More Virtuous Smoothing

Luze Xu[1], Jon Lee[1], Daphne Skipper[2]

[1]Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI, USA

[2]Department of Mathematics
U.S. Naval Academy, Annapolis, MD, USA

Optimization and Discrete Geometry : Theory and Practice
24-27 April, 2018
Tel Aviv University, Israel

10 years ago, Shmuel organized a very nice discrete optimization (and hummus) day at Haifa

Discrete Optimization Day, 13:30--18:00, Tuesday, August 5, 2008, Bloomfield 526, Technion

- **13:30--14:15, Robert Weismantel, "Mixed Integer Optimization: a Geometric View" (slides)**
- **14:15--14:45, Yael Berstein, "Randomized Algorithms for some Nonlinear Combinatorial Optimization Problems" (slides)**
- **15:00--15:45, Jon Lee, "Matrix Computing for Combinatorial Optimization" (slides)**
- **15:45--16:15, Tal Raviv, "Efficient Algorithm for Maximizing the Expected Profit from a Serial Production Line with Inspection Stations and Rework" (slides)**
- **16:45--17:30, Antoine Deza, "Polytopes and Arrangements: Diameter and Curvature" (slides)**
- **17:30--18:00, Uri Rothblum, "Accuracy Certificates for Computational Problems with Convex Structure" (slides)**

Overview

1. Introduction
   - Motivation
   - Prior Work
   - Definition of δ-smoothing
2. General behaviors of δ-smoothing
   - Increasing and concave
   - Controlled derivative at 0
   - Monotonicity of $g_1 = g'(0)$ in δ
3. Lower bound for f
   - Lower bounding
   - Role of increasing and concave
4. Better Bound

Table of Contents

Motivation

- Most Mixed-Integer Nonlinear Optimization (MINLO) software, aiming at global optimization of factorable mathematical-optimization formulations, apply spatial branch-and-bound or a variant (e.g., SCIP, Baron, Couenne, Antigone. . .)
- As a first step, problem functions are "factored" (i.e., fully decomposed) via a small library of low-dimensional nonlinear functions (e.g., $\sin(x)$, $\log(x)$, a^x, x/y, xy, xyz, x^p $(0 < p < 1)$. . .) together with affine functions.
- It is helpful, for robustness, if the library functions are sufficiently smooth over their domains, i.e., typically twice continuously differentiable, so that typical nonlinear-optimization algorithms may be reliably applied (e.g., Wächter and Biegler [2006]).

Motivation

The issue can also be grappled with algorithmically by (purely continuous) nonlinear-optimization solvers through parameter setting. For example, Wächter explains:

> *"Problem modification: Ipopt seems to perform better if the feasible set of the problem has a nonempty relative interior. Therefore, by default, Ipopt relaxes all bounds (including bounds on inequality constraints) by a very small amount (on the order of* 10^{-8}*) before the optimization is started. In some cases, this can lead to problems, and this features can be disabled by setting bound_relax_factor to 0."*

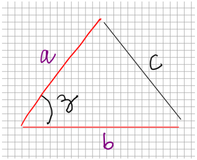
Motivation

Consider $f(w) := \sqrt{w}$ on the domain $[0, +\infty)$. Notice how in this case Ipopt's default value for this parameter `bound_relax_factor` is not robust for even function evaluation, on the modified domain $[-10^{-8}, +\infty)$. And for the suggested nondefault parameter setting (0), $\sqrt{w}$ is not differentiable at 0 (in the actual domain). So, we are led back to modeling advice:

> *"Therefore, it can be useful to replace the argument of a function with a limited range of definition by a variable with appropriate bounds. For example, instead of "$\log(h(x))$", use "$\log(y)$" with a new variable $y \geq \epsilon$ (with a small constant $\epsilon > 0$) and a new constraint $h(x) - y = 0$."*

We note that this kind of advice might be problematic in the context of integer variables, where precise zero may be important in constraints implementing some logic, and for this reason, our study is particularly relevant to MINLO.

Example, ESTP



Lemma (simple)

Among triangles with edge lengths a, b, c and corresponding (opposite) angles x, y, z, with c and z fixed, the one maximizing $a + b$ is isosceles (that is $a = b$, $x = y$).

Theorem ([D'Ambrosio, Fampa, Lee, and Vigerske, 2015, Theorem 9])

For all $n \geq 2$, we have

$$y_{ik} y_{jk} \left(\|x^k - a^i\| + \|x^k - a^j\| \right) \leq \frac{2}{\sqrt{3}} \|a^i - a^j\|, \ \forall\ i, j \in P, \ i < j, \ k \in S.$$

Motivation

The motivating application for our work is root functions $f(w) := w^p$, with $0 < p < 1$, which are smooth everywhere on their domains $[0, +\infty)$, except at $w = 0$.

