NONSMOOTH ANALYSIS OF LORENTZ INVARIANT FUNCTIONS

HRISTO S. SENDOV*

Abstract. A real valued function $g(x,t)$ on $\mathbb{R}^n \times \mathbb{R}$ is called *Lorentz invariant* if $g(x,t) = g(Ux,t)$ for all $n \times n$ orthogonal matrices U and all (x,t) in the domain of g. In other words, g is invariant under the linear orthogonal transformations preserving the *Lorentz cone*: $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,|\, t \geq \|x\|\}$. It is easy to see that every Lorentz invariant function can be decomposed as $g = f \circ \beta$, where $f : \mathbb{R}^2 \to \mathbb{R}$ is a symmetric function and β is the root map of the *hyperbolic polynomial* $p(x,t) = t^2 - x_1^2 - \cdots - x_n^2$. We investigate variety of important variational and non-smooth properties of g and characterize them in terms of the symmetric function f.

Key words and phrases: non-smooth analysis, convex analysis, hyperbolic polynomials, Lorentz cone, second-order cone, Clarke subdifferential, regular subdifferential, limiting subdifferential, proximal subdifferential, lower semicontinuous.

AMS 2000 Subject Classification. Primary 49J52; 58C20;
Secondary 58C25, 58E30.

1. Introduction and notation. Denote the set of all orthogonal $n \times n$ matrices by $O(n)$. Let the function $g(x,t)$ be defined on an open subset of $\mathbb{R}^n \times \mathbb{R}$, taking values in $\mathbb{R}$. The inner product of two vectors, (x,t) and (y,r) in $\mathbb{R}^n \times \mathbb{R}$ is $\langle (x,t),(y,r) \rangle = x^T y + tr$. Throughout the entire paper we assume that

$$g(Ux,t) = g(x,t), \quad \text{for all } U \in O(n), \tag{1.1}$$

and all (x,t) in the domain of g. We call a function g with property (1.1) *Lorentz invariant* because it is invariant under the linear orthogonal transformations preserving the *Lorentz cone* $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,|\, t \geq \|x\|\}$. A set $\Omega \subseteq \mathbb{R}^n \times \mathbb{R}$ is called *Lorentz invariant* if $(x,t) \in \Omega$ implies that $(Ux,t) \in \Omega$ for every $U \in O(n)$. Define the map

$$\beta(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^2,$$
$$\beta(x,t) = \frac{1}{\sqrt{2}}(t + \|x\|, t - \|x\|).$$

The rational behind the map β is the following. Consider the polynomial $p(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined by $p(x,t) = t^2 - x_1^2 - \cdots - x_n^2$ and let $d := (0,...,0,\sqrt{2}) \in \mathbb{R}^n \times \mathbb{R}$. Then, the coordinates of $\beta(x,t)$ are the roots of the polynomial $\lambda \mapsto P((x,t) - \lambda d)$. In general, a homogeneous polynomial $p(x) : \mathbb{R}^n \to \mathbb{R}$ with degree of homogeneity m, for which there is a direction $d \in \mathbb{R}^n$, $p(d) \neq 0$, such that $\lambda \mapsto p(x - \lambda d)$ has m real roots for every $x \in \mathbb{R}^n$, is called *hyperbolic*. In 1997, Güler [6], pointed out the relevance of these polynomials for optimization. Further information and developments can be found in [2], [13], [12], [18].

Let the function $f(a,b)$ be defined on an open subset of $\mathbb{R}^2$ and assume that it is *symmetric*, that is $f(a,b) = f(b,a)$ for all (a,b) in its domain. Necessarily, the domain of f is a *symmetric subset* of $\mathbb{R}^2$, that is $(a,b) \in A \Rightarrow (b,a) \in A$. The following easy lemma establishes the connection between g, β and f.

Lemma 1.1 (Lorentz invariant functions). *The next two properties of a function $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ are equivalent:*

*DEPARTMENT OF STATISTICAL AND ACTUARIAL SCIENCES; UNIVERSITY OF WESTERN ONTARIO; 1151 RICHMOND STREET NORTH; WESTERN SCIENCE CENTRE, ROOM 262; LONDON, ONTARIO; CANADA N6A 5B7. EMAIL: HSSENDOV@STATS.UWO.CA. RESEARCH SUPPORTED BY NSERC.

(i) g is Lorentz invariant;

(ii) $g = f \circ \beta$ for some symmetric function $f : \mathbb{R}^2 \to \mathbb{R}$.

If $g = f \circ \beta$ we say that f is the symmetric function corresponding to g. This correspondence is one-to-one and given g the corresponding symmetric function is

$$f(a, b) = g\Big(\frac{a-b}{\sqrt{2}}, 0, ..., 0, \frac{a+b}{\sqrt{2}}\Big). \tag{1.2}$$

That (1.2) defines a symmetric function in (a, b) is quaranteed by (1.1).

The aim of this paper is to establish variety of important for optimization variational and non-smooth properties of the function $g = f \circ \beta$ and how they arise from the corresponding properties of f. By deriving a wide range or nonsmooth formulae we hope this work to be a useful reference source. This work completes the similar investigations of spectral functions [8], [9] [11], [7]; and singular value functions [10], [14], [15]. Optimization problems over the Lorentz cone, also known as the second order cone, have wide range of applications, see for example [16]. With the development of the non-smooth Newton's method and various smoothing techniques the non-smooth properties of functions associated with the Lorentz cone have been of interest lately. For example, the strong semismoothness of the projection onto the Lorentz cone have been established in [23, Proposition 4.3]. A formula for the Bouligand subdifferential of the projection onto the Lorentz cone is derived in [24, Lemma 14]. Our paper is based on results that first appeared in the author's Ph.D. dissertation [21].

We conclude this section with an elementary fact.

LEMMA 1.2. *The composition $f \circ \beta$ is lower semicontinuous if and only if f is.*

Throughout the entire work, the functions g, β and f will have the properties described in this section.

2. Fenchel conjugation. For a function $F : \mathbb{R}^n \to (-\infty, +\infty]$, the *Fenchel conjugate* $F^* : \mathbb{R}^n \to [-\infty, +\infty]$ is the function

$$F^*(y) = \sup_{x \in \mathbb{R}^n} \{x^T y - F(x)\}.$$

It is well known that F^* is lower semicontinuous and convex [19]. In this section we prove the following formula.

PROPOSITION 2.1. *We always have*

$$(f \circ \beta)^* = f^* \circ \beta. \tag{2.1}$$

Proof. Let $y \neq 0$. In the third equality below, we use the fact that f is symmetric to see that the given supremum is the same as the supremum over the set $\{(a, b) \in \mathbb{R}^2 \,|\, a - b \geq 0\}$. From the definition we have

$$\begin{aligned}
(f \circ \beta)^*(y, r) &= \sup_{(x,t) \in \mathbb{R}^{n+1}} \{\langle (y, r), (x, t)\rangle - (f \circ \beta)(x, t)\} \\
&= \sup_{(a,b) \in \mathbb{R}^2} \sup_{\substack{(x,\,t) \ \text{s.t.} \\ t \,+\, \|x\| \,=\, a\sqrt{2} \\ t \,-\, \|x\| \,=\, b\sqrt{2}}} \{\langle (y, r), (x, t)\rangle - f(a, b)\} \\
&= \sup_{(a,b) \in \mathbb{R}^2} \Big\{\Big\langle (y, r), \Big(\frac{y}{\|y\|}\frac{a-b}{\sqrt{2}}, \frac{a+b}{\sqrt{2}}\Big)\Big\rangle - f(a, b)\Big\} \\
&= \sup_{(a,b) \in \mathbb{R}^2} \Big\{\|y\|\frac{a-b}{\sqrt{2}} + r\frac{a+b}{\sqrt{2}} - f(a, b)\Big\}
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{(a,b)\in\mathbb{R}^2} \left\{ \left\langle \left(\frac{r+\|y\|}{\sqrt{2}}, \frac{r-\|y\|}{\sqrt{2}} \right), (a,b) \right\rangle - f(a,b) \right\} \\
&= (f^* \circ \beta)(y,r).
\end{aligned}
$$

The case $y = 0$ is easy. □

An alternative proof of this result uses Theorem 5.5 and the example in Section 7.5 in [2], where the proposition has been generalized to the subclass of so called *isometric* hyperbolic polynomials. In [1, Theorem 6.1] the proposition has been shown to hold for symmetric functions composed with the eigenvalues of the elements of formally real Jordan algebras.

3. Convexity and convex subdifferentials.

3.1. Convexity. THEOREM 3.1. *The composition $f \circ \beta$ is convex and lower semicontinuous if and only if f is convex and lower semicontinuous.*

Proof. Suppose f is convex and lower semicontinuous. If $f \equiv +\infty$ then $f\circ\beta \equiv +\infty$ and the theorem is clear. Suppose f assumes some finite values. Then, using the convexity one can show that $f > -\infty$ and by [19, Theorem 12.2] we have $f^{**} = f$. Since f^* is symmetric, we use (2.1) in $f \circ \beta = f^{**} \circ \beta = (f^* \circ \beta)^*$, to conclude that $f \circ \beta$ is convex and lower semicontinuous. The opposite direction follows from (1.2) and Lemma 1.2. □

The proof of above theorem can be also deduced from Theorem 3.9 and the example in Section 7.5 in [2]. Even though the proof of Theorem 3.1 is quite elegant, a direct approach removes the condition that f be lower semicontinuous.

THEOREM 3.2. *The composition $f \circ \beta$ is convex if and only if f is convex.*

Proof. If $f \circ \beta$ is convex then f is by (1.2). Suppose now that f is convex with domain C. The domain of $f\circ\beta$ is $\beta^{-1}(C)$. Let (x,t), $(y,r) \in \beta^{-1}(C)$ and let $\alpha \in [0,1]$. Since $(t+\|x\|, t-\|x\|), (r+\|y\|, r-\|y\|) \in \sqrt{2}C$ and C is symmetric and convex we find that the points

$$
\begin{gathered}
(\alpha t + (1-\alpha)r + \alpha\|x\| + (1-\alpha)\|y\|, \alpha t + (1-\alpha)r - \alpha\|x\| - (1-\alpha)\|y\|), \\
(\alpha t + (1-\alpha)r - \alpha\|x\| - (1-\alpha)\|y\|, \alpha t + (1-\alpha)r + \alpha\|x\| + (1-\alpha)\|y\|)
\end{gathered}
$$

are both in $\sqrt{2}C$. Denote the first displayed point by $a\sqrt{2}$ and the second by $b\sqrt{2}$. Since

$$
-\alpha\|x\| - (1-\alpha)\|y\| \le \|\alpha x + (1-\alpha)y\| \le \alpha\|x\| + (1-\alpha)\|y\|,
$$

there is a $\beta \in [0,1]$ such that for the point

$$
c\sqrt{2} := (\alpha t + (1-\alpha)r + \|\alpha x + (1-\alpha)y\|, \alpha t + (1-\alpha)r - \|\alpha x + (1-\alpha)y\|)
$$

we have $c = \beta a + (1-\beta)b \in C$. Thus

$$
\begin{aligned}
f(c) &\le \beta f(a) + (1-\beta)f(b) = f(a) \\
&\le \alpha f((t+\|x\|, t-\|x\|)/\sqrt{2}) + (1-\alpha)f((r+\|y\|, r-\|y\|)/\sqrt{2}),
\end{aligned}
$$

where we used the fact that $f(a) = f(b)$ and that f is convex. □

The proof of Theorem 3.2 shows the following property.

