SECOND-ORDER METHODS FOR NONCONVEX OPTIMIZATION: THEORY AND COMPLEXITY ANALYSIS

HONG WANG

Ph.D

The Hong Kong Polytechnic University

2017

The Hong Kong Polytechnic University

Department of Applied Mathematics

Second-order Methods for Nonconvex Optimization: Theory and Complexity Analysis

Hong Wang

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

July 2016

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signature)

Hong WANG (Name of student)

Declaration of Relationship with External Examiner

With the support of the PhD attachment program of PolyU, I visited University of California, Los Angles from 16 March 2015 to 17 July 2015. Prof. Wotao Yin was my host supervisor during that period of four months. Under the guidance of Prof. Yin, I studied a research topic which was named "*Decentralized consensus optimization over directed graph*" and it was proposed by Prof. Yin. This topic is not related to this PhD thesis.

I did not have any academic contact or discussions with Prof. Yin since I completed the attachment program and left America. I completed my PhD thesis under the guidance of my supervisor, Prof. Chen, after I came back to Hong Kong. In 2016 July, Prof. Yin became one of my External Examiners.

I hereby declare that I did not have any co-authorship research output with Prof. Yin in the past, and will not have any co-authorship research output with Prof. Yin in the next 5 years. I declare that I do not have any relationship with Prof. Yin related to this PhD thesis. Prof. Yin does not have any contributions to this PhD thesis.

_____(Signature)

Hong WANG (Name of student)

Dedicated to my parents.

Abstract

We consider the second-order methods for solving two classes of nonconvex minimization problem arising from diverse applications of optimization. By second-order methods, we refer to those methods involving second-order information of the objective function. The first class of nonconvex problem is the so-called Affine Rank Minimization Problem (ARMP), whose aim is to minimize the rank of a matrix over a given affine set. The other one is the Partially Separable Minimization Problem (PSMP), which is to minimize the objective function with a partially separable structure over a given convex set. This thesis hence can be sharply divided into two distinct parts.

In the first part, we focus on exploring the ARMP utilizing the matrix factorization reformulation. Under some particular situations, we show that the corresponding factorization models are of the property that all second-order stationary points are global minimizers. By presuming such property holds, we propose an algorithm framework which outputs the global solution of the ARMP after solving a series of its factorization models with different ranks to the second-order necessary optimality. Finally, we put forward a conjecture that the reduction between the global minima of the low-rank approximation with consecutive ranks is monotonically decreasing with the increase of the rank. If this conjecture holds, we can accelerate the estimation of the optimal rank by an adaptive technique, and hence significantly increase the efficiency of the overall performance of our framework in solving ARMP.

In the second part of this thesis, we mainly study the PSMP over a convex constraint. We first propose an adaptive regularization algorithm for solving PSMP, in which the expense of using high-order models is mitigated by the use of the partially separable structure. We then show that the algorithm using an order p model needs at most $O(\epsilon^{-\frac{p+1}{p}})$ evaluations of the objective function and its derivatives to arrive at an ϵ-approximate first-order stationary point. The complexity in terms of ϵ is unaffected by the use of structure. An extension of the main idea is also presented for the case where the objective function might be non-Lipschitz continuous. We apply the algorithm with an adaptive cubic regularization term to solving the problem of data fitting involving the q-quasi norm for $q \in (0, 1)$ and it turns out that even for non-Lipschitz case, the complexity bound $O(\epsilon^{-\frac{3}{2}})$ can be retained.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Prof. Xiaojun Chen, my chief supervisor, for her motherly patient care and generous support throughout my PhD career. She pushed me to go ahead, taught me to value my time, trained me how to do presentations and many more things than I could have expected from a supervisor. Special gratitude to her for supporting me to apply for the PhD outgoing program so that I could visit University of California, Los Angeles during my PhD period. A big thanks goes to Dr. An Li, husband of Prof. Chen, for treating us delicious food and good wine during Christmas/Spring Festival dinners and other important days, which colored my memory of the past three years.

Special thanks are due to Prof. Xin Liu (Academy of Mathematics and System Sciences, Chinese Academy of Sciences) for leading me into the field of matrix optimization and showing me the charm of matrix. I also appreciate that he invited me to visit him and recommended me to Prof. Wotao Yin, which provided me with the opportunities to learn from him and Prof. Yin. He is a great collaborator and friend. He spent a tremendous amount of time discussing with me on my research which dragged me back from dilemma again and again. Without his help and guidance, it is impossible for me to finish my first PhD project, which led to the first part of this thesis.

I would also like to thank Prof. Ya-xiang Yuan (Academy of Mathematics and System Sciences, Chinese Academy of Sciences) for his insightful comments and

suggestions on the first part of this thesis. Furthermore, I really appreciate the opportunities to visit his research group in Beijing with support from his grants. I learned a lot from those precious visiting experiences. And it was really happy and unforgettable to hike and play cards with Prof. Yuan and his group.

My sincere acknowledgements go to Prof. Wotao Yin (University of California, Los Angeles) for giving me the opportunity to visit his research group. His vision and passion have urged me a lot. He also spent plenty of time discussing with me, from which I learned how to think and gradually approach to the target when facing a new topic.

I am very grateful to Prof. Xiaoping Xue and Prof. Wei Bian from Harbin Institute of Technology, my former supervisor and academic sister in Harbin, respectively. Without their support and help, it would be impossible for me to be here to conduct my PhD study. Also, Prof. Wei Bian provides me with a lot of help on my research, including proofreading of my draft, offering me her MATLAB code and so forth.

Prof. Philippe Toint (University of Namur) visited my supervisor Prof. Xiaojun Chen for three weeks at the beginning of 2016, during which we started a new project and hence we spent a lot of time together on discussions. The second part of this thesis is mainly derived from our joint work. My special appreciations are due to him for his insightful ideas, patient guidance and generous help.

I would also like to acknowledge Prof. Zhaosong Lu (Simon Fraser University) and Prof. Fengmin Xu (Xi'an Jiaotong University) for their kind help and special care. Prof. Lu is a well-read scholar. I learned a lot from several discussions with him both in our group meetings and in private. It was great kind of Prof. Xu to give me insightful suggestions on my future career. I really appreciate it.

Over the past three years, it is my great honor to have the opportunities to visit Chinese Academy of Sciences and University of California, Los Angeles. I was lucky to meet and make many new friends during those visits, including Dr. Qian Dong,

Mr. Ran Gu, Mr. Cheng Chen, Mr. Bin Gao, Mr. Liang Zhao in Beijing and Dr. Ming Yan, Dr. Zhiming Peng, Mr. Zhenbo Lu in Los Angeles. Specially, my appreciations go to Mr. Cheng Chen and Mr. Zhenbo Lu for their kind help when I was in Beijing and Los Angeles, respectively.

My sincere and deep appreciations are due to my academic brothers and sisters: Prof. Congpei An, Dr. Yanfang Zhang, Dr. Zaikun Zhang, Prof. Hailin Sun, Dr. Yang Zhou, Ms. Qiyu Wang, Mr. Lei Yang, Mr. Bo Wen, Mr. Yun Shi and Mr. Guidong Liu for their constant and substantial support.

My special thanks go to my former and current officemates for their company and help throughout my stay. The list is very long and dull, but deserves to be here: Dr. Chuanxin Bian, Dr. Haiyang Jin, Ms. Xianping Wu, Mr. Haibin Chen, Ms. Qun Wang, Mr. Shujun Wang in DE 405, Dr. Zhibao Li, Dr. Boshi Tian, Mr. Jin Yang, Ms. Qiujin Peng, Ms. Yan Wei in HJ 609, and Ms. Yue Shi, Ms. Meiling Hao, Ms. Chenchen Zu, Ms. Huili Zhang, Ms. Kaihui Liu, Ms. Chen Ouyang, Ms. Fei Fang and Mr. Qingzheng Wang in TU 834. Here I would particularly like to single out Dr. Chuanxin Bian and Dr. Yang Zhou, who gave me generous help when I was fresh in Hong Kong.

Last but far from least, my deepest gratitude and love belong to my mother Shulan Wu, my father Nengsong Wang, my brother Yong Wang and all my family for their unconditional love and consistent support through my life.

Contents

List of Figures

List of Tables

List of Notations

$\mathbb{R}$	The set of real numbers
$\mathbb{R}^n$	The set of n-dimensional real vectors
$\mathbb{R}^{m\times n}$	The set of $m \times n$ real matrices
$\text{rank}(X)$	The rank of X
$\text{tr}(X)$	The trace of X
$\|X\|$	The spectral norm (maximum singular value) of X
$\|X\|_*$	The nuclear norm of X
$\|X\|_F$	The Frobenius norm of X
$\lambda_i(X)$	The ith largest eigenvalue of X
$\sigma_i(X)$	The ith largest singular value of X
$I(I_n)$	The identity matrix (of n-dimension)
$\mathcal{B}(\mathbb{R}^{n\times m}, \mathbb{R}^p)$	The set of all linear operator mapping $\mathbb{R}^{n\times m}$ to $\mathbb{R}^p$
$\mathcal{A}$ ($\mathcal{A}^{\text{T}}$)	A linear operator (resp. its adjoint operator)
$N(\mathcal{A})$	The nullspace of operator $\mathcal{A}$
$R(\mathcal{A})$	The range space of operator $\mathcal{A}$
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	Finite-dimensional real Hilbert spaces
$\mathcal{X}'$	The dual space of a Hilbert space $\mathcal{X}$

Chapter 1

Introduction

In this thesis, we study second-order methods for nonconvex optimization problems that involves two classes of problems. The first one is minimizing the rank of a matrix over an affine set. We refer to this as the Affine Rank Minimization Problem (ARMP), which arises in a diverse set of areas such as control, system identification, statistics, and signal processing, to name a few.

The other one is minimizing an objective function over a convex set, which is partially separable in terms of decision variables but may be nonconvex and non-Lipschitz. We refer to it as the Partially Separable Minimization Problem (PSMP). The partially separable structure studied here is ubiquitous in applications of optimization. It subsumes sparse optimization as a special case, which has been widely used in image restoration and statistics.

For these two classes of problems, we will investigate thoroughly their structures and optimality conditions. Based on that, we will devise second-order methods for solving them, respectively. By second-order methods, we here mean that the methods that utilize the second-order information of the problem, i.e., the second-order derivatives of the objective function.

1.1 Affine Rank Minimization Problem (ARMP)

The affine rank minimization problem (ARMP), whose aim is to find an unknown matrix of low rank subject to a given system of linear equality constraints, has increasingly gained a great amount of attention in recent years. Such a prevailing trend can not only be ascribed that ARMP can be viewed as, in some sense, an extension of matrix completion problem [14, 46, 72, 83, 84] and compressive sensing [13, 34, 84], but also be attributed to the wide applications of ARMP, the good property of low-rank and the rich content in matrix topics.

First of all, ARMP abstracts a lot of problems arisen from a diverse set of application domains, such as machine learning, dimensionality reduction and control theory. Specific applications of ARMP include dynamic system identification [61] and state covariance estimation [26, 58, 103] in control theory, collaborative filtering [55] and multi-class learning [1, 33] in machine learning, phase retrieval [10], Euclidean embedding and dimensionality reduction in sensor networks [51, 78, 90, 91], Direction of Arrival (DOA) estimation in radar signal processing [94] and index coding in information theory [32]. And see [53] for more discussions of relative topics.

The property of low-rankness provides us with a special and efficient perspective in handling a lot of diversely large-scale problems since it extracts the essential characteristics of a wide range of concepts. For instance, dimensionality, complexity or order can often be expressed as the rank of some appropriate matrix. See [31, 35, 60] for more details. Moreover, analysis of the low-rank minimization model may also be extended to many other specific low-rank models, such as Principle Component Analysis (PCA), robust PCA [11], Non-negative matrix factorization (NMF) [57], tensor completion [38, 71] and so forth.

Last but not least, the topic of matrix is just like a tropical forest consisting of various populations. It usually involves a blend of many disciplines, such as linear

algebra, optimization, information theory, probability, graph theory and algebraic geometry and thus the investigation of rank minimization problem will not only be advantageous to interdisciplinary collaboration and also may lead to new spark of thoughts.

1.1.1 Approaches for Solving ARMP

As mentioned at the very beginning, there are two elements, i.e., low-rankness and consistency with linear equality constraints, in modeling the matrix rank minimization. An ARMP can usually be cast as the following rank minimization problem:

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad & \operatorname{rank}(X) \\ \text{subject to} \quad & \mathcal{A}(X)=\mathbf{b}, \end{aligned} \tag{1.1}$$

where $\mathbf{b}$ is a p-dimensional column vector and $\mathcal{A}:\mathbb{R}^{n\times m}\to\mathbb{R}^p$ is a linear operator[1]. More specifically, operator $\mathcal{A}$ can be defined as follows,

$$\mathcal{A}(X)=\left(\langle A_1,X\rangle,\langle A_2,X\rangle,\ldots,\langle A_p,X\rangle\right)^{\mathrm{T}},$$

where $A_i\in\mathbb{R}^{n\times m}$ $(i=1,2,\ldots,p)$ are called the column matrices of operator $\mathcal{A}$, and $\langle M_1,M_2\rangle:=\operatorname{tr}(M_1^{\mathrm{T}}M_2)$ denotes the inner product of two matrices with same size.

It should be noted that if there exists a nonzero vector $\mathbf{c}\in\mathbb{R}^p$ such that $\mathcal{A}^{\mathrm{T}}(\mathbf{c})=0$, then the linear operator $\mathcal{A}$ is row linearly dependent which means the redundancy of the constraint $\mathcal{A}(x)=\mathbf{b}$. Here $\mathcal{A}^{\mathrm{T}}$ denotes the adjoint operator of $\mathcal{A}$ (see Section 2.2 of Chapter 2 for details), which actually has an explicit expression (to be deduced in Section 2.2 of Chapter 2). Due to the linearity of operator $\mathcal{A}$, a collection of independent constraints can be selected and the rest of constraints can be expressed as the linear combination of the selected independent constraints.

Hence, without loss of generality, to simply invoke the second-order optimality condition in the further analysis, it is reasonable for us to concentrate on the cases

[1] Any linear operator from Euclidean space to Euclidean space is bounded.

in which the operator $\mathcal{A}$ is of full row rank[2], namely, $\mathcal{A}^{\mathrm{T}}(\mathbf{c}) = 0$ implies $\mathbf{c} = 0$. For the convenience of later analysis, we generalize it as the following assumption and assume it always holds throughout this thesis.

Assumption 1.1. *The operator $\mathcal{A}$ is of full row rank, that is, $\mathcal{A}^{\mathrm{T}}(\mathbf{c}) = 0$ implies $\mathbf{c} = 0$ for any $\mathbf{c} \in \mathbb{R}^p$.*

The rank function is a discrete function and is difficult to deal with directly. There are usually two different approaches to handle problem (1.1). The first one is to relax (1.1) to a nuclear norm minimization problem

$$\begin{aligned} \underset{X \in \mathbb{R}^{n\times m}}{\text{minimize}} \quad & \|X\|_* \\ \text{subject to} \quad & \mathcal{A}(X) = \mathbf{b}, \end{aligned} \tag{1.2}$$

where $\|X\|_*$ is called the nuclear norm or the Ky-Fan n-norm of X, i.e., the sum of all singular values of X; see, *e.g.*, [48].

Fazel et al. in [34, 35] introduced the nuclear norm (i.e., the sum of singular value of a matrix) as the surrogate of the rank function for problem (1.1) and thus obtain the relaxed minimization (1.2). The nuclear norm of a matrix can be seen as the ℓ_1 norm of the vector consisting of all singular values of that matrix while the rank of the matrix is ℓ_0 norm of that vector. Hence, the ℓ_1 norm relaxation techniques can be naturally extended to rank minimization case.

On the side of theory, a series of excellent works [14, 46, 72, 83, 84] established the recoverability for matrix completion problems. It has been shown that for matrix completion problem (1.4), solving the corresponding problem (1.2) will exactly recover matrix M with high probability if M of rank r satisfies the so-called incoherence condition and $O(rn\log^2 n)$ entries of it are uniformly revealed. Negahban et al. in [74, 75] also gave the so-called restricted strong convexity property to guarantee

[2] See Definition 2.3 for details.

exact reconstruction for matrix completion. [74, 79, 80, 85] extended the null space conditions to matrix case from its vector counterpart and established exact recovery theory.

On the side of computation, the nuclear norm is a convex function which can be optimized efficiently and globally. In fact, problem (1.2) can be reformulated as semidefinite program (SDP) and hence solved to global minimizer by standard SDP solvers if the problem scale is not large (say, the dimension of matrix smaller than 500). For large-scale problem, quite a few efficient algorithms have been proposed as well over the past ten years, such as the SVT (singular value thresholding) algorithm [9, 49], the Bregman iterative algorithm [40, 64], the alternating direction method of multipliers [20, 59, 100] and accelerated proximal gradient method [97].

In the CS, the performance of ℓ_p $(0 < p < 1)$ approximation is often better than that of ℓ_1 relaxation in terms of recovery condition and result. Following the same idea, the so-called Schatten p-norm [54, 65] minimization, which is the ℓ_p norm of the vector consisting of all the singular values of a given matrix, is utilized to approximate the rank minimization problem. However, the Schatten p-norm minimization is difficult to solve. A tractable way is the iterative reweighted method, which can be as well applied to ℓ_0 minimization of the singular values directly, see [36, 56, 62, 69, 70]. In [63], the authors proposed a block coordinate descent algorithm to solve the penalized rank minimization problem directly. In [98], the authors extended orthogonal matching pursuit method from the vector case to the matrix case, and obtain an efficient and scalable approach for matrix completion problems.

The second approach is to re-express the decision variable matrix X as the decomposition form $X = YZ^{\mathrm{T}}$ where $Y \in \mathbb{R}^{n\times k}$ and $Z \in \mathbb{R}^{m\times k}$ according to the fact that any matrix with rank no more than k can be decomposed into the product of two matrices of rank k. In this case, the constraint of $\mathrm{rank}(X) \leqslant k$ is embedded into the objective function. Then (1.1) can be reformulated as the following unconstrained

optimization problem

$$\underset{Y\in\mathbb{R}^{n\times k},Z\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad f(X,Y) := \|\mathcal{A}(YZ^{\mathrm{T}}) - \mathbf{b}\|_2^2. \tag{1.3}$$

Decomposition based formulation (1.3) has attracted increasing attention in the recommender systems field and served as the foundation of many algorithms for the Netflix Prize [55, 96].

In the era of big data, tremendous amounts of data arisen from online merchants, mobile devices, sensors and so forth, need to be analyzed and handled, which brings about a lot of large-scale problems. However, the singular value decomposition computation, which is prohibitively expensive, is heavily involved when one applies nuclear norm based method to search the solution of problem (1.1) and it takes up the main computational cost per iteration.

To get rid of the excessive invoking of the singular value decomposition calculation and solve large-scale problems more efficiently, people turn to another direction, namely, methods based on model (1.3), or the matrix factorization-based methods. LMaFit [99], 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 nuclear norm relaxation model.

Another method based on factorization model (1.3) is called alternating squares (ALS) method which alternatively fixed Y and Z and solves the linear least squares with respect to Y and Z, respectively, see [47, 50] for details.

Matrix factorization is also used to tackle SDP problems. For instance, [8, 52] introduced an equivalent factorization model for SDP through the Cholesky decomposition. Mishra in [68] 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 (1.3) is nonconvex. More specifically, it is a quartic polynomial optimization problem. It may contain exponentially many local minimizers or saddle points. Hence, solving problem (1.3) to the global optimality is usually unachievable.

Recently, Candès and Li in [12] proposed a so-called Wirtinger Flow (WF) method to solve the phase retrieval problem, which is, like (1.3), essentially a quadratic 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 [93] 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 [13] and the random observations similar to [12], the authors of [93] showed their framework can converge to a global solution linearly.

1.1.2 Relation to Matrix Completion

Many interesting models arisen from various application domains requires to seeking an eligible matrix with the lowest possible rank given a partial subset of that matrix, which is the so-called matrix completion problem. For instance, in machine learning scenarios, one is usually required to estimate the missing entries of a low-rank covariance matrix of a process based on partial observations. It was first put forward by the famous online movie-provider company Netflix [2]. The past five years have seen the surge of interest in studying the matrix completion problem due to

its wide applicability in many areas. The model of low-rank matrix completion is also ubiquitous in collaborative filtering, latent semantic analysis and factor analysis [86, 92].

Suppose the coefficients of matrix M are partially revealed as $\Omega \subset \{(i,j)|1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$ with $|\Omega| = p$. And one is required to fill in the unknown entries of M such that the resulting matrix has the as low as possible rank according the the partially observation Ω. The matrix completion problem will solve the following problem

$$\begin{aligned} \text{minimize} \quad & \text{rank}(X) \\ \text{subject to} \quad & X_{ij} = M_{ij}, \ (i,j) \subset \Omega. \end{aligned} \tag{1.4}$$

In many application settings, the entries of Ω are assumed to identically yield certain probability distribution (such as Bernoulli distribution, uniform distribution or Gaussian distribution).

Problem (1.1) includes problem (1.4) as a special case. To see this, it is sufficient to illustrate that the constraint of (1.4) is a special case of that of (1.1). Suppose that E_{ij} is a matrix with all entries equal to 0 except that the (i,j)-th entry is 1. Let $b_t^\Omega = \langle E_{i_t j_t}, M\rangle$ and $A_t^\Omega = E_{i_t j_t}$ for all $t = 1, 2, \cdots, p$ with $p = |\Omega|$. Then the constraint of (1.4) can be reformulated as the form of the constraint of (1.1), that is,

$$\mathcal{A}^\Omega(X) = \mathbf{b}^\Omega,$$

here $\mathcal{A}^\Omega(\cdot) = \left(\langle A_1^\Omega, \cdot\rangle, \ldots, \langle A_t^\Omega, \cdot\rangle, \ldots, \langle A_p^\Omega, \cdot\rangle\right)^{\mathrm{T}}$ and $\mathbf{b}^\Omega = (b_1^\Omega, \ldots, b_t^\Omega, \ldots, b_p^\Omega)^{\mathrm{T}}$.

1.1.3 Relation to Compressed Sensing

Over the past decade, compressed sensing (CS) has attracted considerable attention in communities of applied mathematics, computer science and electrical engineering. We have mentioned that ARMP can be regarded as a natural extension of CS in the

matrix domain. To see this, just let matrix X variable in problem (1.1) be diagonal, the ARMP reduces to finding a sparse vector in an affine subspace, which is exactly the so-called compressed sensing problem.

In compressed sensing, the goal is to reconstruct a sparse or compressible representation $\mathbf{x} \in \mathbb{R}^{n_1}$ from a small set of linear, nonadaptive observations. This process can be represented mathematically as

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{x}\|_0 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}, \end{aligned} \tag{1.5}$$

where $\|\mathbf{x}\|_0 := \#\{k : x_k \neq 0, k = 1, \ldots, n_1\}$ is the number of non-zero entries in $\mathbf{x}$, A is an $n_2 \times n_1$ matrix and $\mathbf{y} \in \mathbb{R}^{n_2}$. The matrix A, which maps $\mathbb{R}^{n_1}$ into $\mathbb{R}^{n_2}$, designates the dimensionality reduction. Usually, n_1 is very large and n_2 is typically much smaller than n_1.

Nevertheless, it is difficult to solve problem (1.5) directly due to the nonconvexity of the objective function from the perspective of both theory and computation. Actually, it is a combinatorial optimization problem and one can show that even finding a approximating solution is NP-hard for a general matrix A [72, 73]. Over the past decade, a variety of heuristic methods have been proposed to solve CS. We refer to [37, 66, 82] for projection pursuit and [22, 30, 81] for orthogonal matching pursuit.

1.2 Partially Separable Minimization Problem

Partially separable optimization was proposed by Griewank and Toint in [45], studied for more than twenty years (see [21, 39, 41, 67, 104] for instance) and extensively used in the popular CUTE(st) testing environment [42] as well as in the LANCELOT [29] and FILTRANE [43] packages, amongst others. In particular, the design of trust-region algorithms exploiting the partially separable decomposition was investigated

by Conn, Gould, Sartenaer and Toint in [28, 27] and Shahabuddin [89].

We here consider the following partially separable convexly constrained nonlinear optimization problem:

$$\underset{\mathbf{x}\in\mathcal{F}}{\text{minimize}} \quad f(\mathbf{x}) = \sum_{i=1}^{m} f_i(U_i\mathbf{x}) = \sum_{i=1}^{m} f_i(x_i) \tag{1.6}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{F}$ is a non-empty closed convex set and for $i \in \{1, \ldots, m\}$, $x_i := U_i\mathbf{x}$ with U_i a (fixed) $n_i \times n$ matrix and $n_i \leqslant n$. Without loss of generality, we assume that the union of the ranges of the U_i^{T} matrices spans $\mathbb{R}$, which can also be equivalently expressed as the intersection of the nullspaces of the U_i is reduced to the origin. The partially separable structure defined in this problem is ubiquitous in applications of optimization. It is most useful in the frequent case where $n_i \ll n$ and subsumes that of sparse optimization (in the special case where the rows of each U_i are selected rows of the identity matrix). Moreover the decomposition in (1.6) has the advantage of being invariant for linear changes of variables (only the U_i matrices vary).

Using the partially separable nature of a function f can be very useful if one wishes to use derivative of f of order larger than one in the context of the p-th order Taylor series

$$T_{f,p}(\mathbf{x}, \mathbf{s}) = f(\mathbf{x}) + \sum_{j=1}^{p} \frac{1}{j!} \nabla_{\mathbf{x}}^{j} f(\mathbf{x})[\mathbf{s}]^j, \tag{1.7}$$

the notation $T[s]^i$ meaning that the tensor T is applied to i copies of the vector $\mathbf{s}$. Indeed, it can be verified that

$$\nabla_{\mathbf{x}}^{1} f(\mathbf{x})[\mathbf{s}] = \sum_{i=1}^{m} (U_i \nabla_{x_i} f_i(x_i))^{\mathrm{T}} \mathbf{s} = \sum_{i=1}^{m} \nabla_{x_i} f_i(x_i)[U_i\mathbf{s}],$$

$$\nabla_{\mathbf{x}}^{2} f(\mathbf{x})[\mathbf{s}]^2 = \sum_{i=1}^{m} \mathbf{s}^{\mathrm{T}} (U_i^{\mathrm{T}} \nabla_{x_i}^{2} f_i(x_i)) \mathbf{s} = \sum_{i=1}^{m} \nabla_{x_i}^{2} f_i(x_i)[U_i\mathbf{s}]^2$$

and, in general, that

$$\nabla^j_{\mathbf{x}} f(\mathbf{x})[\mathbf{s}]^j = \sum_{i=1}^{m} \nabla^j_{x_i} f_i(x_i)[U_i\mathbf{s}]^j \quad (j \geqslant 1).$$

This last expression indicates that only the m tensors $\{\nabla^j_{x_i} f_i(x_i)\}_{i=1}^m$ of dimension n_i^j needs to be computed and stored, a very substantial gain compared to the n^j-dimensional $\nabla^j_{\mathbf{x}} f(\mathbf{x})$ when (as is common) $n_i \ll n$ for all i. It can therefore be argued that exploiting derivative tensors of order larger than 2 — and thus using the high-order Taylor series (1.7) as a local model of $f(\mathbf{x}+\mathbf{s})$ in the neighborhood of $\mathbf{x}$ — is practically feasible if f is partially separable.

Interestingly, the use of high-order Taylor models for optimization was recently investigated by Birgin *et. al.* [6] in the context of adaptive regularization algorithms for unconstrained problems. Their proposal belongs to this emerging class of methods pioneered by Griewank [44], Nesterov and Polyak [77] and Cartis, Gould and Toint [16, 17] for the unconstrained case and by these last authors in [18] for the convexly constrained case of interest here. Such methods are distinguished by their excellent evaluation complexity, in that they need at most $O(\epsilon^{-\frac{p+1}{p}})$ evaluations of the objective function and their derivatives to produce an ϵ-approximate first-order critical point, compared to the $O(\epsilon^{-2})$ evaluations which might be necessary for the steepest descent and Newton's methods (see [15] and [76] for details).

However, most adaptive regularization methods rely on a non-separable regularization term in the model of the objective function, making the exploitation of structure difficult.[3]

[3] The only exception that we are only aware of is the unpublished note [95] in which a p-th order Taylor model is coupled with a regularization term involving the (totally separable) q-th power of the q norm ($q \geqslant 1$).

1.3 Contributions of the Thesis

This thesis can be sharply divided into two distinct parts. The first part includes Chapter 2 and Chapter 3. These two chapters focus on the first class of nonconvex problem, i.e., the Affine Rank Minimization Problem (ARMP). And the second part includes Chapter 4 and Chapter 5, which are devoted to the Partially Separable Minimization Problem (PSMP).

It is worthwhile to mention that the notations in these two distinct parts are totally independent, which can be distinguished without any ambiguity in due courses.

According the structure of the thesis, our contributions are natural to be divided into two different parts.

Throughout Chapter 2 and Chapter 3, we mainly discuss problem (1.3). We observed that some local optimal solvers can often find a global solution of problem (1.3) by starting from a randomly chosen initial guess even if the linear operator $\mathcal{A}$ of (1.3) does not involve any stochastic property. One main contribution of this thesis is to theoretically investigate the relationship between the global optimality of problem (1.3) and its second-order optimality under certain scenarios, which can partly explain the above mentioned phenomenon.

In order to study this phenomenon, we introduce a concept call "SNIG" condition, which is the abbreviation of "Second-order Necessary optimality Implies Global optimality". Its exact definition will be introduced in the coming Chapter 2 (see Definition 2.1). Based on SNIG condition, the contribution of the first part can generalized as follows:

- we find several scenarios where the SNIG condition holds and prove it, which includes
  - the SNIG condition always holds over the cone consisting of all the rank

deficient points in $\mathbb{R}^{n\times m} \times \mathbb{R}^{n\times m}$,

- the SNIG condition holds when the number of observation entries is equal to the total number of unknowns,
- the SNIG condition holds if the operator $\mathcal{A}$ enjoys the special structure mentioned in [105], more specifically, $\mathcal{A}$ maps a matrix to a part of it and the missing part is a block of the matrix;

- we propose a framework which utilizes the SNIG condition to solve the affine rank minimization problem and some numerical examples are also presented.

Chapter 4 and Chapter 5 then focus on the partially separable minimization problem (1.6). In this part, the contributions are twofold:

- we first introduce an algorithm using a partially separable regularization term and to show that its evaluation complexity retains the excellent bound of $O(\epsilon^{-(p+1)/p})$ evaluations while exploiting structure;
- we then extend the algorithm and complexity to the (non-Lipschitz) problem of data fitting involving the so-called q-quasi norm for $q \in (0, 1)$. Such problems were already investigated by Bian, Chen and Ye in [3, 4, 5] in the context of trust-region methods.

1.4 Organization of the Thesis

The thesis is structured as follows.

- Chapter 2 first gives the exact definition of the so-called SNIG condition, then present some preliminaries and finally three different scenarios where the SNIG condition can hold.

- Chapter 3 focuses on the algorithm framework for solving the affine rank minimization problem (1.1), which is based on the matrix factorization model (1.3) using the property of SNIG condition.

- Chapter 4 is devoted on the partially separable algorithm with cubic regularization and the complexity analysis of function value evaluations for this algorithm.

- Chapter 5 extends the partially separable algorithm and its complexity of function value evaluations to the non-Lipshchitz problem of data fitting with the q-quasi norm for $q \in (0, 1)$.

- Chapter 6 ends the whole thesis with some concluding remarks.

Chapter 2

Theoretical Analysis of SNIG

The purpose of this chapter is to discuss the "Second-order Necessary optimality Implies Global optimality" (SNIG) condition and the theoretical results when using the SNIG condition to analyze problem (1.3). We first begin with the exact definition of the SNIG condition. Then we present some preliminaries which will be used in the later theoretical analysis, which include the concepts of the adjoint operator of a linear operator, Fréchet differentiability in abstract space and permutation matrix. The first two concepts are both of importance in deducing the optimality conditions. And both of them are well-known results which can be found in any standard textbook for functional analysis and optimization theory ([88] and [7] for instance). The third one serve as an essential role in our theoretical analysis. Subsequently, we introduce the optimality conditions of problem (1.3). Finally we manifest three distinct scenarios where the SNIG condition holds and end up with this chapter with an example in which the SNIG condition fails to hold.

Problem (1.3) is essentially a Nonlinear Least Squares problem. For the sake of convenience, we abbreviate it as (NLS-k) and restate it below. That is, consider the following unconstrained matrix factorization problem

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

where $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}} \in \mathbb{R}^p$ is a column vector, $\mathcal{A} \in \mathcal{B}(\mathbb{R}^{n\times m}, \mathbb{R}^p)$ is a bounded linear operator mapping $n \times m$ matrices to p-dimensional Euclidean space which is defined specifically by

$$\mathcal{A}(X) = (\langle A_1, X\rangle, ..., \langle A_p, X\rangle)^{\mathrm{T}}, \tag{2.1}$$

with $A_i \in \mathbb{R}^{n\times m}$ $(i = 1, ..., p)$ are the p column matrices of $\mathcal{A}$. And $\langle W_1, W_2\rangle := \mathrm{tr}(W_1^{\mathrm{T}} W_2)$ designates the inner product of two matrices W_1 and W_2 with the same size.

2.1 SNIG: Second-order Necessary optimality Implies Global optimality

Clearly, the Nonlinear Least Squares problem (1.3), i.e., problem (NLS-k) is bounded from below. We first introduce the following notation for the sake of simplicity. Denote by

$$\Pi^{(n,m,p)} := \mathcal{B}(\mathbb{R}^{n\times m}, \mathbb{R}^p) \cap \{\mathcal{A} \mid \mathcal{A}^{\mathrm{T}}(\mathbf{c}) \neq 0,\ \forall \mathbf{c} \neq 0, \mathbf{c} \in \mathbb{R}^p\}$$

the set consisting of all linear bounded operators satisfying Assumption 1.1.

Further, we present another another assumption as follows.

Assumption 2.1. *Problem* (NLS-k) *has zero residual solution.*

By "zero residual solution" here, we mean those solutions in which their corresponding optimal values vanish. It is clear that the global optimality of (NLS-k) becomes checkable: $f(Y, Z) = 0$ once Assumption 2.1 holds. For the convenience of our analysis, throughout this thesis, we also assume Assumption 2.1 holds.

Let $K := \{1, 2, \cdots, \min(n, m)\}$. Note that once a triplet $(\mathcal{A}, \mathbf{b}, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ is given, a specific instance of problem (NLS-k) is immediately determined. For convenience, once we say $(\mathcal{A}, \mathbf{b}, k)$ satisfies Assumption 2.1, it refers to the fact

that problem (NLS-k) satisfies Assumption 2.1. Moreover, if $(\mathcal{A}, \mathbf{b}, k)$ satisfies Assumption 2.1, there must exist at least one rank-k matrix $W \in \mathbb{R}^{n\times m}$ such that $\mathcal{A}(W) = \mathbf{b}$. Denote

$$r^* = \min_W \{\text{rank}(W) : \mathcal{A}(W) = b\},$$

we have $r^* \leqslant k \leqslant \min(n, m)$.

Although solving the nonlinear least squares problem (NLS-k) to the global optimality is NP-hard in general, obtaining a second-order stationary point can be achieved in polynomial time and there is no gap between them in quite some scenarios.

Definition 2.1. *Given a triplet $(\mathcal{A}, \mathbf{b}, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ satisfying Assumption 2.1. Let $\mathcal{C}$ be a subset of $\mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$. Then, if for any $(Y, Z) \in \mathcal{C}$ satisfying the second-order necessary optimality condition of (NLS-k), (Y, Z) is a global optimizer of (NLS-k), we call the SNIG (Second-order Necessary optimality Implies Global optimality) condition holds at the triplet $(\mathcal{A}, \mathbf{b}, k)$ over $\mathcal{C}$. Particularly, if $\mathcal{C} = \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$, we say the SNIG condition holds at the triplet $(\mathcal{A}, \mathbf{b}, k)$.*

2.2 Preliminaries

Denote by $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ three finite-dimensional real Hilbert spaces. And $\mathcal{X}'$ designates the dual space of $\mathcal{X}$.

Adjoints in Hilbert Space

Suppose that $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ is the set consisting of all bounded linear operators mapping $\mathcal{X}$ to $\mathcal{Y}$ and $\mathcal{A} \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$. Let $y \in \mathcal{Y}$ be a fixed vector, and consider the following functional:

$$g_y : \mathcal{X} \to \mathbb{R}, \text{ where } g_y := \langle \mathcal{A}x, y\rangle.$$

The the functional g_y is

- linear, since both $\mathcal{A}$ and $\langle \cdot, y \rangle$ are linear;
- bounded, since $|g_y(x)| \leqslant \|\mathcal{A}x\|_{\mathcal{Y}} \|y\|_{\mathcal{Y}} \leqslant \|\mathcal{A}\| \|x\|_{\mathcal{X}} \|y\|_{\mathcal{Y}}$ with $\mathcal{A}$ bounded.

In other words, $g_y \in \mathcal{X}'$.

Proposition 2.1 (Riesz Representation Theorem)**.** *If $\mathcal{H}$ is a Hilbert space and $g : \mathcal{H} \to \mathbb{R}$ is a linear and bounded functional. Then there exist a unique $y \in \mathcal{H}$ such that*

$$g(x) = \langle x, y \rangle, \ \forall x \in H.$$

Proof. Readers can find the proof in any standard textbook concerning functional analysis, see [88] for instance. □

Definition 2.2. *By Riesz representation theorem, for each $y \in \mathcal{Y}$ there exists a unique $z = z_y \in \mathcal{X}$ such that*

$$g_y(x) = \langle x, z_y \rangle, \ \forall x \in \mathcal{X}.$$

So we can define legitimately a mapping $\mathcal{A}^{\mathrm{T}} : \mathcal{Y} \to \mathcal{X}$, called the adjoint of $\mathcal{A}$, by the relationship $z_y = \mathcal{A}^{\mathrm{T}}(y)$.

The defining property of $\mathcal{A}^{\mathrm{T}}$ is then:

$$\langle \mathcal{A}x, y \rangle_{\mathcal{Y}} = \langle x, \mathcal{A}^{\mathrm{T}}(y) \rangle_{\mathcal{X}}. \tag{2.2}$$

Moreover, $\mathcal{A}^{\mathrm{T}}$ further has the following properties:

- $\mathcal{A}^{\mathrm{T}}$ is the only mapping from $\mathcal{Y}$ to $\mathcal{X}$ that satisfies (2.2);
- $\mathcal{A}^{\mathrm{T}}$ is linear and bounded, i.e., $\mathcal{A}^{\mathrm{T}} \in \mathcal{B}(\mathcal{Y}, \mathcal{X})$.

The *range* of $\mathcal{A}$ is denoted by

$$R(\mathcal{A}) := \{y \in \mathcal{Y} | y = \mathcal{A}(x) \text{ for } x \in \mathcal{X}\},$$

and the *nullspace* of $\mathcal{A}$ is denoted by

$$N(\mathcal{A}) := \{x \in \mathcal{X} | \mathcal{A}(x) = 0\}.$$

For a linear operator $\mathcal{A} \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, we have that $N(\mathcal{A}) \subset \mathcal{X}$, $R(\mathcal{A}) \subset \mathcal{Y}$ and $\{R(\mathcal{A})\}^{\perp} = N(\mathcal{A}^{\mathrm{T}})$.

Definition 2.3. *If $R(\mathcal{A}) = \mathcal{Y}$, then $\mathcal{A}$ is called full row rank.*

Proposition 2.2. *Suppose that $\mathbf{y} = (y_1, y_2, \ldots, y_p)^{\mathrm{T}} \in \mathbb{R}^p$ and $\mathcal{A} \in \mathcal{B}(\mathbb{R}^{n\times m}, \mathbb{R}^p)$ defined by (2.1). Then the adjoint of $\mathcal{A}$ can be written explicitly, namely,*

$$\mathcal{A}^{\mathrm{T}}(\mathbf{y}) = \sum_{i=1}^{p} y_i A_i. \tag{2.3}$$

Proof. Since $\mathcal{A}$ is of full row rank, for any $\mathbf{y} \in \mathbb{R}^p$, there must exist at least one $X \in \mathbb{R}^{n\times m}$ such that $\mathbf{y} = \mathcal{A}(X)$, i.e., $y_i = \langle A_i, X\rangle$ for all $i = 1, 2, \cdots, p$.

