Supplementary File for Fair and Diverse DPP-Based Data Summarization

A. Figures

Figure 1 provides example sets of images displaying tradeoffs between fairness and geometric diversity and highlights our goal to produce a subset of images that is visually distinct and demographically varied, as depicted in the bottom row.

Figure 2 demonstrates how volume represents diversity. The vectors represent features of elements in the ground set, and the diversity can be seen to be captured by the volume of the parallelepiped formed by the vectors.

Figure 3 represents an iteration of the Algorithm 1. The algorithm selects a partition and samples the vector from that partition. The figure shows the effect of removing the projection of the sampled vector from other vectors.

Figure 4 gives an example to motivate the β-balanced condition. Suppose matrix V has vectors v_1, v_2, v_3, v_4 as rows, and partitions V_{X_1} contains v_1, v_2 and V_{X_2} contains v_3, v_4.
Negative Example (A) : For $v_1 = (2, 0), v_2 = (2, \varepsilon)$, $v_3 = (0, 2), v_4 = (\varepsilon, 2)$, as ε goes to zero, both non-zero singular values of V approach $2\sqrt{2}$. However for both V_{X_1} and V_{X_2}, the smallest singular value approaches 0 as ε decreases.
Positive Example (B) : For $v_1 = (2, 0), v_2 = (2, 3)$, $v_3 = (0, 2), v_4 = (3, 2)$, the singular values of V are 5.38 and 2.23. The singular values of both V_{X_1} and V_{X_2} are 3.81 and 1.57, which is more than half of the corresponding singular values of V. Therefore X_1, X_2 is β-balanced for $\beta = 2$.

Figure 1. . The top row of images is diverse in the geometric sense but not fair with respect to gender of race. The second row of images seems fair with respect to these sensitive features but is not diverse in the feature space. The bottom row is visually distinct and demographically varied.

B. Appendix

B.1. Proof of Lemma 1

Proof. We need to show that $q^\star$, as defined below, is the optimal (closest to $\tilde{q}$ in KL-distance) distribution over $\mathcal{C}$

$$q^\star(S) = \begin{cases} \alpha \cdot \tilde{q}(S) & \text{for } S \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha = 1/\sum_{S\in\mathcal{C}} \tilde{q}(S)$. Note first that $D_{KL}(q^\star||\tilde{q}) = \log\alpha$. Consider any distribution q over $\mathcal{C}$, it remains to show that $D_{KL}(q||\tilde{q}) \geq \log\alpha$. We have

$$\begin{aligned} D_{KL}(q||\tilde{q}) &= \sum_{S\in\mathcal{C}} q_S \log\frac{q_S}{\tilde{q}_S} \\ &= \sum_{S\in\mathcal{C}} q_S \log\frac{q_S}{\alpha\tilde{q}_S} + \log\alpha \\ &= D_{KL}(q||q^\star) + \log\alpha \\ &\geq \log\alpha, \end{aligned}$$

since $D_{KL}(q||q^\star) \geq 0$. Therefore, the minimum possible value of $D_{KL}(q||\tilde{q})$ is $\log\alpha$, which is achieved for $q = q^\star$.

■

B.2. Low rank approximation

We use the following low rank approximation lemma in the proof of Theorem 1.

Lemma 1 (Low Rank Approximation, see e.g. (Golub & Van Loan, 2012))**.** *For a matrix $A \in \mathbb{R}^{m\times n}$, with $m \geq n$, let $A = \sum_{j=1}^m \sigma_j u_j z_j^\top$ be its singular value decomposition. Then $A' = \sum_{j=1}^k \sigma_j u_j z_j^\top$ is the best rank k approximation of A, i.e., $\min_{B:\, rank(B)=k} \|A - B\|_F^2$ is achieved for $B = A'$ and attains the value $\sum_{j=k+1}^n \sigma_j^2$.*

