HAL open science

Matrix Decomposition on Graphs: A Functional View

Abhishek Sharma, Maks Ovsjanikov

To cite this version:

Abhishek Sharma, Maks Ovsjanikov. Matrix Decomposition on Graphs: A Functional View. 2021. hal-02871840v3

HAL Id: hal-02871840

https://hal.science/hal-02871840v3

Preprint submitted on 5 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Matrix Decomposition on Graphs: A Functional View

Abhishek Sharma [1] **Maks Ovsjanikov** [1]

Abstract

We propose a functional view of matrix decomposition problems on graphs such as geometric matrix completion and graph regularized dimensionality reduction. Our unifying framework is based on the key idea that using a reduced basis to represent functions on the product space is sufficient to recover a low rank matrix approximation even from a sparse signal. We validate our framework on several real and synthetic benchmarks (for both problems) where it either outperforms state of the art or achieves competitive results at a fraction of the computational effort of prior work.

1. Introduction

The assumption that high-dimensional data samples lie on or close to a smooth low-dimensional manifold is exploited as a regularizer or prior in many machine learning algorithms. Often, the low-dimensional manifold information is exploited via a graph structure between the data samples. As a result, graphs are often used as regularizers in various machine learning problems such as dimensionality reduction (Jiang et al., 2013), hashing (Liu et al., 2011) or matrix completion (Kalofolias et al., 2014) to name a few. In this article, we focus on dimensionality reduction and matrix completion and propose a principled framework that gives a unified solution to both problems by modelling the extra geometric information available in terms of graphs.

Dimensionality reduction: Given a data matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ with n m-dimensional data vectors, most prior work related to PCA (Abdi & Williams, 2010) can be broadly categorised in two themes: 1) matrix factorization approach of the classical PCA and its variants 2) matrix subtraction approach of robust PCA (Candès et al., 2011) and its variants. The former learns the projection $Q \in \mathbb{R}^{d\times n}$ of M in a d-dimensional linear space characterized by an orthonormal basis $U \in \mathbb{R}^{m\times d}$. Several followup works (Jiang et al., 2013; Zhang & Zhao, 2013; Jin et al., 2015; Tao et al., 2014) have shown that the clustering quality of PCA can be significantly improved by incorporating the data manifold information in the form of some underlying graph structure.

Instead of relying on matrix factorization, the second line of work directly estimates clean low rank approximation $\boldsymbol{X}$ of data matrix $\boldsymbol{M}$ by separating noise with a matrix additive model. Along these line, Fast Robust PCA on graphs (FRPCAG (Shahid et al., 2016)) introduces a joint notion of reduced rank for the rows and columns of a data matrix and proposes to jointly minimize the Dirichlet energy on the row and column graphs:

$$\min_{\boldsymbol{X}} \|\boldsymbol{M}-\boldsymbol{X}\|_1 + \gamma_1 \operatorname{tr}(\boldsymbol{X}\boldsymbol{L}_1\boldsymbol{X}^\top) + \gamma_2 \operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{L}_2\boldsymbol{X}). \tag{1}$$

Here $\boldsymbol{L}_1, \boldsymbol{L}_2$ are Laplacian matrices of graphs built, respectively, from the rows and columns of the data matrix $\boldsymbol{M}$. Conceptually, minimizing the Dirichlet energy, $\operatorname{tr}(\boldsymbol{X}\boldsymbol{L}_1\boldsymbol{X}^\top)$, promotes smoothness of $\boldsymbol{X}$ by penalizing high frequency components of the signals on corresponding graphs. The authors of FRPCAG (Shahid et al., 2016) demonstrate theoretically that under certain assumptions this minimization is connected with the spectrum of the underlying low rank matrix $\boldsymbol{X}$. Building on this idea, we instead directly constrain the low rank approximation by decomposing it using the first few eigenvectors of row and column graph Laplacians $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ and optimizing over the coupling matrix $\boldsymbol{C}$ only. Our approach, similar in spirit to the matrix factorization approach (Boyarski et al., 2020; Arora et al., 2019), leads to explicit control over the resulting rank of the matrix, and as we demonstrate below, superior performance and significantly simpler optimization problems.

Matrix completion deals with the recovery of missing values of a matrix of which we have only measured a subset of the entries. In general, without any constraints, this problem is ill-posed. However if the rank of underlying matrix is small, it is common to find the lowest rank matrix that agrees with known measurements (Candès & Recht, 2009). Under this low rank assumption, the problem is very similar to dimensionality reduction and can be rewritten as,

$$\min_{\boldsymbol{X}} \operatorname{rank}(\boldsymbol{X}) + \frac{\mu}{2}\|(\boldsymbol{X}-\boldsymbol{M})\odot\boldsymbol{S}\|_F^2. \tag{2}$$

Here $\boldsymbol{X}$ stands for the unknown matrix, $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ for

[1]LIX, Ecole Polytechnique, IPP Paris. Correspondence to: sharma@lix.polytechnique.fr <>.

