Recognizing Emotions from Abstract Paintings using Non-Linear Matrix Completion

Xavier Alameda-Pineda[1], Elisa Ricci[2,3], Yan Yan[1], Nicu Sebe[1]

[1]University of Trento, Via Sommarive 9, 38123 Trento, Italy
[2]Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy
[3]University of Perugia, Via Duranti 93, 06123, Perugia, Italy

{xavier.alamedapineda,yan.yan,niculae.sebe}@unitn.it, eliricci@fbk.eu

Abstract

Advanced computer vision and machine learning techniques tried to automatically categorize the emotions elicited by abstract paintings with limited success. Since the annotation of the emotional content is highly resource-consuming, datasets of abstract paintings are either constrained in size or partially annotated. Consequently, it is natural to address the targeted task within a transductive framework. Intuitively, the use of multi-label classification techniques is desirable so to synergically exploit the relations between multiple latent variables, such as emotional content, technique, author, etc. A very popular approach for transductive multi-label recognition under linear classification settings is matrix completion. In this study we introduce non-linear matrix completion (NLMC), thus extending classical linear matrix completion techniques to the non-linear case. Together with the theory grounding the model, we propose an efficient optimization solver. As shown by our extensive experimental validation on two publicly available datasets, NLMC outperforms state-of-the-art methods when recognizing emotions from abstract paintings.

1. Introduction

Beyond the automatic recognition of objective properties of images, in the past few years the computer vision research community successfully invested large efforts in the systematic characterization of subjective properties from visual cues. Image aesthetics [27, 45, 26], portrait beauty assessment [31], meaningful texture selection [7], memorability gaugement [13], emotion recognition [28] and creativity [30] are examples of such subjective vision-based recognition tasks. Remarkably, researchers made a tremendous progress in the automatic analysis of artworks targeting a diverse range of tasks, such as inferring paintings styles [23], studying the influences between artists and art movements [37], distinguishing authentic drawings from
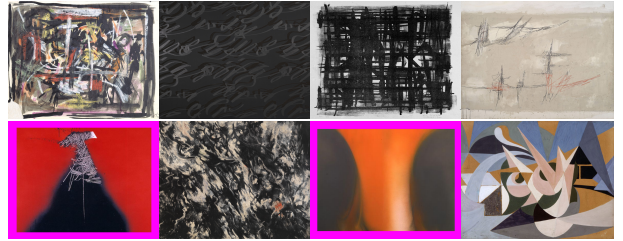

Figure 1. Sample abstract paintings of the MART dataset: which one does elicite in you a positive or a negative emotion?

imitations [12], automatically generating artworks [34] and assessing evoked emotions [32, 46]. Figure 1 shows examples of abstract paintings from the MART dataset eliciting positive and negative emotions: which one does what?[1]

The particular case of the automatic analysis of modern art is exciting and challenging for the research community, since the artist aims to convey strong and deep emotional content to the observer. Indeed, artists immersed into the abstract art movement tend to enhance the non-figurative component and to express "only internal truths, renouncing in consequence all consideration of external form" [18]. Subsequently, when analysing modern art paintings, it is of utmost importance to study the relationship between visual features (*e.g.* colour, shapes, textures) and evoked emotions. In other words, it is crucial, and even more intriguing, to design vision-based learning models able to exploit this link and to predict the emotion evoked by a particular painting. It is therefore unsurprising that there exist several attempts to develop computational approaches for analysing people's emotional experience in reaction to modern artworks [25, 44, 46, 33, 32]. Most of these studies rely on advanced computer vision and machine learning approaches for emotionally categorising artworks [25, 46, 32] and for inferring the parts of the paintings responsible for evoking specific feelings [44]. Other studies investigate how the tex-

---

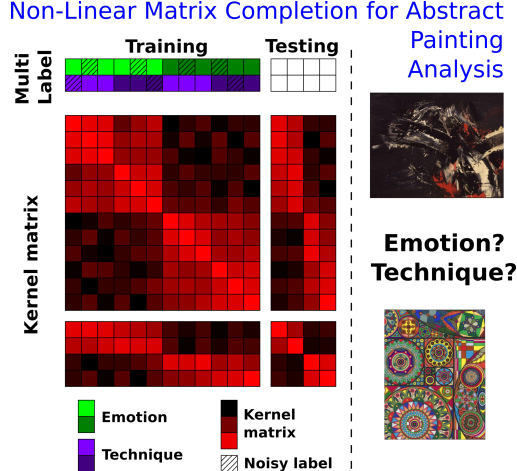[1]The answer is: Filled boxes correspond to negative emotions.

Figure 2. Overview of the proposed non-linear matrix completion framework for abstract painting analysis. Multiple and possibly noisy labels (*i.e.* emotion and painting technique) are estimated from the kernel matrix within a transductive setting, where both the training and testing features (and thus the full kernel matrix) are available at training time.

tual component (the title and the description of a painting) influences the observer's perceptual experience [33].

Importantly, the automatic analysis of modern artworks is challenging for different reasons. First of all, annotating artworks with perceptual attributes is extremely time consuming and requires the development of *ad-hoc* crowd-sourcing platforms together with further post-processing to account for inter-subject and painting variability. Even if conducted under controlled settings, the outcome of this process may lead to noisy and/or missing labels. Second, while the textual component has shown to influence the observer's perception, the title and the description of paintings are very heterogeneous, pithy and not always available. Third, the emotional content is strongly related to other characteristics of the painting, such as the painting technique, the artist or even the creation year. Consequently, we believe that the automatic analysis of modern artworks should be done (i) in a transductive setting so all the visual features, including those of the unlabeled samples, are used for learning and (ii) using multi-label methods able to exploit the relations between different latent variables.

Up to the authors' knowledge one of the most successful transductive multi-label learning frameworks is matrix completion (MC) [9]. Previous works have proven MC to be an effective approach for different computer vision tasks such as multi-label image classification with noisy labels [5], image retrieval and tagging [41, 6], manifold correspondence finding [20] and head/body pose estimation [2, 1]. Until now, matrix completion approaches were tied to assume a linear classification model. In this paper we introduce the first method for non-linear classification with matrix completion and name it non-linear matrix comple-

tion (NLMC). Figure 2 shows an overview of the proposed NLMC approach: multiple and possible noisy training labels are used together with the full (training and testing) kernel matrix to estimate the testing labels. Intuitively, we extend the linear MC to non-linear kernels, where the implicit kernel features may be of infinite dimension, providing all the necessary theoretical background. We show that the problem can be cast into a finite-dimension optimization problem, for which we only need the kernel matrices (and not the features themselves). Finally, we report the method's performance with an extensive set of experiments conducted on two publicly available datasets for emotion recognition from abstract paintings.

Contributions. This paper has several contributions:

- Introducing the very first non-linear learning approach within the well-known matrix completion philosophy and its application to the emotion recognition problem from abstract paintings (we provide all the necessary theoretical background to support the formalization of the method).

- Casting the learning problem using the implicit (potentially infinite-dimensional) features into a non-linear optimization problem for which only the (finite-dimensional) kernel matrix is required, and not the implicit features.