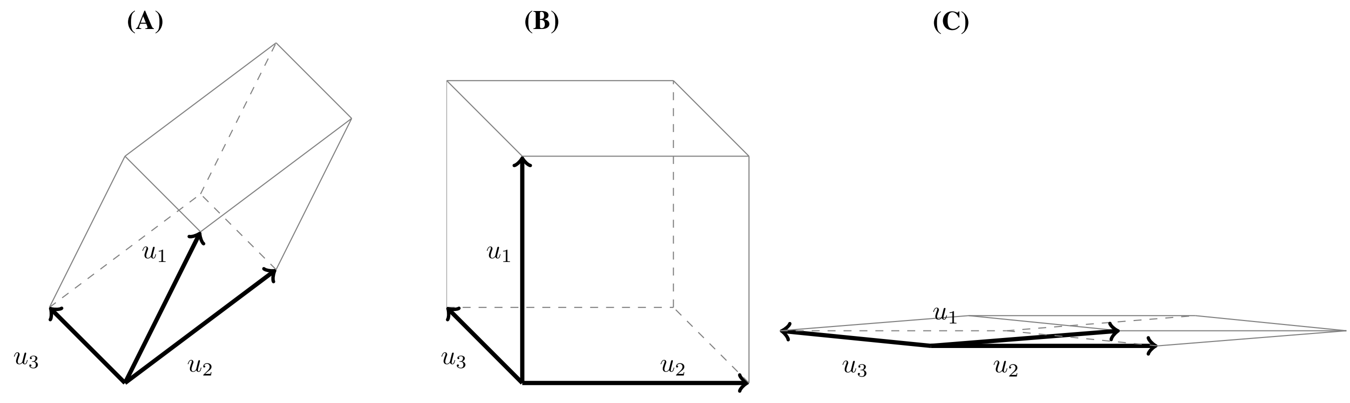

Figure 2. **(A)** depicts how diversity relates to the volume of the parallelepiped formed by the feature vectors: more the volume, more the diversity. All the vectors in **(B)** are pairwise orthogonal and their collection has a large determinant and, hence, the parallelepiped has a large volume. The parallelepiped in **(C)**, has a low volume which tends to zero as the angle between u_1, u_2 decreases or between u_2, u_3 increases. For a matrix with these vectors as rows, the determinant will be small, since the orthogonal projection of u_1 on u_2 is very small, and similarly for u_2, u_3. If they become parallel, the determinant becomes zero since one row is then linearly dependent on another.

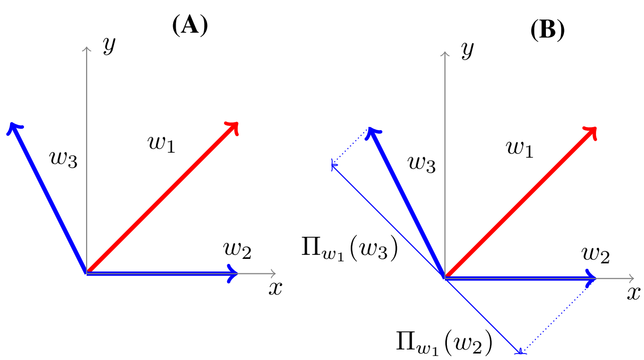

Figure 3. This figure represents an iteration of the Algorithm 1 for input $X = \{1, 2, 3\}$, $V_{X_1} = \{w_1\}$ (red) and $V_{X_2} = \{w_2, w_3\}$ (blue). If the algorithm selects the partition X_1 and samples the vector w_1, it removes the projection of w_1 from w_2 and w_3 to obtain $\Pi_{w_1}(w_2)$ and $\Pi_{w_1}(w_3)$.

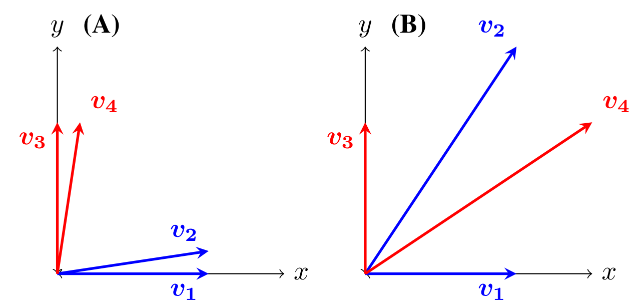

Figure 4. This figure gives a negative (A) and a positive example (B) of β-balanced condition, as described above. The colors represent the partitions. For the positive example, the partitions are β-balanced with $\beta = 2$.

B.3. Proof of Lemma 2

Proof. We will prove this lemma by induction. For the base case where there is just one row in W, $\det(WW^\top)$ is equal to $\|w_1\|^2$ which is equal to $\|\Pi_{H_1} w_1\|^2$.

