J Glob Optim (2010) 48:311–321
DOI 10.1007/s10898-009-9495-y

Refined optimality conditions for differences of convex functions

Tuomo Valkonen

Received: 3 July 2008 / Accepted: 11 November 2009 / Published online: 21 November 2009

Abstract We provide a necessary and sufficient condition for strict local minimisers of differences of convex (DC) functions, as well as related results pertaining to characterisation of (non-strict) local minimisers, and uniqueness of global minimisers.

Keywords Diff-convex · Optimality

Mathematics Subject Classification (2000) 90C26 · 90C46

1 Introduction

Let g and h be convex functions on $\mathbb{R}^m$, both proper and closed with h finite-valued. Define the difference of these functions as $f := g - h$. As shown in particular by Hiriart-Urruty et al. [4], a necessary and sufficient condition for $\hat{y} \in \mathbb{R}^m$ to be a global minimiser of f is that

$$\partial_\varepsilon h(\hat{y}) \subset \partial_\varepsilon g(\hat{y}) \quad \text{for all } \varepsilon \geq 0. \tag{1}$$

For local optimality, Dür [2] has showed the sufficiency of the existence of $\overline{\varepsilon} > 0$ such that (1) holds for all $\varepsilon \in [0, \overline{\varepsilon})$. This condition is, however, not necessary for local optimality. In this paper, we show that necessity follows under the additional constraint of the set of "mutual linearity" of g and h around $\hat{y}$, being the singleton $\{\hat{y}\}$.

We also show that the condition on mutual linearity along with strict inclusion in (1) for $\varepsilon \in (0, \overline{\varepsilon})$—but importantly not necessarily for $\varepsilon = 0$—is both necessary *and* sufficient for strict local optimality. Also, when f is level-bounded, it turns out that strict inclusion for all $\varepsilon > 0$ and a singleton mutual linearity set is both necessary and sufficient for the uniqueness of $\hat{y}$ as a global minimiser.

T. Valkonen (✉)
Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland
e-mail: tujumava@jyu.fi

The rest of this paper is organised as follows. In Sect. 2 we introduce notation and concepts employed in the later analysis. Then in Sect. 3 we study the aforementioned characterisations of strict local optimality. In Sect. 4 we adapt these results to get refined characterisations of non-strict local optimality, while in Sect. 5 we consider the uniqueness of global minimisers. We conclude the paper in Sect. 6 with a brief discussion of related sensitivity and level-boundedness results. The results have been extracted and improved from the author's Ph.D thesis [9].

2 Definitions

We use the following notations. The support function of a non-empty convex set $A \subset \mathbb{R}^m$ is denoted by $\sigma(x; A) := \sup\{\langle z, x\rangle \mid z \in A\}$, and the normal cone at $x \in A$ is defined as $N_A(x) := \{z \in \mathbb{R}^m \mid \langle z, x' - x\rangle \le 0 \text{ for all } x' \in A\}$. The star-difference is defined for two sets A and B as

$$A \stackrel{*}{-} B := \left\{z \in \mathbb{R}^m \mid z + B \subset A\right\}.$$

Note that this set is closed and convex if A is. The closure, boundary, interior, and relative interior of a set A are denoted, respectively, by $\operatorname{cl} A$, $\operatorname{bd} A$, $\operatorname{int} A$ and $\operatorname{ri} A$.

The effective domain of a function $f : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is denoted $\operatorname{dom} f := \{y \in \mathbb{R}^m \mid f(y) < \infty\}$. This set is non-empty for our functions of interest. Similarly, the c-level set of f is defined as $\operatorname{lev}_c f := \{y \in \mathbb{R}^m \mid f(y) \le c\}$, and f is said to be level-bounded if $\operatorname{lev}_c f$ is a bounded set for every $c \in \mathbb{R}$.

Throughout the rest of this paper, unless otherwise explicitly mentioned, we utilise the following assumption.

Assumption 1 The functions denoted by g and h are convex, proper, and closed with $g : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ and $h : \mathbb{R}^m \to \mathbb{R}$. The symbol f denotes a function $\mathbb{R}^m \to (-\infty, \infty]$ that can be decomposed as $f = g - h$. Such a function is called a difference of convex functions, or diff-convex.

The ε-subdifferential of g at $y \in \mathbb{R}^m$ is defined as the set $\partial_\varepsilon g(y)$ of $z \in \mathbb{R}^m$ that satisfy

$$g(y') - g(y) \ge \langle z, y' - y\rangle - \varepsilon \quad \text{for all } y' \in \mathbb{R}^m.$$

Our general reference for many of the basic properties of ε-subdifferentials listed below is provided by [5].

Defining the convex graphs

$$G_g(y) := \{(z, \varepsilon) \mid \varepsilon \ge 0,\ z \in \partial_\varepsilon g(y)\},$$

we have the expression

$$\begin{aligned} g(y + v) - g(y) &= \sup\{\langle v, z\rangle - \varepsilon \mid \varepsilon > 0,\ z \in \partial_\varepsilon g(y)\} \\ &= \sup\{\sigma(v; \partial_\varepsilon g(y)) - \varepsilon \mid \varepsilon > 0\} \\ &= \sigma((v, -1); G_g(y)). \end{aligned} \tag{2}$$

This will be useful throughout the paper.

Let us also recall the definition of the linearisation error,

$$e_g\left(y'; y, z\right) := g\left(y'\right) - g(y) - \langle z, y' - y\rangle, \tag{3}$$

and the subdifferential transportation formula:

$$\text{if } z \in \partial_\eta g(y), \quad \text{then } z \in \partial_\varepsilon g\left(y'\right) \quad \text{for } \varepsilon \geq \eta + e_g\left(y'; y, z\right). \tag{4}$$

We may then define the region of mutual linearity of g and h around y as

$$L(y) := \left\{y' \in \mathbb{R}^m \mid e_g(y'; y, z) = e_h(y'; y, z) = 0 \quad \text{for some } z \in \partial g(y) \cap \partial h(y)\right\}.$$

Finally, we denote

$$C_\varepsilon(y) := \partial_\varepsilon g(y) \overset{*}{-} \partial_\varepsilon h(y).$$

The condition $0 \in C_\varepsilon(\hat{y})$ (resp. $0 \in \operatorname{int} C_\varepsilon(\hat{y})$) is then the same as $\partial_\varepsilon h(\hat{y}) \subset \partial_\varepsilon g(\hat{y})$ (resp. $\partial_\varepsilon h(\hat{y}) \subset \operatorname{int} \partial_\varepsilon g(\hat{y})$), since $\partial_\varepsilon h(\hat{y})$ is compact by our standing assumption on h being finite-valued. Thus $0 \in \bigcap_{\varepsilon>0} C_\varepsilon(y)$ is equivalent to the necessary and sufficient global optimality condition $\partial_\varepsilon h(y) \subset \partial_\varepsilon g(y)$ for all $\varepsilon > 0$. According to [6], $\partial f(y) = \bigcap_{\varepsilon>0} C_\varepsilon(y)$, providing the connection to yet another characterisation of optimality.

