SIAM J. OPTIM.
Vol. 24, No. 2, pp. 766–794
© 2014 Society for Industrial and Applied Mathematics

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ON THE MOREAU–YOSIDA REGULARIZATION OF THE VECTOR k-NORM RELATED FUNCTIONS*

BIN WU†, CHAO DING‡, DEFENG SUN§, AND KIM-CHUAN TOH†

Abstract. Matrix optimization problems (MOPs) involving the Ky Fan k-norm arise frequently from many applications. In order to design algorithms to solve large scale MOPs involving the Ky Fan k-norm, we need to understand the first and second order properties of the Moreau–Yosida regularization of the Ky Fan k-norm function and the indicator function of the Ky Fan k-norm ball. According to the general theory on spectral functions, in this paper we shall conduct a thorough study on the Moreau–Yosida regularization of the vector k-norm function and the indicator function of the vector k-norm ball. In particular, we show that the proximal mappings associated with these two vector k-norm related functions both admit fast and analytically computable solutions. Moreover, we propose algorithms of low computational cost to compute the directional derivatives of these two proximal mappings and then completely characterize their Fréchet differentiability. The work here thus builds the fundamental tools needed in the design of proximal point based algorithms for solving large scale MOPs involving the Ky Fan k-norm as well as in the study of the sensitivity and stability analysis of these problems.

Key words. Moreau–Yosida regularization, Ky Fan k-norm, metric projector

AMS subject classifications. 90C25, 90C30, 65K05, 49J52

DOI. 10.1137/110827144

1. Introduction.

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

$$\chi_f(x) := \min_{z \in \mathcal{Z}} \left\{ f(z) + \frac{1}{2}\|z - x\|_{\mathcal{Z}}^2 \right\}. \tag{1}$$

The unique optimal solution $P_f(x)$ to (1) is called the proximal point of x associated with f, and $P_f : \mathcal{Z} \to \mathcal{Z}$ is called the proximal mapping. It is known that $P_f(\cdot)$ is globally Lipschitz continuous with modulus 1 [36, Proposition 12.19] and $\chi_f(\cdot)$ is continuously differentiable on $\mathcal{Z}$ [26] (see also [35, Theorem 31.5]) with

$$\nabla \chi_f(x) = x - P_f(x), \quad x \in \mathcal{Z}. \tag{2}$$

Denote the Fenchel conjugate of f by $f^*(x) := \sup_{z \in \mathcal{Z}} \{\langle x, z \rangle_{\mathcal{Z}} - f(z)\}$ for any $x \in \mathcal{Z}$. A particularly useful property for the Moreau–Yosida regularization is the following

*Received by the editors March 10, 2011; accepted for publication (in revised form) January 2, 2014; published electronically May 1, 2014.
http://www.siam.org/journals/siopt/24-2/82714.html

†Department of Mathematics, National University of Singapore, Singapore 119076 (wubin@nus.edu.sg, mattohkc@nus.edu.sg).
‡National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, P. R. China (dingchao@amss.ac.cn). Part of this work was done while the author was with the Department of Mathematics, National University of Singapore. The research of this author was supported by the National Science Foundation for Distinguished Young Scholars of China (grant 11301515).
§Department of Mathematics and Risk Management Institute, National University of Singapore, Singapore 119076 (matsundf@nus.edu.sg). The research of this author was supported in part by the Academic Research Fund under grant R-146-000-149-112.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Moreau decomposition [26] (see also [35, Theorem 31.5]): any $x \in \mathcal{Z}$ can be uniquely decomposed into

$$x = P_f(x) + P_{f^*}(x)\,. \tag{3}$$

Note that if f is an indicator function of some closed convex set $C \subseteq \mathcal{Z}$, then $P_f(\cdot) \equiv \Pi_C(\cdot)$ is the metric projector over C.

Let $\mathbb{R}^{n\times m}$ (assuming $n \le m$) denote the linear space of all $n \times m$ real matrices equipped with the inner product $\langle X, Y\rangle := \mathrm{Tr}(X^T Y)$ for X and Y in $\mathbb{R}^{n\times m}$, where "Tr" denotes the trace of a squared matrix. Let k be a given integer satisfying $1 \le k \le n$. We are interested in the Moreau–Yosida regularization of the following two Ky Fan k-norm related functions:

(i) $f(\cdot) = \|\cdot\|_{(k)}$, the Ky Fan k-norm, defined as the sum of its k largest singular values of any matrix $Z \in \mathbb{R}^{n\times m}$, i.e.,

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

where $\sigma_1(Z) \ge \sigma_2(Z) \ge \cdots \ge \sigma_n(Z)$ are the singular values of Z arranged in nonincreasing order;

(ii) $f(\cdot) = \delta_{\mathcal{B}^r_{n,m}}(\cdot)$, the indicator function of the Ky Fan k-norm ball with radius $r > 0$, where

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

The above two Ky Fan k-norm related functions appear frequently in matrix optimization problems (MOPs) of the following form:

$$\begin{array}{ll} \min & h(X) + f(X) \\ \text{s.t.} & \mathcal{A}(X) = b\,, \end{array} \tag{4}$$

where $h : \mathbb{R}^{n\times m} \to \mathbb{R}$ is a continuously differentiable function, f is one of the two Ky Fan k-norm related functions defined on $\mathbb{R}^{n\times m}$, $\mathcal{A} : \mathbb{R}^{n\times m} \to \mathbb{R}^p$ is a linear operator, and $b \in \mathbb{R}^p$ is a given vector. Suppose that $\mathcal{A}$ is onto and that $X \in \mathbb{R}^{n\times m}$ solves problem (4). Then according to [36, Theorems 8.15 and 6.14], there exist some multipliers $y \in \mathbb{R}^p$ and $\Gamma \in \mathbb{R}^{n\times m}$ such that the following Karush–Kuhn–Tucker (KKT) conditions hold:

$$\left\{ \begin{array}{l} \nabla h(X) - \mathcal{A}^* y - \Gamma = 0\,, \\ \mathcal{A}(X) - b = 0\,, \\ 0 \in \Gamma + \partial f(X)\,, \end{array} \right.$$

which, due to [26] (see also [35, Theorem 31.5]), can be equivalently rewritten as

$$F(X, y, \Gamma) := \begin{bmatrix} \nabla h(X) - \mathcal{A}^* y - \Gamma \\ \mathcal{A}(X) - b \\ X - P_{\tau f}(X - \tau\Gamma) \end{bmatrix} = 0$$

for any given $\tau > 0$, where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$. This close connection reveals that solving problem (4) relies on our understanding of the proximal mappings associated with the Ky Fan k-norm related functions.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The MOPs involving the Ky Fan k-norm function have many important applications arising from different areas such as engineering, statistics, finance, scientific computing, and machine learning. One typical example is to minimize the Ky Fan k-norm of a continuously differentiable matrix-valued function for the symmetric case [28, 31] and the nonsymmetric case [41]. Another example is the problem of finding the fastest mixing Markov chain (FMMC) on an undirected graph studied in [6, 7]. The FMMC problem can be posed as minimizing the second largest singular value of a symmetric and doubly stochastic matrix with a given sparse pattern. Since the largest singular value of any symmetric stochastic matrix is 1, the objective function in the FMMC problem is equivalent to the Ky Fan 2-norm function. More examples involving the Ky Fan k-norm function come from recent research on solving structured low rank matrix approximation problems [10], which aim to find an optimal matrix whose rank is not greater than a given positive integer. One may refer to [14, 16] to see how the Ky Fan k-norm function arises naturally in their proposed majorized penalty approach for solving the latter. For the special case that $k = 1$ or $k = n$, one can refer to the introduction of [13] and references therein for more examples of MOPs with the spectral or nuclear norm.

It is well known (cf. [3, section 4.2]) that the Ky Fan k-norm function is semidefinite representable (SDr); i.e., for any $(t, X) \in \mathbb{R} \times \mathbb{R}^{n\times m}$,

$$\|X\|_{(k)} \le t \quad \Longleftrightarrow \quad \left\{ \begin{array}{l} t - kz - \langle Z, I_{n+m}\rangle \ge 0\,, \\[2mm] Z - \left[\begin{array}{cc} 0 & X \\ X^T & 0 \end{array} \right] + zI_{n+m} \succeq 0\,, \\[2mm] Z \in \mathcal{S}^{n+m}_{+} \text{ and } z \in \mathbb{R}\,. \end{array} \right.$$

As a consequence, one popular approach for solving MOPs with the Ky Fan k-norm is to reformulate these problems as semidefinite programming (SDP) problems with expanded dimensions (cf. [28, 1]) and to apply the well-developed interior point methods (IPMs) based SDP solvers, such as SeDuMi [37] and SDPT3 [39]. 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 particularly the case when $m \gg n$ (or $n \gg m$ assuming $m \le n$). Even if $m \approx n$ (e.g., the symmetric case), the expansion of variable dimensions will inevitably lead to extra computational costs. Thus, approaches other than IPMs for solving MOPs with the Ky Fan k-norm are needed.

Our ideas for solving these problems are built on our experiences in using semismooth Newton-CG based methods to solve the nearest correlation matrix problems [32, 15] and the large scale SDP problems [43, 44]. The success of these methods depends heavily on the efficient computability of the closed form solution and the explicit directional derivative of the metric projector over the cone of symmetric and positive semidefinite matrices (SDP cone) or, equivalently, the proximal mapping associated with the SDP cone indicator function. Moreover, these semismooth Newton-CG based methods [32, 15, 43, 44] also exploit another important property—the strong semismoothness of this metric projector established in [38]. Therefore, in order to develop efficient semismooth Newton-CG based methods for solving MOPs involving the Ky Fan k-norm, the most natural step for one to take is to study the analogues of the proximal mappings associated with the Ky Fan k-norm related functions as for the metric projector over the SDP cone. Another equally important motivation for studying the proximal mappings associated with the Ky Fan k-norm related functions is

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

about the sensitivity and stability analysis of problem (4). In particular, the study on the Lipschitz-like properties such as the strong regularity [34] of the solution mapping for the perturbed version of problem (4) inevitably needs the variational properties of these proximal mappings.

For dealing with the above-mentioned computational and theoretical issues in solving MOPs with the Ky Fan k-norm, we must understand the proximal mappings associated with the Ky Fan k-norm related functions. In particular, we need

- to efficiently compute the proximal mapping $P_{\|\cdot\|_{(k)}}$ and the metric projector $\Pi_{\mathcal{B}^r_{n,m}}$;
- to obtain explicit formulas for the directional derivatives of $P_{\|\cdot\|_{(k)}}$ and $\Pi_{\mathcal{B}^r_{n,m}}$; and
- to characterize the Fréchet differentiability and establish the strong semismoothness of $P_{\|\cdot\|_{(k)}}$ and $\Pi_{\mathcal{B}^r_{n,m}}$.

