Riemannian Optimization for Euclidean Distance Geometry

Chandler Smith

Tufts University

chandler.smith@tufts.edu

April 25, 2024

Overview

1. Euclidean Distance Geometry (EDG) Problem
   - Matrix Completion
   - Existing Riemannian Technique
2. Novel Algorithmic Approach
3. Results

Motivation

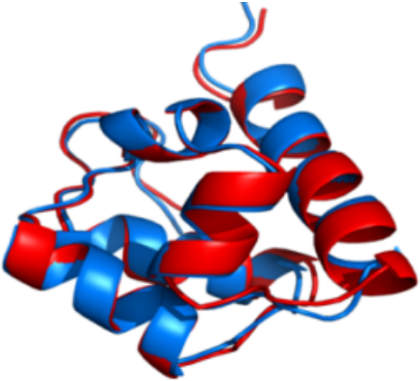

Figure: 3D image of a protein (Parhizkar, 2012)

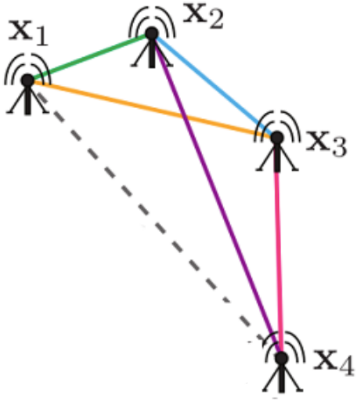

Figure: Example of a sensor network (Cucuringu et al., 2012)

Full EDG Problem I

- Let $\{\mathbf{p}_k\}_{k=1}^n \subset \mathbb{R}^d$ for some $d < n$, and define $\mathbf{P} \in \mathbb{R}^{n \times d}$

$$\mathbf{P} = \begin{bmatrix} - & \mathbf{p}_1^T & - \\ - & \mathbf{p}_2^T & - \\ & \vdots & \\ - & \mathbf{p}_n^T & - \end{bmatrix}, \quad \mathbf{P} \cdot \mathbf{1} = \mathbf{0}$$

- The squared distance between $\mathbf{p}_i$ and $\mathbf{p}_j$ is

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

- Define the Distance Matrix

$$\mathbf{D} = [d_{ij}^2] \in \mathbb{R}^{n \times n}$$

- One can show $\mathrm{rank}(\mathbf{D}) \leq d + 2$.

Full EDG Problem II

- Can also define the Gram Matrix

$$\mathbf{X} = \mathbf{P}\mathbf{P}^T = [\mathbf{p}_i^T \mathbf{p}_j] \in \mathbb{R}^{n \times n}$$

- Notice that for orthogonal $\mathbf{O} \in \mathbb{R}^{d \times d}$ that $\mathbf{P}\mathbf{P}^T = (\mathbf{P}\mathbf{O})(\mathbf{P}\mathbf{O})^T = \mathbf{X}$, so $\mathbf{P} \mapsto \mathbf{X}$ not injective. Additionally, notice that $\mathrm{rank}(\mathbf{X}) \leq d$
- Interested in the problem where we have access to $\mathbf{D}$ and want to compute $\mathbf{P}$ up to orthogonal transformation.
- Closed form relationships between $\mathbf{D}$ and $\mathbf{X}$ exist!

$$\mathbf{D} = \mathrm{diag}(\mathbf{X})\mathbf{1}^T + \mathbf{1}\mathrm{diag}(\mathbf{X})^T - 2\mathbf{X} \tag{1}$$

$$\mathbf{X} = -\frac{1}{2}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\mathbf{D}\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) \tag{2}$$

Multidimensional Scaling (MDS)

Using (2), we can compute **P** as follows from **D**:

$$\mathbf{X} = \overbrace{\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_d \\ | & | & & | \end{bmatrix}}^{n \times d} \overbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}}^{d \times d} \overbrace{\begin{bmatrix} \text{—} & \mathbf{u}_1^T & \text{—} \\ \text{—} & \mathbf{u}_2^T & \text{—} \\ & \vdots & \\ \text{—} & \mathbf{u}_d^T & \text{—} \end{bmatrix}}^{d \times n}$$

$$= \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_d \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_d^{1/2} \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_d^{1/2} \end{bmatrix} \begin{bmatrix} \text{—} & \mathbf{u}_1^T & \text{—} \\ \text{—} & \mathbf{u}_2^T & \text{—} \\ & \vdots & \\ \text{—} & \mathbf{u}_d^T & \text{—} \end{bmatrix}}_{\mathbf{P}^T}$$