3 Strict local optimality

We may now state the main result. Recall that throughout the paper, the symbols f, g, and h denote functions that satisfy Assumption 1.

Theorem 1 *The point $\hat{y} \in \mathbb{R}^m$ is a strict local minimiser of $f := g - h$ if and only if $L(\hat{y}) = \{\hat{y}\}$ and the following subdifferential inclusion is satisfied:*

$$\textit{there exists } \bar{\varepsilon} > 0, \textit{ such that } 0 \in \operatorname{int} C_\varepsilon(\hat{y}) \ \textit{for each } \varepsilon \in (0, \bar{\varepsilon}). \tag{SDI}$$

(Note that we do not explicitly require $\hat{y} \in \operatorname{dom} g$ for necessity, because the condition $L(\hat{y}) = \{\hat{y}\}$ already forces this.)

We begin the proof with a few technical lemmas. Lemma 1 followed by Lemma 2 forms the core of the necessity proof, while Lemma 3 provides a lower estimate of $f(y) - f(\hat{y})$ that quickly yields the sufficiency proof.

Lemma 1 *Suppose $(z_h, \varepsilon_h) \in G_h(\hat{y}) \setminus \operatorname{int} G_g(\hat{y})$. Then there exists $(z_g, \varepsilon_g) \in \operatorname{bd} G_g(\hat{y})$, $\alpha \geq 0$, and $(v, \delta) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ with $\delta \in \{0, -1\}$ and $\|v\| \geq 1 + \delta$, such that $(z_h, \varepsilon_h) = (z_g, \varepsilon_g) + \alpha(v, \delta)$. We additionally have $\|v\| > 0$ if $\varepsilon_h > 0$.*

Proof Since $\hat{y} \in \operatorname{dom} g$, the set $G_g(\hat{y})$ is non-empty. It is also closed and convex, and the function $w : (z, \varepsilon) \mapsto \|(z_h, \varepsilon_h) - (z, \varepsilon)\|^2/2$ is convex, continuous, and level-bounded. Therefore, w has a minimiser (z_g, ε_g) over $G_g(\hat{y})$ that satisfies the optimality condition $-\nabla w(z_g, \varepsilon_g) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$, or

$$(z_h - z_g, \varepsilon_h - \varepsilon_g) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g). \tag{5}$$

Clearly also $(z_g, \varepsilon_g) \in \operatorname{bd} G_g(\hat{y})$ due to $(z_h, \varepsilon_h) \notin \operatorname{int} G_g(\hat{y})$.

The inclusion $(z_g, \varepsilon_g) \in G_g(\hat{y})$ means $z_g \in \partial_{\varepsilon_g} g(\hat{y})$. The definition of the ε-subdifferential implies that, in fact, $z_g \in \partial_\varepsilon g(\hat{y})$ for all $\varepsilon \geq \varepsilon_g$. Thus, by choosing $z = z_g$ and $\varepsilon > \varepsilon_g$, we find that the inequality $\langle (v, \delta), (z - z_g, \varepsilon - \varepsilon_g)\rangle \leq 0$ cannot hold for all $(z, \varepsilon) \in G_g(\hat{y})$ unless $\delta \leq 0$. Hence we obtain that

$$(v, \delta) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g) \quad \text{implies} \quad \delta \leq 0. \tag{6}$$

Combining (5) and (6), we find $\varepsilon_h \leq \varepsilon_g$. If actually $\varepsilon_h < \varepsilon_g$, we may set $\alpha = \varepsilon_g - \varepsilon_h$, and find $(v, \delta) = ((z_h - z_g)/\alpha, -1) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ by dividing (5) by α. Clearly both conditions $(z_h, \varepsilon_h) = (z_g, \varepsilon_g) + \alpha(v, \delta)$ and $\|v\| \geq 1 + \delta$ are satisfied by this choice. If, on the other hand, $\varepsilon_g = \varepsilon_h$, there are two cases to consider for the choice of (v, δ) and α:

(i) If also $z_g = z_h$, we take $\alpha = 0$. Since (z_g, ε_g) is a boundary point of the convex set $G_g(\hat{y})$, there exists some $(0, 0) \neq (v, \delta) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$. If $\delta = 0$, this implies $v \neq 0$, so we may normalise (v, δ) to satisfy $\|v\| = 1$, thus fulfilling the requirement $\|v\| \geq 1 + \delta$. Otherwise, $\delta < 0$ by (6), so we normalise (v, δ) to have $\delta = -1$, which trivially fulfills $\|v\| \geq 1 + \delta$.
(ii) If $z_g \neq z_h$, we take $\alpha = \|z_h - z_g\|$, $v = (z_h - z_g)/\alpha$, and $\delta = 0$. All the requirements are again clearly fulfilled by this choice.

