Supplementary Material for: Estimation and Quantization of Expected Persistence Diagrams

A. Proofs of Section 3

We let $\mu(f)$ denote the integral of some function $f : \Omega \to \mathbb{R}$ against the measure μ.

Lemma 3. *Let P be a probability measure on $\mathcal{M}^p$ such that $\mathbb{E}_P[\mathrm{Pers}_p(\mu)] < \infty$. Let $(\mu_n)_{n\geq 1}$ be a sequence of i.i.d. variables of law P and let $\overline{\mu}_n = \frac{1}{n}(\mu_1 + \cdots + \mu_n)$. Then,*

$$\mathrm{OT}_p(\overline{\mu}_n, \mathbf{E}(P)) \xrightarrow[n\to\infty]{} 0 \text{ almost surely.} \tag{A.1}$$

Proof of Lemma 3. By the strong law of large numbers applied to the function $\|\cdot - \partial\Omega\|^p$, we have $\mathrm{Pers}_p(\overline{\mu}_n) \to \mathrm{Pers}_p(\mathbf{E}(P))$ almost surely. Also, for any continuous function $f : \Omega \to \mathbb{R}$ with compact support, we have $\overline{\mu}_n(f) \to \mathbf{E}(P)(f)$ almost surely. This convergence also holds almost surely for any countable family $(f_i)_i$ of functions. Applying this result to a countable convergence-determining class for the vague convergence, we obtain that $(\overline{\mu}_n)_n$ converges vaguely towards $\mathbf{E}(P)$ almost surely. We conclude thanks to (Divol & Lacombe, 2020, Thm 3.7). □

Before proving Theorem 1, we give a general upper bound on the distance OT_p between two measures in $\mathcal{M}^p$. The bound is based on a classical multiscale approach to control a transportation distance between two measures, appearing for instance in (Singh & Póczos, 2018). Let $J \in \mathbb{N}$. For $k \geq 0$, let $B_k = \{x \in A_L,\ \|x - \partial\Omega\| \in (L2^{-(k+1)}, L2^{-k}]\}$. The sets $\{B_k\}_{k\geq 0}$ form a partition of A_L. We then consider a sequence of nested partitions $\{\mathcal{S}_{k,j}\}_{j=1}^J$ of B_k, where $\mathcal{S}_{k,j}$ is made of $N_{k,j}$ squares of side length $\varepsilon_{k,j} = L2^{-(k+1)}2^{-j}$. See also Figure 7. Let $\mu_{|B_k}$ be the measure μ restricted to B_k and $\mu_k = \frac{\mu_{|B_k}}{\mu(B_k)}$ be the conditional probability on B_k. If $\mu(B_k) = 0$, we let μ_k be any fixed measure, for instance the uniform distribution on B_k.

Lemma 4. *Let μ, ν be two measures in $\mathcal{M}^p$, supported on A_L. Then, for any $J \geq 0$, with $c_p = 2^{-p/2}(1 + 1/(2^p - 1))$,*

$$\mathrm{OT}_p^p(\mu,\nu) \leq 2^{p/2}L^p \sum_{k\geq 0} 2^{-kp}\Big(2^{-Jp}(\mu(B_k)\wedge\nu(B_k)) + c_p|\mu(B_k)-\nu(B_k)| + \sum_{\substack{1\leq j\leq J\\ S\in\mathcal{S}_{k,j-1}}} 2^{-jp}|\mu(S)-\nu(S)|\Big).$$

Proof. Denote by m_k the quantity $\mu(B_k)\wedge\nu(B_k)$. Let $\pi_k \in \Pi(\mu_k,\nu_k)$ be an optimal plan (in the sense of W_p) between the probability measures μ_k and ν_k. If $\mu(B_k) \leq \nu(B_k)$, then $\mu(B_k)\pi_k$ transports mass between $\mu_{|B_k}$ and $\frac{\mu(B_k)}{\nu(B_k)}\nu_{|B_k}$. We then build an admissible plan between $\frac{\mu(B_k)}{\nu(B_k)}\nu_{|B_k}$ and $\nu_{|B_k}$ by transporting $\left(1 - \frac{\mu(B_k)}{\nu(B_k)}\right)\nu_{|B_k}$ to the diagonal, with cost bounded by $\left(1 - \frac{\mu(B_k)}{\nu(B_k)}\right)\nu(B_k)(L2^{-k})^p$. Acting in a similar way if $\nu(B_k) \leq \mu(B_k)$, we can upper bound $\mathrm{OT}_p^p(\mu,\nu)$ by

$$\sum_{k\geq 0}\left(m_k W_p^p(\mu_k,\nu_k) + L^p 2^{-kp}|\mu(B_k)-\nu(B_k)|\right). \tag{A.2}$$