Procrustes Problem

- Let $\tilde{\mathbf{P}}$ be a recovered point matrix through MDS, and $\mathbf{P}$ be the true distance. We can align these point clouds by solving

$$\underset{\mathbf{O}\in\mathcal{O}(d)}{\text{minimize}}\ \|\mathbf{P} - \tilde{\mathbf{P}}\mathbf{O}\|_F \tag{3}$$

- Can solve through a variational argument and see that, if $\tilde{\mathbf{P}}\mathbf{P}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, then

$$\underset{\mathbf{O}\in\mathcal{O}(d)}{\text{argmin}}\ \|\mathbf{P} - \tilde{\mathbf{P}}\mathbf{O}\|_F = \mathbf{V}\mathbf{U}^T \tag{4}$$

- Say we have access to a subset of $\{\mathbf{p}_i\}_{i\in I}$, and define $\mathbf{P}_I = [\mathbf{p}_{i_1}^T ... \mathbf{p}_{i_j}^T]^T$, $\tilde{\mathbf{P}}_I = [\tilde{\mathbf{p}}_{i_1}^T ... \tilde{\mathbf{p}}_{i_j}^T]^T$.
- Can solve $\mathbf{O} = \mathbf{V}\mathbf{U}^T$ for $\tilde{\mathbf{P}}_I^T\mathbf{P}_I = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, and then align $\tilde{\mathbf{P}}$ with $\mathbf{P}$ via $\tilde{\mathbf{P}}\mathbf{O}$.

Partial EDG Problem I

- Problem is more interesting when considering a subset $\Omega \subset [n] \times [n]$ where if $(i,j) \in \Omega$, d_{ij}^2 is known.
- Define $\mathcal{P}_\Omega : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ as follows:

$$\mathcal{P}_\Omega(\cdot) = \sum_{(i,j)\in\Omega} \langle \cdot, \mathbf{E}_{ij} \rangle \mathbf{E}_{ij} \tag{5}$$

  where $\langle \mathbf{A}, \mathbf{B} \rangle = \mathrm{Trace}(\mathbf{A}^T\mathbf{B}) = \sum_{i,j} A_{ij}B_{ij}$ and $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^T$.
- Only $L = \frac{n(n-1)}{2}$ unique entries in $\mathbf{D}$, and $\mathbf{D} = \mathbf{D}^T$
- For $\alpha = (i,j),\ i < j$, if we define $\mathbf{E}_\alpha := \mathbf{E}_{ij} + \mathbf{E}_{ji}$, we can rewrite $\mathcal{P}_\Omega$ as

$$\mathcal{P}_\Omega(\mathbf{D}) = \sum_{\alpha\in\Omega} \langle \mathbf{D}, \mathbf{E}_\alpha \rangle \mathbf{E}_\alpha \tag{6}$$

- Goal becomes, can we reconstruct $\mathbf{D}$ from viewing $\mathcal{P}_\Omega(\mathbf{D})$?

Matrix Completion I

- Given a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with $\mathrm{rank}(\mathbf{A}) = r \ll n$, lots of repeated information in entries of $\mathbf{A}$.
- If $r = 1$, need at least 1 entry in each row/column. If sampling at random, expected number of samples to achieve this is $\approx n\log(n)$
- What about matrices like the following?

$$\underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & 0 \end{bmatrix}}_{\text{coherent}} \qquad \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & & \ddots & \\ 1 & & & 1 \end{bmatrix}}_{\text{incoherent}} \tag{7}$$

- Mathematical notion of entrywise "diffuseness" necessary to exclude pathological cases

Matrix Completion II

- If we wanted to compute the ground truth matrix $\mathbf{M}$, we can define the following optimization routine

$$\underset{\mathbf{X}\in\mathbb{R}^{n\times n}}{\text{minimize}} \ \text{rank}(\mathbf{X}) \ \text{subject to} \ \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}) \tag{8}$$

- Rank is non-smooth, optimization is combinatorially hard. Replace rank with a convex surrogate function $\|\mathbf{Y}\|_\star = \sum_{i=1}^n \sigma_i$ where σ_i is the i-th largest singular value of $\mathbf{Y}$
- Defining new convex routine yields

$$\underset{\mathbf{X}\in\mathbb{R}^{n\times n}}{\text{minimize}} \ \|\mathbf{X}\|_\star \ \text{subject to} \ \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M}) \tag{9}$$

which has lots of nice convergence guarantees.

