Smoothing max-k-sum functions

Michael J. Todd

January 3, 2016

School of Operations Research and Information Engineering, Cornell University

http://people.orie.cornell.edu/∼miketodd/todd.html

U.S.-Mexico Optimization Workshop, Merida, January 2016

1. Problem

Given: $f_1, \ldots, f_n : \Re^d \to \bar{\Re} := \Re \cup \{\infty\}$ smooth (and convex),

find smooth (and convex) approximations to

$$f^1(x) := \max_i \{f_i(x)\}$$

and

$$f^k(x) := \max_{|K|=k} \sum_{i \in K} f_i(x).$$

This is related to Owl norms (Bogdan et al., Zeng and Figueiredo). For a vector $y \in \Re^n$, let $y_{[k]}$ denote the kth largest component. Then if $f(x) \geq 0$ and $w_1 \geq w_2 \geq \cdots \geq w_n \geq w_{n+1} := 0$,

$$\Omega_w(f(x)) = \sum_h w_h f_{[h]}(x) = \sum_{j=1}^{n} (w_j - w_{j+1}) f^j(x).$$

This includes the ℓ_1 ($w_j = 1$ all j) and ℓ_∞ ($w_1 = 1, w_2 = \cdots = w_n = 0$) norms.

Motivation

Let f_i be a (self-concordant) barrier function for a closed convex set C_i in $\Re^d$, with barrier parameter ν_i, $i = 1, \ldots, n$.
Example: $-\ln(b_i - a_i^T x)$ for $\{x \in \Re^d : a_i^T x \leq b_i\}$ with $\nu_i = 1$.

Then $\sum_{i=1}^n f_i(x)$ is a barrier function for $C := \cap_{i=1}^n C_i$, with barrier parameter $\nu := \sum_{i=1}^n \nu_i$.
Example: $-\sum_{i=1}^n \ln(b_i - a_i^T x)$ for the polyhedron $P := \{x \in \Re^d : Ax \leq b\}$ with $\nu = n$.

Perhaps a suitable smooth approximation to f^d can also be used as a barrier function, with a smaller barrier parameter.
We know there is a barrier function for the set C with barrier parameter of order d (but it is not easily computable).

Barrier functions can be used in the efficient optimization of a linear function over the corresponding set, and the complexity depends on the barrier parameter.

2. First idea

"Smear" in the domain:
Approximate a nonsmooth function f via a convolution or as an expectation:

$$\begin{aligned}\hat{f}(x) &:= E_z f(x-z) \\ &= \int f(x-z)\phi(z)dz,\end{aligned}$$

where ϕ is the probability density function of a localized random variable $z \in \Re^d$.

However, this shrinks the domain dom $f := \{x : f(x) < \infty\}$, inappropriate for a barrier function.

3. Our idea

"Smear" in the range: Let $\xi_1, \ldots, \xi_n$ be iid random variables and set

$$\bar{f}^1(x) := E_{\xi_1,\ldots,\xi_n} \max_i \{f_i(x) - \xi_i\} + E\xi,$$

$$\bar{f}^k(x) := E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} (f_i(x) - \xi_i) + kE\xi.$$

These functions inherit the smoothness (and convexity) of the f_i's. Moreover, they inherit the domains of the nonsmooth functions. To enable fairly efficient evaluation, we choose Gumbel random variables: $P(\xi > x) = \exp(-\exp(x))$, $E\xi = -\gamma$.

4. Evaluation

We are interested in $q_k := E((f(x) - \xi)_{[k]})$.

$$q_k = \sum_{i \notin J, |J|=k-1} \int_{\xi_i} \Pi_{j \in J} P(f_j(x) - \xi_j \geq f_i(x) - \xi_i) \cdot$$

$$\Pi_{h \neq i, h \notin J} P(f_h(x) - \xi_h \leq f_i(x) - \xi_i) \cdot (f_i(x) - \xi_i) \cdot \exp(\xi_i - e^{\xi_i}) d\xi_i$$

$$= \ldots$$

$$= \sum_{|K|<k} (-1)^{k-|K|-1} \binom{n- \mid K \mid -1}{k- \mid K \mid -1} \ln \sum_{i \notin K} \exp(f_i(x)) + \gamma.$$

(Here $\binom{0}{0} := 1$.) But $\bar{f}^k(x)$ is just the sum of the first k $(q_j - \gamma)$'s. Since the alternating sum of the binomial coefficients simplifies, we get:

Theorem 1

$$\bar{f}^k(x) = \sum_{|K|<k} (-1)^{k-|K|-1} \binom{n- \mid K \mid -2}{k- \mid K \mid -1} \ln \sum_{i \notin K} \exp(f_i(x)).$$

□

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

We have reduced the work from an n-dimensional integration to a sum over $O(n^{k-1})$ terms.

