Supplementary Material

The sufficiently scattered condition

In some prior works (Huang et al., 2014; 2016; 2018), the sufficiently scattered condition is presented as follows:

Definition 2 (sufficiently scattered [dual cone form].)**.** Let $\mathrm{cone}(\boldsymbol{M})^*$ denote the polyhedral cone $\{\boldsymbol{x} : \boldsymbol{M}^\top \boldsymbol{x} \geq 0\}$ and $\mathcal{K}^*$ denote the elliptical cone $\{\boldsymbol{x} : \|\boldsymbol{x}\| \leq \mathbf{1}^\top \boldsymbol{x}\}$. Matrix $\boldsymbol{M}$ is sufficiently scattered if

i) $\mathrm{cone}(\boldsymbol{M})^* \subseteq \mathcal{K}^*$,
ii) $\mathrm{cone}(\boldsymbol{M})^* \cap \mathrm{bd}\mathcal{K}^* = \{\alpha \boldsymbol{e}_\ell : \alpha \geq 0, \ell = 1, ..., k\}$.

As for our definition, if we drop the $\mathbf{1}^\top \boldsymbol{x} = 1$ constraint, it can be equivalently written in the following cone form:

Definition 3 (sufficiently scattered [primal cone form].)**.** Let $\mathrm{cone}(\boldsymbol{M})$ denote the polyhedral cone $\{\boldsymbol{M}\boldsymbol{\beta} : \boldsymbol{\beta} \geq 0\}$ and $\mathcal{K}$ denote the elliptical cone $\{\boldsymbol{x} \in \mathbb{R}^k : \|\boldsymbol{x}\| \leq \frac{1}{\sqrt{k-1}} \mathbf{1}^\top \boldsymbol{x}\}$. Matrix $\boldsymbol{M}$ is sufficiently scattered if

i) $\mathrm{cone}(\boldsymbol{M}) \supseteq \mathcal{K}$,
ii) $\mathrm{bd}\,\mathrm{cone}(\boldsymbol{M}) \cap \mathcal{K} = \{\alpha(\mathbf{1} - \boldsymbol{e}_\ell) : \alpha \geq 0, \ell = 1, ..., k\}$.

To understand the equivalence of Definitions 2 and 3, it is essential to invoke the concept of the dual cone, denoted with a superscript *

$$\mathcal{C}^* = \{\boldsymbol{x} : \boldsymbol{y}^\top \boldsymbol{x} \geq 0, \forall\, \boldsymbol{y} \in \mathcal{C}\}.$$

One can verify that $\mathrm{cone}(\boldsymbol{M})$ and $\mathrm{cone}(\boldsymbol{M})^*$ are indeed dual to each other, and so are $\mathcal{K}$ and $\mathcal{K}^*$.

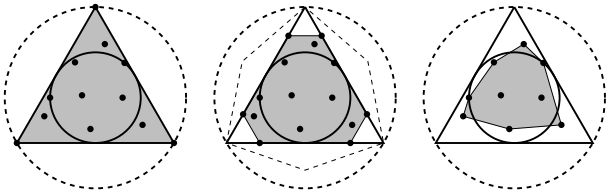
We now prove that Definitions 2 and 3 are indeed primal-dual representations of the same condition. For two convex cones $\mathcal{C}_1$ and $\mathcal{C}_2$, if $\mathcal{C}_1 \subseteq \mathcal{C}_2$, then $\mathcal{C}_1^* \supseteq \mathcal{C}_2^*$. It clearly shows the equivalence of the first requirements in Definitions 2 and 3. As for the second requirement, we claim the following:

- Requirement ii) in Definition 2 asks that all extreme rays of $\mathrm{cone}(\boldsymbol{M})^*$ lie strictly inside $\mathcal{K}^*$, except for the coordinate directions $\alpha \boldsymbol{e}_\ell$'s, which lie on the boundary of $\mathcal{K}^*$;
- Requirement ii) in Definition 3 asks that all facets of $\mathrm{cone}(\boldsymbol{M})$ lie strictly outside $\mathcal{K}$, except for the facets spanned by $k-1$ coordinate vectors, which touch the boundary of $\mathcal{K}$ at $\alpha(\mathbf{1} - \boldsymbol{e}_\ell)$.