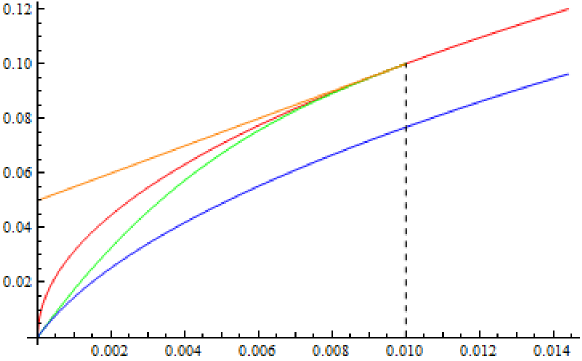
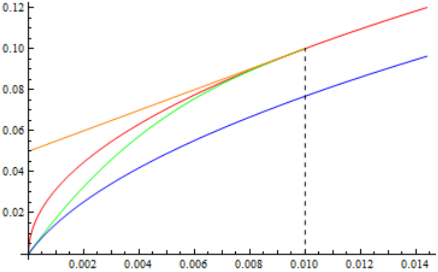


Figure: given function $f := w^p$, Linear extrapolation, Shift: $\sqrt{w+\lambda} - \sqrt{\lambda}$, our smoothing g

Motivation



- Linear extrapolation is bad because it is far from w^p when w is near 0, and it is not twice differentible at δ;
- Shift smoothing is twice differentiable everywhere, follows the general trends of f (i.e., concave, increasing), but it is bad because it is not close to w^p when w is not near 0;
- Our Smoothing has the nice properties of the shift, but aims to find a better lower bound.

Prior Work

- The inception of this approach is from D'Ambrosio, Fampa, Lee, and Vigerske [2014], which grappled with handling square-root functions ($p = 1/2$) arising in formulations of the Euclidean Steiner Problem.
- That successful approach was to replace the part of the root function on $[0, \delta]$, for some small (but not extremely small) $\delta > 0$, with a homogeneous cubic, matching the function and its first two derivatives at δ.
- We showed that the new piecewise function g is (i) increasing and concave, (ii) underestimates the square root, and (iii) dominates the simple shift smoothing $h(w) := \sqrt{w + \lambda} - \sqrt{\lambda}$, when the parameters δ (for g) and λ (for h) are chosen "fairly" — i.e., so that $g'(0) = h'(0)$, and hence both smoothing have the same numerical stability.

Prior Work

- Lee and Skipper [2017] extended this idea of D'Ambrosio et al. [2014, 2015], with the following main results:
  (i) a rather general sufficient condition on f (which includes all root functions and more) so that our smoothing g is increasing and concave;
  (ii) for root functions of the form $f(w) = w^{1/q}$, with integer $q \geq 2$, our smoothing g underestimates f;
  (iii) for root functions of the form $f(w) = w^{1/q}$, with integer $2 \leq q \leq 10,000$, our smoothing g 'fairly dominates' the shift smoothing h; i.e., when g and h are chosen so that $g'(0) = h'(0)$.

Prior Work

- Lee and Skipper [2017] extended this idea of D'Ambrosio et al. [2014, 2015], with the following main results:
  - (i) a rather general sufficient condition on f (which includes all root functions and more) so that our smoothing g is increasing and concave;
  - (ii) for root functions of the form $f(w) = w^{1/q}$, with integer $q \geq 2$, our smoothing g underestimates f;
  - (iii) for root functions of the form $f(w) = w^{1/q}$, with integer $2 \leq q \leq 10,000$, our smoothing g 'fairly dominates' the shift smoothing h; i.e., when g and h are chosen so that $g'(0) = h'(0)$.
- We generalize all of these results, using very different techniques: ~~algebra~~ analysis

Definition

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Suppose that $\delta > 0$ is in the domain of f.

We say that such an f satisfies the minimal δ-smoothing requirements if $f(0) = 0$, and f is twice differentiable at δ.

Definition

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Suppose that $\delta > 0$ is in the domain of f.

We say that such an f satisfies the minimal δ-smoothing requirements if $f(0) = 0$, and f is twice differentiable at δ.

Suppose that such an f satisfies the minimal δ-smoothing requirements. Then the δ-smoothing of f is the piecewise-defined function

$$g(w) := \begin{cases} g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3, & 0 \leq w \leq \delta; \\ f(w), & \delta < w < U, \end{cases}$$

having

$$g(0) = f(0); g(\delta) = f(\delta); g'(\delta) = f'(\delta); g''(\delta) = f''(\delta).$$

Definition

Suppose that such an f satisfies the minimal δ-smoothing requirements. Then the δ-smoothing of f is the piecewise-defined function

$$g(w) := \begin{cases} g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3, & 0 \le w \le \delta; \\ f(w), & \delta < w < U, \end{cases}$$

with

$$\begin{aligned} g_1 &= \frac{3f(\delta)}{\delta} - 2f'(\delta) + \frac{\delta f''(\delta)}{2} && (= g'(0)); \\ g_2 &= -\frac{6f(\delta)}{\delta^2} + \frac{6f'(\delta)}{\delta} - 2f''(\delta) && (= g''(0)); \\ g_3 &= \frac{6f(\delta)}{\delta^3} - \frac{6f'(\delta)}{\delta^2} + \frac{3f''(\delta)}{\delta} && (= g'''(w),\ \text{for } w \in [0, \delta]). \end{aligned}$$

Preliminary

$$g(w) := \begin{cases} g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3, & 0 \le w \le \delta; \\ f(w), & \delta < w < U, \end{cases}$$

- First, we will show that under certain conditions, g is increasing and concave when f is.
- This is useful in global optimization because we can use tangents to overestimate g and secants to underestimate g.
- SCIP has a new feature to allow user (through AMPL suffixes) to tell SCIP that such a g is, for example, increasing and concave.

