

Recovering Wasserstein Distance Matrices from Few Measurements

Muhammad Rana¹, Abiy Tasissa³, HanQin Cai⁴, Yakov Gavriyelov¹, and Keaton Hamm^{1,2}

¹Department of Mathematics, University of Texas at Arlington

²Division of Data Science, University of Texas at Arlington

³Department of Mathematics, Tufts University, Medford, MA 02155, USA.

⁴Department of Statistics and Data Science and Department of Computer Science,
University of Central Florida

Abstract

This paper proposes two algorithms for estimating square Wasserstein distance matrices from a small number of entries. These matrices are used to compute manifold learning embeddings like multidimensional scaling (MDS) or Isomap, but contrary to Euclidean distance matrices, are extremely costly to compute. We analyze matrix completion from upper triangular samples and Nyström completion in which $\mathcal{O}(d \log(d))$ columns of the distance matrices are computed where d is the desired embedding dimension, prove stability of MDS under Nyström completion, and show that it can outperform matrix completion for a fixed budget of sample distances. Finally, we show that classification of the OrganCMNIST dataset from the MedMNIST benchmark is stable on data embedded from the Nyström estimation of the distance matrix even when only 10% of the columns are computed.

1 Introduction

The bare representation of data in many applications is high-dimensional [1]. In today’s era of big data, this high dimensionality poses significant challenges and necessitates specialized methods. Manifold learning refers to a class of techniques that compress data representations into lower-dimensional forms, making them more suitable for analysis and computation. A widely used linear technique is principal component analysis (PCA). However, linear approaches can be restrictive when the data lie on or near a nonlinear manifold. To address this, nonlinear manifold learning techniques aim to obtain compact representations while preserving local geometric structures [2]. We note that, in the standard way they are used, all of the aforementioned methods typically treat data as vectors in a Euclidean space.

While many applications yield vector data, there are a number of domains that produce data that may be best modeled as distributions or measures. For example, gene expression and natural language processing data can be represented as distributions over gene networks or dictionaries, respectively, and image data can be considered as discrete distributions over a 2-dimensional pixel grid. Recent works have used this notion to improve understanding of data geometry and performance on tasks such as classification or clustering [3, 4, 5]. One of the advantages of treating images, for example, as measures, is that spatial relationships are not obscured as they are under vectorization. Moreover, the natural notion of a metric structure on the space of measures is that of the Wasserstein metric from optimal transport (OT) [6].

Several works have explored manifold learning using probability measures and the Wasserstein metric, e.g., [5, 7]. This approach has shown improvement in task performance and data modeling, but it comes with the significant drawback of slow computation time. In particular, computing one Wasserstein distance exactly for discrete measures supported on m points costs $\Omega(m^3 \log(m))$ flops. To form neighbor embeddings for N data measures requires computing the square distance matrix $D_{ij} = W_2(\mu_i, \mu_j)^2$, carrying total complexity $\mathcal{O}(N^2 m^3 \log(m))$.

Mitigating this high computational cost can be approached in two ways: 1) speeding up a single Wasserstein distance computation, and 2) computing fewer entries of D to form an approximation of it. Many methods from computational OT explore the first approach, e.g., through entropic regularization, multiscale methods, or Nyström approximations to the cost matrix in the discrete formulation of the Kantorovich problem (see [6] for a survey). In contrast, relatively few works have explored the problem of estimating an entire Wasserstein distance matrix directly from samples.

The focus of this paper is to propose algorithms for estimating square Wasserstein distance matrices for use in neighbor embeddings by computing as few entries as possible. Our approach is based on exploiting the approximate low-rank structure of Wasserstein squared distances: if the matrix D is approximately rank r , we propose two algorithms based on matrix completion and the Nyström method which allow us to estimate D from $\mathcal{O}(r \log(r))$ fully observed columns and maintain stability of the MDS embedding.