So why spend time on this if methods exist?

- Scalability
- Convergence guarantees for non-convex methods
- Distance matrices do not play nicely with existing methods
- Have not leveraged knowledge of dimension of point cloud

Riemannian Approach I

- Common approach in constrained optimization is if the constraint is smooth, can consider algorithms as unconstrained gradient descent on a manifold.
- Algorithmic intuition is that you step in an "allowable" direction on a manifold (in the tangent space) by some step size, then pull back on to the manifold

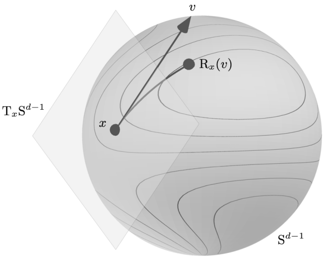

Figure: Tangent space of a sphere with a retraction map (Boumal, 2023)

Riemannian Approach II

- In (Wei et al., 2020), the following approach was established. The set

$$\mathcal{M} := \{\mathbf{X} \in \mathbb{R}^{n\times n} \mid \operatorname{rank}(\mathbf{X}) = d\}$$

is an embedded manifold in $\mathbb{R}^{n\times n}$. This allows us to define the following algorithm.

$$\underset{\mathbf{X}\in\mathbb{R}^{n\times n}}{\text{minimize}} \ \langle \mathbf{X} - \mathbf{M}, \mathcal{P}_\Omega(\mathbf{X} - \mathbf{M})\rangle \ \text{subject to} \ \operatorname{rank}(\mathbf{X}) = d \tag{10}$$

- At a point $\mathbf{X} \in \mathcal{M}$. the tangent space

$$T_{\mathbf{X}}\mathcal{M} = \{\mathbf{U}\mathbf{Z}_1^T + \mathbf{Z}_2\mathbf{V}^T | \mathbf{Z}_1, \mathbf{Z}_2 \in \mathbb{R}^{n\times d}\} \tag{11}$$

where $\mathbf{X} = \mathbf{U\Sigma V}$ for $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times d}$.

Riemannian Approach III

Using the previous ideas, Wei et al. define the following algorithm

RGrad

Initialize with $\mathbf{X}_0 = \texttt{SVD}_r(\mathcal{P}_\Omega(\mathbf{M})) = \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{V}_0^T = \sum_{\imath=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$. For $l = 0, 1, 2...$ do

1. $\mathbf{G}_l = \mathcal{P}_\Omega(\mathbf{X}_l - \mathbf{M})$
2. $\alpha_l = \frac{\|\mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l\|_F^2}{\langle \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l, \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l \rangle}$
3. $\mathbf{W}_l = \mathbf{X}_l + \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l$
4. $\mathbf{X}_{l+1} = \texttt{SVD}_r(\mathbf{W}_l)$

where $\mathbb{T}_l$ is the tangent space at the l-th iteration and $\mathbb{T}$ is the tangent space at the true solution. Some assumptions on $|\Omega|$ required to prove convergence.

Translating to optimization over Gram matrices

- We only have access to d_{ij}^2 for $(i,j) \in \Omega$.
- Recall that

$$\begin{aligned} d_{ij}^2 &= \mathbf{p}_i^T \mathbf{p}_i + \mathbf{p}_j^T \mathbf{p}_j - 2\mathbf{p}_i^T \mathbf{p}_j \\ &= X_{ii} + X_{jj} - X_{ij} - X_{ji} \end{aligned}$$

- The relevant set will be

$$\mathbb{S} = \{\mathbf{Y} \in \mathbb{R}^{n \times n} | \mathbf{Y} = \mathbf{Y}^T, \mathbf{Y} \cdot \mathbf{1} = \mathbf{0}\}, \quad \dim(\mathbb{S}) = L$$

- Let $\alpha = (i,j)$ for $i < j$. Define $\mathbf{w}_\alpha = \mathbf{E}_{ii} + \mathbf{E}_{jj} - \mathbf{E}_{ij} - \mathbf{E}_{ji}$. It follows now that

$$d_{ij}^2 = \langle \mathbf{X}, \mathbf{w}_\alpha \rangle \tag{12}$$

- One can show that $\{\mathbf{w}_\alpha\}_{\alpha \in \mathbb{I}}$ forms a basis for $\mathbb{S}$ but $\langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle \neq \delta_{\alpha\beta}$.

