Journal of Global Optimization (2005)32: 401–407
DOI 10.1007/s10898-004-3134-4
© Springer 2005

On the Liu–Floudas Convexification of Smooth Programs*

SANJO ZLOBEC
Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6 (e-mail: zlobec@math.mcgill.ca)

(Received 23 March 2004; accepted in revised form 1 September 2004)

Abstract. It is well known that a twice continuously differentiable function can be convexified by a simple quadratic term. Here we show that the convexification is possible also for every Lipschitz continuously differentiable function. This implies that the Liu–Floudas convexification works for, loosely speaking, almost every smooth program occurring in practice.

AMS Subject Classifications. 90C30, 90C31

Key words: Global optimum, convexification, Lipschitz continuously differentiable function

1. Introduction

Using the fact that every twice continuously differentiable function can be convexified, Liu and Floudas have shown in [3] that every nonlinear program with these functions can be transformed into an equivalent partly linear-convex program. One can study such programs by, e.g., parametric programming [6, 8]. In this paper we show that the convexification of functions and the transformation to partly linear-convex programs is possible also for Lipschitz continuously differentiable functions.

2. Convexification of Functions

An arbitrary twice continuously differentiable function f: $R^n \to R$ is made convex after adding to it a quadratic of the form $c(x) = -\alpha x^{\mathrm{T}} x$ where α is a sufficiently small number. We will now extend this claim to Lipschitz continuously differentiable functions. Assume that f is a differentiable function defined in an open convex region C of the n-dimensional Euclidean space R^n. Denote by $\nabla f(x)$ its derivative at x, represented as an n-tuple row vector (gradient); hence $\nabla^{\mathrm{T}} f(x)$ is a column, i.e., a vector in R^n. Using the norm $||x|| = (x^{\mathrm{T}} x)^{1/2}$, we say that f is Lipschitz continuously differentiable in C if

*Research supported by an NSERC of Canada grant.

$$||\nabla^T f(x) - \nabla^T f(y)|| \leqslant \mathrm{L}\|x - y\| \tag{2.1}$$

for all $x,y \in C$ and some constant $\mathrm{L} \geqslant 0$ We call L a Lipschitz constant. This constant is not unique because if L satisfies (2.1), so does every $L \geqslant \mathrm{L}$.

REMARK. Lipschitz continuous differentiability is a stronger notion than continuous differentiability. There exist functions that are continuously, even twice continuously, differentiable but not Lipschitz continuously differentiable, e.g., the scalar function $f(t) = 1/t$ on $0 < t < 1$. On the other hand, there are functions that are Lipschitz continuously differentiable but not twice continuously differentiable, e.g., $f(t) = t^2 \cdot \mathrm{sgn}(t)$ on the interval $-1 < t < 1$. So here we study a different class of functions from those that are twice continuously differentiable. Lipschitz continuously differentiable functions are extensively used in numerical optimization, e.g., [4]. We denote the Euclidean inner product by $(u,v) = u^T v$.

THEOREM 2.1. *If f: $R^n \to R$ is Lipschitz continuously differentiable on a convex set C with some Lipschitz constant L, then $\varphi(\mathrm{x}) = f(x) - 1/2\ \alpha \mathrm{x}^T x$ is a convex function on C for every $\alpha \leqslant -\mathrm{L}$.*

Proof. The proof is essentially different from the twice continuously differentiable case. Take x and y in C, $x \neq y$. Then

$$|(\nabla^T f(x) - \nabla^T f(y), x - y)| \leqslant \|\nabla^T f(x) - \nabla^T f(y)\| \cdot \|x - y\|,$$

by the Cauchy–Schwarz inequality

$$\leqslant \mathrm{L}\|x - y\|^2, \text{ by (2.1).}$$

Hence

$$|(\nabla^T f(x) - \nabla^T f(y), x - y)/\ \|x - y\|^2| \leqslant \mathrm{L}.$$

Using the absolute value property, this implies

$$-L \leqslant (\nabla^T f(x) - \nabla^T f(y), x - y)/\|x - y\|^2 \leqslant \mathrm{L}. \tag{2.2}$$

