On the Moreau-Yosida regularization of the vector k-norm related functions

Bin Wu*, Chao Ding†, Defeng Sun‡ and Kim-Chuan Toh§

This version: March 08, 2011

Abstract

In this paper, we conduct a thorough study on the first and second order properties of the Moreau-Yosida regularization of the vector k-norm function, the indicator function of its epigraph, and the indicator function of the vector k-norm ball. We start with settling the vector k-norm case via applying the existing breakpoint searching algorithms to the metric projector over its dual norm ball. In order to solve the other two cases, we propose algorithms of low computational cost for the metric projectors over four basic polyhedral convex cones. These algorithms are then used to compute the metric projector over the epigraph of the vector k-norm function (or the vector k-norm ball) and its directional derivative. Moreover, we completely characterize the differentiability of the proximal point mappings of the three vector k-norm related functions. The work done in this paper serves as a key step to understand the Moreau-Yosida regularization of the matrix Ky Fan k-norm related functions and thus provides us with fundamental tools to use the proximal point algorithms to solve large scale matrix optimization problems involving the matrix Ky Fan k-norm function.

Key Words: Moreau-Yosida regularization, the vector k-norm function, metric projector
AMS subject classifications: 90C25, 90C30, 65K05, 49J52

1 Introduction

In this paper, we aim to study the Moreau-Yosida regularization of the vector k-norm function, the indicator function of its epigraph, and the indicator function of the vector k-norm ball. Besides its own interest, this research will pave the way for understanding the corresponding Moreau-Yosida regularization of the matrix Ky Fan k-norm related functions. For any $X \in$

*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore (wubin@nus.edu.sg).

†Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore (dingchao@nus.edu.sg).

‡Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore (matsundf@nus.edu.sg).

§Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Republic of Singapore (mattohkc@nus.edu.sg).

$\mathbb{R}^{m\times n}$ (assuming $m \geq n$), the Ky Fan k-norm $\|X\|_{(k)}$ of X is defined as the sum of its k largest singular values, i.e.,

$$\|X\|_{(k)} := \sum_{i=1}^{k} \sigma_i(X)\,,$$

where $1 \leq k \leq n$ and $\sigma_1(X) \geq \sigma_2(X) \geq \ldots \geq \sigma_n(X)$ are the singular values of X arranged in the non-increasing order. The Ky Fan k-norm $\|\cdot\|_{(k)}$ reduces to the spectral or the operator norm $\|\cdot\|_2$ if $k = 1$ and the nuclear norm $\|\cdot\|_*$ if $k = n$, respectively.

The Ky Fan k-norm function appears frequently in matrix optimization problems (MOPs). One such example is to minimize the Ky Fan k-norm function of matrices over a convex set. An early research related to these problems was on the minimization of the sum of the k largest eigenvalues of linearly constrained symmetric matrices in connection with graph partitioning problems [10]. The problem of minimizing the Ky Fan k-norm of a continuously differentiable matrix-valued function was studied for the symmetric case [21, 24] and the nonsymmetric case [31], respectively. In [4, 5], the authors considered the problem of finding the fastest mixing Markov chain (FMMC) on a graph, which can be posed as minimizing the second largest singular value of symmetric and doubly stochastic matrices with a given sparse pattern. Since the largest singular value of any symmetric stochastic matrix is 1, the FMMC problem can be recast as minimizing the Ky Fan 2-norm. MOPs involving the Ky Fan k-norm function also come from recent research on structured low rank matrix approximation [8], which aims at finding a matrix in $\mathbb{R}^{m\times n}$, whose rank is not greater than a given positive integer l, such that it approximates a given target matrix with certain structures. By using the penalty approach to handling the rank constraint proposed in [12, 13], one can see that the Ky Fan k-norm function, where $l \leq k \leq n$, arises naturally in the subproblems of such low rank approximation problems. For the special case that $k = 1$ or $k = n$, one can refer to [11] and references therein for more examples of MOPs with the spectral or nuclear norm function.

One popular approach for solving MOPs with the Ky Fan k-norm function is to reformulate these problems as semidefinite programming (SDP) problems with expanded dimensions (cf. [21, 1]) and apply the well developed interior point methods (IPMs) based SDP solvers, such as SeDuMi [29] and SDPT3 [30]. This approach is fine as long as the sizes of the reformulated problems are not large. For large scale problems, this approach becomes impractical, if possible at all. This is particular the case when $m \gg n$ (or $n \gg m$ if assuming $m \leq n$). Even if $m \approx n$ (e.g., the symmetric case), the expansion of variable dimensions will inevitably lead to extra computational cost. Thus, IPMs do not seem to be viable for MOPs involving the Ky Fan k-norm function and different approaches are needed.

Our idea for solving these problems is built on the classical proximal point algorithms (PPAs) [27, 28]. The reason for doing so is because we have witnessed a lot of interests in applying augmented Lagrangian methods, or in general PPAs, to large scale SDP problems during the last several years, e.g., [25, 20, 33, 34, 32]. The success of the PPAs depends crucially on the closed form solution and the differential properties of the metric projector over the cone of symmetric and positive semi-definite matrices (in short, SDP cone), or equivalently, the solution of the Moreau-Yosida regularization of the indicator function of the SDP cone. It is therefore apparent that in order to make it possible to apply the PPAs to MOPs involving the Ky Fan k-norm function, we need to understand the Moreau-Yosida regularization of the Ky Fan k-

norm related functions. As evidenced in [11], the key step for achieving this is to first study the counterparts of the vector k-norm related functions.

Let $g : \mathcal{X} \to (-\infty, +\infty]$ be a closed proper convex function defined on a finite dimensional real Euclidean space $\mathcal{X}$ equipped with an inner product $\langle \cdot\,,\,\cdot \rangle_{\mathcal{X}}$ and its induced norm $\|\cdot\|_{\mathcal{X}}$. The Moreau-Yosida regularization of g at $x \in \mathcal{X}$ is defined by

$$\min_{y \in \mathcal{X}} \left\{ g(y) + \frac{1}{2}\|y - x\|_{\mathcal{X}}^2 \right\}.$$

From the strong convexity of the objective function in the above problem, we know that for any $x \in \mathcal{X}$, the above problem has a unique solution, which is called the proximal point of x associated with g and denoted by $P_g(x)$, i.e.,

$$P_g(x) := \arg\min_{y \in \mathcal{X}} \left\{ g(y) + \frac{1}{2}\|y - x\|_{\mathcal{X}}^2 \right\}.$$

Define the conjugate of g by

$$g^*(x) := \sup_{y \in \mathcal{X}} \left\{ \langle x\,,\, y \rangle_{\mathcal{X}} - g(y) \right\}, \quad x \in \mathcal{X}.$$

A particular elegant and useful property on the Moreau-Yosida regularization is the following Moreau decomposition (cf. [26, Theorem 31.5]): any $x \in \mathcal{X}$ can be uniquely decomposed into

$$x = P_g(x) + P_{g^*}(x). \tag{1}$$

The vector k-norm function $f_{(k)}(\cdot) \equiv \|\cdot\|_{(k)} : \mathbb{R}^n \to \mathbb{R}$ is defined as the sum of the k largest components in absolute value of any vector in $\mathbb{R}^n$. It is not difficult to see that the vector k-norm function $\|\cdot\|_{(k)}$ is a norm, and it includes the ℓ_∞ norm ($k = 1$) and the ℓ_1 norm ($k = n$). Direct calculation shows that the dual norm of $\|\cdot\|_{(k)}$ (cf. [3, Exercise IV.1.18]) is given by

$$\|z\|_{(k)^*} = \max\left\{ \|z\|_{(1)}, \frac{1}{k}\|z\|_{(n)} \right\} = \max\left\{ \|z\|_\infty, \frac{1}{k}\|z\|_1 \right\}. \tag{2}$$

The epigraph of $f_{(k)}$, denoted by $\operatorname{epi} f_{(k)}$, is given by

$$\operatorname{epi} f_{(k)} := \left\{ (t, z) \in \mathbb{R} \times \mathbb{R}^n \mid t \geq f_{(k)}(z) \right\}.$$

Denote the balls of radius $r > 0$ with respect to the k-norm and its dual norm respectively by

$$\mathcal{B}^r_{(k)} := \{ z \in \mathbb{R}^n \mid \|z\|_{(k)} \leq r \} \quad \text{and} \quad \mathcal{B}^r_{(k)^*} := \{ z \in \mathbb{R}^n \mid \|z\|_\infty \leq r,\ \|z\|_1 \leq kr \}.$$

In this paper, we will take an initial step to study the Moreau-Yosida regularization of $f_{(k)}$, $\delta_{\operatorname{epi} f_{(k)}}$ and $\delta_{\mathcal{B}^r_{(k)}}$, where δ_C is the indicator function of a given set $C \subset \mathbb{R}^n$, i.e., $\delta_C(z) = 0$ if $z \in C$ and $\delta_C(z) = +\infty$ otherwise. From (1) and simple calculations, we can see that the study on the Moreau-Yosida regularization of $f_{(k)}$ is equivalent to studying the Moreau-Yosida regularization of $f^*_{(k)} \equiv \delta_{\mathcal{B}^1_{(k)^*}}$. Thus in this paper, we will study the metric projectors over $\mathcal{B}^r_{(k)^*}$, $\operatorname{epi} f_{(k)}$ and $\mathcal{B}^r_{(k)}$. The metric projector over $\mathcal{B}^r_{(k)^*}$ and its directional derivative are relatively easy to compute, since the metric projector over the intersection of an ℓ_∞-ball and an ℓ_1-ball has been well studied in literatures, e.g., [16, 17, 6, 7, 23, 19, 18]. The metric projectors over $\mathrm{epi}\, f_{(k)}$ (or $\mathcal{B}^r_{(k)}$) and their directional derivatives are much more involved.

The remaining parts of this paper are organized as follows. In Section 2, we give some preliminaries, in particular on the breakpoint searching algorithms and the vector k-norm function. In Section 3, we study the projector over $\mathcal{B}^r_{(k)*}$ including its directional derivative and Fréchet differentiability. Section 4 is devoted to computing the projectors over four basic polyhedral convex cones. This will facilitate our subsequent analysis on the projector over $\mathrm{epi}\, f_{(k)}$ in Section 5. In Section 6, we list some important results on the projector over $\mathcal{B}^r_{(k)}$, which are simpler but parallel to those of the projector over $\mathrm{epi}\, f_{(k)}$. In Section 7, we make our conclusions including several possible extensions of our work done in this paper.

Notation. For any given positive integer n, denote $[n] := \{1, \ldots, n\}$. For any $z \in \mathbb{R}^n$, let $z^\downarrow$ be the vector of components of z being arranged in the non-increasing order $z^\downarrow_1 \geq \cdots \geq z^\downarrow_n$. We use $|z|$ to denote the vector in $\mathbb{R}^n$ whose i-th component is $|z_i|$, $i = 1, \ldots, n$. Let $\mathrm{sgn}(z)$ be the sign vector of z, i.e., $(\mathrm{sgn})_i(z) = 1$ if $z_i \geq 0$ and -1 otherwise. For any index set $\mathcal{I} \subseteq \{1, \ldots, n\}$, we use $|\mathcal{I}|$ to represent the cardinality of $\mathcal{I}$, i.e., the number of elements contained in $\mathcal{I}$. Moreover, we use $z_\mathcal{I} \in \mathbb{R}^{|\mathcal{I}|}$ to denote the sub-vector of z obtained by removing all the components of z not in $\mathcal{I}$. The standard inner product between two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ is defined as $\langle x\, ,\, y \rangle := \sum_{i=1}^n x_i y_i$. For any x and $y \in \mathbb{R}^n$, the notation $x \leq y$ ($x < y$) means that $x_i \leq y_i$ ($x_i < y_i$) for all $i = 1, \ldots, n$, and the notation $x \perp y$ means that $\langle x\, ,\, y \rangle = 0$. The Hardamard product between vectors is denoted by "$\circ$", i.e., for any x and $y \in \mathbb{R}^n$ the i-th component of $w := x \circ y \in \mathbb{R}^n$ is $w_i = x_i y_i$. For any closed convex set $C \subseteq \mathbb{R}^n$, let $\Pi_C(\cdot)$ be the metric projector over C under the standard inner product in $\mathbb{R}^n$. That is, for any $x \in \mathbb{R}^n$, $\Pi_C(x)$ is the unique optimal solution to the following convex optimization problem

$$\min \left\{ \frac{1}{2}\|y - x\|^2 \mid y \in C \right\}.$$

In addition, let e be the vector of all ones, whose dimension should be clear from the context.

2 Preliminaries

In this section, we will first recall the breakpoint searching algorithms for computing the metric projector over a simple polyhedral set consisting one linear equality (or inequality) constraint with simple bounds, and then collect some preliminary results for the vector k-norm functions.

2.1 The breakpoint searching algorithms

In this subsection, we consider the problem of projecting a vector onto the intersection of a hyperplane (or a half space) and a generalized box in $\mathbb{R}^m$.

Assume that $r \in \mathbb{R}$ and $b \in \mathbb{R}^m$ are given. Let l and u be two m-dimensional generalized vectors such that $-\infty \leq l_i \leq u_i \leq +\infty$, $i = 1, \ldots, m$. Denote the polyhedral convex set $\mathcal{S}$ by

$$\mathcal{S} := \{ z \in \mathbb{R}^m \mid \langle b\, ,\, z \rangle = r\, ,\ l \leq z \leq u \}.$$

Then for any given $v \in \mathbb{R}^m$, $\Pi_{\mathcal{S}}(v)$ is the unique optimal solution to the following convex optimization problem of one linear equation with simple bounds

$$\begin{array}{ll} \min & \frac{1}{2}\,\|z - v\|^2 \\ \text{s.t.} & \langle b\,,\, z\rangle = r\,,\;\; l \le z \le u\,. \end{array} \tag{3}$$

When $l, u \in \mathbb{R}^m$, problem (3) is a special case of the continuous quadratic knapsack problem (or the singly linearly constrained quadratic program subject to lower and upper bounds). Specialized algorithms for problem (3), which are based on breakpoint searching (BPS), aim at solving its Karush-Kuhn-Tucker (KKT) system by finding a Lagrange multiplier corresponding to the linear equality constraint. Among these BPS algorithms, the $O(m)$ methods [6, 7, 23, 19, 18] make use of medians of breakpoint subsets, while the $O(m \log m)$ methods [17] sort the breakpoints initially. In particular, for the case that $v = v^{\downarrow}$, b is the vector of all ones, $l = 0$ and $u = +\infty$, we may apply the simple BPS algorithm in [16] to problem (3) to obtain the unique solution $\bar{z}$ by finding the integer $\bar{j}$ such that

$$\bar{j} = \max\left\{\, j \in [m] \mid v_j > \Big(\sum_{i=1}^{j} v_i - r\Big)/j \,\right\},$$

and setting $\bar{z} = \Pi_{\mathbb{R}^m_+}(v - \bar{\lambda}\, b)$ with $\bar{\lambda} = \big(\sum_{i=1}^{\bar{j}} v_i - r\big)/\bar{j}$. This is especially useful for the matrix case where the singular values are usually arranged in the non-increasing order.

BPS algorithms can also be used to solve the convex optimization problem of one linear inequality constraint with simple bounds

$$\begin{array}{ll} \min & \frac{1}{2}\,\|z - v\|^2 \\ \text{s.t.} & \langle b\,,\, z\rangle \le r\,,\;\; l \le z \le u \end{array} \tag{4}$$

as follows: we first check whether the first constraint of problem (4) is satisfied at the metric projection of v onto the generalized box $\{z \in \mathbb{R}^m \mid l \le z \le u\}$; if the first constraint is violated, problem (4) reduces to problem (3), which can be solved by BPS algorithms.

In order to discuss the differentiability of the metric projectors over the polyhedral convex sets $\mathcal{B}^r_{(k)^*}$, epi $f_{(k)}$ and $\mathcal{B}^r_{(k)}$, we need the following proposition which characterizes the directional derivative of the metric projector over a polyhedral convex set [14, 22].

Proposition 2.1 *Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a polyhedral convex set and $\Pi_{\mathcal{C}}(\cdot)$ be the metric projector over $\mathcal{C}$. Assume that $x \in \mathbb{R}^n$ is given. Let $\bar{x} := \Pi_{\mathcal{C}}(x)$. Denote the critical cone of $\mathcal{C}$ at x by $\overline{\mathcal{C}} := \mathcal{T}_{\mathcal{C}}(\bar{x}) \cap (x - \bar{x})^{\perp}$, where $\mathcal{T}_{\mathcal{C}}(\bar{x})$ is the tangent cone of $\mathcal{C}$ at $\bar{x}$. Then for any $h \in \mathbb{R}^n$, the directional derivative of $\Pi_{\mathcal{C}}(\cdot)$ at x along h is given by*

$$\Pi'_{\mathcal{C}}(x; h) = \Pi_{\overline{\mathcal{C}}}(h)\,.$$

2.2 The vector k-norm function

For any given integer k with $1 \le k \le n$, the vector k-norm function $\|\cdot\|_{(k)} : \mathbb{R}^n \to \mathbb{R}$ takes the form of

$$\|z\|_{(k)} = \sum_{i=1}^{k} |z|_i^{\downarrow} . \tag{5}$$

Define the positively homogeneous convex function $s_{(k)}(\cdot) : \mathbb{R}^n \to \mathbb{R}$ by

$$s_{(k)}(z) = \sum_{i=1}^{k} z_i^{\downarrow} . \tag{6}$$

Let the convex sets $\phi_{n,k}$, $\psi_{n,k}$, $\phi_{n,k}^{\le}$ and $\phi_{n,k}^{>}$ of $\mathbb{R}^n$ be defined respectively by

$$\begin{aligned}
\phi_{n,k} &:= \left\{ w \in \mathbb{R}^n \mid 0 \le w \le e,\ \langle e\,,\, w\rangle = k \right\}, \\
\psi_{n,k} &:= \left\{ w \in \mathbb{R}^n \mid w = u - v,\ 0 \le u \le e,\ 0 \le v \le e,\ \langle e\,,\, u+v\rangle = k \right\}, \\
\phi_{n,k}^{\le} &:= \left\{ w \in \mathbb{R}^n \mid 0 \le w \le e,\ \langle e\,,\, w\rangle \le k \right\} \text{ and } \phi_{n,k}^{>} := \left\{ w \in \mathbb{R}^n \mid 0 \le w \le e,\ \langle e\,,\, w\rangle > k \right\}.
\end{aligned}$$

Then, it is not difficult to check that (cf. [21]) for any $z \in \mathbb{R}^n$,

$$s_{(k)}(z) \;=\; \sup \left\{ \langle \mu\,,\, z\rangle \mid \mu \in \phi_{n,k} \right\}, \tag{7}$$

$$\|z\|_{(k)} \;=\; \sup \left\{ \langle \mu\,,\, z\rangle \mid \mu \in \psi_{n,k} \right\}. \tag{8}$$

Lemma 2.1 *$\phi_{n,k}^{\le} \subseteq \psi_{n,k}$ and $\phi_{n,k}^{>} \cap \psi_{n,k} = \emptyset$. Consequently, for any $w \in \mathbb{R}^n_+$, $w \in \psi_{n,k}$ if and only if $w \in \phi_{n,k}^{\le}$.*

Proof. We only need to show that $\phi_{n,k}^{\le} \subseteq \psi_{n,k}$. Suppose that $w \in \phi_{n,k}^{\le}$. If $\sum_{i=1}^n w_i = k$, it is obvious that $w \in \psi_{n,k}$. If $\sum_{i=1}^n w_i < k$, it is easy to see that $\mathcal{J} := \left\{ j \in [n] \mid \sum_{i=1}^{j} 2(1-w_i) \ge k - \sum_{i=1}^n w_i \right\} \ne \emptyset$. Let $j_0 := \min \mathcal{J}$. Define $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ by

$$\left\{ \begin{array}{lll} u_i := 1\,, & v_i := 1 - w_i\,, & i = 1, \ldots, j_0 - 1\,, \\ u_{j_0} := w_{j_0} + \Delta\,, & v_{j_0} := \Delta, & \\ u_i := w_i\,, & v_i := 0\,, & i = j_0 + 1, \ldots, n\,, \end{array} \right.$$

where $\Delta = [k - \sum_{i=1}^n w_i - \sum_{i=1}^{j_0-1} 2(1-w_i)]/2 \in (0\,, 1 - w_{j_0}]$. Then, $w = u - v \in \psi_{n,k}$. $\square$

Suppose that $z \in \mathbb{R}^n$ satisfies $z = z^{\downarrow}$. We may assume that z has the following structure:

$$z_1 \ge \cdots \ge z_{k_0} > z_{k_0+1} = \cdots = z_k = \cdots = z_{k_1} > z_{k_1+1} \ge \cdots \ge z_n\,, \tag{9}$$

where k_0 and k_1 are integers such that $0 \le k_0 < k \le k_1 \le n$ with the conventions that $k_0 = 0$ if $z_1 = z_k$ and that $k_1 = n$ if $z_k = z_n$. Then, the following lemma completely characterizes the subdifferential of $s_{(k)}(\cdot)$ at such z (cf. [21]).

Lemma 2.2 *Suppose that $z \in \mathbb{R}^n$ satisfies $z = z^{\downarrow}$ with structure (9). Then*

$$\partial s_{(k)}(z) = \left\{ \mu \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \mu_i = 1,\ i = 1, \ldots, k_0,\ \mu_i = 0,\ i = k_1 + 1, \ldots, n\,, \\ \text{and } (\mu_{k_0+1}, \ldots, \mu_{k_1}) \in \phi_{k_1 - k_0, k - k_0} \end{array} \right\}.$$

Assume that $z \in \mathbb{R}^n$ satisfying $z = |z|^{\downarrow}$ with the structure:

$$z_1 \geq \cdots \geq z_{k_0} > z_{k_0+1} = \cdots = z_k = \cdots = z_{k_1} > z_{k_1+1} \geq \cdots \geq z_n \geq 0\,, \tag{10}$$

where k_0 and k_1 are integers such that $0 \leq k_0 < k \leq k_1 \leq n$ with the conventions that $k_0 = 0$ if $z_1 = z_k$ and that $k_1 = n$ if $z_k = z_n$. Then, the subdifferential of $\| \cdot \|_{(k)}$ at such z is characterized by the following lemma (cf. [21, 31]).

Lemma 2.3 *Suppose that $z \in \mathbb{R}^n$ satisfies $z = |z|^{\downarrow}$ with structure (10). If $z_k > 0$, then*

$$\partial \|z\|_{(k)} = \left\{ \mu \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \mu_i = 1,\ i = 1, \ldots, k_0,\ \mu_i = 0,\ i = k_1 + 1, \ldots, n\,, \\ \text{and } (\mu_{k_0+1}, \ldots, \mu_{k_1}) \in \phi_{k_1 - k_0, k - k_0} \end{array} \right\}.$$

Otherwise, i.e., if $z_k = 0$, then

$$\partial \|z\|_{(k)} = \left\{ \mu \in \mathbb{R}^n \mid \mu_i = 1,\ i = 1, \ldots, k_0,\ (\mu_{k_0+1}, \ldots, \mu_n) \in \psi_{n-k_0, k-k_0} \right\}.$$

The next three lemmas are useful for simplifying problems in the subsequent sections. The first one is an inequality concerning the rearrangement of two vectors [15, Theorems 368 & 369].

