On max-k-sums

Michael J. Todd *

October 2, 2017

Abstract

The max-k-sum of a set of real scalars is the maximum sum of a subset of size k, or alternatively the sum of the k largest elements. We study two extensions: First, we show how to obtain smooth approximations to functions that are pointwise max-k-sums of smooth functions. Second, we discuss how the max-k-sum can be defined on vectors in a finite-dimensional real vector space ordered by a closed convex cone.

1 Introduction

Given $y_1, \ldots, y_n \in \mathbf{R}$, we define their max-$k$-sum to be

$$M^k(y) := \max_{|K|=k} \sum_{i \in K} y_i = \sum_{j=1}^{k} y_{[j]}, \tag{1}$$

where $y_{[1]}, \ldots, y_{[n]}$ are the y_i's listed in nonincreasing order. We similarly define their min-k-sum to be

$$m^k(y) := \min_{|K|=k} \sum_{i \in K} y_i = \sum_{j=n-k+1}^{n} y_{[j]}. \tag{2}$$

(Throughout, we use both sub- and superscripts for indexing. Powers are denoted by placing their arguments in parentheses.)

Clearly, these generalize the maximum and minimum of a set of numbers ($k = 1$). They arise for instance in limiting a quantile of a distribution as in the conditional value at risk (Rockafellar and Uryasev [17]), when the distribution is modelled by equally likely scenarios, as well as in certain penalties for peak demands in electricity modelling (Zakeri, Craigie, Philpott, and Todd [20]). Max-k-sums are also special cases (when y is nonnegative) of *OWL* (ordered weighted l_1-) norms, which have recently been introduced in Bogdan et al. [3] and Zeng and Figueiredo [21, 22] as regularizers for statistical learning problems. Conversely, OWL norms can be seen as nonnegative combinations of max-k-sums. The sum of the k largest eigenvalues of a symmetric matrix arises in applications and also as a tool in the variational study of the kth largest eigenvalue: see for example Hiriart-Urruty and Ye, [7].

*School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA. E-mail mjt7@cornell.edu.

We are interested in extending these notions in two ways. First, we might define a function by taking pointwise the max-k-sum or min-k-sum of a set of smooth functions. We define the max-k-sum and min-k-sum of functions $f_1, \ldots, f_n$ on $\mathbf{R}^d$ to be

$$F^k(t) := M^k(f_1(t), \ldots, f_n(t)) \tag{3}$$

and

$$f^k(t) := m^k(f_1(t), \ldots, f_n(t)). \tag{4}$$

It is immediate that, even if the f_i's are smooth, these derived functions may not be. (The cover of this journal depicts the max-1-sum of two linear functions.) It is well known that, if the f_i's are convex, so is F^k, while if they are concave, so is f^k. We are interested in finding smooth approximations to these functions, with predetermined accuracies. We would also like these approximations to inherit the above convexity/concavity properties, and also the sign reversal property:

$$\text{if } (g_1, \ldots, g_n) = -(f_1, \ldots, f_n), \quad f^k(t) = -G^k(t)$$

and the summability property:

$$F^k(t) + f^{n-k}(t) = \sum_{i=1}^{n} f_i(t)$$

for any $1 \leq k \leq n-1$. We can also extend this to $k = 0$ and $k = n$ by naturally defining max- and min-0-sums to be identically 0. (Henceforth, we use $\sum_i$ to denote summation over the range $i = 1$ to n.) There are also natural invariance properties, under translation and positive scalar multiplication, which we would also like to be preserved.

Second, we would like to define max- and min-k-sums for sets $y_1, \ldots, y_n$ of vectors in a finite-dimensional real vector space which is partially ordered by a closed convex cone, for example the space of real symmetric or complex Hermitian matrices with the Löwner partial order: $X \succeq Y$ iff $X - Y$ is positive semidefinite. We discuss various approaches to extend these notions, either directly or in their smoothed versions as above, and ask whether these extended definitions satisfy the same properties as in the real scalar case.

Clearly, the max- and min-k-sum of a set of real scalars can be computed in $O(n \ln n)$ time by sorting the entries. We are interested in the complexity of computing the smoothed versions or the extensions to various partially ordered vector spaces.

Our original motivation was in deriving efficient self-concordant barriers for interior-point methods. It is well-known that, if each f_i is a ν_i-self-concordant barrier for the closed convex set $C_i \subseteq \mathbf{R}^d$, and $C := \cap_i C_i$ has a nonempty interior, then $\sum_i f_i$ is a ν-self-concordant barrier for C, with $\nu = \sum_i \nu_i$. On the other hand, there exists a self-concordant barrier for C with parameter $O(d)$. (See Nesterov and Nemirovskii [14, 12].) Our hope was that a suitable smoothing of the max-d-sum of the f_i's (where all ν_i's are $O(1)$) would be such a self-concordant barrier. Unfortunately, the smoothings we produce do not seem to be self-concordant or to deal gracefully with replication of the C_i's (Nemirovskii [15]).

In the next section, we consider smoothing the max-k-sum using randomization. Section 3 addresses an alternative smoothing technique that perturbs an optimization representation of the sum. In Section 4 we discuss the extensions of max-k- and min-k-sums to vector spaces ordered by a closed convex cone, building on the constructions in Sections 2 and 3. Nice properties hold for symmetric cones, for example the second-order cone

or the cone of positive semidefinite matrices in the space of real symmetric or complex Hermitian matrices.

Henceforth, to simplify the notation, we consider smoothing or extending the function taking y to $M^k(y)$ or $m^k(y)$; composing the result with the function from $t \in \mathbf{R}^d$ to $f(t) := (f_1(t), \ldots, f_n(t))$ will then deal with the general case. Our results are generallly stated for M^k and m^k, and their translations to F^k and f^k are immediate and will usually not be explicitly stated.

2 Smoothing using randomization

A natural way to smooth a nonsmooth function $h : \mathbf{R}^d \to \mathbf{R}$, such as F^k above, is to take its convolution with a smooth function, i.e., to consider

$$\tilde{h}(t) := \int_{\mathbf{R}^d} h(t-s)\phi(s)ds,$$

where ϕ is a smooth probability density function concentrated around 0. However, this smoothing may be hard to compute, and in the barrier case (where h is $+\infty$ outside some open convex set C) it does not have the same effective domain (the set where its value is less than $+\infty$) as the generating nonsmooth function. We can view this approach as "smearing" in the domain of h or of the f_i's. The first method we consider instead "smears" in the range of these functions, which yields a computable function that maintains many properties of the max-k-sum.

(When we deal with M^k, there is no distinction between smearing in the domain and smearing in the range, which is why the discussion above was phrased for max-k-sums of functions.)

We add noise to each component of y, take the max-k-sum, and then take expectations. We also compensate for the expectation of the noise. Hence let $\xi_1, \ldots, \xi_n$ be independent identically distributed random variables distributed like the continuous random variable Ξ, and define

$$\bar{M}^k(y) := E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i\in K} (y_i - \xi_i) + kE\Xi. \tag{5}$$

For our smooth version of the min-k-sum, we correspondingly define

$$\bar{m}^k(y) := E_{\xi_1,\ldots,\xi_n} \min_{|K|=k} \sum_{i\in K} (y_i - \xi_i) + kE\Xi. \tag{6}$$

We record some simple properties of these approximations.

Proposition 1 *We have*

(a) (0- and n-consistency) $\bar{M}^0(y) = \bar{m}^0(y) = 0$, *and* $\bar{M}^n(y) = \bar{m}^n(y) = \sum_i y_i$.

(b) (summability) For $0 \le k \le n$, $\bar{M}^k(y) + \bar{m}^{n-k}(y) = \sum_i y_i$.

(c) (translation invariance) For $0 \le k \le n$ *and* $\eta \in \mathbf{R}$, *with* $\mathbf{1} := (1, \ldots, 1)$ *we have* $\bar{M}^k(y + \eta\mathbf{1}) = \bar{M}^k(y) + k\eta$, $\bar{m}^k(y + \eta\mathbf{1}) = \bar{m}^k(y) + k\eta$.

(d) (approximation)

$$M^k(y) \le \bar{M}^k(y) \le M^k(y) + \bar{M}^k(0) \le M^k(y) + \min(k\bar{M}^1(0), -(n-k)\bar{m}^1(0))$$

and

$$m^k(y) \ge \bar{m}^k(y) \ge m^k(y) + \bar{m}^k(0) \ge m^k(y) - \min((n-k)\bar{M}^1(0), -k\bar{m}^1(0)).$$

Proof: (a) follows since there is only one subset of cardinality 0, the empty set, and only one of cardinality n, the full set.

For (b), note that

$$\begin{aligned} E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i\in K}(y_i-\xi_i) + E_{\xi_1,\ldots,\xi_n} \min_{|L|=n-k} \sum_{i\in L}(y_i-\xi_i) &= \\ E_{\xi_1,\ldots,\xi_n}[\max_{|K|=k} \sum_{i\in K}(y_i-\xi_i) + \min_{|L|=n-k} \sum_{i\in L}(y_i-\xi_i)] &= \\ E_{\xi_1,\ldots,\xi_n} \sum_i (y_i-\xi_i) &= \sum_i y_i - nE\Xi. \end{aligned}$$

(c) also follows from the linearity of the expectation, since after translation, each sum of k entries of $y_i - \xi_i$ is translated by $k\eta$.

Finally, let $\hat{K}$ be the subset attaining the maximum in $M^k(y)$. Then for any realization of $(\xi_1,\ldots,\xi_n)$, we can choose $K=\hat{K}$, giving a sum of $M^k(y) - \sum_{i\in\hat{K}} \xi_i$, with expectation $M^k(y) - kE\xi$. The definition of $\bar{M}^k(y)$ allows any choice of K to achieve the maximum, and so the left-hand inequality of the first part of (d) follows. Next,

$$\max_{|K|=k} \sum_{i\in K}(y_i-\xi_i) \le \max_{|K|=k} \sum_{i\in K} y_i + \max_{|K|=k} \sum_{i\in K}(-\xi_i) \le M^k(y) + k \max_i(-\xi_i),$$

and this yields $\bar{M}^k(y) \le M^k(y) + \bar{M}^k(0) \le M^k(y) + k\bar{M}^1(0)$. A similar argument holds for $\bar{m}^k(y)$, yielding $m^k(y) \ge \bar{m}^k(y) \ge m^k(y) + \bar{m}^k(0) \ge m^k(y) + k\bar{m}^1(0)$. Finally, using summability for $\bar{M}^k$ shows that $\bar{M}^k(0) = -\bar{m}^{n-k}(0) \le -(n-k)\bar{m}^1(0)$, and similarly $\bar{m}^k(0) = -\bar{M}^{n-k}(0) \ge -(n-k)\bar{M}^1(0)$, and thus we obtain (d).

