Smoothing maximum functions

Michael J. Todd

October 2, 2014

School of Operations Research and Information Engineering, Cornell University

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

ICERM, October 2014
1. **Problem**

**Given:** \( f_1, \ldots, f_n : \mathbb{R}^d \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\} \) 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 (Rob Nowak, Monday): if \( f(x) \geq 0 \) and \( w_1 \geq w_2 \geq \cdots \geq w_n \geq w_{n+1} := 0 \),

\[
  \Omega_w(f(x)) = \sum_{j=1}^{n} (w_j - w_{j+1}) f^j(x).
\]
Motivation

Let $f_i$ be a (self-concordant) barrier function for a closed convex set $C_i$ in $\mathbb{R}^d$, with barrier parameter $\nu_i$, $i = 1, \ldots, n$.

Example: $-\ln(b_i - a_i^T x)$ for $\{x \in \mathbb{R}^d : a_i^T x \leq b_i\}$ with $\nu_i = 1$.

Then $\sum_{i=1}^n f_i(x)$ is a barrier function for $C := \bigcap_{i=1}^n C_i$, 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 $P := \{x \in \mathbb{R}^d : Ax \leq b\}$ 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:

\[ \hat{f}(x) := E_y f(x - z) \]
\[ = \int f(x - z) \phi(z) dz, \]

where $\phi$ is the probability density function of a localized random variable $z \in \mathbb{R}^d$.

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

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

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

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

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

For a vector $y \in \mathbb{R}^n$, let $y[k]$ denote the $k$th largest component. We are interested in $q_k := E((f - \xi)[k])$.

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

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

= \ldots

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

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

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

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

Such functions have been used as potential functions in theoretical computer science, starting with Shahrokhi-Matula and Grigoriadis-Khachiyan. They also appear in the economic literature on consumer choice, dating back to the 1960s (e.g., Luce and Suppes). See also the smoothing of the absolute value function (John Wright, Tuesday).
\( k = 2 \): Here \( K \) can be the empty set or any singleton, and we find

\[
\bar{f}^2(x) = \bar{f}^1(x) + \ln \left( \sum_{i>1} \exp(f_i(x)) \right) + \\
\sum_{i>1} \ln \left( 1 - \frac{\exp(f_i(x))}{\sum_h \exp(f_h(x))} \right) \\
= \ln \left( \sum_{i \neq 2} \exp(f_i(x)) \right) + \\
\ln \left( \sum_{i \neq 1} \exp(f_i(x)) \right) + \\
\sum_{i>2} \ln \left( 1 - \frac{\exp(f_i(x))}{\sum_h \exp(f_h(x))} \right).
\]
\( k = 3 \): We have

\[
\overline{f}^3(x) = \overline{f}^2(x) + \ln \left( \sum_{i>2} \exp(f_{[i]}(x)) \right) + \sum_{1<i<j} \ln \left( 1 - \frac{\exp(f_{[i]}(x)) + \exp(f_{[j]}(x))}{\sum_h \exp(f_h(x))} \right) + \sum_{j>2} \left[ \ln \left( 1 - \frac{\exp(f_{[j]}(x))}{\sum_{h>1} \exp(f_{[h]}(x))} \right) - \sum_{i>1} \ln \left( 1 - \frac{\exp(f_{[i]}(x))}{\sum_h \exp(f_h(x))} \right) \right]
\]
\[
= \ln \left( \sum_{i \neq 2, i \neq 3} \exp(f_i(x)) \right) + \\
\ln \left( \sum_{i \neq 1, i \neq 3} \exp(f_i(x)) \right) + \\
\ln \left( \sum_{i > 2} \exp(f_i(x)) \right) + \\
\sum_{1 < i < j > 3} \ln \left( 1 - \frac{\exp(f_i(x)) + \exp(f_j(x))}{\sum_h \exp(f_h(x))} \right) + \\
\sum_{j > 3} \left[ \ln \left( 1 - \frac{\exp(f_j(x))}{\sum_{h > 1} \exp(f_h(x))} \right) - \sum_{i > 1} \ln \left( 1 - \frac{\exp(f_i(x))}{\sum_h \exp(f_h(x))} \right) \right].
\]
6. 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 $\alpha^{-1}$, apply the formulae above, and then scale the result by $\alpha$.

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 the properties of these smooth approximations.

We have

$$f^k(x) \leq \bar{f}^k(x) \leq f^k(x) + k \ln n,$$

(the upper bound is close to tight if the $f_i(x)$’s are close), but we want more information on their derivatives of order up to three.