SIAM J. OPTIM.
Vol. 29, No. 2, pp. 1240–1259

MORE VIRTUOUS SMOOTHING*

LUZE XU†, JON LEE†, AND DAPHNE SKIPPER‡

Abstract. In the context of global optimization of mixed-integer nonlinear optimization formulations, we consider smoothing univariate functions f that satisfy $f(0) = 0$, f is increasing and concave on $[0, +\infty)$, f is twice differentiable on all of $(0, +\infty)$, but $f'(0)$ is undefined or intolerably large. The canonical examples are root functions $f(w) := w^p$ for $0 < p < 1$. We consider the earlier approach of defining a smoothing function g that is identical with f on $(\delta, +\infty)$, for some chosen $\delta > 0$, then replacing the part of f on $[0, \delta]$ with the unique homogeneous cubic, matching f, f', and f'' at δ. The parameter δ is used to control (i.e., upper bound) the derivative at 0 (which controls it on all of $[0, +\infty)$ when g is concave). Our main results are as follows: (i) we weaken an earlier sufficient condition to give a necessary and sufficient condition for the piecewise function g to be increasing and concave; (ii) we give a general sufficient condition for $g'(0)$ to be decreasing in the smoothing parameter δ, and under the same condition we demonstrate that the worst-case error of g as an estimate of f is increasing in δ; (iii) we give a general sufficient condition for g to underestimate f; (iv) we give a general sufficient condition for g to dominate the simple "shift smoothing" $h(w) := f(w + \lambda) - f(\lambda)$ $(\lambda > 0)$ when the parameters δ and λ are chosen "fairly," i.e., so that $g'(0) = h'(0)$. In doing so, we solve two natural open problems of Lee and Skipper [*J. Global Optim.*, 69 (2017), pp. 677–697] concerning (iii) and (iv) for root functions.

Key words. global optimization, mixed-integer nonlinear optimization, spatial branch-and-bound, concave, nondifferentiable, smoothing, piecewise

AMS subject classifications. 90C26, 90C30, 65K05

DOI. 10.1137/18M1172831

1. Introduction.

1.1. Motivation. Most mixed-integer nonlinear optimization (MINLO) software, aiming at global optimization of so-called factorable mathematical-optimization formulations, apply the spatial branch-and-bound algorithm or some close relative of it (e.g., `BARON` [TS02], `ANTIGONE` [MF14], open-source `Couenne` [BLL+09], and free-for-academic-use `SCIP` [Ach09]). As a first step, problem functions are "factored" (i.e., fully decomposed) via a small library of low-dimensional nonlinear functions (typically, functions in one, two, or three variables) together with affine functions of an arbitrary number of variables. It is helpful, for reliability, 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., [WB06]). For functions that are not already sufficiently smooth, it is standard practice for modelers to replace "bad" functions by smoother approximating functions (e.g., [BDL+06], [BDL+12], and [GMS13]). But the issue can also be grappled

*Received by the editors February 26, 2018; accepted for publication (in revised form) February 6, 2019; published electronically April 30, 2019.
http://www.siam.org/journals/siopt/29-2/M117283.html
Funding: The second author was supported in part by ONR grant N00014-17-1-2296. Additionally, part of this work was done while the second author was visiting the Simons Institute for the Theory of Computing, which was partially supported by the DIMACS/Simons Collaboration on Bridging Continuous and Discrete Optimization through NSF grant CCF-1740425.
†Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117 (xuluze@umich.edu, jonxlee@umich.edu).
‡Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402-1300 (skipper@usna.edu).

with algorithmically by (purely continuous) nonlinear-optimization solvers through parameter setting. For example, Wächter explains (see [Wäc09]):

> *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 [sic] can be disabled by setting `bound_relax_factor` to 0.

Consider $f(w) := \sqrt{w}$ on the domain $[0, +\infty)$. Notice how in this case, using IPOPT's default value for the parameter `bound_relax_factor`, the function cannot even be evaluated on the entire modified domain $[-10^{-8}, +\infty)$. And for the suggested nondefault parameter setting (0), $\sqrt{w}$ is not differentiable at 0, which is in the domain. Techniques like smoothly extending f so that $f(w) := -\sqrt{-w}$ for $w < 0$ suffer from still not being differentiable at 0. So, we are led back to modeling advice (see [Wäc09]):

> 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 (e.g., see [DFLV15] and [DFLV18], where binary variables and square roots of continuous variables both occur in some inequalities; it turns out that underestimation for the square-root smoothing is critical for validity of the resulting inequalities), and for this reason, our study is particularly relevant to MINLO.

Notably, the MINLO software `SCIP` has incorporated features (see [GGH+16]) to accommodate a more sophisticated approach (see [DFLV15], [DFLV18], and [LS17]), tackling issues of nonsmoothness while at the same time *working within a paradigm that aims at seeking global optimality for nonconvex problems* (see [LS17, section 1] for details). Extending the approach of *virtuous smoothing* from [LS17, section 1] is the subject of what follows.

The practical convergence behavior of the different nonlinear-optimization algorithms that are relied on by MINLO solvers is the subject of intense investigation; see [MKV17] for a recent experimental comparison. As we have indicated, the issue of nonsmoothness pertains to an aspect of the theoretical and practical behavior of various nonlinear-optimization solvers employed by MINLO solvers. Our work is aimed at developing a mathematical framework for improving the behavior. However, because our interest is especially in global optimization, we seek some control on how solutions employing smooth approximators relate to solutions employing the functions that they replace—for example, lower or upper bounding. In cases where our approximators lower bound the functions that they replace, we can compare a pair of lower bounding functions when one dominates the other on its domain.

1.2. Prior work. 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 $w = 0$. The inception of this approach is from [DFLV15, DFLV18], which grappled with handling square-root functions ($p = 1/2$) arising in formulations of the Euclidean Steiner tree 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. By construction, the new piecewise function g is twice differentiable on $(0, +\infty)$. The parameter δ is used to control (i.e., upper bound) the derivative at 0. In [DFLV15, DFLV18] D'Ambrosio et al. 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 smoothings have the same numerical stability.

In [LS17], we extended this idea of [DFLV15, DFLV18], 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.

Regarding (i), the property is useful because we want g to behave like the function f that it replaces. Furthermore, the concavity of g means that controlling its derivative at 0 implies that it is controlled on all of $[0, +\infty)$. We are now able to extend (i) to get a necessary and sufficient condition. Regarding (ii) and (iii), the results requiring that p have the form $1/q$ for an integer $q \geq 2$ (and for (iii) even $q \leq 10{,}000$, which required some computer algebra for each q) were limited by the algebraic proof techniques that we employed—making a transformation to then be able to apply methods that work for analyzing polynomials (e.g., Descartes's rule of signs). We left in [LS17] as substantial open problems extending (ii) and (iii) to *all* root functions. In what follows, we resolve these open problems and generalize the theorems quite a bit further, by employing methods of analysis instead of algebraic methods.

1.3. Definition of δ-smoothing.

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.