It remains to show that $\|v\| > 0$ if $\varepsilon_h > 0$. If $\delta = 0$, this trivially follows from the bound $\|v\| \geq 1 + \delta$. If $\delta = -1$, it suffices to show that $(v, -1) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ implies $v \neq 0$. This follows from the definition of $N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ due to $\varepsilon_g \geq \varepsilon_h > 0$ and the inclusion $\partial_\varepsilon g(\hat{y}) \subset \partial_{\varepsilon_g} g(\hat{y})$ for $\varepsilon \in [0, \varepsilon_g]$. □

Lemma 2 *Suppose $\hat{y} \in \operatorname{dom} g$, and (z_h, ε_h), (z_g, ε_g), (v, δ) and α are as in the statement of Lemma 1. Let $y_\lambda := \hat{y} + \lambda v$ for $\lambda \geq 0$. When $\delta = -1$, we then have $f(y_\lambda) \leq f(\hat{y}) + (1 - \lambda)\varepsilon_h - \lambda\alpha$ for $\lambda \in [0, 1]$. In the case $\delta = 0$, we have $f(y_\lambda) \leq f(\hat{y}) + \varepsilon_h - \lambda\alpha$ for all $\lambda \geq 0$.*

Proof In both of the cases $\delta = 0$ and $\delta = -1$, applying $(z_h, \varepsilon_h) = (z_g, \varepsilon_g) + \alpha(v, \delta)$, and $z_h \in \partial_{\varepsilon_h} h(\hat{y})$, we find

$$\begin{aligned} h(y_\lambda) - h(\hat{y}) &\geq \lambda\langle z_h, v\rangle - \varepsilon_h \\ &= \lambda(\langle z_h, v\rangle - \varepsilon_h) - (1 - \lambda)\varepsilon_h \\ &= \lambda\left(\langle z_g, v\rangle - \varepsilon_g + \alpha\left(\|v\|^2 - \delta\right)\right) - (1 - \lambda)\varepsilon_h. \end{aligned}$$

Because $\|v\|^2 - \delta \geq 1$ due to $\|v\| \geq 1 + \delta$ and $\delta \in \{0, -1\}$, we thus obtain

$$h(y_\lambda) - h(\hat{y}) \geq \lambda(\langle z_g, v\rangle - \varepsilon_g) + \lambda\alpha - (1 - \lambda)\varepsilon_h. \tag{7}$$

Consider now the case $\delta = -1$. By the expression (2) and the property $(v, -1) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ from Lemma 1, we find that for $y_1 = \hat{y} + v$, we have

$$g(y_1) - g(\hat{y}) = \sigma((v, -1); G_g(\hat{y})) = \langle z_g, v\rangle - \varepsilon_g.$$

Thus, by the convexity of g and the expression $y_\lambda = \lambda y_1 + (1 - \lambda)\hat{y}$, we find

$$g(y_\lambda) - g(\hat{y}) \leq \lambda(g(y_1) - g(\hat{y})) = \lambda\left(\langle z_g, v\rangle - \varepsilon_g\right) \quad \text{for} \quad \lambda \in [0, 1]. \tag{8}$$

Combining the inequalities (8) and (7), we hence obtain the claimed

$$f(y_\lambda) - f(\hat{y}) \leq (1 - \lambda)\varepsilon_h - \lambda\alpha \quad \text{for} \quad \lambda \in [0, 1].$$

It remains to consider the case $\delta = 0$. Since $(v, 0) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ with $\|v\| > 0$ by Lemma 1, we find that $\langle z_g, v\rangle \geq \langle z, v\rangle$ over all $(z, \varepsilon) \in G_g(\hat{y})$. Consequently, (2) yields

$$\begin{aligned} g(y_\lambda) - g(\hat{y}) &= \sup\{\lambda\langle z, v\rangle - \varepsilon \mid \varepsilon > 0,\ z \in \partial_\varepsilon g(y)\} \\ &\leq \sup\{\lambda\langle z_g, v\rangle - \varepsilon \mid \varepsilon > 0,\ z \in \partial_\varepsilon g(\hat{y})\} \\ &= \lambda\langle z_g, v\rangle. \end{aligned} \tag{9}$$

Since, by Lemma 1, $\varepsilon_g = \varepsilon_h$ when $\delta = 0$, the inequality (7) reduces into

$$h(y_\lambda) - h(\hat{y}) \geq \lambda\langle z_g, v\rangle + \lambda\alpha - \varepsilon_h.$$

Combining this with (9) yields the claim $f(y_\lambda) \leq f(\hat{y}) + \varepsilon_h - \lambda\alpha$ for all $\lambda \geq 0$. □

Lemma 3 *Suppose* $\hat{y} \in \operatorname{dom} g$, $y \in \mathbb{R}^m$, *and* $z \in \partial h(y)$. *Suppose also* $\overline{\varepsilon} > \varepsilon_h := e_h(\hat{y}; y, z)$. *Then*

$$f(y) - f(\hat{y}) \geq \sup\left\{\sigma(y - \hat{y}; C_\varepsilon(\hat{y})) - (\varepsilon - \varepsilon_h) \mid \varepsilon \in [\varepsilon_h, \overline{\varepsilon})\right\}. \tag{10}$$

Additionally, provided that $y \neq \hat{y}$, *that* $L(\hat{y}) = \{\hat{y}\}$, *and that (SDI) holds for* $\hat{y}$ *and* $\overline{\varepsilon}$, *then* $f(y) > f(\hat{y})$.

Proof By (2), we have

$$h(y) - h(\hat{y}) \geq \sigma(y - \hat{y}; \partial_{\varepsilon_h} h(\hat{y})) - \varepsilon_h. \tag{11}$$

By the definition (3), $\varepsilon_h = e_h(\hat{y}; y, z)$ reads as

$$h(y) - h(\hat{y}) = \langle y - \hat{y}, z\rangle - \varepsilon_h. \tag{12}$$

But $z \in \partial h(y)$ and the subdifferential transportation formula (4) imply that $z \in \partial_{\varepsilon_h} h(\hat{y})$. Therefore, by (12), we have equality in (11), i.e.,

$$h(y) - h(\hat{y}) = \sigma(y - \hat{y}; \partial_{\varepsilon_h} h(\hat{y})) - \varepsilon_h. \tag{13}$$