There is a one-to-one correspondence between extreme rays of $\mathrm{cone}(\boldsymbol{M})^*$ and facets of $\mathrm{cone}(\boldsymbol{M})$, which are both defined by $k-1$ columns of $\boldsymbol{M}$. Let $\boldsymbol{v}$ denote an extreme ray of $\mathrm{cone}(\boldsymbol{M})^*$, then $\boldsymbol{M}^\top \boldsymbol{v} \geq 0$ and $k-1$ of them holds as equalities; those $k-1$ columns defines a facet of $\mathrm{cone}(\boldsymbol{M})$, and any point $\boldsymbol{x}$ in that facet satisfies that $\boldsymbol{x}^\top \boldsymbol{v} = 0$.

Now if $\boldsymbol{v}$ is not a coordinate direction, Definition 2 asks that $\|\boldsymbol{v}\| < \mathbf{1}^\top \boldsymbol{v}$. It can be rearranged as

$$\frac{\mathbf{1}^\top \boldsymbol{v}}{\|\mathbf{1}\|\|\boldsymbol{v}\|} > \frac{1}{\sqrt{k}},$$



(a) Pure node (b) Sufficiently scattered (c) Not identifiable

Figure 5. Same illustration as in Figure 3 with dual cones added (in dash).

which means the angle between $\boldsymbol{v}$ and $\mathbf{1}$ is less than $\arccos(1/\sqrt{k})$. Therefore the angle between $\mathbf{1}$ and any point on the corresponding facet is greater than $\pi - \arccos(1/\sqrt{k})$, or equivalently

$$\frac{\mathbf{1}^\top \boldsymbol{x}}{\|\mathbf{1}\|\|\boldsymbol{x}\|} < \sqrt{\frac{k-1}{k}} \quad \Longleftrightarrow \quad \|\boldsymbol{x}\| > \frac{1}{\sqrt{k-1}} \mathbf{1}^\top \boldsymbol{x},$$

meaning all points on that facet lie strictly outside $\mathcal{K}$.

The other direction is true as well: If the smallest angle between $\mathbf{1}$ and any point on the facet is greater than $\pi - \arccos(1/\sqrt{k})$, then the angle between $\mathbf{1}$ and $\boldsymbol{v}$ is less than $\arccos(1/\sqrt{k})$, meaning $\boldsymbol{v}$ lie strictly inside $\mathcal{K}^*$. This shows the equivalence of the second requirement in Definitions 2 and 3. A geometric illustration with the dual cones shown is given in Figure 5.

Proof of Theorem 1

Denote an optimal solution of (8) as $(\boldsymbol{\Xi}_\star, \boldsymbol{M}_\star)$, then clearly

$$|\det \boldsymbol{\Xi}_\star| \leq |\det \widetilde{\boldsymbol{\Xi}}^\natural| \quad \Leftrightarrow \quad |\det \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural| \geq 1. \tag{13}$$

Furthermore, since both $(\boldsymbol{\Xi}_\star, \boldsymbol{M}_\star)$ and $(\widetilde{\boldsymbol{\Xi}}^\natural, \boldsymbol{M}_2^\natural)$ are feasible, we have

$$\begin{aligned} &\boldsymbol{\Xi}_\star \boldsymbol{M}_\star = \widetilde{\boldsymbol{\Xi}}^\natural \boldsymbol{M}_2^\natural, \boldsymbol{M}_\star \geq 0 \\ &\Rightarrow \quad \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural \boldsymbol{M}_2^\natural \geq 0 \qquad (14) \\ &\boldsymbol{e}_k^\top \boldsymbol{\Xi}_\star = \mathbf{1}^\top, \boldsymbol{e}_k^\top \widetilde{\boldsymbol{\Xi}}^\natural = \mathbf{1}^\top, \\ &\Rightarrow \quad \mathbf{1}^\top \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural = \mathbf{1}^\top. \qquad (15) \end{aligned}$$

