In this document, we provide detailed technical proofs of our main results, as well as the additional results (differentially private subspace clustering), experiments and dicussions that do not fit into the paper due to space constraint.

Appendix A contains proofs for our main results. The proofs are sorted in the order that their corresponding statements appear in the paper. Appendix B formalizes our claims in the paper about attribute privacy and the corresponding utility theorem and includes additional discussions on the difficulty of a stronger user-level privacy claim. Appendix C contains numerical simulations on the performance of compressed SSC under fully random models. Appendix D summarizes a few concentration bounds that we used in the paper.

Lastly, for readers' easy reference, we compile a table of symbols and notations used.

A. Proofs of the main results

A.1. Proofs of propositions in Section 3

In this section we prove that a subspace embedding enjoys the property detailed in Proposition 1. We also prove that both random Gaussian projection and uniform row sampling are subspace embeddings with respect to $\mathcal{B} = \{\mathrm{span}(\mathcal{U}^{(\ell)} \cup \mathcal{U}^{(\ell')}); \ell, \ell' \in [k]\} \cup \{\boldsymbol{x}_i, \boldsymbol{z}_i; i \in [N]\}$.

Proof of Proposition 1. Fix $\ell, \ell' \in \{1, \cdots, k\}$ and let $\mathcal{U} = \mathrm{span}(\mathcal{U}^{(\ell)} \cup \mathcal{U}^{(\ell')})$ denote the subspace spanned by the union of the two subspaces $\mathcal{U}^{(\ell)}$ and $\mathcal{U}^{(\ell')}$. By assumption, the rank of $\mathcal{U}^{(\ell)} \cup \mathcal{U}^{(\ell')}$, r', satisfies $r' \leq r_\ell + r_{\ell'} \leq 2r$. For any $\boldsymbol{x} \in \mathcal{U}^{(\ell)}$ and $\boldsymbol{y} \in \mathcal{U}^{(\ell')}$ we have

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{4} \left(\|\boldsymbol{x} + \boldsymbol{y}\|_2^2 - \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \right); \tag{A.1}$$

subsequently,

$$\left| \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \langle \boldsymbol{\Psi}\boldsymbol{x}, \boldsymbol{\Psi}\boldsymbol{y} \rangle \right| \leq \frac{1}{4} \left(\left| \|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{\Psi}(\boldsymbol{x} + \boldsymbol{y})\|^2 \right| + \left| \|\boldsymbol{x} - \boldsymbol{y}\|^2 - \|\boldsymbol{\Psi}(\boldsymbol{x} - \boldsymbol{y})\|^2 \right| \right). \tag{A.2}$$

Since $\boldsymbol{\Psi}$ is a subspace embedding, the following holds for all $\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y} \in \mathrm{span}(\mathcal{U}^{(\ell)} \cup \mathcal{U}^{(\ell')})$:

$$\begin{aligned} (1-\epsilon)^2 \|\boldsymbol{x} + \boldsymbol{y}\|^2 \leq \|\boldsymbol{\Psi}(\boldsymbol{x} + \boldsymbol{y})\|^2 \leq (1+\epsilon)^2 \|\boldsymbol{x} + \boldsymbol{y}\|^2, \\ (1-\epsilon)^2 \|\boldsymbol{x} - \boldsymbol{y}\|^2 \leq \|\boldsymbol{\Psi}(\boldsymbol{x} - \boldsymbol{y})\|^2 \leq (1+\epsilon)^2 \|\boldsymbol{x} - \boldsymbol{y}\|^2. \end{aligned}$$

The bound for $\left| \langle \boldsymbol{x}, \boldsymbol{y} \rangle - \langle \boldsymbol{\Psi}\boldsymbol{x}, \boldsymbol{\Psi}\boldsymbol{y} \rangle \right|$ then follows by noting that $(1-\epsilon)^2 \geq 1 - 3\epsilon$, $(1+\epsilon)^2 \leq 1 + 3\epsilon$ and $\|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2(\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2)$. Finally, a union bound over all k^2 subspaces and $2N$ data points yields the proposition. □

Proof of Proposition 2. Fix $\mathcal{U} \subseteq \mathbb{R}^d$ to be any subspace of dimension at most r' and let $\mathbf{U} \in \mathbb{R}^{d \times r}$ be an orthonormal basis of $\mathcal{U}$. Let $\widetilde{\boldsymbol{\Psi}} = \sqrt{p}\boldsymbol{\Psi}$ denote the unnormalized version of $\boldsymbol{\Psi}$. Since each entry in $\widetilde{\boldsymbol{\Psi}}$ follows i.i.d. standard Gaussian distribution and $\mathbf{U}$ is orthogonal, the projected matrix $\widetilde{\boldsymbol{\Psi}}\mathbf{U} \in \mathbb{R}^{p \times r'}$ follows an entrywise standard Gaussian distribution, too. By Lemma 28 (taking $t = \sqrt{2\delta}$ and scale the matrix by $1/\sqrt{p}$), the singular values of the Gaussian random matrix $\boldsymbol{\Psi}$ obey

$$1 - \sqrt{\frac{r'}{p}} - \sqrt{\frac{2\log(1/\delta)}{p}} \leq \sigma_{r'}(\boldsymbol{\Psi}) \leq \sigma_1(\boldsymbol{\Psi}) \leq 1 + \sqrt{\frac{r'}{p}} + \sqrt{\frac{2\log(1/\delta)}{p}} \tag{A.3}$$

with probability at least $1 - \delta$. Let $\epsilon := \sqrt{\frac{r'}{p}} + \sqrt{\frac{2\log(1/\delta)}{p}}$, then with the same probability, (supposing $\boldsymbol{x} = \mathbf{U}\boldsymbol{\alpha} \in \mathcal{U}$)

$$\begin{aligned} \left| \|\boldsymbol{x}\|_2^2 - \|\boldsymbol{\Psi}\boldsymbol{x}\|_2^2 \right| &= \left| \boldsymbol{\alpha}^\top \mathbf{U}^\top \mathbf{U} \boldsymbol{\alpha} - \boldsymbol{\alpha}^\top \mathbf{U}^\top \boldsymbol{\Psi}^\top \boldsymbol{\Psi} \mathbf{U} \boldsymbol{\alpha} \right| \\ &\leq \|\boldsymbol{\alpha}\|_2^2 \|\mathbf{U}^\top \mathbf{U} - \mathbf{U}^\top \boldsymbol{\Psi}^\top \boldsymbol{\Psi} \mathbf{U}\|_2 \\ &= \|\boldsymbol{x}\|_2^2 \|\mathbf{I}_{r' \times r'} - \mathbf{U}^\top \boldsymbol{\Psi}^\top \boldsymbol{\Psi} \mathbf{U}\|_2 \\ &\leq \epsilon \|\boldsymbol{x}\|_2^2. \end{aligned} \tag{A.4}$$

Subsequently,

$$(1-\epsilon)\|\boldsymbol{x}\| \leq \sqrt{1-\epsilon}\|\boldsymbol{x}\| \leq \|\boldsymbol{\Psi}\boldsymbol{x}\| \leq \sqrt{1+\epsilon}\|\boldsymbol{x}\| \leq (1+\epsilon)\|\boldsymbol{x}\|. \tag{A.5}$$

□

Proof of Proposition 5. Let $\Omega \subseteq \{1, \cdots, d\}$, $|\Omega| = p$ be the subsampling indices of $\boldsymbol{\Omega}$. By definition, $\Pr[\Omega(j) = i] = 1/d$ for every $i \in \{1, \cdots, d\}$ and $j \in \{1, \cdots, p\}$. Fix any subspace $\mathcal{U} \subseteq \mathbb{R}^d$ of dimension at most r' with incoherence level bounded by $\mu(\mathcal{U}) \leq \mu_0$. Let $\mathbf{U} \in \mathbb{R}^{d \times r'}$ be an orthonormal basis of $\mathcal{U}$. By definition, $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_{r' \times r'}$.