Hence, for every $\alpha \leqslant -\mathrm{L}$, we have

$$\alpha \leqslant (\nabla^T f(x) - \nabla^T f(y), x - y)/\|x - y\|^2$$

and then

$$\alpha||x - y||^2 \leqslant (\nabla^T f(x) - \nabla^T f(y), x - y).$$

This is the same as

$$(\nabla^T f(x) - \nabla^T f(y) - \alpha(x - y), x - y) \geqslant 0. \tag{2.3}$$

Using the definition of φ, we know that its derivative is $\nabla\varphi\ (x) = \nabla f(x) - \alpha x^T$. Therefore (2.3) is

$$(\nabla^T \varphi(x) - \nabla^T \varphi(y), x - y) \geqslant 0. \tag{2.4}$$

The inequality holds also if $x = y \in C$, so it holds for every x and y in C. This means that φ is convex on C. □

EXAMPLE 2.2. Consider f: $R \to R$ defined by $f(t) = t^3$ with $C = [-2, 2]$. Every Lipschitz parameter satisfies L∃12. Hence $\varphi(t) = t^3 - 1/2\ \alpha t^2$ is convex on C for every $\alpha = -\mathrm{L} \leqslant -12$.

The proof of Theorem 2.1 uses the left-hand side inequality in (2.2). If its right-hand side is used for any $\alpha \geqslant \mathrm{L}$ then one obtains (2.4) with a reverse inequality. Hence we have the following result.

THEOREM 2.3. *If f: $R^n \to R$ is a Lipschitz continuously differentiable on a convex set C with a Lipschitz constant* L, *then $\varphi(x) = f(x) - 1/2\ \alpha x^T x$ is a concave function on C for every $\alpha \geqslant \mathrm{L}$.*

Theorem 2.1 can be used to estimate a Lipschitz constant without using derivatives.

THEOREM 2.4. *If f: $R^n \to R$ is Lipschitz continuously differentiable on a convex set C then every Lipschitz constant* L *satisfies*

$$-\mathrm{L} \leqslant \{2/[\lambda(1-\lambda)\|x-y\|^2]\}\{\lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y)\} \leqslant \mathrm{L}$$

for every $x \in C$ and $y \in C$, $y \neq x$, and for every $0 < \lambda < 1$.

Proof. Let us specify $\alpha = -\mathrm{L}$ in $\varphi(x)$. Then $\varphi(x) = f(x) + 1/2\ \mathrm{L}\ x^T x$ is convex on C, i.e.,

$$\varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y)$$

for every x and y in C and for every $0 \leqslant \lambda \leqslant 1$. This is, after substitution,

$$f(\lambda x + (1-\lambda)y) + 1/2\ \mathrm{L}\ \|\lambda x + (1-\lambda)y\|^2 \leqslant \lambda\{f(x) + 1/2\ \mathrm{L}\ \|x\|^2\} + (1-\lambda)\{f(y) + 1/2\mathrm{L}\|y\|^2\}.$$

Hence

$$\begin{aligned} f(\lambda x + (1-\lambda)y) - \lambda f(x) - (1-\lambda)f(y) &\leqslant -1/2\ \mathrm{L}\{\|\lambda x + (1-\lambda)y\|^2 \\ &\quad - \lambda\|x\|^2 - (1-\lambda)\|y\|^2\} \\ &= 1/2\ \mathrm{L}\lambda(1-\lambda)\|x-y\|^2 \end{aligned}$$

after squaring and rearranging. Now a division by $\lambda(1-\lambda)\ ||x-y||^2 > 0$ yields the left hand-side inequality. Similarly one uses Theorem 2.3 to estimate the right hand side. □

COROLLARY 2.5. *If f: $R^n \to$ R is Lipschitz continuously differentiable on a convex set C then every Lipschitz constant L satisfies*

$$-\mathrm{L} \leqslant [8/\|x-y\|^2]\cdot\{[f(x)+f(y)]/2 - f((x+y)/2)\} \leqslant \mathrm{L} \tag{2.5}$$

for every $x \in C$ and $y \in C$, $y \neq x$.

Proof. Specify $\lambda = 1/2$ in Theorem 2.4.