Lemma 2.4 *For $x, y \in \mathbb{R}^n$,*

$$\langle x, y \rangle \leq \langle x^{\downarrow}, y^{\downarrow} \rangle\,,$$

where the inequality holds if and only if there exists a permutation π of $\{1, \ldots, n\}$ such that $x_\pi = x^{\downarrow}$ and $y_\pi = y^{\downarrow}$.

Lemma 2.5 *Suppose that $w \in \phi_{n,k}$ with $w = w^{\downarrow} = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \ldots, n\}$ such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. Then $|\beta_1| \leq k \leq |\beta_1| + |\beta_2|$ and*

$$\left\{ z \in \mathbb{R}^n \mid s_{(k)}(z) \leq \langle w\,,\, z \rangle \right\} = \left\{ z \in \mathbb{R}^n \mid z_{i_1} \geq z_{i_2} = z_{i_2'} \geq z_{i_3},\ \forall\, i_1 \in \beta_1,\ i_2, i_2' \in \beta_2,\ i_3 \in \beta_3 \right\}.$$

Proof. We only need to show that the relation "$\subseteq$" holds. Since for any $z \in \mathbb{R}^n$ satisfying $s_{(k)}(z) \leq \langle w\,,\, z \rangle$, w solves problem (7), we obtain from the KKT conditions for (7) that

$$z_{\beta_1} = \xi_{\beta_1} + \lambda e_{\beta_1}\,, \quad z_{\beta_2} = \lambda e_{\beta_2} \quad \text{and} \quad z_{\beta_3} = -\xi_{\beta_3} + \lambda e_{\beta_3}\,,$$

for some $\xi_{\beta_1} \in \mathbb{R}_+^{|\beta_1|}$, $\xi_{\beta_3} \in \mathbb{R}_+^{|\beta_3|}$ and $\lambda \in \mathbb{R}$. Then the conclusion follows. $\square$

Lemma 2.6 *Suppose that $w \in \phi_{n,k}^{\leq}$ with $w = w^{\downarrow} = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \ldots, n\}$ such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. If $\sum_{i=1}^n w_i = k$, then $|\beta_1| \leq k \leq |\beta_1| + |\beta_2|$ and*

$$\left\{ z \in \mathbb{R}^n \mid \|z\|_{(k)} \leq \langle w\,,\, z \rangle \right\} = \left\{ z \in \mathbb{R}^n \mid z_{i_1} \geq z_{i_2} = z_{i_2'} \geq |z_{i_3}| \geq 0,\ \forall\, i_1 \in \beta_1,\ i_2, i_2' \in \beta_2,\ i_3 \in \beta_3 \right\}.$$

Otherwise, i.e., if $\sum_{i=1}^n w_i < k$, then

$$\left\{ z \in \mathbb{R}^n \mid \|z\|_{(k)} \leq \langle w\,,\, z \rangle \right\} = \left\{ z \in \mathbb{R}^n \mid z_{\beta_1} \geq 0\,,\ z_{\beta_2} = 0\,,\ z_{\beta_3} = 0 \right\}.$$

Proof. We only need to show that the relation "$\subseteq$" holds in both cases. Assume that $z \in \mathbb{R}^n$ satisfies $\|z\|_{(k)} \le \langle w\,,\, z\rangle$. From (8) and Lemma 2.1, we know that w solves the following problem

$$\sup\left\{\, \langle \mu\,,\, z\rangle \mid \mu \in \phi_{n,k}^{\le} \,\right\}. \tag{11}$$

Then the KKT conditions for (11) yield that

$$z_{\beta_1} = \xi_{\beta_1} + \lambda e_{\beta_1}\,, \quad z_{\beta_2} = \lambda e_{\beta_2} \quad \text{and} \quad z_{\beta_3} = -\xi_{\beta_3} + \lambda e_{\beta_3}\,, \tag{12}$$

for some $\xi_{\beta_1} \in \mathbb{R}_+^{|\beta_1|}$, $\xi_{\beta_3} \in \mathbb{R}_+^{|\beta_3|}$ and λ satisfying $0 \le (k - \langle e\,,\, w\rangle) \perp \lambda \in \mathbb{R}_+$. Define $\hat{z} \in \mathbb{R}^n$ by $\hat{z}_{\beta_1\cup\beta_2} := z_{\beta_1\cup\beta_2}$ and $\hat{z}_{\beta_3} := -z_{\beta_3}$. Then $\|\hat{z}\|_{(k)} \le \langle w\,,\, \hat{z}\rangle$. By following the same way to obtain (12), we derive that

$$\hat{z}_{\beta_1} = \hat{\xi}_{\beta_1} + \hat{\lambda} e_{\beta_1}\,, \quad \hat{z}_{\beta_2} = \hat{\lambda} e_{\beta_2} \quad \text{and} \quad \hat{z}_{\beta_3} = -\hat{\xi}_{\beta_3} + \hat{\lambda} e_{\beta_3}\,, \tag{13}$$

for some $\hat{\xi}_{\beta_1} \in \mathbb{R}_+^{|\beta_1|}$, $\hat{\xi}_{\beta_3} \in \mathbb{R}_+^{|\beta_3|}$ and $\hat{\lambda}$ satisfying $0 \le (k - \langle e\,,\, w\rangle) \perp \hat{\lambda} \in \mathbb{R}_+$. Then the conclusions follow from (12) and (13). □

3 Projection over the ball defined by the dual norm $\|\cdot\|_{(k)^*}$

Let k be an given integer with $1 \le k \le n$ and r be a given positive number. In this section, we will study the metric projector over the ball $\mathcal{B}_{(k)^*}^r$ defined by the dual norm with radius r, i.e.,

$$\mathcal{B}_{(k)^*}^r = \left\{\, z \in \mathbb{R}^n \mid \|z\|_{(k)^*} \le r \,\right\} = \left\{\, z \in \mathbb{R}^n \mid \|z\|_\infty \le r\,,\ \|z\|_1 \le kr \,\right\}.$$

Since $f_{(k)}^* \equiv \delta_{\mathcal{B}_{(k)^*}^1}$, we know from (1) that the research on the metric projector over $\mathcal{B}_{(k)^*}^1$ is equivalent to that on the Moreau-Yosida regularization of the k-norm function $f_{(k)}$.

3.1 Computing $\Pi_{\mathcal{B}_{(k)^*}^r}(\cdot)$

For any given $x \in \mathbb{R}^n$, $\Pi_{\mathcal{B}_{(k)^*}^r}(x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\,\|y - x\|^2 \\ \text{s.t.} & \|y\|_\infty \le r\,,\ \|y\|_1 \le kr\,, \end{array} \tag{14}$$

which can be equivalently rewritten as

$$\begin{array}{ll} \min & \dfrac{1}{2}\,\|y - |x|\|^2 \\ \text{s.t.} & 0 \le y \le re\,,\ \langle e\,,\, y\rangle \le kr \end{array} \tag{15}$$

in the sense that $\bar{y} \in \mathbb{R}^n$ solves problem (15) if and only if $\mathrm{sgn}(x) \circ \bar{y}$ solves problem (14). From the discussions in Section 2.1, we know that BPS algorithms can be used to solve problem (15). The computational cost for computing $\Pi_{\mathcal{B}_{(k)^*}^r}(x)$ can be achieved within $O(n)$ arithmetic operations.

3.2 The differentiability of $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$

Assume that $x \in \mathbb{R}^n$ is given. Let π be a permutation of $\{1,\ldots,n\}$ such that $|x|^{\downarrow} = |x|_{\pi}$, i.e., $|x|^{\downarrow}_i = |x|_{\pi(i)}$, $i = 1,\ldots,n$, and π^{-1} be the inverse of π. By using Lemma 2.4, one can equivalently reformulate problem (14) as

$$\begin{array}{ll} \min & \frac{1}{2}\|y - |x|^{\downarrow}\|^2 \\ \text{s.t.} & \|y\|_\infty \le r\,,\ \|y\|_1 \le kr \end{array} \tag{16}$$

in the sense that $\bar{y} \in \mathbb{R}^n$ solves problem (16) (note that $\bar{y} = |\bar{y}|^{\downarrow} \ge 0$ in this case) if and only if $\mathrm{sgn}(x) \circ \bar{y}_{\pi^{-1}}$ solves problem (14). The KKT conditions for (16) are given as follows:

$$\begin{cases} 0 = y - |x|^{\downarrow} + \lambda_1\mu + \lambda_2\nu\,, & \text{for some } \mu \in \partial\|y\|_\infty \text{ and } \nu \in \partial\|y\|_1\,, \\ 0 \le \big(r - \|y\|_\infty\big) \perp \lambda_1 \ge 0 & \text{and} \quad 0 \le \big(kr - \|y\|_1\big) \perp \lambda_2 \ge 0\,, \end{cases} \tag{17}$$

where λ_1 and λ_2 are the corresponding Lagrange multipliers. Note that the constraints of problem (16) can be equivalently replaced by finitely many linear constraints. Then, by using [26, Corollary 28.3.1] and the fact that problem (16) has a unique solution, we know that the KKT system (17) has a solution $(\bar{y}, \bar{\lambda}_1, \bar{\lambda}_2)$ and $\bar{y}$ is the unique optimal solution to problem (16). Let $\bar{\mu} \in \partial\|\bar{y}\|_\infty$ and $\bar{\nu} \in \partial\|\bar{y}\|_1$ such that $(\bar{y}, \bar{\lambda}_1, \bar{\lambda}_2, \bar{\mu}, \bar{\nu})$ satisfies (17).

For convenience, we will use $\mathcal{B}_*$ to denote $\mathcal{B}^r_{(k)^*}$ in the following discussion. Let $\bar{x} := \Pi_{\mathcal{B}_*}(x)$. Then $\bar{y} = |\bar{x}|^{\downarrow}$ and $\bar{x} = \mathrm{sgn}(x) \circ \bar{y}_{\pi^{-1}}$. Denote the critical cone of $\mathcal{B}_*$ at x by $\overline{\mathcal{C}}$, i.e.,

$$\overline{\mathcal{C}} := \mathcal{T}_{\mathcal{B}_*}(\bar{x}) \cap (x - \bar{x})^{\perp}\,,$$

where $\mathcal{T}_{\mathcal{B}_*}(\bar{x})$ is the tangent cone of $\mathcal{B}_*$ at $\bar{x}$. Let $g(z) := \|z\|_{(k)^*}$, $g_\infty(z) := \|z\|_\infty$, and $g_1(z) := \|z\|_1$, $z \in \mathbb{R}^n$. Note that $g(z) = \max\{g_\infty(z), g_1(z)/k\}$. From [9, Theorem 2.4.7], we know that

$$\mathcal{T}_{\mathcal{B}_*}(z) = \{\, d \in \mathbb{R}^n \mid g'(z; d) \le 0 \,\}\,, \tag{18}$$

where for any $d \in \mathbb{R}^n$, $g'(z; d)$ is the directional derivative of g at z along d. Moreover, since g_∞ and g_1 are finite convex functions, from [26, Theorem 23.4], we know that for any $d \in \mathbb{R}^n$,

$$g'_\infty(z; d) = \sup\big\{\langle \mu\,,\, d\rangle \mid \mu \in \partial g_\infty(z)\big\}\,, \tag{19}$$

$$g'_1(z; d) = \sup\big\{\langle \mu\,,\, d\rangle \mid \mu \in \partial g_1(z)\big\}\,. \tag{20}$$

Denote

$$\hat{d} := \mathrm{sgn}(x) \circ d\,, \quad d \in \mathbb{R}^n\,.$$

We characterize the critical cone $\overline{\mathcal{C}}$ of $\mathcal{B}_*$ at x by considering the following four cases:

Case 1: $\|\bar{x}\|_\infty < r$ and $\|\bar{x}\|_1 < kr$. In this case, $\bar{x} = x$ and $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \mathbb{R}^n$.

Case 2: $\|\bar{x}\|_\infty = r$ and $\|\bar{x}\|_1 < kr$. In this case, $\bar{\lambda}_2 = 0$ and $g(\bar{x}) = g_\infty(\bar{x}) = r$. Thus, $g'(\bar{x}; \cdot) = g'_\infty(\bar{x}; \cdot)$. Let

$$\alpha := \big\{ i \in [n] \mid |\bar{x}|_i = r \big\} \quad \text{and} \quad \beta := [n] \backslash \alpha\,.$$

Note that $\alpha \neq \emptyset$. By using Lemma 2.3, (7), (18) and (19), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \{\, d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \,\}. \tag{21}$$

Case 2.1: $\bar{x} = x$. Then $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \{\, d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \,\}$.

Case 2.2: $\bar{x} \neq x$. Then from (17) and Lemma 2.3, we know that $\bar{\lambda}_1 > 0$ and $|x| - |\bar{x}| = \bar{\lambda}_1 \hat{\mu}$, where $\hat{\mu} := \bar{\mu}_{\pi^{-1}} \in \||\bar{x}|\|_\infty$ satisfying $0 \leq \hat{\mu}_\alpha \in \phi_{|\alpha|,1}$ and $\hat{\mu}_\beta = 0$. Since $\|\bar{x}\|_\infty = r$, we obtain that $\bar{\lambda}_1 = \sum_{i\in\alpha} |x|_i - |\alpha| r$. Hence, we can derive $\hat{\mu}$ from $\hat{\mu} = (|x| - |\bar{x}|)/\bar{\lambda}_1$. Then we have

$$(x - \bar{x})^\perp = \big(\mathrm{sgn}(x) \circ (|x|^\downarrow)_{\pi^{-1}} - \mathrm{sgn}(x) \circ (|\bar{x}|^\downarrow)_{\pi^{-1}}\big)^\perp = \big(\mathrm{sgn}(x) \circ \hat{\mu}\big)^\perp,$$

which, together with (21), yields that

$$\overline{\mathcal{C}} = \{\, d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0\,,\ \langle \hat{\mu}_\alpha\,, \hat{d}_\alpha \rangle = 0 \,\} = \{\, d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0\,,\ \hat{d}_{\alpha_2} \leq 0 \,\}, \tag{22}$$

where

$$\alpha_1 := \{i \in \alpha \mid \hat{\mu}_i > 0\} \quad \text{and} \quad \alpha_2 := \alpha \backslash \alpha_1\,.$$

Case 3: $\|\bar{x}\|_\infty < r$ and $\|\bar{x}\|_1 = kr$. In this case, $\bar{\lambda}_1 = 0$, $g(\bar{x}) = g_1(\bar{x})/k = r$. Thus, $g'(\bar{x};\cdot) = g_1'(\bar{x};\cdot)/k$. Let

$$\alpha := \big\{ i \in [n] \mid |\bar{x}|_i > 0 \big\} \quad \text{and} \quad \beta := [n] \backslash \alpha\,.$$

By using Lemma 2.3, (7), (8), (18) and (20), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \begin{cases} \{\, d \in \mathbb{R}^n \mid \langle e\,, \hat{d} \rangle \leq 0 \,\}, & \text{if } \beta = \emptyset\,, \\ \{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,, \hat{d}_\alpha \rangle + \|\hat{d}_\beta\|_1 \leq 0 \,\}, & \text{if } \beta \neq \emptyset\,. \end{cases} \tag{23}$$

Case 3.1: $\bar{x} = x$. Then $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{B}_*}(\bar{x})$, which is given by (23).

Case 3.2: $\bar{x} \neq x$. Then from (17), Lemma 2.3 and Lemma 2.1, we know that $\bar{\lambda}_2 > 0$ and $|x| - |\bar{x}| = \bar{\lambda}_2 \hat{\nu}$, where $\hat{\nu} := \bar{\nu}_{\pi^{-1}} \in \partial \||\bar{x}|\|_1$ satisfying that $\hat{\nu} = e$ if $\beta = \emptyset$, and that $\hat{\nu}_\alpha = e_\alpha$ and $0 \leq \hat{\nu}_\beta \leq e_\beta$ if $\beta \neq \emptyset$. Since $\|\bar{x}\|_1 = kr$, we obtain that $\bar{\lambda}_2 = (\|x\|_1 - kr)/n$ if $\beta = \emptyset$, and $\bar{\lambda}_2 = \sum_{i\in\alpha} (|x|_i - |\bar{x}|_i)/|\alpha|$ if $\beta \neq \emptyset$ (note that $\bar{x} \neq 0$ and thus $\alpha \neq \emptyset$). Hence, we can derive $\hat{\nu}$ from $\hat{\nu} = (|x| - |\bar{x}|)/\bar{\lambda}_2$. Then we have

$$(x - \bar{x})^\perp = \big(\mathrm{sgn}(x) \circ (|x|^\downarrow)_{\pi^{-1}} - \mathrm{sgn}(x) \circ (|\bar{x}|^\downarrow)_{\pi^{-1}}\big)^\perp = \big(\mathrm{sgn}(x) \circ \hat{\nu}\big)^\perp,$$

which, together with (23), yields that

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \langle e\,, \hat{d} \rangle = 0 \,\}, & \text{if } \beta = \emptyset\,, \\ \{\, d \in \mathbb{R}^n \mid \|\hat{d}_\beta\|_1 \leq \langle \hat{\nu}_\beta\,, \hat{d}_\beta \rangle\,,\ \langle e_\alpha\,, \hat{d}_\alpha \rangle + \langle \hat{\nu}_\beta\,, \hat{d}_\beta \rangle = 0 \,\}, & \text{if } \beta \neq \emptyset\,. \end{cases}$$

Let

$$\beta_1 := \{i \in \beta \mid \hat{\nu}_i = 1\} \quad \text{and} \quad \beta_2 := \beta \backslash \beta_1\,.$$

Then by Lemma 2.6, we have

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \langle e\,, \hat{d} \rangle = 0 \,\}, & \text{if } \beta = \emptyset\,, \quad (24) \\ \{\, d \in \mathbb{R}^n \mid \langle e\,, \hat{d} \rangle = 0\,,\ \hat{d}_{\beta_1} \geq 0\,,\ \hat{d}_{\beta_2} = 0 \,\}, & \text{if } \beta \neq \emptyset\,. \quad (25) \end{cases}$$

Case 4: $\|\bar{x}\|_\infty = r$ and $\|\bar{x}\|_1 = kr$. Let

$$\alpha := \big\{ i \mid |\bar{x}|_i = r \big\} , \quad \beta := \big\{ i \mid 0 < |\bar{x}|_i < r \big\} \quad \text{and} \quad \gamma := [n] \backslash (\alpha \cup \beta) .$$

Note that $\alpha \neq \emptyset$. From (17), Lemma 2.3 and Lemma 2.1, we know that $|x| - |\bar{x}| = \bar{\lambda}_1 \hat{\mu} + \bar{\lambda}_2 \hat{\nu}$, where $\hat{\mu} := \bar{\mu}_{\pi^{-1}} \in \partial \| |\bar{x}| \|_\infty$ satisfying that $0 \le \hat{\mu}_\alpha \in \phi_{|\alpha|,1}$ and $\hat{\mu}_{\beta \cup \gamma} = 0$, and $\hat{\nu} := \bar{\nu}_{\pi^{-1}} \in \partial \| |\bar{x}| \|_1$ satisfying that $\hat{\nu} = e$ if $\gamma = \emptyset$, and that $\hat{\nu}_{\alpha \cup \beta} = e_{\alpha \cup \beta}$ and $0 \le \hat{\nu}_\gamma \le e_\gamma$ if $\gamma \neq \emptyset$. Thus we have

$$\begin{cases} r e_\alpha = |x|_\alpha - \bar{\lambda}_1 \hat{\mu}_\alpha - \bar{\lambda}_2 e_\alpha , \quad |\bar{x}|_\beta = |x|_\beta - \bar{\lambda}_2 e_\beta , \quad 0 = |x|_\gamma - \bar{\lambda}_2 \hat{\nu}_\gamma , \\ 0 \le \hat{\mu}_\alpha \le e_\alpha , \ \sum_{i \in \alpha} \hat{\mu}_i = 1 , \quad 0 \le \hat{\nu}_\gamma \le e_\gamma , \quad \bar{\lambda}_1 \ge 0 , \ \bar{\lambda}_2 \ge 0 , \ \bar{\lambda}_1^2 + \bar{\lambda}_2^2 \neq 0 \end{cases} \tag{26}$$

and

$$(x - \bar{x})^\perp = \begin{cases} \{ d \in \mathbb{R}^n \mid \bar{\lambda}_1 \langle \hat{\mu}_\alpha , \hat{d}_\alpha \rangle + \bar{\lambda}_2 \langle e , \hat{d} \rangle = 0 \} , & \text{if } \gamma = \emptyset , \\ \{ d \in \mathbb{R}^n \mid \bar{\lambda}_1 \langle \hat{\mu}_\alpha , \hat{d}_\alpha \rangle + \bar{\lambda}_2 \big(\langle e_{\alpha \cup \beta} , \hat{d}_{\alpha \cup \beta} \rangle + \langle \hat{\nu}_\gamma , \hat{d}_\gamma \rangle \big) = 0 \} , & \text{if } \gamma \neq \emptyset . \end{cases} \tag{27}$$

Denote

$$\alpha_1 := \{ i \in \alpha \mid \hat{\mu}_i > 0 \} , \quad \alpha_2 := \alpha \backslash \alpha_1 , \quad \gamma_1 := \{ i \in \gamma \mid \hat{\nu}_i = 1 \} \quad \text{and} \quad \gamma_2 := \gamma \backslash \gamma_1 .$$

Since in this case $g(\bar{x}) = g_\infty(\bar{x}) = g_1(\bar{x})/k = r$, we know that $g'(\bar{x}; \cdot) = \max\{ g'_\infty(\bar{x}; \cdot), g'_1(\bar{x}; \cdot)/k \}$. Then by using Lemma 2.3, (7), (8), (18), (19) and (20), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \begin{cases} \{ d \in \mathbb{R}^n \mid \hat{d}_\alpha \le 0 , \ \langle e , \hat{d} \rangle \le 0 \} , & \text{if } \gamma = \emptyset , \\ \{ d \in \mathbb{R}^n \mid \hat{d}_\alpha \le 0 , \ \langle e_{\alpha \cup \beta} , \hat{d}_{\alpha \cup \beta} \rangle + \| \hat{d}_\gamma \|_1 \le 0 \} , & \text{if } \gamma \neq \emptyset . \end{cases} \tag{28}$$

If $\beta \neq \emptyset$, we obtain from (26) that $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha|(r + \bar{\lambda}_2)$ and $\bar{\lambda}_2 = \sum_{i \in \beta} (|x|_i - |\bar{x}|_i)/|\beta|$. If $\beta = \emptyset$, from (26) we know that $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha|(r + \bar{\lambda}_2)$. In order to derive $\bar{\lambda}_1$ and $\bar{\lambda}_2$ from $|x|$ and $|\bar{x}|$ for the latter case according to (26), we need to consider the following five cases.