For a given positive integer q, denote the set of all $q \times q$ orthogonal matrices by $\mathcal{O}^q$. For any given $X \in \mathbb{R}^{n\times m}$ ($n \le m$), consider its singular value decomposition (SVD):

$$X = \overline{U}\,[\Sigma(X)\ 0]\,\overline{V}^T = \overline{U}\,[\Sigma(X)\ 0]\,\left[\overline{V}_1\ \overline{V}_2\right]^T = \overline{U}\Sigma(X)\overline{V}_1^T\,,$$

where $\sigma(X) := (\sigma_1(X), \dots, \sigma_n(X))^T \in \mathbb{R}^n$, $\Sigma(X) := \mathrm{Diag}(\sigma(X)) \in \mathbb{R}^{n\times n}$, $\overline{U} \in \mathcal{O}^n$, and $\overline{V} = \left[\overline{V}_1\ \overline{V}_2\right] \in \mathcal{O}^m$ with $\overline{V}_1 \in \mathbb{R}^{n\times m}$ and $\overline{V}_2 \in \mathbb{R}^{m\times(m-n)}$. Let $f : \mathbb{R}^{n\times m} \to (-\infty, +\infty]$ represent either the Ky Fan k-norm function $\|\cdot\|_{(k)}$ or the Ky Fan k-norm ball indicator function $\delta_{\mathcal{B}^r_{n,m}}$. Correspondingly, let $g : \mathbb{R}^n \to (-\infty, +\infty]$ be either the vector k-norm function $g_{(k)}$ defined as the sum of the k largest entries in absolute value of any vector in $\mathbb{R}^n$ or the vector k-norm ball indicator function $\delta_{\mathcal{B}^r_{(k)}}$. Since f is unitarily invariant (cf. [22]), according to the von Neumann trace inequality [27], it is not difficult to see that $\chi_f(X) = \chi_g(\sigma(X))$ for any $X \in \mathbb{R}^{n\times m}$. This implies that χ_f is a nonsymmetric spectral function and χ_g is the corresponding absolutely symmetric function (cf. [22]). From (2), we can see that the study on the proximal mapping P_f and its first order differential property is equivalent to the study on the first and second order differential properties of the nonsymmetric spectral function χ_f. Before we proceed further, let us recall some basic results on spectral functions. For the symmetric case, the first and second order differential properties of spectral functions have been comprehensively studied in the literature. In particular, Lewis [23] showed that a symmetric spectral function is differentiable at a symmetric matrix if and only if the corresponding symmetric function is differentiable at the vector of eigenvalues, in which case the gradient of the symmetric spectral function is determined by that of the corresponding symmetric function. Subsequently, Lewis and Sendov [24] further proved that similar equivalence also holds for the second order differentiability. Furthermore, Qi and Yang [33] established that the gradient of a differentiable symmetric spectral function is directionally differentiable at a symmetric matrix provided that the gradient of the corresponding symmetric function is Hadamard directionally differentiable at the vector of eigenvalues. It was also proved in [33] that the gradient of a differentiable symmetric spectral function is (strongly) semismooth at a symmetric matrix if and only if the gradient of the corresponding symmetric function is (strongly) semismooth at the vector of eigenvalues. For the nonsymmetric case, the similar first order differential property of spectral functions was studied by Lewis [22, Theorem 3.1]. This, together with (2), implies that the proximal point $P_f(X)$ of X

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

associated with f is given by

$$P_f(X) = \overline{U}\left[\operatorname{diag}\left(P_g(\sigma(X))\right) \quad 0\right]\overline{V}^T .$$

Then, according to the general results in [12, Theorems 3.4 and 3.6], we can obtain the explicit formulas for the directional derivatives and characterize the Fréchet differentiability of the proximal mapping $P_{\|\cdot\|_{(k)}}$ and the metric projector $\Pi_{\mathcal{B}^r_{n,m}}$ by studying their vector counterparts $P_{g_{(k)}}$ and $\Pi_{\mathcal{B}^r_{(k)}}$. Moreover, since $P_{g_{(k)}}$ and $\Pi_{\mathcal{B}^r_{(k)}}$ are both piecewise affine, we know from [12, Theorem 3.12] that $P_{\|\cdot\|_{(k)}}$ and $\Pi_{\mathcal{B}^r_{n,m}}$ are both strongly semismooth everywhere. From these observations, in this paper we shall

- show that the proximal mapping $P_{g_{(k)}}$, together with its directional derivative, and the metric projector $\Pi_{\mathcal{B}^r_{(k)}}$ admit fast and exactly computable solutions;
- propose algorithms of low computational cost to compute the directional derivative of $\Pi_{\mathcal{B}^r_{(k)}}$; and
- characterize the Fréchet differentiability of $P_{g_{(k)}}$ and $\Pi_{\mathcal{B}^r_{(k)}}$.

Though our study on the Moreau–Yosida regularization of the vector k-norm related functions is mainly for the purpose of solving MOPs with the Ky Fan k-norm, the research on the vector case is also of interest itself. For example, in a recent work [2], the authors studied the Moreau–Yosida regularization of the square of the k-support norm, whose dual norm is the ℓ_2-norm of the k largest entries in the absolute value of a vector.

The remainder of this paper is organized as follows. In section 2, we give some preliminaries, mainly on the vector k-norm function. In sections 3 and 4, we study the proximal mapping $P_{g_{(k)}}$ of the vector k-norm function $g_{(k)}$ and the metric projector $\Pi_{\mathcal{B}^r_{(k)}}$ over the vector k-norm ball $\mathcal{B}^r_{(k)}$, respectively. From these two sections, we will see that given $\sigma(X)$, the computational cost for $P_{g_{(k)}}(\sigma(X))$ and $\Pi_{\mathcal{B}^r_{(k)}}(\sigma(X))$ is significantly lower than that for computing an SVD of X, which is extremely crucial from the computational point of view. In section 5, we make our conclusions including several possible extensions of our work done in this paper.

Notation. Without causing any ambiguity, denote by $\|\cdot\|_{(k)}$ both the Ky Fan k-norm and the vector k-norm, and by $\|\cdot\|_{(k)^*}$ both of their dual norms if the context is clear. In addition, we also use $g_{(k)}$ to denote the vector k-norm function. 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 entries of z being arranged in the nonincreasing order $z_1^\downarrow \geq \cdots \geq z_n^\downarrow$. We use $|z|$ to denote the vector in $\mathbb{R}^n$ whose ith entry is $|z_i|$. Let $\operatorname{sgn}(z)$ be the sign vector of z given by $(\operatorname{sgn})_i(z) = 1$ if $z_i \geq 0$ and -1 otherwise for $i = 1, \ldots, n$. 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 subvector of z obtained by removing all the entries 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 Hadamard product between vectors is denoted by "$\circ$"; i.e., for any x and $y \in \mathbb{R}^n$ the ith entry of $w := x \circ y \in \mathbb{R}^n$ is $w_i = x_i y_i$. Let e be the vector of all ones, whose dimension should be clear from the context. For any closed convex set C in a finite dimensional real Euclidean space $\mathcal{Z}$, let $\Pi_C(\cdot)$ be the metric projector over C under the standard inner product in $\mathcal{Z}$. That is, for any $x \in \mathcal{Z}$, $\Pi_C(x)$ is the unique optimal solution to the convex optimization problem $\min\left\{\frac{1}{2}\|y - x\|_{\mathcal{Z}}^2 \,|\, y \in C\right\}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. Preliminaries. In this section, we will collect some preliminary results for the vector k-norm functions. Let k be a given integer satisfying $1 \le k \le n$. The vector k-norm of any $z \in \mathbb{R}^n$ is defined as $\|z\|_{(k)} := \sum_{i=1}^k |z|_i^{\downarrow}$. The vector k-norm includes the ℓ_∞-norm ($k = 1$) and the ℓ_1-norm ($k = n$). Direct calculation shows that the dual norm of the vector k-norm (cf. [5, Exercises IV.1.18 and IV.2.12]) is given by $\|z\|_{(k)^*} = \max\{\|z\|_{(1)}, \frac{1}{k}\|z\|_{(n)}\} = \max\{\|z\|_\infty, \frac{1}{k}\|z\|_1\}$ for any $z \in \mathbb{R}^n$. In addition, let $s_{(k)}(z)$ denote the sum of the k largest entries of any $z \in \mathbb{R}^n$, i.e., $s_{(k)}(z) := \sum_{i=1}^k z_i^{\downarrow}$. Define the following four convex sets: $\phi_{n,k} := \{w \in \mathbb{R}^n \,|\, 0 \le w \le e,\ \langle e, w\rangle = k\}$, $\psi_{n,k} := \{w \in \mathbb{R}^n \,|\, w = u - v,\ 0 \le u \le e,\ 0 \le v \le e,\ \langle e, u + v\rangle = k\}$, $\phi_{n,k}^{\le} := \{w \in \mathbb{R}^n \,|\, 0 \le w \le e,\ \langle e, w\rangle \le k\}$, and $\phi_{n,k}^{>} := \{w \in \mathbb{R}^n \,|\, 0 \le w \le e,\ \langle e, w\rangle > k\}$. With these notations, it is not difficult to check that (cf. [28]) for any $z \in \mathbb{R}^n$,

$$s_{(k)}(z) = \sup\left\{\langle \mu, z\rangle \,|\, \mu \in \phi_{n,k}\right\}, \tag{5}$$

$$\|z\|_{(k)} = \sup\left\{\langle \mu, z\rangle \,|\, \mu \in \psi_{n,k}\right\}. \tag{6}$$

The following lemma gives the relationship between the convex sets defined above.

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 need only 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 $\mathcal{J} := \{j \in [n] \,|\, \sum_{i=1}^j 2(1 - w_i) \ge k - \sum_{i=1}^n w_i\} \ne \emptyset$. Let $j_0 := \min \mathcal{J}$. Define $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$ by

$$\begin{cases} u_i := 1\,, & v_i := 1 - w_i\,, \quad i = 1, \ldots, j_0 - 1\,, \\ u_{j_0} := w_{j_0} + \Delta\,, & v_{j_0} := \Delta, \\ u_i := w_i\,, & v_i := 0\,, \qquad\quad i = j_0 + 1, \ldots, n\,, \end{cases}$$

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, we have $w = u - v \in \psi_{n,k}$. ∎

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

$$z_1 \ge \cdots \ge z_{k_0} > z_{k_0+1} = \cdots = z_k = \cdots = z_{k_1} > z_{k_1+1} \ge \cdots \ge z_n\,, \tag{7}$$

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 $k_1 = n$ if $z_k = z_n$. Then, the following lemma completely characterizes the subdifferential of $s_{(k)}(\cdot)$ at such z (cf. [28]).

LEMMA 2.2. *Suppose that $z \in \mathbb{R}^n$ satisfies $z = z^{\downarrow}$ with structure* (7). *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 $z \in \mathbb{R}^n$ satisfies $z = |z|^{\downarrow}$ with the structure

$$z_1 \ge \cdots \ge z_{k_0} > z_{k_0+1} = \cdots = z_k = \cdots = z_{k_1} > z_{k_1+1} \ge \cdots \ge z_n \ge 0\,, \tag{8}$$

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 $k_1 = n$ if $z_k = z_n$. Then, the subdifferential of $\|\cdot\|_{(k)}$ at such z is characterized by the following lemma (cf. [40, 28]).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA 2.3. *Suppose that $z \in \mathbb{R}^n$ satisfies $z = |z|^{\downarrow}$ with structure* (8). *If $z_k > 0$, then*

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

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

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

The next three lemmas are useful for simplifying problems in the subsequent sections. The first is an inequality concerning the rearrangement of two vectors [18, Theorems 368 and 369].

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

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

where the inequality holds if and only if there exists a permutation π of $\{1, \dots, n\}$ 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 $\{1, \dots, n\}$ 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 $|\beta_1| \le k \le |\beta_1| + |\beta_2|$ and*

$$\begin{aligned} &\left\{ z \in \mathbb{R}^n \,|\, s_{(k)}(z) \le \langle w, z \rangle \right\} \\ &\quad = \left\{ z \in \mathbb{R}^n \,|\, z_{i_1} \ge z_{i_2} = z_{i_2'} \ge z_{i_3}\ \forall\ i_1 \in \beta_1,\ i_2, i_2' \in \beta_2,\ i_3 \in \beta_3 \right\}. \end{aligned}$$

Proof. We need only show that the relation "$\subseteq$" holds. Since for any $z \in \mathbb{R}^n$ satisfying $s_{(k)}(z) \le \langle w\,,\, z \rangle$, w solves problem (5), we obtain from the KKT conditions for (5) 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. ∎

LEMMA 2.6. *Suppose that $w \in \phi_{n,k}^{\le}$ with $w = w^{\downarrow} = (w_{\beta_1}, w_{\beta_2}, w_{\beta_3})$, where $\{\beta_1, \beta_2, \beta_3\}$ is a partition of $\{1, \dots, n\}$ 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 $|\beta_1| \le k \le |\beta_1| + |\beta_2|$ and*

$$\begin{aligned} &\left\{ z \in \mathbb{R}^n \,|\, \|z\|_{(k)} \le \langle w, z \rangle \right\} \\ &\quad = \left\{ z \in \mathbb{R}^n \,|\, z_{i_1} \ge z_{i_2} = z_{i_2'} \ge |z_{i_3}| \ge 0\ \forall\ i_1 \in \beta_1,\ i_2, i_2' \in \beta_2,\ i_3 \in \beta_3 \right\}. \end{aligned}$$

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

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

Proof. We need only 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 (6) and Lemma 2.1, we know that w solves the following problem:

$$\sup \left\{ \langle \mu, z \rangle \,|\, \mu \in \phi_{n,k}^{\le} \right\}. \tag{9}$$

Then the KKT conditions for (9) yield that

$$z_{\beta_1} = \xi_{\beta_1} + \lambda e_{\beta_1}\,, \quad z_{\beta_2} = \lambda e_{\beta_2}\,, \quad \text{and} \quad z_{\beta_3} = -\xi_{\beta_3} + \lambda e_{\beta_3} \tag{10}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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 $\|\hat{z}\|_{(k)} \le \langle w, \hat{z}\rangle$. The same way we obtained (10), we derive that

$$\hat{z}_{\beta_1} = \hat{\xi}_{\beta_1} + \hat{\lambda} e_{\beta_1}\,, \quad \hat{z}_{\beta_2} = \hat{\lambda} e_{\beta_2}\,, \quad \text{and} \quad \hat{z}_{\beta_3} = -\hat{\xi}_{\beta_3} + \hat{\lambda} e_{\beta_3} \tag{11}$$

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 (10) and (11). $\square$

In order to discuss the differentiability of the metric projectors over the polyhedral convex sets $\mathcal{B}^r_{(k)^*}$ (defined in section 3) and epi $g_{(k)}$, we need the following proposition, which characterizes the directional derivative of the metric projector over a polyhedral convex set [17, 29].

