5-1-2018

Quadratic Optimization for Nonsmooth Optimization Algorithms: Theory and Numerical Experiments

Baoyu Zhou
Lehigh University, baoyu.zhou@outlook.com

Quadratic Optimization for Nonsmooth Optimization Algorithms: Theory and Numerical Experiments

by

Baoyu Zhou

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Industrial and Systems Engineering

Lehigh University

May 2018

This thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science.

Date

Thesis Advisor

Co-Advisor

Chairperson of Department

Acknowledgements

This research is supported by Lehigh University. I am thankful to Frank E. Curtis and Andreas Wächter, who provided me with selfless technical support. I would like to show my gratitude to Yue Hu and Xingyu Wang, who always give me reasonable suggestions on my research. I gratefully thank Yihe Zhuo, who gives me both courage and spiritual support. I especially appreciate the assistance from Mu Tao and Sheng Zhou, who provide me with their endless love. At any time, they never let me down and give me as much support as they can.

Contents

List of Tables

Abstract

Nonsmooth optimization arises in many scientific and engineering applications, such as optimal control, neural network training, and others. Gradient sampling and bundle methods are two efficient types of algorithms for solving nonsmooth optimization problems. Quadratic optimization (commonly referred to as QP) problems arise as subproblems in both types of algorithms. This thesis introduces an algorithm for solving the types of QP problems that arise in such methods. The proposed algorithm is inspired by one proposed in a paper written by Krzysztof C. Kiwiel in the 1980s. Improvements are proposed so that the algorithm may solve problems with additional bound constraints, which are often required in practice. The solver also allows for general quadratic terms in the objective. Our QP solver has been implemented in C++. This thesis not only covers the theoretical background related to the QP solver; it also contains the results of numerical experiments on a wide range of randomly generated test problems.

Chapter 1

Introduction

In various nonsmooth optimization algorithms such as in the class of bundle methods (see [3]), QP subproblems arise in the following form:

$$\begin{aligned} &\min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}} z + \frac{1}{2}(x-x_1)^T H(x-x_1) \\ &\text{s.t. } f_j + g_j^T(x-x_j) \le z \text{ for all } j \in \{1,\ldots,m\} \text{ and } \|x-x_1\|_\infty \le \delta, \end{aligned} \tag{1.1}$$

where $H \in \mathbb{R}^{n\times n}$ is a given symmetric positive definite matrix, $\{f_j\}_{j=1}^m$ is a set of scalar values, $\{g_j\}_{j=1}^m$ and $\{x_j\}_{j=1}^m$ are sets of real vectors in $\mathbb{R}^n$, and δ is a nonnegative scalar. The dual of this problem can be written as

$$\begin{aligned} &\min_{(\omega,\gamma)\in\mathbb{R}^m\times\mathbb{R}^n} \frac{1}{2}(G\omega+\gamma)^T W(G\omega+\gamma) - b^T\omega + \delta\|\gamma\|_1 \\ &\text{s.t.} \quad \mathbb{1}^T\omega = 1 \text{ and } \omega \ge 0, \end{aligned} \tag{1.2}$$

where $W = H^{-1}$, G is a matrix whose columns are the elements of $\{g_j\}_{j=1}^m$, and b is a real vector whose values are easily computed using the values $\{(f_j, g_j, x_j)\}_{j=1}^m$. In this thesis, we present, analyze, and test an algorithm for solving the QP (1.2). Letting (ω_*, γ_*) be a solution of this problem, the solution of (1.1) can be recovered as $x_* = x_1 - W(G\omega_* + \gamma_*)$.

Theoretical background related to the proposed QP solver is introduced in Chapter 2, including KKT conditions of problem (1.2) and information about Cholesky factorization updates and Eigen-decompositions. The basis for the proposed method is that proposed by Kiwiel in [1]. We have improved the algorithm to make it feasibly solve QP problems with bound constraints and general quadratic terms. Details of the algorithm are mentioned in Chapter 3. Chapter 4 covers a proof of convergence of the algorithm. Results of numerical experiments using a C++ implementation of our QP solver are given in Chapter 5. We state concluding remarks in Chapter 6.

Chapter 2

Mathematical Background

In this chapter, we provide background on mathematical concepts that are needed to understand our QP solver and our implementation of it.

2.1 Optimality Conditions

One can solve the optimization problem (1.2) by solving its optimality conditions, i.e., conditions such that, if they are satisfied, then one confirms that a solution of problem (1.2) has been obtained. The optimality conditions for (1.2) are referred to as Karush-Kuhn-Tucker (KKT) conditions [2]. For problem (1.2), these conditions take the form

$$\omega^T(b - G^T W(G\omega + \gamma)) - b_j + g_j^T W(G\omega + \gamma) \geq 0 \quad \text{for all} \quad j \in \{1, \ldots, m\}, \tag{2.1a}$$

$$\delta \mathbb{1} + W(G\omega + \gamma) \geq 0, \tag{2.1b}$$

$$\delta \mathbb{1} - W(G\omega + \gamma) \geq 0, \tag{2.1c}$$

$$1 - \mathbb{1}^T \omega = 0, \quad \omega \geq 0, \tag{2.1d}$$

$$\max\{\gamma, 0\}^T(\delta \mathbb{1} + W(G\omega + \gamma)) = 0, \quad \text{and} \tag{2.1e}$$

$$\max\{-\gamma, 0\}^T(\delta \mathbb{1} - W(G\omega + \gamma)) = 0. \tag{2.1f}$$

Here, (2.1a)–(2.1c) and (2.1d) are known as primal and dual feasibility conditions while (2.1e)–(2.1f) are complementarity conditions.

One can rewrite (2.1) by breaking it down in terms of which elements of the variable vectors are positive or negative. In particular, for a solution (ω, γ) of (2.1), let $\mathcal{S} \subseteq \{1, \ldots, m\} =: \mathcal{J}$ be the indices corresponding to positive elements of ω, let $\mathcal{P} \subseteq \{1, \ldots, n\} =: \mathcal{I}$ be the indices corresponding to positive elements of γ, and let $\mathcal{N} \subseteq \mathcal{I}$ be the indices corresponding to negative elements of γ. Notice that there are no negative elements of ω since (2.1d) requires that $\omega \geq 0$. Also, notice that, by definition, $\mathcal{P} \cap \mathcal{N} = \emptyset$. With these definitions and defining $\bar{\gamma}_{\mathcal{P}\cup\mathcal{N}}$ such that

$$[\bar{\gamma}_{\mathcal{P}\cup\mathcal{N}}]_i = \begin{cases} \gamma_i & \text{if } i \in \mathcal{P} \cup \mathcal{N} \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

the KKT conditions (2.1) become

$$\omega_{\mathcal{S}}^T(b_{\mathcal{S}} - G_{:,\mathcal{S}}^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}})) - b_j + g_j^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}}) \geq 0 \quad \text{for all} \quad j \in \mathcal{J}, \tag{2.3a}$$

$$\delta + [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}})]_i \geq 0 \quad \text{for all} \quad i \in \mathcal{I}, \tag{2.3b}$$

$$\delta - [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}})]_i \geq 0 \quad \text{for all} \quad i \in \mathcal{I}, \tag{2.3c}$$

$$\mathbb{1}^T\omega_{\mathcal{S}} = 1, \tag{2.3d}$$

$$\gamma_i(\delta + [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}})]_i) = 0 \quad \text{for all} \quad i \in \mathcal{P}, \quad \text{and} \tag{2.3e}$$

$$\gamma_i(\delta - [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \bar{\gamma}_{\mathcal{P}\cup\mathcal{N}})]_i) = 0 \quad \text{for all} \quad i \in \mathcal{N}. \tag{2.3f}$$

Simplifying further since $\gamma_{\mathcal{P}} \neq 0$ and $\gamma_{\mathcal{N}} \neq 0$, one obtains that

$$\omega_{\mathcal{S}}^T(b_{\mathcal{S}} - G_{:,\mathcal{S}}^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}})) - b_j + g_j^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}}) \geq 0 \quad \text{for all} \quad j \in \mathcal{J}, \tag{2.4a}$$

$$\delta + [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}})]_i \geq 0 \quad \text{for all} \quad i \in \mathcal{I} \setminus \mathcal{P}, \tag{2.4b}$$

$$\delta - [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}})]_i \geq 0 \quad \text{for all} \quad i \in \mathcal{I} \setminus \mathcal{N}, \tag{2.4c}$$

$$\mathbb{1}^T\omega_{\mathcal{S}} = 1, \tag{2.4d}$$

$$\delta + [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}})]_i = 0 \quad \text{for all} \quad i \in \mathcal{P}, \quad \text{and} \tag{2.4e}$$

$$\delta - [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}})]_i = 0 \quad \text{for all} \quad i \in \mathcal{N}. \tag{2.4f}$$

2.2 Cholesky Factorization Updates

As we will see later, our QP algorithm involves updating estimates of the optimal index sets $\mathcal{S}$, $\mathcal{P}$, and $\mathcal{N}$ corresponding to (2.3) in each iteration. When some element is added to/removed from $\mathcal{S}$, $\mathcal{P}$, or $\mathcal{N}$, an implementation of our algorithm can be made efficient by adding/removing rows and columns of a Cholesky factorization of a particular matrix.

For a symmetric positive definite M, there exists a unique upper triangular R such that

$$M = R^T R. \tag{2.5}$$

In the following subsections, we discuss how to update R to $\bar{R}$ such that $\bar{M} = \bar{R}^T \bar{R}$, where $\bar{M}$ is obtained by adding (deleting) a row/column pair to (from) M. We also show how to update a solution of $R^T x = y$ to a solution of $\bar{R}^T \bar{x} = \bar{y}$ by adding (deleting) an element of y corresponding to the row/column pair.

2.2.1 Row/column addition

Consider partitioning M and R such that

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} = \begin{bmatrix} R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}. \tag{2.6}$$

Consider also partitions for $\bar{M}$ and $\bar{R}$ so that

$$\bar{M} = \begin{bmatrix} M_{11} & a & M_{12} \\ a^T & b & c^T \\ M_{12}^T & c & M_{22} \end{bmatrix} = \begin{bmatrix} \bar{R}_{11}^T & 0 & 0 \\ u^T & v & 0 \\ \bar{R}_{12}^T & w & \bar{R}_{22}^T \end{bmatrix} \begin{bmatrix} \bar{R}_{11} & u & \bar{R}_{12} \\ 0 & v & w^T \\ 0 & 0 & \bar{R}_{22} \end{bmatrix}. \tag{2.7}$$

Considering (2.7), we have $M_{11} = \bar{R}_{11}^T \bar{R}_{11}$. After combining with (2.6), we have

$$\bar{R}_{11} = R_{11}. \tag{2.8}$$

Reconsidering (2.7), we then have

$$R_{11}^T u = a \tag{2.9}$$

and

$$v = \sqrt{b - \|u\|^2}. \tag{2.10}$$

Moreover, we have

$$M_{12}^T = \bar{R}_{12}^T \bar{R}_{11} = \bar{R}_{12}^T R_{11}, \tag{2.11}$$

meaning that from (2.6) we know

$$\bar{R}_{12} = R_{12}. \tag{2.12}$$

From (2.7), we have

$$c = \bar{R}_{12}^T u + wv = R_{12}^T u + wv, \tag{2.13}$$

which means

$$w = \frac{1}{v}(c - R_{12}^T u). \tag{2.14}$$

At last, we need to determine $\bar{R}_{22}$ such that

$$\begin{aligned} \bar{R}_{22}^T \bar{R}_{22} &= M_{22} - \bar{R}_{12}^T \bar{R}_{12} - ww^T \\ &= M_{22} - R_{12}^T R_{12} - ww^T \\ &= R_{22}^T R_{22} - ww^T. \end{aligned} \tag{2.15}$$

This shows that $\bar{R}_{22}$ can be obtained from R_{22} with a rank-one update.

Now let's consider the linear system $R^T x = y$. We can partition x and y such that

$$\begin{bmatrix} R_{11}^T & 0 \\ R_{12}^T & R_{22}^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}. \tag{2.16}$$

Similarly, we can partition the linear system $\bar{R}^T\bar{x} = \bar{y}$ such that

$$\begin{bmatrix} R_{11}^T & 0 & 0 \\ u^T & v & 0 \\ R_{12}^T & w & \bar{R}_{22}^T \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_0 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_0 \\ y_2 \end{bmatrix}. \tag{2.17}$$

Combining (2.16) and (2.17), then we have

$$\bar{x}_1 = x_1 \tag{2.18}$$

and

$$\begin{aligned} \bar{x}_0 &= \frac{1}{v}(y_0 - u^T\bar{x}_1) \\ &= \frac{1}{v}(y_0 - u^T x_1). \end{aligned} \tag{2.19}$$

We also have

$$\begin{aligned} \bar{R}_{22}^T\bar{x}_2 &= y_2 - R_{12}^T\bar{x}_1 - w\bar{x}_0 \\ &= y_2 - R_{12}^T x_1 - w\bar{x}_0 \\ &= R_{22}^T x_2 - w\bar{x}_0. \end{aligned} \tag{2.20}$$

Hence, we can use (2.20) to compute $\bar{x}_2$.

