Similarities between PSD Matric completion and Euclidean Distance Matrix completion

Sudeep Raja Putta
sputta@cs.umass.edu

1 Introduction

Recently Ge et al. (2017) showed that non-convex optimization problems such as Low rank Matrix Completion, Matrix Sensing and Robust PCA possess a well behaved optimization landscape: 1) all local optima are also globally optimal and 2) no high order saddle points exists. In their work, they prove that the non-convex objectives of these problems are strict saddles.

We consider the problem of Euclidean Distance Matrix Completion (EDMC). Our goal will be to show that there are several similarities between PSD Matrix completion (PSDMC) and EDMC. Though we were unable to show that EDMC's objective function is a strict saddle, we always converged to the global optimum when using gradient descent in our extensive experiments.

2 Notation

For a vector $\mathbf{v}$, $\|\mathbf{v}\|$ denotes its ℓ_2 norm. For a matrix $\mathbf{M}$, $\|\mathbf{M}\|$ denotes its spectral norm and $\|\mathbf{M}\|_F$ denotes its Frobenius norm. $\langle \mathbf{u}, \mathbf{v} \rangle$ denotes the inner-product of vectors and for matrices $\langle \mathbf{M}, \mathbf{N} \rangle = \text{tr}(\mathbf{M}\mathbf{N}^\top) = \sum_{i,j} \mathbf{M}_{ij}\mathbf{N}_{ij}$. We use $\mathbf{M} : \mathcal{H} : \mathbf{N}$ to denote the quadratic form $\langle \mathbf{M}, \mathcal{H}(\mathbf{N}) \rangle$.

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n \in \mathbb{R}^{n \times k}$ be a set of n points in k dimensions. The squared Euclidean distance between the two points $\mathbf{x}_i$ and $\mathbf{x}_j$ is d_{ij}

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \mathbf{x}_i^\top \mathbf{x}_i + \mathbf{x}_j^\top \mathbf{x}_j - 2\mathbf{x}_i^\top \mathbf{x}_j$$

Let $\mathbf{X}$ be the $n \times k$ matrix where $\mathbf{x}_i$ is the ith row. Let $\mathbf{D}$ be the Euclidean Distance Matrix of these points. The entry of $\mathbf{D}$ at location (i, j) is $d_{i,j}$. $\mathbf{D}$ can be computed using the following equation.

$$\mathbf{D} = (\text{diag}(\mathbf{X}\mathbf{X}^\top))\mathbf{e}^\top + \mathbf{e}(\text{diag}(\mathbf{X}\mathbf{X}^\top))^\top - 2\mathbf{X}\mathbf{X}^\top$$

Here $\text{diag}(\mathbf{A})$ is the diagonal vector of $\mathbf{A}$ and $\mathbf{e}$ is the vector of size n consisting of all ones. Let $\mathcal{K}(\mathbf{A}) = (\text{diag}(\mathbf{A}))\mathbf{e}^\top + \mathbf{e}(\text{diag}(\mathbf{A}))^\top - \mathbf{A} - \mathbf{A}^\top$. So $\mathbf{D} = \mathcal{K}(\mathbf{X}\mathbf{X}^\top)$.

Let $\mathbf{D}^\star$ be the EDM of unknown point matrix $\mathbf{X}^\star \in \mathbb{R}^{n \times k}$, i.e $\mathbf{D}^\star = \mathcal{K}(\mathbf{X}^\star \mathbf{X}^{\star\top})$. Let $\Omega \subseteq [n] \times [n]$ be the entries of $\mathbf{D}^\star$ which are observed. Since $\mathbf{D}^\star$ is a zero diagonal symmetric matrix, $(i, i) \notin \Omega$ and $(i, j) \in \Omega \iff (j, i) \in \Omega$ for all $i \neq j$ and $i, j \in [n]$. For any matrix $\mathbf{M}$, let $\mathbf{M}_\Omega$ be the matrix whose entries outside of Ω are set to 0. Let $\mathcal{H} = (\mathbf{e}\mathbf{e}^\top)_\Omega$.

The objective is to find $\mathbf{X} \in \mathbb{R}^{n \times k}$ such that the following function is minimized.

$$f(\mathbf{X}) = \frac{1}{2p}\|(\mathcal{K}(\mathbf{X}\mathbf{X}^\top) - \mathbf{D}^\star)_\Omega\|_F^2 = \frac{1}{2p}(\mathcal{K}(\mathbf{X}\mathbf{X}^\top) - \mathbf{D}^\star) : \mathcal{H} : (\mathcal{K}(\mathbf{X}\mathbf{X}^\top) - \mathbf{D}^\star)$$

3 Similarities between PSDMC and EDMC

PSDMC to the left and EDMC to the right.