Under Review

the ground truth matrix, $\boldsymbol{S}$ is a binary mask representing the input support, and $\odot$ denotes the Hadamard product. Various problems in collaborative filtering can be posed as a *matrix completion* problem (Kalofolias et al., 2014; Rao et al., 2015), where for example the columns and rows represent users and items, respectively, and matrix values represent a score determining the preference of user for that item. Often, additional structural information is available in the form of column and row graphs representing similarity of users and items, respectively. Most of the prior work that incorporates geometric structure into matrix completion problems is either based on highly engineered frameworks, e.g., (Monti et al., 2017) or non convex formulation with several hyperparameters (Boyarski et al., 2020) thereby making the overall optimization harder to optimize. Instead, our simple formulation based on the functional map representation (Ovsjanikov et al., 2012), with a single regularizer, mitigates the problems associated with (Boyarski et al., 2020).

Contributions. Our contributions are threefold. First, we propose a novel unified view of geometric matrix completion and graph regularized dimensionality reduction that is convex and smooth. Second, conceptually, our matrix decomposition formulation explicitly imposes and optimizes for a low rank approximation and, as we demonstrate below, is empirically more accurate in recovering a low rank matrix approximation than competitive baselines. Third, we propose a novel regularization inspired from the functional map literature that is shown to be competitive with a combination of several regularizers on various real world datasets.

2. Related work

Matrix completion and graph regularized PCA have been studied with many viewpoints and thus, exhaustive coverage of prior work is beyond the scope of this paper. In this section, we first briefly cover related work and then describe prior work that is directly related to our work. We refer to (Shahid et al., 2016) for more details on PCA and related formulations.

Geometric matrix completion. A prominent relaxation of the rank operator in Eq. (2) is to constrain the space of solutions to be smooth w.r.t. some geometric structure of the matrix rows and columns. There exist several prior works on geometric matrix completion that exploit geometric information (Berg et al., 2017; Kalofolias et al., 2014; Rao et al., 2015) such as graphs encoding relations between rows and columns. More recent works leverage deep learning on geometric domains (Berg et al., 2017; Monti et al., 2017) to extract relevant information from geometric data such as graphs. While these techniques achieve state-of-the-art results, their design is highly engineered and thus, non-intuitive and often lacks a proper theoretical foundation.

Graph Regularized Dimensionality Reduction. Jiang et. al. proposed Graph Laplacian PCA (GLPCA) (Jiang et al., 2013) which imposes the graph regularization of principal components using the Dirichlet term for clustering in the low dimensional space. Similarly, the models proposed in (Jiang et al., 2013; Zhang & Zhao, 2013; Jin et al., 2015; Tao et al., 2014) leverage the graph structure to learn enhanced class structures. All of these methods still suffer from non-convexity (Jiang et al., 2013; Jin et al., 2015; Tao et al., 2014). RPCAG (Shahid et al., 2015) is convex but uses the nuclear norm relaxation that involves an expensive SVD step inhibiting its scalabilty to large datasets.

The idea of using two graph regularization terms has also been applied in co-clustering (Gu & Zhou, 2009), Non negative matrix factorization (Shang et al., 2012) and more recently in the context of low-rank representation (Yin et al., 2015). The co-clustering & NMF based models which use such a scheme (Gu & Zhou, 2009), (Shang et al., 2012) suffer from non-convexity and the works of (Yin et al., 2015) use a nuclear-norm formulation making it difficult to scale. Note that there also exist methods that learn a union of low dimensional subspaces where each class belongs to a different subspace (Elhamifar & Vidal, 2013; Vidal & Favaro, 2014) but they are not directly related to our approach.

Low Rank Estimators. In classical matrix completion or estimation literature, there is large body of work that assumes the underlying signal matrix $\boldsymbol{M}$ to be low rank and then tries to estimate it using truncated SVD methods (Koltchinskii et al., 2011; Klopp, 2011; Donoho & Gavish, 2013; Chatterjee, 2012; Klopp, 2015) as it is the best approximation of a given rank r to the data in the least squares sense. Most of these work estimate this unknown rank and provide bounds on optimality of hard thresholded SVD in an asymptotic framework. Our method is not directly related to these work and we explain it in more detail in the methodology section 4.

Functional Maps. Our work is mainly inspired from the functional map framework (Ovsjanikov et al., 2012), which is used ubiquitously in non-rigid shape correspondence, and has been extended to handle challenging partial matching cases, e.g. (Litany et al., 2017). This framework has recently been adapted for geometric matrix completion in (Boyarski et al., 2020), where the authors propose to build a functional map between graphs of rows and columns. However, they 1) impose several non convex regularization terms each with a scaling hyperparameter and some even with different initialization 2) explore a huge range of hyperparameter space. Moreover, their framework is tailored towards geometric matrix completion and does not target separability of data in some lower dimensional space.

3. Preliminaries

In this section, we cover some preliminaries about product graphs and functional maps.

