^T \left(\boldsymbol{E}_1 \boldsymbol{\Lambda}_1 \boldsymbol{E}_1^T \right)^{-1} \boldsymbol{E}_1 \boldsymbol{\Lambda}_1 \right) \\ &\quad \cdot \left(\begin{pmatrix} 2\boldsymbol{\zeta} \\ \boldsymbol{0} \end{pmatrix} + \frac{1}{\gamma} \boldsymbol{G}^T (\boldsymbol{\nu}_k - \boldsymbol{\xi}_k) \right) \\ \boldsymbol{\nu}_{k+1} &= P_{C_2}(\boldsymbol{G}\boldsymbol{s}_{k+1} + \boldsymbol{\xi}_k) \\ \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + \boldsymbol{G}\boldsymbol{s}_{k+1} - \boldsymbol{\nu}_{k+1} \end{aligned}\right.$$

with $\gamma > 0$ and any initialization $\boldsymbol{s}_0 \in \mathbb{R}^{3n+3}$, $\boldsymbol{\nu}_0 \in \mathbb{R}^{3n+2}$ and $\boldsymbol{\xi}_0 \in \mathbb{R}^{3n+2}$, where $\boldsymbol{I}$ is the identity matrix

$$\boldsymbol{\Lambda}_1 := \left(\begin{pmatrix} 2\boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{O} & 2\lambda \boldsymbol{A} \end{pmatrix} + \frac{1}{\gamma} \boldsymbol{G}^T \boldsymbol{G} \right)^{-1},$$

[1] For any nonempty closed convex set C, the proximity operator of the indicator function of C is equivalent to the metric projection onto C, i.e.,

$$\operatorname{prox}_{\iota_C}(\boldsymbol{x}) = P_C(\boldsymbol{x}) := \operatorname*{argmin}_{\boldsymbol{y} \in C} \|\boldsymbol{y} - \boldsymbol{x}\|^2.$$

and P_{C_2} is defined as

$$[P_{C_2}(\boldsymbol{\nu})]_i := \begin{cases} [\boldsymbol{\nu}]_i & \text{if } [\boldsymbol{\nu}]_i \geq 0, \\ 0 & \text{if } [\boldsymbol{\nu}]_i < 0. \end{cases}$$

Finally, we obtain $s^* \in \mathcal{S}_4^2(\Delta_n)$ uniquely from $\boldsymbol{s}^*$ by computing the coefficients of s^* (see (3) in Section 3.2).

3 PDF Estimation by Quartic C^2-Spline

In this section, we propose two estimation methods of a twice continuously differentiable probability density function $f^* : \mathbb{R} \to \mathbb{R}_+$ by using the ideas of the positive quartic C^2-spline. In this situation, we cannot observe values of f^* but construct a histogram from observed samples which are generated from f^* (or we can obtain only the histogram which is published by e.g., the public institution). Hence we reconstruct f^* from the histogram based on the observed samples.

3.1 Reconstruction of PDF from Its Histogram

Assume that $\{\eta_k\}_{k=1}^K$ are the observed samples and a grid $\Delta_n := \{x_i\}_{i=0}^n$ is used as the bins of the histogram s.t. $x_0 < x_1 < \ldots < x_n$, $x_0 < \min\{\eta_k\}$, and $x_n > \max\{\eta_k\}$. Let K_i $(i = 0, 1, \ldots, n-1)$ be the number of η_k which belongs to ith bin $[x_i, x_{i+1})$, i.e., $x_i \leq \eta_k < x_{i+1}$. Then the histogram roughly approximates samples, and we can expect

$$\frac{K_i}{K} \approx \int_{x_i}^{x_{i+1}} f^*(x)\, dx.$$

Therefore we reconstruct $f^* \in C^2(\mathbb{R})$ by a function $s \in C^2(\mathbb{R})$ satisfying the following properties.