2.2.2 Row/column deletion

Consider the original matrix $M = R^T R$ with a partition such that:

$$\begin{bmatrix} M_{11} & m_{12} & M_{13} \\ m_{12}^T & m_{22} & m_{23}^T \\ M_{13}^T & m_{23} & M_{33} \end{bmatrix} = \begin{bmatrix} R_{11}^T & 0 & 0 \\ r_{12}^T & r_{22} & 0 \\ R_{13}^T & r_{23} & R_{33}^T \end{bmatrix} \begin{bmatrix} R_{11} & r_{12} & R_{13} \\ 0 & r_{22} & r_{23}^T \\ 0 & 0 & R_{33} \end{bmatrix}. \tag{2.21}$$

After deleting the row/column pair, we have $\tilde{M} = \tilde{R}^T\tilde{R}$ as:

$$\begin{bmatrix} M_{11} & M_{13} \\ M_{13}^T & M_{33} \end{bmatrix} = \begin{bmatrix} \tilde{R}_{11}^T & 0 \\ \tilde{R}_{12}^T & \tilde{R}_{22}^T \end{bmatrix} \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ 0 & \tilde{R}_{22} \end{bmatrix}. \tag{2.22}$$

Combining (2.21) and (2.22), we easily have

$$\tilde{R}_{11} = R_{11} \text{ and } \tilde{R}_{12} = R_{13}. \tag{2.23}$$

We also have

$$\begin{aligned} \tilde{R}_{22}^T \tilde{R}_{22} &= M_{33} - \tilde{R}_{12}^T \tilde{R}_{12} \\ &= M_{33} - R_{13}^T R_{13} \\ &= r_{23} r_{23}^T + R_{33}^T R_{33}. \end{aligned} \tag{2.24}$$

For the linear equation $R^T x = y$, which could be written as

$$\begin{bmatrix} R_{11}^T & 0 & 0 \\ r_{12}^T & r_{22} & 0 \\ R_{13}^T & r_{23} & R_{33}^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \tag{2.25}$$

we want to solve the new linear equation

$$\begin{bmatrix} \tilde{R}_{11}^T & 0 \\ \tilde{R}_{12}^T & \tilde{R}_{22}^T \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_3 \end{bmatrix}. \tag{2.26}$$

From (2.23), (2.25), and (2.26), we have

$$\tilde{x}_1 = x_1, \tag{2.27}$$

and

$$\begin{aligned} \tilde{R}_{22}^T \tilde{x}_2 &= y_3 - \tilde{R}_{12}^T \tilde{x}_1 \\ &= y_3 - R_{13}^T x_1 \\ &= r_{23} x_2 + R_{33}^T x_3. \end{aligned} \tag{2.28}$$

The relations in (2.24) and (2.28) could be written together in a linear system as

$$\begin{bmatrix} \tilde{R}_{22}^T & 0 \end{bmatrix} \begin{bmatrix} \tilde{R}_{22} & \tilde{x}_2 \\ 0 & * \end{bmatrix} = \begin{bmatrix} r_{23} & R_{33}^T \end{bmatrix} \begin{bmatrix} r_{23}^T & x_2 \\ R_{33} & x_3 \end{bmatrix}. \tag{2.29}$$

We can apply Givens rotations to $\begin{bmatrix} r_{23}^T & x_2 \\ R_{33} & x_3 \end{bmatrix}$ to calculate $\tilde{R}_{22}$ and $\tilde{x}_2$ simultaneously.

The original matrix $\begin{bmatrix} r_{23}^T & x_2 \\ R_{33} & x_3 \end{bmatrix}$ is in the form as

$$\begin{bmatrix} a & * & \cdots & * & * \\ b & * & \cdots & * & * \\ 0 & * & * & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & * & * \end{bmatrix}.$$

After the first rotation, the matrix would be changed into the form as

$$\begin{bmatrix} \tilde{a} & * & \cdots & * & * \\ 0 & c & \cdots & * & * \\ 0 & d & * & \vdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & * & * \end{bmatrix},$$

where (a, b) in first column would be changed into $(\tilde{a}, 0)$.

After the second rotation, (c, d) in second column would be changed into $(\tilde{c}, 0)$. Following this way, eventually, we would have the final matrix as

$$\begin{bmatrix} \tilde{R}_{22} & \tilde{x}_2 \\ 0 & * \end{bmatrix}.$$

2.3 Eigen-decomposition

To test the performance of our proposed QP solver, we generate challenging problems. One way we used to do this is to make W in (1.2) more ill-conditioned. We did this in the following manner.

We can always write a symmetric positive definite matrix W_0 as

$$W_0 = Q\Lambda_0 Q^T, \tag{2.30}$$

where Q is a orthonormal matrix and Λ_0 is a diagonal matrix with eigenvalues of W_0. Supposing that such a Q has been computed, when we want W with some condition number τ, we first make $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ and $\tau = \frac{\lambda_n}{\lambda_1}$. Then we could set

$$W = Q\Lambda Q^T. \tag{2.31}$$

In this manner, we could generate a random symmetric positive definite matrix W with a particular condition number of τ as we want.

Chapter 3

Algorithm

This chapter contains two sections. We first mention some basic ideas of our algorithm. Then we introduce the detailed algorithm step by step.

3.1 Algorithm Basics

In each iterate of the algorithm, we compute the solution of the linear system

$$\begin{bmatrix} G_{:,\mathcal{S}}^T W G_{:,\mathcal{S}} & G_{:,\mathcal{S}}^T W_{:,\mathcal{P}} & G_{:,\mathcal{S}}^T W_{:,\mathcal{N}} & \mathbb{1} \\ W_{\mathcal{P},:} G_{:,\mathcal{S}} & W_{\mathcal{P},\mathcal{P}} & W_{\mathcal{P},\mathcal{N}} & 0 \\ W_{\mathcal{N},:} G_{:,\mathcal{S}} & W_{\mathcal{N},\mathcal{P}} & W_{\mathcal{N},\mathcal{N}} & 0 \\ \mathbb{1}^T & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \omega_{\mathcal{S}} \\ \gamma_{\mathcal{P}} \\ \gamma_{\mathcal{N}} \\ z \end{bmatrix} = \begin{bmatrix} b_{\mathcal{S}} \\ -\delta\mathbb{1} \\ \delta\mathbb{1} \\ 1 \end{bmatrix} \tag{3.1}$$

based on current choices of the index sets $\mathcal{S}$, $\mathcal{P}$, and $\mathcal{N}$.

There are two important things in the algorithm:

1) We update our choices of the index sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ in each iteration. Some specific indices are appended to/deleted from the sets to improve the objective function value. The algorithm terminates when the current index sets $\mathcal{S}$, $\mathcal{P}$, and $\mathcal{N}$ are optimal. Then we can also get a solution (ω, γ) satisfying (2.1), which is an optimal solution for (1.2) as well.

2) In each iteration, before we try to append some $j \notin \mathcal{S}$ into index sets $\mathcal{S}$, we should always

make sure the matrix in (3.1) will be invertible. We can do a rank-deficiency check by solving

$$\left(\begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}} & W_{:,\mathcal{N}} \\ W_{\mathcal{P},:} & W_{\mathcal{P},\mathcal{P}} & W_{\mathcal{P},\mathcal{N}} \\ W_{\mathcal{N},:} & W_{\mathcal{N},\mathcal{P}} & W_{\mathcal{N},\mathcal{N}} \end{bmatrix} \begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \mathbb{1}^T & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \omega_{\mathcal{S}} \\ \gamma_{\mathcal{P}} \\ \gamma_{\mathcal{N}} \end{bmatrix}$$
$$= \begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}} & W_{:,\mathcal{N}} \\ W_{\mathcal{P},:} & W_{\mathcal{P},\mathcal{P}} & W_{\mathcal{P},\mathcal{N}} \\ W_{\mathcal{N},:} & W_{\mathcal{N},\mathcal{P}} & W_{\mathcal{N},\mathcal{N}} \end{bmatrix} \begin{bmatrix} g_j \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix}, \tag{3.2}$$

and then check whether

$$\mathbb{1}^T \omega_{\mathcal{S}} = 1 \text{ and } G_{:,\mathcal{S}} \omega_{\mathcal{S}} = g_j. \tag{3.3}$$

We can solve (3.1) and (3.2) by Cholesky factorization update mentioned in Chapter 2. Maintaining an upper triangular matrix R satisfying

$$R^T R = \begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}} & W_{:,\mathcal{N}} \\ W_{\mathcal{P},:} & W_{\mathcal{P},\mathcal{P}} & W_{\mathcal{P},\mathcal{N}} \\ W_{\mathcal{N},:} & W_{\mathcal{N},\mathcal{P}} & W_{\mathcal{N},\mathcal{N}} \end{bmatrix} \begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \mathbb{1}^T & 0 & 0 \end{bmatrix}, \tag{3.4}$$

then by solving

$$\begin{cases} R^T \begin{bmatrix} \omega_1 \\ \gamma_{1,\mathcal{P}} \\ \gamma_{1,\mathcal{N}} \end{bmatrix} = \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix} \\ R^T \begin{bmatrix} \omega_2 \\ \gamma_{2,\mathcal{P}} \\ \gamma_{2,\mathcal{N}} \end{bmatrix} = \begin{bmatrix} b_{\mathcal{S}} \\ -\delta\mathbb{1} \\ \delta\mathbb{1} \end{bmatrix} \\ z = \left(\left\| \begin{bmatrix} \omega_1 \\ \gamma_{1,\mathcal{P}} \\ \gamma_{1,\mathcal{N}} \end{bmatrix} \right\|^2 + \begin{bmatrix} \omega_1 \\ \gamma_{1,\mathcal{P}} \\ \gamma_{1,\mathcal{N}} \end{bmatrix}^T \begin{bmatrix} \omega_2 \\ \gamma_{2,\mathcal{P}} \\ \gamma_{2,\mathcal{N}} \end{bmatrix} - 1 \right) \Big/ \left\| \begin{bmatrix} \omega_1 \\ \gamma_{1,\mathcal{P}} \\ \gamma_{1,\mathcal{N}} \end{bmatrix} \right\|^2 \\ R \begin{bmatrix} \omega_{\mathcal{S}} \\ \gamma_{\mathcal{P}} \\ \gamma_{\mathcal{N}} \end{bmatrix} = (1-z) \begin{bmatrix} \omega_1 \\ \gamma_{1,\mathcal{P}} \\ \gamma_{1,\mathcal{N}} \end{bmatrix} + \begin{bmatrix} \omega_2 \\ \gamma_{2,\mathcal{P}} \\ \gamma_{2,\mathcal{N}} \end{bmatrix} \end{cases} \tag{3.5}$$

and

$$\begin{cases} R^T \begin{bmatrix} \omega_3 \\ \gamma_{3,\mathcal{P}} \\ \gamma_{3,\mathcal{N}} \end{bmatrix} = \begin{bmatrix} G_{:,\mathcal{S}} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}} & W_{:,\mathcal{N}} \\ W_{\mathcal{P},:} & W_{\mathcal{P},\mathcal{P}} & W_{\mathcal{P},\mathcal{N}} \\ W_{\mathcal{N},:} & W_{\mathcal{N},\mathcal{P}} & W_{\mathcal{N},\mathcal{N}} \end{bmatrix} \begin{bmatrix} g_j \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix} \\ R \begin{bmatrix} \omega_{\mathcal{S}} \\ \gamma_{\mathcal{P}} \\ \gamma_{\mathcal{N}} \end{bmatrix} = \begin{bmatrix} \omega_3 \\ \gamma_{3,\mathcal{P}} \\ \gamma_{3,\mathcal{N}} \end{bmatrix}, \end{cases} \tag{3.6}$$

we can get solutions of (3.1) and (3.2) separately.

If (3.3) is satisfied, then the matrix in (3.1) would not be invertible. In such a case, we perform a swap to use such j to substitute some $j' \in \mathcal{S}$. If (3.3) is not satisfied, then we can skip the swap step and simply append the newly index to $\mathcal{S}$.

