MATH. SCAND. 41 (1977), 159–174

ON SECOND DERIVATIVES OF CONVEX FUNCTIONS[1]

R. M. DUDLEY

Abstract.

A Schwartz distribution T on R^k is a convex function iff its second derivative D^2T is a nonnegative $k \times k$ matrix-valued Radon measure μ. Such a μ is absolutely continuous with respect to $(k-1)$-dimensional Hausdorff measure. Neither Lebesgue decomposition of measures nor Riesz decomposition of subharmonic functions preserves convexity. Pointwise second derivatives are also considered.

1. Introduction.

L. Schwartz [15, p. 54] showed that a distribution on R^1 is a convex function iff its second derivative is a nonnegative Radon measure. In that case (but not for $k>1$, as will be seen) every Radon measure $\mu \geqq 0$ is the second derivative of a convex function.

We recall some definitions. Given a subset A of a finite-dimensional real vector space V such that A is included in the closure of its interior, $\mathscr{D}(A)$ will denote the space of all C^∞ real-valued functions on V with compact support included in A. Given a function f on a subset of V and $v \in V$, the directional derivative of f at x in the direction v is defined (when it exists) by

$$D_v f(x) = \lim_{h \downarrow 0} (f(x+hv)-f(x))/h .$$

We say $f_n \to f$ in $\mathscr{D}(A)$ iff there is a compact $K \subset A$ such that all f_n have supports in K, and for any m and $v(1),\ldots,v(m) \in V$,

$$D_{v(1)} \ldots D_{v(m)}(f_n - f) \to 0$$

uniformly on A. Then the space of distributions $\mathscr{D}'(A)$ is the set of all real linear forms T on $\mathscr{D}(A)$ such that $T(f_n) \to T(f)$ whenever $f_n \to f$ in $\mathscr{D}(A)$.

If we choose a translation-invariant (Lebesgue, Haar) measure dx on V, then any locally integrable function f defines a distribution $[f]$ by

---

[1] Support of this research by the Statens Naturvidenskabelige Forskningsråd (Denmark) is gratefully acknowledged.

Received June 24, 1976.

$$[f](g) = \int fg\,dx, \qquad g \in \mathscr{D}(A).$$

We will say $T_n \to T$ in $\mathscr{D}'(A)$ iff for every $g \in \mathscr{D}(A)$, $T_n(g) \to T(g)$.

Given an open set U, a *Radon measure* on U is a set function defined on all Borel sets with compact closure included in U, and countably additive on the Borel subsets of each compact $K \subset U$.

Let U be a connected open set in V. A measurable function $f\colon U \to \mathsf{R}$ is called *convex* iff

$$f(tx+(1-t)y) \leqq tf(x)+(1-t)f(y) \qquad \text{for } 0 \leqq t \leqq 1$$

whenever $tx+(1-t)y \in U$ for all $t \in [0,1]$.

If μ is a measure absolutely continuous with respect to ν ($\nu(B)=0$ implies $\mu(B)=0$), we write $\mu \prec \nu$. If μ and ν are singular we write $\mu \perp \nu$. Lebesgue measure on R^k will be called λ^k.

We will define Hausdorff measures [9, pp. 169–171]. The *diameter* of a set S is

$$\operatorname{diam} S := \sup\{|x-y| : x, y \in S\}.$$

Let $c(m) := \pi^{m/2}/(2^m \Gamma(1+(m/2)))$. If $A \subset R^k$ and $\delta > 0$, let

$$\varphi_{m,\delta}(A) := c(m) \inf\left\{ \sum_{S \in G} (\operatorname{diam} S)^m : G \text{ countable}, \right.$$

$$\left. A \subset \bigcup_{S \in G} S, \text{ and } \operatorname{diam} S \leqq \delta \ \forall S \in G \right\}.$$

Let

$$H_m(A) := \sup_{\delta > 0} \varphi_\delta(A) = \lim_{\delta \downarrow 0} \varphi_\delta(A).$$

Then for $0 \leqq m \leqq k$, H_m is a regular measure on the Borel sets of R^k, called m-dimensional *Hausdorff measure*. $H_k = \lambda^k$ on R^k [9, p. 174].

For any real-valued signed measure μ we have a Jordan decomposition $\mu = \mu^+ - \mu^-$. Let $|\mu| := \mu^+ + \mu^-$.

