

Rare Eclipses in Quantised Random Embeddings of Disjoint Convex Sets: a Matter of Consistency?

Valerio Cambareri, Chunlei Xu and Laurent Jacques

ISPGROUP, ICTEAM/ELEN, Université catholique de Louvain, Louvain-la-Neuve, Belgium.

E-mail: {valerio.cambareri, chunlei.xu, laurent.jacques}@uclouvain.be

Abstract—We study the problem of verifying when two disjoint closed convex sets remain separable after the application of a quantised random embedding, as a means to ensure exact classification from the signatures produced by this non-linear dimensionality reduction. An analysis of the interplay between the embedding, its quantiser resolution and the sets' separation is presented in the form of a convex problem; this is completed by its numerical exploration in a special case, for which the phase transition corresponding to exact classification is easily computed.

I. PROBLEM STATEMENT

Non-linear dimensionality reduction techniques play an important role in simplifying statistical learning on very large-scale datasets. Among such techniques, we focus on *quantised random embeddings* obtained by a non-linear map A applied to $\mathbf{x} \in \mathcal{K} \subset \mathbb{R}^n$, that is

$$\mathbf{y} = A(\mathbf{x}) := \mathcal{Q}_\delta(\Phi \mathbf{x} + \xi) \quad (1)$$

with $\Phi \in \mathbb{R}^{m \times n}$ a random *sensing matrix*, $\mathcal{Q}_\delta(\cdot) := \delta \lfloor \cdot / \delta \rfloor$ a *uniform scalar quantiser* of resolution $\delta > 0$ (applied component-wise), and the *signature* $\mathbf{y} \in \delta \mathbb{Z}^m$. In (1), the *dithering*¹ $\xi \sim \mathcal{U}^m([0, \delta])$ is a well-known means to stabilise the action of the quantiser [1], [2].

The non-linear map (1) is a non-adaptive dimensionality reduction that yields compact signatures for storage and transmission, while retaining a notion of *quasi-isometry* that enables the approximation of $\mathbf{x}$ [2], [3]. Consequently, distance-based learning tasks preserve their accuracy if run on $A(\mathcal{K})$ rather than $\mathcal{K}$, provided some requirements are met on m , δ , the distribution of Φ and the “dimension” of $\mathcal{K}$ as measured, e.g., by its *Gaussian mean width* $w(\mathcal{K}) := \sup_{\mathbf{x} \in \mathcal{K}} |\mathbf{g}^\top \mathbf{x}|$ with $\mathbf{g} \sim \mathcal{N}(0, 1)$ (see, e.g., [2]). In this context we aim to show that, given two classes described by some sets $\mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{K} : \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ and $\mathbf{x} \in \mathcal{C}_1 \cup \mathcal{C}_2 \subset \mathcal{K}$, classifying whether $\mathbf{x}$ belongs to $\mathcal{C}_1$ or $\mathcal{C}_2$ is still possible from $\mathbf{y} = A(\mathbf{x})$. For linear embeddings such as $\mathbf{y} = \Phi \mathbf{x}$, Bandeira *et al.* [4] approach the above classification problem as follows.

Problem 1 (Rare Eclipse Problem (from [4])). *Let $\mathcal{C}_1, \mathcal{C}_2 \subset \mathbb{R}^n : \mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$ be closed convex sets, $\Phi \sim \mathcal{N}^{m \times n}(0, 1)$. Given $\eta \in (0, 1)$, find the smallest m so that $p_0 := \mathbb{P}[\Phi \mathcal{C}_1 \cap \Phi \mathcal{C}_2 = \emptyset] \geq 1 - \eta$.*

Prob. 1 amounts to ensuring for all $\mathbf{x}' \in \mathcal{C}_1, \mathbf{x}'' \in \mathcal{C}_2$ that their images $\Phi \mathbf{x}' \neq \Phi \mathbf{x}''$. Using the *difference set* $\mathcal{C}^- := \mathcal{C}_1 - \mathcal{C}_2 = \{\mathbf{z} := \mathbf{x}' - \mathbf{x}'' : \mathbf{x}' \in \mathcal{C}_1, \mathbf{x}'' \in \mathcal{C}_2\}$ we see the above problem equals

$$\mathbb{P}[\forall \mathbf{z} \in \mathcal{C}^-, \Phi \mathbf{z} \neq \mathbf{0}_m] = 1 - \mathbb{P}[\exists \mathbf{z} \in \mathcal{C}^- : \Phi \mathbf{z} = \mathbf{0}_m] \geq 1 - \eta.$$

This requires a bound on the probability that the kernel of Φ “collides” with $\mathcal{C}^-$, i.e., $\mathbb{P}[\text{Ker}(\Phi) \cap \mathcal{C}^- \neq \emptyset] \leq \eta$, and [4] shows that η is small if m is large compared to the “dimension” of $\mathcal{C}^-$ as measured by $w_\cap^2 := w^2((\mathbb{R}_+ \mathcal{C}^-) \cap \mathbb{S}^{n-1})$ with $\mathbb{R}_+ \mathcal{C}^-$ the cone generated by $\mathcal{C}^-$.