(a) $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha| r$ and $\bar{\lambda}_2 = 0$. For this case, it is sufficient and necessary that $\bar{\lambda}_1 > 0$, $0 \le \hat{\mu}_\alpha \le e_\alpha$ and $|x|_\gamma = 0$, which are equivalent to the conditions that $r < \sum_{i \in \alpha} |x|_i / |\alpha|$, $r \le \min_{i \in \alpha} |x|_i$, $\max_{i \in \alpha} |x|_i \le \sum_{i \in \alpha} |x|_i - (|\alpha| - 1) r$ and $|x|_\gamma = 0$.

(b) $\bar{\lambda}_1 = 0$ and $\bar{\lambda}_2 = \sum_{i \in \alpha} |x|_i / |\alpha| - r$. For this case, it is sufficient and necessary that $\bar{\lambda}_2 > 0$, $|x|_\alpha = (r + \bar{\lambda}_2) e_\alpha$ and $0 \le \hat{\nu}_\gamma \le e_\gamma$, which are equivalent to the conditions that $r < \sum_{i \in \alpha} |x|_i / |\alpha|$, $|x|_j = \sum_{i \in \alpha} |x|_i / |\alpha|$ for $j \in \alpha$, and $\max_{i \in \gamma} |x|_i \le \sum_{i \in \alpha} |x|_i / |\alpha| - r$.

(c) $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha| \min_{i \in \alpha} |x|_i$ and $\bar{\lambda}_2 = \min_{i \in \alpha} |x|_i - r$. For this case, it is sufficient and necessary that $\bar{\lambda}_1 > 0$, $\bar{\lambda}_2 > 0$, $\alpha_2 \neq \emptyset$, $0 \le \hat{\mu}_\alpha \le e_\alpha$ and $0 \le \hat{\nu}_\gamma \le e_\gamma$, which are equivalent to the conditions that $r < \min_{i \in \alpha} |x|_i < \sum_{i \in \alpha} |x|_i / |\alpha|$, $\max_{i \in \alpha} |x|_i \le \sum_{i \in \alpha} |x|_i - (|\alpha| - 1) r$ and $\max_{i \in \gamma} |x|_i \le \min_{i \in \alpha} |x|_i - r$.

(d) $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha| (r + \max_{i \in \gamma} |x|_i)$ and $\bar{\lambda}_2 = \max_{i \in \gamma} |x|_i$. For this case, it is sufficient and necessary that $\bar{\lambda}_1 > 0$, $\bar{\lambda}_2 > 0$, $\gamma_1 \neq \emptyset$, $0 \le \hat{\mu}_\alpha \le e_\alpha$ and $0 \le \hat{\nu}_\gamma \le e_\gamma$, which are equivalent to the conditions that $0 < \max_{i \in \gamma} |x|_i < \sum_{i \in \alpha} |x|_i / |\alpha| - r$, $\gamma \neq \emptyset$, $\max_{i \in \gamma} |x|_i \le \min_{i \in \alpha} |x|_i - r$ and $\max_{i \in \alpha} |x|_i \le \sum_{i \in \alpha} |x|_i - (|\alpha| - 1)(r + \max_{i \in \gamma} |x|_i)$.

(e) $\bar{\lambda}_1 = \sum_{i \in \alpha} |x|_i - |\alpha| (r + \bar{\lambda}_2)$ and $\bar{\lambda}_2 \in (\bar{\lambda}_2^{\min}, \bar{\lambda}_2^{\max}) \neq \emptyset$, where $\bar{\lambda}_2^{\min} := \max\{0, \max_{i \in \gamma} |x|_i\}$ and $\bar{\lambda}_2^{\max} := \min_{i \in \alpha} |x|_i - r$. For this case, it is sufficient and necessary that $\alpha_2 = \gamma_1 = \emptyset$ and $\max\{0, \max_{i \in \gamma} |x|_i\} < \min_{i \in \alpha} |x|_i - r$ (it is not difficult to see that $\bar{\lambda}_1 > 0$ and $\bar{\lambda}_2 > 0$).

Therefore in all the cases, $\bar{\lambda}_1$ and $\bar{\lambda}_2$ can be derived from $|x|$ and $|\bar{x}|$, which implies that $\hat{\mu}$ and $\hat{\nu}$ can be determined by $|x|$ and $|\bar{x}|$. Then we consider the following four subcases.

Case 4.1: $\bar{\lambda}_1 = \bar{\lambda}_2 = 0$ (i.e., $\bar{x} = x$). Then $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{B}_*}(\bar{x})$, which is given by (28).

Case 4.2: $\bar{\lambda}_1 > 0$ and $\bar{\lambda}_2 = 0$. Then from (27), (28) and the fact that $\hat{\mu}_\alpha \geq 0$, we obtain

$$\overline{\mathcal{C}} = \begin{cases} \{ d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0 \,,\ \hat{d}_{\alpha_2} \leq 0 \,,\ \langle e \,, \hat{d} \rangle \leq 0 \} \,, & \text{if } \gamma = \emptyset \,, \quad (29) \\ \{ d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0 \,,\ \hat{d}_{\alpha_2} \leq 0 \,,\ \langle e_{\alpha\cup\beta} \,, \hat{d}_{\alpha\cup\beta} \rangle + \|\hat{d}_\gamma\|_1 \leq 0 \} \,, & \text{if } \gamma \neq \emptyset \,. \quad (30) \end{cases}$$

Case 4.3: $\bar{\lambda}_1 = 0$ and $\bar{\lambda}_2 > 0$. Then from (27), (28), the structure of $\hat{\nu}_\gamma$ and Lemma 2.6, we derive that

$$\overline{\mathcal{C}} = \begin{cases} \{ d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \,,\ \langle e \,, \hat{d} \rangle = 0 \} \,, & \text{if } \gamma = \emptyset \,, \quad (31) \\ \{ d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \,,\ \hat{d}_{\gamma_1} \geq 0 \,,\ \hat{d}_{\gamma_2} = 0 \,,\ \langle e \,, \hat{d} \rangle = 0 \} \,, & \text{if } \gamma \neq \emptyset \,. \quad (32) \end{cases}$$

Case 4.4: $\bar{\lambda}_1 > 0$ and $\bar{\lambda}_2 > 0$. Let $d \in \overline{\mathcal{C}}$ be arbitrarily chosen. Since $\hat{\mu}_\alpha \geq 0$ and $\hat{d}_\alpha \leq 0$, we know that $\langle \hat{\mu}_\alpha \,, \hat{d}_\alpha \rangle \leq 0$. If $\gamma = \emptyset$, by noting that $\langle e \,, \hat{d} \rangle \leq 0$ in (28), we can see that $\bar{\lambda}_1 \langle \hat{\mu}_\alpha \,, \hat{d}_\alpha \rangle + \bar{\lambda}_2 \langle e \,, \hat{d} \rangle = 0$ if and only if $\langle \hat{\mu}_\alpha \,, \hat{d}_\alpha \rangle = 0$ and $\langle e \,, \hat{d} \rangle = 0$. If $\gamma \neq \emptyset$, by using (8) and the structure of $\hat{\nu}_\gamma$, we obtain $\langle e_{\alpha\cup\beta} \,, \hat{d}_{\alpha\cup\beta} \rangle + \langle \hat{\nu}_\gamma \,, \hat{d}_\gamma \rangle \leq 0$ from (28). Hence in this case, $\bar{\lambda}_1 \langle \hat{\mu}_\alpha \,, \hat{d}_\alpha \rangle + \bar{\lambda}_2 \big(\langle e_{\alpha\cup\beta} \,, \hat{d}_{\alpha\cup\beta} \rangle + \langle \hat{\nu}_\gamma \,, \hat{d}_\gamma \rangle \big) = 0$ if and only if $\langle \hat{\mu}_\alpha \,, \hat{d}_\alpha \rangle = 0$ and $\langle e_{\alpha\cup\beta} \,, \hat{d}_{\alpha\cup\beta} \rangle + \langle \hat{\nu}_\gamma \,, \hat{d}_\gamma \rangle = 0$. Then from (27), (28), the fact that $\hat{\mu}_\alpha \geq 0$, the structure of $\hat{\nu}_\gamma$ and Lemma 2.6, we derive that

$$\overline{\mathcal{C}} = \begin{cases} \{ d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0 \,,\ \hat{d}_{\alpha_2} \leq 0 \,,\ \langle e \,, \hat{d} \rangle = 0 \} \,, & \text{if } \gamma = \emptyset \,, \quad (33) \\ \{ d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0 \,,\ \hat{d}_{\alpha_2} \leq 0 \,,\ \hat{d}_{\gamma_1} \geq 0 \,,\ \hat{d}_{\gamma_2} = 0 \,,\ \langle e \,, \hat{d} \rangle = 0 \} \,, & \text{if } \gamma \neq \emptyset \,. \quad (34) \end{cases}$$

From the characterization of the critical cone $\overline{\mathcal{C}}$ of $\mathcal{B}_*$ at x described in the above four cases, we know that BPS algorithms can be used to compute $\Pi_{\overline{\mathcal{C}}}(\cdot)$. Since $\mathcal{B}_*$ is a polyhedral convex set, by Proposition 2.1 introduced in Section 2.1, we know that for any $h \in \mathbb{R}^n$ the directional derivative of $\Pi_{\mathcal{B}_*}(\cdot)$ at x along h is given by

$$\Pi'_{\mathcal{B}_*}(x; h) = \Pi_{\overline{\mathcal{C}}}(h) \,.$$

The complete characterization of the critical cone $\overline{\mathcal{C}}$ and the directional derivative of $\Pi_{\mathcal{B}_*}(\cdot)$ allow us to derive the sufficient and necessary conditions for the Fréchet differentiability of $\Pi_{\mathcal{B}_*}(\cdot)$ in the following theorem. Since its proof can be obtained in a similar way to that of Theorem 5.2 to be given latter, we omit it here.

Theorem 3.1 *The metric projector $\Pi_{\mathcal{B}_*}(\cdot)$ is differentiable at $x \in \mathbb{R}^n$ if and only if x satisfies one of the following eight conditions, where $\bar{x} = \Pi_{\mathcal{B}_*}(x)$:*

(i) $\|\bar{x}\|_\infty < r$ *and* $\|\bar{x}\|_1 < kr$;

(ii) $\|\bar{x}\|_\infty = r$, $\|\bar{x}\|_1 < kr$, $\bar{x} \neq x$ *and* $r < \min_{i\in\alpha} |x|_i$, *where* $\alpha = \big\{ i \in [n] \mid |\bar{x}|_i = r \big\}$;

(iii) $\|\bar{x}\|_\infty < r$, $\|\bar{x}\|_1 = kr$, $\bar{x} \neq x$ *and* $|\bar{x}| > 0$;

(iv) $\|\bar{x}\|_\infty < r$, $\|\bar{x}\|_1 = kr$, $\bar{x} \neq x$, $\min_{1\leq i\leq n} |\bar{x}|_i = 0$ *and* $\max_{i\notin\alpha} |x|_i < \sum_{i\in\alpha}(|x|_i - |\bar{x}|_i)/|\alpha|$, *where* $\alpha = \{i \in [n] \mid |\bar{x}|_i > 0\}$ *(note that* $\alpha \neq \emptyset$ *since* $\bar{x} \neq 0$*);*

(v) $\|\bar{x}\|_\infty = r$, $\|\bar{x}\|_1 = kr$, $\bar{x} \neq x$, $|\bar{x}| > 0$, $\beta \neq \emptyset$ *and* $r + \sum_{i\in\beta}(|x|_i - |\bar{x}|_i)/|\beta| < \min_{i\in\alpha} |x|_i$, *where* $\alpha = \{i \in [n] \mid |\bar{x}|_i = r\}$ *and* $\beta = \{i \in [n] \mid 0 < |\bar{x}|_i < r\}$*;*

(vi) $\bar{x} \neq x$ *and* $\bar{x}_i = r$ *for* $i = 1, \ldots, n$ *(note that this condition holds only when* $k = n$*);*

(vii) $\|\bar{x}\|_\infty = r$, $\|\bar{x}\|_1 = kr$, $\bar{x} \neq x$, $\min_{1\leq i\leq n} |\bar{x}|_i = 0$, $\beta \neq \emptyset$ *and* $\max_{i\notin\alpha\cup\beta} |x|_i < \sum_{i\in\beta}(|x|_i - |\bar{x}|_i)/|\beta| < \min_{i\in\alpha} |x|_i - r$, *where* α *and* β *are the index sets given in* (v)*;*

(viii) $\|\bar{x}\|_\infty = r$, $\|\bar{x}\|_1 = kr$, $\bar{x} \neq x$, $\min_{1\leq i\leq n} |\bar{x}|_i = 0$, $\beta = \emptyset$ *and* $\max_{i\notin\alpha\cup\beta} |x|_i < \min_{i\in\alpha} |x|_i - r$, *where* α *and* β *are the index sets given in* (v)*.*

4 Projections over four basic polyhedral convex cones

In this section, we will mainly focus on the metric projectors over four basic polyhedral convex cones. The results obtained in this section are not only of their own interest, but also are crucial for studying the metric projector over the epigraph of the vector k-norm function.

Let m, p and q be integers such that $1 \leq q \leq p \leq m$. Suppose that $w = w^{\downarrow} \in \phi_{p,q}$. Then we may assume that $w = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \cdots, p\}$ such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. Let $e \in \mathbb{R}^{m-p}$ be the vector of all ones. Denote the polyhedral convex cones $\mathcal{C}_1(m,p,q)$, $\mathcal{C}_2(m,p,q)$, $\mathcal{C}_3(m,p,q,w)$ and $\mathcal{C}_4(m,p,q,w)$ by

$$\mathcal{C}_1(m,p,q) := \{ (\zeta, y, z) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e\, ,\, y\rangle + \|z\|_{(q)} \leq \zeta \}, \tag{35}$$

$$\mathcal{C}_2(m,p,q) := \{ (\zeta, y, z) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e\, ,\, y\rangle + s_{(q)}(z) \leq \zeta \}, \tag{36}$$

$$\mathcal{C}_3(m,p,q,w) := \{ (\zeta, y, z) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \|z\|_{(q)} \leq \langle w\, ,\, z\rangle\, ,\ \langle e\, ,\, y\rangle + \langle w\, ,\, z\rangle = \zeta \}, \tag{37}$$

$$\mathcal{C}_4(m,p,q,w) := \{ (\zeta, y, z) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \mid s_{(q)}(z) \leq \langle w\, ,\, z\rangle\, ,\ \langle e\, ,\, y\rangle + \langle w\, ,\, z\rangle = \zeta \}. \tag{38}$$

In the following discussion, we will drop m, p, q and w from $\mathcal{C}_1(m,p,q)$, $\mathcal{C}_2(m,p,q)$, $\mathcal{C}_3(m,p,q,w)$ and $\mathcal{C}_4(m,p,q,w)$ when their dependence on m, p, q and w can be seen clearly from the context.

In order to propose our algorithms to compute the metric projectors over $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$, we need the following two subroutines. In the algorithms for computing the projectors over $\mathcal{C}_1$ and $\mathcal{C}_3$, `Subroutine 1` and `Subroutine 2` aim to check whether $\bar{z}_q = 0$ and $\bar{z}_q > 0$, respectively, where $\bar{z}_q$ is the q-th component of the z variable of the optimal solutions to problem (42) and problem (63). Moreover, `Subroutine 2` also serves as a main step in the algorithms for computing the projectors over $\mathcal{C}_2$ and $\mathcal{C}_4$.

Subroutine 1 : `function` $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_1(\eta, u, v, \tilde{v}, q_0, \tilde{s}, \texttt{opt})$

- **Input:** $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\tilde{v}$ *is a* $(q+1)$*-tuple,* q_0 *is an integer satisfying* $0 \leq q_0 \leq q-1$, $\tilde{s} \in \mathbb{R}^{p+1}$, $\texttt{opt} = 1$ *or* 0.

- **Main Step:** *Set* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = (\eta, u, v, 0)$. *Compute*

$$\lambda = (\tilde{s}_{q_0} + \langle e\,,\, u\rangle - \eta)/(\|e\|^2 + q_0 + 1)\,.$$

If $\texttt{opt} = 1$, $\lambda > 0$, $\tilde{v}_{q_0} > \lambda \geq \tilde{v}_{q_0+1}$ *and* $\lambda \geq (\tilde{s}_p - \tilde{s}_{q_0})/(q - q_0)$, *set* $\texttt{flag} = 1$*; If* $\texttt{opt} = 0$, $\tilde{v}_{q_0} > \lambda \geq \tilde{v}_{q_0+1}$ *and* $\lambda \geq (\tilde{s}_p - \tilde{s}_{q_0})/(q - q_0)$, *set* $\texttt{flag} = 1$. *If* $\texttt{flag} = 1$, *set* $\bar{\lambda} = \lambda$ *and*

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, & \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z}_i = v_i - \bar{\lambda}\,, & i = 1, \ldots, q_0, \\ \bar{z}_i = 0\,, & i = q_0 + 1, \ldots, p\,. \end{cases}$$

Subroutine 2 : $\texttt{function}\ (\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, v, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, \texttt{opt})$

- **Input:** $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\tilde{v}^-$ *and* $\tilde{v}^+$ *are two tuples of length* $q+1$ *and* $p - q + 2$, q_0 *and* q_1 *are integers satisfying* $0 \leq q_0 < q \leq q_1 \leq p$, $\tilde{s} \in \mathbb{R}^{p+1}$, $\texttt{opt} = 1$ *or* 0.

- **Main Step:** *Set* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = (\eta, u, v, 0)$. *Compute* $\rho = (q_1 - q_0)(\|e\|^2 + q_0 + 1) + (q - q_0)^2$ *and*

$$\begin{cases} \theta = \big((\|e\|^2 + q_0 + 1)(\tilde{s}_{q_1} - \tilde{s}_{q_0}) - (q - q_0)(\tilde{s}_{q_0} + \langle e\,,\, u\rangle - \eta)\big)/\rho\,, \\ \lambda = \big((q - q_0)(\tilde{s}_{q_1} - \tilde{s}_{q_0}) + (q_1 - q_0)(\tilde{s}_{q_0} + \langle e\,,\, u\rangle - \eta)\big)/\rho\,. \end{cases}$$

If $q_0 = 0$ *and* $q_1 = p$, *set* $\texttt{flag} = 1$. *Otherwise, if* $\texttt{opt} = 1$, $\lambda > 0$, $\tilde{v}^-_{q_0} > \theta + \lambda \geq \tilde{v}^-_{q_0+1}$ *and* $\tilde{v}^+_{q_1} \geq \theta > \tilde{v}^+_{q_1+1}$, *set* $\texttt{flag} = 1$*; if* $\texttt{opt} = 0$, $\tilde{v}^-_{q_0} > \theta + \lambda \geq \tilde{v}^-_{q_0+1}$ *and* $\tilde{v}^+_{q_1} \geq \theta > \tilde{v}^+_{q_1+1}$, *set* $\texttt{flag} = 1$. *If* $\texttt{flag} = 1$, *set* $\bar{\theta} = \theta$, $\bar{\lambda} = \lambda$ *and*

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, & \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z}_i = v_i - \bar{\lambda}\,, & i = 1, \ldots, q_0, \\ \bar{z}_i = \bar{\theta}\,, & i = q_0 + 1, \ldots, q_1, \\ \bar{z}_i = v_i\,, & i = q_1 + 1, \ldots, p\,. \end{cases}$$

4.1 Projection over $\mathcal{C}_1$

For any given $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\Pi_{\mathcal{C}_1}(\eta, u, v)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta - \eta)^2 + \|y - u\|^2 + \|z - v\|^2\big) \\ \text{s.t.} & \langle e\,,\, y\rangle + \|z\|_{(q)} \leq \zeta\,. \end{array} \tag{39}$$

Let π_1 be a permutation of $\{1, \ldots, p\}$ such that $|v|^{\downarrow} = |v|_{\pi_1}$, i.e., $|v|^{\downarrow}_i = |v|_{\pi_1(i)}$, $i = 1, \ldots, p$, and ${\pi_1}^{-1}$ be the inverse of π_1. Denote $|v|^{\downarrow}_0 := +\infty$ and $|v|^{\downarrow}_{p+1} := 0$. Define $\tilde{s} \in \mathbb{R}^{p+1}$ by

$$\tilde{s}_0 := 0\,, \quad \tilde{s}_j := \sum_{i=1}^{j} |v|^{\downarrow}_i\,, \quad j = 1, \ldots, p\,. \tag{40}$$

Let $\tilde{v}^-$ and $\tilde{v}^+$ be two tuples of length $q+1$ and $p-q+2$ such that

$$\begin{cases} \tilde{v}_0^- := +\infty\,, & \tilde{v}_i^- := |v|_i^{\downarrow}\,,\ i=1,\ldots,q\,, \\ \tilde{v}_{p+1}^+ := 0\,, & \tilde{v}_i^+ := |v|_i^{\downarrow}\,,\ i=q,\ldots,p\,. \end{cases} \tag{41}$$

According to Lemma 2.4, problem (39) can be equivalently reformulated as

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta-\eta)^2 + \|y-u\|^2 + \|z-|v|^{\downarrow}\|^2\big) \\ \text{s.t.} & \langle e\,,\,y\rangle + \|z\|_{(q)} \leq \zeta \end{array} \tag{42}$$

in the sense that $(\bar{\zeta},\bar{y},\bar{z}) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (42) (note that $\bar{z} = |\bar{z}|^{\downarrow} \geq 0$ in this case) if and only if $(\bar{\zeta},\bar{y},\operatorname{sgn}(v)\circ\bar{z}_{\pi_1^{-1}})$ solves problem (39).