Our algorithm is formally stated in the next subsection. In Step 1, we initialize the solution estimate by selecting index sets $(\mathcal{S}, \mathcal{P}, \mathcal{N})$, whose corresponding solution of (3.1) satisfies positive-negative sign constraints. Generally speaking, we can always set $\mathcal{S}$ including only one element and both $\mathcal{P}$ and $\mathcal{N}$ being empty to satisfy our goal. Then we go to Step 2.

In Step 2, we always check whether the current feasible solution satisfies KKT conditions. If so, we can conclude that our current solution is the optimal solution; if not, we could find such $j \notin \mathcal{S}$ or $i \notin \mathcal{P} \cup \mathcal{N}$ that by including the index into $(\mathcal{S}, \mathcal{P}, \mathcal{N})$, we can potentially improve the objective function value. Then we go to Step 3.

Step 3 is mainly the pre-test for our new index candidate, to check whether appending the index into current index sets would result in rank-deficiency or not. If so, we go to Step 5 to do column exchange to remove some index from a current index set to avoid rank-deficiency; if not, we would go to Step 4 to append the index into the current index sets without an exchange.

No matter if the algorithm performs a column exchange or column augmentation, we always go to Step 6 to solve (3.1) corresponding to the new index sets, say $(\mathcal{S}_1, \mathcal{P}_1, \mathcal{N}_1)$. If the solution satisfies $\omega_{\mathcal{S}_1} > 0$, $\gamma_{\mathcal{P}_1} > 0$ and $\gamma_{\mathcal{N}_1} < 0$, we get a feasible solution for the new index sets, meaning we can go back to Step 2 again to check whether the (new) current solution is optimal or not. If the solution violates one of inequalities above, we have to go to Step 7 to remove an index (or more) from the index sets corresponding to a zero element after taking a convex combination of the previous and current trial solution. Then we go back to Step 6 again to check whether the solution of (3.1) corresponding to the new index sets satisfies those inequalities or not. If yes, we go to Step 2 to continue on the next iterate; if not, we have to go to Step 7 to delete more indices.

3.2 Algorithm Details

We provide details of the algorithm for the QP Solver as follows:

Algorithm 1 QP Solver

1: **procedure** : (Initialization) Choose $(\mathcal{S}, \mathcal{P}, \mathcal{N})$ such that the solution $(\hat{\omega}_\mathcal{S}, \hat{\gamma}_\mathcal{P}, \hat{\gamma}_\mathcal{N}, \hat{z})$ of (3.1) has $\hat{\omega}_\mathcal{S} \geq 0$, $\hat{\gamma}_\mathcal{P} \geq 0$, and $\hat{\gamma}_\mathcal{N} \leq 0$. Set the remaining elements of these vectors to zero.

2: (Termination check) if (2.4) holds, then termination. Otherwise, if $i \notin \mathcal{P} \cup \mathcal{N}$ exists such that (2.4b) or (2.4c) violates, then choose i. Otherwise, choose $j \notin \mathcal{S}$ such that (2.4a) violates.

3: (Rank-deficiency check) Solve (3.2) for $(\tilde{\omega}_\mathcal{S}, \tilde{\gamma}_\mathcal{P}, \tilde{\gamma}_\mathcal{N})$. If

$$\begin{bmatrix} \mathbb{1}^T \\ G_{:,\mathcal{S}} \end{bmatrix} \tilde{\omega}_\mathcal{S} = \begin{bmatrix} 1 \\ g_j \end{bmatrix},$$

then go to step 5; otherwise, continue.

4: (Column augmentation) Append i to $\mathcal{P}$ or $\mathcal{N}$ or j to $\mathcal{S}$ as appropriate based on the choice in Step 2. Append a zero element to $\hat{\omega}_\mathcal{S}$, $\hat{\gamma}_\mathcal{P}$ or $\hat{\gamma}_\mathcal{N}$ corresponding to the newly added index. Go to Step 6.

5: (Column exchange) Replace $\hat{\omega}_\mathcal{S}$ by $\hat{\omega}_\mathcal{S} - t\tilde{\omega}_\mathcal{S}$ where

$$t \leftarrow \min_k \{\hat{\omega}_k / \tilde{\omega}_k : \tilde{\omega}_k > 0, k \in \mathcal{S}\}.$$

Find some index corresponding to a zero element of $\hat{\omega}_\mathcal{S}$. Delete this element along with the corresponding index from $\mathcal{S}$. Append j to $\mathcal{S}$. Append an element with value t to $\hat{\omega}_\mathcal{S}$ corresponding to the newly added index.

6: (Subproblem solution) Solve (3.1) for $(\bar{\omega}_\mathcal{S}, \bar{\gamma}_\mathcal{P}, \bar{\gamma}_\mathcal{N}, \bar{z})$. If $\bar{\omega}_\mathcal{S} > 0$, $\bar{\gamma}_\mathcal{P} > 0$ and $\bar{\gamma}_\mathcal{N} < 0$, then set $(\hat{\omega}_\mathcal{S}, \hat{\gamma}_\mathcal{P}, \hat{\gamma}_\mathcal{N}) = (\bar{\omega}_\mathcal{S}, \bar{\gamma}_\mathcal{P}, \bar{\gamma}_\mathcal{N})$ and go to Step 2; otherwise, continue.

7: (Column deletion) Replace $(\hat{\omega}_\mathcal{S}, \hat{\gamma}_\mathcal{P}, \hat{\gamma}_\mathcal{N})$ by $t(\bar{\omega}_\mathcal{S}, \bar{\gamma}_\mathcal{P}, \bar{\gamma}_\mathcal{N}) + (1-t)(\hat{\omega}_\mathcal{S}, \hat{\gamma}_\mathcal{P}, \hat{\gamma}_\mathcal{N})$ where

$$t_1 \leftarrow \min\{1, \min_k \{\hat{\omega}_k / (\hat{\omega}_k - \bar{\omega}_k) : \bar{\omega}_k < 0, k \in \mathcal{S}\}\},$$

$$t_2 \leftarrow \min\{1, \min_k \{\hat{\gamma}_k / (\hat{\gamma}_k - \bar{\gamma}_k) : \bar{\gamma}_k < 0, k \in \mathcal{P}\}\},$$

$$t_3 \leftarrow \min\{1, \min_k \{\hat{\gamma}_k / (\hat{\gamma}_k - \bar{\gamma}_k) : \bar{\gamma}_k > 0, k \in \mathcal{N}\}\},$$

$$\text{and } t \leftarrow \min\{t_1, t_2, t_3\}.$$

Find some index corresponding to a zero element of $(\hat{\omega}_\mathcal{S}, \hat{\gamma}_\mathcal{P}, \hat{\gamma}_\mathcal{N})$. Delete this element along with the corresponding index from $\mathcal{S}$, $\mathcal{P}$, or $\mathcal{N}$. Go to Step 6.

Chapter 4

Convergence of the Algorithm

In this chapter, we are going to prove that under some assumptions, the algorithm always finds the optimal solution of (1.2) in a finite number of steps.

Our first lemma shows that if the current index sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ are optimal, then the solution obtained in Step 6 by solving (3.1) corresponding to $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ is optimal.

Lemma 4.1. *If $\mathcal{S}$, $\mathcal{P}$, and $\mathcal{N}$ are chosen as the index sets corresponding to the solution of* (2.4)*, the inequalities in* (2.4b) *and* (2.4c) *are satisfied strictly, and the inequalities in* (2.4a) *are satisfied strictly only if $j \in \mathcal{J} \backslash \mathcal{S}$, then by solving the linear system* (3.1) *and setting $\omega_{\mathcal{J}\backslash\mathcal{S}} \leftarrow 0$ and $\gamma_{\mathcal{I}\backslash(\mathcal{P}\cup\mathcal{N})} \leftarrow 0$, one obtains the solution (ω, γ) for* (2.1)*.*

Proof. For (ω^*, γ^*) satisfying KKT conditions (2.1), knowing optimal sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ under assumptions we have made, combining (2.2), we would have a unique $(\omega^*_{\mathcal{S}}, \gamma^*_{\mathcal{P}\cup\mathcal{N}})$ satisfying (2.4). We first want to prove there is some $(\omega^*_{\mathcal{S}}, \gamma^*_{\mathcal{P}}, \gamma^*_{\mathcal{N}}, z^*)$ related to $(\omega^*_{\mathcal{S}}, \gamma^*_{\mathcal{P}\cup\mathcal{N}})$, being the solution of (3.1). Let $z^* = \omega^{*T}_{\mathcal{S}}(b_{\mathcal{S}} - G^T_{:,\mathcal{S}} W(G_{:,\mathcal{S}}\omega^*_{\mathcal{S}} + \gamma^*_{\mathcal{P}\cup\mathcal{N}}))$ and consider (2.4a), because of the assumption made in the statement of the lemma, we know that

$$z^* = b_j - g_j^T W(G_{:,\mathcal{S}}\omega^*_{\mathcal{S}} + \gamma^*_{\mathcal{P}\cup\mathcal{N}}) \text{ for all } j \in \mathcal{S}. \tag{4.1}$$

This leads to

$$\begin{aligned} b_{\mathcal{S}} &= z^*\mathbb{1} + G^T_{:,\mathcal{S}} W(G_{:,\mathcal{S}}\omega^*_{\mathcal{S}} + \gamma^*_{\mathcal{P}\cup\mathcal{N}}) \\ &= z^*\mathbb{1} + G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}}\omega^*_{\mathcal{S}} + G^T_{:,\mathcal{S}} W_{:,\mathcal{P}}\gamma^*_{\mathcal{P}} + G^T_{:,\mathcal{S}} W_{:,\mathcal{N}}\gamma^*_{\mathcal{N}}. \end{aligned} \tag{4.2}$$

Hence, $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ satisfies the first equation in (3.1).

Consider the last three equations in (2.4). First, (2.4d) is the last equation in (3.1). Second, (2.4e) is the same as

$$\begin{aligned} 0 &= \delta\mathbb{1} + [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*)]_{\mathcal{P}} \\ &= \delta\mathbb{1} + W_{\mathcal{P},:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*) \\ &= \delta\mathbb{1} + W_{\mathcal{P},:}G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + W_{\mathcal{P},\mathcal{P}}\gamma_{\mathcal{P}}^* + W_{\mathcal{P},\mathcal{N}}\gamma_{\mathcal{N}}^*, \end{aligned} \tag{4.3}$$

which means $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ satisfies the second equation in (3.1). In a similar way, (2.4f) is the same as

$$\begin{aligned} 0 &= \delta\mathbb{1} - [W(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*)]_{\mathcal{N}} \\ &= \delta\mathbb{1} - W_{\mathcal{N},:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*) \\ &= \delta\mathbb{1} - (W_{\mathcal{N},:}G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + W_{\mathcal{N},\mathcal{P}}\gamma_{\mathcal{P}}^* + W_{\mathcal{N},\mathcal{N}}\gamma_{\mathcal{N}}^*), \end{aligned} \tag{4.4}$$

so the third equation in (3.1) is satisfied by $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ as well. Overall, (2.4d)-(2.4f) are the same as the last three equations in (3.1) and $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ is a solution of (3.1).

Let $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z)$ be the solution of (3.1), then we know $z = \omega_{\mathcal{S}}^T(b_{\mathcal{S}} - G_{:,\mathcal{S}}^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}}))$. To finish the whole proof, all we need to show is

$$z > b_j - g_j^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}}), \text{ for all } j \in \mathcal{J}\backslash\mathcal{S} \tag{4.5a}$$

$$W_{i,:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}}) > -\delta, \text{ for all } i \in \mathcal{I}\backslash\mathcal{P} \tag{4.5b}$$

$$W_{i,:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma_{\mathcal{P}\cup\mathcal{N}}) < \delta, \text{ for all } i \in \mathcal{I}\backslash\mathcal{N}. \tag{4.5c}$$

Assume (4.5) is violated by $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z)$, which also means $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z) \neq (\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$. Because $\mathcal{S}$, $\mathcal{P}$, and $\mathcal{N}$ are all optimal sets, then for $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$, we have

$$z^* > b_j - g_j^T W(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*), \text{ for all } j \in \mathcal{J}\backslash\mathcal{S} \tag{4.6a}$$

$$W_{i,:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*) > -\delta, \text{ for all } i \in \mathcal{I}\backslash\mathcal{P} \tag{4.6b}$$

$$W_{i,:}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}^* + \gamma_{\mathcal{P}\cup\mathcal{N}}^*) < \delta, \text{ for all } i \in \mathcal{I}\backslash\mathcal{N}. \tag{4.6c}$$

Because $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z)$ and $(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ are both solutions of (3.1), then $(\omega_{\mathcal{S}}^\alpha, \gamma_{\mathcal{P}}^\alpha, \gamma_{\mathcal{N}}^\alpha, z^\alpha) = \alpha(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z) + (1-\alpha)(\omega_{\mathcal{S}}^*, \gamma_{\mathcal{P}}^*, \gamma_{\mathcal{N}}^*, z^*)$ are all the solutions of (3.1) for any $\alpha \in (0, 1)$. The inequalities in (4.6) are strict, so when α goes down to 0, we always have $(\omega_{\mathcal{S}}^{\alpha}, \gamma_{\mathcal{P}}^{\alpha}, \gamma_{\mathcal{N}}^{\alpha}, z^{\alpha})$ satisfying (4.6). So $(\omega_{\mathcal{S}}^{\alpha}, \gamma_{\mathcal{P}}^{\alpha}, \gamma_{\mathcal{N}}^{\alpha}, z^{\alpha})$ is a solution of (2.4). Because there is only a unique solution satisfying (2.4), we can conclude $(\omega_{\mathcal{S}}^{\alpha}, \gamma_{\mathcal{P}}^{\alpha}, \gamma_{\mathcal{N}}^{\alpha}, z^{\alpha}) = (\omega_{\mathcal{S}}^{*}, \gamma_{\mathcal{P}}^{*}, \gamma_{\mathcal{N}}^{*}, z^{*})$, which means $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z) = (\omega_{\mathcal{S}}^{*}, \gamma_{\mathcal{P}}^{*}, \gamma_{\mathcal{N}}^{*}, z^{*})$, leading to a contradiction.