□

While we have translation invariance for our smoothed max-k-sums, we do not have positive scaling invariance: in general, $\bar{M}^k(\alpha y) \ne \alpha\bar{M}^k(y)$ for positive α. The reason is that we need to scale the random variables as well as y. If we show the dependence of $\bar{M}$ on the random variables, and write $\bar{M}^k(y;\Xi)$ when the ξ's are independently and identically distributed like Ξ, and similarly for $\bar{m}^k$, we have immediately

$$\bar{M}^k(\alpha y;\alpha\Xi) = \alpha\bar{M}^k(y,\Xi) \text{ and } \bar{m}^k(\alpha y;\alpha\Xi) = \alpha\bar{m}^k(y;\Xi) \tag{7}$$

for positive scalars α. For negative α, we obtain

$$\bar{M}^k(\alpha y;\alpha\Xi) = \alpha\bar{m}^k(y,\Xi) \text{ and } \bar{m}^k(\alpha y;\alpha\Xi) = \alpha\bar{M}^k(y;\Xi).$$

Setting $\alpha=-1$, we get the *sign reversal property*: $\bar{m}^k(y;\Xi) = -\bar{M}^k(-y;-\Xi)$.

In order to proceed, we need to choose convenient distributions for the ξ's. It turns out that, to obtain simple formulae for $\bar{M}^k$ for small k, it is suitable to choose Gumbel random variables, so that $P(\Xi > z) = \exp(-\exp(z))$, with probability density function $\exp(z-\exp(z))$ and expectation $E\Xi = -\gamma$, the negative of the Euler-Mascheroni constant. (If we are interested in $\bar{M}^k$ for k near n, we can choose negative Gumbel variables; then by the equations above with $\alpha=-1$, we can alternatively calculate $\bar{m}^k$ for Gumbel variables, or by summability $\bar{M}^{n-k}$ for Gumbel variables, which will again be simple to compute.)

Henceforth, we assume each ξ_i is an independent Gumbel random variable. Let q_k denote the expectation of the kth largest $y_i - \xi_i$.

Proposition 2

$$q_k = \sum_{|K|<k} (-1)^{k-|K|-1} \binom{n-|K|-1}{k-|K|-1} \ln\left(\sum_{i\notin K} \exp(y_i)\right) + \gamma$$

(here we employ the convention that $\binom{0}{0} = 1$*).*

Proof: We know that q_k is some $y_i - \xi_i$, with a set J of $k-1$ indices j with $y_j - \xi_j$ at least $y_i - \xi_i$, and the remaining indices h with $y_h - \xi_h$ at most $y_i - \xi_i$. Hence, summing over all possible i's and J's, we obtain

$$\begin{aligned} q_k &= \sum_i \sum_{J:|J|=k-1, i\notin J} \int_{\xi_i} \Pi_{j\in J} P(y_j - \xi_j \geq y_i - \xi_i) \\ &\quad \Pi_{h\notin J\cup\{i\}} P(y_h - \xi_h \leq y_i - \xi_i)(y_i - \xi_i)\exp(\xi_i - \exp(\xi_i))d\xi_i. \end{aligned}$$

Now each term in the first product is $1 - \exp(-\exp(y_j - y_i + \xi_i))$, and each term in the second product is $\exp(-\exp(y_h - y_i + \xi_i))$. Let us expand the first product, and consider the term where the "1" is taken for $j \in K$, with K a subset of J and hence of cardinallity less than k. For this K, and a particular i, there are $\binom{n-|K|-1}{k-|K|-1}$ choices for the remaining indices in J. For each such J, the summand is the same. Hence, noting that $\exp(-\exp(\xi_i))$ can be written as $\exp(-\exp(y_i - y_i + \xi_i))$, we see that

$$q_k = \sum_{K:|K|<k} \binom{n-|K|-1}{k-|K|-1} (-1)^{k-|K|-1} \sum_{i\notin K} \int_{\xi_i} (y_i - \xi_i)\exp(\xi_i)\exp(-\sum_{h\notin K} \exp(y_h - y_i + \xi_i))d\xi_i.$$

The last exponential above can be written as $\exp(-\exp(\xi_i - y_i)\exp(\ln\sum_{h\notin K}\exp(y_h))) = \exp(-\exp(\xi_i - z_i))$, where

$$z_i := y_i - \ln\left(\sum_{h\notin K} \exp(y_h)\right).$$

Thus the integral above can be simplified to

$$\int_{\xi_i} (y_i - \xi_i)\exp(\xi_i - \exp(\xi_i - z_i))d\xi_i = \int_{\xi_i} ([y_i - z_i] - [\xi_i - z_i])\exp(z_i)\exp(\xi_i - z_i - \exp(-\xi_i - z_i))d\xi_i.$$

Now $\exp(\xi_i - z_i - \exp(\xi_i - z_i))$ is the probability density of a translated Gumbel variable, so the integral evaluates to

$$(y_i - z_i)\exp(z_i) - (-\gamma)\exp(z_i) = \left[\ln\left(\sum_{h\notin K}\exp(y_h)\right) + \gamma\right]\exp(z_i).$$

From the definition of z_i, $\exp(z_i) = \exp(y_i)/\sum_{h\notin K}\exp(y_h)$, so $\sum_{i\notin K}\exp(z_i) = 1$. Thus the sum of the integrals is $\ln(\sum_{h\notin K}\exp(y_h)) + \gamma$, and we almost have the result of the proposition, except that γ appears in each term, rather than added on once. It therefore suffices to show that

$$\sum_{K:|K|<k} \binom{n-|K|-1}{k-|K|-1}(-1)^{k-|K|-1} = 1,$$

or, writing j for $|K|$ and p for $k-1$,

$$\sum_{j=0}^{p}(-1)^{p-j}\binom{n}{j}\binom{n-1-j}{p-j}=1. \tag{8}$$

The following argument is due to Arthur Benjamin (private communication), using the Description-Involution-Exception technique of Benjamin and Quinn [1].

Let us count the number of ways the numbers 1 through n can be colored red, white, and blue, with p numbers colored red or blue and the first non-red number colored white. If there are j red numbers, then these can be chosen in $\binom{n}{j}$ ways, and the blue numbers can be chosen in $\binom{n-1-j}{p-j}$ ways (since the first non-red number must be white). So the sum above, without the sign term, exactly counts these configurations.

Now we introduce an involution on these configurations, by toggling the last non-white number between red and blue. Thus RWBR becomes RWBB and vice versa. This maintains the number of red or blue numbers, and changes the parity of the number of red and the number of blue colors, and hence the sign of $(-1)^{p-j}$. Moreover, we still have a configuration of the required form, except in the single case RR...RWWW...W, where the last red number cannot be changed to blue. Hence all the terms above cancel with their pairs under the involution, except for this one configuration, and thus we see that the alternating sum is 1.

□

From this, we easily obtain

Theorem 1 *a)*

$$\bar{M}^k(y)=\sum_{|K|<k}(-1)^{k-|K|-1}\binom{n-|K|-2}{k-|K|-1}\ln\left(\sum_{i\notin K}\exp(y_i)\right)$$

(here $\binom{-1}{0}:=1$*, and otherwise* $\binom{p}{q}:=0$ *if* $p<q$*).*

b)

$$M^k(y)\le\bar{M}^k(y)\le M^k(y)+k\ln n.$$

Proof: a) Indeed, $\bar{M}^k(y)$ is just the sum of the first k $(q_j-\gamma)$'s, and the alternating sum of the binomial coefficients for a fixed K simplifies, using the identity $\binom{m}{p}=\binom{m-1}{p}+\binom{m-1}{p-1}$, giving the expression above. (The conventions for the binomial coefficients for $p<q$ are designed to make this work for all cases.)

b) Note that, for $k=1$, the only term corresponds to $K=\emptyset$, so that $\bar{M}^1(0)=\ln\sum_i\exp(0)=\ln n$, so the bound follows from Proposition 1.

□

From (a), we see immediately that $\bar{M}^k$ is a smooth function. As a sanity check, we see that the expression in (a) is an empty sum for $k=0$, while for $k=n$, the only nonzero terms correspond to sets K with cardinality $n-1$, and so we obtain $\sum_i y_i$ as desired.

Examples.

$k=1$: As above the only term corresponds to $K=\emptyset$, and we find

$$\bar{M}^1(y)=\ln\left(\sum_i\exp(y_i)\right).$$

This function is sometimes called the *soft maximum* of the y_i's, and dates back to the economic literature on consumer choice: see, e.g., Luce and Suppes [10]. In this context it corresponds to the maximum utility of a consumer whose utilities for n objects are perturbed by noise. The term soft maximum also sometimes refers to the weight vector

$$\left(\frac{\exp(y_i)}{\sum_h \exp(y_h)} \right),$$

which corresponds to the probability with which such a consumer would choose each item. This weight vector is also the gradient of $\bar{M}^1$.

$\bar{M}^1$ has also been used as a penalty function in nonlinear programming (without the logarithm) — see for example Murphy [11] and Bertsekas [2] — and as a potential function in theoretical computer science, starting with Shahrokhi-Matula [18].

We remark that Tunçel and Nemirovskii [19] have noted that $\bar{M}^1$ is not a self-concordant function.

$k = 2$: Now K can be the empty set or any singleton, and we obtain

$$\begin{aligned} \bar{M}^2(y) &= -(n-2)\ln\left(\sum_i \exp(y_i)\right) + \sum_i \ln\left(\sum_{j\neq i} \exp(y_j)\right) \\ &= \ln\left(\sum_{h\neq 2} \exp(y_{[h]})\right) + \ln\left(\sum_{h\neq 1} \exp(y_{[h]})\right) + \\ &\quad \sum_{i\geq 3} \ln\left(1 - \frac{\exp(y_{[i]})}{\sum_h \exp(y_h)}\right). \end{aligned}$$

The second expression is recommended for accurate computation. Note that, if the components of y are well separated, the sum in the first (second) term is dominated by $\exp(y_{[1]})$ ($\exp(y_{[2]})$) and the terms in the last sum are all small. Observe also that, even for $\bar{M}^1$, it is worth ordering the components of y first and then evaluating all terms of the form $\sum_{i\in S} \exp(y_i)$ by summing from the smallest to the largest, to avoid roundoff error. Also, terms of the form $\ln(1-z)$ should be carefully evaluated in case z is small in absolute value; see the MATLAB function log1p, for example.