LEMMA 3.3. *Let $C \subseteq \mathbb{R}^2$ be a convex and symmetric set. Then*

$$
\beta^{-1}(C) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \,|\, \beta(x,t) \in C\}
$$

is convex and Lorentz invariant.

3.2. Convex subdifferentials. Let $f : \mathbb{R}^2 \to (-\infty, +\infty]$ be convex. For every point (a, b) such that $f(a, b) < +\infty$ we define the *subdifferential* of f at (a, b) to be the set

$$\partial f(a, b) = \{(a', b') | f(c, d) - f(a, b) \geq \langle (a', b'), (c, d) - (a, b)\rangle, \forall (c, d)\}.$$

It is easy to see that $f(a, b) + f^*(a', b') = \langle (a, b), (a', b')\rangle$ if and only if $(a', b') \in \partial f(a, b)$. The set $\partial f(a, b)$ is a singleton $\{(a', b')\}$ if and only if f is differentiable at the point (a, b) with gradient $\nabla f(a, b) = (a', b')$, see [19, Theorem 25.1].

The following result gives a formula for the subgradient of the composition $f \circ \beta$.

THEOREM 3.4. *Suppose $f : \mathbb{R}^2 \to (-\infty, +\infty]$ is convex and lower semicontinuous. Then $(y, r) \in \partial(f \circ \beta)(x, t)$ if and only if $\beta(y, r) \in \partial f(\beta(x, t))$ and $x^T y = \|x\|\|y\|$.*

Proof. Suppose first $(y, r) \in \partial(f \circ \beta)(x, t)$. Then using formula (2.1) we get

$$\begin{aligned}
\|x\|\|y\| + rt &\geq x^T y + rt = \langle (y, r), (x, t)\rangle \\
&= (f \circ \beta)(x, t) + (f \circ \beta)^*(y, r) \\
&= (f \circ \beta)(x, t) + (f^* \circ \beta)(y, r) \\
&= f\left(\frac{t + \|x\|}{\sqrt{2}}, \frac{t - \|x\|}{\sqrt{2}}\right) + f^*\left(\frac{r + \|y\|}{\sqrt{2}}, \frac{r - \|y\|}{\sqrt{2}}\right) \\
&\geq \big((t + \|x\|)(r + \|y\|) + (t - \|x\|)(r - \|y\|)\big)/2 \\
&= \|x\|\|y\| + rt.
\end{aligned}$$

Thus, we have equalities everywhere: (a) $\beta(y, r) \in \partial f(\beta(x, t))$ and (b) $x^T y = \|x\|\|y\|$. In the other direction the proof is clear by reversing the steps above. $\square$

For a generalization of this proposition to formally real Jordan algebras see [1, Corollary 6.2].

4. Differentiability. The partial derivatives of the function f with respect to its first and second argument are denoted by f_1' and f_2' respectively.

THEOREM 4.1. *The composition $f \circ \beta$ is differentiable at the point (x, t) if and only if f is differentiable at $\beta(x, t)$. In that case we have the formulae*

$$\nabla_x (f \circ \beta)(x, t) = \begin{cases} \frac{f_1'(\beta(x,t)) - f_2'(\beta(x,t))}{\sqrt{2}\|x\|} x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

and

$$\frac{d}{dt}(f \circ \beta)(x, t) = \frac{1}{\sqrt{2}}(f_1'(\beta(x, t)) + f_2'(\beta(x, t))).$$

Proof. Suppose first that f is differentiable at the point $\beta(x, t)$. If $x \neq 0$ the theorem and the formulae are trivial and follow from the chain rule. So let us assume now that $x = 0$. Let $h = (\bar{h}, h_{n+1}) \in \mathbb{R}^n \times \mathbb{R}$ and

$$d := \big(0, ..., 0, (f_1'(\beta(x, t)) + f_2'(\beta(x, t)))/\sqrt{2}\big) \in \mathbb{R}^n \times \mathbb{R}.$$

Then

$$\lim_{h \to 0} \frac{|(f \circ \beta)((0, t) + (\bar{h}, h_{n+1})) - (f \circ \beta)((0, t)) - d^T h|}{\|h\|} =$$

$$\lim_{h\to 0} \frac{|f(\beta(\bar h, t+h_{n+1})) - f(\beta(0,t)) - h_{n+1}(f_1'(\beta(0,t)) + f_2'(\beta(0,t)))/\sqrt{2}|}{\|h\|}.$$

The fact that f is differentiable at $\beta(0,t) = (t/\sqrt{2}, t/\sqrt{2})$ gives

$$f(\beta(\bar h, t+h_{n+1})) \sim f(\beta(0,t)) + f_1'(\beta(0,t))\frac{h_{n+1} + \|\bar h\|}{\sqrt{2}} + f_2'(\beta(0,t))\frac{h_{n+1} - \|\bar h\|}{\sqrt{2}},$$

where $\sim$ indicates that the difference of both sides is of order $o(\|h\|)$. Using the fact that for a symmetric function f, $f_1'(\beta(0,t)) = f_2'(\beta(0,t))$ and substituting above we see that the limit is zero, that is, $\nabla(f\circ\beta)(0,t) = d$.

The proof in the other direction is easy using formula (1.2). □

THEOREM 4.2. *Let f be symmetric and defined on an open symmetric subset of $\mathbb{R}^2$. Then $f\circ\beta$ is continuously differentiable at the point (x,t) if and only if f is continuously differentiable at $\beta(x,t)$.*

Proof. Suppose that f is continuously differentiable at $\beta(x,t)$. The theorem is clear if $x \neq 0$. So suppose $x = 0$. Let $\{(x^k, t^k)\}$ be a sequence of points in $\mathbb{R}^n \times \mathbb{R}$ approaching $(0,t)$. We need only prove that $\nabla(f\circ\beta)(x^k,t^k)$ approaches $\nabla(f\circ\beta)(0,t)$. We consider two cases. The general case easily follows by combining them.

Case 1. If $x^k = 0$ for all k. Then using the formula in Theorem 4.1 we obtain

$$\begin{aligned}\lim_{k\to\infty} \nabla(f\circ\beta)(0,t^k) &= \lim_{k\to\infty}\left(0,...,0,\frac{1}{\sqrt{2}}\big(f_1'(\beta(0,t^k)) + f_2'(\beta(0,t^k))\big)\right)\\ &= \nabla(f\circ\beta)(0,t),\end{aligned}$$

by the continuity of ∇f at $\beta(0,t)$.

Case 2. If $x^k \neq 0$ for all k. Using again the formula in Theorem 4.1 for the derivative with respect to t we obtain

$$\lim_{k\to\infty} (f\circ\beta)_t'(x^k,t^k) = \lim_{k\to\infty}\frac{1}{\sqrt{2}}\big(f_1'(\beta(x^k,t^k)) + f_2'(\beta(x^k,t^k))\big) = (f\circ\beta)_t'(0,t).$$

For the derivative with respect to x_i we get

$$\lim_{k\to\infty} (f\circ\beta)_{x_i}'(x^k,t^k) = \lim_{k\to\infty}\frac{x_i^k}{\sqrt{2}\|x^k\|}\big(f_1'(\beta(x^k,t^k)) - f_2'(\beta(x^k,t^k))\big) = 0,$$

because $x_i^k/\|x^k\|$ is bounded and the continuity of ∇f at $\beta(0,t)$ gives us

$$\lim_{k\to\infty} (f_1'(\beta(x^k,t^k)) - f_2'(\beta(x^k,t^k)) = f_1'(\beta(0,t)) - f_2'(\beta(0,t)) = 0.$$

The last equality follows from the fact that f is symmetric.

The opposite direction of the theorem is easy using (1.2). □

5. The decomposition functions. In this section we define the functions d_z and d_z^* and summarize some of their properties that will be used frequently. We call them *decomposition* functions because they will be used to describe how the subgradients of $f\circ\beta$ are decomposed into subgradients of f.

DEFINITION 5.1. *For every nonzero vector z in $\mathbb{R}^n$ we define the map*

$$d_z : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^2,$$

$$d_z(y,t) = \frac{1}{\sqrt{2}}\left(t + \frac{z^T y}{\|z\|}, t - \frac{z^T y}{\|z\|}\right).$$

In cases when the direction (y,t) is fixed and clear from the context we simply write d_z in stead of $d_z(y,t)$.

DEFINITION 5.2. *For every nonzero vector z in $\mathbb{R}^n$ we define the map*

$$d_z^* : \mathbb{R}^2 \to \mathbb{R}^n \times \mathbb{R},$$
$$d_z^*(a,b) = \left(\frac{z}{\|z\|}\frac{a-b}{\sqrt{2}}, \frac{a+b}{\sqrt{2}}\right).$$

The following lemma collects a few elementary properties of the maps d_z and d_z^*. The proof is omitted.

LEMMA 5.3. *Let z and w be nonzero vectors in $\mathbb{R}^n$.*

(i) The maps $d_z(\cdot)$ and $d_z^*(\cdot)$ are linear and adjoint to each other.

(ii) For every point (γ_1, γ_2) in $\mathbb{R}^2$

$$d_w d_z^*(\gamma_1,\gamma_2) = \frac{1+\delta}{2}(\gamma_1,\gamma_2) + \frac{1-\delta}{2}(\gamma_2,\gamma_1),$$

where $\delta = \frac{w^T z}{\|w\|\|z\|} \in [-1,1]$. In particular, when $w = z$ we have

$$d_z d_z^*(\gamma_1,\gamma_2) = (\gamma_1,\gamma_2).$$

(iii) For every point (y,r) in $\mathbb{R}^n \times \mathbb{R}$ such that $y = az$ for some $a \in \mathbb{R}$

$$d_z^* d_z(y,r) = (y,r).$$

LEMMA 5.4. *Let A and B be symmetric subsets of $\mathbb{R}^2$. The sets*

$$\mathcal{D}(A) = \{d_z^*(\gamma_1,\gamma_2) | (\gamma_1,\gamma_2) \in A,\ z \neq 0\}, \text{ and}$$
$$\mathcal{C}(A) = \{(y,r) | d_z(y,r) \in A,\ \forall z \neq 0\},$$

satisfy the following properties.

(i) If A is convex then

(a) If (x,t) is in $\mathcal{D}(A)$, then $(\delta x, t)$ is in $\mathcal{D}(A)$ for every $\delta \in [-1,1]$.

(b) $\mathcal{D}(A)$ is a convex set.

(c) $\mathcal{D}(A) = \mathcal{C}(A)$.

(d) If B is also convex, then $\operatorname{cl}(\mathcal{D}(A) + \mathcal{D}(B)) = \operatorname{cl}\mathcal{D}(A+B)$.

(ii) For any A we have

(a) $\operatorname{conv}\mathcal{D}(A) = \mathcal{D}(\operatorname{conv} A)$.

(b) $\mathcal{D}(\operatorname{cl} A) = \operatorname{cl}\mathcal{D}(A)$.

