5 Dense Regime

The argument used to show that the $p$-th Betti numbers grow only sublinearly in the sparse regime extends to the critical case, showing that

\[
\begin{align*}
\text{Exp}[\beta_p(D(X_n; r))] & = \Theta(n), \\
\text{Exp}[\beta_p(\bar{C}(X_n; r))] & = \Theta(n),
\end{align*}
\]

whenever $p \geq 1$ and $r = \Theta(n^{-\frac{2}{3}})$. As before, the assumption is that the points are sampled i.i.d. with common distribution defined by the probability density function $f : \mathbb{R}^d \to \mathbb{R}$, which is bounded and Lebesgue measurable. The behavior in the dense regime is different, and we need new ideas to study the expected Betti numbers.

Expected Betti numbers. As defined earlier, the dense regime is characterized by $r = \omega(n^{-\frac{2}{3}})$. To prove bounds on the Betti numbers, we need stronger assumptions on the probability density function. Specifically, we assume that $f$ be uniform on a compact convex region with non-empty interior in $\mathbb{R}^d$, and that the boundary of that region is a smooth submanifold. We write $\text{vol} f$ for the $d$-dimensional volume of this region. It seems indeed likely that the properties stated below do not hold with any assumptions on $f$ in addition to those introduced in Section 3.

VIETORIS-RIPS, DENSE REGIME. For $2 \leq d$, $1 \leq p$, and $r = \omega(n^{-\frac{2}{3}})$, the expected $p$-th Betti number of $R(X_n; r)$ is

\[
\text{Exp}[\beta_p] = O \left( (mr^d)^p \cdot e^{-c n r^d} \cdot n \right),
\]

for some positive constant $c$.

We will see that $c$ depends only on $\text{vol} f$. It will be convenient to write $N = nr^d$, and since $r = \omega(n^{-\frac{2}{3}})$, we have $N = \omega(1)$. Noting that $N^p$ is asymptotically smaller than one over $e^{-c N}$, for every positive constant $c$ and every fixed integer $p$, we see that expected Betti number grows sublinearly. The proof of the bound on the expected Betti number uses an elementary geometric lemma and a fundamental result on discrete Morse theorem. Similar to homology, we introduce the background on Morse theory in a separate section.

The geometric lemma. Consider a sequence of $\ell + 1$ points in $\mathbb{R}^d$ that satisfy $\|y_0\| \leq \|y_1\| \leq \ldots \leq \|y_\ell\|$. In addition, we assume that $\|y_i - y_j\| \leq 2$, for all $0 \leq i < j \leq \ell$, except for $i = 0$ and $j = 1$ for which we have $\|y_0 - y_1\| > 2$. Note that this implies $\|y_1\| > 1$, else the distance of $y_0$ from the origin would exceed 1, thus violating $\|y_0\| \leq \|y_1\|$. A configuration that satisfies the distance assumptions is illustrated in Figure 10.

![Figure 10: Four points satisfying the stated assumption in $\mathbb{R}^2$.](image)

GEOMETRIC LEMMA. For $\ell \geq 1$, let $y_0, y_1, \ldots, y_\ell$ be points in $\mathbb{R}^d$ that satisfy the distance assumptions stated above. Then

\[
I(1) = B(0, \|y_1\|) \cap \bigcap_{i=1}^{\ell} B(y_i, 2)
\]

has $d$-dimensional volume bounded from below by a positive constant $\varepsilon_d$.

PROOF. For $\ell = 1$, the claim is clear because we only have two balls, and $B(0, \|y_1\|)$ contains the center of the other ball on its boundary. The volume is larger than what we would get for $\|y_1\| = 1$, which is the volume of the $d$-dimensional unit ball.

For the rest of the proof, assume $\ell \geq 2$, which implies $\|y_0 - y_1\| < 4$ since $y_2$ and all further points are at distance at most 2 from both $y_0$ and $y_1$. Let $\bar{y} = \frac{1}{2}(y_0 + y_1)$, which satisfies $\|\bar{y} - y_0\| = \|y - y_1\| > 1$. We will show that $\bar{y}$ is not too far from any of the $y_i$. Indeed, for every $2 \leq i \leq \ell$, the angle at $y_i$ formed by $y_0 - y_i$ and $y_1 - y_i$ is larger than $60^\circ$, and $\|\bar{y} - y_i\| \leq \sqrt{3}$. Now set $\varepsilon = 2 - \sqrt{3}$ and note that

\[
B(\bar{y}, \varepsilon) \subset B(y_i, 2)
\]

for all $2 \leq i \leq \ell$. It follows that $I(1)$ contains the common intersection of $B(0, \|y_1\|), B(y_i, 2)$, and $B(\bar{y}, \varepsilon)$. By construction, $\bar{y}$ is contained in all three balls. Since $y_1$ lies on
the boundary of \( B(0, \|y_i\|) \), the angle at which the sphere bounding this ball meets the sphere bounding \( B(y_i, 2) \) is at least 60°, and because \( \varepsilon \) is the smallest of the three radii, roughly one sixth of \( B(y, \varepsilon) \) is contained in the intersection of the other two balls. The claim follows.