$k = 3$: Here K can be the empty set, any singleton, or any pair, and we find

$$\begin{aligned} \bar{M}^3(y) &= \binom{n-2}{2} \ln\left(\sum_i \exp(y_i)\right) - (n-3)\sum_i \ln\left(\sum_{j\neq i} \exp(y_j)\right) \\ &+ \sum_{i<j} \ln\left(\sum_{h\neq i,j} \exp(y_h)\right) \\ &= \ln\left(\sum_{h\neq 2,3} \exp(y_{[h]})\right) + \ln\left(\sum_{h\neq 1,3} \exp(y_{[h]})\right) + \ln\left(\sum_{h\geq 3} \exp(y_{[h]})\right) \\ &+ \sum_{1\leq i\leq 3} \sum_{j\geq 4} \ln\left(1 - \frac{\exp(y_{[j]})}{\sum_{h\neq i} \exp(y_{[h]})}\right) - \frac{n-2}{2} \sum_{i\geq 4} \ln\left(1 - \frac{\exp(y_{[i]})}{\sum_h \exp(y_h)}\right) \\ &+ \frac{1}{2} \sum_{4\leq i\neq j\geq 4} \ln\left(1 - \frac{\exp(y_{[j]})}{\sum_{h\neq i} \exp(y_{[h]})}\right). \end{aligned}$$

Again, the last expression is recommended for numerically stable computation.

□

We remark that the formula for the soft maximum is fundamental here; in fact, the other smoothed max k-sums can be derived from it as follows. It is not hard to prove that

$$M^k(y) = \sum_{|K|<k} (-1)^{k-|K|-1} \binom{n-|K|-2}{k-|K|-1} \max\{y_i : i \notin K\}. \tag{9}$$

Indeed, it is enough to establish this when the y_i's are all distinct (because we can take a limit of such cases) and in decreasing order (because both sides are invariant under permutation of the components of y). Then y_j never appears on the right-hand side for $j > k$, while for $1 \le j \le k$ it appears when K consists of $\{1, \ldots, j-1\}$ together with $h := |K| - j + 1$ elements of the set $\{j+1, \ldots, n\}$ of cardinality $n - j$. Thus h runs from 0 to $k - j$ (since $|K| < k$) and for each such h there are $\binom{n-j}{h}$ sets K, each with the same coefficient, and since $|K| = h + j - 1$, the total coefficient of y_j is

$$\sum_{h=0}^{k-j} (-1)^{k-j-h} \binom{n-j-h-1}{k-j-h} \binom{n-j}{h}.$$

This is 1 by (8) with $n - j$, $k - j$, and h replacing n, p, and j. Now if we replace each maximum on the right-hand side of (9) by its soft maximum, we obtain the smoothed max-k-sum by Theorem 1. This derivation is relatively short, but fails to establish the other properties of the smoothed max-k-sum.

In general, evaluating $\bar{M}^k$ takes $O((n)^{k-1})$ time ($O(n \ln n)$ for $k = 1, 2$ if the above presorting strategy is used). The same is true for $\bar{m}^{n-k}$, while $\bar{M}^{n-k}$ and $\bar{m}^k$ can be evaluated in this amount of time if negative Gumbel random variables are used.

If we want a more accurate (but rougher) approximation to M^k, we can use Gumbel random variables scaled by α between 0 and 1. If we denote these by $\bar{M}^k_\alpha$, we find by (7) that

$$\bar{M}^k_\alpha(y) = \alpha \bar{M}^k(y/\alpha),$$

and by Proposition 1 we find

$$M^k(y) \le \bar{M}^k_\alpha(y) \le M^k(y) + \alpha k \ln n.$$

The derivative of $\bar{M}^k_\alpha$ is that of $\bar{M}^k$ at an argument scaled by $1/\alpha$, and thus the Lipschitz constant for the derivative is increased by this factor.

Of course, these results transfer directly to functions: for example, if we define $\bar{F}^k(t) := \bar{M}^k(f_1(t) \ldots, f_n(t))$, it will be a smooth function of t and we have

$$F^k(t) \le \bar{F}^k(t) \le F^k(t) + k \ln n$$

for all t, and similarly $F^k(t) \le \bar{F}^k_\alpha(t) \le F^k(t) + \alpha k \ln n$ for $\bar{F}^k_\alpha$ defined in the obvious way. We similarly define $\bar{f}^k(t) := \bar{m}^k(f_1(t), \ldots, f_n(t))$. Then we have:

Proposition 3 *If all f_i's are convex, so is $\bar{F}^k$. If all are concave, so is $\bar{f}^k$.*

Proof: We have

$$\bar{F}^k(t) = E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} (f_i(t) - \xi_i) + kE\Xi.$$

For each K and fixed ξ_i's, the sum above is convex, as the sum of convex functions. So the maximum is also convex, and finally taking expectations preserves convexity. The proof is analogous for $\bar{f}^k$. □

3 Smoothing by perturbing an optimization formulation

Now we discuss smoothing M^k and m^k by expressing them as values of optimization problems and then adding perturbations. It is clear that the maximum of the y_i's can be written as either the smallest number that exceeds each y_i or as the largest convex combination of the y_i's. Thus it is the optimal value of

$$P(M^1): \quad \min\{x : x \geq y_i \text{ for all } i\}$$

and

$$D(M^1): \quad \max\{\sum_i u_i y_i : \sum_i u_i = 1,\ u_i \geq 0 \text{ for all } i\}.$$

These are probably the simplest and most intuitive dual linear programming problems of all! The corresponding problems for the minimum are $P(m^1) : \max\{x : x \leq y_i \text{ for all } i\}$ and $D(m^1) : \min\{\sum_i u_i y_i : \sum_i u_i = 1, u_i \geq 0 \text{ for all } i\}$.

For the max-k-sum, it is perhaps simplest to generalize $D(M^1)$ to get

$$D(M^k): \quad \max\{\sum_i u_i y_i : \sum_i u_i = k,\ 0 \leq u_i \leq 1 \text{ for all } i\},$$

whose dual is

$$P(M^k): \quad \min\{kx + \sum_i z_i : x + z_i \geq y_i, z_i \geq 0, \text{ for all } i\}.$$

The validity of these is easily confirmed by exhibiting feasible solutions for each with equal objective values: $u_i = 1$ for the indices i corresponding to the k largest components of y, $u_i = 0$ otherwise, and $x = y_{[k]}$, $z_i = \max(y_i - x, 0)$ for each i. For min-k-sums, we have the corresponding problems $D(m^k) : \min\{\sum_i u_i y_i : \sum_i u_i = k, 0 \leq u_i \leq 1 \text{ for all } i\}$, $P(m^k) : \max\{kx - \sum_i s_i : x - s_i \leq y_i, s_i \geq 0, \text{ for all } i\}$.

Note that $P(M^k)$ and $D(M^k)$ do not coincide with $P(M^1)$ and $D(M^1)$ when $k = 1$, but it is easily seen that these two formulations are equivalent. This is immediate for the dual problems, because the extra upper bounds on the u_i's are redundant. For the primal problems, if x is feasible for $P(M^1)$, then x with $z_i = 0$ for all i (briefly, $(x, (0))$, is feasible in $P(M^k)$ for $k = 1$ with the same objective value. Conversely, if $(x, (z_i))$ is feasible for $P(M^k)$ with $k = 1$, then $x + \sum_i z_i$ is feasible for $P(M^1)$ with the same objective value.

Now we introduce perturbations, as in Nesterov [13]. We do this for $D(M^k)$. Let U denote its feasible region, and let g^* denote a closed proper smooth σ-strongly convex *prox-function* on U ($\sigma > 0$), so that g^* is convex with $g^*(v) \geq g^*(u) + \nabla g^*(u)^T(v - u) + (\sigma/2)\|v - u\|^2$ for all $u, v \in U$. Here $\|\cdot\|$ is some appropriate norm on $\mathbf{R}^n$, with dual norm $\|z\|^* := \max\{u^T z : \|u\| \leq 1\}$. We allow g^* to be finite on some convex subset of $\text{aff}(U) = \{u : \sum_i u_i = k\}$ containing U but require it to be $+\infty$ off aff(U). (The reason for using g^* instead of g will appear in the next section.) Then g^* attains its maximum and minimum over the compact set U; assume the minimum is attained at u_0 (unique by strong convexity) and without loss of generality that this minimum is 0; and let the maximum be Δ. We sometimes write U^k or g^{*k} if we want to stress that they depend on k. Then we let $\hat{M}^k(y)$ be the optimal value of

$$\hat{D}(M^k): \quad \max\{\sum_i u_i y_i - g^*(u) : u \in U\},$$

and similarly $\hat{m}^k(y)$ be that of

$$\hat{D}(m^k): \quad \min\{\sum_i u_i y_i + g^*(u) : u \in U\}.$$

If we want to highlight the fact that these depend on the prox-function, we write $\hat{M}^k(y; g^*)$ and $\hat{m}^k(y; g^*)$, etc. We now have

Proposition 4 *a) $\hat{M}^k(y)$ is differentiable, with gradient*

$$\nabla \hat{M}^k(y) = u_{g^*}(y),$$

where $u_{g^*}(y)$ is the unique maximizer of $\hat{D}(M^k)$. Moreover, its gradient is Lipshitz continuous with constant $1/\sigma$:

$$\|\nabla \hat{M}^k(y) - \nabla \hat{M}^k(z)\|^* \leq \|y - z\|/\sigma.$$

b) (0- and n-consistency)

$$\hat{M}^0(y) = \hat{m}^0(y) = 0, \text{ and } \hat{M}^n(y) = \hat{m}^n(y) = \sum_i y_i.$$

c) (sign reversal) $\hat{m}^k(y) = -\hat{M}^k(-y)$.
d) (summability) If $g^{*n-k}(u) = g^{*k}(\mathbf{1} - u)$, where $\mathbf{1} := (1, \ldots, 1)$,

$$\hat{M}^k(y) + \hat{m}^{n-k}(y) = \sum_i y_i.$$

e) (translation invariance) For $\eta \in \mathbf{R}$,

$$\hat{M}^k(y + \eta\mathbf{1}) = \hat{M}^k(y) + k\eta, \quad \hat{m}^k(y + \eta\mathbf{1}) = \hat{m}^k(y) + k\eta.$$

f) (positive scaling invariance) If $\alpha > 0$,

$$\hat{M}^k(\alpha y; \alpha g^*) = \alpha \hat{M}^k(y; g^*), \quad \hat{m}^k(\alpha y; \alpha g^*) = \alpha \hat{m}^k(y; g^*).$$

g) (approximation)

$$M^k(y) - \Delta \leq \hat{M}^k(y) \leq M^k(y), \quad m^k(y) \leq \hat{m}^k(y) \leq m^k(y) + \Delta.$$

