European Journal of Operational Research 256 (2017) 35–43

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Continuous Optimization

Matrix completion under interval uncertainty

Jakub Mareček[a,*], Peter Richtárik[b], Martin Takáč[c]

[a] IBM Research – Ireland, B3 F14 IBM Campus Damastown, Dublin 15, Ireland
[b] University of Edinburgh School of Mathematics, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, Scotland, UK
[c] Lehigh University, Harold S. Mohler Laboratory, 200 West Packer Avenue, Bethlehem, PA 18015-1582, USA

ARTICLE INFO

Article history:
Received 5 September 2015
Accepted 10 July 2016
Available online 14 July 2016

Keywords:
Matrix completion
Robust optimization
Collaborative filtering
Coordinate descent
Large-scale optimization

ABSTRACT

Matrix completion under interval uncertainty can be cast as a matrix completion problem with element-wise box constraints. We present an efficient alternating-direction parallel coordinate-descent method for the problem. We show that the method outperforms any other known method on a benchmark in image in-painting in terms of signal-to-noise ratio, and that it provides high-quality solutions for an instance of collaborative filtering with 100,198,805 recommendations within 5 minutes on a single personal computer.

1. Introduction

There has been much recent interest in non-convex optimization problems in statistics, data mining, and machine learning communities. Clearly, non-convex optimization is also at the heart of operations research (Olafsson, Li, & Wu, 2008), where considerable advances are being made, e.g., in decomposition approaches to non-convex optimization, and robust optimization (Gabrel, Murat, & Thiele, 2014). In this paper, we present a decomposition approach to a robust variant of matrix completion, a key problem in data science, with numerous applications ranging from image processing to recommender systems. This shows the value of advances in operations research to data science.

After an informal overview highlighting some key applications, we introduce the problem formally in Section 2. In Section 3, we present our algorithm and its convergence analysis. In Section 4, we present our computational results: In terms of statistical performance, our approach with an explicit consideration of the uncertainty, outperforms a number of previously proposed approaches to matrix completion, on a well-known benchmark. On the computational side, our implementation runs within minutes on a standard personal computer even on instances with a 480, 189 × 17, 770 matrix with 100,198,805 non-zero entries, which had been previously (Gemulla, Nijkamp, Haas, & Sismanis, 2011; Li, Tata, & Sismanis, 2013; Makari, Teflioudi, Gemulla, Haas, & Sismanis, 2015; Teflioudi, Makari, & Gemulla, 2012) solved on substantial clusters of computers in comparable times. We conclude with a variety of suggestions for future work.

1.1. An informal overview

When dimensions of a matrix X and some of its elements $X_{i,j}$, $(i, j) \in \mathcal{E}$ are known, the matrix completion problem is to find the unknown elements. Without imposing any further requirements on X, there are infinitely many solutions. Nevertheless, a matrix completion that minimizes the rank:

$$\min_Y \ \mathrm{rank}(Y) \quad \text{subject to} \quad Y_{i,j} = X_{i,j}, \quad (i, j) \in \mathcal{E}, \tag{1}$$

provides the simplest explanation for the known elements, in many applications. There is a long history of work on the problem, c.f. (Chistov & Grigoriev, 1984; Koren, Bell, & Volinsky, 2009; Sarwar, Karypis, Konstan, & Riedl, 2000; Ye, 2005), with thousands of papers published annually since 2010.

Although we cannot provide a complete overview, let us note that Fazel (2002) suggested to replace the rank, which is the count of non-zero elements of the spectrum, with the nuclear norm, which is the sum of the spectrum. The minimization of the nuclear norm can be cast as a semidefinite programming (SDP) problem and approaches based on the nuclear-norm have proven very successful in theory (Candès & Recht, 2009) and very popular in practice. Cai, Candès, and Shen (2010), Sarwar et al. (2000) study the Singular Value Thresholding (SVT) algorithm. This, however, required the computation of a singular value decomposition (SVD) in each iteration. A number of other approaches, e.g., augmented Lagrangian methods (Tomioka, Suzuki, Sugiyama, & Kashima, 2010), appeared, but those would require a truncated SVD or a number

* Corresponding author.
E-mail addresses: jakub.marecek@ie.ibm.com (J. Mareček), peter.richtarik@ed.ac.uk (P. Richtárik), Takac.MT@gmail.com (M. Takáč).

http://dx.doi.org/10.1016/j.ejor.2016.07.014

of iterations (Jaggi & Sulovský, 2010; Lee & Bresler, 2010; Shalev-Shwartz, Gonen, & Shamir, 2011; Wang et al., 2014) of the power method. Even considering the recent progress in randomized methods for approximating SVD, (Halko, Martinsson, & Tropp, 2011), the approximation becomes very time-consuming as the dimensions of matrices grow.

