

ROP inception: signal estimation with quadratic random sketching

Rémi Delogne, Vincent Schellekens, and Laurent Jacques*

ISPGROUP, INMA, ICTEAM, UCLouvain, Belgium.

17th May 2022

Abstract

Rank-one projections (ROP) of matrices and quadratic random sketching of signals support several data processing and machine learning methods, as well as recent imaging applications, such as phase retrieval or optical processing units. In this paper, we demonstrate how signal estimation can be operated directly through such quadratic sketches—equivalent to the ROPs of the “lifted signal” obtained as its outer product with itself—without explicitly reconstructing that signal. Our analysis relies on showing that, up to a minor debiasing trick, the ROP measurement operator satisfies a generalised sign product embedding (SPE) property. In a nutshell, the SPE shows that the scalar product of a signal sketch with the *sign* of the sketch of a given pattern approximates the square of the projection of that signal on this pattern. This thus amounts to an insertion (an *inception*) of a ROP model inside a ROP sketch. The effectiveness of our approach is evaluated in several synthetic experiments.

1 Introduction

More and more algorithms in signal processing, optimisation, matrix algebra and machine learning rely on *random projections*. They are used to relax the computational burden of ever-growing data flows with data-agnostic dimensionality reduction procedures, such as linear random projections for data embedding or compressive sensing [2], while preserving specific information. Random projections are also used (in combination with nonlinear maps) to unfold (or embed) a dataset in larger dimensional feature spaces [10]. For example, this expanded domain can be more amenable to data separability or to clustering than the initial space, or endowed with specific kernels induced by the projection [14].

The *estimation* of specific properties (or functions) of signals is crucial to many data processing techniques, hence the ability to perform it in the feature (“projected”) domain is of strong interest. For instance, one may be interested in processing video streams (*e.g.*, for traffic monitoring, industrial quality control or video surveillance) to deduce localised data characteristics, *i.e.*, restricted to a given area of the original field-of-view (*e.g.*, for traffic density estimation or object detection [1]). Data processing tools based on random projection should thus comply with this goal, which can be challenging if the area-of-interest is unknown *a priori*, or subject to change.

In this work, we tackle the question of performing *signal estimation*—here restricted to the estimation of some linear function of a signal—from a projected data stream provided by

*LJ is funded by the Belgian FNRS. Part of this work is funded by the FNRS under Grant no T.0136.20 (Learn2Sense).

quadratic random projections of signals (see Sec. 2). This specific data-agnostic projection is related to the sketching¹ technique of rank-one projections (ROP) [8, 6]. Our work is motivated by a recent optical machine which performs ultra-rapid and low-power computations of such quadratic sketches [10]. However, with this technology, these computations are performed in a black-box manner, which prohibits us to *explicitly* access the random construction supporting the ROP model inside the localisation procedure—they are thus *hidden* to us (*i.e.*, we do not have access to the adjoint ROP operator). Our main contribution amounts to showing theoretically that, up to some controlled distortion, signal estimation can be operated directly on quadratic signal sketches without reconstructing the signal (and hence avoiding often costly reconstruction methods [7, 6, 5, 8]). More specifically, the square of this comparison—which is a ROP itself—can be approximated by projecting the sketched signal on the *sign* of the sketched pattern, *i.e.*, summarising this last sketch to only the *sign* its components (see Sec. 3). We thus achieve a sort of *inception* of a ROP inside a ROP model. This result is thus similar to the techniques pursued in [3], where specific signal processing tasks are shown to be computable directly from compressive measurements. Our theoretical analysis is achieved from a generalisation of the sign product embedding (SPE) property, initially developed in one-bit compressive sensing [9, 5].

To demonstrate the efficacy of our approach, we consider several synthetic signal estimation scenarios in Sec. 4. A first experiment quantifies the approximation error induced by our approach by considering a pattern that is orthogonal to the signal of interest. Next, we show how we can localise a rotating disk in one of the four quadrants of an image using only the image ROPs. Finally, we propose to classify MNIST handwritten digit images [4] directly in the DROP domain, with accuracy comparable to a direct processing of these images.

2 Sketching with hidden rank one projections

In this work, we consider the *quadratic data sketching* mechanism, which consists in taking a series of m quadratic measurements $(\mathbf{a}_i^\top \mathbf{x})^2$ of a signal of interest $\mathbf{x} \in \mathbb{R}^n$, with a set of m random vectors $\{\mathbf{a}_i\}_{i=1}^m \subset \mathbb{R}^n$. The sketching operator \mathcal{A}^v applied to the vector \mathbf{x} is thus defined as

$$\mathcal{A}^v : \mathbf{x} \in \mathbb{R}^n \mapsto \mathcal{A}^v(\mathbf{x}) := ((\mathbf{a}_i^\top \mathbf{x})^2)_{i=1}^m := ((\mathbf{a}_1^\top \mathbf{x})^2, \dots, (\mathbf{a}_m^\top \mathbf{x})^2) \in \mathbb{R}_+^m. \quad (1)$$