Proof: Part (a) follows from Nesterov [13], while part (g) is a direct consequence of the bounds on g, and also appears in [13]. Parts (c), (e), and (f) are immediate from trivial properties of optimization problems. Part (b) is a consequence of the facts that U^0 and U^n are singletons and that therefore g^* evaluated at the unique feasible point is zero. For part (d), note that

$$\begin{aligned}
\hat{M}^k(y) - \sum_i y_i &= \max\{\sum_i (u_i - 1)y_i - g^{*k}(u) : u \in U^k\} \\
&= \max\{-\sum_i (1 - u_i)y_i - g^{*n-k}(\mathbf{1} - u) : \mathbf{1} - u \in U^{n-k}\} \\
&= -\min\{\sum v_i y_i + g^{*n-k}(v) : v \in U^{n-k}\} = -\hat{m}^{n-k}(y).
\end{aligned}$$

□

As in the previous section, we can define $\hat{F}^k(t) := \hat{M}^k(f_1(t), \ldots, f_n(t))$, and similarly $\hat{f}^k(t)$. We now show

Proposition 5 *If all f_i's are convex, so is $\hat{F}^k$, while if all are concave, so is $\hat{f}^k$.*

Proof: We only prove the first statement; the second follows similarly. Suppose $0 \leq \lambda \leq 1$. Then, if u is an optimal solution to $\hat{D}(M^k(f_1((1-\lambda)s+\lambda t), \ldots, f_n((1-\lambda)s+\lambda t)))$, we obtain

$$\begin{aligned}
\hat{F}^k((1-\lambda)s+\lambda t) &= \sum_i u_i f_i((1-\lambda)s+\lambda t) - g^*(u) \\
&\leq (1-\lambda)\left(\sum_i u_i f_i(s) - g^*(u)\right) + \lambda\left(\sum_i u_i f_i(t) - g^*(u)\right) \\
&\leq (1-\lambda)\hat{F}^k(s) + \lambda\hat{F}^k(t),
\end{aligned}$$

since u is also feasible for the problems for s and for t, establishing convexity. □

Once again, we can use scaling to achieve a rougher but more accurate smooth approximation: if $\hat{M}^k_\alpha(y) := \hat{M}^k(y; \alpha g^*)$ for $0 < \alpha < 1$, then (f) and (g) yield

$$M^k(y) - \alpha\Delta \leq \hat{M}^k_\alpha(y) \leq M^k(y),$$

and analogously for a scaled version of $\hat{m}^k$. Note that αg^* is $\alpha\sigma$-strongly convex, so the bound on the Lipschitz constant for the gradient goes up by a factor $1/\alpha$.

It is straightfoward to derive the dual of $\hat{D}(M^k)$. We provide the details in a more general setting in the next section. Here g denotes the convex conjugate of g^*, with

$$g(w) := \max_u \{u^T w - g^*(u)\}.$$

Then $g(0) = 0$. (We assumed g^* was closed and proper so that g^* is also the convex conjugate of g.) Here is the dual (compare with $P(M^k)$ above):

$$\hat{P}(M^k): \quad \min\{kx + \sum_i z_i + g(w) : x + z_i \geq y_i - w_i, z_i \geq 0, \text{ for all } i\}.$$

If we choose g^* to be identically zero on $\mathrm{aff}(U)$, which satisfies all of our requirements except strong convexity, then $g(\lambda\mathbf{1}) = k\lambda$ for $\lambda \in \mathbf{R}$ and $g(w) = +\infty$ otherwise, and our perturbed problems reduce to the original problems $P(M^k)$ and $D(M^k)$. We will see in the next section (Proposition 6) that we can require $\sum_i w_i = 0$ without loss of generality.

We now choose particular prox-functions to see how such smooth approximations can be calculated.

3.1 Quadratic prox-function

First we choose $\|\cdot\|$ to be the Euclidean norm $\|\cdot\|_2$ on $\mathbf{R}^n$, which is self-dual, and take

$$g^*(u) := g^{*k}(u) := \frac{\beta}{2}(\|u\|_2)^2 - \frac{\beta(k)^2}{2n}$$

on $\mathrm{aff}(U)$, where $\beta > 0$, and $+\infty$ otherwise. (We could take $\beta = 1$, but we allow the redundant scaling to ease our generalization in the next section.) Then $\sigma = \beta$, and $\Delta = \beta k(n-k)/[2n]$. Note also that $g^{*k}(u) = (\beta/2)(\|u - (k/n)\mathbf{1}\|_2)^2$, so that $g^{*n-k}(v) = g^{*k}(\mathbf{1} - v)$ for $v \in U^{n-k}$; hence the summability property holds.

Now $\hat{M}^k(y)$ can be computed in $O(n)$ time after we calculate $u := u_{g^*}(y)$. The necessary and sufficient optimality conditions for $D(\hat{M}^k)$ give

$$y_i - \beta u_i = x - s_i + z_i, \text{ for all } i,$$

$$x \in \mathbf{R}, \sum_i u_i = k, s_i \geq 0, u_i \geq 0, s_i u_i = 0, z_i \geq 0, u_i \leq 1, z_i(1-u_i) = 0, \text{ for all } i. \quad (10)$$

From this, with $\lambda := x/\beta$, we obtain

$$u_i = \text{mid}(0, y_i/\beta - \lambda, 1) \text{ for all } i,$$

where $\text{mid}(\cdot,\cdot,\cdot)$ denotes the middle of its three arguments. Note that each u_i is a nonincreasing function of λ equal to 1 for sufficiently small and to 0 for sufficiently large λ. We seek a value of λ so that $\sum_i u_i = k$. We therefore sort the $2n$ numbers y_i/β and $y_i/\beta - 1$. Suppose for simplicity that the y_i's are in nonincreasing order. Then for λ in some subinterval formed by these $2n$ numbers, the first say $h-1$ u_i's equal 1, the last say $n-j$ equal 0, and the remaining components equal $y_i/\beta - \lambda$. (It is possible that $j = h-1$, so that all u_i's are 1 or 0, and h must be $k+1$. So we first check if $y_k \geq y_{k+1} + \beta$; if so we can choose λ between y_{k+1}/β and $y_k/\beta - 1$ and we have our solution. Henceforth we assume this is not the case, so that $h \leq k$ and some u_i's are fractional.)

Then the sum is

$$(h-1) + \sum_h^j y_i/\beta - (j-h+1)\lambda,$$

so that if this is the correct subinterval, λ is equal to $[(k-h+1-\sum_h^j y_i/\beta]/[j-h+1]$. If this value falls in the subinterval, we have our optimal solution; if not, we know we should be in a subinterval to the left (if λ is too small) or to the right (if too large) of the current one. Thus we can perform a binary search, requiring at most $\log_2(2k)$ steps. Each step can be performed in $O(1)$ time if we precompute all the cumulative sums $\sum_h^n y_i/\beta$ in $O(n)$ time, because the partial sums required are just the differences of two such cumulative sums. The total time required is then dominated by the sort, in $O(n \ln n)$ time.

We do not give formulas for the resulting $\hat{M}^k(y)$, but note that, if the components of y are well-separated, so that the gaps are all at least β, then the exceptional case above occurs for any k, and the resulting u puts a weight 1 on the k largest y_h's and 0 on the rest, so that $\hat{M}^k(y) = M^k(y) - \Delta$. This contrasts with the result of our approach using randomization, and that of the following subsections, which are dependent on all the y_i's.

For this g^*, it is easy to see that $g(w) = (\|w - \bar{w}\mathbf{1}\|_2)^2/[2\beta] + k\bar{w} + \beta(k)^2/[2n]$, where $\bar{w} = \sum_i w_i/n$, so using the fact that we can assume $\sum_i w_i = 0$, we have

$$\hat{P}(M^k): \min\{kx + \sum_i z_i + \frac{1}{2\beta}(\|w\|_2)^2 + \frac{\beta(k)^2}{2n} : x + z_i \geq y - w_i, z_i \geq 0 \text{ for all } i, \sum_i w_i = 0\}$$

and

$$\hat{D}(M^k): \quad \max\{\sum_i u_i y_i - \frac{\beta}{2}(\|u\|_2)^2 + \frac{\beta(k)^2}{2n} : \sum_i u_i = k,\ 0 \leq u_i \leq 1 \text{ for all } i\}.$$

3.2 Single-sided entropic prox-function

Next we choose the 1-norm $\|\cdot\|_1$ on $\mathbf{R}^n$, with dual norm the ∞-norm, and take

$$g^*(u) := g^{*k}(u) := \sum_i u_i \ln u_i + k \ln\left(\frac{n}{k}\right)$$

for u satisfying $\sum_i u_i = k$, $u \geq 0$, $+\infty$ otherwise. Note that we do not take g^* infinite outside U^k, which would make g more complicated. This prox-function is appropriate for the simplex (Nesterov [13]) and also suitable for the slightly more complicated U^k. Then $\sigma = 1/k$ and $\Delta = k \ln(n/k)$ (by an extension of the argument in Nesterov [13]). Now we do not have $g^{*n-k}(u) = g^{*k}(\mathbf{1} - u)$, but we could use the alternative single-sided entropic prox-function $\bar{g}^{*k}(u) := g^{*n-k}(\mathbf{1} - u)$ for $\hat{m}^k$ to restore the summability property.

Once again $\hat{M}^k(y)$ can be computed in $O(n)$ time from $u := u_{g^*}(y)$. The necessary and sufficient optimality conditions for $\hat{D}(M^k)$ now read

$$y_i - 1 - \ln u_i = x - s_i + z_i, \text{ for all } i$$

together with (10), so that with $\lambda = x + 1$, we obtain

$$u_i = \min(\exp(y_i - \lambda), 1) \text{ for all } i,$$

where again λ is chosen so that $\sum_i u_i = k$. We assume that the y_i's are sorted in nonincreasing order, so that for some h, the first $h-1$ u_h's are one, and the others equal to $(k-h+1)\exp(y_i)/\sum_h^n \exp(y_j)$, where $\exp(y_h) \leq \exp(\lambda) = \sum_h^n \exp(y_i)/(k-h+1) \leq \exp(y_{h-1})$. The appropriate index h can again be found by binary search, so that the total time required is $O(n \ln n)$.

Let us evaluate the corresponding $\hat{M}^k(y)$, assuming that the critical index is h. Then the first $h-1$ terms in both the linear term and the perturbation yield $\sum_1^{h-1} y_i$, while the remainder give

$$\sum_h^n u_i(y_i - [y_i - \lambda]) = \sum_h^n u_i \lambda = (k - h + 1)\lambda,$$

so we obtain

$$\hat{M}^k(y) = \sum_1^{h-1} y_i + (k-h+1)\ln\left(\sum_h^n \exp(y_i)\right) - (k-h+1)\ln(k-h+1) - k\ln\left(\frac{n}{k}\right).$$

Note that if $k = 1$, then h must also be 1, and we get $\ln(\sum_i \exp(y_i)) - \ln n$, differing only by a constant from our approximation using randomization. However, for larger k, our formula is quite different, and much easier to evaluate.

