PET REBINNING WITH REGULARIZED DENSITY SPLINES

Aleix Boquet-Pujadas†, Pol del Aguila Pla†‡, and Michael Unser†

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

ABSTRACT

PET reconstruction algorithms have long relied on sinogram rebinning. However, as detectors grow smaller in a recent wave of cutting-edge scanners, individual sensors no longer accrue hundreds of photons. Instead, most detect a single photon or none at all, effectively turning sinogram data into point-cloud measurements. The highly heterogeneous sensitivity of these scanners is another issue. We approach sinogram rebinning in the face of these challenges with a density-estimation framework that promotes knot sparsity in an underlying spline basis.

Index Terms—Positron emission tomography, Poisson process, Density estimation, Hessian-Schatten norm.

1. INTRODUCTION

Modern small-animal PET devices achieve sub-mm detail. The race for higher and higher resolution has led experimental scanners to the half-millimeter mark through the use of crystal scintillators with pitches as small as 0.32 mm, 0.43 mm, or 0.5 mm [1–3]. Silicon-based sensors appear ready to take over now, promising pixel arrays at the 100 μm scale [4]. However, the efficiency of these new scanners does not scale in cubic proportion to their pitch. The resulting sinograms are therefore much sparser than those acquired in standard PET, with the vast majority of lines of response (LOR) accruing either a single photon or none at all. In addition, the diverse geometries of new scanner architectures and the physics behind their stopping power [4] give rise to sensitivity maps with highly heterogeneous detection probabilities (Figure 1). These two peculiarities hinder the applicability of common PET reconstruction pipelines to these new scanners.

One concept to revise in light of these challenges is rebinning. PET algorithms have long relied on rebinning to bypass the computational burden of direct reconstruction methods when forward models grow prohibitively large [5]. For instance, the potential size of sinogram spaces in state-of-the-art scanners (10^{15} possible LORs for the scanner in Figure 1) can preclude some computations in approaches that rely on the popular method of ordered subset expectation maximization. For these reasons, rebinning continues to be used in commercial scanners, in medical studies, and in new proto-
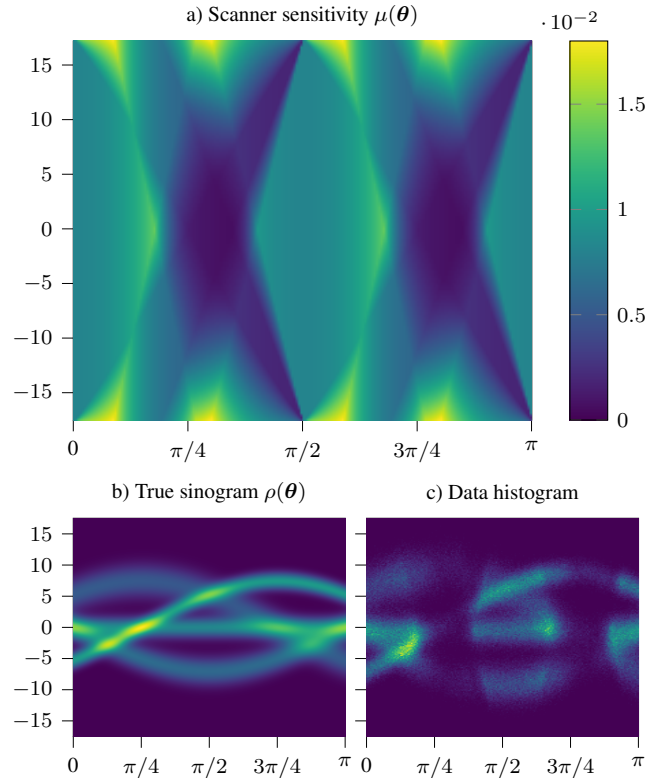

Fig. 1. **a)** Irregular sensitivity $\mu(\boldsymbol{\theta})$ over a 2D sinogram ($\boldsymbol{\theta} = (r, \theta)$) of a state-of-the-art experimental PET scanner made of silicon detectors [4]. The simulated scanner is rectangular. It has a field of view of 34 mm, and detectors with a pitch of 100 μm. **b)** Sinogram of the image in Figure 4. **c)** Histogram of the data sampled from (b) with sensitivity (a). The total number of counts is $5 \cdot 10^5$.

type scanners [1, 6]. Rebinning has also been incorporated into time-of-flight PET systems [7]. List-mode processing is another very common alternative.