Let $(V, |\cdot|)$ and $(W, \|\cdot\|)$ be two normed linear spaces. Let $L(V, W)$ denote the set of all continuous linear maps of V into W. For a function f from a subset U of V into W, the (Fréchet) derivative $f'(x)$ at x is defined (when it exists) as an element of $L(V, W)$ such that

$$\|f(y)-f(x)-f'(x)(y-x)\| = o(|y-x|)$$

as $y \to x \in U$, $y \in U$. For V and W finite-dimensional, as they will be throughout this paper, the choice of norm does not matter. We use the usual Euclidean norm.

If f is convex on an open set $U \subset V$, $D_v f(x)$ exists for all $x \in U$ and $v \in V$. If $V = \mathsf{R}^k$ and all the first partial derivatives $\partial f/\partial x_j$, $j = 1,\ldots,k$, exist at a point x (i.e. if $D_{v(j)} f(x) = -D_{-v(j)} f(x)$ where $v(j)$ is the jth unit vector), then $f'(x)$ exists and its components are $\partial f/\partial x_j$ [13, p. 101].

On each line parallel to the x_j axis, $\partial f/\partial x_j$ exists except at most on a countable set. Let

$$E := \{x : f'(x) \text{ is undefined}\} .$$

Then $\lambda^k(E) = 0$. In fact, $E \subset \bigcup_{n \geqq 1} K_n$ where K_n are compact and $H_{k-1}(K_n) < \infty$ (Anderson and Klee [3, Theorem 3.1 p. 353]).

If f is convex, it is locally Lipschitzian [14, p. 86 Theorem 10.4], [13, p. 93]. In other words, $D_v f(x)$ is bounded for x and v in compact sets, so $f'(x)$ is locally bounded (cf. the proof of Theorem 2.1, Claim IV below). This fact will be used several times in the sequel.

The second Fréchet derivative is defined by taking limits, if they exist, as $y \to x \in U \setminus E$, $y \in U \setminus E$. (For $y \in E$, we could also use any element of the subgradient at y [14] in place of $f'(y)$.) Even on R^1, there are continuous convex functions f such that E is dense. Then f'' would be nowhere defined if the definition was not adapted, e.g. as above. In this sense existence of $f''(x)$ a.e. for f convex on R^k was proved by Busemann and Feller [7] for $k = 2$ and by A. D. Alexandrov [2] for $k \geqq 3$.

Second partial derivatives $\partial^2 f/\partial x_i \partial x_j$ will also be defined as limits through the set where $\partial f/\partial x_j$ exists. Their matrix for $1 \leqq i, j \leqq k$ is called the *Hessian* wherever it exists.

2. The set of convex functions is closed.

THEOREM 2.1. *For any connected open $U \subset V$, $\{[f] : f$ convex on $U\}$ is sequentially closed, i.e. if $[f_n] \to T \in \mathscr{D}'(U)$, f_n convex, then $T = [f]$ for a unique convex f.*

PROOF. Any convex g is continuous (e.g. [13, p. 93]). Thus $[g]$ is a distribution. Convexity is a local property, i.e. if g is convex on open sets U_j for each j, it is convex on their union. So we may assume U is convex.

CONVENTION. In what follows, *tasuma* means "taking a subsequence, we may assume ...".

Let M be a countable dense set in U. Tasuma f_n converge pointwise on M to an extended real-valued function f. Let $f := \limsup_{n \to \infty} f_n$ on U.

CLAIM I. *$f(x) < +\infty$ for all $x \in U$.*

PROOF. If not, tasuma there is a $p \in U$ such that $f_n(p) \to +\infty$. We take a k-simplex $S \subset U$, i.e. a convex hull of $k+1$ points such that none is in the linear variety spanned by the others, with one vertex at p. Then tasuma there is a vertex q of S such that

1) $f_n(q) = \max(f_n(s) : s \in S)$ for all n.

Let Y be the reflection of S in the point q (we may assume $Y \subset U$). Then by convexity, $f_n(t) \geqq f_n(q)$ for all t in Y, so $f_n \to +\infty$ uniformly on Y, contradicting $[f_n] \to T$ (take $g \in \mathscr{D}(Y)$ with $g \geqq 0$, not identically 0). So Claim I is proved.