We assume that, while the operator \mathcal{A}^v can be computed, we cannot explicitly access to the random vectors $\{\mathbf{a}_i\}_{i=1}^m$. This restriction is indeed required by a recent optical technology, named Optical Processing Unit (OPU) [10]. An OPU allows us to compute all the components of $\mathcal{A}^v(\mathbf{x})$ in a reproducible way using the physical properties of multiple scattering of coherent light in random media, which is thus extremely fast and power-efficient (even if $m \simeq n$). In this context, the vectors $\{\mathbf{a}_i\}_{i=1}^m$ are fixed, but *hidden* to us. Moreover, following the observations made in [10], we assume that each random vector \mathbf{a}_i is i.i.d. as a Gaussian random vector² $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, with identity covariance \mathbf{I}_n .

As observed in the context of phase retrieval [7], the operator \mathcal{A}^v amounts to a rank-one projection of the *lifted signal*, *i.e.*, the rank-one matrix $\mathbf{X} = \mathbf{x}\mathbf{x}^\top \in \mathbb{R}^{n \times n}$, onto the rank-one random matrices $\{\mathbf{A}_i := \mathbf{a}_i\mathbf{a}_i^\top\}_{i=1}^m \subset \mathbb{R}^{n \times n}$, as defined by the equivalence

$$\mathcal{A}^v(\mathbf{x}) = ((\mathbf{a}_i^\top \mathbf{x})^2)_{i=1}^m = (\mathbf{a}_i^\top \mathbf{x}\mathbf{x}^\top \mathbf{a}_i)_{i=1}^m = (\langle \mathbf{A}_i, \mathbf{X} \rangle)_{i=1}^m =: \mathcal{A}(\mathbf{X}),$$

¹In this work, the term “*sketch*” designates a generic data transformation (feature map) without restriction to the special case of dimensionality reduction; the dimension of a signal sketch can thus be larger than the input signal space dimension.

²This optical projection is actually modelled by quadratic projections over complex random vectors, but we here work in the real field for the sake of simplicity.

where $\langle \cdot, \cdot \rangle$ is the Frobenius inner product [8, 6]. We thus use “quadratic sketch” and “ROP measurements” interchangeably.

As the ROP operator is biased—*i.e.*, given a signal \mathbf{x} , there is no constant $c > 0$ such that $\frac{c}{m}\mathbb{E}\|\mathcal{A}^v(\mathbf{x})\|^2$ equals $\|\mathbf{x}\|^4 = \|\mathbf{X}\|_F^2$ —it is useful to introduce the *debiased* ROP operator (DROP) [8]

$$\mathcal{B} : \mathbf{x} \in \mathbb{R}^n \mapsto \mathcal{B}(\mathbf{x}) = (\mathcal{A}_{2i}^v(\mathbf{x}) - \mathcal{A}_{2i+1}^v(\mathbf{x}))_{i=1}^m. \quad (2)$$

An OPU can trivially implement this operation by applying finite pixel differences in its focal plane. It can be shown that $\frac{1}{4m}\mathbb{E}\|\mathcal{B}(\mathbf{x})\|^2 = \|\mathbf{x}\|^4$ [8, Lemma 4]. While the approximation $\frac{1}{4m}\|\mathcal{B}(\mathbf{x})\|^2 \approx \|\mathbf{x}\|^4$ is hard to achieve at reasonable values of m — \mathcal{B} respects the restricted isometry property (RIP) in only a few restrictive settings [8, 6]—we leverage below another useful property of \mathcal{B} .

3 Signal estimation in the DROP domain

Our objective is to show that we can do (approximate) signal estimation—when this estimation amounts to estimating a linear function of the signal—directly from random quadratic signal sketches, and thus without knowing the observed signal. This is possible by demonstrating that the DROP operator respects, with high probability, the (local) sign product embedding (SPE) property [9], whose proof is postponed to App. A. This SPE is here instantiated on a signal space \mathcal{S} consisting of k sparse signals of $\Sigma_k := \{\mathbf{v} \in \mathbb{R}^n : |\text{supp}(\mathbf{v})| \leq k\}$. However, by rotational symmetry of the Gaussian distribution, the proposition below is also valid for sets of k sparse signals in an orthonormal basis $\Psi \in \mathbb{R}^{n \times n}$ (such as the wavelet or Fourier bases).