Dual Basis

- Can define dual, or bi-orthogonal, basis $\{\mathbf{v}_\alpha\}_{\alpha\in\mathbb{I}}$ such that $\langle \mathbf{w}_\alpha, \mathbf{v}_\beta \rangle = \delta_{\alpha\beta}$.
- Defining $\mathbf{H} = [\langle \mathbf{w}_\alpha, \mathbf{w}_\beta \rangle] \in \mathbb{R}^{L\times L}$ allows us to analytically compute the dual basis $\{\mathbf{v}_\alpha\}$

$$\mathbf{v}_\alpha = \sum_{\beta\in\mathbb{I}} \mathbf{H}^{-1}_{\alpha\beta}\mathbf{w}_\beta \tag{13}$$

- The structure of the dual basis also allows any $\mathbf{Y} \in \mathbb{S}$ to be decomposed as

$$\mathbf{Y} = \sum_{\alpha\in\mathbb{I}} \langle \mathbf{Y}, \mathbf{w}_\alpha \rangle \mathbf{v}_\alpha = \sum_{\alpha\in\mathbb{I}} \langle \mathbf{Y}, \mathbf{v}_\alpha \rangle \mathbf{w}_\alpha \tag{14}$$

- Using these ideas, we define a new sampling operator $\mathcal{R}_\Omega : \mathbb{S} \to \mathbb{S}$

$$\mathcal{R}_\Omega(\cdot) = \sum_{\alpha\in\Omega} \langle \cdot, \mathbf{w}_\alpha \rangle \mathbf{v}_\alpha \tag{15}$$

Defining Algorithm I

One problem with $\mathcal{R}_\Omega$ is that it is not self-adjoint.
Defining the function

$$f(\mathbf{Y}) = \langle \mathbf{Y} - \mathbf{X}, \mathcal{R}_\Omega(\mathbf{Y} - \mathbf{X}) \rangle \tag{16}$$

We can see that

$$\nabla_{\mathbf{Y}} f(\mathbf{Y}) = \frac{1}{2}(\mathcal{R}_\Omega(\mathbf{Y} - \mathbf{X}) + \mathcal{R}_\Omega^\star(\mathbf{Y} - \mathbf{X})) \tag{17}$$

But this requires knowledge of $\mathcal{R}_\Omega^\star(\mathbf{X})$, and

$$\mathcal{R}_\Omega^\star(\cdot) = \sum_{\alpha \in \Omega} \langle \cdot, \mathbf{v}_\alpha \rangle \mathbf{w}_\alpha \tag{18}$$

and we don't have access to $\langle \mathbf{X}, \mathbf{v}_\alpha \rangle$

Defining Algorithm II

One option is to define a self-adjoint surrogate $\mathcal{R}_\Omega^\star \mathcal{R}_\Omega$ as follows:

$$\mathcal{R}_\Omega^\star \mathcal{R}_\Omega(\cdot) = \sum_{\alpha,\beta\in\Omega} \langle \cdot, \mathbf{w}_\alpha \rangle \langle \mathbf{v}_\alpha, \mathbf{v}_\beta \rangle \mathbf{w}_\beta \tag{19}$$

This allows us to define a new objective function

$$\underset{\mathbf{Y}\in\mathbb{R}^{n\times n}}{\text{minimize}} \; \langle \mathbf{Y}-\mathbf{X}, \mathcal{R}_\Omega^\star \mathcal{R}_\Omega(\mathbf{Y}-\mathbf{X}) \rangle \; \text{subject to} \; \text{rank}(\mathbf{Y}) = r \tag{20}$$

$\mathcal{R}_\Omega^\star \mathcal{R}_\Omega$ Algorithm

Initialize with $\mathbf{X}_0 = \texttt{EVD}_r(\mathcal{R}_\Omega^\star \mathcal{R}_\Omega(\mathbf{X})) = \mathbf{U}_0 \mathbf{\Sigma}_0 \mathbf{U}_0^T$ For $l = 0, 1, 2...$ do

1. $\mathbf{G}_l = \mathcal{R}_\Omega^\star \mathcal{R}_\Omega(\mathbf{X}_l - \mathbf{X})$
2. $\alpha_l = \frac{\|\mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l\|_F^2}{\langle \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l, \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l \rangle}$
3. $\mathbf{W}_l = \mathbf{X}_l + \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l$
4. $\mathbf{X}_{l+1} = \texttt{EVD}_r(\mathbf{W}_l)$