Preliminary

$$g(w) := \begin{cases} g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3, & 0 \le w \le \delta; \\ f(w), & \delta < w < U, \end{cases}$$

- First, we will show that under certain conditions, g is increasing and concave when f is.
- This is useful in global optimization because we can use tangents to overestimate g and secants to underestimate g.
- SCIP has a new feature to allow user (through AMPL suffixes) to tell SCIP that such a g is, for example, increasing and concave.
- We will show that under certain conditions, the first and second derivatives of g is controlled when f is increasing and concave.
- This is useful for robustness of the nonlinear-optimization algorithms.

Table of Contents

Increasing and concave

Necessary and sufficient condition

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \{w \in \mathbb{R} \ : \ w > 0\} \cup \{+\infty\}$. Suppose that $\delta > 0$ is in the domain of f. Assume that f satisfies the minimal δ-smoothing requirements. Suppose further that

- f is increasing and differentiable on $[\delta, U)$;
- f' is nonincreasing (resp., decreasing) on $[\delta, U)$.

Then g is increasing and concave (strictly concave) on $[0, U)$ if and only if

$$f''(\delta) \geq \frac{3}{\delta}\left(f'(\delta) - \frac{f(\delta)}{\delta}\right) \quad (\Leftrightarrow g_2 \leq 0\ ). \qquad (T^*_\delta)$$

Controlled derivative at 0

Properties

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Suppose that $\delta > 0$ is in the domain of f. Assume that f satisfies the minimal δ-smoothing requirements. Suppose further that

- f is continuous on $[0, \delta]$ and thrice differentiable on $(0, \delta]$,
- f''' is decreasing on $(0, \delta]$.

Then f has the following properties:

1. $\lim_{w \to 0^+} f'(w) > g_1 = g'(0)$;
2. $\lim_{w \to 0^+} f''(w) < g_2 = g''(0)$;
3. $\lim_{w \to 0^+} f'''(w) > g_3 = g'''(0)$;
4. $f'''(\delta) < g_3$.

Controlled derivative at 0

1. $\lim\limits_{w \to 0^+} f'(w) > g_1 = g'(0)$;
2. $\lim\limits_{w \to 0^+} f''(w) < g_2 = g''(0)$;

When f is increasing and concave and $g_2 \leq 0$, g is increasing and concave. In this case, property (1) implies that g' is more controlled near 0 than f', and property (2) implies that g'' is more controlled near 0 than f''.

Controlled derivative at 0

1. $\lim_{w\to 0^+} f'(w) > g_1 = g'(0)$;
2. $\lim_{w\to 0^+} f''(w) < g_2 = g''(0)$;

When f is increasing and concave and $g_2 \leq 0$, g is increasing and concave. In this case, property (1) implies that g' is more controlled near 0 than f', and property (2) implies that g'' is more controlled near 0 than f''.

Useful later:

3. $\lim_{w\to 0^+} f'''(w) > g_3 = g'''(0)$;
4. $f'''(\delta) < g_3$.

Monotonicity of $g_1 = g'(0)$ in δ

Decreasing in δ

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Assume that f satisfies the minimal δ-smoothing requirements for all $\delta > 0$ in the domain of f. Suppose further that

- f is continuous on $[0, U)$ and thrice differentiable on $(0, U)$;
- f''' is decreasing on $(0, U)$.

Then $g_1(\delta)$ is decreasing on $(0, U)$.

Monotonicity of $g_1 = g'(0)$ in δ

Decreasing in δ

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Assume that f satisfies the minimal δ-smoothing requirements for all $\delta > 0$ in the domain of f. Suppose further that

- f is continuous on $[0, U)$ and thrice differentiable on $(0, U)$;
- f''' is decreasing on $(0, U)$.

Then $g_1(\delta)$ is decreasing on $(0, U)$.

This is a very useful property, because we can then easily find a value for δ to achieve a target value for $g_1 = g'(0)$ using a simple univariate search.

Monotonicity of $g_1 = g'(0)$ in δ

Proof.

It is easy to check that

$$\frac{dg_1(\delta)}{d\delta} = \frac{\delta}{2}(f'''(\delta) - g_3(\delta)).$$

We want $f'''(\delta) - g_3(\delta) < 0$ on $(0, U)$, so we can conclude that $g_1(\delta)$ is decreasing on $(0, U)$. For a fixed $\delta \in (0, U)$, by the property (4), we have $f'''(\delta) - g_3(\delta) < 0$, which gives us $g_1'(\delta) < 0$ on $(0, U)$. □

Examples

Example

Let $f(w) := w^p$, for some $0 < p < 1$. Then $g_1(\delta)$ is decreasing for $\delta \in (0, +\infty)$.

Proof.