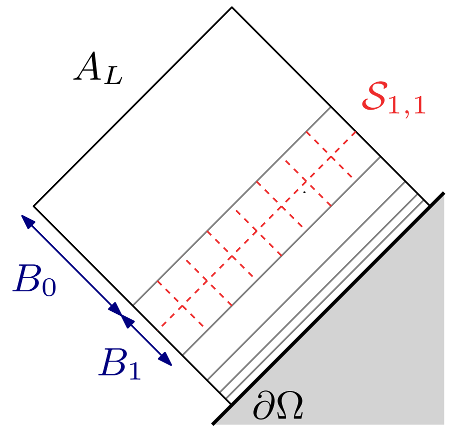

Figure 7. Partition of A_L used in the proof of Theorem 1

Lemma 6 in (Singh & Póczos, 2018) shows that

$$W_p^p(\mu_k,\nu_k) \leq 2^{p/2}L^p 2^{-(k+1)p}\Big(2^{-Jp} + \sum_{\substack{1\leq j\leq J\\ S\in\mathcal{S}_{k,j-1}}} 2^{-jp}|\mu_k(S)-\nu_k(S)|\Big). \tag{A.3}$$

Furthermore, one can check that for any $S \subset B_k$

$$m_k|\mu_k(S)-\nu_k(S)| \leq |\mu(S)-\nu(S)| + \frac{\nu(S)\wedge\mu(S)}{\mu(B_k)\vee\nu(B_k)}|\mu(B_k)-\nu(B_k)|.$$

By summing over $S \in \mathcal{S}_{k,j-1}$, we obtain that

$$m_k\sum_{S\in\mathcal{S}_{k,j-1}}|\mu_k(S)-\nu_k(S)| \leq |\mu(B_k)-\nu(B_k)| + \sum_{S\in\mathcal{S}_{k,j-1}}|\mu(S)-\nu(S)|. \tag{A.4}$$

Using $\sum_{j=1}^J 2^{-pj} \leq 2^{-p}/(1-2^{-p})$, and putting together inequalities (A.2), (A.3) and (A.4), one obtains the inequality of Lemma 4. □

Before proving Theorem 1, we state a useful inequality. Let $\mu \in \mathcal{M}^q_{M,L}$ and let $B \subset \Omega$ be at distance ℓ from the diagonal $\partial\Omega$. Then,

$$\mu(B) = \int_B \frac{\|x-\partial\Omega\|^q}{\|x-\partial\Omega\|^q}\mathrm{d}\mu(x) \leq M\ell^{-q}. \tag{A.5}$$

Proof of Theorem 1. Consider a distribution $P \in \mathcal{P}^q_{M,L}$. Remark first that for any measure $\mu \in \mathcal{M}^q_{M,L}$, we have $\mu(B_k) \leq ML^{-q}2^{kq}$ one by (A.5). Let μ be a random persistence measure of law P and $\overline{\mu}_n$ be the empirical EPD associated to a n-sample of law P. By the Cauchy-Schwartz inequality, given a Borel set $A \subset \Omega$, we have

$$\mathbb{E}|\overline{\mu}_n(A) - \mathbf{E}(P)(A)| \leq \sqrt{\frac{\mathbb{E}[\mu(A)^2]}{n}}. \tag{A.6}$$

The Cauchy-Schwartz inequality also yields, as $|\mathcal{S}_{k,j-1}| = 2^{k+1}4^{j-1}$,

$$\begin{aligned}\sum_{S\in\mathcal{S}_{k,j-1}} \mathbb{E}|\hat{\mu}_n(S) - \mathbf{E}(P)(S)| &\leq \sum_{S\in\mathcal{S}_{k,j-1}} \sqrt{\frac{\mathbb{E}[\mu(S)^2]}{n}} \\ &\leq \sqrt{\frac{\mathbb{E}\left[\sum_{S\in\mathcal{S}_{k,j-1}}\mu(S)^2\right]}{n}|\mathcal{S}_{k,j-1}|} \\ &\leq \sqrt{\frac{\mathbb{E}\left[\mu(B_k)^2\right]}{n}|\mathcal{S}_{k,j-1}|} \leq \frac{ML^{-q}2^{kq}}{\sqrt{n}}2^{\frac{k+1}{2}}2^{j-1}.\end{aligned}$$

Note also that $\sum_{S\in\mathcal{S}_{k,j-1}} \mathbb{E}|\hat{\mu}_n(S) - \mathbf{E}(P)(S)| \leq 2\mathbf{E}(P)(B_k) \leq 2ML^{-q}2^{kq}$ and that $\overline{\mu}_n(B_k) \wedge \mathbf{E}(P)(B_k) \leq ML^{-q}2^{kq}$. By using those three previous inequalities, Lemma 4 and inequality (A.6), we obtain that $\mathbb{E}[\mathrm{OT}_p^p(\overline{\mu}_n, \mathbf{E}(P))]$ is smaller than

$$\begin{aligned}&2^{p/2}ML^{p-q}\sum_{k\geq 0}2^{-kp}\Big(2^{-Jp}2^{kq} + \frac{c_p}{\sqrt{n}}2^{kq} \\ &\qquad + \sum_{j=1}^{J}2^{-jp}2^{kq}\left(2 \wedge \frac{2^{\frac{k+1}{2}}2^{j-1}}{\sqrt{n}}\right)\Big) \\ &\leq c_{p,q}ML^{p-q}\Big(2^{-Jp} + \frac{1}{\sqrt{n}} + U\Big),\end{aligned}$$