So (4.5) is satisfied by $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z)$, for the unique solution of original problem, we have $(\omega_{\mathcal{S}}, \gamma_{\mathcal{P}}, \gamma_{\mathcal{N}}, z) = (\omega_{\mathcal{S}}^{*}, \gamma_{\mathcal{P}}^{*}, \gamma_{\mathcal{N}}^{*}, z^{*})$. Hence, by setting $\omega_{\mathcal{J}\setminus\mathcal{S}} \leftarrow 0$ and $\gamma_{\mathcal{I}\setminus(\mathcal{P}\cup\mathcal{N})} \leftarrow 0$, we would get the solution of (2.1). $\square$

Our second lemma is to prove that under some appropriate assumptions, through column augmentation process, the objective function value would always be improved by iterates of the algorithm.

Lemma 4.2. *Let the current iterate have index sets as $\mathcal{S}_0$, $\mathcal{P}_0$ and $\mathcal{N}_0$, which are not optimal. Assume the newly added index passed the rank-deficiency check (Step 3) and we go to Step 4 (column augmentation) in the algorithm. After doing Step 4, we always go through Step 6 and Step 7 (if necessary), then we will get $(\hat{\omega}_{\mathcal{S}_0'}, \hat{\gamma}_{\mathcal{P}_0'}, \hat{\gamma}_{\mathcal{N}_0'})$ as a feasible solution of* (3.1) *with a no worse objective value, where $\mathcal{S}_0'$, $\mathcal{P}_0'$ and $\mathcal{N}_0'$ are the index sets at next iterate. If we finish the current iterate without column deletion processes, or by solving* (3.1) *at Step 6, the solution's element corresponding to the newly added index being positive for $\mathcal{S}_0$ and $\mathcal{P}_0$ or being negative for $\mathcal{N}_0$, we will always get the $(\hat{\omega}_{\mathcal{S}_0'}, \hat{\gamma}_{\mathcal{P}_0'}, \hat{\gamma}_{\mathcal{N}_0'})$ with a better objective value.*

Proof. We first want to show that given index sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$, and $(\omega_{\mathcal{S}}', \gamma_{\mathcal{P}}', \gamma_{\mathcal{N}}', z')$ as the solution of (3.1), then by setting

$$[\omega']_j = \begin{cases} [\omega_{\mathcal{S}}']_j \text{ , if } j \in \mathcal{S}; \\ 0 \qquad\quad \text{, otherwise;} \end{cases} \quad \text{and } [\gamma']_i = \begin{cases} [\gamma_{\mathcal{P}\cup\mathcal{N}}']_i \text{ , if } i \in \mathcal{P}\cup\mathcal{N}; \\ 0 \qquad\qquad \text{, otherwise,} \end{cases} \tag{4.7}$$

and

$$\omega_j = \begin{cases} [\omega_{\mathcal{S}}]_j \text{ , if } j \in \mathcal{S}; \\ 0 \qquad\quad \text{, otherwise;} \end{cases} \quad \text{and } \gamma_i = \begin{cases} [\gamma_{\mathcal{P}\cup\mathcal{N}}]_i \text{ , if } i \in \mathcal{P}\cup\mathcal{N}; \\ 0 \qquad\qquad \text{, otherwise,} \end{cases} \tag{4.8}$$

where $\mathbb{1}^T\omega_{\mathcal{S}} = 1$, $\omega_{\mathcal{S}} \geq 0$, $\gamma_{\mathcal{P}} \geq 0$ and $\gamma_{\mathcal{N}} \leq 0$, if $\omega_{\mathcal{S}}' > 0$, $\gamma_{\mathcal{P}}' > 0$ and $\gamma_{\mathcal{N}}' < 0$, we would get a no worse objective value with (ω', γ') than with (ω, γ); or else, for some $t \in [0, 1)$, we would have

$(\omega'', \gamma'') = t(\omega', \gamma') + (1-t)(\omega, \gamma)$ where $\omega''_{\mathcal{S}} \geq 0$, $\gamma''_{\mathcal{P}} \geq 0$ and $\gamma''_{\mathcal{N}} \leq 0$, has a no worse objective value than with (ω, γ).

First, in condition of $\omega'_{\mathcal{S}} > 0$, $\gamma'_{\mathcal{P}} > 0$ and $\gamma'_{\mathcal{N}} < 0$, we have the objective function value with (ω', γ') as

$$\begin{aligned} f' &= \frac{1}{2}(G\omega' + \gamma')^T W (G\omega' + \gamma') - b^T \omega' + \delta \|\gamma'\|_1 \\ &= \frac{1}{2}(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma') - b_{\mathcal{S}}^T \omega'_{\mathcal{S}} + \delta \|\gamma'\|_1 \\ &= \frac{1}{2}\omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma'_{\mathcal{P}} + \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma'_{\mathcal{N}} + \frac{1}{2}\gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma'_{\mathcal{N}} \\ &+ \frac{1}{2}\gamma'^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma'_{\mathcal{N}} - b^T_{\mathcal{S}} \omega'_{\mathcal{S}} + \delta \mathbb{1}^T \gamma'_{\mathcal{P}} - \delta \mathbb{1}^T \gamma'_{\mathcal{N}}. \end{aligned} \tag{4.9}$$

We also have the objective function value with (ω, γ) as

$$\begin{aligned} f &= \frac{1}{2}(G\omega + \gamma)^T W (G\omega + \gamma) - b^T \omega + \delta \|\gamma\|_1 \\ &= \frac{1}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma)^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) - b_{\mathcal{S}}^T \omega_{\mathcal{S}} + \delta \|\gamma\|_1 \\ &= \frac{1}{2}\omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega_{\mathcal{S}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma_{\mathcal{P}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma_{\mathcal{N}} + \frac{1}{2}\gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma_{\mathcal{P}} + \gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma_{\mathcal{N}} \\ &+ \frac{1}{2}\gamma^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma_{\mathcal{N}} - b^T_{\mathcal{S}} \omega_{\mathcal{S}} + \delta \mathbb{1}^T \gamma_{\mathcal{P}} - \delta \mathbb{1}^T \gamma_{\mathcal{N}}. \end{aligned} \tag{4.10}$$

Because $(\omega'_{\mathcal{S}}, \gamma'_{\mathcal{P}}, \gamma'_{\mathcal{N}}, z')$ satisfies (3.1) corresponding to index sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$, then we have

$$b^T_{\mathcal{S}} \omega'_{\mathcal{S}} = \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma'_{\mathcal{P}} + \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma'_{\mathcal{N}} + z' \tag{4.11a}$$

$$-\delta \mathbb{1}^T \gamma'_{\mathcal{P}} = \gamma'^T_{\mathcal{P}} W_{\mathcal{P},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma'_{\mathcal{N}} \tag{4.11b}$$

$$\delta \mathbb{1}^T \gamma'_{\mathcal{N}} = \gamma'^T_{\mathcal{N}} W_{\mathcal{N},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \gamma'^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma'^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma'_{\mathcal{N}} \tag{4.11c}$$

$$b^T_{\mathcal{S}} \omega_{\mathcal{S}} = \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma'_{\mathcal{P}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma'_{\mathcal{N}} + z' \tag{4.11d}$$

$$-\delta \mathbb{1}^T \gamma_{\mathcal{P}} = \gamma^T_{\mathcal{P}} W_{\mathcal{P},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma'_{\mathcal{N}} \tag{4.11e}$$

$$\delta \mathbb{1}^T \gamma_{\mathcal{N}} = \gamma^T_{\mathcal{N}} W_{\mathcal{N},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \gamma^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma'_{\mathcal{N}} \tag{4.11f}$$

Combining (4.9), (4.10) and (4.11), we would have

$$\begin{aligned}
f' - f &= -\frac{1}{2}\omega_{\mathcal{S}}'^T G_{:,\mathcal{S}}^T W G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - z' - \gamma_{\mathcal{P}}'^T W_{\mathcal{P},:} G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - \gamma_{\mathcal{N}}'^T W_{\mathcal{N},:} G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - \frac{1}{2}\gamma_{\mathcal{P}}'^T W_{\mathcal{P},\mathcal{P}}\gamma_{\mathcal{P}}' \\
&\quad - \gamma_{\mathcal{P}}'^T W_{\mathcal{P},\mathcal{N}}\gamma_{\mathcal{N}}' - \frac{1}{2}\gamma_{\mathcal{N}}'^T W_{\mathcal{N},\mathcal{N}}\gamma_{\mathcal{N}}' - \frac{1}{2}\omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W G_{:,\mathcal{S}}\omega_{\mathcal{S}} - \omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W_{:,\mathcal{P}}\gamma_{\mathcal{P}} \\
&\quad - \omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W_{:,\mathcal{N}}\gamma_{\mathcal{N}} - \frac{1}{2}\gamma_{\mathcal{P}}^T W_{\mathcal{P},\mathcal{P}}\gamma_{\mathcal{P}} - \gamma_{\mathcal{P}}^T W_{\mathcal{P},\mathcal{N}}\gamma_{\mathcal{N}} - \frac{1}{2}\gamma_{\mathcal{N}}^T W_{\mathcal{N},\mathcal{N}}\gamma_{\mathcal{N}} + \omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W G_{:,\mathcal{S}}\omega_{\mathcal{S}}' \\
&\quad + \omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W_{:,\mathcal{P}}\gamma_{\mathcal{P}}' + \omega_{\mathcal{S}}^T G_{:,\mathcal{S}}^T W_{:,\mathcal{N}}\gamma_{\mathcal{N}}' + z' + \gamma_{\mathcal{P}}^T W_{\mathcal{P},:} G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma_{\mathcal{P}}^T W_{\mathcal{P},\mathcal{P}}\gamma_{\mathcal{P}}' + \gamma_{\mathcal{P}}^T W_{\mathcal{P},\mathcal{N}}\gamma_{\mathcal{N}}' \\
&\quad + \gamma_{\mathcal{N}}^T W_{\mathcal{N},:} G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma_{\mathcal{N}}^T W_{\mathcal{N},\mathcal{P}}\gamma_{\mathcal{P}}' + \gamma_{\mathcal{N}}^T W_{\mathcal{N},\mathcal{N}}\gamma_{\mathcal{N}}' \\
&= -\frac{1}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - G_{:,\mathcal{S}}\omega_{\mathcal{S}})^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - G_{:,\mathcal{S}}\omega_{\mathcal{S}}) - \frac{1}{2}(\gamma' - \gamma)^T W(\gamma' - \gamma) \\
&\quad - (G_{:,\mathcal{S}}\omega_{\mathcal{S}}' - G_{:,\mathcal{S}}\omega_{\mathcal{S}})^T W(\gamma' - \gamma) \\
&= -\frac{1}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma' - G_{:,\mathcal{S}}\omega_{\mathcal{S}} - \gamma)^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma' - G_{:,\mathcal{S}}\omega_{\mathcal{S}} - \gamma) \\
&\leq 0,
\end{aligned} \tag{4.12}$$

because W is a positive definite matrix. Hence, $f' - f = 0$ if and only if $G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma' = G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma$.

