SKETCHED RT3D: HOW TO RECONSTRUCT BILLIONS OF PHOTONS PER SECOND

*Julián Tachella, Michael P. Sheehan, Mike E. Davies**

School of Engineering
University of Edinburgh
Edinburgh, UK

ABSTRACT

Single-photon light detection and ranging (lidar) captures depth and intensity information of a 3D scene. Reconstructing a scene from observed photons is a challenging task due to spurious detections associated with background illumination sources. To tackle this problem, there is a plethora of 3D reconstruction algorithms which exploit spatial regularity of natural scenes to provide stable reconstructions. However, most existing algorithms have computational and memory complexity proportional to the number of recorded photons. This complexity hinders their real-time deployment on modern lidar arrays which acquire billions of photons per second. Leveraging a recent lidar sketching framework, we show that it is possible to modify existing reconstruction algorithms such that they only require a small sketch of the photon information. In particular, we propose a sketched version of a recent state-of-the-art algorithm which uses point cloud denoisers to provide spatially regularized reconstructions. A series of experiments performed on real lidar datasets demonstrates a significant reduction of execution time and memory requirements, while achieving the same reconstruction performance than in the full data case.

***Index Terms*—** Inverse problems, single-photon lidar, 3D reconstruction, compressive learning

1. INTRODUCTION

Single-photon lidar technology enables multiple important applications, ranging from autonomous driving [1, 2] to tropical archaeology [3]. This sensing modality can provide long-range information [4] with millimetre precision [5] while using eye-safe laser power levels. Depth information is obtained by measuring the round-trip time-of-arrival (ToA) of laser pulses using a time-correlated single-photon counting system. Recovering 3D information from the photon detections can be very challenging in scenarios where the number of collected photons associated with the signal of interest is very low or in the presence of strong ambient illumination which generates spurious detections. Leveraging the spatial regularity of natural scenes, several algorithms [6–12] have been proposed to provide stable 3D reconstructions in these settings. Existing reconstruction algorithms require access to either the ToA of all the detected photons or a ToA histogram. Both memory requirements and computational complexity scale at least linearly with these parameters [13] (see fig. 1). This dependency hinders their use in modern lidar arrays, which record an ever-increasing number of photons per second [14].

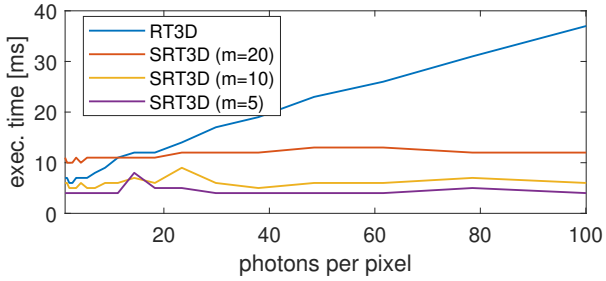

Fig. 1: Execution time of RT3D [13] and the proposed sketched SRT3D as a function of the mean number of photons per pixel. RT3D suffers from a linear complexity, whereas SRT3D only depends on the number of sketches m.

A simple solution to the computational and memory bottlenecks consists of aggregating the fine-resolution timing data into histograms with a small number of bins, at the cost of sacrificing depth resolution [15]. This approach results in a trade-off between compression and temporal resolution, leading to suboptimal reconstructions which do not make full use of the high resolution potential of the lidar device. In contrast, a novel sketching based approach was recently proposed in [16] as a solution to the data transfer bottleneck that does not suffer from an inherent trade-off between compression and temporal resolution. A compact representation of the ToA data is computed using the detected photon timestamps in an online manner for each individual pixel. The size of this representation scales with the parameters of the ToA model (i.e. the positions and intensities of the objects)

*This work was supported by the ERC Advanced grant, project C-SENSE, (ERC-ADG-2015-694888). Mike E. Davies is also supported by a Royal Society Wolfson Research Merit Award. Code is available at `https://gitlab.com/Tachella/real-time-sp-lidar`.