For this prox-function, we find after some computation that

$$g(w) = k \ln\left(\sum_i \exp(w_i)\right) - k \ln n.$$

Thus, again using the fact that we can assume $\sum_i w_i = 0$, we have

$$\hat{P}(M^k): \min\{kx + \sum_i z_i + k\ln\left(\sum_i \exp(w_i)\right) - k\ln n : x + z_i \geq y - w_i, z_i \geq 0 \text{ for all } i, \sum_i w_i = 0\}$$

and

$$\hat{D}(M^k): \quad \max\{\sum_i u_i y_i - \sum_i u_i \ln u_i - k\ln\left(\frac{n}{k}\right) : \sum_i u_i = k,\ 0 \leq u_i \leq 1 \text{ for all } i\}.$$

3.3 Double-sided entropic prox-function

Finally we choose the 1-norm again and take

$$g^*(u) := g^{*k}(u) := \sum_i [u_i \ln u_i + (1-u_i)\ln(1-u_i)] + k \ln\left(\frac{n}{k}\right) + (n-k)\ln\left(\frac{n}{n-k}\right),$$

which is also appropriate for U^k. Then $\sigma = 1/k + 1/(n-k)$ and $\Delta = k\ln(n/k) + (n-k)\ln(n/[n-k])$, again by using an extension of the arguments in Nesterov [13]. We have regained the property that $g^{*n-k}(u) = g^{*k}(\mathbf{1}-u)$, so the summability property holds with no adjustments.

In this case, the necessary and sufficient optimality conditions for $\hat{D}(M^k)$ read

$$y_i - \ln u_i + \ln(1-u_i) = x - s_i + z_i, \text{ for all } i$$

together with (10), so that with $\lambda := \exp(x)$, we have

$$\frac{u_i}{1-u_i} = \frac{\exp(y_i)}{\lambda}, \text{ or } u_i = \frac{\exp(y_i)}{\lambda + \exp(y_i)} \text{ for all } i.$$

Now we have to make a nonlinear search to find the appropriate λ so that $\sum_i u_i = k$. This is straightforward, but we do not have a finite procedure and thus no computational complexity.

For this prox-function, determining g at a particular w also requires such a nonlinear search to find the maximizing u, and so we do not state the corresponding perturbed primal problem, but the perturbed dual is

$$\begin{array}{lll} \hat{D}(M^k): & \max & \sum_i u_i y_i - \sum_i [u_i \ln u_i + (1-u_i)\ln(1-u_i)] - k\ln\left(\frac{n}{k}\right) - (n-k)\ln\left(\frac{n}{n-k}\right) \\ & & \sum_i u_i = k, \\ & & 0 \le u_i \le 1 \quad \text{for all } i. \end{array}$$

4 Max-k-sums and min-k-sums in a general cone

Let E be a finite-dimensional real vector space, and let E^* denote its dual, the space of all linear functions on E. We use $\langle u, x\rangle$ to denote the result of $u \in E^*$ acting on $x \in E$. Let $\mathcal{K}$ be a closed convex cone in E that is pointed ($\mathcal{K} \cap (-\mathcal{K}) = \{0\}$) and has a nonempty interior. Then its dual cone $\mathcal{K}^* \subseteq E^*$, defined by

$$\mathcal{K}^* := \{u \in E^* : \langle u, x\rangle \ge 0 \text{ for all } x \in \mathcal{K}\},$$

shares the same properties, and $\mathcal{K}^{**} = \mathcal{K}$. $\mathcal{K}$ and $\mathcal{K}^*$ define partial orders in E and E^* by

$$x \succeq z, x, z \in E \text{ means } x - z \in \mathcal{K}, \quad u \succeq^* v, u, v \in E^* \text{ means } u - v \in \mathcal{K}^*.$$

We also write $z \preceq x$ and $v \preceq^* u$ with the obvious definitions.

Suppose $y_1, \ldots, y_n \in E$. We would like to define the max-k-sum and the min-k-sum of the y_i's in E (and maybe smooth approximations to them) to conform with their definitions in $\mathbf{R}$. We write (y_i) for $(y_1, \ldots, y_n) \in E^n$ for ease of notation. Our prime examples for E and $\mathcal{K}$ are:

a) $\mathbf{R}$ and $\mathbf{R}_+$;

b) $\mathbf{R}^p$ and $\mathbf{R}^p_+$;

c) the space of real (complex) symmetric (Hermitian) $d \times d$ matrices, and the subset of positive semidefinite matrices; and

d) $\mathbf{R}^{1+p}$ and the second-order cone $\{(\xi; x) \in \mathbf{R}^{1+p} : \xi \ge \|x\|_2\}$.

4.1 Symmetric cones

The approach to smoothing we took in Section 2 presupposes that unsmoothed max-k-sums and min-k-sums have already been defined, but we can use the formulae resulting from that analysis in a subclass of cones including those above. This requires that the functions ln and exp be defined in E. We therefore recall the definition of *symmetric cones*, which are those that are self-dual (there is an isomorphism between $\mathcal{K}$ and $\mathcal{K}^*$) and homogeneous (for all $x, y \in \text{ int}(\mathcal{K})$, there is an automorphism of $\mathcal{K}$ taking x to y). Such cones have been characterized (they are the direct products of cones of the forms above as well as positive semidefinite matrices over quaternions and one exceptional cone), and a detailed study appears in Faraut and Koranyi [4]. Moreover, these cones coincide with the cones of squares of Euclidean Jordan algebras [4], which are vector spaces endowed with a commutative bilinear product $\circ : E \times E \to E$ satisfying certain properties. For our purposes, the salient facts are these:

i) There is a unit $e \in E$ with $e \circ x = x$ for all $x \in E$.

ii) An idempotent in E is a nonzero element c with $c^2 := c \circ c = c$. A complete system of orthogonal idempotents is a set $\{c_1, \ldots, c_m\}$ of idempotents with $c_i \circ c_j = 0$ for $i \neq j$ and $c_1 + \cdots + c_m = e$. For every $x \in E$, there is a unique complete system of orthogonal idempotents as above and a unique set of distinct real numbers $\lambda_1, \ldots, \lambda_m$ with

$$x = \lambda_1 c_1 + \cdots + \lambda_m c_m.$$

This is called the spectral decomposition of x, and the λ's are the eigenvalues of x. It is easy to see that the cone of squares $\mathcal{K}$ consists of those elements with nonnegative eigenvalues, and its interior those with positive eigenvalues. We define a primitive idempotent as one that cannot be written as the sum of two (necessarily orthogonal) idempotents. The idempotents c_j above may not be primitive, but if not they can be subdivided and we arrive at another (not necessarily unique) decomposition:

$$x = \mu_1 d_1 + \cdots + \mu_p d_p$$

with the d_j's primitive idempotents, and each μ_j one of the λ_i's. We can define the trace of element x as the sum of these μ_j's (and hence the sum of its eigenvalues with multiplicities), and an inner product on E by $(x, z) := \text{trace}(x \circ z)$, so that orthogonal above also implies orthogonal in the inner product. In fact the converse holds also for vectors in the cone: indeed if $x, z \in \mathcal{K}$, $(x, z) = 0$ if and only if $x \circ z = 0$: see, for instance, Faybusovich [5].

In examples (a) and (b), $\circ$ is (componentwise) product, e is a vector of ones, the unit coordinate vectors form a complete system of orthogonal idempotents, as do the sums of unit coordinate vectors in any partition of the index set $\{1, \ldots, p\}$, from which the spectral decomposition of any x is immediate. In example (c) we let $X \circ Z := (XZ + ZX)/2$. Then E is the identity matrix, any projection matrix is an idempotent, and the spectral decomposition of X corresponds to its distinct eigenvalues and the projections onto their corresponding eigenspaces. Finally, in (d), we set $(\xi; x) \circ (\zeta; z) := (\xi\zeta + x^T z; \xi z + \zeta x)$. Then $e = (1; 0)$, and any pair $\{(1/2; z/2), (1/2; -z/2)\}$ for $\|z\|_2 = 1$ forms a complete system of orthogonal idempotents, as does e alone. Any $(\xi; x)$ with x nonzero can be written as a linear combination of such a pair of elements, with $z = x/\|x\|_2$, and any $(\xi; 0)$ is a multiple of e alone. It is easy to check that the cones of squares in these examples are as given above. Moreover, the norm $\|x\| := (x, x)^{1/2}$ is the Euclidean norm for examples

(a) and (b), the Frobenius norm for example (c), and $\sqrt{2}$ times the Euclidean norm for example (d). We can also define the 1- (∞-) norm of x as the sum (maximum) of the absolute values of the μ_j's in the decomposition above.

We can now define exp and ln. For any $x \in E$ with spectral decomposition as above, we set

$$\exp(x) := \exp(\lambda_1)c_1 + \cdots + \exp(\lambda_m)c_m,$$

and if moreover all eigenvalues of x are positive, we set

$$\ln(x) := \ln(\lambda_1)c_1 + \cdots + \ln(\lambda_m)c_m.$$

Note that these two functions are inverses of one another.

Then for any set $y_1, \ldots, y_n$ of elements of E, each $\exp(y_i)$, and hence their sum, lies in $\mathrm{int}(\mathcal{K})$, and so

$$\bar{M}^1((y_i)) := \ln\left(\sum_i \exp(y_i)\right)$$

is defined. Moreover, for any j, $\sum_i \exp(y_i) \succeq \exp(y_j)$, and since by Löwner [9] and Koranyi [8] the function ln is operator monotone, we have

$$\bar{M}^1((y_i)) \succeq y_j$$

for all j. We can similarly define $\bar{M}^k((y_i))$ for any k using part (a) of Theorem 1, but we do not have a proof that this dominates all sums $\sum_{j\in K} y_j$ with $|K| = k$. The computational cost of computing this expression is that of $O((n)^{k-1})$ spectral decompositions ($O(n)$ for $k = 1$), together with some lesser algebra.

4.2 Definition by optimization formulations

We now return to the general case, but we will also discuss the case of symmetric cones when refined results are possible. We first seek to extend $P(M^k)$ and $D(M^k)$. If we translate them directly, we see that the objective function of $P(M^k)$ becomes a vector in E, and we need to replace the one in the constraints of $D(M^k)$. We therefore choose some

$$v \in \mathrm{int}(\mathcal{K}^*),$$

and define

$$P(M^k((y_i)): \quad \min\{k\langle v, x\rangle + \sum_i \langle v, z_i\rangle : x + z_i \succeq y_i,\ z_i \succeq 0,\ \text{for all } i\}$$

and

$$D(M^k((y_i))): \quad \max\{\sum_i \langle u_i, y_i\rangle : \sum_i u_i = kv,\ 0 \preceq^* u_i \preceq^* v \text{ for all } i\}.$$

Once again we can use simplified problems when $k = 1$: we can eliminate the z_i's in the primal problem and the upper bounds on the u_i's in the dual. The equivalence is straightforward. Note that x and the z_i's lie in E, while the u_i's lie in E^*. Let $U := U^k$ denote the feasible region of $D(M^k((y_i)))$ in E^{*n}.

