ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746780

BONA FIDE RIESZ PROJECTIONS FOR DENSITY ESTIMATION

Pol del Aguila Pla†‡ and Michael Unser‡

† CIBM Center for Biomedical Imaging, Switzerland
‡ Biomedical Imaging Group, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

ABSTRACT

The projection of sample measurements onto a reconstruction space represented by a basis on a regular grid is a powerful and simple approach to estimate a probability density function. In this paper, we focus on Riesz bases and propose a projection operator that, in contrast to previous works, guarantees the bona fide properties for the estimate, namely, nonnegativity and total probability mass 1. Our bona fide projection is defined as a convex problem. We propose solution techniques and evaluate them. Results suggest an improved performance, specifically in circumstances prone to rippling effects.

***Index Terms*—** Non-negativity, Riesz bases, generalized sampling, convex optimization.

1. INTRODUCTION

The estimation of probability density functions (pdf) pervades most problems in statistics and machine learning. For instance, the Bayes classifier achieves optimal classification, but requires an estimate of the pdf conditioned to each class. Similarly, any regression problem can be trivially solved provided a good estimate of the joint pdf between outcomes and covariates is available. Practically, pdf estimation remains one of the most common tools in data science [1–3], with its basic version (a histogram) being the entry point to any exploratory data analysis. As a result, the field remains active despite its long history [4–7].

The mathematical structure of the problem of pdf estimation is very similar to that of image reconstruction for imaging modalities that operate in the limited-photon regime, e.g., the construction of a sinogram from positron emission tomography measurements. From the observation of the empirical measure p_δ generated by N independent identically distributed samples $x_n \sim \mathcal{X}$ of a continuous random variable $\mathcal{X}$, with

$$p_\delta = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n}\,, \qquad (1)$$

one aims to recover the probability density function $f : \mathbb{R} \to \mathbb{R}_+$, of which we assume $f \in \mathrm{L}_2(\mathbb{R})$.

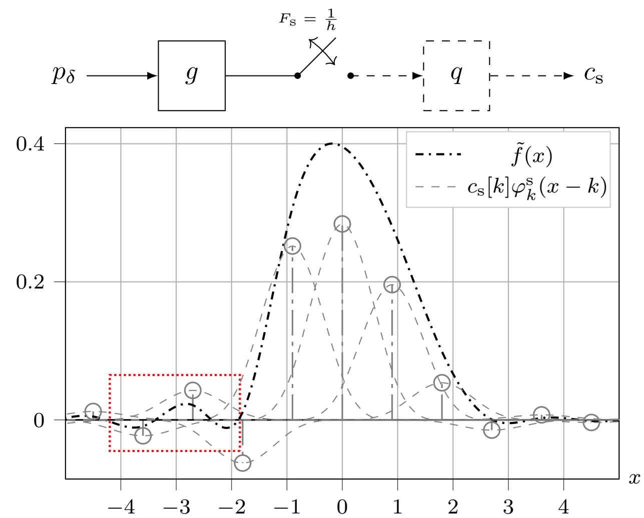

Fig. 1. Projection onto the space of uniform splines of degree 3 of the empirical estimate p_δ in (1) for $h = 0.9$ and $N = 100$ samples of a standard normal random variable. The projection is implemented by a generalized sampling system composed of a continuous filter g and a digital correction filter q. The resulting estimate is not a bona fide pdf, which is the problem we address in this paper.

In [8], our group proposed a pdf estimator that relies on the theory of generalized sampling using Riesz bases. However, in general, the resulting estimates are not bona fide pdfs. Although they integrate to 1, they are not guaranteed to be nonnegative (see Fig. 1). Recently, Cui *et al.* [5, 6] rediscovered the same estimator and studied it in much detail, but did not provide a technique to generate bona fide estimates. In this paper, we present a simple technique based on convex optimization to obtain better estimates that are bona fide pdfs within the same framework.