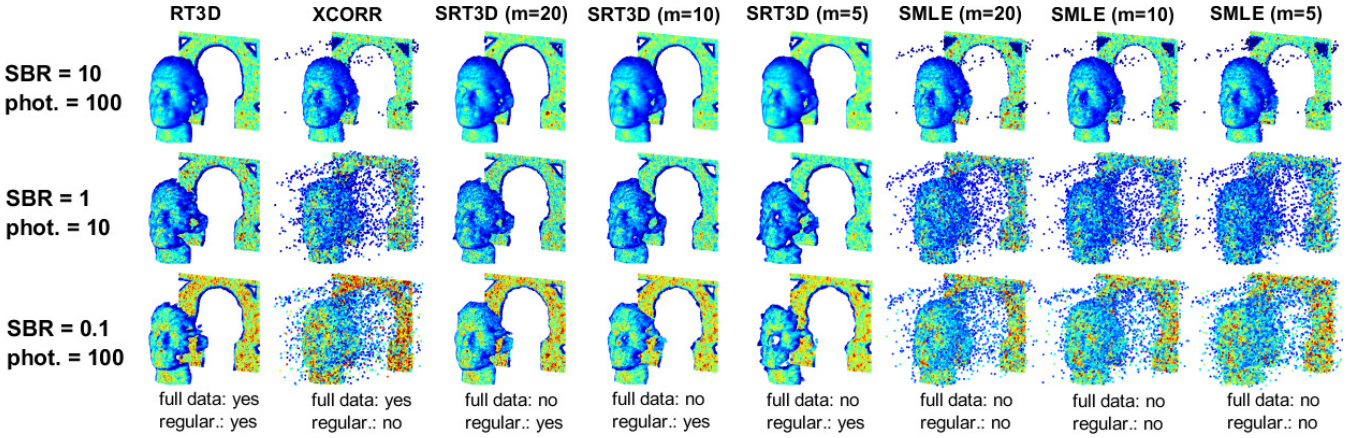

Fig. 2: 3D reconstructions obtained by the proposed sketched RT3D algorithm for different sketch sizes m and other competing algorithms. The proposed SRT3D method incorporates spatial regularization, providing stable reconstructions in settings with low SBR or low number of measured photons.

and is independent of both the temporal resolution and the number of photons. However, the reconstruction methods for sketched data presented in [16] do not incorporate any spatial regularization (pixels are processed independently), and thus can provide unstable reconstruction when the signal-to-background ratio (SBR) is low (see fig. 2).

In this paper, we extend the sketched lidar framework to spatially regularized reconstruction algorithms. In particular, we present a novel sketched-based real-time algorithm whose complexity is agnostic of the number of recorded photons. The proposed algorithm paves the way for real-time 3D reconstruction of complex scenes for any number of recorded photons. The main contributions of the paper are:

- We propose a novel approach to regularized 3D reconstruction with memory and computation requirements that are independent of the total number of observed photons.
- We evaluate the robustness of the sketched reconstruction algorithm using real lidar datasets, demonstrating that the proposed approach achieves the same reconstruction performance as in the full data case.

The paper is organised as follows. In Section 2, we introduce the sketched lidar framework. In Section 3, we present a novel reconstruction algorithm with spatial regularization which only relies on sketched data. In Section 4, we analyse the performance of the sketched algorithm on synthetic and real datasets. Conclusions are discussed in Section 5.

2. PIXELWISE SKETCHED LIDAR

An individual single-photon lidar pixel is associated with n time-stamps indicating the ToA of individual photons, which are denoted by x_p for $1 \leq p \leq n$. Assuming that there are K distinct reflecting surfaces in the field of view of the pixel, we let $\boldsymbol{t} = [t_1, \ldots, t_K]^\top$ be a vector containing the depths of the surfaces, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^\top$ where α_k is the probability that the detected photon from the kth surface. We denote by α_0 the probability that the detected photon is due to the background illumination. The ToA of the pth photon detected follows a mixture distribution [17]