Denote $\boldsymbol{V} = \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural$ with ℓ-th row denoted as $\boldsymbol{v}_\ell^\top$. Then (14) means $\boldsymbol{v}_\ell$'s lie in $\mathrm{cone}(\boldsymbol{M}_2^\natural)$; since $\boldsymbol{M}_2^\natural$ is sufficiently scattered, $\boldsymbol{v}_\ell \in \mathcal{K}^*$, meaning

$$\|\boldsymbol{v}_\ell\| \leq \mathbf{1}^\top \boldsymbol{v}_\ell.$$

Using Hadamard's inequality,

$$|\det \boldsymbol{B}| \leq \prod_{\ell=1}^{k} \|\boldsymbol{v}_\ell\|,$$

and we further have that

$$\prod_{\ell=1}^{k} \|\boldsymbol{v}_\ell\| \le \prod_{\ell=1}^{k} \mathbf{1}^\top \boldsymbol{v}_\ell \tag{16a}$$

$$\le \left(\frac{\sum_{\ell=1}^{k} \mathbf{1}^\top \boldsymbol{v}_\ell}{k} \right)^k \tag{16b}$$

$$= \left(\frac{\mathbf{1}^\top \boldsymbol{V} \mathbf{1}}{k} \right)^k = 1, \tag{16c}$$

where (16a) stems from $\boldsymbol{v}_\ell \in \mathcal{K}^*$, (16b) is due to arithmetic-geometric mean inequality, and (16c) is due to (15). This means

$$|\det \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural| \le 1. \tag{17}$$

Combining (17) and (13), we conclude that

$$|\det \boldsymbol{\Xi}_\star^{-1} \widetilde{\boldsymbol{\Xi}}^\natural| = 1.$$

Furthermore, all inequalities in (16) hold as equalities, which means $\boldsymbol{v}_\ell$'s lie on the boundary of $\mathcal{K}^*$. The second requirement of sufficiently scattered (dual form Definition 2) imply that $\boldsymbol{v}_\ell$'s can only take coordinate vectors, meaning $\boldsymbol{V}$ is a permutation matrix $\boldsymbol{\Pi}$, and we have

$$\boldsymbol{M}_2^\natural = \boldsymbol{\Pi} \boldsymbol{M}_\star, \qquad \widetilde{\boldsymbol{\Xi}}^\natural = \boldsymbol{\Xi}_\star \boldsymbol{\Pi}^\top.$$

Q.E.D.

Proof of Theorem 2

Define a vector $\widetilde{\boldsymbol{f}} \in \mathbb{R}^k$ for a specific $\boldsymbol{X}$ with the m-th element equal to

$$\widetilde{f}_m = (-1)^{\ell+m} \det \boldsymbol{X}_{\ell m},$$

then the co-factor expansion tells us

$$\det \boldsymbol{X} = \boldsymbol{x}_\ell^\top \widetilde{\boldsymbol{f}},$$

where $\boldsymbol{x}_\ell^\top$ is the ℓ-th row of $\boldsymbol{X}$. Therefore, at a particular point $\boldsymbol{X}$, and we look at problem (10) at the ℓ-th row of $\boldsymbol{X}$, the subproblem is

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{maximize}} \quad & (\widetilde{\boldsymbol{f}}^\top \boldsymbol{z})^2 \\ \text{subject to} \quad & \boldsymbol{z}^\top \widetilde{\boldsymbol{Y}} \ge 0, \boldsymbol{z}^\top \widetilde{\boldsymbol{Y}} \mathbf{1} = 1. \end{aligned}$$

The objective is a convex quadratic, so if we take a linear approximation at $\boldsymbol{x}_\ell$ it defines a global lowerbound to the objective

$$(\boldsymbol{x}_\ell^\top \widetilde{\boldsymbol{f}}) \widetilde{\boldsymbol{f}}^\top \boldsymbol{z} = (\det \boldsymbol{X}) \widetilde{\boldsymbol{f}}^\top \boldsymbol{z},$$

with equality holds when $\boldsymbol{z} = \boldsymbol{x}_\ell$. This is where the linear programming sub-problem (12) comes from.

