Sketched over-parametrized projected gradient descent for sparse spike estimation

Pierre-Jean Bénard, Yann Traonmilin, Jean-François Aujol

Sketched over-parametrized projected gradient descent for sparse spike estimation

Pierre-Jean Bénard[1,*], Yann Traonmilin[1], Jean-François Aujol[1]

***Abstract*—We consider the problem of recovering off-the-grid spikes from linear measurements in the context of Single Molecule Localization Microscopy (SMLM). State of the art model-based methods such as Over-Parametrized Continuous Orthogonal Matching Pursuit (OP-COMP) with Projected Gradient Descent (PGD) have been shown to successfully recover those signals. The computational cost of these methods scales linearly with the number of measurements. When this number of measurements is large with respect to the dimensionality of the signal, we propose to reduce it with a so-called sketching operator. Based on recent results on compressive sensing in the space of measures, we approximate the ideal sketching operator (benefiting from theoretical recovery guarantees), in the context of SMLM. This sketching method coupled to OP-COMP with PGD shows significant improvements in calculation time in realistic synthetic microscopy experiments.**

***Index Terms*—spike super-resolution, sketching, non-convex optimization**

I. Introduction

The off-the-grid super-resolution problem consists in recovering spikes from linear measurements in a continuous setting. Among its many applications, we focus in this article on Single Molecule Localization Microscopy (SMLM). Let x_0 be an off-the-grid sparse signal (modeling the position and intensity of illuminated molecules) over $\mathbb{R}^d$. Such signals can be modeled as a sum of K Dirac measures:

$$x_0 = \sum_{i=1}^{K} a_i \delta_{t_i} \tag{1}$$

where $a = (a_1, \dots, a_K) \in \mathbb{R}^K$ are the amplitudes and $t = (t_1, \dots, t_K) \in \mathbb{R}^{K \times d}$ are the positions of the spikes. We observe the signal through a linear operator A from the space $\mathcal{M}$ of finite signed measures over $\mathbb{R}^d$ to $\mathbb{C}^m$ with m the number of measurements. The result of this observation is $y = Ax_0 \in \mathbb{C}^m$. Note that we consider the noiseless case in this article for the sake of clarity.

A way to recover the true signal x_0 from its observation y is to find a minimizer of a non-convex least-square problem:

$$x^* \in \underset{x \in \Sigma_{K,\epsilon}}{\arg\min} \|Ax - y\|_2^2 \tag{2}$$

where $\Sigma_{K,\epsilon}$, defined by (4), is a set modeling a separation constraint between spikes. In practice, it is possible to consider the minimization problem parametrized by amplitude and positions θ:

$$\theta^* \in \underset{\theta \in \Theta_{K,\epsilon}}{\arg\min}\, g(\theta) \quad \text{with} \quad g(\theta) := \|A\phi(\theta) - y\|_2^2. \tag{3}$$

where $\phi(\theta) = \sum_{i=1}^{K} b_i \delta_{s_i}$. The separation between spikes is modeled as the following constraint set

$$\Theta_{K,\epsilon} = \Big\{ \theta = (b_1, \dots, b_K, s_1, \dots, s_K) \in \mathbb{R}^{K(d+1)}, \\ \forall i, j \in \{1, \dots, K\}, i \neq j, \|s_i - s_j\|_2 > \epsilon \Big\}. \tag{4}$$

The corresponding low dimensional model in the space of measures is simply $\Sigma_{K,\epsilon} = \phi(\Theta_{K,\epsilon})$.

Theoretical guarantees for the recovery of x_0 with (2) have been given in [1], [2]. As an example, when the linear operator A models random Fourier measurements at frequencies drawn with a Gaussian distribution and $\dim(y) \gtrapprox \mathcal{O}(K^2 d \operatorname{polylog}(K, d))$, we have $x^* = x_0$. In practice, state-of-the-art Sliding Continuous Orthogonal Matching Pursuit (SCOMP) [3] and Over-Parametrized COMP with Projected Gradient Descent (OP-COMP + PGD) [4] have been successful at minimizing (2). Recovery guarantees are given in [5], [6], [7]. We can also cite the Sliding Frank-Wolfe (SFW) algorithm [8] (similar in practice to SCOMP), that solves a regularized version of (2).