Employing the properties of the support function and the definition of $C_\varepsilon(\hat{y})$ as a set satisfying $\partial_\varepsilon h(\hat{y}) + C_\varepsilon(\hat{y}) \subset \partial_\varepsilon g(\hat{y})$, we furthermore get

$$\begin{aligned} g(y) - g(\hat{y}) &= \sup\left\{\sigma(y - \hat{y}; \partial_\varepsilon g(\hat{y})) - \varepsilon \mid \varepsilon > 0\right\} \\ &\geq \sup\left\{\sigma(y - \hat{y}; \partial_\varepsilon g(\hat{y})) - \varepsilon \mid \varepsilon \in [\varepsilon_h, \overline{\varepsilon})\right\} \\ &\geq \sup\left\{\sigma(y - \hat{y}; C_\varepsilon(\hat{y})) + \sigma(y - \hat{y}; \partial_\varepsilon h(\hat{y})) - \varepsilon \mid \varepsilon \in [\varepsilon_h, \overline{\varepsilon})\right\} \\ &\geq \sup\left\{\sigma(y - \hat{y}; C_\varepsilon(\hat{y})) + \sigma(y - \hat{y}; \partial_{\varepsilon_h} h(\hat{y})) - \varepsilon \mid \varepsilon \in [\varepsilon_h, \overline{\varepsilon})\right\}. \end{aligned} \tag{14}$$

The compactness of $\partial_\varepsilon h(\hat{y})$ ensures that the final estimate in (14) is well-defined as $-\infty$ when $C_\varepsilon(\hat{y}) = \emptyset$. Thus, subtracting (13) from (14), we get the claimed (10).

It remains to show that $y \neq \hat{y}$ together with $L(\hat{y}) = \{\hat{y}\}$ and (SDI) imply $f(y) > f(\hat{y})$. Setting $\varepsilon = \varepsilon_h$ in (10) gives

$$f(y) - f(\hat{y}) \geq \sigma(y - \hat{y}; C_{\varepsilon_h}(\hat{y}))).$$

Thus, the claim is established if $0 \in \operatorname{int} C_{\varepsilon_h}(\hat{y})$. By (SDI), this is the case when $\varepsilon_h \in (0, \overline{\varepsilon})$. We have $\varepsilon_h < \overline{\varepsilon}$ by assumption, so only the case $\varepsilon_h = 0$ remains to be dealt with. Then $e_h(y; \hat{y}, z) = -e_h(\hat{y}; y, z) = -\varepsilon_h = 0$, wherefore $z \in \partial h(\hat{y})$ by the subdifferential transportation formula. Since (SDI) implies $0 \in C_0(\hat{y})$, we therefore have $z \in \partial g(\hat{y})$ as well. If $y \neq \hat{y}$, the condition $L(\hat{y}) = \{\hat{y}\}$ thus forces $e_g(y; \hat{y}, z) > 0$. But then we have $e_g(y; \hat{y}, z) - e_h(y; \hat{y}, z) > 0$, which is just $f(y) > f(\hat{y})$ written in another way. □

We now have the necessary ingredients to prove Theorem 1.

Proof (Theorem 1) Necessity. We may assume that $\hat{y} \in \operatorname{dom} g$, since otherwise $\hat{y}$ cannot minimise f *strictly*, not even locally. If we had $\hat{y} \neq y \in L(\hat{y})$, then by the definition of the mutual linearity set $L(\hat{y})$, the segment $[\hat{y}, y] \subset L(\hat{y})$. Since f is a constant on the set $L(\hat{y})$, and $\hat{y}$ was assumed a strict local minimiser, we must therefore have $L(\hat{y}) = \{\hat{y}\}$.

To prove the necessity of (SDI), we assume the contrary, i.e., that there exists a sequence $\varepsilon_{h,k} \searrow 0$ ($k = 0, 1, 2, \ldots$), such that $0 \notin \operatorname{int} C_{\varepsilon_{h,k}}(\hat{y})$. By the compactness of $\partial_{\varepsilon_{h,k}} h(\hat{y})$, this may be rewritten $\partial_{\varepsilon_{h,k}} h(\hat{y}) \not\subset \operatorname{int} \partial_{\varepsilon_{h,k}} g(\hat{y})$. We therefore have the existence of some $z_{h,k} \in \partial_{\varepsilon_{h,k}} h(\hat{y}) \setminus \operatorname{int} \partial_{\varepsilon_{h,k}} g(\hat{y})$. This may be restated $(z_{h,k}, \varepsilon_{h,k}) \in G_h(\hat{y}) \setminus \operatorname{int} G_g(\hat{y})$. Consequently Lemma 1 provides $(z_{g,k}, \varepsilon_{g,k}) \in \operatorname{bd} G_g(\hat{y})$ and $(v_k, \delta_k) \in N_{G_g(\hat{y})}(z_{g,k}, \varepsilon_{g,k})$ with $\delta_k \in \{0, -1\}$ and $\|v_k\| \geq 1 + \delta_k$, as well as $\alpha_k \geq 0$ such that $(z_{h,k}, \varepsilon_{h,k}) = (z_{g,k}, \varepsilon_{g,k}) + \alpha_k(v_k, \delta_k)$.

First, we consider the case that (for a subsequence) $\|v_k\| \to 0$. We may then assume that $\delta_k = -1$, as this must be the case for large enough k due to the property $\|v_k\| \geq 1 + \delta_k$. An application of Lemma 2 with $\lambda = 1$ then shows that $f(\hat{y} + v_k) \leq f(\hat{y})$. But $\hat{y} + v_k \to \hat{y}$, which provides a contradiction to strict local minimality at $\hat{y}$.

We may therefore assume that $\|v_k\| \geq \theta$ for some $\theta > 0$. Since $\varepsilon_{h,k} \searrow 0$, and h has bounded ε-subdifferentials due to $\operatorname{dom} h = \mathbb{R}^m$, the sequence $\{z_{h,k}\}_{k=0}^{\infty}$ is bounded. By possibly switching to a subsequence, it may therefore be assumed convergent to some $z_h \in \partial h(\hat{y})$. Since $(z_{g,k}, \varepsilon_{g,k})$ is, by construction, the closest point in the non-empty set $G_g(\hat{y})$ to $(z_{h,k}, \varepsilon_{h,k})$, the sequence $\{(z_{g,k}, \varepsilon_{g,k})\}_{k=0}^{\infty}$ is also bounded, and may likewise be assumed convergent to some $(z_g, \varepsilon_g) \in \operatorname{bd} G_g(\hat{y})$. Observe that these considerations force $\{\alpha_k(v_k, \delta_k)\}_{k=0}^{\infty} = \{(z_{h,k}, \varepsilon_{h,k}) - (z_{g,k}, \varepsilon_{g,k})\}_{k=0}^{\infty}$ to be likewise convergent.