2. SAMPLING AND RECONSTRUCTION

A classical problem in signal processing is that of the sampling and reconstruction of continuous-domain signals [9,10]. In short, a function $f \in \mathrm{L}_2(\mathbb{R})$ is observed through a filter with impulse response $g \in \mathrm{L}_2(\mathbb{R})$ and sampled regularly at $x = kh$ for $k \in \mathbb{Z}$. The problem is then to obtain the best approximation $\tilde{f} \in \mathrm{L}_2(\mathbb{R})$ from the collected samples $c_a \in \ell_2$, defined as $c_\mathrm{a}[k] = (g * f)(kh), \forall k \in \mathbb{Z}$.

An option is to find a projection of f onto a synthesis space

$$V_\mathrm{s} = \left\{ \sum_{k\in\mathbb{Z}} c_\mathrm{s}[k]\varphi_k^\mathrm{s} : c_\mathrm{s} \in \ell_2 \right\} \subset \mathrm{L}_2(\mathbb{R}), \qquad (2)$$

where the synthesis function $\varphi^\mathrm{s} \in \mathrm{L}_2(\mathbb{R})$ is scaled and shifted to define $\varphi_k^\mathrm{s}(x) \triangleq \sqrt{1/h}\,\varphi^\mathrm{s}(x/h - k)$. For convenience, we define the analysis functions φ_k^a *mutatis mutandis* with respect to $\varphi^\mathrm{a}(x) = \sqrt{h}\,g(-hx)$ and note that $c_\mathrm{a}[k] = \langle \varphi_k^\mathrm{a}, f\rangle$. For the remainder of this paper, we assume that $\int_\mathbb{R} \varphi^\mathrm{a}(x)\mathrm{d}x = \int_\mathbb{R} \varphi^\mathrm{s}(x)\mathrm{d}x = 1$ and that $\sum_{k\in\mathbb{Z}} \varphi^\mathrm{s}(x-k) = 1$ (partition of unity). Provided that $\{\varphi_k^\mathrm{s}\}_{k\in\mathbb{Z}}$ and $\{\varphi_k^\mathrm{a}\}_{k\in\mathbb{Z}}$ are Riesz bases, the coefficients $c_\mathrm{s} \in \ell_2$ that yield a consistent reconstruction $\tilde{f} \in V_\mathrm{s}$, in the sense that $\langle \varphi_k^\mathrm{a}, \tilde{f}\rangle = c_\mathrm{a}[k]$, can be obtained using a discrete filter [9, 11] such that $c_\mathrm{s}[k] = (q * c_\mathrm{a})[k]$, where q is the convolutional inverse of the analysis-synthesis correlation sequence $r_{\mathrm{a,s}}[k] = \langle \varphi_k^\mathrm{a}, \varphi_0^\mathrm{s}\rangle$. This follows directly from the consistent reconstruction condition because

$$\begin{aligned}\langle \varphi_k^\mathrm{a}, \tilde{f}\rangle &= \sum_{k'\in\mathbb{Z}} c_\mathrm{s}[k']\langle \varphi_k^\mathrm{a}, \varphi_{k'}^\mathrm{s}\rangle = \sum_{k'\in\mathbb{Z}} c_\mathrm{s}[k'] r_{\mathrm{a,s}}[k-k'] \\ &= (r_{\mathrm{a,s}} * c_\mathrm{s})[k]\,. \end{aligned} \qquad (3)$$

The function $\tilde{f}$ so constructed is a projection of f onto V_s, because $\mathrm{P}_{V_\mathrm{s}} : \mathrm{L}_2(\mathbb{R}) \to V_\mathrm{s}$ defined as $f \mapsto \tilde{f}$ fulfills that $\mathrm{P}_{V_\mathrm{s}}\{\tilde{f}\} = \tilde{f}$. Particularly, if $V_\mathrm{s} = V_\mathrm{a}$, with V_a defined analogously to (2), then $\tilde{f}$ is the minimum L_2-norm approximation of f in V_s (*i.e.*, its orthogonal projection). The polynomial-reproduction properties of φ^s [12] then characterize the behavior of $\|\mathrm{P}_{V_\mathrm{s}}\{f\} - f\|_{\mathrm{L}_2(\mathbb{R})}$ as h gets small, known as the order of approximation.