$$\pi(x_p|\boldsymbol{\alpha}, \boldsymbol{t}) = \sum_{k=1}^{K} \alpha_k \pi_s(x_p|t_k) + \alpha_0 \pi_b(x_p), \qquad (1)$$

where $\sum_{k=0}^{K} \alpha_k = 1$. The distribution of the photons originating from the signal is defined by $\pi_s(x_p|t) = h(x_p - t)/H$, where the impulse response of the system and its associated integral are denoted by h and $H = \sum_{t=1}^{T} h(t)$, respectively. The distribution of photons originating from background sources is in general uniformly distributed, $\pi_b(x_p) = 1/T$ over the interval $[0, T-1]$ (more complex models can also be accommodated). The parameters in eq. (1) are summarized by the tuple $\theta = (\boldsymbol{t}, \boldsymbol{\alpha}, \alpha_0)$.

Timing information is generally represented either as a list of n ToA values [12] or aggregated into a ToA histogram with $T' \leq T$ temporal bins [15]. Modern lidar devices acquire hundreds or thousands of photons n in a short time frame, which hinders the use of ToA lists as the memory requirement scales linearly with the n. Histograms can alleviate the memory requirement by aggregating photon detections into coarse temporal bins, i.e.

$$y_\ell = \sum_{p=1}^{n} \mathbb{1}_{x_p \in [t_\ell, t_\ell + \Delta t]} \qquad (2)$$

for $\ell = 1, \ldots, T'$. However, this strategy sacrifices depth resolution, leading to suboptimal reconstructions which do not exploit the full potential of the device.

In order to alleviate the trade-off between memory requirement and depth resolution, [16] have recently introduced a novel representation of the timing information, whose size is independent of the number of acquired photons and does not incur in a significant loss of depth resolution. The compact representation, or so-called sketch, is computed as

$$z_\ell = \frac{1}{n}\sum_{p=1}^{n} e^{\mathrm{i}\omega_\ell x_p}, \tag{3}$$

for $\ell = 1, \dots, m$ where $\mathrm{i} = \sqrt{-1}$ and m is the number of sketches. As with the coarse histogram, the sketch has the favourable property that it can be updated in an online fashion with each incoming photon throughout the duration of the acquisition time. Thereafter, only the resultant sketch $\mathbf{z} = [z_1, \dots, z_\ell]^\top \in \mathbb{C}^m$ needs to be stored and/or transferred off-chip to further estimate the parameters θ of the observation model.

The sketch is equivalent to the empirical characteristic function sampled at frequencies ω_ℓ, where $\Psi_\theta(\omega) = \mathbb{E}_{x\sim\pi} e^{\mathrm{i}\omega x}$ is the corresponding expected characteristic function (CF) [18]. The CF has the special property that it exists for all probability distributions and captures all the information of the distribution, providing a one-to-one correspondence. The CF of the observation model in (1) is given by

$$\Psi_\theta(\omega) = \sum_{k=1}^{K} \alpha_k \hat{h}(\omega) e^{\mathrm{i}\omega t} + \alpha_0 \,\mathrm{sinc}(\omega T/2), \tag{4}$$

where $\hat{h}$ denotes the (discrete) Fourier transform of the impulse response function h. It is well documented in the empirical characteristic function (ECF) literature e.g. [19], that a sketch $\mathbf{z}$ computed over a finite dataset $\mathcal{X} = \{x_1, \dots, x_n\}$, satisfies the central limit theorem. Formally, the sketch $\mathbf{z}$ converges asymptotically to a Gaussian random variable

$$\mathbf{z} \xrightarrow{\text{dist}} \mathcal{N}\big(\Psi_\theta, n^{-1}\Sigma_\theta\big), \tag{5}$$

where $\Sigma_\theta \in \mathbb{C}^{m\times m}$ depends on $\Psi_\theta(\omega)$. Thus, the sketched lidar inference task reduces to solving the following optimization problem

$$\arg\min_\theta n\|\mathbf{z} - \Psi_\theta\|^2_{\mathbf{W}}, \tag{6}$$

where $\mathbf{W} \in \mathbb{C}^{m\times m}$ is a positive definite Hermitian weighting matrix, which can be chosen as the identity matrix for reduced computational load or the precision matrix $\mathbf{W} = \Sigma_\theta^{-1}$ for higher statistical efficiency.