Now, if $\{v_k\}_{k=0}^{\infty}$ contains a bounded subsequence, we take α and (v, δ) as limits of some common subsequence of $\{\alpha_k\}_{k=0}^{\infty}$ and $\{(v_k, \delta_k)\}_{k=0}^{\infty}$. Clearly then $\delta \in \{0, -1\}$ and $\alpha \geq 0$, while $\|v\| \geq \max\{1 + \delta, \theta\} > 0$.

Otherwise, if $\{v_k\}_{k=0}^{\infty}$ is unbounded, we take (v, δ) as a limit of a subsequence of the renormalised sequence $\{(v_k, \delta_k)/\|v_k\|\}_{k=0}^{\infty}$, and α as a limit of a common subsequence of $\{\alpha_k \|v_k\|\}_{k=0}^{\infty} = \left\{\left\|z_{g,k} - z_{h,k}\right\|\right\}_{k=0}^{\infty}$. Clearly then $\delta = 0$, $\|v\| = 1$ and $\alpha \geq 0$.

In both of the above cases, recalling that $N_{G_g(\hat{y})}$ is outer-semicontinuous, we observe that $(v, \delta) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$. From the convergence of $\alpha_k(v_k, \delta_k) = (z_{h,k}, \varepsilon_{h,k}) - (z_{g,k}, \varepsilon_{g,k})$, we also find that $\alpha(v, \delta) = (z_h, 0) - (z_g, \varepsilon_g)$. With $\varepsilon_h := 0$, the assumptions of Lemma 2 are therefore satisfied by (z_h, ε_h), (z_g, ε_g), (v, δ), and α. Consequently, in both of the cases $\delta = 0$ and $\delta = -1$, we have $f(y_\lambda) \leq f(\hat{y}) - \lambda\alpha \leq f(\hat{y})$ for $\lambda \in [0, 1]$ and $y_\lambda := \hat{y} + \lambda v$. Since $\|v\| > 0$, letting $\lambda \searrow 0$ provides the contradiction to f having strict local minimum at $\hat{y}$.

Sufficiency. Suppose $y_k \to \hat{y}$ ($k = 0, 1, 2, \ldots; y_k \neq \hat{y}$). We may then choose some $z_k \in \partial h(y_k)$ since $\operatorname{dom} h = \mathbb{R}^m$. For sufficiently large k, we have $e_h(\hat{y}; y_k, z_k) < \overline{\varepsilon}$, since $\{z_k\}_{k=0}^{\infty}$ is bounded, h is continuous, and $y_k \to \hat{y}$. Therefore, for large k, the second claim of Lemma 3 applies with $y = y_k$ and $z = z_k$, yielding $f(y_k) > f(\hat{y})$. This shows that $\hat{y}$ is a strict local minimiser of f, thus completing the proof. □

Remark 1 Note that (SDI) ensures $0 \in C_0(\hat{y})$ by closedness of the subdifferentials, but we *do not require* $0 \in \operatorname{int} C_0(\hat{y})$, which in itself if sufficient for strict local optimality, as shown by, e.g., [3], or [7] in a more general setting.

Corollary 1 *The diff-convex function* $f : \mathbb{R}^m \to (-\infty, \infty]$ *has a strict local minimum at* $\hat{y} \in \mathbb{R}^m$ *if and only if* (SDI) *holds for* every *decomposition* $f = g - h$.

Proof The necessity of (SDI) holding for every decomposition at a strict local minimiser is immediate from Theorem 1. Sufficiency follows from Theorem 1 by choosing a decomposition $f = g - h$ with $L(\hat{y}) = \{\hat{y}\}$. This can be done by taking an arbitrary decomposition and adding the function $y \mapsto \theta \left\|y - \hat{y}\right\|^2$ for arbitrary $\theta > 0$ to both g and h, forming the functions g^θ and h^θ. Then $\partial g^\theta(\hat{y}) = \partial g(\hat{y})$, whence for all $z \in \partial g^\theta(\hat{y})$ and $y \neq \hat{y}$,

$$g^\theta(y) - g^\theta(\hat{y}) = g(y) - g(\hat{y}) + \theta \left\|y - \hat{y}\right\| > g(y) - g(\hat{y}) \geq \langle z, y - \hat{y} \rangle.$$

This says that $e_{g^\theta}(y; \hat{y}, z) > 0$, showing that $L(\hat{y}) = \{\hat{y}\}$ for the decomposition $f = g^\theta - h^\theta$. Since (SDI) holds by assumption, Theorem 1 now proves strict local optimality. □

We may also rephrase Corollary 1 in terms of strictly convex functions:

Corollary 2 *The diff-convex function $f : \mathbb{R}^m \to (-\infty, \infty]$ has a strict local minimum at $\hat{y} \in \mathbb{R}^m$ if and only if* (SDI) *holds for* some *decomposition $f = g - h$ such that either g or h is strictly convex.*

Proof The functions g^θ and h^θ constructed in the proof of Corollary 1 are strictly convex, showing necessity. To show sufficiency of (SDI) holding for some decomposition with either g or h strictly convex, it suffices by Theorem 1 to show that $L(\hat{y}) = \{\hat{y}\}$ then. To see this, we note that if g is strictly convex, then the sets $N(y) := \{y' \in \mathbb{R}^m \mid e_g(y'; y, z) = 0,\ z \in \partial g(y)\}$ are the singletons $\{y\}$ for $y \in \mathbb{R}^m$. Similar reasoning holds with h in place of g. This implies that $L(\hat{y}) = \{\hat{y}\}$, establishing sufficiency. □

The next example demonstrates that the condition $L(\hat{y}) = \{\hat{y}\}$ cannot be dropped, that is, (SDI) is not sufficient alone.