From this standpoint, extending such existing results on Prob. 1 to non-linear maps as (1) is non-trivial. Applying A to each closed convex set $\mathcal{C}_1, \mathcal{C}_2$ would produce two countable sets $A(\mathcal{C}_1), A(\mathcal{C}_2) \subset \delta \mathbb{Z}^m$, and assessing if they still “collide” is our key question below.

Problem 2 (Quantised Eclipse Problem). *In the setup of Prob. 1, given $\eta \in (0, 1)$, find the smallest m so that $\mathbb{P}[A(\mathcal{C}_1) \cap A(\mathcal{C}_2) = \emptyset] \geq 1 - \eta$, i.e., $p_\delta := \mathbb{P}[\forall \mathbf{x}' \in \mathcal{C}_1, \mathbf{x}'' \in \mathcal{C}_2, A(\mathbf{x}') \neq A(\mathbf{x}'')] \geq 1 - \eta$.*

Note that the event in Prob. 2 requires $\mathbb{P}[\exists \mathbf{x}' \in \mathcal{C}_1, \mathbf{x}'' \in \mathcal{C}_2 : A(\mathbf{x}') = A(\mathbf{x}'')] \leq \eta$, i.e., a bound on the probability of existence of

two *consistent* vectors (through the mapping A) that do not belong to the same set.

We here leverage the quantised restricted isometry property (QRIP) introduced in [2] to estimate η and the conditions on m . The QRIP establishes some conditions on m that ensure $\frac{1}{m} \|A(\mathbf{x}') - A(\mathbf{x}'')\|_1 \geq (c' - \varepsilon) \|\mathbf{z}\| - c\delta\varepsilon' \geq (c' - \varepsilon)\sigma - c\delta\varepsilon' =: \bar{H}$ for some controllable distortions $\varepsilon, \varepsilon' > 0$, with $\sigma \leq \|\mathbf{z}\|$ and some constants $c, c' > 0$. Thus $A(\mathbf{x}') \neq A(\mathbf{x}'')$ if $\bar{H} > 0$. In particular, we deduce the following proposition whose proof is postponed to an extended version of this work.

Proposition 1. *In the setup of Prob. 1, given $\delta > 0$, $\eta \in (0, 1)$, $\sigma := \min_{\mathbf{z} \in \mathcal{C}^-} \|\mathbf{z}\|$ and the mapping A defined in (1), if*

$$m \gtrsim (w_\cap^2 + \frac{\delta^2}{\sigma^2} n) \log(1 + \frac{rm}{\delta}) + C \log \frac{1}{\eta}, \quad (2)$$

with $C > 1$ not depending on m and n , then $p_\delta \geq 1 - \eta$.

Numerically testable but stronger conditions ensuring $p_\delta > 1 - \eta$ in Prob. 2 can be deduced as follows. We first note that if $\Phi \mathbf{z} = \mathbf{0}_m$ for a given Φ and any $\mathbf{z} \in \mathcal{C}^-$, i.e., $\text{Ker}(\Phi) \cap \mathcal{C}^- \neq \emptyset$, then $p_\delta = 0$ for all $\delta > 0$ since then $\Phi \mathbf{x}' + \xi = \Phi \mathbf{x}'' + \xi$. Second, since $A(\mathbf{x}') = A(\mathbf{x}'')$ induces $\|\Phi \mathbf{z}\|_\infty \leq \delta$ for $\mathbf{z} := \mathbf{x}' - \mathbf{x}'' \in \mathcal{C}^-$, proving $p_\delta = \mathbb{P}[\forall \mathbf{z} \in \mathcal{C}^-, \|\Phi \mathbf{z}\|_\infty > \delta] \geq 1 - \eta$ will solve Prob. 2 since $p_\delta \geq \bar{p}_\delta$.