We have:

$$\begin{aligned} f'(w) &= pw^{p-1}; \\ f''(w) &= p(p-1)w^{p-2}; \\ f'''(w) &= p(p-1)(p-2)w^{p-3}; \\ f^{(4)}(w) &= p(p-1)(p-2)(p-3)w^{p-4}. \end{aligned}$$

Because $0 < p < 1$, $f^{(4)}(w) < 0$ on $(0, +\infty)$, which implies f''' is decreasing on $(0, +\infty)$, thus the monotonicity theorem applies. $\square$

Examples

Example

Let $f(w) := \mathrm{ArcSinh}(\sqrt{w}) = \log\left(\sqrt{w} + \sqrt{1+w}\right)$, for $w \geq 0$.

Proof.

Calculate the following derivatives of f on $(0, +\infty)$:

$$f'(w) = \frac{1}{2\sqrt{w(w+1)}}; \qquad f''(w) = -\frac{2w+1}{4\left(w(w+1)\right)^{\frac{3}{2}}};$$

$$f'''(w) = \frac{8w^2+8w+3}{8\left(w(w+1)\right)^{\frac{5}{2}}}; \qquad f^{(4)}(w) = -\frac{48w^3+72w^2+54w+15}{16\left(w(w+1)\right)^{\frac{7}{2}}}.$$

For $w \in (0, +\infty)$, clearly $f^{(4)}(w) < 0$, which implies f''' is decreasing on $(0, +\infty)$, thus the monotonicity theorem applies. □

Table of Contents

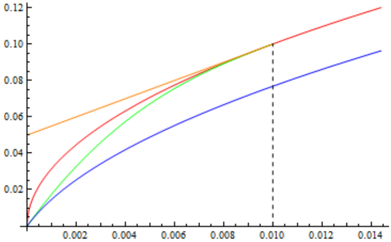
Lower Bound

Lower bound theorem

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Suppose that $\delta > 0$ is in the domain of f. Assume that f satisfies the minimal δ-smoothing requirements. Assume further that

- f is continuous on $[0, \delta]$;
- f''' exists and is decreasing on $(0, \delta]$.

Then $g(w) < f(w)$ for all $w \in (0, \delta)$.



Lower Bound Examples

Example

Let $f(w) := w^p$, for some $0 < p < 1$. For all $\delta > 0$, the δ-smoothing g lower bounds f on $[0, +\infty)$. This generalizes the result in Lee and Skipper [2017] that states that g is a lower bound for root functions of the form $f(w) = w^{1/q}$, for integer $q \geq 2$.

Example

Consider $f(w) := \mathrm{ArcSinh}(\sqrt{w}) = \log\left(\sqrt{w} + \sqrt{1+w}\right)$, for $w \geq 0$. For all $\delta > 0$, the δ-smoothing g lower bounds f on $[0, +\infty)$.

Proof for the Lower Bound

- Use the technique of error estimation for "osculating interpolation" to prove that

$$K(w) = \frac{f(w) - g(w)}{w(w-\delta)^3} < 0, \text{ for } w \in (0, \delta).$$

Proof for the Lower Bound

- Use the technique of error estimation for "osculating interpolation" to prove that

$$K(w) = \frac{f(w) - g(w)}{w(w-\delta)^3} < 0, \text{ for } w \in (0, \delta).$$

- For some fixed $w \in (0, \delta)$, denote $K := K(w)$ for simplicity, and introduce a new function F :

$$F(x) := f(x) - g(x) - Kx(x-\delta)^3.$$

Proof for the Lower Bound

- Use the technique of error estimation for "osculating interpolation" to prove that

$$K(w) = \frac{f(w) - g(w)}{w(w-\delta)^3} < 0, \text{ for } w \in (0, \delta).$$

- For some fixed $w \in (0, \delta)$, denote $K := K(w)$ for simplicity, and introduce a new function F :

$$F(x) := f(x) - g(x) - Kx(x-\delta)^3.$$

- By the definition of K, we have $F(w) = 0$. Also from the relationships between f and g, we have $F(0) = F(\delta) = F'(\delta) = F''(\delta) = 0$. It is easy to see that $0, w, \delta$ are three zeros for $F(x)$.

Proof for the Lower Bound

- Use the technique of error estimation for "osculating interpolation" to prove that

$$K(w) = \frac{f(w) - g(w)}{w(w-\delta)^3} < 0, \text{ for } w \in (0, \delta).$$

- For some fixed $w \in (0, \delta)$, denote $K := K(w)$ for simplicity, and introduce a new function F :

$$F(x) := f(x) - g(x) - Kx(x-\delta)^3.$$

- By the definition of K, we have $F(w) = 0$. Also from the relationships between f and g, we have $F(0) = F(\delta) = F'(\delta) = F''(\delta) = 0$. It is easy to see that $0, w, \delta$ are three zeros for $F(x)$.
- Because $F(x)$ is continuous on $[0, \delta]$ and differentiable on $(0, \delta)$, applying Rolle's Theorem (twice), there exists $0 < w_1 < w < \eta_1 < \delta$ such that $F'(w_1) = F'(\eta_1) = 0$.