With these methods, the computational cost depends on the number of measurements m. In particular, this poses a problem in the context of SMLM where, in some modalities, the measurements are a collection of sampled microscope images at different angles. The number of measurements is then much larger than the number of parameters that we need to estimate. An idea is to compress the measurements into a "sketch" in order to accelerate the computation time while still recovering the ground truth signal with a good accuracy.

Originated from compressed sensing, sketching methods permit fast operations on large amounts of data. Recently, Gribonval et al. proposed a sketching framework as a linear inverse problem in the space of measures [1], [2], [9]. It has been successfully applied to estimate sums of Dirac measures [10], Gaussian mixture parameters [3], [11], [12], [13], parameters of neural networks defining the underlying density of a database [14], [12] or source location [15], [16].

We propose to adapt the sketching framework to a SMLM context where the measurement operator in the space of measure is already designed.

Contributions: In this paper, our main contribution is an accelerated method for the recovery of sparse signals with Over-Parametrized Continuous Orthogonal Matching Pursuit

*Contact: pierre-jean.benard@math.u-bordeaux.fr, [1]CNRS, Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400 Talence, France.
This work was supported by the French National Research Agency (ANR) under reference ANR-20-CE40-0001 (EFFIREG project). Experiments presented in this paper were carried out using the PlaFRIM experimental testbed (Inria, CNRS , Univ. Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine).

with Projected Gradient Descent and applicable to similar algorithms like Sliding Continuous Orthogonal Matching Pursuit and Sliding Frank-Wolfe. We propose to compress the measurements y by approximating an ideal sketching operator in the space of measures, leading to an effective reduction of m. This method greatly improves the global computation times OP-COMP with PGD. Indeed, we provide synthetic experiments in the context of microscopy that show an improvement up to 80% in calculation time compared to non-sketched linear operator with no loss in accuracy.

II. Sketched super-resolution for SMLM

Measurements are defined by a collection of functions $(\alpha_l)_{l=1}^m$ representing the response of the l-th sensor. We have that

$$Ax_0 = A\sum_{i=1}^{K} a_i\delta_{t_i} = \sum_{i=1}^{K} a_i \int_{\mathbb{R}^d} \alpha_l \mathrm{d}\delta_{t_i}. \tag{5}$$

The integral in (5) is the duality product $\langle \delta_{t_i}, \alpha_l\rangle$ between a finite signed measure over $\mathbb{R}^d$ and a function $\alpha_l : \mathbb{R}^d \mapsto \mathbb{C}$. One can cite random or uniform Fourier sampling ($\alpha_l(t) = \mathrm{e}^{-j\langle \omega_l, t\rangle}$) as an example of a linear operator.

In the field of microscopy, the measurements y are images captured by a microscope. Their associated linear operators model the Point Spread Function (PSF) of the microscope. For the recovery of the amplitudes and positions of cells in 3D, we model the operator A with the functions:

$$\alpha_{i,k}(t) = \frac{f_k(z)}{2\pi\sigma_x\sigma_y} \times \mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)} \tag{6}$$

with $t = (x, y, z) \in \mathbb{R}^3$, $\sigma_x, \sigma_y \in \mathbb{R}^+$. The $(x_i, y_i) \in \Gamma \subset \mathbb{R}^2$ are the position of the i-th pixel on a finite grid Γ at a given incidence angle k, i.e., we consider a Gaussian approximation of the impulse response of the microscope. The functions $f_k : \mathbb{R} \mapsto \mathbb{R}$ take the depth (the z component) as input and model the impact of the incidence angle. One can cite the multi-angle total internal reflection fluorescence (MA-TIRF) operator [8] which can be approximated by (6).

Sketching Gaussian measurements in the space of measures: The goal of the sketching operator is to compress measurements while keeping the relevant information from the original observation y. Thanks to the framework of Gribonval et al. [1], it can be defined on the infinite dimensional space of measures $\mathcal{M}$. Consider an ideal 2D Gaussian measurement process A^{id} defined by

$$\alpha_i^{id}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y}\mathrm{e}^{-\left(\frac{(x_i-x)^2}{2\sigma_x^2}+\frac{(y_i-y)^2}{2\sigma_y^2}\right)} \tag{7}$$