The KKT conditions for (42) have the following form:

$$\begin{cases} 0 = \zeta-\eta-\lambda\,, \quad 0 = y-u+\lambda e\,, \\ 0 = z-|v|^{\downarrow}+\lambda\mu\,, \quad \text{for some } \mu \in \partial\|z\|_{(q)}\,, \\ 0 \leq \big(\zeta - \langle e\,,\,y\rangle - \|z\|_{(q)}\big) \perp \lambda \geq 0\,, \end{cases} \tag{43}$$

where λ is the corresponding Lagrange multiplier. It is not difficult to see that the constraint of problem (42) can be equivalently replaced by finitely many linear constraints. Then from [26, Corollary 28.3.1] and the fact that the optimal solution to problem (42) is unique, we know that the KKT system (43) has a unique solution $(\bar{\zeta},\bar{y},\bar{z},\bar{\lambda})$ and $(\bar{\zeta},\bar{y},\bar{z})$ is also the unique optimal solution to problem (42). If $\eta \geq \langle e\,,\,u\rangle + \|v\|_{(q)}$, then $(\bar{\zeta},\bar{y},\bar{z},\bar{\lambda}) = (\eta,u,|v|^{\downarrow},0)$. Otherwise, i.e., if $\eta < \langle e\,,\,u\rangle + \|v\|_{(q)}$, we have that

$$\bar{\lambda} > 0 \quad \text{and} \quad \bar{\zeta} = \langle e\,,\,\bar{y}\rangle + \|\bar{z}\|_{(q)}\,. \tag{44}$$

Lemma 4.1 *Assume that $(\eta,u,v) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ is given, where $\eta < \langle e\,,\,u\rangle + \|v\|_{(q)}$. Then, $(\bar{\zeta},\bar{y},\bar{z},\bar{\lambda}) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}_+$ solves the KKT system (43) with $\bar{z}_q = 0$ if and only if $(\bar{\zeta},\bar{y},\bar{z},\texttt{flag}) = \mathtt{S}_1(\eta,u,|v|^{\downarrow},\tilde{v}^-,\bar{q}_0,\tilde{s},1)$ with $\texttt{flag} = 1$ for some integer $\bar{q}_0$ satisfying $0 \leq \bar{q}_0 \leq q-1$.*

Proof. By noting that $\bar{z} = |\bar{z}|^{\downarrow}$, we know that $\bar{z}_q = 0$ if and only if there exists an index $\bar{q}_0$ satisfying $0 \leq \bar{q}_0 \leq q-1$ such that

$$\bar{z}_1 \geq \cdots \geq \bar{z}_{\bar{q}_0} > \bar{z}_{\bar{q}_0+1} = \cdots = \bar{z}_q = \cdots = \bar{z}_p = 0 \tag{45}$$

with the convention that $\bar{q}_0 = 0$ if $\bar{z}_1 = \bar{z}_q$. Then according to Lemma 2.3, (44) and (45), the KKT conditions (43) are equivalent to

$$\begin{cases} \bar{\zeta} = \eta+\bar{\lambda}\,, \quad \bar{y} = u-\bar{\lambda}e\,, \\ \bar{z} = |v|^{\downarrow}-\bar{\lambda}\bar{\mu}\,, \quad \bar{z}_{\bar{q}_0} > \bar{z}_{\bar{q}_0+1} = \cdots = \bar{z}_p = 0\,, \\ \bar{\mu}_i = 1\,,\ i=1,\ldots,\bar{q}_0\,, \quad \big(\bar{\mu}_{\bar{q}_0+1},\ldots,\bar{\mu}_p\big) \in \psi_{p-\bar{q}_0,q-\bar{q}_0}\,, \\ \bar{\zeta} = \langle e\,,\,u-\bar{\lambda}e\rangle + \displaystyle\sum_{i=1}^{\bar{q}_0}|v|_i^{\downarrow} - \bar{q}_0\bar{\lambda}\,, \quad \bar{\lambda} > 0\,. \end{cases} \tag{46}$$

By solving (46), we obtain that

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, \quad \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z}_i = |v|^{\downarrow}_i - \bar{\lambda}\,,\ i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = |v|^{\downarrow}_i - \bar{\lambda}\bar{\mu}_i = 0\,,\ \ i = \bar{q}_0 + 1, \ldots, p\,, \\ \bar{\lambda} = \dfrac{1}{\|e\|^2 + \bar{q}_0 + 1}\big(\sum_{i=1}^{\bar{q}_0} |v|^{\downarrow}_i + \langle e\,,\,u\rangle - \eta\big)\,. \end{cases} \tag{47}$$

From Lemma 2.1 and the observation that $\bar{\mu} \geq 0$, we know that

$$\big(\bar{\mu}_{\bar{q}_0+1}, \ldots, \bar{\mu}_p\big) \in \psi_{p-\bar{q}_0, q-\bar{q}_0} \quad \text{if and only if} \quad \big(\bar{\mu}_{\bar{q}_0+1}, \ldots, \bar{\mu}_p\big) \in \phi^{\leq}_{p-\bar{q}_0, q-\bar{q}_0}\,.$$

Then due to the structure of $|v|^{\downarrow}$, (46) is equivalent to

$$\bar{\lambda} > 0\,, \quad |v|^{\downarrow}_{\bar{q}_0} > \bar{\lambda} \geq |v|^{\downarrow}_{\bar{q}_0+1} \quad \text{and} \quad \bar{\lambda} \geq \frac{1}{q - \bar{q}_0} \sum_{i=\bar{q}_0+1}^{p} |v|^{\downarrow}_i \tag{48}$$

with $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda})$ being given by (47). Hence from (47), (48) and the compatibility of the KKT conditions (43), for the case that $\eta < \langle e\,,\,u\rangle + \|v\|_{(q)}$, we can see that $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}_+$ solves the KKT system (43) with $\bar{z}_q = 0$ if and only if $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_1(\eta, u, |v|^{\downarrow}, \tilde{v}^-, \bar{q}_0, \tilde{s}, 1)$ with $\texttt{flag} = 1$ for some integer $\bar{q}_0$ satisfying $0 \leq \bar{q}_0 \leq q - 1$. ☐

Lemma 4.2 *Assume that* $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ *is given, where* $\eta < \langle e\,,\,u\rangle + \|v\|_{(q)}$*. Then,* $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}_+$ *solves the KKT system (43) with* $\bar{z}_q > 0$ *if and only if* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, |v|^{\downarrow}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 1)$ *with* $\texttt{flag} = 1$ *for some integers* $\bar{q}_0$ *and* $\bar{q}_1$ *satisfying* $0 \leq \bar{q}_0 \leq q - 1$ *and* $q \leq \bar{q}_1 \leq p$*.*

Proof. By noting that $\bar{z} = |\bar{z}|^{\downarrow}$, we know that $\bar{z}_q > 0$ if and only if there exist indices $\bar{q}_0$ and $\bar{q}_1$ such that $0 \leq \bar{q}_0 \leq q - 1$, $q \leq \bar{q}_1 \leq p$ and

$$\bar{z}_1 \geq \cdots \geq \bar{z}_{\bar{q}_0} > \bar{z}_{\bar{q}_0+1} = \cdots = \bar{z}_q = \cdots = \bar{z}_{\bar{q}_1} > \bar{z}_{\bar{q}_1+1} \geq \cdots \geq \bar{z}_p \geq 0 \tag{49}$$

with the conventions that $\bar{q}_0 = 0$ if $\bar{z}_1 = \bar{z}_q$ and that $\bar{q}_1 = p$ if $\bar{z}_q = \bar{z}_p$. Denote $\bar{\theta} := \bar{z}_q$. Then by using Lemma 2.3, (44) and (49), we can equivalently rewrite the KKT conditions (43) as

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, \quad \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z} = |v|^{\downarrow} - \bar{\lambda}\bar{\mu}\,, \quad \bar{z}_{\bar{q}_0} > \bar{\theta} > \bar{z}_{\bar{q}_1+1}\,, \quad \bar{z}_i = \bar{\theta},\ i = \bar{q}_0 + 1, \ldots, \bar{q}_1, \\ \bar{\mu}_i = 1\,,\ i = 1, \ldots, \bar{q}_0\,, \quad \big(\bar{\mu}_{\bar{q}_0+1}, \ldots, \bar{\mu}_{\bar{q}_1}\big) \in \phi_{\bar{q}_1 - \bar{q}_0, q - \bar{q}_0}\,, \quad \bar{\mu}_i = 0\,,\ i = \bar{q}_1 + 1, \ldots, p\,, \\ \bar{\zeta} = \langle e\,,\,u - \bar{\lambda} e\rangle + \sum_{i=1}^{\bar{q}_0} |v|^{\downarrow}_i - \bar{q}_0\bar{\lambda} + (q - \bar{q}_0)\,\bar{\theta}\,, \quad \bar{\lambda} > 0\,. \end{cases} \tag{50}$$

By solving (50), we obtain that

$$\begin{cases} \bar{\theta} = \Big((\|e\|^2 + \bar{q}_0 + 1) \sum_{i=\bar{q}_0+1}^{\bar{q}_1} |v|^{\downarrow}_i - (q - \bar{q}_0)\big(\sum_{i=1}^{\bar{q}_0} |v|^{\downarrow}_i + \langle e\,,\,u\rangle - \eta\big)\Big)/\bar{\rho}\,, \\ \bar{\lambda} = \Big((q - \bar{q}_0) \sum_{i=\bar{q}_0+1}^{\bar{q}_1} |v|^{\downarrow}_i + (\bar{q}_1 - \bar{q}_0)\big(\sum_{i=1}^{\bar{q}_0} |v|^{\downarrow}_i + \langle e\,,\,u\rangle - \eta\big)\Big)/\bar{\rho} \end{cases} \tag{51}$$

and

$$
\begin{cases}
\bar{\zeta} = \eta + \bar{\lambda}\,, & \bar{y} = u - \bar{\lambda} e\,, \\
\bar{z}_i = |v|^{\downarrow}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\
\bar{z}_i = |v|^{\downarrow}_i - \bar{\lambda}\bar{\mu}_i = \bar{\theta}\,, & i = \bar{q}_0 + 1, \ldots, \bar{q}_1, \\
\bar{z}_i = |v|^{\downarrow}_i\,, & i = \bar{q}_1 + 1, \ldots, p\,,
\end{cases}
\tag{52}
$$

where $\bar{\rho} = (\bar{q}_1 - \bar{q}_0)(\|e\|^2 + \bar{q}_0 + 1) + (q - \bar{q}_0)^2$. Then, due to the structure of $|v|^{\downarrow}$, (50) is equivalent to

$$
\bar{\lambda} > 0\,, \quad |v|^{\downarrow}_{\bar{q}_0} > \bar{\theta} + \bar{\lambda} \geq |v|^{\downarrow}_{\bar{q}_0+1} \quad \text{and} \quad |v|^{\downarrow}_{\bar{q}_1} \geq \bar{\theta} > |v|^{\downarrow}_{\bar{q}_1+1} \tag{53}
$$

with $\bar{\theta}$ and $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda})$ being given by (51) and (52). Hence from (51), (52), (53) and the compatibility of the KKT conditions (43), for the case that $\eta < \langle e\,,\, u\rangle + \|v\|_{(q)}$, we can see that $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}_+$ solves the KKT system (43) with $\bar{z}_q > 0$ if and only if $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, |v|^{\downarrow}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 1)$ with $\texttt{flag} = 1$ for some integers $\bar{q}_0$ and $\bar{q}_1$ satisfying $0 \leq \bar{q}_0 \leq q - 1$ and $q \leq \bar{q}_1 \leq p$. □

Algorithm 1 : Computing $\Pi_{\mathcal{C}_1}(\eta, u, v)$.

Step 0. (Preprocessing) *If $\langle e\,,\, u\rangle + \|v\|_{(q)} \leq \eta$, output $\Pi_{\mathcal{C}_1}(\eta, u, v) = (\eta, u, v)$ and stop. Otherwise, sort $|v|$ to obtain $|v|^{\downarrow}$, pre-compute $\tilde{s}$ by (40), evaluate $\tilde{v}^-$ and $\tilde{v}^+$ by (41), set $q_0 = q - 1$ and go to* **Step 1**.

Step 1. (Searching for the case that $\bar{z}_q = 0$) *Call* `Subroutine 1` *with* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_1(\eta, u, |v|^{\downarrow}, \tilde{v}^-, q_0, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$, *go to* **Step 3**. *Otherwise, if $q_0 = 0$, set $q_0 = q - 1$ and $q_1 = q$, and go to* **Step 2**; *if $q_0 > 0$, replace q_0 by $q_0 - 1$ and repeat* **Step 1**.

Step 2. (Searching for the case that $\bar{z}_q > 0$) *Call* `Subroutine 2` *with* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, |v|^{\downarrow}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$, *go to* **Step 3**. *Otherwise, if $q_1 < p$, replace q_1 by $q_1 + 1$ and repeat* **Step 2**; *if $q_0 > 0$ and $q_1 = p$, replace q_0 by $q_0 - 1$, set $q_1 = q$, and repeat* **Step 2**.

Step 3. *Output $\Pi_{\mathcal{C}_1}(\eta\,,\, u\,,\, v) = (\bar{\zeta}\,,\, \bar{y}\,,\, \mathrm{sgn}(v) \circ \bar{z}_{\pi_1^{-1}})$ and stop.*

Proposition 4.1 *Assume that $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given. Then the metric projection $\Pi_{\mathcal{C}_1}(\eta, u, v)$ of (η, u, v) onto $\mathcal{C}_1$ can be computed by Algorithm 1. Moreover, the computational cost of Algorithm 1 is $O(p \log p + q(p - q + 1) + m)$, where the sorting cost is $O(p \log p)$ and the searching cost is $O(q(p - q + 1))$.*

Proof. If $\langle e\,,\, u\rangle + \|v\|_{(q)} \leq \eta$, it is easy to see that $\Pi_{\mathcal{C}_1}(\eta, u, v) = (\eta, u, v)$. Otherwise, i.e., if $\langle e\,,\, u\rangle + \|v\|_{(q)} > \eta$, by combining Lemma 4.1 and Lemma 4.2, we know that Algorithm 1 solves the KKT system (43) with the solution $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda})$. Consequently, $(\bar{\zeta}, \bar{y}, \bar{z})$ is the unique optimal solution to problem (42). Thus, we obtain that $\Pi_{\mathcal{C}_1}(\eta\,,\, u\,,\, v) = (\bar{\zeta}\,,\, \bar{y}\,,\, \mathrm{sgn}(v) \circ \bar{z}_{\pi_1^{-1}})$.

Note that `Subroutine 1` and `Subroutine 2` both cost $O(1)$. Since the total number of calls to `Subroutine 1` and `Subroutine 2` is $O(q(p - q + 1))$, we know that searching the solution costs $O(q(p - q + 1))$, which, together with the pre-computing cost of $O(p)$, the initial sorting cost of $O(p \log p)$ and the final evaluation cost of $O(m)$ (note that $(\bar{\zeta}, \bar{y}, \bar{z})$ is evaluated only when $\texttt{flag} = 1$), implies that the computational cost of Algorithm 1 is $O(p \log p + q(p - q + 1) + m)$. □

4.2 Projection over $\mathcal{C}_2$

For any given $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\Pi_{\mathcal{C}_2}(\eta, u, v)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta-\eta)^2 + \|y-u\|^2 + \|z-v\|^2\big) \\ \text{s.t.} & \langle e\,,\,y\rangle + s_{(q)}(z) \le \zeta\,. \end{array} \tag{54}$$

Let π_2 be a permutation of $\{1,\ldots,p\}$ such that $v^{\downarrow} = v_{\pi_2}$, i.e., $v_i^{\downarrow} = v_{\pi_2(i)}$, $i = 1,\ldots,p$, and ${\pi_2}^{-1}$ be the inverse of π_2. Denote $v_0^{\downarrow} := +\infty$ and $v_{p+1}^{\downarrow} := -\infty$. Define $\tilde{s} \in \mathbb{R}^{p+1}$ by

$$\tilde{s}_0 := 0\,, \quad \tilde{s}_j := \sum_{i=1}^{j} v_i^{\downarrow}\,, \quad j = 1,\ldots,p\,. \tag{55}$$

Let $\tilde{v}^-$ and $\tilde{v}^+$ be two tuples of length $q+1$ and $p-q+2$ such that

$$\begin{cases} \tilde{v}_0^- := +\infty\,, & \tilde{v}_i^- := v_i^{\downarrow}\,,\ i = 1,\ldots,q\,, \\ \tilde{v}_{p+1}^+ := -\infty\,, & \tilde{v}_i^+ := v_i^{\downarrow}\,,\ i = q,\ldots,p\,. \end{cases} \tag{56}$$

Then by using Lemma 2.4, one can equivalently reformulate problem (54) as

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta-\eta)^2 + \|y-u\|^2 + \|z-v^{\downarrow}\|^2\big) \\ \text{s.t.} & \langle e\,,\,y\rangle + s_{(q)}(z) \le \zeta \end{array} \tag{57}$$

in the sense that $(\bar{\zeta}, \bar{y}, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ solves problem (57) (note that $\bar{z} = \bar{z}^{\downarrow}$ in this case) if and only if $(\bar{\zeta}, \bar{y}, \bar{z}_{{\pi_2}^{-1}})$ solves problem (54).

The KKT conditions for (57) are given as follows:

$$\begin{cases} 0 = \zeta - \eta - \lambda\,, \quad 0 = y - u + \lambda e\,, \\ 0 = z - |v|^{\downarrow} + \lambda\mu\,, \quad \text{for some } \mu \in \partial s_{(q)}(z)\,, \\ 0 \le \big(\zeta - \langle e\,,\,y\rangle - s_{(q)}(z)\big) \perp \lambda \ge 0\,, \end{cases} \tag{58}$$

where λ is the corresponding Lagrange multiplier. Note that the constraint of problem (57) can be equivalently replaced by finitely many linear constraints. Then by using [26, Corollary 28.3.1] and the fact that problem (57) has a unique solution, we know that the KKT system (58) has a unique solution $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda})$ and $(\bar{\zeta}, \bar{y}, \bar{z})$ is also the unique optimal solution to problem (57). If $\eta \ge \langle e\,,\,u\rangle + s_{(q)}(v)$, then $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda}) = (\eta, u, v^{\downarrow}, 0)$. Otherwise, i.e., if $\eta < \langle e\,,\,u\rangle + s_{(q)}(v)$, we have that

$$\bar{\lambda} > 0 \quad \text{and} \quad \bar{\zeta} = \langle e\,,\,\bar{y}\rangle + s_{(q)}(\bar{z})\,. \tag{59}$$

Lemma 4.3 *Assume that $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given, where $\eta < \langle e\,,\,u\rangle + s_{(q)}(v)$. Then, $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}_+$ solves the KKT system (58) if and only if $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_2(\eta, u, v^{\downarrow}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 1)$ with $\texttt{flag} = 1$ for some integers $\bar{q}_0$ and $\bar{q}_1$ satisfying $0 \le \bar{q}_0 \le q-1$ and $q \le \bar{q}_1 \le p$.*

Proof. By using Lemma 2.2 and (59), we can obtain the proof in a similar way to that of Lemma 4.2. We omit it here. $\square$

Algorithm 2 : Computing $\Pi_{\mathcal{C}_2}(\eta, u, v)$**.**

Step 0. (Preprocessing) *If* $\langle e\,,\, u\rangle + s_{(q)}(v) \leq \eta$*, output* $\Pi_{\mathcal{C}_2}(\eta,u,v) = (\eta,u,v)$ *and stop. Otherwise, sort* v *to obtain* $v^{\downarrow}$*, pre-compute* $\tilde{s}$ *by (55), evaluate* $\tilde{v}^{-}$ *and* $\tilde{v}^{+}$ *by (56), set* $q_0 = q-1$ *and* $q_1 = q$*, and go to* **Step 1**.

Step 1. (Searching) *Call* `Subroutine 2` *with* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, v^{\downarrow}, \tilde{v}^{-}, \tilde{v}^{+}, q_0, q_1, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$*, go to* **Step 2**. *Otherwise, if* $q_1 < p$*, replace* q_1 *by* q_1+1 *and repeat* **Step 1***; if* $q_0 > 0$ *and* $q_1 = p$*, replace* q_0 *by* $q_0 - 1$*, set* $q_1 = q$*, and repeat* **Step 1**.

Step 2. *Output* $\Pi_{\mathcal{C}_2}(\eta\,,\, u\,,\, v) = (\bar{\zeta}\,,\, \bar{y}\,,\, \bar{z}_{\pi_2^{-1}})$ *and stop.*

By using Lemma 4.3, we have the following proposition. Since its proof can be obtained in a similar way to that of Proposition 4.1, we omit it here.

Proposition 4.2 *Assume that* $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^{p}$ *is given. Then the metric projection* $\Pi_{\mathcal{C}_2}(\eta, u, v)$ *of* (η, u, v) *onto* $\mathcal{C}_2$ *can be computed by Algorithm 2. Moreover, the computational cost of Algorithm 2 is* $O(p\log p + q(p-q+1) + m)$*, where the sorting cost is* $O(p\log p)$ *and the searching cost is* $O(q(p-q+1))$*.*

4.3 Projection over $\mathcal{C}_3$

Let $w = w^{\downarrow} \in \phi_{p,q}$ be given. For any given $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^{p}$, $\Pi_{\mathcal{C}_3}(\eta, u, v)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta-\eta)^2 + \|y-u\|^2 + \|z-v\|^2\big) \\ \text{s.t.} & \|z\|_{(q)} \leq \langle w\,,\, z\rangle\,, \\ & \langle e\,,\, y\rangle + \langle w\,,\, z\rangle = \zeta\,. \end{array} \tag{60}$$

Suppose that $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \cdots, p\}$ such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. Let $\text{psgn}(v) \in \mathbb{R}^p$ be the vector such that $\text{psgn}_i(v) = 1$, $i \in \beta_1 \cup \beta_2$, and $\text{psgn}_i(v) = \text{sgn}_i(v)$, $i \in \beta_3$. Let π_3 be a permutation of $\{1, \ldots, p\}$ such that $(v_{\beta_1})^{\downarrow} = (v_{\pi_3})_{\beta_1}$, $v_{\beta_2} = (v_{\pi_3})_{\beta_2}$, $|v_{\beta_3}|^{\downarrow} = |(v_{\pi_3})_{\beta_3}|$, and ${\pi_3}^{-1}$ be the inverse of π_3. Let $\hat{v} := \big(\text{psgn}(v) \circ v\big)_{\pi_3}$, i.e., $\hat{v}_i = \big(\text{psgn}(v) \circ v\big)_{\pi_3(i)}$, $i = 1, \ldots, p$. Denote $\hat{v}_0 := +\infty$ and $\hat{v}_{p+1} := 0$. Define $\tilde{s} \in \mathbb{R}^{p+1}$ by

$$\tilde{s}_0 := 0\,, \quad \tilde{s}_j := \sum_{i=1}^{j} \hat{v}_i\,, \quad j = 1, \ldots, p\,. \tag{61}$$

Let $\tilde{v}^{-}$ and $\tilde{v}^{+}$ be two tuples of length $q+1$ and $p-q+2$ such that

$$\begin{cases} \tilde{v}_0^{-} := +\infty\,, \ \ \tilde{v}_{|\beta_1|+1}^{-} = \tilde{v}_q^{-} := -\infty\,, \quad \tilde{v}_i^{-} := \hat{v}_i\,,\ i = 1, \ldots, q\,,\ i \neq |\beta_1|+1 \text{ and } q\,, \\ \tilde{v}_q^{+} = \tilde{v}_{|\beta_1|+|\beta_2|}^{+} := +\infty\,, \ \ \tilde{v}_{p+1}^{+} := 0\,, \quad \tilde{v}_i^{+} := \hat{v}_i\,,\ i = q, \ldots, p\,,\ i \neq q \text{ and } |\beta_1|+|\beta_2|\,. \end{cases} \tag{62}$$

Then by using Lemma 2.4, Lemma 2.6 and the assumption that $w = w^{\downarrow} \in \phi_{p,q}$, one can equivalently reformulate problem (60) as

$$
\begin{array}{ll}
\min & \frac{1}{2}\big((\zeta-\eta)^2 + \|y-u\|^2 + \|z-\hat{v}\|^2\big) \\
\text{s.t.} & z_i \geq z_q\,,\ i = 1,\ldots,|\beta_1|\,, \\
& z_i = z_q\,,\ i = |\beta_1|+1,\ldots,|\beta_1|+|\beta_2|\,,\ i \neq q\,, \\
& z_i \leq z_q\,,\ i = |\beta_1|+|\beta_2|+1,\ldots,p\,, \\
& z_p \geq 0\,,\ \langle e\,,\,y\rangle + \langle w\,,\,z\rangle = \zeta
\end{array}
\tag{63}
$$

in the sense that $(\bar{\zeta},\bar{y},\bar{z}) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (63) (note that $\bar{z} = |\bar{z}|^{\downarrow}$ in this case) if and only if $(\bar{\zeta},\bar{y},\mathrm{psgn}(v)\circ\bar{z}_{\pi_3^{-1}})$ solves problem (60). Note that $|\beta_1| \leq q \leq |\beta_1|+|\beta_2|$, and $q = |\beta_1|+|\beta_2|$ if and only if $q = |\beta_1|$, i.e., $\beta_2 = \emptyset$.