3. RIESZ PROJECTIONS FOR PDF ESTIMATION

This approach to signal reconstruction can be extended to the estimation of pdfs [8]. In particular, if one chooses $\varphi^\mathrm{a}, \varphi^\mathrm{s} \in \mathcal{C}_0(\mathbb{R})$, where $\mathcal{C}_0(\mathbb{R})$ is the space of continuous functions that decay at infinity equipped with the uniform norm $\|\cdot\|_\infty$, then $\langle \delta_x, \varphi^\mathrm{a}\rangle = \varphi^\mathrm{a}(x)$ is well defined. Consequently, the approach described above can be directly applied on the empirical estimate p_δ, as portrayed in Figure 1, to obtain an estimate $\tilde{f}$ of the pdf. For $\varphi^\mathrm{a} = \varphi^\mathrm{s} = \beta^0(x)$, this results in a traditional histogram. Here, β^m for $m \in \mathbb{N}$ is the uniform B-spline of degree m. We illustrate in Figure 2 the measurement process $\langle p_\delta, \varphi_k^\mathrm{a}\rangle$ when β^0 and β^1 are used as φ^a. In the general case, Blu and Unser [8] showed that $\tilde{f} = \mathrm{P}_{V_\mathrm{s}}\{p_\delta\}$ is an L_2-consistent estimator for f. They also characterized thoroughly its expected L_2-error averaged over all possible shifts

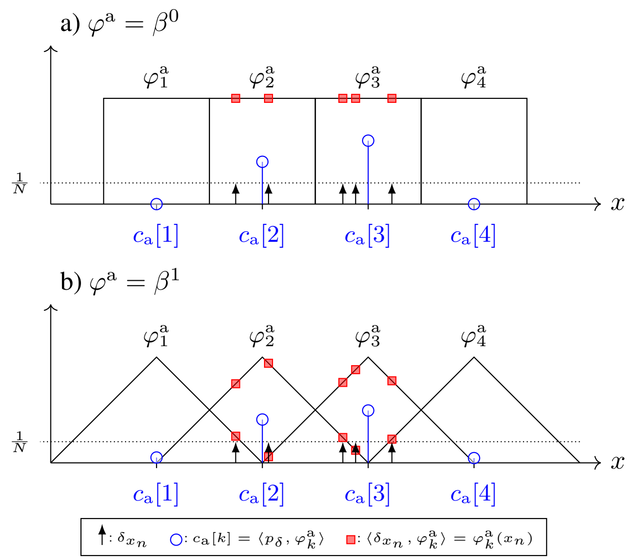

Fig. 2. Measurement procedure $c_\mathrm{a}[k] = \langle p_\delta, \varphi_k^\mathrm{a}\rangle$ for the cases $\varphi^\mathrm{a} = \beta^0$ and $\varphi^\mathrm{a} = \beta^1$. The coefficient $c_\mathrm{a}[k]$ is constructed as an average of the contributions $\varphi_k^\mathrm{a}(x_n)$ of all the samples over the support of φ_k^a.

within h

$$\tilde{\eta}^2 = \frac{1}{h}\int_0^h \mathrm{E}\left\{ \left\| f_\tau - \mathrm{P}_{V_\mathrm{s}}\left\{ \frac{1}{N}\sum_{n=1}^N \delta_{x_n+\tau} \right\} \right\|_{\mathrm{L}_2(\mathbb{R})}^2 \right\} \mathrm{d}\tau. \qquad (4)$$

There, $f_\tau(x) \triangleq f(x-\tau)$ is used for convenience.