Example 1 Define the real functions

$$g(y) := \begin{cases} 0, & y \in (-1, 1), \\ |y| - 1, & \text{otherwise}, \end{cases} \quad \text{and} \quad h(y) := g(y/2) \quad \text{for } y \in \mathbb{R}.$$

Then clearly $y = 0$ is a non-strict global minimiser of $f = g - h$. But $\partial_\varepsilon h(0) = \partial_\varepsilon g(0)/2$, wherefore (SDI) holds, although strict optimality does not.

4 Non-strict local optimality

We now consider necessary conditions for local optimality, improving the sufficiency analysis of [2].

Theorem 2 *For the point $\hat{y} \in \mathbb{R}^m$ to be a local minimiser of $f := g - h$, it is sufficient that*

$$\text{there exists } \bar{\varepsilon} > 0, \quad \text{such that } 0 \in C_\varepsilon(\hat{y}) \text{ for each } \varepsilon \in [0, \bar{\varepsilon}). \tag{SDI$'$}$$

If $L(\hat{y}) = \{\hat{y}\}$, this condition is also necessary.

Proof As mentioned, sufficiency has been shown in [2], but it can also be shown analogously to the sufficiency proof of Theorem 1, as follows. We take $y_k \to \hat{y}$ ($k = 0, 1, 2, \ldots$; $y_k \neq \hat{y}$), some $z_k \in \partial h(y_k)$, and set $\varepsilon_{h,k} := e_h(\hat{y}; y_k, z_k)$. As before, for sufficiently large k, we have $\varepsilon_{h,k} < \bar{\varepsilon}$. Therefore, simply by employing $0 \in C_{\varepsilon_{h,k}}(\hat{y})$ in (10), Lemma 3 shows that $f(y_k) \geq f(\hat{y})$ for such large k. Thus $\hat{y}$ is a local minimiser. Note that a singleton mutual linearity set is not needed in this case.

Necessity likewise follows by adapting the necessity proof of Theorem 1. So, we assume that (SDI$'$) does not hold for any $\bar{\varepsilon} > 0$, and for $k = 0, 1, 2, \ldots$, take $\varepsilon_{h,k} \searrow 0$ and $(z_{h,k}, \varepsilon_{h,k}) \in G_h(\hat{y}) \setminus G_g(\hat{y})$ with $z_{h,k} \to z_h$. By application of Lemma 1 we again get $(z_{g,k}, \varepsilon_{g,k}) \in \operatorname{bd} G_g(\hat{y})$, $\alpha_k \geq 0$, and $(v_k, \delta_k) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ satisfying $\delta_k \in \{0, -1\}$, $\|v_k\| \geq 1 + \delta_k$, and $(z_{h,k}, \varepsilon_{h,k}) = (z_{g,k}, \varepsilon_{g,k}) + \alpha_k(v_k, \delta_k)$. As in the proof of Theorem 1, by possibly switching to a subsequence and letting $k \to \infty$, we may find α, (v, δ) and (z_g, ε_g) also satisfying these conditions for (z_h, ε_h) with $\varepsilon_h := 0$.

We first consider the case that (for a subsequence) $\|v_k\| \to 0$. Thanks to the property $\|v_k\| \geq 1+\delta_k$, we may again assume $\delta_k = -1$. Presently actually $\alpha_k > 0$, because $(z_{h,k}, \varepsilon_{h,k})$ is outside the closed set $G_g(\hat{y})$ containing $(z_{g,k}, \varepsilon_{g,k})$. An application of Lemma 2 therefore shows that $f(\hat{y}+v_k) \leq f(\hat{y})-\alpha_k < f(\hat{y})$. But we had $\|v_k\| \to 0$, so this is in contradiction to local optimality.

Suppose then that $\|v_k\| \geq \theta > 0$. Since $\varepsilon_h = 0$, in both of the cases $\delta = 0$ and $\delta = -1$, an application of Lemma 2 shows that $f(y_\lambda) \leq f(\hat{y}) - \lambda\alpha$ for $\lambda \in [0, 1]$ and $y_\lambda := \hat{y} + \lambda v$. If $\alpha > 0$, we thus reach the desired contradiction to local optimality. So suppose $\alpha = 0$. Since $(z_g, \varepsilon_g) = (z_h, 0) + \alpha(v, \delta)$, we in particular have $\varepsilon_g = 0$, and $z_g = z_h \in \partial g(\hat{y}) \cap \partial h(\hat{y})$. From (8) for $\delta = -1$ and (9) for $\delta = 0$, we find $g(y_\lambda) - g(\hat{y}) \leq \lambda\langle z_g, v\rangle$ for $\lambda \in [0, 1]$. The inclusion $z_g \in \partial g(\hat{y})$ provides the opposite inequality, establishing the equality

$$g(y_\lambda) - g(\hat{y}) = \lambda\langle z_g, v\rangle \quad \text{for} \quad \lambda \in [0, 1]. \tag{15}$$

This says $e_g\left(y_\lambda; \hat{y}, z_g\right) = 0$. Since, by assumption, $y_\lambda \notin L(\hat{y})$ for $\lambda \neq 0$, we must therefore have $e_h(y_\lambda; \hat{y}, z_g) \neq 0$. But $z_g \in \partial h(\hat{y})$, so we get the strict inequality $e_h(y_\lambda; \hat{y}, z_g) > 0$, i.e.,

$$h(y_\lambda) - h(\hat{y}) > \lambda\langle z_g, v\rangle \quad \text{for} \quad \lambda \in (0, 1]. \tag{16}$$

Combining (15) and (16) shows that $f(y_\lambda) < f(\hat{y})$ for $\lambda \in (0, 1]$. Since $\|v\| > 0$ implies $y_\lambda \neq \hat{y}$ when $\lambda \neq 0$, this provides the desired contradiction to local optimality. □

Similarly to Corollary 2, we get the following result.

