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 \( \mu \).

**Lemma 3.** Let \( P \) be a probability measure on \( \mathcal{M}^p \) such that \( \mathbb{E}[\mu(P)] < \infty \). Let \((\mu_n)_{n \geq 1}\) be a sequence of i.i.d. variables of law \( P \) and let \( \mathbb{P}_n = \frac{1}{n} (\mu_1 + \cdots + \mu_n) \). Then,\
\[
\text{OT}_p(\mathbb{P}_n, \mathbb{E}(P)) \xrightarrow{\mathbb{P}} 0 \text{ almost surely.} \tag{A.1}
\]

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

Before proving Theorem 1, we give a general upper bound on the distance \( \text{OT}_p \) 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 \( \{S_{k,j}\}_{j=1} \) of \( B_k \), where \( S_{k,j} \) is made of \( N_{k,j} \) squares of side length \( \sqrt{2^{-(k+1)} - 2^{-k}} \).

See also Figure 7. Let \( \mu|_{B_k} \) be the measure \( \mu \) restricted to \( B_k \) and \( \mu^* = \frac{\mu(B_k)}{\mu(B_k)} \) be the conditional probability on \( B_k \). If \( \mu(B_k) = 0 \), we let \( \mu^* \) be any fixed measure, for instance the uniform distribution on \( B_k \).

**Lemma 4.** Let \( \mu, \nu \) be two measures in \( \mathcal{M}^p \), supported on \( A_L \). Then, for any \( J \geq 0 \), with \( c_p = 2^{-p/2}(1 + 1/(2^p - 1)), \)
\[
\text{OT}_p^2(\mu, \nu) \leq 2^{p/2} L^p \sum_{k=0}^J 2^{-kp} \left( 2^{-Jp}(\mu(B_k) \land \nu(B_k)) + c_p |\mu(B_k) - \nu(B_k)| + \sum_{1 \leq j < J} 2^{-jp}|\mu(S) - \nu(S)| \right). \tag{A.2}
\]

**Proof.** Denote by \( m_k \) the quantity \( \mu(B_k) \land \nu(B_k) \). Let \( \pi_k \in \Pi(\mu_k, \nu_k) \) be an optimal plan (in the sense of \( W_p \)) between the probability measures \( \mu_k \) and \( \nu_k \). If \( \mu(B_k) \leq \nu(B_k) \), then \( \mu(B_k) \pi_k \) transports mass between \( \mu(B_k) \) and \( \nu(B_k) \) against the measure \( \mu(B_k) \pi_k \).

We then build an admissible plan between \( \mu(B_k) \pi_k \) and \( \nu(B_k) \) by transporting \( \left( 1 - \frac{\nu(B_k)}{\mu(B_k)} \right) \nu(B_k) \) 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 \( \text{OT}_p(\mu, \nu) \)
\[
\sum_{k \geq 0} \left( m_k W_p(\mu_k, \nu_k) + L^p 2^{-k} |\mu(B_k) - \nu(B_k)| \right). \tag{A.3}
\]

Furthermore, one can check that for any \( S \subset B_k \)
\[
|\mu_k(S) - \nu_k(S)| \leq \frac{\nu(S) - \nu_k(S)}{\mu_k(B_k) - \nu_k(S)} |\mu_k(B) - \nu_k(B)|. \tag{A.4}
\]

By summing over \( S \in S_{k,j-1} \), we obtain that
\[
m_k \sum_{S \in S_{k,j-1}} |\mu_k(S) - \nu_k(S)| \leq |\mu(B_k) - \nu(B_k)| + \sum_{S \in S_{k,j-1}} |\mu(S) - \nu(S)|. \tag{A.4}
\]

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. \( \square \)