CLAIM II. *$f(x) > -\infty$ for all $x \in U$.*

PROOF. If not, then $f_n(u) \to -\infty$ for some $u \in U$. Take another simplex $P \subset U$ with u in the interior of P. Since $\limsup f_n < +\infty$ at each vertex of P, convexity implies that for any neighborhood W of u with $\bar{W}$ included in the interior of P, and any $N < +\infty$, $f_n(w) < -N$ for all $w \in W$ if n is large. Again this contradicts convergence of $[f_n]$, and proves Claim II.

So f is finite valued, and convex, hence continuous.

CLAIM III. *$f_n \to f$ everywhere on U.*

PROOF. If not, let $\liminf f_n(v) < f(v)$. Tasuma $f_n(v) \to y < f(v)$. Then $\limsup f_n$ is convex, finite valued by I and II, hence continuous, contradicting $f_n \to f$ on M, proving III.

CLAIM IV. *$f_n \to f$ uniformly on compact subsets of U.*

PROOF. If not, take $x_n \to x \in U$ with $f_n(x_n) \not\to f(x)$. Tasuma

$$f_n(x_n) \to y \in [-\infty, \infty], \qquad y \neq f(x),$$

so

$$|f_n(x_n) - f_n(x)|/|x_n - x| \to +\infty .$$

We can assume the x_n are all in some simplex $S \subset U$ with a vertex at x. If $y > f(x)$, then the f_n are unbounded on S, specifically on segments from x through x_n to the opposite face of S, by convexity. But this contradicts boundedness of f_n on the vertices of S. So $y < f(x)$. Let Y be the reflection of S through x. We can assume $Y \subset U$. Convexity implies that f_n are unbounded on Y, again a contradiction, proving Claim IV.

By IV, $T = [f]$, proving Theorem 2.1.

In this paper we do not need a topology on $\mathscr{D}'$ except for convergence of sequences. The usual topology on $\mathscr{D}'$ can be defined as the finest locally convex topology consistent with the convergence of sequences defined above. For a fixed compact $K \subset U$, $\mathscr{D}(K)$ with relative topology is a metrizable Montel space. Thus by a result of J. H. Webb [17, Prop. 5.7] any sequentially closed set in $\mathscr{D}'(K)$ is closed. Since convexity is a local property, $\{[f] : f \text{ convex}\}$ is closed in $\mathscr{D}'(U)$, although not every sequentially closed set in $\mathscr{D}'(U)$ is closed [8].

Now let W be a finite-dimensional real vector space with usual topology. Let $\mathscr{D}'(U, W)$ be the set of all (sequentially) continuous linear maps from $\mathscr{D}(U)$ into W. For any $f \in \mathscr{D}(U)$, $v \to D_v f$ is linear from V into $\mathscr{D}(U)$. Given any $T \in \mathscr{D}'(U, W)$, we define its (Fréchet) derivative DT by

$$(DT)(f)(v) := -T(D_v f).$$

Then $DT \in \mathscr{D}'(U, L(V, W))$, where $L(V, W)$ is the vector space of all linear maps from V into W. We also write

$$(D_v T)(f) := (DT)(f)(v).$$

In particular, given a coordinate system $(x_1, \ldots, x_k)$ on V, we have

$$(\partial T/\partial x_i)(f) := -T(\partial f/\partial x_i),$$

the usual definition of partial derivatives of a distribution.

If $T \in \mathscr{D}'(U, \mathsf{R})$, the second derivative $D^2 T := D(DT)$ maps $\mathscr{D}(U)$ into $L(V, V')$, where $V' := L(V, \mathsf{R})$. Here $L(V, V')$ can be identified with the space $B(V)$ of all bilinear forms b: $V \times V \to \mathsf{R}$, via $b(v, w) = a(v)(w)$ for $a \in L(V, V')$, $v, w \in V$.

3. Positive second derivatives.

Let $B^+(V)$ denote the set of all nonnegative symmetric bilinear forms on $V \times V$, that is, the set of all $b \in B(V)$ such that $b(v, v) \geqq 0$ and $b(v, w) = b(w, v)$ for all $v, w \in V$. If $S \in \mathscr{D}'(U, B(V))$ and $S(g) \in B^+(V)$ for each $g \in \mathscr{D}(U)$ with $g \geqq 0$, we write $S \gg 0$.