From (2.2), it follows that

$$\langle Z, \mathcal{A}^{\mathrm{T}}(\mathbf{y})\rangle = \langle \mathcal{A}(Z), \mathbf{y}\rangle = \langle \mathcal{A}(Z), \mathcal{A}(X)\rangle, \ \forall Z \in \mathbb{R}^{n\times m}. \tag{2.4}$$

On the other hand,

$$\langle Z, \sum_{i=1}^{p} y_i A_i\rangle = \sum_{i=1}^{p} y_i \langle Z, A_i\rangle = \langle \mathcal{A}(Z), \mathcal{A}(X)\rangle, \ \forall Z \in \mathbb{R}^{n\times m}. \tag{2.5}$$

Combining (2.4) and (2.5), we have that for any $Z \in \mathbb{R}^{n\times m}$, there holds

$$\langle Z, \mathcal{A}^{\mathrm{T}}(\mathbf{y})\rangle = \langle Z, \sum_{i=1}^{p} y_i A_i\rangle,$$

which implies that

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

The uniqueness of $\mathcal{A}^{\mathrm{T}}(\mathbf{y})$ is directly follows from Riesz representation theorem. The proof is completed.

□

Fréchet Derivatives

To begin with, we first give the definition of Fréchet differentiability as follows. For more details of differentiability in abstract spaces, we refer readers to [7, Section 2.2].

Definition 2.4. *A mapping $\Psi : \mathcal{X} \to \mathcal{Y}$ is said to be Fréchet-differentiable at $x \in \mathcal{X}$ if there exists a linear and continuous operator, denoted by $\mathcal{J}\Psi(x)$ such that*

$$\Psi(x + \Delta x) - \Psi(x) - \mathcal{J}\Psi(x)[\Delta x] = o(\|\Delta x\|), \quad \Delta x \in \mathcal{X}.$$

Suppose that $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is a mapping. If F is Fréchet-differentiable at $(x, y) \in (\mathcal{X}, \mathcal{Y})$, then we use $\mathcal{J}F(x, y)$ to denote the Fréchet derivative of F at (x, y) and $\mathcal{J}_x F(x, y)$ to denote the partial Fréchet derivative of F at (x, y) with respect to x.

Let the gradient of F

$$\nabla F(x, y) := \mathcal{J}F(x, y)^{\mathrm{T}}$$

be the adjoint of $\mathcal{J}F(x, y)$ (respectively, $\nabla_x F(x, y) := \mathcal{J}_x F(x, y)^{\mathrm{T}}$, the adjoint of $\mathcal{J}_x F(x, y)$). If F is further twice Fréchet differentiable at $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we define

$$\mathcal{J}^2 F(x, y) := \mathcal{J}(\mathcal{J}F)(x, y),$$

$$\mathcal{J}^2_{xx} F(x, y) := \mathcal{J}_x(\mathcal{J}_x F)(x, y),$$

$$\nabla^2 F(x, y) := \mathcal{J}(\nabla F)(x, y),$$

$$\nabla^2_{xx} F(x, y) := \mathcal{J}_x(\nabla_x F)(x, y).$$

Permutation Matrix

In this section, we give the definition of permutation matrix and list some properties of it without proof, which will act an important role in the later theoretical analysis.

Definition 2.5. *A matrix $P \in \mathbb{R}^{n \times n}$ is called a permutation matrix if every row and every column contains one* 1 *and zeros otherwise.*

Then we give some properties of the permutation matrix which may be useful in the following analysis.

Proposition 2.3. *The following statements hold:*

1. *If P is a permutation matrix, then P^{T} is also a permutation matrix.*

2. *If $P \in \mathbb{R}^{n\times n}$ is a permutation matrix, then*

$$(P^{\mathrm{T}}P)_{ij} = \sum_{k=1}^{n} P_{ki}P_{kj} = \delta_{ij}$$

   with δ_{ij} the Kronecker delta, and thus $P^{\mathrm{T}}P = I$ which implies that permutation matrix is orthogonal.

3. *If $\pi : \{1, 2, \cdots, n\} \to \{1, 2, \cdots, n\}$ is a permutation, given in two-line form by*

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \cdots & \pi(n) \end{pmatrix}$$

   then the matrix $P = (p_{ij})$ with

$$p_{ij} = \begin{cases} 1, & \text{if } j = \pi(i) \\ 0, & \text{otherwise} \end{cases}$$

   is a permutation matrix. Indeed, every permutation matrix is of this form. Particularly, identity matrix is a permutation matrix.

4. *If P_π is the permutation matrix corresponding to the permutation π, then $(P_\pi^{-1})_{ij} = 1$ if and only if $j = \pi(i)$. Hence the permutation matrix P_π^{-1} corresponds to the permutation π^{-1}. Hence, we have that*

$$P_\pi^{-1} = P_{\pi^{-1}} = P_\pi^{\mathrm{T}}.$$

5. *If* $P_\pi = (p_{ij}) \in \mathbb{R}^{n\times n}$ *and* $Q_\sigma = (q_{ij}) \in \mathbb{R}^{m\times m}$ *are two permutation matrices respectively corresponding to permutation* π *and* σ*, then let* $W = (w_{ij})$ *be a* $n \times m$ *matrix, we have*

$$(P_\pi W)_{ij} = \sum_{k=1}^{n} p_{ik} w_{kj} = W_{\pi(i)j}$$

for all $(i, j) \in [n] \times [m]$*. This shows that multiplying a permutation matrix from left reorders the rows of* W*. Moreover, we have*

$$(WQ_\sigma)_{ij} = \sum_{k=1}^{m} w_{ik} q_{kj} = W_{i\sigma^{-1}(j)}$$

and hence multiplying a permutation matrix from the right reorders the columns of W*.*

2.3 Optimality Conditions

With the above preparation in hand, we are now ready to present the optimality condition of problem (NLS-k). In this section, we first give the optimality conditions of formulation (NLS-k) and then take a further look at the SNIG condition.

2.3.1 First- and Second-order Optimality Conditions

According to the discussion on Fréchet derivative in the above Subsection 2.2, it is not difficult to know that the gradient of the objective function in problem (NLS-k) can be expressed as

$$\nabla f(Y, Z) = \begin{bmatrix} \nabla_Y f(Y, Z) \\ \nabla_Z f(Y, Z) \end{bmatrix},$$

where

$$\nabla_Y f(Y, Z) = \mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}}) - \mathbf{b})Z;$$

$$\nabla_Z f(Y, Z) = (\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}}) - \mathbf{b}))^{\mathrm{T}}Y.$$

Here $\mathcal{A}^{\mathrm{T}} : \mathbb{R}^p \to \mathbb{R}^{n\times m}$ denotes the adjoint operator of $\mathcal{A}$. It follows from Proposition 2.2 that for any $\mathbf{y} = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$

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

Furthermore, the Hessian of $f(Y, Z)$ can be expressed as

$$\nabla^2 f(Y,Z) \quad = \quad \begin{bmatrix} \nabla^2_{YY} f(Y,Z) & \nabla^2_{YZ} f(Y,Z) \\ \nabla^2_{ZY} f(Y,Z) & \nabla^2_{ZZ} f(Y,Z) \end{bmatrix}, \tag{2.6}$$

where

$$\nabla^2_{YY} f(Y,Z)[S_Y] = \mathcal{A}^{\mathrm{T}}(\mathcal{A}(S_Y Z^{\mathrm{T}}))Z; \tag{2.7a}$$

$$\nabla^2_{YZ} f(Y,Z)[S_Z] = \mathcal{A}^{\mathrm{T}}(\mathcal{A}(Y S_Z^{\mathrm{T}}))Z + \mathcal{A}^{\mathrm{T}}((\mathcal{A}(Y Z^{\mathrm{T}}) - b))S_Z; \tag{2.7b}$$

$$\nabla^2_{ZY} f(Y,Z)[S_Y] = (\mathcal{A}^{\mathrm{T}}(\mathcal{A}(S_Y Z^{\mathrm{T}})))^{\mathrm{T}} Y + (\mathcal{A}^{\mathrm{T}}(\mathcal{A}(Y Z^{\mathrm{T}}) - b)))^{\mathrm{T}} S_Y; \tag{2.7c}$$

$$\nabla^2_{ZZ} f(Y,Z)[S_Z] = (\mathcal{A}^{\mathrm{T}}(\mathcal{A}(Y S_Z^{\mathrm{T}})))^{\mathrm{T}} Y, \tag{2.7d}$$

for all $S_Y \in \mathbb{R}^{n\times k}$ and $S_Z \in \mathbb{R}^{m\times k}$.

Since (NLS-k) is s a twice continuously differentiable unconstrained optimization problem, we can directly give its first-order and second-order necessary optimality condition as follows, respectively.

Definition 2.6. *A pair of matrices $(Y^*, Z^*) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ is called a stationary point of (NLS-k) if $\nabla f(Y^*, Z^*) = 0$.*

Proposition 2.4. *Let (Y^*, Z^*) be a local minimizer of (NLS-k), then it must be a stationary point and $\nabla^2 f(Y^*, Z^*)$ is positive semi-definite. Namely,*

$$||\mathcal{A}(Y^* S_Z^{\mathrm{T}} + S_Y (Z^*)^{\mathrm{T}})||_2^2 + 2\mathrm{tr}(S_Y^{\mathrm{T}} \mathcal{A}^{\mathrm{T}}(\mathcal{A}(Y^*(Z^*)^{\mathrm{T}} - UV^{\mathrm{T}}))S_Z) \geqslant 0, \tag{2.8}$$

for all $S_Y \in \mathbb{R}^{n\times k}$, $S_Z \in \mathbb{R}^{m\times k}$ where (U, V) is a fixed point satisfying $\mathcal{A}(UV^{\mathrm{T}}) = \mathbf{b}$.

Proof. The first part directly follows from the first-order optimality condition. Next, we give the proof of the second part. The Hessian $\nabla^2 f(Y^*, Z^*)$ is positive semi-definite due to the second-order optimality necessary condition. We derive from the positive semi-definiteness of $\nabla^2 f(Y^*, Z^*)$ and the relation (2.6) that

$$\begin{aligned} &\langle S_Y, \nabla^2_{YY} f(Y^*, Z^*)[S_Y]\rangle + \langle S_Y, \nabla^2_{YZ} f(Y^*, Z^*)[S_Z]\rangle \\ &+\langle S_Z, \nabla^2_{ZY} f(Y^*, Z^*)[S_Y]\rangle + \langle S_Z, \nabla^2_{ZZ} f(Y^*, Z^*)[S_Z]\rangle \geqslant 0 \end{aligned} \tag{2.9}$$

holds for all $S_Y \in \mathbb{R}^{n\times k}$ and $S_Z \in \mathbb{R}^{m\times k}$. Substituting relations (2.7) into (2.9), we can obtain (2.8) which completes the proof. □

Definition 2.7. *A pair of matrices $(Y^*, Z^*) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ is called a second-order stationary point of* (NLS-k)*, if it is a stationary point and the second-order necessary optimality condition* (2.8) *holds for all $S_Y \in \mathbb{R}^{n\times k}$, $S_Z \in \mathbb{R}^{m\times k}$.*

Obviously, a local minimizer must be a second-order stationary point, but not necessarily vice versa.

2.3.2 A Further Look at the SNIG Condition

Now let us back to the SNIG condition. Our original motivation is actually to find the solution of problem (1.1) without heavily involving singular value decomposition, which will prohibitively expensive from the perspective of computation when handling large-scale problems.

Actually, to obtain formulation (1.3) or (NLS-k), the affine rank minimization problem is first cast as the following Linear Least Squares problem with a rank constraint:

$$\begin{aligned} \underset{X\in\mathbb{R}^{n\times m}}{\text{minimize}} \quad & e_k(X) := \|\mathcal{A}(X) - \mathbf{b}\|_2^2 \\ \text{subject to} \quad & \operatorname{rank}(X) \leqslant k, \end{aligned} \tag{LLS-k}$$

where k is an positive integer. For brevity, we denote this Linear Least Squares problem as (LLS-k).

But the global minimizer of (LLS-k) with a fixed k is not attainable in practice due to the nonsmooth and nonconvex rank constraint. So we use matrix factorization to embed such rank constraint into the objective and hence consider problem (NLS-k). (NLS-k) is still a nonconvex problem that may exist many local minimizers or saddle points, which the gap between global minimizer and local minimizer are nonzero.

Based on quite a few preliminary analyses, both theoretically and numerically, we observed that there exist a lot of scenarios where the gap between the global minimizer and the second-order stationary point is zero. The SNIG condition actually focus on the gap between the global minimizer and the second-order stationary point of (NLS-k). And it extracts the scenarios where the gap between the second-order stationary point and the global optimizer vanishes. It somewhat resembles the convexity which rule out the cases at which the local minimizers and the global minimizer has no gap. And hence, for any problem instance with triplet satisfying the SNIG condition, its global minimizer can be attained by invoking any local optimization method which terminates at a second-order stationary point, which is the basic idea of our propose algorithm framework in Chapter 3.

2.4 Three Scenarios where the SNIG Condition Holds

The main purpose of this section is to present our theoretical results, i.e., three scenarios of triplet $(\mathcal{A}, \mathbf{b}, k)$ under which the SNIG condition holds. We first give the special case of rank deficient second-order stationary point and then the case in which the operator $\mathcal{A}$ is a bijection. Finally, we focus on the case that the operator is a projection mapping taking a special structure like the structure investigated in [105].

A simple example is first presented to illustrate the idea of our proof techniques. Then we give the main result. Subsequently, a special case is analyzed carefully and detailedly. The ultimate conclusion is achieved by permutating the corresponding projection matrix with respect to the aforementioned special case.

2.4.1 Rank Deficient Second-order Stationary Point

In this section, we show that the SNIG condition holds at a rather general scenario of triplet $(\mathcal{A}, \mathbf{b}, k)$ over a particular subset of $\mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$.

Theorem 2.1. *Let $(\mathcal{A}, \mathbf{b}, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ be a triplet satisfying Assumption 2.1. Then the SNIG condition holds at the triplet $(\mathcal{A}, \mathbf{b}, k)$ over the set*

$$\bar{\mathcal{C}} := \{(X, Y) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k} \mid \textit{at least one of } X \textit{ and } Y \textit{ is rank deficient}\}.$$

Proof. Let $(Y, Z) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ be any second-order stationary point of (NLS-k) satisfying that at least one of Y and Z is rank deficient.

Denote (U, V) as a global optimizer of (NLS-k). It follows from Assumption 2.1 and $k \geqslant r^*$ that $f(U, V) = 0$. Namely, $\mathcal{A}(UV^{\mathrm{T}}) = \mathbf{b}$ holds. Therefore we can rewrite the objective function of (NLS-k) as

$$f(Y, Z) = \frac{1}{2}||\mathcal{A}(YZ^{\mathrm{T}}) - \mathcal{A}(UV^{\mathrm{T}})||_2^2. \tag{2.10}$$

Without loss of generality, we assume that Y is rank deficient, i.e., there exists a nonzero vector $\tilde{y} \in \mathbb{R}^k$ satisfying $Y\tilde{y} = 0$. Without loss of generality, we assume $\tilde{y}_l \neq 0$ for some $l \in \{1, 2, \ldots, k\}$. We prove the conclusion by contradiction. Suppose there exists $(s, t) \in \{(i, j) \mid 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m\}$ satisfying $\eta = (\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})))_{st} \neq 0$.

Then, we set $S_Y \in \mathbb{R}^{n\times k}$ as follows

$$(S_Y)_{i_1 i_2} = \begin{cases} \eta, & \text{if } i_1 = s \text{ and } i_2 = l, \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

Let $S_Z = \tilde{z}\tilde{y}^{\mathrm{T}}$, and set $\tilde{z} \in \mathbb{R}^m$ as follows

$$\tilde{z}_j = \begin{cases} -\xi\tilde{y}_l, & \text{if } j = t, \\ 0, & \text{otherwise}, \end{cases} \tag{2.12}$$

where $\xi > ||\mathcal{A}(r_s z_l^{\mathrm{T}})||_2^2/2\tilde{y}_l^2$. Here $r_s \in \mathbb{R}^n$ is the s-th column of identity matrix I_n and z_l denotes the l-th column of Z.

Recall Proposition 2.4, we have

$$||\mathcal{A}(YS_Z^{\mathrm{T}} + S_Y Z^{\mathrm{T}})||_2^2 + 2\mathrm{tr}(S_Y^{\mathrm{T}}\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}))S_Z) \geqslant 0, \tag{2.13}$$

for all $S_Y \in \mathbb{R}^{n\times k}$, $S_Z \in \mathbb{R}^{m\times k}$ and $||S_Y||_{\mathrm{F}}^2 + ||S_Z||_{\mathrm{F}}^2 \neq 0$.

Plugging (2.11) and (2.12) into (2.13), we obtain

$$\begin{aligned} & ||\mathcal{A}(YS_Z^{\mathrm{T}} + S_Y Z^{\mathrm{T}})||_2^2 + 2\mathrm{tr}(S_Y^{\mathrm{T}}\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}))S_Z) \\ = \; & ||\mathcal{A}(Y\tilde{y}\tilde{z}^{\mathrm{T}} + S_Y Z^{\mathrm{T}})||_2^2 + 2\mathrm{tr}(S_Y^{\mathrm{T}}\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}))\tilde{z}\tilde{y}^{\mathrm{T}}) \\ = \; & ||\mathcal{A}(S_Y Z^{\mathrm{T}})||_2^2 + 2\mathrm{tr}(\tilde{y}^{\mathrm{T}}S_Y^{\mathrm{T}}\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}))\tilde{z}) \\ = \; & \eta^2||\mathcal{A}(r_s z_l^{\mathrm{T}})||_2^2 - 2\xi\eta^2\tilde{y}_l^2 < 0. \end{aligned}$$

Hence, the second-order necessary optimality condition is violated, which is contrary to the fact that (Y, Z) is a second-order stationary point. Therefore $\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})) = 0$, which implies $\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}) = 0$ due to the full rankness of $\mathcal{A}$. Namely, $f(Y, Z) = 0$. We complete the proof. □

Theorem 2.1 tells us that for any (NLS-k) with zero-residual, any rank deficient second-order stationary point is a global minimizer. In this chapter, we show this result and utilize it to prove our subsequent results in the following two sections. Actually, this result might be much more useful than we have explored here. Specifically speaking, if we can design an algorithm which always terminates at rank deficient second-order stationary point if the problem has, then we can use Theorem 2.1 to directly check the global optimality.

2.4.2 The Scenario When $\mathcal{A}$ is a Special Bijection

In this section, we consider the case that $p = nm$ and a special case of affine constraint $\mathcal{A}(X) = \mathbf{b}$, namely

$$C^{\mathrm{T}} X D = B$$

where $C \in \mathbb{R}^{n\times n}$, $D \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times m}$ are given.

Consider $p = nm$ and the special class of triplet $(\mathcal{A}^{(C,D)}, \mathbf{b}^{(B)}, k)$ where

$$A^{(C,D)}_{i+(j-1)n} = C_i D_j^{\mathrm{T}} \quad \text{and} \quad b^{(B)}_{i+(j-1)n} = B_{i,j}, \quad i = 1, ..., n,\ j = 1, ..., m, \tag{2.14}$$

with C_i and D_j the i-th column of C and j-th column of D, respectively. Then it is not difficult to verify that

$$\|\mathcal{A}^{(C,D)}(X) - b^{(B)}\|_2^2 = \|C^{\mathrm{T}} X D - B\|_{\mathrm{F}}^2. \tag{2.15}$$

C and D are Identity Matrices

When $C = I_n$ and $D = I_m$ where I_q is the q-dimensional identity matrices, the function of the operator $\mathcal{A}^{(I_n,I_m)}$ is to select all elements of a given matrix and stack them column by column as a column vector. The the triplet $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}, k)$ satisfies the SNIG condition. Intuitively, this is right without doubt since we of course can reconstruct a matrix by observing all coefficients of that matrix. But in order to abstract the idea and methodology, we begin our investigation with such simply and obviously special case.

Lemma 2.1. *Suppose that the triplet $(\mathcal{A}^{(I_n,I_m)}, b^{(B)}, k)$ defined by (2.14) with $C = I_n$ and $D = I_m$ satisfies Assumption 2.1. Then the SNIG condition holds at the triplet $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}^{(B)}, k)$.*

Proof. Suppose that the triplet $(\mathcal{A}^{(I_n,I_m)}, b^{(B)}, k)$ defined by (2.14) with $C = I_n$ and $D = I_m$ satisfies Assumption 2.1 that $k \geqslant \mathrm{rank}(B)$ and there exist $U \in \mathbb{R}^{n\times k}$ and

$V \in \mathbb{R}^{m\times k}$ such that $\mathbf{b}^{(B)} = \mathcal{A}^{(I_n,I_m)}(UV^{\mathrm{T}})$. Namely, (U,V) is a global optimizer of (NLS-k) corresponding to the triplet $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}^{(B)}, k)$. By (2.15), the objective function of (NLS-k) can be reformulated as

$$f(Y,Z) = \frac{1}{2}\|YZ^{\mathrm{T}} - B\|_{\mathrm{F}}^2.$$

Assume that (Y,Z) is a second-order stationary point of problem (NLS-k) with respect to $(\mathcal{A}^{(I_n,I_m)}, b^{(B)}, k)$. The first-order optimality condition can be written as

$$\begin{cases} Y^{\mathrm{T}}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}) = \mathbf{0}, \\ (YZ^{\mathrm{T}} - UV^{\mathrm{T}})Z = \mathbf{0}. \end{cases} \tag{2.16}$$

After rearranging, we have

$$\begin{cases} Y^{\mathrm{T}}YZ^{\mathrm{T}} = Y^{\mathrm{T}}UV^{\mathrm{T}}, \\ YZ^{\mathrm{T}}Z = UV^{\mathrm{T}}Z. \end{cases} \tag{2.17}$$

Recall Theorem 2.1, the triplet $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}, k)$ satisfies the SNIG condition if either Y or Z is rank deficient. So we now only need to consider the case that both of Y and Z are of full rank, which implies that both $Y^{\mathrm{T}}Y$ and $Z^{\mathrm{T}}Z$ are nonsingular. Then it directly from (2.17) that Y and Z have the following expressions

$$\begin{cases} Y = UV^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}, \\ Z = ((Y^{\mathrm{T}}Y)^{-1}Y^{\mathrm{T}}UV^{\mathrm{T}})^{\mathrm{T}}. \end{cases} \tag{2.18}$$

The relation (2.17) also gives us the following equality

$$Y^{\mathrm{T}}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})Z = 0, \tag{2.19}$$

which can be reformulated as

$$Y^{\mathrm{T}}YZ^{\mathrm{T}}Z = Y^{\mathrm{T}}UV^{\mathrm{T}}Z. \tag{2.20}$$

Then $Y^{\mathrm{T}}U$ and $V^{\mathrm{T}}Z$ are also nonsingular which follows from (2.20) and the non-singularity of $Y^{\mathrm{T}}Y$ and $Z^{\mathrm{T}}Z$.

Hence, the following relation holds from (2.20)

$$(Z^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}Y)^{-1} = (V^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}U)^{-1}. \tag{2.21}$$

Combining (2.18) and (2.21), we obtain

$$YZ^{\mathrm{T}} - UV^{\mathrm{T}} = U(V^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}Y)^{-1}Y^{\mathrm{T}}U - I)V^{\mathrm{T}} = 0.$$

Namely, $f(Y, Z) = 0$. We complete the proof. □

C and D are Diagonal Matrices

The same conclusion of Lemma 2.1 also holds when C and D are both nonsingular diagonal matrices. To see this, suppose that

$$C = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & c_n \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & d_n \end{bmatrix},$$

with $c_i \neq 0$ $(\forall i = 1, \ldots, n)$ and $d_j \neq 0$ $(\forall j = 1, \ldots, m)$, then

$$\begin{aligned} \mathcal{A}^{(C,D)}(X) &= \left(\langle A_1^{(C,D)}, X\rangle, \ldots, \langle A_n^{(C,D)}, X\rangle, \ldots, \langle A_{1+(m-1)n}^{(C,D)}, X\rangle, \ldots, \langle A_{nm}^{(C,D)}, X\rangle\right)^{\mathrm{T}} \\ &= \left(C_1^{\mathrm{T}}XD_1, \ldots, C_n^{\mathrm{T}}XD_1, \ldots, C_1^{\mathrm{T}}XD_m, \ldots, C_n^{\mathrm{T}}XD_m\right)^{\mathrm{T}} \\ &= \left(c_1d_1\mathbf{e}_1^{\mathrm{T}}X_1, \ldots, c_nd_1\mathbf{e}_n^{\mathrm{T}}X_m, \ldots, c_1d_m\mathbf{e}_1^{\mathrm{T}}X_m, \ldots, c_nd_m\mathbf{e}_n^{\mathrm{T}}X_m\right)^{\mathrm{T}}, \end{aligned} \tag{2.22}$$

where $\mathbf{e}_i$ is the i-th column of identity matrix I_n for $i = 1, \ldots, n$ and X_j is the j-th column of matrix X for $j = 1, \ldots, m$.

It follows from Proposition 2.2 and (2.22) that

$$\begin{aligned} (\mathcal{A}^{(C,D)})^{\mathrm{T}}(\mathcal{A}^{(C,D)}(X)) &= \sum_{i=1}^{n}\sum_{j=1}^{m}(c_id_j\mathbf{e}_i^{\mathrm{T}}X_m)C_iD_j^{\mathrm{T}} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m}(c_i^2d_j^2\mathbf{e}_i^{\mathrm{T}}X_m)E_{ij} \\ &= C^2XD^2, \end{aligned}$$

where E_{ij} is a n by m matrix with the ij-th position equal to 1 and 0 otherwise, $C^2 = \mathrm{Diag}(c_1^2, \ldots, c_n^2)$ and $D^2 = \mathrm{Diag}(d_1^2, \ldots, d_m^2)$ are both diagonal matrices.

Then the first-order optimality condition of (NLS-k) with respect to $(\mathcal{A}^{(C,D)}, \mathbf{b}, k)$ can expressed as (for simplicity, we below use $\mathcal{A}$, C and D to denote $\mathcal{A}^{(C,D)}$, C^2 and D^2 in the analysis)

$$\begin{cases} \mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}))Z = C(YZ^{\mathrm{T}}D - CUV^{\mathrm{T}})DZ = (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}})Z = 0 \\ Y^{\mathrm{T}}(\mathcal{A}^{\mathrm{T}}(\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})) = Y^{\mathrm{T}}C(YZ^{\mathrm{T}}D - CUV^{\mathrm{T}})D = Y^{\mathrm{T}}(\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}}) = 0 \end{cases} \tag{2.23}$$

where $U \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{m\times k}$ such that $\mathcal{A}(UV^{\mathrm{T}}) = \mathbf{b}$ and

$$\tilde{Y} = CY, \quad \tilde{Z}^{\mathrm{T}} = Z^{\mathrm{T}}D, \quad \tilde{U} = CU, \quad \tilde{V}^{\mathrm{T}} = V^{\mathrm{T}}D. \tag{2.24}$$

As a result, the optimality condition (2.23) can imply

$$\mathcal{A}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}) = 0,$$

whose proof is totally same with the counterpart of Lemma 2.1, we here omit the details.

On the other hand, the singularity of $Y^{\mathrm{T}}\tilde{Y}$ and $\tilde{Z}^{\mathrm{T}}Z$ is determined by the singularity of Y and Z. This is because it follows from (2.24) that

$$\begin{cases} Y^{\mathrm{T}}\tilde{Y} = Y^{\mathrm{T}}CY = Y^{\mathrm{T}}C^{\frac{1}{2}}C^{\frac{1}{2}}Y = (C^{\frac{1}{2}}Y)^{\mathrm{T}}(C^{\frac{1}{2}}Y) \\ \tilde{Z}^{\mathrm{T}}Z = Z^{\mathrm{T}}DZ = Z^{\mathrm{T}}D^{\frac{1}{2}}D^{\frac{1}{2}}Z = (D^{\frac{1}{2}}Z)^{\mathrm{T}}(D^{\frac{1}{2}}Z) \end{cases}$$

where $C^{\frac{1}{2}} = \mathrm{Diag}(|c_1|, \ldots, |c_n|)$ and $D^{\frac{1}{2}} = \mathrm{Diag}(|d_1|, \ldots, |d_m|)$.

Combining the above two conclusions, we can ultimately conclude that the same result holds of Lemma 2.1 holds when C and D are diagonal and nonsingular matrices.

C and D are Nonsingular Matrices

Applying the same idea, we actually can extend the conclusion of Lemma (2.1) to more general case. Below, we present the result which illustrate that $(\mathcal{A}^{(C,D)}, \mathbf{b}^{(B)}, k)$ shall satisfy the SNIG condition only if C and D are nonsingular and $k \geqslant r^*$.

Theorem 2.2. *Suppose that the triplet $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$ defined by (2.14) satisfies Assumption 2.1. If both $C \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{m\times m}$ are nonsingular, then the SNIG condition holds at the triplet $(\mathcal{A}^{(C,D)}, \mathbf{b}^{(B)}, k)$.*

Proof. It follows from the triplet $(\mathcal{A}^{(C,D)}, \mathbf{b}^{(B)}, k)$ satisfying Assumption 2.1 that $k \geqslant \operatorname{rank}(B)$ and there exist $U \in \mathbb{R}^{n\times k}$ and $V \in \mathbb{R}^{m\times k}$ such that $\mathbf{b}^{(B)} = \mathcal{A}^{(C,D)}(C^{\mathrm{T}}(UV^{\mathrm{T}})D)$. Namely, (U, V) is a global optimizer of (NLS-k) corresponding to the triplet $(\mathcal{A}^{(C,D)}, \mathbf{b}^{(B)}, k)$. By (2.15), the objective function of (NLS-k) can be reformulated as

$$f(Y, Z) = \frac{1}{2}\|C^{\mathrm{T}}(YZ^{\mathrm{T}})D - B\|_{\mathrm{F}}^2.$$

Assume that (Y, Z) is a second-order stationary point of problem (NLS-k) with respect to $(\mathcal{A}^{(C,D)}, b^{(B)}, k)$. The first-order optimality condition can be written as

$$\begin{cases} Y^{\mathrm{T}}CC^{\mathrm{T}}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})DD^{\mathrm{T}} = \mathbf{0}, \\ CC^{\mathrm{T}}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})DD^{\mathrm{T}}Z = \mathbf{0}. \end{cases} \tag{2.25}$$

Denote $\tilde{Y} = CC^{\mathrm{T}}Y$, $\tilde{Z}^{\mathrm{T}} = Z^{\mathrm{T}}DD^{\mathrm{T}}$, $\tilde{U} = CC^{\mathrm{T}}U$, $\tilde{V}^{\mathrm{T}} = V^{\mathrm{T}}DD^{\mathrm{T}}$, (2.25) can be rearranged as

$$\begin{cases} (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}})Z = \mathbf{0}, \\ Y^{T}(\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}}) = \mathbf{0}. \end{cases}$$

First we consider the case that both $Y^{\mathrm{T}}\tilde{Y}$ and $\tilde{Z}^{\mathrm{T}}Z$ are nonsingular, we have

$$\begin{cases} \tilde{Y} = \tilde{U}\tilde{V}^{\mathrm{T}}Z(\tilde{Z}^{\mathrm{T}}Z)^{-1}, \\ \tilde{Z} = ((Y^{\mathrm{T}}\tilde{Y})^{-1}Y^{\mathrm{T}}\tilde{U}\tilde{V}^{\mathrm{T}})^{\mathrm{T}}, \end{cases} \tag{2.26}$$

and then

$$Y^{\mathrm{T}}(\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}})Z = \mathbf{0} \quad \Rightarrow \quad Y^{\mathrm{T}}\tilde{Y}\tilde{Z}^{\mathrm{T}}Z = Y^{\mathrm{T}}\tilde{U}\tilde{V}^{\mathrm{T}}Z,$$

which indicates that $Y^{\mathrm{T}}\tilde{U}$ and $\tilde{V}^{\mathrm{T}}Z$ are both nonsingular as well. Moreover, we can obtain

$$(\tilde{U}^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}\tilde{V})^{-1} = (\tilde{Z}^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}\tilde{Y})^{-1}. \tag{2.27}$$

Together with (2.26) and (2.27), we have

$$\begin{aligned}(\mathcal{A}^{(C,D)})^{\mathrm{T}}(\mathcal{A}^{(C,D)}(YZ^{\mathrm{T}} - UV^{\mathrm{T}})) &= \tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}} \\ &= \tilde{U}\tilde{V}^{\mathrm{T}}Z(\tilde{Z}^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}\tilde{Y})^{-1}Y^{\mathrm{T}}\tilde{U}\tilde{V}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}} \\ &= \tilde{U}(\tilde{V}^{\mathrm{T}}Z(\tilde{Z}^{\mathrm{T}}Z)^{-1}(Y^{\mathrm{T}}\tilde{Y})^{-1}Y^{\mathrm{T}}\tilde{U} - I_k)\tilde{V}^{\mathrm{T}} \\ &= \mathbf{0},\end{aligned}$$

which indicates $\mathcal{A}^{(C,D)}(YZ^{\mathrm{T}} - UV^{\mathrm{T}}) = \mathbf{0}$ due to the full rankness of $\mathcal{A}^{(C,D)}$.

Finally, we notice that $Y^{\mathrm{T}}\tilde{Y} = (CY)^{\mathrm{T}}(CY)$ and $\tilde{Z}^{\mathrm{T}}Z = (DZ)^{\mathrm{T}}(DZ)$. It then follows from the nonsingularity of C and D that the nonsingularity of $Y^{\mathrm{T}}\tilde{Y}$ or $\tilde{Z}^{\mathrm{T}}Z$ implies the rank deficiency of Y or Z. Then we recall Theorem 2.1, and complete the proof.

□

2.4.3 The Scenario When $\mathcal{A}$ Takes a Special Form

In this section we present a scenario of SNIG condition holding, where $\mathcal{A}$ is a projection mapping enjoying a special structure which has been investigated in [105]. To better illustrate the idea behind our proof, we begin with discussing a simple case in which the rank of the unknown matrix is 1, and then extend the argument to its general parallel.

A Simple Case: Rank of M is 1

As discussed in Chapter 1, formulation (NLS-k) includes matrix completion as a special case. Suppose that P_Ω is the projection operator mapping a matrix onto the observation set Ω, now consider the following problem.

$$\underset{Y\in\mathbb{R}^{n\times k},Z\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad \|P_\Omega(XY^{\mathrm{T}}-M)\|_{\mathrm{F}}^2, \tag{2.28}$$

where M is the target matrix to be recovered and k is an integer to refer to the rank estimation of matrix M.

Suppose that $\mathbf{e}$ is the vector with all entries equal to one of certain dimension.

i) $M\in\mathbb{R}^{6\times 6}$ with the last entry missing:

In problem (2.28), set $k=1$, $M=\mathbf{e}\mathbf{e}^{\mathrm{T}}\in\mathbb{R}^{6\times 6}$ and $\Omega=\{1,\cdots,6\}\times\{1,\cdots,6\}\backslash\{(6,6)\}$.

Denote that

$$\mathbf{x}=\begin{bmatrix}\mathbf{x}_1\\ a\end{bmatrix},\quad \mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ b\end{bmatrix},$$

and rewrite the vector $\mathbf{e}$ in the form of block matrix correspondingly,

$$\mathbf{e}=\begin{bmatrix}\mathbf{e}_1\\ 1\end{bmatrix}.$$

Then we have that

$$\mathbf{x}\mathbf{y}^{\mathrm{T}}-M=\begin{bmatrix}\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & b\mathbf{x}_1-\mathbf{e}_1\\ a\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1^{\mathrm{T}} & ab-1\end{bmatrix}$$

and

$$P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}}-M)=\begin{bmatrix}\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}} & b\mathbf{x}_1-\mathbf{e}_1\\ a\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1^{\mathrm{T}} & 0\end{bmatrix}.$$

Hence, the gradient of the objective function can be calculated as follows.

$$P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}}-M)\mathbf{y}=\begin{bmatrix}(\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}})\mathbf{y}_1+(b\mathbf{x}_1-\mathbf{e}_1)b\\ (a\mathbf{y}_1^{\mathrm{T}}-\mathbf{e}_1^{\mathrm{T}})\mathbf{y}_1\end{bmatrix}, \tag{2.29}$$

and

$$P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} (\mathbf{y}_1\mathbf{x}_1^{\mathrm{T}} - \mathbf{e}_1\mathbf{e}_1^{\mathrm{T}})\mathbf{x}_1 + (a\mathbf{y}_1 - \mathbf{e}_1)a \\ (b\mathbf{x}_1^{\mathrm{T}} - \mathbf{e}_1^{\mathrm{T}})\mathbf{x}_1 \end{bmatrix}. \tag{2.30}$$

Plugging (2.29) and (2.30) into the following first-order optimality condition and write in a compact form,

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)\mathbf{y} = \mathbf{0}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x} = \mathbf{0}, \end{cases}$$

we can obtain that

$$\begin{cases} (\mathbf{y}^{\mathrm{T}}\mathbf{y})\mathbf{x}_1 = (\mathbf{e}^{\mathrm{T}}\mathbf{y})\mathbf{e}_1, & \text{(2.31a)} \\ (\mathbf{y}_1^{\mathrm{T}}\mathbf{y}_1)a = \mathbf{e}_1^{\mathrm{T}}\mathbf{y}_1, & \text{(2.31b)} \\ (\mathbf{x}^{\mathrm{T}}\mathbf{x})\mathbf{y}_1 = (\mathbf{e}^{\mathrm{T}}\mathbf{x})\mathbf{e}_1, & \text{(2.31c)} \\ (\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1)b = \mathbf{e}_1^{\mathrm{T}}\mathbf{x}_1. & \text{(2.31d)} \end{cases}$$

From (2.31a) and (2.31c), we know that $\mathbf{x}_1, \mathbf{y}_1 \in \mathrm{span}\{\mathbf{e}_1\}$. So we can assume that

$$\begin{cases} \mathbf{x}_1 = c\mathbf{e}_1, & \text{(2.32a)} \\ \mathbf{y}_1 = d\mathbf{e}_1. & \text{(2.32b)} \end{cases}$$

Substituting (2.32a) and (2.32b) into (2.31d) and (2.31b), respectively, we obtain that

$$\begin{cases} bc = 1, \\ ad = 1. \end{cases} \tag{2.33}$$

Recall that $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ a \end{bmatrix} = \begin{bmatrix} c\mathbf{e}_1 \\ a \end{bmatrix}$, this together with (2.31c) and (2.32b), we know that

$$(5c^2 + a^2)d = 5c + a. \tag{2.34}$$

Multiplying d in both sides of (2.34), and then combine with $ad = 1$, we obtain that

$$cd = 1. \tag{2.35}$$

(2.33) and (2.35) jointly give that $ab = 1$. And hence we obtain that $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$, that is, $(\mathbf{x}, \mathbf{y})$ is the global optimizer of problem (2.28).

ii) $M \in \mathbb{R}^{n\times m}$ **with the last entry missing:**

The same conclusion in case i can be extended any rank-1 rectangular matrix M in $\mathbb{R}^{n\times m}$ with the last entry unobserved. Some details in the below analysis will be dropped due to the same analysis idea.

In problem (2.28), set $k = 1$, $M = \mathbf{e}\mathbf{e}^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ and $\Omega = \{1, \cdots, n\}\times\{1, \cdots, m\}\backslash\{(n, m)\}$.

Given two vectors $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ such that $M = \mathbf{u}\mathbf{v}^{\mathrm{T}}$. Rewriting all vectors in block form, that is,

$$\mathbf{u} = \begin{bmatrix}\mathbf{u}_1 \\ u\end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix}\mathbf{v}_1 \\ v\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix}\mathbf{x}_1 \\ x\end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix}\mathbf{y}_1 \\ y\end{bmatrix}.$$

Then,

$$\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \begin{bmatrix}\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_1 - v\mathbf{u}_1 \\ x\mathbf{y}_1^{\mathrm{T}} - u\mathbf{v}_1^{\mathrm{T}} & xy - uv\end{bmatrix},$$

$$P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M) = \begin{bmatrix}\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_1 - v\mathbf{u}_1 \\ x\mathbf{y}_1^{\mathrm{T}} - u\mathbf{v}_1^{\mathrm{T}} & 0\end{bmatrix}.$$

The first-order optimality condition can be expressed as follows.

