## A. Sublinear sampling

**Proof of Proposition 1.**

*Proof.* We have from Equation 3 that:

$$\sum_{j\in S} {K^Y}_{jj} = \sum_{j\in S}\Big[K_{jj} - \sum_{y_1,y_2\in Y}[(K_Y)^{-1}]_{y_1y_2}K_{jy_1}K_{jy_2}\Big]$$
$$= \sum_{j\in S}K_{jj} - \sum_{y_1,y_2\in Y}[(K_Y)^{-1}]_{y_1y_2}\sum_{j\in S}K_{jy_1}K_{jy_2}\ . \tag{11}$$

The first summation reduces to:

$$\sum_{j\in S}K_{jj} = \sum_{j\in S}\sum_{i\in E}\gamma_i G_{ij}^2 = \sum_{i\in E} z_i^{(S)} = \mathbf{1}^\top \boldsymbol{z}_E^{(S)}, \tag{12}$$

where $\boldsymbol{z}_E^{(S)}$ is a $|E|\times 1$ vector and $\mathbf{1}$ is the all-1s vector.

The innermost summation reduces to:

$$\sum_{j\in S}K_{jy_1}K_{jy_2} = \sum_{j\in S}\Big[\sum_{i\in E}H_{ij}G_{iy_1}\Big]\Big[\sum_{i\in E}H_{ij}G_{iy_2}\Big]$$
$$= \sum_{i_1\in E}\sum_{i_2\in E}G_{i_1y_1}G_{i_2y_2}\sum_{j\in S}H_{i_1j}H_{i_2j}$$
$$= \sum_{i_1\in E}\sum_{i_2\in E}G_{i_1y_1}G_{i_2y_2}A_{i_1i_2}^{(S)}$$
$$= G_{Ey_1}^\top A_E^{(S)} G_{Ey_2}\ , \tag{13}$$

where $G_{Ey_1}$ is a $|E|\times 1$ vector containing rows $E$ and column $y_1$ of $G$, and $A_E^{(S)}$ is the $|E|\times|E|$ matrix containing rows and columns $E$ from $A^{(S)}$. The proposition result follows from plugging Equation 12 and Equation 13 into Equation 11. □

**Proof of Theorem 1.**

*Proof.* Algorithm 2 describes the construction of the tree. It takes as inputs:

- $\gamma$, which can be computed in $\mathcal{O}(D^3)$ time via an eigendecomposition of the dual kernel $C = BB^\top$,
- $G$ (Equation 4), which can be computed via dot products in $\mathcal{O}(ND^2)$ time, and
- $H$ (Equation 4), which can be computed from $G$ in $\mathcal{O}(ND)$ time.

The complexity of constructing a leaf node is dominated by the $\mathcal{O}(D^2)$ Line 4, where $A$ is computed. There are $N$ leaf nodes, so overall leaf node construction requires $\mathcal{O}(ND^2)$ time. The complexity of constructing an internal (non-leaf) node is dominated by the $\mathcal{O}(D^2)$ Line 12, where its children's $A$ matrices are summed. There are also $N$ internal nodes in the tree. Thus, the overall complexity of tree construction is $\mathcal{O}(ND^2)$.

Algorithm 3 describes how to sample given a tree and the quantities $\lambda$, $G$, and $H$ that were pre-computed.

The initial step of selecting an elementary DPP is identical to that of the standard dual sampling algorithm (Algorithm 1), and requires $\mathcal{O}(D)$ time.

Complexity of sampling a single item: As shown by Proposition 1, the probabilities required for moving down one level in the tree are computable in $\mathcal{O}(|E|^2|Y|)$ time, where $Y$ is the set of items that have been sampled so far and $|E|$ is the size of the selected elementary DPP. Thus, the ComputeMarginal subroutine in Algorithm 3 requires $\mathcal{O}(|E|^2|Y|)$ time. To sample a single item, this subroutine gets called twice for each of the $\mathcal{O}(\log N)$ levels of the tree. Thus, the complexity of sampling a single item after $|Y|$ others have already been selected is: $\mathcal{O}(|E|^2|Y|\log N)$.