where $U = \sum_{k\geq 0}\sum_{j=1}^{J}2^{k(q-p)}2^{-jp}\left(1 \wedge \frac{2^{\frac{k}{2}}2^j}{\sqrt{n}}\right)$. To bound U, we remark that if $k \geq \log_2(n)$, then the minimum in the definition of U is equal to 1. Therefore, letting $b_J = 1$ if $p > 1$ and $b_J = J$ if $p = 1$, we find that U is smaller than

$$\begin{aligned}&\sum_{k=0}^{\log_2(n)}\sum_{j=1}^{J}\frac{2^{k(q-p+1/2)}2^{(1-p)j}}{\sqrt{n}} + \sum_{k\geq\log_2(n)}\sum_{j=1}^{J}2^{-kp}2^{-jp} \\ &\leq c_p b_J \sum_{k<\log_2(n)}\frac{2^{k(q+1/2-p)}}{\sqrt{n}} + c_p n^{-p} \\ &\leq c_{p,q}b_J(n^{-1/2} \vee n^{q-p}).\end{aligned}$$

Eventually, if $p > 1$, we may set $J = +\infty$ and obtain a bound of order $ML^{p-q}(n^{-1/2} + n^{q-p})$. If $p = 1$, we choose $J = (q-p)(\log n)/(2p)$ to obtain a rate of order $n^{-1/2} + n^{q-p}\log n$. □

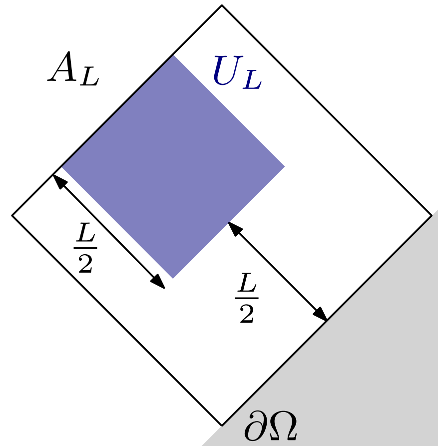

Figure 8. In the box U_L, the distance ρ is equal to the Euclidean distance.

Proof of Theorem 2. As $\mathcal{P}^{q,s}_{L,M,T} \subset \mathcal{P}^q_{L,M}$, we have $\mathcal{R}_n(\mathcal{P}^q_{L,M}) \geq \mathcal{R}_n(\mathcal{P}^{q,s}_{L,M,T})$. Therefore, Theorem 3, whose proof is found below, directly implies Theorem 2. □

Proof of Theorem 3. We first consider the case $q = 0$. If μ, ν are two measures on Ω of mass smaller than M, then $\mathrm{OT}_p(\mu,\nu) = W_{p,\rho}(\Phi(\mu), \Phi(\nu))$ (Divol & Lacombe, 2020, Prop. 3.15), where ρ is the distance on $\tilde{\Omega} := \Omega \cup \{\partial\Omega\}$ defined by $\forall x, y \in \tilde{\Omega}$,

$$\rho(x,y) = \min(\|x-y\|, d(x,\partial\Omega) + d(y,\partial\Omega))$$

and $\Phi(\mu) = \mu + (2M - |\mu|)\delta_{\partial\Omega}$. Remark that $\rho(x,y) = \|x-y\|$ if $x, y \in U_L$, where $U_L \subset A_L$ is any ℓ_1-ball of radius $L/\sqrt{8}$ at distance $L/2$ from the diagonal, see Figure 8. As Φ is a bijection, the minimax rates for the estimation of $\mathbf{E}(P)$ is therefore equal to

$$\inf_{\Phi(\hat{\mu}_n)} \sup_{P\in\mathcal{P}^{0,s}_{L,M,T}} \mathbb{E}[W^p_{p,\rho}(\Phi(\hat{\mu}_n), \Phi(\mathbf{E}(P)))].$$

Let $\mathcal{Q}$ be the set of probability measures on U_L whose densities belong to $B^s_{p',q'}$ with associated norm smaller than T/M. Then, $\mathcal{P}^{0,s}_{M,L,T}$ contains in particular the set of all distributions P for which $\mu \sim P$ satisfies $\Phi(\mu) = M\delta_x$ and x is sampled according to some law $\tau \in \mathcal{Q}$. For such a distribution P, one has $\Phi(\mathbf{E}(P)) = M\tau$, so that the minimax rate is larger than

$$\inf_{\hat{a}_n}\sup_{\tau\in\mathcal{Q}}\mathbb{E}[W^p_p(\hat{a}_n, M\tau)],$$

where the infimum is taken on all measurable functions based on K observations of the form $M\delta_{x_i}$ with $x_1, \ldots, x_n$ a n-sample of law $\tau \in \mathcal{Q}$. Hence, we have shown that the minimax rate for the estimation of $\mathbf{E}(P)$ with respect to OT_p is larger up to a factor M than the minimax rate for the estimation of $\tau \in \mathcal{Q}$ given n i.i.d. observations of law τ. As the minimax rate for this problem is known to be larger than $L^p/\sqrt{n}$ (Weed & Berthet, 2019, Thm. 5), we obtain the conclusion in the case $q = 0$.