Proof for the Lower Bound

$$F(x) := f(x) - g(x) - Kx(x-\delta)^3.$$

- Noting that $F'(\delta) = 0$ and that $F'(x)$ is differentiable on $[w_1, \delta]$, we apply Rolle's Theorem (twice more, now to F') to get $w_1 < w_2 < \eta_1 < \eta_2 < \delta$ such that $F''(w_2) = F''(\eta_2) = 0$.

Proof for the Lower Bound

$$F(x) := f(x) - g(x) - Kx(x-\delta)^3.$$

- Noting that $F'(\delta) = 0$ and that $F'(x)$ is differentiable on $[w_1, \delta]$, we apply Rolle's Theorem (twice more, now to F') to get $w_1 < w_2 < \eta_1 < \eta_2 < \delta$ such that $F''(w_2) = F''(\eta_2) = 0$.
- Using Rolle's Theorem (twice again, now on F''; we have $F''(\delta) = 0$ and $F''(x)$ is differentiable on $[w_2, \delta]$), we get $w_2 < w_3 < \eta_2 < \eta_3 < \delta$ such that $F'''(w_3) = F'''(\eta_3) = 0$.

Proof for the Lower Bound

$$F(x) := f(x) - g(x) - Kx(x - \delta)^3.$$

- Noting that $F'(\delta) = 0$ and that $F'(x)$ is differentiable on $[w_1, \delta]$, we apply Rolle's Theorem (twice more, now to F') to get $w_1 < w_2 < \eta_1 < \eta_2 < \delta$ such that $F''(w_2) = F''(\eta_2) = 0$.
- Using Rolle's Theorem (twice again, now on F''; we have $F''(\delta) = 0$ and $F''(x)$ is differentiable on $[w_2, \delta]$), we get $w_2 < w_3 < \eta_2 < \eta_3 < \delta$ such that $F'''(w_3) = F'''(\eta_3) = 0$.
- Now, $F'''(x) = f'''(x) - g_3 - K(24x - 18\delta)$. Applying $F'''(w_3) = F'''(\eta_3)$ and $f'''(w_3) > f'''(\eta_3)$, we can conclude that $K(24w_3 - 18\delta) > K(24\eta_3 - 18\delta)$. But this last inequality holds only when $K < 0$. $\square$

Increasing and concave

Our lower bound theorem suggests that there could be f that are not increasing and concave for which the δ-smoothing of f is a lower bound for f.

So we have the natural question: do we automatically satisfy (T_δ^*) when the lower bound theorem applies to functions that are increasing and concave?

Increasing and concave

Our lower bound theorem suggests that there could be f that are not increasing and concave for which the δ-smoothing of f is a lower bound for f.
So we have the natural question: do we automatically satisfy (T_δ^*) when the lower bound theorem applies to functions that are increasing and concave?

Observation

For an increasing concave function f, the hypotheses of our lower bound theorem do not imply that the smoothing g is increasing and concave, i.e., (T_δ^*) is not implied by the hypotheses of our lower bound theorem, even for increasing concave f.

Example

- Consider the function

$$f(w) := \begin{cases} a_5 w^5 + a_4 w^4 + a_3 w^3 + a_2 w^2 + a_1 w, & 0 \le w \le w_0; \\ a\sqrt{w-c} + b, & w > w_0. \end{cases}$$

- After fixing the values of the parameters δ, w_0, a_2, a_3, a_4, and a_5 so that $\frac{f''(w_0)}{f'''(w_0)} \le 0$, we ensure continuity and thrice differentiability of f at w_0 by calculating the remaining parameters as follows:

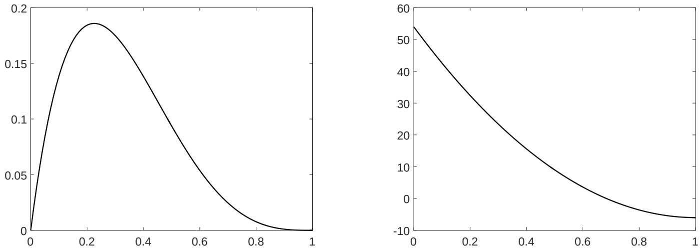
$$\begin{aligned} c &= w_0 + \frac{3f''(w_0)}{2f'''(w_0)}; \\ a_1 &= -2f''(w_0)(w_0 - c) - (5a_5 w_0^4 + 4a_4 w_0^3 + 3a_3 w_0^2 + 2a_2 w_0); \\ a &= \frac{8f'''(w_0)(w_0 - c)^{\frac{5}{2}}}{3}; \\ b &= f(w_0) - a\sqrt{w_0 - c}. \end{aligned}$$

Example

For $\delta \leq w_0$, we have the δ-smoothing $g(w) = g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3$, where

$$
\begin{aligned}
g_1 &= 3a_5\delta^4 + a_4\delta^3 + a_1; \\
g_2 &= -16a_5\delta^3 - 6a_4\delta^2 + 2a_2; \\
g_3 &= 36a_5\delta^2 + 18a_4\delta + 6a_3.
\end{aligned}
$$