Note that almost all the terms disappear for $k = n$, and we get $\bar{f}^n(x) = f^n(x)$ as expected.

5. Examples

$k = 1$: Here only $K = \emptyset$ contributes to the sum, so we obtain

$$\bar{f}^1(x) = \ln \left(\sum_i \exp(f_i(x)) \right).$$

Such functions have been used as potential functions in theoretical computer science, starting with Shahrokhi-Matula and Grigoriadis-Khachiyan in the 1990s. They also appear in the economic literature on consumer choice, dating back to the 1960s (e.g., Luce and Suppes).
This function is sometimes called the soft maximum of the f_i's. This term is also used for the weight vector

$$\left(\frac{\exp(f_i(x))}{\sum_h \exp(f_h(x))} \right).$$

Note that the gradient of $\bar{f}^1$ is the weighted combination of those of the f_i's using these weights.

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

$$\begin{aligned}\bar{f}^2(x) &= -(n-2)\ln\left(\sum_i \exp(f_i(x))\right) + \sum_i \ln\left(\sum_{j\neq i}\exp(f_j(x))\right) \\ &= \ln\left(\sum_{h\neq 2}\exp(f_{[h]}(x))\right) + \ln\left(\sum_{h\neq 1}\exp(f_{[h]}(x))\right) + \\ &\quad \sum_{i>2}\ln\left(1 - \frac{\exp(f_{[i]}(x))}{\sum_h \exp(f_h(x))}\right).\end{aligned}$$

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

$$\begin{aligned} \bar{f}^3(x) &= \binom{n-2}{2} \ln\left(\sum_i \exp(f_i(x))\right) - (n-3)\sum_i \ln\left(\sum_{j\neq i} \exp(f_j(x)\right) \\ &+ \sum_{i<j} \ln\left(\sum_{h\neq i,j} \exp(f_h(x))\right) \\ &= \ln\left(\sum_{h\neq 2,3} \exp(f_{[h]}(x))\right) + \ln\left(\sum_{h\neq 1,3} \exp(f_{[h]}(x))\right) + \ln\left(\sum_{h>2} \exp(f_{[h]}(x))\right) \\ &+ \sum_{1\leq i\leq 3} \sum_{j>3} \ln\left(1 - \frac{\exp(f_{[j]}(x)}{\sum_{h\neq i} \exp(f_{[h]}(x))}\right) \\ &+ \frac{1}{2} \sum_{3<i\neq j>3} \ln\left(1 - \frac{\exp(f_{[j]}(x))}{\sum_{h\neq i} \exp(f_{[h]}(x))}\right) - \frac{n-2}{2} \sum_{i>3} \ln\left(1 - \frac{\exp(f_{[i]}(x))}{\sum_h \exp(f_h(x))}\right). \end{aligned} \tag{2}$$

6. Bounds

Theorem 2

$$f^k(x) \le \bar{f}^k(x) \le f^k(x) + k \ln n.$$

Proof

$$\begin{aligned}
\bar{f}^k(x) &= E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} (f_i(x) - \xi_i) + kE\xi \\
&\ge E_{\xi_1,\ldots,\xi_n} \sum_{i \in \hat{K}} (f_i(x) - \xi_i) + kE\xi \\
&= f^k(x) - kE\xi + kE\xi = f^k.
\end{aligned}$$

$$\begin{aligned}
\bar{f}^k(x) &= E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} (f_i(x) - \xi_i) + kE\xi \\
&\le E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} f_i(x) + E_{\xi_1,\ldots,\xi_n} \max_{|K|=k} \sum_{i \in K} (-\xi_i) + kE\xi \\
&\le f^k(x) + kE_{\xi_1,\ldots,\xi_n} \max_i (-\xi_i) + kE\xi \\
&= f^k(x) + k \ln n.
\end{aligned}$$

□

7. Final remarks

If we want a closer (but "rougher") approximation, we can scale the Gumbel random variables by $\alpha < 1$, or equivalently, scale the functions f_i by α^{-1}, apply the formulae above, and then scale the result by α.

If the f_i's differ by orders of magnitude, the above expressions need to be carefully evaluated, but at the same time, we may be able to ignore many of the terms.

Still to do: study further properties of these smooth approximations.
(Unfortunately, even $\bar{f}^1$ turns out not to be self-concordant in the case of log barrier functions for linear constraints (Tuncel-Nemirovsky).)

Interesting fact: for symmetric matrices F_i,

$$\ln\left(\sum_h \exp(F_h)\right) \succeq F_i, \quad i = 1, ..., m.$$

What is $\lim_{\alpha \downarrow 0} \alpha \ln \left(\sum_i \exp(\alpha^{-1} F_i)\right)$?