We define accordingly a *consistency margin* $\tau := \|\Phi \mathbf{z}^*\|_\infty$ with

$$\mathbf{z}^* := \argmin_{\mathbf{z} \in \mathcal{K}} \|\Phi \mathbf{z}\|_\infty \text{ s.t. } \mathbf{z} \in \mathcal{C}^- := \mathcal{C}_1 - \mathcal{C}_2, \quad (3)$$

i.e., as a property of Φ and $\mathcal{C}^-$ so that, if $\tau > \delta$, we necessarily have $A(\mathbf{x}') \neq A(\mathbf{x}'')$ for $\mathbf{x}', \mathbf{x}''$ in different classes, i.e., $\bar{p}_\delta := \mathbb{P}[\tau > \delta]$. Intuitively, $\mathbf{z}^*$ is related to the minimal separation σ between $\mathcal{C}_1$ and $\mathcal{C}_2$.

Note that (3) is clearly convex if $\mathcal{K}$ and $\mathcal{C}^-$ are convex. We anticipate that the construction of a *certificate* for this problem will provide a bound on $\tau > \delta$ when $\mathcal{C}^-$ is known, and analyse an exemplary case afterwards.

II. NUMERICAL TEST FOR TWO DISJOINT ℓ_2 -BALLS

We consider the simple, yet broadly applicable convex case of two balls $\mathcal{C}_1 := r_1 \mathbb{B}_{\ell_2}^n + \mathbf{c}_1$ and $\mathcal{C}_2 := r_2 \mathbb{B}_{\ell_2}^n + \mathbf{c}_2$. Then, $\mathcal{C}^- = r \mathbb{B}_{\ell_2}^n + \mathbf{c}$ with $\mathbf{c} := \mathbf{c}_1 - \mathbf{c}_2$ and $r := r_1 + r_2$ (see Fig. 1). In this context $\|\mathbf{c}\| = \sigma + r$ and $\frac{w_\cap}{\sqrt{n}} \lesssim \frac{r}{\|\mathbf{c}\|} \leq \frac{r}{\sigma}$ in Prop. 1 [4].

By solving random instances of this problem² w.r.t. $\Phi \sim \mathcal{N}^{m \times n}(0, 1)$, with $n = 2^8$ and $m \in [2^1, 2^8]$, we are able to compute the consistency margin for each Φ on $\mathcal{C}^-$, which is varied by fixing $r = 2$ and taking $\sigma = \|\mathbf{c}\| - r \in [2^0, 2^9]$. Then, we collect $\tau_{\min}$, i.e., the smallest τ resulting from 2^7 trials for each configuration (Fig. 2a), and also estimate on the same trials the probability $\bar{p}_\delta = \mathbb{P}[\tau > \delta := 1]$ in Fig. 2b.

Fig. 2a reports several level curves of $\tau_{\min}$. For each curve, the event $A(\mathcal{C}_1) \cap A(\mathcal{C}_2) = \emptyset$ holds if $\delta := \tau_{\min}$. While this condition is necessary but not sufficient, these level curves are compatible with the points $(\frac{m}{n}, \sigma)$ that satisfy the rule $m \approx \frac{\delta^2}{\sigma^2} n$ (up to log factors) induced by (2) in Prop. 1. Fig. 2b displays a sharp *phase transition* in the contours of $\bar{p}_\delta$. Despite the fact that $p_\delta \geq \bar{p}_\delta$, the contours are also approximately aligned with the iso-probability curves that can be deduced from (2), i.e., $m - c \frac{r^2 + \delta^2}{\sigma^2} n \approx C \log(\frac{1}{\eta})$, with $\bar{p}_\delta \approx 1 - \eta$ for some $C, c > 1$.

III. CONCLUSION AND OPEN QUESTIONS

The fundamental limits of learning tasks with embeddings are being tackled in several studies [5]–[8]. Our contribution expands the requirements for exact classification from the signatures produced by two closed convex sets after quantised random embeddings. We shall also specify this analysis to low-complexity structured sets $\mathcal{K}$ (e.g., selecting disjoint “clusters” of sparse signals).

²By uniformity of $\text{Ker}(\Phi)$, $\Phi \sim \mathcal{N}^{m \times n}(0, 1)$ over the Grassmannian at the origin, it is legitimate to fix a randomly drawn direction $\mathbf{c}/\|\mathbf{c}\|$ for the simulations.

The authors are funded by the F.R.S.-FNRS and by the project ALTERSENSE (MIS-FNRS). All authors have equally contributed to the realisation of this work.