Because φ^s satisfies the partition of unity, we have that $\int_{-\infty}^{\infty} \tilde{f}(x)\mathrm{d}x = 1$. Regretfully, this does not guarantee that $\tilde{f}$ will be a bona fide pdf, and we often encounter $\tilde{f}(x) < 0$ for some $x \in \mathbb{R}$. When $\varphi^\mathrm{a} = \varphi^\mathrm{s} = \beta^m$ for some $m \in \mathbb{N}$, the scheme in Figure 1 yields an orthogonal projection. Nonetheless, for degrees $m \geq 1$, the resulting estimate may exhibit negative values due to ripples in the estimate. For example, this is seen inside the dotted box in Figure 1. To address this shortcoming of the method, we propose the following projection operator based on an optimization model.

Definition 1 (Bona Fide Projection). The bona fide projection onto V_s is the operator $\mathrm{P}_{\mathcal{BF}} : \mathcal{M}(\mathbb{R}) \to V_\mathrm{s}$ such that $p_\delta \mapsto \tilde{f}_+$ with

$$\begin{aligned}\tilde{f}_+ &= \arg\min_{\breve{f}\in V_\mathrm{s}} \left\{ \|\langle p_\delta, \varphi_k^\mathrm{a}\rangle - \langle \varphi_k^\mathrm{a}, \breve{f}\rangle\|_{\ell_2}^2 \right\}, \\ &\quad \text{such that } \breve{f}(x) \geq 0, \forall x \in \mathbb{R}, \text{ and } \int_{-\infty}^{\infty} \breve{f}(x)\mathrm{d}x = 1\,. \end{aligned} \qquad (5)$$

Here, $\mathcal{M}(\mathbb{R})$ is the space of bounded Radon measures. It is the continuous dual of $\mathcal{C}_0(\mathbb{R})$ and contains Dirac delta distributions. Because the set of constraints is convex and the cost function is strictly convex, (5) has a unique solution.

We may express succinctly the set of constraints by defining the subset of bona fide pdfs in V_{s} as $\mathcal{BF} = \{\breve{f} \in V_{\mathrm{s}} : \breve{f}(x) \geq 0, \forall x \in \mathbb{R}, \text{and } \int_{-\infty}^{\infty} \breve{f}(x)\mathrm{d}x = 1\}$. Lemma 1 establishes that (5) indeed does define a projection operator, precisely onto $\mathcal{BF} \subset V_{\mathrm{s}}$.

Lemma 1 (Projection). The operator $\mathrm{P}_{\mathcal{BF}}$ defined in (5) is a projection operator onto $\mathcal{BF}$.

Proof: Let $\tilde{f}_+ \in \mathcal{BF}$. Because $\mathcal{BF} \subset V_{\mathrm{s}} \bigcap \mathrm{L}_1(\mathbb{R})$ by construction ($\|\tilde{f}_+\|_{\mathrm{L}_1} = 1$ for any $\tilde{f}_+ \in \mathcal{BF}$), we have that $\mathcal{BF} \subset \mathcal{M}(\mathbb{R})$ (see [13]), and thus $\mathrm{P}_{\mathcal{BF}}\{\tilde{f}_+\}$ is well defined. Let $c_{\mathrm{s}}^{\tilde{f}_+}$ be the unique coefficients of $\tilde{f}_+$ on the basis $\{\varphi_k^{\mathrm{s}}\}_{k\in\mathbb{Z}}$, and $c_{\mathrm{s}}^{\breve{f}}$ analogously for $\breve{f}$. Then, the cost function in (5) is $\|r_{\mathrm{a,s}} * (c_{\mathrm{s}}^{\tilde{f}_+} - c_{\mathrm{s}}^{\breve{f}})\|_{\ell_2}^2$. Because $r_{\mathrm{a,s}}$ has a convolutional inverse and $\tilde{f}_+ \in \mathcal{BF}$, the unique solution that achieves cost 0 and fulfills the constraints in (5) is $\breve{f} = \tilde{f}_+$. Therefore, $\mathrm{P}_{\mathcal{BF}}\{\tilde{f}_+\} = \tilde{f}_+$ and $\mathrm{P}_{\mathcal{BF}}$ is a projection operator onto $\mathcal{BF}$. ■