Proposition 1. *Given a fixed unit vector $\mathbf{u} \in \mathbb{R}^n$, $\kappa = \pi/4$, and a distortion $0 < \delta < 1$, provided that*

$$m \geq C\delta^{-2}k \log\left(\frac{n}{k\delta}\right), \quad (3)$$

*then, with probability exceeding $1 - C \exp(-c\delta^2 m)$, for all k -sparse signals $\mathbf{x} \in \Sigma_k := \{\mathbf{v} \in \mathbb{R}^n : |\text{supp}(\mathbf{v})| \leq k\}$, \mathcal{B} respects the SPE over Σ_k , *i.e.*,*

$$\left| \frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle - \langle \mathbf{u}, \mathbf{x} \rangle^2 \right| \leq \delta \|\mathbf{x}\|^2, \quad (4)$$

with sign the sign operator applied componentwise on vectors.

This proposition states that, provided that m is large compared to the dimension of the signal space \mathcal{S} (here $\mathcal{S} = \Sigma_k$ and we thus need $m = \Omega(\delta^{-2}k)$ up to log factors), for any vector $\mathbf{x} \in \mathcal{S}$, projecting its sketch $\mathcal{B}(\mathbf{x})$ on $\text{sign}(\mathcal{B}(\mathbf{u})) \in \{\pm 1\}^m$ is a proxy for $\langle \mathbf{u}, \mathbf{x} \rangle^2$ —a ROP of $\mathbf{x}\mathbf{x}^\top$ by $\mathbf{u}\mathbf{u}^\top$. The spirit of this result is thus similar to the approach of [3] in linear compressive sensing [3], where a set of signal processing techniques (including signal estimation) are proved to be applicable in the compressed signal measurements. Let us emphasise that the error of the approximation (4) is bounded by $\delta \|\mathbf{x}\|^2$, that is $O(\sqrt{k/m} \|\mathbf{x}\|^2)$ (up to log factors) by saturating (3). Therefore, the equation (4) is useful only if $\langle \mathbf{u}, \mathbf{x} / \|\mathbf{x}\| \rangle^2$ is sufficiently large compared to δ . Incidentally, this allows us to estimate δ as a function of m by taking a vector \mathbf{u} orthogonal to \mathbf{x} (see Sec. 4). Notice that, by a simple union bound argument, Eq. 4 in Prop. 1 can be shown to hold with the same probability bound for all \mathbf{u} in a finite set of S unit vectors provided $m \geq C\delta^{-2}(k \log(\frac{n}{k\delta}) + \log S)$ (see Cor. 3 in App. A).

4 Experiments

This section contains three experiments demonstrating how one can directly carry out signal estimation in the range of the DROP operator. Our first experiment aims at measuring the