Before proving Theorem 1, we state a useful inequality. Let \( \mu \in \mathcal{M}^p_{\text{MLL}} \) and let \( B \subset \Omega \) be at distance \( \ell \) from the diagonal \( \partial \Omega \). Then,
\[
\mu(B) = \int_B \frac{\| x - \partial \Omega \|}{\| x - \partial \Omega \|^q} \| x - \partial \Omega \| \, d\mu(x) \leq M \ell^{-q}. \tag{A.5}
\]
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 $\mu(B_k) \leq ML^{-q/2}k^q$ one by (A.5). Let $\mu$ be a random persistence measure of law $P$ and $\hat{\pi}_n$ 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
\[
\mathbb{E}[\hat{\pi}_n(A) - \mathbb{E}(P)(A)] \leq \sqrt{\mathbb{E}[\mu(A)^2]}. \tag{A.6}
\]
The Cauchy-Schwartz inequality also yields, as $|S_{k,j-1}| = 2^{k+1}4j^{-1},$
\[
\sum_{S \in S_{k,j-1}} \mathbb{E}[\hat{\mu}_n(S) - \mathbb{E}(P)(S)] \leq \sum_{S \in S_{k,j-1}} \sqrt{\mathbb{E}[\mu(S)^2]/n} \leq \sqrt{\mathbb{E}[\mu(B_k)^2]/n} |S_{k,j-1}| \leq ML^{-q/2}k^q. \tag{A.6}
\]
Note also that $\sum_{S \in S_{k,j-1}} \mathbb{E}[\hat{\mu}_n(S) - \mathbb{E}(P)(S)] \leq 2\mathbb{E}(P)(B_k) \leq 2ML^{-q/2}k^q$ and that $\mathbb{E}(P)(B_k) \leq ML^{-q/2}k^q$. By using those three previous inequalities, Lemma 4 and inequality (A.6), we obtain that $\mathbb{E}[\mathbb{O}^p(\hat{\pi}_n, \mathbb{E}(P))]$ is smaller than
\[
2p/2 ML^{p-q} \sum_{k \geq 0} 2^{-kp} \left( 2^{-j}p2^{jk} + \frac{c_p2^{jk}}{\sqrt{n}} \right) + \sum_{j=1}^{p} 2^{-j}p2^{jk} \left( 2 \wedge \frac{2^{k+1}2j-1}{\sqrt{n}} \right) \leq c_p,q ML^{p-q} \left( 2^{-jp} + \frac{1}{\sqrt{n}} + U \right),
\]
where $U = \sum_{k \geq 0} \sum_{j=1}^{p} 2^{k(q-1)}2^{-jp} \left( 1 \wedge \sqrt{n}^{q/2} \right)$. 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
\[
\sum_{k \geq \log_2(n)}^{\log_2(n)} \sum_{j=1}^{p} 2^{k(q-1/2)}2^{1(p-1)} + \sum_{k \geq \log_2(n)}^{\log_2(n)} \sum_{j=1}^{p} 2^{-kp}2^{-jp} \leq c_p b_j \sum_{k \geq \log_2(n)}^{\log_2(n)} \sum_{j=1}^{p} 2^{k(q+1/2-p)} + \sum_{k \geq \log_2(n)}^{\log_2(n)} \sum_{j=1}^{p} 2^{-kp}2^{-jp} \leq c_p' b_j (n^{-1/2} \vee n^{-q_p}).
\]
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 $n^{-1/2} + n^{-q_p} \log(n)$. \hfill \Box

Figure 8. In the box $U_L$, the distance $\rho$ is equal to the Euclidean distance.

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

Proof of Theorem 3. We first consider the case $q = 0$. If $\mu, \nu$ are two measures on $\Omega$ of mass smaller than $M$, then $\mathbb{O}_p(\mu, \nu) = W_{p,\rho}(\Phi(\mu), \Phi(\nu))$ (Divol & Lacombe, 2020, Prop. 3.15), where $\rho$ is the distance on $\Omega := \Omega \cup \{\partial\}$ defined by $\forall x, y \in \Omega$,
\[
\rho(x, y) = \min(||x - y||, d(x, \partial \Omega) + d(y, \partial \Omega)),
\]
and $\Phi(\mu) = \mu + (2M - |\mu|)\delta_{\Omega}$. Remark that $\rho(x, y) = ||x - y||$ if $x, y \in U_L$, where $U_L \subset A_L$ is any $\ell_1$-ball of radius $L/\sqrt{8}$ at distance $L/2$ from the diagonal, see Figure 8. As $\Phi$ is a bijection, the minimax rates for the estimation of $\mathbb{E}(P)$ is therefore equal to
\[
\inf_{\hat{\mu}} \sup_{\Phi(\mu) \in \mathcal{P}^q_{\mathcal{P}^q_{L,M,T}}} \mathbb{E}[W_{p,\rho}(\hat{\mu}, \Phi(\mathbb{E}(P)))].
\]
Let $\mathcal{Q}$ be the set of probability measures on $U_L$ whose densities belong to $B^q_{\mathcal{P}^q_{L,M,T}}$ with associated norm smaller than $T/M$. Then, $\mathcal{P}^q_{\mathcal{P}^q_{L,M,T}}$ 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 $\Phi(\mathbb{E}(P)) = M\tau$, so that the minimax rate is larger than
\[
\inf_{\hat{\mu}} \sup_{\Phi(\mu) \in \mathcal{P}^q_{\mathcal{P}^q_{L,M,T}}} \mathbb{E}[W_{p,\rho}(\hat{\mu}, M\tau)],
\]
where the infimum is taken on all measurable functions based on $K$ observations of the form $M\delta_x$, 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 $\mathbb{E}(P)$ with respect to $\mathbb{O}_p$ 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 $\tau$. 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 $P^0_{M',L}$ is included in $P^0_{M',L,T}$. In particular, the minimax rate on $P^0_{M,L,T}$ is larger than the minimax rate on $P^0_{M',L,T}$, which is larger than $c \frac{M'L^{-q}}{\sqrt{M}} = c \frac{ML^{-q}}{\sqrt{M}}$ for some constant $c > 0$.