The sampling scheme proposed in [16] chooses the frequencies $\omega_\ell = 2\pi\ell/T$ for $\ell \in [1, T-1]$, such that Ψ_θ in eq. (4) is only sampled at regions where $\mathrm{sinc}(\omega T/2) = 0$, and the resulting sketch is effectively blind to background noise. Fundamentally, the size of the sketch m scales with the degree of parameters in the observation model (i.e. the number of surfaces in the scene) and, crucially, is independent of both the depth resolution and the number of photons n. Sketching therefore enables significant compression of the ToA data without sacrificing temporal resolution or estimation accuracy. In the next section, we show how spatial regularization can also be incorporated to the sketching framework.

3. MULTIPIXEL SKETCHED LIDAR

Single-photon lidar devices acquire an array of $N_r \times N_c$ pixels. We encompass all the parameters in a scene into $\boldsymbol{\theta} = (\theta_{1,1}, \dots, \theta_{N_r,N_c})$. Due to the spatial regularity of natural scenes, parameters in neighbouring pixels are generally strongly correlated. This prior knowledge is exploited by several 3D reconstruction algorithms to improve the quality with respect to simple pixelwise depth estimation. Most algorithms solve the following optimization problem [2]

$$\arg\min_{\boldsymbol{\theta}} \sum_{i,j}^{N_r,N_c} f_{\boldsymbol{y}_{i,j}}(\theta_{i,j}) + \rho(\boldsymbol{\theta}) \tag{7}$$

where $\boldsymbol{y}_{i,j}$ denotes the observed histogram at the (i,j)th pixel, $f_{\boldsymbol{y}_{i,j}}(\theta_{i,j})$ are per-pixel data fidelity terms (negative log-likelihood of the ToA list or histogram observation models), and $\rho(\boldsymbol{\theta})$ is a spatial regularization term which encodes the prior information about the spatial regularity of typical scenes. There has been significant efforts dedicated to the design of powerful regularizations $\rho(\boldsymbol{\theta})$. The RT3D algorithm [13] exploits the plug-and-play framework [20] together with a fast computer-graphics point cloud denoiser to design a regularizer that can capture the geometry of complex scenes with $K > 1$ surfaces per pixel, while achieving real-time performance. However, existing algorithms (including RT3D) require multiple evaluations of the data fidelity terms, and thus suffer from large memory requirements and a computational complexity which is at least linear in the number of photon detections or histogram bins [13].

Here we propose to replace the histogram-based loss in eq. (7) for the more compact sketch loss in eq. (6), while leveraging the spatial regularization penalty of existing methods. The proposed objective can be expressed as

$$\arg\min_{\boldsymbol{\theta}} \sum_{i,j}^{N_r,N_c} n_{i,j}\|\mathbf{z}_{i,j} - \Psi_{\theta_{i,j}}\|^2_{\mathbf{W}_{i,j}} + \rho(\boldsymbol{\theta}) \tag{8}$$

where $\mathbf{z}_{i,j}$ is the sketch associated with the (i,j)th pixel. The number of detected photons $n_{i,j}$ controls the trade-off between the data-fidelity and regularization terms. As the number of detected photons increases, the data fidelity term dominates eq. (8), which tends to the non-regularized problem in eq. (6). In order to perform real-time reconstruction with an arbitrary number of photon detections, we propose a sketched version of the RT3D algorithm, which we name SRT3D. The proposed algorithm replaces the histogram-based likelihood of RT3D for the sketched loss of eq. (8) with $\mathbf{W}_{i,j}$ set as the identity matrix for reduced computational load.