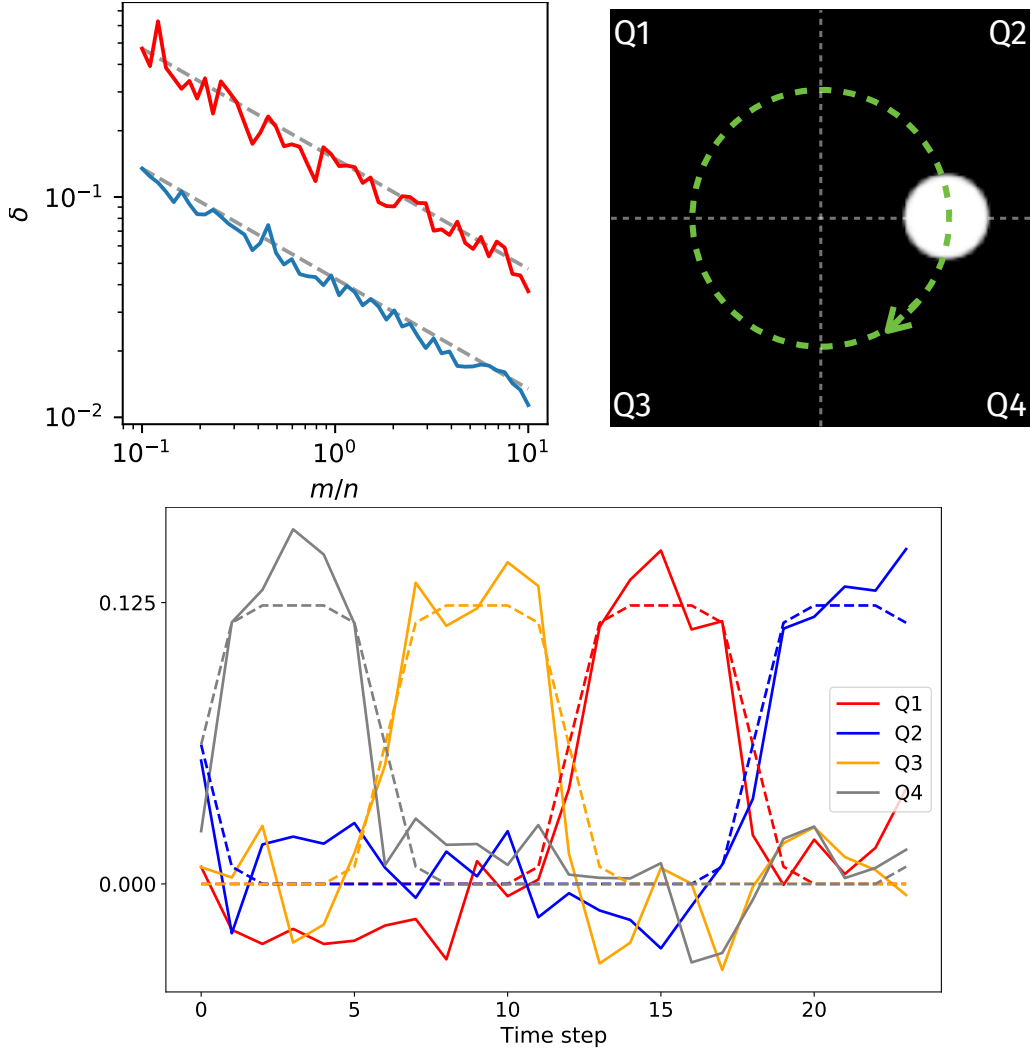


Figure 1: (Top-left) Estimation of δ in (4). (Top-right) Synthetic dynamic image sequence; a rotating white disk on a black background. (Bottom) The estimation of the quadrant occupancy functions per quadrant.

approximation error (δ) in (4). For this purpose, we took two unit vectors $\mathbf{x}, \mathbf{u} \in \mathbb{R}^n$ picked uniformly at random on the sphere \mathbb{S}^{n-1} with the additional constraint that \mathbf{u} must be orthogonal to \mathbf{x} . In this case, the term $\langle \mathbf{u}, \mathbf{x} \rangle^2$ vanishes in (4) and one can estimate δ by Monte Carlo simulations over several trials of $\frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle$. We show in Fig. 1(top-left) the estimated δ over 100 trials (average and maximum value) for $n = 1000$ and varying ratio $m/n \in [0.1, 10]$. This estimation confirms that δ decays as $O(m^{-1/2})$ when m increases. Moreover, on average, $\delta \simeq 0.1$ as soon as $m/n > 0.2$.

As a second experiment, we test the possibility to perform localised detection in a simple synthetic video; a sequence of 24 images $\{\mathbf{x}_t\}_{t=0}^{23}$ of size 128×128 ($n = 16384$) representing a white disk rotating on a black (zero) background (see Fig. 1(top-right)). Our objective is to detect the passage of the disk in each of the four image quadrants by only processing the DROP measurements $\{\mathcal{B}(\mathbf{x}_t)\}_{t=0}^{23} \subset \mathbb{R}^m$. We have thus created 4 normalised patterns $\{\mathbf{u}_j\}_{j=1}^4$, with \mathbf{u}_j being constant in the j -th quadrant and 0 outside, for each $1 \leq j \leq 4$. We show in Fig. 1(bottom) the time evolution of $q_j^{\text{est}}(t) := \frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u}_j)), \mathcal{B}(\mathbf{x}_t) \rangle$ (continuous curves) for $m/n = 0.5$ in comparison with the *quadrant occupancy functions* $q_j(t) := \langle \mathbf{u}_j, \mathbf{x}_t \rangle^2$ (dashed lines). The colour coding is given in the legend. We observe that each curves $q_j^{\text{est}}(t)$ provides

a fair estimation of $q_j(t)$. Moreover, the amplitude of each curve $q_j^{\text{est}}(t)$ reduces when the disk is not in the associated quadrant. A rough detection of the quadrant occupancy could thus be established by appropriately thresholding each time signal $q_j^{\text{est}}(t)$.

As final illustration, we perform a toy example (naive) classification of images taken in the labelled MNIST dataset $\mathcal{X} := \{(\mathbf{x}_k, t_k)\}_{k=1}^{N=70\,000} \subset \mathbb{R}^{28 \times 28} \times \{0, \dots, 9\}$ [4] (with normalisation $\|\mathbf{x}_k\| = 1$). We want to compare the classification performed in the direct, pixel domain, to the one operated in the sketched domain. We start by randomly splitting \mathcal{X} into a training \mathcal{X}_{tr} and a test set \mathcal{X}_{te} according to a split of 60 000 and 10 000 images, respectively. From \mathcal{X}_{tr} we compute the 10 centroids $\{\mathbf{c}_j\}_{j=0}^9 \subset \mathbb{R}^{28 \times 28}$ of each class of digits and we define 10 vectors $\mathbf{u}_j = \mathbf{c}_j / \|\mathbf{c}_j\|$. In the direct domain, the estimated label of an image \mathbf{x}_k is then defined as $\hat{t}_k := \arg \max_j \langle \mathbf{u}_j, \mathbf{x}_k \rangle^2$. In the sketched domain, the label estimate of a test image \mathbf{x}_k is computed as $\hat{t}_k^{\text{sk}} := \arg \max_j \frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u}_j)), \mathcal{B}(\mathbf{x}_k) \rangle$, which should be close to \hat{t}_k according to Prop. 1. We reach the following average testing accuracy (100 trials, \pm standard deviation in %) for both the direct and the sketched classifications:

	Direct	$m = 200$	$m = 400$	$m = 800$	$m = 1600$
Accuracy [%]	81.2	69.3 \pm 1.9	75.2 \pm 1.25	78.6 \pm 0.91	80.4 \pm 0.79

These results confirm that, up to some distortion, one can apply this naive classification procedure directly in the DROP measurements of the dataset. As predicted by our analysis, this approximation improves when m increases.

To prospectively question the possibility to reach better classification accuracy in the DROP domain, we consider the dataset $\mathcal{X}^{\mathcal{B}} = \{(\mathcal{B}(\mathbf{x}_k), t_k)\}_{k=1}^{N=70\,000}$, and compute the centroids $\{\mathbf{c}_j^{\mathcal{B}}\}_{j=0}^9 \subset \mathbb{R}^m$ in the training split of $\mathcal{X}^{\mathcal{B}}$. We then compare two possible classification methods. The first proceeds as a direct classification in the DROP domain, *i.e.*, we compute the label estimate $\hat{t}_k^{\mathcal{B}} := \arg \max_j \langle \mathbf{c}_j^{\mathcal{B}}, \mathcal{B}(\mathbf{x}_k) \rangle$. The second approach assumes that, for each centroid, there exists a vector \mathbf{v}_j such that $\mathbf{c}_j^{\mathcal{B}} \approx \mathcal{B}(\mathbf{v}_j)$. Under this assumption, we rather estimate our label by applying a *sign* operation to each $\mathbf{c}_j^{\mathcal{B}}$ —which is more aligned to Prop. 1—and we compute $\hat{t}_k^{\mathcal{B}, \text{sign}} := \arg \max_j \langle \text{sign}(\mathbf{c}_j^{\mathcal{B}}), \mathcal{B}(\mathbf{x}_k) \rangle$. We reach the following average testing accuracy for the two approaches (for $m = 800$, 500 trials, \pm standard deviation in %):

	Direct	Estimate \hat{t}_k^{sk}	Estimate $\hat{t}_k^{\mathcal{B}}$	Estimate $\hat{t}_k^{\mathcal{B}, \text{sign}}$
Accuracy [%]	81.2	78.5 \pm 0.7	63.4 \pm 3.3	84.2 \pm 0.6

Keeping only the sign of the estimated centroids provides better testing accuracy than the unsigned method. In fact, the signed approach outperforms the testing accuracy of the direct approach by about 3%.

5 Conclusion and perspectives

Our developments have shown that signal estimation is possible by directly processing the quadratic measurements of a signal. We achieved this by showing that such ROP measurements satisfy the sign product embedding with high probability provided m is sufficiently large compared to the signal space dimension. Our theoretical results were backed up by several synthetic signal estimation and classification experiments. In future works, we plan to reproduce this experiment on an actual OPU to further reduce the computational cost of the procedure, by also leveraging of the one-bit nature of the signed sketches.

A Proofs

In this appendix, we start by providing a proof of the isotropy of the DROP operator \mathcal{B} . A more general proof is provided in [8, Lemma 4] (when \mathcal{B} is extended to the mapping of matrices of any rank and for sub-gaussian random vectors \mathbf{a}_i), but we find useful to provide the short alternate proof below for the sake of self-containedness.

Proposition 2 (DROP isotropy). *Given $\mathbf{x} \in \mathbb{R}^n$, we have*

$$\frac{1}{m} \mathbb{E} \|\mathcal{B}(\mathbf{x})\|^2 = 4 \|\mathbf{x}\|^4. \quad (5)$$

Proof. By rotational symmetry of the Gaussian distribution and by homogeneity of (5) in $\|\mathbf{x}\|^4$, it is enough to prove Prop. 2 for $\mathbf{x} = \mathbf{e}_1 := (1, 0, \dots, 0)^\top$. In this case, for two independent random vectors $\mathbf{a}, \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, we easily show that $\frac{1}{m} \mathbb{E} \|\mathcal{B}(\mathbf{x})\|^2 = \mathbb{E}[(\mathbf{a}^\top \mathbf{x})^2 - (\mathbf{b}^\top \mathbf{x})^2]^2 = \mathbb{E}[a_1^2 - b_1^2]^2 2(\mathbb{E}a_1^4 - \mathbb{E}a_1^2 b_1^2) = 4 = 4 \|\mathbf{x}\|^4$. \square

We now prove the central result of this work, Prop. 1, that is restated below for convenience