For the general case $q > 0$, we remark that if $M' = ML^{-q}$ then $\mathcal{P}^{0,s}_{M',L}$ is included in $\mathcal{P}^{q,s}_{M,L,T}$. In particular, the minimax rate on $\mathcal{P}^{q,s}_{M,L,T}$ is larger than the minimax rate on $\mathcal{P}^{0,s}_{M',L,T}$, which is larger than $c\frac{M'L^p}{\sqrt{n}} = c\frac{ML^{p-q}}{\sqrt{n}}$ for some constant $c > 0$. □

Remark 3 (Case $p = \infty$)**.** *It can be shown that for $p = \infty$, the minimax rate is larger than $c_a n^{-a}$, $\forall a > 0$. This is a consequence of an inequality between the OT_∞ distance and the distance between the support of the measures, for which minimax rates are known (Hardle et al., 1995). This means that no reasonable estimator exists on $\mathcal{P}L, M^\infty$: some additional conditions should be added, while standard assumptions in the support estimation literature seem artificial in our context (as in Remark 1).*

B. Delayed proofs from Section 4.1

Proof of Lemma 2. Fix a codebook $\mathbf{c} = (c_1 \dots c_k)$. Let $T_\mathbf{c} : x \mapsto c_j$ if $x \in V_j(\mathbf{c})$ $(1 \leq j \leq k)$ and $\mathrm{proj}_{\partial\Omega}(x)$ if $x \in V_{k+1}(\mathbf{c})$, where $\mathrm{proj}_{\partial\Omega}(x)$ denotes the orthogonal projection of a point $x \in \Omega$ on the diagonal $\partial\Omega$. Let π be the pushforward of μ by the map $x \mapsto (x, T_\mathbf{c}(x))$, extended on $\overline{\Omega} \times \overline{\Omega}$ by $\pi(U, \overline{\Omega}) = 0$ for $U \subset \partial\Omega$ (intuitively, π pushes the mass of μ on their nearest neighbor in $\{c_1 \dots c_{k+1}\}$). One has, for $A, B \subset \Omega$, $\pi(A, \overline{\Omega}) = \mu((\mathrm{id}, T_c)^{-1}(A, \overline{\Omega})) = \mu(A)$, and $\pi(\overline{\Omega}, B) = \mu(T_c^{-1}(B)) = \sum_j \mu(V_j(\mathbf{c}))\mathbf{1}\{c_j \in B\}$, that is π is an admissible between the measures μ and $\sum_j \mu(V_j(\mathbf{c}))\delta_{c_j}$. Hence,

$$\mathrm{OT}_p^p\left(\mu, \sum_j \mu(V_j(\mathbf{c}))\delta_{c_j}\right) \leq \int_\Omega \min_{1\leq j\leq k+1} \|x - c_j\|^p \mathrm{d}\mu(x).$$

Let $(m_1 \dots m_k)$ be a vector of non-negative weights, let $\nu = \sum_{j=1}^k m_j \delta_{c_j}$, and π be an admissible transport plan between μ and ν. One has

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} \|x - y\|^p \mathrm{d}\pi(x, y) &= \sum_{j=1}^{k+1} \int_{\overline{\Omega}} \|x - c_j\|^p \mathrm{d}\pi(x, c_j) \\ &\geq \sum_{j=1}^{k+1} \int_{\overline{\Omega}} \min_{j'} \|x - c_{j'}\|^p \mathrm{d}\pi(x, c_j) \\ &\geq \int_{\overline{\Omega}} \min_{j'} \|x - c_{j'}\|^p \mathrm{d}\mu(x) \\ &\geq \mathrm{OT}_p^p\left(\mu, \sum_{j=1}^k \mu(V_j(\mathbf{c}))\delta_{c_j}\right).\end{aligned}$$

Taking the infimum over π gives the conclusion. □

We now turn to the proof of Proposition 4. For technical reasons, we extend the function R_k to $\overline{\Omega}^k$, by noting that if $c_j \in \partial\Omega$, then the Voronoï cell $V_j(\mathbf{c})$ is empty by definition, see (4.1).

Lemma 5. *Let $\mathbf{c} \in \overline{\Omega}^k$ be such that there exists $1 \leq j \leq k$ with $\mu(V_j(\mathbf{c}^*)) = 0$. Then, $R_k(\mathbf{c}) > R_k^*$.*

In particular, if two centroids of a codebook $\mathbf{c}$ are equal or if a centroid c_j of $\mathbf{c}$ belongs to $\partial\Omega$, then the condition of the above lemma is satisfied, so that the $\mathbf{c}$ cannot be optimal. This proves the second part of Proposition 4.