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)\mathbf{y} = \begin{bmatrix}(\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 + (y\mathbf{x}_1 - v\mathbf{u}_1)y \\ (x\mathbf{y}_1^{\mathrm{T}} - u\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1\end{bmatrix} = \mathbf{0}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x} = \begin{bmatrix}(\mathbf{y}_1\mathbf{x}_1^{\mathrm{T}} - \mathbf{v}_1\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1 + (x\mathbf{y}_1 - u\mathbf{v}_1)x \\ (y\mathbf{x}_1^{\mathrm{T}} - v\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1\end{bmatrix} = \mathbf{0}. \end{cases}$$

That is,

$$\begin{cases} (\mathbf{y}^{\mathrm{T}}\mathbf{y})\mathbf{x}_1 = (\mathbf{v}^{\mathrm{T}}\mathbf{y})\mathbf{u}_1, \\ (\mathbf{y}_1^{\mathrm{T}}\mathbf{y}_1)x = (\mathbf{v}_1^{\mathrm{T}}\mathbf{y}_1)u, \\ (\mathbf{x}^{\mathrm{T}}\mathbf{x})\mathbf{y}_1 = (\mathbf{u}^{\mathrm{T}}\mathbf{x})\mathbf{v}_1, \\ (\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1)y = (\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_1)v. \end{cases} \tag{2.36}$$

Let $\mathbf{x}_1 = \alpha\mathbf{u}_1$ and $\mathbf{y}_1 = \beta\mathbf{v}_1$, then the above equation (2.36) can be simplified as

$$\alpha(\beta^2\|\mathbf{v}_1\|^2 + y^2)\mathbf{u}_1 = (\beta\|\mathbf{v}_1\|^2 + vy)\mathbf{u}_1,$$

$$\beta x = u,$$

$$\beta(\alpha^2\|\mathbf{u}_1\|^2 + x^2)\mathbf{v}_1 = (\alpha\|\mathbf{u}_1\|^2 + ux)\mathbf{v}_1,$$

$$\alpha y = v,$$

which implies that $\alpha\beta = 1$ and $xy = uv$. Therefore, $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$.

iii) $M \in \mathbb{R}^{n\times m}$ with one arbitrary entry missing:

A natural question is if the missing entry in M has to be at (n, m)-th position. The answer might be not so apparent but fortunately no. That is, the only missing entry can be at any position of M.

In problem (2.28), set $k = 1$, $M = \mathbf{e}\mathbf{e}^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ and the (i_0, j_0)-th entry is missing, i.e., the observation set is $\Omega = \{1, \cdots, n\} \times \{1, \cdots, m\}\backslash\{(i_0, j_0)\}$ with $1 \leqslant i_0 \leqslant n,\ 1 \leqslant j_0 \leqslant m$.

Given $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ such that $M = \mathbf{u}\mathbf{v}^{\mathrm{T}}$. Rewrite all vectors in the form of block matrix, that is,

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ u \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ v \\ \mathbf{v}_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ x \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ y \\ \mathbf{y}_2 \end{bmatrix},$$

where u and x are the i_0-th entry of $\mathbf{u}$ and $\mathbf{x}$, v and y are the j_0-th entry of $\mathbf{v}$ and $\mathbf{y}$, respectively. Then we have that

$$\begin{aligned} \mathbf{x}\mathbf{y}^{\mathrm{T}} - M &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_1 - v\mathbf{u}_1 & \mathbf{x}_1\mathbf{y}_2^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_2^{\mathrm{T}} \\ x\mathbf{y}_1^{\mathrm{T}} - u\mathbf{v}_1^{\mathrm{T}} & xy - uv & x\mathbf{y}_2^{\mathrm{T}} - u\mathbf{v}_2^{\mathrm{T}} \\ \mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_2 - v\mathbf{u}_2 & \mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_2^{\mathrm{T}} \end{bmatrix}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M) &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_1 - v\mathbf{u}_1 & \mathbf{x}_1\mathbf{y}_2^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_2^{\mathrm{T}} \\ x\mathbf{y}_1^{\mathrm{T}} - u\mathbf{v}_1^{\mathrm{T}} & 0 & x\mathbf{y}_2^{\mathrm{T}} - u\mathbf{v}_2^{\mathrm{T}} \\ \mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}} & y\mathbf{x}_2 - v\mathbf{u}_2 & \mathbf{x}_2\mathbf{y}_2^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_2^{\mathrm{T}} \end{bmatrix}. \end{aligned} \tag{2.37}$$

Hence, the first-order optimality condition can be expressed as follows.

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y}^\mathrm{T} - M)\mathbf{y} = \begin{bmatrix} (\mathbf{x}_1\mathbf{y}_1^\mathrm{T} - \mathbf{u}_1\mathbf{v}_1^\mathrm{T})\mathbf{y}_1 + (y\mathbf{x}_1 - v\mathbf{u}_1)y + (\mathbf{x}_1\mathbf{y}_2^\mathrm{T} - \mathbf{u}_1\mathbf{v}_2^\mathrm{T})\mathbf{y}_2 \\ (x\mathbf{y}_1^\mathrm{T} - u\mathbf{v}_1^\mathrm{T})\mathbf{y}_1 + (x\mathbf{y}_2^\mathrm{T} - u\mathbf{v}_2^\mathrm{T})\mathbf{y}_2 \\ (\mathbf{x}_2\mathbf{y}_1^\mathrm{T} - \mathbf{u}_2\mathbf{v}_1^\mathrm{T})\mathbf{y}_1 + (y\mathbf{x}_2 - v\mathbf{u}_2)y + (\mathbf{x}_2\mathbf{y}_2^\mathrm{T} - \mathbf{u}_2\mathbf{v}_2^\mathrm{T})\mathbf{y}_2 \end{bmatrix} = \mathbf{0}, \\ P_\Omega(\mathbf{x}\mathbf{y}^\mathrm{T} - M)^\mathrm{T}\mathbf{x} = \begin{bmatrix} (\mathbf{y}_1\mathbf{x}_1^\mathrm{T} - \mathbf{v}_1\mathbf{u}_1^\mathrm{T})\mathbf{x}_1 + (x\mathbf{y}_1 - u\mathbf{v}_1)x + (\mathbf{y}_1\mathbf{x}_2^\mathrm{T} - \mathbf{v}_1\mathbf{u}_2^\mathrm{T})\mathbf{x}_2 \\ (y\mathbf{x}_1^\mathrm{T} - v\mathbf{u}_1^\mathrm{T})\mathbf{x}_1 + (y\mathbf{x}_1^\mathrm{T} - v\mathbf{u}_2^\mathrm{T})\mathbf{x}_2 \\ (\mathbf{y}_2\mathbf{x}_1^\mathrm{T} - \mathbf{v}_2\mathbf{u}_1^\mathrm{T})\mathbf{x}_1 + (x\mathbf{y}_2 - u\mathbf{v}_2)x + (\mathbf{y}_2\mathbf{x}_2^\mathrm{T} - \mathbf{v}_2\mathbf{u}_2^\mathrm{T})\mathbf{x}_2 \end{bmatrix} = \mathbf{0}. \end{cases} \tag{2.38}$$

That is,

$$\begin{cases} (\mathbf{y}^\mathrm{T}\mathbf{y})\mathbf{x}_1 = (\mathbf{v}^\mathrm{T}\mathbf{y})\mathbf{u}_1, \quad (\mathbf{y}^\mathrm{T}\mathbf{y})\mathbf{x}_2 = (\mathbf{v}^\mathrm{T}\mathbf{y})\mathbf{u}_2, & \text{(2.39a)} \\ (\mathbf{y}_1^\mathrm{T}\mathbf{y}_1 + \mathbf{y}_2^\mathrm{T}\mathbf{y}_2)x = (\mathbf{v}_1^\mathrm{T}\mathbf{y}_1 + \mathbf{v}_2^\mathrm{T}\mathbf{y}_2)u, & \text{(2.39b)} \\ (\mathbf{x}^\mathrm{T}\mathbf{x})\mathbf{y}_1 = (\mathbf{u}^\mathrm{T}\mathbf{x})\mathbf{v}_1, \quad (\mathbf{x}^\mathrm{T}\mathbf{x})\mathbf{y}_2 = (\mathbf{u}^\mathrm{T}\mathbf{x})\mathbf{v}_2, & \text{(2.39c)} \\ (\mathbf{x}_1^\mathrm{T}\mathbf{x}_1 + \mathbf{x}_2^\mathrm{T}\mathbf{x}_2)y = (\mathbf{u}_1^\mathrm{T}\mathbf{x}_1 + \mathbf{u}_2^\mathrm{T}\mathbf{x}_2)v. & \text{(2.39d)} \end{cases}$$

Let $\mathbf{x}_1 = \alpha\mathbf{u}_1$, $\mathbf{x}_2 = \alpha\mathbf{u}_2$, $\mathbf{y}_1 = \beta\mathbf{v}_1$ and $\mathbf{y}_2 = \beta\mathbf{v}_2$, then equation (2.39) gives the following relationship.

$$\begin{cases} \alpha(\beta^2(\|\mathbf{v}_1\|_2^2 + \|\mathbf{v}_2\|_2^2) + y^2) = \beta(\|\mathbf{v}_1\|_2^2 + \|\mathbf{v}_2\|_2^2) + vy, \\ \beta x = u, \\ \beta(\alpha^2(\|\mathbf{u}_1\|_2^2 + \|\mathbf{u}_2\|_2^2) + x^2) = \alpha(\|\mathbf{u}_1\|_2^2 + \|\mathbf{u}_2\|_2^2) + ux, \\ \alpha y = v, \end{cases} \tag{2.40}$$

which implies that $\alpha\beta = 1$ and $xy = uv$, and hence gives that $\mathbf{x}\mathbf{y}^\mathrm{T} - M = \mathbf{0}$.

iv) How many entries can be removed to guarantee exact recovery?

According to the above argument, we know that for any rank-1 matrix $M \in \mathbb{R}^{n\times m}$, $(P_\Omega, M, 1)$ satisfies the SNIG condition if the observation set contains $nm - 1$ elements. Then next needs to be considered are the following questions.

- How many at most entries can be removed?

- What the relationship of positions among those entries?

Along the idea of writing the vector in block form which has been used repeatedly above, we assume that the target matrix $M = \mathbf{u}\mathbf{v}^{\mathrm{T}}$ with $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ two given column vectors. And suppose also that $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ is the first-order stationary point of problem (2.28). Then we will rewrite $\mathbf{u}$, $\mathbf{v}$, $\mathbf{x}$ and $\mathbf{y}$ in different forms in term of block vector as necessary.

Remove the last two entries in the last column. In this case, the observation set $\Omega = [n] \times [m] \backslash \{(n-1, m), (n, m)\}$. Denote that

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ v \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ y \end{bmatrix},$$

with $\mathbf{u}_1, \mathbf{x}_1 \in \mathbb{R}^{n-2}$, $\mathbf{u}_2, \mathbf{x}_2 \in \mathbb{R}^2$, $\mathbf{v}_1, \mathbf{y}_1 \in \mathbb{R}^{m-1}$ and $v, y \in \mathbb{R}$. Then we have that

$$M = \mathbf{u}\mathbf{v}^{\mathrm{T}} = \begin{bmatrix} \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & \mathbf{u}_1 v \\ \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}} & \mathbf{u}_2 v \end{bmatrix}, \quad \mathbf{x}\mathbf{y}^{\mathrm{T}} = \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} & \mathbf{x}_1 y \\ \mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} & \mathbf{x}_2 y \end{bmatrix}, \tag{2.41}$$

and hence

$$\begin{aligned} \mathbf{x}\mathbf{y}^{\mathrm{T}} - M &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & \mathbf{x}_1 y - \mathbf{u}_1 v \\ \mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}} & \mathbf{x}_2 y - \mathbf{u}_2 v \end{bmatrix}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M) &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & \mathbf{x}_1 y - \mathbf{u}_1 v \\ \mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}} & 0 \end{bmatrix}. \end{aligned} \tag{2.42}$$

Thus the first-order optimality condition can be expressed explicitly as follows,

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)\mathbf{y} = \begin{bmatrix} (\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 + (\mathbf{x}_1 y - \mathbf{u}_1 v)y \\ (\mathbf{x}_2\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_2\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 \end{bmatrix} = \mathbf{0}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} (\mathbf{y}_1\mathbf{x}_1^{\mathrm{T}} - \mathbf{v}_1\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1 + (\mathbf{y}_1\mathbf{x}_2^{\mathrm{T}} - \mathbf{v}_1\mathbf{u}_2^{\mathrm{T}})\mathbf{x}_2 \\ (y\mathbf{x}_1^{\mathrm{T}} - v\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1 \end{bmatrix} = \mathbf{0}, \end{cases} \tag{2.43}$$

which can be further reformulated as

$$\begin{cases} \mathbf{x}_1(\mathbf{y}^{\mathrm{T}}\mathbf{y}) = \mathbf{u}_1(\mathbf{v}^{\mathrm{T}}\mathbf{y}), & (2.44a) \\ \mathbf{x}_2(\mathbf{y}^{\mathrm{T}}\mathbf{y}) = \mathbf{u}_2(\mathbf{v}^{\mathrm{T}}\mathbf{y}), & (2.44b) \\ \mathbf{y}(\mathbf{x}^{\mathrm{T}}\mathbf{x}) = \mathbf{v}(\mathbf{u}^{\mathrm{T}}\mathbf{x}), & (2.44c) \\ y(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1) = v(\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_1). & (2.44d) \end{cases}$$

Due to (2.44a), (2.44b) and (2.44c), we can set $\mathbf{x} = \alpha\mathbf{u}$, and $\mathbf{y}_1 = \beta\mathbf{v}_1$. Substitute back $\mathbf{x}_1 = \alpha\mathbf{u}_1$ into (2.44d), we obtain $\alpha y = v$. Then substitute $\mathbf{x} = \alpha\mathbf{u}$, $\mathbf{y} = \begin{bmatrix}\beta\mathbf{v}_1^{\mathrm{T}} & y\end{bmatrix}^{\mathrm{T}}$ and $\alpha y = v$ into (2.44b), we have $\alpha\beta = 1$. Therefore, $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$ can be easily achieved just by simple verification.

Remove two entries in secondary diagonal. In this case, the observation set $\Omega = [n] \times [m] \backslash \{(n-1, m), (n, m-1)\}$. Denote that

$$\mathbf{u} = \begin{bmatrix}\mathbf{u}_1 \\ u_1 \\ u_2\end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix}\mathbf{v}_1 \\ v_1 \\ v_2\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix}\mathbf{x}_1 \\ x_1 \\ x_2\end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix}\mathbf{y}_1 \\ y_1 \\ y_2\end{bmatrix},$$

with $\mathbf{u}_1, \mathbf{x}_1 \in \mathbb{R}^{n-2}$, $\mathbf{v}_1, \mathbf{y}_1 \in \mathbb{R}^{m-2}$ and $u_i, v_i, x_i, y_i \in \mathbb{R}$ $(i = 1, 2)$. Then

$$\begin{aligned} \mathbf{x}\mathbf{y}^{\mathrm{T}} - M &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & \mathbf{x}_1 y_1 - \mathbf{u}_1 y_1 & \mathbf{x}_1 y_2 - \mathbf{u}_1 v_2 \\ x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}} & x_1 y_1 - u_1 v_1 & x_1 y_2 - u_1 v_2 \\ x_2\mathbf{y}_1 - u_2\mathbf{v}_1^{\mathrm{T}} & x_2 y_1 - u_2 v_1 & x_2 y_2 - u_2 v_2 \end{bmatrix}, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M) &= \begin{bmatrix} \mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} & \mathbf{x}_1 y_1 - \mathbf{u}_1 y_1 & \mathbf{x}_1 y_2 - \mathbf{u}_1 v_2 \\ x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}} & x_1 y_1 - u_1 v_1 & 0 \\ x_2\mathbf{y}_1 - u_2\mathbf{v}_1^{\mathrm{T}} & 0 & x_2 y_2 - u_2 v_2 \end{bmatrix}. \end{aligned} \tag{2.45}$$

The first-order optimality condition is

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)\mathbf{y} = \begin{bmatrix} (\mathbf{x}_1\mathbf{y}_1^{\mathrm{T}} - \mathbf{u}_1\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 + (\mathbf{x}_1 y_1 - \mathbf{u}_1 v_1)y_1 + (\mathbf{x}_1 y_2 - \mathbf{u}_1 v_2)y_2 \\ (x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 + (x_1 y_1 - u_1 v_1)y_1 \\ (x_2\mathbf{y}_1^{\mathrm{T}} - u_2\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 + (x_2 y_2 - u_2 v_2)y_2 \end{bmatrix} = 0, \\ P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} (\mathbf{y}_1\mathbf{x}_1^{\mathrm{T}} - \mathbf{v}_1\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1 + (\mathbf{y}_1 u_1 - \mathbf{v}_1 u_1)x_1 + (\mathbf{y}_1 x_2 - \mathbf{v}_1 u_2)x_2 \\ (y_1\mathbf{x}_1^{\mathrm{T}} - v_1\mathbf{u}_1^{\mathrm{T}})\mathbf{x}_1 + (x_1 y_1 - u_1 v_1)x_1 \\ (y_2\mathbf{x}_1 - v_2\mathbf{u}_1)\mathbf{x}_1 + (x_2 y_2 - u_2 v_2)x_2 \end{bmatrix} = 0, \end{cases}$$

that is,

$$\begin{cases} \mathbf{x}_1(\mathbf{y}^{\mathrm{T}}\mathbf{y}) = \mathbf{u}_1(\mathbf{v}^{\mathrm{T}}\mathbf{y}), & (2.46a) \\ x_1(\mathbf{y}_1^{\mathrm{T}}\mathbf{y}_1) + x_1 y_1^2 = u_1(\mathbf{v}_1^{\mathrm{T}}\mathbf{y}) + u_1 v_1 y_1, & (2.46b) \\ x_2(\mathbf{y}_1^{\mathrm{T}}\mathbf{y}_1) + x_2 y_2^2 = u_2(\mathbf{v}_1^{\mathrm{T}}\mathbf{y}_1) + u_2 v_2 y_2, & (2.46c) \\ \mathbf{y}_1(\mathbf{x}^{\mathrm{T}}\mathbf{x}) = \mathbf{v}_1(\mathbf{u}^{\mathrm{T}}\mathbf{x}), & (2.46d) \\ y_1(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1) + x_1^2 y_1 = v_1(\mathbf{u}_1^{\mathrm{T}}\mathbf{x}_1) + u_1 v_1 x_1, & (2.46e) \\ y_2(\mathbf{x}_1^{\mathrm{T}}\mathbf{x}_1) + x_2^2 y_2 = v_2(\mathbf{u}_1\mathbf{x}_1) + u_2 v_2 x_2. & (2.46f) \end{cases}$$

Unfortunately, it cannot be concluded that $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$ from the above (2.46).

Remove the last column. In this case, the observation set $\Omega = [n]\times[m]\backslash\{(i,m)|1 \leqslant i \leqslant n\}$. Denote that

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_1 \\ v \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ y \end{bmatrix}.$$

Then we have that

$$\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \begin{bmatrix} x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}} & x_1 y - u_1 v \\ x_2\mathbf{y}_1^{\mathrm{T}} - u_2\mathbf{v}_1^{\mathrm{T}} & x_2 y - u_2 v \\ \vdots & \vdots \\ x_n\mathbf{y}_1^{\mathrm{T}} - u_n\mathbf{v}_1^{\mathrm{T}} & x_n y - u_n v \end{bmatrix}, \quad \text{and} \quad P_\Omega(\mathbf{x}\mathbf{y} - M) = \begin{bmatrix} x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}} & 0 \\ x_2\mathbf{y}_1^{\mathrm{T}} - u_2\mathbf{v}_1^{\mathrm{T}} & 0 \\ \vdots & \vdots \\ x_n\mathbf{y}_1^{\mathrm{T}} - u_n\mathbf{v}_1^{\mathrm{T}} & 0 \end{bmatrix}.$$

So the first-order optimality condition can be written as follows.

$$\begin{cases} P_\Omega(\mathbf{x}\mathbf{y} - M)\mathbf{y} = \begin{bmatrix} (x_1\mathbf{y}_1^{\mathrm{T}} - u_1\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 \\ (x_2\mathbf{y}_1^{\mathrm{T}} - u_2\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 \\ \vdots \\ (x_n\mathbf{y}_1^{\mathrm{T}} - u_n\mathbf{v}_1^{\mathrm{T}})\mathbf{y}_1 \end{bmatrix} = 0, \\ P_\Omega(\mathbf{x}\mathbf{y} - M)^{\mathrm{T}}\mathbf{x} = \begin{bmatrix} \sum_{i=1} x_i^2\mathbf{y}_1 - \sum_{i=1}^n x_i u_i \mathbf{v}_1 \\ 0 \end{bmatrix} = 0, \end{cases} \tag{2.47}$$

which indicates that $\beta x_i = u_i$ $(i = 1, \cdots, n)$ and $\alpha\beta = 1$ if we set $\mathbf{x} = \alpha\mathbf{u}$ and $\mathbf{y}_1 = \beta\mathbf{v}_1$. But we have no idea about the relation between v and y since the second row of $P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)^{\mathrm{T}}\mathbf{x}$ vanishes. Therefore, we could not conclude that $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$.

Nevertheless, the above analysis show that we can obtain $\alpha y = v$ if there exists at least one nonzero entry in the last column of $P_\Omega(\mathbf{x}\mathbf{y}^{\mathrm{T}} - M)$. And hence we can conclude $\mathbf{x}\mathbf{y}^{\mathrm{T}} - M = \mathbf{0}$. Therefore, we can claim that it can be removed at most $n - 1$ entries in the last column to guarantee exact recovery for the case that the rank of M is one.

Main Results and Proof Outline

Actually, using the same idea and proof techniques, we can extend the argument on rank one matrix M in Subsection 2.4.3 to more general case, i.e., the rank of matrix M is not necessarily restricted to one. And similar conclusion can be achieved as well.

Consider the triplet $(\mathcal{A}^\Omega, \mathbf{b}^\Omega, k) \in \Pi^{(n,m,p)} \times \mathbb{R}^p \times K$ with $(\mathcal{A}^\Omega, \mathbf{b}^\Omega)$ defined by

$$\mathcal{A}^\Omega(X) = (\langle A_1^\Omega, X\rangle, \ldots, \langle A_p^\Omega, X\rangle)^{\mathrm{T}}, \quad \mathbf{b}^\Omega = (b_1^\Omega, \ldots, b_p^\Omega)^{\mathrm{T}}, \tag{2.48}$$

where $\Omega \subset \{(i,j)|1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$, $p = |\Omega|$ is the cardinality of Ω, and $A_t^\Omega = E_{i_t j_t}$ $((i_t, j_t) \in \Omega, 1 \leqslant t \leqslant p)$. Here $E_{ij} \in \mathbb{R}^{n\times m}$ is a matrix with (i,j)-th entry equal to 1 and 0 otherwise. We call Ω the observation-index set.

Any triplet $(\mathcal{A}^\Omega, \mathbf{b}^\Omega, k)$ determines a concrete instance of (NLS-k) as the following

$$\underset{Y\in\mathbb{R}^{n\times k}, Z\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad f(Y,Z) := \frac{1}{2}||\mathcal{A}^\Omega(YZ^{\mathrm{T}}) - \mathbf{b}^\Omega||_2^2. \tag{2.49}$$

If the triplet $(\mathcal{A}^\Omega, \mathbf{b}^\Omega, k)$ satisfies Assumption 2.1, as mentioned before, there exists at least one matrix M such that

$$\mathcal{A}^\Omega(M) = \mathbf{b}^\Omega,$$

or equivalently,

$$M_{i_t j_t} = b_t^{\Omega}, \quad (i_t, j_t) \in \Omega, \ 1 \leqslant t \leqslant p.$$

We call such matrix M an observation matrix with respect to the triplet $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$.

Suppose the observation-index set Ω in (2.48) has the following form

$$\Omega = I_1 \cup I_2 \cup I_3, \tag{2.50}$$

where

$$\begin{aligned} I_1 &= \{(i_s, j_t) | 1 \leqslant s \leqslant \tilde{n}, 1 \leqslant t \leqslant \tilde{m}\}, \\ I_2 &= \{(i_s, j_t) | 1 \leqslant s \leqslant \tilde{n}, \tilde{m}+1 \leqslant t \leqslant m\}, \\ I_3 &= \{(i_s, j_t) | \tilde{n}+1 \leqslant s \leqslant n, 1 \leqslant t \leqslant \tilde{m}\}, \end{aligned} \tag{2.51}$$

for given $\tilde{n} \in \{1, \cdots, n\}$, $\tilde{m} \in \{1, \cdots, m\}$. Clearly, Ω defined by (2.50) consists of $\tilde{n}$ rows and $\tilde{m}$ columns of the observation matrix M.

Our main results can be stated as follows.

Theorem 2.3. *Let $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega})$ be defined as (2.48) and Ω be in the form of (2.50). Suppose the triplet $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$ satisfies Assumption 2.1 and M is its corresponding observation matrix with rank r. If the rank of the submatrices M_{I_i} indexed by I_i satisfies $\operatorname{rank}(M_{I_i}) = \operatorname{rank}(M) = r$ where $I_i\,(i = 1, 2, 3)$ is given by (2.51), then the SNIG condition holds at the triplet $(\mathcal{A}^{\Omega}, b^{\Omega}, k)$.*

Remark 2.1. *We notice that the scenario discussed in Theorem 2.3 is much more general than the one discussed in Theorem 2.2. The number of observations (the number of known entries of the observation matrix M) here is $p = \tilde{m}n + \tilde{n}(m - \tilde{m}) < mn$. The smallest choice of p is $k(m+n)$ when $\tilde{m} = \tilde{n} = k$, which is much less than the lowest requirement, i.e. $(m+n)\log(m+n)$, on the number of observations to guarantee the exact recovery.*

The proof of Theorem 2.3 will be divided into two parts:

i) Firstly, we prove that the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ with the observation set $\tilde{\Omega}$ being a special case of (2.50);

ii) Secondly, we show that for any Ω defined as (2.48), there exists $\tilde{\Omega}$ being of the structure mentioned above and the same cardinality with Ω, and satisfying that problems (NLS-k) determined by $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$ and $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ share the same optimality properties, i.e. function values, optimality conditions.

The Situation that $\tilde{\Omega}$ is of Special Structure

In particular, suppose the special triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ satisfies Assumption 2.1 where $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}})$ is defined by (2.48) and the observation-index set is of the form

$$\tilde{\Omega} = \tilde{I}_1 \cup \tilde{I}_2 \cup \tilde{I}_3 \tag{2.52}$$

where

$$\begin{aligned} \tilde{I}_1 &= \{(i,j) | 1 \leqslant i \leqslant \tilde{n}, 1 \leqslant j \leqslant \tilde{m}\}, \\ \tilde{I}_2 &= \{(i,j) | 1 \leqslant i \leqslant \tilde{n}, \tilde{m}+1 \leqslant j \leqslant m\}, \\ \tilde{I}_3 &= \{(i,j) | \tilde{n}+1 \leqslant i \leqslant n, 1 \leqslant j \leqslant \tilde{m}\}, \end{aligned} \tag{2.53}$$

which contains all the indices of the first $\tilde{n}$ rows and the first $\tilde{m}$ columns of the corresponding observation matrix, say $\tilde{M}$.

Hence, problem (2.49) can further be reformulated as,

$$\underset{\tilde{Y}\in\mathbb{R}^{n\times k}, \tilde{Z}\in\mathbb{R}^{m\times k}}{\text{minimize}} \quad \tilde{f}(\tilde{Y}, \tilde{Z}) := \frac{1}{2}||\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}||_2^2. \tag{2.54}$$

Lemma 2.2. *Let $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}})$ be defined by (2.48) and $\tilde{\Omega}$ be in the form of (2.52). Suppose the triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ satisfies Assumption 2.1 and $\tilde{M}$ is the corresponding observation matrix with rank $\tilde{r}$. If the rank of the submatrices $\tilde{M}_{\tilde{I}_i}$ indexed by $\tilde{I}_i$ satisfies $\mathrm{rank}(\tilde{M}_{\tilde{I}_i}) = \mathrm{rank}(\tilde{M}) = \tilde{r}$ where $\tilde{I}_i\,(i = 1, 2, 3)$ is given by (2.53), then the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$.*

Proof. Suppose $(\tilde{U}, \tilde{V}) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ is the global minimizer of (2.54), namely, $\tilde{M} = \tilde{U}\tilde{V}^{\mathrm{T}}$ and $\mathbf{b}^{\tilde{\Omega}} = \mathcal{A}^{\tilde{\Omega}}(\tilde{U}\tilde{V}^{\mathrm{T}})$. Let $(\tilde{Y}, \tilde{Z}) \in \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times k}$ be a second-order stationary point of problem (2.54). Next, we are going to prove the lemma by the following two folds.

I: Either $\tilde{Y}$ or $\tilde{Z}$ is rank deficient. In this case, the SNIG condition holding at $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ directly follows from Theorem 2.1 and hence the proof is completed.

II: Both of $\tilde{Y}$ and $\tilde{Z}$ are of full column rank. According to the structure of $\tilde{\Omega}$ in the form of (2.52), we rewrite the matrices $\tilde{U}$, $\tilde{V}$, $\tilde{Y}$ and $\tilde{Z}$ as follows

$$\tilde{U} = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, \quad \tilde{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad \tilde{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix},$$

where $U_1, Y_1 \in \mathbb{R}^{\tilde{n}\times k}$, $U_2, Y_2 \in \mathbb{R}^{(n-\tilde{n})\times k}$, $V_1, Z_1 \in \mathbb{R}^{\tilde{m}\times k}$ and $V_2, Z_2 \in \mathbb{R}^{(m-\tilde{m})\times k}$. Then it follows from straightforward calculations that

$$(\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}) = (\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}}) = \begin{bmatrix} Y_1Z_1^{\mathrm{T}} - U_1V_1^{\mathrm{T}} & Y_1Z_2^{\mathrm{T}} - U_1V_2^{\mathrm{T}} \\ Y_2Z_1^{\mathrm{T}} - U_2V_1^{\mathrm{T}} & \mathbf{0} \end{bmatrix}.$$

Hence, the first-order optimality condition of (2.54) can be expressed as follows,

$$\begin{cases} ((\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}))\tilde{Z} = \begin{bmatrix} (Y_1Z_1^{\mathrm{T}} - U_1V_1^{\mathrm{T}})Z_1 + (Y_1Z_2^{\mathrm{T}} - U_1V_2^{\mathrm{T}})Z_2 \\ (Y_2Z_1^{\mathrm{T}} - U_2V_1^{\mathrm{T}})Z_1 \end{bmatrix} = \mathbf{0}, \\ ((\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}))^{\mathrm{T}}\tilde{Y} = \begin{bmatrix} (Z_1Y_1^{\mathrm{T}} - V_1U_1^{\mathrm{T}})Y_1 + (Z_1Y_2^{\mathrm{T}} - V_1U_2^{\mathrm{T}})Y_2 \\ (Z_2Y_1^{\mathrm{T}} - V_2U_1^{\mathrm{T}})Y_1 \end{bmatrix} = \mathbf{0} \end{cases}$$

that is

$$\begin{cases} (Y_1Z_1^{\mathrm{T}} - U_1V_1^{\mathrm{T}})Z_1 + (Y_1Z_2^{\mathrm{T}} - U_1V_2^{\mathrm{T}})Z_2 = \mathbf{0}, & \text{(2.55a)} \\ (Y_2Z_1^{\mathrm{T}} - U_2V_1^{\mathrm{T}})Z_1 = \mathbf{0}, & \text{(2.55b)} \\ (Z_1Y_1^{\mathrm{T}} - V_1U_1^{\mathrm{T}})Y_1 + (Z_1Y_2^{\mathrm{T}} - V_1U_2^{\mathrm{T}})Y_2 = \mathbf{0}, & \text{(2.55c)} \\ (Z_2Y_1^{\mathrm{T}} - V_2U_1^{\mathrm{T}})Y_1 = \mathbf{0}. & \text{(2.55d)} \end{cases}$$

By rearranging (2.55a) and (2.55c), we have

$$\begin{cases} Y_1(\tilde{Z}^{\mathrm{T}}\tilde{Z}) = U_1(\tilde{V}^{\mathrm{T}}\tilde{Z}), & (2.56a) \\ Z_1(\tilde{Y}^{\mathrm{T}}\tilde{Y}) = V_1(\tilde{U}^{\mathrm{T}}\tilde{Y}). & (2.56b) \end{cases}$$

Combining (2.56), (2.55b) and (2.55d) together, we obtain

$$\begin{cases} Y_1 = U_1(\tilde{V}^{\mathrm{T}}\tilde{Z})(\tilde{Z}^{\mathrm{T}}\tilde{Z})^{-1}, & (2.57a) \\ Z_1 = V_1(\tilde{U}^{\mathrm{T}}\tilde{Y})(\tilde{Y}^{\mathrm{T}}\tilde{Y})^{-1}, & (2.57b) \\ Y_2(Z_1^{\mathrm{T}}Z_1) = U_2(V_1^{\mathrm{T}}Z_1), & (2.57c) \\ Z_2(Y_1^{\mathrm{T}}Y_1) = V_2(U_1^{\mathrm{T}}Y_1). & (2.57d) \end{cases}$$

As we know, the second-order necessary optimality condition of problem (2.54) can be formulated as

$$\|\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}S_{\tilde{Z}}^{\mathrm{T}} + S_{\tilde{Y}}\tilde{Z}^{\mathrm{T}})\|_2^2 + 2\mathrm{tr}(S_{\tilde{Y}}^{\mathrm{T}}((\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{U}\tilde{V}^{\mathrm{T}})))S_{\tilde{Z}}) \geqslant 0, \tag{2.58}$$

for all $S_{\tilde{Y}} \in \mathbb{R}^{n\times k}$, $S_{\tilde{Z}} \in \mathbb{R}^{m\times k}$.

We further prove our argument through discussing the following four cases of different structures of (Y_1, Z_1),

i. Y_1 is rank deficient and Z_1 is of full column rank;

ii. Y_1 is of full column rank and Z_1 is rank deficient;

iii. both Y_1 and Z_1 are rank deficient;

iv. both Y_1 and Z_1 are of full column rank.

i. Y_1 is rank deficient and Z_1 is of full column rank. Due to the rank deficiency of Y_1, there exists a nonzero vector $\tilde{\mathbf{y}} = (y_1, \ldots, y_k)^{\mathrm{T}} \in \mathbb{R}^k$ satisfying $Y_1\tilde{\mathbf{y}} = 0$. Without loss of generality, we assume that $y_l \neq 0$ for some $l \in \{1, \cdots, k\}$.

Then we can conclude that $Y_1 Z_2^{\mathrm{T}} = U_1 V_2^{\mathrm{T}}$. Otherwise, suppose that there exists $(s,t) \in \{1,\cdots,\tilde{n}\} \times \{\tilde{m}+1,\cdots,m\}$ satisfying $\eta = (Y_1 Z_2^{\mathrm{T}} - U_1 V_2^{\mathrm{T}})_{st} \neq 0$.

Then we set

$$S_{\tilde{Y}} = \begin{bmatrix} S_{Y_1} \\ \mathbf{0} \end{bmatrix} \text{ and } S_{\tilde{Z}} = \begin{bmatrix} \mathbf{0} \\ S_{Z_2} \end{bmatrix} \tag{2.59}$$

where

$$(S_{Y_1})_{i_1 i_2} = \begin{cases} \eta, & \text{if } i_1 = s \text{ and } i_2 = l, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad S_{Z_2} = \tilde{z}\tilde{y}^{\mathrm{T}}$$

with $\tilde{\mathbf{z}} \in \mathbb{R}^{m-\tilde{m}}$ given by

$$\tilde{z}_j = \begin{cases} -\xi y_l, & \text{if } j = t, \\ 0, & \text{otherwise,} \end{cases}$$

and $\xi > \|\mathcal{A}^{\Omega}(\mathbf{r}_s \mathbf{z}_l^{\mathrm{T}})\|_2^2 / 2y_l^2$. Here $\mathbf{r}_s \in \mathbb{R}^{\tilde{n}}$ is the s-th column of identity matrix $I_{\tilde{n}}$ and $\mathbf{z}_l$ denotes the l-th column of Z.

Substituting (2.59) into (2.58), we obtain

$$\begin{aligned} &\|\mathcal{A}^{\Omega}(Y S_{Z_2}^{\mathrm{T}} + S_{Y_1} Z^{\mathrm{T}})\|_2^2 + 2\mathrm{tr}(S_{Y_1}^{\mathrm{T}}(Y_1 Z_2^{\mathrm{T}} - U_1 V_2^{\mathrm{T}}) S_{Z_2}) \\ =&\|\mathcal{A}^{\Omega}(S_{Y_1} Z^{\mathrm{T}})\|_2^2 + 2\mathrm{tr}(\tilde{y}^{\mathrm{T}} S_{Y_1}^{\mathrm{T}}(Y_1 Z_2^{\mathrm{T}} - U_1 V_2^{\mathrm{T}})\tilde{z}) \\ =&\eta^2 \|\mathcal{A}^{\Omega}(r_s z_l^{\mathrm{T}})\|_2^2 - 2\xi\eta^2 \tilde{y}_l^2 < 0, \end{aligned}$$

which implies that the second-order necessary condition is violated at $(\tilde{Y}, \tilde{Z})$. Therefore, it holds that $Y_1 Z_2^{\mathrm{T}} = U_1 V_2^{\mathrm{T}}$. Together with (2.55a), (2.55b), we obtain $\tilde{Y}(Z_1^{\mathrm{T}} Z_1) = \tilde{U}(V_1^{\mathrm{T}} Z_1)$.

From the assumption $\mathrm{rank}(\tilde{M}_{\tilde{I}_1}) = \tilde{r} = k$, we know $\mathrm{rank}(U_1 V_1^{\mathrm{T}}) = \tilde{r} = k$, which implies U_1 and V_1 are full column rank. Thus, $V_1^{\mathrm{T}} Z_1$ is nonsingular. Consequently,

$$Y_1 = U_1(V_1^{\mathrm{T}} Z_1)(Z_1^{\mathrm{T}} Z_1)^{-1}$$

is of full column rank, which contradicts to the assumption that Y_1 is rank deficient.

ii. Y_1 is of full column rank and Z_1 is rank deficient. We can show the same contradiction as case i. in the same manner and hence omit the detailed proof.

iii. Both Y_1 and Z_1 are rank deficient. By the same argument in case i, it follows from the second-order optimality (2.58) that

$$\begin{cases} Y_1 Z_2^{\mathrm{T}} = U_1 V_2^{\mathrm{T}}, \\ Y_2 Z_1^{\mathrm{T}} = U_2 V_1^{\mathrm{T}}, \end{cases} \tag{2.60}$$

due to the rank deficiency of Y_1 and Z_1.