Product graphs Let $\mathcal{G} = (V, E, W)$ be a (weighted) graph with its vertex set V and edge set E and adjacency matrix denoted by W. Graph Laplacian $\boldsymbol{L}$ is given by $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, where $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{W})$ is the *degree matrix*. $\boldsymbol{L}$ is symmetric and positive semi-definite and therefore admits a spectral decomposition $\boldsymbol{L} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$. Let $\mathcal{G}_1 = (V_1, E_1, W_1)$, $\mathcal{G}_2 = (V_2, E_2, W_2)$ be two graphs, with $\boldsymbol{L}_1 = \boldsymbol{\Phi}\boldsymbol{\Lambda}_1\boldsymbol{\Phi}^\top$, $\boldsymbol{L}_2 = \boldsymbol{\Psi}\boldsymbol{\Lambda}_2\boldsymbol{\Psi}^\top$ being their corresponding graph Laplacians. We define the Cartesian product of $\mathcal{G}_1$ and $\mathcal{G}_2$, denoted by $\mathcal{G}_1 \square \mathcal{G}_2$, as the graph with vertex set $V_1 \times V_2$.

Functional maps. Let $\boldsymbol{X}$ be a function defined on $\mathcal{G}_1 \square \mathcal{G}_2$. It can be encoded as a matrix of size or $|V_1| \times |V_2|$. Then it can be represented using the bases $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ of the individual graph Laplacians, $\boldsymbol{C} = \boldsymbol{\Phi}^\top \boldsymbol{X} \boldsymbol{\Psi}$. In the shape processing community, such $\boldsymbol{C}$ is called a *functional map*, as it it used to map between the functional spaces of $\mathcal{G}_1$ and $\mathcal{G}_2$. One of the advantages of working with the functional map representation $\boldsymbol{C}$ rather than the matrix $\boldsymbol{X}$ is that its size is typically much smaller, and is only controlled by the size of the basis, independent of the number of nodes in G_1 and G_2, resulting in simpler optimization problems. Moreover, the projection onto a basis also provides a strong regularization, which can itself be beneficial for both shape matching, and, as we show below, matrix completion.

4. Low Rank Matrix Decomposition on Graphs

We assume that we are given a set of samples in some matrix $\boldsymbol{M} \in \mathbb{R}^{m \times n}$. In addition, we construct two graphs $\mathcal{G}_\mathrm{r}, \mathcal{G}_\mathrm{c}$, encoding relations between the rows and the columns, respectively. We represent the Laplacians of these graphs and their spectral decompositions by $\boldsymbol{L}_\mathrm{r} = \boldsymbol{\Phi}\boldsymbol{\Lambda}_\mathrm{r}\boldsymbol{\Phi}^\top$, $\boldsymbol{L}_\mathrm{c} = \boldsymbol{\Psi}\boldsymbol{\Lambda}_\mathrm{c}\boldsymbol{\Psi}^\top$. For matrix completion problem, the matrix $\boldsymbol{M}$ is not completely known so we are also given a binary indicator mask $\boldsymbol{S}$ that indicates 1 for measured samples and 0 for missing ones. We minimize the objective function of the following form:

$$\min_{\boldsymbol{X}} \; E_\mathrm{data}(\boldsymbol{X}) + \mu E_\mathrm{reg}(\boldsymbol{X}) \tag{3}$$

with E_data denoting a data term of the form

$$E_\mathrm{data}(\boldsymbol{X}) = \|(\boldsymbol{X} - \boldsymbol{M}) \odot \boldsymbol{S}\|_F^2 \,, \tag{4}$$

As observed in (Boyarski et al., 2020), we can decompose $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$. Remarkably, the data term itself, as we show in our experiments, when expressed through the functional map i.e.$\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ already recovers low-rank matrices and outperforms the recent approach of (Boyarski et al., 2020) on synthetic geometric experiments for matrix completion and obtains competitive results on dimensionality reduction tasks. Before we explain the choice and motivation of our regularizer E_reg, we explain next why the data term itself already works remarkably well on rank constrained geometric problems.

4.1. Motivation and Analysis

Our first observation is that by using a reduced basis to represent a function $\boldsymbol{X}$ on the product space $\mathcal{G}_1 \square \mathcal{G}_2$ already provides a strong regularization, which can be sufficient to recover a low rank matrix approximation even from a sparse signal.

Specifically, suppose that we constrain $\boldsymbol{X}$ to be a matrix such that $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ for some matrix $\boldsymbol{C}$. Note that if $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ have k columns each then $\boldsymbol{C}$ must be a $k \times k$ matrix. We would like to argue that solving Eq. (4) under the constraint that $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ will recover the underlying ground truth signal if it is low rank and satisfies an additional condition that we call basis consistency.

For this suppose that the ground truth hidden signal $\boldsymbol{M}$ has rank r. Consider its singular value decomposition $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$. As $\boldsymbol{M}$ has rank r, $\boldsymbol{\Sigma}$ is a diagonal matrix with r non-zero entries. We will call $\boldsymbol{M}$ *basis-consistent* with respect to $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ if the first r left singular vectors U_r (i.e., those corresponding to non-zero singular values) lie in the span of $\boldsymbol{\Phi}$, and the first r right singular vectors V_r lie in the span of $\boldsymbol{\Psi}$. In other words, there exist some matrices $\boldsymbol{R}, \boldsymbol{Q}$ s.t. $U_r = \boldsymbol{\Phi}\boldsymbol{R}$ (note that this implies $k \geq r$) and $V_r = \boldsymbol{\Psi}\boldsymbol{Q}$.