Second, if some inequality among $\omega_{\mathcal{S}}' > 0$, $\gamma_{\mathcal{P}}' > 0$ and $\gamma_{\mathcal{N}}' < 0$ is violated, we have the objective function value with (ω'', γ'') as

$$\begin{aligned}
f'' &= \frac{1}{2}(G\omega'' + \gamma'')^T W (G\omega'' + \gamma'') - b^T\omega'' + \delta\|\gamma''\|_1 \\
&= \frac{1}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}}'' + \gamma'')^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}}'' + \gamma'') - b_{\mathcal{S}}^T\omega_{\mathcal{S}}'' + \delta\|\gamma''\|_1 \\
&= \frac{1}{2}(t(G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma') + (1-t)(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma))^T W (t(G_{:,\mathcal{S}}\omega_{\mathcal{S}}' + \gamma') + (1-t)(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma)) \\
&\quad - (t b_{\mathcal{S}}^T\omega_{\mathcal{S}}' + (1-t) b_{\mathcal{S}}^T\omega_{\mathcal{S}}) + t(\delta\mathbb{1}^T\gamma_{\mathcal{P}}' - \delta\mathbb{1}^T\gamma_{\mathcal{N}}') + (1-t)(\delta\mathbb{1}^T\gamma_{\mathcal{P}} - \delta\mathbb{1}^T\gamma_{\mathcal{N}})
\end{aligned} \tag{4.13}$$

Combining (4.10), (4.13) and (4.11), we would have

$$
\begin{aligned}
f'' - f = &\frac{t^2}{2}(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma') + t(1-t)(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) \\
&+ \frac{t^2 - 2t}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma)^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) - t(b_{\mathcal{S}}^T \omega'_{\mathcal{S}} - b_{\mathcal{S}}^T \omega_{\mathcal{S}}) \\
&+ t((\delta \mathbb{1}^T \gamma'_{\mathcal{P}} - \delta \mathbb{1}^T \gamma'_{\mathcal{N}}) - (\delta \mathbb{1}^T \gamma_{\mathcal{P}} - \delta \mathbb{1}^T \gamma_{\mathcal{N}})) \\
= &\frac{t^2}{2}(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma') - t^2(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) \\
&+ \frac{t^2}{2}(G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma)^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) + t((G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma')^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) \\
&- (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma)^T W (G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma) - \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega'_{\mathcal{S}} - \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma'_{\mathcal{P}} - \omega'^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma'_{\mathcal{N}} \\
&+ \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{P}} \gamma'_{\mathcal{P}} + \omega^T_{\mathcal{S}} G^T_{:,\mathcal{S}} W_{:,\mathcal{N}} \gamma'_{\mathcal{N}} - \gamma'^T_{\mathcal{P}} W_{\mathcal{P},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} - \gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma'_{\mathcal{P}} \\
&- \gamma'^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma'_{\mathcal{N}} - \gamma'^T_{\mathcal{N}} W_{\mathcal{N},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} - \gamma'^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{P}} \gamma'_{\mathcal{P}} - \gamma'^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma'_{\mathcal{N}} + \gamma^T_{\mathcal{P}} W_{\mathcal{P},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} \\
&+ \gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma^T_{\mathcal{P}} W_{\mathcal{P},\mathcal{N}} \gamma'_{\mathcal{N}} + \gamma^T_{\mathcal{N}} W_{\mathcal{N},:} G_{:,\mathcal{S}} \omega'_{\mathcal{S}} + \gamma^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{P}} \gamma'_{\mathcal{P}} + \gamma^T_{\mathcal{N}} W_{\mathcal{N},\mathcal{N}} \gamma'_{\mathcal{N}}) \\
= &(\frac{t^2}{2} - t)(G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma' - G_{:,\mathcal{S}}\omega_{\mathcal{S}} - \gamma)^T W (G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma' - G_{:,\mathcal{S}}\omega_{\mathcal{S}} - \gamma) \\
\leq &\, 0,
\end{aligned}
\tag{4.14}
$$

because W is positive definite and $\frac{t^2}{2} - t \leq 0$ for $t \in [0, 1)$. Hence, $f'' - f = 0$ if $G_{:,\mathcal{S}}\omega'_{\mathcal{S}} + \gamma' = G_{:,\mathcal{S}}\omega_{\mathcal{S}} + \gamma$ or $t = 0$.

The arguments above have shown that after proceeding through Step 6 and Step 7 in turn, we would always at least get a no worse objective function value.

Then we consider the column augmentation process. We know that at current iterate's Step 2, positive and negative index sets for ω and γ are $\mathcal{S}_0$, $\mathcal{P}_0$ and $\mathcal{N}_0$. Let $(\omega_{\mathcal{S}_0}, \gamma_{\mathcal{P}_0}, \gamma_{\mathcal{N}_0}, z_0)$ be the solution of (3.1) corresponding to index sets $\mathcal{S}_0$, $\mathcal{P}_0$ and $\mathcal{N}_0$, then we know $\omega_{\mathcal{S}_0} > 0$, $\gamma_{\mathcal{P}_0} > 0$ and $\gamma_{\mathcal{N}_0} < 0$. Because they are not optimal sets, we know there is some $j_0 \notin \mathcal{S}_0$ violating (2.4a) or $i_1 \notin \mathcal{P}_0$ violating (2.4b) or $i_2 \notin \mathcal{N}_0$ violating (2.4c), which passes Step 3 and goes to Step 4. Following by Step 4, we append the corresponding index into index sets and append a zero element to $(\omega_{\mathcal{S}_0}, \gamma_{\mathcal{P}_0}, \gamma_{\mathcal{N}_0})$ corresponding to the newly added index, then we get the newly index sets to be $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$ and $(\omega'_{\mathcal{S}_1}, \gamma'_{\mathcal{P}_1}, \gamma'_{\mathcal{N}_1})$, where

$$[\omega'_{\mathcal{S}_1}]_j = \begin{cases} [\omega_{\mathcal{S}_0}]_j & \text{, if } j \in \mathcal{S}_0; \\ 0 & \text{, if } j = j_0; \end{cases} \quad \text{and } [\gamma'_{\mathcal{P}_1}]_i = \begin{cases} [\gamma_{\mathcal{P}_0}]_i & \text{, if } i \in \mathcal{P}_0; \\ 0 & \text{, if } i = i_1; \end{cases} \quad \text{and } [\gamma'_{\mathcal{N}_1}]_i = \begin{cases} [\gamma_{\mathcal{N}_0}]_j & \text{, if } i \in \mathcal{N}_0; \\ 0 & \text{, if } i = i_2. \end{cases} \tag{4.15}$$

Let $(\omega_{\mathcal{S}_1}, \gamma_{\mathcal{P}_1}, \gamma_{\mathcal{N}_1}, z_1)$ be the solution of (3.1) corresponding to index sets $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$. Set

$$[\gamma_n]_i = \begin{cases} [\gamma_{\mathcal{P}_n \cup \mathcal{N}_n}]_i & \text{, if } i \in \mathcal{P}_n \cup \mathcal{N}_n; \\ 0 & \text{, otherwise;} \end{cases} \quad \text{and } [\gamma'_n]_i = \begin{cases} [\gamma'_{\mathcal{P}_n \cup \mathcal{N}_n}]_i & \text{, if } i \in \mathcal{P}_n \cup \mathcal{N}_n; \\ 0 & \text{, otherwise;} \end{cases} \tag{4.16}$$

and

$$[\omega_n]_j = \begin{cases} [\omega_{\mathcal{S}_n}]_j & \text{, if } j \in \mathcal{S}_n; \\ 0 & \text{, otherwise;} \end{cases} \quad \text{and } [\omega'_n]_j = \begin{cases} [\omega'_{\mathcal{S}_n}]_j & \text{, if } j \in \mathcal{S}_n; \\ 0 & \text{, otherwise.} \end{cases} \tag{4.17}$$

If $\omega_{\mathcal{S}_1} > 0$, $\gamma_{\mathcal{P}_1} > 0$ and $\gamma_{\mathcal{N}_1} < 0$, no column deletion processes are needed, the index sets of next iterate are directly $\mathcal{S}'_0 = \mathcal{S}_1$, $\mathcal{P}'_0 = \mathcal{P}_1$ and $\mathcal{N}'_0 = \mathcal{N}_1$. Moreover, we also have $\omega'_0 = \omega_1$ and $\gamma'_0 = \gamma_1$., then the only thing needs to be proved is (ω_1, γ_1) has a better objective value than (ω_0, γ_0).

Assume we append j_0 into $\mathcal{S}_0$, then we know $\mathcal{P}_1 = \mathcal{P}_0$ and $\mathcal{N}_1 = \mathcal{N}_0$. Because $(\omega_{\mathcal{S}_1}, \gamma_{\mathcal{P}_1}, \gamma_{\mathcal{N}_1}, z_1)$ is a solution of (3.1) corresponding to $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$, then we know $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, or else, we would have

$$\begin{aligned} z_1 &= b_{j_1} - g_{j_1}^T W(G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1) \\ &= b_{j_1} - g_{j_1}^T W(G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1) \\ &= b_{j_1} - g_{j_1}^T W(G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} + \gamma_0) \\ &= z_0, \end{aligned} \tag{4.18}$$

for some $j_1 \in \mathcal{S}_0$. And then we have

$$\begin{aligned} b_{j_0} - g_{j_0}^T W(G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} + \gamma_0) &= b_{j_0} - g_{j_0}^T W(G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1) \\ &= b_{j_0} - g_{j_0}^T W(G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1) \\ &= z_1 \\ &= \omega_{\mathcal{S}_0}^T \mathbb{1} z_0 \\ &= \omega_{\mathcal{S}_0}^T (b_{\mathcal{S}_0} - G_{:,\mathcal{S}_0}^T W(G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} + \gamma_0)), \end{aligned} \tag{4.19}$$

which contradicts the assumption we have made that j_0 violates (2.4a). So $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, because appending zero elements to $\omega_{\mathcal{S}_0}$ does not change objective function value, then from what we have shown above, we know (ω_1, γ_1) has a better objective function value than (ω_0, γ_0).

Or if we append i_1 into $\mathcal{P}_0$, then we know $\mathcal{S}_1 = \mathcal{S}_0$ and $\mathcal{N}_1 = \mathcal{N}_0$. Because $(\omega_{\mathcal{S}_1}, \gamma_{\mathcal{P}_1}, \gamma_{\mathcal{N}_1}, z_1)$ is a solution of (3.1) corresponding to $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$, then we know $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, or else, we would have

$$\begin{aligned} 0 &> \delta + W_{i_1,:}(G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} + \gamma_0) \\ &= \delta + W_{i_1,:}(G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1) \\ &= \delta + W_{i_1,:}(G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1) \\ &\geq 0, \end{aligned} \tag{4.20}$$

which leads to a contradiction. So $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, because appending zero elements to $\gamma_{\mathcal{P}_0}$ does not change objective function value, then from what we have shown above, we know (ω_1, γ_1) has a better objective function value than (ω_0, γ_0).

In a similar way, if we append i_2 into $\mathcal{N}_0$, then we know $\mathcal{S}_1 = \mathcal{S}_0$ and $\mathcal{P}_1 = \mathcal{P}_0$. Because $(\omega'_{\mathcal{S}_1}, \gamma'_{\mathcal{P}_1}, \gamma'_{\mathcal{N}_1})$ is a solution of (3.1) corresponding to $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$, then we know $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq$

$G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, or else, we would have

$$\begin{aligned} 0 &> \delta - W_{i_2,:}(G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} + \gamma_0) \\ &= \delta - W_{i_2,:}(G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1) \\ &= \delta - W_{i_2,:}(G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1) \\ &\geq 0, \end{aligned} \tag{4.21}$$

which leads to a contradiction. So $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega'_{\mathcal{S}_1} + \gamma'_1$, because appending zero elements to $\gamma_{\mathcal{N}_0}$ does not change objective function value, then from what we have shown above, we know (ω_1, γ_1) has a better objective function value than (ω_0, γ_0).

So if after the column augmentation process, we directly get index sets of the next iterate, we can always make an improvement on the objective function value.

If some inequality among $\omega_{\mathcal{S}_1} > 0$, $\gamma_{\mathcal{P}_1} > 0$ and $\gamma_{\mathcal{N}_1} < 0$ is violated, which means we need go further into column deletion processes.