- Reporting an extensive experimental campaign on the only two publicly available datasets for emotion recognition from abstract paintings.[2] In this regard, we compare the accuracy in two tasks: emotion recognition and joint painting technique and emotion recognition, showing the advantage of the proposed approach over state-of-the-art methods on transductive learning and on emotion recognition from abstract paintings.

2. Related Work

In this section we discuss previous works related to (i) visual learning models for emotional analysis of abstract paintings and (ii) matrix completion.

2.1. Painting Emotion Recognition

Nowadays, there is an increasing research interest on developing computational models for emotion analysis of modern art paintings. Previous works investigated the role of several visual features (*e.g.* color, shape and texture) when predicting the emotion conveyed by the artworks to the observers. Yanulevskaya *et al.* [43] proposed an emotion categorization approach based on the aggregation of local image statistics and Support Vector Machines (SVM).

---

[2]We are in private communication with the authors of [3], but the dataset is still unavailable due to copyright issues.

Machajdik *et al.* [25] introduced a unified framework to classify artworks emotionally combining low-level visual features and high-level concepts from psychology and art theory. In [44] a bag-of-visual-words model combined with SVM was proposed to classify abstract paintings into those eliciting positive or negative emotions. Moreover, a back-projection technique was introduced to identify which parts of the paintings evoke positive or negative emotions. Zhao *et al.* [46] proposed an approach for emotional classification of images where specific visual features were designed according to art principles, trying to capture information about balance, harmony, variety and movement. Amirshani *et al.* [3] documented the research effort towards collecting datasets of paintings with annotations based on human perceptual scores. Sartori *et al.* [33] introduced a joint learning framework for abstract painting emotion recognition which integrates both visual and text information. To our knowledge, no previous works have tackled the problem of emotional categorization of modern art paintings considering a multi-label transductive classification framework.

2.2. Matrix Completion

As discussed in the introduction, the recognition of emotions elicited by abstract paintings would definitely benefit from a transductive multi-label learning framework. While several multi-label classification methods under a supervised setting with lots of training data have been developed, for instance [22, 21], research on transductive multi-label classification has received much less attention. Indeed, Kong *et al.* [19] proposed a method based on label set propagation for small datasets, Hariharan *et al.* [10] introduced a max-margin formulation for zero-shot learning applications, Hueber *et al.* [11] derived a generative regressor with missing data, and Goldberg *et al.* [9] studied the matrix completion framework in depth.

This is the technical scope of the present paper, since matrix completion is particularly advantageous when data and labels are noisy or in the case of missing data. Previous research studies in computer vision exploited matrix completion for several applications, such as multi-label image classification [5, 24], image retrieval [41], facial analysis [40, 35] and joint head and body pose estimation [2]. Other works focused on developing algorithms to efficiently solve the MC optimization problem [38, 42]. A recent study [17] extended matrix completion to incorporate an underlying graph structure inducing a weighted relationship between the columns and the rows of the matrix. Importantly, previous works considered matrix completion only for linear classification. The proposed non-linear matrix completion is, up to our knowledge, the first approach for non-linear multi-label classification in a transductive setting able to deal with noisy or missing labels/features.

3. Emotion Analysis using NLMC

The main goal of this study is to analyse the emotional experience of people looking at modern art. Intuitively, the emotional experience is linked to other characteristics of the paintings, such as the painting technique or the style. Therefore, for the sake of completeness, we chose to address this problem within a multi-label framework. Furthermore, given that the annotation cost of the perceived emotion is considerably high, a transductive scenario seems to be the most appropriate. Thus, we extract visual features from m training paintings $\mathbf{X}_0 = [\mathbf{x}_1, \ldots, \mathbf{x}_m]$ and from n testing paintings $\mathbf{X}_1 = [\mathbf{x}_{m+1}, \ldots, \mathbf{x}_{p=m+n}]$, $\mathbf{x}_i \in \mathbb{R}^l$ $\forall i \in \{1, \ldots, p\}$, where l is the dimension of the feature space. The multi-labels are denoted by $\mathbf{y}_i \in \mathbb{R}^k$, and for the sake of clarity we assume they are available for the training set $\mathbf{Y}_0 = [\mathbf{y}_1, \ldots, \mathbf{y}_m]$ while unknown for the testing set $\mathbf{Y}_1 = [\mathbf{y}_{m+1}, \ldots, \mathbf{y}_p]$, although (non-linear) matrix completion can naturally handle partially available labels. In practice, being in a transductive setting means that visual features of all (training *and* testing) samples are used at training time. In the following, we first describe the features extracted from the abstract paintings, to later on state the non-linear matrix completion model and the optimisation problem, and finally propose an efficient solver that leads to the optimum labels.

3.1. Visual features for abstract paintings

We use the state-of-the-art color-based visual features recently proposed in [32]. These color-based image features are inspired from Itten's understanding of the relevance of colors and their combinations [14]. The complete description of the features can be found in [32], but we sketch the three-stage pipeline used to extract them.

Firstly, we use the Color Naming Method (CNM) of [39] to create a visual vocabulary inspired from 11 linguistic labels, considered to be the ones human beings use to express colors. This color nomenclature has shown to be more robust to photometric changes than previous approaches. CNM allows us to map each of the pixels in a painting to a new color value taking into account a small neighborhood of the pixel. Once this has been done independently per each artwork, the new pixel values of all paintings are jointly clustered in the maximum number of possible colors in Itten's model (*i.e.* 180 [14]). After that all pixels are quantized to the centroids obtained from the clustering algorithm. In a second stage, the images quantized accordingly to Itten's color model are segmented using [8]. The two main parameters of this method are the standard deviation of the Gaussian filter and the observation scale. We used the same values as in [32] for both parameters, without observing significant performance variations around this working point. In a third stage, we use two color-based features, namely: color co-occurrence features and patch-

based color-combination features. These features are designed to capture different relations between colors: two-color combinations, the amount of colors, the position of colors and the distance between colors. A fully detailed description of the Itten-inspired features can be found in [32].

3.2. Non-Linear Matrix Completion

Recently, the computer vision community developed interesting methods for classification in the framework of matrix completion [9, 5, 2]. In all MC approaches, the joint label-feature matrix is considered:

$$\begin{bmatrix} \mathbf{Y}_0 & \mathbf{Y}_1 \\ \mathbf{X}_0 & \mathbf{X}_1 \end{bmatrix} \in \mathbb{R}^{(k+l)\times(m+n)}. \tag{1}$$

In previous MC studies the classifier taken under consideration was linear, and the following assumption held:

$$[\mathbf{Y}_0\,\mathbf{Y}_1] = \mathbf{W}\,[\mathbf{X}_0\,\mathbf{X}_1], \tag{2}$$

with $\mathbf{W} \in \mathbb{R}^{k\times l}$ being the classifier's parameters. One of the prominent features of this formulation is that the classifier's parameters are not explicitly computed and the unknown labels are directly estimated. The classifier imposes a linear dependency between the rows of the matrix. Therefore, the matrix completion problem is usually cast into a rank-minimization problem.

Up to the present, nobody considered non-linear classification under the MC framework. In this study we present a formal methodology together with an optimization method so to combine the classification capabilities of the matrix completion framework with the representation power of non-linear kernels applied to the visual features so to efficiently estimate the evoked emotion of abstract paintings.

