A Discussion of Other Attacks

Attack of Jagielski (2020) It has been pointed out Jagielski (2020) that for $k_{\mathsf{priv}} = 2, k_{\mathsf{pub}} = 0$, given a single synthetic image one can discern large regions of the constituent private images simply by taking the entrywise absolute value of the synthetic image. The reason is the pixel values of a natural image are mostly continuous, i.e. nearby pixels typically have similar values, so the entrywise absolute value of the InstaHide image should be similarly continuous. That said, natural images have enough discontinuities that this breaks down if one mixes more than just two images, and as discussed above, this attack is not applicable when the individual private features are i.i.d. like in our setting.

Attack of Carlini et al. (2020) A month after this submission, Carlini et al. Carlini et al. (2020) independently gave an attack breaking the InstaHide challenge originally released by the authors of Huang et al. (2020b). In that challenge, the public dataset was ImageNet, the private dataset consisted of $n_{\mathsf{priv}} = 100$ natural images, and $k_{\mathsf{priv}} = 2$, $k_{\mathsf{pub}} = 4$, $m = 5000$. They were able to produce a visually similar copy of each private image.

Most of their work goes towards recovering which private images contributed to each synthetic image. Their first step is to train a neural network on the public dataset to compute a *similarity matrix* with rows and columns indexed by the synthetic dataset, such that the (i, j)-th entry approximates the indicator for whether the pair of private images that are part of synthetic image i overlaps with the pair that is part of synthetic image j. Ignoring the rare event that two private images contribute to two distinct synthetic images, and ignoring the fact that the accuracy of the neural network for estimating similarity is not perfect, this similarity matrix is precisely our Gram matrix in the $k_{\mathsf{priv}} = 2$ case.

The bulk of Carlini et al.'s work Carlini et al. (2020) is focused on giving a heuristic for factorizing this Gram matrix. They do so essentially by greedily decomposing the graph with adjacency matrix given by the Gram matrix into n_{priv} cliques (plus some k-means post-processing) and regarding each clique as consisting of synthetic images which share a private image in common. They then construct an $m \times n_{\mathsf{priv}}$ bipartite graph as follows: for every synthetic image index i and every private image index j, connect i to j if for four randomly chosen elements $i_1, ..., i_4 \in [m]$ of the j-th clique, the (i, i_ℓ)-th entries of the Gram matrix are nonzero. Finally, they compute a min-cost max-flow on this instance to assign every synthetic image to exactly $k_{\mathsf{priv}} = 2$ private images.

It then remains to handle the contribution from the public images. Their approach is quite different from our sparse PCA-based scheme. At a high level, they simply pretend the contribution from the public images is mean-zero noise and set up a nonconvex least-squares problem to solve for the values of the constituent private images.

Comparison to Our Generative Model Before we compare our algorithmic approach to that of Carlini et al. (2020), we mention an important difference between the setting of the InstaHide challenge and the one studied in this work, namely the way in which the random subset of public/private images that get combined into a synthetic image is sampled. In our case, for each synthetic image, the subset is chosen independently and uniformly at random from the collection of all subsets consisting of k_{priv} private images and k_{pub} public images. For the InstaHide challenge, batches of n_{priv} synthetic images get sampled one at a time via the following process: for a given batch, sample two random permutations π_1, π_2 on n_{priv} elements and let the t-th synthetic image in this batch be given by combining the private images indexed by $\pi_1(t)$ and $\pi_2(t)$. Note that this process ensures that every private image appears *exactly* $2m/n_{\mathsf{priv}}$ times, barring the rare event that $\pi_1(t) = \pi_2(t)$ for some t in some batch. It remains to be seen to what extent the attack of Carlini et al. (2020) degrades in the absence of this sort of regularity property in our setting.

Comparison to Our Attack The main commonality between our approach and that of Carlini et al. (2020) is to identify the question of extracting private information from the Gram matrix as the central algorithmic challenge.

How we compute this Gram matrix differs. We use the relationship between covariance of a folded Gaussian and covariance of a Gaussian, while Carlini et al. (2020) use the public dataset to train a neural network on public data to approximate the Gram matrix.

How we use this matrix also differs significantly. We do not produce a candidate factorization but instead pinpoint a collection of synthetic images such that we can provably ascertain that each one comprises k_{priv} private images from the same set of $k_{\mathsf{priv}}+2$ private images. This allows us to set up an appropriate piecewise linear system of size $O(k_{\mathsf{priv}})$ with a provably unique solution and solve for the $k_{\mathsf{priv}}+2$ private images.

An exciting future direction is to understand how well the heuristic in Carlini et al. (2020) scales with k_{priv}. Independent of the connection to InstaHide, it would be very interesting from a theoretical standpoint if one could show that their heuristic provably solves the multi-task phase retrieval problem defined in Problem 2 in time scaling only polynomially with k_{priv} (i.e. the sparsity of the vectors $w_1, \ldots, w_m$ in the notation of Problem 2).

B Recovering Private Images From a Gaussian Dataset

In this section we prove our main algorithmic result:

Theorem B.1 (Main). *Let $S \subsetneq [n]$, and let $n_{\mathsf{pub}} = |S|$ and $n_{\mathsf{priv}} = |S^c|$. Let $k = k_{\mathsf{pub}} + k_{\mathsf{priv}}$. If $d \geq \Omega(\mathrm{poly}(k_{\mathsf{pub}}, k_{\mathsf{priv}}) \cdot \log(n_{\mathsf{pub}} + n_{\mathsf{priv}}))$ and $m \geq \Omega\left(n_{\mathsf{priv}}^{k_{\mathsf{priv}} - \frac{2}{k_{\mathsf{priv}}+1}} k^{\mathrm{poly}(k_{\mathsf{priv}})}\right)$, then with high probability over $\mathbf{X}$ and the sequence of randomly chosen selection vectors $w_1, \ldots, w_m \sim \mathcal{D}$, there is an algorithm which takes as input the synthetic dataset $\{y^{\mathbf{X}, w_i}\}_{i \in [m]}$ and the columns of $\mathbf{X}$ indexed by S, and outputs $k_{\mathsf{priv}}+2$ distinct images $\widetilde{x}_1, \ldots, \widetilde{x}_{k_{\mathsf{priv}}+2}$ for which there exist $k_{\mathsf{priv}}+2$ distinct private images $x_{i_1}, \ldots, x_{i_{k_{\mathsf{priv}}+2}}$ satisfying $|\widetilde{x}_j| = |x_{i_j}|$ for all $j \in [k_{\mathsf{priv}}+2]$. Furthermore, the algorithm runs in time*

$$O(dm^2 + dn_{\mathsf{pub}}^2 + n_{\mathsf{pub}}^{2\omega+1}).$$

where $\omega \approx 2.373$ is the exponent of matrix multiplication.

Remark B.2. Here we give some interpretation to the quantitative guarantees of Theorem B.1:

- The number of pixels d only needs to depend logarithmically on the number of public/private images and polynomially in the sparsity $k_{\mathsf{pub}}, k_{\mathsf{priv}}$, which will be some small positive integer (e.g. $k_{\mathsf{pub}} + k_{\mathsf{priv}} = 4$ or 8 in Huang et al. (2020a), $k_{\mathsf{pub}} + k_{\mathsf{priv}} = 4$ or 6 in Huang et al. (2020b) and $k_{\mathsf{pub}} + k_{\mathsf{priv}} = 2$ in the implementation of MixUp in Zhang et al. (2018)), so the regime in which Theorem B.1 applies is quite realistic.

- Note that we can achieve recovery even when $m = o(n_{\mathsf{priv}}^{k_{\mathsf{priv}}})$. The reason this is significant is that as soon as $m = \Omega(n_{\mathsf{priv}}^{k_{\mathsf{priv}}})$, all possible combinations of k private images are used. While it is still not immediately clear how to recover private images once this has happened, we regard the fact that we can do so well before this point to be one of the most interesting aspects of our result. Finally, we remark that the runtime is largely dominated by the $O(m^2)$ term coming from forming an $m \times m$ matrix whose (i,j)-th entry turns out to equal $\langle w_i, w_j \rangle$ for all $i, j \in [m]$. In fact, naive implementations of the most sophisticated part of our algorithm (see Sections B.3, B.4, B.5, and B.6) require time $\omega(m^2)$, and getting these parts of the algorithm to run in $O(m^2)$ time turns out to be quite subtle.

B.1 Learning the Public Coordinates via Gaussian Phase Retrieval

In this section we give a procedure which, given any synthetic image $y^{\mathbf{X}, w}$, recovers the entire support of $[w]_S$. The algorithm is inspired by existing algorithms for sparse phase retrieval, with the catch that we need to handle the fact that we only get to observe the public subset of coordinates of any of the vectors p_j. Our algorithm, LearnPublic is given in Algorithm 1 below.

We first show that the population version of the matrix $\widetilde{\mathbf{M}}$ formed in Step 1 is a rank-1 projector whose top eigenvector is in the direction of $[w]_S$.