The KKT conditions for (63) have the following form:

$$
\begin{cases}
0 = \zeta - \eta - \lambda\,, \quad 0 = y - u + \lambda e\,, \\
0 = z - \hat{v} - \sum_{i=1}^{q-1}\xi_i(e^i - e^q) - \sum_{i=q+1}^{p}\xi_i(e^q - e^i) - \xi_0 e^p + \lambda\, w\,, \\
0 \leq (z_i - z_q) \perp \xi_i \geq 0\,,\ i = 1,\ldots,|\beta_1|\,, \\
z_i = z_q\,,\ i = |\beta_1|+1,\ldots,|\beta_1|+|\beta_2|\,,\ i \neq q\,, \\
0 \leq (z_q - z_i) \perp \xi_i \geq 0\,,\ i = |\beta_1|+|\beta_2|+1,\ldots,p\,, \\
0 \leq z_p \perp \xi_0 \geq 0\,, \quad \langle e\,,\,y\rangle + \langle w\,,\,z\rangle = \zeta\,,
\end{cases}
\tag{64}
$$

where $\lambda \in \mathbb{R}$ and $\xi = (\xi_0,\xi_1,\ldots,\xi_{q-1},\xi_{q+1},\ldots,\xi_p) \in \mathbb{R}^p$ are the corresponding Lagrange multipliers, and $e^i \in \mathbb{R}^p$, $i = 1,\ldots,p$, is the i-th standard basis whose components are all 0 except its i-th component being 1. Since problem (63) has only finitely many linear constraints, then by using [26, Corollary 28.3.1] and the fact that the optimal solution to problem (63) is unique, we know that the KKT system (64) has a unique solution $(\bar{\zeta},\bar{y},\bar{z},\bar{\xi},\bar{\lambda})$ and $(\bar{\zeta},\bar{y},\bar{z})$ is the unique optimal solution to problem (63).

Lemma 4.4 *Assume that* $(\eta,u,v) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given. Then,* $(\bar{\zeta},\bar{y},\bar{z},\bar{\xi},\bar{\lambda}) \in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}$ *solves the KKT system (64) with* $\bar{z}_q = 0$ *if and only if* $(\bar{\zeta},\bar{y},\bar{z},\mathtt{flag}) = \mathtt{S}_1(\eta,u,\hat{v},\tilde{v}^-,\bar{q}_0,\tilde{s},0)$ *with* $\mathtt{flag} = 1$ *for some integer* $\bar{q}_0$ *satisfying* $0 \leq \bar{q}_0 \leq \min\{q-1,|\beta_1|\}$.

Proof. By noting that $\bar{z} = |\bar{z}|^{\downarrow}$, we know that $\bar{z}_q = 0$ if and only if there exists an index $\bar{q}_0$ satisfying $0 \leq \bar{q}_0 \leq q-1$ such that

$$
\bar{z}_1 \geq \cdots \geq \bar{z}_{\bar{q}_0} > \bar{z}_{\bar{q}_0+1} = \cdots = \bar{z}_q = \cdots = \bar{z}_p = 0
\tag{65}
$$

with the convention that $\bar{q}_0 = 0$ if $\bar{z}_1 = \bar{z}_q$. From (64) we know that $\bar{q}_0 \leq |\beta_1|$, which implies that $0 \leq \bar{q}_0 \leq \min\{q-1,|\beta_1|\}$. According to (65), the KKT conditions (64) are equivalent to

$$
\begin{cases}
\bar{\zeta} = \eta + \bar{\lambda}\,, \quad \bar{y} = u - \bar{\lambda}e\,, \\
\bar{z} = \hat{v} + \sum_{i=\bar{q}_0+1}^{q-1}\bar{\xi}_i(e^i - e^q) + \sum_{i=q+1}^{p}\bar{\xi}_i(e^q - e^i) + \bar{\xi}_0 e^p - \bar{\lambda}w\,, \\
\bar{z}_{\bar{q}_0} > 0 = \bar{z}_i\,,\ i = \bar{q}_0+1,\ldots,p\,, \quad \langle e\,,\,\bar{y}\rangle + \langle w\,,\,\bar{z}\rangle = \bar{\zeta}\,, \\
\bar{\xi}_i = 0\,,\ i = 1,\ldots,\bar{q}_0, \quad \bar{\xi}_i \geq 0\,,\ i = 0, \bar{q}_0+1,\ldots,|\beta_1|, |\beta_1|+|\beta_2|+1,\ldots,p\,.
\end{cases}
\tag{66}
$$

Note that $w = w^{\downarrow} = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. If $\beta_2 \neq \emptyset$, then $0 \leq \bar{q}_0 \leq |\beta_1| < q < |\beta_1| + |\beta_2| \leq p$ and $\bar{z}$ is given by

$$\begin{cases} \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} = 0\,, & i = \bar{q}_0 + 1, \ldots, |\beta_1|\,, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} w_i = 0\,, & i = |\beta_1| + 1, \ldots, q-1\,, \\ \bar{z}_q = \hat{v}_q - \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i + \sum_{i=q+1}^{p} \bar{\xi}_i - \bar{\lambda} w_q = 0\,, & \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i - \bar{\lambda} w_i = 0\,, & i = q+1, \ldots, |\beta_1| + |\beta_2|\,, \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i = 0\,, & i = |\beta_1| + |\beta_2| + 1, \ldots, p-1\,, \\ \bar{z}_p = \begin{cases} \hat{v}_p - \bar{\xi}_p - \bar{\lambda} w_p + \bar{\xi}_0 = 0\,, & \text{if } p = |\beta_1| + |\beta_2|\,, \\ \hat{v}_p - \bar{\xi}_p + \bar{\xi}_0 = 0\,, & \text{if } p > |\beta_1| + |\beta_2|\,. \end{cases} & \end{cases}$$

Otherwise, i.e., if $\beta_2 = \emptyset$, then $0 \leq \bar{q}_0 < |\beta_1| = q = |\beta_1| + |\beta_2| \leq p$ and $\bar{z}$ is given by

$$\begin{cases} \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} = 0\,, & i = \bar{q}_0 + 1, \ldots, q-1\,, \\ \bar{z}_q = \hat{v}_q - \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i + \sum_{i=q+1}^{p} \bar{\xi}_i - \bar{\lambda} = 0\,, & \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i = 0\,, & i = q+1, \ldots, p-1\,, \\ \bar{z}_p = \begin{cases} \hat{v}_q - \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i - \bar{\lambda} + \bar{\xi}_0 = 0\,, & \text{if } p = q\,, \\ \hat{v}_p - \bar{\xi}_p + \bar{\xi}_0 = 0\,, & \text{if } p > q\,. \end{cases} & \end{cases}$$

Since $\sum_{i=\bar{q}_0+1}^{p} w_i = q - \bar{q}_0$, by solving (66) we obtain that

$$\bar{\lambda} = \frac{1}{\|e\|^2 + \bar{q}_0 + 1} \big(\sum_{i=1}^{\bar{q}_0} \hat{v}_i + \langle e\,,\, u \rangle - \eta \big)\,, \quad \bar{\xi}_0 = (q - \bar{q}_0)\bar{\lambda} - \sum_{i=\bar{q}_0+1}^{p} \hat{v}_i \tag{67}$$

and

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, & \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = 0\,, & i = \bar{q}_0 + 1, \ldots, p\,. \end{cases} \tag{68}$$

Then due to the structure of $\hat{v}$, (66) is equivalent to

$$\begin{cases} \hat{v}_{\bar{q}_0} > \bar{\lambda} \geq \hat{v}_{\bar{q}_0+1} \ \text{ and } \ \bar{\lambda} \geq \sum_{i=\bar{q}_0+1}^{p} \hat{v}_i/(q - \bar{q}_0)\,, & \begin{cases} \text{if } \beta_2 \neq \emptyset\,,\ \bar{q}_0 < |\beta_1|\,, \\ \text{if } \beta_2 = \emptyset\,,\ \bar{q}_0 < q-1\,, \end{cases} \\ \hat{v}_{\bar{q}_0} > \bar{\lambda} \geq \sum_{i=\bar{q}_0+1}^{p} \hat{v}_i/(q - \bar{q}_0)\,, & \begin{cases} \text{if } \beta_2 \neq \emptyset\,,\ \bar{q}_0 = |\beta_1|\,, \\ \text{if } \beta_2 = \emptyset\,,\ \bar{q}_0 = q-1 \end{cases} \end{cases} \tag{69}$$

with $(\bar{\zeta}, \bar{y}, \bar{z})$ and $\bar{\lambda}$ being given by (67) and (68) (note that $\bar{\xi}$ can be obtained from $\bar{z}$ and $\bar{\lambda}$). Hence from (67), (68), (69) and the compatibility of the KKT conditions (64), we can see that $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\xi}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}$ solves the KKT system (64) with $\bar{z}_q = 0$ if and only if $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_1(\eta, u, \hat{v}, \tilde{v}^-, \bar{q}_0, \tilde{s}, 0)$ with $\texttt{flag} = 1$ for some integer $\bar{q}_0$ satisfying $0 \leq \bar{q}_0 \leq \min\{q-1, |\beta_1|\}$. □

Lemma 4.5 *Assume that* $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ *is given. Then,* $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\xi}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}$ *solves the KKT system (64) with* $\bar{z}_q > 0$ *if and only if* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, \hat{v}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 0)$ *with* $\texttt{flag} = 1$ *for some integers* $\bar{q}_0$ *and* $\bar{q}_1$ *satisfying* $0 \le \bar{q}_0 \le \min\{q-1, |\beta_1|\}$ *and* $\max\{q, |\beta_1| + |\beta_2|\} \le \bar{q}_1 \le p$.

Proof. By noting that $\bar{z} = |\bar{z}|^{\downarrow}$, we know that $\bar{z}_q > 0$ if and only if there exist indices $\bar{q}_0$ and $\bar{q}_1$ such that $0 \le \bar{q}_0 \le q-1$, $q \le \bar{q}_1 \le p$ and

$$\bar{z}_1 \ge \cdots \ge \bar{z}_{\bar{q}_0} > \bar{z}_{\bar{q}_0+1} = \cdots = \bar{z}_q = \cdots = \bar{z}_{\bar{q}_1} > \bar{z}_{\bar{q}_1+1} \ge \cdots \ge \bar{z}_p \ge 0 \tag{70}$$

with the conventions that $\bar{q}_0 = 0$ if $\bar{z}_1 = \bar{z}_q$ and that $\bar{q}_1 = p$ if $\bar{z}_q = \bar{z}_p$. From (64) we know that $\bar{q}_0 \le |\beta_1|$ and $\bar{q}_1 \ge |\beta_1| + |\beta_2|$, which imply that $0 \le \bar{q}_0 \le \min\{q-1, |\beta_1|\}$ and $\max\{q, |\beta_1| + |\beta_2|\} \le \bar{q}_1 \le p$. Denote $\bar{\theta} := \bar{z}_q$. By using (70), we can equivalently rewrite the KKT conditions (64) as

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}, \quad \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z} = \hat{v} + \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i (e^i - e^q) + \sum_{i=q+1}^{\bar{q}_1} \bar{\xi}_i (e^q - e^i) + \bar{\xi}_0 e^p - \bar{\lambda} w\,, \\ \bar{z}_{\bar{q}_0} > \bar{\theta} > \bar{z}_{\bar{q}_1+1}\,, \quad \bar{z}_i = \bar{\theta}\,,\ i = \bar{q}_0+1, \ldots, \bar{q}_1\,, \\ \bar{\xi}_i = 0\,,\ i = 1, \ldots, \bar{q}_0, \bar{q}_1+1, \ldots, p\,, \\ \bar{\xi}_i \ge 0\,,\ i = \bar{q}_0+1, \ldots, |\beta_1|, |\beta_1|+|\beta_2|+1, \ldots, \bar{q}_1\,, \\ 0 \le \bar{z}_p \perp \bar{\xi}_0 \ge 0\,, \quad \langle e\,, \bar{y}\rangle + \langle w\,, \bar{z}\rangle = \bar{\zeta}\,. \end{cases} \tag{71}$$

Note that $w = w^{\downarrow} = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. If $\beta_2 \ne \emptyset$, then $0 \le \bar{q}_0 \le |\beta_1| < q < |\beta_1| + |\beta_2| \le \bar{q}_1 \le p$ and $\bar{z}$ is given by

$$\begin{cases} \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} = \bar{\theta}\,, & i = \bar{q}_0+1, \ldots, |\beta_1|\,, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} w_i = \bar{\theta}\,, & i = |\beta_1|+1, \ldots, q-1\,, \\ \bar{z}_q = \hat{v}_q - \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i + \sum_{i=q+1}^{\bar{q}_1} \bar{\xi}_i - \bar{\lambda} w_q = \bar{\theta}\,, & \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i - \bar{\lambda} w_i = \bar{\theta}\,, & i = q+1, \ldots, |\beta_1|+|\beta_2|\,, \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i = \bar{\theta}\,, & i = |\beta_1|+|\beta_2|+1, \ldots, \bar{q}_1, \\ \bar{z}_i = \hat{v}_i\,, & i = \bar{q}_1+1, \ldots, p-1\,, \\ \bar{z}_p = \begin{cases} \bar{\theta}\,, & \text{if } p = \bar{q}_1\,, \\ \hat{v}_p + \bar{\xi}_0\,, & \text{if } p > \bar{q}_1\,. \end{cases} & \end{cases}$$

Otherwise, i.e., if $\beta_2 = \emptyset$, then $0 \le \bar{q}_0 < |\beta_1| = q = |\beta_1| + |\beta_2| \le \bar{q}_1 \le p$ and $\bar{z}$ is given by

$$\begin{cases} \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = \hat{v}_i + \bar{\xi}_i - \bar{\lambda} = \bar{\theta}\,, & i = \bar{q}_0+1, \ldots, q-1\,, \\ \bar{z}_q = \hat{v}_q - \sum_{i=\bar{q}_0+1}^{q-1} \bar{\xi}_i + \sum_{i=q+1}^{\bar{q}_1} \bar{\xi}_i - \bar{\lambda} = \bar{\theta}\,, & \\ \bar{z}_i = \hat{v}_i - \bar{\xi}_i = \bar{\theta}\,, & i = q+1, \ldots, \bar{q}_1, \\ \bar{z}_i = \hat{v}_i\,, & i = \bar{q}_1+1, \ldots, p-1\,, \\ \bar{z}_p = \begin{cases} \bar{\theta}\,, & \text{if } p = \bar{q}_1\,, \\ \hat{v}_p + \bar{\xi}_0\,, & \text{if } p > \bar{q}_1\,. \end{cases} & \end{cases}$$

If $\bar{q}_1 = p$, then $\bar{z}_p = \bar{\theta}$. Since $0 \le \bar{z}_p \perp \bar{\xi}_0 \ge 0$ and $\bar{\theta} > 0$, we know that $\bar{\xi}_0 = 0$. Otherwise, i.e., if $\bar{q}_1 < p$, then $\bar{z}_p = \hat{v}_p + \bar{\xi}_0$. Since $0 \le \bar{z}_p \perp \bar{\xi}_0 \ge 0$ and $\hat{v}_p \ge 0$, we know that $\bar{z}_p = \hat{v}_p$ and $\bar{\xi}_0 = 0$ if $\hat{v}_p > 0$, and that $\bar{z}_p = \bar{\xi}_0 = 0$ if $\hat{v}_p = 0$. Thus, in any case, we have that $\bar{\xi}_0 = 0$. Since $\sum_{i=\bar{q}_0+1}^{\bar{q}_1} w_i = q - \bar{q}_0$, by solving (71) we obtain that

$$\begin{cases} \bar{\theta} = \Big((\|e\|^2 + \bar{q}_0 + 1) \displaystyle\sum_{i=\bar{q}_0+1}^{\bar{q}_1} \hat{v}_i - (q - \bar{q}_0)\big(\sum_{i=1}^{\bar{q}_0} \hat{v}_i + \langle e\,,\, u \rangle - \eta \big) \Big) / \bar{\rho}\,, \\ \bar{\lambda} = \Big((q - \bar{q}_0) \displaystyle\sum_{i=\bar{q}_0+1}^{\bar{q}_1} \hat{v}_i + (\bar{q}_1 - \bar{q}_0)\big(\sum_{i=1}^{\bar{q}_0} \hat{v}_i + \langle e\,,\, u \rangle - \eta \big) \Big) / \bar{\rho} \end{cases} \tag{72}$$

and

$$\begin{cases} \bar{\zeta} = \eta + \bar{\lambda}\,, & \bar{y} = u - \bar{\lambda} e\,, \\ \bar{z}_i = \hat{v}_i - \bar{\lambda}\,, & i = 1, \ldots, \bar{q}_0, \\ \bar{z}_i = \bar{\theta}\,, & i = \bar{q}_0 + 1, \ldots, \bar{q}_1, \\ \bar{z}_i = \hat{v}_i\,, & i = \bar{q}_1 + 1, \ldots, p\,, \end{cases} \tag{73}$$

where $\bar{\rho} = (\bar{q}_1 - \bar{q}_0)(\|e\|^2 + \bar{q}_0 + 1) + (q - \bar{q}_0)^2$. Then, due to the structure of $\hat{v}$, (71) is equivalent to

$$\begin{cases} \hat{v}_{\bar{q}_0} > \bar{\theta} + \bar{\lambda} \ge \hat{v}_{\bar{q}_0+1} \text{ and } \hat{v}_{\bar{q}_1} \ge \bar{\theta} > \hat{v}_{\bar{q}_1+1}, & \begin{cases} \text{if } \beta_2 \ne \emptyset,\ \bar{q}_0 < |\beta_1|\,,\ \bar{q}_1 > |\beta_1| + |\beta_2|, \\ \text{if } \beta_2 = \emptyset,\ \bar{q}_0 < q - 1\,,\ \bar{q}_1 > q\,, \end{cases} \\ \hat{v}_{\bar{q}_0} > \bar{\theta} + \bar{\lambda} \text{ and } \hat{v}_{\bar{q}_1} \ge \bar{\theta} > \hat{v}_{\bar{q}_1+1}, & \begin{cases} \text{if } \beta_2 \ne \emptyset,\ \bar{q}_0 = |\beta_1|\,,\ \bar{q}_1 > |\beta_1| + |\beta_2|, \\ \text{if } \beta_2 = \emptyset,\ \bar{q}_0 = q - 1\,,\ \bar{q}_1 > q\,, \end{cases} \\ \hat{v}_{\bar{q}_0} > \bar{\theta} + \bar{\lambda} \ge \hat{v}_{\bar{q}_0+1} \text{ and } \bar{\theta} > \hat{v}_{\bar{q}_1+1}, & \begin{cases} \text{if } \beta_2 \ne \emptyset,\ \bar{q}_0 < |\beta_1|\,,\ \bar{q}_1 = |\beta_1| + |\beta_2|, \\ \text{if } \beta_2 = \emptyset,\ \bar{q}_0 < q - 1\,,\ \bar{q}_1 = q\,, \end{cases} \\ \hat{v}_{\bar{q}_0} > \bar{\theta} + \bar{\lambda} \text{ and } \bar{\theta} > \hat{v}_{\bar{q}_1+1}, & \begin{cases} \text{if } \beta_2 \ne \emptyset,\ \bar{q}_0 = |\beta_1|\,,\ \bar{q}_1 = |\beta_1| + |\beta_2|, \\ \text{if } \beta_2 = \emptyset,\ \bar{q}_0 = q - 1\,,\ \bar{q}_1 = q \end{cases} \end{cases} \tag{74}$$

with $(\bar{\zeta}, \bar{y}, \bar{z})$ and $\bar{\lambda}$ being given by (72) and (73) (note that $\bar{\xi}$ can be obtained from $\bar{z}$ and $\bar{\lambda}$). Hence from (72), (73), (74) and the compatibility of the KKT conditions (64), we can see that $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\xi}, \bar{\lambda}) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}$ solves the KKT system (64) with $\bar{z}_q > 0$ if and only if $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_2(\eta, u, \hat{v}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 0)$ with $\texttt{flag} = 1$ for some integers $\bar{q}_0$ and $\bar{q}_1$ satisfying $0 \le \bar{q}_0 \le \min\{q-1, |\beta_1|\}$ and $\max\{q, |\beta_1| + |\beta_2|\} \le \bar{q}_1 \le p$. □

Algorithm 3 : Computing $\Pi_{\mathcal{C}_3}(\eta, u, v)$.

Step 0. (Preprocessing) *Calculate* $\hat{v} = \big(\mathrm{psgn}(v) \circ v\big)_{\pi_3}$*, pre-compute* $\tilde{s}$ *by (61), evaluate* $\tilde{v}^-$ *and* $\tilde{v}^+$ *by (62), set* $q_0 = \min\{q-1, |\beta_1|\}$ *and go to* **Step 1**.

Step 1. (Searching for the case that $\bar{z}_q = 0$) *Call* `Subroutine 1` *with* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_1(\eta, u, \hat{v}, \tilde{v}^-, q_0, \tilde{s}, 0)$*. If* $\texttt{flag} = 1$*, go to* **Step 3**. *Otherwise, if* $q_0 = 0$*, set* $q_0 = \min\{q - 1, |\beta_1|\}$ *and* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and go to* **Step 2**; *if* $q_0 > 0$*, replace* q_0 *by* $q_0 - 1$ *and repeat* **Step 1**.

Step 2. **(Searching for the case that $\bar{z}_q > 0$)** *Call* `Subroutine 2` *with* $(\bar{\zeta}, \bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_2(\eta, u, \hat{v}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 0)$*. If* $\texttt{flag} = 1$*, go to* **Step 3***. Otherwise, if* $q_1 < p$*, replace* q_1 *by* $q_1 + 1$ *and repeat* **Step 2***; if* $q_0 > 0$ *and* $q_1 = p$*, replace* q_0 *by* $q_0 - 1$*, set* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and repeat* **Step 2***.*

Step 3. *Output* $\Pi_{\mathcal{C}_3}(\eta\,,\, u\,,\, v) = (\bar{\zeta}\,,\, \bar{y}\,,\, \mathrm{psgn}(v) \circ \bar{z}_{\pi_3{}^{-1}})$ *and stop.*

Proposition 4.3 *Assume that* $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ *is given. Then the metric projection* $\Pi_{\mathcal{C}_3}(\eta, u, v)$ *of* (η, u, v) *onto* $\mathcal{C}_3$ *can be computed by Algorithm 3. Moreover, the computational cost of Algorithm 3 is* $O(|\beta_1| \log |\beta_1| + |\beta_3| \log |\beta_3| + q(p - q + 1) + m)$*, where the sorting cost is* $O(|\beta_1| \log |\beta_1| + |\beta_3| \log |\beta_3|)$ *and the searching cost is* $O(q(p - q + 1))$*.*

Proof. By combining Lemma 4.4 and Lemma 4.5, we know that Algorithm 3 solves the KKT system (64) with the solution $(\bar{\zeta}, \bar{y}, \bar{z}, \bar{\xi}, \bar{\lambda})$ (note that $\bar{\xi}$ can be derived from $(\bar{\zeta}, \bar{y}, \bar{z})$ and $\bar{\lambda}$). Consequently, $(\bar{\zeta}, \bar{y}, \bar{z})$ is the unique optimal solution to problem (63). Thus, we obtain that $\Pi_{\mathcal{C}_3}(\eta\,,\, u\,,\, v) = (\bar{\zeta}\,,\, \bar{y}\,,\, \mathrm{psgn}(v) \circ \bar{z}_{\pi_3{}^{-1}})$.