Proof. **Part (i)a**. Let $(x,t) = d_z^*(\gamma_1,\gamma_2)$ for some (γ_1,γ_2) in A and $z \neq 0$. Since the set A is symmetric and convex, (γ_2,γ_1) is in A and for every $\alpha \in [0,1]$ the convex combination $(\alpha\gamma_1 + (1-\alpha)\gamma_2, \alpha\gamma_2 + (1-\alpha)\gamma_1)$ is in A. Thus,

$$\begin{aligned} d_z^*\big(\alpha\gamma_1 + (1-\alpha)\gamma_2, \alpha\gamma_2 + (1-\alpha)\gamma_1\big) &= \left(\frac{z}{\|z\|}\frac{\gamma_1-\gamma_2}{\sqrt{2}}(2\alpha-1), \frac{\gamma_1+\gamma_2}{\sqrt{2}}\right) \\ &= (x(2\alpha-1), t) \in \mathcal{D}, \end{aligned}$$

for all $\alpha \in [0,1]$. Now set $\delta := 2\alpha - 1$.

Part (i)b. Since A is convex, for any two points (γ_1,γ_2) and (δ_1,δ_2) in A and $\mu\in[0,1]$, we have that $(\mu\gamma_1+(1-\mu)\delta_1, \mu\gamma_2+(1-\mu)\delta_2)$ is in A. Thus, for every $z\neq 0$

$$\left(\frac{z}{\|z\|}\frac{\mu(\gamma_1-\gamma_2)+(1-\mu)(\delta_1-\delta_2)}{\sqrt{2}}, \frac{\mu(\gamma_1+\gamma_2)+(1-\mu)(\delta_1+\delta_2)}{\sqrt{2}}\right)\in\mathcal{D}. \tag{5.1}$$

Take two points (x^1,t^1) and (x^2,t^2) in $\mathcal{D}$ and a number $\mu\in(0,1)$. We want to show that $(\mu x^1+(1-\mu)x^2, \mu t^1+(1-\mu)t^2)$ is also in $\mathcal{D}$. Suppose

$$(x^1,t^1)=d^*_{z^1}(\gamma_1,\gamma_2), \qquad (x^2,t^2)=d^*_{z^2}(\delta_1,\delta_2)$$

for some (γ_1,γ_2) and (δ_1,δ_2) in A, $z^1\neq 0$ and $z^2\neq 0$. Set

$$z_\mu := \mu\frac{\gamma_1-\gamma_2}{\sqrt{2}}\frac{z^1}{\|z^1\|}+(1-\mu)\frac{\delta_1-\delta_2}{\sqrt{2}}\frac{z^2}{\|z^2\|},$$

and notice that

$$\|z_\mu\|\le\mu\frac{|\gamma_1-\gamma_2|}{\sqrt{2}}+(1-\mu)\frac{|\delta_1-\delta_2|}{\sqrt{2}}.$$

Then

$$\mu(x^1,t^1)+(1-\mu)(x^2,t^2)=\left(z_\mu, \frac{\mu(\gamma_1+\gamma_2)+(1-\mu)(\delta_1+\delta_2)}{\sqrt{2}}\right). \tag{5.2}$$

If $z_\mu=0$ then from (5.1) and part (i)a with $\delta=0$ we see that

$$\mu(x^1,t^1)+(1-\mu)(x^2,t^2)\in\mathcal{D}.$$

Suppose now $z_\mu\neq 0$. Choose one of the points (γ_1,γ_2), (γ_2,γ_1) in A and one of the points (δ_1,δ_2), (δ_2,δ_1) in A so that, using part (i)a, inclusion (5.1) becomes

$$\left(\frac{z}{\|z\|}\frac{\mu|\gamma_1-\gamma_2|+(1-\mu)|\delta_1-\delta_2|}{\sqrt{2}}\delta, \frac{\mu(\gamma_1+\gamma_2)+(1-\mu)(\delta_1+\delta_2)}{\sqrt{2}}\right)\in\mathcal{D},$$

for all $z\neq 0$ and $\delta\in(0,1)$. Let δ be a number in $(0,1)$ such that

$$\frac{\mu|\gamma_1-\gamma_2|+(1-\mu)|\delta_1-\delta_2|}{\sqrt{2}}\delta=\|z_\mu\|.$$

Putting all together we obtain that (5.2) is in $\mathcal{D}$, showing that $\mathcal{D}$ is a convex set.

Part (i)c. Suppose $(y,r)\in\mathcal{C}$, then $d_z(y,r)\in A$ for all $z\neq 0$. Apply Lemma 5.3 part (iii) with $a=0$ and any z if $y=0$; or with $a=1$ and $z=y$ if $y\neq 0$ to obtain

$$(y,r)=d^*_z d_z(y,r)=d^*_z(d_z(y,r))\in\mathcal{D}.$$

This shows that $\mathcal{C}\subseteq\mathcal{D}$.

Suppose now $(y,r)\in\mathcal{D}$. That is, $(y,r)=d^*_z(\gamma_1,\gamma_2)$ for some (γ_1,γ_2) in A and some $z\neq 0$. Let $\hat{z}$ be an arbitrary nonzero vector and set $\delta:=\frac{z^T\hat{z}}{\|z\|\|\hat{z}\|}\in[-1,1]$. Then by Lemma 5.3 part (ii) we have

$$d_{\hat{z}}(y,r)=d_{\hat{z}}d^*_z(\gamma_1,\gamma_2)=\frac{1+\delta}{2}(\gamma_1,\gamma_2)+\frac{1-\delta}{2}(\gamma_2,\gamma_1)\in A,$$

because A is symmetric and convex. So $\mathcal{D} \subseteq \mathcal{C}$.

Part (i)d. By part (i)b we have that both $\mathcal{D}(A)+\mathcal{D}(B)$ and $\mathcal{D}(A+B)$ are convex sets. It is clear that the latter set is contained in the former:

$$\mathrm{cl}\,(\mathcal{D}(A) + \mathcal{D}(B)) \supseteq \mathrm{cl}\,\mathcal{D}(A + B).$$

In order to show that the sets are equal it suffices to show that their support functions are equal. Fix any $x \in \mathbb{R}^n$ and suppose first that $x \neq 0$. In the first and last equality below, we use the fact that A and B are symmetric sets.

$$\begin{aligned}
&\max\{\langle (x,t), (d^*_{z^1}(\gamma_1,\gamma_2) + d^*_{z^2}(\delta_1,\delta_2))\rangle \,|\, (\gamma_1,\gamma_2) \in A,\ (\delta_1,\delta_2) \in B,\ z^1 \neq 0,\ z^2 \neq 0\} \\
&\quad = \max\{\langle (x,t), (d^*_x(\gamma_1,\gamma_2) + d^*_x(\delta_1,\delta_2))\rangle \,|\, (\gamma_1,\gamma_2) \in A,\ (\delta_1,\delta_2) \in B\} \\
&\quad = \max\{\langle (x,t), d^*_x(\gamma_1+\delta_1,\gamma_2+\delta_2)\rangle \,|\, (\gamma_1,\gamma_2) \in A,\ (\delta_1,\delta_2) \in B\} \\
&\quad = \max\{\langle (x,t), d^*_x(\alpha_1,\alpha_2)\rangle \,|\, (\alpha_1,\alpha_2) \in A+B\} \\
&\quad = \max\{\langle (x,t), d^*_z(\gamma_1,\gamma_2)\rangle \,|\, (\gamma_1,\gamma_2) \in A+B,\ z \neq 0\}.
\end{aligned}$$

The case $x = 0$ is easy.

Part (ii)a. The inclusion $A \subseteq \mathrm{conv}\,A$ implies $\mathcal{D}(A) \subseteq \mathcal{D}(\mathrm{conv}\,A)$. Since the set $\mathcal{D}(\mathrm{conv}\,A)$ is convex by part (i)b, we obtain $\mathrm{conv}\,\mathcal{D}(A) \subseteq \mathcal{D}(\mathrm{conv}\,A)$. The opposite inclusion $\mathcal{D}(\mathrm{conv}\,A) \subseteq \mathrm{conv}\,\mathcal{D}(A)$ is easy.

Part (ii)b. Let $\{d^*_{x^k}(\gamma_1^k,\gamma_2^k)\}$ be a sequence in $\mathcal{D}(A)$ approaching a vector (z,s). Since the unit sphere in $\mathbb{R}^n$ is compact, we can find a subsequence, denoted again by k, such that $x^k/\|x^k\|$ converges to a unit vector x. For this subsequence we have $|\gamma_1^k - \gamma_2^k| \to \sqrt{2}\|z\|$ and $\gamma_1^k + \gamma_2^k \to \sqrt{2}s$. Consequently, $\{(\gamma_1^k,\gamma_2^k)\}$ is bounded so there is a subsequence, denoted again by k, for which $(\gamma_1^k,\gamma_2^k) \to (\gamma_1,\gamma_2) \in \mathrm{cl}\,A$. So, the sequence $\{d^*_{x^k}(\gamma_1^k,\gamma_2^k)\}$ approaches $d^*_x(\gamma_1,\gamma_2)$ which is in $\mathcal{D}(\mathrm{cl}\,A)$. This shows that for an arbitrary set A we have the inclusion $\mathcal{D}(\mathrm{cl}\,A) \supseteq \mathrm{cl}\,\mathcal{D}(A)$. The opposite inclusion is easy. □

6. Clarke subdifferential - the Lipschitz case. Suppose that h is a real-valued function defined on some subset of $\mathbb{R}^m$. We say that h is *locally Lipschitz* at x in $\mathbb{R}^m$ if there exists a scalar K such that

$$|h(x'') - h(x')| \leq K\|x'' - x'\|, \quad \text{for all } x'', x' \text{ close to } x.$$

For locally Lipschitz functions the *Clarke directional derivative* [4] at the point x in the direction v is defined to be

$$h^\circ(x;v) = \limsup_{y\to x;\,\lambda\downarrow 0} \frac{h(y+\lambda v) - h(y)}{\lambda}.$$

For y close to x and λ to 0, the difference quotient in the definition of $h^\circ(x;v)$ is bounded above by $K|v|$. Thus, $h^\circ(x;v)$ is well defined and finite. We need the following formula for the Clarke directional derivative that can be found in [4, p. 64]:

$$h^\circ(x;v) = \limsup_{y\to x} \{\langle \nabla h(y), v\rangle \,|\, y \text{ is such that } \nabla h(y) \text{ exists}\},^{1} \tag{6.1}$$

for every pair $(x;v)$. In other words, there exists a sequence $\{x^k\}$ in $\mathbb{R}^m$ approaching x such that f is differentiable at each x_n and

$$\langle \nabla h(x^k), v\rangle \to h^\circ(x;v). \tag{6.2}$$

[1] By the Rademacher's theorem, locally Lipschitz functions are differentiable almost everywhere.

The *Clarke subdifferential* $\partial^c h(x)$ is defined as

$$\partial^c h(x) = \{\xi \mid \langle v, \xi\rangle \le h^\circ(x; v) \text{ for all } v\}.$$

In can be shown that the set $\partial^c h(x)$ is compact, nonempty and convex. If h is convex and finite on a neighbourhood of x then $\partial^c h(x) = \partial h(x)$ and if h is continuously differentiable at x then $\partial^c h(x) = \{\nabla h(x)\}$. In this sense the Clarke subdifferential generalizes both the convex subdifferential and the gradient of a C^1 function. Finally, Proposition 2.1.2 in [4] shows that the Clarke directional derivative is the *support function* of the Clarke subdifferential:

$$h^\circ(x; v) = \max\{\langle v, \xi\rangle \mid \xi \in \partial^c h(x)\}. \tag{6.3}$$

Now, we return to the symmetric, bivariate function f, which we now require to be locally Lipschitz. It is not difficult to see that f is locally Lipschitz if and only if $f \circ \beta$ is. We are going to present a formula expressing the Clarke subdifferential of $f \circ \beta$ in terms of the Clarke subdifferential of f.