Proposition (1). *Given a fixed unit vector $\mathbf{u} \in \mathbb{R}^n$, $\kappa = \pi/4$, and a distortion $0 < \delta < 1$, provided that*

$$m \geq C \delta^{-2} k \log\left(\frac{n}{k\delta}\right), \quad (3)$$

then, with probability exceeding $1 - C \exp(-c\delta^2 m)$, for all k -sparse signals $\mathbf{x} \in \Sigma_k := \{\mathbf{v} \in \mathbb{R}^n : |\text{supp}(\mathbf{v})| \leq k\}$, \mathcal{B} respects the SPE over Σ_k , i.e.,

$$\left| \frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle - \langle \mathbf{u}, \mathbf{x} \rangle^2 \right| \leq \delta \|\mathbf{x}\|^2, \quad (4)$$

with sign the sign operator applied componentwise on vectors.

Proof. Let us first show that the expectation of $\frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle$ is actually equal to $\langle \mathbf{u}, \mathbf{x} \rangle^2$.

By rotational symmetry of the Gaussian distribution and homogeneity of this expectation in $\|\mathbf{x}\|^2$, it is enough to prove it for $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{x} = c\mathbf{e}_1 + s\mathbf{e}_2$, with $c = \cos \theta$, $s = \sin \theta$, and θ the angle between \mathbf{x} and \mathbf{u} .

In this case, for two independent random vectors $\mathbf{a}, \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, we easily show that

$$\begin{aligned} \frac{1}{m} \mathbb{E} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle &= \mathbb{E} \text{sign}(a_1^2 - b_1^2) [(ca_1 + sa_2)^2 - (cb_1 + sb_2)^2] \\ &= \mathbb{E} \text{sign}(a_1^2 - b_1^2) [c^2(a_1^2 - b_1^2) + s^2(a_2^2 - b_2^2) + 2cs(a_1 a_2 - b_1 b_2)] \\ &= c^2 \mathbb{E} \text{sign}(a_1^2 - b_1^2) (a_1^2 - b_1^2) = c^2 \mathbb{E} |a_1 - b_1| |a_1 + b_1|. \end{aligned}$$

Since $a_1 - b_1 \sim \mathcal{N}(0, 2)$ and $a_1 + b_1 \sim \mathcal{N}(0, 2)$ are decorrelated, and thus independent, we have $\mathbb{E} |a_1 - b_1| |a_1 + b_1| = 2 \mathbb{E}(2^{-1/2} |a_1 - b_1|) \mathbb{E}(2^{-1/2} |a_1 + b_1|) = \frac{4}{\pi} = 1/\kappa$, which proves the claim since $c^2 = \cos^2 \theta = \langle \mathbf{u}, \mathbf{x} \rangle^2$.

Second, we study the concentration of $\frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle$ around its mean for a fixed unit vector \mathbf{u} and still assuming that $\|\mathbf{x}\|^2 = 1$ by homogeneity of (4).

One can write

$$S := \kappa \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle - m \langle \mathbf{u}, \mathbf{x} \rangle^2,$$

as the sum $S = \sum_{i=1}^m Z_i$ with

$$Z_i := s_i(\mathbf{u}) [(\mathbf{a}_{2i}^\top \mathbf{x})^2 - (\mathbf{a}_{2i+1}^\top \mathbf{x})^2],$$

and $s_i(\mathbf{u}) = \text{sign}[(\mathbf{a}_{2i}^\top \mathbf{u})^2 - (\mathbf{a}_{2i+1}^\top \mathbf{u})^2]$. The random variables Z_i are all i.i.d. and sub-exponential since their sub-exponential norm is bounded as

$$\|Z_i\|_{\psi_1} = \|(\mathbf{a}_{2i}^\top \mathbf{x})^2 - (\mathbf{a}_{2i+1}^\top \mathbf{x})^2\|_{\psi_1} \leq 2\|(\mathbf{a}_{2i}^\top \mathbf{x})^2\|_{\psi_1} \leq 4\|\mathbf{a}_{2i}^\top \mathbf{x}\|_{\psi_2}^2 \leq C,$$

see [11, Def. 5.13 & Lemma 5.14].

Therefore, from [11, Cor. 5.17], for any $\delta \geq 0$, we get the concentration result $\mathbb{P}[|S| \geq \delta m] \leq 2 \exp(-c \min(\delta^2, \delta)m)$, or

$$\mathbb{P}\left[\left|\frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u})), \mathcal{B}(\mathbf{x}) \rangle - \langle \mathbf{u}, \mathbf{x} \rangle^2\right| \geq \delta\right] \leq 2 \exp(-c \min(\delta^2, \delta)m). \quad (6)$$

Finally, we extend the previous concentration result for a fixed unit vector \mathbf{u} and for all vectors of Σ_k . By the same homogeneity argument, it is enough to prove the final claim of Prop. 1 for all unit k -sparse signals \mathbf{x} of $\Sigma_k \cap \mathbb{S}^{n-1}$.