2 Technical background

2.1 Optimal transport and MDS

The Kantorovich formulation of the quadratic Wasserstein metric on discrete probability measures $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$ and $\nu = \sum_{i=1}^m \beta_i \delta_{y_i}$ is

$$W_2(\mu, \nu) := \min_{P \in \Gamma(\mu, \nu)} \langle C, P \rangle \quad (1)$$

where $\Gamma(\mu, \nu) = \{P \in \mathbb{R}_+^{m \times n} : P\mathbf{1}_m = \alpha, P^\top \mathbf{1}_n = \beta\}$ are couplings with marginals μ and ν , and $C_{ij} = \|x_i - y_j\|^2$ [6]. This problem computes an optimal matching of mass between two discrete measures. Given measures $\{\mu_i\}_{i=1}^N \subset W_2(\mathbb{R}^n)$, the Wassmap algorithm [5, 7] computes $D_{ij} = W_2(\mu_i, \mu_j)^2$, a double centering of it $B = -\frac{1}{2}H D H$ where $H = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, with truncated SVD $B_d = V_d \Sigma_d V_d^\top$. The embedding vectors in \mathbb{R}^d are then rows of $V_d \Sigma_d$.

2.2 Matrix completion

Given a low-rank matrix with missing entries, the problem of estimating the unknown entries is known as the matrix completion problem. This problem has been studied extensively due to its broad applications, most notably in recommendation systems [8]. A naive formulation seeks to minimize the rank of the matrix subject to the constraints imposed by the observed entries; however, rank minimization is computationally hard in general [9]. A common convex relaxation is to minimize the nuclear norm, which serves as a surrogate for rank. In this convex setting, seminal theoretical results show that, under random sampling, an $N \times N$ rank- r matrix can be exactly recovered from $\mathcal{O}(\nu r N \log N)$ observed entries, where ν is an incoherence parameter capturing the conditioning of the underlying matrix [10]. Given the computational challenges of convex methods, non-convex algorithms have also been explored, and similar recovery guarantees have been established in this setting [11]. In this paper, we focus on the completion of distance matrices. In the case of Euclidean distance matrices, it can be shown that the squared distance matrix

has rank at most $d + 2$, where d is the embedding dimension of the underlying points [12]. This structural property enables the use of low-rank optimization techniques for Euclidean distance matrix completion [13, 14].

2.3 Nyström method

Consider a symmetric matrix in block form $D = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, where $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{m \times n}$. We are interested in the regime $m \ll n$, where the aim is to estimate the block C using only the smaller blocks A and B . Suppose that $\text{rank}(A) = \text{rank}(D)$. Then, one can show that the columns of the submatrix $\begin{pmatrix} B \\ C \end{pmatrix}$ lie in the span of the columns of $\begin{pmatrix} A \\ B^\top \end{pmatrix}$. Specifically, there exists an $X \in \mathbb{R}^{m \times n}$ such that $B = AX$ and $C = B^\top X$. If A is invertible, then $X = A^{-1}B$, which yields the exact relation $C = B^\top A^{-1}B$. In the more general case where A is not invertible, we may instead use the Moore–Penrose pseudoinverse to obtain $X = A^\dagger B$ leading to $C = B^\top A^\dagger B$. This procedure is known as the Nyström method [15].

While it is not a completion algorithm per se, the Nyström method can be used to complete a distance matrix. In particular, if only the columns containing A and B^\top are computed, then the entries of C can be recovered, or in the general case approximated, by $B^\top A^\dagger B$.

Notations For a distance matrix D , owing to symmetry and zero diagonal entries, we only consider samples from the strictly upper-triangular part. We denote by $\Omega \subset \{(i, j) : i \in [N], j > i\}$ the set of sampled index pairs, with cardinality $|\Omega| = m$. Given a matrix D and an index set $I \subset [N] := \{1, \dots, N\}$, we write $D(:, I)$ for the submatrix of D consisting of the columns indexed by I . Analogously, $D(I, :)$ denotes the submatrix formed by the rows indexed by I . $P_\Omega(D)_{ij} = D_{ij}$ if $(i, j) \in \Omega$ and is 0 otherwise. If $\text{rank}(D) = r$ and $D = V_r \Sigma_r V_r^\top$, then the incoherence ν of D is $\sqrt{n/r} \max_i \|V_r(i, :)\|$.