Given this definition, it is easy to see that all basis-consistent matrices with rank $r \leq k$ can be represented by some functional map $\boldsymbol{C}$. In other words, given $\boldsymbol{Y}$ that is basis-consistent, there is some functional map $\boldsymbol{C}$ s.t. $Y = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^T$. Conversely any $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^T$ has rank at most k and must be basis-consistent by construction.

Second, suppose we are optimizing Eq (4) under the constraint $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ and that the optimum, i.e., the ground truth matrix $\boldsymbol{M}$, is basis-consistent. Then since the energy $E_\mathrm{data}(C)$ is convex, given k^2 known samples to fully constrain the corresponding linear system, we are guaranteed to recover the optimum low-rank basis-consistent matrix. We note briefly that the argument above can also be made approximate, when the ground truth matrix is not exactly, but only approximately basis consistent, by putting appropriate error bounds.

This simple observation suggests that by restricting $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$ and optimizing over the matrices $\boldsymbol{C}$ instead of $\boldsymbol{X}$ already provides a strong regularization that can help recover appropriate low-rank signals even without any other

regularization. Further, it avoids solving complex optimization problems involving iterative SVD, since $\boldsymbol{C}$ becomes the only free variable, which can be optimized directly. For problems such as geometric matrix completion, we observe that a weak additional regularization is often sufficient to obtain state-of-the-art results.

More formally, we state our result as follows

Proposition 1 *we recover an optimal low rank matrix with high probability as long as the underlying latent matrix $\boldsymbol{X}$ is basis consistent.*

Proof: The proof is based on the main result (Theorem 1 in (Candès & Recht, 2009)) in low rank exact matrix recovery method. (Candès & Recht, 2009) prove that there is a unique rank k matrix that agrees with the sampled values with high probability and thus, recovers this underlying hidden signal matrix. Our method also recovers a rank k matrix by construction. Since our problem is convex, our method will recover the best rank k matrix that is within the span of the eigenfunctions. If the underlying matrix is basis consistent, then our method will recover the same exact matrix as a low rank exact recovery method (by definition of basis consistency).

In Proposition 1, we assume basis consistency over $\boldsymbol{X}$. Note that if we could assume basis consistency over data matrix $\boldsymbol{M}$, our results can then be derived from truncated-SVD methods for low rank matrix completion from noisy observations (Klopp, 2011; Chatterjee, 2012; Donoho & Gavish, 2013). The link to our approach then comes simply from the Eckart-Young theorem on the optimality of SVD for low-rank recovery.

Link to Dimensionality Reduction. For dimensionality reduction, we optimize the data term alone i.e. $E_{\text{data}}(\boldsymbol{X}) = \|(\boldsymbol{X} - \boldsymbol{M})\|_F^2$. The resulting low rank matrix is then considered a new representation of original data matrix $\boldsymbol{M}$ and later used for clustering and classification.

Differences from FRPCAG (Shahid et al., 2016) We do not target the Robust PCA problem (Candès et al., 2011) as done in FRPCAG. FRPCAG obtains a low rank approximation by minimizing Dirichlet energy on the two graphs and thereby, implicitly obtains a low rank approximation. In contrast, we explicitly factorize the data matrix. As shown in our experiments below, this explicit control over the resulting low rank of matrix, by optimizing over $\boldsymbol{C}$, yields superior clustering results over FRPCAG.

4.2. Functional Regularization

Our E_{reg} contains a single regularization term on the functional map induced between row space and column space described next.

Laplacian Commutativity as a Regularizer We propose to use the simplest possible regularizer, which furthermore leads to a convex optimization problem and can achieve state-of-the-art results. For this we borrow a condition that is prominent in the functional map literature (Ovsjanikov et al., 2016). Namely, in the context of surfaces, the functional map is often expected to *commute with the Laplace-Beltrami operator*:

$$E_{\text{reg}} = \left\|\boldsymbol{C}\boldsymbol{\Lambda}_{\text{r}} - \boldsymbol{\Lambda}_{\text{c}}\boldsymbol{C}\right\|^2, \tag{5}$$

where $\boldsymbol{\Lambda}_{\text{r}}$ and $\boldsymbol{\Lambda}_{\text{c}}$ are diagonal matrices of Laplacian eigenvalues of the source graph (row graph) and target graph (column graph).

For shape matching problems, this constraint helps to find better mappings because functional maps that commute with the Laplacian must arise from near isometric point-to-point correspondences (Rosenberg, 1997; Ovsjanikov et al., 2012). More broadly, commutativity with the Laplacian imposes a diagonal structure of the functional map, which intuitively promotes preservation of low frequency eigenfunctions used in the basis. In the context of matrix completion this can be interpreted simply as approximate preservation of global low frequency signals defined on the two graphs.