To this aim, we assume the data is mapped into a h-dimensional space through an unknown feature mapping ϕ. For this we define $\phi_i = \phi(\mathbf{x}_i)$, and from them $\mathbf{\Phi}_0$, and $\mathbf{\Phi}_1$ analogously to $\mathbf{X}_0$ and $\mathbf{X}_1$. Importantly, (i) the dimension of the new feature space is possibly infinite ($h \leq \infty$), (ii) the new features are unknown and (iii) only the kernel matrices, that is $\mathbf{K}_{00} = \mathbf{\Phi}_0^\top\mathbf{\Phi}_0 \in \mathbb{R}^{m\times m}$, $\mathbf{K}_{01} = \mathbf{\Phi}_0^\top\mathbf{\Phi}_1 \in \mathbb{R}^{m\times n}$ and $\mathbf{K}_{11} = \mathbf{\Phi}_1^\top\mathbf{\Phi}_1 \in \mathbb{R}^{n\times n}$, are available. The new label-feature matrix is defined as:

$$\widetilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{Y}_0 & \mathbf{Y}_1 \\ \mathbf{\Phi}_0 & \mathbf{\Phi}_1 \end{bmatrix}. \tag{3}$$

Given the linear relationship between the labels and the new features $[\mathbf{Y}_0\,\mathbf{Y}_1] = \widetilde{\mathbf{W}}\,[\mathbf{\Phi}_0\,\mathbf{\Phi}_1]$, we seek for a low-rank approximation of the matrix $\widetilde{\mathbf{Z}}$:

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} \mathrm{rank}(\mathbf{Z}) \quad \text{s.t.} \quad P_\Omega(\mathbf{Z} - \widetilde{\mathbf{Z}}) = \mathbf{0}, \tag{4}$$

where P_Ω has the effect of a binary mask over the set of all features and training labels, ensuring that the testing visual features $\mathbf{\Phi}_1$ are also used at training time.

While the optimization problem in (4) is well-defined for finite matrices, defining the rank of a matrix with an infinite number of rows is not immediate. Lemma 1 in the supplementary material proves that the rank operator is defined for such matrices (under mild conditions), and subsequently (4) is well-defined with at least one feasible solution: $\mathbf{Z} = \widetilde{\mathbf{Z}}$.

Since minimizing the rank is an NP problem, classical studies on matrix completion used the nuclear norm, that is the sum of the singular values. Again Lemma 1 provides the background to define the singular values (and therefore the nuclear norm) of a matrix with infinite number of rows. Very importantly, and this is the main theoretical result of this study, Theorem 1 proves that the nuclear norm is the tightest convex envelope of the rank (as with finite matrices) and that the following problem is equivalent to (4):

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}} \|\mathbf{Z}\|_* \quad \text{s.t.} \quad P_\Omega(\mathbf{Z} - \widetilde{\mathbf{Z}}) = \mathbf{0}. \tag{5}$$

Furthermore, also inspired by [29], we impose the decomposition $\mathbf{Z} = \mathbf{L}\mathbf{Q}^\top$, where $\mathbf{L} \in \mathbb{R}^{(k+h)\times r}$ and $\mathbf{Q} \in \mathbb{R}^{p\times r}$, and therefore the optimization problem rewrites:

$$(\mathbf{L}^*, \mathbf{Q}^*) = \arg\min_{\mathbf{L},\mathbf{Q}} \|\mathbf{L}\|_\mathcal{F}^2 + \|\mathbf{Q}\|_\mathcal{F}^2$$
$$\text{s.t.} \quad P_\Omega(\mathbf{L}\mathbf{Q}^\top - \widetilde{\mathbf{Z}}) = \mathbf{0}. \tag{6}$$

Importantly, as noted in [29]–Lemma 5.1, the problems (5) and (6) are equivalent for a *sufficiently large* value of r.

Finally, we impose an additional constraint in order to avoid a persistent issue of matrix decomposition techniques (non-negative matrix factorization is another example): the scale ambiguity problem. Indeed, if the j-th column of $\mathbf{L}$ and $\mathbf{Q}$ is multiplied/divided by the same scalar, the final approximation does not change. This ambiguity induces many *identical* local minima in the objective function, thus confusing any optimization solver. Typically, one imposes some kind of normalization on one of the matrices of the decomposition. Without loss of generality, we chose to impose that $\mathbf{Q}$ is orthogonal:[3] $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_r$, where $\mathbf{I}_r$ is the $r \times r$ identity matrix. The optimization problem rewrites:

$$(\mathbf{L}^*, \mathbf{Q}^*) = \arg\min_{\mathbf{L},\mathbf{Q}} \|\mathbf{L}\|_\mathcal{F}^2 + \|\mathbf{Q}\|_\mathcal{F}^2$$
$$\text{s.t. } P_\Omega(\mathbf{L}\mathbf{Q}^\top - \widetilde{\mathbf{Z}}) = \mathbf{0} \text{ and } \mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_r. \tag{7}$$

3.3. Recognizing evoked emotions

In order to recognize the emotions evoked from the abstract paintings, we need to solve the previous optimization problem. For clarity purposes, we transfer the label-feature (resp. the training-testing) structure into $\mathbf{L}$ (resp. $\mathbf{Q}$):

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_0 \\ \mathbf{L}_1 \end{bmatrix} \begin{matrix} \in \mathbb{R}^{k\times r} \\ \in \mathbb{R}^{h\times r} \end{matrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_0 \\ \mathbf{Q}_1 \end{bmatrix} \begin{matrix} \in \mathbb{R}^{m\times r} \\ \in \mathbb{R}^{n\times r} \end{matrix}. \tag{8}$$

---

[3] If $\mathbf{Q} = \mathbf{U}\mathbf{D}\mathbf{V}$ is the SVD of $\mathbf{Q}$ we define $\tilde{\mathbf{L}} = \mathbf{L}\mathbf{V}^\top\mathbf{D}$ and $\tilde{\mathbf{Q}} = \mathbf{U}$, so that $\mathbf{L}\mathbf{Q}^\top = \tilde{\mathbf{L}}\tilde{\mathbf{Q}}^\top$ and $\tilde{\mathbf{Q}}$ is orthogonal.

Once the optimal solution is found, the unknown labels (emotion and painting technique) are estimated using:

$$\mathbf{Y}_1^* = \mathbf{L}_0^*(\mathbf{Q}_1^*)^\top, \qquad (9)$$

where * denotes optimality.

In an ideal scenario, where the visual features and the label annotations were noiseless and fully trustworthy, the equality constrain $P_\Omega(\mathbf{L}\mathbf{Q}^\top - \widetilde{\mathbf{Z}}) = 0$ would be appropriate. In the current scenario, where annotations as well as the extracted visual features may be noisy, relaxing the original problem by means of a measure of how close are estimations to the available observations is intuitively more appropriate. Hence, the objective function rewrites:

$$F(\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{L}_0, \mathbf{L}_1) := \|\mathbf{Y}_0 - \mathbf{L}_0\mathbf{Q}_0^\top\|_\mathcal{F}^2 + \|\mathbf{\Phi}_0 - \mathbf{L}_1\mathbf{Q}_0^\top\|_\mathcal{F}^2$$
$$+ \|\mathbf{\Phi}_1 - \mathbf{L}_1\mathbf{Q}_1^\top\|_\mathcal{F}^2 + \lambda(\|\mathbf{L}_0\|_\mathcal{F}^2 + \|\mathbf{L}_1\|_\mathcal{F}^2), \qquad (10)$$

being λ a regularization parameter (we used $\|\mathbf{Q}\|_\mathcal{F}^2 = r$).