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*

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

3. The proximal mapping of the vector k-norm function.

In this section, we will study the proximal mapping $P_{g_{(k)}}$ associated with the vector k-norm function $g_{(k)}$, where k is a given integer with $1 \le k \le n$. As a consequence of the Moreau decomposition (3) and the fact that $g^*_{(k)} \equiv \delta_{\mathcal{B}^1_{(k)^*}}$, it is evident that for any $x \in \mathbb{R}^n$,

$$P_{g_{(k)}}(x) = x - P_{\delta_{\mathcal{B}^1_{(k)^*}}}(x) = x - \Pi_{\mathcal{B}^1_{(k)^*}}(x)\,, \tag{12}$$

where $\mathcal{B}^r_{(k)^*}$ is the ball with radius $r > 0$ defined by the dual norm of the vector k-norm, i.e.,

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

From (12), we can see that all the research on the proximal mapping $P_{g_{(k)}}(\cdot)$ associated with the vector k-norm function $g_{(k)}$ is equivalent to that on the metric projector $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$ over $\mathcal{B}^1_{(k)^*}$. Thus, we will mainly focus on studying $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$ in the following discussion. We first point out that $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$ can be computed via solving a well-studied quadratic program within $O(n)$ arithmetic operations. Then in order to use Proposition 2.1 to derive the directional derivative and to characterize the Fréchet differentiability of $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$ at any given $x \in \mathbb{R}^n$, we will provide a complete characterization of the critical cone of $\mathcal{B}^r_{(k)^*}$ at this point. From this characterization, we will see that the directional derivative of $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$ at x, i.e., the metric projector over the corresponding critical cone, can also be computed within $O(n)$ arithmetic operations.

3.1. 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 convex optimization problem

$$\begin{array}{ll} \min & \dfrac{1}{2}\|y - x\|^2 \\ \text{s.t.} & \|y\|_\infty \le r\,,\ \|y\|_1 \le kr\,, \end{array} \tag{13}$$

which can be equivalently rewritten as

$$\begin{array}{ll} \min & \dfrac{1}{2}\|y - |x|\|^2 \\ \text{s.t.} & 0 \le y \le re\,,\ \langle e\,,\, y\rangle \le kr \end{array} \tag{14}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

in the sense that $\bar{y} \in \mathbb{R}^n$ solves problem (14) if and only if $\mathrm{sgn}(x) \circ \bar{y}$ solves problem (13).

Remark 3.1. Note that (14) is a special case of projecting a vector onto a simple polyhedral set consisting of one linear equality (or inequality) constraint with lower and upper bounds. Specialized algorithms for this kind of problem, which aim at solving its KKT system by finding a Lagrange multiplier corresponding to the linear constraint based on breakpoint searching (BPS), have already been well studied in the literature. Among these BPS algorithms, the $O(n \log n)$ methods [19, 20] sort the breakpoints initially, while the $O(n)$ methods [8, 9, 30, 25, 21] make use of medians of breakpoint subsets.

From Remark 3.1, we can see that the computational cost of computing $\Pi_{\mathcal{B}^r_{(k)^*}}(x)$ can be achieved within $O(n)$ arithmetic operations.

3.2. The differentiability of $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$. In this subsection, we proceed to consider the directional derivative and the Fréchet differentiability of $\Pi_{\mathcal{B}^r_{(k)^*}}(\cdot)$. Note that $\mathcal{B}^r_{(k)^*}$ is a polyhedral convex set. In order to make use of Proposition 2.1, we first need to characterize the critical cone of $\mathcal{B}^r_{(k)^*}$ at a given point.

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 let π^{-1} be the inverse of π. By using Lemma 2.4, one can equivalently reformulate problem (13) as

$$\begin{array}{ll} \min & \frac{1}{2}\, \|y - |x|^\downarrow\|^2 \\ \text{s.t.} & \|y\|_\infty \le r\,,\ \ \|y\|_1 \le kr \end{array} \tag{15}$$

in the sense that $\bar{y} \in \mathbb{R}^n$ solves problem (15) (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 (13). The KKT conditions for (15) are given as follows:

$$\left\{ \begin{array}{ll} 0 = y - |x|^\downarrow + \lambda_1 \mu + \lambda_2 \nu & \text{for some } \mu \in \partial \|y\|_\infty \text{ and } \nu \in \partial \|y\|_1\,, \\ 0 \le \big(r - \|y\|_\infty\big) \perp \lambda_1 \ge 0\,, & \text{and} \quad 0 \le \big(kr - \|y\|_1\big) \perp \lambda_2 \ge 0\,, \end{array} \right. \tag{16}$$

where λ_1 and λ_2 are the corresponding Lagrange multipliers. Note that the constraints of problem (15) can be equivalently replaced by finitely many linear constraints. Then, by using [35, Corollary 28.3.1] and the fact that problem (15) has a unique solution, we know that the KKT system (16) has a solution $(\bar{y}, \bar{\lambda}_1, \bar{\lambda}_2)$ and $\bar{y}$ is the unique optimal solution to problem (15). 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 (16).

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 [11, Theorem 2.4.7], we know that for any $z \in \mathbb{R}^n$ with $g(z) = r$, it holds that

$$\mathcal{T}_{\mathcal{B}_*}(z) = \{\, d \in \mathbb{R}^n \mid g'(z; d) \le 0 \,\}\,, \tag{17}$$

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 [35, Theorem 23.4], we know that

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for any $d \in \mathbb{R}^n$,

$$g'_\infty(z; d) = \sup \left\{ \langle \mu \,,\, d \rangle \mid \mu \in \partial g_\infty(z) \right\}, \tag{18}$$

$$g'_1(z; d) = \sup \left\{ \langle \mu \,,\, d \rangle \mid \mu \in \partial g_1(z) \right\}. \tag{19}$$

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

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

Note that $\alpha \neq \emptyset$. By using Lemma 2.3, (5), (17), and (18), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \{ \, d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \, \} \,. \tag{20}$$

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 (16) 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

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

which, together with (20), yields that

$$\overline{\mathcal{C}} = \{ \, d \in \mathbb{R}^n \mid \hat{d}_\alpha \leq 0 \,,\ \langle \hat{\mu}_\alpha \,,\, \hat{d}_\alpha \rangle = 0 \, \} = \{ \, d \in \mathbb{R}^n \mid \hat{d}_{\alpha_1} = 0 \,,\ \hat{d}_{\alpha_2} \leq 0 \, \} \,, \tag{21}$$

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

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

By using Lemma 2.3, (5), (6), (17), and (19), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \begin{cases} \{ \, d \in \mathbb{R}^n \mid \langle e \,,\, \hat{d} \rangle \leq 0 \, \} & \text{if } \beta = \emptyset \,, \\ \{ \, d \in \mathbb{R}^n \mid \langle e_\alpha \,,\, \hat{d}_\alpha \rangle + \|\hat{d}_\beta\|_1 \leq 0 \, \} & \text{if } \beta \neq \emptyset \,. \end{cases} \tag{22}$$

Case 3.1: $\bar{x} = x$. Then $\overline{\mathcal{C}} = \mathcal{T}_{\mathcal{B}_*}(\bar{x})$, which is given by (22).

Case 3.2: $\bar{x} \neq x$. Then from (16), 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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

which, together with (22), yields that

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

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

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \langle e\,,\, \hat{d}\rangle = 0 \,\} & \text{if } \beta = \emptyset\,, \\ \{\, d \in \mathbb{R}^n \mid \langle e\,,\, \hat{d}\rangle = 0\,,\ \hat{d}_{\beta_1} \ge 0\,,\ \hat{d}_{\beta_2} = 0 \,\} & \text{otherwise.} \end{cases}$$

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 (16), 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

(23)
$$\begin{cases} re_\alpha = |x|_\alpha - \bar{\lambda}_1 \hat{\mu}_\alpha - \bar{\lambda}_2 e_\alpha\,, \quad |\bar{x}|_\beta = |x|_\beta - \bar{\lambda}_2 e_\beta\,, \quad 0 = |x|_\gamma - \bar{\lambda}_2 \hat{\nu}_\gamma\,, \\ 0 \le \hat{\mu}_\alpha \le e_\alpha\,,\ \sum_{i\in\alpha} \hat{\mu}_i = 1\,, \quad 0 \le \hat{\nu}_\gamma \le e_\gamma\,, \quad \bar{\lambda}_1 \ge 0\,,\ \bar{\lambda}_2 \ge 0\,,\ \bar{\lambda}_1^2 + \bar{\lambda}_2^2 \neq 0 \end{cases}$$

and

(24)
$$(x - \bar{x})^\perp = \begin{cases} \{\, d \in \mathbb{R}^n \mid \bar{\lambda}_1 \langle \hat{\mu}_\alpha\,,\, \hat{d}_\alpha\rangle + \bar{\lambda}_2 \langle e\,,\, \hat{d}\rangle = 0 \,\} & \text{if } \gamma = \emptyset, \\ \{\, d \in \mathbb{R}^n \mid \bar{\lambda}_1 \langle \hat{\mu}_\alpha\,,\, \hat{d}_\alpha\rangle + \bar{\lambda}_2 \big(\langle e_{\alpha\cup\beta}\,,\, \hat{d}_{\alpha\cup\beta}\rangle + \langle \hat{\nu}_\gamma\,,\, \hat{d}_\gamma\rangle\big) = 0 \,\} & \text{otherwise.} \end{cases}$$

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, (5), (6), (17), (18), and (19), we obtain that

$$\mathcal{T}_{\mathcal{B}_*}(\bar{x}) = \begin{cases} \{\, d \in \mathbb{R}^n \mid \hat{d}_\alpha \le 0\,,\ \langle e\,,\, \hat{d}\rangle \le 0 \,\} & \text{if } \gamma = \emptyset\,, \\ \{\, d \in \mathbb{R}^n \mid \hat{d}_\alpha \le 0\,,\ \langle e_{\alpha\cup\beta}\,,\, \hat{d}_{\alpha\cup\beta}\rangle + \|\hat{d}_\gamma\|_1 \le 0 \,\} & \text{otherwise.} \end{cases} \tag{25}$$

If $\beta \neq \emptyset$, we obtain from (23) 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 (23) 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 (23), 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 $|x|_\gamma = 0$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(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 $\bar\lambda_1 > 0$, $\bar\lambda_2 > 0$, $\alpha_2 \ne \emptyset$, $0 \le \hat\mu_\alpha \le e_\alpha$, 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 $\bar\lambda_1 > 0$, $\bar\lambda_2 > 0$, $\gamma_1 \ne \emptyset$, $0 \le \hat\mu_\alpha \le e_\alpha$, and $0 \le \hat\nu_\gamma \le e_\gamma$, which are equivalent to the conditions that $0 < \max_{i\in\gamma} |x|_i < \sum_{i\in\alpha} |x|_i/|\alpha| - r$, $\gamma \ne \emptyset$, $\max_{i\in\gamma} |x|_i \le \min_{i\in\alpha} |x|_i - r$, 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 $\bar\lambda_2 \in (\bar\lambda_2^{\min}, \bar\lambda_2^{\max}) \ne \emptyset$, 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 (25).