Proof of Lemma 5. Let $\mathbf{c} = (c_1, \dots, c_k) \in \overline{\Omega}^k$. Assume without loss of generality that $\mu(V_1(\mathbf{c})) = 0$. Let $\mathbf{c}_0 = (c_2, \dots, c_k) \in \overline{\Omega}^{k-1}$ (that is, $\mathbf{c}$ where we removed the first centroid). Assume first that $\mu(V_{k+1}(\mathbf{c})) > 0$, that is there is some mass transported onto the diagonal. Consider a compact subset $A \subset V_{k+1}(\mathbf{c})$ such that $\mu(A) > 0$ and the diameter $\mathrm{diam}(A)$ of A is smaller than the distance $d(A, \partial\Omega)$ between A and $\partial\Omega$. Let $c' \in A$ and observe that, for $x \in A$, $\|x - c'\| < \|x - \partial\Omega\|$. Therefore,

$$\int_A \|x - c'\|^p \mathrm{d}\mu(x) < \int_A \|x - \partial\Omega\|^p \mathrm{d}\mu(x).$$

Consider the measure $\nu = \hat{\mu}(\mathbf{c}_0) + \mu(A)\delta_{c'}$. Then

$$\begin{aligned}\mathrm{OT}_p^p(\nu, \mu) \leq & \sum_{j=1}^k \int_{V_j(\mathbf{c})} \|x - c_j\|^p \mathrm{d}\mu(x) \\ & + \int_{V_{k+1}(\mathbf{c})\setminus A} \|x - \partial\Omega\|^p \mathrm{d}\mu(x) + \int_A \|x - c'\|^p \mathrm{d}\mu(A) \\ < & R_k(\mathbf{c}),\end{aligned}$$

thus $\mathbf{c}$ cannot be optimal. We can thus assume that $\mu(V_{k+1}(\mathbf{c})) = 0$, in which case we can reproduce the proof of (Graf & Luschgy, 2007, Thm 4.1), which gives that $\mathbf{c}$ cannot be optimal either in that case, yielding the conclusion. □

Lemma 6. *R_k is continuous.*

Proof of Lemma 6. For a given $x \in \overline{\Omega}$, the map $\mathbf{c} \mapsto \min_i \|x - c_i\|^p$ is continuous and upper bounded by $\|x - \partial\Omega\|^p$. Thus, R_k is continuous by dominated convergence as we have finite Pers_p. □

Lemma 7. *Let $0 \leq \lambda < R_{k-1}^*$. Then, the set $\{\mathbf{c} \in \overline{\Omega}^k,\ R_k(\mathbf{c}) \leq \lambda\}$ is compact.*

Proof of Lemma 7. Fix $\lambda < R_{k-1}^*$. The set is closed by continuity of R_k, so that it suffices to show that it is bounded. Let $\mathbf{c}$ be such that $R_k(\mathbf{c}) \leq \lambda$. Pick L such that $\int_{A_L} \|x - \partial\Omega\|^p \mathrm{d}\mu(x) \geq \lambda$ and $\int_{A_L^c} \|x - \partial\Omega\|^p \mathrm{d}\mu(x) < R_{k-1}^* - \lambda$.

Such a L exists since $\int_\Omega \|x - \partial\Omega\|^p \mathrm{d}\mu(x) = \mathrm{Pers}_p(\mu) = R_0^* \geq R_{k-1}^*$. Then, all the c_js must be in A_{2L}. Indeed, assume without loss of generality that $c_1 \in A_{2L}^c$. Then $V_1(\mathbf{c}) \subset A_L^c$, as any point in A_L is closer to the diagonal than to c_1. Therefore,

$$\begin{aligned}
R_{k-1}^* \leq & \sum_{j=2}^{k+1} \int_{V_j(\mathbf{c})} \|x - c_j\|^p \mathrm{d}\mu(x) \\
& + \int_{V_1(\mathbf{c})} \min_{j \in \{2 \dots k+1\}} \|x - c_j\|^p \mathrm{d}\mu(x) \\
\leq & R_k(\mathbf{c}) + \int_{V_1(\mathbf{c})} \|x - \partial\Omega\|^p \mathrm{d}\mu(x) \\
\leq & R_k(\mathbf{c}) + \int_{A_L^c} \|x - \partial\Omega\|^p \mathrm{d}\mu(x) \\
< & \lambda + R_{k-1}^* - \lambda = R_{k-1}^*,
\end{aligned}$$

leading to a contradiction. □

Proof of Proposition 4. We show by recursion on $0 \leq m \leq k$ that $R_m^* < R_{m-1}^*$ and that $\mathbf{C}_m$ is a non-empty compact set (with the convention $R_{-1}^* = +\infty$. The initialization holds as $R_0^* = \mathrm{Pers}_p(\mu) < +\infty$ with the empty codebook being optimal. We now prove the induction step. Let $\mathbf{c} = (c_1, \dots, c_{m-1}) \in \mathbf{C}_{m-1}$. Consider $\mathbf{c}' = (c_1, c_1, c_2, \dots, c_{m-1})$. Then, $\mu(V_1(\mathbf{c}')) = 0$, so that $R_{m-1}^* = R_{m-1}(\mathbf{c}) = R_m(\mathbf{c}') > R_m^*$ by Lemma 5. Furthermore, pick $\lambda \in (R_m^*, R_{m-1}^*)$. Then, R_m^* is equal to the infimum of R_m on the set $\{\mathbf{c} \in \overline{\Omega}^k,\ R_m(\mathbf{c}) \leq \lambda\}$, which is compact according to Lemma 7. As the function R_k is continuous, the set of minimizers $\mathbf{C}_m$ is a non-empty compact set, concluding the induction step. □