DEFINITION 1. *We say that such an f satisfies the* minimal δ-smoothing requirements *if $f(0) = 0$ and f is twice differentiable at δ.*

In the spirit of [LS17] (though we note that they always assumed $U = +\infty$), we will define a "δ-smoothing" of f.

DEFINITION 2. *Suppose that such an f satisfies the minimal δ-smoothing requirements. Then the* δ-smoothing *of f is the piecewise-defined function*

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

with

$$\begin{aligned} g_1 &:= \frac{3f(\delta)}{\delta} - 2f'(\delta) + \frac{\delta f''(\delta)}{2}, \\ g_2 &:= -\frac{6f(\delta)}{\delta^2} + \frac{6f'(\delta)}{\delta} - 2f''(\delta), \\ g_3 &:= \frac{6f(\delta)}{\delta^3} - \frac{6f'(\delta)}{\delta^2} + \frac{3f''(\delta)}{\delta}. \end{aligned}$$

Obviously the function g and its coefficients $g_1, \ g_2, \ g_3$ depend on δ, but to avoid clutter we do not indicate this in the notation.

Although the coefficients g_i (in the cubic portion of g) have a rather complicated specification, it is easy to check that the cubic portion of g is the unique minimum-degree polynomial having

$$\begin{aligned} g(0) &= f(0) = 0, \\ g(\delta) &= f(\delta), \\ g'(\delta) &= f'(\delta), \\ g''(\delta) &= f''(\delta). \end{aligned}$$

We chose the precise form of the homogeneous cubic, for later convenience, so that the coefficients g_i satisfy

$$\begin{aligned} &g_1 = g'(0), \\ &g_2 = g''(0), \\ &g_3 = g'''(w) \text{ for } w \in [0, \delta]. \end{aligned}$$

As in [LS17], our main motivation is situations in which, like root functions, $f'(0)$ is undefined or intolerably large. As we will see, the parameter δ is used to control the derivative of g at 0. Also motivated by root functions, we are particularly interested in functions f that are continuous, increasing, and concave on their domains. MINLO solvers like `BARON`, `SCIP`, and `ANTIGONE` are improving their performance by growing their set of low-dimensional library functions as a means of getting stronger relaxations. This can lead to stronger relaxations than simply combining relaxations across function compositions. So we seek general methods for smoothing that can be readily applied. Although our first challenging motivation is root functions, there are other natural functions that occur for which our methods apply. For example, the concave *entropy function*

$$f(w) := \begin{cases} -w\log(w), & 0 < w \le 1, \\ 0, & w = 0 \end{cases}$$

is continuous on $[0, 1]$, but its derivative blows up at 0. Our methods apply here, and we have plans to implement our smoothing for it in `ANTIGONE`. Another example is the concave and increasing *incremental entropy function*

$$f(w) := \begin{cases} w\log(1 + \frac{1}{w}), & w > 0, \\ 0, & w = 0. \end{cases}$$

To eliminate any chance of confusion, throughout, for a real interval I and a function $\phi : I \to \mathbb{R}$, f is *increasing* if $\phi(w_1) < \phi(w_2)$ for all $w_1 < w_2 \in I$, and *nondecreasing* if $\phi(w_1) \le \phi(w_2)$ for all $w_1 < w_2 \in I$[1] (and similarly for *decreasing* and *nonincreasing*). The function ϕ is *strictly concave* if $\phi(\lambda w_1 + (1-\lambda)w_2) > \lambda\phi(w_1) + (1-\lambda)\phi(w_2)$ for all $w_1 < w_2 \in I$ and all $0 < \lambda < 1$, and *concave* if $\phi(\lambda w_1 + (1-\lambda)w_2) \ge \lambda\phi(w_1) + (1-\lambda)\phi(w_2)$ for all $w_1 < w_2 \in I$ and all $0 < \lambda < 1$ (and similarly for *strictly convex* and *convex*).

In section 2, we weaken the sufficient condition [LS17, Theorem 2] to now give a necessary and sufficient condition for g to be increasing and concave. We also provide conditions under which $g'(0)$ has desirable behaviors. In section 3, we give a

[1] We do note that many authors prefer the term ‘strictly increasing’ for what we call ‘increasing’, and ‘increasing’ for what we call ‘nondecreasing’.

sufficient condition for g to underestimate f, greatly generalizing [LS17, Theorem 9]. Additionally, in section 3, we analyze the dependence on our smoothing parameter δ of the worst-case behavior of g as an approximation of f. In section 4, we give a general sufficient condition for g to dominate the simple "shift smoothing" $h(w) := f(w+\lambda) - f(\lambda)$ $(\lambda > 0)$, when the parameters δ (for g) and λ (for h) are chosen "fairly," i.e., so that $g'(0) = h'(0)$, greatly generalizing [LS17, Theorem 10]. Via our main results in sections 3 and 4, we solve two natural open problems of [LS17] concerning root functions, and in fact extend those results significantly beyond root functions. In section 5, we make some brief concluding remarks.

2. General behaviors of δ-smoothing.

In this section, we explore general properties of δ-smoothings that are not directly related to bounding f (which we will take up in section 3). In section 2.1 we provide a necessary and sufficient condition on an increasing and concave f under which its δ-smoothing g is also increasing and concave. In section 2.2, we provide properties relating the behaviors of g' and f' near zero, when f''' is decreasing. In section 2.3, we show that f''' being decreasing is a sufficient condition for $g_1 = g'(0)$ to be decreasing in the smoothing parameter δ—a property which is practically useful in choosing a good value for δ.

2.1. Increasing and concave.

In the context of global optimization, it is desirable for the δ-smoothing g of a function f to share properties with f beyond those inherent in the definition of g. For example, when f is a root function, f is increasing and concave. In this way, g can be algorithmically treated by global-optimization software in a way that is consistent with the treatment of f (e.g., tangents for overestimating and secants for underestimating). Furthermore, concavity of g implies that controlling

$$g'(0) = g_1 = \frac{3f(\delta)}{\delta} - 2f'(\delta) + \frac{\delta f''(\delta)}{2}$$

(by choosing $\delta > 0$ appropriately) has the effect of controlling $g'(w)$ on all of its nonnegative domain.

In [LS17], we gave the following lower bound on the negative curvature of f at δ as a sufficient condition for g to be increasing and concave on $[0, \delta]$.

THEOREM 3 (see [LS17, Theorem 2]). *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 $[\delta, U)$.*

If

$$(T_\delta) \qquad f''(\delta) \geq \frac{2}{\delta}\left(f'(\delta) - \frac{f(\delta)}{\delta}\right) \quad \left(\Leftrightarrow g_3 \geq 0\right),$$

then the δ-smoothing g of f is increasing and concave (resp., strictly concave) on $[0, U)$.

If we make the further mild assumption that f is differentiable on $(0, \delta)$, then by Rolle's theorem, there is a $u \in (0, \delta)$ so that

$$\frac{f(\delta) - f(0)}{\delta - 0} = f'(u).$$

If we make the still further mild assumption that f' is nonincreasing on $(0, \delta]$, then we can conclude that $f'(\delta) - \frac{f(\delta)}{\delta} \leq 0$. Then the intuition for (T_δ) is that if the curvature of f at δ is not too negative, then the function can make it to the origin staying increasing and concave.