Leveraging (3), the optimization problem (5) can be equivalently stated in terms of the coefficients c_{s} of $\breve{f}$ as

$$\min_{c_{\mathrm{s}}\in\ell_2} \left\{\|c_{\mathrm{a}} - r_{\mathrm{a,s}} * c_{\mathrm{s}}\|_{\ell_2}^2\right\}$$
$$\text{such that } \sum_{k\in\mathbb{Z}} c_{\mathrm{s}}[k]\varphi_k^{\mathrm{s}}(x) \geq 0, \forall x \in \mathbb{R} \text{ and } \sum_{k\in\mathbb{Z}} c_{\mathrm{s}}[k] = 1\,.$$

However, the enforcement of the non-negativity constraint is generally a hard problem for all but the simplest bases $\{\varphi_k^{\mathrm{s}}\}_{k\in\mathbb{Z}}$. For example, for spline functions of degree m, this entails as many as m semidefinite constraints in the polynomial coefficients describing each segment $[k, k+1]$ (see [14–16]). In this paper, we take a general approach that is valid for any φ^{s} and relies only on linear constraints and convolution. It is inexact but can be made arbitrarily close to (5) at the cost of increased computational complexity. Specifically, for a given $M \in \mathbb{N}$, we impose that $\breve{f}(k/M) \geq 0, \forall k \in \mathbb{Z}$. The constrained values are readily computed as $\breve{f}(k/M) = (c_{\mathrm{s}}^{\uparrow M} * \varphi^{\mathrm{s},M})[k]$, where i) $c_{\mathrm{s}}^{\uparrow M}$ is a sequence of coefficients upsampled from c_{s}, so that it contains $(M-1)$ zeros between $c_{\mathrm{s}}[k]$ and $c_{\mathrm{s}}[k+1]$, for every $k \in \mathbb{Z}$, and ii) $\varphi^{\mathrm{s},M}[k] = \varphi^{\mathrm{s}}(k/M)$, which corresponds to a short finite-impulse-response filter because φ^{s} is often chosen with a small support. Therefore, the final optimization problem becomes

$$\min_{c_{\mathrm{s}}\in\ell_2} \left\{\|c_{\mathrm{a}} - (r_{\mathrm{a,s}} * c_{\mathrm{s}})\|_2^2\right\}$$
$$\text{such that } (c_{\mathrm{s}}^{\uparrow M} * \varphi^{\mathrm{s},M})[q] \geq 0, \forall q \in \mathbb{Z} \text{ and } \sum_{k\in\mathbb{Z}} c_{\mathrm{s}}[k] = 1\,. \tag{6}$$

Problem (6) has a quadratic-programming structure and can be solved by a number of standard iterative techniques.

As we shall see in Section 4, and specifically in Figures 3 and 4, empirical results suggest that the localized non-negative constraints of (6) lead to estimates that are in $\mathcal{BF}$ and result in smaller approximation errors and a smoother behavior than the unconstrained solution.

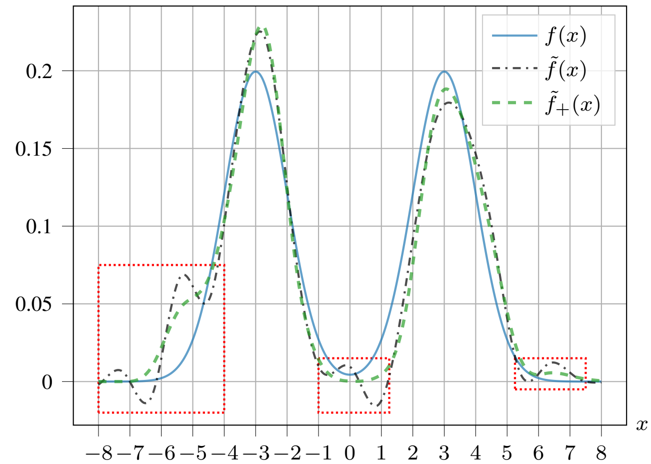

Fig. 3. Estimates $\tilde{f} = \mathrm{P}_{V_{\mathrm{s}}}\{p_\delta\}$ and $\tilde{f}_+ = \mathrm{P}_{\mathcal{BF}}\{p_\delta\}$ obtained from $N = 100$ samples of a Gaussian mixture, $h = 0.9$, and $\varphi^{\mathrm{a}} = \varphi^{\mathrm{s}} = \beta^3$. Evidently, $\tilde{f}_+$ is a better estimate of f. Most importantly, it is also a bona fide pdf.