The above result gives an estimate for L of an arbitrary Lipschitz continuously differentiable function f in terms of the difference between the midpoint of values and the value of midpoints of f. The estimate is symmetric in the sense that the variables x and y are interchangeable. We will now use this result to estimate the Lipschitz parameter of a non-convex function.

EXAMPLE 2.6. Consider $f(t) = -t^2/(1+t)$ on $C = [0,\ 1]$. Estimate (2.5) becomes

$$-\mathrm{L} \leqslant 4/[(1+s)(1+t)(2+s+t)] \leqslant \mathrm{L}$$

for every $0 \leqslant s, t \leqslant 1$. The largest value of the function is assumed at $s = t = 0$. Hence we estimate $\mathrm{L} \geqslant 2$. A convexification of f on C is $\varphi(t) = f(t) + t^2 = t^3/(1+t)$.

3. Convexification of Programs

In this section we will use the Liu–Floudas transformation to convexify smooth programs. Let us consider the general non-linear program

$$\text{(NP)} \quad \begin{array}{l} \text{Min } f(x) \\ f^i(x) \leqslant 0, \quad i \in P, \quad x \in C. \end{array}$$

Here all functions are assumed to be twice continuously differentiable or Lipschitz continuously differentiable on an open set containing a non-empty compact convex set C and P is a finite index set. Let us also assume that (NP) has a unique globally optimal solution x^*. We know that each function in (NP) can be convexified. If, say, $f(x)$ is convexified by $-1/2\ \alpha x^{\mathrm{T}}x$, $\alpha \leqslant -\mathrm{L}$, then so can be the function $f(x) + 1/2\ x^{\mathrm{T}}\theta$, where $\theta \in R^n$ is a fixed vector. Let us associate with (NP), for every $\varepsilon \geqslant 0$, the following class of partly linear-convex programs

$$(\mathrm{LF}(x,\theta;\varepsilon)) \quad \begin{array}{l} \mathrm{Min}_{(x,\theta)} \varphi(x,\theta) = f(x) + 1/2\alpha(\theta^T x - x^T x) \\ \varphi^i(x,\theta) = f^i(x) + 1/2\alpha_i(\theta^T x - x^T x) \leqslant 0, \quad i \in P,\ \|x-\theta\| \leqslant \varepsilon,\ x \in C. \end{array}$$

Here α and α_i, $i \in P$ are some fixed sufficiently small numbers for which $\varphi(x,\theta)$ and $\varphi^i(x,\theta)$ are convexifications (in the variable x) of $f(x)$ and $f^i(x)$, respectively, $i \in P$. For every θ and ε, LF($x,\theta;\varepsilon$) is a convex program in x, while for every x and ε, it is a linear program in θ. For the purpose of having linear programs in θ we use l_∞ or l_1 norm rather than the Euclidean norm. Such programs are called partly linear-convex. They can also be considered as simple convex models and many results on their optimality and stability from, e.g., [6, 8] are applicable. With the introduction of $\varepsilon \geqslant 0$ the original Liu-Floudas transformation (introduced in [3] for $\varepsilon = 0$) assumes a parametric form. Its purpose is: (i) to "drive" the optimal solutions ($x^0(\varepsilon)$, $\theta^0(\varepsilon)$) of LF($x,\theta;\varepsilon$) to the optimal solution of LF($x,\theta;0$) and (ii) to allow a bigger feasible set (and more flexibility) in the search for global optima of the programs LF($x,\theta;\varepsilon$). One could introduce $\varepsilon \geqslant 0$ also on the right-hand sides of all implicit constraints without changing the statements given below.