This sufficient condition is met by all root functions and more (see [LS17, Examples 6 and 7]). Of course we may be concerned that the sufficient condition (T_δ) is too strong, and the following example demonstrates that (T_δ) is not necessary for g to be increasing and concave.

Example 4. For $\epsilon > 0$, let

$$f(w) := \begin{cases} -\dfrac{1}{24\epsilon\sqrt{\epsilon}(1+\epsilon)} w^3 + \dfrac{1+5\epsilon}{8\epsilon\sqrt{\epsilon}} w, & 0 \leq w \leq 1+\epsilon, \\ \sqrt{w-1} + \dfrac{1+8\epsilon-5\epsilon^2}{12\epsilon\sqrt{\epsilon}}, & w > 1+\epsilon. \end{cases}$$

It is straightforward to verify that $f(0) = 0$ and that f is twice-differentiable, increasing, and concave on $[0, +\infty)$.

Now, let $\delta := 1$. Because f is a cubic function on $[0, \delta]$, g is the same as f on $[0, \delta]$, which means that g is also increasing and concave. However, in this case, $g'''(\delta) = -\frac{1}{4\epsilon\sqrt{\epsilon}(1+\epsilon)} < 0$, contradicting (T_δ).

Finally, one could argue that this example is unfair, because we are not actually smoothing anything at 0. But, following the idea in [LS17, section 2.2], we could add a very small positive multiple of $\sqrt{w}$ to this $f(w)$, and then we would get a legitimate example, nonsmooth at 0. ∎

Next, we will see that by modifying the condition (T_δ) we obtain a necessary and sufficient condition for g to be increasing and concave on $[0, \delta]$. In fact, we will see that the condition is precisely motivated by Example 4.

THEOREM 5. *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 (resp., strictly concave) on $[0, U)$ if and only if

$$(T_\delta^*) \qquad f''(\delta) \geq \frac{3}{\delta}\left(f'(\delta) - \frac{f(\delta)}{\delta}\right) \quad \left(\Leftrightarrow g_2 \leq 0 \right).$$

Proof. Necessity is obvious because $g''(0) = g_2 \leq 0$. For sufficiency, under (T_δ^*), we have $g''(0) \leq 0$. Along with the fact that $g''(\delta) = f''(\delta)$ is nonpositive (negative) and that $g''(w)$ is linear (in w), we have that $g'(w)$ is nonincreasing (decreasing) to $g'(\delta) = f'(\delta) > 0$, and therefore g is concave (strictly concave) and increasing on $[0, \delta]$. Note that the assumptions on f imply that g is concave (strictly concave) and increasing on $[\delta, U)$. The conclusion follows. ∎

As the function in Example 4 has $g_2 = 0$, it satisfies property (T_δ^*) as an equation.

2.2. Controlled derivative at 0. The primary goal of δ-smoothing is to approximate f by a smooth function g having derivative controlled at zero. In Proposition 6, we present several properties relating the behaviors of derivatives of f and g at the ends of the interval $[0, \delta]$ in the event that f''' exists and is decreasing on $(0, \delta]$. As we will see, for increasing and concave f, we get conditions under which both the first and second derivatives of g are more controlled near zero than those of f. Looking ahead, Proposition 6 will also be used in section 2.3 and section 3.4 to prove the monotonicity of g_1 and $\|f - g\|_\infty$ in the smoothing parameter δ, and in section 4 to demonstrate that g is a tighter lower bound for "root-like functions" than the natural "shift smoothing."

PROPOSITION 6. *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 continuous on $[0, \delta]$ and thrice differentiable on $(0, \delta]$,*
- *f''' is decreasing on $(0, \delta]$.*

Then f has the following properties:
(1) $\lim\limits_{w \to 0^+} f'(w) > g_1 = g'(0)$,
(2) $\lim\limits_{w \to 0^+} f''(w) < g_2 = g''(0)$,
(3) $\lim\limits_{w \to 0^+} f'''(w) > g_3 = g'''(0)$,
(4) $f'''(\delta) < g_3$.

Proof. Clearly $f \neq g$ because f''' is decreasing on $[0, \delta]$ while g''' is constant on $[0, \delta]$. Define $F := f - g$ on $[0, \delta]$ and let $J := (0, \delta)$. Then $F(0) = 0$ and $F^{(i)}(\delta) = 0$ for $i = 0, 1, 2$. Because f''' is decreasing on $(0, \delta]$, $F''' = f''' - g_3$ is also decreasing on $(0, \delta]$.

Suppose property (4) does not hold, i.e., $F'''(\delta) \geq 0$. Then on J, $F''' > 0$ or, equivalently, F'' is increasing. Because $F''(\delta) = 0$, $F'' < F''(\delta) = 0$ and F' is decreasing on J. Because $F'(\delta) = 0$, $F' > F'(\delta) = 0$ and F is increasing on J. Noting that $F(0) = F(\delta) = 0$, we have $F \equiv 0$, i.e., $f = g$.

Suppose property (3) does not hold, i.e., $\lim_{w \to 0^+} F'''(w) \leq 0$, so that $F''' \leq \lim_{w \to 0^+} F'''(w) \leq 0$ on J. Following a similar argument as above, on interval J, F'' is decreasing, F' is increasing, and F is decreasing. Again we arrive at the trivial case: $f = g$.

Suppose property (2) does not hold, i.e., $\lim_{w \to 0^+} F''(w) \geq 0$. From properties (3) and (4), we know that F'' is first increasing and then decreasing on J. Thus, $F'' \geq 0$ on J. As above, we find that F' is increasing and F is decreasing on J, leading again to the trivial case.

Suppose property (1) does not hold, i.e., $\lim_{w \to 0^+} F'(w) \leq 0$. Property (2), along with the facts that F'' is first increasing and then decreasing on J and $F''(\delta) = 0$, implies that F' is first decreasing and then increasing on J. Therefore, $F' \leq 0$, and F is nonincreasing on J, so that $f = g$. ∎

When f is increasing and concave and $g_2 \leq 0$, g is increasing and concave by Theorem 5. 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 $-f''$. Of course, via δ we have control over both $g'(0)$ and $-g''(0)$.

2.3. Monotonicity of $g_1 = g'(0)$ in δ. For a particular increasing and concave f, it may seem intuitive that $g_1 = g'(0)$ should be decreasing in the smoothing parameter δ for $\delta > 0$ in the domain of f. This would be a very useful property, because then we could easily find a value for δ to achieve a target value for g_1 using a simple univariate search. As we explore the tendency of g_1 with respect to δ, it is useful to emphasize the functional dependence of g_1 on δ by writing $g_1(\delta)$.

It is straightforward to calculate the derivative of this function:

$$\frac{dg_1(\delta)}{d\delta} = -\frac{3}{\delta^2}f(\delta) + \frac{3}{\delta}f'(\delta) - \frac{3}{2}f''(\delta) + \frac{\delta}{2}f'''(\delta).$$

Unfortunately, for concrete functions f, it may not be so practical to check that this derivative is nonpositive for $\delta > 0$ in the domain of f. So, to establish such monotonicity in a *practically verifiable manner*, we need to make some appropriate hypotheses.