Example



(a) $f(w) - g(w)$ (b) $f'''(w) - g_3$

Figure: $a_5 = 1,\ a_4 = -5,\ a_3 = 0,\ w_0 = 3,\ a_2 = -3$

Example

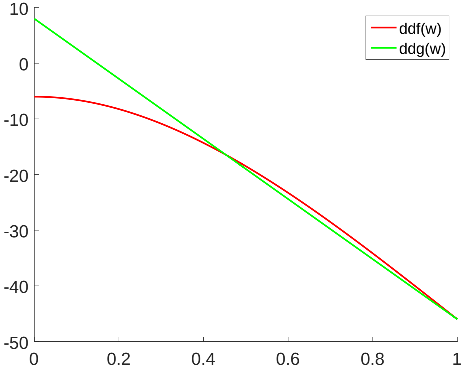


Figure: $a_5 = 1,\ a_4 = -5,\ a_3 = 0,\ w_0 = 3,\ a_2 = -3$

Observations on the example

- f is increasing and concave
- $g_2 \not\leq 0$, so the necessary and sufficient condition for g to be increasing and concave fails (g is not concave near 0)
- But the condition for g to lower bound f holds
- Note that f''' is decreasing on $(0, \delta]$, which, as we have seen, implies that $\lim_{w \to 0^+} f''(w) < g_2 = g''(0)$ (g has more controlled second derivative than f near 0). In the graph we see that we need the limit.

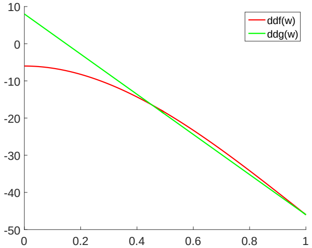


Table of Contents

Better Bound

- g is defined based on δ, and h is defined based on λ, a fair comparison is achieved by making $h'(0) = g'(0)$. (the same maximum derivative at 0)
- Let $h'(0) = f'(\lambda) = g'(0) = g_1$ $= 3f(\delta)/\delta - 2f'(\delta) + \delta f''(\delta)/2$.

  Then we have

$$\hat{\lambda} := (f')^{-1} \left(3f(\delta)/\delta - 2f'(\delta) + \delta f''(\delta)/2\right).$$

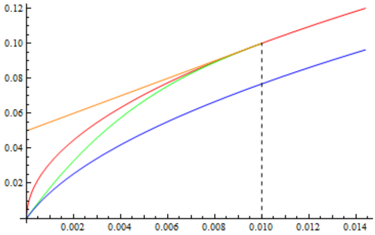


Figure: Blue: Shift smoothing $h(w) = f(w + \lambda) - f(\lambda)$, Green: our smoothing $g(w)$

Better Bound

Better Bound Theorem

Let f be a univariate function having a domain $I = [0, U)$, where $U \in \mathbb{R}^+ \cup \{+\infty\}$. Suppose that $U \geq \delta/2 > 0$. Assume that f satisfies the minimal δ-smoothing requirements. Assume further that

- f is continuous, increasing, and strictly concave on its domain;
- f is thrice differentiable on $(0, U)$.

Moreover, suppose that

(I) f''' is decreasing on $(0, 2\delta)$;

(II) $f'''(w) \geq 0$, for $w \in (0, 2\delta)$.

Then

$$h(w) := f(w + \hat{\lambda}) - f(\hat{\lambda}) \leq g(w),$$

for w in the domain of f.

Examples

Example

Let $f(w) := w^p$, for some $0 < p < 1$. For all $\delta > 0$, the δ-smoothing g upper bounds the shift smoothing h on $[0, +\infty)$. This generalizes the result in Lee and Skipper [2017], which states that our smoothing g 'fairly dominates' the shift smoothing h for root functions of the form $f(w) = w^{1/q}$, with integer $2 \le q \le 10,000$.

Example

Consider $f(w) := \mathrm{ArcSinh}(\sqrt{w}) = \log\left(\sqrt{w} + \sqrt{1+w}\right)$, for $w \geq 0$. For all $\delta > 0$, the δ-smoothing g upper bounds the shift smoothing h on $[0, +\infty)$.

Thank You For Your Attention

Luze Xu, Jon Lee, and Daphne Skipper. More virtuous smoothing. arXiv preprint arXiv:1802.09112, 2018.

Andreas Wächter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming, 106(1):25–57, 2006.

Claudia D'Ambrosio, Marcia Fampa, Jon Lee, and Stefan Vigerske. On a nonconvex MINLP formulation of the Euclidean Steiner tree problems in n-space. Technical report, Optimization Online, 2014. `http://www.optimization-online.org/DB_HTML/2014/09/4528.html` , Contains missing proofs of results in D'Ambrosio et al. [2015].

Claudia D'Ambrosio, Marcia Fampa, Jon Lee, and Stefan Vigerske. On a nonconvex MINLP formulation of the Euclidean Steiner tree problem in n-space. In E. Bampis, editor, Experimental Algorithms, volume 9125 of Lecture Notes in Computer Science, pages 122–133. Springer International Publishing, 2015. doi: 10.1007/978-3-319-20086-6_10.

Jon Lee and Daphne Skipper. Virtuous smoothing for global optimization. Journal of Global Optimization, 69(3):677–697, 2017.