In the spirit of physical collimators, rebinning techniques interpolate 3D LOR data onto equally spaced 2D slices with the aim to leverage reconstruction algorithms originally developed for 2D tomography [5, 8]. In the process, each 2D sinogram is cast on a uniform grid to reproduce the set of parallel rays necessary for the filtered back-projection (FBP) algorithm or for some fast iterative methods. The sensitivity of

the scanner is addressed during preprocessing via pointwise division of a reference scan.

With the marked decrease of the ratio of photons per pixel in new scanners, however, function interpolation becomes less adequate. In a regime of point-cloud measurements, we argue that rebinning is instead better regarded from the perspective of density estimation, in consideration of the randomness underlying the photon-emission process. Moreover, there is a need for a robust approach to compensate for the scanner sensitivity when the values of the corresponding maps span a whole order of magnitude (Figure 1).

Our aim is to offer a rebinning method with enough flexibility to adapt to new scanner designs in the face of big, sparse data. We propose a density-estimation framework based on a regularized optimization problem. It promotes knot sparsity in an underlying spline representation and considers the sensitivity as a weighting measure. The results are probability densities with the computational advantage of the small support of B-splines [9]. Our method is applicable to other density-estimation problems, as well as to traditional PET scanners if quantization were accounted for [10]. It can be used for the usual 3D to 2D rebinning, but also for 2D to 2D and 3D to 3D—e.g., when 3D parallel reconstruction is feasible.

2. HESSIAN-SCHATTEN DENSITY SPLINES

We define the density of interest $\rho : \Theta \subset \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ in sinogram space (Figure 1b) as the extension

$$\rho(\boldsymbol{\theta}; \mathbf{c}) = \exp\left(\sum_{\mathbf{n} \in \mathcal{N}} \mathbf{c}_{\mathbf{n}} \Lambda(\boldsymbol{\theta} - \mathbf{n})\right) \quad (1)$$

of a logspline of degree 1 (c.f. [11]) to a d-dimensional periodic domain, where $\boldsymbol{\theta} \in \Theta$ refers to a specific LOR in the sinogram. The exponential guarantees non-negativity. Expression (1) is parameterized by a set of coefficients $\mathbf{c} \in \mathcal{C} = \prod_{k=1}^{d} \mathbb{R}^{N_k}$ indexed on a uniform unit grid, $\mathbf{n} \in \mathcal{N} = \prod_{k=1}^{d} \{0, 1, \ldots, N_k - 1\}$. The underlying linear B-spline is defined as

$$\Lambda(\boldsymbol{\theta}) = \prod_{k=1}^{d} \max\left(1 - |\boldsymbol{\theta}_k \bmod (N_k - 1)|, 0\right). \quad (2)$$

The density ρ is not directly accessible because it is experimentally obscured by the sensitivity map $\mu : \Theta \to \mathbb{R}_{\geq 0}$, which characterizes the probability of detecting an event within a LOR as determined by the stopping power of the scanner (Figure 1a). We therefore formulate the probability density of observing a LOR as

$$\rho_{\mathrm{obs}}(\boldsymbol{\theta}; \mathbf{c}) = I^{-1} \rho(\boldsymbol{\theta}; \mathbf{c}) \mu(\boldsymbol{\theta}), \quad (3)$$

where $I = \int_{\Theta} \rho(\boldsymbol{\theta}) \mu(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$. Note that choosing μ is equivalent to choosing a measure and that this avoids the point-wise division typical of other weighted estimation methods.
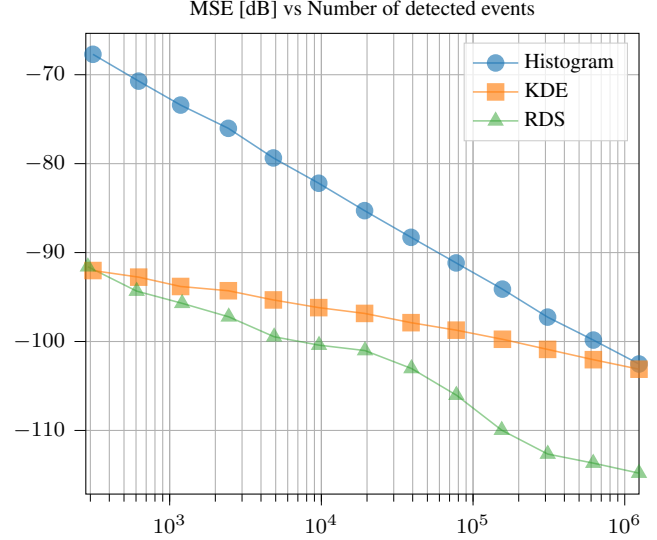

Fig. 2. MSE of the estimate of the sinogram for a weight-aware histogram and a KDE estimates with automatic bin and bandwidth selection, compared to RDS with $\alpha = 5$ on a (200×168) grid. Note that the MSE values are low because the density values are necessarily bound to a relatively small range.