THEOREM 7. *Let f be a univariate function having a domain $I := [0, U)$, where $U \in \{w \in \mathbb{R} \ : \ w > 0\} \cup \{+\infty\}$. 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 $(0, U)$.

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 Proposition 6, we have $f'''(\delta) - g_3(\delta) < 0$, which gives us $g_1'(\delta) < 0$ on $(0, U)$. ∎

Applying Theorem 7, it is now a simple matter to verify that when f is a root function, $g'(0)$ behaves as expected with respect to the parameter δ.

COROLLARY 8. *Let $f(w) := w^p$ for some $0 < p < 1$. Then $g_1(\delta)$ is decreasing on $(0, +\infty)$.*

Proof. We must verify that f satisfies the hypothesis of Theorem 7. Consider the following derivatives of f on $(0, +\infty)$:

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

Because $0 < p < 1$, $f^{(4)}(w) < 0$ on $(0, +\infty)$, which implies f''' is decreasing on $(0, +\infty)$, and thus Theorem 7 applies. ∎

The next example demonstrates that Theorem 7 applies to functions that are not root functions.

Example 9. Let $f(w) := \text{ArcSinh}(\sqrt{w}) = \log(\sqrt{w} + \sqrt{1+w})$ for $w \geq 0$. Checking the hypotheses of Theorem 7, we calculate the following derivatives of f on $(0, +\infty)$:

$$\begin{aligned}
f'(w) &= \frac{1}{2\sqrt{w(w+1)}}, \\
f''(w) &= -\frac{2w+1}{4\,(w(w+1))^{\frac{3}{2}}}, \\
f'''(w) &= \frac{8w^2+8w+3}{8\,(w(w+1))^{\frac{5}{2}}}, \\
f^{(4)}(w) &= -\frac{48w^3+72w^2+54w+15}{16\,(w(w+1))^{\frac{7}{2}}}.
\end{aligned}$$

For $w \in (0, +\infty)$, it is easy to verify that $f^{(4)}(w) < 0$, which implies f''' is decreasing on $(0, +\infty)$. By Theorem 7, $g_1(\delta)$ is decreasing for $\delta \in (0, +\infty)$.

3. Lower bound for f.

In section 3.1, we establish Theorem 10: g provides a lower bound for a broad class of functions f which includes all root functions, solving an open problem from [LS17]. We provide an example to demonstrate that this class of functions contains functions beyond root functions. In section 3.2, we present variations on the hypotheses of Theorem 10, along with supporting examples. In section 3.3, we veer briefly from root-like functions to provide an example of a function f that is neither increasing nor concave, but for which g serves as a lower bound. In other words, we show that Theorem 10 does not require f to be increasing and concave. Also we give an example to show that for an increasing and concave function f, (T^*_δ) is not necessary for Theorem 10. In section 3.4, we demonstrate that the worst-case error of g as an approximation of f is increasing with respect to δ under the same conditions as Theorem 7.

3.1. Lower bounding.

Because the δ-smoothing g is simply f on (δ, U), we restrict our attention to lower bounding on the interval $[0, \delta]$. The parameter δ provides control over $g'(0)$, and in a predictable manner under the hypotheses of Theorem 7. As δ vanishes, g tends to f, but the choice of δ is dictated by the numerical tolerance of the software with respect to the value of $g'(0)$. The following theorem shows that g provides a lower bound for a broad class of functions f which is neither necessarily increasing nor concave (examples are in section 3.3), but includes all root functions.

THEOREM 10. *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. Assume further that*

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

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

Proof. This is a special case of "osculating interpolation" (also know as Hermite interpolation; see [BF11], for example). We are going to 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 $w \in (0, \delta)$, define $K := K(w)$ for simplicity, and introduce a new function F with respect to x as

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

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 $(0, \delta)$, according to Rolle's theorem, there exists $0 < w_1 < w < \eta_1 < \delta$ such that $F'(w_1) = F'(\eta_1) = 0$. Noting that $F'(\delta) = 0$ and that $F'(x)$ is differentiable on $[w_1, \delta]$, we apply Rolle's theorem and get $w_1 < w_2 < \eta_1 < \eta_2 < \delta$ such that $F''(w_2) = F''(\eta_2) = 0$. Using Rolle's theorem again on $F''(x)$, with $F''(\delta) = 0$ and $F''(x)$ 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 $K < 0$. ∎

It is easy to see that Theorem 10 has a counterpart when f''' is increasing rather than decreasing, by applying Theorem 10 to $-f$.

COROLLARY 11. *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. Assume further that*

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

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

Returning to our primary motivation, the following corollary demonstrates that Theorem 10 generalizes the result in [LS17], which states that g is a lower bound for root functions of the form $f(w) = w^{1/q}$ for integer $q \geq 2$.

COROLLARY 12. *Let $f(w) := w^p$ for some $0 < p < 1$. For all $\delta > 0$, if g is the δ-smoothing of f, then $g(w) \leq f(w)$ for $w \geq 0$.*

Proof. According to Corollary 8, we can simply verify that f''' is decreasing on $(0, \delta]$, and thus Theorem 10 applies. ∎

The next example demonstrates that there are other increasing and concave functions (besides root functions) to which Theorem 10 applies.

Example 13. Consider $f(w) := \mathrm{ArcSinh}(\sqrt{w}) = \log(\sqrt{w} + \sqrt{1 + w})$ for $w \geq 0$. We demonstrate that f satisfies the conditions of Theorem 10, so that g lower bounds f on $[0, +\infty)$. From Example 9, we can easily verify that f''' is decreasing on $(0, \delta]$, and thus f satisfies the conditions of Theorem 10. ∎

3.2. More possibilities for a lower bound. We digress again to provide results that take us beyond root functions. In particular, there are other possibilities for f''' (besides decreasing) to ensure that g is a lower bound on f. For example, in Theorem 14 below, if we have f''' first decreasing and then increasing on $(0, \delta]$, we can add conditions almost identical to properties (1)–(4) of Proposition 6 to ensure a lower-bounding g.

THEOREM 14. *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. Assume further that*

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

Moreover, suppose that
(1) $\lim_{w \to 0^+} f'(w) > g_1$,
(2) $\lim_{w \to 0^+} f''(w) < g_2$,
(3) $\lim_{w \to 0^+} f'''(w) > g_3$,
($4^{\leq}$) $f'''(\delta) \leq g_3$;
then $f(w) \geq g(w)$ *for all* $w \in [0, +\infty)$.

Proof. According to the definition of g, we have $g(0) = 0$, $g^{(i)}(\delta) = f^{(i)}(\delta)$ for $i = 0, 1, 2$. We consider the function $F(w) := f(w) - g(w)$, for $w \in [0, \delta]$, which has

$$F(0) = F(\delta) = F'(\delta) = F''(\delta) = 0.$$

In what follows, we begin with the third derivative of F and work our way to the conclusion that $F(w) > 0$ for $w \in (0, \delta)$.