Case 4.2: $\bar\lambda_1 > 0$ *and* $\bar\lambda_2 = 0$. Then from (24), (25), and the fact that $\hat\mu_\alpha \ge 0$, we obtain

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha_1} = 0\,,\ \hat d_{\alpha_2} \le 0\,,\ \langle e\,, \hat d\rangle \le 0 \,\} & \text{if } \gamma = \emptyset, \\ \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha_1} = 0\,,\ \hat d_{\alpha_2} \le 0\,,\ \langle e_{\alpha\cup\beta}\ \hat d_{\alpha\cup\beta}\rangle + \|\hat d_\gamma\|_1 \le 0 \,\} & \text{otherwise.} \end{cases}$$

Case 4.3: $\bar\lambda_1 = 0$ *and* $\bar\lambda_2 > 0$. Then from (24), (25), the structure of $\hat\nu_\gamma$, and Lemma 2.6, we derive that

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha} \le 0\,,\ \langle e\,, \hat d\rangle = 0 \,\} & \text{if } \gamma = \emptyset, \\ \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha} \le 0\,,\ \hat d_{\gamma_1} \ge 0\,,\ \hat d_{\gamma_2} = 0\,,\ \langle e\,, \hat d\rangle = 0 \,\} & \text{otherwise.} \end{cases}$$

Case 4.4: $\bar\lambda_1 > 0$ *and* $\bar\lambda_2 > 0$. Let $d \in \overline{\mathcal{C}}$ be arbitrarily chosen. Since $\hat\mu_\alpha \ge 0$ and $\hat d_\alpha \le 0$, we know that $\langle \hat\mu_\alpha\,, \hat d_\alpha\rangle \le 0$. If $\gamma = \emptyset$, by noting that $\langle e\,, \hat d\rangle \le 0$ in (25), 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 $\gamma \ne \emptyset$, by using (6) and the structure of $\hat\nu_\gamma$, we obtain $\langle e_{\alpha\cup\beta}\,, \hat d_{\alpha\cup\beta}\rangle + \langle \hat\nu_\gamma\,, \hat d_\gamma\rangle \le 0$ from (25). 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 (24), (25), the fact that $\hat\mu_\alpha \ge 0$, the structure of $\hat\nu_\gamma$, and Lemma 2.6, we derive that

$$\overline{\mathcal{C}} = \begin{cases} \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha_1} = 0\,,\ \hat d_{\alpha_2} \le 0\,,\ \langle e\,, \hat d\rangle = 0 \,\} & \text{if } \gamma = \emptyset\,, \\ \{\, d \in \mathbb{R}^n \mid \hat d_{\alpha_1} = 0\,,\ \hat d_{\alpha_2} \le 0\,,\ \hat d_{\gamma_1} \ge 0\,,\ \hat d_{\gamma_2} = 0\,,\ \langle e\,, \hat d\rangle = 0 \,\} & \text{otherwise.} \end{cases}$$

From the characterization of the critical cone $\overline{\mathcal{C}}$ of $\mathcal{B}_*$ at x described in the above four cases, it is not difficult to see that the BPS algorithms mentioned in Remark 3.1 can be used to compute $\Pi_{\overline{\mathcal{C}}}(\cdot)$ within $O(n)$ arithmetic operations. Since $\mathcal{B}_*$ is a polyhedral convex set, by Proposition 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

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

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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 4.2 in subsection 4.2, we omit it here.

THEOREM 3.1. *Let $x \in \mathbb{R}^n$ be given. The metric projector $\Pi_{\mathcal{B}_*}(\cdot)$ is differentiable at x 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* $\alpha = \{i \in [n] \mid |\bar{x}|_i = r\}$;

(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. Projection over the vector *k*-norm ball.

Let k be an integer satisfying $1 \leq k \leq n$, and let r be a positive number. In this section, we will study the metric projector $\Pi_{\mathcal{B}^r_{(k)}}$ over the vector k-norm ball $\mathcal{B}^r_{(k)}$ with radius r, i.e.,

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

First, we design a fast algorithm to exactly compute the solution of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$. Second, we completely characterize the critical cone of $\mathcal{B}^r_{(k)}$ at any given point and then make use of Proposition 2.1 to obtain an explicit formula for the directional derivative and the characterization for the Fréchet differentiability of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$. Finally, we propose algorithms of low computational cost for computing the metric projectors over four basic polyhedral convex cones that come from the characterization of the critical cone of $\mathcal{B}^r_{(k)}$ at the given point. These algorithms allow us to efficiently compute the directional derivative of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$.

4.1. 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:

$$\min_{y\in\mathbb{R}^n} \left\{ \frac{1}{2}\|y - x\|^2 \mid \|y\|_{(k)} \leq r \right\}. \tag{26}$$

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 let 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 1. Computing $\Pi_{\mathcal{B}^r_{(k)}}(x)$.

Step 0 (Preprocessing). *If $\|x\|_{(k)} \leq r$, output $\Pi_{\mathcal{B}^r_{(k)}}(x) = x$ and stop. Otherwise, sort $|x|$ in nonincreasing order to obtain $|x|^{\downarrow}$, and set $k_0 = k - 1$. If $k_0 > 0$, go to* Step 1; *if $k_0 = 0$, set* `flag` $= 0$ *and $k_1 = k$, and go to* Step 2.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Step 1 (Searching for the case that $\bar{y}_k = 0$). *Compute* $\lambda = \big(\sum_{i=1}^{k_0} |x|_i^{\downarrow} - r\big)/k_0$. *If* $\lambda > 0$, $\lambda \geq \sum_{i=k_0+1}^{n} |x|_i^{\downarrow}/(k-k_0)$, *and* $|x|_{k_0}^{\downarrow} > \lambda \geq |x|_{k_0+1}^{\downarrow}$, *set* $\bar{k}_0 = k_0$, $\bar{\lambda} = \lambda$, *and*

$$\begin{cases} \bar{y}_i = |x|_i^{\downarrow} - \bar{\lambda}\,, & i = 1, \ldots, \bar{k}_0, \\ \bar{y}_i = 0\,, & i = \bar{k}_0 + 1, \ldots, n\,, \end{cases}$$

and go to Step 3. *Otherwise, if* $k_0 = 1$, *set* $\texttt{flag} = 0$, $k_0 = k-1$, *and* $k_1 = k$, *and go to* Step 2*; if* $k_0 > 1$, *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*

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

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* $\Pi_{\mathcal{B}_{(k)}^r}(x) = \operatorname{sgn}(x) \circ \bar{y}_{\pi^{-1}}$ *and* $\bar{\lambda}$. *Then stop.*

The following proposition validates Algorithm 1 and shows its low complexity by considering the KKT system of a reformulation of problem (26).

PROPOSITION 4.1. *Assume that* $x \in \mathbb{R}^n$ *is given. Then the metric projection* $\Pi_{\mathcal{B}_{(k)}^r}(x)$ *of* x *onto* $\mathcal{B}_{(k)}^r$ *can be computed by Algorithm* 1 *at a 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* $O(k(n-k+1))$.

Proof. By applying Lemma 2.4, we can equivalently reformulate (26) as

$$\min_{y \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - |x|^{\downarrow}\|^2 \mid \|y\|_{(k)} \leq r \right\} \tag{27}$$

in the sense that $\bar{y} \in \mathbb{R}^n$ solves problem (27) (note that $\bar{y} = |\bar{y}|^{\downarrow} \geq 0$ in this case) if and only if $\operatorname{sgn}(x) \circ \bar{y}_{\pi^{-1}}$ solves problem (26). The KKT conditions for (27) take the form of

$$\begin{cases} 0 = y - |x|^{\downarrow} + \lambda\,\mu \quad \text{for some } \mu \in \partial \|y\|_{(k)}\,, \\ 0 \leq (r - \|y\|_{(k)}) \perp \lambda \geq 0\,, \end{cases} \tag{28}$$

where λ is the corresponding Lagrange multiplier. Note that the constraint of problem (27) can be equivalently replaced by finite many linear constraints. Then from [35,

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Corollary 28.3.1] and the fact that the optimal solution to problem (26) is unique, we know that the KKT system (28) has a unique solution $(\bar{y}, \bar{\lambda})$ and $\bar{y}$ is also the unique solution to problem (27). If $\|x\|_{(k)} \le r$, then $(\bar{y}, \bar{\lambda}) = (|x|^{\downarrow}, 0)$. Thus, without loss of generality, we may assume that $\|x\|_{(k)} > r$. Consequently,

$$\bar{\lambda} > 0 \quad \text{and} \quad \|\bar{y}\|_{(k)} = r\,. \tag{29}$$

We next consider the following two cases.

Case 1: $\bar{y}_k = 0$. In this case, there exists an index $0 \le \bar{k}_0 \le k-1$ such that

$$\bar{y}_1 \ge \cdots \ge \bar{y}_{\bar{k}_0} > \bar{y}_{\bar{k}_0+1} = \cdots = \bar{y}_k = \cdots = \bar{y}_n = 0\,, \tag{30}$$

with the convention that $\bar{k}_0 = 0$ if $\bar{y}_1 = \bar{y}_k$. Then according to Lemma 2.3, (29), and (30), the KKT system (28) is equivalent to

$$\begin{cases} \bar{y} = |x|^{\downarrow} - \bar{\lambda}\bar{\mu} \ \text{ with } \ \bar{y}_{\bar{k}_0} > \bar{y}_{\bar{k}_0+1} = \cdots = \bar{y}_n = 0\,, \\ \bar{\mu}_i = 1 \ \text{ for } \ i = 1, \ldots, \bar{k}_0\,, \quad \big(\bar{\mu}_{\bar{k}_0+1}, \ldots, \bar{\mu}_n\big)^T \in \psi_{n-\bar{k}_0, k-\bar{k}_0}\,, \\ r = \displaystyle\sum_{i=1}^{\bar{k}_0} |x|^{\downarrow}_i - \bar{k}_0\bar{\lambda} \ \text{ and } \ \bar{\lambda} > 0\,. \end{cases} \tag{31}$$

By solving (31), we obtain that

$$\begin{cases} \bar{y}_i = |x|^{\downarrow}_i - \bar{\lambda}\,, \ i = 1, \ldots, \bar{k}_0, \\ \bar{y}_i = |x|^{\downarrow}_i - \bar{\lambda}\bar{\mu}_i = 0\,, \ i = \bar{k}_0 + 1, \ldots, n\,, \\ \bar{\lambda} = \big(\sum_{i=1}^{\bar{k}_0} |x|^{\downarrow}_i - r\big)/\bar{k}_0\,. \end{cases} \tag{32}$$

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

$$\big(\bar{\mu}_{\bar{k}_0+1}, \ldots, \bar{\mu}_n\big)^T \in \psi_{n-\bar{k}_0, k-\bar{k}_0} \ \text{ if and only if } \ \big(\bar{\mu}_{\bar{k}_0+1}, \ldots, \bar{\mu}_n\big)^T \in \phi^{\le}_{n-\bar{k}_0, k-\bar{k}_0}\,. \tag{33}$$

Thus, (31) is equivalent to

$$\bar{\lambda} > 0\,, \ |x|^{\downarrow}_{\bar{k}_0} > \bar{\lambda} \ge |x|^{\downarrow}_{\bar{k}_0+1}\,, \ \text{ and } \ \bar{\lambda} \ge \frac{1}{k - \bar{k}_0} \sum_{i=\bar{k}_0+1}^{n} |x|^{\downarrow}_i\,, \tag{34}$$

with $(\bar{y}, \bar{\lambda})$ being given by (32). From (32), (34), and the compatibility of the KKT system (28), we can see that Algorithm 1 is well defined for the case that $\bar{y}_k = 0$; that is, Step 1 of Algorithm 1 produces the unique solution to the KKT system (28) when $\bar{y}_k = 0$.