It follows from the relationship (2.60) and $\mathrm{rank}(\tilde{M}_{\tilde{I}_i}) = \mathrm{rank}(\tilde{M}) = \tilde{r}$ $(i = 1, 2, 3)$ that $\tilde{r}(= \mathrm{rank}(\tilde{U}\tilde{V}^{\mathrm{T}}) = \mathrm{rank}(\tilde{M})) < k$. Hence there exist $\bar{U} \in \mathbb{R}^{n\times\tilde{r}}$ and $\bar{V} \in \mathbb{R}^{m\times\tilde{r}}$ such that $\bar{U}\bar{V}^{\mathrm{T}} = \tilde{U}\tilde{V}^{\mathrm{T}} = \tilde{M}$. Denote $\bar{U}$ and $\bar{V}$ as

$$\bar{U} = \begin{bmatrix} \bar{U}_1 \\ \bar{U}_2 \end{bmatrix} \quad \text{and} \quad \bar{V} = \begin{bmatrix} \bar{V}_1 \\ \bar{V}_2 \end{bmatrix}.$$

Similar to (2.57) and (2.60), we obtain

$$\begin{cases} Y_1 = \bar{U}_1(\bar{V}^{\mathrm{T}}\tilde{Z})(\tilde{Z}^{\mathrm{T}}\tilde{Z})^{-1}, & \text{(2.61a)} \\ Z_1 = \bar{V}_1(\bar{U}^{\mathrm{T}}\tilde{Y})(\tilde{Y}^{\mathrm{T}}\tilde{Y})^{-1}, & \text{(2.61b)} \\ Y_2(Z_1^{\mathrm{T}}Z_1) = \bar{U}_2(\bar{V}_1^{\mathrm{T}}Z_1), & \text{(2.61c)} \\ Z_2(Y_1^{\mathrm{T}}Y_1) = \bar{V}_2(\bar{U}_1^{\mathrm{T}}Y_1). & \text{(2.61d)} \end{cases}$$

and

$$\begin{cases} Y_1 Z_2^{\mathrm{T}} = \bar{U}_1 \bar{V}_2^{\mathrm{T}}, \\ Y_2 Z_1^{\mathrm{T}} = \bar{U}_2 \bar{V}_1^{\mathrm{T}}. \end{cases} \tag{2.62}$$

It follows from (2.61b), the second equation of (2.62) and the full column rankness of $\bar{U}_2$, $\bar{V}_1$ that

$$\mathrm{span}\{Z_1\} = \mathrm{span}\{\bar{V}_1\}. \tag{2.63}$$

Define $\bar{H} := Y_1 Z_1^{\mathrm{T}} - \bar{U}_1 \bar{V}_1^{\mathrm{T}} \in \mathbb{R}^{\tilde{n}\times\tilde{m}}$. Due to (2.63), it holds that $\bar{H}\mathbf{y} = 0$, for any $\mathbf{y} \perp \mathrm{span}\{Z_1\}$. On the other hand, for any $\mathbf{y} \in \mathrm{span}(Z_1)$, there exists $\lambda \in \mathbb{R}^{\tilde{m}}$ satisfying $\mathbf{y} = Z_1\lambda$. Using (2.55a) and the first equation of (2.62), we have $\bar{H}\mathbf{y} = 0$. Consequently, we can conclude that $\bar{H}\mathbf{y} = 0$, for all $\mathbf{y} \in \mathbb{R}^{\tilde{m}}$, which implies $\bar{H} = 0$. Together with (2.62), we have $(\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}) = 0$. Due to the full rankness of $A^{\tilde{\Omega}}$, we have $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}} = \mathbf{0}$.

iv. Both Y_1 and Z_1 are of full column rank. Define

$$\hat{H} = Y_2 Z_1^{\mathrm{T}} - U_2 V_1^{\mathrm{T}} \in \mathbb{R}^{(n-\tilde{n})\times\tilde{m}}.$$

The full column rankness of Z_1 and equation (2.57b) imply

$$\mathrm{span}\{Z_1\} = \mathrm{span}\{V_1\}. \tag{2.64}$$

Therefore, we can prove $\hat{H}\mathbf{y} = 0$ holds for any $\mathbf{y} \perp \mathrm{span}\{Z_1\}$. By the optimality condition (2.55b), we have $\hat{H}\mathbf{y} = 0$ holds for any $\mathbf{y} \in \mathrm{span}\{Z_1\}$. Thus, for any $\mathbf{y} \in \mathbb{R}^{\tilde{m}}$, it holds that $\hat{H}\mathbf{y} = 0$, which further implies

$$\hat{H} = 0. \tag{2.65}$$

Similarly,

$$Y_1 Z_2^{\mathrm{T}} - U_1 V_2 = 0. \tag{2.66}$$

Together with (2.55a), we have $(Y_1 Z_1^{\mathrm{T}} - U_1 V_1^{\mathrm{T}})Z_1 = 0$. On the other hand, (2.64) implies $(Y_1 Z_1^{\mathrm{T}} - U_1 V_1^{\mathrm{T}})\mathbf{z} = 0$ holds for any $\mathbf{z} \perp \mathrm{span}\{Z_1\}$. Therefore,

$$Y_1 Z_1^{\mathrm{T}} - U_1 V_1^{\mathrm{T}} = 0. \tag{2.67}$$

Collecting (2.65), (2.66) and (2.67), we obtain $(\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}) = 0$. Due to the full rankness of $\mathcal{A}^{\tilde{\Omega}}$, we have $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}} = \mathbf{0}$.

To sum up, we conclude that if $(\tilde{Y}, \tilde{Z})$ is a second-order stationary point of problem (2.54), then $\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}} = \mathbf{0}$, i.e., $(\tilde{Y}, \tilde{Z})$ is a global minimizer of problem

(2.54), which implies that the SNIG condition holds at the triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$. We complete the proof. □

Remark 2.2. *Let $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}})$ be defined by (2.48) with $b_i^{\tilde{\Omega}} = 1$ $(i = 1, \ldots, p)$ and $\tilde{\Omega}$ of the form*

$$\tilde{\Omega} = \{(i,j)|1 \leqslant i \leqslant 2, 1 \leqslant j \leqslant 2\} \cup \{(i,j)|1 \leqslant i \leqslant 2, 3 \leqslant j \leqslant 4\} \cup \{(i,j)|3 \leqslant i \leqslant 4, 1 \leqslant j \leqslant 2\},$$

which implies $n = m = 4$ and $p = |\Omega| = 12$. Set

$$\tilde{U} = \tilde{V} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}^{\mathrm{T}},$$

and denote $\tilde{M} = \tilde{U}\tilde{V}^{\mathrm{T}}$ as the obervation matrix, we can easily verify that $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ satisfies Assumption 2.1 and the SNIG condition holds at $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, 2)$.

Set

$$\tilde{Y} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 1 \end{bmatrix}^{\mathrm{T}}, \quad \tilde{Z} = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 \end{bmatrix}^{\mathrm{T}}.$$

Clearly, $(\tilde{Y}, \tilde{Z})$ is also a global minimizer of (2.54). However, $\tilde{Y}\tilde{Z}^{\mathrm{T}} \neq \tilde{M}$, which means the exact recovery does not hold at $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$.

Permutation between (2.49) and (2.54)

Let Ω be any observation-index set defined by (2.50), and $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega})$ be defined by (2.48). Suppose M is the corresponding observation matrix of triplet $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$, which implies that Assumption 2.1 is satisfied. Denote $S = (S_{ij})$ by

$$S_{ij} = \begin{cases} 1, & (i,j) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

It is not difficult to verify that there exist two permutation matrices, say P and Q, such that

$$\tilde{S} = PSQ^{\mathrm{T}}, \tag{2.68}$$

and the observation-index set $\tilde{\Omega}$ corresponding to $\tilde{S}$, i.e.

$$\tilde{S}_{ij} = \begin{cases} 1, & (i,j) \in \tilde{\Omega}, \\ 0, & \text{otherwise}, \end{cases}$$

satisfies (2.52). Let $\mathcal{A}^{\tilde{\Omega}}$ be defined by (2.48), and $\mathbf{b}^{\tilde{\Omega}}$ be defined by

$$b_t^{\tilde{\Omega}} = \tilde{M}_{i_t j_t}, \qquad (i_t, j_t) \in \tilde{\Omega}, t = 1, 2, ..., p := |\Omega|,$$

where $\tilde{M} = PMQ^{\mathrm{T}}$.

Problem (NLS-k) determined by $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$ and $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ can be reformulated as

$$\underset{Y \in \mathbb{R}^{n\times k}, Z \in \mathbb{R}^{m\times k}}{\text{minimize}} \quad f(Y,Z) = \frac{1}{2}\|S \circ (YZ^{\mathrm{T}} - M)\|_{\mathrm{F}}^2, \tag{2.69}$$

and

$$\underset{\tilde{Y} \in \mathbb{R}^{n\times k}, \tilde{Z} \in \mathbb{R}^{m\times k}}{\text{minimize}} \quad \tilde{f}(\tilde{Y},\tilde{Z}) = \frac{1}{2}\|\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M})\|_{\mathrm{F}}^2, \tag{2.70}$$

respectively, where $\circ$ designates the Hadamard product of two matrices with the same size.

We first reveal the relationship between the rank of the submatrices of M and $\tilde{M}$.

Lemma 2.3. *Let M be the observation matrix with respect to $(\mathcal{A}^{\Omega}, \mathbf{b}^{\Omega}, k)$ and $\tilde{M} = PMQ^{\mathrm{T}}$ with P and Q defined by (2.68). Suppose that I_i and $\tilde{I}_i$ are defined by (2.51) and (2.53) for $i = 1, 2, 3$, then the submatrix M_{I_i} and $\tilde{M}_{\tilde{I}_i}$ have the same rank, that is,*

$$\mathrm{rank}(M_{I_i}) = \mathrm{rank}(\tilde{M}_{\tilde{I}_i}), \ i = 1, 2, 3.$$

Proof. It is well known that multiplying a permutation matrix from the left and right side of a matrix is only reorder the rows and columns of that matrix. From $\tilde{M} = PMQ^{\mathrm{T}}$, we know that actually $\tilde{M}_{\tilde{I}_i} = M_{I_i}$, $\forall i = 1, 2, 3$. Thus, it holds that $\mathrm{rank}(\tilde{M}_{\tilde{I}_i}) = \mathrm{rank}(M_{I_i})$ $(\forall i = 1, 2, 3)$. □

To obtain the equivalence of (2.69) and (2.70), we need the following relationship.

Lemma 2.4. *Let P and Q be two permutation matrices and $S \in \mathbb{R}^{n\times m}$ be a $0-1$ matrix. Then we have*

$$S \circ W = P^{\mathrm{T}}(PSQ \circ PWQ)Q^{\mathrm{T}}, \ \forall\, W \in \mathbb{R}^{n\times m}. \tag{2.71}$$

The proof of Lemma 2.4 directly follows from the definition of Hadamard product and the basic properties of permutation matrix, and hence is omitted here.

Now, we arrive at our main theorem.

Theorem 2.4. *Problem (2.69) and problem (2.70) share the following optimality properties:*

1. *$f(Y,Z) = \tilde{f}(\tilde{Y},\tilde{Z})$, for any $Y \in \mathbb{R}^{n\times k}$, $Z \in \mathbb{R}^{m\times k}$, $\tilde{Y} = PY$ and $\tilde{Z} = QZ$;*
2. *if (Y^*, Z^*) is a second-order stationary point of problem (2.69), then (PY^*, QZ^*) is a second-order stationary point of problem (2.70), and vice versa.*

Proof. 1.

$$\begin{aligned}
\tilde{f}(\tilde{Y},\tilde{Z}) =& \frac{1}{2}\|\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M})\|_{\mathrm{F}}^2 \\
=& \frac{1}{2}\|PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}}\|_{\mathrm{F}}^2 \\
=& \frac{1}{2}\|P^{\mathrm{T}}(PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}})Q\|_{\mathrm{F}}^2 \\
=& \frac{1}{2}\|S \circ (YZ^{\mathrm{T}} - M)\|_{\mathrm{F}}^2 = f(Y,Z).
\end{aligned}$$

The third equality is due to the orthogonality of P and Q, and the fourth results from (2.71).

2. What we need to prove is that (Y^*, Z^*) satisfying the optimality conditions of problem (2.69) implies that (PY^*, QZ^*) satisfying the optimality conditions of

problem (2.70), and vice versa. For simplicity of notations, we below drop the corner mark "$*$" in (Y^*, Z^*) and (PY^*, QZ^*). Then the first order optimality conditions are as follows.

$$\text{Problem (2.69):} \quad \begin{cases} (S \circ (YZ^{\mathrm{T}} - M))Z = 0, \\ (S \circ (YZ^{\mathrm{T}} - M))^{\mathrm{T}}Y = 0. \end{cases} \tag{2.72}$$

$$\text{Problem (2.70):} \quad \begin{cases} (\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))\tilde{Z} = 0, \\ (\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))^{\mathrm{T}}\tilde{Y} = 0. \end{cases} \tag{2.73}$$

It is not difficult to verify that (2.72)$\Leftrightarrow$ (2.73). Indeed,

"$\Rightarrow$"

$$\begin{cases} (S \circ (YZ^{\mathrm{T}} - M))Z = 0, \\ (S \circ (YZ^{\mathrm{T}} - M))^{\mathrm{T}}Y = 0, \end{cases}$$

$$\xRightarrow{(2.71)} \begin{cases} (P^{\mathrm{T}}(PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}})Q)Z = 0, \\ (P^{\mathrm{T}}(PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}})Q)^{\mathrm{T}}Y = 0, \end{cases}$$

$$\xRightarrow{P,Q\,\text{nonsingular}} \begin{cases} \tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M})\tilde{Z} = 0, \\ (\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))^{\mathrm{T}}\tilde{Y} = 0. \end{cases}$$

"$\Leftarrow$"

$$\begin{cases} (\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))\tilde{Z} = 0, \\ (\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))^{\mathrm{T}}\tilde{Y} = 0, \end{cases}$$

$$\xRightarrow{P,Q\,\text{nonsigular}} \begin{cases} P^{\mathrm{T}}(PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}})QZ = 0, \\ Q^{\mathrm{T}}(PSQ^{\mathrm{T}} \circ P(YZ^{\mathrm{T}} - M)Q^{\mathrm{T}})^{\mathrm{T}}PY = 0, \end{cases}$$

$$\xRightarrow{(2.71)} \begin{cases} (S \circ (YZ^{\mathrm{T}} - M))Z = 0, \\ (S \circ (YZ^{\mathrm{T}} - M))^{\mathrm{T}}Y = 0. \end{cases}$$

Next, we show that the second-order optimality conditions of problem (2.69)

and (2.70) are equivalent. Their second-order optimality conditions are equivalent to the following inequalities.

$$\text{Problem (2.69):} \quad \|S\circ(YS_Z^{\mathrm{T}}+S_YZ^{\mathrm{T}})\|_{\mathrm{F}}^2+2\mathrm{tr}(S_Y^{\mathrm{T}}(S\circ(YZ^{\mathrm{T}}-M))S_Z)\geqslant 0 \tag{2.74}$$

holds for any $S_Y\in\mathbb{R}^{n\times k}$ and $S_Z\in\mathbb{R}^{m\times k}$.

$$\text{Problem (2.70):} \quad \|\tilde{S}\circ(\tilde{Y}S_{\tilde{Z}}^{\mathrm{T}}+S_{\tilde{Y}}Z^{\mathrm{T}})\|_{\mathrm{F}}^2+2\mathrm{tr}(S_{\tilde{Y}}^{\mathrm{T}}(\tilde{S}\circ(\tilde{Y}\tilde{Z}-\tilde{M}))S_{\tilde{Z}})\geqslant 0 \tag{2.75}$$

holds for any $S_{\tilde{Y}}=PS_Y$ and $S_{\tilde{Z}}=QS_Z$.

It is also not difficult to verify that

$$(2.74)\text{ holds for }S_Y\in\mathbb{R}^{n\times k}\text{ and }S_Z\in\mathbb{R}^{m\times k}\Leftrightarrow(2.75)\text{ holds for }S_{\tilde{Y}}\text{ and }S_{\tilde{Z}}.$$

Actually,

$$\begin{aligned}\|S\circ(YS_Z^{\mathrm{T}}+S_YZ^{\mathrm{T}})\|_{\mathrm{F}}^2 &\overset{(2.71)}{=\!=}\|P^{\mathrm{T}}(PSQ^{\mathrm{T}}\circ P(YS_Z^{\mathrm{T}}+S_YZ^{\mathrm{T}})Q^{\mathrm{T}})Q\|_{\mathrm{F}}^2\\ &\overset{P,Q\,\text{orthogonal}}{=\!=\!=}\|\tilde{S}\circ(\tilde{Y}S_{\tilde{Z}}^{\mathrm{T}}+S_{\tilde{Y}}\tilde{Z}^{\mathrm{T}})\|_{\mathrm{F}}^2,\end{aligned}$$

and

$$\begin{aligned}\mathrm{tr}(S_Y^{\mathrm{T}}(S\circ(YZ^{\mathrm{T}}-M))S_Z) &\overset{(2.71)}{=\!=}\mathrm{tr}(S_Y^{\mathrm{T}}[P^{\mathrm{T}}(PSQ^{\mathrm{T}}\circ P(YZ^{\mathrm{T}}-M)Q^{\mathrm{T}})Q]S_Z)\\ &=\mathrm{tr}(S_{\tilde{Y}}(\tilde{S}\circ(\tilde{Y}\tilde{Z}^{\mathrm{T}}-\tilde{M}))S_{\tilde{Z}}).\end{aligned}$$

Hence, we have that

$$\begin{aligned}&\|S\circ(YS_Z^{\mathrm{T}}+S_YZ^{\mathrm{T}})\|_{\mathrm{F}}^2+2\mathrm{tr}(S_Y^{\mathrm{T}}(S\circ(YZ^{\mathrm{T}}-M))S_Z)\\ =&\|\tilde{S}\circ(\tilde{Y}S_{\tilde{Z}}^{\mathrm{T}}+S_{\tilde{Y}}Z^{\mathrm{T}})\|_{\mathrm{F}}^2+2\mathrm{tr}(S_{\tilde{Y}}^{\mathrm{T}}(\tilde{S}\circ(\tilde{Y}\tilde{Z}-\tilde{M}))S_{\tilde{Z}})\end{aligned} \tag{2.76}$$

holds for any $S_Y\in\mathbb{R}^{n\times k}$, $S_Z\in\mathbb{R}^{m\times k}$ and $S_{\tilde{Y}}=PS_Y$ and $S_{\tilde{Z}}=QS_Z$.

Equation (2.76) implies the equivalence of the second-order optimality between (2.69) and (2.70). We hence achieve the proof.

□

Proof of Theorem 2.3

Proof. From Lemma 2.3, $\text{rank}(\tilde{M}_{\tilde{I}_i}) = \text{rank}(M_{I_i})$ $(i = 1, 2, 3)$. Then for any triplet $(\mathcal{A}^\Omega, \mathbf{b}^\Omega, k)$, we can find a corresponding triplet $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ according to the procedure introduced in the previous subsection, and hence $(\mathcal{A}^{\tilde{\Omega}}, \mathbf{b}^{\tilde{\Omega}}, k)$ satisfies the assumptions of Theorem 2.3 and Theorem 2.2. Suppose that $(\tilde{Y}, \tilde{Z})$ is a second-order stationary point of problem (2.70), it follows from Theorem 2.4 that $(Y, Z) = (P^{\mathrm{T}}\tilde{Y}, Q^{\mathrm{T}}\tilde{Z})$ is a second-order stationary point of problem (2.69). By (2.71) and Theorem 2.2, we have

$$\begin{aligned}(\mathcal{A}^\Omega)^{\mathrm{T}}(\mathcal{A}^\Omega(YZ^{\mathrm{T}}) - \mathbf{b}^\Omega) &= S \circ (YZ^{\mathrm{T}} - M) = P^{\mathrm{T}}(\tilde{S} \circ (\tilde{Y}\tilde{Z}^{\mathrm{T}} - \tilde{M}))Q \\ &= P^{\mathrm{T}}((\mathcal{A}^{\tilde{\Omega}})^{\mathrm{T}}(\mathcal{A}^{\tilde{\Omega}}(\tilde{Y}\tilde{Z}^{\mathrm{T}}) - \mathbf{b}^{\tilde{\Omega}}))Q = \mathbf{0},\end{aligned}$$

which implies $\mathcal{A}^\Omega(YZ^{\mathrm{T}}) - \mathbf{b}^\Omega = \mathbf{0}$ due to the full rankness of $\mathcal{A}$, namely, $f(Y, Z) = 0$. Therefore, the SNIG condition holds the triplet $(\mathcal{A}^\Omega, \mathbf{b}^\Omega, k)$. The proof is completed. □

Chapter 3

An Algorithmic Framework for Solving ARMP

The purpose of this chapter is to investigate an algorithm framework for solving the affine rank minimization problem based our SNIG condition. To start with, we establish the equivalence between (LLS-k) and (NLS-k) in the sense of sharing the same local minima, and then illustrate some properties of the SNIG condition. Then we propose a new algorithm framework based on the SNIG condition for finding the solution of the affine rank minimization problem. Finally, we put forward a conjecture about the reduction between the global minima of the problems with consecutive ranks, and show how this conjecture plays an important role in our propose algorithm framework.

3.1 Relationship between the Linear Least Squares problem (LLS-k) and the Nonlinear Least Squares problem (NLS-k)

The only difference between (LLS-k) and (NLS-k) is that (NLS-k) introduces the variable change $X = YZ^{\mathrm{T}}$. By continuity, it is not difficult to verify that if X is a local minimizer of (LLS-k), each pair of (Y, Z) satisfying $X = YZ^{\mathrm{T}}$ must be a local minimizer of (NLS-k). However, the converse may not hold. In another word, the

change of variables may perhaps introduce extraneous local minima. The following lemma points out that the above situation cannot happen.

Lemma 3.1. *Suppose that $\bar{Y} \in \mathbb{R}^{n\times k}$, $\bar{Z} \in \mathbb{R}^{m\times k}$, and $\bar{X} = \bar{Y}\bar{Z}^{\mathrm{T}}$. Then $\bar{X}$ is a local minimizer of (LLS-k), if and only if $(\bar{Y}, \bar{Z})$ is a local minimizer of (NLS-k).*

Proof. Using the continuity of the map $(Y, Z) \mapsto YZ^{\mathrm{T}}$, we can obtain that if $\bar{X}$ is a local minimizer of (LLS-k), $(\bar{Y}, \bar{Z})$ is a local minimizer of (NLS-k).

Now we consider the case when $(\bar{Y}, \bar{Z})$ is a local minimizer of (NLS-k). Suppose $\bar{X} = \bar{Y}\bar{Z}^{\mathrm{T}}$ is not a local minimizer of (LLS-k) at this time. Then there exists a sequence $\{X^j\}$ such that each X^j is feasible to (LLS-k), $e_k(X^j) < e_k(\bar{X})$ for each j and $\lim\limits_{j\to\infty} X^j = \bar{X}$.

Let $Y^j\Sigma^j(V^j)^{\mathrm{T}}$ be the k-dominant singular value decomposition of X^j. Namely, the columns of $Y^j \in \mathbb{R}^{n\times k}$ and $V^j \in \mathbb{R}^{m\times k}$ are the left and right singular vectors of X^j, respectively, associated with the k largest singular values of X^j which are the diagonal entries of the diagonal matrix $\Sigma^j \in \mathbb{R}^{k\times k}$. Let $Z^j = V^j\Sigma^j$. We have $X^j = Y^j(Z^j)^{\mathrm{T}}$ due to the feasibility of X^j ($\mathrm{rank}(X^j) = k$). According to the definition, it holds that $||Y^j||_2 = 1$ and $||Z^j||_2 = ||X^j||_2$. Since $\{X^j\}$ is bounded, we obtain the boundedness of $\{(Y^j, Z^j)\}$. Hence, there exists a subsequence of $\{(Y^j, Z^j)\}_{j\in\mathcal{J}}$ converging to some (Y, Z). Clearly, we have $YZ^{\mathrm{T}} = \bar{X}$. It follows from $e_k(Y^j(Z^j)^{\mathrm{T}}) = e_k(X^j) < e_k(\bar{X}) = e_k(YZ^{\mathrm{T}})$ that (Y, Z) is not a local minimizer and hence $(\bar{Y}, \bar{Z})$ is not either due to the fact that $YZ^{\mathrm{T}} = \bar{Y}\bar{Z}^{\mathrm{T}}$, which leads to the contrary. We complete the proof. □

3.2 The Algorithm Framework for ARMP

In this subsection, we propose a complete algorithm framework for (1.1) through solving a series of (NLS-k). We start with the relationship between (LLS-k) and (1.1).

Proposition 3.1. *Suppose Assumption 2.1 holds and r^* is the solution of (1.1). For any given k, suppose X^* is a global minimizer of (LLS-k), then the following statements are true:*

1. *If $e_k(X^*) = 0$, then $k \geqslant r^*$.*
2. *If $e_k(X^*) > 0$, then $k < r^*$.*

Proof. Suppose $e_k(X^*) = 0$. Then we have $\mathcal{A}(X^*) = b$ holds, namely $\mathcal{A}(X^*)$ is a feasible point of (1.1). If $k \geqslant r^*$ is not true, then $\operatorname{rank}(X^*) \leqslant k < r^*$, which is contrary to the definition of r^*. Hence, $k \geqslant r^*$.

Now we come to the second part. Suppose that $k \geqslant r^*$, then combining with Assumption 2.1, we obtain that the global minimum of (LLS-k) should be zero. Namely, the second part holds. □

It directly follows from Proposition 3.1 that

$$e_k^* = 0, \qquad e_{k-1}^* > 0 \tag{3.1}$$

holds when $k = r^*$, where e_k^* and e_{k-1}^* denote the global minima of (LLS-k) and (LLS-$(k-1)$), respectively. This means that we can obtain r^* through solving (LLS-k) several times with respect to different k until criterion (3.1) holds.

Proposition 3.1 provides a way to check whether an estimate rank is optimal or not. However, it is not practically useful in general, because solving each (NLS-k) to global optimality is required. The following proposition combines Proposition 3.1 with the SNIG condition, and then obtain the following stronger result.

Proposition 3.2. *Suppose $\mathcal{A} \in \Pi^{(n,m,p)}$, $\mathbf{b} \in \mathbb{R}^p$, and $(\mathcal{A}, \mathbf{b}, k)$ satisfies the SNIG condition for any $k \geqslant r^*$. If (Y^*, Z^*) is a second-order stationary point of (NLS-k), then the following statements are true:*

1. *If $f(Y^*, Z^*) = 0$, then $k \geqslant r^*$.*

2. If $f(Y^*, Z^*) > 0$, then $k < r^*$.

The proof directly follows from Proposition 3.1, Definition 2.1 and Lemma 3.1, and hence is omitted.

Based on the above proposition, suppose f_k^* and f_{k-1}^* are the function values of a second-order stationary point of (NLS-k) and (NLS-$(k-1)$), respectively. Then $f_k^* = 0$ and $f_{k-1}^* > 0$ if and only if $k = r^*$. Hence, we can design the following algorithm framework for solving (1.1).

Algorithm 3.1: Factorization Framework for Affine Rank Minimization

```
   /* Initialization */
1  Input A ∈ Π^(n,m,p) and b ∈ R^p; Set initial rank interval Φ := {1, ..., min(n, m)}.
2  Initialize k ∈ Φ, Y^0 ∈ R^(n×k) and Z^0 ∈ R^(m×k).
   /* Main Iteration */
3  Solve (NLS-k) to a second-order stationary point (Y*, Z*).
4  if f(Y*, Z*) = 0 then
5      Delete all elements greater than k from Φ;
6  else
7      Delete all elements less equal than k from Φ.
8  if |Φ| > 1 then
9      Select k from Φ by a preset strategy;
10     Goto Step 3;
11 else
12     Terminate and return Y*(Z*)^T.
```

The preset strategy in Step 9 of Algorithm 3.1 can be enumerating from 1 to r^* or the bisection method. For each of them, the number of total times that we need to call the inner local optimization solver in the worst-case can easily be determined as follows.

Theorem 3.1. *Suppose that the SNIG condition holds at $(\mathcal{A}, \mathbf{b}, k)$ for any $k \geqslant r^*$. Algorithm 3.1 terminates after invoking Step 3 at most r^*, or $\log(\min\{n, m\})$ times, if the Step 9 utilizes the enumeration or bisection, respectively.*

Proof. We skip the detailed proof, because it is a direct corollary of Proposition 3.2. □

In Step 3 of Algorithm 3.1, we can invoke any local optimization algorithm that terminates when the second-order necessary optimality condition is satisfied. A practically efficient choice is the trust-region Newton method, see [23, 28, 101, 102] for details.

Recently, to estimate the worst-case iteration complexity or function value evaluation complexity becomes more and more popular for some first-order methods. There are some works discussing the worst-case complexity for the second-order methods. For instance, [5] showed the $O(\epsilon^{-\frac{3}{2}})$ complexity for a second order interior point method to solve the constrained non-Lipschitz nonconvex optimization problem to an ϵ second order stationary point. In [19], the authors proved the $O(\epsilon^{-3})$ complexity for ARC method to solve a general unconstrained nonconvex problem. Since trust-region Newton method is a special case of ARC, it should share a comparable worst-case complexity.

In what follows, we illustrate by a simple numerical experiment how Algorithm 3.1 works with trust-region Newton method as the local solver and the bisection strategy is taken to choose the estimate rank sequence $\{k\}$. For simplicity, $\mathcal{K}$ denotes the whole set of k selected. We consider the following matrix completion problems. The target matrix $M \in \mathbb{R}^{n\times m}$ with rank r is randomly generated with standard Gaussian distribution. The sampling $0-1$ matrix S of p 1-entries is also randomly generated with uniform distribution. We denote the ratio between the number of measurements and the number of entries in the matrix by $\mathrm{sr} = p/(nm)$ (sampling ratio). Hence, there are four parameters to determine a test instance: (n, m, sr, r). We choose three different small size instances with parameters $(40, 40, 0.9, 20)$, $(40, 80, 0.6, 10)$ and $(200, 200, 0.3, 6)$, respectively. See Figure 3.1 for details. As introduced in [19], the

following stopping criterion is used in our algorithm.

$$||\nabla f(Y,Z)||_{\mathrm{F}}^2 \leqslant \epsilon, \qquad \lambda_{\min}(\nabla^2 f(Y,Z)) \geqslant -\sqrt{\epsilon}, \tag{3.2}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue of a matrix.

Each curve is to illustrate how the total iteration number Iter_k changes to meet reach different tolerances of ϵ with a particular $k \in \mathcal{K}$. We display several curves associated with different $k \in \mathcal{K}$, which are in the sequence chosen by our algorithm.

According to Figure 3.1, we notice that the trust-region Newton method does share a similar worst-case iteration complexity with ARC method. Moreover, we observe that most of the estimated ranks k are greater or much greater than the real rank r^*. Since the computational cost per iteration of the trust-region Newton method significantly expands with the increasing of k, to invoke the local solver in the case of big k is definitely not desired. In this sense, the bisection strategy is not a good option for Algorithm 3.1. Clearly, the enumeration strategy is not either. Since it requires too many unnecessary inner layer loops.

In the next subsection, we would like to introduce a clever strategy for Step 9 of Algorithm 3.1. By which, we need solve much fewer (NLS-k) than enumeration before finding the optimal rank r^*, and the estimated ranks k become much less than those required in bisection strategy.

3.3 The Global Minima Reduction Conjecture

Let $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}^B)$ be defined by (2.14) where I_n denotes the $n \times n$ identity matrix and B is any give matrix. In this case, (LLS-k) is nothing but the k-dominant singular value decomposition of the given data matrix B.

Proposition 3.3. *Let e_k^* be the global minimum of* (LLS-k) *with* $(\mathcal{A}^{(I_n,I_m)}, \mathbf{b}^{(B)})$

defined by (2.14). *We have*

$$e_k^* = \sum_{i=k+1}^{\min\{n,m\}} \sigma_i^2(B), \quad k = 1, 2, ..., \min\{n, m\},$$

where $\sigma_i(B)$ *denotes the* i*-th largest singular value of matrix* B.

The proof of the above proposition directly follows from the equivalence between (LLS-k) with such particular $\mathcal{A}^{(I_n, I_m)}$ and the singular value decomposition.

Denote $e_0^* = \|\mathcal{A}(0) - \mathbf{b}\|_{\mathrm{F}}^2 = \sum_{i=1}^{\min\{n,m\}} \sigma_i^2(B)$, then we have

$$e_{k-1}^* - e_k^* = \sigma_k^2(B),$$

and the following monotonically decreasing reduction:

$$e_0^* - e_1^* \geqslant e_1^* - e_2^* \geqslant \cdots \geqslant e_{\min\{n,m\}-1} - e_{\min\{n,m\}}. \tag{3.3}$$

A natural conjecture is whether the relationship (3.3) holds for any combination of $\mathcal{A}$ and $\mathbf{b}$.

Conjecture 3.1. *Let* e_k^* *be the global minimum of* (LLS-k) *with any given* $(\mathcal{A}, \mathbf{b})$. *The reduction between* e_{k-1}^* *and* e_k^* *is monotonically decreasing. Namely,* (3.3) *holds.*

Now let us come back to Step 9 of Algorithm 3.1. Suppose we know that the real rank r^* is very small, namely, $r^* \ll \min\{n, m\}$. Then to sequentially choose $k = 1, 2, ...$ is often better than to use the bisection technique. If the conjecture holds, then we can further optimize our choice. Suppose we have e_{j-1}^* and e_j^*, then the next k can be $j + \left\lceil \frac{e_j^*}{e_{j-1}^* - e_j^*} \right\rceil$, where $\lceil x \rceil$ denotes the smallest integer not less than x. This is because of the following proposition.

Proposition 3.4. *If Conjecture 3.1 holds, then*

$$r^* \geqslant j + \left\lceil \frac{e_j^*}{e_{j-1}^* - e_j^*} \right\rceil, \quad j = 1, 2, ..., \min\{n, m\} - 1. \tag{3.4}$$

Proof. It follows from the relationship $e^*_{r^*} = 0$ and Conjecture 3.1 that

$$e^*_j = e^*_j - e^*_{r^*} = \sum_{i=j+1}^{r^*} (e^*_i - e^*_{i-1}) \leqslant (r^* - j)(e^*_j - e^*_{j-1}),$$

which gives us the inequality (3.4). We complete the proof. □

We numerically check Conjecture 3.1 by the same matrix completion problems introduced in the last subsection. In order to obtain a high quality global minimum of (LLS-k), we invoke the trust-region Newton approach to solve the nonlinear least square problem (NLS-k) initiated from N randomly generated starting points. The "computed global minimum" is set to be the smallest value among all the function values of the N computed solutions. Theoretically, when N goes to infinity, the computed global minimum equals to e_k with probability one. Empirically, we set $N = \max\{m, n\}$ and believe the computed global minimum is good enough for these small size problem. The reductions between the computed global minima of two consecutive ranks are illustrated in Figure 3.2.

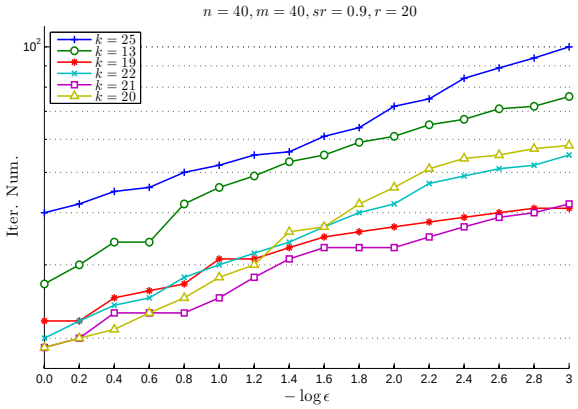
According to Figure 3.2, we can learn that Conjecture 3.1 holds in the testing matrix completion problems. It is reasonable to believe that Conjecture 3.1 holds in a large bunch of linear constrained rank minimization problems. We will further analyze Conjecture 3.1 to obtain better understanding in our future work.

Finally, we check whether the overall performance of Algorithm 3.1 can be improved by using Conjecture 3.1. We still use the aforementioned three matrix completion problems. We use the enumeration with the acceleration by Proposition 3.4 in the Step 9 of Algorithm 3.1. For simplicity, we call it the modified enumeration strategy. We observe how the total weighted iteration number

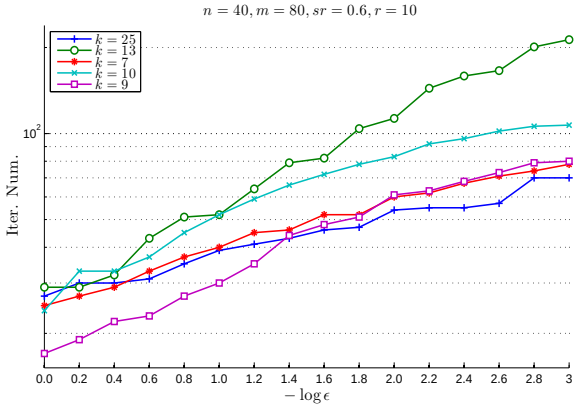
$$\sum_{k \in \mathcal{K}} \text{Iter}_k \cdot \frac{k}{r^*}$$

changes with different tolerance of ϵ.

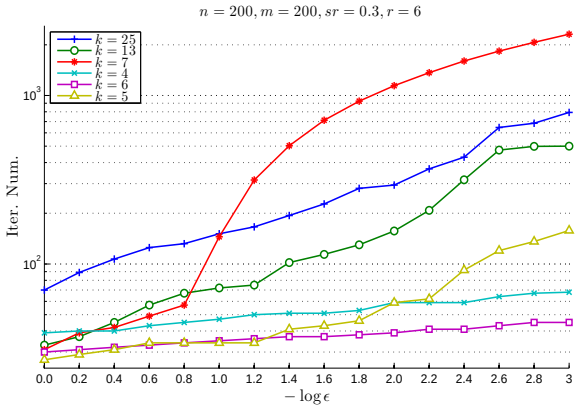
Figure 3.3 verifies the worst-case iteration complexity of the trust-region Newton method again, and also shows the set $\mathcal{K}$ when the new strategy is used in Step 9 of Algorithm 3.1. Figure 3.4 demonstrates the advantage of the modified enumeration strategy comparing with the bisection strategy, since the former one requires much less weighted total iteration numbers than the latter one especially when r^* is far less than the scale of $\min\{m, n\}$.