First, we note that $F'''(w) = f'''(w) - g_3$ is a first decreasing and then increasing function with

$$\lim_{w \to 0^+} F'''(w) > 0 \quad \text{and} \quad F'''(\delta) \leq 0.$$

Therefore, there exists exactly one root of F''' in $(0, \delta)$, which we denote by w_0.

From this, we conclude that F'' is increasing on $[0, w_0]$ and decreasing on $[w_0, \delta]$, so that $F''(w_0) > F''(\delta) = 0$. Combining this fact with

$$\lim_{w \to 0^+} F''(w) = \lim_{w \to 0^+} (f''(w) - g_3 w - g_2) = \lim_{w \to 0^+} (f''(w) - g_2) < 0,$$

we see that F'' has exactly one root in $(0, w_0)$, which we denote by w_1. In summary, we have

$$F''(w) \begin{cases} < 0, & 0 < w < w_1, \\ = 0, & w \in \{w_1, \delta\}, \\ > 0, & w_1 < w < \delta. \end{cases}$$

Applying these results, we conclude that F' is decreasing on $[0, w_1]$ to a minimum of $F'(w_1) < F'(\delta) = 0$. Because

$$\lim_{w \to 0^+} F'(w) = \lim_{w \to 0^+} \left(f'(w) - \frac{1}{2} g_3 w^2 - g_2 w - g_1 \right) = \lim_{w \to 0^+} (f'(w) - g_1) > 0,$$

we see that F' has exactly one root in $(0, w_1)$, which we denote by w_2, and

$$F'(w) \begin{cases} > 0, & 0 < w < w_2, \\ = 0, & w \in \{w_2, \delta\}, \\ < 0, & w_2 < w < \delta. \end{cases}$$

By properties of its derivative, $F(w)$ is increasing on $[0, w_2]$ and decreasing on $[w_2, \delta]$. Because $F(0) = F(\delta) = 0$, we have that $F(w) = f(w) - g(w) > 0$ for $w \in (0, \delta)$. Recalling that $f(w) = g(w)$ for $w \in \{0\} \cup [\delta, \infty)$, we conclude that $g \leq f$ on $[0, +\infty)$. ∎

Remark 15. If f''' is decreasing, then the hypotheses of Theorem 10 imply properties (1)–(4) of Proposition 6. By employing these properties, we can use the same proof technique from Theorem 14 to prove Theorem 10. As in the proof of Theorem 14, we can prove $f \geq g$ by considering the function $F := f - g$. The third derivative, $F'''(w) = f'''(w) - g_3$, is decreasing with $\lim_{w \to 0^+} F'''(w) = \lim_{w \to 0^+} f'''(w) - g_3 > 0$

and $F'''(\delta) = f'''(\delta) - g_3 < 0$. Therefore, there exists exactly one root of F''' in $(0, \delta)$, which we denote by w_0. The rest of the proof is the same as that of Theorem 14. From the proof, we can find the roots $w_0,\ w_1,\ w_2$ of the derivatives of the function F and the same characterization for the derivatives as Theorem 14. We require this characterization in the proofs of Theorems 22 and 23.

In order to demonstrate the applicability of Theorem 14, we construct Example 17 using the general form described in Example 16 below. Inspired by Example 4, we build a continuous piecewise-defined function specified as a quintic on $[0, w_0]$, and a shifted square root function on $(w_0, +\infty)$. We will use the same general form again in Example 20.

Example 16. Consider the function

$$f(w) := \begin{cases} a_5w^5 + a_4w^4 + a_3w^3 + a_2w^2 + a_1w, & 0 \le w \le w_0, \\ a\sqrt{w-c} + b, & w > w_0. \end{cases}$$

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:

$$\begin{aligned} c &= w_0 + \frac{3f''(w_0)}{2f'''(w_0)}, \\ a_1 &= -2f''(w_0)(w_0 - c) - (5a_5w_0^4 + 4a_4w_0^3 + 3a_3w_0^2 + 2a_2w_0), \\ a &= \frac{8f'''(w_0)(w_0-c)^{\frac{5}{2}}}{3}, \\ b &= f(w_0) - a\sqrt{w_0 - c}. \end{aligned}$$

For $\delta \le w_0$, we have the δ-smoothing $g(w) = g_1w + \frac{1}{2}g_2w^2 + \frac{1}{6}g_3w^3$, where

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

(The requirement that $\frac{f''(w_0)}{f'''(w_0)} \le 0$ ensures that $\sqrt{w-c}$ is real valued for $w > w_0$.) $\square$

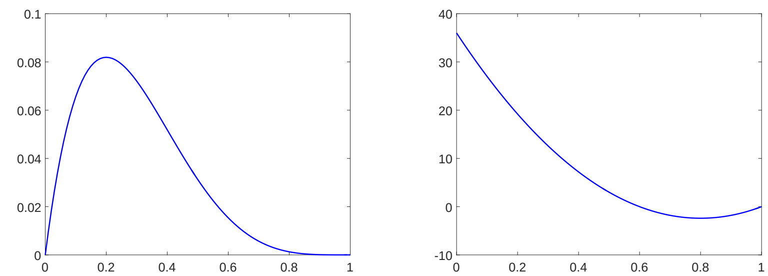
And now we are ready to build a function that satisfies the hypotheses of Theorem 14.

Example 17. Following Example 16, let

$$f(w) := \begin{cases} a_5w^5 + a_4w^4 + a_3w^3 + a_2w^2 + a_1w, & w \le w_0, \\ a\sqrt{w-c} + b, & w > w_0. \end{cases}$$

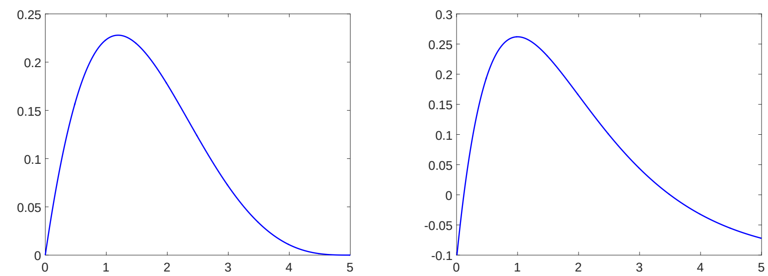
We seek parameters of f for which $g_2 \le 0$ and all conditions of Theorem 14 are satisfied. For $0 \le w \le w_0$, we have $f'''(w) = 60a_5w^2 + 24a_4w + 6a_3$ and $f''''(w) = 120a_5w + 24a_4$. In order to have $f'''(w)$ first decreasing and then increasing on $[0, \delta]$, we require $a_5 > 0$, $a_4 < 0$, and $5a_5\delta + a_4 > 0$.

