Journal of Computational Mathematics
Vol.36, No.3, 2018, 374–390.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1707-m2016-0796

SNIG PROPERTY OF MATRIX LOW-RANK FACTORIZATION MODEL*

Hong Wang
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China
Email: hong.wang@connect.polyu.hk
Xin Liu
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
Email: liuxin@lsec.cc.ac.cn
Xiaojun Chen
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China
Email: xiaojun.chen@polyu.edu.hk
Ya-xiang Yuan
State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
Email:yyx@lsec.cc.ac.cn

Abstract

Recently, the matrix factorization model attracts increasing attentions in handling large-scale rank minimization problems, which is essentially a nonconvex minimization problem. Specifically, it is a quadratic least squares problem and consequently a quartic polynomial optimization problem. In this paper, we introduce a concept of the SNIG ("Second-order Necessary optimality Implies Global optimality") condition which stands for the property that any second-order stationary point of the matrix factorization model must be a global minimizer. Some scenarios under which the SNIG condition holds are presented. Furthermore, we illustrate by an example when the SNIG condition may fail.

Mathematics subject classification: 15A18, 15A83, 65K05, 90C26
Key words: Low rank factorization, Nonconvex optimization, Second-order optimality condition, Global minimizer.

1. Introduction

1.1. Problem description

Consider the following matrix factorization problem

$$\underset{Y\in\mathbb{R}^{n\times k},Z\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad f(Y,Z):=\frac{1}{2}||\mathcal{A}(YZ^\top)-b||_2^2=\frac{1}{2}\sum_{i=1}^{p}\left(\langle A_i,YZ^\top\rangle-b_i\right)^2, \tag{1.1}$$

where $b\in\mathbb{R}^p$ is a column vector, $\mathcal{A}\in\mathcal{B}(\mathbb{R}^{n\times m},\mathbb{R}^p)$ is a linear operator mapping $n\times m$ matrices onto p-dimensional Euclidean space. Namely,

$$\mathcal{A}(X)=\left(\langle A_1,X\rangle,...,\langle A_p,X\rangle\right)^\top,$$

* Received November 22, 2016 / Revised version received May 2, 2017 / Accepted July 23, 2017 / Published online March 28, 2018 /

where $A_i \in \mathbb{R}^{n\times m}(i = 1, ..., p)$ are the p column matrices of $\mathcal{A}$ and $\langle W_1, W_2\rangle := \mathrm{tr}(W_1^\top W_2)$ designates the inner product of two matrices W_1 and W_2 with the same size. Denoting the adjoint operator of $\mathcal{A}$ by $\mathcal{A}^\top : \mathbb{R}^p \mapsto \mathbb{R}^{n\times m}$, it is not difficult to verify that

$$\mathcal{A}^\top(y) = \sum_{i=1}^{p} y_i A_i.$$

1.2. Existing works

Model (1.1) appears to be a practical and efficient way for solving low-rank optimization problem. It is also arisen from many areas of scientific and engineering applications including matrix completion, principle component analysis (PCA) and others [1, 8, 21]. LMaFit [28], for instance, using a series of matrix factorization models with different k (the approximation of the optimal rank) to describe the matrix completion problem, turns out to be an efficient and robust alternative to the convex relaxation model [3, 7, 11, 18] based on nuclear norm relaxation [4–6, 12, 19, 25]. Matrix factorization is also used to tackle semidefinite programs (SDP) problems. For instance, [2, 14] introduced an equivalent factorization model for SDP through the Cholesky decomposition. Mishra in [19] used a factorization to make the trace norm differentiable in the search space and the duality gap numerically computable, which is a similar approach to SVD.

However, the factorization model (1.1) is nonconvex. More specifically, it is a quartic polynomial optimization problem. It may contain exponential number of local minimizers or saddle points. Hence, solving problem (1.1) to the global optimality is usually unachievable.

Recently, Candès and Li [9] proposed a so-called Wirtinger Flow (WF) method to solve the phase retrieval problem, which is, like (1.1), essentially a quardratic least squares problem and quartic polynomial problem. The WF algorithm consists of two phases, one is a careful initialization stage realized by a spectral method, and the other is the local minimization stage invoking a gradient descent algorithm with a restricted stepsize. The authors proved that if the random sampling vectors obey certain distribution and there is no noise in the observation, the sequence generated by the gradient descent scheme will converge linearly to a global solution with high probability. Sun and Luo in [22] applied a similar idea to analyze the matrix completion problems described by factorization formulation, in which an initialization step is followed by a general first-order algorithm framework. Under the standard assumptions on incoherence condition [4] and the random observations similar to [9], the authors of [22] showed their framework can converge to a global solution linearly. Ge et. al [10] proved that matrix completion problem, a special case of (1.1), does not have spurious local minimum under the positive definiteness and randomness assumptions on the target matrix, i.e. the observation vector b in our model (1.1).

1.3. Our contributions

Even if the linear operator $\mathcal{A}$ of problem (1.1) does not involve any random property, it is observed that some local optimal solvers can often find a global solution of (1.1) by starting from a randomly chosen initial point. In this paper, we theoretically investigate the relationship between the global optimality of problem (1.1) and its second-order optimality under certain scenarios, which can partly explain the above mentioned phenomenon.

Note that if there exists a nonzero vector $c \in \mathbb{R}^p$ such that $\mathcal{A}^\top(c) = \mathbf{0}$, the linear operator $\mathcal{A}$ is row linearly dependent which implies the redundancy of the observations $\mathcal{A}(YZ^\top) = b$.