Now, recall from Section 2 that, having selected an elementary DPP with $|E|$ items, any sample $Y$ from this elementary DPP will have $|Y| = |E|$. Thus, we have to call the SampleItem subroutine $|E|$ times. This gives an overall complexity of $\mathcal{O}(|E|^4\log N)$ for sampling from the elementary DPP. □

## B. Personalized sampling

**Proof of Proposition 2.**

*Proof.* The proof is similar to that of Prop. 1; this time we need to expand all inner products and analyze each sum separately. As previously, we have:

$$\sum_{j\in S}{\hat{K}^Y}_{jj} = \sum_{j\in S}\hat{K}_{jj} - \sum_{y_1,y_2\in Y}[(\hat{K}_Y)^{-1}]_{y_1y_2}\sum_{j\in S}\hat{K}_{jy_1}\hat{K}_{jy_2}\ . \tag{14}$$

The first summation reduces to:

$$\sum_{j\in S}\hat{K}_{jj} = \sum_{i\in E}\hat{\gamma}_i\sum_{j\in S}\left(\hat{\boldsymbol{b}}_j^\top\hat{\boldsymbol{v}}_i\right)^2$$
$$= \sum_{i\in E}\hat{\gamma}_i\sum_{\ell_1,\ell_2=1}^{D}\sum_{j\in S}\hat{b}_{\ell_1j}\hat{v}_{\ell_1i}\hat{b}_{\ell_2j}\hat{v}_{\ell_2i}$$
$$= \sum_{i\in E}\hat{\gamma}_i\sum_{\ell_1,\ell_2=1}^{D}w_{\ell_1}\hat{v}_{\ell_1i}w_{\ell_2}\hat{v}_{\ell_2i}\sum_{j\in S}b_{\ell_1j}b_{\ell_2j}$$
$$= \sum_{i\in E}\hat{\gamma}_i\sum_{\ell_1,\ell_2=1}^{D}w_{\ell_1}\hat{v}_{\ell_1i}w_{\ell_2}\hat{v}_{\ell_2i}\Sigma_{\ell_1\ell_2}^{(S)}\ . \tag{15}$$

Let $W$ denote the $D \times D$ diagonal matrix with weights $\boldsymbol{w}$ on its diagonal and let $\hat{\Gamma}$ denote the $D \times D$ diagonal matrix with the inverse eigenvalues $\hat{\gamma}$ on its diagonal. Let $\hat{V}_{:,E}$ denote all rows of $\hat{V}$ and columns indexed by $E$. Define $M = \hat{V}_{:,E}^\top W$. Then Equation 15 above can be written as the following matrix product:

$$\sum_{j \in S} \hat{K}_{jj} = \mathbf{1}^\top \left[ (M^\top \hat{\Gamma}_E M) \circ \Sigma^{(S)} \right] \mathbf{1} \,. \tag{16}$$

Similarly, the innermost summation reduces to:

$$\begin{aligned}
\sum_{j \in S} \hat{K}_{jy_1} \hat{K}_{jy_2} &= \sum_{j \in S} \left[ \sum_{i \in E} \hat{\gamma}_i \hat{\boldsymbol{b}}_j^\top \hat{\boldsymbol{v}}_i \hat{\boldsymbol{b}}_{y_1}^\top \hat{\boldsymbol{v}}_i \right] \left[ \sum_{i \in E} \hat{\gamma}_i \hat{\boldsymbol{b}}_j^\top \hat{\boldsymbol{v}}_i \hat{\boldsymbol{b}}_{y_2}^\top \hat{\boldsymbol{v}}_i \right] \\
&= \sum_{i_1, i_2 \in E} \hat{\gamma}_{i_1} \hat{\gamma}_{i_2} \sum_{j \in S} \hat{\boldsymbol{b}}_j^\top \hat{\boldsymbol{v}}_{i_1} \hat{\boldsymbol{b}}_{y_1}^\top \hat{\boldsymbol{v}}_{i_1} \hat{\boldsymbol{b}}_j^\top \hat{\boldsymbol{v}}_{i_2} \hat{\boldsymbol{b}}_{y_2}^\top \hat{\boldsymbol{v}}_{i_2} \\
&= \sum_{i_1, i_2 \in E} \hat{\gamma}_{i_1} \hat{\gamma}_{i_2} \hat{\boldsymbol{b}}_{y_1}^\top \hat{\boldsymbol{v}}_{i_1} \hat{\boldsymbol{b}}_{y_2}^\top \hat{\boldsymbol{v}}_{i_2} \\
&\quad \times \sum_{j \in S} \Big( \sum_{\ell_1 = 1}^{D} w_{\ell_1} b_{\ell_1 j} \hat{v}_{\ell_1 i_1} \Big) \Big( \sum_{\ell_2 = 1}^{D} w_{\ell_2} b_{\ell_2 j} \hat{v}_{\ell_2 i_2} \Big) \\
&= \sum_{i_1, i_2 \in E} \hat{\gamma}_{i_1} \hat{\gamma}_{i_2} \hat{\boldsymbol{b}}_{y_1}^\top \hat{\boldsymbol{v}}_{i_1} \hat{\boldsymbol{b}}_{y_2}^\top \hat{\boldsymbol{v}}_{i_2} \\
&\quad \times \sum_{\ell_1, \ell_2 = 1}^{D} w_{\ell_1} \hat{v}_{\ell_1 i_1} w_{\ell_2} \hat{v}_{\ell_2 i_2} \Sigma_{\ell_1 \ell_2}^{(S)} \\
&= B_{:,y_1}^\top M^\top \hat{\Gamma}_E M \Sigma^{(S)} M^\top \hat{\Gamma}_E M B_{:,y_2} \,,
\end{aligned} \tag{17}$$