for any $(x_i, y_i) \in \mathbb{R}^2$, i.e., we have access to the full convolution of Dirac measures with a Gaussian impulse response. We can consider a *sketching* operator $\mathcal{S} : L^2(\mathbb{R}^2) \to \mathbb{C}^{m_\mathcal{S}}$ from the space of square integrable functions to the space $\mathbb{C}^{m_\mathcal{S}}$ of reduced dimension $m_\mathcal{S}$. We call $\mathcal{S}A^{id}$ the corresponding *sketched* ideal operator. One way to sketch the linear operator A^{id} is to take random Fourier measurements with a finite number $m_\mathcal{S}$ of well-chosen random frequencies $\omega_l \in \mathbb{R}^d$. This operation, for a given frequency ω_l and any $x = \sum_{r=1}^K b_r\delta_{s_r}$ is written as

$$(\mathcal{S}(A^{id}x))_l = [\mathcal{F}(A^{id}x)](w_l) = \sum_{r=1}^{K} b_r[\mathcal{F}(A^{id}\delta_{s_r})](w_l) \tag{8}$$

where $\mathcal{F}$ is the continuous Fourier transform, i.e.

$$[\mathcal{F}(A^{id}\delta_{s_r})](w_l) = \int_{(x_i,y_i)\in\mathbb{R}^2} \alpha_i(s_r)e^{-j\langle\omega_l,(x_i,y_i)\rangle}dx_idy_i. \tag{9}$$

It was shown in [2] that it is possible to recover positions and amplitudes of sufficiently separated Gaussians with a choice of random Gaussian frequencies with $m_\mathcal{S} \gtrapprox \mathcal{O}(K^2 d\,\mathrm{polylog}(K, d))$ frequencies by minimizing, the sketched data-fit

$$g_\mathcal{S}(\theta) := \|\mathcal{S}A^{id}\phi(\theta) - \mathcal{S}A^{id}x_0\|_2^2 \tag{10}$$

under a separation constraint. Unfortunately, we do not have access to ideal Gaussian measurements A^{id} in practice: they are sampled on a grid. However, in SMLM, this grid is generally chosen to satisfy the Shannon-Nyquist limits of the instruments, i.e., there is little loss of information by the discretization process.

While we can calculate for any estimate $x = \phi(\theta)$ the sketched ideal measurements $\mathcal{S}A^{id}\phi(\theta)$, we cannot calculate exactly $\mathcal{S}A^{id}x_0$ as only Ax_0 is available. Instead, we approximate $\mathcal{S}A^{id}x_0$ with a discretized version called $\mathcal{S}_d y = \mathcal{S}_d A x_0$ directly obtained from the measurements y. We propose to minimize the approximated functional

$$g_{\mathcal{S}_d}(\theta) := \|\mathcal{S}A^{id}\phi(\theta) - \mathcal{S}_d y\|_2^2. \tag{11}$$

The approximated operator $\mathcal{S}_d$ performs the *truncated discrete time Fourier Transform* of y on the grid Γ:

$$(\mathcal{S}_d y)_l = \frac{1}{|\Gamma|}\sum_{i=1}^{|\Gamma|} y_i \mathrm{e}^{-j\langle\omega_l,(x_i,y_i)\rangle} \tag{12}$$

with (x_i, y_i) the position of the i-th cell on the grid Γ, y_i its associated measurement and $|\Gamma|$ the number of cells in Γ. Note that to compute $\mathcal{S}_d y$, we iterate over the grid Γ which we wanted to avoid originally. However, this calculation of $\mathcal{S}_d y$ needs to be done only once. Thus, this has a low impact on the whole computation time when compared to using the classical linear operator A within the optimization process.

The introduction of an approximation $\mathcal{S}_d$ of the sketching operator $\mathcal{S}$ brings the issue that $\arg\min_{\theta\in\Theta} g_{\mathcal{S}_d} \neq \arg\min_{\theta\in\Theta} g_\mathcal{S}$. Thanks to the following Lemma, we guarantee that the ideal sketched datafit (where $A^{id}x_0$ would be available) is approximately minimized by the solution of the approximated sketched datafit functional.