¹A matrix denoted by $\mathbf{M} \sim X^{d_1 \times d_2}$ has entries $M_{ij} \sim_{i.i.d.} X$ for a r.v. X .

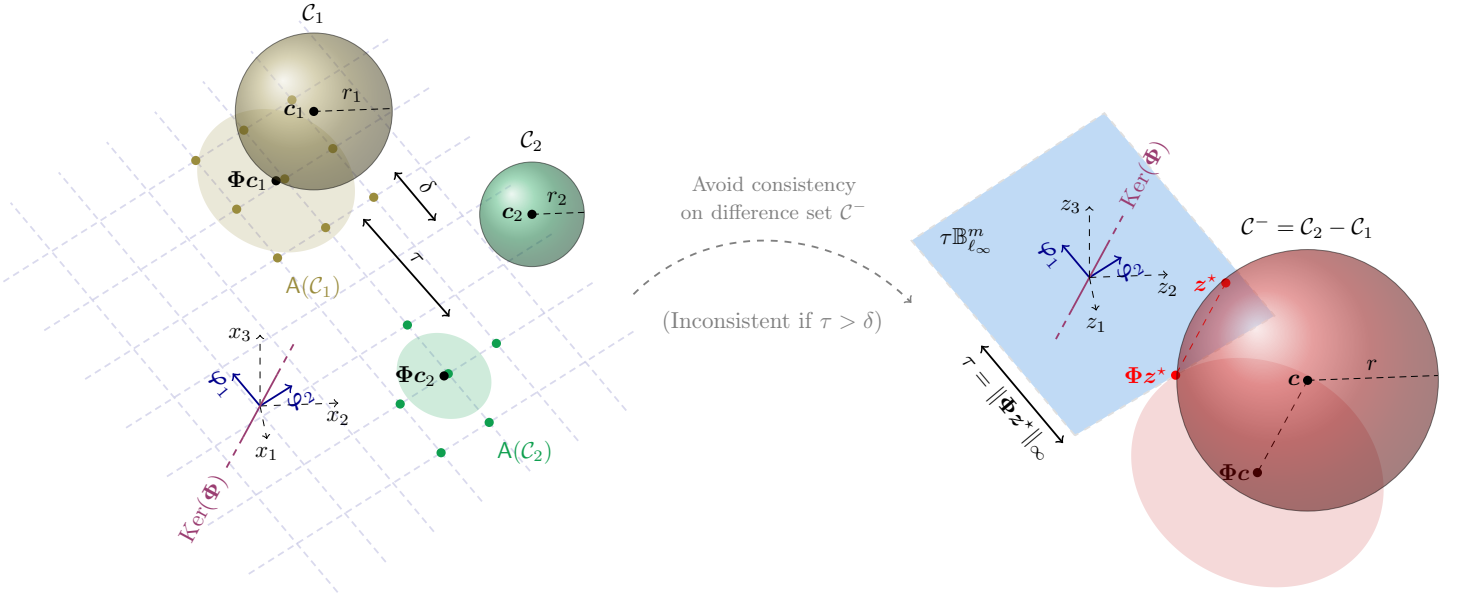


Figure 1. Geometrical intuition on the quantised eclipse problem for two disjoint ℓ_2 -balls and $n = 3$, $m = 2$: (left) C_1 and C_2 are projected on Φ , identified by the unit vectors φ_1, φ_2 ; on these directions, we construct the lattice $\delta\mathbb{Z}^m$, with a shift ξ of the origin due to dithering; the finite sets $A(C_1)$, $A(C_2)$ are also reported, along with the consistency margin τ ; (right) ensuring that $A(C_1) \cap A(C_2) = \emptyset$ requires that any $z \in C^-$ is so that its image under Φ has $\|\Phi z\|_\infty > \tau$; taking the smallest of such values on the difference set yields the consistency margin, which is $\tau = 0$ when $\text{Ker}(\Phi) \cap C^- \neq \emptyset$.