Lemma B.3. *Let w be a unit vector. Let $\widetilde{\mathbf{M}} \in \mathbb{R}^{n \times n}$ be defined as*

$$\widetilde{\mathbf{M}} \triangleq \frac{1}{d} \sum_{j=1}^{d} (y_j^2 - 1) \cdot \left([p_j]_S \cdot [p_j]_S^\top - \mathsf{Id} \right)$$

Algorithm 1: LEARNPUBLIC($\{([p_j]_S, y_j)\}_{j\in[d]}$)

Input: Samples $([p_1]_S, y_1), ..., ([p_d]_S, y_d)$
Output: $\mathrm{supp}([w]_S)$ with probability at least $1-\delta$, provided $d \geq \mathrm{poly}(k_{\mathsf{pub}})/\log(n/\delta)$

Form the matrix $\widetilde{\mathbf{M}} \triangleq \frac{1}{d}\sum_{j=1}^{d}(y_j^2-1)\cdot\big([p_j]_S\cdot[p_j]_S^\top - \mathsf{Id}\big)$.

Solve the semidefinite program (SDP) (this step takes $n_{\mathsf{pub}}^{2\omega+1}$ via Jiang et al. (2020))

$$\max_{Z\succeq 0}\langle Z, \widetilde{\mathbf{M}}\rangle \text{ subject to } \mathrm{Tr}(Z)=1, \sum_{i,j}|Z_{i,j}| \leq k_{\mathsf{pub}} \tag{2}$$

Compute the top eigenvector $\widetilde{w}$ of Z.

return coordinates of the k entries of $\widetilde{w}$ with the largest magnitudes.

Then $\mathbf{E}[\widetilde{\mathbf{M}}] = \frac{1}{2}[w]_S[w]_S^\top$.

Proof. First, it is obvious that the expectation of $\widetilde{\mathbf{M}}$ can be written as

$$\mathbf{E}[\widetilde{\mathbf{M}}] = \mathop{\mathbf{E}}_{p\sim\mathcal{N}(0,I_d)}[(\langle w,p\rangle^2-1)\cdot(p_Sp_S^\top - \mathsf{Id})].$$

For any vector $v\in\mathbb{R}^n$ with $\|v\|_2=1$, we can compute $v^\top\mathbf{E}[\widetilde{\mathbf{M}}]v$

$$\begin{aligned}
v^\top\mathbf{E}[\widetilde{\mathbf{M}}]v &= v^\top\mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(p_Sp_S^\top-\mathsf{Id})]v\\
&= \mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\langle[v]_S,p\rangle^2-1)]\\
&= \mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\|[v]_S\|_2^2\langle[v]_S/\|[v]_S\|_2,p\rangle^2-1)]\\
&= \mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\|[v]_S\|_2^2\langle[v]_S/\|[v]_S\|_2,p\rangle^2-\|[v]_S\|_2^2)]\\
&\quad+\mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\|[v]_S\|_2^2-1)]\\
&=: A_1+A_2
\end{aligned}$$

where the second step follows from $\|v\|_2^2=1$.

For the first term in the above equation, we have

$$\begin{aligned}
A_1 &= \mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\|[v]_S\|_2^2\langle[v]_S/\|[v]_S\|_2,p\rangle^2-\|[v]_S\|_2^2)]\\
&= \|[v]_S\|_2^2\mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\langle[v]_S/\|[v]_S\|_2,p\rangle^2-\|[v]_S\|_2^2)]\\
&= 2\|[v]_S\|_2^2\mathop{\mathbf{E}}_p[\phi_2(\langle w,p\rangle)\cdot\phi_2(\langle[v]_S/\|[v]_S\|_2,p\rangle)]\\
&= 2\|[v]_S\|_2^2\langle w,[v]_S/\|[v]_S\|_2\rangle^2\\
&= 2\langle w,[v]_S\rangle^2
\end{aligned}$$

where the third step follows from the fact that w and $[v]_S/\|[v]_S\|_2$ are unit vectors, ϕ_2 denotes the normalized degree-2 Hermite polynomial $\phi_2(z)\triangleq\frac{1}{\sqrt{2}}(z^2-1)$, and the last step follows from the standard fact that $\mathbf{E}_{g\sim\mathcal{N}(0,I_d)}[\phi_i(\langle g,v_1\rangle)\phi_j(\langle g,v_2\rangle)] = \langle v_1,v_2\rangle^i$ if $i=j$ and 0 otherwise.

For the second term, we have

$$A_2 = \mathop{\mathbf{E}}_p[(\langle w,p\rangle^2-1)\cdot(\|[v]_S\|_2^2-1)] = (\|[v]_S\|_2^2-1)\cdot\mathop{\mathbf{E}}_p[\langle w,p\rangle^2-1] = 0.$$

Thus, we have

$$A_1+A_2 = 2\langle w,[v]_S\rangle^2.$$

In particular, for $v=[w]_S/\|[w]_S\|_2$, the above quantity is $2\|[w]_S\|_2^2$, while for $v\perp[w]_S$, the above quantity is 0. Thus we complete the proof. □

Finally, we complete the proof of correctness of LEARNPUBLIC. Here we leverage the fact that we are running an SDP (the canonical SDP for sparse PCA) to show that as long as d is at least polynomially large in k_{pub} and *logarithmically large in* n, with high probability we can recover $\mathrm{supp}([w]_S)$.

Lemma B.4 (Learning the public coordinates)**.** *For any $\delta > 0$, if $d \geq \mathrm{poly}(k_{\mathsf{pub}})/\log(n/\delta)$, then with probability at least $1-\delta$ over the randomness of $\mathbf{X}$, we have that the coordinates output by* LEARNPUBLIC$(\{([p_j]_S, y_j)\}_{j\in[d]}$ *for $y_j \triangleq |\langle p_j, w\rangle|$ are exactly equal to* $\mathrm{supp}([w]_S)$.

Proof. Let Z be the solution to the SDP in (2), and define $w_* \triangleq [w]_S/\|[w]_S\|$. Because w_* is a feasible solution for the SDP, by optimality of Z we get that

$$\begin{aligned} 0 &\leq \langle Z - w_* w_*^\top, \widetilde{\mathbf{M}}\rangle \\ &= \langle Z - w_* w_*^\top, \mathbf{E}[\widetilde{\mathbf{M}}]\rangle + \langle Z - w_* w_*^\top, \widetilde{\mathbf{M}} - \mathbf{E}[\widetilde{\mathbf{M}}]\rangle \\ &= \frac{\|[w]_S\|^2}{2} \underbrace{\langle Z - w_* w_*^\top, w_* w_*^\top\rangle}_{①} + \underbrace{\langle Z - w_* w_*^\top, \widetilde{\mathbf{M}} - \mathbf{E}[\widetilde{\mathbf{M}}]\rangle}_{②}, \end{aligned} \tag{3}$$

where in the last step we used Lemma B.3.

Because $\|Z\|_F \leq \mathrm{Tr}(Z) = 1 = \|x_*\|$, we may upper bound ① by $-\frac{1}{2}\|Z - w_* w_*^\top\|_F^2$. For ②, note that because the entrywise L_1 norm of Z and $x_* x_*^\top$ are both upper bounded by k, by Holder's we can upper bound ② by $2k_{\mathsf{pub}} \cdot \|\widetilde{\mathbf{M}} - \mathbf{E}[\widetilde{\mathbf{M}}]\|_{\max}$. Standard concentration (see e.g. Neykov et al. (2016)) implies that as long as $d \geq \log(n/\delta)/\eta^2$, then $\|\widetilde{\mathbf{M}} - \mathbf{E}[\widetilde{\mathbf{M}}]\|_{\max} \leq \eta$. We conclude from (3) that

$$0 \leq -\frac{\|[w]_S\|^2}{4}\|Z - w_* w_*^\top\|_F^2 + 2k_{\mathsf{pub}}\eta,$$

so $\|Z - w_* w_*^\top\|_F^2 \leq 8k_{\mathsf{pub}}\eta/\|[w]_S\|^2 \geq 8\eta k_{\mathsf{pub}}^2$, where in the last step we used that if w has at least one public coordinate, then $\|[w]_S\|^2 \geq 1/k_{\mathsf{pub}}$. By Davis-Kahan, this implies that the top eigenvector $\widetilde{w}$ of Z satisfies $\|\widetilde{w} - w_*\|^2 \leq 8\eta k_{\mathsf{pub}}^2$. As the nonzero entries of w_* are at least $1/\sqrt{k_{\mathsf{pub}}}$, by taking $\eta = O(1/k_{\mathsf{pub}}^3)$ we ensure that $\|\widetilde{w} - w_*\|_\infty \leq \|\widetilde{w} - w_*\|_2 < 1/2\sqrt{k_{\mathsf{pub}}}$, so the largest entries of $\widetilde{w}$ in magnitude will be in the same coordinates as the nonzero entries of w_*. □

B.2 RECOVERING THE GRAM MATRIX VIA FOLDED GAUSSIANS

We now turn to the second step of our overall recovery algorithm: recovering the $m \times m$ Gram matrix whose (i,j)-th entry is $\mathrm{supp}(w_i) \cap \mathrm{supp}(w_j)$. For this section and the next four sections, we will assume that $S = \emptyset$, i.e. that all images are private. For brevity, let $k \triangleq k_{\mathsf{priv}}$. This turns out to be without loss of generality. Given that in the case where $S \neq \emptyset$ we can recover the public coordinates of any selection vector using LEARNPUBLIC, passing to the case of general S will be a simple matter of subtracting the contribution of the public coordinates from the entries of the Gram matrix obtained by GRAMEXTRACT to reduce to the case of $S = \emptyset$. We will elaborate on this in the final proof of Theorem B.1.