3.1 Objective function

$$\min_{X \in \mathbb{R}^{n \times k}} \frac{1}{2p} \|(\mathbf{X}\mathbf{X}^\top - \mathbf{M}^\star)_\Omega\|_F^2 + Q(\mathbf{X})$$

Here $\mathbf{M}^\star = \mathbf{X}^\star \mathbf{X}^{\star\top}$. Let $M = \mathbf{X}\mathbf{X}^\top$.

$$\min_{X \in \mathbb{R}^{n \times k}} \frac{1}{2p} \|(\mathcal{K}(\mathbf{X}\mathbf{X}^\top) - \mathbf{D}^\star)_\Omega\|_F^2 + Q(\mathbf{X})$$

Here $\mathbf{D}^\star = \mathcal{K}(\mathbf{X}^\star \mathbf{X}^{\star\top})$. Let $D = \mathcal{K}(\mathbf{X}\mathbf{X}^\top)$.

3.2 Definition 6 in (Ge et al., 2017)

Given the matrices $\mathbf{X}, \mathbf{X}^\star$, define their difference as $\Delta = \mathbf{X} - \mathbf{X}^\star \mathbf{R}_\mathbf{X}$, where $\mathbf{R}_\mathbf{X} \in \mathbb{R}^{k \times k}$ such that

$$\mathbf{R}_\mathbf{X} = \underset{RR^\top = R^\top R = I}{argmin} \|\mathbf{X} - \mathbf{X}^\star \mathbf{R}\|_F^2$$

Given the matrices $\mathbf{X}, \mathbf{X}^\star$, define their difference as $\Delta = \mathbf{J}\mathbf{X} - \mathbf{J}\mathbf{X}^\star \mathbf{R}_\mathbf{X}$, where $\mathbf{R}_\mathbf{X} \in \mathbb{R}^{k \times k}$ such that

$$\mathbf{R}_\mathbf{X} = \underset{RR^\top = R^\top R = I}{argmin} \|\mathbf{J}\mathbf{X} - \mathbf{J}\mathbf{X}^\star \mathbf{R}\|_F^2$$

Here $\mathbf{J}$ is the centering matrix.

$$\mathbf{J} = \mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^\top$$

The product $\mathbf{J}\mathbf{X}$ centers the set of points around the origin. This does not change the distances between the points, so $\mathcal{K}(\mathbf{X}\mathbf{X}^\top) = \mathcal{K}(\mathbf{J}\mathbf{X}\mathbf{X}^\top\mathbf{J})$. Let $\mathbf{M}_c = \mathbf{J}\mathbf{X}\mathbf{X}^\top\mathbf{J}$ and $\mathbf{M}_c^\star = \mathbf{J}\mathbf{X}^\star\mathbf{X}^{\star\top}\mathbf{J}$ be the centered gram matrices of $\mathbf{X}$ and $\mathbf{X}^\star$. A few crucial relationships are:

$$\mathcal{K}(\mathbf{M}_c) = (\mathrm{diag}(\mathbf{M}_c))\mathbf{e}^\top + \mathbf{e}(\mathrm{diag}(\mathbf{M}_c))^\top - 2\mathbf{M}_c = \mathbf{D}$$

$$\mathcal{T}(\mathbf{D}) = -\frac{1}{2}\mathbf{J}\mathbf{D}\mathbf{J} = \mathbf{M}_c$$

3.3 Lemma 6 in Ge et al. (2017)

Given matrices $\mathbf{X}, \mathbf{X}^\star \in \mathbb{R}^{n \times k}$, let $\mathbf{M} = \mathbf{X}\mathbf{X}^\top$, $\mathbf{M}^\star = \mathbf{X}^\star(\mathbf{X}^\star)^\top$ and $\sigma_r^\star$ is the smallest singular value of $\mathbf{M}^*$, and let Δ be defined as above, then we have $\|\Delta\Delta^\top\|_F^2 \leq 2\|\mathbf{M} - \mathbf{M}^\star\|_F^2$, and $\sigma_r^\star\|\Delta\|_F^2 \leq \frac{1}{2(\sqrt{2}-1)}\|\mathbf{M} - \mathbf{M}^\star\|_F^2$.

Given matrices $\mathbf{X}, \mathbf{X}^\star \in \mathbb{R}^{n \times k}$, let $\mathbf{M}_c = \mathbf{J}\mathbf{X}\mathbf{X}^\top\mathbf{J}$, $\mathbf{M}_c^\star = \mathbf{J}\mathbf{X}^\star(\mathbf{X}^\star)^\top\mathbf{J}$ and $\sigma_r^\star$ is the smallest singular value of $\mathbf{M}_c^*$, and let Δ be defined as above, then we have $\|\Delta\Delta^\top\|_F^2 \leq 2\|\mathbf{M}_c - \mathbf{M}_c^\star\|_F^2$, and $\sigma_r^\star\|\Delta\|_F^2 \leq \frac{1}{2(\sqrt{2}-1)}\|\mathbf{M}_c - \mathbf{M}_c^\star\|_F^2$.