For any $\boldsymbol{x} \in \mathcal{U}$, there exists $\boldsymbol{\alpha} \in \mathbb{R}^{r'}$ such that $\boldsymbol{x} = \mathbf{U}\boldsymbol{\alpha}$. Subsequently, we have

$$\left| \|\boldsymbol{x}\|^2 - \|\boldsymbol{\Omega}\boldsymbol{x}\|^2 \right| = \left| \boldsymbol{\alpha}^\top \boldsymbol{\alpha} - \boldsymbol{\alpha}^\top (\boldsymbol{\Omega}\mathbf{U})^\top (\boldsymbol{\Omega}\mathbf{U}) \boldsymbol{\alpha} \right| \leq \|\boldsymbol{\alpha}\|^2 \cdot \|\mathbf{I} - (\boldsymbol{\Omega}\mathbf{U})^\top (\boldsymbol{\Omega}\mathbf{U})\|. \tag{A.6}$$

Our next objective is to bound the norm $\|\mathbf{I} - (\boldsymbol{\Omega}\mathbf{U})^\top (\boldsymbol{\Omega}\mathbf{U})\|$ with high probability. First let $\mathbf{U}_\Omega := (\boldsymbol{u}_{\Omega(1)}, \cdots, \boldsymbol{u}_{\Omega(p)}) = \sqrt{\frac{p}{d}} (\boldsymbol{\Omega}\mathbf{U})^\top$ be the unnormalized version of subsampled orthogonal operators. By definition we have

$$\|(\boldsymbol{\Omega}\mathbf{U})^\top (\boldsymbol{\Omega}\mathbf{U}) - \mathbf{I}\| = \frac{d}{p} \left\| \mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I} \right\|. \tag{A.7}$$

With Eq. (A.7), we can use noncommutative Matrix Berstein inequality (Gross et al., 2010; Recht, 2011) to bound $\|\mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I}\|$ and subsequently obtain an upper bound for the rightmost term in Eq. (A.6). The proof is very similar to the one presented in (Balzano et al., 2010; Krishnamurthy & Singh, 2014), where an upper bound for $\|(\mathbf{U}_\Omega \mathbf{U}_\Omega^\top)^{-1}\|$ is obtained. More specifically, let $\mathbf{B}_1, \cdots, \mathbf{B}_p$ be i.i.d. random matrices such that $\mathbf{B}_j = \boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top - \frac{1}{d}\mathbf{I}$. We then have

$$\mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I} = \sum_{j=1}^{p} \mathbf{B}_j \tag{A.8}$$

and furthermore,

$$\mathbb{E}\left[\mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I} \right] = p \left(\sum_{i=1}^{d} \boldsymbol{u}_i \boldsymbol{u}_i^\top - \mathbf{I} \right) = \mathbf{0}. \tag{A.9}$$

To use Matrix Bernstein, we need to upper bound the range and variance parameters of $\mathbf{B}_j$. Under the matrix incoherence assumption Eq. (3.5) the range of $\mathbf{B}_j$ can be bounded as

$$\|\mathbf{B}_j\| \leq \max_i \left\| \boldsymbol{u}_i \boldsymbol{u}_i^\top - \frac{1}{d}\mathbf{I} \right\| \leq \frac{\sqrt{r'^2}\mu_0}{d} + \frac{1}{d} \leq \frac{r'\mu_0}{d} + \frac{1}{d} \leq \frac{2r'\mu_0}{d} =: R. \tag{A.10}$$

The last inequality is due to the fact that $1 \leq \mu(\mathbf{U}) \leq \frac{d}{r'}$ for any subspace $\mathcal{U}$ of rank r'. For the variance, we have

$$\begin{aligned}
\|\mathbb{E}[\mathbf{B}_j^\top \mathbf{B}_j]\| = \|\mathbb{E}[\mathbf{B}_j \mathbf{B}_j^\top]\| &= \left\| \mathbb{E}\left[\left(\boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top - \frac{1}{d}\mathbf{I} \right) \left(\boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top - \frac{1}{d}\mathbf{I} \right) \right] \right\| \\
&= \left\| \mathbb{E}\left[\boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top \boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top \right] - \frac{1}{d^2}\mathbf{I} \right\| \\
&\leq \left\| \mathbb{E}\left[\boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top \boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top \right] \right\| + \frac{1}{d^2} \\
&\leq \frac{\mu_0 \sqrt{r'^2}}{d^2} \|\mathbb{E}[\boldsymbol{u}_{\Omega(j)} \boldsymbol{u}_{\Omega(j)}^\top]\| + \frac{1}{d^2} \\
&\leq \frac{\mu_0 r'}{d^2} + \frac{1}{d^2} \leq \frac{2\mu_0 r'}{d^2}.
\end{aligned}$$

As a result, we can define $\sigma^2 := 2\mu_0 r'/d^2$ such that $\sigma^2 \geq \max\{\|\mathbb{E}[\mathbf{B}_j \mathbf{B}_j^\top]\|, \|\mathbb{E}[\mathbf{B}_j^\top \mathbf{B}_j]\|\}$ for every j. Using Lemma 27, for every $t > 0$ we have

$$\Pr\left[\|\mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I}\| \geq t \right] \leq 2r \exp\left(-\frac{t^2/2}{\sigma^2 p + Rp/3} \right) = 2r' \exp\left(-\frac{t^2/2}{\frac{2\mu_0 r'}{d^2} p + \frac{2\mu_0 r'}{d} t/3} \right). \tag{A.11}$$

For $\epsilon < 1$ set $t = \frac{p}{d}\epsilon$ and $p = 8\epsilon^{-2}\mu_0 r' \log(2r'/\delta)$. Then with probability $\geq 1 - \delta$ we have

$$\left\| \mathbf{U}_\Omega \mathbf{U}_\Omega^\top - \frac{p}{d}\mathbf{I} \right\| \leq \frac{p}{d}\epsilon. \tag{A.12}$$

The proof is then completed by multiplying both sides in Eq. (A.12) by $\frac{d}{p}$.

□

A.2. Proof of the main theorems in Section 4

In this section we give rigorous proofs of the three key lemmas in Section 4. We also prove Theorem 15 and 18, which are simple corollaries of Lemma 12, 14 and 16.

Proof of Lemma 12. Fix $\ell \in [k]$ and one column $\boldsymbol{x}_i$ in $\mathbf{X}$. Let $\mathcal{U}^{(\ell)}$ and $\widetilde{\mathcal{U}}^{(\ell)}$ denote the low-rank subspaces to which $\boldsymbol{x}_i$ belongs before and after compression. That is, $\widetilde{\mathcal{U}}^{(\ell)} = \{\boldsymbol{\Psi}\boldsymbol{x} : \boldsymbol{x} \in \mathcal{U}^{(\ell)}\}$.

First note that $(1-2\lambda)^2 \leq \|\boldsymbol{\nu}\|^2 \leq 1/(2\lambda)$. $\|\boldsymbol{\nu}\| \geq 1-2\lambda$ because $\langle \boldsymbol{x}, \boldsymbol{\nu}\rangle - 2\lambda\|\boldsymbol{\nu}\|^2 \leq \|\boldsymbol{\nu}\|$ and putting $\boldsymbol{\nu} = \boldsymbol{x}$ we obtain a solution with value $1-2\lambda$. On the other hand, $\langle \boldsymbol{x}, \boldsymbol{\nu}\rangle - 2\lambda\|\boldsymbol{\nu}\|^2 \leq \|\boldsymbol{\nu}\| - 2\lambda\|\boldsymbol{\nu}\|^2$ and putting $\boldsymbol{\nu} = \mathbf{0}$ we obtain a solution with value 0. Also, under the noiseless setting $\boldsymbol{\nu} \in \mathcal{U}^{(\ell)}$, if $\boldsymbol{x} \in \mathcal{U}^{(\ell)}$.

Define $\tilde{\boldsymbol{\nu}}' = \frac{\sqrt{1-\epsilon}}{1+\epsilon\max(1,\|\boldsymbol{\nu}\|)} \cdot \tilde{\boldsymbol{\nu}}$, where $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Psi}\boldsymbol{\nu}$. Let $f(\boldsymbol{\nu}) = \langle \boldsymbol{\nu}, \boldsymbol{x}\rangle - \frac{\lambda}{2}\|\boldsymbol{\nu}\|_2^2$ and $\tilde{f}(\tilde{\boldsymbol{\nu}}') = \langle \tilde{\boldsymbol{\nu}}', \tilde{\boldsymbol{x}}'\rangle - \frac{\lambda}{2}\|\tilde{\boldsymbol{\nu}}'\|_2^2$ denote the values of the optimization problems. The first step is to prove that $\tilde{\boldsymbol{\nu}}$ is feasible and nearly optimal to the projected optimization problem; that is, $\tilde{f}(\tilde{\boldsymbol{\nu}}')$ is close to $\tilde{f}(\boldsymbol{\nu}^*)$.

We first show that $\tilde{\boldsymbol{\nu}}'$ is a feasible solution with high probability. By Proposition 1, the following bound on $|\tilde{\boldsymbol{x}}_i^\top \tilde{\boldsymbol{\nu}}|$ holds:

$$|\tilde{\boldsymbol{x}}_i^\top \tilde{\boldsymbol{\nu}}| \leq |\boldsymbol{x}_i, \boldsymbol{\nu}| + \epsilon \cdot \frac{\|\boldsymbol{x}_i\| + \|\boldsymbol{\nu}\|}{2} \leq 1 + \epsilon\max(1, \|\boldsymbol{\nu}\|). \quad \forall \boldsymbol{x}_i \in \mathbf{X}. \tag{A.13}$$

Furthermore, with probability $\geq 1-\delta$

$$\|\tilde{\boldsymbol{x}}_i\|_2^2 \geq (1-\epsilon)\|\boldsymbol{x}_i\|_2^2 = 1-\epsilon. \tag{A.14}$$

Consequently, by the definition of $\tilde{\boldsymbol{\nu}}'$ one has

$$\|\widetilde{\mathbf{X}}'^\top \tilde{\boldsymbol{\nu}}'\|_\infty \leq \frac{1}{\sqrt{1-\epsilon}} \cdot \frac{\sqrt{1-\epsilon}}{1+\epsilon\max(1,\|\boldsymbol{\nu}\|)} \|\widetilde{\mathbf{X}}^\top \tilde{\boldsymbol{\nu}}\|_\infty \leq 1. \tag{A.15}$$