Given selection vectors $w_1, ..., w_m$, define the matrix $W \in \mathbb{R}^{m\times d}$ to have rows consisting of these vectors, so that the Gram matrix we are after is simply given by $WW^\top$. Recall that the $m \times d$ matrix whose rows consist of $y^{\mathbf{X},w_1}, ..., y^{\mathbf{X},w_m}$ can be written as

$$\mathbf{Y} \triangleq \begin{pmatrix} |\langle p_1, w_1\rangle| & \cdots & |\langle p_d, w_1\rangle| \\ \vdots & \ddots & \vdots \\ |\langle p_1, w_m\rangle| & \cdots & |\langle p_d, w_m\rangle| \end{pmatrix},$$

and as each entry of $\mathbf{X}$ is an independent standard Gaussian, the columns of $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{m\times d}$ can be regarded as independent draws from $\mathcal{N}^{\mathsf{fold}}(0, WW^\top)$, where W is defined above. Let $\mathbf{\Sigma}^{\mathsf{fold}}$ denote the covariance of this folded Gaussian distribution. It is known that one can recover information about the covariance $WW^\top$ of the original Gaussian distribution from the covariance $\mathbf{\Sigma}^{\mathsf{fold}}$ of its folded counterpart:

Lemma B.5 (Page 7 in Kan & Robotti (2017)). *Given a Gaussian $\mathcal{N}(0, \Sigma)$, the covariance $\mathbf{\Sigma}^{\mathsf{fold}} \in \mathbb{R}^{m \times m}$ of the corresponding folded Gaussian distribution $\mathcal{N}^{\mathsf{fold}}(0, \Sigma)$ is given by $\mathbf{\Sigma}^{\mathsf{fold}}_{i,i} = \Sigma_{i,i}$ and, for $i \neq j$,*

$$\mathbf{\Sigma}^{\mathsf{fold}}_{i,j} = \Sigma_{i,j}(4\Phi_2(0,0;\rho_{i,j}) - 1) + 4\Sigma^{1/2}_{i,i}\Sigma^{1/2}_{j,j}(1-\rho^2_{i,j})\phi_2(0,0;\rho_{i,j}) - \frac{2}{\pi}\Sigma^{1/2}_{i,i}\Sigma^{1/2}_{j,j}$$

where $\rho_{i,j} \triangleq \Sigma_{i,j}/(\Sigma^{1/2}_{i,i}\Sigma^{1/2}_{j,j})$.

We can apply Lemma B.5 in our specific setting to obtain the following relationship between $WW^\top$ and the covariance of $\mathcal{N}^{\mathsf{fold}}(0, WW^\top)$:

Corollary B.6. *If $\Sigma = WW^\top \in \mathbb{R}^{m \times m}$ for some matrix $W \in \mathbb{R}^{m \times n}$ where the rows of W are unit vectors, then the covariance $\mathbf{\Sigma}^{\mathsf{fold}} \in \mathbb{R}^{m \times m}$ of the corresponding folded Gaussian distribution $\mathcal{N}^{\mathsf{fold}}(0, \Sigma)$ is given by*

$$\mathbf{\Sigma}^{\mathsf{fold}}_{i,j} = \begin{cases} 1, & \text{if } i = j; \\ \Psi(\langle w_i, w_j \rangle), & \text{if } i \neq j. \end{cases}$$

where $\Psi(z) \triangleq \frac{2}{\pi}(z \cdot \arcsin(z) + \sqrt{1-z^2} - 1)$.

Proof. Because the rows of W are unit vectors, we have that $\Sigma_{i,j} = \rho_{i,j} = \langle w_i, w_j \rangle$ for all $i, j \in [m]$. To compute the off-diagonal entries of $\mathbf{\Sigma}^{\mathsf{fold}}$, note that by definition of CDF and PDF,

$$\phi_2(0,0;\langle w_i, w_j\rangle) = \frac{1}{2\pi\sqrt{1-\langle w_i, w_j\rangle^2}}, \quad \Phi_2(0,0;\langle w_i, w_j\rangle) = \frac{1}{4} + \frac{\arcsin\langle w_i, w_j\rangle}{2\pi}.$$

The claim follows. □

Algorithm 2: GRAMEXTRACT($\{y^{\mathbf{X},w_i}\}_{i\in[m]}, \eta$)

Input: InstaHide dataset $\{y^{\mathbf{X},w_i}\}_{i\in[m]}$), accuracy parameter η

Output: Matrix $\mathbf{M}$ equal to the Gram matrix $k \cdot WW^\top$, scaled to have integer entries (see Lemma B.7)

1 $\eta^* \leftarrow O(\eta^2)$.

2 Let $z_1, ..., z_d \in \mathbb{R}^m$ be the vectors given by

$$(z_j)_i = y^{\mathbf{X},w_i}_j.$$

for all $i \in [m], j \in [d]$.

3 Form the empirical estimates

$$\widehat{\mu} = \frac{1}{d}\sum_{i=1}^{d} z_i \qquad \widehat{\mathbf{\Sigma}} = \frac{1}{d}\sum_{i=1}^{d}(z_i - \widehat{\mu})(z_i - \widehat{\mu})^\top$$

and define $\widehat{\mathbf{\Sigma}}'$ to be the matrix obtained by applying the function clip_{η^*} entrywise to $\widehat{\mathbf{\Sigma}}$.

4 Let $\widetilde{\mathbf{\Sigma}}$ be the matrix obtained by applying Ψ^{-1} entrywise to $\widehat{\mathbf{\Sigma}}'$.

5 Let $\mathbf{\Sigma}^*$ denote the matrix obtained by entrywise rounding every entry of $\widetilde{\mathbf{\Sigma}}$ to the nearest multiple of $1/k$.

6 **return** $k \cdot \mathbf{\Sigma}^*$.

We now show that provided the number of pixels is moderately large, we can recover the matrix *exactly*, regardless of the choice of selection vectors $w_1, ..., w_m \in \mathbb{R}^n$. The full algorithm, GRAMEXTRACT, is given in Algorithm 2 above.

Lemma B.7 (Extract Gram matrix). *Suppose $d = \Omega(\log(m/\delta)/\eta^4)$. For random Gaussian image matrix $\mathbf{X}$ and arbitrary $w_1, ..., w_m \in \mathbb{S}^{d-1}_{\geq 0}$, let $\widetilde{\mathbf{\Sigma}}$ be the matrix computed in Step 4 of* GRAMEXTRACT *($\{y^{\mathbf{X},w_i}\}_{i\in[m]}, \eta$), and let $\mathbf{\Sigma}^*$ be the output. Then with probability $1-\delta$ over the randomness of $\mathbf{X}$, we have that $|\widetilde{\mathbf{\Sigma}}_{i,i'} - \langle w_i, w_{i'}\rangle| \leq \eta$ for all $i, i' \in [m]$. In particular, if $\eta = 1/2k$, the conditioned on this happening, $\mathbf{\Sigma}^* = k \cdot WW^\top$.*

To prove this, we will need the following helper lemma about Ψ^{-1}.

Lemma B.8. *There is an absolute constant $c > 0$ such that for any $0 < \eta < 1$ and $\widehat{z}, z \geq \eta$,*

$$|\Psi^{-1}(\widehat{z}) - \Psi^{-1}(z)| \leq \frac{c}{\sqrt{\eta}} \cdot |\widehat{z} - z|.$$

Proof. Noting that $\Psi'(z) = 2\arcsin(x)/\pi$, we get that the derivative of Ψ^{-1} at z is given by $\frac{1}{\Psi'(\Psi^{-1}(z))} = \frac{\pi}{2\arcsin(\Psi^{-1}(z))}$. One can verify numerically that for $0 \leq x \leq 1$, $\frac{x^2}{\pi} \leq \Psi(x) \leq \frac{1.2x^2}{\pi}$, so in particular $\sqrt{\pi z/1.2} \leq \Psi^{-1}(z) \leq \sqrt{\pi z}$. The derivative of Ψ^{-1} at z is therefore upper bounded by $O(1/\arcsin(\sqrt{\pi z/1.2})) \leq O(\sqrt{1.2/(\pi z)})$. In particular, for $z \geq \eta$, this is at most $O(1/\sqrt{\eta})$. In other words, over $\eta \leq z \leq 1$, Ψ^{-1} is $O(1/\sqrt{\eta})$-Lipschitz as claimed. □

Up to this point we have not used the randomness of the process generating the selection vectors $w_1, ..., w_m$. Note that without leveraging this, there exist choices of W for which it is information-theoretically impossible to discern anything. Indeed, consider a situation where $w_1, ..., w_m \in \mathbb{S}^{d-1}_{\geq 0}$ have pairwise disjoint supports. In this case all we know is that the columns of $\mathbf{Y}$ are independent standard Gaussian vectors, as $WW^\top = \mathsf{Id}$. We now proceed to the most involved component of our proof, where we exploit the randomness of the selection vectors.