A major computational break-through came in the form of the alternating least squares (ALS) algorithms (Rennie & Srebro, 2005; Srebro, Rennie, & Jaakkola, 2004). Initially, the algorithm has been used as a heuristic for finding stationary points of the non-convex problem (Bell & Koren, 2007; Haldar & Hernando, 2009; Mnih & Salakhutdinov, 2007; Rennie & Srebro, 2005; Srebro et al., 2004), where a single iteration had complexity $O(|\mathcal{E}|r^2)$, for $|\mathcal{E}|$ observations and rank r, c.f., p. 60 in Keshavan (2012). Keshavan, Montanari, and Oh (2010) and Keshavan (2012), however, proved its exponential rate of convergence to the global optimum with high probability, under probabilistic assumptions common in the compressed sensing community. Independently, Cai et al. (2010) analyzed matrix completion with an arbitrary convex constraint. Further, more technical analyses of the convergence to the global optimum have been performed by Jain, Netrapalli, and Sanghavi (2013).

Many studies of matrix completion consider the uncertainty, in some form. A number of analyses (Jain et al., 2013; Keshavan, 2012; Keshavan et al., 2010) consider the use of the standard rank-minimization for the reconstruction of low-rank $m \times n$ matrix XY^T from $XY^T + W$, where $X \in \mathbb{R}^{m\times r}$, $Y \in \mathbb{R}^{n\times r}$, $W \in \mathbb{R}^{m\times n}$ with elements of W being bounded i.i.d. random variables, which are sub-Gaussian and have bounded expectation. A number of further analyses (Candès, Li, Ma, & Wright, 2011; Wright, Ganesh, Rao, Peng, & Ma, 2009) considered the use of the standard rank-minimization for the reconstruction of low-rank $m \times n$ matrix XY^T from $XY^T + S$, where X, Y are as above and W has a small number of non-zero entries. Chen, Xu, Caramanis, and Sanghavi (2011) consider some columns being corrupted. Although we are not aware of any studies of matrix completion under interval uncertainty, interval-based uncertainty has been considered in related problems. Alaíz, Dinuzzo, and Sra (2013) consider the min-max variant of the problem of finding the nearest correlation matrix, i.e., the problem of finding the closest matrix within the set of symmetric positive definite matrices with the unit diagonal to an uncertainty set, with respect to the Frobenius norm. Li, Ma, and Pong (2014) studied interval uncertainty in certain semidefinite programming problems, which can be used to encode the nuclear-norm minimization.

We present an explicit extension of matrix completion towards interval uncertainty, which has applications in image in-painting, collaborative filtering, and beyond. The algorithm we present for solving the problem can be seen as a coordinate-wise version of the ALS algorithm, which does not require the approximation of the spectrum of the matrix. Before we proceed to describe the actual algorithm, we provide a motivating overview of the possible applications.

1.2. Collaborative filtering under uncertainty

Collaborative filtering is a well-established application of matrix completion problems (Srebro, 2004), largely thanks to the success of the Netflix Prize. There is a matrix, where each row corresponds to one user and each column corresponds to a product or service. Considering that every user rates only a modest number of products or services, there are only a small number of entries of the matrix known. Our extension is motivated by the fact, that one user may provide two different ratings for one and the same product at two different times, depending on the current mood and other circumstances at the two times. One may hence want to consider an interval $[\underline{x}, \overline{x}]$ instead of a fixed value x of the rating, e.g., $[x - \epsilon, x + \epsilon]$. Further, when one knows the scale [0, M] the rating x is chosen from, one can consider $[\max\{0, x - \epsilon\}, \min\{x + \epsilon, M\}]$. Hence, if intervals are known for elements $X_{i,j}$ of a matrix X indexed by $(i, j) \in \mathcal{I}$, one may want to solve:

$$\min_{Y_{i,j}\in[0,M]} \max_{X_{i,j}\in[\underline{X_{i,j}},\overline{X_{i,j}}]\forall(i,j)\in\mathcal{I}} \operatorname{rank}(Y) \quad \text{subject to } Y_{i,j} = X_{i,j}, \quad \forall (i,j) \in \mathcal{I}. \tag{2}$$

