On max-$k$-sums

Michael J. Todd *

September 5, 2016

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

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

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

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

(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 \cite{Rockafellar2000} as well as in certain penalties for peak demands in electricity modelling \cite{Papageorgiou2000}. 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. \cite{Bogdan2013} and Zeng and Figueiredo \cite{Zeng2014, Zeng2015} as regularizers for statistical learning problems. Conversely, OWL norms can be seen as nonnegative combinations of max-$k$-sums.

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 $\mathbb{R}^d$ to be

$$
F^k(t) := M^k(f_1(t), \ldots, f_n(t))
$$

\*School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA.
E-mail mjt7@cornell.edu.
and
\[ f^k(t) := m^k(f_1(t), \ldots, f_n(t)). \] (4)

It is immediate that, even if the \( f_i \)'s are smooth, these derived functions may not be. 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, that we would 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 or 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 \( \nu_i \)-self-concordant barrier for the closed convex set \( C_i \subseteq \mathbb{R}^d \), and \( C := \cap_i C_i \) has a nonempty interior, then \( \sum_i f_i \) is a \( \nu \)-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 [12, 10].) Our hope was that a suitable smoothing of the max-\( d \)-sum of the \( f_i \)'s (where all \( \nu_i \)'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 [13].

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 functions taking \( y \) to \( M^k(y) \) or \( m^k(y) \); composing the result with the function from \( t \in \mathbb{R}^d \) to \( f(t) := (f_1(t), \ldots, f_n(t)) \) will then deal with the general case. Our results are generally 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 \( g : \mathbb{R}^d \to \mathbb{R} \), such as \( F^k \) above, is to take its convolution with a smooth function, i.e., to consider

\[
\tilde{g}(t) := \int_{\mathbb{R}^d} g(t-s)\phi(s)ds,
\]

where \( \phi \) is a smooth probability density function concentrated around 0. However, this smoothing may be hard to compute, and in the barrier case (where \( g \) 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 the \( f_i \)’s. The first method we consider instead “smears” in the range, which is why the discussion above was phrased for max-\(k\)-sums of functions.

(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 with expectation \( E \xi \), and define

\[
M^k(y) := E_{\xi_1, \ldots, \xi_n} \max_{|K|=k} \sum_{i \in K} (y_i - \xi_i) + kE\xi.
\]

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

\[
\tilde{m}^k(y) := E_{\xi_1, \ldots, \xi_n} \min_{|K|=k} \sum_{i \in K} (y_i - \xi_i) + kE\xi.
\]

We record some simple properties of these approximations.

**Proposition 1** We have

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

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

(c) (translation invariance) For \( 0 \leq k \leq n \) and \( \eta \in \mathbb{R} \), for \( \mathbf{1} := (1, \ldots, 1) \) we have

\[ \tilde{M}^k(y + \eta \mathbf{1}) = \tilde{M}^k(y) + k\eta, \tilde{m}^k(y + \eta \mathbf{1}) = \tilde{m}^k(y) + k\eta. \]

(d) (approximation)

\[
M^k(y) \leq \tilde{M}^k(y) \leq M^k(y) + \tilde{m}^k(0) \leq M^k(y) + \min(k\tilde{M}^1(0), -(n-k)\tilde{m}^1(0))
\]

and

\[
m^k(y) \geq \tilde{m}^k(y) \geq m^k(y) + \tilde{m}^k(0) \geq m^k(y) - \min((n-k)\tilde{M}^1(0), -k\tilde{m}^1(0)).
\]

**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

\[
E_{\xi_1, \ldots, \xi_n} \max_{|K|=k} \sum_{i \in K} (y_i - \xi_i) + E_{\xi_1, \ldots, \xi_n} \min_{|L|=n-k} \sum_{i \in L} (y_i - \xi_i) =
\]

\[
E_{\xi_1, \ldots, \xi_n} \left[ \max_{|K|=k} \sum_{i \in K} (y_i - \xi_i) + \min_{|L|=n-k} \sum_{i \in L} (y_i - \xi_i) \right] =
\]

\[
E_{\xi_1, \ldots, \xi_n} \sum_{i} (y_i - \xi_i) = \sum_{i} y_i - nE\xi.
\]
(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 $K$ 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 $M^k(y) - k\bar{E}\xi$. 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,