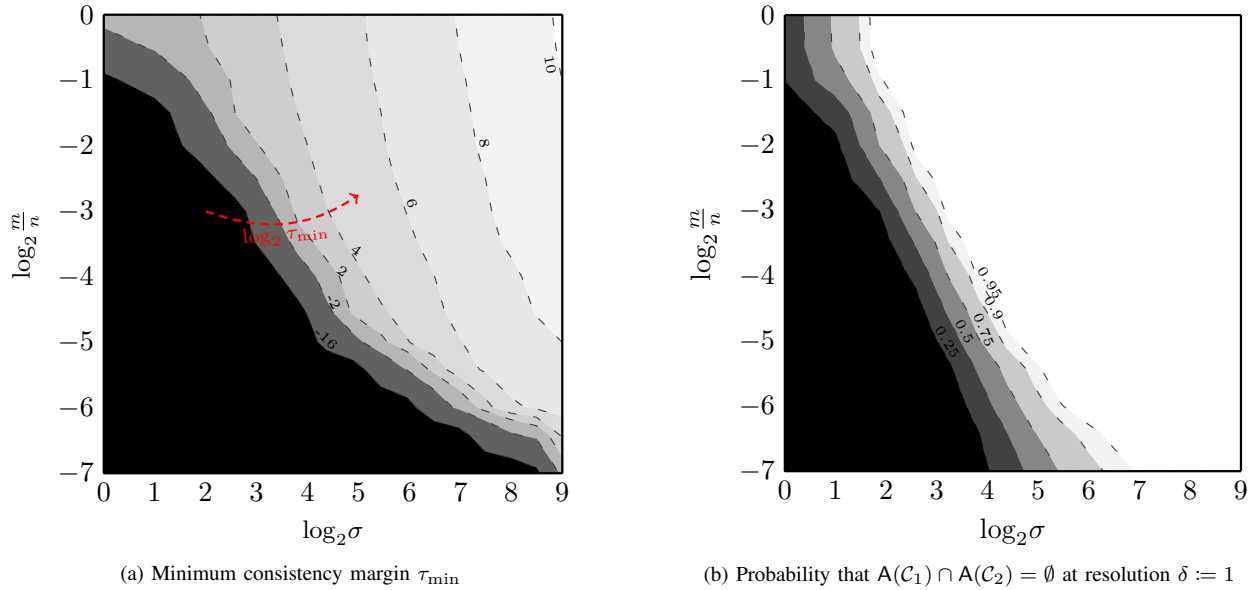


Figure 2. Empirical phase transitions of the quantised eclipse problem for the case of two disjoint ℓ_2 -balls; for several random instances of Φ and as a function of σ and the dimensionality reduction rate $\frac{m}{n}$, we report (a) the contours of $\log_2 \tau_{\min}$; (b) the contours of $\bar{p}_\delta = \mathbb{P}[\tau > \delta] \approx 1 - \eta$ for $\delta := 1$. In (a), the level curves of $\tau_{\min}$ are compatible, up to log factors, with the points $\{(\frac{m}{n}, \sigma) : m \approx \delta^2 n / \sigma^2\}$ deduced from (2) in Prop. 1. In (b), the level curves of $\bar{p}_\delta$ are also approximately aligned with the iso-probability curves $m - c \frac{\tau^2 + \delta^2}{\sigma^2} n \approx C \log(\frac{1}{\eta})$, also deduced from (2), once we set $\bar{p}_\delta \approx 1 - \eta \in \{0.25, 0.5, 0.75, 0.9, 0.95\}$ for some $C, c > 1$.

REFERENCES

- [1] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE transactions on information theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [2] L. Jacques and V. Cambareri, "Time for dithering: fast and quantized random embeddings via the restricted isometry property," *arXiv preprint arXiv:1607.00816*, 2016.
- [3] A. Moshtaghpour, L. Jacques, V. Cambareri, K. Degraux, and C. De Vleeschouwer, "Consistent Basis Pursuit for Signal and Matrix Estimates in Quantized Compressed Sensing," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 25–29, 2016.
- [4] A. S. Bandeira, D. G. Mixon, and B. Recht, "Compressive classification and the rare eclipse problem," *arXiv preprint arXiv:1404.3203*, 2014.
- [5] R. Giryes, G. Sapiro, and A. M. Bronstein, "Deep Neural Networks with Random Gaussian Weights: A Universal Classification Strategy?" *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3444–3457, 2015.
- [6] H. Reboredo, F. Renna, R. Calderbank, and M. R. Rodrigues, "Bounds on the Number of Measurements for Reliable Compressive Classification," *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 5778–5793, 2016.
- [7] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for Large-Scale Learning of Mixture Models," *arXiv preprint arXiv:1606.02838*, 2016.
- [8] N. Keriven, N. Tremblay, Y. Traonmilin, and R. Gribonval, "Compressive K-means," *arXiv preprint arXiv:1610.08738*, 2016.