Given these above definitions, our problem defined in Eq. (3) reduces to

$$\min_{\boldsymbol{C}} \|(\boldsymbol{X} - \boldsymbol{M}) \odot \boldsymbol{S}\|_F^2 + \mu * \left\|\boldsymbol{C}\boldsymbol{\Lambda}_{\text{r}} - \boldsymbol{\Lambda}_{\text{c}}\boldsymbol{C}\right\|^2 \quad \text{where } \boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top \tag{6}$$

As noted in several works, isometry between two spaces is a key to functional map representation. Assuming isometry between real world graphs is however over optimistic. Thus, one way to work under relaxed isometry condition is to instead align the eigen basis with additional transformation matrix to achieve diagonal functional map matrix (Litany et al., 2017; Boyarski et al., 2020). In practice, we observe faster convergence if we replace $\boldsymbol{C}$ with $\boldsymbol{P}\boldsymbol{C}\boldsymbol{Q}^\top$, and let all three $\boldsymbol{P}, \boldsymbol{C}$ and $\boldsymbol{Q}$ be free variables.

Differences from SGMC (Boyarski et al., 2020) Even though both methods, ours and SGMC build on the functional map framework, there is a fundamental difference between the two. SGMC focus is on high complexity functional map based model (large values of $\boldsymbol{C}$, multiple resolutions of $\boldsymbol{C}, \boldsymbol{P}, \boldsymbol{Q}$) and thus, requires a variety of (non-convex) regularizers. In contrast, our core idea is on the low rank matrix recovery based on the functional map based decomposition alone $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{C}\boldsymbol{\Psi}^\top$(See 'Ours-FM' baseline in experiments Section 5.2).

To outline the differences more precisely, in addition to Dirichlet energy on the two graphs, (Boyarski et al., 2020) also introduces two regularization on the transformation matrix $\boldsymbol{P}, \boldsymbol{Q}$. Additionaly, (Boyarski et al., 2020) also uss

a multi-resolution spectral loss named SGMC-Zoomout (SGMC-Z) (Melzi et al., 2019) with its own hyperparameters (step size between different resolutions) besides several scalars to weigh different regularizations.

4.3. Implementation

The optimization is carried out using gradient descent in Tensorflow (Abadi et al., 2015).

Graphs Construction Following (Shahid et al., 2016), we use two types of graphs G_1 and G_2 in our proposed model. The graph G_1 is constructed between the data samples or the columns of the data matrix and the graph G_2 is constructed between the features or the rows of the data matrix. The graphs are undirected and built using a standard K-nearest neighbor strategy. We connect each x_i to its K nearest neighbors x_j where K is 10. This is followed by the graph weight matrix A computation as

$$A_{ij} = \begin{cases} \exp\left(-\frac{\|(x_i - x_j)\|_2^2}{\sigma^2}\right) & \text{if } x_j \text{ is connected to } x_i \\ 0 & \text{otherwise.} \end{cases}$$

Initialization Similar to (Boyarski et al., 2020), we initialize the $\boldsymbol{P}$ and $\boldsymbol{Q}$ with an identity matrix with size equal to that of underlying matrix $\boldsymbol{M}$ corresponding to respective dataset and $\boldsymbol{C}$ by projecting $\boldsymbol{X} \odot \boldsymbol{S}$ on the first eigen vector of $\boldsymbol{L}_{\mathrm{c}}$ and $\boldsymbol{L}_{\mathrm{r}}$.

Hyperparameters For all experiments, we set μ and learning rate to be .00001 for all the experiments. We report ize of $\boldsymbol{C}$ explicitly in each experiment below. For geometric matrix completion, we divide the number of available entries in the matrix randomly into training and validation set in a 95 to 5 ratio respectively. .

5. Results

This section is divided into subsections. The goal of Subsection 5.1 is to validate our dimension reduction framework for the task of clustering and classification. In Subsection 5.2, we evaluate our model performance for matrix completion problem on both synthetic and real world datasets.

5.1. Graph Regularized Dimensionality Reduction

5.1.1. Datasets

We use 4 well-known benchmarks and perform our clustering experiments on following databases: ORL, BCI, COIL20, and MFEAT. ORL[1] is a face database with small pose variations. COIL20 [2] is a dataset of objects with significant pose changes. MFeat[3] consists of features extracted from handwritten numerals whereas BCI database comprises of features extracted from a Brain Computer Interface setup [4].

5.1.2. Baselines

We compare the clustering performance of our model with 5 other dimensionality reduction models. Apart from classical PCA, the rest of the models exploit graph information.

Models using graph structure: We compare 1) Graph Laplacian PCA (GLPCA)(Jiang et al., 2013) 2) Laplacian Eigenmaps (LE) 3) Robust PCA on graphs RPCAG (Shahid et al., 2015) 4) Fast Robust PCA on graphs FRPCAG (Shahid et al., 2016) 5) Our proposed model. Note that RPCAG and FRPCAG are closest to our approach and known to outperform other graph regularized models such as Manifold Regularized Matrix Factorization (MMF) (Zhang & Zhao, 2013), Non-negative Matrix Factorization (NMF)(Lee & Seung, 1999), Graph Regularized Non-negative Matrix Factorization (GNMF) (Cai et al., 2011). We obtain FRPCAG and RPCAG results by running the open source implementation provided by the authors on the four datasets. Note that we run the clustering on all the samples of COIL20 and all the features of MFEAT whereas FRPCAG only use a subset of them in their paper. FRPCAG contains two hyperparameters, namely weighing scalars for Dirichlet energy. For these scalars, we pick the best value from the set (1,10,50,100) based on empirical performance. For PCA, we chose the first 40 principal components from a set (30, 40, 50). For our method, the only hyper-parameter is the dimensionality of matrix $\boldsymbol{C}$. We pick the best value out of (50, 100).We pre-process the datasets to zero mean and unit standard deviation along the features.