Proof for the Better Bound

- With condition *(I)*, f''' is decreasing, so we have all the properties of controlled derivatives. First, we consider the existence and uniqueness of $\hat{\lambda}$. Condition *(II)* and property 4 imply that $g_3 > f'''(\delta) \geq 0$, and so $g_1 - f'(\delta) = \frac{1}{2}g_3\delta^2 - \delta f''(\delta) > 0$. Therefore,

$$\lim_{w \to 0^+} f'(w) > g_1 > f'(\delta),$$

  and because $f'(w)$ is decreasing, there exists exactly one $\hat{\lambda}$ in $(0, \delta)$ such that $f'(\hat{\lambda}) = g_1$.

Proof for the Better Bound

- With condition *(I)*, f''' is decreasing, so we have all the properties of controlled derivatives. First, we consider the existence and uniqueness of $\hat{\lambda}$. Condition *(II)* and property 4 imply that $g_3 > f'''(\delta) \geq 0$, and so $g_1 - f'(\delta) = \frac{1}{2} g_3 \delta^2 - \delta f''(\delta) > 0$. Therefore,

$$\lim_{w \to 0^+} f'(w) > g_1 > f'(\delta),$$

  and because $f'(w)$ is decreasing, there exists exactly one $\hat{\lambda}$ in $(0, \delta)$ such that $f'(\hat{\lambda}) = g_1$.
- Now consider the function $H := g - h$, which has

$$\begin{aligned}
H(w) &= g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3 - f(w + \hat{\lambda}) + f(\hat{\lambda}); \\
H'(w) &= g_1 + g_2 w + \frac{1}{2} g_3 w^2 - f'(w + \hat{\lambda}); \\
H''(w) &= g_2 + g_3 w - f''(w + \hat{\lambda}); \\
H'''(w) &= g_3 - f'''(w + \hat{\lambda}),
\end{aligned}$$

  where the coefficients of the associated function g are as usual.

Proof for the Better Bound

$$
\begin{aligned}
H(w) &= g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3 - f(w + \hat{\lambda}) + f(\hat{\lambda}); \\
H'(w) &= g_1 + g_2 w + \frac{1}{2} g_3 w^2 - f'(w + \hat{\lambda}); \\
H''(w) &= g_2 + g_3 w - f''(w + \hat{\lambda}); \\
H'''(w) &= g_3 - f'''(w + \hat{\lambda}),
\end{aligned}
$$

- It is now straightforward to verify that $H(0) = H'(0) = 0$, $H(\delta) = f(\delta) - h(\delta) \geq 0$, and $H'(\delta) = f'(\delta) - f'(\delta + \hat{\lambda}) > 0$.

Proof for the Better Bound

$$H(w) = g_1 w + \frac{1}{2} g_2 w^2 + \frac{1}{6} g_3 w^3 - f(w + \hat{\lambda}) + f(\hat{\lambda});$$
$$H'(w) = g_1 + g_2 w + \frac{1}{2} g_3 w^2 - f'(w + \hat{\lambda});$$
$$H''(w) = g_2 + g_3 w - f''(w + \hat{\lambda});$$
$$H'''(w) = g_3 - f'''(w + \hat{\lambda}),$$

- It is now straightforward to verify that $H(0) = H'(0) = 0$, $H(\delta) = f(\delta) - h(\delta) \geq 0$, and $H'(\delta) = f'(\delta) - f'(\delta + \hat{\lambda}) > 0$.
- Noting that $0 < \hat{\lambda} < \delta$, we also have

$$H''(\delta) = f''(\delta) - f''(\delta + \hat{\lambda}) < 0, \text{ (by condition } \mathit{(I)}\text{)},$$
$$H''' \text{ is increasing on } (0, \delta], \text{ (by condition } \mathit{(II)}\text{)},$$

and by condition *(I)* and property 4 together,

$$H'''(\delta) = g_3 - f'''(\delta + \hat{\lambda}) > f'''(\delta) - f'''(\delta + \hat{\lambda}) > 0.$$

Proof for the Better Bound

- Finally, we assert that $H'''(0) < 0$ and $H''(0) > 0$, which we prove below.
- Because H''' is increasing on $[0, \delta]$ with $H'''(0) < 0$ and $H'''(\delta) > 0$, we see that $H''(w)$ is first decreasing and then increasing on $[0, \delta]$. Because $H''(0) > 0$ and $H''(\delta) < 0$, there exists exactly one zero of H'' on $(0, \delta)$, which we label v_1. Thus $H'(w)$ is increasing on $[0, v_1]$ and decreasing on $[v_1, \delta]$. Because $H'(0) = 0$ and $H'(\delta) > 0$, we see that $H(w)$ is increasing on $[0, \delta]$, and so for $w \in [0, \delta]$, $H(w) \geq H(0) = 0$; i.e., $h(w) \leq g(w)$, for $w \geq 0$.

Proof for the Better Bound

- Now we turn our attention to proving that $H'''(0) < 0$ and $H''(0) > 0$. As the conditions of this theorem are a restriction of those of lower bound theorem, we can find the roots of the derivatives of the function $F := f - g$, $0 < w_2 < w_1 < w_0 < \delta$, where w_0 is the root of F''', w_1 is the root of F'', and w_2 is the root of F' as in the previous remark.