Although numerous extensions of matrix completion problems have been studied, e.g. (Mehta, Hofmann, & Nejdl, 2007), the use of robustness to interval uncertainty is novel. It can be seen as an extension of robust optimization (Soyster, 1973) to matrix completion.

1.3. Image in-painting

Further applications can be found in image processing. In in-painting problems, a subset of pixels from an image are given and the goal is to fill in the missing pixels. Rank-constrained matrix completion with equalities, where $\mathcal{I}$ is the index set of all known pixels, has been used numerous times (Candès & Recht, 2009; Goldfarb, Ma, & Wen, 2009; Jaggi & Sulovský, 2010; Jain, Meka, & Dhillon, 2010; Lee & Bresler, 2010; Mazumder, Hastie, & Tibshirani, 2010; Wang et al., 2014) in this setting. If the image comes from real sensors, it the corresponding matrix may have full (numerical) rank, but have quickly decreasing singular values in its spectrum. In such a case, instead of solving the equality-constrained problem (1), one should like to find a low-rank approximation Y^* of X, such that the known entry of X is not far away from Y^*, i.e., $\forall (i, j) \in \mathcal{I}$ we have $Y_{i,j} \approx X_{i,j}$. Let us illustrate this with a small matrix

$$X = \begin{pmatrix} 68.16 & 78.12 & 24.04 \\ 78.12 & 90.09 & 30.03 \\ 24.04 & 30.03 & 20.01 \end{pmatrix},$$

which has rank 3 and its singular values $\Sigma = (167.9945, 10.2553, 0.0102)^T$. It is easy to verify that

$$Y^*(2) = \begin{pmatrix} 68.1546 & 78.1250 & 24.0389 \\ 78.1250 & 90.0853 & 30.0310 \\ 24.0389 & 30.0310 & 20.0098 \end{pmatrix}$$

is the best rank 2 approximation of X in Frobenius norm. Observe that no single element of $Y^*(2)$ is identical to X, but that $Y^*(2) \approx X$. It is an easy exercise to show that for any $X \in \mathbb{R}^{m\times n}$ with singular values $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_{\min\{m,n\}}$, and $Y^*(r)$ as its best rank-r approximation, we have $|X_{i,j} - (Y^*(r))_{i,j}| \le \sum_{i=r+1}^{\min\{m,n\}} \sigma_i =: \mathcal{R}(r)$ for all (i, j). Therefore, one should not use equality constrains in (1), but rather inequalities $|Y_{i,j} - X_{i,j}| \le \mathcal{R}(r), \forall (i, j) \in \mathcal{I}$. Notice that this approach is not the same as minimizing $\sum_{(i,j)\in\mathcal{I}} (X_{i,j} - Y_{i,j})^2$ over all rank r matrices, because we do not penalize the elements of Y, which are already close to X. It is also different from the usual treatment of noise in the observations (Candès & Plan, 2010). One could rather formulate this as the minimization of $\sum_{(i,j)\in\mathcal{I}} \max\{0, |X_{i,j} - Y_{i,j}| - \mathcal{R}(r)\}^2$ over all rank r matrices. Further, one knows the range of values allowed, e.g., [0, 1] for a common encoding of gray-scale images. This can hence be seen as "side information" which, as we will show in Section 4, improves the recovery of a low-rank approximation considerably. Further still, one could assume that the intensity should be at least 0.8, if pixels are missing within a light region of the image, or similar domain-specific heuristics.

A number of other applications, e.g., in the recovery of structured matrices (Chen & Chi, 2013), in certain forecasting problems with periodic time series and side information, and in sparse principal component analysis with priors on the principal components can be envisioned. Some are discussed in Section 5. Now, let us introduce our notation and formalize the problem.