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

Cambareri, Valerio ; Xu, Chunlei ; Jacques, Laurent

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.

Document type: Communication à un colloque (Conference Paper)

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

Valerio Cambareri, Chunlei Xu and Laurent Jacques

ISGGroup, 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 $x \in \mathbb{R}^n$, that is

$$y = A(x) := Q_A(\Phi x + \xi)$$  \hspace{1cm} (1)

with $\Phi \in \mathbb{R}^{m \times n}$ a random sensing matrix, $Q_A(\cdot) := \delta[\varepsilon, \delta]$ a uniform scalar quantiser of resolution $\delta > 0$ (applied component-wise), and the signature $y \in \delta \mathbb{Z}^m$. In (1), the dithering $\xi \sim \mathcal{N}(0, [\delta])$ 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 $x$ [2], [3]. Consequently, distance-based learning tasks preserve their accuracy if run on $A(x)$ rather than $x$, provided some requirements are met on $m$, $\delta$, the distribution of $\Phi$ and the “dimension” of $\mathbb{K}$ as measured, e.g., by its Gaussian mean width $w(\mathbb{K}) := \sup_{x \in \mathbb{K}} \|g x\|$ with $g \sim \mathcal{N}(0, 1)$ (see, e.g., [2]). In this context we aim to show that, given two classes described by some sets $C_1, C_2 \subset \mathbb{K}$: $C_1 \cap C_2 = \emptyset$ and $x \in C_1 \cup C_2 \subset \mathbb{K}$, classifying whether $x$ belongs to $C_1$ or $C_2$ is still possible from $y = A(x)$. For linear embeddings such as $y = \Phi x$, Bandeira et al. [4] approach the above classification problem as follows.

Problem 1 (Rare Eclipse Problem from [4]). Let $C_1, C_2 \subset \mathbb{R}$ : $\cap = \emptyset$ be closed convex sets, $\sim \mathcal{N}^{m \times n}(0, 1)$. Given $\eta \in (0, 1)$, find the smallest $m$ so that $p_0 := \mathbb{P}(C_1 \cap C_2 = \emptyset) \geq 1 - \eta$

Prob. 1 amounts to ensuring for all $x' \in C_1$, $x'' \in C_2$ that their images $\Phi x' \neq \Phi x''$. Using the difference set $\sim := C_1 - C_2 = \{x := x' - x'' : x \in C_1, x'' \in C_2\}$ we see the above problem equals

$$\mathbb{P}\{\forall z \in \sim, \Phi z \neq 0_m\} = 1 - \mathbb{P}\{\exists z \in \sim : \Phi z = 0_m\} \geq 1 - \eta$$

This requires a bound on the probability that the kernel of $\Phi$ “collides” with $\sim$, i.e., $\mathbb{P}(\ker(\Phi) \cap \sim = \emptyset) \leq \eta$, and [4] shows that $\eta$ is small if $m$ is large compared to the “dimension” of $\sim$ as measured by $w(\sim) := \mathbb{E}(\|\Phi x\|_2 \in [0, 1])$ with $\sim$ the cone generated by $C_1 - C_2$.

From this standpoint, extending such existing results on Prob. 1 to non-linear maps as (1) is non-trivial. Applying $\Phi$ to each closed convex set $C_1$ would produce two countable sets $\{A(C_1) = \{z \in \mathbb{R}^m : A(x) = \Phi (z)\} \subset \mathbb{Z}^m$, 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 $\mathbb{P}(A(C_1) \cap A(C_2) = \emptyset) \geq 1 - \eta$, i.e.,

$$p := \mathbb{P}\{\exists x' \in C_1, x'' \in C_2, A(x') = A(x'')\} \geq 1 - \eta$$

Note that the event in Prob. 2 requires $\mathbb{P}(\exists x' \in C_1, x'' \in C_2 : A(x') = A(x'')) \leq \eta$, 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 $\eta$ and the conditions on $m$. The QRIP establishes some conditions on $m$ that ensure $\frac{1}{\sqrt{2m}} \|A(x') - A(x'')\| \geq (c' - \epsilon) \|x' - x''\| - \epsilon \delta$ where $c'$ and $\epsilon$ are some constants, $c, \epsilon > 0$. Thus $A(x') \neq A(x'')$ if $H > 0$. 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 $\delta > 0$, $\eta \in (0, 1)$, $\sigma := \min_{z \in \mathbb{R}^n} \|z\|$ and the mapping $A$ defined in (1), if

$$m \geq \left(\frac{\sigma^2}{2} + 2 \delta \eta \right) \log \left(1 + \frac{2 \sigma^2}{\eta^2} + C \log \frac{1}{\eta}\right)$$

then $C > 1$ not depending on $m$ and $\eta$, then $\eta \leq 1 - \eta$.