4. EMPIRICAL RESULTS

We implemented $\mathrm{P}_{V_{\mathrm{s}}}$ as described in Sections 2 and 3 in `Python 3.9.5`, leveraging the NumPy [17] and SciPy [18] libraries for computations. We implemented an approximated $\mathrm{P}_{\mathcal{BF}}$ as described by (6), which was solved with $M = 10$ using `CVXPY` [19]. An open-source repository that contains all implementations, including those to generate all the figures in this paper, is available through GitHub[1].

A comparison of [8] and the proposed method on 100 samples of an equal mixture of $\mathcal{N}(3, 1)$ and $\mathcal{N}(-3, 1)$ with $h = 0.9$ and $\varphi^{\mathrm{a}} = \varphi^{\mathrm{s}} = \beta^3$ results in the estimates $\tilde{f} = \mathrm{P}_{V_{\mathrm{s}}}\{p_\delta\}$ and $\tilde{f}_+ = \mathrm{P}_{\mathcal{BF}}\{p_\delta\}$ shown in Figure 3. There, one can immediately appreciate that $\tilde{f}_+$ fulfills the constraints of $\mathcal{BF}$, at least within visual tolerance. Furthermore, as highlighted by the dotted boxes in the figure, $\tilde{f}_+$ is a much better estimate overall, exhibiting a much less ripples at the locations where the true distribution f changes sharply. This applies even when the resulting ripples in $\tilde{f}$ do not produce negative values (lower-right box).

A more detailed study of the expected shift-averaged error incurred by both $\tilde{f}$ and $\tilde{f}_+$ is included in Figure 4 and compared to the theoretical predictions of [8] for $\tilde{f}$. Both methods are evaluated on 100 samples of a standard normal distribution $\mathcal{N}(0, 1)$ for $h \in [0.8, 1.6]$. To approximate $\tilde{\eta}$ empirically, the error is computed by numerical integration over x and is averaged over 120 realizations. For each realization, the data and the probability density function are shifted by all multiples of 0.025 between 0 and h, and the results are averaged to approximate the integral over τ in (4). The results suggest that $\tilde{f}_+$ improves on $\tilde{f}$ by roughly 1 dB for all reasonable values of h. While this improved error is certainly an advantage, the

[1] https://github.com/poldap/rpde

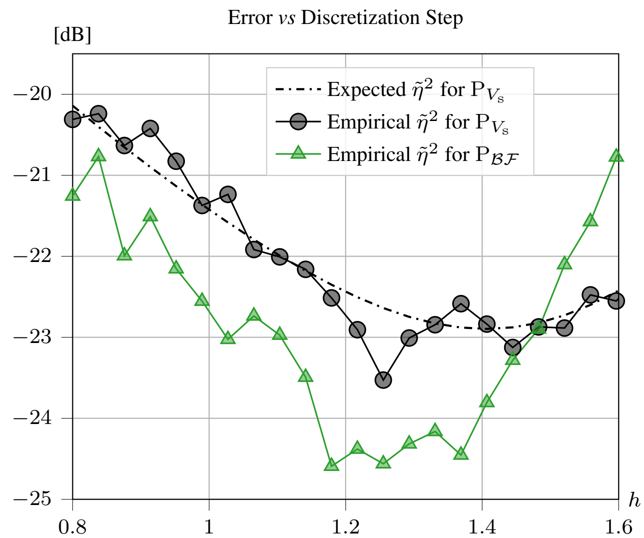

Fig. 4. Performance metric $\tilde{\eta}$ in (4) evaluated for $\mathrm{P}_{V_\mathrm{s}}$ and $\mathrm{P}_{\mathcal{BF}}$ applied to $N = 100$ samples of a standard Gaussian. Seen here as function of the grid size h and for $\varphi^\mathrm{a} = \varphi^\mathrm{s} = \beta^3$.