As it is often the case in kernel approaches, the features cannot be explicitly computed. Consequently, the explicit computation of $\mathbf{L}_1$ is also unfeasible. We solve this issue by taking the derivative of the cost function with respect to $\mathbf{L}_1$ and replace its optimal value into the cost function. After this step, the feature matrices are not needed any more, but only the kernel matrices. Since $\mathbf{L}_1$ can potentially have an infinte number of rows, we cannot reason on the rules of finite-dimensional calculus. The theoretical foundations allowing us to take the derivative of a function with respect to a matrix with an infinite number of rows lie in the field of functional analysis and are detailed in the supplementary material. We obtain the following result:

$$\frac{\partial F}{\partial \mathbf{L}_1} = -2\mathbf{\Phi}_0\mathbf{Q}_0 + 2\mathbf{L}_1\mathbf{Q}_0^\top\mathbf{Q}_0 - 2\mathbf{\Phi}_1\mathbf{Q}_1 + 2\mathbf{L}_1\mathbf{Q}_1^\top\mathbf{Q}_1 + 2\lambda\mathbf{L}_1,$$

and by canceling this derivative we obtain:

$$\mathbf{L}_1^* = (\mathbf{\Phi}_0\mathbf{Q}_0 + \mathbf{\Phi}_1\mathbf{Q}_1)(\lambda\mathbf{I}_r + \mathbf{Q}_0^\top\mathbf{Q}_0 + \mathbf{Q}_1^\top\mathbf{Q}_1)^{-1}$$
$$= \frac{1}{\lambda + 1}(\mathbf{\Phi}_0\mathbf{Q}_0 + \mathbf{\Phi}_1\mathbf{Q}_1), \qquad (11)$$

where we used the normalization of $\mathbf{Q}$. By replacing $\mathbf{L}_1$ with its optimal value in the objective function we obtain:

$$F(\mathbf{Q}_0, \mathbf{Q}_1, \mathbf{L}_0) = -2\mathrm{Tr}(\mathbf{Y}_0^\top\mathbf{L}_0\mathbf{Q}_0^\top) + \mathrm{Tr}(\mathbf{L}_0^\top\mathbf{L}_0\mathbf{Q}_0^\top\mathbf{Q}_0)$$
$$+ \lambda\mathrm{Tr}(\mathbf{L}_0^\top\mathbf{L}_0) - \frac{1}{\lambda + 1}\mathrm{Tr}(\mathbf{Q}^\top\mathbf{K}\mathbf{Q}), \quad (12)$$

where $\mathbf{K} = [\mathbf{K}_{ij}]_{ij}$ is the kernel matrix constructed from the visual features, i.e. $\mathbf{K}_{ij} = \mathbf{\Phi}_i^\top\mathbf{\Phi}_j$. We can also take the (regular) derivative with respect to $\mathbf{L}_0$:

$$\frac{\partial F}{\partial \mathbf{L}_0} = -2\mathbf{Y}_0\mathbf{Q}_0 + 2\mathbf{L}_0\mathbf{Q}_0^\top\mathbf{Q}_0 + 2\lambda\mathbf{L}_0 \qquad (13)$$

which cancels out when:

$$\mathbf{L}_0^* = \mathbf{Y}_0\mathbf{Q}_0(\mathbf{Q}_0^\top\mathbf{Q}_0 + \lambda\mathbf{I}_r)^{-1}. \qquad (14)$$

By plugging back this value into F we obtain:

$$F(\mathbf{Q}_0, \mathbf{Q}_1) = -\mathrm{Tr}(\mathbf{Q}_0^\top\mathbf{Y}_0^\top\mathbf{Y}_0\mathbf{Q}_0(\mathbf{Q}_0^\top\mathbf{Q}_0 + \lambda\mathbf{I}_r)^{-1})$$
$$- \frac{1}{\lambda + 1}\mathrm{Tr}(\mathbf{Q}^\top\mathbf{K}\mathbf{Q}). \qquad (15)$$

We reduced the original optimisation problem to:

$$(\mathbf{Q}_0^*, \mathbf{Q}_1^*) = \arg \min_{\mathbf{Q}_0, \mathbf{Q}_1} F(\mathbf{Q}_0, \mathbf{Q}_1)$$
$$\text{s.t.} \quad \mathbf{Q}_0^\top\mathbf{Q}_0 + \mathbf{Q}_1^\top\mathbf{Q}_1 = \mathbf{I}_r, \qquad (16)$$

that can be solved with an interior-point algorithm [36, 4] where the estimation of the Hessian is done with finite differences and the iteration step is computed using conjugate gradient. In that case, the precision of such solver increases drastically if the gradient of the objective function is provided:

$$\frac{\partial F}{\partial \mathbf{Q}_1} = -\frac{2}{1 + \lambda}(\mathbf{K}_{01}\mathbf{Q}_0 + \mathbf{K}_{11}\mathbf{Q}_1), \qquad (17)$$

$$\frac{\partial F}{\partial \mathbf{Q}_0} = 2\mathbf{Q}_0\left(\mathbf{Q}_0^\lambda\right)^{-1}\mathbf{Q}_0^\top\mathbf{Y}_0^\top\mathbf{Y}_0\mathbf{Q}_0\left(\mathbf{Q}_0^\lambda\right)^{-1}$$
$$- 2\mathbf{Y}_0^\top\mathbf{Y}_0\mathbf{Q}_0\left(\mathbf{Q}_0^\lambda\right)^{-1} - \frac{2}{1 + \lambda}(\mathbf{K}_{00}\mathbf{Q}_0 + \mathbf{K}_{01}\mathbf{Q}_1), (18)$$

where $\mathbf{Q}_0^\lambda = \mathbf{Q}_0^\top\mathbf{Q}_0 + \lambda\mathbf{I}_r$. Once the optimal solution $(\mathbf{Q}_0^*, \mathbf{Q}_1^*)$ is found, we can compute $\mathbf{L}_0^*$ with (14) and estimate the emotions and painting style using (9).

4. Experimental Validation

We validate the proposed non-linear matrix completion approach on single-label and multi-label classification. The complete experimental validation consists on five different tests: (i) the evaluation of the color-based features using a standard classification method, (ii) the recognition of the emotional experience of people observing abstract paintings (as in [32] we aim to recognize if an abstract painting elicits a positive or a negative feeling), (iii) the parameter sensitivity analysis of the proposed NLMC method on this task, (iv) the ability of NLMC to jointly estimate the emotion elicited and the painting technique (among *Acrylic*, *Oil*, *Tempera* and *Lithography*) and (v) the ability of NLMC to do that from only one example.