The following elementary lemma shows that the Clarke directional derivative of $h \circ \beta$ is invariant under Lorentz orthogonal transformations of the argument and the direction.

Lemma 6.1. *Let (x, t) be a point in the domain of $f \circ \beta$, let (y, r) be a direction and let U be an orthogonal matrix. Then*

$$(f \circ \beta)^\circ((x, t); (y, r)) = (f \circ \beta)^\circ((Ux, t); (Uy, r)).$$

Theorem 6.2 (Clarke directional derivative). *Let $(0, t)$ be a point in the domain of $f \circ \beta$ and let (y, r) be any direction. Then if $x = 0$*

$$(f \circ \beta)^\circ((0, t); (y, r)) = \max\{f^\circ(\beta(0, t); d_z(y, r)) | z \in \mathbb{R}^n,\ z \neq 0\}. \tag{6.4}$$

Note 6.3. *For the Clarke directional derivative at a point (x, t) with $x \neq 0$ see Corollary 6.6.*

Proof. By (6.2), there is a sequence of points $\{(x^k, t^k)\}$ approaching $(0, t)$ such that

$$(f \circ \beta)^\circ((0, t); (y, r)) = \lim_{k\to\infty} \langle \nabla(f \circ \beta)(x^k, t^k), (y, r)\rangle.$$

In order to evaluate $\nabla(f \circ \beta)$ using Theorem 4.1 we need to consider two cases depending on whether x^k is zero or not. The general situation follows from these two cases by considering subsequences.

Case 1.a. Suppose $x^k = 0$ for all k. Let $\beta^k := \beta(0, t^k)$ and note that $f_1'(\beta^k) = f_2'(\beta^k)$. Fix an arbitrary nonzero vector $z \in \mathbb{R}^n$. Then

$$\begin{aligned}
(f \circ \beta)^\circ((0, t); (y, r)) &= \lim_{k\to\infty} \langle \nabla(f \circ \beta)(0, t^k), (y, r)\rangle \\
&= \lim_{k\to\infty} \left\langle \left(0, ..., 0, \frac{f_1'(\beta^k) + f_2'(\beta^k)}{\sqrt{2}}\right), (y, r)\right\rangle \\
&= \lim_{k\to\infty} \langle \nabla f(\beta^k), \beta(0, r)\rangle \\
&= \lim_{k\to\infty} \big\langle \nabla f(\beta^k), d_z(y, r)\big\rangle
\end{aligned}$$

$$\leq f^\circ(\beta(0,t); d_z(y,r))).$$

In the last inequality we used (6.1).

Case 1.b. Suppose $x^k \neq 0$ for all k, $\lim_{k\to\infty} \frac{x^k}{\|x^k\|} = \frac{z}{\|z\|}$ and let $\beta^k := \beta(x^k, t^k)$. Then, we have

$$\begin{aligned}
(f\circ\beta)^\circ((0,t);(y,r)) &= \lim_{k\to\infty} \langle \nabla(f\circ\beta)(x^k,t^k),(y,r)\rangle \\
&= \lim_{k\to\infty} \left\langle \left(\frac{f_1'(\beta^k) - f_2'(\beta^k)}{\sqrt{2}\|x^k\|} x^k, \frac{f_1'(\beta^k) + f_2'(\beta^k)}{\sqrt{2}} \right), (y,r) \right\rangle \\
&= \lim_{k\to\infty} \frac{f_1'(\beta^k) - f_2'(\beta^k)}{\sqrt{2}\|x^k\|} (x^k)^T y + \frac{f_1'(\beta^k) + f_2'(\beta^k)}{\sqrt{2}} r \\
&= \lim_{k\to\infty} f_1'(\beta^k)\left(\frac{r}{\sqrt{2}} + \frac{(x^k)^T y}{\sqrt{2}\|x^k\|} \right) + f_2'(\beta^k)\left(\frac{r}{\sqrt{2}} - \frac{(x^k)^T y}{\sqrt{2}\|x^k\|} \right) \\
&= \lim_{k\to\infty} \big\langle \nabla f'(\beta^k), d_z(y,r) \big\rangle \\
&\leq f^\circ(\beta(0,t); d_z(y,r)),
\end{aligned}$$

where, in substituting $x^k/\|x^k\|$ by $z/\|z\|$ in the last equality, we used the fact that since f is locally Lipschitz the sequence $\{(f_1'(\beta^k), f_2'(\beta^k)\}$ is bounded. All this shows that if $x = 0$ then

$$(f\circ\beta)^\circ((0,t);(y,r)) \leq \sup\{f^\circ(\beta(0,t); d_z(y,r) | z \in \mathbb{R}^n,\ z \neq 0\}.$$

To show the opposite inequality, fix a nonzero vector $z \in \mathbb{R}^n$. There is a sequence of points $\{(a_k, b_k)\}$ approaching $\beta(0,t)$ such that

$$f^\circ(\beta(0,t); d_z(y,r)) = \lim_{n\to\infty} \langle \nabla f(a_k,b_k), d_z(y,r)\rangle.$$

There is an infinite subsequence $\{(a_{k'}, b_{k'})\}$ of $\{(a_k,b_k)\}$ that satisfies one of the three possibilities

(i) $a_{k'} = b_{k'}$ for all k'.

(ii) $a_{k'} > b_{k'}$ for all k'.

(iii) $a_{k'} < b_{k'}$ for all k'.

For this subsequence we still have

$$f^\circ(\beta(0,t); d_z(y,r)) = \lim_{k'\to\infty} \langle \nabla f(a_{k'},b_{k'}), d_z(y,r)\rangle.$$

Without loss of generality we may assume that $\{(a_k,b_k)\}$ satisfies one of the three possibilities and consider them separately.

Case 2.a. Suppose $a_k = b_k$ for all k. Note that in this case we have $f_1'(a_k,a_k) = f_2'(a_k,a_k)$. Thus,

$$\begin{aligned}
f^\circ(\beta(0,t); d_z(y,r)) &= \lim_{k\to\infty} \langle \nabla f(a_k,a_k), d_z(y,r)\rangle \\
&= \lim_{k\to\infty} \frac{f_1'(a_k,a_k) + f_2'(a_k,a_k)}{\sqrt{2}} r \\
&= \lim_{k\to\infty} \langle \nabla(f\circ\beta)(0,a_k),(y,r)\rangle \\
&\leq (f\circ\beta)^\circ((0,t);(y,r)).
\end{aligned}$$

Case 2.b. Suppose $a_k > b_k$ for all k. Define the sequence of vectors

$$z^k := \left(\frac{a_k - b_k}{2}, 0, ..., 0\right) \in \mathbb{R}^n,$$

(notice that $\|z^k\| = (a_k - b_k)/2$) and let U be an orthogonal matrix such that

$$\lim_{k\to\infty} \frac{Uz^k}{\|z^k\|} = \frac{z}{\|z\|}. \tag{6.5}$$

In the third equality below we use the fact that the Lipschitzness of f implies that the sequence $\{f_1'(a_k, b_k) - f_2'(a_k, b_k)\}$ is bounded, thus in the limit we can replace $z/\|z\|$ by $Uz^k/\|z^k\|$. We calculate

$$\begin{aligned}
f^\circ(\beta(0,t); d_z(y,r)) &= \lim_{k\to\infty} \langle \nabla f(a_k, b_k), d_z(y, r)\rangle \\
&= \lim_{k\to\infty} \left\langle \frac{f_1'(a_k, b_k) - f_2'(a_k, b_k)}{\sqrt{2}\|z\|} z, y \right\rangle + \frac{f_1'(a_k, b_k) + f_2'(a_k, b_k)}{\sqrt{2}} r \\
&= \lim_{k\to\infty} \left\langle \frac{f_1'(a_k, b_k) - f_2'(a_k, b_k)}{\sqrt{2}\|z^k\|} z^k, U^T y \right\rangle + \frac{f_1'(a_k, b_k) + f_2'(a_k, b_k)}{\sqrt{2}} r \\
&= \lim_{k\to\infty} \left\langle \nabla(f\circ\beta)\left(\frac{a_k - b_k}{\sqrt{2}}, 0, ..., 0, \frac{a_k + b_k}{\sqrt{2}}\right), (U^T y, r)\right\rangle \\
&\le (f\circ\beta)^\circ((0,t); (U^T y, r)) \\
&= (f\circ\beta)^\circ((0,t); (y, r)).
\end{aligned}$$

In the last equality we used Lemma 6.1.

Case 2.c. Suppose $a_k < b_k$ for all k. Define the sequence of vectors

$$z^k := \left(\frac{b_k - a_k}{2}, 0, ..., 0\right) \in \mathbb{R}^n,$$

(notice that $\|z^k\| = (b_k - a_k)/2$) and proceed analogously to the previous case. □

It is straightforward to check that for every $(y, r) \in \mathbb{R}^n \times \mathbb{R}$ and every nonzero $x \in \mathbb{R}^n$ we have

$$\lim_{(x',t')\to(x,t),\, \mu\downarrow 0} \frac{\beta((x', t') + \mu(y, r)) - \beta(x', t')}{\mu} = d_x(y, r).$$

Applying [3, Theorem 6.2.3] to the Lipschitz map $\beta(x, t)$ we obtain the following result.

Lemma 6.4. *If $x \neq 0$ then $\beta(x, t)$ is strictly differentiable and its strict derivative is the linear map d_x. That is*

$$\lim_{\substack{(x',t'),\,(x'',t'')\to(x,t)\\ (x',t')\neq(x'',t'')}} \frac{\beta(x', t') - \beta(x'', t'') - d_x(x' - x'', t' - t'')}{\|(x' - x'', t' - t'')\|} = 0.$$

We now turn our attention to the problem of characterizing the Clarke subdifferential $\partial^c(f\circ\beta)(x, t)$.

Theorem 6.5. *The Clarke subgradient at (x, t) of a Lorentz invariant function $f\circ\beta$, locally Lipschitz at (x, t), is given by the formulae* *(i) if $x \neq 0$ then*

$$\partial^c(f \circ \beta)(x,t) = \{d_x^*(\gamma_1,\gamma_2) \,|\, (\gamma_1,\gamma_2) \in \partial^c f(\beta(x,t))\};$$

(ii) if $x = 0$ then

$$\partial^c(f \circ \beta)(0,t) = \{d_z^*(\gamma_1,\gamma_2) \,|\, (\gamma_1,\gamma_2) \in \partial^c f(\beta(0,t)),\ z \neq 0\}.$$

Proof. **Case (i)** Suppose that $x \neq 0$. Then, by Lemma 6.4, β is strictly differentiable at (x,t) with strict derivative d_x. Moreover, d_x is a surjective linear map. So we can apply the chain rule for the Clarke subdifferential [4, Theorem 2.3.10], which in our situation holds with equality:

$$\partial^c(f \circ \beta)(x,t) = \partial^c f(\beta(x,t)) \circ d_x.$$

Now, if $(v,p) \in \partial^c(f \circ \beta)(x,t)$ and $(y,r) \in \mathbb{R}^n \times \mathbb{R}$, then there is a subgradient $(\gamma_1,\gamma_2) \in \partial^c f(\beta(x,t))$ such that

$$\langle (v,p),(y,r)\rangle = ((\gamma_1,\gamma_2) \circ d_x)(y,r) = \langle (\gamma_1,\gamma_2), d_x(y,r)\rangle = \langle d_x^*(\gamma_1,\gamma_2),(y,r)\rangle,$$

where the last equality follows by Lemma 5.3. So

$$\partial^c(f \circ \beta)(x,t) \subseteq \{d_x^*(\gamma_1,\gamma_2) | (\gamma_1,\gamma_2) \in \partial^c f(\beta(x,t))\},$$

the other inclusion is now clear.