Next, we compute a lower bound on $\tilde{f}(\tilde{\boldsymbol{\nu}}')$, which serves as a lower bound for $\tilde{f}(\boldsymbol{\nu}^*)$ because $\boldsymbol{\nu}^*$ is the optimal solution to the dual optimization problem on the projected data.

$$\begin{aligned}
\tilde{f}(\tilde{\boldsymbol{\nu}}') &= \langle \tilde{\boldsymbol{x}}', \tilde{\boldsymbol{\nu}}'\rangle - \frac{\lambda}{2}\|\tilde{\boldsymbol{\nu}}'\|_2^2 \\
&\geq \sqrt{\frac{1-\epsilon}{1+\epsilon}} \frac{\langle \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{\nu}}\rangle}{1+\epsilon\max(1,\|\boldsymbol{\nu}\|)} - \frac{\lambda}{2}(1-\epsilon)\|\tilde{\boldsymbol{\nu}}\|_2^2 \\
&\geq (1-\epsilon)(1-\epsilon\max(1,\|\boldsymbol{\nu}\|))\left(\langle \boldsymbol{x}, \boldsymbol{\nu}\rangle - \epsilon\max(1,\|\boldsymbol{\nu}\|)\right) - \frac{\lambda}{2}(1-\epsilon)(1+\epsilon)\|\boldsymbol{\nu}\|^2 \\
&\geq \langle \boldsymbol{x}, \boldsymbol{\nu}\rangle - \epsilon\max(1,\|\boldsymbol{\nu}\|) - \epsilon\left(\langle \boldsymbol{x}, \boldsymbol{\nu}\rangle - \epsilon\max(1,\|\boldsymbol{\nu}\|)\right) - \frac{\lambda}{2}\|\boldsymbol{\nu}\|^2 \\
&\geq f(\boldsymbol{\nu}) - 2\epsilon\max(1,\|\boldsymbol{\nu}\|).
\end{aligned} \tag{A.16}$$

On the other hand, since $\boldsymbol{\nu}^* \in \widetilde{\mathcal{U}}^{(\ell)}$, there exists $\bar{\boldsymbol{\nu}} \in \mathcal{U}^{(\ell)}$ such that $\boldsymbol{\nu}^* = \boldsymbol{\Psi}\bar{\boldsymbol{\nu}}$. Let $\bar{\boldsymbol{\nu}}'$ be a scaled version of $\bar{\boldsymbol{\nu}}$ so that it is a feasible solution to the optimization problem in Eq. (4.1) before projection. Using essentially similar analysis one can show that $f(\bar{\boldsymbol{\nu}}') \geq \tilde{f}(\boldsymbol{\nu}^*) - 2\epsilon\max(1,\|\boldsymbol{\nu}^*\|)$. Consequently, the following bound on the gap between $\tilde{f}(\tilde{\boldsymbol{\nu}}')$ and $\tilde{f}(\boldsymbol{\nu}^*)$ holds:

$$\left|\tilde{f}(\tilde{\boldsymbol{\nu}}') - \tilde{f}(\boldsymbol{\nu}^*)\right| \leq 4\epsilon\max(1,\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|). \tag{A.17}$$

Because the dual problem in Eq. (4.1) is strongly convex with parameter λ (this holds for both the projected and the original problem), we can bound the perturbation of dual directions $\|\tilde{\boldsymbol{\nu}}' - \boldsymbol{\nu}^*\|$ by the bounds on their values $|\tilde{f}(\tilde{\boldsymbol{\nu}}') - \tilde{f}(\boldsymbol{\nu}^*)|$ as

$$\|\tilde{\boldsymbol{\nu}}' - \boldsymbol{\nu}^*\|_2 \leq \sqrt{\frac{2|\tilde{f}(\tilde{\boldsymbol{\nu}}') - \tilde{f}(\boldsymbol{\nu}^*)|}{\lambda}} \leq \sqrt{\frac{8\epsilon\max(1,\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|)}{\lambda}}. \tag{A.18}$$

Next, note that $\tilde{\boldsymbol{\nu}}', \boldsymbol{\nu}^* \in \widetilde{\mathcal{U}}^{(\ell)}$. Also note that for any two vector $\boldsymbol{a}, \boldsymbol{b}$ the following holds:

$$\left\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} - \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\right\| = \left\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|} - \frac{\boldsymbol{b}}{\|\boldsymbol{a}\|} + \frac{\boldsymbol{b}}{\|\boldsymbol{a}\|} - \frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\right\|$$

$$\begin{aligned} &\leq \frac{\|\boldsymbol{a}-\boldsymbol{b}\|}{\|\boldsymbol{a}\|}+\frac{\|\boldsymbol{b}\|\cdot\big|\|\boldsymbol{a}\|-\|\boldsymbol{b}\|\big|}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} \\ &\leq \frac{\|\boldsymbol{a}-\boldsymbol{b}\|}{\|\boldsymbol{a}\|}+\frac{\|\boldsymbol{a}-\boldsymbol{b}\|}{\|\boldsymbol{a}\|} \\ &= \frac{2\|\boldsymbol{a}-\boldsymbol{b}\|}{\|\boldsymbol{a}\|}. \end{aligned}$$

By symmetry we also have $\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}-\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\|\leq\frac{2\|\boldsymbol{a}-\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$. Therefore,

$$\left\|\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}-\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|}\right\|\leq\frac{2\|\boldsymbol{a}-\boldsymbol{b}\|}{\max(\|\boldsymbol{a}\|,\|\boldsymbol{b}\|)}. \tag{A.19}$$

Now we can bound $\|\tilde{\boldsymbol{v}}'-\boldsymbol{v}^*\|$ as follows:

$$\begin{aligned} \|\tilde{\boldsymbol{v}}'-\boldsymbol{v}^*\| &= \left\|\frac{\tilde{\boldsymbol{\nu}}'}{\|\tilde{\boldsymbol{\nu}}'\|}-\frac{\boldsymbol{\nu}^*}{\|\boldsymbol{\nu}^*\|}\right\| \\ &\leq \frac{2\|\tilde{\boldsymbol{\nu}}'-\boldsymbol{\nu}^*\|}{\max(\|\tilde{\boldsymbol{\nu}}'\|,\|\boldsymbol{\nu}^*\|)}\leq\frac{2\|\tilde{\boldsymbol{\nu}}'-\boldsymbol{\nu}^*\|}{\max(\|\boldsymbol{\nu}\|/4,\|\boldsymbol{\nu}^*\|)} \\ &\leq \frac{16\sqrt{2\epsilon\max(1,\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|)}}{\sqrt{\lambda}\max(\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|)}\leq 16\sqrt{\frac{2\epsilon}{\lambda}\max\left(1,\frac{1}{\|\boldsymbol{\nu}\|},\frac{1}{\|\boldsymbol{\nu}^*\|}\right)} \\ &\leq 16\sqrt{\frac{2\epsilon}{\lambda(1-2\lambda)}}\leq 32\sqrt{\frac{\epsilon}{\lambda}}. \end{aligned}$$

Note that after normalization $\tilde{\boldsymbol{v}}'$ is exactly the same with $\tilde{\boldsymbol{v}}$. Subsequently, for any $\boldsymbol{y}\in\mathbf{X}\backslash\mathbf{X}^{(\ell)}$ we have

$$\begin{aligned} \big|\langle\boldsymbol{v},\boldsymbol{y}\rangle-\langle\boldsymbol{v}^*,\tilde{\boldsymbol{y}}'\rangle\big| &\leq \big|\langle\tilde{\boldsymbol{v}}',\tilde{\boldsymbol{y}}'\rangle-\langle\boldsymbol{v}^*,\tilde{\boldsymbol{y}}'\rangle\big|+\big|\langle\tilde{\boldsymbol{v}}',\tilde{\boldsymbol{y}}'\rangle-\langle\boldsymbol{v},\boldsymbol{y}\rangle\big| \\ &\leq \|\tilde{\boldsymbol{v}}'-\boldsymbol{v}^*\|\|\tilde{\boldsymbol{y}}'\|+\big|\langle\tilde{\boldsymbol{v}},\tilde{\boldsymbol{y}}'\rangle-\langle\boldsymbol{v},\boldsymbol{y}\rangle\big| \\ &\leq \|\tilde{\boldsymbol{v}}'-\boldsymbol{v}^*\|+\left|\frac{1}{\|\boldsymbol{\Psi}\boldsymbol{v}\|\|\boldsymbol{\Psi}\boldsymbol{y}\|}\langle\boldsymbol{\Psi}\boldsymbol{v},\boldsymbol{\Psi}\boldsymbol{y}\rangle-\langle\boldsymbol{v},\boldsymbol{y}\rangle\right| \\ &\leq 32\sqrt{\frac{\epsilon}{\lambda}}+\left(1-\frac{1}{\|\boldsymbol{\Psi}\boldsymbol{v}\|\|\boldsymbol{\Psi}\boldsymbol{y}\|}\right)\|\boldsymbol{\Psi}\boldsymbol{v}\|\|\boldsymbol{\Psi}\boldsymbol{y}\|+\big|\langle\boldsymbol{\Psi}\boldsymbol{v},\boldsymbol{\Psi}\boldsymbol{y}\rangle-\langle\boldsymbol{v},\boldsymbol{y}\rangle\big| \\ &\leq 32\sqrt{\frac{\epsilon}{\lambda}}+\left(1-\frac{1}{1+\epsilon}\right)(1+\epsilon)+\epsilon \\ &= 32\sqrt{\frac{\epsilon}{\lambda}}+2\epsilon. \end{aligned}$$