B.3 Solving a Large System of Equations

In this section we show that if we can pinpoint a collection of selection vectors corresponding to all size-k subsets of some set of $k+2$ private images, then we can solve a certain system of equations to uniquely (up to sign) recover those private images. We will need the following basic notion corresponding to the fact that this system has only one unique solution, up to sign.

Definition B.9 (Generic solution of system of equations)**.** For any m and any vector $v = (v_S)_{S \in \mathcal{C}^k_{[m]}} \in \mathbb{R}^{\binom{m}{k}}$, we say that v is *generic* if there are at most two solutions to the system

$$\left|\sum_{i \in S} a_i\right| = v_S \qquad \forall S \in \mathcal{C}^k_{[m]}$$

in the variables $\{a_i\}_{i \in [m]}$. Note that there are exactly two solutions $\{a'_i\}$ and $\{a''_i\}$ to this system if and only if $a'_i = -a''_i$ for all $i \in [m]$ and $a'_i \neq 0$ for some $i \in [m]$.

We now show that for Gaussian images, the abovementioned system of equations almost surely has a unique solution up to sign.

Lemma B.10 (Vector of Gaussian subset sums is generic)**.** *Let $g_1, ..., g_m$ be independent draws from $\mathcal{N}(0,1)$. For any m satisfying $m \geq k+2$, the vector $v = (v_S)_{S \in \mathcal{C}^k_{[m]}}$ given by $v_S \triangleq \sum_{i \in S} g_i$ is generic almost surely (with respect to the randomness of $g_1, ..., g_m$).*

Proof. First note that the entries of v are all nonzero almost surely. For v to not be generic, there must exist another vector v' whose entrywise absolute value satisfies $|v| = |v'|$ but for which $v' \neq v, -v$ and for which there exists $h_1, ..., h_m$ satisfying $\sum_{i \in S} h_i = v'_S$ for all $S \in \mathcal{C}^k_{[m]}$. This would imply there exist indices S, T for which $v'_S = v_S$ and $v'_T = -v_T$.

By the assumption that $m \geq k+2$ (and recalling that $k > 1$ in our setup), we have that $\binom{m}{k} > m$. In particular, the set of vectors $w = (w_S)_{S \in \mathcal{C}^k_{[m]}}$ for which there exist numbers $\{g'_i\}$ such that $w_S = \sum_{i \in S} g'_i$ for all S is a proper subspace U of $\mathbb{R}^{\binom{m}{k}}$. Let $\ell_1, ..., \ell_a$ be a basis for the set of vectors ℓ satisfying $\langle \ell, w \rangle = 0$ for all $w \in U$. Note that there is at least one nonzero generic vector in U, for instance, the vector w^* given by $w^*_S = \mathbb{1}[i \in S]$ (here we again use the fact that $m \geq k+2$).

Letting $\mathbf{D} \in \mathbb{R}^{\binom{m}{k} \times \binom{m}{k}}$ denote the diagonal matrix whose S-th diagonal entry is equal to v_S/v'_S, note that the existence of $h_1, ..., h_m$ above implies that v additionally satisfies $\langle \mathbf{D}\ell_i, v \rangle = 0$ for all $i \in [a]$. But there must be some i for which $\mathbf{D}\ell_i$ does not lie in the span of $\ell_1, ..., \ell_a$, or else we would conclude that for any $w \in U$, the vector w' whose S-th entry is $w_S \cdot v_S/v'_S$ would also lie

in U. Because of the existence of indices S, T for which $v'_S = v_S$ and $v'_T = -v_T$, we know that $w \neq w', -w'$, so we would erroneously conclude that w is not generic for any $w \in U$, contradicting the fact that the vector w^* defined above is generic.

We conclude that there is some i for which $\mathbf{D}\ell_i$ lies outside the span of $\ell_1, \ldots, \ell_a$. But then the fact that $\langle \mathbf{D}\ell_i, v \rangle = 0$ for this particular i implies that the variables g_i satisfy some nontrivial linear relation. This almost surely cannot be the case because $g_1, ..., g_m$ are independent draws from $\mathcal{N}(0, 1)$. □

B.4 Locating a Set of Useful Selection Vectors

In the previous section we showed that we just need to find a set of selection vectors from among the rows of W that correspond to size-k subsets of some set of $k + 2$ private images. Here we show that such a collection of selection vectors is uniquely identified, up to trivial ambiguities, by their pairwise inner products.

Lemma B.11 (Uniquely identifying a family of subsets)**.** *Let $\mathcal{F} = \{T_S\}_{S \in \mathcal{C}^k_{[k+2]}}$ be a collection of subsets of $[n]$ for which $|T_S \cap T_{S'}| = |S \cap S'|$ for all $S, S' \in \mathcal{C}^k_{[k+2]}$. Then there is some subset $U \subseteq [n]$ of size $k + 2$ for which $\{T_S\} = \mathcal{C}^k_U$ as (unordered) sets.*

1	2	3	4		
		3	4	5	6
1		3	4	5	
1		3	4		6
	2	3	4	5	
	2	3	4		6
1	2	3		5	
1	2		4	5	
1	2	3			6
1	2		4		6
1		3		5	6
1			4	5	6
	2	3		5	6
	2		4	5	6
1	2			5	6

Table 1: Illustration of the sequence of subsets constructed in the proof of Lemma B.11 for $k = 4$. Red and blue denote S_0 and S_1, purple denotes $S_{a,b}$ for $a \in \{1, 2\}$, $b \in \{k + 1, k + 2\}$, green denotes the $4k - 8$ sets S'', and gold denotes the $\binom{k-2}{k-4} = 1$ set S'''.

Proof. For the reader's convenience, we illustrate the sequence of subsets constructed in the following proof in Table 1.

Suppose without loss of generality that $\mathcal{F}$ contains the sets $S_{1,2} \triangleq \{1, ..., k\}$ and $S_{k+1,k+2} \triangleq \{3, ..., k + 2\}$ (the indexing will become clear momentarily). We will show that $\{T_S\} = \mathcal{C}^k_U$ for $U = [k + 2]$.

Let $S^* \triangleq S_0 \cap S_1$. For any $S' \in \mathcal{C}^k_{[k+2]}$ satisfying $|S_0 \cap S'| = |S_1 \cap S'| = k - 1$, observe that S' must contain S^* and one element from each of $S_0 \backslash S_1 = \{1, 2\}$ and $S_1 \backslash S_0 = \{k + 1, k + 2\}$, so there are four such choices of S', call them $\{S_{a,b}\}_{a \in \{1,2\}, b \in \{k+1,k+2\}}$, and $\mathcal{F}$ must contain all of them.

Now consider any subset $S'' \subset [k + 2]$ for which, for some $b \neq b' \in \{k + 1, k + 2\}$, we have that $|S'' \cap S_{1,2}| = |S'' \cap S_{1,b}| = |S'' \cap S_{2,b}| = k - 1$, and $|S'' \cap S_{k+1,k+2}| = |S'' \cap S'_{1,b'}| = |S'' \cap S'_{2,b'}| = k - 2$. Observe that it must be that $|S'' \cap S^*| = k - 3$ and that S'' contains $\{1, 2\}$, so there are $2 \cdot \binom{k-2}{k-3} = 2k - 4$ such choices of S'', and $\mathcal{F}$ must contain all of them. We can similarly consider S'' for which, for some $a \neq a' \in \{1, 2\}$, we have that $|S'' \cap S_{k+1,k+2}| = |S'' \cap S_{a,k+1}| =$

$|S'' \cap S_{a,k+2}| = k-1$, and $|S'' \cap S_{1,2}| = |S'' \cap S'_{a',k+1}| = |S'' \cap S'_{a',k+2}| = 2k-4$, for which there are again $2k-4$ choices of S'', and $\mathcal{F}$ must contain all of them.

Alternatively, if $\mathcal{F}$ contained $k-2$ subsets S'' satisfying $|S'' \cap S_{1,2}| = |S'' \cap S_{b,k+1}| = |S'' \cap S_{b,k+2}| = k-1$ for some $b \in \{1,2\}$, then it would have to be that any such S'' contains the $k-1$ elements of $\{b, 3, \ldots, k\}$, and therefore the intersection between any pair of such S'' must be equal to $k-1$, violating the constraint that $|T_S \cap T_{S'}| = |S \cap S'|$ for all $S, S' \in \mathcal{C}^k_{[k+2]}$. The same reasoning applies to rule out the case where $\mathcal{F}$ contains $k-2$ subsets S'' satisfying $|S'' \cap S_{k+1,k+2}| = |S'' \cap S_{1,b}| = |S'' \cap S_{2,b}| = k-1$ for some $b \in \{k+1, k+2\}$.