Lemma 1. *Let T_Γ the linear sampling operator on a grid Γ. Let A^{id} an ideal measurement operator and $A = T_\Gamma A^{id}$ its discretization on a grid Γ. Let $y = A\phi(\theta_0)$ an observation of a signal $\phi(\theta_0)$ on Γ. Let $\mathcal{S}$ and $\mathcal{S}_d$ be a sketching operator and its discretized approximation on the grid Γ.*

Suppose that

$$\|\mathcal{S} - \mathcal{S}_d T_\Gamma\|_{\mathrm{op}}^{A^{id},\Sigma} := \sup_{x\in\Sigma} \frac{\|\mathcal{S}A^{id}x - \mathcal{S}_d T_\Gamma A^{id}x\|_2}{\|A^{id}x\|_2} \leq \eta \tag{13}$$

and let $\theta^* \in \arg\min_{\theta\in\Theta} g_{\mathcal{S}_d}(\theta)$. *Then we have that, for all* $\theta \in \Theta$,

$$\sqrt{g_{\mathcal{S}}(\theta^*)} \leq \sqrt{\min_{\theta\in\Theta} g_{\mathcal{S}}(\theta)} + 2\eta\|y\|_2. \tag{14}$$

Proof. The proof of this Lemma can be found in the supplementary material. □

The main hypothesis in Lemma 1 uses the operator norm to quantify the difference between the sketching operator $\mathcal{S}$ and its approximation $\mathcal{S}_d$. The bound η exists if the discretization respects the Shannon-Nyquist condition which can be approximately true with small η for Gaussian observations (up to the truncation error), and when $\mathcal{S}A^{id}$ and A^{id} guarantee recovery of x_0 (e.g., under a restricted isometry property, see [1]).

Application to SMLM: As given by the expression of measurement functionals in (6), several 2D images representing different incidence angles are captured. We propose to sketch independently each angle. We give an explicit expression of the sketching operation in 2D, angle by angle with the following Lemma. Let us call A_k^{id} the ideal Gaussian measurements corresponding to incidence angle k. We just need the continuous Fourier transform of $A_k^{id}x_0$ to implement the sketching in practice.

Lemma 2. *Let* $\alpha_{i,k}$ *defined in* (6), *with* $t = (x, y, z) \in \mathbb{R}^3$, $\sigma = (\sigma_x, \sigma_y) \in \mathbb{R}^2$. *Then the Fourier transform of ideal Gaussian measurements is*

$$[\mathcal{F}(A_k^{id}\delta_t)](w_l) = f_k(z)\mathrm{e}^{-\frac{\langle\omega_l^2,\sigma^2\rangle}{2} - j\langle\omega_l,(x,y)\rangle} \tag{15}$$

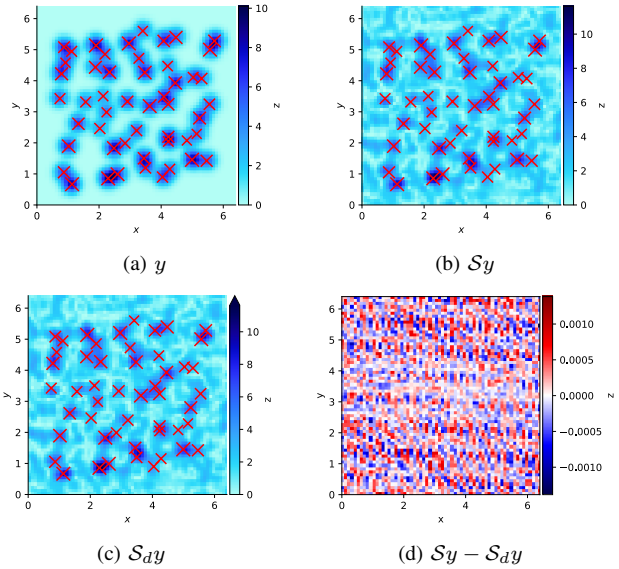
with $\omega = (\omega_x, \omega_y) \in \mathbb{R}^2$.

Proof. This Lemma is a direct consequence of the fact that the Fourier transform of a Gaussian is a Gaussian (proof in supplementary material). □

With Lemma 2 and with (12), we have all the tools to perform recoveries of a signal from sketched observations (in particular gradients of the sketched functional are easily calculated, see supplementary material).