Let W' be the matrix with $\{w_1, \ldots, w_{k-1}\}$ as rows. Assume that the statement is true for $k-1$ rows, i.e.,

$$\det(W'W'^\top) = \prod_{i=1}^{k-1} \|\Pi_{H_i} w_i\|^2 .$$

Then for W we have,

$$WW^\top = \begin{bmatrix} w_k \\ W' \end{bmatrix} \begin{bmatrix} w_k^\top & W'^\top \end{bmatrix} = \begin{bmatrix} \|w_k\|^2 & W'^\top w_k \\ w_k^\top W' & W'W'^\top \end{bmatrix} .$$

The first row of this matrix is

$$\begin{bmatrix} w_k^\top w_k & w_k^\top w_{k-1} & \ldots & w_k^\top w_1 \end{bmatrix} .$$

Note that elementary row product or addition transformations do not change the determinant. We will apply these transformation to make the entries of first row and first column go to zero.

Let (i) denote the i-th row of the above matrix and $WW^\top_{(i,j)}$ denote the (i, j) entry. Then the transformation

$$(1) - \frac{w_k^\top w_{k-1}}{w_{k-1}^\top w_{k-1}} (2)$$

will make the $WW^\top_{(1,2)}$ entry go to zero. For the rest of the elements,

$$WW^\top_{(1,i)} = w_k^\top w_{k-i+1} - \frac{w_k^\top w_{k-1}}{w_{k-1}^\top w_{k-1}} w_{k-1}^\top w_{k-i+1}$$
$$= w_{k-i+1}^\top \Pi_{w_{k-1}}(w_k).$$

In particular,

$$WW^\top_{(1,1)} = w_k^\top w_k - \frac{w_k^\top w_{k-1}}{w_{k-1}^\top w_{k-1}} w_{k-1}^\top w_k.$$
$$= w_k^\top \Pi_{w_{k-1}}(w_k).$$

We continue this way and next apply the transformation

$$(1) - \frac{w_{k-2}^\top \Pi_{w_{k-1}}(w_k)}{w_{k-2}^\top w_{k-2}}(3).$$

This will make the $WW^\top_{(1,3)}$ entry go to zero and by the similar analysis as above we get $WW^\top_{(1,i)} = w_{k-i+1}^\top \Pi_{H_2'}(w_k)$, where H_i' is the subspace spanned by the vectors $\{w_{k-1}, \ldots, w_{k-i}\}$. After applying $k-1$ row transformations of the form

$$(1) - \frac{w_{k-j+1}^\top \Pi_{H_{j-1}'}(w_k)}{w_{k-j+1}^\top w_{k-j+1}}(j)$$

we get that the entries $WW^\top_{(1,i)} = 0$, for $i \neq 1$ and

$$WW^\top_{(1,1)} = w_k^\top \Pi_{H_k'}(w_k) = \left\| \Pi_{H_k'}(w_k) \right\|^2.$$

Note that $H_k' = H_k$ defined in the statement of the lemma.

We can apply similar column operations to make all the entries of the first column, except $WW^\top_{(1,1)}$, go to zero. Since these elementary operations do not affect the determinant, we get

Therefore

$$\det(WW^\top) = \det \begin{bmatrix} \|w_k\|^2 & W'^\top w_k \\ w_k^\top W' & WW'^\top \end{bmatrix}$$
$$= \det \begin{bmatrix} \|\Pi_{H_k}(w_k)\|^2 & 0 \\ 0 & W'W'^\top \end{bmatrix}.$$

Using the induction hypothesis we get,

$$\det(WW^\top) = \|\Pi_{H_k}(w_k)\|^2 \cdot \det(W'W'^\top)$$
$$= \prod_{i=1}^{k} \|\Pi_{H_i}(w_i)\|^2.$$

■

B.4. Proof of Lemma 3

Proof. Consider two forms of the characteristic polynomial of the matrix $-VV^\top \in \mathbb{R}^{m \times m}$, i.e.,

$$\det(xI + VV^\top) = \prod_{i=1}^{m}(x + \sigma_i^2),$$

where $\sigma_1, \ldots, \sigma_m$ are the singular values of V.