Finally, consider the set of all subsets S''' distinct from the ones exhibited thus far, and for which $|S''' \cap S_0| = |S''' \cap S_1| = |S''' \cap S_{a,b}| = k-2$ for all $a \in \{1,2\}, b \in \{k+1, k+2\}$ and $|S''' \cap S''$ for at least one of the $4k-8$ subsets constructed two paragraphs above. Observe that any S''' distinct from the ones exhibited thus far which satisfies the first constraint must either contain S^* and two elements outside of $\{1, ..., k+4\}$, or must satisfy $|S''' \cap S^*| = k-4$ and contain $\{1, 2, k+1, k+2\}$. In the former case, such an S''' would violate the second constraint. As for the latter case, there are $\binom{k-2}{k-4}$ such choices of S''', and $\mathcal{F}$ must therefore contain all of them. We have now produced $4k-2+\binom{k-2}{k-4} = \binom{k+2}{k}$ unique subsets, all belonging to $\mathcal{C}^k_{[k+2]}$, and $\mathcal{F}$ is of size $\binom{k+2}{k}$, concluding the proof. □

B.5 Existence of a Floral Submatrix

Recall the notion of a *floral* submatrix from Definition 3.1. In this section we show that with high probability $\mathbf{M}$ contains a floral principal submatrix. In the language of sets, this means that with high probability over a sufficiently long sequence of randomly chosen size-k subsets of $[n]$, there is a collection of $\binom{k+2}{k}$ subsets in the sequence which together comprise all size-k subsets of some $U \subseteq [n]$ of size $k+2$. Quantitatively, we have the following:

Lemma B.12 (Existence of a floral submatrix)**.** *Let $m \geq \Omega(k^{O(k^3)} n^{k-\frac{2}{k+1}})$. If sets $T_1, ..., T_m$ are independent draws from the uniform distribution over $\mathcal{C}^k_n$, then with probability at least $9/10$, there is some $U \in \mathcal{C}^{k+2}_{[n]}$ for which every element of $\mathcal{C}^k_U$ is present among $T_1, ..., T_m$.*

Proof. Let $L = \binom{k+2}{k} = \frac{1}{2}(k+2)(k+1)$. Define

$$Z \triangleq \sum_{i_1 < \cdots < i_L \in [m]} \mathbb{1}\left[\{T_{i_1}, ..., T_{i_L}\} = \mathcal{C}^k_U \text{ for some } U \in \mathcal{C}^{k+2}_{[n]}\right].$$

By linearity of expectation, $\mathbf{E}[Z]$ is equal to $\binom{m}{L}$ times the probability that $\{T_1, ..., T_L\} = \mathcal{C}^k_U$ for some $U \in \mathcal{C}^{k+2}_{[n]}$. The latter probability is equal to $\binom{n}{k+2} \cdot L! \cdot \binom{n}{k}^{-L}$, so we conclude that

$$\begin{aligned}
\mathbf{E}[Z] &= \binom{m}{L} \cdot \binom{n}{k+2} \cdot L! \cdot \binom{n}{k}^{-L} \\
&\geq m^L \cdot \frac{n^{k+2}}{n^{kL}} \cdot \frac{L! \cdot (k!)^L}{L^L \cdot (k+2)^{k+2}} \\
&\geq \Omega\left(m^L n^{k+2-kL}\right) \geq \Omega(1),
\end{aligned}$$

where in the penultimate step we used that $\frac{L! \cdot (k!)^L}{L^L \cdot (k+2)^{k+2}}$ is nonnegative and increasing over $k \geq 2$, and in the last step we used that $m \geq \Omega\left(n^{k-\frac{2}{k+1}}\right)$.

We now upper bound $\mathbf{E}[Z^2]$. Consider a pair of distinct summands $(i_1, ..., i_L)$ and $(i'_1, ..., i'_L)$. Without loss of generality, we may assume these are $(1, ..., L)$ and $(s+1, ..., L)$ for some $0 \leq s \leq L$. In order for $\{T_1, ..., T_L\} = \mathcal{C}^k_U$ and $\{T_{L-s+1}, ..., T_{2L-s+1}\} = \mathcal{C}^k_{U'}$ for some $U, U' \in \mathcal{C}^{k+2}_{[n]}$, it must be that $\{T_{L-s+1}, ..., T_L\} = \mathcal{C}^k_{U \cap U'}$. Note that if $|U \cap U'| = k+2$, then $U = U'$ and therefore s must be 0. So if $s > 0$, it must be that $|U \cap U'| \in \{k, k+1\}$.

In either case, the probability that $\{T_1, ..., T_{L-s+1}\} = \mathcal{C}_U^k \backslash \mathcal{C}_{U\cap U'}^k$, $\{T_{L+1}, ..., T_{2L-s+1}\} = \mathcal{C}_U^k \backslash \mathcal{C}_{U\cap U'}^k$, and $\{T_{L-s+1}, ..., T_L\} = \mathcal{C}_{U\cap U'}^k$ is

$$(L-s)!^2 \cdot s! \cdot \binom{n}{k}^{-2L+s} \leq L!^2 \cdot (k/n)^{2kL-ks}$$

If $|U \cap U'| = k$, then s must be 1, and there are

$$\binom{n}{k} \cdot \binom{n-k-2}{2} \cdot \binom{n-k-4}{2} \leq n^{k+4}$$

choices for (U, U'). If $|U \cap U'| = k+1$ then s must be $k+1$ and there are and there are

$$\binom{n}{k+1} \cdot (n-k-1) \cdot (n-k-2) \leq n^{k+3}$$

choices for (U, U').

Finally, note that there are $\binom{m}{L}$ pairs of summands $(i_1, ..., i_l), (i'_1, ..., i'_L)$ for which $s = 0$ (namely the ones for which $i_j = i'_j$ for all j), $m \cdot \binom{m-1}{L-1} \cdot \binom{m-L}{L-1} \leq \Theta(m)^{2L-1} \cdot L!^2$ pairs for which $s = 1$, and $\binom{m}{k+1} \cdot \binom{m-k-1}{L-k-1} \cdot \binom{m-L}{L-k-1} \leq \Theta(m)^{2L-k-1} \cdot L!^2$ for which $s = k+1$. Putting everything together, we conclude that

$$\mathbf{E}[Z^2] = \mathbf{E}[Z] + \Theta(m)^{2L-1} \cdot L!^4 \cdot n^{k+4} \cdot (k/n)^{2kL-k} + \Theta(m)^{2L-k-1} \cdot L!^4 \cdot n^{k+3} \cdot (k/n)^{2kL-k(k+1)}$$

$$\leq \mathbf{E}[Z]^2 \cdot \left(1 + O(1/m) \cdot L!^4 \cdot k^{2kL-k} + O(1/m^{k+1}) \cdot L!^4 \cdot k^{2kL-k(k+1)} \cdot n^{k^2-1}\right)$$
$$\leq (1.01\,\mathbf{E}[Z])^2,$$

where in the last step we used that $L \leq k^2$ and that $n^{k^2-1}/m^{k+1} \leq 1$ because $m \geq k^{\Omega(k^3)} n^{k-1}$.

By Paley-Zygmund, we conclude that

$$\mathbb{P}[Z > 0.01\,\mathbf{E}[Z]] \geq 0.99^2 \cdot \frac{\mathbf{E}[Z]^2}{\mathbf{E}[Z^2]} \geq 9/10,$$

as desired, upon picking constant factors appropriately. □

Lemma B.12 implies that with probability at least 9/10 over the randomness of the mixup vectors $w_1, ..., w_m$, if $m \geq \Omega(k^{O(k^3)} n^{k-\frac{2}{k+1}})$, then there is a subset of $[m]$ for which the corresponding principal submatrix of $WW^\top$ is floral. By Lemma B.7, with high probability $\mathbf{M} = k \cdot WW^\top$, so this is also the case for the output of GRAMEXTRACT.

B.6 FINDING A FLORAL SUBMATRIX

As mentioned in Section 3, to find a floral principal submatrix of $\mathbf{M}$, one option is to enumerate over all subsets of size $\binom{k+2}{k}$ of $[m]$, which would take $n^{O(k^3)}$ time. We now give a much more efficient procedure for identifying a floral principal submatrix of $\mathbf{M}$, whose runtime is dominated by the time it takes to write down the entries of $\mathbf{M}$. At a high level, the reason we can obtain such dramatic savings is that the underlying graph defined by the large entries of $WW^\top$ is quite sparse, i.e. vertices of the graph typically have degree independent of k.

We will need the following basic notion:

Definition B.13. Given $i \in [m]$ and integer $0 \leq t \leq k$, let $\mathcal{N}_i^t \triangleq \{j : \langle w_i, w_j \rangle = t/k\}$. For any $j \in \mathcal{N}_i^t$, we refer to i and j as t-neighbors (this relation is obviously commutative).