THEOREM 3.1. *Given $T \in \mathscr{D}'(U, \mathsf{R})$, $T = [f]$ for some convex f iff $D^2 T \gg 0$.*

Before proving the theorem we will give some alternate forms of it. For any finite-dimensional real vector space Y, a Y-valued *Radon measure* μ (as defined in section 1) defines a distribution $[\mu] \in \mathscr{D}'(U, Y)$ by $[\mu](g) := \int g \, d\mu$. Then:

PROPOSITION 3.2. *$A \gg 0$ iff $A = [\mu]$ for some Radon measure μ with values in $B^+(V)$.*

PROOF. The "if" part is immediate. For the converse, if $A \gg 0$, and K is a compact subset of U, there is an $h \in \mathscr{D}(U)$ with $h \equiv 1$ on K and $h \geqq 0$ [15, p. 22]. Then for $f, g \in \mathscr{D}(K)$,

$$-(\sup|f-g|)h \leqq f-g \leqq (\sup|f-g|)h\,.$$

Thus A is continuous on $\mathscr{D}(K)$ for the supremum norm, so $A=[\mu]$ for some Radon measure μ [15, p 25]. Then by dominated convergence, $\mu(C) \in B^+(V)$ for any compact $C \subset U$, and hence for any Borel set C with compact closure in U.

In terms of a basis of V, giving coordinates $x_1, \ldots, x_k$, elements of $B^+(V)$ become nonnegative definite symmetric matrices. Thus the following is equivalent to Theorem 3.1:

THEOREM 3.3. *Given $T \in \mathscr{D}'(U, \mathsf{R})$, $T=[f]$ for some convex f iff $\{\partial^2 T/\partial x_i \partial x_j\}$ is a matrix $\{\mu_{ij}\}$ of real-valued Radon measures μ_{ij} on U, $i, j=1, \ldots, k$, such that for each Borel set B with compact closure included in U, $\{\mu_{ij}(B)\}$ is a nonnegative definite matrix.*

PROOF. We use regularization [15, pp. 165–169]. Let $\varphi_n \in \mathscr{D}(V)$ be an approximate identity, $[\varphi_n] \to \delta_0$ where $\delta_0(f) := f(0)$, with $\varphi_n \geqq 0$ and the support of φ_n decreasing to $\{0\}$. We take the convolutions $T*\varphi_n$. Then if $T=[f]$, f convex, $T*\varphi_n=[f*\varphi_n]$ where $f*\varphi_n$ are C^∞ convex functions. Thus $D^2(f*\varphi_n)$ are C^∞ functions with values in $B^+(V)$ [13, pp. 100, 103]. Thus

$$D^2([f]*\varphi_n) = [D^2(f*\varphi_n)] \gg 0\,.$$

The set $\{A \in \mathscr{D}'(U, B(V)) : A \gg 0\}$ is clearly sequentially closed, $[f*\varphi_n] \to [f]$, and D^2 is a continuous linear map, so $D^2 T \gg 0$.

Conversely if $D^2 T \gg 0$, then

$$(D^2 T)*\varphi_n = D^2(T*\varphi_n) \gg 0\,,$$

so that the C^∞ functions $f_n := T*\varphi_n$ are convex [13, pp. 100, 103]. Since $[f_n] \to T$ in $\mathscr{D}'$, $T=[f]$ for some convex f by Theorem 2.1.

4. Disintegrations.

We quote for later use a theorem implied by known ones. e.g. [6, Proposition 13 pp. 39–40], [11]

THEOREM 4.1. *Let C, D be compact metric spaces and h a continuous map of C into D. Let $\mu \geqq 0$ be a Radon measure on C and $\nu=\mu \circ h^{-1}$ on D. Then there is a* *map $y \to \mu_y$ from D into Radon measures $\geqq 0$ on C with $\mu_y(C) = \mu_y(h^{-1}\{y\}) = 1$ for all y and $\mu = \int \mu_y \, d\nu(y)$, that is for every μ-integrable Borel function f,*

$$\int_C f \, d\mu = \int_D \left[\int_C f(x) \, d\mu_y(x) \right] d\nu(y) .$$

If ϱ is another Radon measure with $\nu \prec \varrho$, so that $\nu = g\varrho$ for $g = d\nu/d\varrho$, we can write

$$\mu = \int g(y) \mu_y \, d\varrho(y) .$$

If μ is a signed measure, we take its Jordan decomposition $\mu = \mu^+ - \mu^-$ and let $|\mu| = \mu^+ + \mu^-$, $\nu = |\mu| \circ h^{-1}$. Then since $\mu^+ \prec |\mu|$ and $\mu^- \prec |\mu|$, there is a disintegration

$$\mu = \int \mu_y \, d\nu(y) . \tag{4.2}$$

We call the measures μ_y *fiber* measures for μ. They are determined only a.e. (ν).