- $s(x) \geq 0$ for all $x \in [x_0, x_n]$.
- $s(x) = 0$ for all $x \in (-\infty, x_0] \cup [x_n, \infty)$.
- $\displaystyle\int_{x_i}^{x_{i+1}} s(x)\, dx \approx \frac{K_i}{K}$ $(i = 0, \ldots, n-1)$.

Then from $s = 0$ over $(-\infty, x_0]$ and $[x_n, \infty)$, we only have to construct s over $[x_0, x_n] =: \Omega$. Moreover from $s \in C^2(\mathbb{R})$, s has to satisfy $s(x_0) = s(x_n) = s'(x_0) = s'(x_n) = s''(x_0) = s''(x_n) = 0$. The following problem is a generalization of the positive spline smoothing in Problem 1.

Problem 2 (PDF estimation by positive spline smoothing) *Find $s^* \in \mathcal{S}_4^2(\Delta_n)$ minimizing*

$$\sum_{i=0}^{n-1} \left| \frac{K_i}{K} - \int_{x_i}^{x_{i+1}} s(x)\, dx \right|^2 + \lambda \int_{x_0}^{x_n} |s''(x)|^2\, dx$$

subject to

$$s(x) \geq 0 \text{ for all } x \in \Omega, \qquad \int_{x_0}^{x_n} s(x)\, dx = 1$$

and

$$s(x_0) = s(x_n) = s'(x_0) = s'(x_n) = s''(x_0) = s''(x_n) = 0,$$

where $\lambda > 0$.

If we can expect

$$\frac{K_i}{K} = \int_{x_i}^{x_{i+1}} f^*(x)\, dx,$$

then the formulation in Problem 2 is not enough, and the following formulation is suitable for such cases.

Problem 3 (PDF estimation by positive spline interpolation) *Find $s^* \in \mathcal{S}_4^2(\Delta_n)$ minimizing*

$$\int_{x_0}^{x_n} |s''(x)|^2\, dx$$

subject to

$$s(x) \geq 0 \text{ for all } x \in \Omega,$$

$$\int_{x_i}^{x_{i+1}} s(x)\, dx = \frac{K_i}{K} \ (i = 0, \ldots, n-1),$$

and

$$s(x_0) = s(x_n) = s'(x_0) = s'(x_n) = s''(x_0) = s''(x_n) = 0.$$

Note that function s^* automatically satisfies the condition of the probability density function because

$$\int_{-\infty}^{\infty} s^*(x)\, dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} s^*(x)\, dx = \sum_{i=0}^{n-1} \frac{K_i}{K} = 1.$$

In the next subsection, we solve Problems 2 and 3 under (2) by the ADMM in the same manner as Section 2.3.

3.2 Problems 2 and 3 under Sufficient Condition

Suppose that a spline function $s \in \mathcal{S}_4^2(\Delta_n)$ is expressed as

$$s(x) = s_i(t) = c_4^i t^4 + c_3^i t^3 + c_2^i t^2 + c_1^i t + c_0^i$$

for $x \in [x_i, x_{i+1}]$ $(i = 0, 1, \ldots, n-1)$, where $t := \frac{x - x_i}{h_i} \in [0, 1]$ and $c_k^i \in \mathbb{R}$ $(k = 0, 1, \ldots, 4)$ are coefficients of polynomial s_i. Then by the relation

$$\begin{pmatrix} c_4^i \\ c_3^i \\ c_2^i \\ c_1^i \\ c_0^i \end{pmatrix} = \frac{1}{12} \begin{pmatrix} & 6h_i & -6h_i & 3h_i^2 & 3h_i^2 \\ & -12h_i & 12h_i & -8h_i^2 & -4h_i^2 \\ & & & 6h_i^2 & \\ & 12h_i & & & \\ 12 & & & & \end{pmatrix} \begin{pmatrix} z_i \\ p_i \\ p_{i+1} \\ P_i \\ P_{i+1} \end{pmatrix}, \tag{3}$$