□

Proof of Lemma 14. For notational simplicity re-define $\mathbf{Y}=\mathbf{Y}_{(-i)}$ and $\widetilde{\mathbf{Y}}'=\widetilde{\mathbf{Y}}'_{(-i)}$ for some fixed data point $\boldsymbol{x}_i^{(\ell)}$. Let $\mathcal{C}$, $\widetilde{\mathcal{C}}$ be the largest Euclidean balls inscribed in $\mathcal{Q}(\mathbf{Y})$ and $\mathcal{Q}(\widetilde{\mathbf{Y}})$. Since both $\mathcal{Q}(\mathbf{Y})$ and $\mathcal{Q}(\widetilde{\mathbf{Y}})$ are symmetric convex bodies with respect to the origin, the centers of $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ are the origin. Let $\tilde{\boldsymbol{c}}$ be any point in $\widetilde{\mathcal{C}}\cap\partial\mathcal{Q}(\widetilde{\mathbf{Y}})$. By definition, $r(\mathcal{Q}(\widetilde{\mathbf{Y}}))=\|\tilde{\boldsymbol{c}}\|$. Since $\tilde{\boldsymbol{c}}\in\widetilde{\mathcal{U}}^{(\ell)}$, we can find $\boldsymbol{c}\in\mathcal{U}^{(\ell)}$ such that $\tilde{\boldsymbol{c}}=\boldsymbol{\Psi}\boldsymbol{c}$. By Proposition 1, we have (with probability $\geq 1-\delta$)

$$\|\tilde{\boldsymbol{c}}\|\geq\frac{1}{\sqrt{1+\epsilon}}\|\boldsymbol{c}\|. \tag{A.20}$$

On the other hand, $\boldsymbol{c}$ is not contained in the interior of $\mathcal{Q}(\mathbf{Y})$. Otherwise, we can find a scalar $a>1$ such that $a\boldsymbol{c}\in\mathcal{Q}(\mathbf{Y})$ and hence $a\tilde{\boldsymbol{c}}\in\mathcal{Q}(\widetilde{\mathbf{Y}})$, contradicting the fact that $\tilde{\boldsymbol{c}}\in\partial\mathcal{Q}(\widetilde{\mathbf{Y}})$. Consequently, we have $\|\boldsymbol{c}\|\geq r(\mathcal{Q}(\mathbf{Y}))$ by definition. Therefore,

$$r(\mathcal{Q}(\widetilde{\mathbf{Y}}))=\|\tilde{\boldsymbol{c}}\|\geq\frac{1}{\sqrt{1+\epsilon}}\|\boldsymbol{c}\|\geq\frac{r(\mathcal{Q}(\mathbf{Y}))}{\sqrt{1+\epsilon}}. \tag{A.21}$$

Next, we need to lower bound $r(\mathcal{Q}(\widetilde{\mathbf{Y}}'))$ in terms of $r(\mathcal{Q}(\widetilde{\mathbf{Y}}))$. This can be easily done by noting that the maximum column norm in $\widetilde{\mathbf{Y}}$ is upper bounded by $\sqrt{1+\epsilon}$. Consequently, we have

$$r(\mathcal{Q}(\widetilde{\mathbf{Y}}')) \geq r\left(\mathcal{Q}\left(\frac{1}{\sqrt{1+\epsilon}}\widetilde{\mathbf{Y}}\right)\right) \geq \frac{r(\mathcal{Q}(\mathbf{Y}))}{1+\epsilon}. \tag{A.22}$$

□

Proof of Lemma 16. Fix $\ell \in \{1, 2, \cdots, k\}$ and a particular column $\boldsymbol{x} = \boldsymbol{x}_i$. Suppose $\boldsymbol{\nu}$ is the optimal solution to the original dual problem in Eq. (4.1). Define $\boldsymbol{\nu}_{\|} = \mathcal{P}_{\mathcal{U}^{(\ell)}}\boldsymbol{\nu}$ and $\boldsymbol{\nu}_{\perp} = \mathcal{P}_{\mathcal{U}^{(\ell)\perp}}\boldsymbol{\nu}$. Let $f(\cdot)$ be the objective value of the dual problem under a specific solution. Then it is easy to observe that

$$f(\boldsymbol{\nu}_{\|}) \geq f(\boldsymbol{\nu}) - \langle \boldsymbol{x}_{\perp}, \boldsymbol{\nu}_{\perp}\rangle \geq f(\boldsymbol{\nu}) - \eta\|\boldsymbol{\nu}_{\perp}\|_2. \tag{A.23}$$

We then cite the following upper bound for $\|\boldsymbol{\nu}_{\perp}\|$, which appears as Eq. (5.16) in (Wang & Xu, 2013).

$$\|\boldsymbol{\nu}_{\perp}\|_2 \leq \lambda\eta\left(\frac{1}{r(\mathcal{Q}(\mathbf{Y}_{-i}^{(\ell)}))}+1\right) \leq \frac{2\lambda\eta}{\rho_\ell}. \tag{A.24}$$

Let $\tilde{\boldsymbol{\nu}} = \boldsymbol{\Psi}\boldsymbol{\nu}_{\|}$ and $\tilde{\boldsymbol{\nu}}' = \frac{\sqrt{1-\epsilon}}{1+(\eta+\epsilon)\max(1,\|\boldsymbol{\nu}\|)} \cdot \tilde{\boldsymbol{\nu}}$. It is easy to verify that $\tilde{\boldsymbol{\nu}}'$ is a feasible solution to the projected dual problem. Define $\eta' := \max_{i=1,\cdots,n}\|\tilde{\boldsymbol{z}}_i\|_2$. Since $\boldsymbol{\Psi}$ is well behaved, $\eta' \leq \sqrt{1+\epsilon}\eta$ with high probability. Applying essentially the same chain of argument as in the proof of Lemma 12 we obtain

$$\begin{aligned}
\tilde{f}(\tilde{\boldsymbol{\nu}}') &= \langle \tilde{\boldsymbol{x}}', \tilde{\boldsymbol{\nu}}'\rangle - \frac{\lambda}{2}\|\tilde{\boldsymbol{\nu}}'\|_2^2 \\
&= \langle \tilde{\boldsymbol{y}}', \tilde{\boldsymbol{\nu}}'\rangle + \langle \tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\nu}}'\rangle - \frac{\lambda}{2}\|\tilde{\boldsymbol{\nu}}'\|_2^2 \\
&\geq \langle \boldsymbol{y}, \boldsymbol{\nu}_{\|}\rangle - \frac{\lambda}{2}\|\boldsymbol{\nu}_{\|}\|^2 - 2(\epsilon+\eta)\max(1,\|\boldsymbol{\nu}\|) - \|\tilde{\boldsymbol{z}}\|_2\|\tilde{\boldsymbol{\nu}}'\|_2 \\
&\geq \langle \boldsymbol{y}, \boldsymbol{\nu}_{\|}\rangle - \frac{\lambda}{2}\|\boldsymbol{\nu}_{\|}\|^2 - 2(\epsilon+\eta)\max(1,\|\boldsymbol{\nu}\|) - \eta' \cdot \frac{\sqrt{(1-\epsilon)(1+\epsilon)}}{1+\epsilon\max(1,\|\boldsymbol{\nu}\|)}\|\boldsymbol{\nu}\| \\
&\geq \langle \boldsymbol{y}, \boldsymbol{\nu}_{\|}\rangle - \frac{\lambda}{2}\|\boldsymbol{\nu}_{\|}\|^2 - 2(\epsilon+\eta)\max(1,\|\boldsymbol{\nu}\|) - \sqrt{1+\epsilon}\eta\|\boldsymbol{\nu}\| \\
&\geq \langle \boldsymbol{x}, \boldsymbol{\nu}_{\|}\rangle - \frac{\lambda}{2}\|\boldsymbol{\nu}_{\|}\|^2 - 2(\epsilon+\eta)\max(1,\|\boldsymbol{\nu}\|) - \sqrt{1+\epsilon}\eta\|\boldsymbol{\nu}\| - \eta\|\boldsymbol{\nu}\| \\
&\geq f(\boldsymbol{\nu}_{\|}) - (2\epsilon+5\eta)\max(1,\|\boldsymbol{\nu}\|) \\
&\geq f(\boldsymbol{\nu}) - \frac{2\lambda\eta^2}{\rho_\ell} - (2\epsilon+5\eta)\max(1,\|\boldsymbol{\nu}\|).
\end{aligned}$$