3 Algorithms

Here we describe two algorithms for estimating the Wasserstein distance matrix to compute embeddings of images. Algorithm 1, **W₂-MC**, randomly computes a given number of entries of the distance matrix, and then uses a matrix completion algorithm to estimate the full squared distance matrix. We leave the completion algorithm as a generic plug-in for existing algorithms, and consider one designed for distance matrices in our experiments. We sample from the strict upper triangle, include the diagonal, and reflect the entries across the diagonal to obtain the observed distance matrix. We apply the same procedure after completion to ensure that the resulting matrix retains the properties of a distance matrix.

Algorithm 2, **W₂-Nyström**, uses the Nyström method to complete the distance matrix. In particular, we compute only a set of columns of the distance matrix $C_{\text{obs}} = D(:, I)$, and estimate $D_{\text{est}} = C_{\text{obs}} U C_{\text{obs}}^\top$ where $U = C_{\text{obs}}(I, :) = D(I, I)$. As in **W₂-MC**, we also enforce the distance matrix structure by enforcing the 0 diagonal.

Note that for large N , the computational cost of both algorithms is dominated by the cost of computing the W_2 distances, and the cost of the matrix completion subroutine **MC** is often negligible by comparison. Thus, algorithms that limit sampling complexity of the matrix are more desirable. For this reason, we do not truncate the rank of the Nyström approximation to D in **W₂-Nyström**, which is usually done to save computation time.

Algorithm 1: W_2 -MC

- 1 **Input:** Data measures $\{\mu_i\}_{i=1}^N$, sample size m , matrix completion subroutine **MC**
 - 2 Randomly sample m entries Ω
 - 3 Compute observed distance matrix $D_{\text{obs}} = P_{\Omega}(D)$, where $D_{ij} = W_2(\mu_i, \mu_j)^2$
 - 4 $\text{diag}(D_{\text{obs}}) = 0$ ▷ enforce 0 diagonal
 - 5 $D_{\text{obs}} = D_{\text{obs}} + D_{\text{obs}}^{\top}$ ▷ reflect across diagonal
 - 6 $D_{\text{est}} = \text{MC}(D_{\text{obs}})$
 - 7 $D_{\text{est}} = \frac{1}{2}(D_{\text{est}} + D_{\text{est}}^{\top})$ ▷ enforce symmetry
 - 8 **Output:** Estimated distance matrix D_{est}
-

Algorithm 2: W_2 -Nyström

- 1 **Input:** Data measures $\{\mu_i\}_{i=1}^N$, number of columns c
 - 2 Randomly sample c columns $I \subset [N]$
 - 3 Compute columns of distance matrix $C_{\text{obs}} = \begin{cases} W_2(\mu_i, \mu_j)^2, & i \in [N], j \in I \\ 0, & \text{otherwise.} \end{cases}$
 - 4 $D_{\text{est}} = C_{\text{obs}} U C_{\text{obs}}^{\top}$, where $U = C_{\text{obs}}(I, :)$
 - 5 $\text{diag}(D_{\text{est}}) = 0$ ▷ enforce 0 diagonal
 - 6 **Output:** Estimated distance matrix D_{est}
-

The Nyström method is well-studied, and using some existing bounds, we are able to prove a stability result for MDS embeddings on Wasserstein distance matrices. We assume that the data measures in $W_2(\mathbb{R}^n)$ are isometric to a set of Euclidean vectors $\{y_i\}_{i=1}^N \subset \mathbb{R}^d$. Examples of such sets have been studied and include common image articulations like translations, scalings, and shears of a fixed base measure [16, 7]. We assume noise (or error) in the computation of the W_2 distances, and consider how the MDS embedding $\{z_i\}_{i=1}^N \subset \mathbb{R}^d$ coming from D_{est} of W_2 -Nyström compares to $\{y_i\}$. MDS embeddings are centered and equivalent up to orthogonal transformation, so we measure error by a Procrustes distance, defined as $\min_{Q \in O(d)} \|Z - QY\|_2$, where $O(d)$ is the set of orthogonal $d \times d$ matrices. Note that Z and Y are matrices whose columns are the embedded vectors.