we have

$$\begin{aligned} \int_{x_i}^{x_{i+1}} s(x)\, dx &= h_i \left(\frac{c_4^i}{5} + \frac{c_3^i}{4} + \frac{c_2^i}{3} + \frac{c_1^i}{2} + c_0^i \right) \\ &= \frac{h_i^2}{10}(p_i - p_{i+1}) + \frac{h_i^3}{20}(P_i + P_{i+1}) - \frac{h_i^2}{4}(p_i - p_{i+1}) \\ &\quad - \frac{h_i^3}{6}(P_i + 2P_{i+1}) + \frac{h_i^3 P_i}{6} + \frac{h_i^2 p_i}{2} + h_i z_i \\ &= h_i z_i + \frac{7h_i^2 p_i}{20} + \frac{3h_i^2 p_{i+1}}{20} + \frac{h_i^3 P_i}{20} - \frac{h_i^3 P_{i+1}}{30} = \boldsymbol{h}_i^T \boldsymbol{s}, \end{aligned}$$

where $\boldsymbol{h}_i \in \mathbb{R}^{3n+3}$ $(i = 0, 1, \ldots, n-1)$ is defined as

$$[\boldsymbol{h}_i]_k := \begin{cases} h_i & \text{if } k = i+1, \\ 7h_i^2/20 & \text{if } k = i+n+2, \\ 3h_i^2/20 & \text{if } k = i+n+3, \\ h_i^3/20 & \text{if } k = i+2n+3, \\ -h_i^3/30 & \text{if } k = i+2n+4, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover by redefining $\boldsymbol{\zeta} \in \mathbb{R}^n$ as

$$\boldsymbol{\zeta} := \left(\frac{K_0}{K}, \frac{K_1}{K}, \ldots, \frac{K_{n-1}}{K}\right)^T,$$

Problem 2 under (2) is expressed as the following convex optimization problem for $\boldsymbol{s}$.

Problem 2.A (Problem 2 under sufficient condition (2)) *Find* $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ *minimizing*

$$\|\boldsymbol{\zeta} - \boldsymbol{H}\boldsymbol{s}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A} \boldsymbol{b}$$

subject to

$$\boldsymbol{E}_2 \boldsymbol{s} = \boldsymbol{0}, \quad \boldsymbol{G}\boldsymbol{s} \geq \boldsymbol{0} \quad \textit{and} \quad \boldsymbol{1}^T \boldsymbol{H}\boldsymbol{s} = 1,$$

where $\boldsymbol{H} := (\boldsymbol{h}_0, \boldsymbol{h}_1, \ldots, \boldsymbol{h}_{n-1})^T \in \mathbb{R}^{n \times (3n+3)}$*,* $\boldsymbol{G}$ *is defined in Problem 1.A, and the condition* $\boldsymbol{E}_2 \boldsymbol{s} = \boldsymbol{0}$ *(s.t.* $\boldsymbol{E}_2 \in \mathbb{R}^{(n+6)\times(3n+3)}$*) is equivalent to (1) and* $s(x_0) = s(x_n) = s'(x_0) = s'(x_n) = s''(x_0) = s''(x_n) = 0$*, i.e.,* $z_0 = z_n = p_0 = p_n = P_0 = P_n = 0$*,* $\boldsymbol{1}$ *denotes the vector whose all components are 1, and the other condition means* $\boldsymbol{1}^T \boldsymbol{H}\boldsymbol{s} = \int_{x_0}^{x_n} s(x)\,dx = 1$.

Then we transform Problem 2.A into an equivalent problem applicable to the ADMM.