4. EXPERIMENTS

We evaluate the proposed SRT3D algorithm on two real datasets: a polystyrene head at a distance of 40 metres [7] and a scene with two people walking behind a camouflage net at a distance of 320 metres [13]. We compare the proposed method with 3 other algorithms: traditional cross-correlation [5] (XCORR), pixelwise sketched reconstruction [16] (SMLE) which doesn't exploit spatial regularization (also with $\mathbf{W}$ set as the identity), and RT3D which accesses the full fine-resolution ToA data. All the experiments were performed using an NVIDIA RTX 3070 laptop GPU.

The polystyrene head dataset has size of 141×141 pixels with $T = 4613$. Most of the pixels in this scene contain exactly $K = 1$ surface. A ground-truth reference was obtained using using the standard cross-correlation algorithm on the raw ToA information (with high number of photons per pixel and high SBR). Using this reference and the observation model in eq. (1), we synthesized multiple datasets for different mean photons per pixel n and SBR levels. Figure 2 shows the 3D reconstructions obtained for SBR levels of 10, 1 and 0.1. All methods perform similarly when the number of photons and SBR are large. Interestingly, only $m = 5$ sketches are enough to provide good reconstructions. However, when the scene contains a low number of photons or low SBR, pixelwise methods fail to provide good reconstructions, whereas both RT3D and SRT3D provide good reconstructions. In this challenging setting, a sketch of size $m = 10$ sufficiently provides a reconstruction that has the same quality as the ones obtained in the full data case. True and false detections, depth absolute error (DAE), and normalised intensity absolute error (IAE) (as defined in [13]) are presented in fig. 3 for an SBR of 1 and different number of photons per pixel.

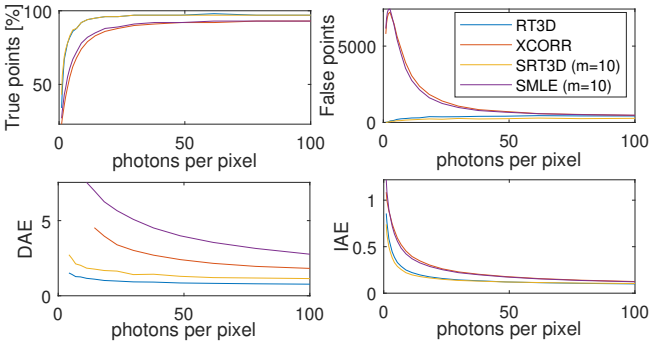

Fig. 3: Performance of the evaluated algorithms for the polystyrene head dataset with SBR=1.

Figure 1 shows the execution time of RT3D and SRT3D as a function of the mean number of photons per pixel in the polystyrene head datasets. The GPU memory requirements of RT3D become prohibitive if the number of observed photons is in the order of hundreds per pixel per frame, whereas the sketched version has a complexity which is agnostic of the number of photons, and can handle any number of photons in real-time. Table 1 shows the execution time of SRT3D for increasing array sizes, demonstrating that the proposed method can process up to 705×705 arrays at 14 frames per second on a laptop computer. The datasets were generated by upsampling the head reference before the synthesis of photon detections.

m/pixels	141^2	282^2	423^2	564^2	705^2
$m = 5$	6	12	28	55	68
$m = 10$	7	18	35	60	88

Table 1: Execution time in milliseconds for different scene sizes in pixels obtained by the proposed sketched RT3D algorithm for sketch sizes m of 5 and 10.

Figure 4 shows the reconstructions by SRT3D and RT3D of one time frame of the camouflage dataset in [13], which is composed of 32×32 pixels with $T = 153$. Most of the pixels in the scene contain $K = 2$ surfaces, which makes the reconstruction task more challenging. However, $m = 10$ sketches are again sufficient to provide the same reconstruction quality as using the full 153 bins. Although the original fine resolution T is not large, the execution time of SRT3D with $m = 10$ was 12 ms, whereas for RT3D it was 20 ms.