For example, choose $\delta = 1$, $a_4 = -4$, $a_5 = 1$. It is straightforward to verify that the conditions of Theorem 14 now hold. Next, we choose $a_3 = 10$ and $w_0 = 2$ to have $f'''(w_0) > 0$, and we choose $a_2 = -50$ to have $f''(w_0) < 0$ and $g_2 = -16a_5\delta^3 - 6a_4\delta^2 + 2a_2 < 0$. Then we compute the remaining parameters $(a_1, a, b, c) = (132, \frac{4\sqrt{6}}{3}, \frac{332}{3}, \frac{11}{6})$. We can see the difference between g and f in Figure 1(a) and the tendency of f''' in Figure 1(b). $\square$



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

FIG. 1. $a_5 = 1$, $a_4 = -4$, $a_3 = 10$, $w_0 = 2$, $a_2 = -50$ *in Example* 16.



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

FIG. 2. $\delta = 5$, $g \leq f$.

In the next example we see yet another possibility for the conditions on f''' under which g is increasing and concave. Interestingly, the same function with a different choice of δ does not satisfy $g \leq f$, but instead provides an example where Corollary 11 applies, and so we can conclude that $g \geq f$.

Example 18. Consider the function $f(w) = -(w+3)e^{-w} + 3$, which has the following derivatives:

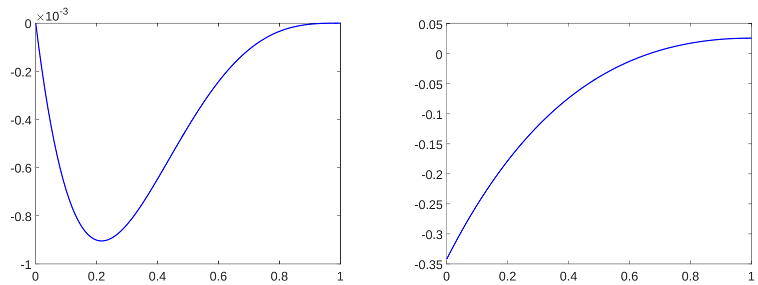
$$\begin{aligned}
f'(w) &= (w+2)e^{-w}, \\
f''(w) &= -(w+1)e^{-w}, \\
f'''(w) &= we^{-w}, \\
f^{(4)}(w) &= -(w-1)e^{-w}.
\end{aligned}$$

Also $f(0) = 0$, and f is increasing, concave, and thrice differentiable on $[0, +\infty)$. Moreover, $f'''(w)$ is increasing on $[0, 1]$ and then decreasing on $[1, +\infty)$.

For $\delta = 5$, $g \leq f$ on their common domain (see Figure 2(a)), even though this function satisfies neither the conditions in Theorem 10 nor those in Theorem 14. Instead, f''' is increasing and then decreasing on $[0, \delta]$.

For $\delta = 1$, f''' is increasing on $[0, \delta]$. We conclude that g upper bounds f (see Figure 3(a)) via Corollary 11.

If we add a small positive multiple of the square root function $\sqrt{w}$ to f, then we can get other possibilities for the tendency of f'''. For example, for $\delta = 5$, $\epsilon = 5 \times 10^{-5}$,



(a) $f(w) - g(w)$ (b) $f'''(w) - g'''(0)$

FIG. 3. $\delta = 1$, $g \geq f$.

$f(w) + \epsilon\sqrt{w}$ satisfies $g \leq f$, while f''' is decreasing, then increasing, then decreasing again. □

3.3. Role of the increasing and concave properties. Theorem 10 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. The following simple example realizes such a scenario.

Example 19. Let $\delta = 1$. Then $f(w) := -w^4 + 6w^2 - 8w$ is decreasing and convex, and satisfies $f^{(4)}(w) = -1 < 0$, which implies that f''' is decreasing on $(0, \delta]$, and Theorem 10 holds; g is a lower bound for f.

Returning to root functions and their relatives, it would be nice if we could count on the lower bounding g to be increasing and concave whenever Theorem 10 applies to an increasing and concave f. In section 2, we gave a necessary and sufficient condition (T^*_δ) $(g_2 \leq 0)$ for g to be increasing and concave. So we have the following natural question: do we automatically satisfy (T^*_δ) when Theorem 10 applies to functions that are increasing and concave? Unfortunately, the answer to this question is "no," as demonstrated by Example 20.

To motivate the development of Example 20, we note that when $f'''(\delta) \geq 0$, property (4) of Proposition 6 implies that $g_3 \geq 0$, as well. So to get an example of a function g that satisfies the hypotheses of Theorem 10 but is not both increasing and concave, we need to have $f'''(\delta) < g_3 < 0$. We impose the required properties in the context of the general form presented in Example 16.

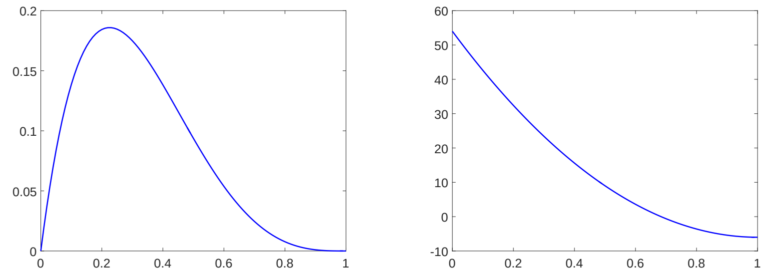
Example 20. Consider the function f described in Example 16. We seek parameters of f for which $g_3 < 0$, and all conditions of Theorem 10 are satisfied. For $0 \leq w \leq w_0$, we have $f'''(w) = 60a_5w^2 + 24a_4w + 6a_3$ and $f''''(w) = 120a_5w + 24a_4$. In order to have $f'''(w)$ decreasing on $[0, \delta]$ and $f'''(\delta) < 0 < f'''(w_0)$ for $\delta < w_0$, we require $a_5 > 0$, $a_4 < 0$, and $5a_5\delta + a_4 \leq 0$.

For example, choose $\delta = 1$, $a_3 = 0$, $a_4 = -5$, $a_5 = 1$. It is straightforward to verify that the conditions of Theorem 10 now hold. Next, we choose $w_0 = 3$ to have $f'''(w_0) > 0$, and we choose $a_2 = -3$ to have $f''(w_0) < 0$. By calculating a_1, a, b, c, we get an example with $g_3 = -54 < 0$.

Note that $g_2 = 8 > 0$, so the associated function g is not concave.

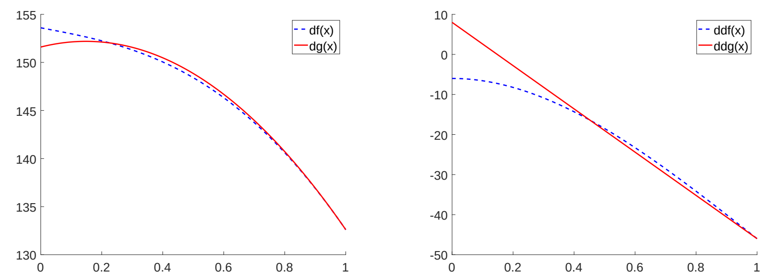
Figure 4 shows the difference between f and g, and also the tendency of f'''. Figure 5 shows the first and second derivatives of $f(w)$ and $g(w)$, which demonstrates that $f(w)$ is increasing and concave, while $g(w)$ is not. □

We encapsulate the result implied by Example 20 in the following observation.