Note that `Subroutine 1` and `Subroutine 2` both have computational cost of $O(1)$. Since the total number of calls to `Subroutine 1` and `Subroutine 2` is $O(q(p - q + 1))$, we know that searching the solution costs $O(q(p - q + 1))$, which, together with the pre-computing cost of $O(p)$, the initial sorting cost of $O(|\beta_1| \log |\beta_1| + |\beta_3| \log |\beta_3|)$ and the final evaluation cost of $O(m)$ (note that $(\bar{\zeta}, \bar{y}, \bar{z})$ is evaluated only when $\texttt{flag} = 1$), implies that the computational cost of Algorithm 3 is $O(|\beta_1| \log |\beta_1| + |\beta_3| \log |\beta_3| + q(p - q + 1) + m)$. □

4.4 Projection over $\mathcal{C}_4$

Let $w = w^{\downarrow} \in \phi_{p,q}$ be given. For any given $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\Pi_{\mathcal{C}_4}(\eta, u, v)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta - \eta)^2 + \|y - u\|^2 + \|z - v\|^2\big) \\ \text{s.t.} & s_{(q)}(z) \leq \langle w\,,\, z\rangle\,, \\ & \langle e\,,\, y\rangle + \langle w\,,\, z\rangle = \zeta\,. \end{array} \tag{75}$$

Suppose that $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \cdots, p\}$ such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0, 1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. Let π_4 be a permutation of $\{1, \ldots, p\}$ such that $(v_{\beta_1})^{\downarrow} = (v_{\pi_4})_{\beta_1}$, $v_{\beta_2} = (v_{\pi_4})_{\beta_2}$, $(v_{\beta_3})^{\downarrow} = (v_{\pi_4})_{\beta_3}$, and $\pi_4{}^{-1}$ be the inverse of π_4. Let $\hat{v} := v_{\pi_4}$, i.e., $\hat{v}_i = v_{\pi_4(i)}$, $i = 1, \ldots, p$. Denote $\hat{v}_0 := +\infty$ and $\hat{v}_{p+1} := -\infty$. Define $\tilde{s} \in \mathbb{R}^{p+1}$ by

$$\tilde{s}_0 := 0\,, \quad \tilde{s}_j := \sum_{i=1}^{j} \hat{v}_i\,, \quad j = 1, \ldots, p\,. \tag{76}$$

Let $\tilde{v}^-$ and $\tilde{v}^+$ be two tuples of length $q + 1$ and $p - q + 2$ such that

$$\left\{ \begin{array}{lll} \tilde{v}_0^- := +\infty\,,\ \tilde{v}_{|\beta_1|+1}^- = \tilde{v}_q^- := -\infty\,, & \tilde{v}_i^- := \hat{v}_i\,,\ i = 1, \ldots, q\,,\ i \neq |\beta_1| + 1 \text{ and } q\,, \\ \tilde{v}_q^+ = \tilde{v}_{|\beta_1|+|\beta_2|}^+ := +\infty\,,\ \tilde{v}_{p+1}^+ := -\infty\,, & \tilde{v}_i^+ := \hat{v}_i\,,\ i = q, \ldots, p\,,\ i \neq q \text{ and } |\beta_1| + |\beta_2|\,. \end{array} \right. \tag{77}$$

Then by using Lemma 2.4, Lemma 2.5 and the structure of w, one can equivalently reformulate problem (75) as

$$\begin{array}{ll} \min & \frac{1}{2}\big((\zeta-\eta)^2+\|y-u\|^2+\|z-\hat{v}\|^2\big) \\ \text{s.t.} & z_i \geq z_q\,,\ i=1,\ldots,|\beta_1|\,, \\ & z_i = z_q\,,\ i=|\beta_1|+1,\ldots,|\beta_1|+|\beta_2|\,,\ i\neq q\,, \\ & z_i \leq z_q\,,\ i=|\beta_1|+|\beta_2|+1,\ldots,p\,, \\ & \langle e\,,\,y\rangle+\langle w\,,\,z\rangle=\zeta \end{array} \tag{78}$$

in the sense that $(\bar{\zeta},\bar{y},\bar{z})\in \mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (78) (note that $\bar{z}=\bar{z}^{\downarrow}$ in this case) if and only if $(\bar{\zeta},\bar{y},\bar{z}_{\pi_4^{-1}})$ solves problem (75). Note that $|\beta_1|\leq q\leq |\beta_1|+|\beta_2|$, and $q=|\beta_1|+|\beta_2|$ if and only if $q=|\beta_1|$, i.e., $\beta_2=\emptyset$.

The KKT conditions for (78) are given as follows:

$$\left\{\begin{array}{l} 0=\zeta-\eta-\lambda\,,\quad 0=y-u+\lambda e\,, \\ 0=z-\hat{v}-\displaystyle\sum_{i=1}^{q-1}\xi_i(e^i-e^q)-\sum_{i=q+1}^{p}\xi_i(e^q-e^i)+\lambda\, w\,, \\ 0\leq (z_i-z_q)\perp \xi_i\geq 0\,,\ i\in\beta_1\,, \\ 0\leq (z_q-z_i)\perp \xi_i\geq 0\,,\ i\in\beta_3\,, \\ z_i=z_q\,,\ i\in\beta_2\backslash\{q\}\,,\quad \langle e\,,\,y\rangle+\langle w\,,\,z\rangle=\zeta\,, \end{array}\right. \tag{79}$$

where $\lambda\in\mathbb{R}$ and $\xi=(\xi_1,\ldots,\xi_{q-1},\xi_{q+1},\ldots,\xi_p)^T\in\mathbb{R}^{p-1}$ are the corresponding Lagrange multipliers, and $e^i\in\mathbb{R}^p$, $i=1,\ldots,p$, is the i-th standard basis whose components are all 0 except its i-th component being 1. Since problem (78) has only finitely many linear constraints, then by using [26, Corollary 28.3.1] and the fact that problem (78) has a unique solution, we know that the KKT system (79) has a unique solution $(\bar{\zeta},\bar{y},\bar{z},\bar{\xi},\bar{\lambda})$ and $(\bar{\zeta},\bar{y},\bar{z})$ is the unique optimal solution to problem (78).

Lemma 4.6 *Assume that* $(\eta,u,v)\in\mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given. Then,* $(\bar{\zeta},\bar{y},\bar{z},\bar{\xi},\bar{\lambda})\in\mathbb{R}\times\mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}^{p-1}\times\mathbb{R}$ *solves the KKT system (79) if and only if* $(\bar{\zeta},\bar{y},\bar{z},\texttt{flag})=\mathtt{S}_2(\eta,u,\hat{v},\tilde{v}^-,\tilde{v}^+,\bar{q}_0,\bar{q}_1,\tilde{s},0)$ *with* $\texttt{flag}=1$ *for some integers* $\bar{q}_0$ *and* $\bar{q}_1$ *satisfying* $0\leq\bar{q}_0\leq\min\{q-1,|\beta_1|\}$ *and* $\max\{q,|\beta_1|+|\beta_2|\}\leq\bar{q}_1\leq p$.

Proof. The proof can be obtained in a similar way to that of Lemma 4.5. We omit it here. □

Algorithm 4 : Computing $\Pi_{\mathcal{C}_4}(\eta,u,v)$.

Step 0. (Preprocessing) *Calculate* $\hat{v}=v_{\pi_4}$, *pre-compute* $\tilde{s}$ *by (76), evaluate* $\tilde{v}^-$ *and* $\tilde{v}^+$ *by (77), set* $q_0=\min\{q-1,|\beta_1|\}$ *and* $q_1=\max\{q,|\beta_1|+|\beta_2|\}$, *and go to* **Step 1**.

Step 1. (Searching) *Call* `Subroutine 2` *with* $(\bar{\zeta},\bar{y},\bar{z},\texttt{flag})=\mathtt{S}_2(\eta,u,\hat{v},\tilde{v}^-,\tilde{v}^+,q_0,q_1,\tilde{s},0)$. *If* $\texttt{flag}=1$, *go to* **Step 2**. *Otherwise, if* $q_1<p$, *replace* q_1 *by* q_1+1 *and repeat* **Step 1**; *if* $q_0>0$ *and* $q_1=p$, *replace* q_0 *by* q_0-1, *set* $q_1=\max\{q,|\beta_1|+|\beta_2|\}$, *and repeat* **Step 1**.

Step 2. *Output* $\Pi_{\mathcal{C}_4}(\eta\,,u\,,v)=(\bar{\zeta}\,,\bar{y}\,,\bar{z}_{\pi_4^{-1}})$ *and stop.*

By using Lemma 4.6, we have the following proposition. Since its proof can be obtained in a similar way to that of Proposition 4.3, we omit it here.

Proposition 4.4 *Assume that* $(\eta, u, v) \in \mathbb{R} \times \mathbb{R}^{m-p} \times \mathbb{R}^p$ *is given. Then the metric projection* $\Pi_{\mathcal{C}_4}(\eta, u, v)$ *of* (η, u, v) *onto* $\mathcal{C}_4$ *can be computed by Algorithm 4. Moreover, the computational cost of Algorithm 4 is* $O(|\beta_1|\log|\beta_1| + |\beta_3|\log|\beta_3| + q(p-q+1) + m)$*, where the sorting cost is* $O(|\beta_1|\log|\beta_1| + |\beta_3|\log|\beta_3|)$ *and the searching cost is* $O(q(p-q+1))$*.*

5 Projection over the epigraph of the vector k-norm function

In this section, we will mainly study the directional differentiability and Fréchet differentiability of the metric projector over $\mathcal{K}_{(k)}$, where $\mathcal{K}_{(k)} := \operatorname{epi} f_{(k)}$, i.e., for any given integer k satisfying $1 \le k \le n$,

$$\mathcal{K}_{(k)} = \{\, (t, z) \in \mathbb{R} \times \mathbb{R}^n \mid t \ge f_{(k)}(z)\,\}\,.$$

In the following discussion, we will drop k from $\mathcal{K}_{(k)}$ if it is clear from the context. Since $\mathcal{K} = \mathcal{C}_1(n, n, k)$, by Proposition 4.1 we have the following proposition.

Proposition 5.1 *Assume that* $(t, x) \in \mathbb{R} \times \mathbb{R}^n$ *is given. Then the metric projection* $\Pi_{\mathcal{K}}(t, x)$ *of* (t, x) *onto* $\mathcal{K}$ *can be computed by Algorithm 1 with computational cost of* $O\big(n\log n + k(n-k+1)\big)$*, where the sorting cost is* $O(n\log n)$ *and the searching cost is* $O(k(n-k+1))$*.*

Next, we consider the directional derivative of $\Pi_{\mathcal{K}}(\cdot,\cdot)$. Let $(\bar{t}, \bar{x}) := \Pi_{\mathcal{K}}(t, x)$, which is computed by Algorithm 1. Note that $\mathcal{K}$ is a polyhedral convex cone. By taking into account Proposition 2.1, we first need to characterize the critical cone $\overline{\mathcal{C}}$ of $\mathcal{K}$ at (t, x), which is defined by

$$\overline{\mathcal{C}} := \mathcal{T}_{\mathcal{K}}(\bar{t}, \bar{x}) \cap \big((t, x) - (\bar{t}, \bar{x})\big)^{\perp}\,,$$

where $\mathcal{T}_{\mathcal{K}}(\bar{t}, \bar{x})$ is the tangent cone of $\mathcal{K}$ at $(\bar{t}, \bar{x})$. Denote

$$\alpha := \{1, \ldots, \bar{k}_0\}\,, \quad \beta := \{\bar{k}_0 + 1, \ldots, \bar{k}_1\} \quad \text{and} \quad \gamma := \{\bar{k}_1 + 1, \ldots, n\}\,,$$

where $\bar{k}_0$ and $\bar{k}_1$ are integers such that $0 \le \bar{k}_0 < k \le \bar{k}_1 \le n$ and

$$|\bar{x}|^{\downarrow}_1 \ge \cdots \ge |\bar{x}|^{\downarrow}_{\bar{k}_0} > |\bar{x}|^{\downarrow}_{\bar{k}_0+1} = \cdots = |\bar{x}|^{\downarrow}_{k} = \cdots = |\bar{x}|^{\downarrow}_{\bar{k}_1} > |\bar{x}|^{\downarrow}_{\bar{k}_1+1} \ge \cdots \ge |\bar{x}|^{\downarrow}_{n} \ge 0$$

with the conventions that $\bar{k}_0 = 0$ if $|\bar{x}|^{\downarrow}_1 = |\bar{x}|^{\downarrow}_k$ and that $\bar{k}_1 = n$ if $|\bar{x}|^{\downarrow}_k = |\bar{x}|^{\downarrow}_n$. Note that $f_{(k)}(z) = \|z\|_{(k)}$, $z \in \mathbb{R}^n$. By using [9, Theorem 2.4.9], we know that

$$\mathcal{T}_{\mathcal{K}}(f_{(k)}(z), z) = \{\, (\zeta, d) \in \mathbb{R} \times \mathbb{R}^n \mid \zeta \ge f'_{(k)}(z; d)\,\}\,. \tag{80}$$

Moreover, from [26, Theorem 23.4], we know that for any $d \in \mathbb{R}^n$,

$$f'_{(k)}(z; d) = \sup\big\{\langle \mu\,,\, d\rangle \mid \mu \in \partial f_{(k)}(z)\big\}\,. \tag{81}$$

Let π be a permutation of $\{1, \ldots, n\}$ such that $|x|^{\downarrow} = |x|_{\pi}$, i.e., $|x|^{\downarrow}_i = |x|_{\pi(i)}$, $i = 1, \ldots, n$, and π^{-1} be the inverse of π. Denote

$$\hat{d} := (\operatorname{sgn}(x) \circ d)_{\pi}, \quad d \in \mathbb{R}^n\,.$$

We characterize the critical cone $\overline{\mathcal{C}}$ of $\mathcal{K}$ at (t,x) by considering the following five cases:

Case 1: $t > \|x\|_{(k)}$. In this case, $(\bar{t},\bar{x}) = (t,x)$ and $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x}) = \mathbb{R}\times\mathbb{R}^n$.

Case 2: $t = \|x\|_{(k)}$ and $|\bar{x}|^{\downarrow}_k = 0$. In this case, $(\bar{t},\bar{x}) = (t,x)$ and $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x})$. From (80), (81), (8) and Lemma 2.3, we have

$$\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x}) = \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le \zeta \right\}. \tag{82}$$

Case 3: $t = \|x\|_{(k)}$ and $|\bar{x}|^{\downarrow}_k > 0$. In this case, $(\bar{t},\bar{x}) = (t,x)$ and $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x})$. By using (80), (81), (7) and Lemma 2.2, we obtain that

$$\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x}) = \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le \zeta \right\}. \tag{83}$$

Case 4: $t < \|x\|_{(k)}$ and $|\bar{x}|^{\downarrow}_k = 0$. In this case, $\bar{t} = t+\bar{\lambda}$ and $\bar{x} = \mathrm{sgn}(x)\circ(|\bar{x}|^{\downarrow})_{\pi^{-1}}$ with $\bar{\lambda} > 0$. From Lemma 2.3 and the proof of Lemma 4.1, we know that $(\mathrm{sgn}(x)\circ(x-\bar{x}))_\pi = |x|^{\downarrow} - |\bar{x}|^{\downarrow} = \bar{\lambda}\,\bar{\mu}$, where $\bar{\mu} = (e_\alpha, w) \in \partial\||\bar{x}|^{\downarrow}\|_{(k)}$ and $w = |w|^{\downarrow} \in \phi^{\le}_{n-\bar{k}_0,k-\bar{k}_0}$. Then we have

$$\begin{aligned}\big((t,x)-(\bar{t},\bar{x})\big)^{\perp} &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid -\bar{\lambda}\,\zeta + \bar{\lambda}\,\langle \mathrm{sgn}(x)\circ\bar{\mu}_{\pi^{-1}}\,,\,d\rangle = 0 \right\} \\ &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta \right\}. \end{aligned} \tag{84}$$

We note that $\mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x})$ is also given by (82). By combining (82) and (84), we derive

$$\begin{aligned}\overline{\mathcal{C}} &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le \zeta\,,\ \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta \right\} \\ &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le \langle w\,,\,\hat{d}_\beta\rangle\,,\ \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta \right\}. \end{aligned} \tag{85}$$

Moreover, if $\langle e_\beta\,,\,w\rangle < k-\bar{k}_0$, by using Lemma 2.6 we can further simplify (85) as

$$\overline{\mathcal{C}} = \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta,\ \hat{d}_{\beta_1}\ge 0,\ \hat{d}_{\beta_2} = 0,\ \hat{d}_{\beta_3} = 0 \right\}. \tag{86}$$

Case 5: $t < \|x\|_{(k)}$ and $|\bar{x}|^{\downarrow}_k > 0$. In this case, $\bar{t} = t+\bar{\lambda}$ and $\bar{x} = \mathrm{sgn}(x)\circ(|\bar{x}|^{\downarrow})_{\pi^{-1}}$ with $\bar{\lambda} > 0$. From Lemma 2.3 and the proof of Lemma 4.2, we know that $(\mathrm{sgn}(x)\circ(x-\bar{x}))_\pi = |x|^{\downarrow} - |\bar{x}|^{\downarrow} = \bar{\lambda}\,\bar{\mu}$, where $\bar{\mu} = (e_\alpha, w, 0_\gamma) \in \partial\||\bar{x}|^{\downarrow}\|_{(k)}$ and $w = |w|^{\downarrow} \in \phi_{\bar{k}_1-\bar{k}_0,k-\bar{k}_0}$. By following the same arguments in Case 3 and Case 4, we know that $\mathcal{T}_{\mathcal{K}}(\bar{t},\bar{x})$ and $\big((t,x)-(\bar{t},\bar{x})\big)^{\perp}$ are given by (83) and (84), respectively. Thus we have

$$\begin{aligned}\overline{\mathcal{C}} &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le \zeta\,,\ \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta \right\} \\ &= \left\{ (\zeta,d)\in\mathbb{R}\times\mathbb{R}^n \mid s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le \langle w\,,\,\hat{d}_\beta\rangle\,,\ \langle e_\alpha\,,\,\hat{d}_\alpha\rangle + \langle w\,,\,\hat{d}_\beta\rangle = \zeta \right\}. \end{aligned} \tag{87}$$

Denote

$$m := |\alpha| + |\beta| = \bar{k}_1\,,\quad p := |\beta| = \bar{k}_1 - \bar{k}_0 \quad\text{and}\quad q := k - \bar{k}_0\,. \tag{88}$$

From Proposition 2.1 and the above characterization of the critical cone $\overline{\mathcal{C}}$, we can derive the directional derivative of $\Pi_{\mathcal{K}}(\cdot,\cdot)$ in the following theorem.

Theorem 5.1 *Assume that $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ is given. For any $(\eta,h) \in \mathbb{R} \times \mathbb{R}^n$, denote $\hat{h} = (\mathrm{sgn}(x) \circ h)_\pi$. The directional derivative of $\Pi_{\mathcal{K}}(\cdot,\cdot)$ at (t,x) along the direction (η,h) is given by*

$$\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = \Pi_{\overline{\mathcal{C}}}(\eta,h) = (\bar{\eta},\bar{h})$$

with $(\bar{\eta},\bar{h}) \in \mathbb{R} \times \mathbb{R}^n$ satisfying

$$\tilde{h} = (\mathrm{sgn}(x) \circ \bar{h})_\pi$$

and

$$(\bar{\eta},\tilde{h}_\alpha,\tilde{h}_\beta,\tilde{h}_\gamma) = \begin{cases} (\eta,\hat{h}_\alpha,\hat{h}_\beta,\hat{h}_\gamma)\,, & \text{if } t > \|x\|_{(k)}\,, \\ \big(\Pi_{\mathcal{C}_1}(\eta,\hat{h}_\alpha,\hat{h}_\beta),\hat{h}_\gamma\big)\,, & \text{if } t = \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_k = 0\,, \\ \big(\Pi_{\mathcal{C}_2}(\eta,\hat{h}_\alpha,\hat{h}_\beta),\hat{h}_\gamma\big)\,, & \text{if } t = \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_k > 0\,, \\ \big(\Pi_{\mathcal{C}_3}(\eta,\hat{h}_\alpha,\hat{h}_\beta),\hat{h}_\gamma\big)\,, & \text{if } t < \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_k = 0\,, \\ \big(\Pi_{\mathcal{C}_4}(\eta,\hat{h}_\alpha,\hat{h}_\beta),\hat{h}_\gamma\big)\,, & \text{if } t < \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_k > 0\,, \end{cases}$$

where $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ are respectively given by (35), (36), (37) and (38), with m,p,q being defined by (88), $w = |x|^{\downarrow}_\beta/(\bar{t}-t)$ for $\mathcal{C}_3$ and $w = (|x|^{\downarrow}_\beta - |\bar{x}|^{\downarrow}_k e_\beta)/(\bar{t}-t)$ for $\mathcal{C}_4$.

From Section 4 and Section 2.1, we know that $\Pi_{\mathcal{C}_1}(\cdot)$, $\Pi_{\mathcal{C}_2}(\cdot)$ and $\Pi_{\mathcal{C}_4}(\cdot)$ can be respectively computed by Algorithm 1, Algorithm 2 and Algorithm 4, while $\Pi_{\mathcal{C}_3}(\cdot)$ can be computed by Algorithm 3 if $\langle e_\beta\,,\,w\rangle = k - \bar{k}_0$ and by BPS algorithms if $\langle e_\beta\,,\,w\rangle < k - \bar{k}_0$.

Based on all the previous discussion in this section, we are ready to characterize the Fréchet differentiability of $\Pi_{\mathcal{K}}(\cdot,\cdot)$ in the next theorem.

Theorem 5.2 *The metric projector $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is differentiable at $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ if and only if (t,x) satisfies one of the following three conditions:*

(i) $t > \|x\|_{(k)}$;

(ii) $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k = 0$, $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} < k - \bar{k}_0$ *and* $\bar{\lambda} > |x|^{\downarrow}_{\bar{k}_0+1}$;

(iii) $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k > 0$, $\bar{\lambda} + \bar{\theta} > |x|^{\downarrow}_{\bar{k}_0+1}$ *and* $\bar{\theta} < |x|^{\downarrow}_{\bar{k}_1}$.

where $(\bar{t},\bar{x}) = \Pi_{\mathcal{K}}(t,x)$, $\bar{\lambda} = \bar{t} - t$ and $\bar{\theta} = |\bar{x}|^{\downarrow}_k$.