The coefficient of x^{m-k} in $\prod_{i=1}^m (x + \sigma_i^2)$ is equal to $\sum_{1 \le i_1 < i_2 < \ldots < i_k \le m} \sigma_{i_1}^2 \sigma_{i_2}^2 \cdot \ldots \cdot \sigma_{i_k}^2$.

Let $\mathcal{W}_k$ be the set of all principal k-minors of $VV^\top$. It is a well known fact in linear algebra that the coefficient of x^{m-k} in $\det(xI + VV^\top)$ is equal to

$$\sum_{W \in \mathcal{W}_k} \det(W) = \sum_{S:|S|=k} \det(V_S V_S^\top).$$

Therefore,

$$\sum_{i_1 < i_2 < \cdots < i_k} \sigma_{i_1}^2 \sigma_{i_2}^2 \cdot \ldots \cdot \sigma_{i_k}^2 = \sum_{S:|S|=k} \det(V_S V_S^\top).$$

■

B.5. Proof of Lemma 4

Proof. We first show that for every part i, the corresponding matrix V_{X_i} has rank at least k. For this, first note that V has at least k non-zero singular values, i.e., $\sigma_k > 0$. This follows from the fact that the number of non-zero singular values determines the rank of V. The rank of V is certainly at least k, since otherwise the diversity of every subset of size k would be zero.

From the β-balance condition it follows that the number of non-zero singular values of V_{X_i} is the same as for V, and hence also the rank of V_{X_i} is at least k, as claimed.

Note now that the set of vectors output by the algorithm has determinant zero if and only if for an iteration j there exists a partition X_i such that $|S \cap X_i| < k_i$ and $\|w_x\| = 0$ for all $x \in X_i$, where $S = \{x_1, \ldots, x_{j-1}\}$.

This is equivalent to saying that all vectors in V_{X_i} belong to the subspace spanned by the vectors in S. Since the size of S is $j-1$, the dimension of the subspace spanned by the vectors in V_S is at most $j-1$. Since, by assumption for every $x \in X_i$ the projection of v_x onto the subspace $\mathrm{span}\{v_y : y \in S\}$ is 0, it implies that the dimension of subspace spanned by vectors in V_{X_i} is less than $j \le k$. This would contradict the claim proved at the very beginning – that this dimension is at least k, hence the lemma follows.

■

B.6. Proof of Theorem 2

To prove Theorem 2 we will use the following matrix concentration inequality.

Theorem 1 (Matrix Chernoff bound, see e.g. (Tropp, 2012))**.** *Given independent, random, Hermitian matrices $M_1, \ldots, M_m$ that satisfy*

$$M_i \succeq 0 \text{ and } \lambda_{\max}(M_i) \le R \quad \text{for all } i$$

it holds

$$\mathbb{P}\big[\lambda_{\min}\big(\sum_{i=1}^{m} M_i\big) \le (1-\delta)\mu_{\min}\big] \le n \cdot e^{-\delta^2 \mu_{\min}/2R}$$

where $0 \le \delta \le 1$, $\mu_{\min} = \lambda_{\min}(\sum_{i=1}^m \mathbb{E}[M_i])$.

Proof of Theorem 2. To use the Matrix Chernoff bound, we design our random experiment in the following way. We are given vectors $v_1, \ldots, v_m \in \mathbb{R}^n$ which are rows of matrix $V \in \mathbb{R}^{m \times n}$. Note that the singular values $\sigma_1 \ge \cdots \ge \sigma_n$ are the eigenvalues of $M := V^\top V = \sum_{i=1}^m v_i v_i^\top$. We will form partitions by putting each vector in X_i with $1/p$ probability.