Case (ii) Suppose that $x = 0$ and define

$$\mathcal{D} := \{d_z^*(\gamma_1,\gamma_2) | (\gamma_1,\gamma_2) \in \partial^c f(\beta(0,t)),\ z \neq 0\}.$$

Two closed, convex sets are equal whenever their support functions are the same. The support function for the set $\operatorname{conv}\mathcal{D}$, evaluated at (y,r), is

$$\begin{aligned}
&\max\{\langle (y,r),(z,s)\rangle \,|\, (z,s) \in \operatorname{conv}\mathcal{D}\}\\
&\quad= \max\{\langle (y,r),(z,s)\rangle \,|\, (z,s) \in \mathcal{D}\}\\
&\quad= \max\{\langle (y,r), d_z^*(\gamma_1,\gamma_2)\rangle \,|\, (\gamma_1,\gamma_2) \in \partial^c f(\beta(0,t)),\ z \neq 0\}\\
&\quad= \max\{\langle d_z(y,r),(\gamma_1,\gamma_2)\rangle \,|\, (\gamma_1,\gamma_2) \in \partial^c f(\beta(0,t)),\ z \neq 0\}\\
&\quad= \max\{\max\{\langle d_z(y,r),(\gamma_1,\gamma_2)\rangle \,|\, (\gamma_1,\gamma_2) \in \partial^c f(\beta(0,t))\} \,|\, z \neq 0\}\\
&\quad= \max\{f^\circ(\beta(0,t); d_z(y,r)) \,|\, z \neq 0\}\\
&\quad= (f \circ \beta)^\circ((0,t);(y,r)),
\end{aligned}$$

where, in the last two equalities, we used (6.3) and Theorem 6.2. By (6.3) again applied to the function $f \circ \beta$ we obtain

$$\operatorname{cl}\operatorname{conv}\mathcal{D} = \partial^c(f \circ \beta)(0,t),$$

because $\partial^c(f \circ \beta)(x,t)$ is a closed convex set [4, Proposition 2.1.2]. The fact that $\operatorname{conv}\mathcal{D} = \mathcal{D}$ follows from Lemma 5.4 part (i)b and the fact that $\mathcal{D}$ is closed follows by the same lemma part (ii)b. □

Corollary 6.6 (Clarke directional derivative). *Let (x,t) be a point in the domain of $f \circ \beta$ and let (y,r) be a direction in $\mathbb{R}^n \times \mathbb{R}$. Then if $x \neq 0$,*

$$(f \circ \beta)^\circ((x,t);(y,r)) = f^\circ(\beta(x,t); d_x(y,r)).$$

Proof. Use again the fact that $(f \circ \beta)^\circ((x,t);(y,r))$ is the support function of $\partial^c(f \circ \beta)(x,t)$, see [4, Proposition 2.1.2]. □

7. Second order properties. Let, in this section, f be twice differentiable at the point (a, b). This means that f is differentiable in a neighbourhood of this point and the first derivative, ∇f, is differentiable again at (a, b). The question that we are going to answer now is whether $g := f \circ \beta$ is twice differentiable at any point (x, t) such that $\beta(x, t) = (a, b)$. Elementary calculus shows that, if $x \neq 0$ then g is twice differentiable. It turns out that this is always the case as we prove in Theorem 7.1. A generalization of Theorem 7.1 and Theorem 7.2 to the setting of formally real Jordan algebras can be found in [25]. Our approach is direct and first appeared in [21].

7.1. Second order differentiability. THEOREM 7.1. *The function $g := f \circ \beta$ is twice differentiable at (x, t) if and only if f is twice differentiable at $\beta(x, t)$. In that case we have*

(i) if $x \neq 0$ then

$$\begin{aligned}
g''_{x_i x_j}(x,t) &= \frac{x_i x_j}{2\|x\|^2}(f''_{11} - f''_{12} - f''_{21} + f''_{22}) + \frac{\delta_{ij}\|x\|^2 - x_i x_j}{\sqrt{2}\|x\|^3}(f'_1 - f'_2),\\
g''_{tx_i}(x,t) &= \frac{x_i}{2\|x\|}(f''_{11} - f''_{12} + f''_{21} - f''_{22}),\\
g''_{x_i t}(x,t) &= \frac{x_i}{2\|x\|}(f''_{11} + f''_{12} - f''_{21} - f''_{22}),\\
g''_{tt}(x,t) &= \frac{1}{2}(f''_{11} + f''_{12} + f''_{21} + f''_{22}),
\end{aligned}$$

where δ_{ij} is 1 if $i = j$ and 0 otherwise;

(ii) if $x = 0$, then

$$\begin{aligned}
g''_{x_i x_j}(0,t) &= \begin{cases} \frac{1}{2}(f''_{11} - f''_{12} - f''_{21} + f''_{22}), & \textit{if } i = j,\\ 0 & \textit{otherwise,} \end{cases}\\
g''_{tx_i}(0,t) &= 0,\\
g''_{x_i t}(0,t) &= 0,\\
g''_{tt}(0,t) &= \frac{1}{2}(f''_{11} + f''_{12} + f''_{21} + f''_{22}).
\end{aligned}$$

All second-order derivatives of f, in both cases, are evaluated at $\beta(x, t)$.

Proof. The 'only if' part follows easily from (1.2).

The verification of part *(i)* is straightforward. For part *(ii)* denote

$$\begin{aligned}
H_{ii} &:= \frac{1}{2}(f''_{11} - f''_{12} - f''_{21} + f''_{22}), \quad \text{for } i = 1, ..., n,\\
H_{tt} &:= \frac{1}{2}(f''_{11} + f''_{12} + f''_{21} + f''_{22}),\\
H &:= \text{Diag}\,(H_{11}, ..., H_{nn}, H_{tt}),
\end{aligned}$$

where the second-order derivatives of f are evaluated at $\beta(0, t)$ and the operator Diag forms a diagonal matrix from its vector argument. Fix an arbitrary sequence $\{h^k\}$ in $\mathbb{R}^n \times \mathbb{R}$ converging to 0 and denote $\bar{h}^k := (h_1^k, ..., h_n^k)^T$. Using Theorem 4.1 we are going to show that the limit of the difference quotient

$$\lim_{k\to\infty} \frac{\|\nabla g(\bar{h}^k, t + h_{n+1}^k) - \nabla g(0,t) - Hh^k\|}{\|h^k\|},$$

is 0. We consider separately each coordinate in the difference quotient. Two cases are necessary: one for the coordinates from 1 to n and one for the $(n+1)^{\text{st}}$ coordinate. The sequence $\{h^k\}$ can be partitioned into two subsequences—one in which $\bar{h}^k = 0$ for all k and one in which $\bar{h}^k \neq 0$ for all k. We will be done if we show that the limit of the difference quotient for each of the two subsequences is zero. That lead us to consider two subcases in each main case.

Case a. Suppose $i \in \{1, ..., n\}$. Then the difference quotient becomes

$$\lim_{k\to\infty} \frac{|g_i'(\bar{h}^k, t + h_{n+1}^k) - g_i'(0,t) - H_{ii}h_i^k|}{\|h^k\|}. \tag{7.1}$$

We use Theorem 4.1 to evaluate the derivatives g_i'. Notice that if $\bar{h}^k = 0$ for all k, then the limit is clearly 0. Thus, suppose $\bar{h}^k \neq 0$ for all k. Then (7.1) becomes

$$\lim_{k\to\infty} \frac{|\frac{\bar{h}^k}{\sqrt{2}\|\bar{h}^k\|}(f_1'(\beta(\bar{h}^k, t+h_{n+1}^k)) - f_2'(\beta(\bar{h}^k, t+h_{n+1}^k))) - \frac{\bar{h}^k}{2}(f_{11}'' - f_{12}'' - f_{21}'' + f_{22}'')|}{\|h^k\|},$$

where the second derivatives of f are evaluated at $\beta(0,t)$. Because f_1' and f_2' exist in a neighbourhood of $\beta(0,t)$ and are differentiable at $\beta(0,t)$ we have

$$f_1'(\beta(\bar{h}^k, t+h_{n+1}^k)) \sim f_1'(\beta(0,t)) + f_{11}''(\beta(0,t))\frac{h_{n+1}^k + \|\bar{h}^k\|}{\sqrt{2}} + f_{12}''(\beta(0,t))\frac{h_{n+1}^k - \|\bar{h}^k\|}{\sqrt{2}},$$

$$f_2'(\beta(\bar{h}^k, t+h_{n+1}^k)) \sim f_2'(\beta(0,t)) + f_{21}''(\beta(0,t))\frac{h_{n+1}^k + \|\bar{h}^k\|}{\sqrt{2}} + f_{22}''(\beta(0,t))\frac{h_{n+1}^k - \|\bar{h}^k\|}{\sqrt{2}},$$

where $\sim$ indicates that the difference of both sides is of order $o(\|h^k\|)$. Because f is symmetric, at the point $\beta(0,t)$ we have $f_1' = f_2'$, $f_{12}'' = f_{21}''$ and $f_{11}'' = f_{22}''$. Substituting the two expansions into the limit shows that it is indeed 0.