Assume beginning from index sets as $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$, after a column deletion process, we get index sets as $\mathcal{S}_2$, $\mathcal{P}_2$ and $\mathcal{N}_2$. Set $(\omega_{\mathcal{S}_2}, \gamma_{\mathcal{P}_2}, \gamma_{\mathcal{N}_2}, z_2)$ as a solution of (3.1) corresponding to $\mathcal{S}_2$, $\mathcal{P}_2$ and $\mathcal{N}_2$. Let $(\omega''_{\mathcal{S}_1}, \gamma''_{\mathcal{P}_1}, \gamma''_{\mathcal{N}_1}) = t(\omega_{\mathcal{S}_1}, \gamma_{\mathcal{P}_1}, \gamma_{\mathcal{N}_1}) + (1-t)(\omega'_{\mathcal{S}_1}, \gamma'_{\mathcal{P}_1}, \gamma'_{\mathcal{N}_1})$, where t takes the value from Step 7 in Algorithm, then we know $\omega''_{\mathcal{S}_1} \geq 0$, $\gamma''_{\mathcal{P}_1} \geq 0$ and $\gamma_{\mathcal{N}_1} \leq 0$. Some zero element lies in $\omega''_{\mathcal{S}_1}$, $\gamma''_{\mathcal{P}_1}$ and $\gamma''_{\mathcal{N}_1}$, which is the only difference between $(\omega''_{\mathcal{S}_1}, \gamma''_{\mathcal{P}_1}, \gamma''_{\mathcal{N}_1})$ and $(\omega'_{\mathcal{S}_2}, \gamma'_{\mathcal{P}_2}, \gamma'_{\mathcal{N}_2})$. Then we have

$$[\omega'_{\mathcal{S}_2}]_j = \begin{cases} [\omega''_{\mathcal{S}_1}]_j & \text{, if } j \in \mathcal{S}_1; \\ 0 & \text{, otherwise;} \end{cases} \quad \text{and } [\gamma'_{\mathcal{P}_2}]_i = \begin{cases} [\gamma''_{\mathcal{P}_1}]_i & \text{, if } i \in \mathcal{P}_1; \\ 0 & \text{, otherwise;} \end{cases} \quad \text{and } [\gamma'_{\mathcal{N}_2}]_i = \begin{cases} [\gamma''_{\mathcal{N}_1}]_j & \text{, if } i \in \mathcal{N}_1; \\ 0 & \text{, otherwise;} \end{cases} \tag{4.22}$$

so $\mathbb{1}^T\omega'_{\mathcal{S}_2} = 1$, $\omega'_{\mathcal{S}_2} \geq 0$, $\gamma'_{\mathcal{P}_2} \geq 0$ and $\gamma'_{\mathcal{N}_2} \leq 0$. After removing the index corresponding to the zero element, we update $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$ to $\mathcal{S}_2$, $\mathcal{P}_2$ and $\mathcal{N}_2$.

Because $\mathbb{1}^T\omega'_{\mathcal{S}_1} = 1$, $\omega'_{\mathcal{S}_1} \geq 0$, $\gamma'_{\mathcal{P}_1} \geq 0$ and $\gamma'_{\mathcal{N}_1} \leq 0$, which satisfies the condition of statement we have proved at the beginning. From what we have proved before, we know that (ω''_1, γ''_1) has a no worse objective function value than (ω_0, γ_0) and (ω''_1, γ''_1) takes the same objective function value with (ω'_2, γ'_2). So (ω'_2, γ'_2) has a no worse function value than (ω_0, γ_0).

After getting $\mathcal{S}_2$, $\mathcal{P}_2$ and $\mathcal{N}_2$, we can always keep updating index sets through the column deletion process till $\mathcal{S}_k$, $\mathcal{P}_k$ and $\mathcal{N}_k$, where $(\omega_{\mathcal{S}_k}, \gamma_{\mathcal{P}_k}, \gamma_{\mathcal{N}_k}, z_k)$ satisfying (3.1) corresponding to $\mathcal{S}_k$, $\mathcal{P}_k$ and $\mathcal{N}_k$ and $\omega_{\mathcal{S}_k} > 0$, $\gamma_{\mathcal{P}_k} > 0$ and $\gamma_{\mathcal{N}_k} < 0$. Because $\mathcal{S}_k$, $\mathcal{P}_k$, $\mathcal{N}_k$ containing finite elements in total, if we can go through column deletion process enough times, there is always some natural number $k_0 \geq 2$ satisfying that $\mathcal{S}_{k_0}$ containing only one element j', where we know $\omega_{j'} = 1$ and both $\mathcal{P}_{k_0}$ and $\mathcal{N}_{k_0}$ being empty. Then we can just set $k = k_0$. Or else, the current iterate's column deletion process would be terminated at some $k = k_1$, where $2 \leq k_1 < k_0$. Knowing the definition of (ω_1'', γ_1''), (ω_2, γ_2) and (ω_2', γ_2'), then in a similar way, we can define $(\omega_{l-1}'', \gamma_{l-1}'')$, (ω_l, γ_l) and (ω_l', γ_l') for all natural number l that $2 \leq l \leq k$. Then we know (ω_l', γ_l') has a no worse objective function value than $(\omega_{l-1}', \gamma_{l-1}')$ and (ω_k, γ_k) has the same objective value with (ω_k', γ_k'). Hence, (ω_k, γ_k) has a no worse performance than (ω_2', γ_2'), which is no worse than (ω_0, γ_0). From the Algorithm, $\mathcal{S}_0' = \mathcal{S}_k$, $\mathcal{P}_0' = \mathcal{P}_k$, $\mathcal{N}_0' = \mathcal{N}_k$, and $\omega_{\mathcal{S}_0'} = \omega_{\mathcal{S}_k}$, $\gamma_{\mathcal{P}_0'} = \gamma_{\mathcal{P}_k}$, $\gamma_{\mathcal{N}_0'} = \gamma_{\mathcal{N}_k}$. Therefore, (ω_0', γ_0') has a no worse objective function value than (ω_0, γ_0).

Consider elements in $(\omega_{\mathcal{S}_1}, \gamma_{\mathcal{P}_1}, \gamma_{\mathcal{N}_1}, z_1)$, which is the solution of (3.1) corresponding to index sets $\mathcal{S}_1$, $\mathcal{P}_1$ and $\mathcal{N}_1$.

If we append $j_0 \notin \mathcal{S}_0$ to get $\mathcal{S}_1$ and $\omega_{j_0} \geq 0$. Then from definition of t in Algorithm Step 7, we know in the first iteration of column deletion process, we would have $t > 0$. Because j_0 violates (2.4a) strictly, we have already shown that in such condition, $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1}' + \gamma_1'$. So after the first iteration of column deletion process, we would know (ω_2', γ_2') having a better objective function value than (ω_0, γ_0). So we would always get a better objective function value at the next iterate comparing to the current one.

Similarly, in the condition of appending $i_1 \notin \mathcal{P}_0$ to get $\mathcal{P}_1$ and $\gamma_{i_1} \geq 0$, or appending $i_2 \notin \mathcal{N}_0$ to get $\mathcal{N}_1$ and $\gamma_{i_2} \leq 0$, we would get $t > 0$ and $G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1} + \gamma_1 \neq G_{:,\mathcal{S}_1}\omega_{\mathcal{S}_1}' + \gamma_1'$ in the first iteration of column deletion process as well. So in these two conditions, we will have a better objective value at the next iterate as well. □

The third lemma is to prove that under some appropriate assumptions, through column exchange process, the objective function value would always be improved by iterates of the algorithm.

Lemma 4.3. *Let the current index sets be $\mathcal{S}_0$, $\mathcal{P}_0$ and $\mathcal{N}_0$, which are not optimal. Assume the newly added index does not pass the rank-deficiency check (Step 3) and we go to Step 5 (column* *exchange) in Algorithm. After doing Step 5, we always go through Step 6 and Step 7 (if necessary), then we will get* $(\hat{\omega}_{\mathcal{S}_0'}, \hat{\gamma}_{\mathcal{P}_0'}, \hat{\gamma}_{\mathcal{N}_0'})$ *as a feasible solution of* (3.1) *with a better objective value, where* $\mathcal{S}_0'$*,* $\mathcal{P}_0'$ *and* $\mathcal{N}_0'$ *are the index sets at next iterate.*

Proof. Given index sets $\mathcal{S}_0$, $\mathcal{P}_0$ and $\mathcal{N}_0$, let $(\omega_{\mathcal{S}_0}, \gamma_{\mathcal{P}_0}, \gamma_{\mathcal{N}_0}, z)$ be the solution of (3.1), and we can get (ω, γ) by setting

$$\omega_j = \begin{cases} [\omega_{\mathcal{S}_0}]_j \text{ , if } j \in \mathcal{S}_0; \\ 0 \qquad \text{, otherwise;} \end{cases} \quad \text{and } \gamma_i = \begin{cases} [\gamma_{\mathcal{P}_0 \cup \mathcal{N}_0}]_i \text{ , if } i \in \mathcal{P}_0 \cup \mathcal{N}_0; \\ 0 \qquad \text{, otherwise.} \end{cases} \tag{4.23}$$

Because the newly added index does not pass the rank-deficiency check, then for some newly added index $j_0 \in \mathcal{J} \backslash \mathcal{S}_0$, we would have $(\tilde{\omega}_{\mathcal{S}_0}, \tilde{\gamma}_{\mathcal{P}_0}, \tilde{\gamma}_{\mathcal{N}_0})$ as the solution of (3.2). Similarly, we can define $(\tilde{\omega}, \tilde{\gamma})$ by setting

$$\tilde{\omega}_j = \begin{cases} [\tilde{\omega}_{\mathcal{S}_0}]_j \text{ , if } j \in \mathcal{S}_0; \\ -1 \qquad \text{, if } j = j_0; \\ 0 \qquad \text{, otherwise;} \end{cases} \quad \text{and } \tilde{\gamma}_i = \begin{cases} [\tilde{\gamma}_{\mathcal{P}_0 \cup \mathcal{N}_0}]_i \text{ , if } i \in \mathcal{P}_0 \cup \mathcal{N}_0; \\ 0 \qquad \text{, otherwise.} \end{cases} \tag{4.24}$$

We know $(\tilde{\omega}_{\mathcal{S}_0}, \tilde{\gamma}_{\mathcal{P}_0}, \tilde{\gamma}_{\mathcal{N}_0})$ should satisfy

$$\begin{aligned} &\left(\begin{bmatrix} G_{:,\mathcal{S}_0} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}_0} & W_{:,\mathcal{N}_0} \\ W_{\mathcal{P}_0,:} & W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} \\ W_{\mathcal{N}_0,:} & W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} \end{bmatrix} \begin{bmatrix} G_{:,\mathcal{S}_0} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \mathbb{1}^T & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \tilde{\omega}_{\mathcal{S}_0} \\ \tilde{\gamma}_{\mathcal{P}_0} \\ \tilde{\gamma}_{\mathcal{N}_0} \end{bmatrix} \\ &= \begin{bmatrix} G_{:,\mathcal{S}_0} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}^T \begin{bmatrix} W & W_{:,\mathcal{P}_0} & W_{:,\mathcal{N}_0} \\ W_{\mathcal{P}_0,:} & W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} \\ W_{\mathcal{N}_0,:} & W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} \end{bmatrix} \begin{bmatrix} g_{j_0} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbb{1} \\ 0 \\ 0 \end{bmatrix}, \end{aligned} \tag{4.25}$$

and

$$\mathbb{1}^T \tilde{\omega}_{\mathcal{S}_0} = 1 \text{ and } G_{:,\mathcal{S}_0} \tilde{\omega}_{\mathcal{S}_0} = g_{j_0} \tag{4.26}$$

at the same time.

Assuming that for current iterate, the matrix $\begin{bmatrix} G_{:,\mathcal{S}_0}^T W G_{:,\mathcal{S}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{P}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{N}_0} & \mathbb{1} \\ W_{\mathcal{P},:} G_{:,\mathcal{S}_0} & W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} & 0 \\ W_{\mathcal{N}_0,:} G_{:,\mathcal{S}_0} & W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} & 0 \\ \mathbb{1}^T & 0 & 0 & 0 \end{bmatrix}$ is with the full column rank. After combining (4.25) and (4.26), we would have

$$\begin{bmatrix} G_{:,\mathcal{S}_0}^T W G_{:,\mathcal{S}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{P}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{N}_0} \\ W_{\mathcal{P}_0,:} G_{:,\mathcal{S}_0} & W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} \\ W_{\mathcal{N}_0,:} G_{:,\mathcal{S}_0} & W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} \end{bmatrix} \begin{bmatrix} \tilde{\omega}_{\mathcal{S}_0} \\ \tilde{\gamma}_{\mathcal{P}_0} \\ \tilde{\gamma}_{\mathcal{N}_0} \end{bmatrix} = \begin{bmatrix} G_{:,\mathcal{S}_0}^T W G_{:,\mathcal{S}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{P}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{N}_0} \\ W_{\mathcal{P}_0,:} G_{:,\mathcal{S}_0} & W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} \\ W_{\mathcal{N}_0,:} G_{:,\mathcal{S}_0} & W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} \end{bmatrix} \begin{bmatrix} \tilde{\omega}_{\mathcal{S}_0} \\ 0 \\ 0 \end{bmatrix}, \tag{4.27}$$

which leads to

$$\begin{bmatrix} G_{:,\mathcal{S}_0}^T W_{:,\mathcal{P}_0} & G_{:,\mathcal{S}_0}^T W_{:,\mathcal{N}_0} \\ W_{\mathcal{P}_0,\mathcal{P}_0} & W_{\mathcal{P}_0,\mathcal{N}_0} \\ W_{\mathcal{N}_0,\mathcal{P}_0} & W_{\mathcal{N}_0,\mathcal{N}_0} \end{bmatrix} \begin{bmatrix} \tilde{\gamma}_{\mathcal{P}_0} \\ \tilde{\gamma}_{\mathcal{N}_0} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{4.28}$$

Because of the full column rank assumption above, we would have $\tilde{\gamma}_{\mathcal{P}_0} = \tilde{\gamma}_{\mathcal{N}_0} = 0$.

