



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

On linear and quadratic Lipschitz bounds for twice continuously differentiable functions

Citation for published version:

Bunin, GA, François, G & Bonvin, D 2014 'On linear and quadratic Lipschitz bounds for twice continuously differentiable functions'.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



On linear and quadratic Lipschitz bounds for twice continuously differentiable functions

Gene A. Bunin, Grégory François, Dominique Bonvin

Submitted: June 17, 2014

Abstract Lower and upper bounds for a given function are important in many mathematical and engineering contexts, where they often serve as a base for both analysis and application. In this short paper, we derive piecewise linear and quadratic bounds that are stated in terms of the Lipschitz constants of the function and the Lipschitz constants of its partial derivatives, and serve to bound the function's evolution over a compact set. While the results follow from basic mathematical principles and are certainly not new, we present them as they are, from our experience, very difficult to find explicitly either in the literature or in most analysis textbooks.

Keywords Lipschitz bounds · Twice continuously differentiable functions · Piecewise linear and piecewise quadratic bounds

1 Overview

We consider a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ of the variables $x \in \mathbb{R}^n$ that is twice continuously differentiable (C^2) over an open set containing the compact set \mathcal{X} . Because f is C^2 over \mathcal{X} , its first and second derivatives on this set exist and must be bounded by the Lipschitz constants

$$\underline{\kappa}_i < \left. \frac{\partial f}{\partial x_i} \right|_x < \bar{\kappa}_i, \quad i = 1, \dots, n \quad (1)$$

$$\underline{M}_{ij} < \left. \frac{\partial^2 f}{\partial x_j \partial x_i} \right|_x < \bar{M}_{ij}, \quad i, j = 1, \dots, n \quad (2)$$

for all $x \in \mathcal{X}$.

The evolution of f between any two points $x_a, x_b \in \mathcal{X}$ may then be bounded as

$$f(x_b) - f(x_a) \geq \sum_{i=1}^n \min \left[\frac{\underline{\kappa}_i}{\bar{\kappa}_i} (x_{b,i} - x_{a,i}), \right], \quad (3)$$

$$f(x_b) - f(x_a) \leq \sum_{i=1}^n \max \left[\frac{\underline{\kappa}_i}{\bar{\kappa}_i} (x_{b,i} - x_{a,i}), \right], \quad (4)$$

where $x_{a,i}$ and $x_{b,i}$ denote the i^{th} elements of the vectors x_a and x_b , respectively.

The bounds (3) and (4) are piecewise linear in x . Alternatively, one may also use the piecewise quadratic bounds

$$f(x_b) - f(x_a) \geq \nabla f(x_a)^T (x_b - x_a) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \min \left[\frac{\underline{M}_{ij}}{\bar{M}_{ij}} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}), \right], \quad (5)$$

$$f(x_b) - f(x_a) \leq \nabla f(x_a)^T (x_b - x_a) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \max \left[\frac{\underline{M}_{ij}}{\bar{M}_{ij}} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}), \right], \quad (6)$$

which are locally less conservative but also require more knowledge in the form of both the gradient and the Lipschitz constants on the partial derivatives of f . While one may generalize this pattern to even higher orders, we will content ourselves with the linear and quadratic cases as we believe these to be sufficient for most applications – see, however, [2] for a discussion of the cubic case.

2 Derivation of the Linear Bounds

To limit our analysis to a single dimension, we will consider the line segment between x_a and x_b . The following one-dimensional parameterization is used:

$$\hat{f}(\gamma) = f(x(\gamma)), \quad (7)$$

with $x(\gamma) = x_a + \gamma(x_b - x_a)$, $\gamma \in [0, 1]$. As f is C^2 , it follows that \hat{f} is as well, which allows us to use the Taylor series expansion between $\gamma = 0$ and $\gamma = 1$, together with the mean-value theorem [1], to state:

$$\hat{f}(1) = \hat{f}(0) + \frac{d\hat{f}}{d\gamma} \Big|_{\tilde{\gamma}}, \quad (8)$$

for some $\tilde{\gamma} \in (0, 1)$. We proceed to define the first-order derivative in terms of the original function f . To do this we apply the chain rule:

$$\frac{d\hat{f}}{d\gamma} \Big|_{\gamma} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x(\gamma)} \frac{dx_i}{d\gamma} \Big|_{\gamma} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x(\gamma)} (x_{b,i} - x_{a,i}). \quad (9)$$

Noting that $\hat{f}(0) = f(x_a)$ and $\hat{f}(1) = f(x_b)$, one may substitute (9) into (8) to obtain

$$f(x_b) = f(x_a) + \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i}). \quad (10)$$