Case b. Suppose $i = n+1$. The arguments are analogous to the previous case. We use again Theorem 4.1 to evaluate the derivative g_t' and then substitute f_1' and f_2' with their first order expansions. □

7.2. Continuity of the Hessian. THEOREM 7.2. *The function $g := f \circ \beta$ is twice continuously differentiable at (x,t) if and only if f is such at $\beta(x,t)$.*

Proof. The 'only if' direction is also easy to obtain from (1.2). The 'if' direction is clear in the case when $x \neq 0$. Thus, we suppose that f is twice continuously differentiable at $\beta(0,t)$ and are going to show that for any sequence of vectors $\{(x^k, t^k)\}$ in $\mathbb{R}^n \times \mathbb{R}$ approaching $(0,t)$, the Hessian $\nabla^2 g(x^k, t^k)$ is approaching $\nabla^2 g(0,t)$. Viewing, for a fixed basis, $\nabla^2 g(0,t)$ as a matrix, we are going to prove the convergence for each entry. We again consider two cases and the general situation follows easily from them. If $x^k = 0$ for all k, then the result follows directly from the continuity of $\nabla^2 f$ at the point $\beta(0,t)$, see Theorem 7.1. If $x^k \neq 0$ for all n then from the continuity of $\nabla^2 f$ at the point $\beta(0,t)$ and the formulae given in Theorem 7.1 we have

$$\lim_{k\to\infty} g_{x_i t}''(x^k, t^k) = \lim_{k\to\infty} g_{t x_i}''(x^k, t^k) = 0$$
$$\lim_{k\to\infty} g_{tt}''(x^k, t^k) = g_{tt}''(0,t),$$

where we also used the fact that since f is symmetric $f_{12}'' = f_{21}''$ and $f_{11}'' = f_{22}''$ at the point $\beta(0,t)$. The interesting part is to show $\lim\limits_{k\to\infty} g_{x_i x_j}''(x^k, t^k) = g_{x_i x_j}''(0,t)$. Denote

$$\beta_{+-}^k := \frac{1}{\sqrt{2}}(t^k + \|x^k\|, t^k - \|x^k\|),$$

$$\beta^k_{++} := \frac{1}{\sqrt{2}}(t^k + \|x^k\|, t^k + \|x^k\|),$$
$$\beta^k_{-+} := \frac{1}{\sqrt{2}}(t^k - \|x^k\|, t^k + \|x^k\|).$$

Because f is symmetric $f_1'(\beta^k_{-+}) = f_2'(\beta^k_{+-})$. This allows us to evaluate the following limit using the Mean Value theorem.

$$\begin{aligned}
&\lim_{k\to\infty} \frac{1}{\sqrt{2}\|x^k\|}\big(f_1'(\beta(x^k,t^k)) - f_2'(\beta(x^k,t^k))\big)\\
&\quad = \lim_{k\to\infty} \frac{1}{\sqrt{2}\|x^k\|}\big(f_1'(\beta^k_{+-}) - f_1'(\beta^k_{++}) + f_1'(\beta^k_{++}) - f_1'(\beta^k_{-+})\big)\\
&\quad = \lim_{k\to\infty}\left(- f_{12}''\left(\frac{t^k+\|x^k\|}{\sqrt{2}}, \nu^k\right) + f_{11}''\left(\mu^k, \frac{t^k+\|x^k\|}{\sqrt{2}}\right)\right)\\
&\quad = \frac{1}{2}\big(f_{11}''(\beta(0,t)) - f_{12}''(\beta(0,t)) - f_{21}''(\beta(0,t)) + f_{22}''(\beta(0,t))\big).
\end{aligned}$$

Above, the numbers ν^k and μ^k are between $\frac{t^k-\|x^k\|}{\sqrt{2}}$ and $\frac{t^k+\|x^k\|}{\sqrt{2}}$ and the last equality uses the continuity of $\nabla^2 f$ and the fact that f is symmetric. Using the formula for $g''_{x_ix_j}$ given in Theorem 7.1 we can see that

$$\begin{aligned}
\lim_{k\to\infty} g''_{x_ix_j}(x^k,t^k) &= \frac{\delta_{ij}}{2}\big(f_{11}''(\beta(0,t)) - f_{12}''(\beta(0,t)) - f_{21}''(\beta(0,t)) + f_{22}''(\beta(0,t))\big)\\
&= g''_{x_ix_j}(0,t).
\end{aligned}$$

This concludes the proof. □

7.3. Positive definite Hessian. We begin with a simple lemma and the main result of this subsection follows after it.

LEMMA 7.3. *Suppose that function f is continuously differentiable on an open convex subset of $\mathbb{R}^2$ and is strictly convex there. For any point (a,b) in its domain with $a > b$ we have $f_1'(a,b) > f_2'(a,b)$.*

THEOREM 7.4. *Suppose that f is twice continuously differentiable at $\beta(x,t)$. Then $\nabla^2(f\circ\beta)$ is positive definite at the point (x,t) if and only if $\nabla^2 f$ is positive definite at $\beta(x,t)$.*

Proof. Suppose that $\nabla^2 f(\beta(x,t))$ is positive definite. We use the formulae in Theorem 7.1 to give a matrix representation of the Hessian of $f\circ\beta$. Define the $(n+1)\times 2$ matrix:

$$X := \frac{1}{\sqrt{2}}\begin{pmatrix} \frac{x}{\|x\|} & -\frac{x}{\|x\|} \\ 1 & 1 \end{pmatrix},$$

and the $(n+1)\times(n+1)$ matrix

$$M := \frac{1}{\sqrt{2}\|x\|}\begin{pmatrix} I_n - \frac{xx^T}{\|x\|^2} & 0 \\ 0 & 0 \end{pmatrix},$$

where I_n is the $n\times n$ identity matrix.

Case I. When $x \neq 0$ the Hessian of $f\circ\beta$ can be written as

$$\nabla^2(f\circ\beta)(x,t) = X\nabla^2 f(\beta(x,t))X^T + M\nabla f(\beta(x,t))\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

For any nonzero vector (y,r) we have

$$(y,r)\big(\nabla^2(f\circ\beta)(x,t)\big)(y,r)^T = \frac{1}{2}d_x(y,r)\nabla^2 f(\beta(x,t))d_x(y,r)^T \\ + \frac{1}{\sqrt{2}\|x\|^3}\big(\|y\|^2\|x\|^2 - (x^T y)^2\big)\big(f_1'(\beta(x,t)) - f_2'(\beta(x,t))\big).$$

Using Lemma 7.3 we see that the above expression is strictly positive.

Case II. In the case when $x = 0$, then the Hessian of $f \circ \beta$ is a diagonal matrix and the fact that it is positive definite can be easily seen.

In the other direction the result follows from (1.2). □

The proof of the next corollary is virtually the same as [22, Theorem 7.2], we omit it.

COROLLARY 7.5. *Let C be a symmetric and convex subset of $\mathbb{R}^2$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be twice continuously differentiable function defined on C. Then*

$$\min_{(a,b)\in C} \lambda_{\min}(\nabla^2 f(a,b)) = \min_{(x,t)\in\beta^{-1}(C)} \lambda_{\min}(\nabla^2(f\circ\beta)(x,t)). \tag{7.2}$$

Above, $\lambda_{\min}$ denotes the smallest eigenvalue of the matrix in its argument.

Multiplying both sides of (7.2) by -1 we obtain

$$\max_{(a,b)\in C} \lambda_{\max}(\nabla^2 f(a,b)) = \max_{(x,t)\in\beta^{-1}(C)} \lambda_{\max}(\nabla^2(f\circ\beta)(x,t)).$$

8. The regular and proximal subdifferentials. Given a function $h : \mathbb{R}^m \to [-\infty, +\infty]$ and a point x in $\mathbb{R}^m$ at which h is finite, a vector y of $\mathbb{R}^m$ is called a *regular subgradient* of h at x if

$$h(x+z) \geq h(x) + \langle y, z\rangle + o(\|z\|) \text{ as } z \to 0.$$

The set of regular subgradients is denoted $\hat{\partial} h(x)$ and is called the *regular subdifferential* of h at x. If h is infinite at x then the set $\hat{\partial} h(x)$ is defined to be empty. It is not difficult to show that it is always a closed and convex set, see [20].

A vector y is called a *proximal subgradient* of the function h at x, a point where $h(x)$ is finite, if there exist $\rho > 0$ and $\delta > 0$ such that

$$h(x+z) \geq h(x) + \langle y, z\rangle - \frac{1}{2}\rho\|z\|^2 \quad \text{when } \|z\| \leq \delta.$$

The set of all proximal subgradients will be denoted $\partial_p h(x)$. If h is infinite at x then the set $\partial_p h(x)$ is defined to be empty. It is not difficult to show that it is always a closed and convex set, see [5].

Let now f be the symmetric, bivariate function on $\mathbb{R}^2$ and $g := f \circ \beta$. We are going to derive a formula for $\hat{\partial} g(x,t)$ in terms of $\hat{\partial} f(\beta(x,t))$. The next lemma lists several properties of the map $\beta(x,t)$ that we need. By $\mathbb{R}^n_{\geq}$ we denote the cone of vectors x in $\mathbb{R}^n$ satisfying $x_1 \geq x_2 \geq ... \geq x_n$.

LEMMA 8.1.

(i) For any vector w in $\mathbb{R}^2_{\geq}$ the function $w^T\beta$ is convex and any point (x,t) in $\mathbb{R}^n \times \mathbb{R}$ satisfies $d_x^*(w) \in \partial(w^T\beta)(x,t)$.

(ii) The directional derivative $\beta'((x,t);(y,r))$ is given by

$$\beta'((x,t);(y,r)) = \begin{cases} d_x(y,r), & \text{if } x \neq 0 \\ \beta(y,r), & \text{if } x = 0. \end{cases}$$

(iii) The map β is Lipschitz with global constant 1.
(iv) Given a point (x,t) in $\mathbb{R}^n \times \mathbb{R}$, all vectors (z,s) close to zero satisfy

$$\beta((x,t)+(z,s)) = \beta(x,t) + \beta'((x,t);(z,s)) + O(\|(z,s)\|^2).$$

Proof.
(i) The convexity is elementary. To check the second half we need to verify that

$$w^T\beta(y,r) - w^T\beta(x,t) \geq \langle d_x^*(w_1,w_2),(y-x,r-t)\rangle,$$

which expanded and simplified is equivalent to

$$\frac{w_1-w_2}{\sqrt{2}}(\|y\|-\|x\|) \geq \frac{x^T(y-x)}{\|x\|}\frac{w_1-w_2}{\sqrt{2}}.$$

After cancelation, the last inequality follows from the Cauchy-Schwarz inequality.
(ii) This part is a straightforward verification.
(iii) For any points (x,t) and (z,s) we have

$$\begin{aligned}
&\|\beta((x,t)+(z,s)) - \beta(x,t)\| \\
&\quad = \frac{1}{\sqrt{2}}\|(t+s+\|x+z\|, t+s-\|x+z\|) - (t+\|x\|, t-\|x\|)\| \\
&\quad = \frac{1}{\sqrt{2}}\|(s+\|x+z\|-\|x\|, s-(\|x+z\|-\|x\|))\| \\
&\quad = \sqrt{s^2+(\|x+z\|-\|x\|)^2} \\
&\quad \leq \sqrt{s^2+\|z\|^2} \\
&\quad = \|(z,s)\|.
\end{aligned}$$

(iv) Suppose first that $x \neq 0$. Then using part (ii) of this lemma and several times the Cauchy-Schwarz inequality we get

$$\begin{aligned}
&\|\beta((x,t)+(z,s)) - \beta(x,t) - \beta'((x,t);(z,s))\|^2 \\
&\qquad = \frac{1}{2}\left\|\left(\|x+z\|-\|x\|-\frac{x^Tz}{\|x\|}, -\|x+z\|+\|x\|+\frac{x^Tz}{\|x\|}\right)\right\|^2 \\
&\qquad = \left(\|x+z\|-\|x\|-\frac{x^Tz}{\|x\|}\right)^2 \\
&\qquad = O(\|z\|^4) = O(\|(z,s)\|^4),
\end{aligned}$$

where the penultimate equality holds since $\nabla\|\cdot\|(x) = \frac{x}{\|x\|}$.
The case $x = 0$ is easy. □

Let L be a subset of $\mathbb{R}^m$ and fix a point x in $\mathbb{R}^m$. An element d belongs to the *contingent cone* to L at x, denoted $K(L|x)$, if either $d = 0$ or there is a sequence $\{x^k\}$ in L approaching x with $(x^k - x)/\|x^k - x\|$ approaching $d/\|d\|$. The *negative polar* of a subset H of $\mathbb{R}^m$ is the set

$$H^- = \{y \in \mathbb{R}^m | \langle x,y\rangle \leq 0 \ \forall x \in H\}.$$

We use the following lemmas from [11], see Propositions 2.1 and 2.2 there.