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

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

$\square$

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

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

(7)

for positive scalars $\alpha$. For negative $\alpha$, 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 $\xi$’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 $\xi_i$ is an independent Gumbel random variable. Let $q_k$ denote the expectation of the $k$th 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 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

$$q_k = \sum_i \sum_{|J|=k-1, i \notin J} \int \Pi_{j \in J} P(y_j - \xi_j \geq y_i - \xi_i) \Pi_{h \notin J, i \in J} P(y_h - \xi_h \leq y_i - \xi_i) \exp(\xi_i - \exp(\xi_i))d\xi_i,$$
Now each term in the first product is 1 − exp(− exp(yj − yi + ξi)) , and each term in the second product is exp(− exp(yh − yi + ξi)). Let us expand the first product, and consider the term where the “1” is taken for j ∈ K, with K a subset of J and hence of cardinality 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. Noting that \( \exp(-\exp(\xi_i)) \) can be written as \( \exp(-\exp(y_i - y_i + \xi_i)) \), we see that

\[
q_k = \sum_{K: |K| < k} \binom{n - |K| - 1}{k - |K| - 1} (-1)^{k-|K|-1} \int \exp(-\exp(y - y_i + \xi_i)) d\xi_i.
\]

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

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

Thus the integral above can be simplified to

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

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 \( \sum_{h \notin K} \exp(z_i) = 1 \). 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 \( \gamma \) 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 \( k - 1 \),

\[
\sum_{j=0}^{p} (-1)^{p-j} \binom{n}{j} \binom{n-1-j}{p-j} = 1.
\]

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.
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.

\[\square\]

From this, we easily obtain

**Theorem 1**  

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

Here \((-1)^0 := 1, \text{ and otherwise } \binom{p}{q} := 0 \text{ if } p < q\).

**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.

\[\square\]

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 [8]. 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 [9] and Bertsekas [2] — and as a potential function in theoretical computer science, starting with Shahrokhi-Matula [15].

We remark that Tunçel and Nemirovskii [16] 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

\[
\tilde{M}^2(y) = -(n - 2) \ln \left( \sum_i \exp(y_i) \right) + \sum_i \ln \left( \sum_{j \neq i} \exp(y_j) \right)
\]

\[
= \ln \left( \sum_{h \neq 2} \exp(y_{[h]}) \right) + \ln \left( \sum_{h \neq 1} \exp(y_{[h]}) \right) + \sum_{i \geq 3} \ln \left( 1 - \frac{\exp(y_{[i]})}{\sum_h \exp(y_h)} \right).
\]

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 $\hat{M}^1$, 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

\[
\tilde{M}^3(y) = \binom{n - 2}{2} \ln \left( \sum_i \exp(y_i) \right) - (n - 3) \sum_i \ln \left( \sum_{j \neq i} \exp(y_j) \right)
\]

\[
+ \sum_{i < j} \ln \left( \sum_{h \neq i,j} \exp(y_h) \right)
\]

\[
= \ln \left( \sum_{h \neq 2,3} \exp(y_{[h]}) \right) + \ln \left( \sum_{h \neq 1,3} \exp(y_{[h]}) \right) + \ln \left( \sum_{h \geq 3} \exp(y_{[h]}) \right)
\]

\[
+ \sum_{1 \leq i \leq 3} \sum_{j \geq 4} \ln \left( 1 - \frac{\exp(y_{[j]})}{\sum_h \exp(y_{[h]})} \right) - \frac{n - 2}{2} \sum_{i \geq 4} \ln \left( 1 - \frac{\exp(y_{[i]})}{\sum_h \exp(y_{[h]})} \right)
\]

\[
+ \frac{1}{2} \sum_{4 \leq i \neq j \leq 4} \ln \left( 1 - \frac{\exp(y_{[j]})}{\sum_h \exp(y_{[h]})} \right).
\]

\[\square\]

In general, evaluating $\tilde{M}^k$ 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 $\hat{m}^{n-k}$, while $\hat{M}^{n-k}$ and $\hat{m}^k$ 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 $\hat{M}^k$, we can use Gumbel random variables scaled by $\alpha$ between 0 and 1. If we denote these by $\tilde{M}^k_\alpha$, we find by (7) that

\[
\tilde{M}^k_\alpha(y) = \alpha \hat{M}^k(y/\alpha),
\]

and by Proposition 1 we find

\[
M^k(y) \leq \tilde{M}^k_\alpha(y) \leq M^k(y) + \alpha k \ln n.
\]
Of course, these results transfer directly to functions: for example, if we define \( \tilde{F}^k(t) := \tilde{M}^k(f_1(t), \ldots, f_n(t)) \), it will be a smooth function of \( t \) and we have
\[
F^k(t) \leq \tilde{F}^k(t) \leq F^k(t) + k \ln n
\]
for all \( t \), and similarly \( F^k(t) \leq \tilde{F}^k_\alpha(t) \leq F^k(t) + \alpha k \ln n \) for \( \tilde{F}^k_\alpha \) defined in the obvious way. We similarly define \( \tilde{f}^k(t) := \tilde{m}^k(f_1(t), \ldots, f_n(t)) \). Then we have:

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

**Proof:** We have
\[
\tilde{F}^k(t) = \mathbb{E}_{\xi_1, \ldots, \xi_n} \max_{|K|=k} \sum_{i \in K} (f_i(t) - \xi_i) + k\mathbb{E}\Xi.
\]
For each \( K \) and fixed \( \xi_i \)'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 \( \tilde{f}^k \). \( \square \)

### 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
\[
P(M^1) : \min \{ x : x \geq y_i \text{ for all } i \}
\]
and
\[
D(M^1) : \max \{ \sum_i u_i y_i : \sum_i u_i = 1, u_i \geq 0 \text{ for all } i \}.
\]
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
\[
D(M^k) : \max \{ \sum_i u_i y_i : \sum_i u_i = k, 0 \leq u_i \leq 1 \text{ for all } i \},
\]
whose dual is
\[
P(M^k) : \min \{ kx + \sum_i z_i : x + z_i \geq y_i, z_i \geq 0, \text{ for all } i \}.
\]

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 $x + \sum z_i$ is feasible for $P(M^1)$ with the same objective value.

Now we introduce perturbations, as in Nesterov [11]. We do this for $D(M^k)$. Let $U$ denote its feasible region, and let $\hat{g}^k$ denote its feasible region, and let

$$\Delta$$

be $\max$ without loss of generality that this $\min$ is 0; and let the

$$U \text{ and } \min$$

over the compact set $P$ be $\Delta$. We sometimes write

$$\hat{g} \text{ instead of } g \text{ will appear in the next section.) Then }$$

$g^*$ 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 $\Delta$. We sometimes write $U^k$ or $g^{*k}$ if we want to stress that they may depend on $k$. Then we let $\hat{M}^k(y)$ be the optimal value of

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

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

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

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 Lipschitz continuous with constant $1/\sigma$.

b) (0- and n-consistency)

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

c) (sign reversal) $\hat{m}^k(y) = -\hat{M}^k(-y)$.

d) (summability) If $g^{\sum_{i=1}^n-i(k)}(u) := g^k(1 - u)$, where $1 := (1, \ldots, 1)$,

$$\hat{M}^k(y) + \hat{m}^{\sum_{i=1}^n-i(k)}(y) = \sum y_i.$$  

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

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

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

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

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 [11], while part (g) is a direct consequence of the bounds on \( g \), and also appears in [11]. 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 \( \hat{g}^* \) evaluated at the unique feasible point is zero. For part (d), note that

\[
\hat{M}^k(y) - \sum_i y_i = \max \left\{ \sum_i (u_i - 1)y_i - g^*(u) : u \in U^k \right\}
\]

\[
= \max \{ - \sum_i (1 - u_i)y_i - g^{\ast_{n-k}}(1 - u) : 1 - u \in U^{n-k} \}
\]

\[
= - \min \{ \sum_i u_i y_i + g^{\ast_{n-k}}(v) : v \in U^{n-k} \} = - \hat{m}^{n-k}(y).
\]

\[\square\]

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 \( \hat{D}(\hat{M}^k(f_1((1 - \lambda)s + \lambda t), \ldots, f_n((1 - \lambda)s + \lambda t))) \), we obtain

\[
\hat{F}^k((1 - \lambda)s + \lambda t) = \sum_i u_i f_i((1 - \lambda)s + \lambda t) - g^*(u)
\]

\[
\leq (1 - \lambda) \left( \sum_i u_i f_i(s) - g^*(u) \right) + \lambda \left( \sum_i u_i f_i(t) - g^*(u) \right)
\]

\[
\leq (1 - \lambda) \hat{F}^k(s) + \lambda \hat{F}^k(t),
\]

since \( u \) is also feasible for the problems for \( s \) and for \( t \), establishing convexity. \( \square \)

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

\[
M^k(y) - \alpha \Delta \leq \hat{M}_\alpha^k(y) \leq M^k(y),
\]

and analogously for a scaled version of \( \hat{m}^k \).