Corollary 3 *The diff-convex function $f : \mathbb{R}^m \to (-\infty, \infty]$ has a local minimum at $\hat{y} \in \mathbb{R}^m$ if and only if* (SDI′) *holds for* some *decomposition $f = g - h$.*

Proof We only have to establish necessity, sufficiency being contained in Theorem 2. Given an arbitrary decomposition $f = g - h$, we consider the decomposition $f = g^\theta - h^\theta$ constructed in the proof of Corollary 1. These functions are strictly convex by construction, wherefore by the proof of Corollary 2, $L(\hat{y}) = \{\hat{y}\}$. Theorem 2 now shows the necessity of (SDI′) for this decomposition. □

Example 2 [2] provides a counterexample to the necessity of (SDI′) without the additional assumption $L(\hat{y}) = \{\hat{y}\}$. This is in the form of

$$g(y) := \begin{cases} 0, & y \leq 1, \\ (y-1)^2, & y > 1, \end{cases} \quad \text{and} \quad h(y) := g(-y).$$

The function f has local minimum at $y = 0$, but

$$\partial_\varepsilon g(0) = \left[0, 2\left(\sqrt{1+\varepsilon} - 1\right)\right], \quad \text{and} \quad \partial_\varepsilon h(0) = -\partial_\varepsilon g(0),$$

whence the condition $\partial_\varepsilon h(0) \subset \partial_\varepsilon g(0)$ does not hold for any $\varepsilon > 0$. However, we have $\{y \in \mathbb{R}^m \mid e_g(y; 0, 0) = 0\} = (-\infty, 1]$, and $\{y \in \mathbb{R}^m \mid e_h(y; 0, 0) = 0\} = [-1, \infty)$, whence $L(0) = [-1, 1]$.

5 Uniqueness of global minimisers

Finally, we represent some results pertaining to global optimality.

Theorem 3 *For the point $\hat{y}$ to be the unique global minimiser of $f := g - h$, it is sufficient that $L(\hat{y}) = \{\hat{y}\}$ and (SDI) holds with $\overline{\varepsilon} = +\infty$. If f is level-bounded, this is also necessary.*

Proof If (SDI) holds with $\overline{\varepsilon} = +\infty$, Lemma 3 yields $f(y) > f(\hat{y})$ for all $y \neq \hat{y}$. This establishes the sufficiency claim.

As for the necessity claim, clearly we must once again have $L(\hat{y}) = \{\hat{y}\}$, because f takes a single value on this set. It remains to show that (SDI) holds with $\overline{\varepsilon} = +\infty$. We do this by assuming that (SDI) is violated by the existence of $(z_h, \varepsilon_h) \in G_h(\hat{y}) \setminus \operatorname{int} G_g(\hat{y})$ with $\varepsilon_h > 0$. As before, we apply Lemma 1 to the pair to get $(z_g, \varepsilon_g) \in \operatorname{bd} G_g(\hat{y})$, $\alpha \geq 0$, and $(v, \delta) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$ with $\delta \in \{0, -1\}$, $\|v\| \geq 1 + \delta$ and $(z_h, \varepsilon_h) = (z_g, \varepsilon_g) + \alpha(v, \delta)$. Note that v is non-zero according to the lemma, since $\varepsilon_h > 0$.

Now, if $\delta = -1$, then Lemma 2 shows that $f(y + v) \leq f(\hat{y})$, providing a contradiction to uniqueness of the minimiser. If, on the other hand, $\delta = 0$, then by Lemma 2, $f(y + \lambda v) \leq f(\hat{y}) + \varepsilon_h$ for $\lambda \geq 0$. Because $v \neq 0$, this is in contradiction to f having been assumed level-bounded. □

The next corollary states the obvious counterpart to Corollaries 1 and 2, and is proved analogously.

Corollary 4 *Suppose $f : \mathbb{R}^m \to (-\infty, \infty]$ is diff-convex and level-bounded. Then it has its unique global minimum at $\hat{y} \in \mathbb{R}^m$ if and only if (SDI) holds with $\overline{\varepsilon} = +\infty$ for* every *decomposition $f = g - h$ or, equivalently,* some *decomposition with either g or h strictly convex.*

The following example demonstrates that the strict subdifferential inclusion condition (SDI) does not necessarily hold without the additional level-boundedness assumption. We denote by $B(x, r)$, the closed ball of radius r around x.

Example 3 Consider

$$g(y) := \|y\|, \quad \text{and} \quad h(y) := \begin{cases} \|y\|^2/4, & \|y\| \leq 2, \\ \|y\| - 1, & \|y\| > 2. \end{cases}$$

Clearly $f(y) = 1$ outside $B(0, 2)$, so the function is not level-bounded, while it has its unique global minimiser at $y = 0$. Indeed, (SDI) does not hold for $\varepsilon \geq 1$, because

$$\partial_\varepsilon g(0) = B(0, 1), \quad \text{and} \quad \partial_\varepsilon h(0) = \begin{cases} B\left(0, \sqrt{\varepsilon}\right), & \varepsilon \leq 1, \\ B(0, 1), & \varepsilon > 1. \end{cases}$$

Remark 2 Theorem 3 could be refined. In the final case of the proof that employs the level-boundedness assumption with $\delta = 0$, we may additionally assume $\alpha = 0$. This is because $f(y_\lambda) \leq f(\hat{y}) + \varepsilon_h - \lambda\alpha$ for $\lambda > 0$ by Lemma 2, so big enough λ would contradict $\hat{y}$ being a global minimiser if we had $\alpha > 0$. Now, due to $\alpha = 0$, we have $(z_g, \varepsilon_g) = (z_h, \varepsilon_h)$. This says that $z \in \partial_\varepsilon g(\hat{y}) \cap \partial_\varepsilon h(\hat{y})$ for $\varepsilon = \varepsilon_g (= \varepsilon_h)$. The procedure of Lemma 1, on the other hand, guarantees that in this case with $\delta = \alpha = 0$, there does not exist $(v, -1) \in N_{G_g(\hat{y})}(z_g, \varepsilon_g)$. Thus it is merely necessary to have

$$z \in \partial_\varepsilon g(\hat{y}) \cap \partial_\varepsilon h(\hat{y}) \implies (v, -1) \notin N_{G_g(\hat{y})}(z, \varepsilon) \quad \text{for all } v \in \mathbb{R}^m \tag{17}$$

along with $L(\hat{y}) = \{\hat{y}\}$ and (SDI′) for $\overline{\varepsilon} = +\infty$.