Define $\hat{H} = \hat{\Gamma}_E M B_{:,Y}$. Then the conditioning summation from Equation 14 can be written as the following matrix product:

$$\begin{aligned}
&\sum_{y_1, y_2 \in Y} [(\hat{K}_Y)^{-1}]_{y_1 y_2} \sum_{j \in S} \hat{K}_{jy_1} \hat{K}_{jy_2} \\
&\qquad = \mathbf{1}^\top \left[ (\hat{K}_Y)^{-1} \circ (\hat{H}^\top M \Sigma^{(S)} M^\top \hat{H}) \right] \mathbf{1} \,.
\end{aligned} \tag{18}$$

The proposition result follows from plugging Equations 16 and 18 into Equation 14. □

**Proof of Theorem 2.**

*Proof.* The proof that tree construction takes $O(ND^2)$ time is analogous to the proof for the non-personalized case. (Replacing $\boldsymbol{z}$ and $A$ by $\Sigma$ in fact saves some time and space.)

Algorithm 4 describes how to sample given that a tree and the non-personalized dual kernel $C$ are pre-computed. Constructing the personalized dual kernel $\hat{C} = WCW$ and computing its eigendecomposition takes $O(D^3)$ time. As shown by Proposition 2, the value returned by the COMPUTEMARGINAL subroutine is computable in $\mathcal{O}(|Y|D^2)$ time, where $Y$ is the set of items that have been sampled so far. Thus, by the same arguments as made in the non-personalized proof, this gives an overall complexity of $\mathcal{O}(k^2 D^2 \log N + D^3)$ for sampling. □

**Extension to other types of personalization.** More generally, we can get a sublinear-time sampling algorithm for any personalization method where, after one-time preprocessing work, we can compute $\sum_{j \in S} \hat{b}_{\ell_1 j} \hat{b}_{\ell_2 j}$ in time independent of $N$. For example, this is the case when we have $\hat{B} = WB + \boldsymbol{c}$ for $\boldsymbol{c} \in \mathbb{R}^D$. This transformation yields $\hat{b}_{\ell_1 j} = w_{\ell_1} b_{\ell_1 j} + c_{\ell_1}$ and so:

$$\begin{aligned}
\sum_{j \in S} \hat{b}_{\ell_1 j} \hat{b}_{\ell_2 j} &= w_{\ell_1} w_{\ell_2} \sum_{j \in S} b_{\ell_1 j} b_{\ell_2 j} + w_{\ell_1} c_{\ell_2} \sum_{j \in S} b_{\ell_1 j} \\
&\quad + w_{\ell_2} c_{\ell_1} \sum_{j \in S} b_{\ell_2 j} + |S| c_{\ell_1} c_{\ell_2} \,,
\end{aligned}$$

which can be computed in time independent of $N$ as long as we also precompute and store $[\sum_{j \in S} b_{ij}]_{i=1}^D$ at each node.