5.1.3. Clustering Metric

We follow the standard evaluation protocol and use clustering purity to evaluate our method. To compute purity, each cluster is assigned to the class which is most frequent in the cluster, and then the accuracy of this assignment is measured by counting the number of correctly assigned and dividing by the total no. of samples. We report the maximum clustering error from 10 runs of k-means and summarize our findings in Table 1.

As shown in Table 1, our model obtains superior or competitive performance over all other baselines.

5.1.4. Classification

We further evaluate our framework on the classification task on the same 4 datasets. We perform classification with PCA,

[1] cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[2] cs.columbia.edu/CAVE/software/softlib/coil-20.php
[3] archive.ics.uci.edu/ml/datasets/Multiple+Features
[4] olivier.chapelle.cc/ssl-book/benchmarks.html

Table 1. Clustering purity on Benchmark Datasets.

Dataset	Samples	PCA	LE	GLPCA	GRPCA	FGRPCA	Ours
ORL	400	57	56	68	74	77	79
COIL20	1404	67	56	66	65	68	71
MFEAT	400	82	90	71	80	85	90
BCI	400	52	52	52	53	52	53

Table 2. Classification accuracy on Benchmark Datasets.

Dataset	PCA	LE	FGRPCA	Ours
ORL	63	56	66	68
COIL20	88	78	88	89
MFEAT	97	94	97	97
BCI	52	48	53	55

LE and our data representations using a KNN classifier. We randomly select 30% of labeled data, and use the rest to evaluate. We repeat this 5 times and summarize the average classification accuracy in Table 2. Our method obtains competitive accuracy compared to other baselines. PCA representation with first 40 components already provides very competitive classification results on several datasets.

5.2. Geometric Matrix Completion experiments

This section is divided into two subsections. The goal of first subsection 5.2.1 is to extensively compare between our approach and Spectral Geometric Matrix Completion **(SGMC)**(Boyarski et al., 2020) on a synthetic example of a community structured graphs. In the second subsection 5.2.2, we compare with all approaches on various real world recommendation benchmarks. Note that we use SGMC and (Boyarski et al., 2020) interchangeably in this section.

5.2.1. Experiments on synthetic datasets

For a fair comparison with (Boyarski et al., 2020), we use graphs taken from the synthetic Netflix dataset. Synthetic Netflix is a small synthetic dataset constructed by (Kalofolias et al., 2014) and (Monti et al., 2017), in which the user and item graphs have strong community structure. Similar to (Boyarski et al., 2020), we use a randomly generated low rank matrix on the product graph $\mathcal{G}_{\mathrm{c}} \square \mathcal{G}_{\mathrm{r}}$ to test the matrix completion accuracy. Synthetic Netflix is useful in conducting controlled experiments to understand the behavior of geometry-exploiting algorithms. We consider the following two baselines:

Ours-FM: This baseline only optimizes for $\boldsymbol{C}$ without any regularization. All results are obtained with $\boldsymbol{C}$ of size 30×30. This value was chosen after a cross validation from a set of $20, 30, 40$.

Table 3. Comparative results to test the dependence on the density of the sampling set for a random rank 10 matrix of size 150×200.

Density in %	Ours	Ours-FM	SGMC
1	2e-2	2e-2	1e-1
5	8e-7	1e-3	5e-4
10	2e-7	5e-5	2e-4
20	1e-7	2e-5	1e-4

Table 4. Comparative results to test the robustness in the presence of noisy graphs.

Noise	Ours	Ours-FM	SGMC
5	1e-3	2e-3	5e-3
10	4e-3	3e-3	1e-2
20	6e-3	6e-3	1e-2

SGMC: All results are obtained with their open source code with their optimal parameters.

Test Error. To evaluate the performance of the algorithms in this section, we report the *root mean squared error*,

$$\mathrm{RMSE}(\boldsymbol{X}, \boldsymbol{S}) = \sqrt{\frac{\|(\boldsymbol{X} - \boldsymbol{M}) \odot \boldsymbol{S}\|_F^2}{\sum_{i,j} \boldsymbol{S}_{i,j}}} \tag{7}$$

computed on the complement of the training set. Here $\boldsymbol{X}$ is the recovered matrix and $\boldsymbol{S}$ is the binary mask representing the support of the set on which the RMSE is computed.

We compare the two approaches on different constraints ranging from rank of the underlying matrix to the sampling density. Note that optimality bounds for classical matrix completion algorithms also depend on constraints such as sampling density, noise variance etc.