Cyclically solving (12) with respect to each row of $\boldsymbol{X}$ falls into the framework of BSUM, proposed by Razaviyayn et al. (2013). Specifically, we have that the constraint set decouples over the rows of $\boldsymbol{X}$, and the objective is a tight lowerbound $(\det \boldsymbol{X})^2$ when restricted to the ℓ-th row of $\boldsymbol{X}$. Because $(\det \boldsymbol{X})$ is a smooth function, it automatically satisfies that the directional derivative of $\boldsymbol{f}^\top \boldsymbol{z}$ is equal to that of $(\det \boldsymbol{X})^2$ at everywhere. Furthermore, it is easy to see that the constraint set is a compact set. According to Razaviyayn et al. (2013), the proposed iterative algorithm is guaranteed to converge to a stationary point as long as each of the LP sub-problems (12) has a unique solution.

Regarding the uniqueness of $\arg\max(12)$, we note that for any LP, the constraint set defines a polyhedron, and a solution always exists on a vertex. A solution is not unique if the objective direction $\boldsymbol{f}$ happens to be normal to one of the edges of the constraint polyhedron. This means there exists a set of $k-1$ columns in $\widetilde{\boldsymbol{Y}}$ such that $\boldsymbol{f}$ lies in their range. As long as columns of $\widetilde{\boldsymbol{Y}}$ appears somewhat incoherent, this will happen with a very small probability. **Q.E.D.**

Proof of Theorem 3

Suppose $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{\Xi}} \boldsymbol{M}_2$ where $\boldsymbol{M}_2$ satisfies the separability assumption. This means all the coordinate vectors $\boldsymbol{e}_1, ..., \boldsymbol{e}_k$ exists in its columns. This means columns of $\widetilde{\boldsymbol{\Xi}}$ exists in columns of $\widetilde{\boldsymbol{Y}}$, which corresponds to the coordinate vectors in the columns of $\boldsymbol{M}_2$. Then $\boldsymbol{z}^\top \widetilde{\boldsymbol{\Xi}} \ge 0$ implies $\boldsymbol{z}^\top \widetilde{\boldsymbol{Y}} \ge 0$, since $\boldsymbol{M}_2 \ge 0$, which means $\boldsymbol{z}^\top \widetilde{\boldsymbol{Y}} \ge 0$ consists of a lot of redundant constraints, and (12) is equivalent to

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{maximize}} \quad & \boldsymbol{f}^\top \boldsymbol{z} \\ \text{subject to} \quad & \boldsymbol{z}^\top \widetilde{\boldsymbol{\Xi}} \ge 0, \boldsymbol{z}^\top \boldsymbol{b} = 1. \end{aligned}$$

Let us denote the dual variable with respect to the inequality constraint as $\boldsymbol{\mu}$, and the equality constraint as λ, then the KKT condition implies

$$\lambda \boldsymbol{b} - \boldsymbol{f} = \widetilde{\boldsymbol{\Xi}} \boldsymbol{\mu}.$$

Because we assume $\widetilde{\boldsymbol{\Xi}}$ is non-singular, it is equivalent to

$$\lambda \widetilde{\boldsymbol{\Xi}}^{-1} \boldsymbol{b} - \widetilde{\boldsymbol{\Xi}}^{-1} \boldsymbol{f} = \boldsymbol{\mu} \ge 0.$$

Therefore, λ should be chosen to make $\lambda \widetilde{\boldsymbol{\Xi}}^{-1} \boldsymbol{b} - \widetilde{\boldsymbol{\Xi}}^{-1} \boldsymbol{f} \ge 0$, and for a scalar λ, the resulting $\boldsymbol{\mu}$ will have only one zero almost surely. According to complementary slackness, then $\boldsymbol{z}^\top \widetilde{\boldsymbol{\Xi}}$ will have only one nonzero entry. This implies that the solution to (12) should be a row of $\widetilde{\boldsymbol{\Xi}}^{-1}$, which is exactly what we want.

Now after running one iteration of CD-MVSI, we would obtain k rows of $\widetilde{\boldsymbol{\Xi}}$. If all k rows of them are present, then we have successfully recovered the ground truth. The only non-ideal case is if one row of $\widetilde{\boldsymbol{\Xi}}$ appear multiple times. However,