Proof. "$\Longleftarrow$" Suppose that $(t,x) \in \mathbb{R} \times \mathbb{R}^n$ satisfies one of the three conditions. Since $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is Lipschitz continuous on $\mathbb{R} \times \mathbb{R}^n$, we know that the Fréchet differentiability and the Gâteaux differentiability of $\Pi_{\mathcal{K}}(\cdot,\cdot)$ coincide (cf. [9]). From Theorem 5.1, we know that $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is directionally differentiable at (t,x). Therefore, we only need to show that the operator $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot)) : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \times \mathbb{R}^n$ is linear. Then we consider the following three cases:

Case 1: $t > \|x\|_{(k)}$. From Theorem 5.1, we know that $\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = (\eta,h)$, for any $(\eta,h) \in \mathbb{R} \times \mathbb{R}^n$. Hence, $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is linear.

Case 2: $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k = 0$, $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} < k - \bar{k}_0$ and $\bar{\lambda} > |x|^{\downarrow}_{\bar{k}_0+1}$. In this case, $w = |x|^{\downarrow}_\beta/\bar{\lambda}$ and $\gamma = \emptyset$. Then, $\langle e_\beta\,,\,w\rangle < k - \bar{k}_0$ and $w < e_\beta$ (i.e., $\beta_1 = \emptyset$). From the characterization of the critical cone $\overline{\mathcal{C}}$, we know that $\overline{\mathcal{C}}$ is given by (86). Since $\beta_1 = \emptyset$, it is obvious that $\overline{\mathcal{C}}$ is a subspace in $\mathbb{R} \times \mathbb{R}^n$. Then from Theorem 5.1, we can see that $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is linear.

Case 3: $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k > 0$, $\bar{\lambda} + \bar{\theta} > |x|^{\downarrow}_{\bar{k}_0+1}$ and $\bar{\theta} < |x|^{\downarrow}_{\bar{k}_1}$. In this case, the critical cone $\overline{\mathcal{C}}$ is given by (87). Since $w = (|x|^{\downarrow}_\beta - \bar{\theta} e_\beta)/\bar{\lambda}$, then $0 < w < e_\beta$, i.e., $\beta_1 \cap \beta_3 = \emptyset$. Then from Lemma 2.5, we know that $\overline{\mathcal{C}}$ is a subspace in $\mathbb{R} \times \mathbb{R}^n$. This, together with Theorem 5.1, shows that $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is linear.

"$\Longrightarrow$" Let $(t,x) \in \mathbb{R} \times \mathbb{R}^{m\times n}$ be given. It suffices to show that $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is not linear in the following four cases:

Case 1: $t = \|x\|_{(k)}$. In this case, the critical cone $\overline{\mathcal{C}}$ is given by (82) and $\gamma = \emptyset$ if $|\bar{x}|^{\downarrow}_k = 0$, while $\overline{\mathcal{C}}$ is given by (83) if $|\bar{x}|^{\downarrow}_k > 0$. Choose $(\eta, h) = (-1, 0) \in \mathbb{R} \times \mathbb{R}^n$. By using Theorem 5.1, Algorithm 1 and Algorithm 2, we obtain that $\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = (\bar{\eta}, \bar{h})$ with

$$\begin{cases} (\bar{\eta}, \bar{h}_\alpha, \bar{h}_\beta) = (\bar{\lambda}_1 - 1\,,\, -\bar{\lambda}_1 e_\alpha\,,\, 0)\,, & \text{if } |\bar{x}|^{\downarrow}_k = 0\,, \\ (\bar{\eta}, \bar{h}_\alpha, \bar{h}_\beta, \bar{h}_\gamma) = (\bar{\lambda}_2 - 1\,,\, -\bar{\lambda}_2 e_\alpha\,,\, \bar{\delta} e_\beta\,,\, 0)\,, & \text{if } |\bar{x}|^{\downarrow}_k > 0\,, \end{cases} \tag{89}$$

where $\bar{\lambda}_1 = 1/(\bar{k}_0 + 1)$, $\bar{\lambda}_2 = (\bar{k}_1 - \bar{k}_0)/\bar{\rho}$, $\bar{\delta} = -(k - \bar{k}_0)/\bar{\rho}$ and $\bar{\rho} = (\bar{k}_1 - \bar{k}_0)(\bar{k}_0 + 1) + (k - \bar{k}_0)^2$. Since $\bar{\lambda}_1 \neq 0$ and $\bar{\lambda}_2 \neq 0$, it is obvious that $\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) = (1,0) \neq -\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big)$. Hence, $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is not linear.

Case 2: $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k = 0$ and $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} = k - \bar{k}_0$. In this case, $w = |x|^{\downarrow}_\beta/\bar{\lambda}$ and $\gamma = \emptyset$. Then, $\langle e_\beta\,,\, w\rangle = k - \bar{k}_0$ and the critical cone $\overline{\mathcal{C}}$ is given by (85). Choose $(\eta, h) = (1, 0) \in \mathbb{R} \times \mathbb{R}^n$. By using Theorem 5.1 and Algorithm 3, we derive that $\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = (\bar{\eta}, \bar{h})$ with

$$\begin{cases} \bar{\eta} = 1 + \bar{\lambda}_1\,, & \bar{h}_\alpha = -\bar{\lambda}_1 e_\alpha\,, \\ (\bar{h}_\beta)_i = -\bar{\lambda}_1\,, & i = 1, \ldots, \bar{q}_0, \\ (\bar{h}_\beta)_i = \bar{\delta}\,, & i = \bar{q}_0 + 1, \ldots, \bar{q}_1, \\ (\bar{h}_\beta)_i = 0\,, & i = \bar{q}_1 + 1, \ldots, |\beta|\,, \end{cases} \tag{90}$$

where $\bar{\lambda}_1 = -(\bar{q}_1 - \bar{q}_0)/\bar{\rho}$, $\bar{\delta} = (q - \bar{q}_0)/\bar{\rho}$, $\bar{\rho} = (\bar{q}_1 - \bar{q}_0)(\bar{k}_0 + \bar{q}_0 + 1) + (q - \bar{q}_0)^2$, $q = k - \bar{k}_0$ and

$$\begin{cases} \bar{q}_0 = |\beta_1|\,,\; \bar{q}_1 = |\beta_1| + |\beta_2|\,, & \text{if } \beta_2 \neq \emptyset\,, \\ \bar{q}_0 = q - 1\,,\; \bar{q}_1 = q\,, & \text{if } \beta_2 = \emptyset\,. \end{cases}$$

On the other hand, by using Theorem 5.1 and Algorithm 3, we have

$$\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) = (\bar{\eta}', \bar{h}') \;\text{ with }\; (\bar{\eta}', \bar{h}'_\alpha, \bar{h}'_\beta) = (-1 + \bar{\lambda}_2\,,\, -\bar{\lambda}_2 e_\alpha\,,\, 0)\,,$$

where $\bar{\lambda}_2 = 1/(\bar{k}_0+1)$. Since $\beta \neq \emptyset$ (at least $k \in \beta$) and $\bar{\delta} > 0$, it is obvious that $\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) \neq -\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big)$. Thus, $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is not linear.

Case 3: $t < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_k = 0$, $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} < k - \bar{k}_0$ and $\bar{\lambda} = |x|^{\downarrow}_{\bar{k}_0+1}$. In this case, $w = |x|^{\downarrow}_\beta/\bar{\lambda}$ and $\gamma = \emptyset$. Then, $\langle e_\beta\,,\, w\rangle < k - \bar{k}_0$ and the critical cone $\overline{\mathcal{C}}$ is given by (86) with $\beta_1 \neq \emptyset$. Choose $(\eta, h) \in \mathbb{R} \times \mathbb{R}^n$ such that $\eta = -|\beta_1|$, $h_\alpha = 0$, $h_{\beta_1} = -e_{\beta_1}$ and $h_{\beta_2\cup\beta_3} = 0$. By using Theorem 5.1 and the BPS algorithms discussed in Section 2.1, we obtain that $\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = (\bar{\eta}, \bar{h})$ with

$$(\bar{\eta}, \bar{h}_\alpha, \bar{h}_\beta) = (-|\beta_1| + \bar{\lambda}_0\,,\, -\bar{\lambda}_0 e_\alpha\,,\, 0)\,,$$

where $\bar\lambda_0 = |\beta_1|/(\bar k_0+1)$. However, $\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) = (-\eta,-h) \neq -\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big)$, so $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is not linear.

Case 4: $t < \|x\|_{(k)}$, $|\bar x|^{\downarrow}_{k} > 0$, and either $\bar\lambda+\bar\theta = |x|^{\downarrow}_{\bar k_0+1}$ or $\bar\theta = |x|^{\downarrow}_{\bar k_1}$. In this case, the critical cone $\overline{\mathcal{C}}$ is given by (87). Since $w = (|x|^{\downarrow}_{\beta} - \bar\theta e_{\beta})/\bar\lambda$, we know that $\beta_1\cup\beta_3 \neq \emptyset$. Choose $(\eta,h) = (1,0) \in \mathbb{R}\times\mathbb{R}^n$. By using Theorem 5.1 and Algorithm 4, we have that $\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big) = (\bar\eta,\bar h)$ with $(\bar\eta,\bar h_{\alpha},\bar h_{\beta},\bar h_{\gamma})$ being given by (90) and $\bar h_{\gamma} = 0$, while $\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) = (\bar\eta',\bar h')$ with $(\bar\eta',\bar h'_{\alpha},\bar h'_{\beta},\bar h'_{\gamma})$ being given by (89). Since $\beta_1\cup\beta_3\neq\emptyset$, we can derive that $\Pi'_{\mathcal{K}}\big((t,x);-(\eta,h)\big) \neq -\Pi'_{\mathcal{K}}\big((t,x);(\eta,h)\big)$ and thus $\Pi'_{\mathcal{K}}((t,x);(\cdot,\cdot))$ is not linear. □

6 Projection over the vector k-norm ball

Let k be an integer satisfying $1 \le k \le n$ and r be a given positive number. In this section, we will list some important results on the metric projector over the vector k-norm ball $\mathcal{B}^r_{(k)}$, i.e.,

$$\mathcal{B}^r_{(k)} = \{\, z\in\mathbb{R}^n \mid \|z\|_{(k)} \le r \,\}\,.$$

Since these results, including the algorithms for computing the solution and directional derivative of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$, and the characterization of its Fréchet differentiability, are simpler but parallel to those on the projector over the epigraph of the vector k-norm function, we omit their proofs.

6.1 Computing $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$

For any given $x\in\mathbb{R}^n$, $\Pi_{\mathcal{B}^r_{(k)}}(x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\,\|y-x\|^2 \\ \text{s.t.} & \|y\|_{(k)} \le r\,. \end{array}$$

Let π be a permutation of $\{1,\ldots,n\}$ such that $|x|^{\downarrow} = |x|_{\pi}$, i.e., $|x|^{\downarrow}_i = |x|_{\pi(i)}$, $i=1,\ldots,n$, and π^{-1} be the inverse of π. Denote $|x|^{\downarrow}_0 = +\infty$ and $|x|^{\downarrow}_{n+1} = 0$. Then we have the following algorithm for computing $\Pi_{\mathcal{B}^r_{(k)}}(x)$.

Algorithm 5 : Computing $\Pi_{\mathcal{B}^r_{(k)}}(x)$.

Step 0. (Preprocessing) *If $\|x\|_{(k)} \le r$, output $\Pi_{\mathcal{B}^r_{(k)}}(x) = x$ and stop. Otherwise, sort $|x|$ in the non-increasing order to obtain $|x|^{\downarrow}$, set $k_0 = k-1$ and go to* **Step 1***.*

Step 1. (Searching for the case that $\bar y_k = 0$) *Compute $\lambda = \big(\sum_{i=1}^{k_0}|x|^{\downarrow}_i - r\big)/k_0$. If $\lambda>0$, $\lambda \ge \sum_{i=k_0+1}^{n}|x|^{\downarrow}_i/(k-k_0)$ and $|x|^{\downarrow}_{k_0} > \lambda \ge |x|^{\downarrow}_{k_0+1}$, set $\bar k_0 = k_0$, $\bar\lambda = \lambda$ and*

$$\begin{cases} \bar y_i = |x|^{\downarrow}_i - \bar\lambda\,, & i=1,\ldots,\bar k_0, \\ \bar y_i = 0\,, & i = \bar k_0+1,\ldots,n\,, \end{cases}$$

and go to **Step 3***. Otherwise, if $k_0 = 0$, set* `flag` $=0$*, $k_0 = k-1$ and $k_1 = k$, and go to* **Step 2***; if $k_0>0$, replace k_0 by k_0-1 and repeat* **Step 1***.*

Step 2. (Searching for the case that $\bar{y}_k > 0$) *Compute* $\rho = k_0(k_1 - k_0) + (k - k_0)^2$ *and*

$$\begin{cases} \theta = \Big(k_0 \sum_{i=k_0+1}^{k_1} |x|_i^{\downarrow} - (k - k_0)\big(\sum_{i=1}^{k_0} |x|_i^{\downarrow} - r\big)\Big)/\rho \,, \\ \lambda = \Big((k - k_0) \sum_{i=k_0+1}^{k_1} |x|_i^{\downarrow} + (k_1 - k_0)\big(\sum_{i=1}^{k_0} |x|_i^{\downarrow} - r\big)\Big)/\rho \,. \end{cases}$$

If $k_0 = 0$ *and* $k_1 = n$*, set* $\texttt{flag} = 1$*. Otherwise, if* $\lambda > 0$*,* $|x|_{k_0}^{\downarrow} > \theta + \lambda \geq |x|_{k_0+1}^{\downarrow}$ *and* $|x|_{k_1}^{\downarrow} \geq \theta > |x|_{k_1+1}^{\downarrow}$*, set* $\texttt{flag} = 1$*. If* $\texttt{flag} = 1$*, set* $\bar{k}_0 = k_0$*,* $\bar{k}_1 = k_1$*,* $\bar{\theta} = \theta$*,* $\bar{\lambda} = \lambda$ *and*

$$\begin{cases} \bar{y}_i = |x|_i^{\downarrow} - \bar{\lambda} \,, & i = 1, \ldots, \bar{k}_0, \\ \bar{y}_i = \bar{\theta} \,, & i = \bar{k}_0 + 1, \ldots, \bar{k}_1, \\ \bar{y}_i = |x|_i^{\downarrow} \,, & i = \bar{k}_1 + 1, \ldots, n \,, \end{cases}$$

and go to **Step 3***. If* $\texttt{flag} = 0$ *and* $k_1 < n$*, replace* k_1 *by* $k_1 + 1$ *and repeat* **Step 2***; if* $\texttt{flag} = 0$*,* $k_0 > 0$ *and* $k_1 = n$*, replace* k_0 *by* $k_0 - 1$*, set* $k_1 = k$*, and repeat* **Step 2***.*

Step 3. *Output* $\Pi_{\mathcal{B}^r_{(k)}}(x) = \mathrm{sgn}(x) \circ \bar{y}_{\pi^{-1}}$ *and* $\bar{\lambda}$*. Then stop.*

Proposition 6.1 *Assume that* $x \in \mathbb{R}^n$ *is given. Then the metric projection* $\Pi_{\mathcal{B}^r_{(k)}}(x)$ *of* x *onto* $\mathcal{B}^r_{(k)}$ *can be computed by Algorithm 5 with computational cost of* $O(n \log n + k(n - k + 1))$*, where the sorting cost is* $O(n \log n)$ *and the searching cost is* $O(k(n - k + 1))$*.*

6.2 The differentiability of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$

In the following discussion, we will use $\mathcal{B}$ to denote $\mathcal{B}^r_{(k)}$ for convenience. For any given $x \in \mathbb{R}^n$, let $\bar{x} := \Pi_{\mathcal{B}}(x)$. Then $|\bar{x}|^{\downarrow} = (\mathrm{sgn}(x) \circ \bar{x})_{\pi}$. Denote

$$\alpha := \{1, \ldots, \bar{k}_0\} \,, \quad \beta := \{\bar{k}_0 + 1, \ldots, \bar{k}_1\} \quad \text{and} \quad \gamma := \{\bar{k}_1 + 1, \ldots, n\} \,,$$

where $\bar{k}_0$ and $\bar{k}_1$ are integers such that $0 \leq \bar{k}_0 < k \leq \bar{k}_1 \leq n$ and

$$|\bar{x}|_1^{\downarrow} \geq \cdots \geq |\bar{x}|_{\bar{k}_0}^{\downarrow} > |\bar{x}|_{\bar{k}_0+1}^{\downarrow} = \cdots = |\bar{x}|_k^{\downarrow} = \cdots = |\bar{x}|_{\bar{k}_1}^{\downarrow} > |\bar{x}|_{\bar{k}_1+1}^{\downarrow} \geq \cdots \geq |\bar{x}|_n^{\downarrow} \geq 0$$

with the conventions that $\bar{k}_0 = 0$ if $|\bar{x}|_1^{\downarrow} = |\bar{x}|_k^{\downarrow}$ and that $\bar{k}_1 = n$ if $|\bar{x}|_k^{\downarrow} = |\bar{x}|_n^{\downarrow}$. Denote

$$m := |\alpha| + |\beta| = \bar{k}_1 \,, \quad p := |\beta| = \bar{k}_1 - \bar{k}_0 \quad \text{and} \quad q := k - \bar{k}_0 \,.$$

For the case that $\|\bar{x}\|_{(k)} = r$, define the polyhedral convex cone $\mathcal{D}_1(m, p, q)$ and $\mathcal{D}_2(m, p, q)$ by

$$\mathcal{D}_1(m, p, q) := \{\, (y, z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e_\alpha \,,\, y\rangle + \|z\|_{(q)} \leq 0 \,\} \,, \tag{91}$$

$$\mathcal{D}_2(m, p, q) := \{\, (y, z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e_\alpha \,,\, y\rangle + s_{(q)}(z) \leq 0 \,\} \,. \tag{92}$$

For the case that $\|x\|_{(k)} > r$, let $w := \bar{\mu}_\beta \in \mathbb{R}^p$, where $\bar{\mu} = (|x|^{\downarrow} - |\bar{x}|^{\downarrow})/\bar{\lambda} \in \partial \||\bar{x}|^{\downarrow}\|_{(k)}$ and $\bar{\lambda}$ is computed by Algorithm 5 (note that $\bar{\lambda} > 0$ in this case). From Lemma 2.3 and Lemma

2.1, we know that $w = |w|^{\downarrow} \in \phi^{\leq}_{\overline{p},q}$ if $|\bar{x}|^{\downarrow}_{k} = 0$, and $w = |w|^{\downarrow} \in \phi_{p,q}$ if $|\bar{x}|^{\downarrow}_{k} > 0$. Therefore, w can be rewritten as $w = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $\{\beta_1, \beta_2, \beta_3\}$ is a partition of β such that $w_i = 1$ for $i \in \beta_1$, $w_i \in (0,1)$ for $i \in \beta_2$, and $w_i = 0$ for $i \in \beta_3$. Define the polyhedral convex cone $\mathcal{D}_3(m,p,q,w)$ and $\mathcal{D}_4(m,p,q,w)$ by

$$\mathcal{D}_3(m,p,q,w) := \{ (y,z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \|z\|_{(q)} \leq \langle w \,,\, z \rangle \,,\ \langle e_\alpha \,,\, y \rangle + \langle w \,,\, z \rangle = 0 \} \,, \quad (93)$$

$$\mathcal{D}_4(m,p,q,w) := \{ (y,z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid s_{(q)}(z) \leq \langle w \,,\, z \rangle \,,\ \langle e_\alpha \,,\, y \rangle + \langle w \,,\, z \rangle = 0 \} \,. \quad (94)$$

In the following discussion, we will drop m, p, q and w from $\mathcal{D}_1(m,p,q)$, $\mathcal{D}_2(m,p,q)$, $\mathcal{D}_3(m,p,q,w)$ and $\mathcal{D}_4(m,p,q,w)$ when their dependence on m, p, q and w can be seen clearly from the context.

Since $\mathcal{B}$ is a polyhedral convex set, by using Proposition 2.1, we have the following two theorems: the first one gives the directional derivative of $\Pi_{\mathcal{B}}(\cdot)$, and the second one completely characterizes the Fréchet differentiability of $\Pi_{\mathcal{B}}(\cdot)$.

Theorem 6.1 *Assume that $x \in \mathbb{R}^n$ is given. For any $h \in \mathbb{R}^n$, denote $\hat{h} := (\mathrm{sgn}(x) \circ h)_\pi$. The directional derivative of $\Pi_{\mathcal{B}}(\cdot)$ at x along the direction h is given by*

$$\Pi'_{\mathcal{B}}(x;h) = \Pi_{\overline{\mathcal{D}}}(h) = \bar{h} \,,$$

where $\bar{h} \in \mathbb{R}^n$ satisfies

$$\tilde{h} = (\mathrm{sgn}(x) \circ \bar{h})_\pi$$

and

$$(\tilde{h}_\alpha, \tilde{h}_\beta, \tilde{h}_\gamma) = \begin{cases} (\hat{h}_\alpha, \hat{h}_\beta, \hat{h}_\gamma) \,, & \text{if } r > \|x\|_{(k)} \,, \\ \big(\Pi_{\mathcal{D}_1}(\hat{h}_\alpha, \hat{h}_\beta), \hat{h}_\gamma\big) \,, & \text{if } r = \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_{k} = 0 \,, \\ \big(\Pi_{\mathcal{D}_2}(\hat{h}_\alpha, \hat{h}_\beta), \hat{h}_\gamma\big) \,, & \text{if } r = \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_{k} > 0 \,, \\ \big(\Pi_{\mathcal{D}_3}(\hat{h}_\alpha, \hat{h}_\beta), \hat{h}_\gamma\big) \,, & \text{if } r < \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_{k} = 0 \,, \\ \big(\Pi_{\mathcal{D}_4}(\hat{h}_\alpha, \hat{h}_\beta), \hat{h}_\gamma\big) \,, & \text{if } r < \|x\|_{(k)} \text{ and } |\bar{x}|^{\downarrow}_{k} > 0 \,. \end{cases}$$

Theorem 6.2 *The metric projector $\Pi_{\mathcal{B}}(\cdot)$ is differentiable at $x \in \mathbb{R}^n$ if and only if x satisfies one of the following three conditions:*

(i) $r > \|x\|_{(k)}$;

(ii) $r < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_{k} = 0$, $\sum_{i\in\beta} |x|^{\downarrow}_{i}/\bar{\lambda} < k - \bar{k}_0$ *and* $\bar{\lambda} > |x|^{\downarrow}_{\bar{k}_0+1}$;

(iii) $r < \|x\|_{(k)}$, $|\bar{x}|^{\downarrow}_{k} > 0$, $\bar{\lambda} + \bar{\theta} > |x|^{\downarrow}_{\bar{k}_0+1}$ *and* $\bar{\theta} < |x|^{\downarrow}_{\bar{k}_1}$,

where $\bar{\lambda}$, $\bar{x} = \Pi_{\mathcal{B}}(x)$ and $\bar{\theta} = |\bar{x}|^{\downarrow}_{k}$ can be computed by Algorithm 5.