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

FIG. 4. $a_5 = 1$, $a_4 = -5$, $a_3 = 0$, $w_0 = 3$, $a_2 = -3$ *in Example* 16.



(a) $f'(w),\ g'(w)$ (b) $f''(w),\ g''(w)$

FIG. 5. $a_5 = 1$, $a_4 = -5$, $a_3 = 0$, $w_0 = 3$, $a_2 = -3$ *in Example* 16.

Observation 21. For an increasing concave function f, the hypotheses of Theorem 10 do not imply that the smoothing g is increasing and concave, i.e., (T_δ^*) is not implied by the hypotheses of Theorem 10, even for increasing concave f.

3.4. Monotonicity of $\|f - g\|_\infty$ in δ. In section 2.3, we demonstrated that the derivative of g at zero is *decreasing* in δ when f''' is decreasing. This is useful for calculating the least value of δ to obtain a target value for $g'(0)$. In this subsection, we demonstrate that the worst-case error of g as an approximation of f is *increasing* in δ, again when f''' is decreasing. This is useful for calculating the greatest value of δ to obtain a target value for the worst-case error of g as an approximation of f. Of course it can be that tolerances for the derivative of g at zero and for the worst-case error of g as an approximation of f can be incompatible (i.e., no valid choice of δ satisfying both). Before continuing, we note that (i) section 5 of [LS17] gives results on the average performance of g when f is a root function, and (ii) Theorem 1, part 6 of [DFLV15] and [DFLV18] give results on the worst-case performance of g when f is the square-root function.

Formally now, we define $F(w) := f(w) - g(w)$, and

$$\|f - g\|_\infty := \max_{w \in [0,\delta]} |f(w) - g(w)| = \max_{w \in [0,\delta]} |F(w)|\,.$$

Note that g and its coefficients $g_1,\ g_2,\ g_3$ are functions of δ, and so $\|f - g\|_\infty$ is also a function of $\delta \in (0, U)$.

THEOREM 22. *Let f be a univariate function having a domain $I := [0, U)$, where $U \in \{w \in \mathbb{R} \ : \ w > 0\} \cup \{+\infty\}$. 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 $\|f - g\|_\infty$ is increasing on $(0, U)$.

Proof. If f''' is decreasing, then by Theorem 10, $F(w) := f(w) - g(w) > 0$ on $(0, \delta)$. Define

$$w_2 := \operatorname{argmax}\{F(w) \ : \ w \in [0, \delta]\}.$$

Then $\|f - g\|_\infty = \max_{w \in [0,\delta]} F(w) = F(w_2)$. As mentioned in Remark 15, we can use the same proof technique from Theorem 14 to prove now that $F'(w_2) = 0$ and $F''(w_2) < 0$.

Clearly w_2 is a function of δ, and we are going to demonstrate that w_2 is actually a differentiable function with respect to δ. For any $\delta \in (0, U)$, let

$$G(x, y) := f'(y) - \frac{g_3(x)}{2} y^2 - g_2(x) y - g_1(x).$$

We have $G(\delta, w_2) = F'(w_2) = 0$, and

$$\frac{\partial G(\delta, w_2)}{\partial y} = f''(w_2) - g''(w_2) = F''(w_2) < 0.$$

By the implicit function theorem, there exists a unique differentiable function $y = w_2(x)$ such that $w_2(\delta) = w_2$ and $G(x, w_2(x)) = 0$ for $x \in N(\delta)$, where $N(\delta)$ is an open interval containing δ.

Therefore,

$$\begin{aligned}\frac{dF(w_2(\delta))}{d\delta} &= f'(w_2)\frac{dw_2(\delta)}{d\delta} - g'(w_2)\frac{dw_2(\delta)}{d\delta} - \frac{dg_3(\delta)}{d\delta}\frac{w_2(\delta)^3}{6} \\ &\quad - \frac{dg_2(\delta)}{d\delta}\frac{w_2(\delta)^2}{2} - \frac{dg_1(\delta)}{d\delta} w_2(\delta) \\ &= -\frac{dg_3(\delta)}{d\delta}\frac{w_2(\delta)^3}{6} - \frac{dg_2(\delta)}{d\delta}\frac{w_2(\delta)^2}{2} - \frac{dg_1(\delta)}{d\delta} w_2(\delta) \\ &= -\frac{w_2(\delta)(w_2(\delta) - \delta)^2}{2\delta}(f'''(\delta) - g_3(\delta)) > 0.\end{aligned}$$

The second equality follows from $F'(w_2) = 0$, the third equality follows from the facts

$$\frac{dg_1(\delta)}{d\delta} = \frac{\delta}{2}(f'''(\delta) - g_3(\delta)), \ \frac{dg_2(\delta)}{d\delta} = -2(f'''(\delta) - g_3(\delta)), \ \frac{dg_3(\delta)}{d\delta} = \frac{3}{\delta}(f'''(\delta) - g_3(\delta)),$$

and the last inequality follows from $w_2(\delta) > 0$ and $f'''(\delta) - g_3(\delta) < 0$ (by Proposition 6). Thus $\|f - g\|_\infty = F(w_2(\delta))$ is increasing on $(0, U)$. ∎

4. Comparison with shift smoothing. We wish to compare our smoothing g with the natural and frequently used *shift smoothing* (for root functions and their relatives): $h(w) := f(w + \lambda) - f(\lambda)$ for $w \in [0, +\infty)$, with $\lambda > 0$ chosen so that $h'(0)$ is numerically tolerable. When the function f that we are considering is globally concave (and because we assume that $f(0) = 0$), f is subadditive, and so h is a lower bound for f on its domain.

Clearly we have $g(0) = h(0) = 0$, and $h(w) \leq f(w) = g(w)$ for $w \geq \delta$, so we are interested in comparing g and h on the interval $(0, \delta)$. Because g is defined based on a choice of δ, and h is defined based on a choice of λ, a fair comparison is achieved by making these choices so that their derivatives at 0 are the same. In this way, both smoothings of f have the same maximum derivative (under the hypotheses of our result (Theorem 23); that is, both smoothings have their derivatives maximized at zero, where f' is assumed to blow up, under the hypotheses of Theorem 23, which imply the hypotheses of Theorem 7).

In order to match derivatives at 0, let $h'(0) = f'(\lambda) = g'(0) = g_1 = 3f(\delta)/\delta - 2f'(\delta) + \delta f''(\delta)/2$. Then we have

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

the value of λ, defined in terms of δ, for which $h'(0) = g_1$.

In [LS17], it is proved that $h \leq g$ for root functions $f = w^p$, with $p = 1/q$ for integers $2 \leq q \leq 10{,}000$. We generalize this result to a class of functions that shares many properties with root functions, and includes all root functions $f(w) := w^p$ for $0 < p < 1$. Note that the conditions of Theorem 23 are more restrictive than those of Theorem 10; here we require that f''' is decreasing on $(0, 2\delta)$, rather than $(0, \delta]$, and we require that $f'''(w) \geq 0$ for $w \in (0, 2\delta)$. This last condition implies that unlike Theorem 10 (see Observation 21), (T^*_δ) is implied by the hypotheses of Theorem 23.