Similarly, one can show that

$$\tilde{f}(\boldsymbol{\nu}^*) \leq f(\boldsymbol{\nu}) + \frac{2\lambda\eta'^2}{\rho_\ell} + (2\epsilon+5\eta')\max(1,\|\boldsymbol{\nu}^*\|) \leq f(\boldsymbol{\nu}) + \frac{3\lambda\eta^2}{\rho_\ell} + (2\epsilon+6\eta)\max(1,\|\boldsymbol{\nu}^*\|). \tag{A.25}$$

Consequently, noting that $\tilde{f}(\tilde{\boldsymbol{\nu}}') \leq \tilde{f}(\boldsymbol{\nu}^*)$ one has

$$\left|\tilde{f}(\boldsymbol{\nu}^*) - \tilde{f}(\tilde{\boldsymbol{\nu}}')\right| \leq \frac{5\lambda\eta^2}{\rho_\ell} + 4(\epsilon+3\eta)\max(1,\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|). \tag{A.26}$$

Since both dual problems (before and after projection) are strongly convex with parameter λ, the following perturbation bound on $\|\boldsymbol{\nu}^* - \tilde{\boldsymbol{\nu}}'\|$ holds:

$$\|\boldsymbol{\nu}^* - \tilde{\boldsymbol{\nu}}'\| \leq \sqrt{\frac{2|\tilde{f}(\boldsymbol{\nu}^*) - \tilde{f}(\tilde{\boldsymbol{\nu}}')|}{\lambda}} \leq \sqrt{\frac{5\eta^2}{\rho_\ell} + \frac{8(\epsilon+3\eta)\max(1,\|\boldsymbol{\nu}\|,\|\boldsymbol{\nu}^*\|)}{\lambda}}. \tag{A.27}$$

Subsequently,

$$\begin{aligned}\|\tilde{\boldsymbol{v}}' - \boldsymbol{v}^*\| &\leq \frac{8\|\tilde{\boldsymbol{\nu}}' - \boldsymbol{\nu}^*\|}{\max(\|\boldsymbol{\nu}\|, \|\boldsymbol{\nu}^*\|)} \\ &\leq 8\sqrt{\frac{5\eta^2}{\rho_\ell \max(\|\boldsymbol{\nu}\|^2, \|\boldsymbol{\nu}^*\|^2)} + \frac{8(\epsilon+3\eta)}{\lambda \max(1, \|\boldsymbol{\nu}\|^2, \|\boldsymbol{\nu}^*\|^2)}} \\ &\leq 8\sqrt{\frac{5\eta^2}{\rho_\ell(1-2\lambda)^2} + \frac{8(\epsilon+3\eta)}{\lambda(1-2\lambda)^2}} \\ &\leq 16\sqrt{\frac{5\eta^2}{\rho_\ell} + \frac{8(\epsilon+3\eta)}{\lambda}}.\end{aligned}$$

Finally, the perturbation of the angle between $\boldsymbol{v}$ and $\boldsymbol{y}$ can be bounded by

$$\left|\langle \boldsymbol{v}, \boldsymbol{y}\rangle - \langle \boldsymbol{v}^*, \tilde{\boldsymbol{y}}'\rangle\right| \leq \|\tilde{\boldsymbol{v}}' - \tilde{\boldsymbol{v}}^*\| + 2\epsilon \leq 16\sqrt{\frac{5\eta^2}{\rho_\ell} + \frac{8(\epsilon+3\eta)}{\lambda}} + 2\epsilon. \tag{A.28}$$

□

Proof of Theorem 15. Let $\tilde{\mu}_\ell, \tilde{\rho}_\ell$ denote the subspace incoherence and inradius of subspace $\mathcal{U}^{(\ell)}$ after dimensionality reduction. Theorem 11 shows that Lasso SSC satisfies the subspace detection property if $\tilde{\mu}_\ell < \tilde{\rho}_\ell$ for every ℓ and $\lambda < \tilde{\rho}$. By Lemma 14, $\tilde{\rho} \geq \rho/2$ with high probability. Note also that $\tilde{\rho}_\ell \geq \frac{\rho_\ell}{1+\epsilon} \geq \rho_\ell(1-\epsilon)$. Subsequently, the following inequality yields $\tilde{\mu}_\ell < \tilde{\rho}_\ell$ for every ℓ:

$$\mu_\ell + 32\sqrt{\epsilon/\lambda} + (2+\rho_\ell)\epsilon < \rho_\ell, \quad \forall \ell = 1, \cdots, k. \tag{A.29}$$

Taking $32\sqrt{\epsilon/\lambda} < \Delta/2$ and $(2+\rho_\ell)\epsilon < \Delta/2$ where $\Delta = \min_\ell(\rho_\ell - \mu_\ell)$, Eq. (A.29) is subsequently satisfied. This yields

$$\epsilon < \min\left\{\frac{\Delta}{2(2+\rho)}, c_1\lambda\Delta^2\right\} \tag{A.30}$$

for some absolute constant c_1. The $\epsilon < 1/2$ term comes from the $\epsilon < 1/\|\boldsymbol{\nu}\|$ condition in Lemma 12. □

Proof of Theorem 18. Define $\tilde{\Delta} := \min_\ell(\tilde{\rho}_\ell - \tilde{\mu}_\ell)$ to be the maximum margin of error after dimensionality reduction. First we prove that with $\lambda = \rho/4 < 1/4$ and the upper bound in Eq. (4.15) we have $\tilde{\Delta} \geq \Delta/2$. Essentially, this requires

$$16\sqrt{\frac{5\eta^2}{\rho_\ell} + \frac{8(\epsilon+3\eta)}{\lambda}} < \frac{\Delta}{4}, \tag{A.31}$$

$$2\epsilon + \rho\epsilon < \frac{\Delta}{4}. \tag{A.32}$$

This amounts to

$$\epsilon < \min\left\{\frac{\Delta}{4(2+\rho)}, \frac{\lambda}{8}\left(c_2\Delta^2 - \frac{5\eta^2}{\rho}\right) - 3\eta\right\}, \tag{A.33}$$

where $c_2 > 0$ is an absolute constant.

Next we verify that Eq. (4.5) are satisfied after dimensionality reduction. Let $\tilde{\eta}$ denote the noise level after projection, that is, $\max_i\{\|\tilde{\boldsymbol{z}}_i\|\} \leq \tilde{\eta}$. Because $\epsilon < 1/3$, by Proposition 1 $\tilde{\eta} \leq 2\eta$ with high probability. Consequently, $\eta < \frac{\rho}{96}$ in Eq. (4.14) implies ($\tilde{\rho} = \min_\ell \tilde{\rho}_\ell$ and $\tilde{\mu} = \max_\ell \tilde{\mu}_\ell$)

$$\tilde{\rho} - 2\tilde{\eta} - \tilde{\eta}^2 \geq \rho(1-\epsilon) - 6\eta \geq \frac{2\rho}{3} - \frac{6\rho}{18} = \frac{\rho}{3} \geq \frac{\rho}{4} = \lambda. \tag{A.34}$$

Hence the upper bound on λ in Eq. (4.5) is satisfied. For the lower bound, note that $\eta \ll 1$, $\tilde{\rho}_\ell < 1$ and hence

$$\frac{\tilde{\eta}(1+\tilde{\eta})(2+\tilde{\rho}_\ell)}{\tilde{\rho}_\ell - \tilde{\mu}_\ell - 2\tilde{\eta}} \leq \frac{6\tilde{\eta}}{\tilde{\Delta}} \leq \frac{12\eta}{\Delta/2} = \frac{24\eta}{\Delta} < \frac{\rho}{4} = \lambda. \tag{A.35}$$

The last inequality is due to Eq. (4.14). □

Proof of Theorem 19. Let the JL transform matrix be $\mathbf{\Psi}$. Since it is a linear transformation, $\boldsymbol{z}_i \sim \mathcal{N}(0, \frac{\sigma^2}{d}\mathbf{I})$ implies that $\mathbf{\Psi}\boldsymbol{z}_i \sim \mathcal{N}(0, \frac{\sigma^2}{d}\mathbf{\Psi}\mathbf{\Psi}^\top)$. Using the fact that this algorithm is invariant to arbitrary unitary transformations, we can apply the rotation that diagonalizes the covariance matrix $\frac{\sigma^2}{d}\mathbf{\Psi}\mathbf{\Psi}^\top$ to every column of the projected (and renormalized) data. This decouples the noise matrix $\mathbf{Z}$ such that every coordinate is independent Gaussian. Moreover, the maximum entrywise variance is upper bounded by

$$\max_{ij}\frac{\sigma_{ij}^2}{p} \le \|\mathbf{\Psi}\|^2\frac{\sigma^2(1+\epsilon)^2}{d} \le \xi^2\frac{d}{p}\frac{\sigma^2(1+\epsilon)^2}{d} \le \xi^2\frac{\sigma^2(1+\epsilon)^2}{p} \le 2\xi^2\frac{\sigma^2}{p},$$

where ϵ is the JL parameter included to acount for the renormalization of the y part. The last inequality holds because $\epsilon > 1/3$ by our assumption.

Applying the same argument as in the proof of Theorem 18 we get $\tilde{\Delta} = \min_\ell(\tilde{\rho}_\ell - \tilde{\mu}_\ell) \ge \Delta/2$ when Eq. (4.17) is satisfied.