Case 2: $\bar{y}_k > 0$. In this case, there exist some indices $\bar{k}_0$ and $\bar{k}_1$ satisfying $0 \le \bar{k}_0 \le k-1$ and $k \le \bar{k}_1 \le n$ such that

$$\bar{y}_1 \ge \cdots \ge \bar{y}_{\bar{k}_0} > \bar{y}_{\bar{k}_0+1} = \cdots = \bar{y}_k = \cdots = \bar{y}_{\bar{k}_1} > \bar{y}_{\bar{k}_1+1} \ge \cdots \ge \bar{y}_n \ge 0\,, \tag{35}$$

with the conventions that $\bar{k}_0 = 0$ if $\bar{y}_1 = \bar{y}_k$ and that $\bar{k}_1 = n$ if $\bar{y}_k = \bar{y}_n$. Denote $\bar{\theta} := \bar{y}_k$. Then by using Lemma 2.3, (29), and (35), we can equivalently rewrite the KKT system (28) as

$$\begin{cases} \bar{y} = |x|^{\downarrow} - \bar{\lambda}\bar{\mu} \ \text{ with } \ \bar{y}_{\bar{k}_0} > \bar{\theta} > \bar{y}_{\bar{k}_1+1} \ \text{ and } \ \bar{y}_i = \bar{\theta}\,, \ i = \bar{k}_0 + 1, \ldots, \bar{k}_1\,, \\ \bar{\mu}_i = 1\,, \ i = 1, \ldots, \bar{k}_0\,, \ \big(\bar{\mu}_{\bar{k}_0+1}, \ldots, \bar{\mu}_{\bar{k}_1}\big)^T \in \phi_{\bar{k}_1-\bar{k}_0, k-\bar{k}_0}\,, \ \bar{\mu}_i = 0\,, \ i = \bar{k}_1 + 1, \ldots, n\,, \\ r = \displaystyle\sum_{i=1}^{\bar{k}_0} |x|^{\downarrow}_i - \bar{k}_0\bar{\lambda} + (k - \bar{k}_0)\bar{\theta}\,, \ \text{ and } \ \bar{\lambda} > 0\,. \end{cases} \tag{36}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

By solving (36), we obtain that

$$\left\{\begin{array}{l} \bar\rho = \bar k_0(\bar k_1 - \bar k_0) + (\bar k - \bar k_0)^2\,, \\ \bar\theta = \left(\bar k_0 \displaystyle\sum_{i=\bar k_0+1}^{\bar k_1} |x|^{\downarrow}_i - (k-\bar k_0)\Big(\sum_{i=1}^{\bar k_0} |x|^{\downarrow}_i - r\Big)\right)/\bar\rho\,, \\ \bar\lambda = \left((k-\bar k_0) \displaystyle\sum_{i=\bar k_0+1}^{\bar k_1} |x|^{\downarrow}_i + (\bar k_1 - \bar k_0)\Big(\sum_{i=1}^{\bar k_0} |x|^{\downarrow}_i - r\Big)\right)/\bar\rho\,, \end{array}\right. \tag{37}$$

and

$$\left\{\begin{array}{ll} \bar y_i = |x|^{\downarrow}_i - \bar\lambda\,, & i = 1,\dots,\bar k_0, \\ \bar y_i = |x|^{\downarrow}_i - \bar\lambda\bar\mu_i = \bar\theta\,, & i = \bar k_0+1,\dots,\bar k_1, \\ \bar y_i = |x|^{\downarrow}_i\,, & i = \bar k_1+1,\dots,n\,. \end{array}\right. \tag{38}$$

Then due to the structure of $|x|^{\downarrow}$, (36) is equivalent to

$$\bar\lambda > 0\,,\ |x|^{\downarrow}_{\bar k_0} > \bar\theta + \bar\lambda \geq |x|^{\downarrow}_{\bar k_0+1}\,, \quad \text{and} \quad |x|^{\downarrow}_{\bar k_1} \geq \bar\theta > |x|^{\downarrow}_{\bar k_1+1}\,, \tag{39}$$

with $\bar\theta$ and $(\bar y, \bar\lambda)$ being given by (37) and (38). From (37), (38), (39), and the compatibility of the KKT system (28), we can see that Algorithm 1 is well defined for the case that $\bar y_k > 0$; that is, Step 2 of Algorithm 1 gives the unique solution to the KKT system (28) when $\bar y_k > 0$.

Therefore, by combining Cases 1 and 2, we prove that Algorithm 1 solves the KKT system (28) with the solution $(\bar y, \bar\lambda)$. Consequently, $\bar y$ is the unique optimal solution to problem (27) and $\Pi_{\mathcal{B}^r_{(k)}}(x) = \operatorname{sgn}(x) \circ \bar y_{\pi^{-1}}$. Furthermore, the complexity of Algorithm 1 can be readily derived by analyzing the sorting cost and the searching cost. ∎

Remark 4.1. The computational cost of computing $\Pi_{\mathcal{B}^r_{(k)}}(x)$ can be further reduced to only $O(k(n-k+1))$ arithmetic operations if $x = |x|^{\downarrow}$, where this is always true for x being the vector of singular values arranged in nonincreasing order.

4.2. The differentiability of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$. In this subsection, we consider the directional derivative and the Fréchet differentiability of $\Pi_{\mathcal{B}^r_{(k)}}(\cdot)$. Without causing any ambiguity, we will use $\mathcal{B}$ to denote $\mathcal{B}^r_{(k)}$ for convenience.

For any given $x \in \mathbb{R}^n$, let $\bar x := \Pi_{\mathcal{B}}(x)$. Note that $\mathcal{B}$ is a polyhedral convex set. By taking into account Proposition 2.1, we first need to characterize the critical cone $\overline{\mathcal{D}}$ of $\mathcal{B}$ at x, which is defined by

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

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

$$\alpha := \{1,\dots,\bar k_0\}\,, \quad \beta := \{\bar k_0+1,\dots,\bar k_1\}\,, \quad \text{and} \quad \gamma := \{\bar k_1+1,\dots,n\}\,,$$

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

$$|\bar x|^{\downarrow}_1 \geq \cdots \geq |\bar x|^{\downarrow}_{\bar k_0} > |\bar x|^{\downarrow}_{\bar k_0+1} = \cdots = |\bar x|^{\downarrow}_k = \cdots = |\bar x|^{\downarrow}_{\bar k_1} > |\bar x|^{\downarrow}_{\bar k_1+1} \geq \cdots \geq |\bar x|^{\downarrow}_n \geq 0\,,$$

with the conventions that $\bar k_0 = 0$ if $|\bar x|^{\downarrow}_1 = |\bar x|^{\downarrow}_k$ and $\bar k_1 = n$ if $|\bar x|^{\downarrow}_k = |\bar x|^{\downarrow}_n$. Recall that $g_{(k)}(z) := \|z\|_{(k)}$, $z \in \mathbb{R}^n$. By using [11, Theorem 2.4.7], we know that

$$\mathcal{T}_{\mathcal{B}}(z) = \{\, d \in \mathbb{R}^n \mid g'_{(k)}(z; d) \leq 0 \,\}\,. \tag{40}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Moreover, from [35, Theorem 23.4], we know that for any $d \in \mathbb{R}^n$,

$$g'_{(k)}(z;d) = \sup\left\{\langle \mu\,,\, d\rangle \mid \mu \in \partial g_{(k)}(z)\right\}. \tag{41}$$

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 let π^{-1} be the inverse of π. Denote

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

We characterize the critical cone $\overline{\mathcal{D}}$ of $\mathcal{B}$ at x by considering the following five cases.

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

Case 2: $\|x\|_{(k)} = r$ *and* $|\bar{x}|^{\downarrow}_k = 0$. In this case, $\bar{x} = x$ and $\overline{\mathcal{D}} = \mathcal{T}_{\mathcal{B}}(\bar{x})$. From (40), (41), (6), and Lemma 2.3, we have

$$\overline{\mathcal{D}} = \mathcal{T}_{\mathcal{B}}(\bar{x}) = \left\{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le 0 \,\right\}.$$

Case 3: $\|x\|_{(k)} = r$ *and* $|\bar{x}|^{\downarrow}_k > 0$. In this case, $\bar{x} = x$ and $\overline{\mathcal{D}} = \mathcal{T}_{\mathcal{B}}(\bar{x})$. By using (40), (41), (5), and Lemma 2.2, we obtain that

$$\overline{\mathcal{D}} = \mathcal{T}_{\mathcal{B}}(\bar{x}) = \left\{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le 0 \,\right\}.$$

Case 4: $\|x\|_{(k)} > r$ *and* $|\bar{x}|^{\downarrow}_k = 0$. In this case, $\bar{x} = \mathrm{sgn}(x)\circ(|\bar{x}|^{\downarrow})_{\pi^{-1}}$ with $\bar{y} = |\bar{x}|^{\downarrow}$ and $\bar{\lambda} > 0$ being given by (32). From Lemma 2.3, (31), and (33), 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

$$\begin{aligned}(x-\bar{x})^{\perp} &= \left\{\, d \in \mathbb{R}^n \mid \bar{\lambda}\,\langle \mathrm{sgn}(x)\circ\bar{\mu}_{\pi^{-1}}\,,\, d\rangle = 0 \,\right\}\\ &= \left\{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \langle w\,,\, \hat{d}_\beta\rangle = 0 \,\right\}.\end{aligned} \tag{42}$$

We note that $\mathcal{T}_{\mathcal{B}}(\bar{x})$ is obtained in Case 2. This, together with (42), yields that

$$\begin{aligned}\overline{\mathcal{D}} &= \left\{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le 0\,,\ \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \langle w\,,\, \hat{d}_\beta\rangle = 0 \,\right\}\\ &= \left\{\, d \in \mathbb{R}^n \mid \|\hat{d}_\beta\|_{(k-\bar{k}_0)} \le \langle w\,,\, \hat{d}_\beta\rangle\,,\ \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \langle w\,,\, \hat{d}_\beta\rangle = 0 \,\right\}.\end{aligned}$$

Case 5: $|x\|_{(k)} > r$ *and* $|\bar{x}|^{\downarrow}_k > 0$. In this case, $\bar{x} = \mathrm{sgn}(x)\circ(|\bar{x}|^{\downarrow})_{\pi^{-1}}$ with $\bar{y} = |\bar{x}|^{\downarrow}$ and $\bar{\lambda} > 0$ being given by (37) and (38). From Lemma 2.3 and (36), 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 $w = |w|^{\downarrow} \in \phi_{\bar{k}_1-\bar{k}_0,k-\bar{k}_0}$. Since $\mathcal{T}_{\mathcal{B}}(\bar{x})$ is obtained in Case 3 and $(x-\bar{x})^{\perp}$ is given by (42), we have

$$\begin{aligned}\overline{\mathcal{D}} &= \left\{\, d \in \mathbb{R}^n \mid \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le 0\,,\ \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \langle w\,,\, \hat{d}_\beta\rangle = 0 \,\right\}\\ &= \left\{\, d \in \mathbb{R}^n \mid s_{(k-\bar{k}_0)}(\hat{d}_\beta) \le \langle w\,,\, \hat{d}_\beta\rangle\,,\ \langle e_\alpha\,,\, \hat{d}_\alpha\rangle + \langle w\,,\, \hat{d}_\beta\rangle = 0 \,\right\}.\end{aligned}$$

Based on the above complete characterization of the critical cone $\overline{\mathcal{D}}$ of $\mathcal{B}$ at any given $x \in \mathbb{R}^n$, we define the polyhedral convex cones $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, and $\mathcal{D}_4$ by

$$\mathcal{D}_1 := \left\{\, (y,z) \in \mathbb{R}^{m-p}\times\mathbb{R}^p \mid \langle e_\alpha\,,\, y\rangle + \|z\|_{(q)} \le 0 \,\right\}, \tag{43}$$

$$\mathcal{D}_2 := \left\{\, (y,z) \in \mathbb{R}^{m-p}\times\mathbb{R}^p \mid \langle e_\alpha\,,\, y\rangle + s_{(q)}(z) \le 0 \,\right\}, \tag{44}$$

$$\mathcal{D}_3 := \left\{\, (y,z) \in \mathbb{R}^{m-p}\times\mathbb{R}^p \mid \|z\|_{(q)} \le \langle w\,,\, z\rangle\,,\ \langle e_\alpha\,,\, y\rangle + \langle w\,,\, z\rangle = 0 \,\right\}, \tag{45}$$

$$\mathcal{D}_4 := \left\{\, (y,z) \in \mathbb{R}^{m-p}\times\mathbb{R}^p \mid s_{(q)}(z) \le \langle w\,,\, z\rangle\,,\ \langle e_\alpha\,,\, y\rangle + \langle w\,,\, z\rangle = 0 \,\right\}, \tag{46}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $m := |\alpha| + |\beta| = \bar{k}_1$, $p := |\beta| = \bar{k}_1 - \bar{k}_0$, $q := k - \bar{k}_0$, $e_\alpha \in \mathbb{R}^{m-p}$ is the vector of all ones, and $w \in \mathbb{R}^p$ is given in Case 4 or Case 5. More specifically, $w = |w|^\downarrow \in \phi^{\leq}_{p,q}$ in (45) for Case 4, while $w = |w|^\downarrow \in \phi_{p,q}$ in (46) for Case 5. In any case, 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 $\{1, \dots, p\}$ 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 $i \in \beta_3$. Furthermore, if $\langle e_\beta\, , \, w\rangle < q$, we can use Lemma 2.6 to simplify (45) as

$$(47)\quad \mathcal{D}_3 = \left\{ (y, z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \mid \langle e_\alpha\, , \, y\rangle + \langle w\, , \, z\rangle = 0,\ z_{\beta_1} \geq 0,\ z_{\beta_2} = 0,\ z_{\beta_3} = 0 \right\}.$$

For this simple case, the BPS algorithms discussed in Remark 3.1 can be used to compute the metric projector over the polyhedral convex cone (47) within $O(m)$ arithmetic operations. For the general cases, we develop algorithms of low computational cost for computing the metric projectors over the polyhedral convex cones (43), (44), (45), and (46) in subsection 4.3. By applying Proposition 2.1, we derive the directional derivative of $\Pi_{\mathcal{B}}(\cdot)$ in the following theorem.