**Remark 3** (Case $p = \infty$). It can be shown that for $p = \infty$, the minimax rate is larger than $c_a n^{-1/a}$, $\forall a > 0$. This is a consequence of an inequality between the OT$_{\infty}$ 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 $P_L, M_\infty$: 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 $c = (c_1, \ldots, c_k)$. Let $T_c : x \mapsto c_j$ if $x \in V_j(c)$ ($1 \leq j \leq k$) and $	ext{proj}_{\partial 0}(x)$ if $x \in V_{k+1}(c)$, where $	ext{proj}_{\partial 0}(x)$ denotes the orthogonal projection of a point $x \in \Omega$ on the diagonal $\partial 0$. Let $\pi$ be the pushforward of $\mu$ by the map $x \mapsto (x, T_c(x))$, extended on $\Omega \times \Omega$ by $\pi(U, \Omega) = 0$ for $U \subset \partial 0$ (intuitively, $\pi$ pushes the mass of $\mu$ on their nearest neighbor in $(c_1, \ldots, c_{k+1})$). One has, for $A, B \subset \Omega, \pi(A, \Omega) = \mu((\text{id}, T_c)^{-1}(A, \Omega)) = \mu(A)$, and $\pi(\Omega, B) = \mu(T_c^{-1}(B)) = \sum_j \mu(V_j(c)) 1\{c_j \in B\}$, that is $\pi$ is an admissible between the measures $\mu$ and $\sum_j \mu(V_j(c)) 1\{c_j \in B\}$. Hence,

$$\text{OT}_p\left(\mu, \sum_j \mu(V_j(c)) 1\{c_j \in B\}\right) \leq \min_{1 \leq j \leq k+1} \|x-c_j\|^p d\mu(x).$$

Let $(m_1, \ldots, m_k)$ be a vector of non-negative weights, let $\nu = \sum_{j=1}^k m_j 1\{c_j \in B\}$, and $\pi$ be an admissible transport plan between $\mu$ and $\nu$. One has

$$\int_{\Omega \times \Omega} \|x-y\|^p d\pi(x,y) = \sum_{j=1}^{k+1} \int_\Omega \|x-c_j\|^p d\pi(x,c_j) \geq \min_{1 \leq j \leq k+1} \|x-c_j\|^p d\pi(x,c_j) \geq \text{OT}_p\left(\mu, \sum_{j=1}^k m_j 1\{c_j \in B\}\right).$$

Taking the infimum over $\pi$ gives the conclusion. \(\square\)

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

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

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

**Proof of Lemma 5.** Let $c = (c_1, \ldots, c_k) \in \Omega^c, k$. Assume without loss of generality that $\mu(V_1(c)) = 0$. Let $c_0 = (c_2, \ldots, c_k) \in \Omega_1$ (that is, $c$ where we removed the first centroid). Assume first that $\mu(V_{k+1}(c)) > 0$, that is there is some mass transported onto the diagonal. Consider a compact subset $A \subset V_{k+1}(c)$ such that $\mu(A) > 0$ and the diameter $d(A, \partial 0)$ between $A$ and $\partial 0$. Let $c' \in A$ and observe that, for $x \in A$, $\|x-c'\| < \|x-\partial 0\|$. Therefore,

$$\int_A \|x-c'\|^p d\mu(x) < \int_A \|x-\partial 0\|^p d\mu(x).$$

Consider the measure $\nu = \delta(c_0) + \mu(A) \delta_{c'}$. Then

$$\text{OT}_p^*(\nu, \mu) \leq \sum_{j=1}^k \int_{V_j(c)} \|x-c_j\|^p d\mu(x) + \int_{V_{k+1}(c) \setminus A} \|x-\partial 0\|^p d\mu(x) + \int_A \|x-c'\|^p d\mu(A) < R_k(c).$$