Theorem 3.1. *Suppose $\{\mu_i\}_{i=1}^N \subset W_2(\mathbb{R}^n)$, and $D_{ij} = W_2(\mu_i, \mu_j)^2$ has incoherence ν . Suppose there exist $\{y_i\}_{i=1}^n \subset \mathbb{R}^d$ such that $D_{ij} = \|y_i - y_j\|^2$. Let $\lambda > 0$, $\delta \in (0, 1)$, and compute $m \gtrsim \lambda \nu^2 (d + 2) \max\{\log(d + 2), \log(1/\delta)\}$ columns uniformly randomly in W_2 -Nyström with error $C_{\text{obs}_{i,j}} = W_2(\mu_i, \mu_j)^2 + E_{ij}$ such that $\|E\|_2 \leq \sigma_{d+2}(D)/3$. If $\|Y^{\dagger}\|_2 \|E\|_2^{1/2} \leq 1/\sqrt{2}$, then $D_{\text{est}} = C_{\text{obs}} U C_{\text{obs}}^{\top}$ yields an MDS embedding $\{z_i\}_{i=1}^N \subset \mathbb{R}^d$ such that*

$$\min_{Q \in O(d)} \|Z - QY\|_2 \leq (1 + \sqrt{2}) \|Y^{\dagger}\|_2 \|E\|_2 \times$$

$$\left(5 \frac{N}{(d+2)(1-\delta)} + 3 \sqrt{\frac{N}{(d+2)(1-\delta)}} + 2 \right)$$

with probability at least $1 - \max\{re^{-\lambda(\delta+(1-\delta)\log(1-\delta))}, \delta\}$.

Proof. A simple modification of the proof of [17, Theorem 3.3] implies the upper bound with $(1 + \sqrt{2}) \|Y^{\dagger}\|_2 \|D - D_{\text{est}}\|_2$. Then [18, Theorem 3] implies that the columns selected would yield

an exact Nyström decomposition of D without the presence of noise. Thus, the hypotheses of [19, Remark 4.7] are satisfied, and yield that $\|D - D_{\text{est}}\|_2 \leq 5\|V_{d+2}(I, :)^{\dagger}\|_2^2 + 3\|V_{d+2}(I, :)^{\dagger}\|_2 + 2$ where I is the set of columns selected by Algorithm 2 and V_{d+2} are the first $d + 2$ left singular vectors of D . Finally, [20, Lemma 3.4] shows that with the given probability, $\|V_{d+2}(I, :)^{\dagger}\|_2 \leq \sqrt{N/((d+2)(1-\delta))}$ for all $\delta \in (0, 1)$. \square

A simple corollary is that if we compute $\mathcal{O}((d+2)\log(d+2))$ columns of D without error, then the MDS embedding is exact up to orthogonal transformation. Exact MDS embeddings of W_2 data were studied in [7].

4 Numerical Experiments

4.1 Fixed sample budget

We first do a fixed sample budget comparison of W_2 -MC and W_2 -Nyström for completing the Wasserstein distance matrix of 2,000 points from the OrganCMNIST dataset from the MedMNIST benchmark [21], comprising coronal plane CT scans of 11 different organs downsampled to size 28×28 . For uniform sampling, we subsample the matrix at different rates $\gamma \in \{3\%, 5\%, 10\%, 20\%, 25\%\}$. We then match the number of observed entries between uniform sampling and the Nyström method by solving the equation $\frac{c(c-1)}{2} + c(n-c) = \gamma \cdot \frac{n(n-1)}{2}$ for c (the number of columns selected by Nyström). This equation equates the number of off-diagonal entries used by both methods. For MC in W_2 -MC, we use the factorized Gram matrix formulation from [14] based on the relationship between the Gram matrix and the square distance matrix:

$$\begin{aligned} \text{Find } & P \in \mathbb{R}^{n \times q}, \\ \text{s.t. } & (PP^{\top})_{ii} + (PP^{\top})_{jj} - 2(PP^{\top})_{ij} = D_{ij}, \\ & P^{\top} \mathbf{1} = 0, \end{aligned} \tag{2}$$

where q denotes the estimate for the rank of the underlying points. The centering constraint can be removed by reparameterization. Defining $\mathcal{A} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{|\Omega|}$ as

$$\mathcal{A}(X)_{\alpha} = X_{ii} + X_{jj} - 2X_{ij}, \quad \alpha = (i, j) \in \Omega,$$

the augmented Lagrangian is

$$L(Q; \Lambda) = \frac{1}{2} \left\| \mathcal{A}(QQ^{\top}) - b + \Lambda \right\|_2^2, \tag{3}$$

with b is the vector of observed distances and $\Lambda \in \mathbb{R}^{|\Omega|}$. The algorithm starts with an initial estimate of Q (randomly initialized in our experiments) and then alternates between two steps: updating Q via Barzilai–Borwein descent, and updating the Lagrange multipliers. For the dataset in consideration, we set the rank estimate to be 200.

Table 1 reports the results averaged over 10 random trials. We observe that for fixed sample budget, the Nyström method consistently yields better accuracy compared to matrix completion.

While both Nyström and typical matrix completion algorithms scale similarly up to logarithmic factors, the Nyström method is a single-step procedure, whereas the cost of the matrix completion algorithm depends on the target accuracy, with the number of iterations required to reach a given error often dominating the runtime. In our experiments, we run W_2 -MC for a sufficient number of iterations, as we consistently observe stable convergence across trials. In practice, choosing a

Rate / # of cols.	W ₂ -Nyström	W ₂ -MC
25% / 268	1.58×10^{-2} $\pm 1.96 \times 10^{-3}$	5.20×10^{-2} $\pm 1.91 \times 10^{-2}$
20% / 211	1.85×10^{-2} $\pm 2.2 \times 10^{-3}$	3.42×10^{-2} $\pm 8.26 \times 10^{-5}$
10% / 103	2.85×10^{-2} $\pm 2.60 \times 10^{-3}$	5.65×10^{-2} $\pm 6.01 \times 10^{-3}$
5% / 51	4.82×10^{-2} $\pm 8.59 \times 10^{-3}$	1.01×10^{-1} $\pm 2.50 \times 10^{-2}$
3% / 30	7.21×10^{-2} $\pm 1.20 \times 10^{-2}$	1.32×10^{-1} $\pm 9.77 \times 10^{-3}$

Table 1: Relative error (mean \pm standard deviation) of Nyström and matrix completion under different sampling rates of a 2000×2000 OrganCMNIST W₂ distance matrix. For the Nyström method, the numbers of columns associated with the sample rates are also reported.

suitable stopping criterion for matrix completion problems could also be challenging. Therefore, this fixed sample budget experiment leads us to conclude that computing full columns of W₂ distance measurements may be preferred for approximating Wasserstein distance matrix. Studying how far this generalizes is the subject of future work.

4.2 Classification stability

We next consider the stability of classification performance under the Nyström approximation as the number of sampled columns varies. Specifically, we compute 5%, 10%, 20%, ..., 100% of the columns of the 2000×2000 OrganCMNIST distance matrix using W₂-Nyström. The resulting approximation D_{est} is embedded via MDS, followed by a 90 – 10 train-test split. For the MDS embedding, we set the dimension to 120, chosen such that the retained singular values account for 97% of the total sum of singular values. Classification is then performed on the embeddings using linear discriminant analysis (LDA), 1-nearest neighbor (KNN1), kernel SVM with Gaussian kernel (SVM_RBF), and random forest. We note that the dataset has 11 underlying classes, and we evaluate the quality of the classification using classification accuracy, defined as the proportion of correctly identified data points. Figure 1 shows classification accuracy as a function of the number of columns sampled.