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

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

(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):

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

If we choose \( g^* \) to be identically zero on \( \text{aff}(U) \), which satisfies all of our requirements except strongly convexity, then \( g(\lambda 1) = k\lambda \) for \( \lambda \in \mathbb{R} \) 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 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 take

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

on \( \text{aff}(U) \), where \( \beta > 0 \). (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 \( g^k(u) = (\beta/2)(\|u - (k/n)1\|)^2 \), so that \( g^{*n-k}(v) = g^k(1-v) \) 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

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

\[ x \in \mathbb{R}, \sum_i u_i = k, s_i \geq 0, u_i \geq 0, s_i u_i = 0, z_i \geq 0, u_i \leq 1, z_i (1 - u_i) = 0, \quad \text{for all } i. \tag{8} \]

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

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

where \( \text{mid}(\cdot, \cdot, \cdot) \) denotes the middle of its three arguments. Note that each \( u_i \) is a non-increasing function of \( \lambda \) equal to 1 for sufficiently small and to 0 for sufficiently large \( \lambda \). We seek a value of \( \lambda \) so that \( \sum_i u_i = k \). We therefore sort the \( 2n \) numbers \( y_i/\beta \) and \( y_i/\beta - 1 \). Suppose for simplicity that the \( y_i \)'s are in nonincreasing order. Then for \( \lambda \) 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 \( \lambda \) between \( y_{k+1}/\beta \) 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

\[ (h - 1) + \sum_{j} y_i/\beta - (j - h + 1)\lambda, \]

so that if this is the correct subinterval, \( \lambda \) is equal to \( [(k - h + 1 - \sum_{j} Y_i/\beta)/[j - h + 1] \). 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 \( \lambda \) 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 \( \sum_i y_i/\beta \) 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 \( \beta \), then the exceptional case above occurs for any \( k \), and the resulting \( u \) puts a weight 1 on the \( k \) largest \( y_i \)'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 \( g^*(w) = (\|w - \bar{w}1\|^2/[2\beta] + k\bar{w} + \beta(k)^2/[2n] \), where \( \bar{w} = \sum_i w_i/n \), so using the fact that we can assume \( \sum_i w_i = 0 \), we have

\[
\hat{P}(M^k) : \min \{kx + \frac{1}{2\beta}(\|w\|^2 + \frac{\beta(k)^2}{2n}) : x + z_i \geq y - w_i, z_i \geq 0 \forall i, \sum_i w_i = 0\}
\]
and
\[ D(M^k) : \max_{\sum_i u_i y_i = k} \left\{ \sum_i u_i y_i - \frac{\beta}{2} (\|u\|)^2 + \frac{\beta(k)^2}{2n} : \sum_i u_i = k, 0 \leq u_i \leq 1 \text{ for all } i \right\}. \]

### 3.2 Single-sided entropic prox-function

Next we choose
\[ g^*(u) := g^k(u) := \sum_i u_i \ln u_i + k \ln \left( \frac{n}{k} \right) \]
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 [11]) 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 [11]; for this and the next subsection, we need to use the \( L_1 \)-norm on \( U^k \)). Now we do not have \( g^{n-k}(u) = g^k(1-u) \), but we could use the alternative single-sided entropic prox-function \( \tilde{g}^k(u) := g^{n-k}(1-u) \) for \( \tilde{n}^k \) 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 (8), so that with \( \lambda = x + 1 \), we obtain
\[ u_i = \min(\exp(y_i - \lambda), 1) \text{ for all } i, \]
where again \( \lambda \) 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 \( h - 1 \) \( u_i \)'s are one, and the others equal to \( (k - h + 1) \exp(y_h)/\sum_i^n \exp(y_j) \), 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

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

### 3.3 Double-sided entropic prox-function

Finally we choose

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

which is again appropriate for \( U^k \). Then \( \sigma = 2/k \) and \( \Delta = k \ln (n/k) + (n - k) \ln (n/[n - k]) \). We have regained the property that \( g^{*n-k} (u) = g^{*k} (1 - u) \), so the summability property holds with no adjustments.