We will also need the following helper lemmas establishing certain deterministic regularity conditions that $WW^\top$ will satisfy with high probability.

Lemma B.14 (Hypergraph sparsity)**.** *For any $\delta > 0$, if $m \geq n^{k-1} \log(1/\delta)$, then with probability at least $1 - 2m\delta$ over the randomness of $w_1, ..., w_m$, we have that for every $j \in [m]$, there are at most $O(m \cdot k^{k+1} \cdot n^{1-k})$ $(k-1)$-neighbors of j, and at most $O(m \cdot k^{k+2} \cdot n^{2-k})$ $(k-2)$-neighbors of j.*

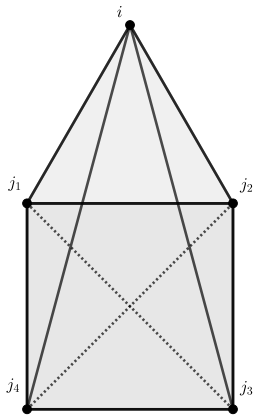


Figure 2: Illustration of a *house* $(i; j_1, j_2, j_3, j_4)$ where the solid lines indicate an entry of $k-1$ in $\mathbf{M}$, while the dotted lines indicate an entry of $k-2$.

Proof. We will union bound over $j \in [m]$, so without loss of generality fix $j = 1$ in the argument below. Let $X_{j'}$ (resp. $Y_{j'}$) denote the indicator for the event that 1 and j' are $(k-1)$-neighbors (resp. $(k-2)$-neighbors). As $w_{j'}$ is sampled independently of w_1, conditioned on w_1 we know that $X_{j'}$ is a Bernoulli random variable with expectation $\mathbf{E}[X_{j'}] = \frac{k(n-k)}{\binom{n}{k}} \le n^{1-k} \cdot k^{k+1}$, where the factor of $k(n-k)$ comes from the number of ways to pick $\mathrm{supp}(w_1) \backslash \mathrm{supp}(w_{j'})$ and $\mathrm{supp}(w_{j'}) \backslash \mathrm{supp}(w_1)$. Similarly, $Y_{j'}$ is a Bernoulli random variable with expectation $\mathbf{E}[Y_{j'}] = \frac{\binom{k}{2}\binom{n-k}{2}}{\binom{n}{k}} \le n^{2-k} \cdot k^{k+2}$. By Chernoff, we conclude that $\sum_{j'>2} X_{j'} > 2n^{1-k} \cdot k^{k+1}$ with probability at most

$$\begin{aligned}\exp\left(-m \cdot D(\mathrm{Ber}(2n^{1-k} \cdot k^{k+1}) \| \mathrm{Ber}(n^{1-k} \cdot k^{k+1}))\right) &\le \exp(-\Omega(mn^{1-k} \cdot k^{k+1})) \\ &\le \exp(-\Omega(mn^{1-k})),\end{aligned}$$

from which the first claim follows. Similarly by Chernoff, $\sum_{j'>2} Y_{j'} > 2n^{2-k} \cdot k^{k+2}$ with probability at most

$$\begin{aligned}\exp\left(-m \cdot D(\mathrm{Ber}(2n^{2-k} \cdot k^{k+2}) \| \mathrm{Ber}(n^{2-k} \cdot k^{k+2}))\right) &\le \exp(-\Omega(mn^{2-k} \cdot k^{k+2})) \\ &\le \exp(-\Omega(mn^{2-k})),\end{aligned}$$

from which the second claim follows.

□

Definition B.15. Given symmetric matrix $\mathbf{M} \in \mathbb{Z}^{m \times m}$ and distinct indices $i, j_1, ..., j_4 \in [m]$ for which $j_1 < j_4$, we say that $(i; j_1, \ldots, j_4)$ is a *house* (see Figure 2) if for all $1 \le a < b \le 4$, $\mathbf{M}_{j_a, j_b} = k-1$ if $(a, b) \in \{(1,2), (2,1), (2,3), (3,4), (1,4)\}$ and $\mathbf{M}_{j_a, j_b} = k-2$ otherwise, and furthermore $\mathbf{M}_{i, j_a} = k-1$ for all $a \in [4]$.

Lemma B.16 (Upper bounding the number of houses)**.** *If $m \ge \Omega(n^{2k/3})$, then with probability at least $9/10$ over the randomness of $w_1, \ldots, w_m$, there are at most $O(k^{5k} \cdot m^5 \cdot n^{-4k+2})$ houses in $\mathbf{M}$.*

Proof. Define

$$Z \triangleq \sum_{i, j_1, \ldots j_4 \text{ distinct}, j_1 < j_4} \mathbb{1}\left[(i; j_1, \ldots, j_4) \text{ is a house}\right].$$

By linearity of expectation, $\mathbf{E}[Z]$ is equal to $m \cdot \binom{m-1}{4} \le m^5$ times the probability that $(1; 2, 3, 4, 5)$ is a house. Note that the only way for $(1; 2, 3, 4, 5)$ to be a house is if there are disjoint subsets $S_1, S-2, T \subseteq [n]$ of size 2, 2, and $k-2$ respectively such that w_1 is supported on $S \cup T$ and each of $w_2, \ldots, w_5$ is supported on $\{s_1, s_2\} \cup T$ where $s_1 \in S_1$, $s_2 \in S_2$. There are

$O\left(\binom{n}{k-2}\cdot\binom{n}{2}^2\right) \le n^{k+2}$ such choices of (S_1, S_2, T), and for each is an $O(\binom{n}{k}^{-5})$ chance that the supports of $w_1, \dots, w_5$ correspond to a given (S_1, S_2, T), so we conclude that

$$\mathbf{E}[Z] = O\left(m^5 \cdot n^{k+2} \cdot \binom{n}{k}^{-5}\right) \le O(k^{5k} \cdot m^5 \cdot n^{-4k+2}).$$

We now upper bound $\mathbf{E}[Z^2]$. Consider a pair of distinct summands $(i; j_1, \dots, j_4)$ and $(i'; j_1', \dots, j_4')$. Recall that they correspond to some (S_1, S_2, T) and (S_1', S_2', T') respectively. Note that if these tuples overlap in any index (e.g. $(1; 2, 3, 4, 5)$ and $(6; 1, 7, 8, 9)$), then $|(S_1 \cup S_2 \cup T) \cap (S_1' \cup S_2' \cup T')| \ge k$. There are at most

$$O\left(\binom{n}{k}\cdot\binom{n-k}{2}\cdot\binom{n-k-2}{2} + \binom{n}{k+1}\cdot\binom{n-k}{1}\cdot\binom{n-k-1}{1} + \binom{n}{k+2}\right) \le O(n^{k+4})$$

pairs of sets $U, U' \subseteq [n]$ of size $k+2$ with intersection of size at least k, and given a set U of size $k+2$, there are $O\left(\binom{k+2}{k-2}\right) \le \mathrm{poly}(k)$ ways of partitioning U into three disjoint sets of size 2, 2, and $k-2$ respectively. We conclude that any pair of distinct summands in the expansion of $\mathbf{E}[Z^2]$ altogether contributes at most $\mathrm{poly}(k) \cdot O(n^{k+4}) \cdot \binom{n}{k}^{-b} \le k^{10k} \cdot n^{-(b-1)k+4}$, where $6 \le b \le 10$ is the number of distinct indices within the tuples $(i; j_1, \dots, j_4)$ and $(i'; j_1', \dots, j_4')$. For any b, there are $\binom{m}{5} \cdot \binom{m-5}{b-5} \le m^b$ such pairs of tuples.

In the special case where $b = 6$, we will use a slightly sharper bound by noting that then, it must be that $S_1 \cup S_2 \cup T$ and $S_1' \cup S_2' \cup T'$ are identical, in which case we can improve the above bound of $O(n^{k+4})$ for the number of pairs U, U' to $O(n^{k+2})$.

We conclude that

$$\mathbf{E}[Z^2] \le \mathbf{E}[Z] + k^{10k} m^6 \cdot n^{-5k+2} + \sum_{b=7}^{10} m^b \cdot n^{-(b-1)k+4} \le O(k^{10k} \cdot m^{10} \cdot n^{-8k+4}).$$

where in the last step we used the fact that $m \ge O(n^{2k/3})$ and $k \ge 2$ to bound the summands corresponding to $b = 6$ and $b = 7$. Finally, by our bounds on $\mathbf{E}[Z]$ and $\mathbf{E}[Z^2]$, we conclude by Chebyshev's that with probability at least $9/10$, there are most $2\,\mathbf{E}[Z] \le O(k^{5k} \cdot m^5 \cdot n^{-4k+2})$ houses in $\mathbf{M}$. □

Lemma B.17 (Finding a floral submatrix)**.** *Suppose $m = \Omega(n^{k-\frac{2}{k+1}})$. With probability at least $3/4$,* FINDFLORALSUBMATRIX$(\mathbf{M})$ *runs in time $O(n^{2k-\frac{4}{k+1}} \cdot \exp(\mathrm{poly}(k)))$ and outputs $\binom{k+2}{k} \times \binom{k+2}{k}$-sized subset $\mathcal{I} \subseteq [m]$ indexing a principal submatrix of $\mathbf{M}$ which is floral, together with a function $F : \mathcal{I} \to \mathcal{C}^k_{[k+2]}$ such that $\mathbf{M}_{j,j'} = |F(j) \cap F(j')|$ for all $j, j' \in \mathcal{I}$.*

Proof. The proof of correctness essentially follows immediately from the proof of Lemma B.11, while the runtime analysis will depend crucially on the sparsity of the underlying weighted graph defined by $\mathbf{M}$, as guaranteed by Lemmas B.14 and B.16. Henceforth, condition on the events of those lemmas holding, which will happen with probability at least $3/4$.