Consider the formation of one such partition X_i. Let Y_j be the random variable taking value $v_j v_j^\top$ with probability $1/p$ and 0 with probability $(1 - 1/p)$. X_i will be all those elements for which we do not sample 0. Then for this instance we have that

$$M_i := V_{X_i}^\top V_{X_i} = \sum_{j=1}^{m} Y_j.$$

Let $u_j := (pV^\top V)^{-\frac{1}{2}} v_j$, $Z_j = u_j u_j^\top$ and $\widetilde{M_i} := \sum_{j=1}^m Z_j$. Then it can be seen that

$$\mathbb{E}\left[\widetilde{M_i}\right] = I.$$

Let $\varepsilon = \delta/2$. Note that

$$(1-\varepsilon) \cdot I \preceq \widetilde{M_i} \Leftrightarrow (1-\varepsilon) \cdot M \preceq pM_i.$$

We know that if $A \preceq B$, then for all j, $\lambda_j(A) \le \lambda_j(B)$ – see e.g. (Bhatia, 2013). Therefore if we show that $(1-\varepsilon) \cdot I \preceq \widetilde{M_i}$, then for all $j \in \{1, \ldots, n\}$,

$$\lambda_j(M_i) \ge \frac{1-\varepsilon}{p} \lambda_j(M).$$

This implies that V_{X_i} will satisfy the β-balanced condition for $\beta = \sqrt{\frac{p}{1-\varepsilon}}$.

To show that $\widetilde{M_i} \succeq (1-\varepsilon) \cdot I$ holds (with decent probability), it is enough to show that $\lambda_{\min}(\widetilde{M_i}) \ge (1-\varepsilon)$.

We will show it using Matrix concentration inequalities. But first we need to bound $\lambda_{\max}(Z_j)$.

$$\lambda_{\max}(Z_j) \le \|u_j\|^2 = p v_j^\top (V^\top V)^{-1} v_j \le \frac{\varepsilon^2}{2\log(np)}$$

Using Theorem 1, we get

$$\begin{aligned}\mathbb{P}\left[\lambda_{\min}\left(\widetilde{M_i}\right) \le (1-\varepsilon)\right] &\le n \cdot e^{-\varepsilon^2/2R} \\ &= n \cdot e^{-\log(np)} = \frac{1}{p}.\end{aligned}$$

From the above two inequalities, we have that

$$\begin{aligned}\mathbb{P}\left[\widetilde{M_i} \succeq (1-\varepsilon) \cdot I\right] &\ge 1 - \mathbb{P}\left[\lambda_{\min}\left(\widetilde{M_i}\right) \le (1-\varepsilon)\right] \\ &\ge 1 - \frac{1}{p}.\end{aligned}$$

Hence the probability that all the partitions satisfy this β-balanced condition, for $\beta = \sqrt{\frac{p}{1-\varepsilon}}$, is atleast

$$\left(1 - \frac{1}{p}\right)^p = \frac{1}{e}.$$

Since $\varepsilon = \delta/2$ and $0 \le \delta \le 1$, it can be seen that

$$\frac{1}{1-\varepsilon} \le 1 + 2\varepsilon = 1 + \delta.$$

Therefore the partition is β-balanced, for $\beta = \sqrt{(1+\delta)p}$, with probability $\ge 1/e$. ■

B.7. Proof of Theorem 3

Recall that

$$\mathcal{B} := \{S \subseteq X : |S \cap X_j| = k_j \text{ for all } j = 1, 2, \ldots, p\},$$

and let

$$\mathcal{C} := \{S \subseteq X : |S| = k\}.$$

We will use the following lemma in the proof.

Lemma 2. *For every $\varepsilon \in (0,1)$, if*

$$\sum_{S \in \mathcal{C} \setminus \mathcal{B}} \det(V_S V_S^\top) \le \varepsilon \sum_{S \in \mathcal{C}} \det(V_S V_S^\top)$$

then

$$D_{KL}(q^\star || q) \le \log \frac{1}{(1-\varepsilon)}.$$

Proof. From the assumption it follows

$$(1-\varepsilon) \sum_{S \in \mathcal{C}} \det(V_S V_S^\top) \le \sum_{S \in \mathcal{B}} \det(V_S V_S^\top).$$

Hence, for all $S \in \mathcal{C}$,

$$\frac{\det(V_S V_S^\top)}{(1-\varepsilon)\sum_{S\in\mathcal{C}}\det(V_S V_S^\top)} \geq \frac{\det(V_S V_S^\top)}{\sum_{S\in\mathcal{B}}\det(V_S V_S^\top)},$$

which translates to

$$\frac{q^*(S)}{q(S)} \leq \frac{1}{(1-\varepsilon)}.$$

Finally, we obtain

$$D_{KL}(q^*||q) = \sum_{S\in\mathcal{B}} q^*(S)\log\frac{q^*(S)}{q(S)} \leq \log\frac{1}{(1-\varepsilon)}.$$

■

Proof of Theorem 3. We start by decomposing the terms in $\sum_{S\in\mathcal{C}\setminus\mathcal{B}}\det(V_S V_S^\top)$ and analyzing each term individually using Lemma 2.