Proof of Corollary 1. The quantities being minimized in the definitions of $D_{\min}$ and $m_{\min}$ are both continuous functions of $\mathbf{c}^*$. As the set $\mathbf{C}_k$ is compact, the minima are attained, and cannot be equal to 0 according to Proposition 4. □

C. Proof of Theorem 5.

In the following, we fix a distribution P supported on $\mathcal{M}_{L,M}^p$ and we consider $\mathbf{c}^*$ be an optimal codebook of $\mathbf{E}(P)$. The different constants encountered in this section all depend on the parameters $p, L, M, k, D_{\min}$ and $m_{\min}$. In particular, we introduce the quantity

$$m_{\max} := \sup_{\mu \in \mathcal{M}_{L,M}^p} \sup_{1 \leq j \leq k} \mu(V_j(\mathbf{c}^*)).$$

Note that $m_{\max} \leq \frac{2^p M}{D_{\min}^p}$ as $\int_{V_j(\mathbf{c}^*)} \mathrm{d}\mu(x) \leq \frac{2^p}{D_{\min}^p} \int_{V_j(\mathbf{c}^*)} \|x - \partial\Omega\|^p \mathrm{d}\mu(x)$.

The proof of Theorem 5 follows the proof of (Chazal et al., 2021, Thm. 5). As a first step, we show that it is enough to prove the following lemma, which relates the loss of $\mathbf{c}^{(t)}$ and the loss of $\mathbf{c}^{(t+1)}$.

Lemma 8. *There exists $R_0 > 0$ such that, if $\|c_j^{(0)} - c_j^*\| \leq R_0$ for $1 \leq j \leq k$, then*

$$\mathbb{E}\|\mathbf{c}^{(t+1)} - \mathbf{c}^*\|^2 \leq \left(1 - \frac{C_0}{t+1}\right) \mathbb{E}\|\mathbf{c}^{(t)} - \mathbf{c}^*\|^2 + \frac{C_1}{(t+1)^2},$$

for some constants $C_0 > 1$, $C_1 > 0$.

Proof of Theorem 5. From Lemma 8, we show by induction that $u_t := \mathbb{E}\|\mathbf{c}^{(t)} - \mathbf{c}^*\|^2$ satisfies $u_t \leq \frac{\alpha}{t+1}$ for $\alpha = C_1/(C_0 - 1)$. This concludes the proof as T is of order $n/\log(n)$. The initialization holds by assumption as long as $R_0 \leq \alpha$, whereas we have by induction

$$\begin{aligned}
u_{t+1} &\leq \left(1 - \frac{C_0}{t+1}\right) \frac{\alpha}{t+1} + \frac{C_1}{(t+1)^2} \\
&\leq \frac{\alpha}{(t+1)^2} (t + 1 - C_0 + C_1/\alpha) = \frac{\alpha t}{(t+1)^2},
\end{aligned}$$

which is smaller than $\alpha/(t+2)$. □

The proof of Lemma 8 is a close adaptation of (Chazal et al., 2021, Lemma 21). The proof of the latter contains tedious computations (that we do not reproduce here) which can be adapted *mutatis mutandis* to our setting once the two following key results are shown. Given a codebook $\mathbf{c}$, we let $p_j(\mathbf{c}) = \mathbf{E}(P)(V_j(\mathbf{c}))$ and similarly, given a n-sample $\mu_1, \dots, \mu_n$ of law P, we let $\hat{p}_j(\mathbf{c}) = \overline{\mu}_n(V_j(\mathbf{c}))$. Note that if $\|\mathbf{c} - \mathbf{c}^*\|$ is small enough, one has $p_j(\mathbf{c}) \leq 2m_{\max}$. Also, we let $w_p(\mathbf{c}, \mu)_j := \mu(V_j(\mathbf{c})) v_p(\mathbf{c}, \mu)_j$ for $\mu \in \mathcal{M}^p$ and $1 \leq j \leq k$. Recall that we assume that the EPD $\mathbf{E}(P)$ satisfies the margin condition (Definition 2) with parameters λ and r_0 around the optimal codebook $\mathbf{c}^*$.

Lemma 9 (Lemma 22 in (Chazal et al., 2021))**.** *Let R_0 be small enough respect to $r_0 D_{\min}^2/L^2$ and let $\mathbf{c}$ be such that $\|\mathbf{c} - \mathbf{c}^*\| \leq R_0$. Then, we have*

$$\sum_{j=1}^{k} |p_j(\mathbf{c}) - p_j(\mathbf{c}^*)| \leq 2\lambda r_0,$$

and

$$\|w_2(\mathbf{c}, \mathbf{E}(P)) - w_2(\mathbf{c}^*, \mathbf{E}(P))\| \leq 7\sqrt{2}\lambda \frac{L^3}{D_{\min}^2} \|\mathbf{c} - \mathbf{c}^*\|.$$

As $w_2(\mathbf{c}^*, \mathbf{E}(P))_j = p_j(\mathbf{c}^*) \mathbf{c}_j^*$, Lemma 9 indicates that the application $w_2(\cdot, \mathbf{E}(P))$ is Lipschitz continuous around an optimal codebook $\mathbf{c}^*$, a key property to show the convergence of the sequence $(\mathbf{c}^{(t)})_t$.