5. Pointwise derivatives.

We will need to connect pointwise and distribution derivatives. The following may well be known. A proof is included for lack of a reference. Let $\mu = \mu_{ac} + \mu_{sing}$ where $\mu_{ac} \prec \lambda^k$ and $\mu_{sing} \perp \lambda^k$.

THEOREM 5.1. *Suppose g is a locally integrable function on $\mathbf{R}^k$ and for some j, $\partial[g]/\partial x_j = \mu$ where μ is a signed Radon measure. Then for some $g_j = g$ a.e. (λ^k), the pointwise derivative $\partial g_j/\partial x_j$ exists a.e. (λ^k) and equals $d\mu_{ac}/d\lambda^k$.*

PROOF. We can assume $j = 1$. Restricted to the cube $C := C_N := [-N, N]^k$, μ is a finite signed measure. Let

$$T(x) := (x_2, \ldots, x_k), \qquad \nu := |\mu| \circ T^{-1} ,$$

and by (4.2) take a disintegration $\mu = \int_D \mu_y \, d\nu(y)$, where $D := [-N, N]^{k-1}$. We can identify each fiber $T^{-1}\{y\}$ with $[-N, N]$. It is enough to prove the theorem on C.

LEMMA 5.2. *$\nu \prec \lambda^{k-1}$ on D.*

PROOF. Let K be compact in D with $\lambda^{k-1}(K) = 0$, take $\varphi_n \in \mathcal{D}(D)$ with $\varphi_n \downarrow 1_K$, and let $f \in \mathcal{D}(C)$. Then

$$\int_C f(x)\varphi_n(Tx)\,d\mu(x) = -\int g(x)\partial(f(x)\varphi_n(Tx))/\partial x_1\,dx$$
$$= -\int g(x)\,\varphi_n(Tx)(\partial f/\partial x_1)\,dx \to 0$$

as $n \to \infty$ by dominated convergence. So $\int f(x)1_K(Tx)\,d\mu(x)=0$. Let L be compact in C with $\mu^+(L)>0=\mu^-(L)$. Let $f=f_m\downarrow 1_L$ to get

$$\int 1_L(x)1_K(Tx)\,d\mu(x) = 0\,.$$

Letting L increase gives $\int 1_K(Tx)\,d\mu^+(x)=0$. Likewise, $\int 1_K(Tx)d\mu^-(x)=0$. Thus

$$\nu(K) = \int 1_K(Tx)\,d|\mu|(x) = 0$$

and the Lemma follows.

Let

$$h := d\nu/d\lambda^{k-1}, \quad j(x) := h(Tx)\mu_{Tx}([-N,x_1])\,.$$

Then $\int_C |j|\,d\lambda^k<\infty$. For any $\varphi \in \mathscr{D}(C)$,

$$(\partial[j]/\partial x_1)(\varphi) = -\int_C h(Tx)\mu_{Tx}([-N,x_1])(\partial\varphi/\partial x_1)\,dx$$
$$= -\int_D h(y)\int_{-N}^{N} \mu_y([-N,x_1])(\partial\varphi/\partial x_1)\,dx_1\,dy$$
$$= \int_D h(y)\int_{-N}^{N} \varphi(t,y)\,d\mu_y(t)\,dy = \int \varphi\,d\mu\,.$$

So $\partial[j]/\partial x_1=\mu$ and $\partial[j-g]/\partial x_1=0$. Hence for some $f\in \mathscr{L}^1(D,\lambda^{k-1})$,

$$g(x) = j(x)+f(Tx) := g_1(x) \quad \text{a.e. on } C \text{ [15, p. 55]}\,.$$

For each $y \in D$, $t \to \mu_y([-N,t])$ is differentiable for almost all $t \in [-N,N]$ by the classical Lebesgue theorem, and

$$d\mu_y([-N,t])/dt = (d\alpha_y/d\lambda^1)(t)$$

where $\alpha_y \prec \lambda^1$ and $\mu_y-\alpha_y\perp\lambda^1$. Thus a.e. (λ^k),