main benefit of $\mathrm{P}_{\mathcal{BF}}$ is that its output is directly usable for any application of pdf estimation, because the estimate $\tilde{f}_+ \in \mathcal{BF}$ is a bona fide pdf.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Aleix Boquet-Pujadas for extensive discussions on the role of non-negativity in spline approximations.

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE), and Geneva University Hospitals (HUG).

6. REFERENCES

[1] A. J. Izenman, “Recent developments in nonparametric density estimation,” *Journal of the American Statistical Association*, vol. 86, no. 413, pp. 205–224, 1991.

[2] T. Takada, “Asymptotic and qualitative performance of nonparametric density estimators: A comparative study,” *Econometrics Journal*, vol. 11, no. 3, pp. 573–592, 2008.

[3] Z. Wang and D. W. Scott, “Nonparametric density estimation for high-dimensional data—Algorithms and applications,” *WIREs Computational Statistics*, vol. 11, no. 4, p. e1461, 2019.

[4] A. Uppal, S. Singh, and B. Poczos, “Nonparametric density estimation & convergence rates for GANS under Besov IPM losses,” in *Advances in Neural Information Processing Systems*, H. Wallach *et al.*, Eds., vol. 32. Curran Associates, Inc., 2019.

[5] Z. Cui, J. L. Kirkby, and D. Nguyen, “Nonparametric density estimation by B-spline duality,” *Econometric Theory*, vol. 36, no. 2, pp. 250–291, 2020.

[6] J. L. Kirkby, A. Leitao, and D. Nguyen, “Nonparametric density estimation and bandwidth selection with B-spline bases: A novel Galerkin method,” *Computational Statistics & Data Analysis*, vol. 159, p. 107202, 2021.

[7] F. Ferraccioli, E. Arnone, L. Finos, J. O. Ramsay, and L. M. Sangalli, “Nonparametric density estimation over complicated domains,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 83, no. 2, pp. 346–368, 2021.

[8] T. Blu and M. Unser, “Quantitative $\mathbf{L}^2$ approximation error of a probability density estimate given by its samples,” in *Proceedings of the Twenty-Ninth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP’04)*, vol. III, Montréal QC, CA, May 17-21 2004, pp. 952–955.

[9] M. Unser, “Sampling—50 Years after Shannon,” *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.

[10] Y. C. Eldar, *Sampling Theory, Beyond Bandlimited Systems*. Cambridge University Press, 2015.

[11] M. Unser and A. Aldroubi, “A general sampling theory for nonideal acquisition devices,” *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 2915–2925, 1994.

[12] T. Blu and M. Unser, “Quantitative Fourier analysis of approximation techniques: Part I—Interpolators and projectors,” *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2783–2795, 1999.

[13] M. Unser, “A note on BIBO stability,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 5904–5913, 2020.

[14] F. Alizadeh, J. Eckstein, N. Noyan, and G. Rudolf, “Arrival rate approximation by nonnegative cubic splines,” *Operations Research*, vol. 56, no. 1, pp. 140–156, 2008.

[15] F. Alizadeh and D. Papp, “Estimating arrival rate of nonhomogeneous Poisson processes with semidefinite programming,” *Annals of Operations Research*, vol. 208, no. 1, pp. 291–308, 2013.

[16] D. Papp and F. Alizadeh, “Shape-constrained estimation using nonnegative splines,” *Journal of Computational and Graphical Statistics*, vol. 23, no. 1, pp. 211–231, 2014.

[17] C. R. Harris *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[18] P. Virtanen *et al.*, “SciPy 1.0: Fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, 2020.

[19] S. Diamond and S. Boyd, “CVXPY: A Python-embedded modeling language for convex optimization,” *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.