To formulate an optimization problem, we proceed by defining a data-fidelity term based on the likelihood of observing a set $\{\boldsymbol{\theta}^m\}_{m=1}^{M}$ of independent samples from the observed density $\rho_{\mathrm{obs}}(\boldsymbol{\theta})$. We express this likelihood as

$$L(\mathbf{c}) = I^{-M} \prod_{m=1}^{M} \rho(\boldsymbol{\theta}^m; \mathbf{c}) \mu(\boldsymbol{\theta}^m), \quad (4)$$

where we consider that the samples $\boldsymbol{\theta}^m$ have been scaled to the uniform grid $\mathcal{N}$. We also want to enforce some *a priori* behavior through the regularization term

$$R(\mathbf{c}) = \sum_{\mathbf{n} \in \mathcal{N}} \sum_{k=1}^{d} |\sigma_k\left(\mathrm{H}\{\log \rho\}(\mathbf{n})\right)|, \quad (5)$$

which approximates the Schatten 1-norm [12] of the Hessian (H) of the logarithm of the density, via its eigenvalues σ_k on the grid. The aim is to automatically sparsify the knots of the underlying linear spline. Together, (4) and (5) combine into a minimization problem for the optimal spline coefficients, with solution

$$\mathbf{c}^* := \arg\min_{\mathbf{c} \in \mathcal{C}} \left\{\alpha R(\mathbf{c}) - \log\left(L(\mathbf{c})\right)\right\}. \quad (6)$$

The log-likelihood is twice continuously differentiable and negative-definite, and the regularization term is convex, making the cost functional in (6) strictly convex. The optimization is done using an accelerated proximal gradient algorithm [13] that combines the gradient of the negative log-likelihood with the proximal operator of the regularization term [12]. The latter is computed using an inner dual scheme with 20 iterations.
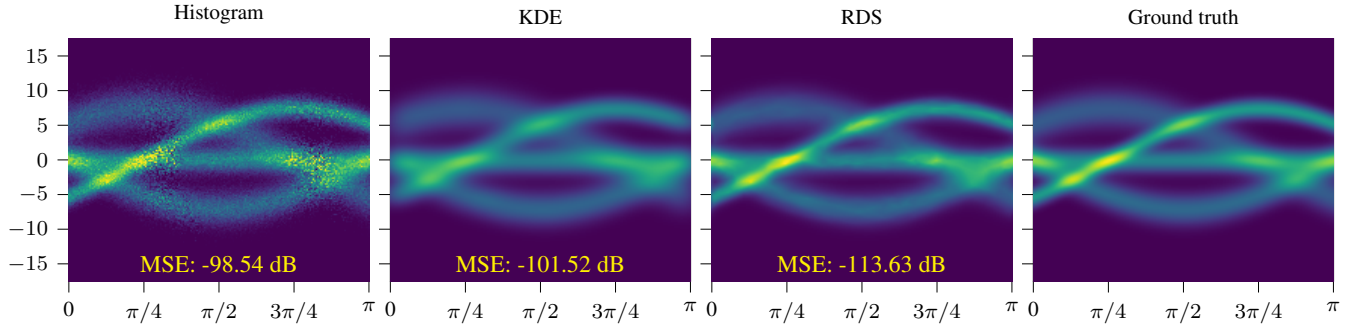
Fig. 3. Sinograms for weight-aware histogram and KDE compared to our RDS method with $5 \cdot 10^5$ detections. The corresponding object in the image domain is shown in the rightmost pane of Figure 4.

The Lipschitz constant of the outer algorithm is updated at every iteration. Hereafter, we refer to the framework described in this section as regularized density splines (RDS).