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

**Lemma 6.** $R_k$ is continuous.

**Proof of Lemma 6.** For a given $x \in \Omega$, the map $c \mapsto \min_i \|x-c_i\|^p$ 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 $\{c \in \Omega^c, R_k(c) \leq \lambda\}$ 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 $c$ be such that $R_k(c) \leq \lambda$. Pick $L$ such that $\int_A \|x-\partial \Omega\|^p d\mu(x) \geq \lambda$ and $\int_A \|x-\partial \Omega\|^p d\mu(x) < R_k^{*-1} - \lambda$.  

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

$$R_{k-1}^* \leq \sum_{j=2}^{k+1} \int_{V_1(c)} \|x - c_j\| \, d\mu(x)$$

$$+ \int_{V_1(c)} \min_{j \in \{2, \ldots, k+1\}} \|x - c_j\| \, d\mu(x)$$

$$\leq R_k(c) + \int_{V_1(c)} \|x - \partial \Omega\| \, d\mu(x)$$

$$\leq R_k(c) + \int_{A_L} \|x - \partial \Omega\| \, d\mu(x)$$

$$< \lambda + R_{k-1}^* - \lambda = R_{k-1}^*,$$

leading to a contradiction. \hfill \Box

**Proof of Proposition 4.** We show by recursion on $0 \leq m \leq k$ that $R_m^* < R_m^* - 1$ and that $C_m$ is a non-empty compact set (with the convention $R_{-1}^* = +\infty$). The initialization holds as $R_0^* = \text{Pers}_p(\mu) < +\infty$ with the empty codebook being optimal. We now prove the induction step. Let $c = (c_1, \ldots, c_{m-1}) \in C_{m-1}$. Consider $c' = (c_1, c_2, \ldots, c_{m-1})$. Then, $\mu(V_1(c')) = 0$, so that $R_{m-1} = R_{m-1}(c) = R_m(c') > R_m^*$ 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 $\{c \in \Pi^k, R_m(c) \leq \lambda\}$, which is compact according to Lemma 7. As the function $R_k$ is continuous, the set of minimizers $C_m$ is a non-empty compact set, concluding the induction step. \hfill \Box

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

### C. Proof of Theorem 5

In the following, we fix a distribution $P$ supported on $\mathcal{M}_{L,M}$ and we consider $c^*$ be an optimal codebook of $E(P)$. 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

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

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

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 $c^{(t)}$ and the loss of $c^{(t+1)}$.

**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

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

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

**Proof of Theorem 5.** From Lemma 8, we show by induction that $u_t := \mathbb{E}\|c^{(t)} - c^*\|^2$ satisfies $u_t \leq \frac{\alpha}{(t+1)^2}$ 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

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

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

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 $c$, we let $p_j(c) := \mathbb{E}(V_j(c))$ and similarly, given a $n$-sample of $P$, we let $\tilde{p}_j(c) := \mathbb{P}(V_j(c))$. Note that if $\|c - c^*\|$ is small enough, one has $p_j(c) \leq 2m_{\max}$. Also, we let $w_p(c, \mu) := \mu(V_j(c))|\mu(c)|$ for $c \in \mathcal{M}^p$ and $1 \leq j \leq k$. Recall that we assume that the EPD $E(P)$ satisfies the margin condition (Definition 2) with parameters $\lambda$ and $r_0$ around the optimal codebook $c^*$.