THEOREM 23. *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 $U \geq 2\delta > 0$. 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, where the shift constant $\hat{\lambda}$ is chosen so that $h'(0) = g_1$, i.e., $\hat{\lambda} = (f')^{-1}(g_1)$.

Proof. With condition (I), f satisfies the hypotheses of Proposition 6, so we have all the properties of Proposition 6. 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 (repeated here for convenience):

$$g_1 = \frac{3f(\delta)}{\delta} - 2f'(\delta) + \frac{\delta f''(\delta)}{2},$$
$$g_2 = -\frac{6f(\delta)}{\delta^2} + \frac{6f'(\delta)}{\delta} - 2f''(\delta),$$
$$g_3 = \frac{6f(\delta)}{\delta^3} - \frac{6f'(\delta)}{\delta^2} + \frac{3f''(\delta)}{\delta}.$$

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 $0 < \hat{\lambda} < \delta$, by condition (II),

$$H''(\delta) = f''(\delta) - f''(\delta + \hat{\lambda}) < 0,$$

by condition (I),

$$H''' \text{ is increasing on } (0, \delta],$$

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

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

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 $w \in I$.

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 Theorem 10, 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 Remark 15.

From Remark 15, 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 $\hat{\lambda} < w_0$. The function f satisfies condition (T_δ) of Theorem 3, 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 $\hat{\lambda} < w_2$ $(< w_0)$.

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 $w_2 < \delta \leq -g_2/g_3$. Now we can obtain the upper bound

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

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 $g_2 > f''(\hat{\lambda})$. ☐

The following corollary demonstrates that Theorem 23 generalizes the result in [LS17], 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 $2 \leq q \leq 10{,}000$.

COROLLARY 24. *Let $f(w) := w^p$ for some $0 < p < 1$. Then $h(w) \leq g(w)$ for all $w \in [0, +\infty)$.*

Proof. According to the derivatives of f in Corollary 12, it is easy to see that $f(w)$ satisfies conditions (I) and (II) of Theorem 23. Therefore the conclusion follows. ☐

Finally, we note that Theorem 23 also applies to the nonroot function that we have explored throughout.

Example 25. Let $f(w) = \mathrm{ArcSinh}(\sqrt{w}) = \log(\sqrt{w} + \sqrt{1+w})$ for $w \geq 0$. Then $h(w) \leq g(w)$ for all $w \geq 0$.

5. Conclusions.

It may seem like a challenge to automatically identify and apply the techniques that we have presented. But in the context of global optimization aimed at factorable formulations, the algorithm/software designer has a limited number of library functions to analyze. Furthermore, even in a fully extensible system, we could automatically apply major parts of our ideas. For example, once a univariate function f has been identified to satisfy $f(0) = 0$, f is increasing and concave on say $[0, +\infty)$, f is twice differentiable on all of $(0, +\infty)$, but $f'(0)$ undefined or intolerably large, the rest of our methodology (i.e., calculating g and identifying its properties) can be applied automatically. A start has been made on making accommodations for our methodology in `SCIP`. Hopefully we will see more advances in such a direction, contributing to the overall goal of making MINLO software more reliable and useful.

From a mathematical point of view, still aiming at potential impact on MINLO software, we could look at functions f with domain a two-variable polyhedron P, where f is nice and smooth on the interior of P, but not differentiable on part of the boundary of P.

Acknowledgment. The authors thank an anonymous reviewer who suggested the problem solved in section 3.4.

REFERENCES

[Ach09] T. Achterberg, *SCIP: Solving constraint integer programs*, Math. Program. Comput., 1 (2009), pp. 1–41.

[BDL+06] C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth, *An MINLP solution method for a water network problem*, in Algorithms—ESA 2006, Lecture Notes in Comput. Sci. 4168, Springer, Berlin, 2006, pp. 696–707.

[BDL+12] C. Bragalli, C. D'Ambrosio, J. Lee, A. Lodi, and P. Toth, *On the optimal design of water distribution networks*, Optim. Eng., 13 (2012), pp. 219–246.

[BF11] R. L. Burden and J. D. Faires, *Numerical Analysis*, Cengage Learning, Boston, MA, 2011.

[BLL+09] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter, *Branching and bounds tightening techniques for non-convex MINLP*, Optim. Methods Softw., 24 (2009), pp. 597–634.

[DFLV15] C. D'Ambrosio, M. Fampa, J. Lee, and S. Vigerske, *On a nonconvex MINLP formulation of the Euclidean Steiner tree problem in n-space*, in Experimental Algorithms, Lecture Notes in Comput. Sci. 9125, E. Bampis, ed., Springer, New York, 2015, pp. 122–133.

[DFLV18] C. D'Ambrosio, M. Fampa, J. Lee, and S. Vigerske, *On a nonconvex MINLP formulation of the Euclidean Steiner tree problem in n-space: Missing proofs*, Optim. Lett. (2018), https://link.springer.com/article/10.1007/s11590-018-1295-1.

[GGH+16] T. Gally, A. M. Gleixner, G. Hendel, T. Koch, S. J. Maher, M. Miltenberger, B. Müller, M. E. Pfetsch, C. Puchert, D. Rehfeldt, S. Schenker, R. Schwarz, F. Serrano, Y. Shinano, S. Vigerske, D. Weninger, M. Winkler, J. T. Witt, and J. Witzig, *The SCIP Optimization Suite* 3.2, Technical report ZR 15-60, Zuse Institute Berlin, 2016; available at http://www.optimization-online.org/DB_HTML/2016/03/5360.html.

[GMS13] I. Gentilini, F. Margot, and K. Shimada, *The travelling salesman problem with neighbourhoods: MINLP solution*, Optim. Methods Softw., 28 (2013), pp. 364–378.

[LS17] J. Lee and D. Skipper, *Virtuous smoothing for global optimization*, J. Global Optim., 69 (2017), pp. 677–697.

[MF14] R. Misener and C. A. Floudas, *ANTIGONE: Algorithms for coNTinuous/Integer Global Optimization of Nonlinear Equations*, J. Global Optim., (2014), pp. 503–526.

[MKV17] B. Müller, R. Kuhlmann, and S. Vigerske, *On the performance of NLP solvers within global MINLP solvers*, in Operations Research Proceedings 2017, Oper. Res. Proc., N. Kliewer, J. Ehmke, and R. Borndörfer, eds., Springer, Cham, 2017, pp. 633–639.

[TS02] M. Tawarmalani and N. V. Sahinidis, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*, Nonconvex Optim. Appl. 65, Springer, New York, 2002.

[Wäc09] A. Wächter, *Short tutorial: Getting started with IPOPT in* 90 *minutes*, in Dagstuhl Seminar Proceedings 09061 – Combinatorial Scientific Computing, U. Naumann, O. Schenk, H. D. Simon, and S. Toledo, eds., Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, Dagstuhl, 2009.

[WB06] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57.