Defining Algorithm III

Another self-adjoint surrogate, $\mathcal{F}_\Omega$, is defined as follows:

$$\mathcal{F}_\Omega(\cdot) = \sum_{\alpha \in \Omega} \langle \cdot, \mathbf{w}_\alpha \rangle \mathbf{w}_\alpha \tag{21}$$

This allows us to define a new objective function

$$\underset{\mathbf{Y} \in \mathbb{R}^{n \times n}}{\text{minimize}} \ \langle \mathbf{Y} - \mathbf{X}, \mathcal{F}_\Omega(\mathbf{Y} - \mathbf{X}) \rangle \ \text{subject to} \ \text{rank}(\mathbf{Y}) = r \tag{22}$$

$\mathcal{F}_\Omega$ Algorithm

Initialize with $\mathbf{X}_0 = \texttt{EVD}_r(\mathcal{F}_\Omega(\mathbf{X})) = \mathbf{U}_0 \mathbf{\Sigma}_0 \mathbf{U}_0^T$ For $l = 0, 1, 2 \ldots$ do

1. $\mathbf{G}_l = \mathcal{F}_\Omega(\mathbf{X}_l - \mathbf{X})$
2. $\alpha_l = \frac{\|\mathcal{P}_{\mathbb{T}_l} \mathbf{G}_l\|_F^2}{\langle \mathcal{P}_{\mathbb{T}_l} \mathbf{G}_l, \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_l} \mathbf{G}_l \rangle}$
3. $\mathbf{W}_l = \mathbf{X}_l + \alpha_l \mathcal{P}_{\mathbb{T}_l} \mathbf{G}_l$
4. $\mathbf{X}_{l+1} = \texttt{EVD}_r(\mathbf{W}_l)$

Defining Algorithm IV

Alternatively, lose quadratic form interpretation, as no reference in algorithm to objective function, as follows:

$\mathcal{R}_\Omega$ Algorithm

Initialize with $\mathbf{X}_0 = \mathtt{EVD}_r(\mathcal{R}_\Omega(\mathbf{X})) = \mathbf{U}_0\mathbf{\Sigma}_0\mathbf{U}_0^T$ For $l = 0, 1, 2...$ do

1. $\mathbf{G}_l = \mathcal{R}_\Omega(\mathbf{X}_l - \mathbf{X})$
2. $\alpha_l = \frac{\|\mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l\|_F^2}{\langle \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l, \mathcal{P}_\Omega \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l\rangle}$
3. $\mathbf{W}_l = \mathbf{X}_l + \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathbf{G}_l$
4. $\mathbf{X}_{l+1} = \mathtt{EVD}_r(\mathbf{W}_l)$

Why might we prefer one over the other? Relies on statistical properties of the respective operators, which is relevant to convergence analysis.

Convergence Analysis

How does one undertake these types of convergence analysis proofs? Want contraction in the Frobenius norm.

$$
\begin{aligned}
\|\mathbf{X}_{l+1} - \mathbf{X}\|_F &= \|\mathbf{X}_{l+1} - \mathbf{W}_l + \mathbf{W}_l - \mathbf{X}\|_F \\
&\leq \|\mathbf{X}_{l+1} - \mathbf{W}_l\|_F + \|\mathbf{W}_l - \mathbf{X}\|_F \\
&\leq 2\|\mathbf{W}_l - \mathbf{X}\|_F \\
&= 2\|(\mathbf{X}_l - \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathcal{R}_\Omega(\mathbf{X} - \mathbf{X}_l)) - \mathbf{X}\|_F \\
&= 2\|(\mathcal{I} - \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathcal{R}_\Omega)(\mathbf{X}_l - \mathbf{X})\|_F \\
&\leq \underbrace{2\|(\mathcal{P}_{\mathbb{T}_l} - \alpha_l \mathcal{P}_{\mathbb{T}_l}\mathcal{R}_\Omega\mathcal{P}_{\mathbb{T}_l})(\mathbf{X}_l - \mathbf{X})\|_F}_{I_1} \\
&\quad + \underbrace{2\|(\mathcal{I} - \mathcal{P}_{\mathbb{T}_l})(\mathbf{X}_l - \mathbf{X})\|_F}_{I_2} \\
&\quad + \underbrace{2|\alpha_l|\|\mathcal{P}_{\mathbb{T}_l}\mathcal{R}_\Omega(\mathcal{I} - \mathcal{P}_{\mathbb{T}_l})(\mathbf{X}_l - \mathbf{X})\|_F}_{I_3}
\end{aligned}
$$

I_1 is the tricky term!