Because $x(\tilde{\gamma}) \in \mathcal{X}$, we may use (1) to bound the individual summation components as

$$x_{b,i} - x_{a,i} \geq 0 \Leftrightarrow \underline{\mathbf{K}}_i(x_{b,i} - x_{a,i}) \leq \frac{\partial f}{\partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i}) \leq \overline{\mathbf{K}}_i(x_{b,i} - x_{a,i}), \quad (11)$$

$$x_{b,i} - x_{a,i} \leq 0 \Leftrightarrow \overline{\mathbf{K}}_i(x_{b,i} - x_{a,i}) \leq \frac{\partial f}{\partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i}) \leq \underline{\mathbf{K}}_i(x_{b,i} - x_{a,i}),$$

or, to account for both cases, as

$$\min \left[\frac{\underline{\mathbf{K}}_i(x_{b,i} - x_{a,i})}{\overline{\mathbf{K}}_i(x_{b,i} - x_{a,i})} \right] \leq \frac{\partial f}{\partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i}) \leq \max \left[\frac{\underline{\mathbf{K}}_i(x_{b,i} - x_{a,i})}{\overline{\mathbf{K}}_i(x_{b,i} - x_{a,i})} \right]. \quad (12)$$

Substituting this result into (10) then yields (3) and (4).

3 Derivation of the Quadratic Bounds

The derivation is similar to that of the linear case, and simply involves taking the Taylor series expansion one degree higher, with

$$\hat{f}(1) = \hat{f}(0) + \frac{d\hat{f}}{d\gamma} \Big|_0 + \frac{1}{2} \frac{d^2\hat{f}}{d\gamma^2} \Big|_{\tilde{\gamma}} \quad (13)$$

for some $\tilde{\gamma} \in (0, 1)$. Applying the chain rule

$$\frac{d\hat{f}}{d\gamma} \Big|_{\gamma} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_{x(\gamma)} \frac{dx_i}{d\gamma} \Big|_{\gamma} = \nabla f(x(\gamma))^T (x_b - x_a) \quad (14)$$

and then differentiating once more with respect to γ yields

$$\begin{aligned} \frac{d^2\hat{f}}{d\gamma^2} \Big|_{\gamma} &= \sum_{i=1}^n \frac{d}{d\gamma} \left(\frac{\partial f}{\partial x_i} \Big|_{x(\gamma)} \frac{dx_i}{d\gamma} \Big|_{\gamma} \right) \\ &= \sum_{i=1}^n \frac{d}{d\gamma} \left(\frac{\partial f}{\partial x_i} \Big|_{x(\gamma)} \right) \frac{dx_i}{d\gamma} \Big|_{\gamma}, \end{aligned} \quad (15)$$

where we have ignored the terms corresponding to $d^2x_i/d\gamma^2$ as all such terms are 0. Applying the chain rule again yields

$$\begin{aligned} \frac{d^2\hat{f}}{d\gamma^2} \Big|_{\gamma} &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\gamma)} \frac{dx_j}{d\gamma} \Big|_{\gamma} \frac{dx_i}{d\gamma} \Big|_{\gamma} \\ &= \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}). \end{aligned} \quad (16)$$

Substituting the results of (14) and (16) into (13), noting that $\hat{f}(0) = f(x_a)$ and $\hat{f}(1) = f(x_b)$, and rearranging then leads to

$$\begin{aligned} f(x_b) - f(x_a) &= \nabla f(x_a)^T (x_b - x_a) + \\ &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}). \end{aligned} \quad (17)$$

The bounds on the quadratic term are derived in a manner analogous to what was done in the linear case:

$$\begin{aligned} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) \geq 0 &\Leftrightarrow \\ \underline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) &\leq \\ \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) &\leq \\ \overline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}), & \end{aligned} \quad (18)$$

$$\begin{aligned} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) \leq 0 &\Leftrightarrow \\ \overline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) &\leq \\ \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) &\leq \\ \underline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}), & \end{aligned}$$

and, taking both cases into account, we obtain

$$\begin{aligned} \min \left[\frac{\underline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})}{\overline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})} \right] &\leq \\ \frac{\partial^2 f}{\partial x_j \partial x_i} \Big|_{x(\tilde{\gamma})} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) &\leq \\ \max \left[\frac{\underline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})}{\overline{\mathbf{M}}_{ij}(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})} \right], & \end{aligned} \quad (19)$$

which may be substituted into (17) to yield (5) and (6).

4 Other Versions

The bounds (3)-(6) allow a good degree of flexibility by considering both lower and upper bounds on the different partial derivatives. Such flexibility may be useful in certain engineering contexts, where *a priori* knowledge about the system in consideration may be used coherently with the lower and upper bounds on the derivatives [3]. However, there are also contexts where these bounds may be needed

for purely conceptual reasons and where simpler versions are desired. For example, one might want to suppose [4]:

$$\begin{aligned}\kappa_i &= \bar{\kappa}_i = -\underline{\kappa}_i, \\ M_{ij} &= \bar{M}_{ij} = -\underline{M}_{ij},\end{aligned}\quad (20)$$

which, if we follow the same steps as before, yields

$$f(x_b) - f(x_a) \geq -\sum_{i=1}^n \kappa_i |x_{b,i} - x_{a,i}|, \quad (21)$$

$$f(x_b) - f(x_a) \leq \sum_{i=1}^n \kappa_i |x_{b,i} - x_{a,i}|, \quad (22)$$

$$\begin{aligned}f(x_b) - f(x_a) &\geq \nabla f(x_a)^T (x_b - x_a) - \\ &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n M_{ij} |(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})|,\end{aligned}\quad (23)$$

$$\begin{aligned}f(x_b) - f(x_a) &\leq \nabla f(x_a)^T (x_b - x_a) + \\ &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n M_{ij} |(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})|.\end{aligned}\quad (24)$$