THEOREM 4.1. *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*

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

with $\bar{h} \in \mathbb{R}^n$ satisfying

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

According to Theorem 4.1 and the complete characterization of the critical cone $\overline{\mathcal{D}}$, we are now ready to characterize the Fréchet differentiability of $\Pi_{\mathcal{B}}(\cdot)$ in the following theorem.

THEOREM 4.2. *For any given $x \in \mathbb{R}^n$, let $\bar{x} := \Pi_{\mathcal{B}}(x)$ and $\bar{\lambda}$ be computed by Algorithm 1. The metric projector $\Pi_{\mathcal{B}}(\cdot)$ is differentiable at x if and only if x satisfies one of the following four conditions:*

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

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

(iii) $\|x\|_{(k)} > r$, $k < \bar{k}_1$, *and* $\max\{0, |x|^\downarrow_{\bar{k}_0+1} - \bar{\lambda}\} < |\bar{x}|^\downarrow_k < |x|^\downarrow_{\bar{k}_1}$;

(iv) $\|x\|_{(k)} > r$, $k = \bar{k}_1$, *and* $|\bar{x}|^\downarrow_k > 0$.

Proof. "$\Longleftarrow$" Suppose that $x \in \mathbb{R}^n$ satisfies one of the four conditions. Since $\Pi_{\mathcal{B}}(\cdot)$ is Lipschitz continuous on $\mathbb{R}^n$, the Fréchet differentiability and the Gâteaux differentiability of $\Pi_{\mathcal{B}}(\cdot)$ coincide (cf. [11]). From Theorem 4.1, we know that $\Pi_{\mathcal{B}}(\cdot)$ is directionally differentiable at x. Therefore, we need only show that the operator $\Pi'_{\mathcal{B}}(x; \cdot) : \mathbb{R}^n \to \mathbb{R}^n$ is linear by using Theorem 4.1 and the complete characterization of the critical cone $\overline{\mathcal{D}}$. We consider the following four cases.

Case 1: $\|x\|_{(k)} < r$. Then $\Pi'_{\mathcal{B}}(x; h) = h$ for any $h \in \mathbb{R}^n$. Hence, $\Pi'_{\mathcal{B}}(x; \cdot)$ is linear.

Case 2: $\|x\|_{(k)} > r$, $|\bar{x}|^\downarrow_k = 0$, $\sum_{i\in\beta} |x|^\downarrow_i / \bar{\lambda} < k - \bar{k}_0$, *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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

characterization Case 4 and Lemma 2.6, we have $\overline{\mathcal{D}} = \{d \in \mathbb{R}^n \mid \langle e_\alpha, \hat{d}_\alpha\rangle + \langle w, \hat{d}_\beta\rangle = 0,\ \hat{d}_{\beta_1} \geq 0,\ \hat{d}_{\beta_2} = 0,\ \hat{d}_{\beta_3} = 0\}$. Since $\beta_1 = \emptyset$, it is obvious that $\overline{\mathcal{D}}$ is a subspace in $\mathbb{R}^n$. Consequently, $\Pi'_{\mathcal{B}}(x;\cdot)$ is linear.

Case 3: $\|x\|_{(k)} > r$, $k < \bar{k}_1$, *and* $\max\{0, |x|^{\downarrow}_{\bar{k}_0+1} - \bar{\lambda}\} < |\bar{x}|^{\downarrow}_k < |x|^{\downarrow}_{\bar{k}_1}$. In this case, $\overline{\mathcal{D}}$ is given by the characterization Case 5. Since $w = (|x|^{\downarrow}_\beta - |\bar{x}|^{\downarrow}_k e_\beta)/\bar{\lambda}$, then $0 < w < e_\beta$, i.e., $\beta_1 \cup \beta_3 = \emptyset$. Then from Lemma 2.5, we know that $\overline{\mathcal{D}}$ is a subspace in $\mathbb{R}^n$. This shows that $\Pi'_{\mathcal{B}}(x;\cdot)$ is linear.

Case 4: $\|x\|_{(k)} > r$, $k = \bar{k}_1$, *and* $|\bar{x}|^{\downarrow}_k > 0$. In this case, $\overline{\mathcal{D}}$ is given by the characterization Case 5 with $w \in \phi_{k-\bar{k}_0, k-\bar{k}_0}$ (i.e., $w = e_\beta$). This implies that $\overline{\mathcal{D}}$ can be further simplified as $\overline{\mathcal{D}} = \{d \in \mathbb{R}^n \mid \langle e_\alpha, \hat{d}_\alpha\rangle + \langle w, \hat{d}_\beta\rangle = 0\}$. Therefore, $\overline{\mathcal{D}}$ is a subspace in $\mathbb{R}^n$ and $\Pi'_{\mathcal{B}}(x;\cdot)$ is linear.

"$\Longrightarrow$" Let $x \in \mathbb{R}^n$ be given. Without loss of generality, we may assume that $x = |x|^{\downarrow}$. Then it suffices to show that $\Pi'_{\mathcal{B}}(x;(\cdot))$ is not linear in the following four cases.

Case 1: $\|x\|_{(k)} = r$. Then $\overline{\mathcal{D}}$ is given by the characterization Case 2 with $\gamma = \emptyset$ if $|\bar{x}|^{\downarrow}_k = 0$ or the characterization Case 3 if $|\bar{x}|^{\downarrow}_k > 0$. Choose $h = (e_\alpha, 0_\beta, 0_\gamma) \in \mathbb{R}^n$. From Theorem 4.1, (52), and (58), we obtain that $\Pi'_{\mathcal{B}}(x;h) = 0$ if $|\bar{x}|^{\downarrow}_k = 0$ and $\Pi'_{\mathcal{B}}(x;h) = ((1-\bar{\lambda})e_\alpha, -\bar{\lambda}\bar{\mu}, 0_\gamma)$ if $|\bar{x}|^{\downarrow}_k > 0$. Since $\alpha \neq \emptyset$ in the former case and $\bar{\lambda}\bar{\mu} \neq 0_\beta$ in the latter case, it is obvious that $\Pi'_{\mathcal{B}}(x;-h) = (-e_\alpha, 0_\beta, 0_\gamma) \neq -\Pi'_{\mathcal{B}}(x;h)$. Hence, $\Pi'_{\mathcal{B}}(x;\cdot)$ is not linear.

Case 2: $\|x\|_{(k)} > r$, $|\bar{x}|^{\downarrow}_k = 0$, *and* $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} = k - \bar{k}_0$. In this case, $\alpha \neq \emptyset$, $\gamma = \emptyset$, and $w = |x|^{\downarrow}_\beta/\bar{\lambda}$. Since $\langle e_\beta\, , \, w\rangle = k - \bar{k}_0$, we know that $\overline{\mathcal{D}}$ is given by the characterization Case 4 with $w = |w|^{\downarrow} \in \phi_{n-\bar{k}_0, k-\bar{k}_0}$. The proof for this case can be completed by following the similar argument in Case 1 and applying Theorem 4.1 with Algorithm 4.

Case 3: $\|x\|_{(k)} > r$, $|\bar{x}|^{\downarrow}_k = 0$, $\sum_{i\in\beta} |x|^{\downarrow}_i/\bar{\lambda} < k - \bar{k}_0$, *and* $\bar{\lambda} = |x|^{\downarrow}_{\bar{k}_0+1}$. In this case, $\alpha \neq \emptyset$, $\gamma = \emptyset$, and $w = |x|^{\downarrow}_\beta/\bar{\lambda}$. Since $\langle e_\beta\, , \, w\rangle < k - \bar{k}_0$, from the characterization Case 4 and Lemma 2.6, we know that $\overline{\mathcal{D}} = \{d \in \mathbb{R}^n \mid \langle e_\alpha, \hat{d}_\alpha\rangle + \langle w, \hat{d}_\beta\rangle = 0,\ \hat{d}_{\beta_1} \geq 0,\ \hat{d}_{\beta_2} = 0,\ \hat{d}_{\beta_3} = 0\}$ with $\beta_1 \neq \emptyset$. By following the similar argument in Case 1 and applying Theorem 4.1 with the BPS algorithms discussed in Remark 3.1, we can complete the proof for this case.

Case 4: $\|x\|_{(k)} > r$, $k < \bar{k}_1$, $|\bar{x}|^{\downarrow}_k > 0$, *and either* $|\bar{x}|^{\downarrow}_k = |x|^{\downarrow}_{\bar{k}_0+1} - \bar{\lambda}$ *or* $|\bar{x}|^{\downarrow}_k = |x|^{\downarrow}_{\bar{k}_1}$. In this case, $\overline{\mathcal{D}}$ is given by the characterization Case 5. Since $w = (|x|^{\downarrow}_\beta - |\bar{x}|^{\downarrow}_k e_\beta)/\bar{\lambda}$, we know that $\beta_1 \cup \beta_3 \neq \emptyset$. Moreover, $k < \bar{k}_1$ implies that $\beta_1 \neq \beta$, while $w \neq 0$ implies that $\beta_3 \neq \beta$. The proof for this case can be completed by following the similar argument in Case 1 and applying Theorem 4.1 with Algorithm 5. $\square$

4.3. Projections over four basic polyhedral convex cones. In this subsection, we will focus on computing the metric projectors over the polyhedral convex cones $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, and $\mathcal{D}_4$ (defined in (43), (44), (45), and (46)), which are closely related to the critical cone of $\mathcal{B}^r_{(k)}$. As can be seen from subsection 4.2, the results obtained in this subsection are crucial and indispensable for studying the differential properties of the metric projector $\Pi_{\mathcal{B}^r_{(k)}}$. Since the developments of these results are highly technical, this subsection could be skipped on the first reading.

To propose our algorithms to compute the metric projectors over these four polyhedral convex cones, we need the following two subroutines. In the algorithms for computing the metric projectors over $\mathcal{D}_1$ and $\mathcal{D}_3$, `Subroutine 1` and `Subroutine 2`

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

aim to check whether $\bar{z}_q = 0$ and $\bar{z}_q > 0$, respectively, where $\bar{z}_q$ is the qth entry of the variable z of the optimal solution to problem (51) or problem (63). In addition, `Subroutine 2` also serves as a main step in the algorithms for computing the metric projectors over $\mathcal{D}_2$ and $\mathcal{D}_4$.

SUBROUTINE 1. $\texttt{function}\ (\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_3(u, v, \tilde{v}, q_0, \tilde{s}, \texttt{opt})$.

- 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* $0 \le q_0 \le q-1$, $\tilde{s} \in \mathbb{R}^{p+1}$, *and* $\texttt{opt} = 1$ *or* 0.
- Main Step: *Set* $(\bar{y}, \bar{z}, \texttt{flag}) = (u, v, 0)$. *Compute*

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

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*

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

SUBROUTINE 2. $\texttt{function}\ (\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, v, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, \texttt{opt})$.

- Input: $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$, *$\tilde{v}^-$ and $\tilde{v}^+$ are tuples of length* $q+1$ *and* $p-q+2$, *q_0 and q_1 are integers satisfying* $0 \le q_0 < q \le q_1 \le p$, $\tilde{s} \in \mathbb{R}^{p+1}$, *and* $\texttt{opt} = 1$ *or* 0.
- Main Step: *Set* $(\bar{y}, \bar{z}, \texttt{flag}) = (u, v, 0)$. *Compute* $\rho = (q_1 - q_0)(\|e\|^2 + q_0) + (q - q_0)^2$ *and*

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

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*

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

In the rest of this subsection, we present algorithms of low complexity for computing the metric projectors over the polyhedral convex cones $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$, and $\mathcal{D}_4$. Since all the proofs are essentially similar to that of Proposition 4.1, we omit them due to the space limitation.