Problem 2.B (Reformulation of Problem 2.A for ADMM) *Find* $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ *minimizing*

$$\|\boldsymbol{\zeta} - \boldsymbol{H}\boldsymbol{s}\|^2 + \lambda \boldsymbol{b}^T \boldsymbol{A}\boldsymbol{b} + \iota_{C_3}(\boldsymbol{s}) + \iota_{C_2}(\boldsymbol{G}\boldsymbol{s})$$

where C_2 *is defined in Problem 2.B and*

$$C_3 := \left\{ \boldsymbol{s} \in \mathbb{R}^{3n+3} \,\middle|\, \boldsymbol{J}_1 \boldsymbol{s} = \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} \ \textit{s.t.}\ \boldsymbol{J}_1 := \begin{pmatrix} \boldsymbol{E}_2 \\ \boldsymbol{1}^T \boldsymbol{H} \end{pmatrix} \right\}.$$

In analogy with Section 2.3, the ADMM solves Problem 2.B by the following iteration:

$$\left\lfloor \begin{aligned} \boldsymbol{s}_{k+1} &= \boldsymbol{\Lambda}_2 \boldsymbol{J}_1^T \left(\boldsymbol{J}_1 \boldsymbol{\Lambda}_2 \boldsymbol{J}_1^T\right)^{-1} \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} \\ &\quad + \boldsymbol{\Lambda}_2 \Big(\boldsymbol{I} - \boldsymbol{J}_1^T \left(\boldsymbol{J}_1 \boldsymbol{\Lambda}_2 \boldsymbol{J}_1^T\right)^{-1} \boldsymbol{J}_1 \boldsymbol{\Lambda}_2\Big) \\ &\qquad \cdot \Big(2\boldsymbol{H}^T \boldsymbol{\zeta} + \frac{1}{\gamma} \boldsymbol{G}^T (\boldsymbol{\nu}_k - \boldsymbol{\xi}_k)\Big) \\ \boldsymbol{\nu}_{k+1} &= P_{C_2}(\boldsymbol{G}\boldsymbol{s}_{k+1} + \boldsymbol{\xi}_k) \\ \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + \boldsymbol{G}\boldsymbol{s}_{k+1} - \boldsymbol{\nu}_{k+1} \end{aligned} \right.$$

with $\gamma > 0$ and any initialization $\boldsymbol{s}_0 \in \mathbb{R}^{3n+3}$, $\boldsymbol{\nu}_0 \in \mathbb{R}^{3n+2}$ and $\boldsymbol{\xi}_0 \in \mathbb{R}^{3n+2}$, where

$$\boldsymbol{\Lambda}_2 := \left(2\boldsymbol{H}^T \boldsymbol{H} + \begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & 2\lambda \boldsymbol{A} \end{pmatrix} + \frac{1}{\gamma} \boldsymbol{G}^T \boldsymbol{G}\right)^{-1}.$$

Similarly, Problem 3 under (2) is expressed as follows.

Problem 3.A (Problem 3 under sufficient condition (2)) *Find* $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ *minimizing*

$$\boldsymbol{b}^T \boldsymbol{A}\boldsymbol{b}$$

subject to

$$\boldsymbol{E}_2 \boldsymbol{s} = \boldsymbol{0}, \quad \boldsymbol{G}\boldsymbol{s} \geq \boldsymbol{0} \quad \textit{and} \quad \boldsymbol{H}\boldsymbol{s} = \boldsymbol{\zeta}.$$

Then we transform Problem 3.A into the ADMM-form.

Problem 3.B (Reformulation of Problem 3.A for ADMM) *Find* $\boldsymbol{s}^* = (\boldsymbol{z}^{*T}, \boldsymbol{b}^{*T})^T \in \mathbb{R}^{3n+3}$ *minimizing*

$$\boldsymbol{b}^T \boldsymbol{A}\boldsymbol{b} + \iota_{C_4}(\boldsymbol{s}) + \iota_{C_2}(\boldsymbol{G}\boldsymbol{s}),$$

where

$$C_4 := \left\{ \boldsymbol{s} \in \mathbb{R}^{3n+3} \,\middle|\, \boldsymbol{J}_2 \boldsymbol{s} = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\zeta} \end{pmatrix} \ \textit{s.t.}\ \boldsymbol{J}_2 := \begin{pmatrix} \boldsymbol{E}_2 \\ \boldsymbol{H} \end{pmatrix} \right\}.$$