THEOREM 3.1. *Consider (NP) and the class of linear-convex programs* LF*($x,\theta;\varepsilon$)*. Then the following statements hold:

(i) *A point* x^* is the optimal solution of (NP) if, and only if, (x^*, θ^*) is the optimal solution of LF*(x,θ;0) where* $x^* = \theta^*$.
(ii) *Program (NP) has a unique optimal solution if, and only if, program* LF*(x,θ;0)* has a unique *optimal solution.*
(iii) *Point* x^* is the optimal solution of (NP) if, and only if,

$$x^* = \lim_{\varepsilon \to 0} x^0(\varepsilon) \text{ and } \theta^* = \lim_{\varepsilon \to 0} \theta^0(\varepsilon) \text{ with } x^* = \theta^*.$$

Proof. (i) This claim is "obvious", e.g., [3]. (ii) This is a consequence of (i) and the fact that (NP) is assumed to have a unique solution. So only (iii) remains to be proved. We give a slightly different proof from the one given in [8]. Choose an arbitrary sequence of $\varepsilon > 0$, $\varepsilon \to 0$. Since C is a compact set, a sequence of optimal solutions ($x^0(\varepsilon)$, $\theta^0(\varepsilon)$) exists. Hence their limits x^* and θ^* exist. Since the functions are continuous, the feasible set mapping is closed. Therefore (x^*, θ^*) is a feasible point of LF(x,θ;0). But this is also the optimal point of the program LF(x,θ;0) . If not, then there would exist another feasible point, say, (u,v) such that

$$\varphi(u,v) < \varphi(x^*,\theta^*)$$

But this point is feasible also for every program LF($x,\theta;\varepsilon$) with $\varepsilon > 0$. Now, by continuity of the objective function, (3.1) implies

$$\varphi(u,v) < \varphi(x,\theta)$$

for every x and θ in some neighborhood of (x^*, θ^*). This violates global optimality of points $(x^0(\varepsilon), \theta^0(\varepsilon))$. Continuity of the norm implies $x^* = \theta^*$. Hence x^* solves (NP) by statement (i). □

The above theorem shows that the optimal solutions of programs with twice continuously differentiable functions, or programs having Lipschitz continuously differentiable functions, are limit points of optimal solutions of suitable linear programs and also of suitable convex programs. It also shows that the feasible set of $\mathrm{LF}(x,\theta;\varepsilon)$ is lower semi-continuous at optimal $\theta = \theta^*$ and $\varepsilon = 0$ relative to feasible perturbations; e.g., [6, 8].The practical success of the convexification depends on the ability to calculate global optima of partly linear-convex programs; e.g., [1, 2, 4–6]. The convexification technique has been used by the author and his students to solve diverse problems: from finding roots of a polynomial and solutions of systems of non-linear equations (see also [7]), and finding optimal steering angles in Zermelo's navigation problems, to solving profit maximization problems in oil industry. In particular, the Oilco NLP problem from [5, pp. 663–666] has been studied in detail in [1]. The problem has 18 variables and 29 constraints, some of these non-convex. Using convexification and LINGO, many globally optimal solutions are found in [1] that are essentially different from the one given in [5]. This has resulted in a lexicographic formulation of the problem.

4. Conclusion

We have shown that every Lipschitz continuously differentiable function can be represented as the difference of a convex and a quadratic concave function. Using this fact every mathematical program with Lipschitz continuously differentiable functions can be reduced, by the Liu–Floudas convexification, to a partly linear-convex program.

5. Acknowledgement

The author is indebted to the referees for their careful reading of the manuscript and for many constructive comments.

References

1. Davidescu, D.M. M.Sc. Thesis, McGill University, Department of Mathematics and Statistics (in progress).
2. Floudas, C.A., (2000) *Deterministic Global Optimization*, Kluwer Academic Publishers, Dordrecht.
3. Liu, W.B. and Floudas C.A., (1993) A remark on the GOP algorithm for global optimization, *Journal of Global Optimization* 3, 519–521.

4. Nocedal, J. and Wright, S.J. (1999), *Numerical Optimization*, Springer Series in Operations Research, New York.
5. Winston, W.L. and Venkataramanan (2003), *Introduction to Mathematical Programming*, Fourth edition, Brooks/Cole, Pacific Grove, CA.
6. Zlobec, S. (2001), *Stable Parametric Programming*, Kluwer Academic Publishers, Dordrecht.
7. Zlobec, S. (2003), Estimating convexifiers in continuous optimization, M*athematical Communications* 8, 129–137.
8. Zlobec, S. (2004), Saddle-point optimality: A look beyond convexity, *Journal of Global Optimization* 29, 97–112.