First note that if one reaches as far as Step 20 in FINDFLORALSUBMATRIX, then by the proof of Lemma B.11, the $\mathcal{I}$ produced in Step 22 indexes a principal submatrix of $\mathbf{M}$ which is floral. The recursive call in Step 24 is applied to a submatrix of $\mathbf{M}$ whose size is independent of n, and it is evident that the time expended past that point is no worse than some $\exp(\mathrm{poly}(k))$, and inductively we know that the resulting F produced in Step 25 when the recursion is complete correctly maps indices $j \in [m]$ to subsets in $\mathcal{C}^k_{[k+2]}$ such that $\mathbf{M}_{j,j'} = |F(j) \cap F(j')|$ for all $j, j' \in \mathcal{I}$.

To carry out the rest of the runtime analysis, it suffices to bound the time expended leading up to the recursive call. Consider any house $(i_0; j_1, j_2, j_3, j_4)$ encountered in Step 5. First note that one can compute $\bigcap_{a=1}^4 \mathcal{N}^{k-1}_{j_a}$ with a basic hash table, so because the first part of Lemma B.14 tells us that with high probability, $|\mathcal{N}^{k-1}_{j_a}| \le O(m \cdot k^{k+1} \cdot n^{1-k})$ for all $a \in [4]$, Step 5 only requires $O(m \cdot k^{k+1} \cdot n^{1-k})$ time. Similarly, for each of the $O(1)$ possibilities in the loop in Step 14, it takes $O(m \cdot k^{k+1} \cdot n^{1-k})$ time to enumerate over $(k-1)$-neighbors of i_z, i_α, i_β in Step 15 and, by

the second part of Lemma B.14, $O(m \cdot k^{k+2} \cdot n^{2-k})$ time to enumerate over $(k-2)$-neighbors of $i_{1-z}, i_\gamma, i_\delta$, and it takes $\mathrm{poly}(k)$ to check that the resulting indices i'' are not all $(k-1)$-neighbors of each other. And once more, in Step 20 it takes $O(m \cdot k^{k+2} \cdot n^{2-k})$ time to enumerate over all indices which are $(k-2)$ neighbors of i_0, i_1 and of every $i'' \in \mathcal{I}''$.

We conclude that for every house $(i_0; j_1, j_2, j_3, j_4)$, FINDFLORALSUBMATRIX expends at most $O(m \cdot k^{k+2} \cdot n^{2-k})$ time checking whether the house can be expanded into a set of indices corresponding to a floral principal submatrix of $\mathbf{M}$. Note that for any $(i_0; j_1, j_2, j_3, j_4)$ encountered in Step 4 which is *not* a house, the algorithm expends $O(1)$ time. As $|\mathcal{N}_{i_0}^{k-1}| \le O(m \cdot n^{1-k} \cdot k^{k+1})$ with high probability for any i_0, there are most $O(m \cdot m^4 \cdot n^{4-4k} \cdot k^{4k+4}) \le O(m^5 \cdot n^{4-4k} \cdot k^{4k+4})$ such tuples which are not houses.

And because Lemma B.16 tells us that with high probability there are $O(k^{5k} \cdot m^5 \cdot n^{-4k+2})$ houses in $\mathbf{M}$, FINDFLORALSUBMATRIX outputs None with low probability. In particular, given that any single house $(i_0; j_1, j_2, j_3, j_4)$ expends $O(m \cdot k^{k+2} \cdot n^{2-k})$ time from Step 9 all the way potentially to Step 24, we conclude that the houses contribute a total of at most $O(k^{5k} \cdot m^5 \cdot n^{-4k+2} \cdot m \cdot k^{k+2} \cdot n^{2-k}) \le O(m^6 \cdot n^{4-5k} \cdot k^{6k+2})$ to the runtime.

Putting everything together, we conclude that FINDFLORALSUBMATRIX runs in time

$$O\left(m^5 \cdot n^{4-4k} \cdot k^{4k+4} + m^6 \cdot n^{4-5k} \cdot k^{6k+2}\right) = O\left(n^{k+4-\frac{10}{k+1}} \cdot k^{O(k)}\right).$$

Lastly, note that $k + 4 - \frac{10}{k+1} \le 2k - \frac{4}{k+1}$ whenever $k \ge 2$, completing the proof. □

B.7 PUTTING EVERYTHING TOGETHER

We are now ready to conclude the proof of correctness of our main algorithm, LEARNPRIVATEIMAGE.

Proof. By Lemma B.4, the subsets S_i computed in Step 3 correctly index the public coordinates of w_i. By Lemma B.7, with high probability over the randomness of $\mathbf{X}$, the matrix $\mathbf{M}$ formed from GRAMEXTRACT in Step 1 of LEARNPRIVATEIMAGE is exactly equal to the Gram matrix $WW^\top$, so after Step 5 and Step 6, $\mathbf{M}$ is equal to the Gram matrix of the vectors $[w_1]_{S^c}, \ldots, [w_m]_{S^c}$, i.e. the restrictions of the selection vectors to the private coordinates. We are now in a position to apply the results of Sections B.3, B.4, B.5, and B.6.

By Lemma B.17, with high probability the output $\mathcal{I}, F$ of FINDFLORALSUBMATRIX in Step 7 satisfies that 1) the principal submatrix of $\mathbf{M}$ indexed by $\mathcal{I}$, a set of indices of size $\binom{k_{\mathsf{priv}}+2}{k_{\mathsf{priv}}}$, is floral, and 2) the function $F : \mathcal{I} \to \mathcal{C}^{k_{\mathsf{priv}}}_{[k_{\mathsf{priv}}+2]}$ satisfies that $|F(i) \cap F(j)| = \mathbf{M}_{i,j}$ for all $i, j \in \mathcal{I}$. By Lemma B.11, because the principal submatrix indexed by $\mathcal{I}$ is floral, there exists some subset $U \subseteq [n]$ of size $k_{\mathsf{priv}} + 2$ for which the supports of the mixup vectors w_j for $j \in \mathcal{I}$ are all the subsets of U of size k_{priv}. Finally, by Lemma B.10 and the fact that the entries of $\mathbf{X}$ are independent Gaussians, for every pixel index $\ell \in [d]$, the solution $\{\widetilde{x}_i^{(\ell)}\}$ to the system in Step 8 satisfies that there is some column x of the original private image matrix $\mathbf{X}$ such that for every $i \in [k_{\mathsf{priv}} + 2]$, $\widetilde{x}_i^{(\ell)}$ is, up to signs, equal to the ℓ-th pixel of x.

Note that the runtime of LEARNPRIVATEIMAGE is dominated by the operations of forming the matrix $\mathbf{M}$ and running FINDFLORALSUBMATRIX, which take time $O(m^2)$ by Lemma B.17. □

B.8 EXAMPLE OF A FLORAL SUBMATRIX

Example B.18. For $k = 2$, the following 6×6 matrix, after dividing every entry by k, is floral:

	$\{1,3\}$	$\{2,4\}$	$\{1,4\}$	$\{1,2\}$	$\{3,4\}$	$\{2,3\}$
$\{1,3\}$	2	0	1	1	1	1
$\{2,4\}$	0	2	1	1	1	1
$\{1,4\}$	1	1	2	1	1	0
$\{1,2\}$	1	1	1	2	0	1
$\{3,4\}$	1	1	1	0	2	1
$\{2,3\}$	1	1	0	1	1	2

Algorithm 3: FINDFLORALSUBMATRIX($\mathbf{M}, k, r$)

Input: Query access to matrix $\mathbf{M} \in \mathbb{R}^{M \times M}$, sparsity level k

Output: $\binom{k+2}{k} \times \binom{k+2}{k}$-sized subset $\mathcal{I} \subseteq [M]$, function $F : \mathcal{I} \to \mathcal{C}^k_{[k+2]}$ (Lemma B.17)