Rank of the underlying matrix. We explore the effect of the rank of the underlying matrix, showing that as the rank increases up to 15, it becomes harder for both methods to recover the matrix. We remark that Ours-FM alone recovers the low rank very effectively. However, on real data, we find the additional regularizer in Ours to be more effective than Ours-FM. We also remark that Ours-FM consistently outperforms SGMC for all rank values. For the training set

Table 5. Comparative results to test the dependence on the rank of the underlying random matrix of size 150×200

Rank	Ours	Ours-FM	SGMC
5	1e-7	2e-5	1e-4
10	2e-7	2e-5	2e-4
12	5e-7	4e-5	9e-4
15	6e-3	1e-3	1e-2

Table 6. Test error on Flixster and Movielens-100K

Model	Flixster	ML-100K
MC (Candès & Recht, 2009)	1.533	0.973
GMC (Kalofolias et al., 2014)	–	0.996
GRALS (Rao et al., 2015)	1.245	0.945
RGCNN (Monti et al., 2017)	0.926	0.929
GC-MC (Berg et al., 2017)	0.917	0.910
Ours-FM	1.02	1.12
DMF(Arora et al., 2019)	1.06	0.922
SGMC	0.900	0.912
SGMC-Z	0.888	0.913
Ours	0.888	0.915

we used 10% of the points chosen at random (same training set for all experiments summarized in Table 5).

Sampling density. We investigate the effect of the number of samples on the reconstruction error. We demonstrate that in the data-poor regime, our regularization is strong enough to recover the matrix, compared to performance achieved by incorporating geometric regularization through SGMC. These experiments are summarized in Table 3. Note that gap between us and SGMC remains high even when the sample density increases to 20%. Even when using 1% of the samples, we perform better than SGMC.

Noisy graphs. We study the effect of noisy graphs on the performance. We follow the same experimental setup as (Boyarski et al., 2020) and perturb the edges of the graphs by adding random Gaussian noise with zero mean and tunable standard deviation to the adjacency matrix. We discard the edges that became negative as a result of the noise, and symmetrize the adjacency matrix. Table 4 demonstrates that our method is robust to substantial amounts of noise in graphs. Surprisingly, Ours-FM demonstrates even stronger resilience to noise.

Runtime Comparison. We also compare our method runtime with SGMC. Source code for both algorithm is Python and Tensorflow based. Our method runs atleast 20 times faster than SGMC when compared on synthetic experiments described above. This is not surprising as SGMC involves optimizing various regularizers and with high values of $\boldsymbol{P}, \boldsymbol{C}, \boldsymbol{Q}$.

5.2.2. Results on recommender systems datasets

In addition to synthetic Netflix, we also validate our method on two more recommender systems datasets for which row and column graphs are available. Movielens-100K (Harper & Konstan, 2016) contains ratings of 1682 items by 943 users whereas Flixter (Jamali & Ester, 2010) contains ratings of 3000 items by 3000 users. All baseline numbers, except Ours-FM, are taken from (Monti et al., 2017) and (Boyarski et al., 2020).

Baselines In addition to **SGMC** and **SGMC(Z)**, we also compare with **DMF**(Arora et al., 2019). This is a matrix factorization approach that was adapted for matrix completion tasks by (Boyarski et al., 2020). Note that this approach does not incorporate any geometric information. We explain some observations from Table 6: First, our baseline, Ours-FM, obtains surprisingly good performance across datasets. This underscores the regularization brought in by the Laplacian eigen-basis of row and column graphs. Second, non geometric model such as DMF shows competitive performance with all the other methods on ML-100K. This suggests that the geometric information is not very useful for this dataset. Third, our proposed algorithm is competitive with the other methods while being simple and interpretable. Lastly, these experimental results validate the effectiveness of our single regularization when compared to the combination of several non-convex regularizations introduced in (Boyarski et al., 2020).

6. Conclusion and Future Work

In this work, we provide a novel unified view of geometric matrix completion and graph regularized dimensionality reduction and establish empirically and theoretically that using a reduced basis to represent a function on the product space of two graphs already provides a strong regularization, that is sufficient to recover a low rank matrix approximation. Moreover, we propose a novel regularization and show, through extensive experimentation on real and synthetic datasets, that our single regularization is very competitive when compared to the combination of several different regularizations proposed before for geometric matrix completion problem.

Extracting geometric information from graph structured data is a core task in several domains from few shot learning (Gidaris & Komodakis, 2019), zero shot learning (Wang et al., 2018) in computer vision, machine learning to knowledge graph based problems in natural language processing since graphs appear everywhere. For future work, we plan to extend our framework to several such large scale problems and also test its robustness to noise and corruptions in input data.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

Abdi, H. and Williams, L. J. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2010.

Arora, S., Cohen, N., Hu, W., and Luo, Y. Implicit regularization in deep matrix factorization. In *Neurips*, 2019.

Berg, R. v. d., Kipf, T. N., and Welling, M. Graph convolutional matrix completion. *arXiv preprint arXiv:1706.02263*, 2017.

Boyarski, A., Vedula, S., and Bronstein, A. Deep matrix factorization with spectral geometric regularization. *arXiv preprint arXiv:1911.07255*, 2020.

Cai, D., He, X., Han, J., and Huang, T. Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence*, 33 8:1548–60, 2011.