We have shown in the proof of **Lemma 4.2** that we would always get a no worse objective function value by proceeding Step 6 and Step 7 by turn, so if we can prove through the column exchange process, we would always make an improvement, which means $(\omega - t\tilde{\omega}, \gamma)$ has a better performance than (ω, γ) on the objective function value, where $t = \min\{\omega_k/\tilde{\omega}_k : \tilde{\omega}_k > 0, k \in \mathcal{S}_0\}$, then we can claim to finish the whole proof.

We have known that the γ part does not change through column exchange process, and moreover, because of

$$\begin{bmatrix} G_{:,\mathcal{S}_0} & g_{j_0} \end{bmatrix} \begin{bmatrix} \omega_{\mathcal{S}_0} - t\tilde{\omega}_{\mathcal{S}_0} \\ t \end{bmatrix} = G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} - tG_{:,\mathcal{S}_0}\tilde{\omega}_{\mathcal{S}_0} + tg_j = G_{:,\mathcal{S}_0}\omega_{\mathcal{S}_0} = \begin{bmatrix} G_{:,\mathcal{S}_0} & g_{j_0} \end{bmatrix} \begin{bmatrix} \omega_{\mathcal{S}_0} \\ 0 \end{bmatrix}, \tag{4.29}$$

which means $G(\omega - t\tilde{\omega}) = G\omega$. So to make a comparison of objective function value, we only need to show $b^T(\omega - t\tilde{\omega}) > b^T\omega$, which equally means $b_{\mathcal{S}_0}^T(\omega_{\mathcal{S}_0} - t\tilde{\omega}_{\mathcal{S}_0}) + tb_{j_0} > b_{\mathcal{S}_0}^T\omega_{\mathcal{S}_0}$. From the definition of t, we know $t > 0$, then we only need to show $b_{j_0} > b_{\mathcal{S}_0}^T\tilde{\omega}_{\mathcal{S}_0}$.

Combining (3.1) and (2.4a), we would know that

$$\begin{aligned} b_{\mathcal{S}_0}^T \tilde{\omega}_{\mathcal{S}_0} &= g_{j_0}^T (W G_{:,\mathcal{S}_0} \omega_{\mathcal{S}_0} + W_{:,\mathcal{P}_0} \gamma_{\mathcal{P}_0} + W_{:,\mathcal{N}_0} \gamma_{\mathcal{N}_0}) + z \\ &= g_{j_0}^T (W G_{:,\mathcal{S}_0} \omega_{\mathcal{S}_0} + W_{:,\mathcal{P}_0} \gamma_{\mathcal{P}_0} + W_{:,\mathcal{N}_0} \gamma_{\mathcal{N}_0}) \\ &\quad + \omega_{\mathcal{S}_0}^T (b_{\mathcal{S}_0} - G_{:,\mathcal{S}_0}^T (W G_{:,\mathcal{S}_0} \omega_{\mathcal{S}_0} + W_{:,\mathcal{P}_0} \gamma_{\mathcal{P}_0} + W_{:,\mathcal{N}_0} \gamma_{\mathcal{N}_0})) \\ &< b_{j_0}, \end{aligned} \tag{4.30}$$

and that's exactly what we want to prove. □

Combining lemmas we have already proved above, in the end, we want to show that under some appropriate assumptions, the algorithm would always be terminated and give the optimal solution of (1.2) in finite steps.

Theorem 4.4. *Beginning with any index set $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$, under assumption that whenever going to Step 7, we always have $t > 0$ in the first iteration of column deletion process. Then following the Algorithm, we would always finish in finite steps with some $(\omega_{\mathcal{S}^*}, \gamma_{\mathcal{P}^*}, \gamma_{\mathcal{N}^*})$, where $\mathcal{S}^*$, $\mathcal{P}^*$ and $\mathcal{N}^*$ are optimal sets, and by setting $\omega_{\mathcal{J} \setminus \mathcal{S}^*} \leftarrow 0$ and $\gamma_{\mathcal{I} \setminus (\mathcal{P}^* \cup \mathcal{N}^*)} \leftarrow 0$, we obtains the solution (ω, γ) for* (2.1).

Proof. If $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ are not optimal sets, then we know there must be some $j_0 \notin \mathcal{S}$ violating (2.4a) or some $i_1 \notin \mathcal{P}$ violating (2.4b) or some $i_2 \notin \mathcal{N}$ violating (2.4c). Then by the inclusion of corresponding index, after we go through Step 4, Step 6 and Step 7 or Step 5, Step 6 and Step 7 in Algorithm, from **Lemma 4.2** and **Lemma 4.3**, we could always update index sets $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ to $\mathcal{S}'$, $\mathcal{P}'$ and $\mathcal{N}'$, where the solution of (3.1) corresponding to such sets after update has a better objective function value. Because there are only finite choices of combination of sets on $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$. Knowing that if current sets are not the optimal one, by solving (3.1) corresponding to such index sets go through the Algorithm, we can always have a better objective function value and a better combination of index sets. So we would always stop with the optimal index set choice. Then from **Lemma 4.1**, we know with the optimal index set $\mathcal{S}^*$, $\mathcal{P}^*$ and $\mathcal{N}^*$, one can obtain the solution (ω, γ) for (2.1) by setting $\omega_{\mathcal{J} \setminus \mathcal{S}^*} \leftarrow 0$ and $\gamma_{\mathcal{I} \setminus (\mathcal{P}^* \cup \mathcal{N}^*)} \leftarrow 0$, which means (ω, γ) is also the optimal solution we try to find. □

Chapter 5

Numerical Test and Analysis

We do numerical tests for checking the performance of the solver, to see whether it can solve problems correctly. Moreover, we also keep challenging it to see its limitation.

To generate test problems, we first need to generate following parameters: W as Hessian inverse matrix, G as gradient list of sample points, δ as bounded radius and b as linear term list corresponding to each sample points.

To generate Hessian inverse matrix W, we first generate a random square matrix P_0 in which all elements have a normal distribution with mean of zero and variation of one. We set a positive definite matrix

$$W_0 = P_0^T P_0. \tag{5.1}$$

Then we can formulate a diagonal matrix Λ_0 with eigenvalues of W_0 on its diagonal, and a matrix Q containing all eigenvectors of W_0 corresponding to eigenvalues in diagonal matrix W_0. The Hessian inverse matrix W as

$$W = Q\Lambda_0 Q^T \tag{5.2}$$

would be a positive definite matrix with same eigenvalues as the diagonal matrix Λ_0. In this way, combining with the eigen-decomposition method discussed in Chapter 2, we can generate a positive definite Hessian inverse matrix

$$W = Q\Lambda Q^T \tag{5.3}$$

with any specific condition number τ we want, where $\tau = \frac{\lambda_n}{\lambda_1}$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ for $\lambda_1 \leq \ldots \leq \lambda_n$.

The objective function we consider is

$$\begin{aligned} f(x) &= \frac{1}{2}x^T Hx + p^T x + q(x) \\ &= \frac{1}{2}x^T Hx + p^T x + \max(Ax + c), \end{aligned} \tag{5.4}$$

where H is the quadratic term parameter as a positive definite matrix, p is the linear term parameter, and $q(x)$ is the maximum of affine terms. Corresponding to each point x, we can calculate the gradient at such a point by

$$g(x) = \frac{\partial f(x)}{\partial x}. \tag{5.5}$$

In such way, we can generate $G = \begin{bmatrix} g_1, & \ldots & , g_m \end{bmatrix}$ as gradient list of sample points. We can also change test problems by changing number of affine terms and number of active terms, which would affect value of p and $q(x)$, and then affect value of G as well.

In our case, we set the bounded radius δ as a constant, and we can calculate the linear term b as

$$b_j = f_j + g_j^T (x - x_j) \text{ for all } j \in \{1, \ldots, m\}. \tag{5.6}$$

With W, G, δ, b as input parameters, we can generate test cases as follows:

	#test	# of Variables	# of Affine	# of Active	# of Points
case 1	$n = 0, 1, \ldots, 19$	100	$2(n+1)$	$n+1$	$10(n+1)$
case 2	$n = 0, 1, \ldots, 19$	100	$5(n+1)$	$3(n+1)$	$10(n+1)$
case 3	$n = 0, 1, \ldots, 19$	100	$8(n+1)$	$5(n+1)$	$10(n+1)$

Table 5.1: Parameter default values in different cases

We can see that the difference between each cases are the number of affine terms and the number of active terms. Moreover, in each case, we change the condition number τ to make a test. Test results are listed in the tables on the following pages.

The "KKT error" term in tables are KKT errors calculated by the solver through sets of $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{N}$ in the last iterate of the algorithm. Assuming no numerical errors generated by

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	53	1	0	2.04959e-14
1	pass	221	99001	0	1.76868e-08
2	pass	258	198001	0	2.92542e-07
3	pass	277	297001	0	3.70545e-08
4	pass	419	396001	0	1.80613e-07
5	pass	561	495001	0	5.75808e-08
6	pass	494	594001	0	1.54755e-06
7	pass	554	693001	0	6.9386e-08
8	pass	466	792001	0	4.89713e-07
9	pass	649	891001	0	1.8024e-07
10	pass	650	990001	0	4.96854e-07
11	pass	781	1089001	0	2.41946e-07
12	pass	666	1188001	0	1.57253e-07
13	pass	763	1287001	0	2.60467e-07
14	pass	909	1386001	0	1.48671e-07
15	pass	969	1485001	0	3.46551e-07
16	pass	809	1584001	0	5.34827e-07
17	pass	825	1683001	0	3.05614e-07
18	pass	586	1782001	0	3.53792e-07
19	pass	822	1881001	0	5.28535e-07

Table 5.2: Results of test $\#n$ of W with condition number $(99000n + 1)$ in case 1

solving the linear system (3.1), the solver calculates "KKT error" based on (2.4a), (2.4b) and (2.4c). However, "KKT error (post)" is obtained in another way. After getting the final solution of (ω, γ), we calculate "KKT error (post)" by mathematical formulas listed in (2.1). So actually, "KKT error (post)" takes the numerical errors of (3.1) into a consideration, and that's the reason why there is always a gap between "KKT error" and "KKT error (post)". When the gap is small, it means there is few numerical errors generated when the solver solves (3.1); when the gap is relatively large, (3.1) would be quite difficult to be solved and there is a few numerical errors generated.

From results of those tables, we can see that if a case with more sample points and more ill-condition Hessian inverse matrix, it would take more iterations to be solved. Moreover, the KKT error's gap would get larger for such cases as well. Influences of the number of affine terms and the number of active terms are not really obvious from those tables, no matter in terms of number of iterations or the difference of KKT errors between two calculation methods. Finally, more future work needs to be done to make our solver be suitable to better tolerance parameters.