$N_{\mathsf{houses}} \leftarrow 0$.
for $i_0 \in [M]$ **do**
  $F(i_0) \leftarrow \{1, ..., k\}$.
  for $j_1, \ldots, j_4$ *in* $\mathcal{N}^{k-1}_{i_0}$ *for which* $j_1 < j_4$ **do**
    if $(i_0; j_1, j_2, j_3, j_4)$ *is a house* **then**
      $N_{\mathsf{houses}} \leftarrow N_{\mathsf{houses}} + 1$.
      if $N_{\mathsf{houses}} \geq \Omega(k^{5k} \cdot M^5 \cdot n^{-4k+2})$ **then**
        return None.
      $\mathcal{I}' \leftarrow \{j_1, j_2, j_3, j_4\}$.
      if $\bigcap_{a=1}^4 \mathcal{N}^{k-1}_{j_a} \backslash \{i_0\} \neq \emptyset$ **then**
        Let i_1 be the (unique) element of $\bigcap_{a=1}^4 \mathcal{N}^{k-1}_{j_a} \backslash \{i_0\}$.
        $\mathcal{I}'' \leftarrow \emptyset$.
        $F(i_1) \leftarrow \{3, \cdots, k+2\}$.
        for $z \in \{0, 1\}$ *and distinct* $\alpha, \beta, \gamma, \delta \in [4]$ *for which* $\alpha < \beta$ *and* i_γ *(resp.* i_δ*) is a* $(k-1)$*-neighbor of* i_α *(resp.* i_β*), and for which* i_0, α, β *are* $(k-1)$*-neighbors and* i_1, γ, δ *are* $(k-1)$*-neighbors* **do**
          if *exactly* $k-2$ *choices of* i'' *which are* $(k-1)$*-neighbors of* i_z, i_α, i_β *and* $(k-2)$*-neighbors of* $i_{1-z}, i_\gamma, i_\delta$*, and which are not all* $(k-1)$*-neighbors of each other* **then**
            Add to $\mathcal{I}''$ all such i''.
          if $|\mathcal{I}''| = 4k - 8$ **then**
            If $z = 0$, set $F(i_\alpha) \leftarrow \{1, 3, \ldots, k, k+1\}$, $F(i_\beta) \leftarrow \{2, 3, \ldots, k, k+1\}$, $F(i_\gamma) \leftarrow \{1, 3, \ldots, k, k+2\}$, and $F(i_\delta) \leftarrow \{2, 3, \ldots, k, k+2\}$.
            If $z = 1$, set $F(i_\alpha) \leftarrow \{1, 3, \ldots, k, k+1\}$, $F(i_\beta) \leftarrow \{1, 3, \ldots, k, k+2\}$, $F(i_{\gamma'}) \leftarrow \{2, 3, \ldots, k, k+1\}$ and $F(i_{\delta'}) \leftarrow \{2, 3, \ldots, k, k+2\}$.
            if *exactly* $\binom{k-2}{k-4}$ *choices of* i''' *which are* $(k-2)$*-neighbors of* i_0*,* i_1*,* i_α*,* i_β*,* i_γ*, and* i_δ*, and which are also* $(k-1)$*-neighbors of at least one* $i'' \in \mathcal{I}''$ **then**
              Let $\mathcal{I}'''$ denote the set of such i'''.
              $\mathcal{I} \leftarrow \{i_0, i_1\} \cup \mathcal{I}' \cup \mathcal{I}'' \cup \mathcal{I}'''$.
              Let $\mathbf{M}_{\mathsf{sub}}$ denote the $\binom{k-2}{k-4} \times \binom{k-2}{k-4}$ submatrix of $\mathbf{M}$ given by restricting to the rows and columns indexed by $\mathcal{I}'''$ and subtracting 4 from every entry.
              $_, G \leftarrow$FINDFLORALSUBMATRIX($\mathbf{M}_{\mathsf{sub}}, k-2$).
              For every $i''' \in \mathcal{I}'''$, set $F(i''') \leftarrow G(i''') \cup \{1, 2, k+1, k+2\}$.
              return $\mathcal{I}$, F.

Algorithm 4: LEARNPRIVATEIMAGE($\{y^{\mathbf{X},w_i}\}_{i\in[m]}$)

Input: InstaHide dataset $\{y^{\mathbf{X},w_i}\}_{i\in[m]}$

Output: Vectors $\widetilde{x}_1, ..., \widetilde{x}_{k+2} \in \mathbb{R}^d$ equal to $k+2$ images (up to signs) from the original private dataset

$\mathbf{M} \leftarrow \frac{1}{k_{\text{priv}}} \cdot$GRAMEXTRACT($\{y^{\mathbf{X},w_i}\}, \frac{1}{2k_{\text{pub}}+2k_{\text{priv}}}$).

for $i \in [m]$ **do**

  $S_i \leftarrow$LEARNPUBLIC($\{([p_j]_S, y_j)\}_{j\in[d]}$).

for $i, j \in [m]$ **do**

  $\mathbf{M}_{i,j} \leftarrow \mathbf{M}_{i,j} - \frac{1}{k_{\text{pub}}}|S_i \cap S_j|$.

$\mathbf{M} \leftarrow k_{\text{priv}} \cdot \mathbf{M}$.

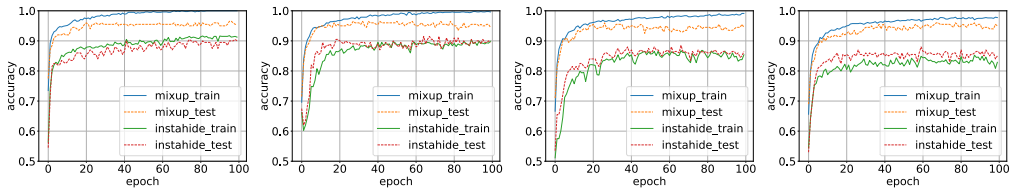
$\mathcal{I}, F \leftarrow$FINDFLORALSUBMATRIX($\mathbf{M}$).

For every pixel index $\ell \in [d]$, solve the system of equations $\left|\sum_{i\in F(j)} \widetilde{x}_i^{(\ell)}\right| = y_\ell^{\mathbf{X},w_j}$ in the variables $\{\widetilde{x}_i^{(\ell)}\}_{i\in[k_{\text{priv}}+2]}$ for all $j \in \mathcal{I}$.

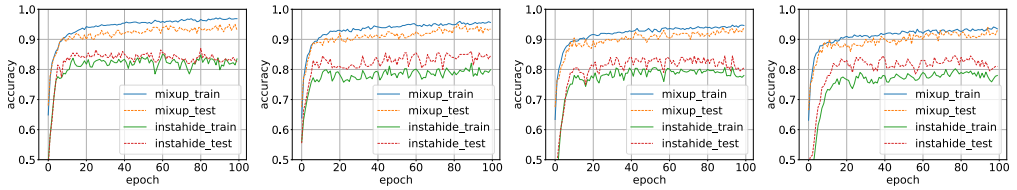
For every image index $i \in [k_{\text{priv}}+2]$, let $\widetilde{x}_i \in \mathbb{R}^d$ denote the image whose ℓ-th pixel is equal to $\widetilde{x}_i^{(\ell)}$.

return $\widetilde{x}_1, ..., \widetilde{x}_{k_{\text{priv}}+2}$.

C ADDITIONAL EXPERIMENTAL RESULTS

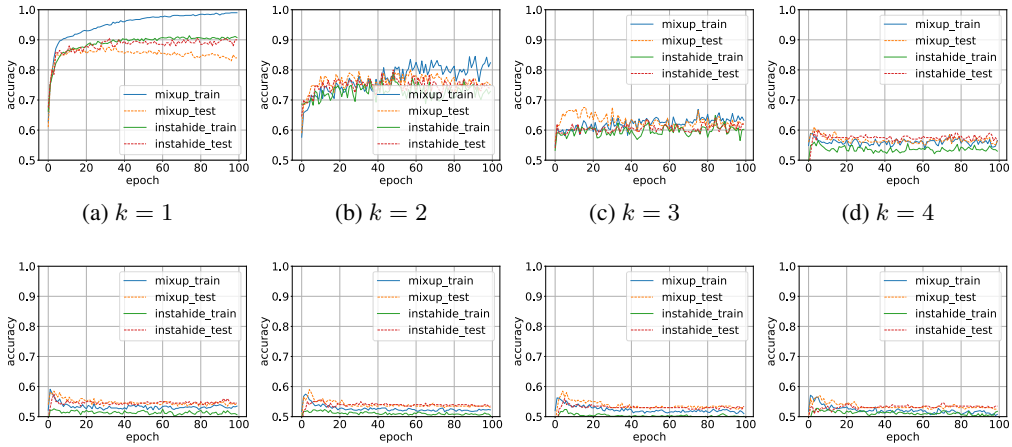


(a) $k = 1$ (b) $k = 2$ (c) $k = 3$ (d) $k = 4$



(e) $k = 5$ (f) $k = 6$ (g) $k = 7$ (h) $k = 8$

Figure 3: Comparing Vanilla, Mixup and Instahide training on Gaussian magnitude dataset with different k.



(a) $k = 1$ (b) $k = 2$ (c) $k = 3$ (d) $k = 4$

(e) $k = 5$ (f) $k = 6$ (g) $k = 7$ (h) $k = 8$

Figure 4: Comparing Vanilla, Mixup and Instahide training on Gaussian dataset with different k.