Lemma 10 (Lemma 24 in (Chazal et al., 2021))**.** *Let* $\mathbf{c}$ *be a codebook such that* $\hat{p}_j(\mathbf{c}) \leq 2m_{\max}$ *(which is always possible if* $\|\mathbf{c}-\mathbf{c}^*\|$ *is small enough). Then, with probability larger than* $1-2ke^{-x}$*, we have, for all* $1 \leq j \leq k$,

$$|\hat{p}_j(\mathbf{c}) - p_j(\mathbf{c})| \leq \sqrt{\frac{4m_{\max}p_j(\mathbf{c})x}{n}} + \frac{2m_{\max}x}{3n}. \quad \text{(C.1)}$$

Moreover, with probability larger than $1-e^{-x}$*, we have*

$$\|w_2(\mathbf{c},\overline{\mu}_n) - w_2(\mathbf{c},\mathbf{E}(P))\| \leq 2m_{\max}L\sqrt{\frac{2k}{n}}\left(1+\sqrt{\frac{x}{2}}\right). \quad \text{(C.2)}$$

The proof of this lemma follows from standard concentration inequalities.

Proof of Lemma 10. Equation (C.1) follows from Bernstein inequality applied to the real-valued random variable $0 \leq \hat{p}_j(\mathbf{c}) \leq 2m_{\max}$, with variance bounded by $\mathbb{E}[\mu(V_j(\mathbf{c}))^2]/n \leq m_{\max}p_j(\mathbf{c})/n$.

For equation (C.2), we introduce the function $f_j : x \mapsto x\mathbf{1}\{x \in V_j(\mathbf{c})\}$, so that $w_2(\mathbf{c},\mu)_j = \mu(f_j)$, the integral of f_j against μ. We have $w_2(\mathbf{c},\mu_n)_j - w_2(\mathbf{c},\mathbf{E}(P))_j = n^{-1}\sum_{i=1}^n(\mu_i(f_j) - \mathbf{E}(P)(f_j))$. Note that $\|\mu_i(f_j)\| \leq \sqrt{2}L \cdot 2m_{\max}$. We write

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n(\mu_i(f_j) - \mathbf{E}(P)(f_j))_j\right\| \\
&\leq \sqrt{\mathbb{E}\left\|\frac{1}{n}\sum_{i=1}^n(\mu_i(f_j) - \mathbf{E}(P)(f_j))_j\right\|^2} \\
&\leq \sqrt{\frac{1}{n}\mathbb{E}\,\|(\mu_1(f_j))_j\|^2} \leq 2\sqrt{\frac{k}{n}}\sqrt{2}Lm_{\max}.
\end{aligned}$$

Also, note that $F(\mu_1,\dots,\mu_n) = \|w_2(\mathbf{c},\mu_n) - w_2(\mathbf{c},\mathbf{E}(P))\|$ satisfies a bounded difference condition of parameter $4\sqrt{2}Lm_{\max}$ (Boucheron et al., 2013, Sec. 6.1). A bounded difference inequality (Boucheron et al., 2013, Thm. 6.2) yields the result. □

The proof of Lemma 9 relies on the following lemma, that essentially tells that the area of misclassified points when using a codebook $\mathbf{c}$ instead of an optimal one $\mathbf{c}^*$ can be controlled linearly in terms of $\|\mathbf{c}^* - \mathbf{c}\|$. Note that this result is well-known when boundaries between the cells are hyperplanes (as it is the case in standard quantization), it remains to treat the case when the boundary is a parabola. Let $d(x,A)$ be the distance from a point $x \in \Omega$ to $A \subset \Omega$.

Lemma 11. *Let* $\mathbf{c}^*$ *be an optimal codebook, and* $\mathbf{c} \in A_L^k$*. Let* $x \in A_L$ *and* $1 \leq j \leq k$*. Assume that* $x \in V_j(\mathbf{c}^*) \cap V_{k+1}(\mathbf{c})$*. Then,* $d(x, \partial V_j(\mathbf{c}^*)) \leq \frac{7L^2}{2D_{\min}^2}\|\mathbf{c}^*-\mathbf{c}\|$*. Symmetrically, if* $x \in V_{k+1}(\mathbf{c}^*) \cap V_j(\mathbf{c})$*, one has* $d(x,\partial V_{k+1}(\mathbf{c}^*)) \leq \frac{7L^2}{2D_{\min}^2}\|\mathbf{c}^* - \mathbf{c}\|$.