3.4 Lemma 7 in (Ge et al., 2017)

$$\langle \nabla f(\mathbf{X}), \mathbf{Z} \rangle = (\mathbf{M} - \mathbf{M}^\star) : \mathcal{H} : (\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) + \langle \nabla Q(\mathbf{X}), \mathbf{Z} \rangle$$

$$\mathbf{Z} : \nabla^2 f(\mathbf{X}) : \mathbf{Z} = 2(\mathbf{M} - \mathbf{M}^\star) : \mathcal{H} : (\mathbf{Z}\mathbf{Z}^\top) + (\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) : \mathcal{H} : (\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) + \mathbf{Z} : \nabla^2 Q(\mathbf{X}) : \mathbf{Z}$$

$$\Delta : \nabla^2 f(\mathbf{X}) : \Delta = \Delta\Delta^\top : \mathcal{H} : \Delta\Delta^\top - 3(\mathbf{M} - \mathbf{M}^\star) : \mathcal{H} : (\mathbf{M} - \mathbf{M}^\star) + 4\langle \nabla f(\mathbf{X}), \Delta \rangle + [\Delta : \nabla^2 Q(\mathbf{X}) : \Delta - 4\langle \nabla Q(\mathbf{X}), \Delta \rangle]$$

In the case where $Q(\mathbf{X}) = 0$ and $\mathcal{H}$ is identity, if gradient is zero:

$$\Delta : \nabla^2 f(\mathbf{X}) : \Delta = \|\Delta\Delta^\top\|_F^2 - 3\|\mathbf{M} - \mathbf{M}^\star\|_F^2$$

Using Lemma 6 from above, $\Delta : \nabla^2 f(\mathbf{X}) : \Delta \leq -\|\mathbf{M} - \mathbf{M}^\star\|_F^2$. Therefore all stationary points with $\mathbf{M} \neq \mathbf{M}^\star$ are saddle points. Hence all local minima satisfy $\mathbf{X}\mathbf{X}^\top = \mathbf{M}^\star$

$$\langle \nabla f(\mathbf{X}), \mathbf{Z} \rangle = (\mathbf{D} - \mathbf{D}^\star) : \mathcal{H} : \mathcal{K}(\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) + \langle \nabla Q(\mathbf{X}), \mathbf{Z} \rangle$$

$$\mathbf{Z} : \nabla^2 f(\mathbf{X}) : \mathbf{Z} = 2(\mathbf{D} - \mathbf{D}^\star) : \mathcal{H} : \mathcal{K}(\mathbf{Z}\mathbf{Z}^\top) + \mathcal{K}(\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) : \mathcal{H} : \mathcal{K}(\mathbf{X}\mathbf{Z}^\top + \mathbf{Z}\mathbf{X}^\top) + \mathbf{Z} : \nabla^2 Q(\mathbf{X}) : \mathbf{Z}$$

$$\Delta : \nabla^2 f(\mathbf{X}) : \Delta = \mathcal{K}(\Delta\Delta^\top) : \mathcal{H} : \mathcal{K}(\Delta\Delta^\top) - 3(\mathbf{D} - \mathbf{D}^\star) : \mathcal{H} : (\mathbf{D} - \mathbf{D}^\star) + 4\langle \nabla f(\mathbf{X}), \Delta \rangle + [\Delta : \nabla^2 Q(\mathbf{X}) : \Delta - 4\langle \nabla Q(\mathbf{X}), \Delta \rangle]$$

In the case where $Q(\mathbf{X}) = 0$ and $\mathcal{H}$ is identity, if gradient is zero:

$$\Delta : \nabla^2 f(\mathbf{X}) : \Delta = \|\mathcal{K}(\Delta\Delta^\top)\|_F^2 - 3\|\mathbf{D} - \mathbf{D}^\star\|_F^2$$

To show a result which is similar to the one on the left, we need an inequality like this: $\|\mathcal{K}(\Delta\Delta^\top)\|_F^2 \leq c\|\mathbf{D} - \mathbf{D}^\star\|_F^2$ where c must be strictly less than 3.
However, we have these relations:

$$4\|\mathbf{M}_c - \mathbf{M}_c^\star\|_F^2 \leq \|\mathbf{D} - \mathbf{D}^\star\|_F^2$$

$$\|\mathcal{K}(\Delta\Delta^\top)\|_F^2 \leq 4n\|\Delta\Delta^\top\|_F^2$$

Using these, we can get

$$\Delta : \nabla^2 f(\mathbf{X}) : \Delta \leq (8n - 12)\|\mathbf{M}_c - \mathbf{M}_c^\star\|_F^2$$

This does not prove the strict-saddle property because of the dependence on n. However, the inequality $\|\mathcal{K}(\Delta\Delta^\top)\|_F^2 \leq 4n\|\Delta\Delta^\top\|_F^2$ was derived by just using the spectral property of $\mathcal{K}$. By imposing extra conditions like $\mathbf{M}$ being PSD matrix and $\mathbf{X}^\top\mathbf{X}^\star\mathbb{R}_\mathbf{X}$ being a symmetric PSD matrix(Used within proof of Lemma 6 in (Ge et al., 2017)), we could derive an inequality which does not depend on n.

References

Ge, R., Jin, C., and Zheng, Y. (2017). No spurious local minima in nonconvex low rank problems: A unified geometric analysis. *arXiv preprint arXiv:1704.00708*.