Given a set $S \subseteq X$, let $S_i := S \cap X_i$. Then $S = \bigcup_{i=1}^p S_i$. Using this, the family $\mathcal{C}\setminus\mathcal{B}$ can be decomposed as

$$\begin{aligned}\mathcal{C}\setminus\mathcal{B} &= \{S \subseteq X \mid \exists j\ \ |S\cap X_j| \neq k_j\}\\ &= \left\{\bigcup_{i=1}^p S_i \mid \forall j\ \ S_j \subseteq X_j \text{ and } \exists j\ \ |S_j| \neq k_j\right\}.\end{aligned}$$

Let $S_{(j_1,\ldots,j_p)}$ denote the following family of subsets

$$S_{(j_1,\ldots,j_p)} := \{S \subseteq X \mid |S\cap X_i| = j_i\}$$

and, for brevity, let $\mathcal{J}$ denote the following set integer tuples (all but $(k_1,k_2,\ldots,k_p)$)

$$\mathcal{J} := \mathbb{N}_{\geq 0}^p \setminus \{(k_1,k_2,\ldots,k_p)\}.$$

Given this notation, we can write the following sum as

$$\sum_{S\in\mathcal{C}\setminus\mathcal{B}}\det(V_S V_S^\top) = \sum_{(j_1,\ldots,j_p)\in\mathcal{J}}\ \sum_{S\in S_{(j_1,\ldots,j_p)}}\det(V_S V_S^\top).$$

We analyze each term of the above summation individually. We start by noting that

$$\det(V_S V_S^\top) \leq \prod_{i=1}^p \det(V_{S_i} V_{S_i}^\top),$$

where for all i, $S_i = S\cap X_i$, this is a simple consequence of the fact that $VV^\top$ is positive semidefinite. Therefore,

$$\sum_{S\in S_{(j_1,\ldots,j_p)}}\det(V_S V_S^\top) \leq \prod_{i=1}^p\ \sum_{S_i\subseteq X_i, |S_i|=j_i}\det(V_{S_i} V_{S_i}^\top).$$

Whenever a set S of cardinality k does not belong to $\mathcal{B}$, for at least one i, we have that $|S_i| = |S\cap X_i| > k_i$. Let us now analyze how does a sum of the form $\sum_{T\subseteq X_i, |T|=j}\det(V_T V_T^\top)$ behave depending on whether $j \leq k_i$ or $j > k_i$.

Case 1. $j \leq k_i$ **:**

$$\begin{aligned}\sum_{T\subseteq X_i,|T|=j}\det(V_T V_T^\top) &= \sum_{1\leq l_1<\cdots<l_j\leq n}\prod_{j'=1}^j \sigma_{i,l_{j'}}^2\\ &\leq \sum_{l=0}^j\binom{k_i}{l}\gamma^{2l}\binom{n-k_i}{j-l}(\gamma\delta)^{2(j-l)}\\ &= \gamma^{2j}\sum_{l=0}^j\binom{k_i}{l}\binom{n-k_i}{j-l}\delta^{2(j-l)}\\ &\leq \gamma^{2j}\sum_{l=0}^j\binom{k_i}{l}(n-k_i)^{j-l}\delta^{2(j-l)}.\end{aligned}$$

Since $\delta < \frac{\varepsilon}{nN_0}$,

$$\sum_{T\subseteq X_i,|T|=j}\det(V_S V_S^\top) \leq \gamma^{2j}2^{k_i}$$

Case 2. $j > k_i$ **:**

$$\begin{aligned}\sum_{T\subseteq X_i,|T|=j}\det(V_T V_T^\top) &= \sum_{1\leq l_1<\cdots<l_j\leq n}\prod_{j'=1}^j \sigma_{i,l_{j'}}^2\\ &\leq \sum_{l=0}^{k_i}\binom{k_i}{l}\gamma^{2l}\binom{n-k_i}{j-l}(\gamma\delta)^{2(j-l)}\\ &= \gamma^{2j}\sum_{l=0}^{k_i}\binom{k_i}{l}\binom{n-k_i}{j-l}\delta^{2(j-l)}\\ &= \gamma^{2j}\sum_{l=0}^{k_i}\binom{k_i}{l}(n-k_i)^{j-l}\delta^{2(j-l)}.\end{aligned}$$