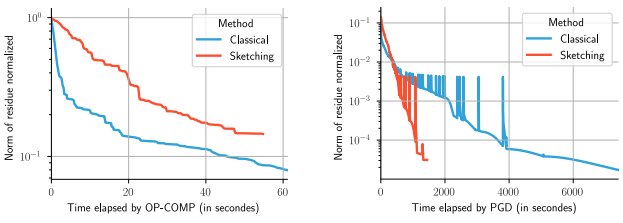
III. Experiments

In this section, we apply our sketching method to the state-of-the art OP-COMP + PGD method [7]. In a first phase of this algorithm, spikes are greedily added based on their correlation with the measurements to reach an over-parametrized initialization (OP-COMP without sliding). From the over-parametrized initialization, a projected descent on all parameters is performed where the projection projects on the separation constraint (i.e. close spikes are merged). We compare the original algorithm to the sketched version where the measurement operator (resp. the measurements) is replaced with the sketched measurement operator (resp. the sketched measurements), i.e., we study the impact of the reduction from m measurements to $m_{\mathcal{S}} < m$ measurements. We recover 50-spikes in 3D. The linear operators A used for these experiments follow the Gaussian approximation of the MA-TIRF model. The input positions are in a cuboid $([0, 6.4] \times [0, 6.4] \times [0, 0.8]$ µm) and amplitudes uniformly are distributed between 1 and 2. The output (the measurement vector y) consists of 4 different 2D grids composed of 64×64 pixels for a total of $m = 4 \times 64 \times 64 = 16\,384$ measurements. Since we perform the sketching of A separately on each 2D grid and for better readability, we note $m_{\mathcal{S}} = 4 \times K \times m_{\text{spike}}$ with $K = 50$ in our experiments and m_{spike} the number of measurements per spike. We first study an example of signal recovery with A and $\mathcal{S}A^{id}$ with $m_{\text{spike}} = 12$. Then, we test $\mathcal{S}A^{id}$ with various values of $m_{\text{spike}} \in \{3, 6, \ldots, 45, 48\}$. Moreover, since the frequencies from the sketched operator are drawn randomly, we repeat the process for each value of m_{spike} 16 times with different frequencies and signals x_0. The code for these experiments is available for download at [17].



(a) y (b) $\mathcal{S}y$ (c) $\mathcal{S}_d y$ (d) $\mathcal{S}y - \mathcal{S}_d y$

Figure 1: Ground truth spikes of x_0 (in red) on one incidence angle on a 2D grid obtained from (a) the observation $y = Ax_0$, (b) the back-projection of ideal sketched observation $\mathcal{S}y = \mathcal{S}A^{id}x_0$, (c) the back-projection of the approximation of the sketched observation $\mathcal{S}_d y = \mathcal{S}_d A x_0$, (d) the difference $\mathcal{S}y - \mathcal{S}_d y$. The small difference indicates that $\mathcal{S}_d y$ is a good approximation of $\mathcal{S}y$.



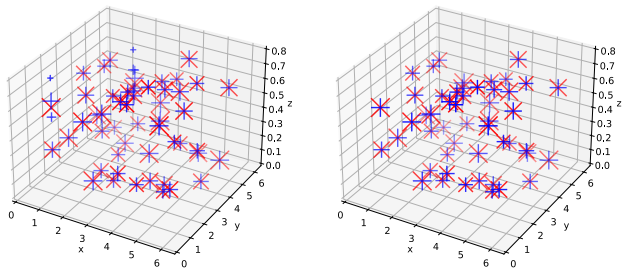
(a) (b)

Figure 2: (a) Norm of the residue r through the iterations of OP-COMP with both A (blue) and $\mathcal{S}A^{id}$ (red) with respect to time elapsed, (b) Norm of the residue r through the iterations of PGD with both A (blue) and $\mathcal{S}A^{id}$ (red) with respect to time elapsed. The total computation time with sketching $\approx 80\%$ shorter than without.

Details with $m_{\text{spike}} = 12$**:** First, we compare, in Figure 1, the

original measurements with the back-projection of the ideal and real sketched measurements on the grid Γ. We note that the difference between ideal and real sketched measurements is very small. In this example, we have $\|y\|_2 \approx 200$ and $\|\mathcal{S}y - \mathcal{S}_d y\|_2 \approx 6 \times 10^{-3}$. This would correspond to $\eta \approx 3 \times 10^{-5}$ in Lemma 1. Hence, this example (and similarly the following experiments) verifies Lemma 1 qualitatively.