(a)

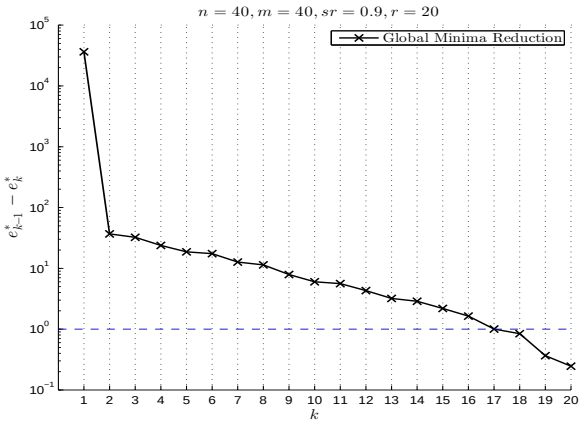


(b)

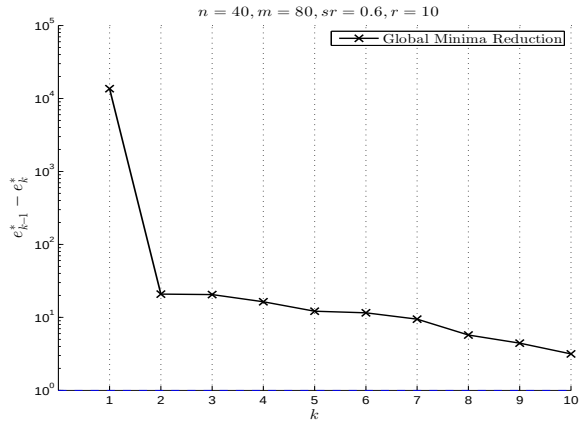


(c)

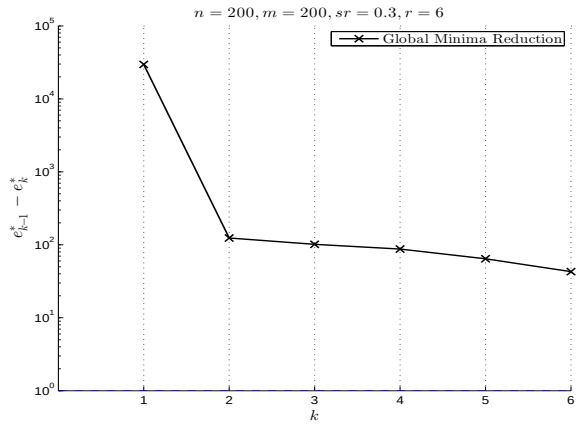
Figure 3.1: iteration numbers required by the trust region Newton method in reaching different tolerance of ϵ



(a)

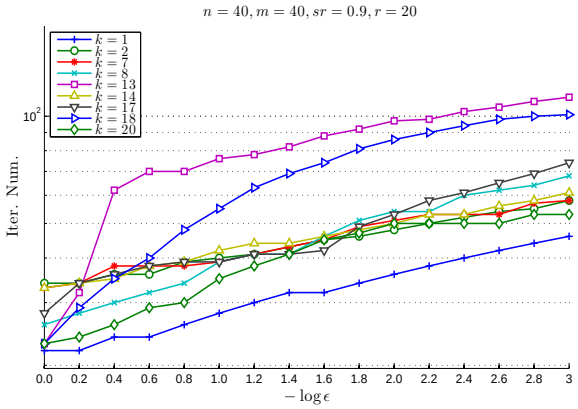


(b)

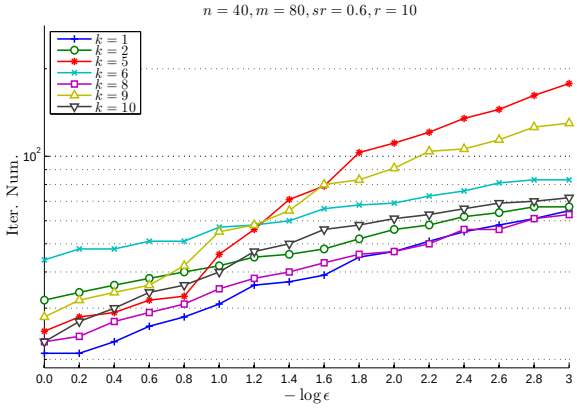


(c)

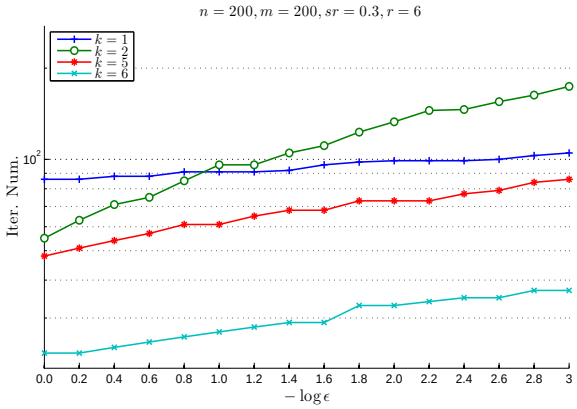
Figure 3.2: reductions between the computed global minima of two consecutive ranks



(a)

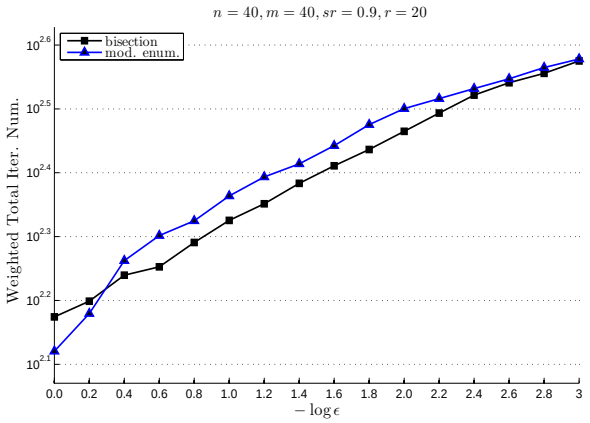


(b)

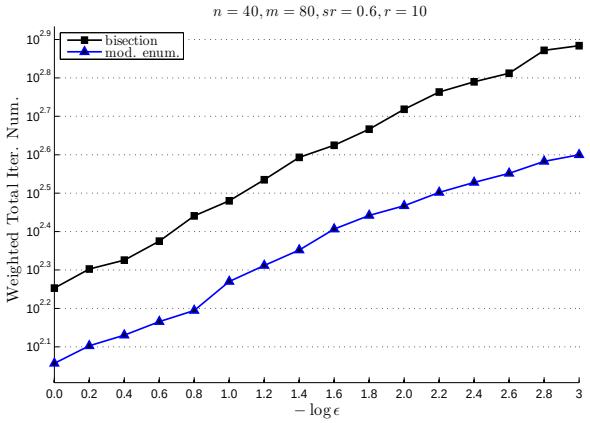


(c)

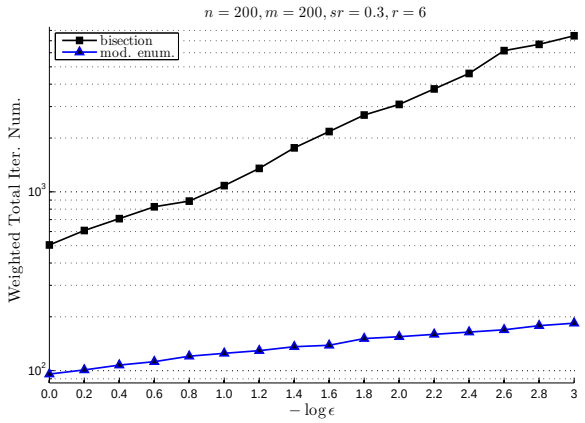
Figure 3.3: iteration numbers required by the trust region Newton method in reaching different tolerance of ϵ



(a)



(b)



(c)

Figure 3.4: weighted total iteration numbers required by using bisection strategy and modified enumeration strategy

Chapter 4

A Regularization Algorithm for PSMP

In this chapter, we turn to the second class of nonconvex optimization problem, i.e., the partially separable convexly constrained nonlinear problem (1.6). We first introduce a partially separable algorithm with adaptive regularization term for solving problem (1.6). Then we present the worst-case complexity evaluation of function value, first- and second-order gradient value. For the sake of convenience, we begin this chapter with restating problem (1.6) as follows:

$$\underset{\mathbf{x}\in\mathcal{F}}{\text{minimize}} \quad f(\mathbf{x}) = \sum_{i=1}^{m} f_i(U_i\mathbf{x}) = \sum_{i=1}^{m} f_i(x_i) \tag{4.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $\mathcal{F}$ is a non-empty closed convex set and for $i \in \{1, \ldots, m\}$, $x_i := U_i\mathbf{x}$ with U_i a (fixed) $n_i \times n$ matrix and $n_i \leqslant n$. And $\{U_i\}_{i=1}^m$ are given m matrices to satisfy that the ranges of the U_i^{T} matrices span $\mathbb{R}$, or equivalently, the intersection of the nullspaces of the U_i is reduced to the origin.

4.1 Details of the Algorithm

Consider problem (4.1) and assume that each element function f_i is p times continuously differentiable and its p-the order derivative tensor $\nabla_{x_i}^p f_i$ is globally Lipschitz

continuous with constant L_i in the sense that, for all $x_i, y_i \in \mathbb{R}^{n_i}$ and some constant $L_i \geqslant 0$,

$$\|\nabla^p_{x_i} f_i(x_i) - \nabla^p_{x_i} f_i(y_i)\|_p \leqslant L_i \|x_i - y_i\| \tag{4.2}$$

where $\|x\|$ is the standard Euclidean norm of x and $\|T\|_p$ is the recursively induced Euclidean norm of the p-th order tensor T (see [6] for details). It can be shown (see Lemma 4.2 below) that this assumption implies that, for $i \in \{1, \ldots, m\}$,

$$f_i(x_i + s_i) = T_{f_i,p}(x_i, s_i) + \frac{1}{p!}\tau_i L_i \|s_i\|^{p+1} \quad \text{with} \quad |\tau_i| \leqslant 1, \tag{4.3}$$

where $s_i = U_i \mathbf{s}$.

Because the quantity $\tau_i L_i$ in (4.3) is usually unknown in practice, it is impossible to use (4.3) directly to model the objective function in a neighborhood of $\mathbf{x}$. However, we may replace this term with an adaptive parameter σ_i, which yields the following $(p+1)$-th order model for the i-th element:

$$m_i(x_i, s_i) = T_{f_i,p} + \frac{1}{p!}\sigma_i \|s_i\|^{p+1}. \tag{4.4}$$

Summing up those element models, we obtain the full model given by

$$m(\mathbf{x}, \mathbf{s}) = \sum_{i=1}^{m} m_i(x_i, s_i) = T_{f,p}(\mathbf{x}, \mathbf{s}) + \frac{1}{p!}\sum_{i=1}^{m} \sigma_i \|s_i\|^{p+1}. \tag{4.5}$$

The algorithm considered in this paper exploits the model (4.5) as follows. At each iteration k, the model (4.5) taken at the iterate $\mathbf{x} = \mathbf{x}_k$ is (approximately) minimized in order to define a step $\mathbf{s}_k$. If the decrease in the objective function value along $\mathbf{s}_k$ is comparable to that predicted by the Taylor model, the trial point $\mathbf{x}_k + \mathbf{s}_k$ is accepted as the new iterate and the regularization parameters $\sigma_{i,k}$ possibly updated. The process is terminated when an approximate local minimizer is found.

If $\mathbf{x} \in \mathcal{F}$ is a local minimizer for problem (4.1) then

$$\chi_{f,1}(\mathbf{x}) := \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1} g(\mathbf{x})^{\mathrm{T}}\mathbf{d} \right| = 0 \tag{4.6}$$

where $g(\mathbf{x}) := \nabla^1_{\mathbf{x}} f(\mathbf{x})$.

Observe that the underlying optimization problem in (4.6), if $\mathcal{F} = \mathbb{R}^n$, $\chi_{f,1}(\mathbf{x}) = \|g(\mathbf{x})\|$ yielding back the familiar first-order necessary condition for unconstrained optimization.

Having defined the criticality measure (4.6), it is natural to use them also for terminating the approximate model minimization: we therefore minimize $m(\mathbf{x}_k, \mathbf{s})$ over $\mathbf{s}$ until

$$\chi_{m,1}(\mathbf{x}_k, \mathbf{s}) \leqslant \kappa \|\mathbf{s}\|^p \tag{4.7}$$

for some constant $\kappa \geqslant 0$, where the derivatives in (4.6) are taken with respect to the second argument of m.

We now introduce some notation useful for describing our algorithm. Define

$$x_{i,k} := U_i \mathbf{x}_k, \quad s_{i,k} := U_i \mathbf{s}_k.$$

Also let

$$\delta f_{i,k} := f_i(x_{i,k}) - f_i(x_{i,k} + s_{i,k}) \quad \text{and} \quad \delta f_k := f(\mathbf{x}_k) - f(\mathbf{x}_k + \mathbf{s}_k) = \sum_{i=1}^{m} \delta f_{i,k},$$

$$\delta m_{i,k} := m_i(x_{i,k}, 0) - m_(x_{i,k}, s_{i,k}) \quad \text{and} \quad \delta m_k := m(\mathbf{x}_k, 0) - m(\mathbf{x}_k, \mathbf{s}_k) = \sum_{i=1}^{m} \delta m_{i,k},$$

$$\delta T_k := T_{f,p}(\mathbf{x}_k, 0) - T_{f,p}(\mathbf{x}_k, \mathbf{s}_k)$$

The partially separable adaptive regularization algorithm is now formally stated as the following Algorithm 4.1.

The requirement that $\rho_k \geqslant \eta$ in both (4.11) and (4.12) is intended to prevent a situation where a particular regularization parameter is increased and another

decreased at a given unsuccessful iteration, followed by the opposite situation at the next iteration, potentially leading to cycling. Other more elaborate mechanisms can be designed to achieve the same goal, such as attempting to reduce a given regularization parameter at most fixed number of times before the occurrence of a successful iteration, but we do not investigate in detail those alternative here.

Algorithm 4.1: Partially Separable Adaptive Regularization

`/* Initialization */`

Input $\mathbf{x}_{-1}$ and $\{\sigma_{i,0}\}_{i=1}^{m} > 0$ as well as constants $0 < \gamma_0 < 1 < \gamma_1 < \gamma_2$, $\eta \in (0,1)$, $\kappa \geqslant 0$, $\sigma_{\min} \in (0, \min_{i=1,\ldots,m} \sigma_{i,0}]$ and $\kappa_{\text{big}} > 1$. Initialize $\mathbf{x}_0$ as the projection of $\mathbf{x}_{-1}$ onto $\mathcal{F}$ and set $k = 0$;

`/* Main iteration */`

Evaluate $f(\mathbf{x}_k)$ and $\{\nabla_x^i f(\mathbf{x}_k)\}_{i=1}^{p}$;

if $\chi_{f,1}(\mathbf{x}_k) \leqslant \epsilon$ **then**

  return $\mathbf{x}_\epsilon = \mathbf{x}_k$ and terminate;

else

  Evaluate a step $\mathbf{s}_k$ such that $\mathbf{x}_k + \mathbf{s}_k \in \mathcal{F}$, $m(\mathbf{x}_k, \mathbf{s}_k) \leqslant m(\mathbf{x}_k, 0)$ and (4.7) holds;

  Compute

$$\rho_k = \frac{\delta f_k}{\delta T_k} \tag{4.8}$$

  and set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ if $\rho_k \geqslant \eta$, or $\mathbf{x}_{k+1} = \mathbf{x}_k$ otherwise;

  if

$$f_i(x_{i,k} + s_{i,k}) > m_i(x_{i,k}, s_{i,k}) \quad i \in \{1, \ldots, m\} \tag{4.9}$$

  then

    set

$$\sigma_{i,k+1} \in [\gamma_1 \sigma_{i,k}, \gamma_2 \sigma_{i,k}]. \tag{4.10}$$

  else if

$$\rho_k \geqslant \eta \text{ and } \delta f_{i,k} \leqslant 0 \text{ and } \delta f_{i,k} < \delta m_{i,k} - \kappa_{\text{big}} |\delta f_k| \tag{4.11}$$

  or

$$\rho_k \geqslant \eta \text{ and } \delta f_{i,k} > 0 \text{ and } \delta f_{i,k} > \delta m_{i,k} + \kappa_{\text{big}} |\delta f_k| \tag{4.12}$$

  then

    set

$$\sigma_{i,k+1} \in [\max[\sigma_{\min}, \gamma_0 \sigma_{i,k}], \sigma_{i,k}]. \tag{4.13}$$

  else

    set

$$\sigma_{i,k+1} = \sigma_{i,k}. \tag{4.14}$$

Increment k by one and go to Step 2;

4.2 Complexity Evaluation of the Algorithm

We start our worst-case analysis by formalizing our assumptions for problem (4.1).

Assumption 4.1. *The feasible set $\mathcal{F}$ is closed, convex and non-empty.*

Assumption 4.2. *Each element function f_i ($i \in \{1, \ldots, m\}$) is p times continuously differentiable in an open set containing $\mathcal{F}$.*

Assumption 4.3. *The p-th derivative of each f_i ($i \in \{1, \ldots, m\}$) is Lipschitz continuous on $\mathcal{F}$, in the sense of (4.2).*

Assumption 4.4. *There exists a constant f_{low} such that $f(\mathbf{x}) \geqslant f_{\text{low}}$ for all $\mathbf{x} \in \mathcal{F}$.*

Note that Assumption 4.4 is necessary for problem (4.1) to be well-defined.

We first observe that our assumption on the partially separable nature of the objective function imply the following useful bounds.

Lemma 4.1. *There exist constants $0 < \varsigma_{\min} \leqslant \varsigma_{\max}$ such that, for all $\mathbf{s} \in \mathbb{R}^n$ and all $r \geqslant 1$,*

$$\varsigma_{\min}^r \|\mathbf{s}\|^r \leqslant \sum_{i=1}^{m} \|s_i\|^r \leqslant m\varsigma_{\max}^r \|\mathbf{s}\|^r. \tag{4.15}$$

Proof. Assume that, for every $\varsigma > 0$ there exists a unit vector $\mathbf{s}_\varsigma$ such that $\max_{i=1,\ldots,m} \|U_i \mathbf{s}_\varsigma\| < \varsigma \|\mathbf{s}_\varsigma\| = \varsigma$. Then taking a sequence of $\{\varsigma_i\}$ converging to zero and using the compactness of the unit sphere, we obtain that the sequence $\{\mathbf{s}_{\varsigma_i}\}$ has at least one limit point $\mathbf{s}_0$ with $\|\mathbf{s}_0\| = 1$ such that $\max_{i=1,\ldots,m} \|U_i \mathbf{s}_0\| = 0$, which is impossible since we assumed that the intersection of the nullspace of the U_i is reduced to the origin. Thus our assumption is false and there is constant $\varsigma_{\min} > 0$ such that, for every $\mathbf{s} \in \mathbb{R}^n$,

$$\max_{i=1,\ldots,m} \|s_i\| = \max_{i=1,\ldots,m} \|U_i \mathbf{s}\| \geqslant \varsigma_{\min} \|\mathbf{s}\|.$$

The first inequality of (4.15) then follows from the fact that

$$\sum_{i=1}^{m} \|s_i\|^r \geqslant \max_{i=1,\ldots,m} \|s_i\|^r \geqslant \varsigma_{\min}^r \|\mathbf{s}\|^r.$$

We have also that

$$\sum_{i=1}^{m} \|s_i\|^r \leqslant m \max_{i=1,\ldots,m} \|U_i\mathbf{s}\|^r \leqslant m \max_{i=1,\ldots,m} (\|U_i\|\|\mathbf{s}\|)^r,$$

which yields the second inequality of (4.15) with $\varsigma_{\max} = \max_{i=1,\ldots,m} \|U_i\|$. □

Taken for $q = 1$, this lemma states that $\sum_{i=1}^{m} \|\cdot\|$ is a norm on $\mathbb{R}^n$ whose equivalence constants with respect to the Euclidean on are $\varsigma_{\min}$ and $m\varsigma_{\max}$. In most applications, these constants are very moderate numbers. We also define

$$L_{\max} := \max_{i=1,\ldots,m} L_i, \quad \text{and} \quad L_f = m\varsigma_{\max}^{p+1} L_{\max} \tag{4.16}$$

and formally state the consequences of Assumption 4.2 and Assumption 4.3.

Lemma 4.2. *Suppose that Assumption 4.2 and Assumption 4.3 hold. Then, for* $k \geqslant 0$,

$$f_i(x_{i,k} + s_{i,k}) = T_{f_i,p}(x_{i,k}, s_{i,k}) + \frac{\tau_{i,k}}{p!} L_i \|s_{i,k}\|^{p+1} \quad \textit{with} \quad |\tau_{i,k}| \leqslant 1, \tag{4.17}$$

for all $i \in \{1, \ldots, m\}$,

$$f(\mathbf{x}_k + \mathbf{s}_k) = T_{f,p}(\mathbf{x}_k, \mathbf{s}_k) + \frac{\tau_k}{p!} L_f \|\mathbf{s}_k\|^{p+1} \quad \textit{with} \quad |\tau_k| \leqslant 1, \tag{4.18}$$

and

$$\|\nabla_{\mathbf{x}}^1 f(\mathbf{x}_k + \mathbf{s}_k) - \nabla_{\mathbf{s}}^1 T_{f,p}(\mathbf{x}_k, \mathbf{s}_k)\| \leqslant \frac{1}{p!} L_f \|\mathbf{s}_k\|^p. \tag{4.19}$$

Proof. The Lipschitz continuity (4.2) implies (4.17) for each $i \in \{1, \ldots, m\}$ (see [6]). Moreover, using (4.15) and the definitions (4.16),

$$\sum_{i=1}^{m} L_i \|s_i\|^{p+1} \leqslant (\max_{i=1,\ldots,m} L_i) \sum_{i=1}^{m} \|s_i\|^{p+1} = L_f \|\mathbf{s}\|^{p+1},$$

and thus (4.17) implies that the p-th derivative tensor of f, $\nabla^p_{\mathbf{x}} f$, is also Lipschitz continuous with constant L_f, from which (4.18) and (4.19) may in turn be derived. □

The definition of the model in (4.5) also implies a simple lower bound on model decrease.

Lemma 4.3. *For all $k \geqslant 0$,*

$$\delta T_k \geqslant \frac{1}{p!} \sigma_{\min} \sum_{i=1}^{m} \|s_{i,k}\|^{p+1}, \tag{4.20}$$

$\mathbf{s}_k \neq 0$ and (4.8) is well-defined.

Proof. The bound directly follows from the observation that

$$T_{f,p}(\mathbf{x}_k, 0) = f(\mathbf{x}_k) = m(\mathbf{x}_k, 0) \geqslant m(\mathbf{x}_k, \mathbf{s}_k) = T_{f,p}(\mathbf{x}_k, \mathbf{s}_k) + \frac{1}{p!} \sum_{i=1}^{m} \sigma_{i,k} \|s_{i,k}\|^{p+1}$$

and (5.61). Moreover, $\chi_{m,1}(\mathbf{x}_k, 0) = \chi_{f,1}(\mathbf{x}_k) > \epsilon$. As a consequence, (4.7) cannot hold for $\mathbf{s}_k = 0$ since termination would have then occured in Step 4 of Algorithm 4.1. Hence, at least one $\|s_{i,k}\|$ is strictly positive because of (4.15) and (4.20) therefore implies that (5.56) is well-defined. □

Our next step is to verify that the regularization parameters cannot grow unbounded.

Lemma 4.4. *Suppose that Assumption 4.2 and Assumption 4.3 hold. Then, for all* $k \geqslant 0$,

$$\sigma_{i,k} \in [\sigma_{\min}, \sigma_{\max}] \tag{4.21}$$

where $\sigma_{\max} := \gamma_2 L_{\max}$.

Proof. Assume that, for some $i \in \{1, \ldots, m\}$ and $k \geqslant 0$, $\sigma_{i,k} \geqslant L_i$. Then (4.17) gives that

$$f_i(x_{i,k} + s_{i,k}) \leqslant T_{f_i,p}(\mathbf{x}_k, \mathbf{s}_k) + \frac{1}{p!} L_i \|s_{i,k}\|^{p+1} = m_i(x_{i,k}, s_{i,k}) + \frac{1}{p!} \|s_{i,k}\|^{p+1} (L_i - \sigma_{i,k})$$

and (5.57) fails, ensuring (4.21) because of the mechanism of the algorithm. □

The final technical ingredient is to ensure that the steps $\mathbf{s}_k$ are sufficiently large compared to the considered criticality measures computed at the next iterate. Let

$$\mathcal{S} = \{k \geqslant 0 \mid \rho_k \geqslant \eta\} \text{ and } \mathcal{U} = \{k \geqslant 0 \mid \rho_k < \eta\}. \tag{4.22}$$

Lemma 4.5. *Suppose that Assumption 4.1-Assumption 4.3 hold. Then, for all* $k \in \mathcal{S}$

$$\|\mathbf{s}_k\| \geqslant \left[\frac{\chi_{f,1}(\mathbf{x}_{k+1})}{2(L_f + \kappa + m\varsigma_{\max}^{p+1}\sigma_{\max})} \right]^{\frac{1}{p}}. \tag{4.23}$$

Proof. Observe first that,

$$\|\nabla_{\mathbf{s}}^1 \|s_i\|^{p+1}\|_1 \leqslant (p+1)\varsigma_{\max}^{p+1} \|\mathbf{s}\|^p, \tag{4.24}$$

where we used the definitions of the induced tensor norm and $\varsigma_{\max}$. This last condition and by definition of the trial point ensure that $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$. Observe now that (4.19), (4.21) and (4.24) imply that

$$\begin{aligned} \|\nabla_{\mathbf{x}}^1 f(\mathbf{x}_{k+1}) - \nabla_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)\|_1 &\leqslant \frac{1}{p!} L_f \|\mathbf{s}_k\|^p + m \frac{(p+1)}{p!} \varsigma_{\max}^{p+1} \sigma_{\max} \|\mathbf{s}_k\|^p \\ &< \left[L_f + m\varsigma_{\max}^{p+1} \sigma_{\max} \right] \|\mathbf{s}_k\|^p, \end{aligned} \tag{4.25}$$

and also that

$$\begin{aligned}\chi_{f,1}(\mathbf{x}_{k+1}) &= |\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}]| \\ &\leqslant |\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] - \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}]| + |\nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}]|,\end{aligned} \tag{4.26}$$

where the first equality defines the vector $\mathbf{d}_{k+1}$ with

$$\|\mathbf{d}_{k+1}\| \leqslant 1. \tag{4.27}$$

Assume now, for the purpose of deriving a contradiction, that (4.23) fails at iteration $k \in \mathcal{S}$. Using (4.25) and (4.27), we obtain that

$$\begin{aligned}&- \nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] + \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}] \\ &\leqslant |\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] - \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}]| \\ &= |(\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1}) - \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k))[\mathbf{d}_{k+1}]| \\ &\leqslant \|\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1}) - \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)\|_1 \|\mathbf{d}_{k+1}\| \\ &< (L_f + m\varsigma_{\max}^{p+1}\sigma_{\max})\|\mathbf{s}_k\|^p.\end{aligned} \tag{4.28}$$

The failure of (4.23) and the first part of (4.26) then imply that

$$-\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] + \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}] < \frac{1}{2}\chi_{f,1}(\mathbf{x}_{k+1}) = -\frac{1}{2} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}],$$

which in turn ensures that

$$\nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}] < \frac{1}{2}\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] < 0.$$

Moreover, $\mathbf{x}_{k+1} + \mathbf{d}_{k+1} \in \mathcal{F}$ by definition of $\chi_{f,1}(\mathbf{x}_{k+1})$, and hence, using (4.27),

$$|\nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}]| \leqslant \chi_{m,1}(\mathbf{x}_k, \mathbf{s}_k). \tag{4.29}$$

We may then substitute this inequality in (4.26) to deduce as above that

$$\begin{aligned}\chi_{f,1}(\mathbf{x}_{k+1}) &\leqslant |\nabla^1_{\mathbf{x}} f(\mathbf{x}_{k+1})[\mathbf{d}_{k+1}] - \nabla^1_{\mathbf{s}} m(\mathbf{x}_k, \mathbf{s}_k)[\mathbf{d}_{k+1}]| + \chi_{m,1}(\mathbf{x}_k, \mathbf{s}_k) \\ &\leqslant (L_f + \kappa + m\varsigma_{\max}^{p+1}\sigma_{\max})\|\mathbf{s}_k\|^p\end{aligned} \tag{4.30}$$

where the last inequality results from (4.28), the identity $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ and (4.7). But this contradicts our assumption that (4.23) fails. Hence (4.23) must hold. □

We are now ready to consider our first complexity result, whose proof uses the definitions

$$\mathcal{S}_k := \{0, \dots, k\} \cap \mathcal{S}, \quad \text{and} \quad \mathcal{U}_k := \{0, \dots, k\} \backslash \mathcal{S}_k \tag{4.31}$$

of the sets of successful and unsuccessful iterations, respectively.

Theorem 4.1. *Suppose that Assumption 4.1-Assumption 4.4 hold. Then Algorithm 4.1 requires at most*

$$\kappa_{\mathcal{S}}(f(\mathbf{x}_0) - f_{\text{low}})\epsilon^{-\frac{p+1}{p}}$$

successful iterations to return a point $\mathbf{x}_\epsilon \in \mathcal{F}$ such that $\chi_{f,1}(\mathbf{x}_\epsilon) \leqslant \epsilon$, for

$$\kappa_{\mathcal{S}} = \frac{p! \left[2(1 + m\varsigma_{\max}^{p+1}\gamma_2)(L_f + \kappa)\right]^{\frac{p+1}{p}}}{\eta\sigma_{\min}\varsigma_{\min}^{p+1}}. \tag{4.32}$$

Proof. Let $k \in \mathcal{S}$ be index of a successful iteration before termination. As a consequence, we obtain, using Assumption 4.4 and Lemma 4.3, that

$$\begin{aligned} f(\mathbf{x}_0) - f_{\text{low}} &\geqslant f(\mathbf{x}_0) - f(\mathbf{x}_{k+1}) \\ &\geqslant \sum_{\ell \in \mathcal{S}_k} [f(\mathbf{x}_\ell) - f(\mathbf{x}_\ell + \mathbf{s}_\ell)] \\ &\geqslant \eta \sum_{\ell \in \mathcal{S}_k} [f(\mathbf{x}_\ell) - T_{f,p}(\mathbf{x}_\ell, \mathbf{s}_\ell)] \\ &\geqslant \frac{\eta}{p!}|\mathcal{S}_k|\sigma_{\min} \min_{\ell \in \mathcal{S}} \left[\sum_{i=1}^{m} \|s_{i,\ell}\|^{p+1}\right]. \end{aligned}$$

Hence we deduce, using (4.15), Lemma 4.5, the definition of $\sigma_{\max}$ in Lemma 4.4 and (4.23), that

$$\begin{aligned} f(\mathbf{x}_0) - f_{\text{low}} &\geqslant \frac{\eta}{p!}|\mathcal{S}_k|\sigma_{\min}\varsigma_{\min}^{p+1} \min_{\ell \in \mathcal{S}_{k,j}} \|\mathbf{s}_\ell\|^{p+1} \\ &\geqslant |\mathcal{S}_k| \frac{\eta\sigma_{\min}\varsigma_{\min}^{p+1}}{p! \left[2(1 + m\varsigma_{\max}^{p+1}\gamma_2)(L_f + \kappa)\right]^{\frac{p+1}{p}}} \epsilon^{\frac{p+1}{p}}. \end{aligned}$$

Thus

$$|\mathcal{S}_k| \leqslant \kappa_{\mathcal{S}}(f(\mathbf{x}_0) - f_{\text{low}})\epsilon^{-\frac{p+1}{p}}$$

where $\kappa_{\mathcal{S}}$ is given by (4.32), and the desired iteration bound then results from the observation that $\epsilon^{-\frac{p+1}{p}} < \epsilon^{-\frac{p+1}{p-1}}$. □

To complete our analysis in terms of evaluations rather than successful iterations, we need to bound the total number of all (successful and unsuccessful) iterations.

Lemma 4.6. *Assume that Assumption 4.2 and Assumption 4.3 hold. Then, for all* $k \geqslant 0$,

$$k \geqslant \kappa^a |\mathcal{S}_k| + \kappa^b,$$

where

$$\kappa^a := 1 + \frac{m|\log \gamma_0|}{\log \gamma_1} \quad \textit{and} \quad \kappa^b := \frac{m}{\log \gamma_1} \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right).$$

Proof. Define

$$\mathcal{I}_{i,k} := \{j \in \{0, \ldots, k\} \mid (5.58) \text{ holds with } k \leftarrow j\},$$

(the set of iterations where $\sigma_{i,j}$ is increased) and

$$\mathcal{D}_{i,k} := \{j \in \{0, \ldots, m\} \mid (5.61) \text{ holds with } k \leftarrow j\} \subset \mathcal{S}_k$$

(the set of iterations where $\sigma_{i,j}$ in decreased), the final inclusion resulting from the condition that $\rho_k \geqslant \eta$ in both (5.59) and (5.60). Observe also that the mechanism of the algorithm, the fact that $\gamma_0 \in (0, 1)$ and Lemma 4.4 impose that, for each $i \in \{1, \ldots, m\}$,

$$\sigma_{\min}\gamma_1^{|\mathcal{I}_{i,k}|}\gamma_0^{|\mathcal{S}_k|} \leqslant \sigma_{0,i}\gamma_1^{|\mathcal{I}_{i,k}|}\gamma_0^{|\mathcal{D}_{i,k}|} \leqslant \sigma_{i,k} \leqslant \sigma_{\max}.$$

Dividing by $\sigma_{\min} > 0$ and taking logarithms yields that, for all $i \in \{1, \ldots, m\}$ and all $k > 0$,

$$|\mathcal{I}_{i,k}| \log \gamma_1 + |\mathcal{S}_k| \log \gamma_0 \leqslant \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right). \tag{4.33}$$

Note now that, if (5.57) fails for all $i \in \{1, \ldots, m\}$, then

$$\delta f_k = \sum_{i=1}^{m} \delta f_{i,k} \geqslant \sum_{i=1}^{m} \delta m_{i,k} = \delta m_k$$

and, in view of (5.56), $\rho_k \geqslant 1 > \eta$, making iteration k successful. Thus, if iteration k is unsuccessful, at least one $\sigma_{i,k}$ is increased with (5.58). Thus

$$|\mathcal{U}_k| \leqslant \sum_{i=1}^{m} |\mathcal{I}_{i,k}| \leqslant m \max_{i=1,\ldots,m} |\mathcal{I}_{i,k}|. \tag{4.34}$$

The desired bound follows from (5.75) and (5.76) by using the fact that $k = |\mathcal{S}_k| + |\mathcal{U}_k| - 1 \leqslant |\mathcal{S}_k| + |\mathcal{U}_k|$, the term -1 in the equality accounting for iteration 0. □

We may now state our main evaluation complexity result.

Theorem 4.2. *Suppose that Assumption 4.1-Assumption 4.3 hold. The Algorithm 4.1 requires at most*

$$\kappa^a \kappa_{\mathcal{S}} (f(\mathbf{x}_0) - f_{\text{low}}) \epsilon^{-\frac{p+1}{p}} + \kappa^b + 1$$

iterations and evaluations of f and its first p derivatives to return a point $\mathbf{x}_\epsilon \in \mathcal{F}$ such that $\chi_{f,1}(\mathbf{x}_\epsilon) \leqslant \epsilon$.

Proof. If termination occurs at iteration 0, the theorem obviously holds. Assume therefore that termination occurs at iteration $k+1$, in which case there must be at least one successful iteration. We may therefore deduce the desired bound from Theorem 4.1, Lemma 4.6 and the fact that each successful iteration involves the evaluation of $f(\mathbf{x}_k)$ and $\{\nabla^i_{\mathbf{x}} f(\mathbf{x}_k)\}_{i=1}^p$, while each unsuccessful iteration only involves that of $f(\mathbf{x}_k)$ and $\{\nabla^i_{\mathbf{x}} f(\mathbf{x}_k)\}_{i=1}^p$. □

Remark 4.1. *This theorem suggests some comments:*

- *This complexity result in $O(\epsilon^{-(p+1)/p})$ evaluations is identical in order to that presented in [6] for the unstructured unconstrained case.*

- *The algorithm (and theory) presented here is considerably simpler than that discussed in [28, 27] in the context of structured trust-regions. In addition, the present assumptions are also weaker. Indeed, not only is the unpleasantly strong assumption of uniform boundedness of all involved derivatives replaced by a single Lipschitz continuity requirement, but an additional condition on long steps (see AA.1s in [28, p.364]) is no longer needed.*

- *The idea of parts 2 and 3 of* (5.59) *and* (5.60) *is simply to identify cases where the model m_i overestimates the element function f_i to an excessive extent, leaving some space for reducing the regularization and hence allowing long steps. Again, other expressions to the same effect are thus acceptable without altering our results.*

Chapter 5

Partially Separable Minimization Problem in Non-Lipschitz Case

In Chapter 4, we consider the partially separable minimization problem with convex constraint, in which the element functions of the objective function are all smooth with high order gradient Lipschitz continuous. In this chapter, we turn to the non-Lipschitz partially separable minimization problem, in which the objective function might be nonsmooth and non-Lipschitz. We restrict the value of p in Chapter 4 to 2 and apply the corresponding algorithm with an adaptive cubic term to solving the following problem and expect to establish corresponding worst-case complexity result:

$$\underset{\mathbf{x}\in\mathcal{F}}{\text{minimize}} \quad f(\mathbf{x}) := f_0(\mathbf{x}) + \sum_{i=1}^{m} \varphi(\|U_i\mathbf{x}\|_1^q) \tag{5.1}$$

where $\mathcal{F}$ is a non-empty closed convex set, $f_0 : \mathbb{R}^n \to \mathbb{R}_+ := [0, \infty)$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$, $U_i \in \mathbb{R}^{m_i \times n}$ for $i = 1, \ldots, m$, $0 < q < 1$ and

$$\|U_i\mathbf{x}\|_1^q = \|x_i\|_1^q = (\sum_{j=1}^{m_i} |x_{i,j}|)^q, \qquad x_i = (x_{i,1}, \ldots, x_{i,m_i})^{\mathrm{T}} \in \mathbb{R}^{m_i}.$$

Problem (5.1) has many applications in engineering and science. Using the non-Lipschitz regularization function in the second term of the objective function f has

remarkable advantages for the restoration of piecewise constant images in image sciences and sparse variable selection in food safety and bioinformatics.

Typical choices of U_i include the identity matrix, first order difference operator, second order difference operator in image sciences and group index matrix in group variable selection.

We assume that f_0 is twice continuously differentiable and its Hessian $\nabla^2 f_0(x)$ is globally Lipschitz continuous with constant L_0. Moreover, φ satisfies the following assumption.

Assumption 5.1. *1. φ is twice continuously differentiable and concave in $(0,\infty)$, and φ'' is globally Lipschitz continuous.*

2. φ is continuous at 0 with $\varphi(0)=0$. And for all $t, t^+ \in (0,\infty)$, there exists a constant c such that

$$0 \leqslant \varphi'(t) \leqslant c \quad \text{and} \quad |\varphi''(t^+) - \varphi''(t)| \leqslant c|t^+ - t|. \tag{5.2}$$

5.1 First-order Necessary Condition

In this section, we first present the first-order necessary condition for the local minimizer of problem (5.1) by a reduced problem of (5.1) in a subspace of $\mathbb{R}^n$, which is derived from the orthogonal decomposition method. Then we give the definition of the first-order and ϵ approximate first-order stationary point of (5.1), which has been studied in [4, 23] for a special case of problem (5.1) where $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} - b \leqslant 0\}$, $m_i = 1, i = 1, \ldots, m$ and $\|\cdot\|_1 = |\cdot|$ with an $\ell \times n$ matrix A and an ℓ-dimensional vector b.

Given a vector $\bar{x} \in \mathbb{R}^n$ and $\epsilon \geqslant 0$, denote

$$I^{\epsilon}_{\bar{\mathbf{x}}} = \{i \in \{1,\ldots,m\} : \|U_i\bar{\mathbf{x}}\|_1 \leqslant \epsilon\}, \qquad J^{\epsilon}_{\bar{\mathbf{x}}} = \{i \in \{1,\ldots,m\} : \|U_i\bar{\mathbf{x}}\|_1 > \epsilon\}$$

$$\mathcal{V}^{\epsilon}_{\bar{\mathbf{x}}} = \{U_i^{\mathrm{T}} : i \in I^{\epsilon}_{\bar{\mathbf{x}}}\}.$$

For convenience, if $\epsilon = 0$, we denote $I_{\bar{\mathbf{x}}} := I^{\epsilon}_{\bar{\mathbf{x}}}$, $J_{\bar{\mathbf{x}}} := J^{\epsilon}_{\bar{\mathbf{x}}}$ and $\mathcal{V}_{\bar{\mathbf{x}}} := \mathcal{V}^{\epsilon}_{\bar{\mathbf{x}}}$. By the assumption that the intersection of the nullspaces of the U_i is reduced to the origin, span$\mathcal{V}_{\bar{\mathbf{x}}} = \mathbb{R}^n$ is equivalent to $\bar{\mathbf{x}} = 0$, and f is non-Lipschitz at $\bar{\mathbf{x}}$ if $\mathcal{V}_{\bar{\mathbf{x}}}$ is nonempty. On the other hand, if $\bar{\mathbf{x}}$ is non-zero, then $\mathcal{V}_{\bar{\mathbf{x}}}$ is a proper subspace of $\mathbb{R}^n$ and hence has a non-empty null space.

Suppose that $Y_{\bar{\mathbf{x}}}$ is an $n \times (n-r)$ matrix whose columns form an orthonormal basis of $\mathcal{V}_{\bar{\mathbf{x}}}$ and $Z_{\bar{\mathbf{x}}}$ is an $n \times r$ matrix whose columns are an orthonormal basis for the null space of $\mathcal{V}_{\bar{\mathbf{x}}}$. If $\bar{\mathbf{x}} = 0$, then $f = f_0$ and span$\mathcal{V}_{\bar{\mathbf{x}}} = \mathbb{R}^n$, which implies that $Z_{\bar{\mathbf{x}}} = 0$ and $Y_{\bar{\mathbf{x}}}$ can be set as the $n \times n$ identity matrix. Otherwise, namely, $\bar{\mathbf{x}} \neq 0$, any $\mathbf{x} \in \mathbb{R}^n$ can be decomposed uniquely as $\mathbf{x} = Y_{\bar{\mathbf{x}}} y + Z_{\bar{\mathbf{x}}} z$, where $\mathbf{y} \in \mathbb{R}^{n-r}$ and $\mathbf{z} \in \mathbb{R}^r$. Due to the non-empty of the null space of $\mathcal{V}_{\bar{\mathbf{x}}}$, $r > 0$ and $Z_{\bar{\mathbf{x}}}$ has r orthonormal columns. By the definition of $Z_{\bar{\mathbf{x}}}$, it is not difficult to verify that

$$U_i Z_{\bar{\mathbf{x}}} = 0,\ \forall i \in I_{\bar{\mathbf{x}}} \quad \text{and} \quad \bar{\mathbf{x}} = Z_{\bar{\mathbf{x}}} \bar{\mathbf{z}}, \tag{5.3}$$

where $\bar{\mathbf{z}} = Z^{\mathrm{T}}_{\bar{\mathbf{x}}} \bar{\mathbf{x}}$ is uniquely determined by $\bar{\mathbf{x}}$ with $Z^{\mathrm{T}}_{\bar{\mathbf{x}}}$ the transpose of $Z_{\bar{\mathbf{x}}}$.