4.1. Datasets

In our experiments we used two publicly available[4] datasets: MART and devArt. The main difference is the
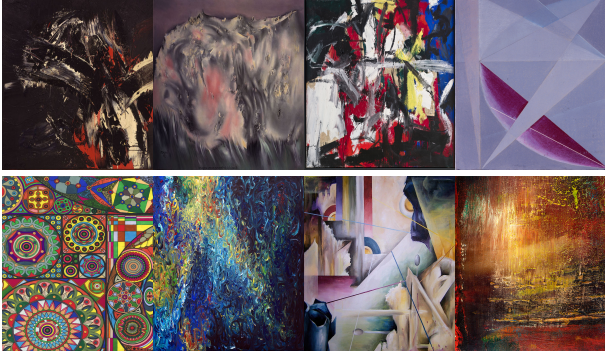
---
[4] http://disi.unitn.it/~sartori/datasets/

Figure 3. Sample paintings from the MART dataset (top row) and from the devArt dataset (bottom row).
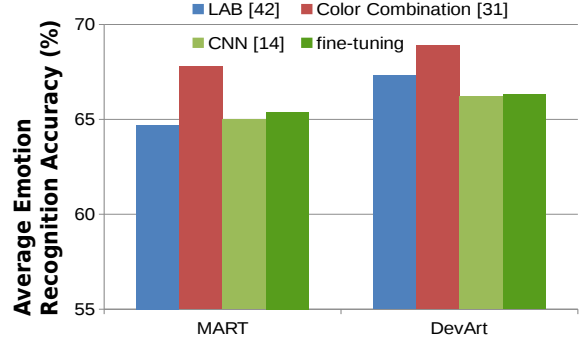

Figure 4. Evaluation of abstract painting emotion recognition for different features.

Table 1. Average emotion recognition accuracy (in %) of all baseline methods in both MART and devArt datasets.

Method	MART	devArt
TSVM [16]	69.2	70.0
LMC [6]	71.8	72.5
Lasso [32]	68.2	70.4
Group Lasso [32]	70.5	72.1
NLMC	**72**.8	**76**.1

artists' background, since MART contains paintings by professional artists and devArt consists on paintings by amateur artists. Each dataset is composed by 500 images. The annotation of the emotional perception was done using the relative score method in [32], which provides positive/negative labels from partial order relations that are answers to the question: "Which painting in the pair looks more positive to you?".

The MART dataset [44] is a collection from the electronic archive of the Museum of Modern and Contemporary Art of Trento and Rovereto (MART). The abstract paintings were produced by almost 80 different professional artists since the beginning of the 20th century until now. Among the painters, there are famous artists like Kandinsky, Albers or Veronesi. Importantly, these artists are known, not only by their exceptional artworks, but also by their theoretical studies on abstract art in terms of color, shapes or texture. An extract of paintings from MART is shown in Figure 3 (top). The painting technique was annotated by an expert.

The devArt dataset is a collection of amateur abstract paintings obtained from the "DeviantArt" online social network[5]. DeviantArt is one of the largest online art communities with more than 280 million artworks and 30 million registered users. We selected the 500 most favored artworks that were under the category Traditional Art/Paintings/Abstract, from 406 different authors. A sample of the devArt dataset is shown in Figure 3 (bottom).

4.2. Results

The features. Abstract art theorists state that color is the most important aspect for abstract painting emotion recognition [14]. Our first experiment serves to find out weather this is also true from a computational point of view. Figure 4 shows the 5-fold cross-validation average emotion recognition accuracy when a SVM classifier is fed with the color combination features [32], the LAB visual words [44] and CNN features [15]. Experimental results shows that the

color combination feature outperforms the two other alternative features. CNN seem not to be effective enough to extract emotion information from the images. Therefore, we only perform experiments with the color-based features.

Emotion Recognition. We compare the average emotion recognition accuracy of the proposed non-linear matrix completion (NLMC) method to four different baselines. On the one hand, to two traditional transductive learning approaches namely: kernel transductive SVM (TSVM [16]) and linear matrix completion (LMC *e.g.* [6]). On the other hand, to two state-of-the-art methods for emotion recognition from abstract paintings: Lasso and Group Lasso, both evaluated in [32]. The decomposition size of NLMC and LMC was cross-validated on $r \in \{2, 3, 4, 5\}$, since no increase of performance was observed for higher values of r. Similarly, for NLMC, LMC and TSVM, the regularization parameter was cross-validated on the set $\{10^{-5}, \ldots, 10^{-1}\}$, and the radial basis function kernel (variance) parameter was cross-validated on the set $\{0.01, 0.0316, 0.1, 0.316, 1.0, 3.16\}$. The regularization parameters of Lasso and Group Lasso were cross-validated in the range $\{0.1, 0.316, 1, 3.16, 10, 31.6, 100\}$ and $\{0.0001, 0.001, 0.01, 0.1, 1\}$ respectively (since for Group Lasso, the regularization parameter represents a proportion of its estimated maximum value[6]). Results in Table 1 are a 12-realization average with a 10-fold cross-validation strategy. The entire software is available at https://github.com/xavirema/nlmc.

---

[5] www.deviantart.com

[6] See the documentation of the publicly available software used for the experiments: http://www.yelab.net/software/SLEP/
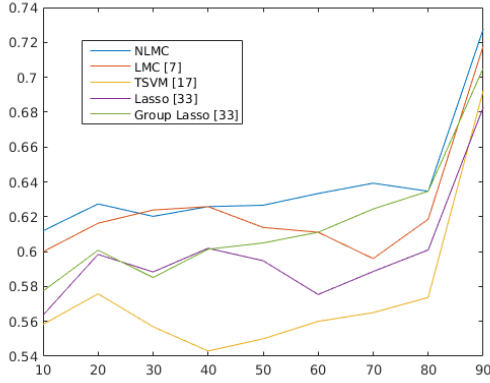
Figure 5. Average emotion recognition accuracy with varying training set size on the MART dataset.
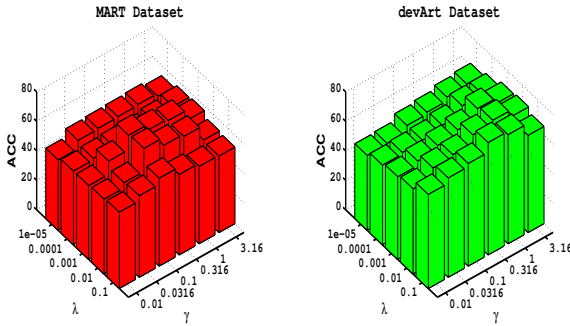

Figure 6. Sensitivity study of parameters of the proposed NLMC method on the MART (left) and devArt (right) datasets.

First of all we observe that the proposed NLMC approach systematically outperforms all baselines in both datasets. Very importantly, this behaviour is observed also another set of experiments aiming to evaluate the performance of the methods when reducing the size of the training set. Since the MART dataset is the most challenging among the two, Figure 5 shows the average recognition accuracy of the methods for different training set sizes. Summarizing, NLMC outperforms, both the state-of-the-art on the emotion recognition task and traditional transductive approaches widely used in the computer vision community. This suggests that NLMC could be beneficial for other computer vision tasks. Finally, Figure 7 shows some qualitative results of the NLMC method on both datasets.