The ADMM solves Problem 3.B by the following iteration:

$$\left\lfloor \begin{aligned} \boldsymbol{s}_{k+1} &= \boldsymbol{\Lambda}_3 \boldsymbol{J}_2^T \left(\boldsymbol{J}_2 \boldsymbol{\Lambda}_3 \boldsymbol{J}_2^T\right)^{-1} \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{\zeta} \end{pmatrix} \\ &\quad + \frac{1}{\gamma} \boldsymbol{\Lambda}_3 \Big(\boldsymbol{I} - \boldsymbol{J}_2^T \left(\boldsymbol{J}_2 \boldsymbol{\Lambda}_3 \boldsymbol{J}_2^T\right)^{-1} \boldsymbol{J}_2 \boldsymbol{\Lambda}_3\Big) \boldsymbol{G}^T (\boldsymbol{\nu}_k - \boldsymbol{\xi}_k) \\ \boldsymbol{\nu}_{k+1} &= P_{C_2}(\boldsymbol{G}\boldsymbol{s}_{k+1} + \boldsymbol{\xi}_k) \\ \boldsymbol{\xi}_{k+1} &= \boldsymbol{\xi}_k + \boldsymbol{G}\boldsymbol{s}_{k+1} - \boldsymbol{\nu}_{k+1} \end{aligned} \right.$$

with $\gamma > 0$ and any initialization $\boldsymbol{s}_0 \in \mathbb{R}^{3n+3}$, $\boldsymbol{\nu}_0 \in \mathbb{R}^{3n+2}$ and $\boldsymbol{\xi}_0 \in \mathbb{R}^{3n+2}$, where

$$\boldsymbol{\Lambda}_3 := \left(\begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & 2\boldsymbol{A} \end{pmatrix} + \frac{1}{\gamma} \boldsymbol{G}^T \boldsymbol{G}\right)^{-1}.$$

Remark 1 *In some cases, in Problem 3.B, there is no vector which satisfies the constraint. For these cases, there is no solution of Problem 3.B, and the iteration based on the ADMM does not converge. On the other hand, in Problem 2.B, there always exists* $\boldsymbol{s}$ *satisfying the constraint, and we can always reconstruct* f^* *by Problem 2.B.*

Remark 2 *In Problem 3, if there exists* $i \in [0, n-1]$ *s.t.* $K_i = 0$*, the condition* $\int_{x_i}^{x_{i+1}} s(x)\,dx = K_i$ *yields* $s(x) = 0$ *for all* $x \in [x_i, x_{i+1}]$*. As a result, by adding the condition* $z_i = z_{i+1} = p_i = p_{i+1} = P_i = P_{i+1} = 0$ *into* $\boldsymbol{E}_2 \boldsymbol{s} = \boldsymbol{0}$*, the convergence speed of the ADMM becomes faster.*

4 Numerical Experiments

Let $\{\eta_k\}_{k=1}^K$ be observed samples generated from a Gaussian mixture defined as

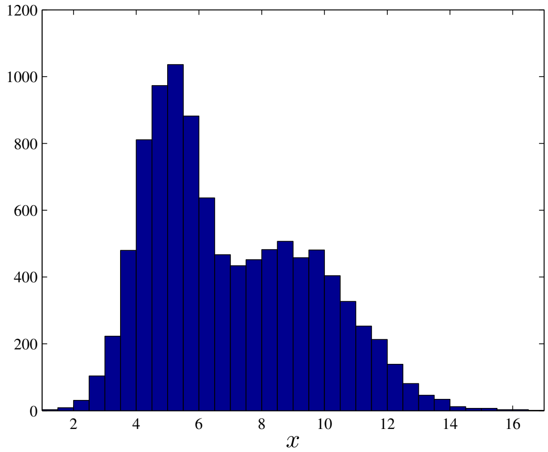
$$f^*(x) := \frac{w}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + \frac{1-w}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}},$$