We compare in Figure 2 and Figure 3 the result of sketched and non-sketched OP-COMP + PGD. With $m_{\text{spike}} = 12$ measurements per spikes representing approximately 15% of the total number of measurements ($m = 16\,384$) obtained with A, the sketched OP-COMP + PGD converges with approximately $38\,000$ iterations in 26 minutes compared to $75\,000$ iteration in 124 minutes for the non-sketched method. This is a gain of almost 80% in computation time, with no loss in performance of the estimation (see next experiment).



(a) x_0 and x^* (b) x_0 and x^* (sketched)

Figure 3: Ground truth (red) and estimation (blue) of the signal with OP-COMP + PGD. (a) Original OP-COMP + PGD, (b) Sketched OP-COMP + PGD. Sketched OP-COMP + PGD recovers spikes with the same precision as OP-COMP + PGD.

Performance of estimation with $m_{\text{spike}} \in \{3, 6, \dots, 45, 48\}$: In Figure 4, we compare recovery performance with the following metric: We denote the RMSE in each spatial dimension to be $RMSE_i = \sqrt{\frac{1}{|TP|}\sum_{j \in TP}\left([t_j]_i - [t_{0,j}]_i\right)^2}$ with i the dimension along which the RMSE is computed, t our estimated positions, t_0 our ground truth positions and TP the set of $|TP|$ paired ground truth spikes and estimated spikes. We construct this set TP by pairing spikes if they are close enough from each other. The chosen radius is 0.02μm (which is smaller than $\epsilon/2$, with ϵ the separation between spikes of the ground truth).

We observe in Figure 4 that the median RMSE along each axis decreases as the number of sketched measurements $m_{\mathcal{S}}$ increases. Starting from approximately 15%, we notice that the median RMSE obtained with our sketched operator $\mathcal{S}A^{id}$ are below the RMSE obtained with the classic operator A (note that the performance should be slightly degraded, but sketching might introduce a smoothing operation that benefits the optimization process).

When comparing the computation time for each value of $m_{\mathcal{S}}$ in Figure 5, we note that the time elapsed for the recovery of x_0 with sketching is linear with respect to $m_{\mathcal{S}}$. With $m_{\mathcal{S}}/m \approx 15\%$, we have an acceleration over the non-sketched method up to 80%. Above $m_{\mathcal{S}}/m \approx 60\%$, its is not worth to use the sketched linear operator compared to the original method (due to additional calculations for sketching).

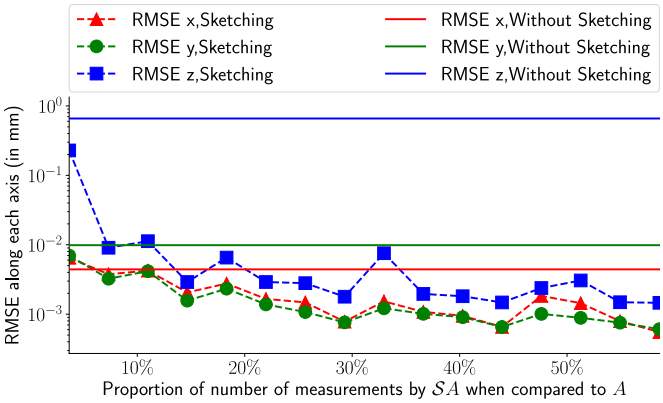


Figure 4: Median of the RMSE along each dimension $\{x, y, z\}$ obtained with OP-COMP with PGD on 50 spikes signals with respect to the compression rate of the sketching. The RMSE along each axis of estimated signals with the sketched method is decreasing when m_{spike} increasing and reaches original performances with 15% compression rate.

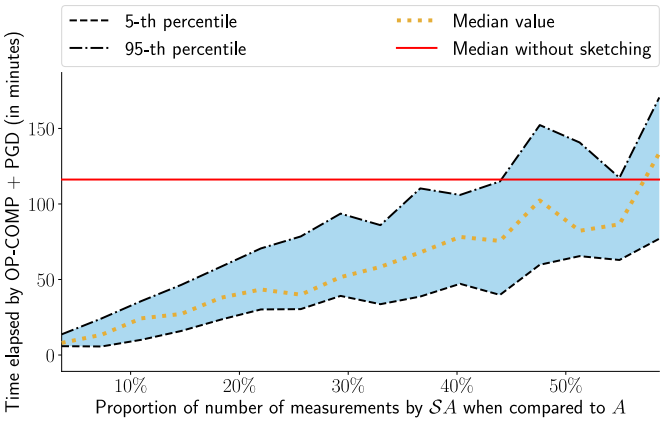


Figure 5: Computation time of OP-COMP + PGD on 50 spikes signals in function of the compression rate. In blue, the span of the results over experiments with its median (yellow), 5-th and 95-th percentile. In red, the median of the results without sketching. The computation time is proportional to $m_{\mathcal{S}}$. With $m_{\mathcal{S}}/m \lesssim 60\%$, the sketched method is faster the original.