Let us observe that, for all $i \in [m]$, $\mathcal{B}(\mathbf{x}) = \mathbf{d}_i^\top \mathbf{X} \mathbf{s}_i$ with $\mathbf{d}_i := \mathbf{a}_{2i} - \mathbf{a}_{2i+1}$, $\mathbf{s}_i := \mathbf{a}_{2i} + \mathbf{a}_{2i+1}$, and $\mathbf{X} = \mathbf{x} \mathbf{x}^\top$. Moreover, the lifted matrix \mathbf{X} has rank one, unit Frobenius norm and is sparse along its rows and columns, or *bi-sparse* [13].

Therefore, given a radius $\epsilon > 0$ to be fixed momentarily, there exists a ϵ -covering $\mathcal{G}_\epsilon \subset \mathcal{G}$ of the lifted set $\mathcal{G} := \{\mathbf{x} \mathbf{x}^\top : \mathbf{x} \in \Sigma_k \cap \mathbb{S}^{n-1}\}$ —such that for any $\mathbf{X} \in \mathcal{G}$ there is a $\mathbf{X}' \in \mathcal{G}_\epsilon$ with same row and column supports and $\|\mathbf{X} - \mathbf{X}'\| \leq \epsilon$ —whose cardinality $|\mathcal{G}_\epsilon|$ does not exceed $\binom{n}{k} (9/\epsilon)^{2k+1}$ [12, Lemma 3.1], which is also crudely bounded by $(\frac{cn}{\epsilon k})^{2k+1}$. Let us set $\bar{\mathbf{u}} := \text{sign}(\mathcal{B}(\mathbf{u}))$. Since

$$\frac{\kappa}{m} \langle \bar{\mathbf{u}}, \mathcal{B}(\mathbf{x}) \rangle - \langle \mathbf{u}, \mathbf{x} \rangle^2 = \frac{\kappa}{m} \langle \bar{\mathbf{u}}, \mathbf{d}_i^\top \mathbf{X} \mathbf{s}_i \rangle - \mathbf{u}^\top \mathbf{X} \mathbf{u},$$

by a standard union bound argument applied to (6), we thus see that, with probability exceeding $1 - 2|\mathcal{G}_\epsilon| \exp(-c \min(\delta^2, \delta)m)$ we have for all $\mathbf{X}' \in \mathcal{G}_\epsilon$

$$\left| \frac{\kappa}{m} \langle \bar{\mathbf{u}}, \mathbf{d}_i^\top \mathbf{X}' \mathbf{s}_i \rangle - \mathbf{u}^\top \mathbf{X}' \mathbf{u} \right| \leq \delta.$$

Therefore, provided $m \geq C \min(\delta, \delta^2)^{-1} k \log(\frac{n}{k\epsilon})$, this last inequality holds with probability exceeding $1 - 2 \exp(-c \min(\delta^2, \delta)m)$.

The rest of the proof assumes that this event holds. Let us define the $n \times n$ matrix

$$\mathbf{Q} := \frac{\kappa}{m} \sum_{i=1}^m \bar{u}_i \mathbf{d}_i \mathbf{s}_i^\top - \mathbf{u} \mathbf{u}^\top.$$

We then have for all $\mathbf{X} \in \mathcal{G}$,

$$\frac{\kappa}{m} \langle \bar{\mathbf{u}}, \mathbf{d}_i^\top \mathbf{X} \mathbf{s}_i \rangle - \mathbf{u}^\top \mathbf{X} \mathbf{u} = \langle \mathbf{Q}, \mathbf{X} \rangle.$$

Moreover, let us set $\rho := \sup_{\mathbf{X} \in \mathcal{G}} |\langle \mathbf{Q}, \mathbf{X} \rangle|$. Then, following a similar argument to [12, Sec. 3.2], given an arbitrary $\mathbf{X} \in \mathcal{G}$, and for $\mathbf{X}' \in \mathcal{G}_\epsilon$ such that $\|\mathbf{X} - \mathbf{X}'\| \leq \epsilon$,

$$|\langle \mathbf{Q}, \mathbf{X} \rangle| \leq |\langle \mathbf{Q}, \mathbf{X}' \rangle| + |\langle \mathbf{Q}, \mathbf{Y} \rangle| \|\mathbf{X} - \mathbf{X}'\| \leq \delta + |\langle \mathbf{Q}, \mathbf{Y} \rangle| \epsilon,$$