3. EMPIRICAL RESULTS AND DISCUSSION

We tested RDS against two other weighted-density estimation techniques in the context of an experimental PET scanner. These were kernel density estimation (KDE) and histogram estimation. The bandwidths of both methods were adjusted using Scott's rule and the maximum among the Freedman-Diaconis' and Sturges' estimators, respectively. The kernel was chosen to be Gaussian for KDE.

To obtain an experimental μ, we assessed the sensitivity of the scanner in the sinogram domain. Measurements showed that the chance of detection of LORs varied by an order of magnitude as a function of their angle and distance to the center (Figure 1a). This heterogeneity had a sizeable effect on the resulting sinograms, which displayed zones with fewer photons than those of an ideal projection under a constant sensitivity (Figure 1c). In addition, only 0.01% of the LORs accrued more than one detection, thereby justifying the use of a density-estimation approach.

We chose a sum of Gaussians as the target radiotracer distribution in the image domain (Figure 4, rightmost pane). This is in favor of KDE, as Gaussianity is preserved over the distance upon projection onto the sinogram domain (Figure 3, rightmost pane). This distribution also leads to a sinogram that is visually interpretable, yet rich enough to test the adaptiveness of the algorithms. We simulated the scanner acquisition for different lengths of time spanning 4 orders of magnitude and, thus, for numbers of detections over a proportional range (Figure 2). The resulting point clouds in sinogram domain (Figure 1c) were then fed into the three methods.

RDS outperformed the two other methods over the full range tested, except at the minimum value of $\approx 10^2$ detections, where the Gaussian kernel in KDE likely made up for the lack of data. For a number of detections relevant to real PET scans (10^5 and higher), the improvement amounted to a full order of magnitude. At these levels, the histogram method reached an accuracy comparable to that of KDE.

The regularization parameter α of the RDS method was kept constant across experiments after having been adjusted at the experimentally relevant mark of $5 \cdot 10^5$ detections; there, the best parameter among three values, corresponding to three orders of magnitude, was selected. This low sensitivity to α likely reflects a sparsifying effect of the Hessian-Schatten norm in our context of linear splines. (We recall that the other two estimation techniques auto-adjust their bandwidths.) Another advantage of our method is its favorable computation time. For $5 \cdot 10^5$ points, the evaluation of the density on a (200×168) parallel grid took 1 ms for RDS compared to 10 min for KDE and 40 ms for the histogram. RDS took 10 s to optimize whereas the other methods were setup in 20 ms. The comparison was on CPU, but we expect a bigger advantage on GPU as RDS is more amenable to parallelization.

Our qualitative assessment of the estimated sinograms (Figure 3) is that RDS is able to adapt to regions with different underlying variance—of the simulated Gaussians—while compensating for the lack of data in regions of less sensitivity. Conversely, KDE over-smoothed some regions and had a reduced accuracy on their boundaries, whereas histogram-based estimation struggled to properly correct for the sensitivity. The improvement in sinogram domain shown by RDS translated into better mean-square errors (MSEs) in the image domain upon reconstruction with the FBP algorithm (Figure 4). Similar improvements were found for a Derenzo phantom (Figure 5), where RDS also reduced the streak artifacts introduced by the FBP. Comparatively better results were found using iterative algorithms because the artifacts introduced by RDS are more naturally compensated by edge-preserving regularization schemes than those of the two competing approaches (data not shown). All implementations in this report were based on NumPy and SciPy.

4. COMPLIANCE WITH ETHICAL STANDARDS

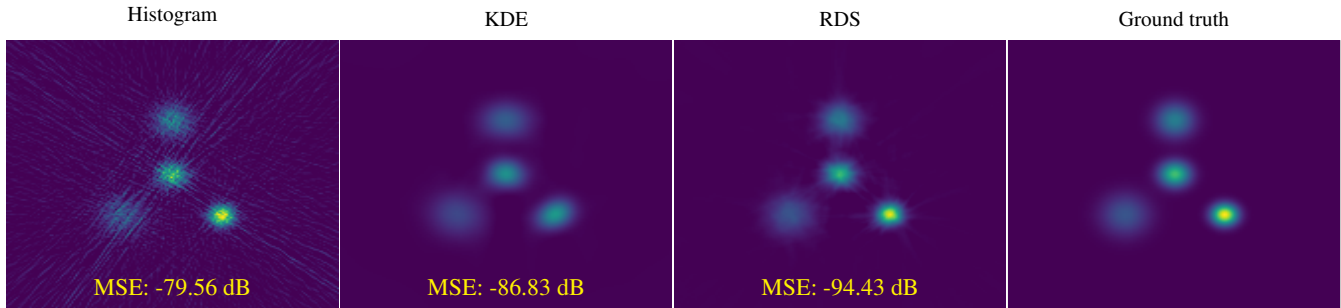This is a numerical simulation study for which no ethical approval was required.