Candès, E., Li, X., Ma, Y., and Wright, J. Robust principal component analysis? *J. ACM*, 58:11:1–11:37, 2011.

Candès, E. J. and Recht, B. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

Chatterjee, S. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43, 12 2012. doi: 10.1214/14-AOS1272.

Donoho, D. and Gavish, M. The optimal hard threshold for singular values is. *IEEE Transactions on Information Theory*, 60, 05 2013. doi: 10.1109/TIT.2014.2323359.

Elhamifar, E. and Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:2765–2781, 2013.

Gidaris, S. and Komodakis, N. Generating classification weights with gnn denoising autoencoders for few-shot learning. *CVPR*, 2019.

Gu, Q. and Zhou, J. Co-clustering on manifolds. In *KDD*, 2009.

Harper, F. M. and Konstan, J. A. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):19, 2016. URL https://grouplens.org/datasets/movielens/.

Jamali, M. and Ester, M. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pp. 135–142. ACM, 2010.

Jiang, B., Ding, C., Luo, B., and Tang, J. Graph-laplacian pca: Closed-form solution and robustness. In *CVPR*, pp. 3492–3498, 2013.

Jin, T., Yu, J., You, J., Zeng, K., Li, C., and Yu, Z. Low-rank matrix factorization with multiple hypergraph regularizer. *Pattern Recognition.*, 48:1011–1022, 2015.

Kalofolias, V., Bresson, X., Bronstein, M., and Vandergheynst, P. Matrix completion on graphs. *arXiv preprint arXiv:1408.1717*, 2014.

Klopp, O. Rank penalized estimators for high-dimensional matrices. *Electron. J. Stat.*, 5:1161–1183, 2011. ISSN 1935-7524. doi: 10.1214/11-EJS637. URL http://dx.doi.org/10.1214/11-EJS637.

Klopp, O. Matrix completion by singular value thresholding: sharp bounds. *Electron. J. Stat.*, 9(2):2348–2369, 2015. ISSN 1935-7524. doi: 10.1214/15-EJS1076. URL http://dx.doi.org/10.1214/15-EJS1076.

Koltchinskii, V., Lounici, K., and Tsybakov, A. B. Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Annals of Statistics*, 39(5):2302–2329, 10 2011. doi: 10.1214/11-AOS894. URL https://doi.org/10.1214/11-AOS894.

Lee, D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

Litany, O., Rodolà, E., Bronstein, A. M., and Bronstein, M. M. Fully spectral partial shape matching. In *Computer Graphics Forum*, volume 36, pp. 247–258. Wiley Online Library, 2017.

Liu, W., Wang, J., Kumar, S., and Chang, S. Hashing with graphs. In *ICML*, 2011.

Melzi, S., Ren, J., Rodola, E., Ovsjanikov, M., and Wonka, P. Zoomout: Spectral upsampling for efficient shape correspondence. In *Siggraph-Asia*, 2019.

Monti, F., Bronstein, M., and Bresson, X. Geometric matrix completion with recurrent multi-graph neural networks. In *Advances in Neural Information Processing Systems*, pp. 3697–3707, 2017.

Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., and Guibas, L. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31(4):30, 2012.

Ovsjanikov, M., Corman, E., Bronstein, M., Rodolà, E., Ben-Chen, M., Guibas, L., Chazal, F., and Bronstein, A. Computing and processing correspondences with functional maps. In *SIGGRAPH ASIA 2016 Courses*, pp. 9. ACM, 2016.

Rao, N., Yu, H.-F., Ravikumar, P. K., and Dhillon, I. S. Collaborative filtering with graph information: Consistency and scalable methods. In *Advances in neural information processing systems*, pp. 2107–2115, 2015.

Rosenberg, S. *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds*, volume 31. Cambridge University Press, 1997.

Shahid, N., Kalofolias, V., Bresson, X., Bronstein, M., and Vandergheynst, P. Robust principal component analysis on graphs. In *ICCV*, pp. 2812–2820, 2015.

Shahid, N., Perraudin, N., Kalofolias, V., Puy, G., and Vandergheynst, P. Fast robust pca on graphs. *IEEE Journal of Selected Topics in Signal Processing*, 10:740–756, 2016.

Shang, F., Jiao, L., and Wang, F. Graph dual regularization non-negative matrix factorization for co-clustering. *Pattern Recognition.*, 45:2237–2250, 2012.

Tao, L., Ip, H., Wang, Y., and Shu, X. Low rank approximation with sparse integration of multiple manifolds for data representation. *Applied Intelligence*, 42:430–446, 2014.

Vidal, R. and Favaro, P. Low rank subspace clustering (lrsc). *Pattern Recognit. Lett.*, 43:47–61, 2014.

Wang, X., Ye, Y., and Gupta, A. Zero-shot recognition via semantic embeddings and knowledge graphs. *CVPR*, 2018.

Yin, M., Gao, J., Lin, Z., Shi, Q., and Guo, Y. Dual graph regularized latent low-rank representation for subspace clustering. *IEEE Transactions on Image Processing*, 24: 4918–4933, 2015.

Zhang, Z. and Zhao, K. Low-rank matrix approximation with manifold regularization. *PAMI*, 35:1717–1729, 2013.