The main conclusion we draw from Figure 1 is that the classification performance is quite stable after computing about 10–20% of the columns of the distance matrix. This result and Theorem 3.1 suggest that the Nyström Wassmap algorithm (MDS on the output of Algorithm 2) is stable after sampling $\mathcal{O}((d+2)\log(d+2))$ columns of D .

5 Conclusion

We showed that the Nyström method yields a good approximation of W₂ distance matrices, leading to stable MDS embeddings. On the OrganCMNIST dataset, we showed that classification accuracy is stable even when only sampling a small fraction of columns, and for a fixed sampling budget, the Nyström method outperforms a distance matrix completion approach. Unlike classical Euclidean distances, where the tight link between Gram and distance matrices grounds the theory of multidimensional scaling, the structural properties of W₂ distance matrices are far less understood. This

t

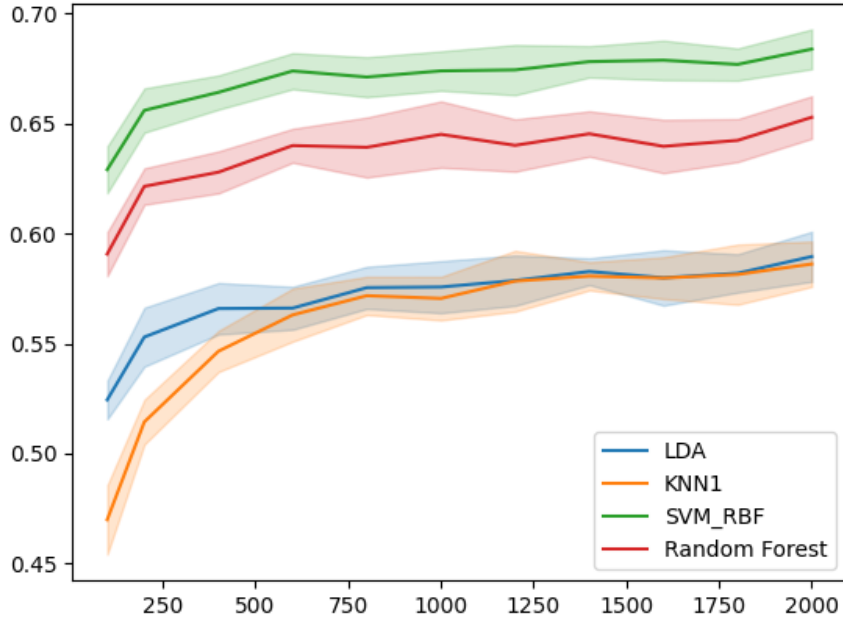


Figure 1: Classification accuracy (mean \pm 1 standard deviation of 100 trials) vs. number of columns computed in Algorithm 2 for 2000 OrganCMNIST data points.

gap makes it especially important to develop principled methods that reduce the cost of computing W_2 distances. In future work, we will explore theoretical conditions under which W_2 distance matrices admit approximate low-rank structure or fast spectral decay, as well as explore hybrid strategies that combine Nyström-selected columns with partial entry sampling similar following ideas similar to [22].

Acknowledgements

Abiy Tasissa acknowledges partial support from the National Science Foundation through grant DMS-2208392. HanQin Cai acknowledges partial support from the National Science Foundation through grant DMS-2304489.

Muhammad Rana and Keaton Hamm were partially supported by a Research Enhancement Program grant from the College of Science at the University of Texas at Arlington and by the Army Research Office under Grant Number W911NF-23-1-0213. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

- [1] Nanning Zheng and Jianru Xue, *Statistical learning and pattern analysis for image and video processing*, Springer Science & Business Media, 2009.