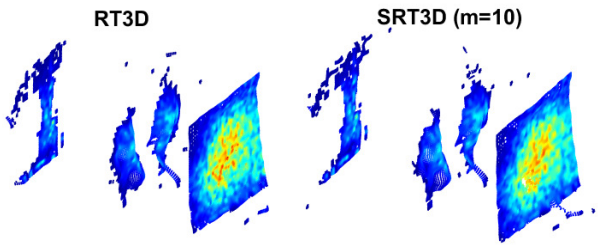

Fig. 4: Reconstruction of the scene in [13] with 2 surfaces per pixel by the original RT3D and its sketched version. Using only $m = 10$ sketches is enough to provide the same reconstruction quality.

5. CONCLUSIONS AND FUTURE WORK

We have extended the sketched lidar framework to incorporate spatial regularization, demonstrating stable reconstructions in the presence of strong ambient illumination. While our results focus on a sketched version of the RT3D algorithm, the ideas presented here can be used to develop sketched versions of other existing regularized methods.

Acknowledgements

We would like to thank the single-photon group of Heriot-Watt University for the head and camouflage data.

6. REFERENCES

[1] J. Hecht, "Lidar for self-driving cars," *Optics and Photonics News*, vol. 29, no. 1, pp. 26–33, 2018.

[2] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 62–71, 2020.

[3] M. A. Canuto, F. Estrada-Belli, T. G. Garrison, S. D. Houston, M. J. Acuña, M. Kováč, D. Marken, P. Nondédéo, L. Auld-Thomas, C. Castanet, D. Chatelain, C. R. Chiriboga, T. Drápela, T. Lieskovský, A. Tokovinine, A. Velasquez, J. C. Fernández-Díaz, and R. Shrestha, "Ancient lowland Maya complexity as revealed by airborne laser scanning of northern Guatemala," *Science*, vol. 361, no. 6409, 2018.

[4] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Opt. Express*, vol. 25, no. 10, pp. 11 919–11 931, May 2017.

[5] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Appl. Opt.*, vol. 48, no. 32, pp. 6241–6251, Nov 2009.

[6] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3-D and reflectivity imaging with single-photon detectors," *IEEE Trans. Comput. Imaging*, vol. 1, no. 2, pp. 112–125, 2015.

[7] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1935–1946, 2016.

[8] ——, "Target detection for depth imaging using sparse single-photon data," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3256–3260.

[9] D. Shin, F. Xu, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "Computational multi-depth single-photon imaging," *Optics express*, vol. 24, no. 3, pp. 1873–1888, 2016.

[10] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imaging*, vol. 3, no. 3, pp. 445–459, Sept 2017.

[11] D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-Photon 3D Imaging with Deep Sensor Fusion," *ACM Trans. Graph. (SIGGRAPH)*, no. 4, 2018.

[12] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Bayesian 3D reconstruction of complex scenes from single-photon lidar data," *SIAM Journal on Imaging Sciences*, 2019.

[13] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J. Tourneret, and S. McLaughlin, "Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," *Nature communications*, vol. 10, no. 1, pp. 1–6, 2019.

[14] S. Royo and M. Ballesta-Garcia, "An overview of lidar imaging systems for autonomous vehicles," *Applied Sciences*, vol. 9, no. 19, 2019.

[15] R. K. Henderson, N. Johnston, H. Chen, D. D. Li, G. Hungerford, R. Hirsch, D. McLoskey, P. Yip, and D. J. S. Birch, "A 192×128 time correlated single photon counting imager in 40nm CMOS technology," in *ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC)*, 2018, pp. 54–57.

[16] M. Sheehan, J. Tachella, and M. E. Davies, "A sketching framework for reduced data transfer in photon counting lidar," *IEEE Trans. on Comp. Imag. (Early Access)*, 2021.

[17] Y. Altmann and S. McLaughlin, "Range estimation from single-photon lidar data using a stochastic em approach," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1112–1116.

[18] M. Carrasco and J. P. Florens, "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, pp. 797–834, 2000.

[19] A. Hall, *Generalized Method of Moments*. Oxford University Press, 11 2007, pp. 230 – 255.

[20] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 945–948.