IV. Discussion/Conclusion

We showed that using a sketched linear operator instead of a classical linear operator when the number of measurements is large can bring significant performance improvement without losing precision on the recovered positions and amplitudes of the spikes. Our results are supported by strong theoretical background and experiments. We also highlighted that the trade-off between computation time and accuracy can be explicitly chosen. We proposed to sketch independently 2D measurements. A possible future work would be to investigate if a full 3D sketching method would improve our results. Also, it would be interesting to investigate the surprising slight improvement in performance that we observed with the sketched method.

REFERENCES

[1] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin, "Compressive Statistical Learning with Random Feature Moments," *Mathematical Statistics and Learning*, vol. 3, no. 2, pp. 113–164, 2021.

[2] ——, "Statistical learning guarantees for compressive clustering and compressive mixture modeling," *Mathematical Statistics and Learning*, vol. 3, no. 2, pp. 165–257, 2021.

[3] N. Keriven, N. Tremblay, Y. Traonmilin, and R. Gribonval, "Compressive k-means," in *ICASSP*. IEEE, 2017, pp. 6369–6373.

[4] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol, "Fast off-the-grid sparse recovery with over-parametrized projected gradient descent," in *2022 30th EUSIPCO*. IEEE, 2022, pp. 2206–2210.

[5] C. Elvira, R. Gribonval, C. Soussen, and C. Herzet, "OMP and continuous dictionaries: Is k-step recovery possible?" in *ICASSP*. IEEE, 2019, pp. 5546–5550.

[6] ——, "When does OMP achieve exact recovery with continuous dictionaries?" *Applied and Computational Harmonic Analysis*, vol. 51, p. 39, 2021.

[7] P.-J. Bénard, Y. Traonmilin, J.-F. Aujol, and E. Soubies, "Estimation of off-the grid sparse spikes with over-parametrized projected gradient descent: theory and application," *Inverse Problems*, vol. 40, no. 5, p. 055010, 2024.

[8] Q. Denoyelle, V. Duval, G. Peyré, and E. Soubies, "The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy," *Inverse Problems*, 2019.

[9] R. Gribonval, A. Chatalic, N. Keriven, V. Schellekens, L. Jacques, and P. Schniter, "Sketching Data Sets for Large-Scale Learning: Keeping only what you need," *IEEE Signal Processing Magazine*, vol. 38, no. 5, pp. 12–36, Sep. 2021, for an extended version of this article that contains additional references and more in-depth discussions on a variety of topics, see the long version published on HAL (https://hal.inria.fr/hal-02909766) and arXiv (2008.01839). [Online]. Available: https://inria.hal.science/hal-03350599

[10] N. Keriven, R. Gribonval, G. Blanchard *et al.*, "Spikes super-resolution with random fourier sampling," in *SPARS workshop*, 2017.

[11] H. Shi, Y. Traonmilin, and J.-F. Aujol, "Compressive learning for patch-based image denoising," *SIAM Journal on Imaging Sciences*, vol. 15, no. 3, pp. 1184–1212, 2022.

[12] ——, "Compressive learning of deep regularization for denoising," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2023, pp. 162–174.

[13] V. Schellekens and L. Jacques, "Compressive classification (machine learning without learning)," in *ITWIST'18 workshop*, 2018.

[14] ——, "Compressive learning of generative networks," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, vol. 28, 2020.

[15] M. Fontaine, C. Vanwynsberghe, A. Liutkus, and R. Badeau, "Scalable source localization with multichannel α-stable distributions," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 11–15.

[16] ——, "Sketching for nearfield acoustic imaging of heavy-tailed sources," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2017, pp. 80–88.

[17] P.-J. Bénard, Y. Traonmilin, and J.-F. Aujol, "Code of the experiments," https://github.com/pjbenard/opCOMP_PGD_sketching, 2024, [Online].