For simplicity of notations, we hereinafter denote

$$\Phi(U_i \mathbf{x}) = \begin{cases} q \nabla_t \varphi(t)_{t=\|U_i \mathbf{x}\|_1^q} \|U_i \mathbf{x}\|_1^{q-1}, & \text{if } U_i \mathbf{x} \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

For a vector $\mathbf{v} \in \mathbb{R}^{m_{\mathbf{v}}}$, define the sign function of v as $\mathrm{sign}(\mathbf{v}) = (\mathrm{sign}(v_1), \ldots, \mathrm{sign}(v_{m_{\mathbf{v}}}))^{\mathrm{T}}$. Then we have

$$\mathrm{sign}(U_i \mathbf{x}) = \mathrm{sign}(x_i) = \begin{bmatrix} \mathrm{sign}(x_{i,1}) \\ \vdots \\ \mathrm{sign}(x_{i,m_i}) \end{bmatrix} = \begin{bmatrix} \mathrm{sign}(u_{i,1} \mathbf{x}) \\ \vdots \\ \mathrm{sign}(u_{i,m_i} \mathbf{x}) \end{bmatrix}.$$

For any $\mathbf{x} \in \mathbb{R}^n$, further denote

$$g(\mathbf{x}) = \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} \Phi(U_i \mathbf{x}) U_i^{\mathrm{T}} \mathrm{sign}(U_i \mathbf{x}), \tag{5.4}$$

which is the gradient of objective function of (5.1) at $\bar{x}$ when $I_{\bar{x}} = \varnothing$.

Lemma 5.1 (first-order necessary condition)**.** *Under Assumption 5.1, if $\bar{\mathbf{x}} \in \mathcal{F}$ is a local minimizer of problem (5.1), then there holds*

$$\chi_{f,1}(\bar{\mathbf{x}}) := \left| \underset{\bar{\mathbf{x}}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\bar{\mathbf{x}}})}{\text{minimize}} g(\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{d} \right| = 0, \tag{5.5}$$

where $R(M)$ denotes the range space of matrix M.

Proof. If $\bar{\mathbf{x}} = 0$, then $Z_{\bar{\mathbf{x}}} = 0$ which means that (5.5) can hold vacuously. Now suppose that $\bar{\mathbf{x}}$ is a nonzero local minimizer of problem (5.1), then $r > 0$ and $Z_{\bar{\mathbf{x}}}$ has r orthonormal columns. Moreover, there exists $\delta_{\bar{\mathbf{x}}}$ such that

$$\begin{aligned}
f(\bar{\mathbf{x}}) &= \underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\{f_0(\mathbf{x}) + \sum_{i=1}^{m}\varphi(\|U_i\mathbf{x}\|_1^q) : \mathbf{x}\in\mathcal{F},\ \|\mathbf{x}-\bar{\mathbf{x}}\|\leqslant\delta_{\bar{\mathbf{x}}}\} \\
&= \underset{\mathbf{y}\in\mathbb{R}^{n-r},\mathbf{z}\in\mathbb{R}^r}{\text{minimize}}\{f_0(Y_{\bar{\mathbf{x}}}\mathbf{y}+Z_{\bar{\mathbf{x}}}\mathbf{z}) + \sum_{i=1}^{m}\varphi(\|U_i(Y_{\bar{\mathbf{x}}}\mathbf{y}+Z_{\bar{\mathbf{x}}}\mathbf{z})\|_1^q) : \\
&\quad Y_{\bar{\mathbf{x}}}\mathbf{y}+Z_{\bar{\mathbf{x}}}\mathbf{z}\in\mathcal{F},\ \|Y_{\bar{\mathbf{x}}}\mathbf{y}+Z_{\bar{\mathbf{x}}}\mathbf{z}-Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}\|\leqslant\delta_{\bar{x}}\} \\
&\leqslant \underset{\mathbf{z}\in\mathbb{R}^r}{\text{minimize}}\{f_0(Y_{\bar{\mathbf{x}}}0+Z_{\bar{\mathbf{x}}}\mathbf{z}) + \sum_{i=1}^{m}\varphi(\|U_i(Y_{\bar{\mathbf{x}}}0+Z_{\bar{\mathbf{x}}}\mathbf{z})\|_1^q) : \\
&\quad Y_{\bar{\mathbf{x}}}0+Z_{\bar{\mathbf{x}}}\mathbf{z}\in\mathcal{F},\ \mathbf{y}=0,\ \|Z_{\bar{\mathbf{x}}}\mathbf{z}-Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}\|\leqslant\delta_{\bar{x}}\} \\
&= \underset{\mathbf{z}\in\mathbb{R}^r}{\text{minimize}}\{f_0(Z_{\bar{\mathbf{x}}}\mathbf{z}) + \sum_{i=1}^{m}\varphi(\|U_iZ_{\bar{\mathbf{x}}}\mathbf{z}\|_1^q) : Z_{\bar{\mathbf{x}}}\mathbf{z}\in\mathcal{F},\ \|Z_{\bar{\mathbf{x}}}\mathbf{z}-Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}\|\leqslant\delta_{\bar{\mathbf{x}}}\} \\
&= \underset{\mathbf{z}\in\mathbb{R}^r}{\text{minimize}}\{f_0(Z_{\bar{\mathbf{x}}}\mathbf{z}) + \sum_{i\in J_{\bar{\mathbf{x}}}}\varphi(\|U_iZ_{\bar{\mathbf{x}}}\mathbf{z}\|_1^q) : Z_{\bar{\mathbf{x}}}\mathbf{z}\in\mathcal{F},\ \|Z_{\bar{\mathbf{x}}}\mathbf{z}-Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}\|\leqslant\delta_{\bar{\mathbf{x}}}\},
\end{aligned}$$

where the last inequality is due to $\varphi(0) = 0$ and (5.3).

Then we introduce a new problem, which is the reduced problem of (5.1) in $\mathbb{R}^r$, namely,

$$\begin{aligned}
\underset{\mathbf{z}\in\mathbb{R}^r}{\text{minimize}} \quad & v(\mathbf{z}) = f_0(Z_{\bar{\mathbf{x}}}\mathbf{z}) + \sum_{i\in J_{\bar{\mathbf{x}}}}\varphi(\|U_iZ_{\bar{\mathbf{x}}}\mathbf{z}\|_1^q), \\
\text{subject to} \quad & Z_{\bar{\mathbf{x}}}\mathbf{z}\in\mathcal{F},
\end{aligned} \tag{5.6}$$

where $v(\mathbf{z})$ is continuously differentiable and the gradient $\nabla v(\mathbf{z})$ is locally Lipschitz continuous in some neighborhood of $\bar{\mathbf{z}}$.

Again using $\varphi(0) = 0$ and (5.3), we obtain that

$$v(\bar{\mathbf{z}}) = f_0(Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}) + \sum_{i\in J_{\bar{\mathbf{x}}}} \varphi(\|U_i Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}\|_1^q) = f(Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}}) = f(\bar{\mathbf{x}}).$$

Therefore, we have

$$v(\bar{\mathbf{z}}) \leqslant \underset{\mathbf{z}\in\mathbb{R}^r}{\text{minimize}}\{v(\mathbf{z}) :\ Z_{\bar{\mathbf{x}}}\mathbf{z} \in \mathcal{F},\ \|Z_{\bar{\mathbf{x}}}(\mathbf{z}-\bar{\mathbf{z}})\| \leqslant \delta_{\bar{\mathbf{x}}}\}.$$

Since $Z_{\bar{\mathbf{x}}}$ is of full column rank, $\bar{\mathbf{z}}$ is a local minimizer of problem (5.6). Hence, we have

$$\nabla v(\bar{\mathbf{z}})^T(\mathbf{z}-\bar{\mathbf{z}}) \geqslant 0, \quad Z_{\bar{\mathbf{x}}}\mathbf{z} \in \mathcal{F}. \tag{5.7}$$

From (5.3), $\varphi(0) = 0$ and the definition of $g(\bar{x})$, we obtain

$$\nabla v(\bar{\mathbf{z}}) = Z_{\bar{\mathbf{x}}}^{\mathrm{T}}\left(\nabla f_0(\bar{\mathbf{x}}) + \sum_{i\in J_{\bar{\mathbf{x}}}} \Phi(U_i Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}})U_i^{\mathrm{T}}\mathrm{sign}(U_i Z_{\bar{\mathbf{x}}}\bar{\mathbf{z}})\right) = Z_{\bar{\mathbf{x}}}^{\mathrm{T}} g(\bar{\mathbf{x}}) \tag{5.8}$$

and

$$Z_{\bar{\mathbf{x}}}(\mathbf{z}-\bar{\mathbf{z}}) = Z_{\bar{\mathbf{x}}}\mathbf{z} - \bar{\mathbf{x}}.$$

Let $\mathbf{d} = Z_{\bar{\mathbf{x}}}(\mathbf{z}-\bar{\mathbf{z}})$. Then (5.7) gives

$$g(\bar{\mathbf{z}})^T d \geqslant 0, \quad \bar{\mathbf{x}} + \mathbf{d} \in \mathcal{F}, \quad \mathbf{d} \in R(Z_{\bar{\mathbf{x}}}). \tag{5.9}$$

Moreover, from

$$\{\mathbf{d} = 0\} \subset \{\mathbf{d} \mid \bar{\mathbf{x}} + \mathbf{d} \in \mathcal{F}, \|\mathbf{d}\| \leqslant 1, \mathbf{d} \in R(Z_{\bar{\mathbf{x}}})\} \subset \{\mathbf{d} \mid \bar{\mathbf{x}} + \mathbf{d} \in \mathcal{F}, \mathbf{d} \in R(Z_{\bar{\mathbf{x}}})\},$$

we obtain (5.5). We complete the proof. □

Definition 5.1 (first-order stationary point)**.** *We call $\bar{\mathbf{x}}$ a first-order stationary point of* (5.1)*, if $\bar{\mathbf{x}}$ satisfies the relation* (5.5) *in Lemma 5.1.*

Definition 5.2 (ϵ approximate first-order stationary point)**.** *For $\epsilon > 0$, we call $\mathbf{x}$ an ϵ approximate stationary point of (5.1), if $\mathbf{x}$ satisfies*

$$\chi_{f,1}(\mathbf{x}) := \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\mathbf{x}}^{\epsilon})} g(\mathbf{x})^{\mathrm{T}}\mathbf{d} \right| \leqslant \epsilon, \tag{5.10}$$

where $R(Z_{\mathbf{x}}^{\epsilon})$ is the range space of $Z_{\mathbf{x}}^{\epsilon}$.

Proposition 5.1. *Let $\mathbf{x}^{\epsilon}$ be an ϵ ($\epsilon > 0$) approximate first-order stationary point of (5.1). Then any cluster point of $\{\mathbf{x}^{\epsilon}\}_{\epsilon>0}$ is a first-order stationary point of (5.1) as $\epsilon \to 0$.*

Proof. Suppose that $\bar{\mathbf{x}}$ is a limit point of $\{\mathbf{x}^{\epsilon_k}\}$ as $k \to \infty$ with $\mathbf{x}^{\epsilon_k}$ being an ϵ_k approximate first-order stationary point of (5.1) and $\epsilon_k \to 0$ as $k \to \infty$.

If $\bar{\mathbf{x}} = 0$, then $\mathrm{span}\{\mathcal{V}_{\bar{\mathbf{x}}}\} = \mathbb{R}^n$ and hence $Z_{\bar{\mathbf{x}}} = 0$, which means that (5.5) holds vacuously and hence $\bar{\mathbf{x}}$ is a first-order stationary point. Hence, we hereinafter assume that $\bar{\mathbf{x}} \neq 0$, that is, $\mathrm{span}\{\mathcal{V}_{\bar{\mathbf{x}}}\}$ is a proper subspace of $\mathbb{R}^n$.

First of all, we claim that there must exist $k_{\bar{\mathbf{x}}} \in \mathcal{N}$ such that $\mathcal{V}_k \subset \mathcal{V}_{\bar{\mathbf{x}}}$ for any $k \geqslant k_{\bar{\mathbf{x}}}$ with $\mathcal{V}_k := \mathcal{V}_{\mathbf{x}^{\epsilon_k}}^{\epsilon_k}$. Otherwise, there will be a subsequence of $\{\mathbf{x}^{\epsilon_k}\}$, say $\{\mathbf{x}^{\epsilon_{k_j}}\}$, such that $\lim_{j\to\infty} \epsilon_{k_j} = 0$ and $\mathcal{V}_{k_j} \nsubseteq \mathcal{V}_{\bar{\mathbf{x}}}$ for all j. Since $\mathcal{V}_{k_j} \subset \{U_1^{\mathrm{T}}, \ldots, U_m^{\mathrm{T}}\}$ which is a set consisting of finite many elements, then there must exist $U_{i_0}^{\mathrm{T}} \in \{U_1^{\mathrm{T}}, \ldots, U_m^{\mathrm{T}}\}$ such that $U_{i_0}^{\mathrm{T}} \in \mathcal{V}_{k_{j_t}}$ but $U_{i_0}^{\mathrm{T}} \notin \mathcal{V}_{\bar{\mathbf{x}}}$, where $\{k_{j_t}\} \subset \{k_j\}$, $t = 1, 2, \cdots$. For convenience, we hereinafter still use $\{k_j\}$ to denote its subsequence $\{k_{j_t}\}$. Hence, we have

$$\|U_{i_0}\mathbf{x}^{\epsilon_{k_j}}\|_1 \leqslant \epsilon_{k_j}.$$

Let j go to ∞, it follows from the above inequality that $\|U_{i_0}\bar{\mathbf{x}}\|_1 = 0$, which contradicts the fact that $U_{i_0}^{\mathrm{T}} \notin \mathcal{V}_{\bar{\mathbf{x}}}$. Thus, we conclude that $\mathcal{V}_k \subset \mathcal{V}_{\bar{\mathbf{x}}}$ for any $k \geqslant k_{\bar{\mathbf{x}}}$ with some $k_{\bar{\mathbf{x}}} \in \mathcal{N}$. Let $Z_k := Z_{\mathbf{x}^{\epsilon_k}}^{\epsilon_k}$, which is a matrix whose columns are an orthonormal basis of the null space of $\mathcal{V}_k$. For any $k \geqslant k_{\bar{\mathbf{x}}}$, $\mathcal{V}_k \subset \mathcal{V}_{\bar{\mathbf{x}}}$ implies $\mathrm{null}\{\mathcal{V}_{\bar{\mathbf{x}}}\} \subset \mathrm{null}\{\mathcal{V}_k\}$.

Therefore, for $k \geqslant k_{\bar{\mathbf{x}}}$, we can choose Z_k such that the columns of Z_k contains all columns of $Z_{\bar{\mathbf{x}}}$.

For any fixed ϵ_k approximate first-order stationary point $\mathbf{x}^{\epsilon_k}$, consider the following two minimization problems.

$$\begin{aligned} \text{minimize} \quad & g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d}, \\ \text{subject to} \quad & \mathbf{x}^{\epsilon_k} + \mathbf{d} \in \mathcal{F},\ \|\mathbf{d}\| \leqslant 1,\ \mathbf{d} \in R(Z_k); \end{aligned} \tag{5.11}$$

$$\begin{aligned} \text{minimize} \quad & g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d}, \\ \text{subject to} \quad & \mathbf{x}^{\epsilon_k} + \mathbf{d} \in \mathcal{F},\ \|\mathbf{d}\| \leqslant 1,\ \mathbf{d} \in R(Z_{\bar{\mathbf{x}}}). \end{aligned} \tag{5.12}$$

Since $\bar{\mathbf{d}} = 0$ is a feasible point of problem (5.11) and (5.12), the minimum values of (5.11) and (5.12) are both nonpositive. Moreover, it follows from $\mathrm{null}\{\mathcal{V}_{\bar{\mathbf{x}}}\} \subset \mathrm{null}\{\mathcal{V}_k\}$ that $R(Z_{\bar{\mathbf{x}}}) \subset R(Z_k)$, which implies that the minimum value of (5.12) is not smaller than that of (5.11).

Hence, from (5.10), we have that for any $\mathbf{x}^{\epsilon_k}$, there holds

$$\left| \underset{\mathbf{x}^{\epsilon_k}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1,\mathbf{d}\in R(Z_{\bar{\mathbf{x}}})}{\text{minimize}} g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d} \right| \leqslant \left| \underset{\mathbf{x}^{\epsilon_k}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1,\mathbf{d}\in R(Z_{\mathbf{x}^{\epsilon_k}})}{\text{minimize}} g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d} \right| \leqslant \epsilon_k. \tag{5.13}$$

Suppose that $\mathbf{d}^{\epsilon_k}$ is a minimizer of problem (5.12), then (5.13) implies that

$$-\epsilon_k \leqslant g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d}^{\epsilon_k} \leqslant 0, \tag{5.14}$$

where $\mathbf{d}^{\epsilon_k}$ should satisfy that $\mathbf{x}^{\epsilon_k} + \mathbf{d}^{\epsilon_k} \in \mathcal{F}$, $\|\mathbf{d}^{\epsilon_k}\| \leqslant 1$ and $\mathbf{d}^{\epsilon_k} \in R(Z_{\bar{\mathbf{x}}})$.

Since $\bar{\mathbf{x}} \neq 0$, it follows from the assumption on U_i $(i = 1, \ldots, m)$ that $J_{\bar{\mathbf{x}}} \neq \varnothing$ and thus $Z_{\bar{\mathbf{x}}}$ is an n by r matrix with $r > 0$. Then for any $\mathbf{d}^{\epsilon_k} \in R(Z_{\bar{\mathbf{x}}})$, there exists $\mathbf{h}^{\epsilon_k} \in R^r$ such that $\mathbf{d}^{\epsilon_k} = Z_{\bar{\mathbf{x}}}\mathbf{h}^{\epsilon_k}$ and $\|\mathbf{h}^{\epsilon_k}\| = \|\mathbf{d}^{\epsilon_k}\| \leqslant 1$ due to the definition of $Z_{\bar{\mathbf{x}}}$.

Note that

$$\begin{aligned} g(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d}^{\epsilon_k} &= \left(\nabla f_0(\mathbf{x}^{\epsilon_k}) + \sum_{i=1}^{m}\Phi(U_i\mathbf{x}^{\epsilon_k})U_i^{\mathrm{T}}\mathrm{sign}(U_i\mathbf{x}^{\epsilon_k})\right)^{\mathrm{T}}\mathbf{d}^{\epsilon_k} \\ &= \nabla f_0(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}\mathbf{d}^{\epsilon_k} + \sum_{i=1}^{m}\Phi(U_i\mathbf{x}^{\epsilon_k})\mathrm{sign}(U_i\mathbf{x}^{\epsilon_k})^{\mathrm{T}}U_i\mathbf{d}^{\epsilon_k} \\ &= \nabla f_0(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}Z_{\bar{\mathbf{x}}}\mathbf{h}^{\epsilon_k} + \sum_{i=1}^{m}\Phi(U_i\mathbf{x}^{\epsilon_k})\mathrm{sign}(U_i\mathbf{x}^{\epsilon_k})^{\mathrm{T}}(U_iZ_{\bar{\mathbf{x}}})\mathbf{h}^{\epsilon_k} \\ &= \nabla f_0(\mathbf{x}^{\epsilon_k})^{\mathrm{T}}Z_{\bar{\mathbf{x}}}\mathbf{h}^{\epsilon_k} + \sum_{i\in J_{\bar{\mathbf{x}}}}\Phi(U_i\mathbf{x}^{\epsilon_k})\mathrm{sign}(U_i\mathbf{x}^{\epsilon_k})^{\mathrm{T}}(U_iZ_{\bar{\mathbf{x}}})\mathbf{h}^{\epsilon_k}. \end{aligned} \tag{5.15}$$

From the compactness of $\{\mathbf{h} \mid \|\mathbf{h}\| \leqslant 1\}$, we know that there must exist a subsequence of $\{\mathbf{h}^{\epsilon_k}\}$ such that $\mathbf{h}^{\epsilon_{k_j}} \to \bar{\mathbf{h}} \in \mathbb{R}^r$ as j goes to infinity. For simplicity of notation, we still use $\{\mathbf{h}^{\epsilon_k}\}$ to denote its subsequence $\{\mathbf{h}^{\epsilon_{k_j}}\}$.

Since for $i \in J_{\bar{\mathbf{x}}}$, we have $\lim_{k\to\infty}\Phi(U_i\mathbf{x}^k) = \Phi(U_i\bar{\mathbf{x}})$, letting k go to infinity in (5.14) and (5.15), we immediately obtain that

$$0 = g(\bar{\mathbf{x}})^{\mathrm{T}}\bar{\mathbf{d}} = \nabla f_0(\bar{\mathbf{x}})^{\mathrm{T}}Z_{\bar{\mathbf{x}}}\bar{\mathbf{h}} + \sum_{i\in J_{\bar{\mathbf{x}}}}\Phi(U_i\bar{\mathbf{x}})\mathrm{sign}(U_i\bar{\mathbf{x}})^{\mathrm{T}}(U_iZ_{\bar{\mathbf{x}}})\bar{\mathbf{h}},$$

where $\bar{\mathbf{d}} = Z_{\bar{\mathbf{x}}}\bar{\mathbf{h}}$, which means that

$$\underset{\bar{\mathbf{x}}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1,\mathbf{d}\in R(Z_{\bar{\mathbf{x}}})}{\text{minimize}} g(\bar{\mathbf{x}})^{\mathrm{T}}\mathbf{d} = g(\bar{\mathbf{x}})^{\mathrm{T}}\bar{\mathbf{d}} = 0.$$

We complete the proof. □

For the special case where $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} - b \leqslant 0\}$, by the similar proof in Lemma 5.1, we can show that if $\bar{\mathbf{x}} \in \mathcal{F}$ is a local minimizer of (5.1), then there exists a nonnegative vector $\lambda \in \mathbb{R}^\ell$ such that

$$Z_{\bar{\mathbf{x}}}^{\mathrm{T}}\left(g(\bar{\mathbf{x}}) + A^{\mathrm{T}}\lambda\right) = 0, \tag{5.16a}$$

$$A\bar{\mathbf{x}} - b \leqslant 0,\ (A\bar{\mathbf{x}} - \mathbf{b})^{\mathrm{T}}\lambda = 0. \tag{5.16b}$$

This is the KKT condition for problem (5.1) with $\mathcal{F} := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} - \mathbf{b} \leqslant 0\}$ and $I_{\bar{\mathbf{x}}} = \varnothing$.

We can use the generalized KKT (5.16) to present the ϵ approximate stationary point given by Definition 5.2.

Proposition 5.2. *For $\epsilon > 0$, given $\mathbf{x} \in \mathcal{F}$, $\mathbf{x}$ is an ϵ approximately stationary point if and only if there exists a nonnegative vector $\lambda \in \mathbb{R}^\ell$ such that*

$$\|(Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}} \left(g(\mathbf{x}) + A^{\mathrm{T}}\lambda\right)\| + |(A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda| \leqslant \epsilon, \tag{5.17}$$

where, $Z_{\mathbf{x}}^\epsilon$ is an $n \times r_{\mathbf{x}}$ matrix whose columns form an orthonormal basis of the null space of $\mathcal{V}_{\mathbf{x}}^\epsilon$.

Proof. Suppose that $\|\mathbf{d}\| \leqslant 1$, $\mathbf{d} \in R(Z_{\mathbf{x}}^\epsilon)$ and $\mathbf{x} + \mathbf{d} \in \mathcal{F}$, then there exists $\mathbf{h} \in \mathbb{R}^{r\mathbf{x}}$ such that $\mathbf{d} = Z_{\mathbf{x}}^\epsilon \mathbf{h}$ and

$$\begin{aligned}
g(\mathbf{x})^{\mathrm{T}}\mathbf{d} = g(\mathbf{x})^{\mathrm{T}} Z_{\mathbf{x}}^\epsilon \mathbf{h} &= \left((Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\right)^{\mathrm{T}} \mathbf{h} - (AZ_{\mathbf{x}}^\epsilon \mathbf{h})^{\mathrm{T}}\lambda \\
&\geqslant \left((Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\right)^{\mathrm{T}} \mathbf{h} + (A\mathbf{x}^\epsilon - \mathbf{b})^{\mathrm{T}}\lambda \\
&\geqslant -\left|\left((Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\right)^{\mathrm{T}} \mathbf{h}\right| + (A\mathbf{x}^\epsilon - \mathbf{b})^{\mathrm{T}}\lambda \\
&\geqslant -\|(Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\|\|\mathbf{h}\| + (A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda \\
&\geqslant -\|(Z_{\mathbf{x}}^\epsilon)^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\| + (A\mathbf{x}^\epsilon - \mathbf{b})^{\mathrm{T}}\lambda,
\end{aligned} \tag{5.18}$$

where the first inequality is from the feasibility of $\mathbf{x} + d$ and the fourth results from $\|\mathbf{h}\| = \|\mathbf{d}\| \leqslant 1$. Since $\bar{\mathbf{d}} = 0$ is a feasible point with $g(\mathbf{x})^{\mathrm{T}}\bar{\mathbf{d}} = 0$, from (5.17) and (5.18), the minimal value of the minimization problem in Definition 5.2 is within $[-\epsilon, 0]$. Hence, $\mathbf{x}$ is an ϵ approximate stationary point.

Conversely, if $\mathbf{x}$ is an ϵ stationary point, consider the following constrained convex

minimization problem

$$\begin{aligned} \text{minimize} \quad & g(\mathbf{x})^{\mathrm{T}} Z\mathbf{h}, \\ \text{subject to} \quad & A(\mathbf{x}+Z\mathbf{h}) \leqslant \mathbf{b}, \\ & \|\mathbf{h}\| \leqslant 1. \end{aligned} \tag{5.19}$$

For simplicity of notation, we denote $Z := Z_{\mathbf{x}}^{\epsilon}$ in (5.19).

Denote by $C_1 = \{\mathbf{h} \mid A(\mathbf{x}+Z\mathbf{h}) \leqslant \mathbf{b}\}$ and $C_2 = \{\mathbf{h} \mid \|\mathbf{h}\| \leqslant 1\}$. Then $\mathbf{h}^*$ is a minimizer of problem (5.19) if and only if

$$0 \in Z^{\mathrm{T}} g(\mathbf{x}) + N_{C_1 \cap C_2}(\mathbf{h}^*), \tag{5.20}$$

where N_C designates the normal cone of the set C. Since $\mathbf{x}$ is feasible, $\bar{\mathbf{h}} = 0$ is a point of C_1 and the relative interior point of C_2 at $\mathbf{h}^*$. It follows from Corollary 23.8.1 in [87] that

$$N_{C_1 \cap C_2}(\mathbf{h}^*) = N_{C_1}(\mathbf{h}^*) + N_{C_2}(\mathbf{h}^*). \tag{5.21}$$

Note $N_{C_1} = \{(AZ)^{\mathrm{T}}\lambda \mid (A(\mathbf{x}+Z\mathbf{h}^*) - \mathbf{b})^{\mathrm{T}}\lambda = 0, \lambda \geqslant 0\}$. Hence, there exists a nonnegative vector $\lambda^* \in \mathbb{R}^{\ell}$ such that $(AZ)^{\mathrm{T}}\lambda^* \in N_{C_1}$ and $(A(\mathbf{x}+Z\mathbf{h}^*) - \mathbf{b})^{\mathrm{T}}\lambda^* = 0$. By (5.20) and (5.21), we obtain that

$$0 \in Z^{\mathrm{T}}(g(\mathbf{x}) + A^{\mathrm{T}}\lambda^*) + N_{C_2}(\mathbf{h}^*). \tag{5.22}$$

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & \left(g(\mathbf{x}) + A^{\mathrm{T}}\lambda^*\right)^{\mathrm{T}} Z\mathbf{h} + (A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda^*, \\ \text{subject to} \quad & \|\mathbf{h}\| \leqslant 1. \end{aligned} \tag{5.23}$$

Note that (5.22) implies that h^* is a minimizer of problem (5.23). Denote by $d(\lambda^*)$ the optimal value of problem (5.23), then we have that

$$\begin{aligned} \mathbf{d}(\lambda^*) &= \left(g(\mathbf{x}) + A^{\mathrm{T}}\lambda^*\right)^{\mathrm{T}} Z\mathbf{h}^* + (A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda^* \\ &= g(\mathbf{x})^{\mathrm{T}} Z\mathbf{h}^* + (A(\mathbf{x}+Z\mathbf{h}^*) - \mathbf{b})^{\mathrm{T}}\lambda^* \\ &= g(\mathbf{x})^{\mathrm{T}} Z\mathbf{h}^*. \end{aligned} \tag{5.24}$$

On the other hand, it follows from Cauchy-Schwartz inequality that for any $\mathbf{h} \in \{\mathbf{h} \mid \|\mathbf{h}\| \leqslant 1\}$

$$(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)^{\mathrm{T}} Z\mathbf{h} \geqslant -\|Z^T(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\|\|\mathbf{h}\| \geqslant -\|Z^T(g(\mathbf{x}) + A^{\mathrm{T}}\lambda)\|,$$

which implies that the optimal value of problem (5.23) is

$$\mathbf{d}(\lambda^*) = (A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda^* - \|Z^T(g(\mathbf{x}) + A^{\mathrm{T}}\lambda^*)\|. \tag{5.25}$$

Since $\mathbf{x}$ is feasible and an ϵ stationary point, we know that $(A\mathbf{x} - \mathbf{b})^{\mathrm{T}}\lambda^* \leqslant 0$ and $|g(\mathbf{x})Z\mathbf{h}^*| \leqslant \epsilon$. It immediately follows from (5.24) and (5.25) that (5.17) holds □

5.2 A Smoothing Method

We construct a twice continuously differentiable smoothing function of the absolute value function, namely,

$$\theta(t, \mu) = \begin{cases} |t|, & \text{if } |t| > \mu, \\ -\frac{1}{8\mu^3}t^4 + \frac{3}{4\mu}t^2 + \frac{3}{8}\mu, & \text{if } |t| \leqslant \mu. \end{cases} \tag{5.26}$$

By straightforward calculation, we can obtain

$$\theta'(t, \mu) = \begin{cases} \text{sign}(t), & \text{if } |t| > \mu, \\ -\frac{1}{2\mu^3}t^3 + \frac{3}{2\mu}t, & \text{if } |t| \leqslant \mu. \end{cases} \tag{5.27a}$$

$$\theta''(t, \mu) = \begin{cases} 0 & \text{if } |t| > \mu, \\ -\frac{3}{2\mu^3}t^2 + \frac{3}{2\mu}, & \text{if } |t| \leqslant \mu. \end{cases} \tag{5.27b}$$

and the Clarke subdifferential of θ''

$$\partial\theta''(t, \mu) = \begin{cases} \{0\}, & \text{if } |t| > \mu \\ [0, \frac{3}{\mu^2}], & \text{if } t = -\mu \\ [-\frac{3}{\mu^2}, 0], & \text{if } t = \mu \\ \{-\frac{3}{\mu^3}t\}, & \text{if } |t| < \mu. \end{cases} \tag{5.28}$$

It is easy to see that for any fixed $\mu > 0$, $\theta(\cdot, \mu)$ is convex, increasing in $[0, \infty)$ and decreasing in $(-\infty, 0]$. Moreover, $\theta(0, \mu) = \min_t \theta(t, \mu) = 3\mu/8$ for $\mu > 0$.

Replacing $|x_{i,j}|$ by $\theta(x_{i,j}, \mu)$ in problem (5.1), we obtain a smoothing function of f as follows

$$\tilde{f}(\mathbf{x}, \mu) := f_0(\mathbf{x}) + \sum_{i=1}^{m} \varphi(\Theta_i^q(x_i, \mu)),$$

where

$$\Theta_i(x_i, \mu) = \sum_{j=1}^{m_i} \theta(x_{i,j}, \mu).$$

Let

$$f_i(x_i) = \varphi(\|U_i \mathbf{x}\|_1^q), \quad i = 1, \ldots, m$$

and

$$\tilde{f}_i(x_i, \mu) = \varphi(\Theta_i^q(x_i, \mu)), \quad i = 1, \ldots, m.$$

Then

$$\tilde{f}(\mathbf{x}, \mu) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \tilde{f}_i(x_i, \mu)$$

satisfies assumptions Assumption 4.1-4.4 with $\mathcal{F} = \mathbb{R}^n$ and $p = 2$. And without loss of generality, we assume $f_{\text{low}} = 0$.

For a fixed $\mu > 0$, we apply Algorithm 2.1 with $p = 2$ to solve

$$\underset{\mathbf{x} \in \mathcal{F}}{\text{minimize}} \quad \tilde{f}(\mathbf{x}, \mu) := \sum_{i=1}^{m} \tilde{f}_i(x_i, \mu). \tag{5.29}$$

Since for any fixed $\mu > 0$, $\theta(t, \mu) > 0$, by Assumption 5.1 on function φ and (5.26), the smoothing function $\tilde{f}(\cdot, \mu)$ satisfies Assumption 4.2-4.4 with $p = 2$.

By Theorem 4.2, Algorithm 4.1 needs at most $O(\epsilon^{-3/2})$ evaluations of the objective function and its first and second derivatives to produce an ϵ approximate first-order

stationary point $\mathbf{x}$ of problem (5.29), namely,

$$\chi_{\tilde{f},1}(\mathbf{x}) := \left| \underset{\mathbf{x}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1}{\text{minimize}} \nabla \tilde{f}(\mathbf{x},\mu)^{\mathrm{T}}\mathbf{d} \right| \leqslant \epsilon. \tag{5.30}$$

However, these constants κ_S and κ^b in Theorem 4.2 are dependent on the Lipschitz constants of $\tilde{f}(\cdot,\mu)$, and its first and second derivatives which are dependent on μ. These constants can go to infinity as $\mu \to 0$.

Now we consider an ϵ approximate first-order stationary point defined in Definition 5.2, which is weaker than (5.30) but reduces to (5.30) when $I_x^\epsilon = \varnothing$. We will show that Algorithm 4.1 needs at most $O(\epsilon^{-3/2})$ evaluations of the smoothing objective function and its first and second derivatives to produce such ϵ approximate first-order stationary point $\mathbf{x}$ of problem (5.1). Moreover, these constants κ_S and κ^b are independent on μ.

It is worth noting that f is not differentiable and not Lipschitz continuous only at

$$D^c = \{\mathbf{x} : \ \|U_i\mathbf{x}\| = 0, \quad \text{for some } i \in \{1,\ldots,m\}\},$$

which means that the Lipschitz constants can go to infinity only some iterates approach to D^c. On the other hand, by Assumption 5.1, φ is continuous at 0, $\varphi'(t) \geqslant 0$, and $\varphi(0) = 0$, which means that we may neglect these terms in $\varphi(\|U_i\mathbf{x}\|_1^q)$ (5.1) when $\|U_i\mathbf{x}\|_1$ is very small. In particular, for the non-Lipschitz continuous problem (5.1), we want to find evaluation complexity for finding an ϵ approximate first-order stationary $\mathbf{x}$ satisfies the generalized ϵ approximate first-order condition (5.10).

In what follows, we consider

$$U_i \in R^{1\times n}, \quad i = 1,\ldots,m.$$

In such case, $f_i : \mathbb{R}_+ \to \mathbb{R}_+, i = 1, \ldots, m$. Moreover, it follows from (5.10) that

$$\begin{aligned}
\chi_{f,1}(\mathbf{x}) &= \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, d\in R(Z_x^\epsilon)} g(\mathbf{x})^{\mathrm{T}}\mathbf{d} \right| \\
&= \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\mathbf{x}}^\epsilon)} \nabla f_0(\mathbf{x})^{\mathrm{T}} d + \sum_{|x_i|>\epsilon} f_i'(x_i) U_i \mathbf{d} \right| \\
&\leqslant \epsilon.
\end{aligned}$$

For the smoothing function $f(\cdot, \mu)$ given by (5.26), we have

$$0 \leqslant \tilde{f}(x, \mu) - f(x) \leqslant c \sum_{|x_i|<\mu} (\frac{3\mu}{8})^q,$$

and if $\mu \leqslant \epsilon$, then $\tilde{f}_i(x_i, \mu) = f_i(x_i)$ and $\tilde{f}'(x_i, \mu) = f_i'(x_i)$ for $|x_i| > \epsilon$. Hence,

$$\begin{aligned}
\chi_{f,1}(\mathbf{x}) &= \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\mathbf{x}}^\epsilon)} \nabla f_0(x)^{\mathrm{T}}\mathbf{d} + \sum_{|x_i|>\epsilon} f_i'(x_i)^{\mathrm{T}} U_i \mathbf{d} \right| \\
&= \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\mathbf{x}}^\epsilon)} \nabla f_0(\mathbf{x})^{\mathrm{T}}\mathbf{d} + \sum_{|x_i|>\epsilon} \tilde{f}_i'(x_i, \mu)^{\mathrm{T}} U_i \mathbf{d} \right| \\
&= \left| \min_{\mathbf{x}+\mathbf{d}\in\mathcal{F}, \|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z_{\mathbf{x}}^\epsilon)} \nabla \tilde{f}(\mathbf{x}, \mu)^{\mathrm{T}}\mathbf{d} \right|
\end{aligned} \tag{5.31}$$

holds for $\mu \in (0, \epsilon]$.

Since $\mathbf{d} = 0$ is a feasible point of the minimization problems in (5.30) and (5.31), the optimal values of both minimization problem are not positive. Hence, we have that if $0 < \mu \leqslant \epsilon$, then

$$\chi_{f,1}(\mathbf{x}) \leqslant \chi_{\tilde{f},1}(\mathbf{x}),$$

which implies that an ϵ approximate first-order stationary point $\mathbf{x}$ of the smoothing minimization problem (5.29) is a generalized ϵ approximate first-order stationary point of the original minimization problem (5.1).

It follows from the first equality of (5.31) that $\chi_{f,1}(\mathbf{x})$ for $\mu \in (0, \epsilon]$ does not involve a smoothing function. We can provide the complexity bound which is independent from the smoothing parameter.

5.3 A Class of Special Cases

In this section, we consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{m} |U_i \mathbf{x}|^q \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n : \|A\mathbf{x} - \mathbf{b}\| \leqslant \sigma\}, \end{aligned} \tag{5.32}$$

where $U_i \in \mathbb{R}^{1\times n}$ for $i = 1, \ldots, m$, $0 < q < 1$, $A \in \mathbb{R}^{\ell\times n}$, $b \in \mathbb{R}^{\ell}$ and $\sigma > 0$ are given. We assume that the feasible set $\mathcal{X}$ is a non-empty closed convex set and $\|\mathbf{b}\| > \sigma$, which means that the zero vector is not a feasible solution.

Let

$$U = \begin{bmatrix} U_1 \\ \vdots \\ U_m \end{bmatrix}, \quad \mathbf{y} = U\mathbf{x}.$$

We consider the following problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{y}) := \sum_{i=1}^{m} |y_i|^q \\ \text{subject to} \quad & \mathbf{y} \in \mathcal{F} := \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} \in R(U),\ \|B\mathbf{y} - \mathbf{b}\| \leqslant \sigma\}, \end{aligned} \tag{5.33}$$

where $B = A(U^{\mathrm{T}}U)^{-1}U^{\mathrm{T}} \in \mathbb{R}^{\ell\times m}$.