Parameter Sensitivity & Complexity Analysis. We also report (see Figure 6) the parameter sensitivity analysis of the proposed NLMC method. The optimal working point of the method is at $\lambda = 0.01, \gamma = 0.1$ for the MART dataset and at $\lambda = 0.1, \gamma = 0.316$ for the devArt dataset. We notice from the two graphs that the two working points are in a fairly flat part of the surface, meaning that the NLMC method is not very sensitive to variations of the parameters around the optimal working point. In ad-

Table 2. Computational complexity (in s) of different methods on the MART dataset (*i.e.* 400 training/100 testing paintings).

Method	Training	Testing
TSVM [16]	0.79	
LMC [6]	2.82	
Lasso [32]	0.03	0.001
Group Lasso [32]	4.21	0.002
NLMC	1.71	

Table 3. Average multi-label recognition accuracy on the MART dataset: emotion and painting technique.

Method	Emotion	Technique
LSP [19]	67.19	40.63
LMC [6]	68.75	34.69
NLMC	**69.38**	**41.87**

dition to evaluate the performance variation of the method with the parameter values, we also compared its computational complexity. Table 2 shows elapsed time of the training and testing phases (joint in transductive approaches) on the MART dataset with 400 training/100 testing paintings. Experiments ran on a regular laptop with an Intel-i5 processor at 2.67 GHz. We remark that, even if NLMC is not the fastest method, it is much faster than the two closest competitors in terms of performance, namely: Group Lasso and LMC. Regarding the convergence, we have experimentally observed that the relative variation of the objective function goe below 10^{-5} after 20 iterations of the algorithm.

Joint Emotion and Technique Recognition. One of the prominent features of MC in general and NLMC in particular is that the model is naturally able to deal with multiple label at once. Hence, we evaluate the performance of the proposed method within a multi-label classification setting addressing two tasks simultaneously: the recognition of the emotion and of the painting technique. Since the painting technique is not annotated in the devArt dataset, this experiment is only conducted on the MART dataset. Moreover, given that TSVM, Lasso and Group Lasso are not multi-label classification techniques, we here compare with the state-of-the-art transductive multi-label approach in [19], called label set propagation (LSP). The neighborhood size parameter of LSP was cross-validated in the range $\{5, \ldots, 12\}$. Finally, in order to guarantee a fair comparison with [19], we rebalanced the dataset, leading to a dataset with 8 samples per combination of emotion and technique label (eight combinations in total). Consequently, the results we report here are not directly comparable to the ones in Table 1. Average accuracy results on both tasks are reported in Table 3. Consistently with the findings in Table 1, the proposed NLMC approach outperforms all baselines.

Table 4. Average multi-label one-shot recognition accuracy on the MART dataset: emotion and painting technique.

Method	Emotion	Technique
LSP [19]	53.13	24.54
LMC [6]	60.27	32.81
NLMC	**64.51**	**33**.48

Importantly, up to the authors' knowledge this is the first study addressing the task of joint recognition of emotion and painting technique from abstract paintings, thus setting the baseline for future research.

One-shot Emotion and Technique Recognition. Naturally, transductive learning approaches can be used in a *one-shot learning* setting, meaning that the training set consists on one sample per class. Hence, we explored the use of NLMC in such application, thus trying to jointly estimate the emotion and the painting technique from only one annotated sample per label combination (eight samples in total). Table 4 shows the results of our experiments comparing NLMC with LMC and LSP. Even if all methods experience an expected small drop in performance, it is clear that NLMC outperforms LMC and LSP both for emotion and technique recognition, consistently with the previous experiments (*i.e.* Tables 1 and 3).

5. Conclusions

In this paper we address the task of recognizing the emotion elicited by abstract paintings. Since annotating the emotion of a painting is a highly resource-consuming task, it is desirable to perform the learning in a transductive setting. Intuitively, the emotion has complex interdependencies with other characteristics of the painting, such as the painting technique. Therefore we propose to address the task with a new multi-label transductive classifier, named non-linear matrix completion. This is the first work addressing matrix completion, which is inherently transductive and multi-label, under the assumption of a non-linear classifier. We derive the theoretical foundation required to set the optimization problem and propose a solver that performs the learning. In order to validate the approach, we conducted experiments on two publicly available dataset and addressed two tasks: emotion recognition and joint emotion and painting technique recognition. Results show systematic improvement over state-of-the-art on transductive single- and multi-label approaches as well as other supervised approaches previously used in emotion recognition of abstract paintings. Future works will focus on extending the proposed framework to handle missing features in order to integrate other sources of information (*e.g.* text) useful for emotional abstract painting analysis.
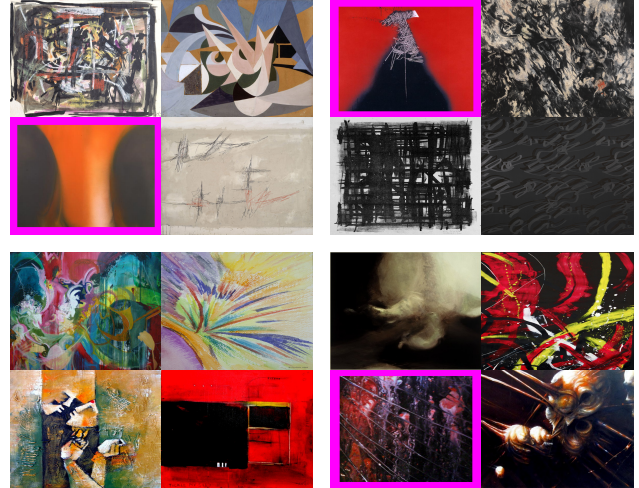

Figure 7. Sample results for abstract painting analysis. Top two groups come from the MART dataset, while bottom two groups from the devArt dataset. Left column corresponds to paintings with positive feelings, while the right column corresponds to negative feelings. Paintings framed in purple corresponds to misclassifications of the proposed NLMC method.

References

[1] X. Alameda-Pineda, J. Staiano, R. Subramanian, L. Batrinca, E. Ricci, B. Lepri, O. Lanz, and N. Sebe. SALSA: A novel dataset for multimodal group behaviour analysis. *TPAMI*, 2016. 2

[2] X. Alameda-Pineda, Y. Yan, E. Ricci, O. Lanz, and N. Sebe. Analyzing free-standing conversational groups: A multimodal approach. In *ACM MM*, 2015. 2, 3, 4

[3] S. A. Amirshahi, G. U. Hayn-Leichsenring, J. Denzler, and C. Redies. Jenaesthetics subjective dataset: Analyzing paintings by subjective scores. In *ECCV Workshops*, 2014. 2, 3

[4] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Math. Prog.*, 89(1):149–185, 2000. 5

[5] R. Cabral, F. De la Torre, J. P. Costeira, and A. Bernardino. Matrix completion for weakly-supervised multi-label image classification. *TPAMI*, 37(1):121–135, 2015. 2, 3, 4

[6] C.-H. Chen, V. M. Patel, and R. Chellappa. Matrix completion for resolving label ambiguity. In *CVPR*, 2015. 2, 6, 7, 8

[7] M.Cimpoi, S.Maji, I.Kokkinos, S.Mohamed, and A.Vedaldi. Describing textures in the wild. In *CVPR*, 2014. 1

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *IJCV*, 59(2):167–181, 2004. 3