The proof is then completed by invoking the second part of Theorem 11 on the compressed problem with the bounded entrywise independent Gausian noise, we get the condition that

$$\sqrt{\frac{\log N}{p}}\sigma(1+\sigma) \le \frac{C}{4\xi^2}\min_{\ell=1,\dots,k}\left\{\rho, r^{-1/2}, \rho_\ell - \mu_\ell\right\}$$

as claimed in (4.6).

Note that for random Gaussian transforms $\mathbf{\Psi}$, by Lemma 28, $\|\mathbf{\Psi}\| \le 3\sqrt{d/p}$ (hence $\xi^2 \le 9$) with high probability. □

B. Privacy preserved subspace clustering

In this section, we formalize the claims on attribute-level differential privacy and the corresponding utility guarantee in the paper.

Privacy Claim In classic statistical privacy literature, transforming data set $\mathbf{X}$ by taking $\tilde{\mathbf{X}} = \mathbf{A}\mathbf{X} + \mathbf{\Delta}$ for some random matrix $\mathbf{A}$ and $\mathbf{\Delta}$ is called *matrix masking*. (Zhou et al., 2009) show that random compression allows the mutual information of the output $\tilde{\mathbf{X}}$ and raw data $\mathbf{X}$ to converge to 0 with rate $O(p/d)$ even when $\mathbf{\Delta} = 0$, their result directly applies to our problem. The guarantee suggests that the amount of information in the compressed output $\tilde{\mathbf{X}}$ about the raw data $\mathbf{X}$ goes to 0 as the ambient dimension d gets large.

On the other hand, if $\mathbf{\Delta} \neq \mathbf{0}$ is an iid Gaussian noise matrix, we can protect the (ε, δ)-differential privacy of every data entry. Such attribute differential privacy notion is defined below.

Definition 20 (Attribute Differential Privacy)**.** *Suppose $\mathcal{O}$ is the set for all possible outcomes. We say a randomized algorithm $\mathcal{A} : \mathbb{R}^{d\times N} \to \mathcal{O}$ is (ε, δ)-differential private at attribute level if*

$$\mathbb{P}(\mathcal{A}(\mathbf{X}) \in \mathcal{S}) \le e^\varepsilon \mathbb{P}(\mathcal{A}(\mathbf{X}') \in \mathcal{S}) + \delta$$

for any measurable outcome $\mathcal{S} \subset \mathcal{O}$, any $\mathbf{X}$ and $\mathbf{X}'$ that differs in only one entry.

This is a well-studied setting in (Kenthapadi et al., 2013). It is weaker than protecting the privacy of individual users, which remains an open question, but much stronger than the average protection via mutual information. In fact, it forbids any feature of an individual user from being identified "for sure" by an adversary with arbitrary side information.

Theorem 21. *Assume the data (and all other users that we need to protect) satisfy column spikiness conditions with parameter μ_0 as in Definition 4. Let $\mathbf{\Psi}$ be a Johnson-Lindenstrauss transform with parameter ϵ. Releasing compressed data $\widetilde{\mathbf{X}}' = \mathrm{Normalize}(\mathbf{\Psi}\mathbf{X}) + \mathcal{N}(0, \sigma^2\mathbf{I}_{p\times d})$ with $\sigma = \frac{1+\epsilon}{1-\epsilon}\sqrt{\frac{32\mu_0\log(1.25/\delta)}{d\varepsilon^2}}$ preserves attribute-level (ε, δ)-differential privacy.*

The proof involves working out the ℓ_2-sensitivity of the operator $\mathrm{Normalize}(\mathbf{\Psi}(\cdot))$ in terms of column incoherence μ_0 and apply "Gaussian Mechanism". By the closeness to post-processing property of differential privacy, the subsequent subspace clustering results protects the same level of privacy. Details are given as follows.

Proof of Theorem 21. Let $\mathbf{X}$ and $\mathbf{X}'$ differs by only one entry, w.l.o.g, assume it is the ith column and jth row,

$$\|\mathbf{\Psi}(\mathbf{X}-\mathbf{X}')\|_F = \|\mathbf{\Psi}(\mathbf{X}_i-\mathbf{X}_i')\|_2 \le \|\mathbf{\Psi}\boldsymbol{e}_j\||\mathbf{X}_{ji}-\mathbf{X}'_{ji}| \le 2\sqrt{\frac{\mu}{d}}\|\mathbf{\Psi}\boldsymbol{e}_j\|.$$

Now we derive the ℓ_2-sensitivity of Normalize$(\mathbf{\Psi}(\cdot))$.

$$\begin{aligned}
&\|\text{Normalize}(\mathbf{\Psi}\mathbf{X}) - \text{Normalize}(\mathbf{\Psi}\mathbf{X}')\|_F \\
=&\left\|\frac{\mathbf{\Psi}\mathbf{X}_i}{\|\mathbf{\Psi}\mathbf{X}_i\|} - \frac{\mathbf{\Psi}\mathbf{X}_i'}{\|\mathbf{\Psi}\mathbf{X}_i'\|}\right\|_2 = \left\|\frac{\mathbf{\Psi}\mathbf{X}_i}{\|\mathbf{\Psi}\mathbf{X}_i\|} - \frac{\mathbf{\Psi}\mathbf{X}_i'}{\|\mathbf{\Psi}\mathbf{X}_i\|} + \frac{\mathbf{\Psi}\mathbf{X}_i'}{\|\mathbf{\Psi}\mathbf{X}_i\|} - \frac{\mathbf{\Psi}\mathbf{X}_i'}{\|\mathbf{\Psi}\mathbf{X}_i'\|}\right\|_2 \\
=&\left\|\frac{\mathbf{\Psi}(\mathbf{X}_i-\mathbf{X}_i')}{\|\mathbf{\Psi}\mathbf{X}_i\|} + \mathbf{\Psi}\mathbf{X}_i'\left(\frac{1}{\|\mathbf{\Psi}\mathbf{X}_i\|} - \frac{1}{\|\mathbf{\Psi}\mathbf{X}_i'\|}\right)\right\| \\
\le&\frac{\|\mathbf{\Psi}(\mathbf{X}_i-\mathbf{X}_i')\|_2}{\|\mathbf{\Psi}\mathbf{X}_i\|} + \|\mathbf{\Psi}\mathbf{X}_i'\|\frac{|\|\mathbf{\Psi}\mathbf{X}_i'\| - \|\mathbf{\Psi}\mathbf{X}_i\||}{\|\mathbf{\Psi}\mathbf{X}_i\|\|\mathbf{\Psi}\mathbf{X}_i'\|} \\
\le&\frac{2\|\mathbf{\Psi}(\mathbf{X}_i-\mathbf{X}_i')\|_2}{\|\mathbf{\Psi}\mathbf{X}_i\|} \le 4\sqrt{\frac{\mu_0}{d}}\frac{\|\mathbf{\Psi}\boldsymbol{e}_j\|}{\|\mathbf{\Psi}\mathbf{X}_i\|} \le 4\sqrt{\frac{\mu_0}{d}}\frac{1+\epsilon}{1-\epsilon}.
\end{aligned}$$

The last step uses the fact that $\mathbf{\Psi}$ is JL with parameter ϵ.

Lemma 22 (Gaussian Mechanism, (Kenthapadi et al., 2013))**.** *Let $\Delta_2 f$ be the ℓ_2 sensitivity of f, Let $\varepsilon \in (0,1)$ be arbitrary. The procedure that output $f(\mathbf{X}) + \mathcal{N}(0,\sigma^2 I)$ with $\sigma \ge \Delta_2 f\sqrt{2\log(1.25/\delta)}/\varepsilon$ is (ε,δ)-differentially private.*

Our claim follows by applying Gaussian Mechanism and the closedness to postprocessing property of data privacy. □

Utility Claim It turns out that if column spikiness μ_0 is a constant, Lasso-SSC is able to provably detect the correct subspace structures, despite privacy constraints.

Corollary 23. *Let the raw data $\mathbf{X}$ be compressed and privatized data $\widetilde{\mathbf{X}}'$ using the above described mechanism. Assume the same set of notations and assumptions in Theorem 15. Suppose $\mathbf{\Psi}$ is a JL transform with parameter ϵ. Let $B := \min_{\ell=1,\dots,k}\{\rho, r^{-1/2}, \rho_\ell - \mu_\ell\}$, and C be the constant in Theorem 15 and 19. If the privacy parameter ε is set to*

$$\varepsilon > \sqrt{\frac{512\mu_0\log(1.25/\delta)}{d}}\max\left\{\frac{(p\log N)^{1/4}}{(CB)^{1/2}}, \frac{\sqrt{\log N}}{CB}\right\}.$$

Then the solution to Lasso-SSC using obeys the subspace detection property with probability $1 - 8/N - \delta$.

The idea is simple. We are now injecting artificial Gaussian noise to a compressed subspace clustering problem with fixed input, and Theorem 19 (Theorem 8 in (Wang & Xu, 2013)) directly addresses that. All we have to do is to replace the geometric quantities in μ_ℓ and ρ_ℓ by their respective bound after compression in Corollary 13 and Lemma 14.