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	53	1	0	2.50111e-12
1	pass	222	297001	0	2.64037e-08
2	pass	258	594001	0	1.55093e-07
3	pass	279	891001	0	1.64297e-07
4	pass	422	1188001	0	1.77956e-07
5	pass	570	1485001	0	1.35086e-07
6	pass	496	1782001	0	2.77206e-07
7	pass	585	2079001	0	7.72346e-04
8	pass	467	2376001	0	7.63876e-07
9	pass	651	2673001	0	2.48374e-07
10	pass	633	2970001	0	2.48772e-06
11	pass	786	3267001	0	1.49023e-06
12	pass	667	3564001	0	1.10207e-06
13	pass	774	3861001	0	8.64568e-07
14	pass	876	4158001	0	5.40701e-07
15	pass	975	4455001	0	8.48488e-07
16	pass	806	4752001	0	4.18675e-07
17	pass	835	5049001	0	5.08263e-07
18	pass	586	5346001	0	1.36594e-06
19	pass	832	5643001	0	1.00412e-06

Table 5.3: Results of test #n of W with condition number $(297000n + 1)$ in case 1

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	53	1	0	2.50111e-12
1	pass	222	495001	0	4.85776e-08
2	pass	258	990001	0	2.91781e-07
3	pass	278	1485001	0	4.41615e-07
4	pass	423	1980001	0	3.43038e-07
5	pass	574	2475001	0	1.48855e-07
6	pass	496	2970001	0	5.34811e-07
7	pass	588	3465001	0	1.76799e-06
8	pass	467	3960001	0	8.92194e-07
9	pass	650	4455001	0	6.8244e-07
10	pass	634	4950001	0	2.16182e-06
11	pass	786	5445001	0	1.59815e-06
12	pass	667	5940001	0	1.06982e-06
13	pass	773	6435001	0	1.83909e-06
14	pass	877	6930001	0	1.14106e-06
15	pass	963	7425001	0	8.52647e-07
16	pass	806	7920001	0	1.11422e-06
17	pass	842	8415001	0	1.49196e-06
18	pass	586	8910001	0	1.26058e-06
19	pass	832	9405001	0	1.10297e-06

Table 5.4: Results of test #n of W with condition number $(495000n + 1)$ in case 1

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	53	1	0	2.41848e-14
1	pass	222	990001	0	1.41042e-07
2	pass	258	1980001	0	2.00477e-07
3	pass	280	2970001	0	3.85716e-07
4	pass	423	3960001	0	5.41879e-07
5	pass	573	4950001	0	2.90797e-07
6	pass	496	5940001	0	1.58849e-06
7	pass	587	6930001	0	6.31762e-06
8	pass	467	7920001	0	1.90159e-06
9	pass	650	8910001	0	9.15481e-07
10	pass	634	9900001	0	7.08061e-06
11	pass	790	10890001	0	3.94528e-06
12	pass	667	11880001	0	2.77469e-06
13	pass	773	12870001	0	4.7428e-06
14	pass	878	13860001	0	3.25919e-05
15	pass	963	14850001	0	1.72691e-06
16	pass	806	15840001	0	3.61096e-06
17	pass	844	16830001	0	2.65478e-06
18	pass	586	17820001	0	6.0632e-06
19	pass	833	18810001	0	3.42443e-06

Table 5.5: Results of test $\#n$ of W with condition number $(990000n + 1)$ in case 1

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	51	1	0	3.60039e-14
1	pass	219	99001	0	1.18862e-08
2	pass	259	198001	0	5.0445e-08
3	pass	331	297001	0	2.65674e-08
4	pass	410	396001	0	8.73165e-08
5	pass	560	495001	0	1.34147e-07
6	pass	446	594001	0	1.15751e-07
7	pass	486	693001	0	2.54205e-05
8	pass	556	792001	0	1.84809e-07
9	pass	537	891001	0	1.59682e-07
10	pass	668	990001	0	1.21861e-07
11	pass	768	1089001	0	1.27333e-07
12	pass	734	1188001	0	2.99128e-07
13	pass	766	1287001	0	6.55171e-07
14	pass	548	1386001	0	5.82181e-06
15	pass	784	1485001	0	1.45456e-07
16	pass	745	1584001	0	3.69139e-06
17	pass	771	1683001	0	5.86195e-07
18	pass	754	1782001	0	9.50517e-07
19	pass	812	1881001	0	2.47392e-07

Table 5.6: Results of test $\#n$ of W with condition number $(99000n + 1)$ in case 2

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	51	1	0	3.60039e-14
1	pass	220	297001	0	6.76507e-08
2	pass	267	594001	0	1.70726e-07
3	pass	332	891001	0	1.63935e-07
4	pass	410	1188001	0	2.15746e-07
5	pass	560	1485001	0	4.80348e-07
6	pass	449	1782001	0	1.40397e-07
7	pass	486	2079001	0	4.74921e-07
8	pass	557	2376001	0	5.01606e-07
9	pass	537	2673001	0	7.94379e-07
10	pass	671	2970001	0	7.41799e-07
11	pass	770	3267001	0	5.27525e-07
12	pass	735	3564001	0	7.31532e-07
13	pass	762	3861001	0	3.44509e-06
14	pass	548	4158001	0	2.15655e-06
15	pass	786	4455001	0	2.18594e-06
16	pass	745	4752001	0	1.095e-06
17	pass	772	5049001	0	2.01354e-06
18	pass	755	5346001	0	2.31477e-05
19	pass	813	5643001	0	5.84047e-07

Table 5.7: Results of test $\#n$ of W with condition number $(297000n + 1)$ in case 2

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	51	1	0	2.20024e-14
1	pass	220	495001	0	2.36311e-07
2	pass	267	990001	0	8.6524e-03
3	pass	332	1485001	0	3.05221e-07
4	pass	412	1980001	0	4.24367e-07
5	pass	561	2475001	0	4.18487e-05
6	pass	451	2970001	0	1.76134e-07
7	pass	486	3465001	0	2.11785e-04
8	pass	557	3960001	0	7.96874e-07
9	pass	538	4455001	0	1.06502e-06
10	pass	673	4950001	0	1.19335e-06
11	pass	770	5445001	0	1.15323e-06
12	pass	735	5940001	0	5.98129e-06
13	pass	762	6435001	0	3.80306e-06
14	pass	548	6930001	0	1.29294e-05
15	pass	810	7425001	0	1.65679e-06
16	pass	768	7920001	0	2.8662e-06
17	pass	772	8415001	0	1.07066e-04
18	pass	758	8910001	0	4.46094e-04
19	pass	811	9405001	0	9.61563e-05

Table 5.8: Results of test $\#n$ of W with condition number $(495000n + 1)$ in case 2

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	51	1	0	1.45138e-14
1	pass	220	990001	0	1.29038e-07
2	pass	267	1980001	0	2.20795e-02
3	pass	332	2970001	0	1.32781e-06
4	pass	412	3960001	0	7.91218e-07
5	pass	561	4950001	0	9.21089e-07
6	pass	507	5940001	0	7.18365e-07
7	pass	486	6930001	0	7.97932e-06
8	pass	557	7920001	0	1.30911e-06
9	pass	538	8910001	0	1.00622e-06
10	pass	673	9900001	0	2.77478e-06
11	pass	770	10890001	0	2.012e-06
12	pass	737	11880001	0	4.70322e-06
13	pass	762	12870001	0	1.3687e-06
14	pass	547	13860001	0	1.88306e-06
15	pass	793	14850001	0	4.47286e-06
16	pass	768	15840001	0	7.70952e-06
17	pass	772	16830001	0	4.66185e-06
18	pass	758	17820001	0	5.29101e-05
19	pass	812	18810001	0	2.14049e-06

Table 5.9: Results of test #n of W with condition number $(990000n + 1)$ in case 2

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	56	1	0	2.27374e-13
1	pass	207	99001	0	1.03075e-08
2	pass	288	198001	0	6.51058e-08
3	pass	405	297001	0	2.65542e-08
4	pass	399	396001	0	1.44411e-07
5	pass	372	495001	0	1.24284e-07
6	pass	522	594001	0	1.54985e-07
7	pass	422	693001	0	2.86496e-07
8	pass	518	792001	0	1.65336e-02
9	pass	408	891001	0	3.27281-07
10	pass	670	990001	0	1.50739e-06
11	pass	586	1089001	0	2.00527e-06
12	pass	583	1188001	0	2.7069e-07
13	pass	604	1287001	0	1.31464e-06
14	pass	676	1386001	0	1.53966e-07
15	pass	707	1485001	0	1.26413e-06
16	pass	800	1584001	0	6.93675e-07
17	pass	526	1683001	0	2.85057e-07
18	pass	785	1782001	0	1.22702e-06
19	pass	819	1881001	0	3.1993e-07

Table 5.10: Results of test #n of W with condition number $(99000n + 1)$ in case 3

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	56	1	0	2.24546e-14
1	pass	209	297001	0	4.93039e-08
2	pass	289	594001	0	1.09413e-07
3	pass	406	891001	0	5.41615e-08
4	pass	397	1188001	0	5.79391e-07
5	pass	373	1485001	0	1.55687e-07
6	pass	526	1782001	0	6.59426e-07
7	pass	417	2079001	0	9.54925e-07
8	pass	519	2376001	0	2.05968e-02
9	pass	411	2673001	0	9.35564e-07
10	pass	673	2970001	0	1.67016e-06
11	pass	585	3267001	0	1.95598e-06
12	pass	576	3564001	0	9.70291e-07
13	pass	605	3861001	0	1.15089e-05
14	pass	676	4158001	0	1.89135e-06
15	pass	709	4455001	0	9.20507e-07
16	pass	799	4752001	0	1.13305e-06
17	pass	531	5049001	0	7.79409e-06
18	pass	786	5346001	0	3.321e-05
19	pass	819	5643001	0	1.54319e-06

Table 5.11: Results of test #n of W with condition number $(297000n + 1)$ in case 3

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	56	1	0	2.24546e-14
1	pass	209	495001	0	1.87143e-07
2	pass	290	990001	0	1.05544e-07
3	pass	412	1485001	0	1.77369e-07
4	pass	397	1980001	0	7.71802e-07
5	pass	373	2475001	0	9.666e-07
6	pass	526	2970001	0	3.36994e-07
7	pass	417	3465001	0	8.523e-05
8	pass	517	3960001	0	8.00754e-02
9	pass	410	4455001	0	9.65887e-07
10	pass	674	4950001	0	6.61818e-07
11	pass	588	5445001	0	1.07093e-06
12	pass	576	5940001	0	1.60855e-06
13	pass	606	6435001	0	3.60294e-05
14	pass	676	6930001	0	9.30145e-06
15	pass	709	7425001	0	5.79007e-06
16	pass	797	7920001	0	8.0034e-06
17	pass	530	8415001	0	1.61888e-05
18	pass	786	8910001	0	3.07879e-06
19	pass	822	9405001	0	2.31198e-06

Table 5.12: Results of test #n of W with condition number $(495000n + 1)$ in case 3

#test	pass/fail	#iter	τ	KKT error	KKT error (post)
0	pass	56	1	0	2.27374e-13
1	pass	209	990001	0	1.34971e-07
2	pass	290	1980001	0	3.24727e-07
3	pass	412	2970001	0	1.21153e-06
4	pass	398	3960001	0	1.32122e-06
5	pass	373	4950001	0	6.33012e-07
6	pass	525	5940001	0	9.83376e-07
7	pass	417	6930001	0	1.90572e-04
8	pass	521	7920001	0	4.38307e-06
9	pass	410	8910001	0	1.91779e-06
10	pass	674	9900001	0	2.54539e-06
11	pass	588	10890001	0	1.74972e-05
12	pass	576	11880001	0	2.66447e-06
13	pass	606	12870001	0	1.504e-04
14	pass	676	13860001	0	4.85031e-06
15	pass	709	14850001	0	4.47918e-06
16	pass	799	15840001	0	7.52245e-06
17	pass	530	16830001	0	3.92362e-06
18	pass	787	17820001	0	3.95073e-06
19	pass	822	18810001	0	5.2485e-06

Table 5.13: Results of test $\#n$ of W with condition number $(990000n + 1)$ in case 3

Chapter 6

Conclusion

In this thesis, we presented, analyzed, and tested a C++ implementation of a QP solver for solving problems arising within nonsmooth optimization methods. We first introduced theoretical background related to our QP Solver. We also talked about basics and details of the algorithm. Most importantly, we proved some theoretical characteristics of the algorithm, namely, that it converges to the optimal solution of the original problem. In the end, we build up some test case problems and run the QP solver on them.

Our future work would still focus on the QP solver itself. We would mainly focus on testing the solver by more cases and diminishing the gap between KKT error and "KKT error (post)".

Bibliography

[1] K. C. Kiwiel. A Method for Solving Certain Quadratic Programming Problems Arising in Nonsmooth Optimization. *IMA Journal of Numerical Analysis*, 6:137–152, 1986.

[2] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer New York, 2006.

[3] A. Ruszczynski. *Nonlinear Optimization*. Princeton University Press, 2006.

Biography

Baoyu Zhou was born in China on July 29th, 1996. He was named by his parents: Sheng Zhou and Tao Mu. Baoyu Zhou was admitted into Shanghai Jiao Tong University, one of the best universities in China, in September 2012. After the four-year undergraduate study, he got his bachelor degree of Mechanical Engineering in August 2016. Just after his graduation from Shanghai Jiao Tong University, he came to Lehigh University and became a member of Industrial & Systems Engineering Department. Hopefully, Baoyu Zhou would get his Master of Science degree in May 2018. Because of his academic performance during the graduate study, Baoyu Zhou earned Lehigh University Fellowship in 2018. In future, he plans to stay at Lehigh and join Ph.D. program in Industrial & Systems Engineering Department for his further studies in Operations Research area.