where $(\mu_1, \sigma_1) = (5, 1)$, $(\mu_2, \sigma_2) = (9, 2)$ and $w = 0.5$. First, we construct a histogram as shown in Fig. 1(a) by using 10000 observed samples $\{\eta_k\}_{k=1}^{10000}$ and a grid $\Delta_n := \{x_i\}_{i=0}^{32}$ s.t. $x_i := 0.5i + 1$. In Fig. 1(b), f^* and three estimates $\hat{f}_{\mathrm{KDE}}$, s^*_{int} and s^*_{smo}, are depicted, where $\hat{f}_{\mathrm{KDE}}$ is computed by the kernel density estimation [1], [2] using Gaussian kernels as

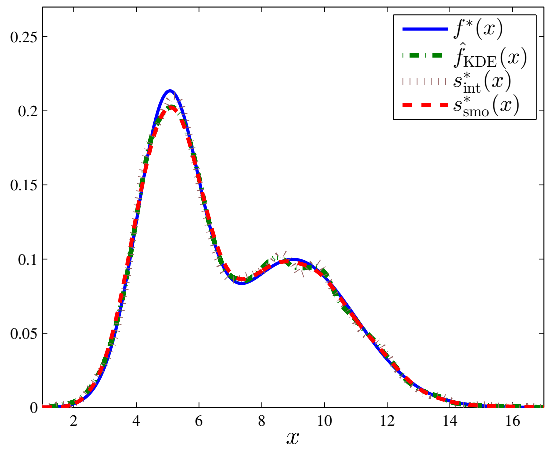
$$\hat{f}_{\mathrm{KDE}}(x) := \frac{1}{K} \sum_{k=1}^{K} \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{(x-\eta_k)^2}{2h^2}}$$

with $h = 0.2371$, and s^*_{int} and s^*_{smo} are respectively the solutions of Problems 3.B and 2.B for $\lambda = 0.015$. In this case, since there are enough samples, the shape of the histogram has a huge similarity to f^*, and all estimates approximate f^* with a high degree of accuracy.

Second, we construct a histogram as shown in Fig. 2(a) by using 500 observed samples $\{\eta_k\}_{k=1}^{500}$ and a grid $\Delta_n := \{x_i\}_{i=0}^{25}$ s.t. $x_i := 0.5i + 2$. In Fig. 2(b), $\hat{f}_{\mathrm{KDE}}$ and s^*_{smo} are respectively computed by $h = 0.5048$ and $\lambda = 0.015$. In this



(a) Histogram constructed by 10000 samples

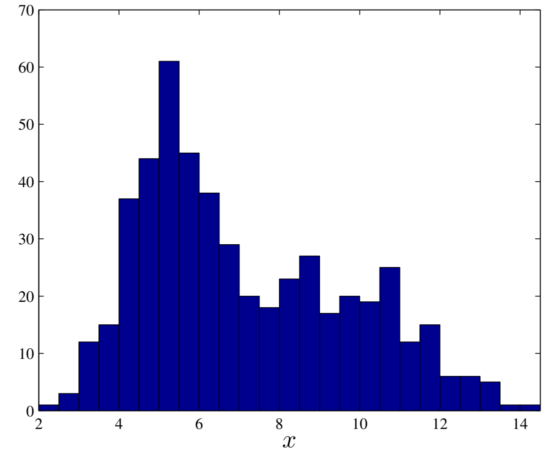


(b) True PDF f^* and three estimates $\hat{f}_{\mathrm{KDE}}$, s^*_{int} and s^*_{smo}