with $\mathbf{Y} = (\mathbf{X} - \mathbf{X}') / \|\mathbf{X} - \mathbf{X}'\|$. Since the symmetric matrix \mathbf{Y} has rank two, we can decompose it as $\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2$ with $\hat{\mathbf{Y}}_i := \mathbf{Y}_i / \|\mathbf{Y}_i\| \in \mathcal{G}$ and $\langle \mathbf{Y}_1, \mathbf{Y}_2 \rangle = 0$. Therefore, $|\langle \mathbf{Q}, \mathbf{Y} \rangle| \leq |\langle \mathbf{Q}, \hat{\mathbf{Y}}_1 \rangle| \|\mathbf{Y}_1\| + |\langle \mathbf{Q}, \hat{\mathbf{Y}}_2 \rangle| \|\mathbf{Y}_2\| \leq 2\rho$ and $|\langle \mathbf{Q}, \mathbf{X} \rangle| \leq \delta + 2\rho\epsilon$. Taking the supremum over \mathbf{X} , this means that $\rho \leq \delta + 2\rho\epsilon$, or $\rho \leq \delta / (1 - 2\epsilon)$. Picking for instance $\epsilon = \delta/4$ and $\delta < 1$, gives $|\langle \mathbf{Q}, \mathbf{X} \rangle| \leq \delta + \delta^2 \leq 2\delta$, and a final rescaling of δ concludes the proof. \square

The following corollary extends Prop. 4 to the case where \mathbf{u} belongs to a finite set of unit vectors.

Corollary 3. *Given a set of S fixed unit vector $\{\mathbf{u}_s\}_{s=1}^S$, $\kappa = \pi/4$, and a distortion $0 < \delta < 1$, provided that $m \geq C\delta^{-2}(k \log(\frac{n}{k\delta}) + \log S)$, we have, with probability exceeding $1 - C \exp(-c\delta^2 m)$, for all k -sparse signals $\mathbf{x} \in \Sigma_k := \{\mathbf{v} \in \mathbb{R}^n : |\text{supp}(\mathbf{v})| \leq k\}$ and all $s \in [S]$,*

$$\left| \frac{\kappa}{m} \langle \text{sign}(\mathcal{B}(\mathbf{u}_s)), \mathcal{B}(\mathbf{x}) \rangle - \langle \mathbf{u}_s, \mathbf{x} \rangle^2 \right| \leq \delta \|\mathbf{x}\|^2. \quad (7)$$

The proof of this corollary consists in bounding the failure of (7) using Prop. 1 and a simple union bound argument.

References

- [1] V. Kastrinaki, M. Zervakis, K. Kalaitzakis, A survey of video processing techniques for traffic applications. *Image and vision computing*, 21(4), 359-381, 2003.
- [2] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [3] P.T. Boufounos, M.B. Wakin, R.G. Baraniuk, M.A. Davenport, Signal processing with compressive measurements, *IEEE Journal of Selected topics in Signal processing*, 4(2), 445-460. 2010.
- [4] L. Deng, The MNSIT database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 2012, vol. 29, no 6, p. 141-142.
- [5] S. Foucart, Flavors of Compressive Sensing. In: Fasshauer, G., Schumaker, L. *Approximation Theory XV: San Antonio 2016. AT 2016. Springer Proceedings in Mathematics & Statistics*, vol 201, Springer, 2017.
- [6] T. Cai, and A. Zhang., ROP: Matrix Recovery via Rank One Projections, *The Annals of Statistics*, vol. 43, no. 1, pp. 102-38, 2015.
- [7] E.J. Candès, T. Strohmer, and V. Voroninski, PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming., *Comm. Pure Appl. Math.*, 66: 1241-1274, 2013.
- [8] Y. Chen, Y. Chi, A.J. Goldsmith, Exact and Stable Covariance Estimation From Quadratic Sampling via Convex Programming, *IEEE Transactions on Information Theory*, vol. 61, no. 7, 2015.
- [9] L. Jacques, K. Degraux, C. De Vleeschouwer, Quantized Iterative Hard Thresholding: Bridging 1-bit and High-Resolution Quantized Compressed Sensing. *Proceedings of International Conference on Sampling Theory and Applications, IEEE2013*, p.105-108, 2013.
- [10] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Drémeau, S. Gigan, and F. Krzakala, Random projections through multiple optical scattering: Approximating kernels at the speed of light. In *IEEE ICASSP 2016* (pp. 6215-6219).
- [11] R. Vershynin, "Introduction to the Non-Asymptotic Analysis of Random Matrices." In Y. C. Eldar & G. Kutyniok (ed.), *Compressed Sensing: Theory and Practice* (pp. 210–268), 2012. Cambridge University Press . ISBN: 9780511794308.
- [12] E. J. Candès, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57(4), 2342-2359. 2011

- [13] S. Foucart, R. Gribonval, L. Jacques, and H. Rauhut, Jointly low-rank and bispase recovery: Questions and partial answers. *Analysis and Applications*, 18(01), 25-48, 2020.
- [14] A. Rahimi, B. Recht, Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. *Advances in neural information processing systems*, 21.B, 2008.