## C. Approximate sampling

**Proof of Proposition 3.**

*Proof.* Under the elementary DPP with marginal kernel $\hat{K}$, given that the elements in the set $Y$ have already been selected, recall that the probability of selecting $j \in S$ after reaching node $S$ is given by:

$$\Pr(j \mid S, Y) = \frac{\hat{K}_{jj}^Y}{\sum_{j' \in S} \hat{K}_{j'j'}^Y} \,,$$

where $\hat{K}^Y$ is the conditional marginal kernel, as defined in Equation 2. Also notice that, as previously mentioned, Equation 2 implies that the diagonal entries of this conditional can be written in terms of the unconditioned kernel as follows:

$$K_{jj}^Y = K_{jj} - \sum_{y_1 y_2 \in Y} [(K_Y)^{-1}]_{y_1 y_2} K_{jy_1} K_{jy_2} \,.$$

For ease of notation, given a set $S$, let:

$$\begin{aligned}
Z_S &:= \sum_{j \in S} \hat{K}_{jj} \,, \qquad \text{and} \\
Z_{SY} &:= \sum_{y_1, y_2 \in Y} [(\hat{K}_Y)^{-1}]_{y_1 y_2} \sum_{j \in S} \hat{K}_{jy_1} \hat{K}_{jy_2} \,.
\end{aligned}$$

Then we have, for all $j \in S$:

$$\begin{aligned}
\Pr(j \mid S, Y) - \frac{1}{|S|} &= \frac{Z_j - Z_{jY}}{Z_S - Z_{SY}} - \frac{1}{|S|} \\
&= \frac{Z_j - Z_{jY} - \frac{1}{|S|}(Z_S - Z_{SY})}{Z_S - Z_{SY}} \\
&\leq \frac{|Z_j - \frac{1}{|S|} Z_S| + |Z_{jY} - \frac{1}{|S|} Z_{SY}|}{Z_S - Z_{SY}} \,.
\end{aligned} \tag{19}$$

Expanding and upper-bounding the first half of the numerator, we have:

$$\hat{Z}_j - \frac{Z_S}{|S|} = \sum_{i\in E} \hat{\gamma}_i(\hat{\boldsymbol{b}}_j^\top \hat{\boldsymbol{v}}_i)^2 - \frac{1}{|S|}\sum_{j'\in S}\sum_{i\in E}\hat{\gamma}_i(\hat{\boldsymbol{b}}_{j'}^\top \hat{\boldsymbol{v}}_i)^2$$
$$= \sum_{i\in E}\hat{\gamma}_i \sum_{\ell_1,\ell_2} w_{\ell_1} w_{\ell_2} \hat{v}_{\ell_1 i}\hat{v}_{\ell_2 i}\left[b_{\ell_1 j}b_{\ell_2 j} - \frac{1}{|S|}\sum_{j'\in S} b_{\ell_1 j'}b_{\ell_2 j'}\right]$$
$$= \sum_{i\in E}\hat{\gamma}_i \sum_{\ell_1,\ell_2} w_{\ell_1} w_{\ell_2} \hat{v}_{\ell_1 i}\hat{v}_{\ell_2 i}\left[\Sigma^{(j)}_{\ell_1\ell_2} - \frac{1}{|S|}\Sigma^{(S)}_{\ell_1\ell_2}\right]$$
$$= \mathbf{1}^\top \left[(W\hat{V}_{:,E}\hat{\Gamma}_E\hat{V}_{:,E}^\top W)\circ\left[\Sigma^{(j)} - \frac{1}{|S|}\Sigma^{(S)}\right]\right]\mathbf{1}$$
$$= \mathbf{1}^\top \left[(M^\top\hat{\Gamma}_E M)\circ\left[\Sigma^{(j)} - \frac{1}{|S|}\Sigma^{(S)}\right]\right]\mathbf{1}$$
$$\leq \mathbf{1}^\top \left[|M^\top\hat{\Gamma}_E M|\circ\tilde{\Sigma}^{(S)}\right]\mathbf{1}\,, \qquad (20)$$

where $M$ is, as previously defined, shorthand for $V_{:,E}W$, and the $|\cdot|$ on the last line is elementwise absolute value.

We now analyze the second term in the numerator of Equation 19. We omit some of the details as they are largely the same manipulations as were done in the personalized sampling proofs. First, recall Equation 18. We will have an expression very similar to this one.

$$Z_{jY} - \frac{1}{|S|}Z_{SY}$$
$$= \sum_{y_1,y_2\in Y}[(\hat{K}_Y)^{-1}]_{y_1y_2}\left[\hat{K}_{jy_1}\hat{K}_{jy_2} - \frac{1}{|S|}\sum_{j'\in S}\hat{K}_{j'y_1}\hat{K}_{j'y_2}\right]$$
$$= \mathbf{1}^\top\left[(\hat{K}_Y)^{-1}\circ\left(\hat{H}^\top M\left[\Sigma^{(j)} - \frac{1}{|S|}\Sigma^{(S)}\right]M^\top\hat{H}\right)\right]\mathbf{1}$$
$$\leq \mathbf{1}^\top\left[|(\hat{K}_Y)^{-1}|\circ\left(|\hat{H}^\top M|\tilde{\Sigma}^{(S)}|M^\top\hat{H}|\right)\right]\mathbf{1}\,. \qquad (21)$$

where the $|\cdot|$ on the last line is elementwise absolute value.