Proof of Corollary 23. The proof involves applying Theorem 19 with $\xi = 1$ and

$$\sigma = \frac{1+\epsilon}{1-\epsilon}\sqrt{\frac{32p\mu_0\log(1.25/\delta)}{d\varepsilon^2}} \le \sqrt{\frac{128p\mu_0\log(1.25/\delta)}{d\varepsilon^2}}$$

according to Theorem 21 and rearranging the expressions in terms of the limit for privacy requirement ε.

Note that the noise here is added after the compression and normalization, but the effect is the same as adding Gaussian noise in the original dimension and scaled orthogonal random projection on a noise. In fact, we can replace $C/4$ with C because there is no renormalization here.

Denote $B := \min_{\ell=1,\dots,k}\{\rho, r^{-1/2}, \rho_\ell - \mu_\ell\}$, and C to be the same as in Theorem 19, the conditions for success is

$$\sigma(1+\sigma) < CB\sqrt{\frac{p}{\log N}}, \tag{B.1}$$

which holds if

$$\sigma < \min\left\{\frac{CB}{2}\sqrt{\frac{p}{\log N}}, \sqrt{\frac{CB}{2}}\frac{p^{1/4}}{(\log N)^{1/4}}\right\}.$$

Substitute the expression of σ into (B.1) and rewrite it in terms of ε, we get our claim:

$$\varepsilon > \sqrt{\frac{512\mu_0 \log(1.25/\delta)}{d}} \max\left\{\frac{(p\log N)^{1/4}}{(CB)^{1/2}}, \frac{\sqrt{\log N}}{CB}\right\}.$$

□

B.1. Discussion of user-level privacy and its impossiblity under perfect subspace detection property

As we described in the main results, attribute-level differential privacy is a much weaker notion of privacy. While it is easy to handle a small group of attributes (in the order of $O(\sqrt{d/p})$ if we consider $B = O(1/r)$) by the composition rule, it does not protect any individual user's complete information. However, this is arguably the best we can do if our measure of utility is in terms of (perfect) subspace detection property.

Let us define formally the user-level differential privacy.

Definition 24 (User-Level Differential Privacy)**.** *We say a randomized algorithm $\mathcal{A} : \mathbb{R}^{d\times N} \to \mathcal{O}$ is (ε, δ)-differential private at attribute level if*

$$\mathbb{P}(\mathcal{A}(\mathbf{X}) \in \mathcal{S}) \leq e^{\varepsilon}\mathbb{P}(\mathcal{A}(\mathbf{X}') \in \mathcal{S}) + \delta$$

for any measurable outcome $\mathcal{S} \subset \mathcal{O}$, any $\mathbf{X}, \mathbf{X}' \in \mathcal{X}^n$ that differs in only one column.

The only difference to the attribute differential privacy is how $\mathbf{X}$ and $\mathbf{X}'$ may differ. Note that we can arbitrarily replace any single point in $\mathbf{X}$ with any $x \in \mathcal{X}$, to form $\mathbf{X}'$.

Proposition 25. *User-level differential privacy is NOT possible for any $0 \leq \varepsilon < \infty$ if we assume perfect subspace detection property, or perfect clustering results. In addition, If an algorithm achieves perfect clustering or subspace detection with probability $1-\delta$, user-level differential privacy is NOT possible for any $\varepsilon < \log\left(\frac{1-\delta}{\delta}\right)$.*

Proof. First of all, if a data point can be arbitrarily chosen, then we can change it entirely into a different subspace. Let's first ignore the gap from subspace detection property and perfect clustering. Assume that the output is the clustering result and it is always correct. then if we arbitrarily change the kth data point from one Subspace A to Subspace B, the result must reflect the change and cluster this data point correctly to its new subspace and the probability of observing an output that has kth data point clustered into Subspace A will change from 1 to 0, which blatantly violates the definition of differential privacy.

The same line of arguments holds if we treat the output as the graph embedding. Note that having subspace detection property for data point k in Subspace A (connected only to a set of points) and having subspace detection for data point k in Subspace B (connected only to another set of points) are two disjoint measurable events. With a perturbation that changes a data point from one subspace to another will blow the likelihood ratio of observing one of these two event to infinity.

The high probability statement holds because

$$\frac{\mathbb{P}(\text{SDP according to } \mathbf{X}|\mathbf{X})}{\mathbb{P}(\text{SDP according to } \mathbf{X}|\mathbf{X}')} \geq \frac{1-\delta}{\delta} \geq e^{\log(\frac{1-\delta}{\delta})}.$$

□

The reason why attribute-level privacy will work is because the promise is much weaker. Also our assumption that the columns are non-spiky ensures that perturbing any attribute of any user will not inject too much error. Intuitively, random projection and the injected dense Gaussian noise makes sure that it is not possible to identify any small changes in one attribute of a single user.

To be fair, the same problem still exists, namely, differential privacy breaks whenever the clustering can be shown to be always correct. What attribute-differential privacy ensures is that it is not possible to tell if a specific attribute of this user used in coming up with the result is actually the same or close to what it truly is.

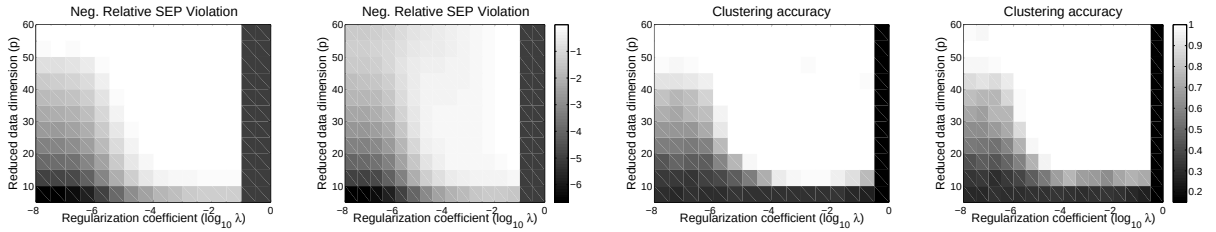

Figure 5. Relative Violation (top) and clustering accuracy (bottom) of Lasso-SSC on noiseless and noisy synthetic datasets. Left: noiseless; right: $\sigma/\sqrt{d} = 0.1$. λ ranges from 10^{-1} to 10^{-8} and the projected data dimension (p) ranges from 5 to 60. For each figure the rightmost columns indicate trivial solutions.

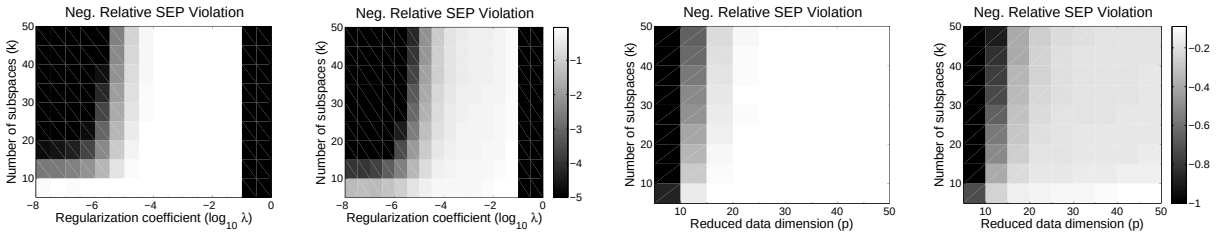

Figure 6. Relative violation of Lasso-SSC on noiseless and noisy synthetic datasets with varying number of clusters (k). Top row: λ ranges from 10^{-8} to 1; data dimension after random projection (p) is set to 25; rightmost columns indicate trivial solutions. Bottom row: p ranges from 5 to 50; λ is set to 10^{-2}. Left: noiseless; right: $\sigma/\sqrt{d} = 0.1$.

User-level privacy for subspace clustering and for privacy in general remains an important open problem. What we know for sure is that, we need to come up with a different/soft measure of utility other than exact clustering or subspace detection property.

C. Numerical results on synthetic datasets

We generate synthetic datasets to verify and extend theoretical findings in this paper. All subspaces and data points within each subspace are generated uniformly at random. We fix the ambient dimension (d) to be 100 and generate 50 data points per cluster. The intrinsic rank of each subspace is fixed to $r = 5$.

In the first set of experiments we generate $K = 10$ clusters and plot the relative violation of SEP as well as clustering accuracy with respect to different λ and p values in Figure 5. It can be shown that when the projected dimension p is smaller than the rank of the union of subspaces (i.e., $p < kr$) the performance of Lasso SSC degrades as λ decreases. This holds even for the noiseless case, which nicely justifies our theoretical findings. Note that when p is large (e.g., $p > kr$) both Lasso SSC and exact SSC ($\lambda \to 0$) succeeds when the input data matrix is not corrupted with noise. On the other hand, when λ is too large we obtain trivial solutions and clustering fails immediately.

In Figure 6 we report the relative violation of SEP and clustering accuracy with varying number of clusters k. It can be seen that even when there are a large number of clusters (e.g., $k = 50$) SEP still holds for a wide range of tuning parameters λ. In addition, the bottom two plots in Figure 6 show that the choice of projection dimension p is insensitive to the number of clusters (k).