Numerically testable but stronger conditions ensuring $p \leq 1 - \eta$ in Prob. 2 can be deduced as follows. We first note that if $\Phi z = 0_m$ for a given $\Phi$ and any $z \in \sim$, i.e., $\ker(\Phi) \cap \sim = \emptyset$, then $p_0 = 0$ for all $\delta > 0$ since then $\Phi x' + \xi = \Phi x'' + \xi$. Second, since $A(x') = A(x'')$ induces $\|\Phi z\|_\infty \leq \delta$ for $z := x' - x'' \in \sim$, proving $p_0 := \mathbb{P}(\Phi z \in \sim, \|\Phi z\|_\infty > \delta) \leq 1 - \eta$ will solve Prob. 2 since $p_0 \geq p_1$

We define accordingly a consistency margin $\tau := \inf_{z \in \sim} \mathbb{E} \|z\|_\infty$, with

$$z^* := \arg\min_{z \in \sim} \mathbb{E} \|z\|_\infty \text{ s.t. } z \in \sim := C_1 - C_2$$

Theorem 1 states (3) is clearly convex if $K$ and $\sim$ are convex. We anticipate that the construction of a certificate for this problem will provide a bound on $\tau > \delta$ when $C$ is known, and analyse an exemplary case afterwards.

II. NUMERICAL TEST FOR TWO DISJOINT $\ell_2$-BALLS

We consider the simple, yet broadly applicable convex case of two balls $C_1 = r_1 \mathbb{B}^2 + c$ and $C_2 = r_2 \mathbb{B}^2 + c$. With $r_1 = r_2$ and $c = r_1 + r_2$ (see Fig. 1). In this context $\|c\| = r + r$ and $\frac{c}{r} \leq \frac{1}{2}$ (Prop. 1). For $\mathbb{R}^2$ and $m = 2$, we are able to compute the consistency margin for each $\Phi$ on $\sim$, which is varied by fixing $r := r_2 + c$ and taking $\sigma := \|c\| - r = r_1 + r_2$. Then, we collect $\tau_m$, i.e., the smallest $r$ resulting from $2^m$ trials for each configuration (Fig. 2a), and also estimate on the same trials the probability $p_0 = \mathbb{P}(\tau > \Delta = 1)$ in Fig. 2b.

Fig. 2a reports several level curves of $\tau_m$. For each curve, the event $A(C_1) \cap A(C_2) = \emptyset$ holds if $\tau \approx \tau_m$. While this condition is necessary but not sufficient, these level curves are compatible with the points $\frac{c}{r} = \frac{1}{2}$ (up to log factors) induced by (1) in Prop. 1. In Fig. 2b displays a sharp phase transition in the contours of $\hat{p}_0$. Despite the fact that $p_0 \geq \hat{p}_0$, the contours are also approximately aligned with the iso-probability curves that can be deduced from (2), i.e., $m \geq \frac{c}{r} \frac{\delta^2}{\hat{c}^2} \approx C \log \frac{1}{\epsilon}$, with $\hat{p}_0 \approx 1 - \eta$ 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 $\mathbb{K}$ (e.g., selecting disjoint “clusters” of sparse signals).

1 A matrix denoted by $M \sim X^{d_1 \times d_2}$ has entries $M_{ij} \sim_i u.d.$ for $i, v, X$.

2 By uniformity of $\ker(\Phi)$, $\Phi \sim \mathcal{N}^{m \times n}(0, 1)$ over the Grassmannian at the origin, it is legitimate to fix a randomly drawn direction $e/\|e\|$ for the simulations.
Figure 1. Geometrical intuition on the quantised eclipse problem for two disjoint $\ell_2$-balls and $n = 3$, $m = 2$: (left) $C_1$ and $C_2$ are projected on $\Phi$, identified by the unit vectors $\varphi_1, \varphi_2$; on these directions, we construct the lattice $\delta Z^m$, with a shift $\xi$ of the origin due to dithering; the finite sets $A(C_1), A(C_2)$ are also reported, along with the consistency margin $\tau$: (right) ensuring that $A(C_1) \cap A(C_2) = \emptyset$ requires that any $z \in C^-$ is so that its image under $\Phi$ 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 $\ell_2$-balls; for several random instances of $\Phi$ and as a function of $\sigma$ and the dimensionality reduction rate $\frac{\log m}{\log n}$, we report (a) the contours of $\log_2 \tau_{\min}$; (b) the contours of $\tilde{\rho}_{\delta} = P[\tau > \delta] \approx 1 - \eta$ for $\delta := 1$. In (a), the level curves of $\tau_{\min}$ are compatible, up to log factors, with the points $\left\{ \left( \frac{\log m}{\log n}, \tau_{\min} \right) : m \approx \delta^2 n / \sigma^2 \right\}$ deduced from (2) in Prop. 1. In (b), the level curves of $\tilde{\rho}_{\delta}$ are also approximately aligned with the iso-probability curves $m - e^{- \frac{\tau_{\min}^2}{\delta^2}} n \approx C \log \left( \frac{1}{\eta} \right)$, also deduced from (2), once we set $\tilde{\rho}_{\delta} \approx 1 - \eta \in \{0.25, 0.5, 0.75, 0.9, 0.95\}$ for some $C, \epsilon > 1$.

REFERENCES