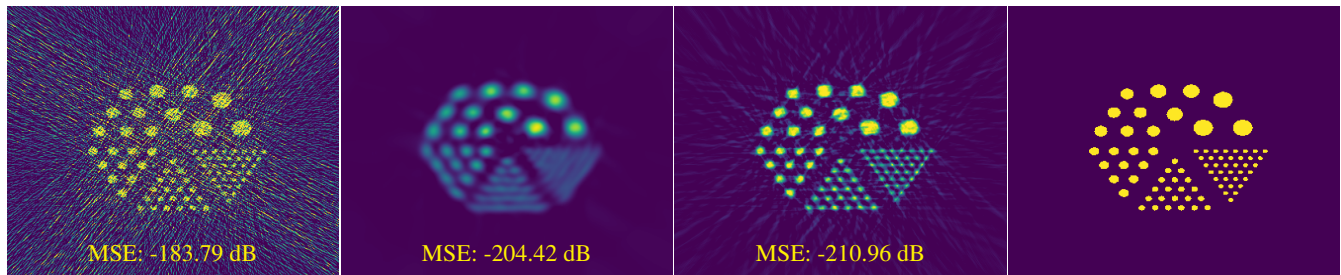**Fig. 4**. FBP reconstruction for the rebinned sinograms of Figure 3.

Fig. 5. FBP reconstruction of a simulated Derenzo phantom.

6. REFERENCES

[1] Seiichi Yamamoto et al., "Development of ultrahigh resolution Si-PM-based PET system using 0.32mm pixel scintillators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 836, pp. 7–12, Nov. 2016.

[2] Yongfeng Yang et al., "A prototype high-resolution small-animal PET scanner dedicated to mouse brain imaging," *Journal of Nuclear Medicine*, vol. 57, no. 7, pp. 1130–1135, July 2016.

[3] Youfang Lai et al., "H2RSPET: A 0.5 mm resolution high-sensitivity small-animal PET scanner, A simulation study," *Physics in Medicine and Biology*, vol. 66, no. 6, pp. 065016, Mar. 2021.

[4] Giuseppe Iacobucci et al., "Efficiency and time resolution of monolithic silicon pixel detectors in SiGe BiCMOS technology," *Journal of Instrumentation*, vol. 17, no. 02, pp. P02019, Feb. 2022.

[5] Michel Defrise et al., "Exact and approximate rebinning algorithms for 3-D PET data," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 145–158, Apr. 1997.

[6] Nobuhiro Tahara et al., "2-deoxy-2-[18F]fluoro-d-mannose positron emission tomography imaging in atherosclerosis," *Nature Medicine*, vol. 20, no. 2, pp. 215–219, Feb. 2014.

[7] Michel Defrise, Vladimir Panin, Christian Michel, and Michael E. Casey, "Continuous and discrete data rebinning in time-of-flight PET," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1310–1322, Sep. 2008.

[8] David W. Townsend, "Positron emission tomography/computed tomography," *Seminars in Nuclear Medicine*, vol. 38, no. 3, pp. 152–166, May 2008, Developments in Instrumentation.

[9] Michael Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[10] Pol del Aguila Pla, Aleix Boquet-Pujadas, and Joakim Jaldén, "Convex quantization preserves logconcavity," *IEEE Signal Processing Letters*, vol. 29, pp. 2697–2701, 2022.

[11] Charles Kooperberg and Charles J. Stone, "A study of logspline density estimation," *Computational Statistics & Data Analysis*, vol. 12, no. 3, pp. 327–347, Nov. 1991.

[12] Stamatios Lefkimmiatis, John Paul Ward, and Michael Unser, "Hessian Schatten-Norm Regularization for Linear Inverse Problems," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1873–1888, May 2013.

[13] Jingwei Liang, Tao Luo, and Carola-Bibiane Schönlieb, "Improving "Fast Iterative Shrinkage-Thresholding Algorithm": Faster, Smarter, and Greedier," *SIAM Journal on Scientific Computing*, vol. 44, no. 3, pp. A1069–A1091, June 2022.