Let $g^* = g^{*k}$ be a closed proper convex prox-function on U. (We have removed the requirements of smooth and σ-strongly convex because we no longer establish smoothness

of the resulting max-k-sum.) We allow g^* to be finite on some subset of $\mathrm{aff}(U) = \{(u_i) \in E^{*n} : \sum_i u_i = kv\}$ containing U but it must be $+\infty$ off $\mathrm{aff}(U)$. Let g be its convex conjugate on E^n (with the natural scalar product $\langle (u_i), (w_i)\rangle := \sum_i \langle u_i, w_i\rangle$). Then we can define perturbed versions of the problems above:

$$\hat{P}(M^k((y_i))): \quad \min\{k\langle v, x\rangle + \sum_i \langle v, z_i\rangle + g((w_i)) : x + z_i \succeq y_i - w_i, z_i \succeq 0, \text{ for all } i\}$$

and

$$\hat{D}(M^k((y_i))): \quad \max\{\sum_i \langle u_i, y_i\rangle - g^*((u_i)) : \sum_i u_i = kv,\ 0 \preceq^* u_i \preceq^* v \text{ for all } i\}.$$

If g^* is identically 0 on $\mathrm{aff}(U)$, so that $g((w_i))$ is $k\langle v, \bar{w}\rangle$ if $w_1 = \cdots = w_n = \bar{w}$ and $g((w_i)) = +\infty$ otherwise, these reduce to the unperturbed problems above.

Let us demonstrate that these are indeed duals. If we start with $\hat{D}(M^k((y_i)))$, and associate multipliers x for the equality and s_i and z_i for the inequality constraints, the dual becomes

$$\begin{array}{rl} \min_{x\in E,(s_i\succeq 0),(z_i\succeq 0)} & \max_{(u_i)} \left(\sum_i [\langle u_i, y_i\rangle + \langle u_i, s_i\rangle + \langle (v-u_i), z_i\rangle - \langle u_i, x\rangle] + k\langle v, x\rangle - g^*((u_i))\right) \\ = & \min_{x,(s_i\succeq 0),(z_i\succeq 0)} \left(k\langle v, x\rangle + \sum_i \langle v, z_i\rangle + g((y_i + s_i - z_i - x))\right). \end{array}$$

It is easy to see that the latter is equivalent to $\hat{P}(M^k((y_i)))$. Conversely, if we start with $\hat{P}(M^k((y_i)))$ and associate multipliers u_i with the (first set of) inequality constraints, the dual is

$$\max_{(u_i\succeq^* 0)} \min_{x,(z_i\succeq 0),(w_i)} (k\langle v, x\rangle + \sum_i [\langle v, z_i\rangle - \langle u_i, x\rangle - \langle u_i, z_i\rangle + \langle u_i, y_i\rangle - \langle u_i, w_i\rangle] + g((w_i))).$$

The minimum is $-\infty$ unless $\sum_i u_i = kv$ and $u_i \preceq^* v$ for all i, and so this reduces to $\hat{D}(M^k)$ above.

Although weak duality is immediate from this derivation, we establish it directly to show the conditions for strong duality to hold. If $(x, (z_i))$ is feasible in $P(M^k)$ and (u_i) in $D(M^k)$, then

$$k\langle v, x\rangle + \sum_i \langle v, z_i\rangle - \sum_i \langle u_i, y_i\rangle = \sum_i [\langle v - u_i, z_i\rangle + \langle u_i, x + z_i - y_i\rangle \geq 0,$$

with equality if and only if

$$\langle v - u_i, z_i\rangle = \langle u_i, x + z_i - y_i\rangle = 0 \text{ for all } i. \tag{11}$$

Similarly, if $(x, (z_i), (w_i))$ is feasible in $\hat{P}(M^k)$ and (u_i) in $\hat{D}(M^k)$, then

$$\begin{array}{rl} & k\langle v, x\rangle + \sum_i \langle v, z_i\rangle + g((w_i)) - \sum_i \langle u_i, y_i\rangle + g^*((u_i)) \\ = & \sum_i [\langle v - u_i, z_i\rangle + \langle u_i, x + z_i - y_i\rangle] + g((w_i)) + g^*((u_i)) \\ \geq & \sum_i [\langle v - u_i, z_i\rangle + \langle u_i, x + z_i - y_i\rangle] + \sum_i \langle u_i, w_i\rangle \\ = & \sum_i [\langle v - u_i, z_i\rangle + \langle u_i, x + z_i - y_i + w_i\rangle] \geq 0, \end{array}$$

with equality throughout if and only if $\langle v - u_i, z_i\rangle = \langle u_i, x + z_i - y_i + w_i\rangle = 0$ for all i and $g((w_i)) + g^*((u_i)) = \sum_i \langle u_i, w_i\rangle$.

Let us demonstrate that all these problems have optimal solutions. First, for $k = 0$, the only feasible solution to $D(M^0((y_i)))$ or $\hat{D}(M^0((y_i)))$ is (0), with objective value 0. (We write (0) for $(0, \ldots, 0) \in E^n$, and similarly $(v) \in E^{*n}$ and $(\eta) \in E^n$ for $\eta \in E$.) Let $\bar{x}$ lie in $\mathrm{int}(\mathcal{K})$, and then choose λ with λ positive and sufficiently large that $\lambda\bar{x} \pm y_i \in \mathrm{int}(\mathcal{K})$ for all i, and replace $\bar{x}$ by $\lambda\bar{x}$. Then $(\bar{x}, (0))$ is feasible in $P(M^0)$ (and $(\bar{x}, (0), (0))$ in $\hat{P}(M^0)$) with objective value 0, and hence optimal. If $k = n$, then the only feasible solution to $D(M^n((y_i)))$ or $\hat{D}(M^n((y_i)))$ is (v), with objective value $\sum_i \langle v, y_i\rangle$. Also, $(-\bar{x}, (\bar{x} + y_i))$, with $\bar{x}$ as above, is feasible in $P(M^n)$ (and $(-\bar{x}, (\bar{x} + y_i), (0))$ in $\hat{P}(M^n)$) with the same objective value, and hence optimal. Now suppose $0 < k < n$. Then $((k/n)v)$ is feasible in $D(M^k)$ and in $\hat{D}(M^k)$, and in fact satisfies all inequality constraints strictly. Moreover, the feasible region of these two problems is compact, and hence an optimal solution exists. Both $P(M^k)$ and $\hat{P}(M^k)$ have feasible solutions as above, and the existence of a strictly feasible solution to their duals implies that they also have optimal solutions, with no duality gap.

We now prove

Proposition 6 *Without loss of generality, we can restrict (w_i) to sum to zero in $\hat{P}(M^k((y_i))$.*

Proof: Let us write w_i as the sum of $w_i' := w_i - \bar{w}$ and $\bar{w}$, where as before $\bar{w} := \sum_i w_i/n$. Then

$$\begin{aligned} g((w_i)) &= \max_u \{\sum_i \langle u_i, \bar{w} + w_i'\rangle - g^*((u_i))\} \\ &= k\langle v, \bar{w}\rangle + \max_u \{\sum_i \langle u_i, w_i'\rangle - g^*((u_i))\} \\ &= k\langle v, \bar{w}\rangle + g((w_i')), \end{aligned}$$

since g^* is only finite if $\sum_i u_i = kv$. So, for any x, (z_i),

$$k\langle v, x\rangle + \sum_i \langle v, z_i\rangle + g((w_i)) = k\langle v, x + \bar{w}\rangle + \sum_i \langle v, z_i\rangle + g((w_i')).$$

Hence if $(x, (z_i), (w_i))$ is feasible in $\hat{P}(M^k)$, so is $(x + \bar{w}, (z_i), (w_i'))$ with the same objective function, and $\sum_i w_i' = 0$. $\square$

Correspondingly we define problems for the min-k-sum and their perturbations; we give only the latter:

$$\hat{P}(m^k((y_i))): \quad \max\{k\langle v, x\rangle - \sum_i \langle v, s_i\rangle - g((w_i)) : x - s_i \preceq y_i + w_i, s_i \succeq 0, \text{ for all } i\}$$

and

$$\hat{D}(m^k((y_i))): \quad \min\{\sum_i \langle u_i, y_i\rangle + g^*((u_i)) : \sum_i u_i = kv,\ 0 \preceq^* u_i \preceq^* v \text{ for all } i\}.$$

Of course, the values of all these problems are scalars, and so will not provide the definitions we need. We therefore set

$$M^k((y_i)) := \{kx + \sum_i z_i : (x, (z_i)) \in \mathrm{Argmin}(P(M^k((y_i))))\}$$

and analogously

$$m^k((y_i)) := \{kx - \sum_i s_i : (x,(s_i)) \in \text{Argmax}(P(m^k((y_i))))\}.$$

(Here Argmin and Argmax denote the sets of all optimal solutions to the problem given.) For the perturbed problems, we add the extra constraint $\sum_i w_i = 0$ to remove the ambiguity from x, and define

$$\hat{M}^k((y_i)) := \{kx + \sum_i z_i : (x,(z_i),(w_i)) \in \text{Argmin}(\hat{P}(M^k((y_i)))), \sum_i w_i = 0\}$$

and

$$\hat{m}^k((y_i)) := \{kx - \sum_i s_i : (x,(s_i),(w_i)) \in \text{Argmax}(\hat{P}(m^k((y_i)))), \sum_i w_i = 0\}.$$

We can now state some properties of these functions.