LEMMA 8.2. *Given a function $f : \mathbb{R}^m \to [-\infty, +\infty]$ and a point x^0 in $\mathbb{R}^m$, any regular subgradient of f at x^0 is polar to the contingent cone of the level set $L = \{x \in E : f(x) \le f(x^0)\}$ at x^0; that is*

$$\hat{\partial} f(x^0) \subset (K(L|x^0))^-.$$

LEMMA 8.3. *If the function $f : \mathbb{R}^m \to [-\infty, +\infty]$ is invariant under a subgroup G of $O(m)$, then any point x in $\mathbb{R}^m$ and transformation g in G satisfy $\hat{\partial} f(gx) = g\hat{\partial} f(x)$. Corresponding results hold for the proximal, approximate, horizon and Clarke subgradients (see next sections).*

We define the action of the orthogonal group $O(n)$ on $\mathbb{R}^n \times \mathbb{R}$ by

$$U.(x,t) = (Ux, t), \text{ for every } U \in O(n).$$

For a fixed point (x,t) in $\mathbb{R}^n \times \mathbb{R}$ we define the orbit

$$O(n).(x,t) = \{(Ux,t) | U \in O(n)\}.$$

If $x \neq 0$, this orbit is just a $n-1$ dimensional sphere with radius $\|x\|$ at level t in $\mathbb{R}^n \times \mathbb{R}$. So it is a $n-1$ dimensional manifold and one can easily calculate that its tangent and normal spaces at the point (x,t) are

$$\begin{aligned} T_{(x,t)}(O(n).(x,t)) &= \{(y,0) | y^T x = 0\}, \text{ and} \\ N_{(x,t)}(O(n).(x,t)) &= \{(ax,b) | (a,b) \in \mathbb{R}^2\}. \end{aligned}$$

If $x = 0$ then

$$\begin{aligned} T_{(0,t)}(O(n).(0,t)) &= \{0\}, \text{ and} \\ N_{(0,t)}(O(n).(0,t)) &= \mathbb{R}^{n+1}. \end{aligned}$$

Now, using these observations and Lemma 8.2 we can say more about $\hat{\partial}(f \circ \beta)(x,t)$ in the case when $x \neq 0$.

LEMMA 8.4. *If $x \neq 0$ and $(y,r) \in \hat{\partial}(f \circ \beta)(x,t)$ then $(y,r) = (ax, r)$ for some $a \in \mathbb{R}$.*

Proof. If $(y,r) \in \hat{\partial}(f \circ \beta)(x,t)$ then by Lemma 8.2 we have

$$\begin{aligned} (y,r) &\in (K(\{(z,s) | (f \circ \beta)(z,s) \le (f \circ \beta)(x,t)\} | (x,t)))^- \\ &\subset (K(O(n).(x,t) | (x,t)))^- \\ &= N_{(x,t)}(O(n).(x,t)). \end{aligned}$$

The claim follows from the expression for the normal space above. ☐

The following is the main theorem of this section.

THEOREM 8.5. *The regular subdifferential of any Lorentz invariant function $f \circ \beta$ at the point (x,t) is given by the formulae*

(i) if $x \neq 0$ then

$$\hat{\partial}(f \circ \beta)(x,t) = \{d_x^*(\gamma_1, \gamma_2) | (\gamma_1, \gamma_2) \in \hat{\partial} f(\beta(x,t))\};$$

(ii) if $x = 0$ then

$$\hat{\partial}(f \circ \beta)(0,t) = \{d_z^*(\gamma_1, \gamma_2) | (\gamma_1, \gamma_2) \in \hat{\partial} f(\beta(0,t)), \, z \neq 0\}.$$

Similar formulae hold for the proximal subdifferential.

Proof. **Case (i)**. This case follows immediately from the chain rule [20, Exercise 10.7].

Case (ii). Let $x = 0$. We are going to show that

$$\hat{\partial}(f \circ \beta)(0, t) = \{(y, r) | d_z(y, r) \in \hat{\partial} f(\beta(0, t)), \forall z \neq 0\}.$$

The stated version follows from Lemma 5.4 part (i)c.

Suppose $(y, r) \in \hat{\partial}(f \circ \beta)(0, t)$, let $z := (z_1, z_2) \in \mathbb{R}^2$ be small and let w be an arbitrary nonzero vector in $\mathbb{R}^n$. Then

$$\begin{aligned} f(\beta(0, t) + (z_1, z_2)) &= (f \circ \beta)\Bigg((0, t) + \Bigg(\frac{w}{\|w\|} \frac{z_1 - z_2}{\sqrt{2}}, \frac{z_1 + z_2}{\sqrt{2}}\Bigg)\Bigg) \\ &\geq (f \circ \beta)(0, t) + \frac{w^T y}{\|w\|} \frac{z_1 - z_2}{\sqrt{2}} + r \frac{z_1 + z_2}{\sqrt{2}} + o(\|z\|) \\ &= f(\beta(0, t)) + \langle d_w(y, r), (z_1, z_2) \rangle + o(\|z\|). \end{aligned}$$

Consequently $d_w(y, r) \in \hat{\partial} f(\beta(0, t))$ for all $w \neq 0$.

In the opposite direction suppose that $d_w(y, r) \in \hat{\partial} f(\beta(0, t))$ for all $w \neq 0$. If $y = 0$ then for any vector $(z, s) \in \mathbb{R}^n \times \mathbb{R}$ close to 0 we have

$$\begin{aligned} (f \circ \beta)((0, t) + (z, s)) &= f(\beta(0, t) + (\beta((0, t) + (z, s)) - \beta(0, t))) \\ &\geq f(\beta(0, t)) + \langle d_w(0, r), (\beta((0, t) + (z, s)) - \beta(0, t)) \rangle + o(\|(z, s)\|) \\ &= f(\beta(0, t)) + rs + o(\|(z, s)\|) \\ &= (f \circ \beta)(0, t) + \langle (0, r), (z, s) \rangle + o(\|(z, s)\|). \end{aligned}$$

so $(0, r) \in \hat{\partial}(f \circ \beta)(0, t)$.

If $y \neq 0$ then for $w = y$ we have $d_y(y, r) \in \hat{\partial} f(\beta(0, t))$. Let $(z, s) \in \mathbb{R}^n \times \mathbb{R}$ be a vector close to 0. Then

$$\begin{aligned} (f \circ \beta)((0, t) + (z, s)) &= f(\beta(0, t) + (\beta((0, t) + (z, s)) - \beta(0, t))) \\ &\geq f(\beta(0, t)) + \langle d_y(y, r), (\beta((0, t) + (z, s)) - \beta(0, t)) \rangle + o(\|(z, s)\|) \\ &= f(\beta(0, t)) + \|y\| \|z\| + rs + o(\|(z, s)\|) \\ &\geq (f \circ \beta)(0, t) + \langle (y, r), (z, s) \rangle + o(\|(z, s)\|). \end{aligned}$$

Consequently $(y, r) \in \hat{\partial}(f \circ \beta)(0, t)$.

The proof for the proximal subdifferential is essentially identical. □

9. The approximate and horizon subdifferential. Given a function $h : \mathbb{R}^m \to [-\infty, +\infty]$ and a point x in $\mathbb{R}^m$ at which h is finite, a vector y of $\mathbb{R}^m$ is called an *approximate subgradient* of h at x if there is a sequence of points $\{x^k\}$ in $\mathbb{R}^m$ approaching x with values $h(x^k)$ approaching $h(x)$ and a sequence of regular subgradients y^k in $\hat{\partial} h(x^k)$ approaching y. The set of all approximate subgradients is called the *approximate subdifferential* $\partial h(x)$. A vector y is called a *horizon subgradient* if either $y = 0$ or there is a sequence of points $\{xv\}$ in $\mathbb{R}^m$ approaching x with values $h(x^k)$ approaching $h(x)$, a sequence $\{t^k\}$ of reals decreasing to zero and a sequence of regular subgradients y^k in $\hat{\partial} h(x^k)$ for which $t^k y^k$ approaches y. The set of all horizon subgradients is denoted $\partial^\infty h(x)$. If h is infinite at x then the set $\partial h(x)$ is defined to be empty and $\partial^\infty h(x)$ to be $\{0\}$.

Recall that we used the same notation, $\partial h(x)$, for the convex subgradient when h is a convex function. There is no danger of confusion because the subdifferentials coincide when h is a proper, convex function, see [20, Proposition 8.12].

THEOREM 9.1. *The approximate subdifferential of any Lorentz invariant function $f \circ \beta$ at the point (x,t) is given by the formulae:*

(i) if $x \neq 0$ then

$$\partial(f \circ \beta)(x,t) = \{d_x^*(a,b) \,|\, (a,b) \in \partial f(\beta(x,t))\};$$

(ii) if $x = 0$ then

$$\partial(f \circ \beta)(0,t) = \{d_z^*(a,b) \,|\, (a,b) \in \partial f(\beta(0,t)),\ z \neq 0\}.$$

Similar formulae hold for the horizon subgradient.

Proof. **Part (i). $\mathbf{x \neq 0}$**. This case follows immediately from the chain rule [20, Exercise 10.7].

Part (ii). $\mathbf{x = 0}$. Suppose $(y,r) \in \partial(f \circ \beta)(0,t)$. By definition, there is a sequence of points $\{(x^k,t^k)\}$ approaching $(0,t)$ with $(f \circ \beta)(x^k,t^k)$ approaching $(f \circ \beta)(0,t)$ and a sequence of regular subgradients $(y^k,r^k) \in \hat{\partial}(f \circ \beta)(x^k,t^k)$ approaching (y,r).

Case 1.a. Suppose $x^k = 0$ for all k. Then Theorem 8.5 says that $(y^k,r^k) = d_{z^k}^*(a_k,a_k)$ such that $(a_k,a_k) \in \hat{\partial} f(\beta(0,t^k))$, for some $z^k \neq 0$. Since (y^k,r^k) approaches (y,r) we get that $y = 0$ and $a_k \to a := r/\sqrt{2}$. So $(0,r) = (0,\sqrt{2}a) = d_z^*(a,a)$ for any $z \neq 0$ and $(a,a) \in \partial f(\beta(0,t))$.

Case 1.b. Suppose $x^k \neq 0$ for all k. Then Theorem 8.5 says that $(y^k,r^k) = d_{x_q}^*(a_k,b_k)$ such that $(a_k,b_k) \in \hat{\partial} f(\beta(x^k,t^k))$. Let us choose a subsequence k' for which $x^{k'}/\|x^{k'}\|$ converges to a unit vector z. Then we have that $|a_{k'} - b_{k'}|$ approaches $\sqrt{2}\|y\|$ and $a_{k'} + b_{k'}$ approaches $\sqrt{2}r$, that is, $(a_{k'},b_{k'})$ is bounded sequence so if necessary we may choose a convergent subsequence k''. Then $(a_{k''},b_{k''}) \to (a,b) \in \partial f(\beta(0,t))$ and $(y,r) = d_z^*(a,b)$.

Case 1.c. Suppose the sequence x^k has infinitely many elements that are equal to 0 and infinitely many elements that are not equal to 0. Let $\{x^k\} = \{x^{k'}\} \cup \{x^{k''}\}$, where $x^{k'} \neq 0$ and $x^{k''} = 0$. We now choose any of the subsequences k' or k'' and apply the corresponding subcase above.