$$\partial g_1/\partial x_1 = h(Tx)(d\alpha_{Tx}/d\lambda^1)(x_1)\,.$$

Both μ_{ac} and μ_{sing} have disintegrations with respect to ν. Since $\nu \prec \lambda^{k-1}$, $\mu_{sing,y} \perp \lambda^1$ for almost all y. The density $d\mu_{ac}/d\lambda^k$ provides a disintegration of μ_{ac} with $\mu_{ac,y} \prec \lambda^1$ a.e. If ν replaces λ^{k-1} in the disintegration, each fiber measure is multiplied by some constant $h(y)$, so we still have $\mu_{ac,y} \prec \lambda^1$ a.e. Then, $\partial g_1/\partial x_1 = d\mu_{ac}/d\lambda^k$ a.e.

Applying Theorem 5.1 to the first derivatives $g = \partial f/\partial x_i$, we find that the Hessian matrix of the pointwise second derivative is equal a.e. to $d(D^2[f]_{ac})/d\lambda^k$.

Now we will see that there is no "Jordan decomposition" of second derivatives which are measures:

PROPOSITION 5.3. *There is a continuous function f on $\mathbf{R}^2$ such that $D^2[f] = [M]$, where M is a locally integrable matrix-valued function, but such that $f \neq g - h$ for any convex g, h, so $[M] \neq [\mu] - [\nu]$ for any nonnegative symmetric matrix-valued Radon measures μ, ν with $D^2 S = [\mu]$, $D^2 T = [\nu]$, $S, T \in \mathscr{D}'(\mathbf{R}^2)$.*

PROOF. Let $f(x,y) := x \cdot \ln(x^2 + y^2)$. Simple calculations give M locally integrable. Here $D[f]$ is a function unbounded near 0, while for g, h convex, $D[g]$ and $D[h]$ are locally bounded. The rest follows.

For $k = 1$, a distribution T satisfies $D^2 T = \mu$ for some signed Radon measure μ iff $T = [f] - [g]$ for some convex f, g (cf. [13, pp. 22–28]).

Let $g(p) := \sum 2^{-j} f(p - p_j)$, where $\{p_j\}$ is a dense sequence in some ball and f is as in Proposition 5.3. Then $D^2[g]$ is a Radon measure but Dg is nowhere continuous on the ball. Thus, to prove the Busemann–Feller–Alexandrov theorem one must use not only that $D^2[f]$ is a measure but also its positivity. Their proofs use a special construction for convex functions (the indicatrix). It would be interesting to find other proofs.

6. Absolute continuity.

We prove absolute continuity of our measures with respect to $(k-1)$-dimensional Hausdorff measure.

THEOREM 6.1. *For any convex function f on an open set $U \subset \mathbf{R}^k$, $D^2[f] \prec H_{k-1}$.*

PROOF. It is enough to prove $(\partial^2[f]/\partial x_1 \partial x_j)(C) = 0$ if $H_{k-1}(C) = 0$, C compact. We apply Theorem 5.1 and Lemma 5.2 to the partial derivatives $g = \partial f/\partial x_j$. Since the projection T does not increase diameters, and we can assume $U = \mathbf{R}^k$, we have

$$\lambda^{k-1}(TC) = H_{k-1}(TC) \leqq H_{k-1}(C) = 0 ,$$

so by Lemma 5.2,

$$(\partial^2[f]/\partial x_1 \partial x_j)(C) \leqq |\partial^2[f]/\partial x_1 \partial x_j|(T^{-1}TC) = \nu(TC) = 0 .$$

The dimension $k-1$ in Theorem 6.1 is best possible (let $f(x) := |x_1|$). However, the proof gives a somewhat stronger result for component measures of $D^2[f]$, since one can have $H_{k-1}(C) = +\infty > 0$ for some C with $\lambda^{k-1}(TC) = 0$.

COROLLARY 6.2. *For f convex on R^k, $k \geqq 2$, $D^2[f]$ has no atoms.*

Let $F = \operatorname{spt} D^2[f]$, that is, F is the smallest closed set such that $(D^2[f])(U) = 0$ where $U = \mathsf{R}^k \backslash F$. Then on the closure of each connected component of U, f is affine. Thus each such component is the interior of its closure. If F is non-empty, it must either have non-empty interior, or disconnenct R^k so that U has two or more components.