Finally, we analyze the denominator of Equation 19.

$$Z_S - Z_{SY} = \sum_{j'\in S}\hat{K}^Y{}_{j'j'}\,.$$

This is exactly the quantity analyzed by Proposition 2, so we obtain a formula analogous to that of Equation 10 here. The proposition result then follows from combining Equations 20 and 21 with Equation 10. □

**Proof of Theorem 3.**

*Proof.* Suppose we have already sampled $y_1,\ldots,y_{\ell-1}$. Let $S_1,\ldots,S_n$ be the path through the tree such that $S_1 = \mathcal{Y}$, $S_n = \{y_\ell\}$ and $S_{i+1}\subset S_i$. To sample $y_\ell$, both algorithms must go down the tree through the same $S_i$ until they reach a node $S_k$ where approximate sampling is possible (at which point the approximate algorithm samples uniformly from $S_k \setminus Y$ and the exact algorithm continues through the remaining $S_{i>k}$).

Let $p(S_k \mid S_{k-1}, y_1,\ldots,y_{\ell-1})$ be the probability under the exact algorithm of moving from $S_{k-1}$ to $S_k$ given already selected items $y_1,\ldots,y_{\ell-1}$, and let $p(y_\ell \mid S_k, y_1,\ldots,y_{\ell-1})$ be the probability of sampling item $y_\ell$ under the exact sampling algorithm starting at node $S_k$ with items $y_1,\ldots,y_{\ell-1}$ already sampled. Define parallel notation for $q$ to represent the probabilities under the approximate sampling algorithm. By hypothesis:

$$p(S_{i+1}\,|\,S_{i<k},y_1,\ldots,y_{\ell-1}) = q(S_{i+1}\,|\,S_{i<k},y_1,\ldots,y_{\ell-1})$$

and

$$\frac{p(y_\ell \mid S_k,y_1,\ldots,y_{\ell-1}) - q(y_\ell \mid S_k,y_1,\ldots,y_{\ell-1})}{q(y_\ell \mid S_k,y_1,\ldots,y_{\ell-1})} \leq \epsilon\,.$$

Thus, as:

$$p(y_\ell\,|y_1,\ldots,y_{\ell-1}) = p(y_\ell \mid S_k,y_1,\ldots,y_{\ell-1})\prod_{i=1}^{k-1}p(S_{i+1}\mid S_i,y_1,\ldots,y_{\ell-1})$$

(and similarly for $q$), it follows that:

$$\frac{p(y_\ell \mid y_1,\ldots,y_{\ell-1}) - q(y_\ell \mid y_1,\ldots,y_{\ell-1})}{q(y_\ell \mid y_1,\ldots,y_{\ell-1})} \leq \epsilon\,.$$

Then, letting $\alpha_\ell = 1$ if element $y_\ell$ was sampled uniformly by the $q$ and 0 otherwise, we have:

$$\left|\frac{p(Y)-q(Y)}{q(Y)}\right|$$
$$= \frac{\left|\prod_{\ell=1}^k p(y_\ell \mid y_1,\ldots,y_{\ell-1}) - \prod_{\ell=1}^k q(y_\ell \mid y_1,\ldots,y_{\ell-1})\right|}{\prod_{\ell=1}^k q(y_\ell \mid y_1,\ldots,y_{\ell-1})}$$
$$\leq \frac{\left|\prod_{\ell=1}^k (1+\alpha_\ell\epsilon)q(y_\ell|y_1,\ldots,y_{\ell-1}) - \prod_{\ell=1}^k q(y_\ell|y_1,\ldots,y_{\ell-1})\right|}{\prod_{\ell=1}^k q(y_\ell \mid y_1,\ldots,y_{\ell-1})}$$
$$\leq (1+\epsilon)^m - 1\,,$$

where $m = |\{\alpha_\ell \mid \alpha_\ell = 1\}|$. □