To show the opposite inclusion, suppose that $(y,r) = d_z^*(a,b)$ for some $(a,b) \in \partial f(\beta(0,t))$ and some $z \neq 0$. By the definition of approximate subgradients there is a sequence (c_k,d_k) approaching $\beta(0,t)$, with $f(c_k,d_k)$ approaching $f(\beta(0,t))$ and a sequence of regular subgradients (a_k,b_k) approaching (a,b) and such that $(a_k,b_k) \in \hat{\partial} f(c_k,d_k)$. We have three possible cases.

Case 2.a. Suppose first that there is an infinite subsequence k' such that $c_{k'} > d_{k'}$ for all k'. Then $d_z^*(c_{k'},d_{k'})$ approaches $d_z^*(\beta(0,t)) = (0,t)$ with $f(c_{k'},d_{k'}) = (f \circ \beta)(d_z^*(c_{k'},d_{k'}))$ approaching $f(\beta(0,t)) = (f \circ \beta)(0,t)$ and regular subgradients $(a_{k'},b_{k'}) \in \hat{\partial} f(\beta(d_z^*(c_{k'},d_{k'})))$. If we set $z^{k'} := \frac{z}{\|z\|}\frac{c_{k'}-d_{k'}}{\sqrt{2}}$, then Theorem 8.5 says that $d_{z^{k'}}^*(a_{k'},b_{k'}) \in \hat{\partial}(f \circ \beta)(d_z^*(c_{k'},d_{k'}))$. Notice that $z^{k'}/\|z^{k'}\|$ converges to $z/\|z\|$, so $d_{z^{k'}}^*(a_{k'},b_{k'})$ approaches $d_z^*(a,b) = (y,r)$, so (y,r) is in $\partial(f \circ \beta)(0,t)$.

Case 2.b. There is an infinite subsequence k' such that $c_{k'} < d_{k'}$ for all k'. We are going to revert to the previous case. We have that $(y,r) = d_{-z}^*(b,a)$ where $(b,a) \in \partial f(\beta(0,t))$ (see Lemma 8.3) and $z \neq 0$. We are given also that the sequence $(d_{k'},c_{k'})$ approaches $\beta(0,t)$, with $f(d_{k'},c_{k'})$ approaching $f(\beta(0,t))$ and the sequence of regular subgradients $(b_{k'},a_{k'})$ approaches (b,a) and is such that $(b_{k'},a_{k'}) \in \hat{\partial} f(d_{k'},c_{k'})$ (by Lemma 8.3 again). The rest is analogous to the previous case.

Case 2.c. Suppose finally that there is an infinite subsequence k' such that $c_{k'} = d_{k'}$ for all k'. Then $d_z^*(c_{k'}, d_{k'})$ approaches $d_z^*(\beta(0,t)) = (0,t)$, with $f(c_{k'}, d_{k'}) = (f \circ \beta)(d_z^*(c_{k'}, d_{k'}))$ approaching $f(\beta(0,t)) = (f \circ \beta)(0,t)$ and regular subgradients $(a_{k'}, b_{k'}) \in \hat{\partial} f(\beta(d_z^*(c_{k'}, d_{k'})))$. But then by Theorem 8.5 we have that $d_z^*(a_{k'}, b_{k'}) \in \hat{\partial}(f \circ \beta)(0, \sqrt{2} d_{q'})$. Since $d_z^*(a_{k'}, b_{k'})$ approaches $d_z^*(a,b)$, we are done.

The proof of the formulae for the horizon subgradient is analogous. □

10. Clarke subgradients - the lower semicontinuous case. A function h is caller *lower semicontinuous* if its epigraph $\operatorname{epi} h = \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} | h(x) \le \alpha\}$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}$. Let $C \subset \mathbb{R}^n$ and $\bar{x} \in C$. A vector $v \in \mathbb{R}^n$ is a *regular normal* to C at $\bar{x}$, written $v \in \hat{N}_C(\bar{x})$, if $\limsup\limits_{x \to \bar{x}} \frac{\langle v, x-\bar{x}\rangle}{|x-\bar{x}|} \le 0$. It is a *normal vector* to C at $\bar{x}$, written $v \in N_C(\bar{x})$, if there is a sequence of points x^k in C approaching $\bar{x}$ and a sequence of regular normals v^k in $\hat{N}_C(x^k)$ approaching v. The set of *Clarke subgradients* of a function h at $\bar{x}$, $\partial^c h(\bar{x})$, is defined by

$$\partial^c h(\bar{x}) = \{v | (v, -1) \in \operatorname{cl}\operatorname{conv} N_{\operatorname{epi} h}(\bar{x}, h(\bar{x}))\}.$$

It can be shown that if h is locally Lipschitz around $\bar{x}$ then this definition coincides with the definition given in Section 6, so there is no danger of confusion, see [20, Theorem 9.13 (b) and Theorem 8.49].

By [20, Theorem 8.9], if h is lower semicontinuous around $\bar{x}$ the following formula holds:

$$N_{\operatorname{epi} h}(\bar{x}, h(\bar{x})) = \{\lambda(v,-1) \,|\, v \in \partial h(\bar{x}),\ \lambda > 0\} \cup \{(v,0) \,|\, v \in \partial^\infty h(\bar{x})\}.$$

The following lemma can be found in [17, Proposition 2.6]. For an independent proof see [15, Lemma 4.1].

LEMMA 10.1. *If h is lower semicontinuous around $\bar{x}$ we have the representation*

$$\partial^c h(\bar{x}) = \operatorname{cl}(\operatorname{conv} \partial h(\bar{x}) + \operatorname{conv} \partial^\infty h(\bar{x})).$$

In particular, when the cone $\partial^\infty h(\bar{x})$ is pointed we have simpler

$$\partial^c h(\bar{x}) = \operatorname{conv} \partial h(\bar{x}) + \operatorname{conv} \partial^\infty h(\bar{x}).$$

It is easy to see that f is lower semicontinuous if and only if $f \circ \beta$ is. Our final result is the following theorem.

THEOREM 10.2. *The Clarke subdifferential of any lower semicontinuous, Lorentz invariant function $f \circ \beta$ at the point (x,t) is given by the formulae:*

(i) if $x \ne 0$ then

$$\partial^c (f \circ \beta)(x,t) = \{d_x^*(a,b) \,|\, (a,b) \in \partial^c f(\beta(x,t))\};$$

(ii) if $x = 0$ then

$$\partial^c (f \circ \beta)(0,t) = \{d_z^*(a,b) \,|\, (a,b) \in \partial^c f(\beta(0,t)),\ z \ne 0\}.$$

Proof. Suppose first that $x = 0$. Let $A := \partial f(\beta(x,t))$ and $B := \partial^\infty f(\beta(x,t))$. Using Lemma 5.4 and Lemma 10.1 we get

$$\partial^c (f \circ \beta)(x,t) = \operatorname{cl}(\operatorname{conv} \partial (f \circ \beta)(x,t) + \operatorname{conv} \partial^\infty (f \circ \beta)(x,t))$$

$$
\begin{aligned}
&= \operatorname{cl}(\operatorname{conv}\mathcal{D}(A) + \operatorname{conv}\mathcal{D}(B)) \\
&= \operatorname{cl}(\mathcal{D}(\operatorname{conv} A) + \mathcal{D}(\operatorname{conv} B)) \\
&= \operatorname{cl}\mathcal{D}(\operatorname{conv} A + \operatorname{conv} B) \\
&= \mathcal{D}(\operatorname{cl}(\operatorname{conv} A + \operatorname{conv} B)) \\
&= \mathcal{D}(\partial^c f(\beta(x,t)).
\end{aligned}
$$

The case $x \neq 0$ is analogous. □

ACKNOWLEDGEMENTS: I am very grateful to an anonymous referee for his very careful reading of the manuscript and generously providing many useful comments.

REFERENCES

[1] M. Baes, Spectral functions on Jordan algebras: differentiability and convexity properties, *CORE Discussion Paper*, http://www.core.ud.ac.be/services/psfiles/dp04/dp2004_16.pdf, 2004.

[2] H. H. Bauschke, O. Güler, A. S. Lewis and H. S. Sendov, Hyperbolic polynomials and convex analysis, *Canad. J. Math.*, 53 (2001), pp. 470–488.

[3] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization*, Springer, New York,CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, second edition, 2006.

[4] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[5] F. H. Clarke, YU. S. Ledyaev, R. J. Stern and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer-Verlag, New York, Inc., 1998.

[6] O. Güler, Hyperbolic polynomials and interior point methods for convex programming, *Math. Oper. Res.*, 22 (1997), pp. 350–377.

[7] J.-B. Hiriart-Urruty and A. S. Lewis, The Clarke and Michel-Penot subdifferentials of the eigenvalues of a symmetric matrix, *Comput. Optim. Appl.*, 13 (1999), pp. 13-23.

[8] A. S. Lewis, Convex analysis on the Hermitian matrices, *SIAM J. Optim.*, 6 (1996), pp. 164–177.

[9] A. S. Lewis, Derivatives of spectral functions, *Math. Oper. Res.*, 21 (1996), 576–588.

[10] A. S. Lewis, The convex analysis of unitarily invariant matrix functions, *J. Convex Anal.*, 2 (1994), pp. 173–183.

[11] A. S. Lewis, Nonsmooth analysis of eigenvalues, *Math. Program.*, 84 (1999), pp. 1–24.

[12] A. S. Lewis, P. A. Parrilo and M. V. Ramana, The Lax conjecture is true, *Proc. Amer. Math. Soc.*, 133 (2005), pp. 2495–2499.

[13] A. S. Lewis and H. S. Sendov, Self-concordant barriers for hyperbolic means, *Math. Program.*, 91 (2001), pp. 1–10.

[14] A. S. Lewis and H. S. Sendov, Nonsmooth analysis of singular values. I. Theory, *Set-Valued Anal.*, 13 (2005), pp. 213–241.

[15] A. S. Lewis and H. S. Sendov, Nonsmooth analysis of singular values. II. Applications, *Set-Valued Anal.*, 13 (2005), pp. 243–264.

[16] M. S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, Applications of second-order cone programming, *Linear Algebra Appl.*, 284 (1998), pp. 193–228.

[17] B. S. Mordukhovich, *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988.

[18] J. Renegar, Hyperbolic programs and their derivative relaxations, *Found. Comput. Math.*, (2006), pp. 59–79.

[19] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.

[20] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer-Verlag, Berlin, 1998.

[21] H. S. Sendov, Variational Spectral Analysis, *University of Waterloo, Ph.D. Thesis*, 2000.

[22] A. S. Lewis and H. S. Sendov, Quadratic expansions of spectral functions, *Linear Algebra Appl.*, 340 (2002), pp. 97–121.

[23] X. D. Chen, D. Sun and J. Sun, Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems, *Comput. Optim. Appl.*, 25 (2003), pp. 39–56.

[24] J. S. Pang, D. Sun and J. Sun, Semismooth homeomorphisms and strong stability of semidefinite and Lorentz cone complementarity problems, *Math. Oper. Res.*, 28 (2003), pp. 39–63.

[25] D. Sun and J. Sun, Löwner's operator and spectral functions in Euclidean Jordan algebras, submitted to *Math. Oper. Res.*, http://www.math.nus.edu.sg/~matsundf/SS4_Dec24.pdf