Note that problems (5.32) and (5.33) are equivalent in the following sense:

- if $\mathbf{x}^*$ is a solution of (5.32), then $\mathbf{y}^* = U\mathbf{x}^*$ is a solution of (5.33);
- if $\mathbf{y}^*$ is a solution of (5.33), then any $\mathbf{x}^*$ satisfying $\mathbf{y}^* = U\mathbf{x}^*$ is a solution of (5.32).

5.3.1 The Smoothing Approximation

In what follows, we consider the complexity of Algorithm 4.1 for solving problem (5.33). We use the following smoothing function of $|t|$

$$\theta(t,\mu) = \begin{cases} |t|, & \text{if } |t| > \mu, \\ \frac{t^2}{2\mu} + \frac{\mu}{2}, & \text{if } |t| \leqslant \mu. \end{cases} \tag{5.34}$$

For any $\mu > 0$, $\theta(\cdot,\mu) : \mathbb{R} \to \mathbb{R}_{++}$ is continuously differentiable, convex, increasing in $[0,\infty)$ and decreasing in $(-\infty, 0]$. Moreover, $\theta(0,\mu) = \min_t \theta(t,\mu) = \mu/2$ for $\mu > 0$.

Hereinafter, when mentioning the smoothing function $\theta(\cdot,\mu)$, we refer to the smoothing function defined by (5.34). The gradient of the smoothing function $\theta(\cdot,\mu)$ is:

$$\theta'(t,\mu) = \begin{cases} \text{sign}(t), & \text{if } |t| \geqslant \mu, \\ \frac{t}{\mu}, & \text{if } |t| < \mu. \end{cases} \tag{5.35}$$

Let

$$f_i(y_i) = |y_i|^q, \quad i = 1,\ldots,m$$

and

$$\tilde{f}_i(y_i,\mu) = \theta^q(y_i,\mu), \quad i = 1,\ldots,m.$$

Then for a fixed $\mu > 0$, we obtain a smoothing version of problem (5.33) as follows:

$$\begin{aligned} \min \quad & \tilde{f}(\mathbf{y},\mu) := \sum_{i=1}^{m} \tilde{f}_i(y_i,\mu) \\ \text{s.t.} \quad & \mathbf{y} \in \mathcal{F} = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} \in R(U),\ \|B\mathbf{y} - b\| \leqslant \sigma\}. \end{aligned} \tag{5.36}$$

Let $I^{\epsilon}_{\mathbf{y}} = \{i \in \{1,\ldots,m\},\ |y_i| \leqslant \epsilon\}$ and $|I^{\epsilon}_{\mathbf{y}}| = r$. Then $Z^{\epsilon}_{\mathbf{y}} = [e_{j_1},...,e_{j_r}] \in \mathbb{R}^{m\times r}$ where e_{j_i} is the j_i column of the $m \times m$ identity matrix and $j_i \in I^{\epsilon}_{\mathbf{y}}$.

For fixed $\mu \in (0, \epsilon]$, we have

$$\begin{aligned}\chi_{f,1}(\mathbf{y}) = \left|\min_{\mathbf{y}+\mathbf{d}\in\mathcal{F},\|\mathbf{y}\|\leqslant 1, d\in R(Z^{\epsilon}_{\mathbf{y}})} g(\mathbf{y})^{\mathrm{T}}\mathbf{d}\right| &= \left|\min_{\mathbf{y}+\mathbf{d}\in\mathcal{F},\|\mathbf{y}\|\leqslant 1} \sum_{|y_i|\geqslant\epsilon} f_i'(y_i)d_i\right| \\ &= \left|\min_{\mathbf{y}+\mathbf{d}\in\mathcal{F},\|\mathbf{y}\|\leqslant 1} \sum_{|y_i|\geqslant\epsilon} \tilde{f}_i'(y_i,\mu)d_i\right| \\ &= \left|\min_{\mathbf{y}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1, \mathbf{d}\in R(Z^{\epsilon}_{\mathbf{y}})} \nabla\tilde{f}(\mathbf{y},\mu)^{\mathrm{T}}\mathbf{d}\right|,\end{aligned} \tag{5.37}$$

where d_i is the i-th element of $\mathbf{d}$ and $g(\mathbf{y}) = \sum_{i=1}^{m} \Phi(x_i)e_i\mathrm{sign}(x_i)$ with

$$\Phi(x_i) = \begin{cases} f_i'(x_i), & \text{if } |y_i| > \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

Then by the same argument with the above section, an ϵ approximate first-order stationary point $\mathbf{y}$ of the smoothing minimization problem (5.36) is a generalized ϵ approximate first-order stationary point of the original minimization problem (5.33) due to

$$\chi_{f,1}(\mathbf{y}) \leqslant \chi_{\tilde{f},1}(\mathbf{y}) = \left|\min_{\mathbf{y}+\mathbf{d}\in\mathcal{F},\|\mathbf{d}\|\leqslant 1} \nabla\tilde{f}(\mathbf{y},\mu)^{\mathrm{T}}\mathbf{d}\right|.$$

So for $\mu \in (0, \epsilon]$, the second equality sign in (5.37) shows that $\chi_{f,1}(\mathbf{y})$ does not involve the smoothing parameter μ as well.

5.3.2 The Model with Adaptive Cubic Regularization

Consider the univariate exponential function t^q for $t \in (0, \infty)$. Since $q(q-1)(q-2)(q-3)t^{q-4} < 0$ holds for all $t > 0$, it is easy to see that for any $t_+, t > 0$, there holds

$$t_+^q \leqslant t^q + qt^{q-1}(t_+ - t) + \frac{q(q-1)t^{q-2}}{2}(t_+ - t)^2 + \frac{q(q-1)(q-2)t^{q-3}}{6}(t_+ - t)^3. \tag{5.38}$$

Let

$$\mathbf{y} = (y_1, \ldots, y_m)^{\mathrm{T}} \in \mathbb{R}^m \quad \text{and} \quad \mathbf{s} = (s_1, \ldots, s_m)^{\mathrm{T}} \in \mathbb{R}^m.$$

Note that $\theta(\cdot, \mu) > 0$, replacing t_+ and t with $\theta(y_i + s_i, \mu)$ and $\theta(y_i, \mu)$ in (5.38), we know that

$$\tilde{f}_i(y_i + s_i, \mu) \leqslant T_{\tilde{f}_i,2}(y_i, s_i, \mu) + \frac{q(q-1)(q-2)\theta^{q-3}(y_i, \mu)}{6}(\delta\theta_i)^3, \tag{5.39}$$

where

$$T_{\tilde{f}_i,2}(y_i, s_i, \mu) = \tilde{f}_i(y_i, \mu) + q\theta^{q-1}(y_i, \mu)\delta\theta_i + \frac{q(q-1)\theta^{q-2}(y_i, \mu)}{2}(\delta\theta_i)^2$$

and

$$\delta\theta_i = \theta(y_i + s_i, \mu) - \theta(y_i, \mu). \tag{5.40}$$

It is worth noting that in $T_{\tilde{f}_i,2}$, we do not use the first and second derivatives of $\tilde{f}$.

It follows from (5.35) that

$$|\delta\theta_i| = |\theta(y_i + s_i, \mu) - \theta(y_i, \mu)| \leqslant |s_i|, \quad i = 1, \ldots, m. \tag{5.41}$$

For fixed μ, denote $v_i = \theta(y_i, \mu)$ and

$$D = \begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_m \end{bmatrix}, \quad \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}.$$

It follows from $\theta(y_i, \mu) \geqslant \frac{\mu}{2} > 0$ that $v_i > 0$ for $i = 1, \ldots, m$ and hence the diagonal matrix D is positive definite. Thus, for any $\mathbf{s} \in \mathbb{R}^m$, there must exist a unique $\alpha \in \mathbb{R}^m$ such that

$$\mathbf{s} = D\alpha. \tag{5.42}$$

Let $\alpha = (\alpha_1, \ldots, \alpha_m)^{\mathrm{T}}$ be given such that (5.42) holds, then $s_i = \alpha_i v_i$.

Combining (5.39), (5.41) and (5.42), we have that for all $i \in \{1, \ldots, m\}$

$$\tilde{f}_i(y_i + s_i, \mu) \leqslant T_{\tilde{f}_i,2}(y_i, \alpha_i, \mu) + \frac{q(q-1)(q-2)\theta^q(y_i, \mu)}{6}|\alpha_i|^3, \tag{5.43}$$

where $s_i = \alpha_i v_i$ and

$$T_{\tilde{f}_i,2}(y_i, \alpha_i, \mu) = \tilde{f}_i(y_i, \mu) + q\theta^{q-1}(y_i, \mu)\delta\theta_i(\alpha_i) + \frac{q(q-1)\theta^{q-2}(y_i, \mu)}{2}(\delta\theta_i(\alpha_i))^2,$$

with

$$\delta\theta_i(\alpha_i) = \theta(y_i + \alpha_i v_i, \mu) - \theta(y_i, \mu). \tag{5.44}$$

It is not difficult to see the objective function $\tilde{f}(\mathbf{y}, \mu)$ is level bounded, i.e., for any $y_0 \in \mathcal{F}$, the level set $\{\|\mathbf{y}\|_\infty : \tilde{f}(\mathbf{y}, \mu) \leqslant \tilde{f}(\mathbf{y}_0, \mu)\}$ is bounded. Then there exists a constant, say $\Gamma > 1$, such that $|\tilde{f}_i(y_i, \mu)|^q \leqslant \Gamma$ for all $i = 1, \ldots, m$.

Therefore, (5.43) further implies that

$$\tilde{f}_i(y_i + s_i, \mu) \leqslant T_{\tilde{f}_i,2}(y_i, \alpha_i, \mu) + \frac{L_i}{3}|\alpha_i|^3, \tag{5.45}$$

where $s_i = \alpha_i v_i$ and $L_i = \frac{1}{4}\Gamma$ due to $q(1-q)(2-q) \leqslant \frac{1}{2}$.

Since $v_i = \theta(y_i, \mu)$ is only determined by y_i and μ for $i \in \{1, \ldots, m\}$, then for fixed $\mathbf{y}$ and μ, it is easy to see that for $s_i = \alpha_i v_i$ with $i \in \{1, \ldots, m\}$

$$\frac{\partial \tilde{f}(\mathbf{y} + D\alpha, \mu)}{\partial \alpha_i} = \frac{\partial \tilde{f}(\mathbf{y} + \mathbf{s}, \mu)}{\partial s_i} v_i,$$

which implies that

$$\nabla_\alpha \tilde{f}(\mathbf{y} + \mathbf{s}, \mu) = D\nabla_{\mathbf{s}} \tilde{f}(\mathbf{y} + \mathbf{s}, \mu) \tag{5.46}$$

if α and $\mathbf{s}$ are given such that (5.42) holds.

Since the quantity L_i in (5.45) is usually unknown in practice, we hence replace this term with an adaptive parameter σ_i, which yields the following third-order model for the i-th element:

$$\tilde{m}_i(y_i, \alpha_i, \mu) = T_{\tilde{f}_i,2}(y_i, \alpha_i, \mu) + \frac{\sigma_i}{3}|\alpha_i|^3. \tag{5.47}$$

Summing up those element models, we obtain the full model given by

$$\tilde{m}(\mathbf{y},\alpha,\mu)=\sum_{i=1}^{m}\tilde{m}_i(y_i,\alpha_i,\mu)=T_{\tilde{f},2}(\mathbf{y},\alpha,\mu)+\frac{1}{3}\sum_{i=1}^{m}\sigma_i|\alpha_i|^3, \tag{5.48}$$

where

$$T_{\tilde{f},2}(\mathbf{y},\alpha,\mu)=\sum_{i=1}^{m}T_{\tilde{f}_i,2}(y_i,\alpha_i,\mu).$$

5.3.3 The Criticality Measure

For fixed $\mathbf{y}$ and $\mu\in(0,\epsilon]$, let $\mathbf{s}=D\alpha$, then the first-order criticality measure of (5.36) in α is as follows,

$$\chi_{\tilde{f},1}(\mathbf{y}+D\alpha):=\left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\nabla\tilde{f}(\mathbf{y}+D\alpha,\mu)^{\mathrm{T}}D\beta\right| \tag{5.49}$$

which is obviously dependent on the smoothing parameter μ.

It follows from Lemma 5.1 that the criticality measure of (5.33) in α can be defined as

$$\chi_{f,1}(y+D\alpha):=\left|\min_{y+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1,D\beta\in R(Z^{\epsilon}_{y+D\alpha})}(Dg(y+D\alpha))^{\mathrm{T}}\beta\right|, \tag{5.50}$$

where $g(x)$ is given by (5.4).

From the definitions of $Z^{\epsilon}_{\mathbf{y}+D\alpha}$ and $g(\mathbf{y}+D\alpha)$, (5.50) can be rewritten as follows

$$\chi_{f,1}(\mathbf{y}+D\alpha)=\left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\sum_{|y_i+\alpha_i v_i|\geqslant\epsilon}f_i'(y_i+\alpha_i v_i)(D\beta)_i\right|. \tag{5.51}$$

By the definition of the smoothing function $\tilde{f}_i(y_i,\mu)$, we can further express (5.51)

as the following form,

$$\begin{aligned}\chi_{f,1}(\mathbf{y}+D\alpha) &= \left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\sum_{|y_i+\alpha_i v_i|\geqslant\epsilon}\tilde{f}_i'(y_i+\alpha_i v_i,\mu)(D\beta)_i\right| \\ &= \left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1, D\beta\in R(Z^{\epsilon}_{\mathbf{y}+D\alpha})}\nabla\tilde{f}(\mathbf{y}+D\alpha,\mu)^{\mathrm{T}}D\beta\right|\end{aligned} \tag{5.52}$$

Compare (5.51) with (5.49), we know that

$$\chi_{f,1}(\mathbf{y}+D\alpha)\leqslant\chi_{\tilde{f},1}(\mathbf{y}+D\alpha),$$

which implies that an ϵ approximate first-order stationary point $\mathbf{x}$ of problem (5.36) is a generalized ϵ approximate first-order stationary point of problem (5.33). (5.51) tells us that $\chi_{f,1}(\mathbf{y}+D\alpha)$ will do not involve the smoothing parameter μ if μ is set to satisfy $0<\mu\leqslant\epsilon$. Furthermore, we have

$$0\leqslant\tilde{f}(\mathbf{y}+D\alpha,\mu)-f(\mathbf{y}+D\alpha)\leqslant\sum_{|y_i+\alpha_i v_i|<\mu}(\frac{\mu}{2})^q.$$

For fixed $\mu\in(0,\epsilon]$, define

$$\begin{aligned}\tilde{\chi}_{\tilde{f},1}(\mathbf{y}+D\alpha) &:= \left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\sum_{|y_i+\alpha_i v_i|\geqslant\epsilon}\tilde{f}_i'(y_i+\alpha_i v_i,\mu)(D\beta)_i\right| \\ &= \left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\sum_{|y_i+\alpha_i v_i|\geqslant\epsilon}f_i'(y_i+\alpha_i v_i)(D\beta)_i\right| \\ &= \left|\min_{\mathbf{y}+D\alpha+D\beta\in\mathcal{F},\|\beta\|\leqslant 1}\sum_{|y_i+\alpha_i v_i|\geqslant\epsilon}(f_i)'_{\alpha_i}(y_i+\alpha_i v_i)\beta_i\right|\end{aligned} \tag{5.53}$$

where $(f_i)'_{\alpha_i}$ denote the derivative of f_i in α_i. And the $\tilde{\chi}_{\tilde{f},1}$ does not involve the smoothing parameter μ for $0<\mu\leqslant\epsilon$ according the above argument.

5.3.4 The Algorithm

The algorithm considered in this section exploits the model (5.48) as follows. At each iteration k, the model (5.48) taken at the iterate $\mathbf{y} = \mathbf{y}_k$ is (approximately) minimized to attain an α_k which can be used to define a step $\mathbf{s}_k$. If the decrease in the objective function value along $\mathbf{s}_k$ is comparable to that predicted by the Taylor model, the trial point $\mathbf{y}_k + \mathbf{s}_k$ is accepted as the new iterate and the regularization parameter $\sigma_{i,k}$ possibly updated. The process is terminated when an approximate local minimizer is found.

Suppose that $\mathbf{y} = \mathbf{y}_k$ is the iterate at iteration k. Let $\mathbf{y}_k = (y_{1,k}, \ldots, y_{m,k})$, $\mathbf{s}_k = (s_{1,k}, \ldots, s_{m,k})$ and $s_{i,k} = \alpha_{i,k} v_{i,k}$ with $v_{i,k} = \theta(y_{i,k}, \mu)$ for $i = 1, \ldots, m$, then $\mathbf{s}_k = D_k \alpha_k$ where

$$D_k = \begin{bmatrix} v_{1,k} & & \\ & \ddots & \\ & & v_{m,k} \end{bmatrix}, \quad \alpha_k = \begin{bmatrix} \alpha_{1,k} \\ \vdots \\ \alpha_{m,k} \end{bmatrix}.$$

Having defined the criticality measure (5.53), it is natural to use it also for terminating the approximate model minimization: we therefore minimize $\tilde{m}(\mathbf{y}_k, \alpha, \mu)$ over α until

$$\|\tilde{\chi}_{\tilde{m},1}(\mathbf{y}_k + D_k \alpha)\| \leqslant \kappa \|\alpha\|^2. \tag{5.54}$$

for some constant $\kappa > 0$ to obtain α_k for constructing the new step as $\mathbf{s}_k = D_k \alpha_k$ and thus the trial point $\mathbf{y}_{k+1} = \mathbf{y}_k + D_k \alpha_k$. In the algorithm, we further add restrictions on the step size and the smoothing parameter as follows.

$$|\alpha_{i,k}| \leqslant \rho, \quad \mu \in (0, \epsilon/(1+\rho)], \tag{5.55}$$

where $\rho \in (0, 1)$.

For $|y_{i,k+1}| \geqslant \epsilon$, we have

$$|y_{i,k}| \geqslant \mu.$$

Otherwise,

$$|y_{i,k+1}| = |y_{i,k} + \alpha_{i,k} v_{i,k}| \leqslant |y_{i,k}| + |\alpha_{i,k}||v_{i,k}| = |y_{i,k}| + |\alpha_{i,k}||\frac{y_i^2}{2\mu} + \frac{\mu}{2}| < \epsilon,$$

which contradicts to $|y_{i,k+1}| \geqslant \epsilon$. Hence, we know that $v_{i,k} = \theta(y_{i,k}, \mu) = |y_{i,k}|$, which implies that

$$y_{i,k+1} y_{i,k} = (y_{i,k} + \alpha_{i,k} y_{i,k}) y_{i,k} = y_{i,k}^2 (1 + \alpha_{i,k}) \geqslant y_{i,k}^2 (1 - \rho) > 0.$$

Denote by

$$\delta \tilde{f}_{i,k} := \tilde{f}_i(y_{i,k}, \mu) - \tilde{f}_i(y_{i,k} + s_{i,k}, \mu) \text{ and } \delta \tilde{f}_k := \tilde{f}(\mathbf{y}_k, \mu) - \tilde{f}(\mathbf{y}_k + \mathbf{s}_k, \mu) = \sum_{i=1}^{m} \delta \tilde{f}_{i,k},$$

with $\mathbf{s}_k = D_k \alpha_k$ and

$$\delta \tilde{m}_{i,k} := \tilde{m}_i(y_{i,k}, 0, \mu) - \tilde{m}_i(y_{i,k}, \alpha_{i,k}, \mu) \text{ and } \delta \tilde{m}_k := m(\mathbf{y}_k, 0, \mu) - m(\mathbf{y}_k, \alpha_k, \mu) = \sum_{i=1}^{m} \delta \tilde{m}_{i,k}.$$

Also let

$$\delta T_k := T_{\tilde{f},2}(\mathbf{y}_k, 0, \mu) - T_{\tilde{f},2}(\mathbf{y}_k, \alpha_k, \mu).$$

The partially separable adaptive regularization algorithm for solving problem

(5.33) is now stated as Algorithm 5.1 as follows.

Algorithm 5.1: Non-Lipschitz Partially Separable Adaptive Regularization

/* Initialization */

Input $\mathbf{y}_{-1}$, $\rho > 0$ and $\mu \in (0, \epsilon/(1+\rho)]$, $\{\sigma_{i,0}\}_{i=1}^m > 0$ as well as constants $0 < \gamma_0 < 1 < \gamma_1 < \gamma_2$, $\eta \in (0,1)$, $\kappa \geqslant 0$, $\sigma_{\min} \in (0, \min_{i=1,\ldots,m} \sigma_{i,0}]$ and $\kappa_{\text{big}} > 1$. Initialize $\mathbf{y}_0$ as the projection of $\mathbf{y}_{-1}$ onto $\mathcal{F}$ and set $k = 0$;

/* Main iteration */

Evaluate $\tilde{f}(\mathbf{y}_k, \mu)$, $\{\theta^{q-1}(y_{i,k}, \mu)\}_{i=1}^m$ and $\{\theta^{q-2}(y_{i,k}, \mu)\}_{i=1}^m$;

if $\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_k) \leqslant \epsilon$ **then**

  return $\mathbf{y}_\epsilon = \mathbf{y}_k$ and terminate;

else

  Minimize $\tilde{m}(\mathbf{y}_k, \alpha, \mu)$ to return an α_k such that $\mathbf{y}_k + D_k\alpha_k \in \mathcal{F}$, $\tilde{m}(\mathbf{y}_k, \alpha_k, \mu) \leqslant \tilde{m}(\mathbf{y}_k, 0, \mu)$, (5.54) and (5.55) hold;

  Compute

$$\rho_k = \frac{\delta \tilde{f}_k}{\delta T_k} \tag{5.56}$$

  and set $\mathbf{y}_{k+1} = \mathbf{y}_k + D_k\alpha_k$ if $\rho_k \geqslant \eta$, or $\mathbf{y}_{k+1} = \mathbf{y}_k$ otherwise;

  if

$$\tilde{f}_i(y_{i,k} + \alpha_{i,k} v_{i,k}, \mu) > \tilde{m}_i(y_{i,k}, \alpha_{i,k}, \mu), \quad \forall i \in \{1, \ldots, m\} \tag{5.57}$$

  then

    set

$$\sigma_{i,k+1} \in [\gamma_1 \sigma_{i,k}, \gamma_2 \sigma_{i,k}]. \tag{5.58}$$

  else if

$$\rho_k \geqslant \eta \text{ and } \delta\tilde{f}_{i,k} \leqslant 0 \text{ and } \delta\tilde{f}_{i,k} < \delta\tilde{m}_{i,k} - \kappa_{\text{big}}|\delta\tilde{f}_k| \tag{5.59}$$

  or

$$\rho_k \geqslant \eta \text{ and } \delta\tilde{f}_{i,k} > 0 \text{ and } \delta\tilde{f}_{i,k} > \delta\tilde{m}_{i,k} + \kappa_{\text{big}}|\delta\tilde{f}_k| \tag{5.60}$$

  then

    set

$$\sigma_{i,k+1} \in [\max[\sigma_{\min}, \gamma_0 \sigma_{i,k}], \sigma_{i,k}]. \tag{5.61}$$

  else

    set

$$\sigma_{i,k+1} = \sigma_{i,k}. \tag{5.62}$$

Increment k by one and go to Step 2;

5.4 Complexity Analysis

To begin with, we cite a well-know result on the equivalence of different norms in Euclidean space.

Proposition 5.3. *Let* $1 \leqslant s \leqslant t < \infty$ *and* $\alpha \in \mathbb{R}^m$, *then*

$$\|\alpha\|_t \leqslant \|\alpha\|_s \leqslant m^{\frac{1}{s}-\frac{1}{t}}\|\alpha\|_t, \tag{5.63}$$

where $\|\alpha\|_r := (\sum_{i=1}^m |\alpha_i|^r)^{\frac{1}{r}}$ *for any* $r \geqslant 1$.

The definition of the model in (5.48) implies a simple lower bound on the model decrease.

Lemma 5.2. *For all* $k \geqslant 0$,

$$\delta T_k \geqslant \frac{1}{3}\sigma_{\min}\sum_{i=1}^m |\alpha_{i,k}|^3, \tag{5.64}$$

$\alpha_k \neq 0$ *and* (5.56) *is well-defined.*

Proof. The bound directly follows from the observation that

$$T_{\tilde{f},2}(\mathbf{y}_k, 0, \mu) = \tilde{f}(\mathbf{y}_k, \mu) = \tilde{m}(\mathbf{y}_k, 0, \mu) \geqslant \tilde{m}(\mathbf{y}_k, \alpha_k, \mu) = T_{\tilde{f},2}(\mathbf{y}_k, \alpha_k, \mu) + \frac{1}{3}\sum_{i=1}^m \sigma_{i,k}|\alpha_{i,k}|^3$$

and (5.61). Moreover, $\tilde{\chi}_{\tilde{m},1}(\mathbf{y}_k) = \tilde{\chi}_{\tilde{f},1}(\mathbf{y}_k) > \epsilon$. As a consequence, (5.54) cannot hold for $\alpha_k = 0$ since termination would have then occurred in Step 3 of Algorithm 5.1. Hence at least one $|\alpha_{i,k}|$ is strictly positive due to (5.63) when $t = 2$ and $s = 1$. Moreover, (5.64) implies that (5.56) is well-defined. □

Our next step is to verify that regularization parameters cannot grow unbounded.

Lemma 5.3. *For all* $k \geqslant 0$,

$$\sigma_{i,k} \in [\sigma_{\min}, \sigma_{\max}] \tag{5.65}$$

where $\sigma_{\max} := \gamma_2\Gamma$.

Proof. Assume that, for some $i \in \{1, \ldots, m\}$ and $k \geqslant 0$, $\sigma_{i,k} \geqslant L_i$. Then (5.45) gives that

$$\tilde{f}_i(y_{i,k}+\alpha_{i,k}v_{i,k}, \mu) \leqslant T_{\tilde{f}_i,2}(y_{i,k}, \alpha_{i,k}, \mu)+\frac{1}{3}L_i|\alpha_{i,k}|^3 = \tilde{m}_i(y_{i,k}, \alpha_{i,k}, \mu)+\frac{1}{3}|\alpha_{i,k}|^3(L_i-\sigma_{i,k})$$

and (5.57) fails, ensuring (5.65) because of the mechanism of the algorithm. □

The following lemma reveals that for $|y_{i,k+1}| \geqslant \epsilon$, the derivatives of the element smoothing functions and the element cubic models can be bounded by $\|\alpha_k\|^2$ if the step size $\alpha_{i,k}$ and the smoothing parameter μ satisfies (5.55).

Lemma 5.4. *Given $\epsilon > 0$, if the step size and the smoothing parameter satisfies (5.55), then there holds that*

$$\left\| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})e_i^{\mathrm{T}} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k}, \alpha_{i,k})e_i^{\mathrm{T}} \right\| \leqslant \left(\frac{1}{4}\Gamma(1-\rho)^{-3} + \sigma_{\max}\right) \|\alpha_k\|^2. \tag{5.66}$$

where e_i is the i-th column of the m-dimensional identity matrix.

Proof. By the above argument, we know that for $|y_{i,k+1}| \geqslant \epsilon$, it holds

$$|y_{i,k}| \geqslant \mu, \quad y_{i,k+1}y_{i,k} > 0$$

if the step size $\alpha_{i,k}$ and the smoothing parameter μ satisfy (5.55).

For simplicity, we consider $y_{i,k} > 0$ and $y_{i,k+1} > 0$. Then for $y_{i,k+1} = y_{i,k} + \alpha_{i,k}v_{i,k} > \epsilon$, we know that

$$\tilde{f}_i(y_{i,k+1}, \mu) = f_i(y_{i,k+1}) = y_{i,k+1}^q,$$

$$\tilde{f}_i(y_{i,k}, \mu) = f_i(y_{i,k}) = y_{i,k}^q,$$

$$v_{i,k} = \theta(y_{i,k}, \mu) = y_{i,k}.$$

It follows that for $y_{i,k+1} \geqslant \epsilon$

$$T_{\tilde{f},2}(y_{i,k}, \alpha_{i,k}, \mu) = y_{i,k}^q + q y_{i,k}^q \alpha_{i,k} + \frac{1}{2} q(q-1) y_{i,k}^q \alpha_{i,k}^2,$$

$$(T_{\tilde{f},2})'_{\alpha_i}(y_{i,k}, \alpha_{i,k}, \mu) = q y_{i,k}^q + q(q-1) y_{i,k}^q \alpha_{i,k}$$

and

$$(\tilde{f}_i)'_{\alpha_i}(y_{i,k+1}, \mu) = (\tilde{f}_i)'_{\alpha_i}(y_{i,k} + \alpha_{i,k} y_{i,k}, \mu) = q(y_{i,k} + \alpha_{i,k} y_{i,k})^{q-1} y_{i,k}.$$

By Taylor's expansion, we can obtain that

$$(y_{i,k} + \alpha_{i,k} y_{i,k})^{q-1} = y_{i,k}^{q-1} + (q-1) y_{i,k}^{q-2} \alpha_{i,k} y_{i,k} + \frac{1}{2}(q-1)(q-2) \hat{y}_{i,k}^{q-3} \alpha_{i,k} y_{i,k}^2$$

where

$$\min(y_{i,k}, y_{i,k+1}) \leqslant \hat{y}_{i,k} \leqslant \max(y_{i,k}, y_{i,k+1}).$$

Hence, we have

$$\begin{aligned}(\tilde{f}_i)'_{\alpha_i}(y_{i,k+1}, \mu) &= q y_{i,k}^q + q(q-1) y_{i,k}^q \alpha_{i,k} + \frac{1}{2} q(q-1)(q-2) \hat{y}_{i,k}^q \hat{y}_{i,k}^{-3} \alpha_{i,k}^2 y_{i,k}^3 \\ &= (T_{\tilde{f}_i,2})'_{\alpha_i}(y_{i,k}, \alpha_{i,k}, \mu) + \frac{1}{2} q(q-1)(q-2) \hat{y}_{i,k}^q \hat{y}_{i,k}^{-3} \alpha_{i,k}^2 y_{i,k}^3.\end{aligned}$$

It is not difficult to see that

$$\hat{y}_{i,k}^{-3} y_{i,k}^3 \leqslant \begin{cases} 1, & \text{if } y_{i,k} \leqslant y_{i,k+1}, \\ y_{i,k+1}^{-3} y_i^3, & \text{otherwise,} \end{cases} \tag{5.67}$$

and

$$y_{i,k+1}^{-3} y_{i,k}^3 = (1 + \alpha_i)^{-3} \leqslant (1 - \rho)^{-3}, \tag{5.68}$$

which still holds for $y_{i,k} < 0$ and $y_{i,k+1} < 0$ by the same argument.

Therefore, for $|y_{i,k+1}| = |y_{i,k} + \alpha_{i,k} y_{i,k}| \geqslant \epsilon$, we have

$$|(f_i)'_{\alpha_i}(y_{i,k} + \alpha_{i,k} y_{i,k}) - \tilde{m}'_{\alpha_i}(y_{i,k}, \alpha_{i,k})| \leqslant \frac{1}{4} \Gamma (1 - \rho)^{-3} \alpha_{i,k}^2 + \sigma_{i,k} \alpha_{i,k}^2,$$

which implies

$$\begin{aligned}
&\left\| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})e_i^{\mathrm{T}} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})e_i^{\mathrm{T}} \right\| \\
&= \left\| \sum_{|y_{i,k+1}|\geqslant\epsilon} \left[(f_i)'_{\alpha_i}(y_{i,k+1}) - (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\right] e_i^{\mathrm{T}} \right\| \\
&\leqslant \left(\frac{1}{4}\Gamma(1-\rho)^{-3} + \sigma_{\max}\right) \|\alpha_k\|^2.
\end{aligned}$$

The above inequality gives the desired result. □

The final technical ingredient is to ensure that the step α_k are sufficiently large compared to the considered criticality measures computed at the next iterate. Let

$$\mathcal{S} = \{k \geqslant 0 \mid \rho_k \geqslant \eta\} \quad \text{and} \quad \mathcal{U} = \{k \geqslant 0 \mid \rho_k < \eta\}.$$

Lemma 5.5. *For all $k \in \mathcal{S}$ and $i = 1, \ldots, m$, if $\mathbf{y}_{k+1} = \mathbf{y}_k + D_k\alpha_k$, then there holds*

$$\|\alpha_k\| \geqslant \left[\frac{\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_{k+1})}{2\left(\frac{1}{4}\Gamma(1-\rho)^{-3} + \sigma_{\max}\right) + \kappa)}\right]^{\frac{1}{2}}. \tag{5.69}$$

Proof. Note that $\mathbf{y}_{k+1} = \mathbf{y}_k + D_k\alpha_k$ with the step size α_k and the smoothing parameter μ satisfying (5.55). It follows from (5.66) in Lemma 5.4 that

$$\left\| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})e_i^{\mathrm{T}} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})e_i^{\mathrm{T}} \right\| \leqslant \left(\frac{1}{4}\Gamma(1-\rho)^{-3} + \sigma_{\max}\right) \|\alpha_k\|^2.$$

It follows that

$$\begin{aligned}\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_{k+1}) &= \left| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1} \right| \\ &\leqslant \left| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1} \right| \qquad (5.70) \\ &\quad + \left| \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1} \right|\end{aligned}$$

where the first equality defines the vector $\beta_{k+1} = (\beta_{1,k+1},\ldots,\beta_{m,k+1})^{\mathrm{T}}$ with

$$\mathbf{y}_{k+1} + D_k\beta_{k+1} \in \mathcal{F}, \quad \|\beta_{k+1}\| \leqslant 1. \qquad (5.71)$$

Assume now, for the purpose of deriving a contradiction, that (5.69) fails at iteration $k \in \mathcal{S}$. Using (5.71), we obtain that

$$\begin{aligned}&- \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1} + \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1} \\ \leqslant &\left| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1} \right| \\ = &\left| \sum_{|y_{i,k+1}|\geqslant\epsilon} \left[(f_i)'_{\alpha_i}(y_{i,k+1}) - (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\right] e_i^{\mathrm{T}}\beta_{k+1} \right| \qquad (5.72) \\ \leqslant &\left\| \sum_{|y_{i,k+1}|\geqslant\epsilon} (f_i)'_{\alpha_i}(y_{i,k+1})e_i^{\mathrm{T}} - \sum_{|y_{i,k+1}|\geqslant\epsilon} (\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})e_i^{\mathrm{T}} \right\| \|\beta_{k+1}\| \\ < &\left(\frac{1}{4}\Gamma(1-\rho)^{-3} + \sigma_{\max}\right)\|\alpha_k\|^2.\end{aligned}$$

The failure of (5.69) and the first part of (5.70) then imply that

$$-\sum_{|y_{i,k+1}|\geqslant\epsilon}(f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1}+\sum_{|y_{i,k+1}|\geqslant\epsilon}(\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1}$$

$$<\frac{1}{2}\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_{k+1})=-\frac{1}{2}\sum_{|y_{i,k+1}|\geqslant\epsilon}(f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1},$$

which in turn ensures that

$$\sum_{|y_{i,k+1}|\geqslant\epsilon}(\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1}<\frac{1}{2}\sum_{|y_{i,k+1}|\geqslant\epsilon}(f_i)'_{\alpha_i}(\mathbf{y}_{k+1})\beta_{i,k+1}.$$

Moreover, $\mathbf{y}_{k+1}+D_k\beta_{k+1}\in\mathcal{F}$ by definition of $\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_{k+1})$, and hence, using (5.71),

$$\left|\sum_{|y_{i,k+1}|\geqslant\epsilon}(\tilde{m}_i)'_{\alpha}(y_{i,k},\alpha_{i,k})\beta_{i,k+1}\right|\leqslant\tilde{\chi}_{\tilde{m},1}(\mathbf{y}_{k+1}).$$

We may then substitute this inequality in (5.70) to deduce as above that

$$\begin{aligned}\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_{k+1})&\leqslant\left|\sum_{|y_{i,k+1}|\geqslant\epsilon}(f_i)'_{\alpha_i}(y_{i,k+1})\beta_{i,k+1}-\sum_{|y_{i,k+1}|\geqslant\epsilon}(\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1}\right|\\&\quad+\left|\sum_{|y_{i,k+1}|\geqslant\epsilon}(\tilde{m}_i)'_{\alpha_i}(y_{i,k},\alpha_{i,k})\beta_{i,k+1}\right|\\&<\left(\frac{1}{4}\Gamma(1-\rho)^{-3}+\sigma_{\max}\right)\|\alpha_k\|^2\end{aligned}\tag{5.73}$$

where the last inequality results from (5.72) and (5.54). But this contradicts our assumption that (5.69) fails. Hence (5.69) must hold. □

We are now ready to consider our first complexity result, whose proof uses the definitions

$$\mathcal{S}_k:=\{0,\dots,k\}\cap\mathcal{S}\quad\text{and}\quad\mathcal{U}_k:=\{0,\dots,k\}\backslash\mathcal{S}_k$$

of the sets of successful and unsuccessful iterations, respectively.

Theorem 5.1. *Algorithm 5.1 requires at most*

$$\kappa_{\mathcal{S}}(\tilde{f}(\mathbf{y}_0,\mu)-\tilde{f}_{\text{low}})\epsilon^{-\frac{3}{2}}$$

successful iterations to return a point $\mathbf{y}_\epsilon \in \mathcal{F}$ *such that* $\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_k) \leqslant \epsilon$, *for*

$$\kappa_{\mathcal{S}} = \frac{3(\frac{1}{2}\Gamma(1-\rho)^{-3}+2\sigma_{\max}+2\kappa)^{\frac{3}{2}}}{\eta\sigma_{\min}m^{-\frac{1}{2}}}. \tag{5.74}$$

Proof. Let $k \in \mathcal{S}$ be index of a successful iteration before termination. As a consequence, we obtain, using Assumption 4.4 and Lemma 5.2, that

$$\begin{aligned}
\tilde{f}(\mathbf{y}_0,\mu)-\tilde{f}_{\text{low}} &\geqslant \tilde{f}(\mathbf{y}_0,\mu)-\tilde{f}(\mathbf{y}_{k+1},\mu)\\
&\geqslant \sum_{\ell\in\mathcal{S}_k}\left[\tilde{f}(\mathbf{y}_\ell,\mu)-\tilde{f}(\mathbf{y}_\ell+\mathbf{s}_\ell,\mu)\right]\\
&\geqslant \eta\sum_{\ell\in\mathcal{S}_k}\left[T_{\tilde{f},2}(\mathbf{y}_k,0,\mu)-T_{\tilde{f},2}(\mathbf{y}_\ell,\alpha_\ell,\mu)\right]\\
&\geqslant \frac{\eta}{3}|\mathcal{S}_k|\sigma_{\min}\min_{\ell\in\mathcal{S}_k}\left[\sum_{i=1}^{m}|\alpha_{i,\ell}|^3\right].
\end{aligned}$$

Hence we deduce, using the second inequality of (5.63) with $s=2$ and $t=3$, Lemma 5.5, the definition of $\sigma_{\max}$ in Lemma 5.3 and (5.69), that

$$\begin{aligned}
\tilde{f}(\mathbf{y}_0,\mu)-\tilde{f}_{\text{low}} &\geqslant \frac{\eta}{3}|\mathcal{S}_k|\sigma_{\min}m^{-\frac{1}{2}}\min_{\ell\in\mathcal{S}_k}\|\alpha_\ell\|^3\\
&\geqslant |\mathcal{S}_k|\frac{\eta\sigma_{\min}m^{-\frac{1}{2}}}{3(\frac{1}{2}\Gamma(1-\rho)^{-3}+2\sigma_{\max}+2\kappa)^{\frac{3}{2}}}\min_{\ell\in\mathcal{S}_k}\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_\ell+D_\ell\alpha_\ell)^{\frac{3}{2}}\\
&\geqslant |\mathcal{S}_k|\frac{\eta\sigma_{\min}m^{-\frac{1}{2}}}{3(\frac{1}{2}\Gamma(1-\rho)^{-3}+2\sigma_{\max}+2\kappa)^{\frac{3}{2}}}\epsilon^{\frac{3}{2}}.
\end{aligned}$$

Thus

$$|\mathcal{S}_k| \leqslant \kappa_{\mathcal{S}}(\tilde{f}(\mathbf{y}_0,\mu)-\tilde{f}_{\text{low}})\epsilon^{-\frac{3}{2}}$$

where $\kappa_{\mathcal{S}}$ is given by (5.74). □

To complete the analysis in terms of evaluations rather than successful iterations, we need to bound the total number of all (successful and unsuccessful) iterations.