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

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

together with (8), so that with \( \lambda := \exp(x) \), we have

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

Now we have to make a nonlinear search to find the appropriate \( \lambda \) 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

\[
\hat{D}(M^k) : \max \sum_i u_i y_i - \sum_i [u_i \ln u_i + (1 - u_i) \ln (1 - u_i)] - k \ln \left( \frac{n}{k} \right) - (n - k) \ln \left( \frac{n}{n - k} \right)
\]

\[
\sum_i u_i = k, \quad 0 \leq u_i \leq 1 \quad \text{for all } i.
\]

### 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 \( K \) be a closed convex cone in \( E \) that is pointed (\( K \cap (-K) = \{0\} \)) and has a nonempty interior. Then its dual cone \( K^* \subseteq E^* \), defined by

\[
K^* := \{ u \in E^* : \langle u, x \rangle \geq 0 \text{ for all } x \in K \},
\]

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

\[
x \geq z, x, z \in E \text{ means } x - z \in K, \quad u \geq^* v, u, v \in E^* \text{ means } u - v \in K^*.
\]
Then corresponding eigenspaces. Finally, in (d), we set \((\xi X)\) decomposition of any \(x\) is immediate. In example (c) we let \(\xi\) the space of real (complex) symmetric (Hermitian) \(d \times d\) matrices, and the subset of positive semidefinite matrices; and

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

### 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 \(K\) and \(K^*\)) and homogeneous (for all \(x, y \in \text{int}(K)\), there is an automorphism of \(K\) 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 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 \(\lambda\)'s are the eigenvalues of \(x\). It is easy to see that the cone of squares \(K\) consists of those elements with nonnegative eigenvalues, and its interior those with positive eigenvalues. We can define the trace of an element as the sum of its eigenvalues 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: indeed if \(x, z \in K\), \((x, z) = 0\) 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 \(\|z\| = 1\) 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 $z = x/\|x\|$, 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.

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 $\text{int}(\mathcal{K})$, 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 [7] and Koranyi [6] 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 $v \in \text{int}(\mathcal{K}^*)$, and define

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

and

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

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 $g^*$ be a closed proper smooth $\sigma$-strongly convex prox-function on $U$ ($\sigma > 0$). We allow $g^*$ to be finite on some subset of $\text{aff}(U) = \{(u_i) \in E^{*n} : \sum_i u_i = kv\}$ containing
Similarly, if \((u_i), (w_i)\) := \(\sum_i \langle u_i, w_i \rangle\). Then we can define perturbed versions of the problems above:

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

and

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

If \(g^*\) is identically 0, so that \(g((w_i)) = k\langle v, \bar{w} \rangle\) 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

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

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

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

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

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

with equality if and only if

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

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

\[
k\langle v, x \rangle + \sum_i \langle v, z_i \rangle + g((w_i)) - \sum_i \langle u_i, y_i \rangle + g^*((u_i)) \geq \sum_i \left[ \langle v - u_i, z_i \rangle + \langle u_i, x + z_i - y_i \rangle \right] + \sum_i \langle u_i, w_i \rangle \geq 0,
\]
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 \( \text{int}(K) \), and then choose \( \lambda \) with \( \lambda \) positive and sufficiently large that \( \lambda \bar{x} \pm \eta_i \in \text{int}(K) \) 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, \( (-\bar{x}, (\bar{x} + y_i)) \) 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 that we can restrict \( (w_i) \) to sum to zero in the primal problem. Indeed, 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

\[
g((w_i)) = \max_u \left\{ \sum_i \langle u_i, \bar{w} + w'_i \rangle - g^*((u_i)) \right\}
= k \langle v, \bar{w} \rangle + \max_u \left\{ \sum_i \langle u_i, w'_i \rangle - g^*((u_i)) \right\}
= k \langle v, \bar{w} \rangle + g((w'_i)),
\]

and 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 \( \sum_i w'_i = 0 \).

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

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

and

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

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

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

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)

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

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

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

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

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 \(\min\{k(v, x) + \sum_i(v, s_i) + g((w_i)) : -x + s_i \geq -y_i - w_i, s_i \geq 0, \text{ for all } i\}\), 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 (10) gives (c). If the stated condition holds, and \(\sum_i w_i = 0\), then (10) gives (c). If the stated condition holds, and \(\sum_i w_i = 0\), then (10) gives (c).