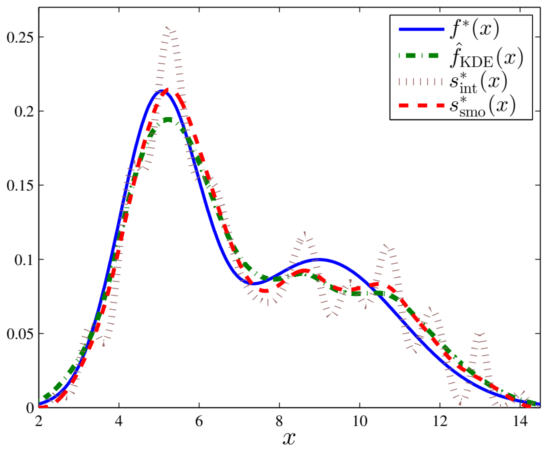
$\hat{f}_{\mathrm{KDE}}$	s^*_{int}	s^*_{smo}
0.03267	0.03773	0.03995

(c) Values of approximation of ℓ_1-norm between f^* and $\hat{f}$ ($\|f-\hat{f}\|_1 \approx \sum_{i=0}^{1600} 0.01|f^*(x_0+0.01i)-\hat{f}(x_0+0.01i)|$)

Figure 1: Experiment result based on 10000 samples



(a) Histogram constructed by 500 samples



(b) True PDF f^* and three estimates $\hat{f}_{\mathrm{KDE}}$, s^*_{int} and s^*_{smo}

$\hat{f}_{\mathrm{KDE}}$	s^*_{int}	s^*_{smo}
0.11777	0.18729	0.11619

(c) Values of approximation of ℓ_1-norm between f^* and $\hat{f}$ ($\|f-\hat{f}\|_1 \approx \sum_{i=0}^{1250} 0.01|f^*(x_0+0.01i)-\hat{f}(x_0+0.01i)|$)

Figure 2: Experiment result based on 500 samples

case, the shape of the histogram is less similar to f^*, and the values of $(K_{i+1}-K_i)$ is bigger than the first experiment. Therefore s^*_{int} does not approximate f^* and is not smooth. On the other hand, s^*_{smo} approximates f^* better than $\hat{f}_{\mathrm{KDE}}$ even though $\hat{f}_{\mathrm{KDE}}$ uses the same kernel, Gaussian, as f^*.

5 Conclusion

In this paper, first we have expanded the idea of the positive quartic C^2-spline interpolation to the positive quartic C^2-spline smoothing. Then we have proposed two estimation methods of the probability density function from the histogram by the positive quartic C^2-spline interpolation and smoothing. The proposed methods are formulated as the convex optimization problems and solved by the ADMM. Numerical experiments show that the method based on the positive quartic C^2-spline smoothing gives good estimates in the both cases of enough samples and not enough samples.

Appendix Alternating Direction Method of Multipliers

The alternating direction method of multipliers (ADMM) solves the following convex optimization problem [20]:

$$\text{Find } \boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \; f(\boldsymbol{x}) + g(\boldsymbol{L}\boldsymbol{x}),$$

where $\boldsymbol{L} \in \mathbb{R}^{m\times n}$ and $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R}\cup\{\infty\}$ are *proper, lower semicontinuous* and *convex.*[2] The

[2] A function $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ is called proper, lower semicontinous, and convex if $\mathrm{dom}(f) := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) < \infty\} \neq \varnothing$, $\mathrm{lev}_{\leq\alpha}(f) := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leq \alpha\}$ is closed for all $\alpha \in \mathbb{R}$, and $f(\lambda\boldsymbol{x}+(1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda)f(\boldsymbol{y})$ for all $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$ and $\lambda \in (0,1)$, respectively.