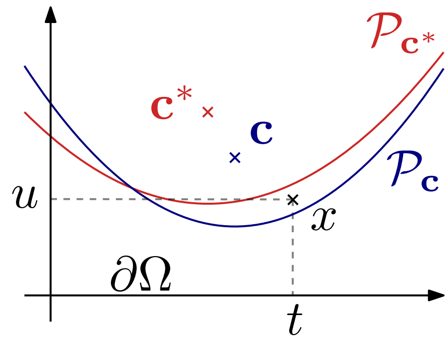

Figure 9. Illustration of the proof of Lemma 9

Proof of Lemma 11. For convenience, we write in this proof the coordinates of points in the basis $(\partial\Omega, \partial\Omega^\perp)$, that $x \in \Omega$ will have coordinates (a,b) where a is the projection of x on $\partial\Omega$ and $b = \|x - \partial\Omega\|$. Also, given $y=(a,b) \in \Omega$, we let $\mathcal{P}_y$ be the parabola with focus y and directrix $\partial\Omega$. To put it another way, if $y=(a,b)$, then $\mathcal{P}_y$ is the image of $\partial\Omega$ by the map

$$f(a,b,\cdot) : t \mapsto \frac{(t-a)^2}{2b} + \frac{b}{2}.$$

One can check that for all $t \in [-L/2, L/2]$, if $b = \|y - \partial\Omega\| \geq D_{\min}$, we have $\left|\frac{\partial f}{\partial a}\right| \leq \frac{L}{D_{\min}}$ and $\left|\frac{\partial f}{\partial b}\right| \leq \frac{1}{2} + \frac{(t-a)^2}{b}\frac{1}{b} \leq \frac{1}{2} + \frac{2L^2}{D_{\min}^2}$.

Let $c_j^* = (a^*, b^*)$ and $c_j = (a,b)$. Let $x = (t,u) \in V_j(\mathbf{c}^*) \cap V_{k+1}(\mathbf{c})$. Then, $u \geq f(a^*,b^*,t)$, whereas $u \leq f(a,b,t)$. The distance $d(x, \partial V_j(\mathbf{c}^*))$ is smaller than $u - f(a^*,b^*,t)$

$$\begin{aligned}
&u - f(a^*,b^*,t) \leq f(a,b,t) - f(a^*,b^*,t) \\
&\leq |f(a^*,b^*,t) - f(a,b^*,t)| + |f(a,b^*,t) - f(a,b,t)| \\
&\leq \int_{a\wedge a^*}^{a\vee a^*}\left|\frac{\partial f}{\partial a}(\alpha,b^*,t)\right|\mathrm{d}\alpha + \int_{b\wedge b^*}^{b\vee b^*}\left|\frac{\partial f}{\partial b}(a,\beta,t)\right|\mathrm{d}\beta \\
&\leq \frac{L}{D_{\min}}|a-a^*| + \left(\frac{1}{2} + \frac{2L^2}{D_{\min}^2}\right)|b-b^*| \\
&\leq \left(\frac{1}{2} + \frac{L}{D_{\min}} + \frac{2L^2}{D_{\min}^2}\right)\|\mathbf{c}-\mathbf{c}^*\| \leq \frac{7}{2}\frac{L^2}{D_{\min}^2}\|\mathbf{c}-\mathbf{c}^*\|,
\end{aligned}$$

which proves the claim. □

Proof of Lemma 9. This proof is inspired from (Levrard et al., 2015, Appendix A.3). Let us prove the first point. One has, with $t = \frac{7L^2}{2D_{\min}^2}\|\mathbf{c}-\mathbf{c}^*\| \leq r_0$,

$$\begin{aligned}
\sum_{j=1}^k |p_j(\mathbf{c}) - p_j(\mathbf{c}^*)| &= \sum_{j=1}^k |\mathbf{E}(P)(V_j(\mathbf{c})) - \mathbf{E}(P)(V_j(\mathbf{c}^*)| \\
&\leq 2\sum_j\sum_{j'\neq j}\mathbf{E}(P)(V_j(\mathbf{c}) \cap V_{j'}(\mathbf{c}^*)) \\
&\leq 2\mathbf{E}(P)[N(\mathbf{c}^*)^t] \leq 2\lambda t \leq 2\lambda r_0.
\end{aligned}$$

where we applied Lemma 11 and the margin condition. To prove the second inequality, remark that $w_2(\mathbf{c}, \mathbf{E}(P))_j = \int_{V_j(\mathbf{c})} x \mathrm{d}\mathbf{E}(P)(x)$. Therefore,

$$\begin{aligned}
&\|w_2(\mathbf{c}, \mathbf{E}(P)) - w_2(\mathbf{c}^*, \mathbf{E}(P))\| \\
&\leq \sum_{j=1}^{k} \|w_2(\mathbf{c}, \mathbf{E}(P))_j - w_2(\mathbf{c}^*, \mathbf{E}(P))_j\| \\
&\leq \sum_{j=1}^{k} \left\| \int_{V_j(\mathbf{c})} x \mathrm{d}\mathbf{E}(P)(x) - \int_{V_j(\mathbf{c}^*)} x \mathrm{d}\mathbf{E}(P)(x) \right\| \\
&\leq 2 \sum_{j} \sum_{j' \neq j} \int_{V_j(\mathbf{c}) \cap V_{j'}(\mathbf{c}^*)} \|x\| \mathrm{d}\mathbf{E}(P)(x) \\
&\leq 2\sqrt{2} L \lambda t \leq 7\sqrt{2} \lambda \frac{L^3}{D_{\min}^2} \|\mathbf{c} - \mathbf{c}^*\|. \qquad \square
\end{aligned}$$