In order to present our algorithms for computing the metric projectors over $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$, we need the following two subroutines, which play the similar roles to `Subroutine` 1 and `Subroutine` 2 in Section 4.

Subroutine 3 : `function` $(\bar{y}, \bar{z}, \texttt{flag}) = \mathrm{S}_3(u, v, \tilde{v}, q_0, \tilde{s}, \texttt{opt})$

- **Input:** $(u,v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\tilde{v}$ *is a* $(q+1)$*-tuple,* q_0 *is an integer satisfying* $0 \le q_0 \le q-1$, $\tilde{s} \in \mathbb{R}^{p+1}$, $\texttt{opt} = 1$ *or* 0.

- **Main Step:** *Set* $(\bar{y}, \bar{z}, \texttt{flag}) = (u, v, 0)$. *Compute*

$$\lambda = (\tilde{s}_{q_0} + \langle e_\alpha \, , \, u \rangle)/(\|e_\alpha\|^2 + q_0) \, .$$

If $\texttt{opt} = 1$, $\lambda > 0$, $\tilde{v}_{q_0} > \lambda \ge \tilde{v}_{q_0+1}$ *and* $\lambda \ge (\tilde{s}_p - \tilde{s}_{q_0})/(q-q_0)$, *set* $\texttt{flag} = 1$*; If* $\texttt{opt} = 0$, $\tilde{v}_{q_0} > \lambda \ge \tilde{v}_{q_0+1}$ *and* $\lambda \ge (\tilde{s}_p - \tilde{s}_{q_0})/(q-q_0)$, *set* $\texttt{flag} = 1$. *If* $\texttt{flag} = 1$, *set* $\bar{\lambda} = \lambda$ *and*

$$\begin{cases} \bar{y} = u - \bar{\lambda} e_\alpha \, , & \\ \bar{z}_i = v_i - \bar{\lambda} \, , & i = 1, \ldots, q_0, \\ \bar{z}_i = 0 \, , & i = q_0 + 1, \ldots, p \, . \end{cases}$$

Subroutine 4 : $\texttt{function}$ $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, v, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, \texttt{opt})$

- **Input:** $(u,v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$, $\tilde{v}^-$ *and* $\tilde{v}^+$ *are two tuples of length* $q+1$ *and* $p-q+2$, q_0 *and* q_1 *are integers satisfying* $0 \le q_0 < q \le q_1 \le p$, $\tilde{s} \in \mathbb{R}^{p+1}$, $\texttt{opt} = 1$ *or* 0.

- **Main Step:** *Set* $(\bar{y}, \bar{z}, \texttt{flag}) = (u, v, 0)$. *Compute* $\rho = (q_1 - q_0)(\|e_\alpha\|^2 + q_0) + (q-q_0)^2$ *and*

$$\begin{cases} \theta = \big((\|e_\alpha\|^2 + q_0)(\tilde{s}_{q_1} - \tilde{s}_{q_0}) - (q-q_0)(\tilde{s}_{q_0} + \langle e_\alpha \, , \, u \rangle) \big)/\rho \, , \\ \lambda = \big((q-q_0)(\tilde{s}_{q_1} - \tilde{s}_{q_0}) + (q_1 - q_0)(\tilde{s}_{q_0} + \langle e_\alpha \, , \, u \rangle) \big)/\rho \, . \end{cases}$$

If $q_0 = 0$ *and* $q_1 = p$, *set* $\texttt{flag} = 1$. *Otherwise, if* $\texttt{opt} = 1$, $\lambda > 0$, $\tilde{v}^-_{q_0} > \theta + \lambda \ge \tilde{v}^-_{q_0+1}$ *and* $\tilde{v}^+_{q_1} \ge \theta > \tilde{v}^+_{q_1+1}$, *set* $\texttt{flag} = 1$*; if* $\texttt{opt} = 0$, $\tilde{v}^-_{q_0} > \theta + \lambda \ge \tilde{v}^-_{q_0+1}$ *and* $\tilde{v}^+_{q_1} \ge \theta > \tilde{v}^+_{q_1+1}$, *set* $\texttt{flag} = 1$. *If* $\texttt{flag} = 1$, *set* $\bar{\theta} = \theta$, $\bar{\lambda} = \lambda$ *and*

$$\begin{cases} \bar{y} = u - \bar{\lambda} e_\alpha \, , & \\ \bar{z}_i = v_i - \bar{\lambda} \, , & i = 1, \ldots, q_0, \\ \bar{z}_i = \bar{\theta} \, , & i = q_0 + 1, \ldots, q_1, \\ \bar{z}_i = v_i \, , & i = q_1 + 1, \ldots, p \, . \end{cases}$$

6.2.1 Projection over $\mathcal{D}_1$

Assume that $(u,v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given. Let π_1 be a permutation of $\{1, \ldots, p\}$ such that $|v|^\downarrow = |v|_{\pi_1}$, i.e., $|v|^\downarrow_i = |v|_{\pi_1(i)}$, $i = 1, \ldots, p$, and ${\pi_1}^{-1}$ be the inverse of π_1. Then we have the following algorithm to compute $\Pi_{\mathcal{D}_1}(u,v)$.

Algorithm 6 : Computing $\Pi_{\mathcal{D}_1}(u,v)$**.**

Step 0. (Preprocessing) *If* $\langle e_\alpha \, , \, u \rangle + \|v\|_{(q)} \le 0$, *output* $\Pi_{\mathcal{D}_1}(u,v) = (u,v)$ *and stop. Otherwise, sort* $|v|$ *to obtain* $|v|^\downarrow$, *pre-compute* $\tilde{s}$ *by (40), evaluate* $\tilde{v}^-$ *and* $\tilde{v}^+$ *by (41), set* $q_0 = q-1$ *and go to* **Step 1***.*

Step 1. (Searching for the case that $\bar{z}_q = 0$) *Call* `Subroutine 3` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_3(u, |v|^{\downarrow}, \tilde{v}^{-}, q_0, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$, *go to* **Step 3**. *Otherwise, if* $q_0 = 0$, *set* $q_0 = q - 1$ *and* $q_1 = q$, *and go to* **Step 2**; *if* $q_0 > 0$, *replace* q_0 *by* $q_0 - 1$ *and repeat* **Step 1**.

Step 2. (Searching for the case that $\bar{z}_q > 0$) *Call* `Subroutine 4` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, |v|^{\downarrow}, \tilde{v}^{-}, \tilde{v}^{+}, q_0, q_1, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$, *go to* **Step 3**. *Otherwise, if* $q_1 < p$, *replace* q_1 *by* $q_1 + 1$ *and repeat* **Step 2**; *if* $q_0 > 0$ *and* $q_1 = p$, *replace* q_0 *by* $q_0 - 1$, *set* $q_1 = q$, *and repeat* **Step 2**.

Step 3. *Output* $\Pi_{\mathcal{D}_1}(u\,,\,v) = (\bar{y}\,,\,\mathrm{sgn}(v) \circ \bar{z}_{\pi_1{}^{-1}})$ *and stop.*

6.2.2 Projection over $\mathcal{D}_2$

Assume that $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given. Let π_2 be a permutation of $\{1, \ldots, p\}$ such that $v^{\downarrow} = v_{\pi_2}$, i.e., $v_i^{\downarrow} = v_{\pi_2(i)}$, $i = 1, \ldots, p$, and $\pi_2{}^{-1}$ be the inverse of π_2. Then we have the following algorithm to compute $\Pi_{\mathcal{D}_2}(u, v)$.

Algorithm 7 : Computing $\Pi_{\mathcal{D}_2}(u, v)$.

Step 0. (Preprocessing) *If* $\langle e_\alpha\,,\,u\rangle + s_{(q)}(v) \le 0$, *output* $\Pi_{\mathcal{D}_2}(u, v) = (u, v)$ *and stop. Otherwise, sort* v *to obtain* $v^{\downarrow}$, *pre-compute* $\tilde{s}$ *by (55), evaluate* $\tilde{v}^{-}$ *and* $\tilde{v}^{+}$ *by (56), set* $q_0 = q - 1$ *and* $q_1 = q$, *and go to* **Step 1**.

Step 1. (Searching) *Call* `Subroutine 4` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, v^{\downarrow}, \tilde{v}^{-}, \tilde{v}^{+}, q_0, q_1, \tilde{s}, 1)$. *If* $\texttt{flag} = 1$, *go to* **Step 2**. *Otherwise, if* $q_1 < p$, *replace* q_1 *by* $q_1 + 1$ *and repeat* **Step 1**; *if* $q_0 > 0$ *and* $q_1 = p$, *replace* q_0 *by* $q_0 - 1$, *set* $q_1 = q$, *and repeat* **Step 1**.

Step 2. *Output* $\Pi_{\mathcal{D}_2}(u\,,\,v) = (\bar{y}\,,\,\bar{z}_{\pi_2{}^{-1}})$ *and stop.*

6.2.3 Projection over $\mathcal{D}_3$

Assume that $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given. If $\langle e_\beta\,,\,w\rangle < q$, according to Lemma 2.6, $\mathcal{D}_3$ defined in (93) can be further simplified as

$$\mathcal{D}_3 = \left\{ (y, z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e_\alpha\,,\,y\rangle + \langle w\,,\,z\rangle = 0,\ z_{\beta_1} \ge 0,\ z_{\beta_2} = 0,\ z_{\beta_3} = 0 \right\}.$$

Then from Section 2.1, we know that BPS algorithms can be used to compute $\Pi_{\mathcal{D}_3}(u, v)$. Next, we consider the case that $\langle e_\beta\,,\,w\rangle = q$. Let $\mathrm{psgn}(v) \in \mathbb{R}^p$ be the vector such that $\mathrm{psgn}_i(v) = 1$, $i \in \beta_1 \cup \beta_2$, and $\mathrm{psgn}_i(v) = \mathrm{sgn}_i(v)$, $i \in \beta_3$. Let π_3 be a permutation of $\{1, \ldots, p\}$ such that $(v_{\beta_1})^{\downarrow} = (v_{\pi_3})_{\beta_1}$, $v_{\beta_2} = (v_{\pi_3})_{\beta_2}$, $|v_{\beta_3}|^{\downarrow} = |(v_{\pi_3})_{\beta_3}|$, and $\pi_3{}^{-1}$ be the inverse of π_3. Let $\hat{v} := \big(\mathrm{psgn}(v) \circ v\big)_{\pi_3}$, i.e., $\hat{v}_i = \big(\mathrm{psgn}(v) \circ v\big)_{\pi_3(i)}$, $i = 1, \ldots, p$. Then the following algorithm can be used to compute $\Pi_{\mathcal{D}_3}(u, v)$ for the case that $\sum_{i=1}^{p} w_i = q$.

Algorithm 8 : Computing $\Pi_{\mathcal{D}_3}(u, v)$ for the case that $\langle e_\beta\,,\,w\rangle = q$.

Step 0. (Preprocessing) *Calculate* $\hat{v} = \big(\mathrm{psgn}(v) \circ v\big)_{\pi_3}$, *pre-compute* $\tilde{s}$ *by (61), evaluate* $\tilde{v}^{-}$ *and* $\tilde{v}^{+}$ *by (62), set* $q_0 = \min\{q - 1, |\beta_1|\}$ *and go to* **Step 1**.

Step 1. **(Searching for the case that $\bar{z}_q = 0$)** *Call* `Subroutine 3` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_3(u, \hat{v}, \tilde{v}^-, q_0, \tilde{s}, 0)$*. If* $\texttt{flag} = 1$*, go to* **Step 3**. *Otherwise, if* $q_0 = 0$*, set* $q_0 = \min\{q - 1, |\beta_1|\}$ *and* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and go to* **Step 2***; if* $q_0 > 0$*, replace* q_0 *by* $q_0 - 1$ *and repeat* **Step 1**.

Step 2. **(Searching for the case that $\bar{z}_q > 0$)** *Call* `Subroutine 4` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, \hat{v}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 0)$*. If* $\texttt{flag} = 1$*, go to* **Step 3**. *Otherwise, if* $q_1 < p$*, replace* q_1 *by* $q_1 + 1$ *and repeat* **Step 2***; if* $q_0 > 0$ *and* $q_1 = p$*, replace* q_0 *by* $q_0 - 1$*, set* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and repeat* **Step 2**.

Step 3. *Output* $\Pi_{\mathcal{D}_3}(u\,,\,v) = (\bar{y}\,,\,\mathrm{psgn}(v) \circ \bar{z}_{\pi_3{}^{-1}})$ *and stop.*

6.2.4 Projection over $\mathcal{D}_4$

Assume that $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ is given. Let π_4 be a permutation of $\{1, \ldots, p\}$ such that $(v_{\beta_1})^\downarrow = (v_{\pi_4})_{\beta_1}$, $v_{\beta_2} = (v_{\pi_4})_{\beta_2}$, $(v_{\beta_3})^\downarrow = (v_{\pi_4})_{\beta_3}$, and $\pi_4{}^{-1}$ be the inverse of π_4. Let $\hat{v} := v_{\pi_4}$, i.e., $\hat{v}_i = v_{\pi_4(i)}$, $i = 1, \ldots, p$. Then we have the following algorithm to compute $\Pi_{\mathcal{D}_4}(u, v)$.

Algorithm 9 : Computing $\Pi_{\mathcal{D}_4}(u, v)$.

Step 0. **(Preprocessing)** *Calculate* $\hat{v} = v_{\pi_4}$*, pre-compute* $\tilde{s}$ *by (76), evaluate* $\tilde{v}^-$ *and* $\tilde{v}^+$ *by (77), set* $q_0 = \min\{q - 1, |\beta_1|\}$ *and* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and go to* **Step 1**.

Step 1. **(Searching)** *Call* `Subroutine 4` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(\eta, u, \hat{v}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 0)$*. If* $\texttt{flag} = 1$*, go to* **Step 2**. *Otherwise, if* $q_1 < p$*, replace* q_1 *by* $q_1 + 1$ *and repeat* **Step 1***; if* $q_0 > 0$ *and* $q_1 = p$*, replace* q_0 *by* $q_0 - 1$*, set* $q_1 = \max\{q, |\beta_1| + |\beta_2|\}$*, and repeat* **Step 1**.

Step 2. *Output* $\Pi_{\mathcal{D}_4}(u\,,\,v) = (\bar{y}\,,\,\bar{z}_{\pi_4{}^{-1}})$ *and stop.*

7 Conclusions

In this paper, we have conducted a thorough study on the first and second order properties of the Moreau-Yosida regularization of the vector k-norm function, the indicator function of its epigraph, and the indicator function of the vector k-norm ball. We believe that these research will play a central role in understanding the Moreau-Yosida regularization of the Ky Fan k-norm related functions and thus constitute the backbone for using the PPAs and other algorithms to solve MOPs involving the Ky Fan k-norm function. Our next step is to study the counterparts of the Ky Fan k-norm related functions and develop softwares based on the PPAs for solving such MOPs. The work done in this paper can be extended to many other situations. Below we briefly list some of them.

i) For any given integer k with $1 \leq k \leq n$, consider the function g given by

$$g(z) := s_{(k)}(z)\,, \quad z \in \mathbb{R}^n\,.$$

It is easy to verify that g^* is the indicator function of $\phi_{n,k}$. Therefore, the similar results on the Moreau-Yosida regularization of the corresponding three functions related to g can be derived in a similar but simpler way to those obtained in this paper.

ii) For any given integer k with $1 \le k \le n$, define

$$g(z) := s_{(k)}(z) + \delta_{\mathbb{R}^n_+}(z) = \|z\|_{(k)} + \delta_{\mathbb{R}^n_+}(z)\,, \quad z \in \mathbb{R}^n\,.$$

Simple calculations show that g^* is the indicator function of $\{z \in \mathbb{R}^n \mid \|z_+\|_{(k)^*} \le 1\}$ with $z_+ := \Pi_{\mathbb{R}^n_+}(z)$. Then, the Moreau-Yosida regularization of the corresponding three functions related to g at any given $x \in \mathbb{R}^n$ can be obtained by considering the counterparts of those three vector k-norm related functions in $\mathbb{R}^{|\alpha|}$ at x_α, where $\alpha := \big\{i \in [n] \mid x_i \ge 0\big\}$.

iii) For any given integer k with $1 \le k \le n$, let

$$g(z) := \frac{1}{2}\|z\|^2_{(k)}\,, \quad z \in \mathbb{R}^n\,.$$

It is easy to see that the Moreau-Yosida regularization of g at x can be derived by projecting $(0, x) \in \mathbb{R} \times \mathbb{R}^n$ onto the epigraph of the vector k-norm function.

iv) Consider the weighted vector k-norm function defined by

$$\|z\|^\omega_{(k)} := \sum_{i=1}^{k} \omega_i |z|^{\downarrow}_i\,, \quad z \in \mathbb{R}^n\,,$$

where k is an integer with $1 \le k \le n$ and $\omega = (\omega_1, \ldots, \omega_k) \in \mathbb{R}^k$ satisfying $\omega_1 \ge \cdots \ge \omega_k > 0$. This function is indeed a norm, and its dual norm is given by

$$\|z\|^\omega_{(k)^*} = \max\left\{\frac{\|z\|_{(1)}}{\omega_1}\,,\ \frac{\|z\|_{(2)}}{\omega_1 + \omega_2}\,,\ \ldots,\frac{\|z\|_{(k-1)}}{\omega_1 + \cdots + \omega_{k-1}}\,,\ \frac{\|z\|_{(n)}}{\omega_1 + \cdots + \omega_k}\right\},$$

which can be readily derived from linear programming theory. In general, the Moreau-Yosida regularization of the weighted vector k-norm related functions in the three forms discussed in this paper are much more complicated, where further research will be needed. However, for the special case that $\omega_1 = \cdots = \omega_{k-1} \ge \omega_k > 0$, it is not difficult to see that the Moreau-Yosida regularization of the weighted vector k-norm related functions have similar properties corresponding to those of the vector k-norm related functions. Also note that for this simpler case, the weighted vector k-norm and its dual norm were considered in robust optimization [2].

References

[1] F. Alizadeh, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, SIAM Journal on Optimization, 5 (1995), pp. 13–51.

[2] D. Bertsimas, D. Pachamanova, and M. Sim, *Robust linear optimization under general norms*, Operations Research Letters, 32 (2004), pp. 510–516.

[3] R. Bhatia, *Matrix Analysis*, Springer-Verlag, New York, 1997.

[4] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM Review, 46 (2004), pp. 667–689.

[5] S. Boyd, P. Diaconis, P. A. Parrilo, and L. Xiao, *Fastest mixing Markov chain on graphs with Symmetries*, SIAM Journal on Optimization, 20 (2009), pp. 792–819.

[6] P. Brucker, *An $O(n)$ algorithm for quadratic knapsack problems*, Operations Research Letters, 3 (1984), pp. 163–166.

[7] P. H. Calamai and J. J. Moré, *Quasi-Newton updates with bounds*, SIAM Journal on Numerical Analysis, 24 (1987), pp. 1434–1441.

[8] M. T. Chu, R. E. Funderlic, and R. J. Plemmons, *Structured low rank approximation*, Linear Algebra and its Applications, 366 (2003), pp. 157–172.

[9] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[10] J. Cullum, W. E. Donath and P. Wolfe, *The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices*, Mathematical Programming Study, 3 (1975), pp. 35–55.

[11] C. Ding, D. F. Sun, and K. C. Toh, *An introduction to a class of matrix cone programming*, Preprint available at `http://www.math.nus.edu.sg/~matsundf/IntroductionMCP_Sep_15.pdf`.

[12] Y. Gao, *Structured low rank matrix optimization problems: a penalized approach*, PhD thesis, Department of Mathematics, National University of Singapore, August 2010.

[13] Y. Gao and D. F. Sun, *A majorized penalty approach for calibrating rank constrained correlation matrix problems*, Preprint available at `http://www.math.nus.edu.sg/~matsundf/MajorPen.pdf`.

[14] A. Haraux, *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities*, Journal of the Mathematical Society of Japan, 29 (1977), pp. 615–631.

[15] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd edition, Cambridge University Press, 1952.

[16] M. Held, P. Wolfe, and H. P. Crowder, *Validation of subgradient optimization*, Mathematical Programming, 6 (1974), pp. 62–88.

[17] K. Helgason, J. Kennington, and H. Lall, *A polynomially bounded algorithm for a singly constrained quadratic program*, Mathematical Programming, 18 (1980), pp. 338–343.

[18] K. C. Kiwiel, *Breakpoint searching algorithms for the continuous quadratic knapsack problem*, Mathematical Programming, 112 (2008), pp. 473–491.

[19] N. Maculan, C. P. Santiago, E. M. Macambira, and M. H. C. Jardim, *An $O(n)$ algorithm for projecting a vector on the intersection of a hyperplane and a box in $\mathbb{R}^n$*, Journal of Optimization Theory and Applications, 117 (2003), pp. 553–574.

[20] J. Malick, J. Povh, F. Rendl, and A. Wiegele, *Regularization methods for semidefinite programming*, SIAM Journal on Optimization, 20 (2009), pp. 336–356.

[21] M. L. Overton and R. S. Womersley, *Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices*, Mathematical Programming, 62 (1993), pp. 321–357.

[22] J. S. Pang, *Newton's method for B-differentiable equations*, Mathematics of Operations Research, 15 (1990), pp. 149–160.

[23] P. M. Pardalos and N. Kovoor, *An algorithm for a singly constrained class of quadratic programs subject to upper and lower bounds*, Mathematical Programming, 46 (1990), pp. 321–328.

[24] G. Pataki, *On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues*, Mathematics of Operations Research, 23 (1998), pp. 339–358.

[25] J. Povh, F. Rendl, and A. Wiegele, *A boundary point method to solve semidefinite programs*, Computing, 78 (2006), pp. 277–286.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[27] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898.

[28] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research, 1 (1976), pp. 97–116.

[29] J. F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optimization Methods and Software, 11/12 (1999), pp. 625–653.

[30] R. H. Tütüncü, K. C. Toh, and M. J. Todd *Solving semidefinite-quadratic-linear programs using SDPT3*, Mathematical Programming, 95 (2003), pp. 189–217.

[31] G. A. Watson, *On matrix approximation problems with Ky Fan k norms*, Numerical Algorithms, 5 (1993), pp. 263–272.

[32] Z. Wen, D. Goldfarb, and W. Yin, *Alternating direction augmented Lagrangian methods for semidefinite programming*, Mathematical Programming Computation, 2 (2010), pp. 203–230.

[33] X. Y. Zhao, *A semismooth Newton-CG augmented Lagrangian method for large scale linear and convex quadratic SDPs*, PhD thesis, Department of Mathematics, National University of Singapore, August 2009.

[34] X. Y. Zhao, D. F. Sun, and K. C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM Journal on Optimization, 20 (2010), 1737–1765.