Proof for the Better Bound

- Now we turn our attention to proving that $H'''(0) < 0$ and $H''(0) > 0$. As the conditions of this theorem are a restriction of those of lower bound theorem, we can find the roots of the derivatives of the function $F := f - g$, $0 < w_2 < w_1 < w_0 < \delta$, where w_0 is the root of F''', w_1 is the root of F'', and w_2 is the root of F' as in the previous remark.
- From the remark, F''' is decreasing on $(0, \delta)$. Therefore, to prove that $H'''(0) = g_3 - f'''(\hat{\lambda}) = g'''(\hat{\lambda}) - f'''(\hat{\lambda}) < 0$, it suffices to show that $\hat{\lambda} < w_0$.

Proof for the Better Bound

- Now we turn our attention to proving that $H'''(0) < 0$ and $H''(0) > 0$. As the conditions of this theorem are a restriction of those of lower bound theorem, we can find the roots of the derivatives of the function $F := f - g$, $0 < w_2 < w_1 < w_0 < \delta$, where w_0 is the root of F''', w_1 is the root of F'', and w_2 is the root of F' as in the previous remark.
- From the remark, F''' is decreasing on $(0, \delta)$. Therefore, to prove that $H'''(0) = g_3 - f'''(\hat{\lambda}) = g'''(\hat{\lambda}) - f'''(\hat{\lambda}) < 0$, it suffices to show that $\hat{\lambda} < w_0$.
- Function f satisfies condition (T_δ^*), so g is concave on $(0, \delta]$, and $f'(\hat{\lambda}) - g'(\hat{\lambda}) = g'(0) - g'(\hat{\lambda}) > 0$. Because F' is positive only to the left of w_2, we have $\hat{\lambda} < w_2$ $(< w_0)$.

Proof for the Better Bound

- To prove that $H''(0) = g_2 - f''(\hat{\lambda}) > 0$, we demonstrate that $g_2 > f''(\hat{\lambda})$, which we accomplish via an inequality that arises as lower and upper bounds on $g'(w_2) - g'(0)$.
  For the lower bound, because $F'''(w) = f'''(w) - g_3 > 0$ on $[0, w_2] \subset [0, w_0)$, we have

  $$f''(w) > f''(\hat{\lambda}) + g_3(w - \hat{\lambda}), \text{ for } w \in [\hat{\lambda}, w_2].$$

  Therefore, the slope of the secant to f'' between the points at $w = \hat{\lambda}$ and $w = w_2$ is at least g_3; i.e.,

  $$g'(w_2) - g'(0) = f'(w_2) - f'(\hat{\lambda}) > \frac{1}{2} g_3 (w_2 - \hat{\lambda})^2 + f''(\hat{\lambda})(w_2 - \hat{\lambda}).$$

Proof for the Better Bound

- To prove that $H''(0) = g_2 - f''(\hat{\lambda}) > 0$, we demonstrate that $g_2 > f''(\hat{\lambda})$, which we accomplish via an inequality that arises as lower and upper bounds on $g'(w_2) - g'(0)$.
  For the lower bound, because $F'''(w) = f'''(w) - g_3 > 0$ on $[0, w_2] \subset [0, w_0)$, we have
  $$f''(w) > f''(\hat{\lambda}) + g_3(w - \hat{\lambda}), \text{ for } w \in [\hat{\lambda}, w_2].$$
  Therefore, the slope of the secant to f'' between the points at $w = \hat{\lambda}$ and $w = w_2$ is at least g_3; i.e.,
  $$g'(w_2) - g'(0) = f'(w_2) - f'(\hat{\lambda}) > \frac{1}{2} g_3 (w_2 - \hat{\lambda})^2 + f''(\hat{\lambda})(w_2 - \hat{\lambda}).$$
- For the upper bound on $g'(w_2) - g'(0)$, we require two observations. First, by condition *(I)* and property 4, we have $g_3 > f'''(\delta) \geq 0$. Second, applying $g_2 + g_3\delta = f''(\delta) \leq 0$, we have $w_2 < \delta \leq -g_2/g_3$.

Proof for the Better Bound

- Now we can obtain the upper bound

$$g'(w_2) - g'(0) = \frac{1}{2}g_3 w_2^2 + g_2 w_2$$
$$\leq \frac{1}{2}g_3(w_2 - \hat{\lambda})^2 + g_2(w_2 - \hat{\lambda}),$$

because this inequality is equivalent to

$$0 \leq -g_3 w_2 - g_2 + g_3\hat{\lambda}/2,$$

which we verify by applying $g_3 > 0$ and $w_2 \leq -g_2/g_3$.
Combining these bounds, we have

$$\frac{1}{2}g_3(w_2-\hat{\lambda})^2+f''(\hat{\lambda})(w_2-\hat{\lambda}) \ < \ g'(w_2)-g'(0) \ \leq \ \frac{1}{2}g_3(w_2-\hat{\lambda})^2+g_2(w_2-\hat{\lambda}),$$

which reduces to the desired $g_2 > f''(\hat{\lambda})$.

□