7. Which matrix-valued measures are $D^2[f]$'s?

The entries of a second derivative matrix satisfy relations such as

$$\partial[\mu_{ij}]/\partial x_r = \partial[\mu_{ir}]/\partial x_j .$$

For $k > 1$ these relations are non-trivial.

THEOREM 7.1. *A system $\{T_{ij}\}_{1 \leqq i, j \leqq k}$ of distributions is of the form $T_{ij} = \partial^2 T/\partial x_i \partial x_j$ for some distribution T iff $T_{ij} = T_{ji}$ and $\partial T_{ij}/\partial x_r = \partial T_{ir}/\partial x_j$ for all $i, j, r = 1, \ldots, k$.*

PROOF. Clearly the conditions on T_{ij} are necessary. To see that they are sufficient, we first find that there are distributions T_i with $T_{ij} = \partial T_i/\partial x_j$ [15, p. 59 Théorème 6]. Then, since $T_{ij} = T_{ji}$, by the same theorem of Schwartz there is a T with $\partial T/\partial x_i = T_i$.

The set of measures $D^2[f]$, f convex, is a convex cone. One can try to study this cone by finding its extremal rays, or equivalently the extremal rays of the cone of convex functions modulo affine functions.

In dimension 1, the following is a corollary of Theorems 3.1 and 7.1, or of a classical result of Blaschke and Pick [5]:

THEOREM 7.2. *On an interval $U \subset \mathbf{R}^1$, the set of all $D^2[f]$, f convex, is the set of all Radon measures $\geqq 0$ on U. The extremal rays consist of point masses $\{a\delta_c : a \geqq 0\}$, $c \in U$.*

In dimension $k \geqq 2$, Corollary 6.2 shows that point masses no longer arise. S. Johansen [12] showed that for $k=2$ (and presumably for all $k \geqq 2$) the extremal rays become a more complicated set, and are dense in the cone. A large class of extremal μ have supports which are connected networks of line segments with exactly 3 segments meeeting at each vertex. We call such a vertex *triple*. On each segment L, $\mu = M_L \lambda_L$ where λ_L is Lebesgue measure in L and M_L is a (nonnegative, symmetric) matrix of rank 1 with $M_L(v)=0$ if v is in the direction of L. The M_L are all determined up to one multiplicative constant by the angles at the vertices. There are further restrictions on the network: e.g., at each vertex, the three segments cannot lie in a sector of opening $\leqq \pi$. If, for example, the network contains a triangle, the three segments outside the triangle, when extended inside, must intersect in a point (where three affine functions are equal) or be parallel.

PROPOSITION 7.3. *There is an extremal convex function on $\mathbf{R}^2$ with a non-triple vertex.*

PROOF. Let

$$f(x,y) := \max(|x+y|, |x-y|, 3|x|-2, 3|y|-2) .$$

The support of $D^2[f]$ is then a network with 9 vertices. All except $(0,0)$ are triple, and can be joined by segments not passing through $(0,0)$. Then the extremality follows as in Johansen [12].

There are more complicated extremal networks, including some with infinitely many vertices in a bounded set. Further complications may arise in higher dimension, infinite dimensions (cf. Asplund [4]) and on curved manifolds (cf. Alexander and Bishop [1]).

8. Lebesgue decomposition.

Let f be convex On $\mathbf{R}^k$ and $D^2[f] = \mu \gg 0$. Then μ has a Lebesgue decomposition $\mu = \mu_{\mathrm{ac}} + \mu_{\mathrm{sing}}$ where $\mu_{\mathrm{ac}} \prec \lambda^k$ and $\mu_{\mathrm{sing}} \perp \lambda^k$, $\mu_{\mathrm{ac}} \gg 0$ and $\mu_{\mathrm{sing}} \gg 0$.

PROPOSITION 8.1. *For $k > 1$, μ_{ac} and μ_{sing} need not be of the form $D^2[g]$ for g convex.*

PROOF. For $k=2$, let $f(x,y)$ be the distance from $\langle x,y\rangle$ to the segment

$$S := \{\langle t,0\rangle : 0\leqq t\leqq 1\} .$$

Then f is convex and $D^2[f]_{\text{sing}}$ is concentrated in S. The support of μ_{ac} is $\{\langle x,y\rangle : x\leqq 0 \text{ or } x\geqq 1\}$. Since f is not affine on $x=0$, neither μ_{ac} nor μ_{sing} is $D^2[g]$ for g convex.