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

$$\begin{array}{rl} \min & \dfrac{1}{2}\big(\|y - u\|^2 + \|z - v\|^2\big) \\ \text{s.t.} & \langle e\,,\, y\rangle + \|z\|_{(q)} \le 0\,. \end{array} \tag{48}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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 let ${\pi_1}^{-1}$ 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

$$\tilde{s}_0 := 0\,, \quad \tilde{s}_j := \sum_{i=1}^{j} |v|^{\downarrow}_i\,, \quad j = 1,\ldots,p\,. \tag{49}$$

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

$$\begin{cases} \tilde{v}^-_0 := +\infty\,, & \tilde{v}^-_i := |v|^{\downarrow}_i\,,\ i = 1,\ldots,q\,, \\ \tilde{v}^+_{p+1} := 0\,, & \tilde{v}^+_i := |v|^{\downarrow}_i\,,\ i = q,\ldots,p\,. \end{cases} \tag{50}$$

According to Lemma 2.4, problem (48) can be equivalently reformulated as

$$\begin{array}{ll} \min & \dfrac{1}{2}\big(\|y-u\|^2 + \|z - |v|^{\downarrow}\|^2\big) \\ \text{s.t.} & \langle e\,,\, y\rangle + \|z\|_{(q)} \le 0 \end{array} \tag{51}$$

in the sense that $(\bar{y},\bar{z}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (51) (note that $\bar{z} = |\bar{z}|^{\downarrow} \ge 0$ in this case) if and only if $(\bar{y}, \operatorname{sgn}(v)\circ \bar{z}_{{\pi_1}^{-1}})$ solves problem (48). The KKT conditions for (51) have the following form:

$$\begin{cases} 0 = y - u + \lambda e\,, \\ 0 = z - |v|^{\downarrow} + \lambda\mu \ \text{ for some } \mu \in \partial\|z\|_{(q)}\,, \\ 0 \le \big(-\langle e\,,\, y\rangle - \|z\|_{(q)}\big) \perp \lambda \ge 0\,, \end{cases} \tag{52}$$

where λ is the corresponding Lagrange multiplier. It is not difficult to see that the constraint of problem (51) can be equivalently replaced by finitely many linear constraints. Then from [35, Corollary 28.3.1] and the fact that the optimal solution to problem (51) is unique, we know that the KKT system (52) has a unique solution $(\bar{y},\bar{z},\bar{\lambda})$ and $(\bar{y},\bar{z})$ is also the unique optimal solution to problem (51). If $\langle e\,,\, u\rangle + \|v\|_{(q)} \le 0$, then $(\bar{y},\bar{z},\bar{\lambda}) = (u, |v|^{\downarrow}, 0)$. Otherwise, i.e., if $\langle e\,,\, u\rangle + \|v\|_{(q)} > 0$, we have that

$$\bar{\lambda} > 0 \quad \text{and} \quad \langle e\,,\, \bar{y}\rangle + \|\bar{z}\|_{(q)} = 0\,. \tag{53}$$

By using Lemma 2.3, (53), and the fact that $\bar{z} = |\bar{z}|^{\downarrow}$, we can solve the KKT system (52) to obtain the following two lemmas.

LEMMA 4.1. *Assume that* $(u,v) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given, where* $\langle e\,,\, u\rangle + \|v\|_{(q)} > 0$. *Then,* $(\bar{y},\bar{z},\bar{\lambda}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}_+$ *solves the KKT system* (52) *with* $\bar{z}_q = 0$ *if and only if* $(\bar{y},\bar{z},\texttt{flag}) = \texttt{S}_1(u, |v|^{\downarrow}, \tilde{v}^-, \bar{q}_0, \tilde{s}, 1)$ *with* $\texttt{flag} = 1$ *for some integer* $\bar{q}_0$ *satisfying* $0 \le \bar{q}_0 \le q-1$.

LEMMA 4.2. *Assume that* $(u,v) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given, where* $\langle e\,,\, u\rangle + \|v\|_{(q)} > 0$. *Then,* $(\bar{y},\bar{z},\bar{\lambda}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}_+$ *solves the KKT system* (52) *with* $\bar{z}_q > 0$ *if and only if* $(\bar{y},\bar{z},\texttt{flag}) = \texttt{S}_2(u, |v|^{\downarrow}, \tilde{v}^-, \tilde{v}^+, \bar{q}_0, \bar{q}_1, \tilde{s}, 1)$ *with* $\texttt{flag} = 1$ *for some integers* $\bar{q}_0$ *and* $\bar{q}_1$ *satisfying* $0 \le \bar{q}_0 \le q-1$ *and* $q \le \bar{q}_1 \le p$.

By combining Lemmas 4.1 and 4.2, we have the following algorithm for computing the projector over $\mathcal{D}_1$ and the following proposition for justifying the proposed algorithm.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 2. Computing $\Pi_{\mathcal{D}_1}(u,v)$.

Step 0 (Preprocessing). *If* $\langle e\,,\,u\rangle + \|v\|_{(q)} \le 0$, *output* $\Pi_{\mathcal{D}_1}(u,v) = (u,v)$, *and stop. Otherwise, sort* $|v|$ *to obtain* $|v|^{\downarrow}$, *precompute* $\tilde{s}$ *by* (49), *evaluate* $\tilde{v}^-$ *and* $\tilde{v}^+$ *by* (50), *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* $(\bar{y},\bar{z},\mathtt{flag}) = \mathtt{S}_3(u,|v|^{\downarrow},\tilde{v}^-,q_0,\tilde{s},1)$. *If* $\mathtt{flag}=1$, *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* $(\bar{y},\bar{z},\mathtt{flag}) = \mathtt{S}_4(u,|v|^{\downarrow},\tilde{v}^-,\tilde{v}^+,q_0,q_1,\tilde{s},1)$. *If* $\mathtt{flag}=1$, *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.*

PROPOSITION 4.2. *Assume that* $(u,v)\in\mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given. Then the metric projection* $\Pi_{\mathcal{D}_1}(u,v)$ *of* (u,v) *onto* $\mathcal{D}_1$ *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 initial sorting cost is* $O(p\log p)$, *the searching cost is* $O(q(p-q+1))$, *and the final evaluation cost is* $O(m)$.

4.3.2. Projection over $\boldsymbol{\mathcal{D}_2}$. For any given $(u,v)\in\mathbb{R}^{m-p}\times\mathbb{R}^p$, $\Pi_{\mathcal{D}_2}(u,v)$ is the unique optimal solution to the following convex optimization problem:

$$\begin{array}{ll} \min & \dfrac{1}{2}\big(\|y-u\|^2 + \|z-v\|^2\big) \\ \text{s.t.} & \langle e\,,\,y\rangle + s_{(q)}(z) \le 0\,. \end{array} \tag{54}$$

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 let $\pi_2{}^{-1}$ 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

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

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

$$\begin{array}{ll} \min & \dfrac{1}{2}\big(\|y-u\|^2 + \|z-v^{\downarrow}\|^2\big) \\ \text{s.t.} & \langle e\,,\,y\rangle + s_{(q)}(z) \le 0 \end{array} \tag{57}$$

in the sense that $(\bar{y},\bar{z})\in\mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (57) (note that $\bar{z}=\bar{z}^{\downarrow}$ in this case) if and only if $(\bar{y},\bar{z}_{\pi_2{}^{-1}})$ solves problem (54). The KKT conditions for (57) are given as follows:

$$\left\{\begin{array}{l} 0 = y - u + \lambda e\,, \\ 0 = z - v^{\downarrow} + \lambda\mu \ \text{ for some } \mu\in\partial s_{(q)}(z)\,, \\ 0 \le \big(-\langle e\,,\,y\rangle - s_{(q)}(z)\big) \perp \lambda \ge 0\,, \end{array}\right. \tag{58}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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 [35, 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 $(\bar{y}, \bar{z}, \bar{\lambda})$ and $(\bar{y}, \bar{z})$ is also the unique optimal solution to problem (57). If $\langle e\,,\,u\rangle + s_{(q)}(v) \le 0$, then $(\bar{y}, \bar{z}, \bar{\lambda}) = (u, v^{\downarrow}, 0)$. Otherwise, i.e., if $\langle e\,,\,u\rangle + s_{(q)}(v) > 0$, we have that

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

By solving the KKT system (58) with Lemma 2.2, (59), and the fact that $\bar{z} = \bar{z}^{\downarrow}$, we obtain the following lemma.

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

According to Lemma 4.3, we propose the following algorithm for computing the projector over $\mathcal{D}_2$. This algorithm and its low computational cost are justified in the next proposition.

ALGORITHM 3. Computing $\Pi_{\mathcal{D}_2}(u, v)$.

Step 0 (Preprocessing). *If* $\langle e\,,\,u\rangle + s_{(q)}(v) \le 0$, *output* $\Pi_{\mathcal{D}_2}(u, v) = (u, v)$, *and stop. Otherwise, sort* v *to obtain* $v^{\downarrow}$, *precompute* $\tilde{s}$ *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* $(\bar{y}, \bar{z}, \texttt{flag}) = \texttt{S}_4(u, v^{\downarrow}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 1)$. *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.*

PROPOSITION 4.3. *Assume that* $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ *is given. Then the metric projection* $\Pi_{\mathcal{D}_2}(u, v)$ *of* (u, v) *onto* $\mathcal{D}_2$ *can be computed by Algorithm* 3. *Moreover, the computational cost of Algorithm* 3 *is* $O(p\log p + q(p-q+1) + m)$, *where the initial sorting cost is* $O(p\log p)$, *the searching cost is* $O(q(p-q+1))$, *and the final evaluation cost is* $O(m)$.

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

$$\begin{array}{ll} \min & \dfrac{1}{2}\big(\|y-u\|^2 + \|z-v\|^2\big) \\ \text{s.t.} & \|z\|_{(q)} \le \langle w\,,\,z\rangle\,, \\ & \langle e\,,\,y\rangle + \langle w\,,\,z\rangle = 0\,. \end{array} \tag{60}$$

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

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

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

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

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

in the sense that $(\bar{y},\bar{z}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ solves problem (63) (note that $\bar{z} = |\bar{z}|^{\downarrow}$ in this case) if and only if $(\bar{y}, \mathrm{psgn}(v)\circ \bar{z}_{\pi_3^{-1}})$ 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:

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

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 ith standard basis whose entries are all 0 except its ith entry, which is 1. Since problem (63) has only finitely many linear constraints, by using [35, 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 $(\bar{y},\bar{z},\bar{\xi},\bar{\lambda})$ and $(\bar{y},\bar{z})$ is the unique optimal solution to problem (63). Then by exploiting the structure of w and $\bar{z}$, we can solve the KKT system (64) to obtain the following two lemmas.

LEMMA 4.4. *Assume that* $(u,v) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given. Then,* $(\bar{y},\bar{z},\bar{\xi},\bar{\lambda}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}$ *solves the KKT system* (64) *with* $\bar{z}_q = 0$ *if and only if* $(\bar{y},\bar{z},\texttt{flag}) = \mathtt{S}_1(u,\hat{v},\tilde{v}^-,\bar{q}_0,\tilde{s},0)$ *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* $(u,v) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ *is given. Then,* $(\bar{y},\bar{z},\bar{\xi},\bar{\lambda}) \in \mathbb{R}^{m-p}\times\mathbb{R}^p\times\mathbb{R}^p\times\mathbb{R}$ *solves the KKT system* (64) *with* $\bar{z}_q > 0$ *if and only if* $(\bar{y},\bar{z},\texttt{flag}) = \mathtt{S}_2(\hat{v},\tilde{v}^-,\tilde{v}^+,\bar{q}_0,\bar{q}_1,\tilde{s},0)$ *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* $\max\{q, |\beta_1|+|\beta_2|\} \leq \bar{q}_1 \leq p$.

By combining Lemmas 4.4 and 4.5, we have the following algorithm for computing the projector over $\mathcal{D}_3$ and the following proposition for justifying the proposed algorithm.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

ALGORITHM 4. Computing $\Pi_{\mathcal{D}_3}(u,v)$ for the case that $\langle e_\beta\,,\,w\rangle = q$.

Step 0 (Preprocessing). *Calculate* $\hat{v} = \big(\mathrm{psgn}(v)\circ v\big)_{\pi_3}$, *precompute* $\tilde{s}$ *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* $(\bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_3(u, \hat{v}, \tilde{v}^-, q_0, \tilde{s}, 0)$. *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 2` *with* $(\bar{y}, \bar{z}, \texttt{flag}) = \mathtt{S}_4(u, \hat{v}, \tilde{v}^-, \tilde{v}^+, q_0, q_1, \tilde{s}, 0)$. *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.*

PROPOSITION 4.4. *Let* $(u,v) \in \mathbb{R}^{m-p}\times\mathbb{R}^p$ *be given. Then the metric projection* $\Pi_{\mathcal{D}_3}(u,v)$ *of* (u,v) *onto* $\mathcal{D}_3$ *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 initial sorting cost is* $O(|\beta_1|\log|\beta_1| + |\beta_3|\log|\beta_3|)$, *the searching cost is* $O(q(p-q+1))$, *and the final evaluation cost is* $O(m)$.