[9] A. Goldberg, B. Recht, J. Xu, R. Nowak, and X. Zhu. Transduction with matrix completion: Three birds with one stone. In *NIPS*, 2010. 2, 3, 4

[10] B. Hariharan, S. Vishwanathan, and M. Varma. Efficient max-margin multi-label classification with applications to zero-shot learning. *Machine learning*, 88(1-2):127–155, 2012. 3

[11] T. Hueber, L. Girin, X. Alameda-Pineda, and G. Bailly. Speaker-adaptive acoustic-articulatory inversion using cascaded gaussian mixture regression. *IEEE/ACM TASLP*, 23(12):2246–2259, 2015. 3

[12] J. M. Hughes, D. J. Graham, and D. N. Rockmore. Quantification of artistic style through sparse coding analysis in the drawings of pieter bruegel the elder. *Proceedings of the National Academy of Sciences*, 107(4):1279–1283, 2010. 1

[13] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What makes a photograph memorable? *TPAMI*, 36(7):1469–1482, 2014. 1

[14] J. Itten. The art of color: The subjective experience and objective rationale of color. In *Wiley*, 1974. 3, 6

[15] Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding, 2013. 6

[16] T. Joachims. Transductive inference for text classification using support vector machines. In *ICML*, 1999. 6, 7

[17] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst. Matrix completion on graphs. In *NIPS*, 2014. 3

[18] W. Kandinsky. Concerning the spiritual in art. *Dover Books on Art History Series*, 1914. 1

[19] X. Kong, M. Ng, and Z.-H. Zhou. Transductive multi-label learning via label set propagation. *TKDE*, 25(3):704–719, 2013. 3, 7, 8

[20] A. Kovnatsky, M. M. Bronstein, X. Bresson, and P. Vandergheynst. Functional correspondence by matrix completion. *CVPR*, 2015. 2

[21] X. Li and Y. Guo. Active learning with multi-label svm classification. In *IJCAI*, 2013. 3

[22] Z. Lin, G. Ding, M. Hu, and J. Wang. Multi-label classification via feature-aware implicit label space encoding. In *ICML*, 2014. 3

[23] G. Liu, Y. Yan, E. Ricci, Y. Yang, Y. Han, S. Winkler, and N. Sebe. Inferring painting style with multi-task dictionary learning. In *IJCAI*, 2015. 1

[24] M. Liu, Y. Luo, D. Tao, C. Xu, and Y. Wen. Low-rank multi-view learning in matrix completion for multi-label image classification. In *AAAI*, 2015. 3

[25] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM MM*, 2010. 1, 3

[26] L. Marchesotti, N. Murray, and F. Perronnin. Discovering beautiful attributes for aesthetic image analysis. *IJCV*, 113(3):246–266, 2015. 1

[27] N. Murray, L. Marchesotti, and F. Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *CVPR*, 2012. 1

[28] K.-C. Peng, T. Chen, A. Sadovnik, and A. Gallagher. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *CVPR*, 2015. 1

[29] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010. 4

[30] M. Redi, N. O'Hare, R. Schifanella, M. Trevisiol, and A. Jaimes. 6 seconds of sound and vision: Creativity in micro-videos. In *CVPR*, 2014. 1

[31] M. Redi, N. Rasiwasia, G. Aggarwal, and A. Jaimes. The beauty of capturing faces: Rating the quality of digital portraits. In *FG*, 2015. 1

[32] A. Sartori, D. Culibrk, Y. Yan, and N. Sebe. Who's afraid of itten: Using the art theory of color combination to analyze emotions in abstract paintings. In *ACM MM*, 2015. 1, 3, 4, 5, 6, 7

[33] A. Sartori, Y. Yan, G. Ozbal, A. Salah, A. Salah, and N. Sebe. Looking at Mondrian's Victory Boogie-Woogie: What do i feel? In *IJCAI*, 2015. 1, 2, 3

[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *CVPR*, 2015. 1

[35] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *CVPR*, 2016. 3

[36] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Math. Prog.*, 107(3):391–408, 2006. 5

[37] Y. Wang and M. Takatsuka. SOM based artistic styles visualization. In *ICME*, 2013. 1

[38] Y.-X. Wang, C. M. Lee, L.-F. Cheong, and K.-C. Toh. Practical matrix completion and corruption recovery using proximal alternating robust subspace minimization. *IJCV*, 111(3):315–344, 2015. 3

[39] J. v. d. Weijer, C. Schmid, J. Verbeek, and D. Larlus. Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory. *TIP*, 18(7):1512–1523, 2009. 3

[40] B. Wu, S. Lyu, B.-G. Hu, and Q. Ji. Multi-label learning with missing labels for image annotation and facial action unit recognition. *Pattern Recognition*, 2015. 3

[41] L. Wu, R. Jin, and A. K. Jain. Tag completion for image retrieval. *TPAMI*, 35(3):716–727, 2013. 2, 3

[42] M. Xu, R. Jin, and Z.-H. Zhou. Speedup matrix completion with side information: Application to multi-label learning. In *NIPS*, 2013. 3

[43] V. Yanulevskaya, J. V. Gemert, K. Roth, A. Herbold, N. Sebe, and J. Geusebroek. Emotional valence categorization using holistic image features. In *ICIP*, 2008. 2

[44] V. Yanulevskaya, J. Uijlings, E. Bruni, A. Sartori, E. Zamboni, F. Bacci, D. Melcher, and N. Sebe. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In *ACM MM*, 2012. 1, 3, 6

[45] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, and X. Li. Fusion of multichannel local and global structural cues for photo aesthetics evaluation. *TIP*, 23(3):1419–1429, 2014. 1

[46] S. Zhao, Y. Gao, X. Jiang, H. Yao, T. Chua, and X. Sun. Exploring principles-of-art features for image emotion recognition. In *ACM MM*, 2014. 1, 3

6. Theoretical Foundations

6.1. The rank of a matrix with infinite rows

Lemma 1 *The rank of $\mathbf{Z} \in \mathcal{Z}_p = \{\mathbf{T} \in \mathbb{R}^{\mathbb{N} \times p}, \|\mathbf{T}\|_{\mathcal{F}} < \infty\}$ exists, is finite and not larger than p. Moreover, such a matrix has p (non-negative) singular values.*

Proof The proof has three steps: 1) prove that both the row and column rank cannot exceed p, 2) sketch that these two ranks coincide and therefore the rank operator is well defined and 3) show that there p singular values.

1. The row rank and the column rank of $\mathbf{Z}$ cannot exceed p. On the one side, the rows are all p-dimensional vectors and therefore they cannot span a space with dimensionality higher than p. On the other side, there are only p columns, and therefore they cannot span a space with dimensionality higher than p.

2. The row and column ranks coincide, and therefore we can write rank$(\mathbf{Z})$, and state $0 \leq$ rank$(\mathbf{Z}) \leq p$. Indeed, it is straightforward to extend the result for finite matrices to any matrix in $\mathcal{Z}_p$.