One may take this one step further and define the bounds with respect to some standard norms. Defining

$$\kappa = \max_{i=1,\dots,n} \kappa_i, \quad (25)$$

the bounds (21) and (22) become

$$f(x_b) - f(x_a) \geq -\kappa \|x_b - x_a\|_1, \quad (26)$$

$$f(x_b) - f(x_a) \leq \kappa \|x_b - x_a\|_1. \quad (27)$$

For Bounds (23) and (24), we may consider the following derivation:

$$\begin{aligned}&\sum_{i=1}^n \sum_{j=1}^n M_{ij} |(x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j})| \\ &\leq \sum_{i=1}^n \sum_{j=1}^n M_{ij} |x_{b,i} - x_{a,i}| |x_{b,j} - x_{a,j}| \\ &\leq \sum_{i=1}^n \sum_{j=1}^n M_{ij} (x_{b,i} - x_{a,i})^2,\end{aligned}\quad (28)$$

which, with

$$M = \max_{i=1,\dots,n} \sum_{j=1}^n M_{ij}, \quad (29)$$

allows for (23) and (24) to be simplified to:

$$f(x_b) - f(x_a) \geq \nabla f(x_a)^T (x_b - x_a) - \frac{1}{2} M \|x_b - x_a\|_2^2, \quad (30)$$

$$f(x_b) - f(x_a) \leq \nabla f(x_a)^T (x_b - x_a) + \frac{1}{2} M \|x_b - x_a\|_2^2. \quad (31)$$

It may also be shown that the bounds (3), (4), (5), (6), (21), (22), (23), (24), (26), (27), (30), and (31) all hold with *strict* inequality whenever $x_a \neq x_b$. This follows from (11) and (18).

We also refer the reader to [2] for more alternatives.

5 Local Bounds

As derived, the presented bounds are valid for any arbitrary pair $x_a, x_b \in \mathcal{X}$, which follows from the validity of the Lipschitz constants over all of \mathcal{X} . In certain applications, this globality may, however, add unnecessary conservatism and thus motivate local relaxations [3]. Noting that the derivations of the bounds only require them to be valid on the line between x_a and x_b , let us define the local Lipschitz constants with respect to these two points in particular as

$$\underline{\kappa}_i^{a,b} < \left. \frac{\partial f}{\partial x_i} \right|_x < \bar{\kappa}_i^{a,b}, \quad i = 1, \dots, n, \quad \forall x \in \mathcal{X}_{a,b}, \quad (32)$$

$$\underline{M}_{ij}^{a,b} < \left. \frac{\partial^2 f}{\partial x_j \partial x_i} \right|_x < \bar{M}_{ij}^{a,b}, \quad i, j = 1, \dots, n, \quad \forall x \in \mathcal{X}_{a,b}, \quad (33)$$

with

$$\mathcal{X}_{a,b} = \{x_a + \gamma(x_b - x_a) : \gamma \in [0, 1]\}. \quad (34)$$

This then yields the corresponding local versions of (3)-(6):

$$f(x_b) - f(x_a) \geq \sum_{i=1}^n \min \left[\frac{\underline{\kappa}_i^{a,b}}{\bar{\kappa}_i^{a,b}} (x_{b,i} - x_{a,i}) \right], \quad (35)$$

$$f(x_b) - f(x_a) \leq \sum_{i=1}^n \max \left[\frac{\underline{\kappa}_i^{a,b}}{\bar{\kappa}_i^{a,b}} (x_{b,i} - x_{a,i}) \right], \quad (36)$$

$$\begin{aligned}f(x_b) - f(x_a) &\geq \nabla f(x_a)^T (x_b - x_a) + \\ &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \min \left[\frac{\underline{M}_{ij}^{a,b}}{\bar{M}_{ij}^{a,b}} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) \right],\end{aligned}\quad (37)$$

$$\begin{aligned}f(x_b) - f(x_a) &\leq \nabla f(x_a)^T (x_b - x_a) + \\ &\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \max \left[\frac{\underline{M}_{ij}^{a,b}}{\bar{M}_{ij}^{a,b}} (x_{b,i} - x_{a,i})(x_{b,j} - x_{a,j}) \right].\end{aligned}\quad (38)$$

References

1. G. A. Korn and T. M. Korn (2000). *Mathematical Handbook for Scientists and Engineers*. Dover Publications.
2. C. Cartis, J. M. Fowkes, and N. I. M. Gould (2013). *Branching and bounding improvements for global optimization algorithms with Lipschitz continuity properties*. Technical Report, University of Edinburgh.
3. G. A. Bunin, G. François, and D. Bonvin (2014). *Implementation techniques for the SCFO experimental optimization framework*. arXiv [math.OC].
4. G. A. Bunin, G. François, and D. Bonvin (2014). *Feasible-side global convergence in experimental optimization*. Submitted to: SIAM J. Optim.