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

$$
\begin{aligned}
\min\quad & \frac{1}{2}\big(\|y-u\|^2 + \|z-v\|^2\big) \\
\text{s.t.}\quad & s_{(q)}(z) \le \langle w\,,\, z\rangle\,, \\
& \langle e\,,\, y\rangle + \langle w\,,\, z\rangle = 0\,.
\end{aligned}
\tag{65}
$$

Due to the structure of w, 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 $\{1,\ldots,p\}$ 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 $(v_{\beta_1})^\downarrow = (v_{\pi_4})_{\beta_1}$, $v_{\beta_2} = (v_{\pi_4})_{\beta_2}$, and $(v_{\beta_3})^\downarrow = (v_{\pi_4})_{\beta_3}$, and let $\pi_4{}^{-1}$ 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

$$
\tilde{s}_0 := 0\,,\quad \tilde{s}_j := \sum_{i=1}^{j}\hat{v}_i\,,\quad j = 1,\ldots,p\,.
\tag{66}
$$

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

$$
\left\{
\begin{array}{ll}
\tilde{v}^-_0 := +\infty, \tilde{v}^-_{|\beta_1|+1} = \tilde{v}^-_q := -\infty, & \tilde{v}^-_i := \hat{v}_i, i = 1,\ldots,q, i\ne|\beta_1|+1 \text{ and } q, \\
\tilde{v}^+_q = \tilde{v}^+_{|\beta_1|+|\beta_2|} := +\infty, \tilde{v}^+_{p+1} := -\infty, & \tilde{v}^+_i := \hat{v}_i, i = q,\ldots,p, i\ne q \text{ and } |\beta_1|+|\beta_2|.
\end{array}
\right.
\tag{67}
$$

Then by using Lemma 2.4, Lemma 2.5, and the structure of w, one can equivalently reformulate problem (65) as

$$
\begin{aligned}
\min\quad & \frac{1}{2}\big(\|y-u\|^2 + \|z-\hat{v}\|^2\big) \\
\text{s.t.}\quad & z_i \ge z_q\,,\ i = 1,\ldots,|\beta_1|\,, \\
& z_i = z_q\,,\ i = |\beta_1|+1,\ldots,|\beta_1|+|\beta_2|\,,\ i\ne q\,, \\
& z_i \le z_q\,,\ i = |\beta_1|+|\beta_2|+1,\ldots,p\,, \\
& \langle e\,,\, y\rangle + \langle w\,,\, z\rangle = 0
\end{aligned}
\tag{68}
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

in the sense that $(\bar y, \bar z) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ solves problem (68) (note that $\bar z = \bar z^{\downarrow}$ in this case) if and only if $(\bar y, \bar z_{\pi_4^{-1}})$ solves problem (65). Note that $|\beta_1| \le q \le |\beta_1| + |\beta_2|$, and $q = |\beta_1| + |\beta_2|$ if and only if $q = |\beta_1|$, i.e., $\beta_2 = \emptyset$. The KKT conditions for (68) are given as follows:

$$
\left\{
\begin{array}{l}
0 = y - u + \lambda e\,, \\
0 = z - \hat v - \displaystyle\sum_{i=1}^{q-1} \xi_i (e^i - e^q) - \sum_{i=q+1}^{p} \xi_i (e^q - e^i) + \lambda\, w\,, \\
0 \le (z_i - z_q) \perp \xi_i \ge 0\,,\ i \in \beta_1\,, \\
0 \le (z_q - z_i) \perp \xi_i \ge 0\,,\ i \in \beta_3\,, \\
z_i = z_q\,,\ i \in \beta_2 \backslash \{q\}\,, \quad \langle e\,,\, y\rangle + \langle w\,,\, z\rangle = 0\,,
\end{array}
\right. \tag{69}
$$

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 ith standard basis. Since problem (68) has only finitely many linear constraints, by using [35, Corollary 28.3.1] and the fact that problem (68) has a unique solution, we know that the KKT system (69) has a unique solution $(\bar y, \bar z, \bar\xi, \bar\lambda)$ and $(\bar y, \bar z)$ is the unique optimal solution to problem (68). Then by taking the structure of w and $\bar z$ into account, we can solve the KKT system (69) to obtain the following lemma.

LEMMA 4.6. *Let $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ be given. Then, $(\bar y, \bar z, \bar\xi, \bar\lambda) \in \mathbb{R}^{m-p} \times \mathbb{R}^p \times \mathbb{R}^{p-1} \times \mathbb{R}$ solves the KKT system* (69) *if and only if $(\bar y, \bar z, \mathtt{flag}) = \mathtt{S}_2(u, \hat v, \tilde v^-, \tilde v^+, \bar q_0, \bar q_1, \tilde s, 0)$ with $\mathtt{flag} = 1$ for some integers $\bar q_0$ and $\bar q_1$ satisfying $0 \le \bar q_0 \le \min\{q-1, |\beta_1|\}$ and $\max\{q, |\beta_1| + |\beta_2|\} \le \bar q_1 \le p$.*

According to Lemma 4.6, we propose the following algorithm for computing the projector over $\mathcal{D}_4$. This algorithm and its low computational cost are justified in the next proposition.

ALGORITHM 5. Computing $\Pi_{\mathcal{D}_4}(u, v)$.

Step 0 (Preprocessing). *Calculate $\hat v = v_{\pi_4}$, precompute $\tilde s$ by* (66)*, evaluate $\tilde v^-$ and $\tilde v^+$ by* (67)*, 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 $(\bar y, \bar z, \mathtt{flag}) = \mathtt{S}_4(\eta, u, \hat v, \tilde v^-, \tilde v^+, q_0, q_1, \tilde s, 0)$. If $\mathtt{flag} = 1$, 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.*

PROPOSITION 4.5. *Let $(u, v) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$ be given. Then the metric projection $\Pi_{\mathcal{D}_4}(u, v)$ of (u, v) onto $\mathcal{D}_4$ can be computed by Algorithm* 5*. Moreover, the computational cost of Algorithm* 5 *is $O(|\beta_1| \log|\beta_1| + |\beta_3| \log|\beta_3| + q(p-q+1) + m)$, where the initial sorting cost is $O(|\beta_1| \log|\beta_1| + |\beta_3| \log|\beta_3|)$, the searching cost is $O(q(p-q+1))$, and the final evaluation cost is $O(m)$.*

5. 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 and the indicator function of the vector k-norm ball. This research constitutes the backbone for designing efficient algorithms to solve MOPs involving the Ky Fan k-norm and also for studying the sensitivity and stability analysis of these problems. The work in this paper can be extended to many other situations. Below we briefly list some of them.

(i) All the corresponding properties for the metric projector over the vector k-norm epigraph cone can be obtained in a similar way to those for the metric projector

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

over the vector k-norm ball. One may refer to the first version of this paper [42] for more details.

(ii) Consider the function defined by $g(z) := s_{(k)}(z)$ for $z \in \mathbb{R}^n$. It is easy to verify that the Fenchel conjugate g^* is the indicator function of $\phi_{n,k}$. Then, the similar results on the proximal mappings associated with the function g and the indicator functions of its level sets can be derived in a similar but simpler way to those obtained in this paper.

(iii) Define $g(z) := \|z\|_{(k)} + \delta_{\mathbb{R}^n_+}(z)$ for $z \in \mathbb{R}^n$. Simple calculations show that the Fenchel conjugate g^* is the indicator function of $\{z \in \mathbb{R}^n \,|\, \|z_+\|_{(k)^*} \le 1\}$, where $z_+ := \Pi_{\mathbb{R}^n_+}(z)$. Consequently, the Moreau–Yosida regularization of the function g and the indicator functions of its level sets at any given $x \in \mathbb{R}^n$ can be obtained by considering the counterparts of the related functions defined in $\mathbb{R}^{|\alpha|}$ at x_α, where $\alpha := \{i \in [n] \,|\, x_i \ge 0\}$.

(iv) Consider the weighted vector k-norm function defined by $\|z\|^{\omega}_{(k)} := \sum_{i=1}^k \omega_i |z|^{\downarrow}_i$ for $z \in \mathbb{R}^n$, where $\omega = (\omega_1, \ldots, \omega_k) \in \mathbb{R}^k$, 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 proximal mappings associated with the weighted vector k-norm related functions are much more complicated, and further study 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 these proximal mappings have similar properties to those obtained in this paper. Also note that for this simple case, the weighted vector k-norm and its dual norm were also considered in robust optimization [4]. Furthermore, it is worth mentioning that the more general setting, where the standard inner product between two vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ is replaced by the weighted inner product in the sense that $\langle x, y\rangle_Q := \langle x, Qy\rangle$ for some symmetric and positive definite matrix Q, also deserves further study.

Acknowledgments. The authors would like to thank the two anonymous referees and the associate editor Professor Jong-Shi Pang for their valuable comments and constructive suggestions, which have greatly helped to improve the quality and presentation of this paper.

REFERENCES

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

[2] A. Argyriou, R. Foygel, and N. Srebro, *Sparse prediction with the k-support norm*, Adv. Neural Inform. Process. Syst., 25 (2012), pp. 1466–1474.

[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, SIAM, Philadelphia, 2001.

[4] D. Bertsimas, D. Pachamanova, and M. Sim, *Robust linear optimization under general norms*, Oper. Res. Lett., 32 (2004), pp. 510–516.

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

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

[7] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Fastest mixing Markov chain on graphs with symmetries*, SIAM J. Optim., 20 (2009), pp. 792–819.

[8] P. Brucker, *An $O(n)$ algorithm for quadratic knapsack problems*, Oper. Res. Lett., 3 (1984), pp. 163–166.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[9] P. H. Calamai and J. J. Moré, *Quasi-Newton updates with bounds*, SIAM J. Numer. Anal., 24 (1987), pp. 1434–1441.

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

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

[12] C. Ding, *An Introduction to a Class of Matrix Optimization Problems*, Ph.D. Thesis, Department of Mathematics, National University of Singapore, 2012; available online at http://www.math.nus.edu.sg/∼matsundf/DingChao_Thesis_final.pdf.

[13] C. Ding, D. F. Sun, and K. C. Toh, *An introduction to a class of matrix cone programming*, Math. Program., to appear; preprint available online at http://www.math.nus.edu.sg/∼matsundf/IntroductionMCP_revised_Mar_26_12.pdf.

[14] Y. Gao, *Structured Low Rank Matrix Optimization Problems: A Penalized Approach*, Ph.D. Thesis, Department of Mathematics, National University of Singapore, 2010; available online at http://www.math.nus.edu.sg/∼matsundf/main_gy.pdf.

[15] Y. Gao and D. F. Sun, *Calibrating least squares semidefinite programming with equality and inequality constraints*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1432–1457.

[16] Y. Gao and D. F. Sun, *A Majorized Penalty Approach for Calibrating Rank Constrained Correlation Matrix Problems*, preprint; available online at http://www.math.nus.edu.sg/∼matsundf/MajorPen_May5.pdf.

[17] A. Haraux, *How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities*, J. Math. Soc. Japan, 29 (1977), pp. 615–631.

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

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

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

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

[22] A. S. Lewis, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal., 2 (1995), pp. 173–183.

[23] A. S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res., 21 (1996), pp. 576–588.

[24] A. S. Lewis and H. S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 368–386.

[25] 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 $\mathbb{R}^n$*, J. Optim. Theory Appl., 117 (2003), pp. 553–574.

[26] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[27] J. von Neumann, *Some matrix-inequalities and metrization of matric-space*, in *Collected Works, Volume* IV, Pergamon, Oxford, 1962, pp. 205–218.

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

[29] J. S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 149–160.

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

[31] G. Pataki, *On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues*, Math. Oper. Res., 23 (1998), pp. 339–358.

[32] H. D. Qi and D. F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 360–385.

[33] H. D. Qi and X. Q. Yang, *Semismoothness of spectral functions*, SIAM J. Matrix Anal. Appl., 25 (2004), pp. 766–783.

[34] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

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

[36] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1998.

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

[38] D. F. Sun and J. Sun, *Semismooth matrix-valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

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

[40] G. A. Watson, *Linear best approximation using a class of polyhedral norms*, Numer. Algorithms, 2 (1992), pp. 321–336.

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

[42] B. Wu, C. Ding, D. F. Sun, and K. C. Toh, *On the Moreau-Yosida Regularization of the Vector k-Norm Related Functions*, preprint; available online at http://www.math.nus.edu.sg/~matsundf/k-norm_08Mar11.pdf.

[43] X. Y. Zhao, *A Semismooth Newton-CG Augmented Lagrangian Method for Large Scale Linear and Convex Quadratic SDPs*, Ph.D. Thesis, Department of Mathematics, National University of Singapore, 2009; available online at http://www.math.nus.edu.sg/~matsundf/main_xyz.pdf.

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

Downloaded 09/01/14 to 137.132.123.69. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.