Lemma 5.6. *For all $k \geqslant 0$,*

$$k \leqslant \kappa^a |\mathcal{S}_k| + \kappa^b,$$

where

$$\kappa^a := 1 + \frac{m|\log \gamma_0|}{\log \gamma_1} \quad \textit{and} \quad \kappa^b := \frac{m}{\log \gamma_1} \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right).$$

Proof. Define

$$\mathcal{I}_{i,k} := \{j \in \{0, \ldots, k\} \mid (5.58) \text{ holds with } k \leftarrow j\},$$

(the set of iterations where $\sigma_{i,j}$ is increased) and

$$\mathcal{D}_{i,k} := \{j \in \{0, \ldots, k\} \mid (5.61) \text{ holds with } k \leftarrow j\} \subset \mathcal{S}_k$$

(the set of iterations where $\sigma_{i,j}$ is decreased), the final inclusion resulting from the condition that $\rho_k \geqslant \eta$ in both (5.59) and (5.60). Observe also that the mechanism of the algorithm, the fact that $\gamma_0 \in (0, 1)$ and Lemma 5.3 impose that, for each $i \in \{1, \ldots, m\}$,

$$\sigma_{\min} \gamma_1^{|\mathcal{I}_{i,k}|} \gamma_0^{|\mathcal{S}_k|} \leqslant \sigma_{0,i} \gamma_1^{|\mathcal{I}_{i,k}|} \gamma_0^{|\mathcal{D}_{i,k}|} \leqslant \sigma_{i,k} \leqslant \sigma_{\max}.$$

Dividing by $\sigma_{\min} > 0$ and taking logarithms yields that, for all $i \in \{1, \ldots, m\}$ and all $k > 0$,

$$|\mathcal{I}_{i,k}| \log \gamma_1 + |\mathcal{S}_k| \log \gamma_0 \leqslant \log\left(\frac{\sigma_{\max}}{\sigma_{\min}}\right). \tag{5.75}$$

Note now that, if (5.57) fails for all $i \in \{1, \ldots, m\}$, then

$$\delta \tilde{f}_k = \sum_{i=1}^{m} \delta \tilde{f}_{i,k} \geqslant \sum_{i=1}^{m} \delta \tilde{m}_{i,k} = \delta \tilde{m}_k$$

and, in view of (5.56), $\rho_k \geqslant 1 > \eta$, making iteration k successful. Thus, if iteration k is unsuccessful, at least one $\sigma_{i,k}$ is increased with (5.58). Thus

$$|\mathcal{U}_k| \leqslant \sum_{i=1}^{m} |\mathcal{I}_{i,k}| \leqslant m \max_{i=1,\ldots,m} |\mathcal{I}_{i,k}|. \tag{5.76}$$

The desired bound follows from (5.75) and (5.76) by using the fact that $k = |\mathcal{S}_k| + |\mathcal{U}_k| - 1 \leqslant |\mathcal{S}_k| + |\mathcal{U}_k|$, the term -1 in the equality accounting for iteration 0. □

We may now state our main evaluation complexity result.

Theorem 5.2. *Algorithm 5.1 requires at most*

$$\kappa^a \kappa_{\mathcal{S}} (\tilde{f}(\mathbf{y}_0, \mu) - \tilde{f}_{\text{low}}) \epsilon^{-\frac{3}{2}} + \kappa^b + 1$$

iterations and evaluations of $\tilde{h}$ and its first 2 derivatives to return a point $\mathbf{y}_\epsilon \in \mathcal{F}$ such that $\tilde{\chi}_{\tilde{f},1}(\mathbf{y}_k) \leqslant \epsilon$.

Proof. If termination occurs at iteration 0, the theorem obviously holds. Assume therefore that termination occurs at iteration $k+1$, in which case there must be at least one successful iteration. We may therefore deduce the desired bound from Theorem 5.1, Lemma 5.6 and the fact that each successful iteration involves the evaluation of $\tilde{f}(\mathbf{y}_k, \mu)$, $\{\theta^{q-1}(y_{i,k} + \alpha_{i,k} v_{i,k}, \mu)\}_{i=1}^m$ and $\{\theta^{q-2}(y_{i,k} + \alpha_{i,k} v_{i,k}, \mu)\}_{i=1}^m$, while each unsuccessful iteration only involves that of $\tilde{f}(y_k, \mu)$, $\{\theta^{q-1}(y_{i,k}, \mu)\}_{i=1}^m$ and $\{\theta^{q-2}(y_{i,k}, \mu)\}_{i=1}^m$. □

5.5 Numerical Experiments

In this section, we use the so-called ℓ_2-ℓ_p problem to illustrate the performance of Algorithm 5.1 and verify the complexity result. All the numerical experiments are conducted by MATLAB R2016a on a Lenovo PC equipped with 64-bit Windows

10 operating system, 3.60 GHz Intel(R) Dual Cores(i7) processor and 32.00 GB of RAM.

We consider the following ℓ_2-$\ell_{\frac{1}{2}}$ minimization problem

$$\underset{\mathbf{y}\in\mathbb{R}^n}{\text{minimize}} \ f(\mathbf{y}) := \|A\mathbf{y}-\mathbf{b}\|_2^2 + \lambda\sum_{i=1}^{n}|y_i|^{\frac{1}{2}} \tag{5.77}$$

where $\lambda > 0$, $A \in \mathbb{R}^{\hat{n}\times n}$ and $\mathbf{b} \in \mathbb{R}^{\hat{n}}$ with $\hat{n}$ an positive integer.

To intuitively observe the behavior of the sequences generated by our algorithm, we first restrict our attention on the 2-dimensional space $\mathbb{R}^2$ and then the high dimensional space.

5.5.1 The Case of 2-dimensional Space

Throughout this case, the parameters listed in Table 5.1 are kept constant during implementation of Algorithm 5.1, where ϵ and μ are the tolerance and the smoothing parameter, respectively.

Table 5.1: Constant Parameters in Implementing Algorithm 5.1

ϵ	ρ	μ	η	γ_0	γ_1	γ_2	κ	κ_{big}
1.00e-4	0.98	0.50e-4	0.20	0.25	1.50	3.00	1.00e2	5.00

Let $n = 2$, $A = (1,1)^{\mathrm{T}}$ and $\mathbf{b} = 1$, then minimization problem (5.77) can be simplified as follows

$$\underset{\mathbf{y}\in\mathbb{R}^2}{\text{minimize}} \ f(\mathbf{y}) = (y_1+y_2-1)^2 + \lambda(\sqrt{|y_1|}+\sqrt{|y_2|}). \tag{5.78}$$

The authors in [25] used this example to explain the optimality conditions.

As well known, for different values of λ, the optimal solutions of (5.78) vary. For $\lambda = 8/(3\sqrt{3})$, $(1/3,0)$ and $(0,1/3)$ are two nonzero points which conform to the first- and second-order optimality conditions proposed in [24]. But only $(0,0)$ is the

unique global minimizer. For $\lambda = 1$, (5.78) has two optimal solutions $(0, 0.7015)$ and $(0.7015, 0)$ with optimal value 0.927.

We generate 20 initial points for $\lambda = 8/(3\sqrt{3})$ and $\lambda = 1$, respectively. All the initial guess y_0 are generated by the MATLAB built-in function `randn(2,1)` and then projected onto the interval $[0, 1]$. To better observe the results, we divide them into two groups with every group 10 initial points and one implementation. In every implementation, the initial regularization parameter $\sigma_0 := (\sigma_{1,0}, \sigma_{2,0})^{\mathrm{T}}$ are generated by the MATLAB statement `4*rand(2,1)` and the value of $\sigma_{\min}$ is minimum value in the entries of σ_0.

From Fig. 5.1 and 5.2, it is not difficult to see that all the sequences starting from the above initial guesses converge to the minimizers of problem (5.78). Particularly, all the sequences converge to the global minimizer for $\lambda = 8/(3\sqrt{3})$. Furthermore, Table 5.2 generalizes the corresponding iteration numbers for solving problem (5.78) to arrive at its optimal solutions.

Table 5.2: Iteration Numbers to Reach Optimality

#Group	λ	#Iter									
1	$\frac{8}{3\sqrt{3}}$	122	100	129	145	125	125	114	115	115	112
	1	215	201	114	139	136	168	137	166	198	169
2	$\frac{8}{3\sqrt{3}}$	143	154	133	158	199	137	153	139	148	141
	1	121	64	113	53	31	56	53	96	94	101

5.5.2 The Case of High Dimensional Space

Now we consider the high dimensional case. That is, in problem (5.77), we set $A = (1, \ldots, 1) \in \mathbb{R}^n$, $\mathbf{b} = 1$ and the dimension n is allowed to be changed. To make sure the numerical results checkable, in every implementation, we use the same procedures with the 2-dimensional case to generate 10 initial guesses for $\lambda = 8/(3\sqrt{3})$

and $\lambda = 1$, respectively. And then we check if the function value of the sequences generated by Algorithm 5.1 converge and what value they converge.

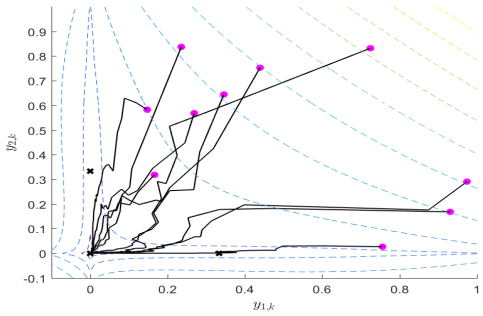
First of all, we keep the parameters listed in Table 5.1 invariant and conduct an implementation for $n = 10$, $n = 100$ and $n = 1000$, respectively. Fig 5.3 are the simulation results, which shows that for small size problem (n=10) the sequences generated from 10 arbitrary initial guesses indeed converge whereas the function values do not arrive at the optimal value for $n = 100$ and $n = 1000$ even if all of them are decreasing before the criticality measure arrives at the prescribed tolerance.

To investigate the above phenomenon, we further conduct another 2 implementation after set the tolerance ϵ to be 1e-6 and 1e-8, respectively. Note that the smoothing parameter $\mu = 0.5\epsilon$. Hence, both ϵ and μ are changed and other parameters listed in Table 5.1 are kept constant. Fig 5.4 and Fig 5.5 are the simulation results. Fig 5.4 shows that the function value of the sequences become to converge for $n = 100$ when the tolerance is improved to 1e-6. And Fig 5.5 encloses that to guarantee convergence for $n = 1000$, the tolerance should be further improved to 1e-8.

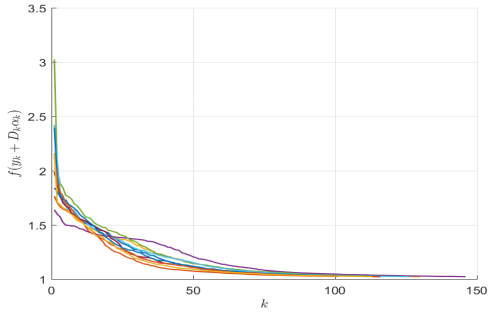
Moreover, we also list the specific iteration numbers for our algorithm to reach the point where the criticality measure is below prescribed precision ϵ, see Table 5.3 for details.

Table 5.3: Iteration Numbers to Reach Optimality in High Dimensional Case

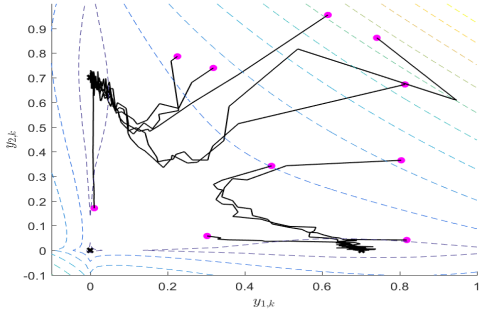
ϵ	λ	n	#Iter									
1e-4	$\frac{8}{3\sqrt{3}}$	10	92	87	95	90	88	94	91	89	89	87
		100	69	70	78	75	66	71	74	70	69	82
		1000	54	56	52	51	54	54	52	55	54	56
	1	10	170	114	120	143	152	148	119	109	144	115
		100	85	84	83	85	92	87	95	87	86	83
		1000	63	70	67	67	70	73	66	71	70	65
1e-6	$\frac{8}{3\sqrt{3}}$	10	357	345	348	358	347	349	351	352	350	354
		100	201	212	200	213	215	199	210	220	210	214
		1000	165	167	174	166	172	172	195	175	168	177
	1	10	579	594	579	781	772	739	678	592	359	792
		100	255	245	252	253	250	263	264	257	262	257
		1000	205	197	204	198	201	190	206	198	207	208
1e-8	$\frac{8}{3\sqrt{3}}$	10	819	824	82	836	830	806	815	832	813	845
		100	708	703	658	706	687	735	667	761	699	666
		1000	483	495	488	507	491	510	498	489	489	489
	1	10	1399	1385	3510	2411	3774	1400	3781	1393	3522	3495
		100	815	827	784	811	797	816	827	800	804	812
		1000	650	631	617	624	674	624	633	609	830	817



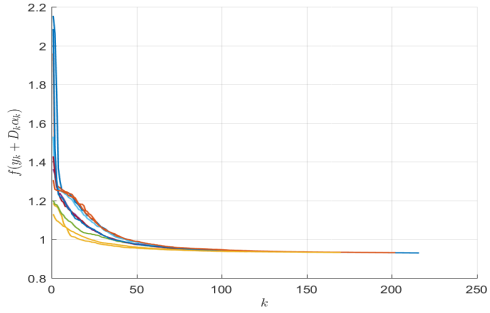
(a) Behavior of $\mathbf{y}_k + D_k\alpha_k$ for $\lambda = 8/3\sqrt{3}$



(b) Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for $\lambda = 8/3\sqrt{3}$

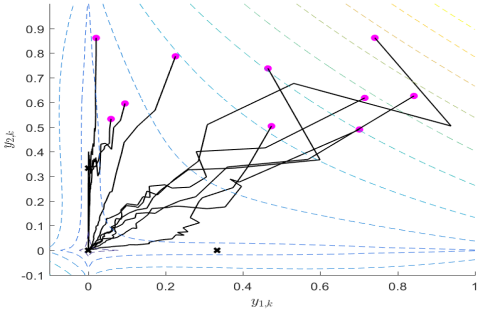


(c) Behavior of $\mathbf{y}_k + D_k\alpha_k$ for $\lambda = 1$

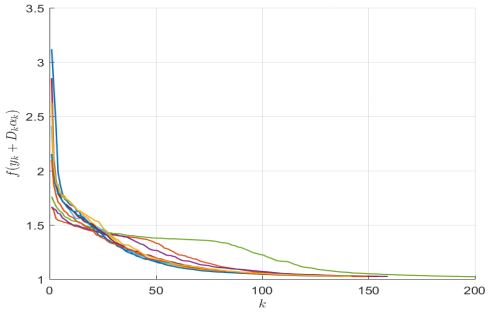


(d) Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for $\lambda = 1$

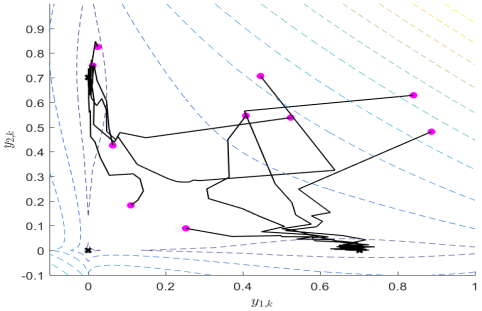
Figure 5.1: Results of Group 1 (the magenta circles mark the initial guesses).



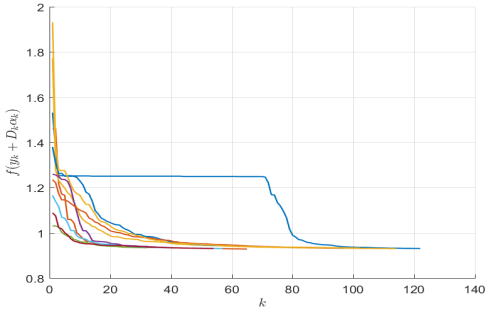
(a) Behavior of $\mathbf{y}_k + D_k\alpha_k$ for $\lambda = 8/3\sqrt{3}$



(b) Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for $\lambda = 8/3\sqrt{3}$

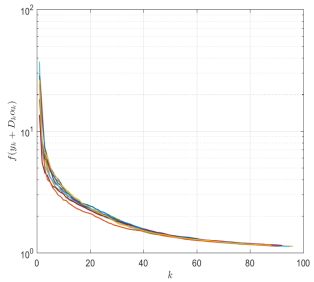


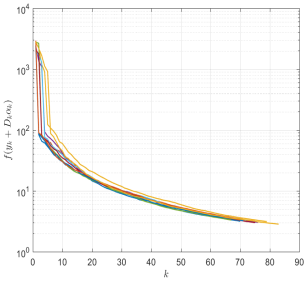
(c) Behavior of $\mathbf{y}_k + D_k\alpha_k$ for $\lambda = 1$

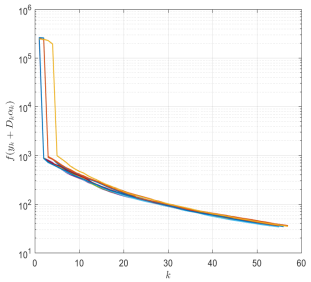


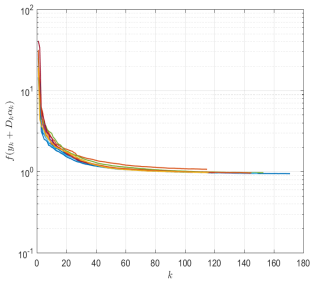
(d) Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for $\lambda = 1$

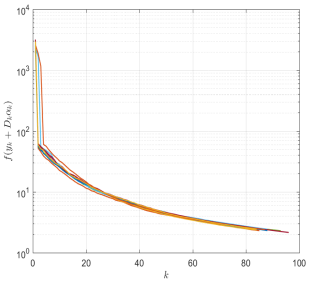
Figure 5.2: Results of Group 2 (the magenta circles mark the initial guesses).


(a) $n = 10$, $\lambda = 8/(3\sqrt{3})$


(b) $n = 100$, $\lambda = 8/(3\sqrt{3})$


(c) $n = 1000$, $\lambda = 8/(3\sqrt{3})$


(d) $n = 10$, $\lambda = 1$


(e) $n = 100$, $\lambda = 1$

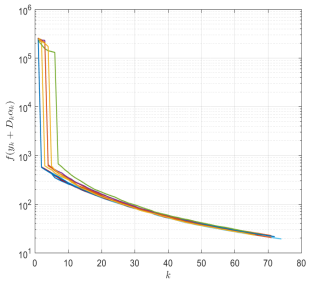
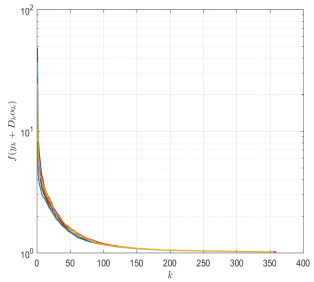
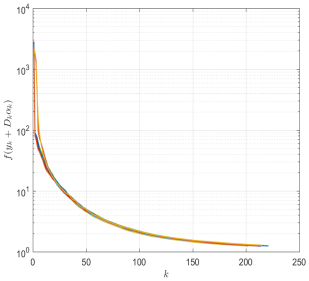
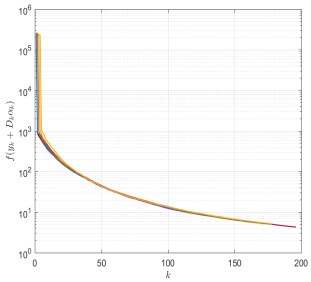
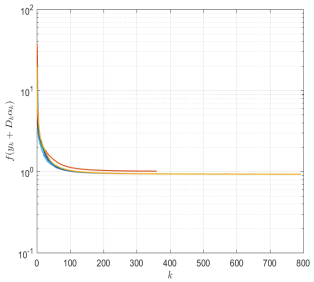

(f) $n = 1000$, $\lambda = 1$

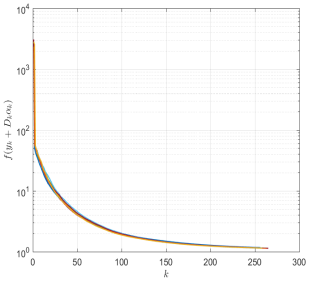
Figure 5.3: Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for ϵ=1e-4


(a) $n = 10$, $\lambda = 8/(3\sqrt{3})$


(b) $n = 100$, $\lambda = 8/(3\sqrt{3})$


(c) $n = 1000$, $\lambda = 8/(3\sqrt{3})$


(d) $n = 10$, $\lambda = 1$


(e) $n = 100$, $\lambda = 1$

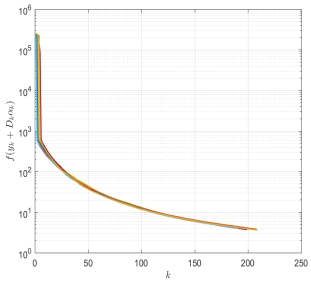
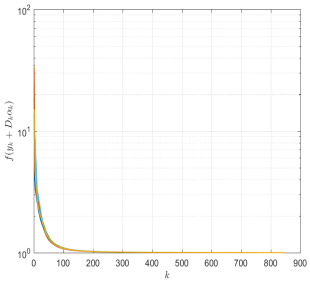
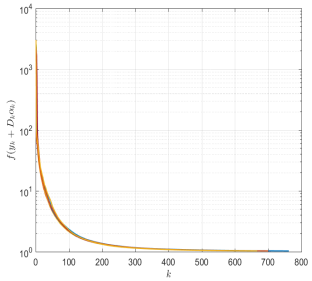

(f) $n = 1000$, $\lambda = 1$

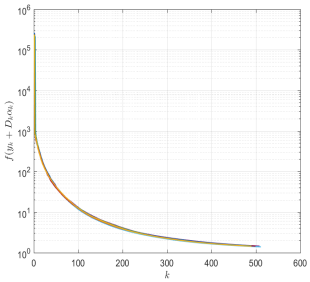
Figure 5.4: Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for ϵ=1e-6



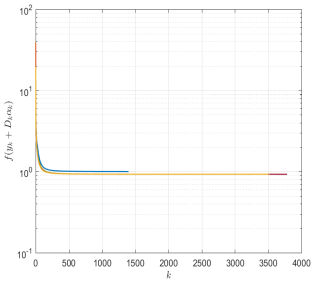
(a) $n = 10$, $\lambda = 8/(3\sqrt{3})$



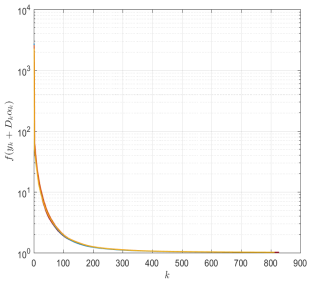
(b) $n = 100$, $\lambda = 8/(3\sqrt{3})$



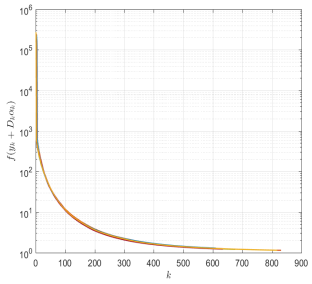
(c) $n = 1000$, $\lambda = 8/(3\sqrt{3})$



(d) $n = 10$, $\lambda = 1$



(e) $n = 100$, $\lambda = 1$



(f) $n = 1000$, $\lambda = 1$

Figure 5.5: Convergence of $f(\mathbf{y}_k + D_k\alpha_k)$ for ϵ=1e-8

Chapter 6

Concluding Remarks

This chapter draws conclusions on the thesis with several concluding remarks and list several possible research directions in the future.

6.1 Affine Rank Minimization Problem

The rank minimization problems with linear constraints are usually NP-hard and hence are difficult to solve. The approaches using the convex approximation models are of theoretical guarantee for the solution quality, however, lack of efficiency in general due to extensive invoking singular value decomposition solvers. Methods based on the low rank approximation or the factorization model are usually numerically efficient, but lack of theoretical guarantee. Because the factorization model is a quartic polynomial and hence to solve it to global optimality is NP-hard. Another difficulty of using factorization model is how to iteratively obtain a suitable rank for factorization which should be close and no less than the global minimum of the rank minimization problem.

In the first part of this thesis (i.e., Chapter 2 and Chapter 6), we point out the property that second-order necessary optimality condition implies global optimality perhaps holds on a large variety of scenarios. However, we also show by a special instance that such property is not always satisfied. By using such SNIG condition,

we propose a new algorithm framework for solving (1.1) by solving a series of factorization model (NLS-k) to the second-order optimality. Any second-order methods which guarantees to terminate at a second-order stationary point can be used in this stage. We further put forward the conjecture that the reductions between the global minima of (LLS-k) with two consecutive ranks are monotonically decreasing. This conjecture, if holds, can significantly increase the efficiency of our algorithm framework.

6.2 Partially Separable Minimization Problem

The partially separable minimization problem with convex constraint is ubiquitous and of importance in many applications of optimization. Such structure also includes sparse optimization as a special case, which has been widely used in image restoration and statistics. And high-order Taylor models for optimization is recently popular in the context of adaptive regularization algorithms for unconstrained problem. Such methods are well-known for their remarkable evaluation complexity, in which they need at most $O(\epsilon^{-\frac{p+1}{p}})$ evaluations of the objective function and their derivatives to arrive at an ϵ-approximate first-order stationary point while the steepest and Newton's methods need the $O(\epsilon^{-2})$. Nevertheless, most adaptive regularization methods rest on a non-separable regularization term, which makes the study of such separable structure difficult.

In the second part of this thesis (i.e., Chapter 4 and Chapter 5), we first devise an algorithm with a partially separable $p+1$ order regularization term. Then under mild conditions, we show that such regularization algorithm also has the remarkable $O(\epsilon^{-\frac{p+1}{p}})$ evaluation complexity of objective function value and its derivatives when utilizing the separable structure. Subsequently, we extend the algorithm with $p = 2$ to solve the problem of data fitting involving the q-quasi norm for $q \in (0, 1)$

which might be nonconvex and non-Lipschitz. Finally, we show that the excellent complexity bound $O(\epsilon^{-\frac{3}{2}})$ can also hold even for the non-Lipschitz case.

6.3 Future Research

The affine rank minimization problem and the partially separable minimization problem are two classes of optimization problems since both of them subsume a lot of problems as their special cases and hence have wide applications in scientific and engineering domains. We will pay further attention to these problems and anticipate that more progresses can be made along the following possible directions.

- For the affine rank minimization problem:
  - we hope to find more triplets $(\mathcal{A}, \mathbf{b}, k)$ at which the SNIG condition holds;
  - we will also consider extend our current results without noise to their counterpart scenarios with the observations contaminated by noises;
  - we will try to improve the overall performance of our algorithm framework by utilizing the trust-region method on the manifold to solve problem (NLS-k) and then compare it with the existing algorithms.
- For the partially separable minimization problem:
  - we will find more accessible numerical examples and compare our algorithm with existing algorithms since our current numerical tests are only designed to check the performance and verify our iteration complexity result;
  - we will also try to improve the performance of our algorithm by parallelizing the regularization parameters updating step in the algorithm, that is, Step 8-13 in Algorithm 5.1.

Bibliography

[1] Y. Amit, M. Fink, N. Srebro, and S. Ullman. Uncovering shared structures in multiclass classification. In *Proceedings of the 24th international conference on Machine learning*, pages 17–24, 2007.

[2] J. Bennett and S. Lanning. The netflix prize. In *Proceedings of KDD cup and workshop*, page 35, 2007.

[3] W. Bian and X. Chen. Worst-case complexity of smoothing quadratic regularization methods for non-Lipschitzian optimization. *SIAM Journal on Optimization*, 23(3):1718–1741, 2013.

[4] W. Bian and X. Chen. Linearly constrained non-Lipschitz optimization for image restoration. *SIAM Journal on Imaging Sciences*, 8(4):2294–2322, 2015.

[5] W. Bian, X. Chen, and Y. Ye. Complexity analysis of interior point algorithms for non-lipschitz and nonconvex minimization. *Mathematical Programming*, 149(1-2):301–327, 2015.

[6] E. Birgin, J. Gardenghi, J. Martınez, S. Santos, and Ph. Toint. Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Report naXys-05-2015, University of Namur, Belgium*, 2015.

[7] J. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013.

[8] S. Burer and R. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[9] J. Cai, E. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[10] E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Review*, 57(2):225–251, 2015.

[11] E. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):11, 2011.

[12] E. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[13] E. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

[14] E. Candès and T. Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

[15] C. Cartis, N.I.M. Gould, and Ph. Toint. On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, 20(6):2833–2852, 2010.

[16] C. Cartis, N.I.M. Gould, and Ph. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming*, 127(2):245–295, 2011.

[17] C. Cartis, N.I.M. Gould, and Ph. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function-and derivative-evaluation complexity. *Mathematical Programming*, 130(2):295–319, 2011.

[18] C. Cartis, N.I.M. Gould, and Ph. Toint. An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity. *IMA Journal of Numerical Analysis*, page drr035, 2012.

[19] C. Cartis, N.I.M. Gould, and Ph. Toint. Complexity bounds for second-order optimality in unconstrained optimization. *Journal of Complexity*, 28(1):93–108, 2012.

[20] C. Chen, B. He, and X. Yuan. Matrix completion via an alternating direction method. *IMA Journal of Numerical Analysis*, 32(1):227–245, 2012.

[21] L. Chen, N. Deng, and J. Zhang. Modified partial-update Newton-type algorithms for unary optimization. *Journal of optimization theory and applications*, 97(2):385–406, 1998.

[22] S. Chen, S. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of control*, 50(5):1873–1896, 1989.

[23] X. Chen, L. Niu, and Y. Yuan. Optimality conditions and a smoothing trust region newton method for nonlipschitz optimization. *SIAM Journal on Optimization*, 23(3):1528–1552, 2013.

[24] X. Chen, F. Xu, and Y. Ye. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. *SIAM Journal on Scientific Computing*, 32(5):2832–2852, 2010.

[25] Xiaojun Chen, Dongdong Ge, Zizhuo Wang, and Yinyu Ye. Complexity of unconstrained ℓ_2-ℓ_p minimization. *Mathematical Programming*, 143(1-2):371–383, 2014.

[26] Y. Chen, M. Jovanovic, and T. Georgiou. State covariances and the matrix completion problem. In *IEEE 52nd Annual Conference on Decision and Control*, pages 1702–1707, 2013.

[27] A. Conn, N.I.M. Gould, A. Sartenaer, and Ph. Toint. Convergence properties of minimization algorithms for convex constraints using a structured trust region. *SIAM Journal on Optimization*, 6(4):1059–1086, 1996.

[28] A. Conn, N.I.M. Gould, and Ph. Toint. *Trust region methods*, volume 1. SIAM, 2000.

[29] A. Conn, N.I.M. Gould, and Ph. Toint. *LANCELOT: a Fortran package for large-scale nonlinear optimization (Release A)*, volume 17. Springer Science & Business Media, 2013.

[30] G. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *Optical Engineering*, 33(7):2183–2191, 1994.

[31] L. El Ghaoui and P. Gahinet. Rank minimization under LMI constraints: A framework for output feedback problems. 1993.

[32] H. Esfahanizadeh, F. Lahouti, and B. Hassibi. A matrix completion approach to linear index coding problem. *arXiv preprint arXiv:1408.3046*, 2014.

[33] A. Evgeniou and M. Pontil. Multi-task feature learning. *Advances in neural information processing systems*, 19:41, 2007.

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

[35] M. Fazel, H. Hindi, and S. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of American Control Conference*, volume 6, pages 4734–4739, 2001.

[36] M. Fornasier, H. Rauhut, and R. Ward. Low-rank matrix recovery via iteratively reweighted least squares minimization. *SIAM Journal on Optimization*, 21(4):1614–1640, 2011.

[37] J. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American statistical Association*, 76(376):817–823, 1981.

[38] S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, 2011.

[39] D. Gay. Automatically finding and exploiting partially separable structure in nonlinear programming problems. *Bell Laboratories, Murray Hill*, 1996.

[40] D. Goldfarb and S. Ma. Convergence of fixed-point continuation algorithms for matrix rank minimization. *Foundations of Computational Mathematics*, 11(2):183–210, 2011.

[41] D. Goldfarb and S. Wang. Partial-update Newton methods for unary, factorable, and partially separable optimization. *SIAM Journal on Optimization*, 3(2):382–397, 1993.

[42] N.I.M. Gould, D. Orban, and Ph. Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3):545–557, 2015.

[43] N.I.M. Gould and Ph. Toint. FILTRANE, a fortran 95 filter-trust-region package for solving nonlinear least-squares and nonlinear feasibility problems. *ACM Transactions on Mathematical Software*, 33(1):3, 2007.

[44] A. Griewank. The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical report, Technical Report NA/12, 1981.

[45] A. Griewank and Ph. Toint. On the unconstrained optimization of partially separable functions. *Nonlinear Optimization*, 1982:247–265, 1981.

[46] D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

[47] M. Hardt. Understanding alternating minimization for matrix completion. In *IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 651–660, 2014.

[48] R. Hom and C. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991.

[49] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *Advances in Neural Information Processing Systems*, pages 937–945, 2010.

[50] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674, 2013.

[51] A. Javanmard and A. Montanari. Localization from incomplete noisy distance measurements. *Foundations of Computational Mathematics*, 13(3):297–345, 2013.

[52] M. Journée, F. Bach, P-A Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization*, 20(5):2327–2351, 2010.

[53] R. Keshavan. *Efficient algorithms for collaborative filtering*. PhD thesis, Stanford University, 2012.

[54] L. Kong and N. Xiu. Exact low-rank matrix recovery via nonconvex schatten p-minimization. *Asia-Pacific Journal of Operational Research*, 30(03):1340010, 2013.

[55] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[56] M. Lai, Y. Xu, and W. Yin. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM Journal on Numerical Analysis*, 51(2):927–957, 2013.

[57] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[58] F. Lin, M. Jovanovic, and T. Georgiou. An ADMM algorithm for matrix completion of partially known state covariances. In *IEEE 52nd Annual Conference on Decision and Control*, pages 1684–1689, 2013.

[59] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems*, pages 612–620, 2011.

[60] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[61] Z. Liu, A. Hansson, and L. Vandenberghe. Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62(8):605–612, 2013.

[62] Z. Lu and Y. Zhang. Iterative reweighted singular value minimization methods for lp regularized unconstrained matrix minimization. *arXiv preprint arXiv:1401.0869*, 2014.

[63] Z. Lu, Y. Zhang, and X. Li. Penalty decomposition methods for rank minimization. *Optimization Methods and Software*, 30(3):531–558, 2015.

[64] S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011.

[65] A. Majumdar and R. Ward. An algorithm for sparse MRI reconstruction by schatten p-norm minimization. *Magnetic resonance imaging*, 29(3):408–417, 2011.

[66] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[67] J. Mareček, P. Richtárik, and M. Takáč. Distributed block coordinate descent for minimizing partially separable functions. In *Numerical Analysis and Optimization*, pages 261–288. 2015.

[68] B. Mishra, G. Meyer, F. Bach, and R. Sepulchre. Low-rank optimization with trace norm penalty. *SIAM Journal on Optimization*, 23(4):2124–2149, 2013.

[69] K. Mohan and M. Fazel. Iterative reweighted least squares for matrix rank minimization. In *48th Annual Allerton Conference on Communication, Control and Computing*, pages 653–661, 2010.

[70] K. Mohan and M. Fazel. Iterative reweighted algorithms for matrix rank minimization. *The Journal of Machine Learning Research*, 13(1):3441–3473, 2012.

[71] C. Mu, B. Huang, J. Wright, and D. Goldfarb. Square deal: Lower bounds and improved relaxations for tensor recovery. In *International Conference on Machine Learning*, pages 73–81, 2014.

[72] S. Muthukrishnan. *Data streams: Algorithms and applications*. Now Publishers Inc, 2005.

[73] B. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[74] S. Negahban, P. Ravikumar, M. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers. *Manuscript, University of California, Berkeley*, 2011.

[75] S. Negahban and M. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1):1665–1697, 2012.

[76] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

[77] Y. Nesterov and B. Polyak. Cubic regularization of Newton method and its global performance. *Mathematical Programming*, 108(1):177–205, 2006.

[78] S. Oh, A. Montanari, and A. Karbasi. Sensor network localization from local connectivity: Performance analysis for the mds-map algorithm. In *IEEE Information Theory Workshop*, pages 1–5, 2010.

[79] S. Oymak and B. Hassibi. New null space results and recovery thresholds for matrix rank minimization. *arXiv preprint arXiv:1011.6326*, 2010.

[80] S. Oymak, A. Khajehnejad, and B. Hassibi. Improved thresholds for rank minimization. In *IEEE International Conference on Speech and Signal Processing*, pages 5988–5991, 2011.

[81] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, pages 40–44, 1993.

[82] S. Qian and D. Chen. Signal representation using adaptive normalized gaussian functions. *Signal processing*, 36(1):1–11, 1994.

[83] B. Recht. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 12:3413–3430, 2011.

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

[85] B. Recht, W. Xu, and B. Hassibi. Null space conditions and thresholds for rank minimization. *Mathematical programming*, 127(1):175–202, 2011.

[86] J. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719, 2005.

[87] R. Rockafellar. *Convex analysis*. Princeton University Press, 2015.

[88] W Rudin. *Real and complex analysis (3rd)*. New York: McGraw-Hill Inc, 1986.

[89] J. Shahabuddin. Structured trust region algorithms for the minimization of nonlinear functions. 1996.

[90] A. Singer. A remark on global positioning from local distances. *Proceedings of the National Academy of Sciences*, 105(28):9507–9511, 2008.

[91] A.M. So and Y. Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, 109(2-3):367–384, 2007.

[92] N. Srebro. *Learning with matrix factorizations*. PhD thesis, Citeseer, 2004.

[93] R. Sun and Z. Luo. Guaranteed matrix completion via nonconvex factorization. In *IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 270–289, 2015.

[94] S. Sun, A. Petropulu, and W. Bajwa. Target estimation in colocated MIMO radar via matrix completion. In *IEEE International Conference on Speech and Signal Processing*, pages 4144–4148, 2013.

[95] W. Sun and Y. Yuan. Solving nonlinear least-squares problems. *Optimization Theory and Methods: Nonlinear Programming*, pages 353–383, 2006.

[96] G. Takács, I. Pilászy, B. Németh, and D. Tikk. Major components of the gravity recommendation system. *ACM SIGKDD Explorations Newsletter*, 9(2):80–83, 2007.

[97] K.C. Toh and S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization*, 6(615-640):15, 2010.

[98] Z. Wang, M. Lai, Z. Lu, W. Fan, H. Davulcu, and J. Ye. Orthogonal rank-one matrix pursuit for low rank matrix completion. *SIAM Journal on Scientific Computing*, 37(1):488–514, 2015.

[99] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4(4):333–361, 2012.

[100] Y. Xu, W. Yin, Z. Wen, and Y. Zhang. An alternating direction algorithm for matrix completion with nonnegative factors. *Frontiers of Mathematics in China*, 7(2):365–384, 2012.

[101] Y. Yuan. *Trust region algorithms for nonlinear equations*. Hong Kong Baptist University, 1994.

[102] Y. Yuan. A review of trust region algorithms for optimization. In *International Congress on Industrial and Applied Mathematics*, volume 99, pages 271–282, 2000.

[103] A. Zare, M. Jovanovic, and T. Georgiou. Completion of partially known turbulent flow statistics. In *American Control Conference*, pages 1674–1679, 2014.

[104] J. Zhang and C. Xu. A class of indefinite dogleg path methods for unconstrained minimization. *SIAM Journal on Optimization*, 9(3):646–667, 1999.

[105] Z. Zhu, A.M. So, and Y. Ye. Fast and near–optimal matrix completion via randomized basis pursuit. In *Proceedings of the Fifth International Congress of Chinese Mathematicians*, volume 51, pages 859–882, 2009.