We will need a scaled version of the lemma, in which the unit distance is replaced by \( r \). In this case, the volume of the intersection of \( \ell + 1 \) balls is at least \( \varepsilon d r^d \).

**Discrete gradient field.** Another ingredient of the proof is a discrete gradient field on the Vietoris-Rips complex, \( K = \mathcal{R}(X_n; r) \); see the next section for general background. To construct this field, we assume that the distances of the vertices from the origin are all different, and we re-index the points such that \( \|x_1\| < \|x_2\| < \ldots < \|x_n\| \). Let \( \sigma = \{x_j_1, x_j_2, \ldots, x_{j+p+1}\} \) be a \( p \)-simplex in \( K \), and assume \( j_1 < j_2 < \ldots < j_{p+1} \). Whenever possible, we pair \( \sigma \) with \( \tau = \{x_{j_0}\} \cup \sigma \) such that \( j_0 < j_1 \) and \( j_0 \) is as small as possible.

We claim that this recipe defines a discrete gradient field on \( K \). To see that the recipe defines a discrete vector field, we note that no simplex belongs to two pairs. Indeed, \( \sigma \) cannot be paired with two \( (p+1) \)-simplices since it prefers the smallest possible index \( j_0 \). It cannot be paired with two \( (p-1) \)-simplices because only one of its \( (j-1) \)-dimensional faces does not include \( x_{j_1} \). Finally, \( \sigma \) cannot be paired with a \( (p-1) \)-simplex \( \varphi \prec \sigma \) and a \((p+1)\)-simplex \( \sigma \prec \tau \) since then \( \varphi \) would be a face of \( \tau \) and therefore prefer \( \{x_{j_0}\} \cup \varphi \) over \( \sigma \).

Finally, we note that the discrete vector field is a discrete gradient field. Indeed, the minimum index of the vertices decreases along any path, which implies that we do not get cycles.

**Probability of critical simplex.** Consider again the simplex \( \sigma = \{x_{j_1}, x_{j_2}, \ldots, x_{j_{p+1}}\} \) in \( K \). It is critical if it satisfies the following two conditions:

(i) there is no common neighbor \( x_j \) of the vertices of \( \sigma \) with \( j < j_1 \), else \( \sigma \) would be paired with the \((p+1)\)-simplex \( \{x_j\} \cup \sigma \), and

(ii) the vertices \( x_{j_2} \) to \( x_{j_{p+1}} \) have a common neighbor \( x_{j_0} \) with \( j_0 < j_1 \), else \( \sigma \) would be paired with the \((p-1)\)-simplex \( \sigma - \{x_{j_1}\} \).

The points \( x_{j_0}, x_{j_1}, \ldots, x_{j_{p+1}} \) satisfy the assumptions of the Geometric Lemma, with \( \ell = p + 1 \) and \( y_i = x_{j_i} \) for \( 0 \leq i \leq \ell \). Consider now the common intersection of the balls,

\[
I(r) = B(0, \|y_i\|) \cap \bigcap_{i=1}^{\ell} B(y_i, 2r). \tag{21}
\]

The Geometric Lemma implies that the volume of \( I(r) \) is at least \( \varepsilon d r^d \), with \( \varepsilon > 0 \) the constant from the Geometric Lemma. If a point \( x_j \) falls into this intersection, then \( \sigma \) would be paired. The probability that no ball is in this common intersection is

\[
\text{Prob}[\sigma \text{ is critical}] \leq \left(1 - \frac{\varepsilon d r^d}{\text{vol } f}\right)^{n-p-1}. \tag{22}
\]

Recall that \( N = nr^d \) and \( N = \omega(1) \) in the dense regime. We therefore have

\[
\left(1 - \frac{\varepsilon d r^d}{\text{vol } f}\right)^{n-p-1} \leq e^{-c N \varepsilon d (n-p-1)} \leq O(\varepsilon^{c N}), \tag{23}
\]

where \( c \) is any constant strictly between 0 and \( \frac{\varepsilon}{\text{vol } f} \).

**Final bookkeeping.** To compute the expected number of critical \( p \)-simplices, we take the total number of subsets of size \( p+1 \) formed by the \( n \) points, and for each, we multiply the probability that the subset forms a simplex in the Vietoris-Rips complex with the conditional probability that this simplex is critical. Given the first point, we get a \( p \)-simplex only if the other \( p \) points lie in the ball of radius \( 2r \) centered at the first point. The probability that this happens is some constant times \( \left(\frac{N}{n}\right)^p \). Hence

\[
\text{Exp}[s_p] \leq \left(\frac{n}{p+1}\right) \left(\frac{N}{n}\right)^p e^{-c N} \tag{25}
\]

\[
\leq O(N^p e^{-c N} n). \tag{26}
\]

We have \( \beta_p \leq s_p \) from the Discrete Morse Inequalities to be presented in Section 6. Hence, \( \text{Exp}[\beta_p] \leq \text{Exp}[s_p] \), and therefore \( \text{Exp}[\beta_p] = O(N^p e^{-c N} n) = o(n) \), as claimed.

Can the analysis for the Vietoris-Rips complexes be generalized to a similar analysis of the Čech complex in the dense regime? In particular, is it true that the expected Betti numbers of the Čech complex grow sublinearly when \( r = \omega(n^{-\frac{1}{2}}) \)?