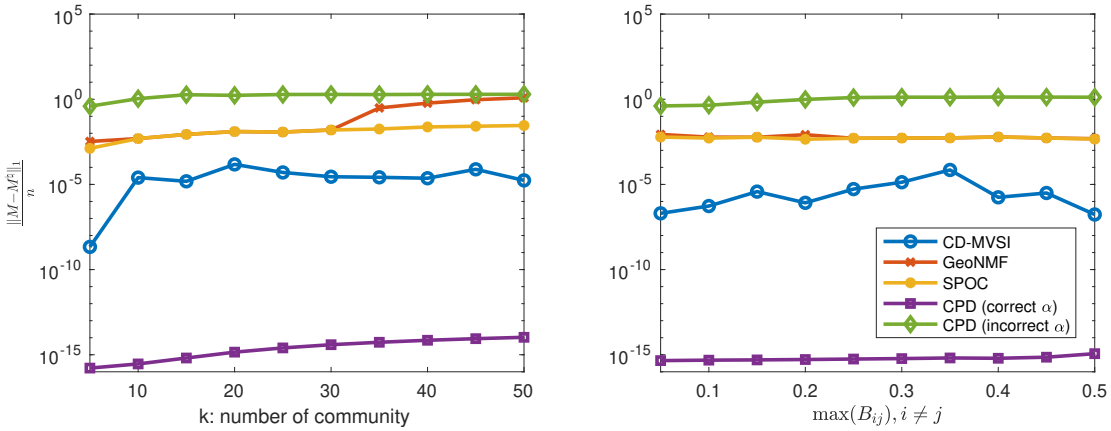


Figure 6. Algorithm performance on synthetic data. Note that the performance of CPD heavily relies on the knowledge of the underlying Dirichlet parameter $\boldsymbol{\alpha}$, which may not be practical.

if that is the case, the resulting $\det \boldsymbol{X} = 0$. Suppose we initialize with a non-singular matrix, say $\boldsymbol{X} = \mathrm{Diag}(\boldsymbol{b})^{-1}$, which is feasible, CD-MVSI is guaranteed to monotonically increase the objective value. Therefore $\det \boldsymbol{X}$ is impossible to be equal to zero, and we indeed manage to recover all k rows of $\widetilde{\boldsymbol{\Xi}}^{-1}$. **Q.E.D.**

Additional synthetic experiments

Here we validate the correctness of various algorithms, *given exact statistics ideal for each method*. We start by generating k-dimensional membership coefficients from a Dirichlet distribution with parameter $(1/k)\mathbf{1}$. Since the Dirichlet parameter for all the components are less than one, it tends to generate points that lie on the boundary of Δ. However, we note that none of the components will be exactly equal to one, which means the sufficiently scattered condition will not be satisfied exactly. Nevertheless, since the points are well spread out, the recovery result is very close to optimal. On the other hand, the separability / pure-node will be grossly violated as k goes larger, and we will see that pure-node-based methods are very vulnerable to such assumption violation.

We fix the number of nodes to be 1000. For GeoNMF and SPOC, the entire underlying matrix $\boldsymbol{M}^\top \boldsymbol{B} \boldsymbol{M}$ is given. For the tensor method, the nodes are divided into three groups of size 300, 300, and 400, and the corresponding $300 \times 300 \times 400$ moment tensor is given. For the tensor method we consider two scenarios: one with the correct Dirichlet parameter $(1/k)\mathbf{1}$ and one with an incorrect Dirichlet parameter $\mathbf{1}$. The goal is to test how sensitive is the method to the accurate knowledge of the prior distribution.

Figure 6 shows the performance of different algorithms under various scenarios. We show the normalized L_1 norm of the estimation error $\|\widehat{\boldsymbol{M}} - \boldsymbol{M}^\natural\|_1/n$. On the left panel, we fix $\boldsymbol{B} = \boldsymbol{I}$, and increase k from 5 to 50. As we can see, CD-MVSI gives acceptable recovery result, even though the sufficiently scattered condition is not exactly satisfied. Both GeoNMF and SPOC give much worse performance compared to CD-MVSI. Tensor method with the correct Dirichlet parameter gives nearly perfect recovery, thanks to its nice identifiability guarantees. However, a correct Dirichlet is absolutely necessary, as an incorrect one gives the worst performance. On the right panel, we fix $k = 10$ and let $\boldsymbol{B}$ take nonzero off-diagonal values, and similar patterns present. We remark that it is relatively easy to approximately satisfy the sufficiently scattered condition, whereas knowing exactly the Dirichlet parameter of the prior is somewhat impractical.