Theorem 2 *a) (0- and n-consistency)*

$$\begin{aligned} M^0((y_i)) = m^0((y_i)) &= \hat{M}^0((y_i)) = \hat{m}^0((y_i)) = 0, \text{ and} \\ M^n((y_i)) = m^n((y_i)) &= \hat{M}^n((y_i)) = \hat{m}^n((y_i)) = \sum_i y_i. \end{aligned}$$

b) (sign reversal) $m^k((y_i)) = -M^k((-y_i))$ *and* $\hat{m}^k((y_i)) = -\hat{M}^k((-y_i))$.

c) (summability)

$$M^k((y_i)) = \{\sum_i y_i\} - m^{n-k}((y_i)),$$

and if $g^{*n-k}((u_i)) = g^{*k}((v - u_i))$,

$$\hat{M}^k((y_i)) = \{\sum_i y_i\} - \hat{m}^{n-k}((y_i)).$$

d) (translation invariance) For $\eta \in E$,

$$M^k((y_i + \eta)) = M^k((y_i)) + \{k\eta\}, \quad m^k((y_i + \eta)) = m^k((y_i)) + \{k\eta\},$$

and

$$\hat{M}^k((y_i + \eta)) = \hat{M}^k((y_i)) + \{k\eta\}, \quad \hat{m}^k((y_i + \eta)) = \hat{m}^k((y_i)) + \{k\eta\}.$$

e) (positive scaling invariance) If $\alpha > 0$,

$$M^k((\alpha y_i)) = \alpha M^k((y_i)), \quad m^k((\alpha y_i)) = \alpha m^k((y_i))$$

and

$$\hat{M}^k((\alpha y_i); \alpha g^*) = \alpha \hat{M}^k((y_i); g^*), \quad \hat{m}^k((\alpha y_i); \alpha g^*) = \alpha \hat{m}^k((y_i); g^*).$$

f) For any K *of cardinality* k, *and any* $x' \in M^k((y_i))$, $x' \succeq \sum_{i \in K} y_i$.

Proof: We generally prove only the results for M^k and $\hat{M}^k$, or only for $\hat{M}^k$ and $\hat{m}^k$; the other cases follow by similar arguments.

For (a), we note that both U^0 and U^n have only one element, (0) and (v) respectively, giving objective values of 0 and $\langle v, \sum_i y_i \rangle$ respectively. In $P(M^0((y_i)))$ and $\hat{P}(M^0((y_i)))$, we can achieve objective value 0 by choosing $(\bar{x}, (0))$ or $(\bar{x}, (0), (0))$, and the only way to achieve this value is to set all z_i's to zero. In $P(M^n((y_i)))$ and $\hat{P}(M^n((y_i)))$, the constraints imply $nx + \sum_i z_i \succeq \sum_i y_i$, and hence $\langle v, nx + \sum_i z_i \rangle \geq \langle v, \sum_i y_i \rangle$, and the only way to achieve equality is to have $nx + \sum_i z_i = \sum_i y_i$. (Note that $g((w_i)) = 0$ for any (w_i) with $\sum_i w_i = 0$.)

Suppose $(x, (s_i), (w_i)) \in \text{Argmax}(\hat{P}(m^k((y_i))))$ with $\sum_i w_i = 0$. Then it is also a minimizing solution to $\min\{k\langle v, -x \rangle + \sum_i \langle v, s_i \rangle + g((w_i)) : -x + s_i \succeq -y_i - w_i, s_i \succeq 0, \text{ for all } i\}$, which implies that $(-x, (s_i), (w_i))$ is an optimal solution to $\hat{P}(M^k(-y))$. Also, $kx - \sum_i s_i = -(k(-x) + \sum_i s_i)$. The argument can be reversed, thus yielding (b).

Next consider (c). If the stated condition holds, and $\sum_i w_i = 0$, then

$$g^{n-k}((w_i)) = \max\{\sum_i \langle u_i, w_i \rangle - g^{*n-k}((u_i))\} = \max\{\sum_i \langle v - u_i, -w_i \rangle - g^{*k}((v - u_i))\} = g^k((-w_i)).$$

Now suppose $(x, (z_i), (w_i)) \in \text{Argmin}(\hat{P}(M^k((y_i))))$ with $\sum_i w_i = 0$, and let $s_i := x + z_i - y_i + w_i \succeq 0$ for all i. Then $y_i - w_i - x + s_i = z_i \succeq 0$ and

$$kx + \sum_i z_i = kx + \sum_i (y_i - w_i - x + s_i) = \sum_i y_i - [(n-k)x - \sum_i s_i]. \tag{12}$$

Moreover, its objective value is

$$\langle v, kx + \sum_i z_i \rangle + g^k((w_i)) = \sum_i \langle v, y_i \rangle - [\langle v, (n-k)x - \sum_i s_i \rangle - g^{n-k}((-w_i))],$$

and since the argument can be reversed, we see that $(x, (s_i), (-w_i)) \in \text{Argmax}(\hat{P}(m^{n-k}((y_i))))$. Then (12) gives (c).

Part (d) follows immediately from the fact that $(x+\eta, (z_i)(w_i))$ is feasible for $\hat{P}(M^k((y_i + \eta)))$ if and only if $(x, (z_i), (w_i))$ is feasible for $\hat{P}(M^k((y_i)))$, with objective value the constant $k\langle v, \eta \rangle$ larger.

Note that $(\alpha g^*)^*((\alpha w_i)) = \max_{(u_i)}\{\alpha \sum_i \langle u_i, w_i \rangle - \alpha g^*((u_i))\} = \alpha g((w_i))$. Thus if $(x, (z_i), (w_i))$ is feasible for $\hat{P}(M^k((y_i); g^*))$, then $(\alpha x, (\alpha z_i), (\alpha w_i))$ is feasible for $\hat{P}(M^k((\alpha y_i); \alpha g^*))$, with objective value a factor α larger. Hence one is optimal if and only if the other is, and this yields (e).

Finally, for any K of cardinality k, and any feasible solution $(x, (z_i))$ to $P(M^k(y))$, $kx + \sum_i z_i \succeq kx + \sum_{i \in K} z_i \succeq \sum_{i \in K} y_i$. This proves (f). Note that we cannot obtain such a result for the perturbed problem, since we cannot control the effect of the w_i's. While the objective values of the problems and their perturbations are closely related, we cannot conclude the same about their sets of optimal solutions.

□

Here is a small example to show that $M^k(y)$ need not be a singleton. Let $E = E^* = \mathbf{R}^3$ with $\langle u, x \rangle := u^T x$. Let $\mathcal{K} := \{x \in E : x_1 \geq |x_2|, x_1 > |x_3|\}$, $v = (1; 0; 0)$, $n = 2$, and $y_1 = (0; 1; 0)$, $y_2 = (0; -1; 0)$. Then $M^1((y_i)) = \{(1; 0; \xi) : -1 \leq \xi \leq 1\}$ Indeed, any x in this set, with $z_1 = z_2 = 0$, is feasible in $P(M^1((y_i)))$ with objective 1, while $u_1 = (1/2; 1/2; 0)$ and $u_2 = (1/2; -1/2; 0)$ are feasible in $D(M^1((y_i)))$ with the same objective.

Given this lack of uniqueness, we find the summability property rather remarkable: the set of max-k-sums corresponds exactly to that of min-$(n-k)$-sums after reflection in the point $\sum_i y_i/2$.

If $E = \mathbf{R}^m$ and $\mathcal{K} = \mathbf{R}^m_+$, then $M^k((y_i))$ is the componentwise max-k-sum. If g^* (and hence g) is separable, $\hat{M}^k((y_i))$ is the componentwise perturbed max-k-sum. Indeed, if E and $\mathcal{K}$ are products of lower-dimensional spaces and cones, then again max-k-sums are obtained by finding max-k-sums for each component, and similarly for the perturbed versions if g^* is separable.

We have suppressed the dependence of M^k and $\hat{M}^k$ on v, but it is present. Suppose $E = E^* = \mathbf{R}^3$ as above, but now $\mathcal{K} := \{x \in E : x_1 \geq \|(x_2; x_3)\|_2\}$, the second-order cone. Then for y_1 and y_2 as above, the feasible region of $P(M^1((y_i)))$ (looking just at $x + z_1 + z_2$) is the intersection of two translated second-order cones with vertices at y_1 and y_2, a convex set with a "ridge" at $\{x = (\xi; 0; \zeta) : \xi \geq (1+\zeta^2)^{1/2}\}$, and different points on this hyperbola will be selected by different vectors $v = (1; 0; \mu)$, $|\mu| < 1$.

This example also shows that, even if $M^1((y_i))$ is a singleton, it may not satisfy the associative law that $M^1(y_1, M^1(y_2, y_3)) = M^1(M^1(y_1, y_2), y_3)$. Indeed, let y_1 be like $M^1(y_2, y_3)$ but defined with a different v so that $y_1 \neq M^1(y_2, y_3)$. Then $y_1 \succeq y_2$, so $M^1(y_1, y_2) = y_1$, and $y_1 \succeq y_3$, so $M^1(M^1(y_1, y_2), y_3) = M^1(y_1, y_3) = y_1$. However, as in the example above, y_1 is not greater than or equal to $M^1(y_2, y_3)$ (defined with v), and so $M^1(y_1, M^1(y_2, y_3)) \neq y_1$.

Computing one element (or all elements) of $M^k((y_i))$ requires finding one (or all) optimal solutions of $P(M^k((y_i)))$, which is a linear cone programming problem over (products of) the cone $\mathcal{K}$. Similarly, for $\hat{M}^k((y_i))$, we need one or all optimal solutions of a similar problem, but with a nonlinear objective function. For example, we can take β to be a positive self-adjoint operator from E^* to E, and define

$$g^*((u_i)) := (1/2)\sum_i \langle u_i, \beta u_i\rangle\rangle - (k)^2\langle v, \beta v\rangle/[2n] \tag{13}$$

on $\mathrm{aff}(U)$ and $+\infty$ otherwise. Then it is easy to see that $g((w_i)) = (1/2)\sum_i \langle \beta^{-1}(w_i - \bar{w}), w_i - \bar{w}\rangle + k\langle v, \bar{w}\rangle + (k)^2\langle v, \beta v\rangle/[2n]$, where $\bar{w} := \sum_i w_i/n$. (Compare with Subsection 3.1.) Then we need to solve a quadratic cone problem over $\mathcal{K}$, but now there is no easy search to find optimal solutions. However, we note that for this prox-function,

$$g^{*k}((u_i)) = (1/2)\sum_i \langle u_i - \frac{k}{n}v, \beta(u_i - \frac{k}{n}v)\rangle$$

on $\mathrm{aff}(U^k)$, so $g^{*n-k}((u_i)) = g^{*k}((v - u_i))$ on $\mathrm{aff}(U^{n-k})$, and thus summability holds.

4.3 Symmetric cones redux

We now return to the special case of symmetric cones for the optimization approach to max-k-sums. As we have noted, such cones have a natural inner product $(x, z) := \mathrm{trace}(x \circ z)$, and so we can identify E^* with E. We can therefore choose v in E, and make the natural choice $v := e$, the unit element. Also, all u_i's now lie in E.

An important consequence of $\mathcal{K}$ being symmetric is uniqueness:

Theorem 3 *In the setting above, the set $M^k((y_i))$ is a singleton. Moreover, if g is strictly convex on $\{(w_i) : \sum_i w_i = 0\}$, then $\hat{M}^k((y_i))$ is a singleton.*

(We remark that, with some regularity conditions on their domains, g is strictly convex as above if and only if g^* is differentiable on the relative interior of its domain in aff(U); see Rockafellar [16], Section 26.)