We now show that this relaxed condition is sufficient as well: Suppose $y \neq \hat{y}$, $z \in \partial h(y)$, and let $\varepsilon := e_h(\hat{y}; y, z)$. Then $z \in \partial_\varepsilon h(\hat{y})$, and by (SDI′) holding for $\overline{\varepsilon} = +\infty$, also $z \in \partial_\varepsilon g(\hat{y})$. Thus the premises of (17) are satisfied. Because $\operatorname{ri} G_g(\hat{y})$ is non-empty,

$\langle (y-\hat{y},-1),(z,\varepsilon)\rangle$ reaching $\sigma((y-\hat{y},-1);G_g(\hat{y}))$ requires $(y-\hat{y},-1)\in N_{G_g(\hat{y})}(z,\varepsilon)$ [see, e.g., 8, Theorem 27.4]. But this is not the case by (17). By (2), we therefore have $g(y)-g(\hat{y})>\langle z,y-\hat{y}\rangle-\varepsilon=h(y)-h(\hat{y})$, where the equality is due to $\varepsilon=e_h(\hat{y};y,z)$. This shows sufficiency.

6 Discussion of some related results

We conclude this paper with a discussion of a few recent results on closely related uses of the inclusions $\partial_\varepsilon h(\hat{y})\subset\partial_\varepsilon g(\hat{y})$: level-boundedness and sensitivity analysis.

Let $\mathcal{R}(\partial g)$ denote the range of ∂g and suppose $\mathcal{R}(\partial h)$ is bounded. According to a result in [9,10], the level sets $\operatorname{lev}_c f:=\{y\in\mathbb{R}^m \mid f(y)\le c\}$ are bounded if $\operatorname{cl}\mathcal{R}(\partial h)\subset\operatorname{int}\mathcal{R}(\partial g)$, and only if $\mathcal{R}(\partial h)\subset\operatorname{int}\mathcal{R}(\partial g)$. This result has particular relevance with regard to Theorem 3. That is, if the subdifferentials of h are bounded and the interior inclusion condition $0\in\operatorname{int}C_\varepsilon(\hat{y})$ holds in a suitable sense "at infinity", it is necessary that it holds for all $\varepsilon>0$, for $\hat{y}$ to be the unique global minimiser.

Also in [9], the use of the sets $C_\varepsilon(\hat{y})$ is studied for analysing the sensitivity of minimisers as the function f is subject to perturbations. This is done by applying and modifying the epigraphical methods of [1] along the same lines as in the specific case of reformulations of the Euclidean TSP in [11]. With the help of the lower estimate (10), it is thus possible to bound minimisers of perturbed functions in scaled polars $C_\varepsilon^\circ(\hat{y}):=\{z\in\mathbb{R}^m \mid \langle z,x\rangle\le 1 \text{ for all } x\in C_\varepsilon(\hat{y})\}$ of the sets $C_\varepsilon(\hat{y})$. One simplified case can be stated as follows: With $f:=g-h$, let $\hat{y}\in\arg\min f$, and assuming (SDI) holds, let $\varepsilon'\in[0,\bar{\varepsilon})$ and D be a closed set with

$$\hat{y}\in D\subset\left\{y\in\mathbb{R}^m \mid e_h\left(\hat{y};y,z\right)\le\varepsilon' \quad \text{for some } z\in\partial h(y)\right\}.$$

Let $\tilde{f}:\mathbb{R}^m\to(-\infty,\infty]$ be another proper lower-semicontinuous function with $y'\in\gamma\text{-}\arg\min_D\tilde{f}$ and choose $\eta>\sup_{D\cap\operatorname{dom}\tilde{f}}\left(f-\tilde{f}\right)-\left(f(\hat{y})-\tilde{f}(\hat{y})\right)$. Then

$$y'\in\hat{y}+\bigcup_{\varepsilon_h\in[0,\varepsilon']}\ \bigcap_{\varepsilon\in[\varepsilon_h,\bar{\varepsilon})}(\eta+\gamma+\varepsilon-\varepsilon_h)\,C_\varepsilon^\circ(\hat{y}).$$

That is, if y' is in a neighbourhood of $\hat{y}$ where the linearisation error of h is not too big, as determined by the choice of ε', then it can be bounded in a smaller neighbourhood determined by scaling the polars of the subdifferential differences $C_\varepsilon(\hat{y})$ by the approximation error $\eta+\gamma$.

Acknowledgments The author would like to thank the two anonymous referees for very helpful suggestions on improving the representation of the manuscript.

References

1. Attouch, H., Wets, R.J.B.: Quantitative stability of variational systems: II. A framework for nonlinear conditioning. SIAM J. Optim. **3**(2), 359–381 (1993)
2. Dür, M.: A parametric characterization of local optimality. Math. Methods Oper. Res. **57**, 101–109 (2003)
3. Hiriart-Urruty, J.B.: Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. In: Convexity and Duality in Optimization: Proceedings of the Symposium on Convexity and Duality in Optimization Held at the University of Groningen, The Netherlands, no. 256 in Lecture notes in Economics and Mathematical Systems, pp. 37–70. Springer, New York 22 June 1984

4. Hiriart-Urruty, J.B.: From convex optimization to non convex optimization, Part I: necessary and sufficent conditions for global optimality. In: Clarke, F., Demyanov, V., Giannessi, F. (eds.) Nonsmooth Optimization and Related Topics, pp. 219–239. Plenum Press, New York (1988)
5. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I–II. Springer, New York (1993)
6. Martínez-Legaz, J.E., Seeger, A.: A formula on the approximate subdifferential of the difference of convex functions. Bull. Aust. Math. Soc. **45**(1), 37–41 (1992)
7. Penot, J.P.: On the minimization of difference functions. J. Global Optim. **12**, 373–382 (1998)
8. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1972)
9. Valkonen, T.: Diff-convex combinations of Euclidean distances: a search for optima. No. 99 in Jyväskylä studies in computing. Ph.D Thesis, University of Jyväskylä (2008)
10. Valkonen, T., Kärkkäinen, T.: Clustering and the perturbed spatial median. Submitted (2008)
11. Valkonen, T., Kärkkäinen, T.: Continuous reformulations and heuristics for the Euclidean travelling salesperson problem. ESAIM Control Optim. Calc. Var. **15**(4), (2009). doi:10.1051/cocv:2008056