- [2] Marina Meilă and Hanyu Zhang, “Manifold learning: What, how, and why,” *Annual Review of Statistics and Its Application*, vol. 11, no. 1, pp. 393–417, 2024.
- [3] James Mathews, Maryam Pouryahya, Caroline Moosmüller, Ioannis G. Kevrekidis, Joseph O. Deasy, and Allen Tannenbaum, “Molecular phenotyping using networks, diffusion, and topology: soft-tissue sarcoma,” *Scientific Reports*, vol. 9, 2019, Article number: 13982.
- [4] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger, “From word embeddings to document distances,” in *International conference on machine learning*. PMLR, 2015, pp. 957–966.
- [5] Wei Wang, John A Ozolek, Dejan Slepčev, Ann B Lee, Cheng Chen, and Gustavo K Rohde, “An optimal transportation approach for nuclear structure-based pathology,” *IEEE transactions on medical imaging*, vol. 30, no. 3, pp. 621–631, 2010.
- [6] Gabriel Peyré and Marco Cuturi, “Computational optimal transport: with applications to data science,” *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.
- [7] Keaton Hamm, Nick Henscheid, and Shujie Kang, “Wassmap: Wasserstein isometric mapping for image manifold learning,” *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 2, pp. 475–501, 2023.
- [8] Andy Ramlatchan, Mengyun Yang, Quan Liu, Min Li, Jianxin Wang, and Yaohang Li, “A survey of matrix completion methods for recommendation systems,” *Big Data Mining and Analytics*, vol. 1, no. 4, pp. 308–323, 2018.
- [9] Raghu Meka, Prateek Jain, Constantine Caramanis, and Inderjit S Dhillon, “Rank minimization via online learning,” in *Proceedings of the 25th International Conference on Machine learning*, 2008, pp. 656–663.
- [10] Emmanuel J Candès, Justin Romberg, and Terence Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [11] Yuejie Chi, Yue M Lu, and Yuxin Chen, “Nonconvex optimization meets low-rank matrix factorization: An overview,” *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5239–5269, 2019.
- [12] Leo Liberti, Carlile Lavor, Nelson Maculan, and Antonio Mucherino, “Euclidean distance geometry and applications,” *SIAM review*, vol. 56, no. 1, pp. 3–69, 2014.
- [13] Nilson JM Moreira, Leonardo T Duarte, Carlile Lavor, and Cristiano Torezzan, “A novel low-rank matrix completion approach to estimate missing entries in euclidean distance matrix,” *Computational and Applied Mathematics*, vol. 37, no. 4, pp. 4989–4999, 2018.
- [14] Abiy Tasissa and Rongjie Lai, “Exact reconstruction of euclidean distance geometry problem using low-rank matrix completion,” *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 3124–3144, 2018.
- [15] Christopher Williams and Matthias Seeger, “Using the nyström method to speed up kernel machines,” *Advances in neural information processing systems*, vol. 13, 2000.

- [16] Varun Khurana, Harish Kannan, Alexander Cloninger, and Caroline Moosmüller, “Supervised learning of sheared distributions using linearized optimal transport,” *Sampling Theory, Signal Processing, and Data Analysis*, vol. 21, no. 1, 2023.
- [17] Alexander Cloninger, Keaton Hamm, Varun Khurana, and Caroline Moosmüller, “Linearized wasserstein dimensionality reduction with approximation guarantees,” *Applied and Computational Harmonic Analysis*, vol. 74, pp. 101718, 2025.
- [18] Ameet Talwalkar and Afshin Rostamizadeh, “Matrix coherence and the nystrom method,” *arXiv preprint arXiv:1408.2044*, 2014.
- [19] Keaton Hamm and Longxiu Huang, “Perturbations of cur decompositions,” *SIAM Journal on Matrix Analysis and Applications*, vol. 42, no. 1, pp. 351–375, 2021.
- [20] Joel A Tropp, “Improved analysis of the subsampled randomized hadamard transform,” *Advances in Adaptive Data Analysis*, vol. 3, no. 01n02, pp. 115–126, 2011.
- [21] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni, “Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific Data*, vol. 10, no. 1, pp. 41, 2023.
- [22] HanQin Cai, Longxiu Huang, Pengyu Li, and Deanna Needell, “Matrix completion with cross-concentrated sampling: Bridging uniform sampling and cur sampling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 8, pp. 10100–10113, 2023.