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

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

and

$$\|w_2(c, E(P)) - w_2(c^*, E(P))\| \leq 7\sqrt{2}\lambda^\frac{L^3}{D_{\min}^2}\|c - c^*\|.$$
Lemma 10 (Lemma 24 in (Chazal et al., 2021)). Let \( c \) be a codebook such that \( \tilde{p}_j(c) \leq 2m_{\max} \) (which is always possible if \( |c - c^*| \) is small enough). Then, with probability larger than \( 1 - 2ke^{-\varepsilon} \), we have, for all \( 1 \leq j \leq k \),

\[
|\tilde{p}_j(c) - p_j(c)| \leq \sqrt{\frac{4m_{\max}p_j(c)x}{n}} + \frac{2m_{\max}x}{3n}, \quad (C.1)
\]

Moreover, with probability larger than \( 1 - e^{-\varepsilon} \), we have

\[
\|w_2(c, \mathcal{P}_n) - w_2(c, \mathcal{E}(P))\| \leq 2m_{\max}L\sqrt{\frac{2k}{n}} \left(1 + \sqrt{\frac{x}{2}}\right). \tag{C.2}
\]

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 \( 0 \leq \tilde{p}_j(c) \leq 2m_{\max} \), with variance bounded by \( \mathbb{E}[(\mu_j(c))^2]/n \leq m_{\max}p_j(c)/n \).

For equation (C.2), we introduce the function \( f_j : x \mapsto x1\{x \in V_j(c)\} \), so that \( w_2(c, \mu_j) = \mu(f_j) \), the integral of \( f_j \) against \( \mu \). We have \( w_2(c, \mu_j) - w_2(c, \mathcal{E}(P))j = n^{-1} \sum_{j=1}^{n} (\mu(f_j) - \mathcal{E}(P)(f_j)) \). Note that \( \|\mu(f_j)\| \leq \sqrt{2L} \cdot m_{\max} \). We write

\[
\mathbb{E} \left[ \frac{1}{n} \sum_{j=1}^{n} (\mu(f_j) - \mathcal{E}(P)(f_j)) \right] \leq \sqrt{\frac{1}{n} \mathbb{E} \left[ \left( \frac{1}{n} \sum_{j=1}^{n} (\mu(f_j) - \mathcal{E}(P)(f_j)) \right)^2 \right]}
\]

\[
\leq \frac{1}{n} \mathbb{E} \left[ \|\mu(f_j)\|_2^2 \right] \leq 2 \sqrt{\frac{2}{n}} \sqrt{2L} m_{\max}.
\]

Also, note that \( F(\mu_1, \ldots, \mu_n) = \|w_2(c, \mu_n) - w_2(c, \mathcal{E}(P))\| \) satisfies a bounded difference condition of parameter \( 4\sqrt{2L} m_{\max} \) (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 \( c \) instead of an optimal one \( c^* \) can be controlled linearly in terms of \( |c^* - c| \). 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 \( c^* \) be an optimal codebook, and \( c \in A_L^k \). Let \( x \in A_L \) and \( 1 \leq j \leq k \). Assume that \( x \in V_j(c^*) \cap V_{k+1}(c) \). Then, \( d(x, \partial V_j(c^*)) \leq \frac{7L^2}{2D_{\min}} |c^* - c| \). Symmetrically, if \( x \in V_{k+1}(c^*) \cap V_j(c) \), one has \( d(x, \partial V_{k+1}(c^*)) \leq \frac{7L^2}{2D_{\min}} |c^* - c| \).

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

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

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

\[
\leq \frac{L}{D_{\min}} |a - a^*| + \left(1 + \frac{1}{2} + \frac{2L^2}{D_{\min}^2} \right) |b - b^*| \leq \left(1 + \frac{L}{D_{\min}} + \frac{2L^2}{D_{\min}^2} \right) |c - c^*| \leq \frac{7L^2}{2D_{\min}^2} |c - c^*|,
\]

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 \( t = \frac{7L^2}{2D_{\min}^2} |c - c^*| \leq r_0 \),

\[
\sum_{j=1}^{k} |p_j(c) - p_j(c^*)| = \sum_{j=1}^{k} |\mathcal{E}(P)(V_j(c)) - \mathcal{E}(P)(V_j(c^*))| \leq 2 \sum_{j} \sum_{j' \neq j} |\mathcal{E}(P)(V_j(c) \cap V_{j'}(c^*))| \leq 2L|\mathcal{E}(P)[N(c^*)]| \leq 2\lambda L \leq 2\lambda r_0.
\]
where we applied Lemma 11 and the margin condition. To prove the second inequality, remark that \( w_2(c, \mathbf{E}(P))_j = \int_{V_j(c)} x d\mathbf{E}(P)(x) \). Therefore,

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