Finally, for any \(K\) of cardinality \(k\), and any feasible solution \((x, (z_i))\) to \(P(M^k(y))\), \(kx + \sum_i z_i \geq kx + \sum_{i \in K} z_i \geq \sum_{i \in K} y_i\). 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.

\[\square\]

Here is a small example to show that \(M^k(y)\) need not be a singleton. Let \(E = E^* = \mathbb{R}^3\) with \(\langle u, x \rangle := u^T x\). Let \(K := \{x \in E : x_1 \geq |x_2|, x_1 > |x_3|\}, v = (1, 0; 0), n = 2, y_1 = (0; 1, 0), y_2 = (0; -1, 0). Then \(M^1((y_i)) = \{(1, 0; \xi) : -1 \leq \xi \leq 1\}\). 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 \(E = \mathbb{R}^m\) and \(K = \mathbb{R}_+^m\), 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 \(K\) 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 \(E = E^* = \mathbb{R}^3\) as above, but now \(K := \{x \in E : x_1 \geq \|(x_2; x_3)\|_2\}\), 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, 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 we have $M^1(M^1(y_1,y_2),y_3) = M^1(y_1,y_3) = y_1$. 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 $K$. 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 $\beta$ to be a positive self-adjoint operator from $E^*$ to $E$, and define $g^*((u_i)) := (1/2)\sum_i(\langle \beta u_i, \beta u_i \rangle) - (k^2/2)\sum_i\langle v_i, v_i \rangle$ on $aff(U)$ and $+\infty$ otherwise. Then it is easy to see that $g((u_i)) = (1/2)\sum_i\langle \beta^{-1}w_i - \bar{w}, w_i - \bar{w} \rangle + k\langle v, \bar{w} \rangle + (k^2/2)\langle v, v \rangle$, where $\bar{w} := \sum_i w_i/n$. (Compare with Subsection 3.1.) Then we need to solve a quadratic cone problem over $K$, but now there is no easy search to find optimal solutions. However, we note that for this prox-function,

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

on $aff(U^k)$, so $g^{*n-k}((u_i)) = g^{*k}((v - u_i))$ on $aff(U^{n-k})$, 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 $(x,z) := \text{trace}(xyz)$, and so we can identify $E^*$ with $E$. We therefore choose $v$ in $E$, and a natural choice is to set $v := e$, the unit element. Also, all $u_i$'s now lie in $E$.

An important consequence of $K$ 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 $g$ is strictly convex as above if and only if $g^*$ is smooth on the relative interior of its domain in $aff(U)$.)

**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 (9), $(e - u_i, z_i) = (u_i, x + z_i - y_i) = 0$ for all $i$. Also, $e - u_i, z_i, u_i$, and $x + z_i - y_i$ all lie in $K$, 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

$$\sum_i u_i \circ y_i = \sum_i u_i \circ (x + z_i) = (\sum_i u_i) \circ x + \sum_i u_i \circ z_i = k e \circ x + \sum_i e \circ z_i = k x + \sum_i z_i.$$  

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

20
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 as above. 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, 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 [4, 6].

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 idempotents $\{e_1, \ldots, e_m\}$ such that all $y_i$’s are linear combinations of these, so that

$$y_i = \lambda_{i1} c_1 + \cdots + \lambda_{im} c_m$$

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

$$\lambda_{i1} c_1 + \cdots + \lambda_{im} c_m,$$

where each $\lambda_j$ is the max-$k$-sum of the $\lambda_{ij}, i = 1, \ldots, n$. This is not hard to show by duality, with each $u_i$ of the form $\mu_1 c_1 + \cdots + \mu_m c_m$ for an appropriate $(0,1)$-vector of coefficients. The situation is analogous to that where $E = R^m$, $K = R^m_+$. 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 $v \in \text{int}(K)$. Indeed, because $K$ is homogeneous, there is a self-adjoint operator $\phi$ taking $e$ to $v$ and $K$ onto itself. Then if we seek $M^k_e((y_i))$ (where the subscript indicates that $e$ has been replaced by $v$), we note that $(x, (z_i)) \in \text{Argmin}(P(M^k_e((y_i))))$ iff $(\phi x, (\phi z_i)) \in \text{Argmin}(P(M^k((\phi y_i))))$, since $(\phi e, kx + \sum_i z_i) = (e, k\phi x + \sum_i \phi z_i)$, and the feasibility conditions are equivalent. Hence

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

### 4.4 Final notes

We might consider

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

21
if the limit exists, as a definition of the maximum of $y_1, \ldots, y_n$. Does this limit exist? Does it equal $M^1((y_i))$?

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