D. Some tail inequalities

Lemma 26 (Matrix Gaussian and Rademacher Series, the general case (Tropp, 2012))**.** *Let $\{\mathbf{B}_k\}_k$ be a finite sequence of fixed matrices with dimensions $d_1 \times d_2$. Let $\{\gamma_k\}_k$ be a finite sequence of i.i.d. standard normal variables. Define the summation random matrix* $\mathbf{Z}$ *as*

$$\mathbf{Z} = \sum_k \gamma_k \mathbf{B}_k. \tag{D.1}$$

Define the variance parameter σ^2 *as*

$$\sigma^2 := \max\{\|\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]\|, \|\mathbb{E}[\mathbf{Z}^\top\mathbf{Z}]\|\}. \tag{D.2}$$

Then for every $t > 0$ *the following concentration inequality holds:*

$$\Pr\left[\|\mathbf{Z}\| \geq t\right] \leq (d_1 + d_2)e^{-t^2/2\sigma^2}. \tag{D.3}$$

Lemma 27 (Noncommutative Matrix Berstein Inequality, (Gross et al., 2010; Recht, 2011))**.** *Let* $\mathbf{B}_1, \cdots, \mathbf{B}_p$ *be independent zero-mean square* $r \times r$ *random matrices. Suppose* $\sigma_j^2 = \max\{\|\mathbb{E}[\mathbf{B}_j\mathbf{B}_j^\top]\|, \|\mathbb{E}[\mathbf{B}_j^\top\mathbf{B}_j]\|\}$ *and* $\|\mathbf{B}_j\| \leq R$ *almost surely for every* j*. Then for any* $t > 0$ *the following inequality holds:*

$$\Pr\left[\left\|\sum_{j=1}^{p}\mathbf{B}_j\right\|_2 > t\right] \leq 2r\exp\left(-\frac{t^2/2}{\sum_{j=1}^{p}\rho_j^2 + Rt/3}\right). \tag{D.4}$$

Lemma 28 (Spectrum bound of a Gaussian random matrix,(Davidson & Szarek, 2001))**.** *Let* A *be an* $m \times n$ $(m > n)$ *matrix with i.i.d standard Gaussian entries. Then, its largest and smallest singular values* $s_1(A)$ *and* $s_n(A)$ *obeys*

$$\sqrt{m} - \sqrt{n} \leq \mathbb{E}s_n(A) \leq \mathbb{E}s_1(A) \leq \sqrt{m} + \sqrt{n},$$

moreover,

$$\sqrt{m} - \sqrt{n} - t \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + \sqrt{n} + t,$$

with probability at least $1 - 2\exp(-t^2/2)$ *for all* $t > 0$*.*

The expectation result is due to Gordon's inequality and the concentration follows from the concentration of measure inequality in Gauss space by the fact that s_1 and s_n are both 1-Lipchitz functions. Take $t = \sqrt{n}$ in the above inequality we get

$$1 - 2\sqrt{\frac{n}{m}} - \epsilon \leq s_n(A/\sqrt{m}) \leq s_1(A/\sqrt{m}) \leq 1 + 2\sqrt{\frac{n}{m}}$$

with probability $1 - 2\exp(-n^2/2)$.

References

Balzano, Laura, Recht, Benjamin, and Nowak, Robert. High-dimensional matched subspace detection when data are missing. In *ISIT*, 2010.

Davidson, Kenneth R and Szarek, Stanislaw J. Local operator theory, random matrices and banach spaces. *Handbook of the geometry of Banach spaces*, 1:317–366, 2001.

Gross, David, Liu, Yi-Kai, Flammia, Steven T, Becker, Stephen, and Eisert, Jens. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

Kenthapadi, Krishnaram, Korolova, Aleksandra, Mironov, Ilya, and Mishra, Nina. Privacy via the johnson-lindenstrauss transform. *Journal of Privacy and Confidentiality*, 5(1):39–71, 2013.

Krishnamurthy, Akshay and Singh, Aarti. On the power of adaptivity in matrix completion and approximation. *Arxiv:1407.3619*, 2014.

Recht, Benjamin. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 12:3413–3430, 2011.

Tropp, Joel. *User-Friendly Tools for Random Matrices: An Introduction*. 2012.

Wang, Yu-Xiang and Xu, Huan. Noisy sparse subspace clustering. *arXiv:1309.1233*, 2013.

Zhou, Shuheng, Lafferty, John, and Wasserman, Larry. Compressed and privacy-sensitive sparse regression. *IEEE Transactions on Information Theory*, 55(2):846–866, 2009.

Table of symbols and notations

Table 1. Summary of symbols and notations

$\lvert\cdot\rvert$	Either absolute value or cardinality
$\lVert\cdot\rVert$; $\lVert\cdot\rVert_2$	2 norm of a vector/spectral norm of a matrix
$\lVert\cdot\rVert_1$	1 norm of a vector
$\lVert\cdot\rVert_\infty$	Infinity norm (maximum absolute value) of a vector
$\langle\cdot,\cdot\rangle$	Inner product of two vectors
$\lVert\mathbf{A}\rVert_{(i)}$	The ith row of matrix $\mathbf{A}$
$\sigma_1(\cdot), \sigma_r(\cdot)$	The largest and rth largest singular value of a matrix
N	Number of data points (number of columns in $\mathbf{X}$)
k	Number of subspaces (clusters)
d	The ambient dimension (number of rows in $\mathbf{X}$)
N_ℓ, r_ℓ for $\ell = 1, \cdots, k$	Number of data points and instrinsic dimension for each subspace
r	Largest intrinsic dimension across all subspaces
$\mathbf{X}$	Observed data matrix
$\mathbf{Y}$	Uncorrupted (noiseless) data matrix
$\mathbf{Z}$	Noise matrix, can be either deterministic or stochastic
$\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}, \widetilde{\mathbf{Z}}$	Projected matrices of $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$
$\widetilde{\mathbf{X}}', \widetilde{\mathbf{Y}}', \widetilde{\mathbf{Z}}'$	Normalized projected matrices of $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$
$\mathcal{U}^{(\ell)}$, $\mathbf{U}^{(\ell)}$	Subspace and its orthonormal basis of the ℓth cluster
$\mathbf{X}_{-i}$, $\mathbf{Y}_{-i}$, $\mathbf{Z}_{-i}$	All columns in $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ except the ith column.
$\mathbf{X}^{(\ell)}$,$\mathbf{Y}^{(\ell)}$,$\mathbf{Z}^{(\ell)}$	All columns in $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ associated with the ℓth subspace
$\mathbf{X}^{(\ell)}_{-i}$,$\mathbf{Y}^{(\ell)}_{-i}$,$\mathbf{Z}^{(\ell)}_{-i}$	All columns in $\mathbf{X}^{(\ell)}, \mathbf{Y}^{(\ell)}, \mathbf{Z}^{(\ell)}$ except the ith column
$\mathcal{Q}(\cdot)$,conv$(\cdot)$	(Symmetric) convex hull of a set of vectors
$r(\cdot)$	Radius of the largest ball inscribed in a convex body
$\mathcal{P}_{\mathcal{U}}(\cdot)$	Projection onto subspace $\mathcal{U}$
p	Target dimension after random projection
ϵ	Approximation error of random projection methods
δ	Failure probability
$\mathbf{\Psi}, \mathbf{\Omega}, \mathbf{\Phi}, \mathbf{S}$	Projection operators for random Gaussian projection, uniform sampling, FJLT and sketching
μ_0	Column space incoherence or column spikiness
μ_ℓ, ρ_ℓ for $\ell = 1, \cdots, k$	Subspace incoherence and inradius for each subspace
$\tilde{\mu}_\ell, \tilde{\rho}_\ell$ for $\ell = 1, \cdots, k$	Subspace incoherence and inradius on the projected data
$f(\cdot), \tilde{f}(\cdot)$	Objective functions of Eq. (4.1) on the original data and projected data
$\boldsymbol{\nu}, \boldsymbol{v}$	Unnormalized and normalized dual direction
$\tilde{\boldsymbol{\nu}}$	Random projection of $\boldsymbol{\nu}$
$\tilde{\boldsymbol{\nu}}'$	A shrinked version of $\tilde{\boldsymbol{\nu}}$ such that it is feasible for Eq. (4.1) on projected data
$\boldsymbol{\nu}^*$	Optimal solution to Eq. (4.1) on projected data
$\bar{\boldsymbol{\nu}}$	A vector in the original space that corresponds to $\boldsymbol{\nu}^*$ after projection
$\bar{\boldsymbol{\nu}}'$	A shrinked version of $\bar{\boldsymbol{\nu}}$ such that it is feasible for Eq. (4.1) on the original data
λ	Regularization coefficient for Lasso SSC
Δ	Margin of error (i.e., $\min_\ell \rho_\ell - \mu_\ell$)
$\eta, \tilde{\eta}$	Noise level for deterministic noise, before and after projection
$\sigma, \tilde{\sigma}$	Noise level for random Gaussian noise, before and after projection
$\mathbf{C}$	Similarity matrix
q	Number of nonzero entries in regression solutions. Used in solution path algorithms.