ADMM iteratively computes

$$\left\lfloor \begin{array}{l} \boldsymbol{x}_{k+1} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}}\ f(\boldsymbol{x}) + \dfrac{1}{2\gamma}\left\|\boldsymbol{\nu}_k - \boldsymbol{L}\boldsymbol{x} - \boldsymbol{\xi}_k\right\|^2 \\ \boldsymbol{\nu}_{k+1} = \operatorname{prox}_{\gamma g}(\boldsymbol{L}\boldsymbol{x}_{k+1} + \boldsymbol{\xi}_k) \\ \boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \boldsymbol{L}\boldsymbol{x}_{k+1} - \boldsymbol{\nu}_{k+1} \end{array} \right. \tag{4}$$

with $\gamma > 0$ and any initialization $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\boldsymbol{\nu}_0 \in \mathbb{R}^m$ and $\boldsymbol{\xi}_0 \in \mathbb{R}^m$, where $\operatorname{prox}_{\gamma g} : \mathbb{R}^m \to \mathbb{R}^m$ denotes the *proximity operator*[3] of γg. Then $(\boldsymbol{x}_k, \boldsymbol{\nu}_k, \boldsymbol{\xi}_k)$ converges to $(\boldsymbol{x}^*, \boldsymbol{L}\boldsymbol{x}^*, \boldsymbol{\xi}^*)$, where $\boldsymbol{\xi}^*$ is a solution of the dual problem.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI (Series of single-year grants) Grant Number 26·920.

REFERENCES

[1] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
[2] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *The Annals of Statistics*, vol. 38, no. 5, pp. 2916–2657, 2010.
[3] A. Criminisi, J. Shotton, and E. Konukoglu "Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning, *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 2–3, pp. 81–227, Mar. 2012.
[4] J. V. Candy, *Bayesian Signal Processing: Classical, Modern and Particle Filtering Methods*. New York, NY: Willey, 2009
[5] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press, 1995.
[6] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification* New York, NY: John Wiley & Sons, 2012.
[7] S. Kar, H. Chen, and P. K. Varshne, "Optimal identical binary quantizer design for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3896 - 3901, Jul. 2012.
[8] D. Simon and D. L. Simon, "Constrained Kalman filtering via density function truncation for turbofan engine health estimation," *International Journal of Systems Science*, vol. 41, no. 2, pp. 159–171, 2010.
[9] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures, 4th ed.* New York, NY: Wiley, 2010.
[10] D. C. Hawort, "Progress in probability density function methods for turbulent reacting flows," *Progress in Energy and Combustion Science*, vol. 36, no. 2, pp. 168–259, Apr. 2010.
[11] M. Unser, "Splines: a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22-38, Nov. 1999.
[12] H. S. Hou and H. Andrews, "Cubic splines for image interpolation and digital filtering," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 26, no. 6, pp. 508–517, Dec. 1978.
[13] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1153–1160 , Dec. 1981.
[14] E. Cohen, T. Lyche, and R. Riesenfeld, "Discrete B-splines and subdivision techniques in computer-aided geometric design and computer graphics," *Computer Graphics and Image Processing*, vol. 14, no. 2, pp. 87–111, Oct. 1980.
[15] G. Farin, *Curves and Surfaces for Computer-Aided Geometric Design: A Practical Code*, 4th ed. New York, NY: Academic Press, 1996.
[16] S. Wold, "Spline functions in data analysis," *Technometrics*, vol. 16, no. 1, pp. 87–111, 1974.
[17] L. C. Marsh and D. R. Cormier, *Spline Regression Models*, ser. Quantitative Applications in the Social Sciences. Thousand Oaks, CA: SAGE, 2001, vol. 137.
[18] W. Heß and J. W. Schmidt, "Positive quartic, monotone quintic C^2-spline interpolation in one and two dimensions," *Journal of Computational and Applied Mathematics*, vol. 55, no. 1, pp. 51–67, Oct. 1994.
[19] G. Wahba, *Spline Models for Observational Data*, ser. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: SIAM, 1990, vol. 59.
[20] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[3]The proximity operator of a proper, lower semicontinous, convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is given by

$$\operatorname{prox}_f(\boldsymbol{x}) := \underset{\boldsymbol{y}\in\mathbb{R}^n}{\operatorname{argmin}}\ f(\boldsymbol{y}) + \tfrac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{x}\right\|^2.$$