3. $\mathbf{Z}$ has p singular values, that are, by definition, the square root of the eigenvalues of the (finite) matrix $\mathbf{Z}^\top \mathbf{Z}$. Since $\mathbf{Z}^\top \mathbf{Z}$ is a symmetric non-negative definite square matrix of size p, $\mathbf{Z}$ has exactly p non-negative singular values. ∎

6.2. The nuclear norm is the tightest convex envelope of the rank also in $\mathcal{Z}_p$

Theorem 1 *On the set $\mathcal{Z}_p^M = \{\mathbf{Z} \in \mathbb{R}^{\mathbb{N} \times p}, \|\mathbf{Z}\|_{\mathcal{F}} < M\}$ the tightest convex envelope of rank$(\mathbf{Z})$ is $g(\mathbf{Z}) = \frac{1}{M}\|\mathbf{Z}\|_*$.*

Proof The proof has two steps. First, we prove that the bi-conjugate of a real-valued function in a Hilbert space ($\mathcal{Z}_p^M$ is one) is the tightest convex envelope. After that, we prove that the bi-conjugate of the rank is the nuclear norm.

We first recall the definition of the *conjugate* of a function:

Definition 1 *Let $\mathcal{H}$ be a Hilbert space, with scalar product denoted by $\langle \cdot, \cdot \rangle$. Let $f \neq \infty$ minorized by an affine function over $\mathcal{H}$, then the conjugate of f is a function f^* defined as:*

$$f^*(s) := \sup\{\langle s, x\rangle - f(x), x \in \text{dom} f\}, s \in \mathcal{H}. \quad (19)$$

Intuitively, dom$f$ is the set of slopes of all affine functions minorizing f over $\mathcal{H}$. From this definition it is easy to see that f^* also satisfies the conjugability conditions. Therefore we can consider the following object f^{**}, referred to as the *biconjugate* of the function f. Importantly, the following lemma proves that the biconjugate of a function is the tightest convex envelope of the function. In other words:

Lemma 2 *For a given f satisfying the conjugability conditions, f^{**} is the pointwise supremum of all the affine functions on $\mathcal{H}$ majorized by f.*

Proof If $\Omega_f \subset \mathcal{H} \times \mathbb{R}$ denotes the set of pairs (y, r) defining affine functions $x \mapsto \langle y, x\rangle - r$ majorized by f, we have:

$$(y, r) \in \Omega_f \Leftrightarrow f(x) \geq \langle y, x\rangle - r \quad \forall x \in \mathcal{H}$$
$$\Leftrightarrow r \geq \sup\{\langle y, x\rangle - f(x), x \in \mathcal{H}\}$$
$$\Leftrightarrow r \geq f^*(y).$$

Then we obtain:

$$\sup_{(y,r) \in \Omega_f} \langle y, x\rangle - r = \sup_{(y,r)}\{\langle y, x\rangle - r, y \in \text{dom} f^*, -r \leq -f^*(y)\}$$
$$= \sup_y\{\langle y, x\rangle - f^*(y), y \in \text{dom} f^*\}$$
$$= f^{**}(x),$$

as we wanted to prove. ∎

The previous result corresponds to part of Theorem X.1.3.5 of [2] extended to any Hilbert space and allows us to write the biconjugate of a function as:

$$f^{**}(x) = \sup_{r,s}\{\langle y, x\rangle - r, \langle y, z\rangle - r \leq f(z), \forall z \in \mathcal{H}\}. \quad (20)$$

Therefore, the biconjugate of a function f is its tightest convex envelope, and this concludes the first part of the proof. In the second part of the proof, we shall see that the biconjugate of the rank is the nuclear norm. First of all we remark that both the rank and the nuclear norm are well defined for matrices in $\mathcal{Z}_p^M$ as proven in Lemma 1 in the main manuscript.

The proof of the second part follows step-by-step the proof of Theorem 1 in [1], and therefore we just sketch the main line of reasoning. The proof starts by focusing on the set $\mathcal{Z}_p^1$, since the generalization to an arbitrary M is straight-forward. Firstly we compute the conjugate of the rank:

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\| \leq 1} (\langle \mathbf{Y}, \mathbf{X}\rangle - \phi(\mathbf{X})), \quad (21)$$

with $\phi(\mathbf{X}) = \text{rank}(\mathbf{X})$. Since the scalar product is defined as $\langle \mathbf{Y}, \mathbf{X}\rangle = \text{Tr}\left(\mathbf{Y}^\top \mathbf{X}\right)$ we write:

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\| \leq 1} \left(\text{Tr}\left(\mathbf{Y}^\top \mathbf{X}\right) - \text{Rank}(\mathbf{X})\right), \quad (22)$$

Using von Neumann's trace theorem, *i.e.*:

$$\text{Tr}\left(\mathbf{Y}^\top \mathbf{X}\right) \leq \sum_{i=1}^{p} \sigma_i(\mathbf{Y})\sigma_i(\mathbf{X}), \quad (23)$$

where σ_i denotes the i-th singular value, we get to the following result:

$$\phi^*(\mathbf{Y}) = \sum_{i=1}^{p} \left(\sigma_i(\mathbf{Y}) - 1 \right)_+ , \qquad (24)$$

where $(y)_+$ is defined as $(y)_+ = \max\{y, 0\}$. The computation of ϕ^{**} from ϕ^* follows a similar reasoning. The main difference is that

$$\phi^{**}(\mathbf{Z}) = \sup_{\mathbf{Y}} \left(\sum_{i=1}^{p} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \left(\sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r \right) \right),$$

diverges for $\|\mathbf{Z}\| > 1$, and therefore can only be computed for $\mathbf{Z} \in \mathcal{Z}_p^1$. In that case we obtain the desired result (see (Fazel 2002) for details):

$$\phi^{**}(\mathbf{Z}) = \|\mathbf{Z}\|_* , \qquad (25)$$

concluding the proof. ∎

6.3. The derivative with respect to $\mathbf{L}_1$

Definition 2 (Fréchet derivative) *Let V and W be two Banach spaces, and $f : U \subset V \to W$, where U is an open set of V. f is Fréchet differentiable at $x \in U$ is there exists an bounded linear operator $T : V \to W$ such that.*

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Th\|_W}{\|h\|_V} = 0, \qquad (26)$$

where $\| \cdot \|_V$ and $\| \cdot \|_W$ denote the norms of V and W.

Since $\mathcal{Z}_p^M$ is a Hilbert (and therefore Banach) space, the previous definition can be applied to our case. In particular we introduce the following lemma:

Lemma 3 *Let $\mathbf{\Phi} \in \mathcal{Z}_p^M$, $\mathbf{Q} \in \mathbb{R}^{m \times r}$ and $\mathbf{L} \in \mathcal{Z}_r^M$. The Fréchet derivative of $f : \mathcal{Z}_r^M \to \mathbb{R}$ defined as $f(\mathbf{L}) = \|\mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top\|_V^2 = \langle \mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top, \mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top \rangle$ is $T = 2 \left(\mathbf{L}\mathbf{Q}^\top \mathbf{Q} - \mathbf{\Phi}\mathbf{Q} \right)$ seen as a linear operator $\mathcal{Z}_r^M \to \mathbb{R}$.*

Proof The proof is straightforward from the definition. ∎

References

[1] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[2] Hiriart-Urruty, J.-B., and Lemaréchal, C. 1996. *Convex Analysis and Minimization Algorithms: Part I Fundamentals and Part II Advanced Theory and Bundle Methods*, volume 305. Springer Science & Business Media.