Restricted Isometry Property

All convergence results in matrix completion rely on the Restricted Isometry Property (RIP).

Idea is simple: In expectation, how much do the sampling operators $\mathcal{K}_\Omega \in \{\mathcal{R}_\Omega^\star \mathcal{R}_\Omega, \mathcal{R}_\Omega, \mathcal{F}_\Omega\}$ deviate from the identity (when restricted to $\mathbb{T}$)? More formally, is

$$\|\mathcal{P}_\mathbb{T} \mathcal{K}_\Omega \mathcal{P}_\mathbb{T} - c\mathcal{P}_\mathbb{T}\| \tag{23}$$

small for some constant $c > 0$?

Theorem (RIP of $\mathcal{R}_\Omega$)

Suppose Ω is a set of entries of size m sampled independently and uniformly with replacement. Then for all $\beta > 1$

$$\left\| \frac{L}{m} \mathcal{P}_\mathbb{T} \mathcal{R}_\Omega \mathcal{P}_\mathbb{T} - \mathcal{P}_\mathbb{T} \right\| \leq \frac{1}{2} \tag{24}$$

with probability $1 - 2n^{1-\beta}$ provided that $m \geq C\beta n \log(n)$ for some $C = O(1)$

Results I

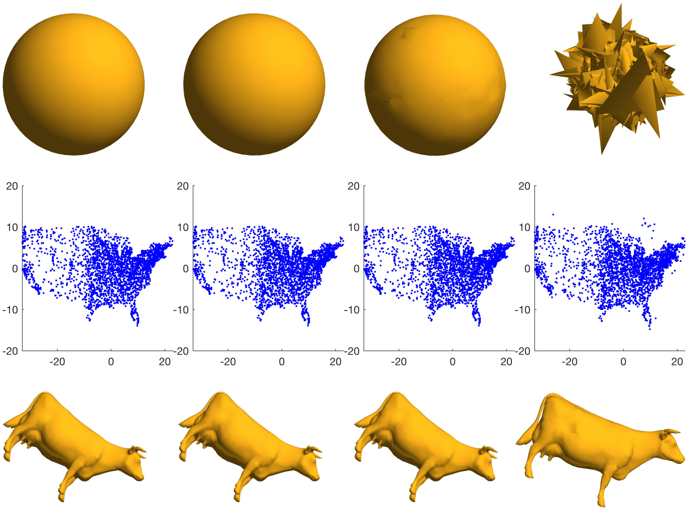

Figure: All results captured using $\mathcal{R}_\Omega^\star \mathcal{R}_\Omega$ algorithm

Results II

Table: Relative recovery error $\|\mathbf{X} - \mathbf{X}_{\text{rev}}\|_F / \|\mathbf{X}\|_F$ between the recovered Gram matrix and the true Gram matrix averaged over 20 trials using the $\mathcal{R}_\Omega$ algorithm.

Dataset \ γ	5%	3%	2%	1%
Sphere (3D, $n = 1002$)	1.5e-07	3.0e-06	8.3e-07	0.46
U.S. Cities (2D, $n = 2920$)	5.9e-08	1.1e-07	2.0e-07	7.3e-07
Cow (3D, $n = 2601$)	6.1e-08	1.3e-07	2.3e-07	8.9e-07
Swiss Roll (3D, $n = 2048$)	8.1e-08	1.3e-07	3.0e-06	0.0035

The End

References

Emmanuel Candes, Justin Romberg, Terence Tao (2004)
Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information

Emmanuel Candes and Benjamin Recht (2008)
Exact Matrix Completion via Convex Optimization

David Gross (2010)
Recovering low-rank matrices from few coefficients in any basis

Benjamin Recht (2011)
A simpler approach to matrix completion

Ke Wei, Jian-Feng Cai, Tony F. Chan, Shingyu Leung (2020)
Guarantees of Riemannian Optimization for Low Rank Matrix Completion

Chandler Smith, Samuel Lichtenberg, Abiy Tasissa, HanQin Cai (2023)
Riemannian Optimization for Euclidean Distance Geometry