Proof: We know that both $P(M^k((y_i)))$ and $D(M^k((y_i)))$ have optimal solutions, say $(x, (z_i))$ and (u_i), with no duality gap, so that by the conditions (11), $(e - u_i, z_i) = (u_i, x + z_i - y_i) = 0$ for all i. Also, $e - u_i$, z_i, u_i, and $x + z_i - y_i$ all lie in $\mathcal{K}$, so by the properties of Euclidean Jordan algebras, we have $(e - u_i) \circ z_i = u_i \circ (x + z_i - y_i) = 0$ for all i. This implies

$$\sum u_i \circ y_i = \sum_i u_i \circ (x + z_i) = (\sum_i u_i) \circ x + \sum_i u_i \circ z_i = ke \circ x + \sum_i e \circ z_i = kx + \sum_i z_i.$$

Hence any optimal solution to $P(M^k((y_i)))$ must have $kx + \sum_i z_i$ equal to $\sum_i u_i \circ y_i$ for some fixed optimal solution to $D(M^k((y_i)))$.

Next consider $\hat{P}(M^k((y_i)))$ and $\hat{D}(M^k((y_i)))$. Since g is strictly convex, the (w_i) part of an optimal solution to the former (with $\sum_i w_i = 0$) must be unique. But if $(x, (z_i))$ is part of an optimal solution to $\hat{P}(M^k((y_i)))$ with this (w_i), then it is also an optimal solution to $P(M^k((y_i - w_i)))$, which as we have shown above is unique. □

Note that the proof shows how the unique element of $M^k((y_i))$ can be obtained from an optimal solution to the dual problem.

In general, computing $M^k((y_i))$ requires the solution of a linear symmetric cone programming problem. Since a quadratic objective can be reformulated in either second-order or semidefinite programming as a linear objective with additional constraints, the same is true of $\hat{M}^k((y_i))$ when g is the quadratic function given above in (13). However, we cannot typically obtain the solution in closed form, even with a simple search. The problem is that the multiplier for the constraint $\sum_i u_i = ke$ is an element of E, not a scalar. We can also use single- or double-sided entropic prox-functions, where each $u_i \ln u_i$ is replaced by $(u_i, \ln u_i)$; these are convex by results of Faybusovich [6]. But again computation is complicated by the need for a search over $x \in E$; moreover, the derivatives of these prox-functions are not always easy, since they involve the Peirce decomposition of the argument: see Faraut and Koranyi [4, 8].

The one special case where the solution can be obtained is the following. We say x and y in E *operator commute* if $x \circ (y \circ z) = y \circ (x \circ z)$ for all $z \in E$. Now suppose $y_1, \ldots, y_n$ pairwise operator commute. Then, by [4], there is a complete set of orthogonal primitive idempotents $\{c_1, \ldots, c_m\}$ such that all y_i's are linear combinations of these, so that

$$y_i = y_i^1 c_1 + \cdots + y_i^m c_m$$

for all i. Then the unique element of $M^k((y_i))$ is

$$y^1 c_1 + \cdots + y^m c_m,$$

where each y^j is the max-k-sum of the $y_i^j, i = 1, \ldots, n$. Indeed, we can construct optimal solutions $(x^j, (z_i^j)$ and (u_i^j) to each problem $P(M^k(y_1^j, \ldots, y_n^j))$ and $D(M^k(y_1^j, \ldots, y_n^j))$, and then set $x := x^1 c_1 + \cdots + x^m c_m$ and similarly for z_i and u_i. These give feasible solutions to $P(M^k((y_i))$ and $D(M^k((y_i))$, and optimality follows from (11). We can do the same for the perturbed versions with suitable perturbations. Let us choose $g^*((u_i)) := (\beta/2) \sum_i (u_i - (k/n)e, u_i - (k/n)e)$ with $\beta > 0$, which is β-strongly convex with respect to the norm $\|(u_i)\| := (\sum_i (u_i, u_i))^{1/2}$. Its conjugate is $g((w_i)) = k(e, \bar{w}) + 1/(2\beta) \sum_i (w_i -$

$\bar{w}, w_i - \bar{w})$ with $\bar{w}$ as before. Then the unique element of $\hat{M}^k((y_i))$ can be obtained as above from the solutions to $\hat{P}(M^k(y_1^j, \ldots, y_n^j)$ and its dual. The key is that each vector constructed in E is a linear combination of the c_j's, and since $(\lambda_1 c_1 + \cdots + \lambda_m c_m, \lambda_1 c_1 + \cdots + \lambda_m c_m) = \sum_j (\lambda_j)^2$, the objective function splits into a component for each j, as does feasibility. We can similarly obtain closed form solutions if we use a single-sided entropy prox-function. The situation is analogous to that where $E = \mathbf{R}^m$, $\mathcal{K} = \mathbf{R}^m_+$.

Suppose we wish to compute $M^1(y_1, y_2)$. Then by translation invariance, this is $(y_1 + y_2)/2 + M^1((y_1 - y_2)/2, (y_2 - y_1)/2)$. Trivially $\pm(y_1 - y_2)/2$ operator commute, and if $(y_1 - y_2)/2 = \lambda_1 c_1 + \cdots + \lambda_m c_m$ is a spectral decomposition, we find

$$M^1(y_1, y_2) = (y_1 + y_2)/2 + |\lambda_1| c_1 + \cdots + |\lambda_m| c_m.$$

As we mentioned above, a natural choice for v is the unit element e. However, if we can compute max-k-sums for $v = e$, we can compute them for any $v \in \mathrm{int}(\mathcal{K})$. Indeed, because $\mathcal{K}$ is homogeneous, there is a self-adjoint operator ϕ taking e to v and $\mathcal{K}$ onto itself. Then if we seek $M_v^k((y_i))$ (where the subscript indicates that e has been replaced by v), we note that $(x, (z_i)) \in \mathrm{Argmin}(P(M_v^k((y_i))))$ iff $(\phi x, (\phi z_i)) \in \mathrm{Argmin}(P(M^k((\phi y_i))))$, since $(\phi e, kx + \sum_i z_i) = (e, k\phi x + \sum_i \phi z_i)$, and the feasibility conditions are equivalent. Hence

$$M_v^k((y_i)) = \phi^{-1} M^k((\phi y_i)).$$

4.4 Final notes

We might consider

$$\lim_{\alpha \downarrow 0} \alpha \ln \left(\sum_i \exp(y_i/\alpha) \right),$$

if the limit exists, as a definition of the maximum of $y_1, \ldots, y_n$. Does this limit exist? If it does, it may not equal the unique element of $M^1((y_i))$; e.g., if

$$y_1 = \frac{1}{25} \begin{bmatrix} 41 & 12 \\ 12 & 34 \end{bmatrix}, \quad y_2 = \frac{1}{25} \begin{bmatrix} 34 & 12 \\ 12 & 41 \end{bmatrix},$$

then some computation shows that

$$\lim_{\alpha \downarrow 0} \alpha \ln \left(\sum_i \exp(y_i/\alpha) \right) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad M^1(y_1, y_2) = \left\{ \frac{1}{25} \begin{bmatrix} 41 & 12 \\ 12 & 41 \end{bmatrix} \right\}.$$

Throughout, we have thought of k (for Kronecker?) as an integer, and this was necessary for our initial definitions and for smoothing by randomization. But for our development using optimization formulations, k could be any real number between 0 and n, and all the proofs go through. Hence for example the max-5/2-sum of $y_1, \ldots, y_6$ is their min-7/2-sum reflected in $\sum_i y_i/2$.

Acknowledgement The author would like to thank Gena Samorodnitsky, Leonid Faybusovich, Rob Freund, Arkadi Nemirovskii, and Yurii Nesterov for very helpful conversations.

References

[1] A. T. Benjamin and J. J. Quinn, *An alternate approach to alternating sums: a method to DIE for*, The College Mathematics Journal, 39 (2008), pp. 191–201.

[2] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Athena Press, Belmont, MA, 1996.

[3] J. Bogdan, E. van den Berg, W. Su, and E. Candes, *Statistical estimation and testing via the sorted ℓ_1 norm*, manuscript, arXiv:1310.1969v2, 2013.

[4] J. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford University Press, Oxford, 1995.

[5] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior-point algorithms*, Journal of Computational and Applied Mathematics, 86 (1997), pp. 149–175.

[6] L. Faybusovich, *E. Lieb convexity inequalities and noncommutative Bernstein inequality in Jordan-algebraic setting*, Theoretical Mathematics & Applications, 6 (2) (2016), pp. 1–35.

[7] J.-B. Hiriart-Urruty and D. Ye, *Sensitivity analysis of all eigenvalues of a symmetric matrix*, Numerische Mathematik 70 (1995), pp. 45–72.

[8] A. Koranyi, *Monotone functions on formally real Jordan algebras*, Mathematische Annalen, 269 (1984), pp. 73–76.

[9] K. Löwner, *Über monotone matrixfunktionen*, Mathematische Zeitschrift, 38 (1934), pp. 177–216.

[10] R. D. Luce and P. Suppes, *Preference, utility, and subjective probability*, in *Handbook of Mathematical Psychology, Vol. III*, R. D. Luce, R. R. Bush, F. Galanter, eds., Wiley, New York, 1965.

[11] F. H. Murphy, *A class of exponential penalty functions*, SIAM Journal on Control, 12 (1974), pp. 679–687.

[12] Yu. E. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, 2004.

[13] Yu. E. Nesterov, *Smooth minimization of nonsmooth functions*, Mathematical Programming, 103 (2005), pp. 127–152.

[14] Yu. E. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Methods in Convex Programming*, SIAM Publications, Philadelphia, 1994.

[15] A. S. Nemirovskii, private communication.

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

[17] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, The Journal of Risk, 2 (2000), pp. 21–41.

[18] F. Shahrokhi and D. W. Matula, *The maximum concurrent flow problem*, Journal of the ACM, 37 (1990), pp. 318–334.

[19] L. Tunçel and A. S. Nemirovskii, *Self-concordant barriers for convex approximations of structured convex sets*, Foundations of Computational Mathematics, 10 (2010), pp. 485–525.

[20] G. Zakeri, D. Craigie, A. Philpott, and M. J. Todd, *Optimization of demand response through peak shaving*, Operations Research Letters, 47 (2014), pp. 97–101.

[21] X. Zeng and M. Figueiredo, *Decreasing weighted sorted ℓ_1 regularization*, IEEE Signal Processing Letters, 21 (2014), pp. 1240–1244.

[22] X. Zeng and M. Figueiredo, *The ordered weighted ℓ_1 norm: atomic formulation, projections, and algorithms*, manuscript, arXiv:1409.4271v3, 2014.