Since $\delta < \frac{\varepsilon}{nN_0}$,

$$\sum_{T\subseteq X_i,|T|=j}\det(V_T V_T^\top) \leq (\frac{\varepsilon}{N_0})^{j-k_i}\gamma^{2j}\sum_{l=0}^{k_i}\binom{k_i}{l}\frac{1}{n^{j-l}}.$$

Since $j > k_i$,

$$\frac{1}{n^{j-l}} \leq \frac{1}{{k_i}^{j-l}} \leq \frac{1}{{k_i}^{k_i-l}\cdot k_i}$$

and

$$\binom{k_i}{l}\frac{1}{n^{j-l}} \leq {k_i}^{k_i-l}\frac{1}{{k_i}^{j-l}\cdot k_i} \leq \frac{1}{k_i}$$

Therefore,

$$\sum_{T\subseteq X_i,|T|=j} \det(V_T V_T^\top) \leq \left(\frac{\varepsilon}{N_0}\right)^{j-k_i} \gamma^{2j} \leq \frac{\varepsilon}{N_0}\gamma^{2j}.$$

Using the above inequalities, we obtain that for every $(j_1,\dots,j_p)\in\mathcal{J}$

$$\sum_{S\in S_{(j_1,\dots,j_p)}} \det(V_S V_S^\top) \leq \frac{\varepsilon}{N_0}\gamma^{2k}2^k.$$

Note that the size of the set of tuples $\mathcal{J}$ is bounded from above by $|\mathcal{J}| \leq \binom{k+p-1}{p-1} = N_0$.

Therefore,

$$\begin{aligned}\sum_{S\in\mathcal{C}\setminus\mathcal{B}} \det(V_S V_S^\top) &= \sum_{(j_1,\dots,j_p)\in\mathcal{J}} \sum_{S\in S_{(j_1,\dots,j_p)}} \det(V_S V_S^\top)\\ &\leq N_0\cdot\frac{\varepsilon}{N_0}\gamma^{2k}2^k = \varepsilon\gamma^{2k}2^k.\end{aligned}$$

It remains to find a lower bound for $\sum_{S\in\mathcal{C}} \det(V_S V_S^\top)$. Using Lemma 3, we obtain

$$\sum_{S\in\mathcal{C}} \det(V_S V_S^\top) = \sum_{1\leq i_1<\cdots<i_k\leq n} \prod_{j=1}^{k} \sigma_{i_k}^2 \geq \binom{n}{k}\cdot\sigma_n^{2k}.$$

By using the inequality $\binom{n}{k} \geq \frac{n^k}{k^k}$ we finally arrive at

$$\sum_{S\in\mathcal{C}} \det(V_S V_S^\top) \geq \left(\frac{n}{k}\sigma_n^2\right)^k.$$

Therefore,

$$\frac{\sum_{S\in\mathcal{C}\setminus\mathcal{B}} \det(V_S V_S^\top)}{\sum_{S\in\mathcal{C}} \det(V_S V_S^\top)} \leq \frac{\varepsilon\gamma^{2k}2^k}{\left(\frac{n}{k}\sigma_n^2\right)^k} \leq \varepsilon\cdot\left(\frac{\sqrt{2}k\gamma^2}{n\sigma_n^2}\right)^k.$$

Using the assumption that $n \geq \sqrt{2}k\cdot\left(\frac{\gamma}{\sigma_n}\right)^2$ we obtain

$$\sum_{S\in\mathcal{C}\setminus\mathcal{B}} \det(V_S V_S^\top) \leq \varepsilon \sum_{S\in\mathcal{C}} \det(V_S V_S^\top)$$

and an application of Lemma 2 finishes the proof. ■

References

Bhatia, R. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

Golub, G. H. and Van Loan, C. F. *Matrix computations*, volume 3. JHU Press, 2012. p. 79.

Tropp, J. A. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12 (4):389–434, 2012.