9. Subharmonic distributions.

Every convex function is subharmonic. Some of the theory of subharmonic distributions (Schwartz [15, pp. 220–221]) could have been used near the end of the proof of Theorem 3.1 to show $T=[f]$ for some f.

Given a nonnegative matrix-valued measure $\mu=\{\mu_{ij}\}$ on $\mathbf{R}^k$, we define its *trace measure* as

$$\operatorname{tr}\mu := \sum_{1\leqq i\leqq k} \mu_{ii} .$$

For any nonnegative symmetric matrix A, the trace $\operatorname{tr} A:=\sum_{1\leqq i\leqq k} A_{ii}$ is the sum of the eigenvalues. Thus if $\operatorname{tr} A=0$, $A=0$. Hence any nonnegative matrix-valued (symmetric) measure μ is absolutely continuous with respect to $\operatorname{tr}\mu$. If μ is a Radon measure, $\operatorname{tr}\mu$ is σ-finite, and we have $\mu=M\operatorname{tr}\mu$ where M is a nonnegative symmetric matrixvalued function with entries bounded by 1.

If f is convex, $\operatorname{tr} D^2[f]=\Delta f$ where Δ is the Laplace operator,

$$\Delta := \sum_{1\leqq i\leqq k} \partial^2/\partial x_i^2 .$$

A subharmonic function has a Riesz decomposition $f=g+h$ where g is a potential and h is a harmonic function [15, p. 219]. If h is also convex, it is affine since $D^2[h]\prec\Delta[h]=0$. But if f is convex, h is not necessarily affine, e.g. if

$$f(x,y) = 2x^2, \quad g(x,y) = x^2+y^2, \quad \text{and} \quad h(x,y) = x^2-y^2 ,$$

on $U:=\{\langle x,y\rangle : x^2+y^2<1\}$.

Let g be the logarithmic potential of a point mass for $k=2$,

$$g(x,y) := -(2\pi)^{-1}\ln(x^2+y^2)^{\frac{1}{2}} .$$

Then $\partial^2[g]/\partial x^2$is not a measure. Thus Theorem 3.1 for convex functions does not extend to subharmonic functions.

10. Bounded variation of $D_v f$ on lines.

We have a theorem for almost all lines and then a counterexample on some lines.

THEOREM 10.1. *If f is convex on $\mathbf{R}^k$, then for any $u, v \in \mathbf{R}^k$ and almost all $w \in \mathbf{R}^k$, the function*

$$g(t) := (D_v f)(w+tu), \qquad t \in \mathbf{R},$$

is locally of bounded variation.

PROOF. We can assume $k=2$. If u and v are linearly dependent then g is monotone. Thus we can assume $u=\langle 1,0\rangle$, $v=\langle 0,1\rangle$. It suffices to consider the square $C: 0 \leqq x \leqq 1$. Let μ be the measure $\partial^2[f]/\partial x \partial y$ restricted to C. By Lemma 5.2 and the proof of Theorem 5.1, since $\partial f/\partial y$ is a bounded function (defined a.e.) on C, the total variation measure $|\mu|$ has a marginal $|\mu| \circ y^{-1} \prec \lambda^1$ on $[0,1]$. Thus by 4.1 and 4.2, μ has a disintegration $\mu = \int_0^1 \mu_y \, dy$. Specifically, we take disintegrations

$$\mu^+ = \int_0^1 \mu_y^+ \, dy, \qquad \mu^- = \int_0^1 \mu_y^- \, dy,$$

and set $\mu_y := \mu_y^+ - \mu_y^-$. Using a Hahn decomposition of μ, we may assume that for each y, $\mu_y^+ \perp \mu_y^-$. Then

$$f(x,y) - f(0,y) - f(x,0) + f(0,0) = \mu([0,x] \times [0,y]),$$

$$D_v f(x,y) - D_v f(0,y) = D_v \mu([0,x] \times [0,y])$$

$$= D_v \int_0^y \mu_z([0,x]) \, dz \qquad \text{for almost all } x, y,$$

which equals $\mu_y([0,x])$ for almost all y, for a given x or for all rational $x \in [0,1]$. Likewise,

$$D_v \int_0^y \mu_z^+([0,x]) \, dz = \mu_y^+([0,x]),$$

$$D_v \int_0^y \mu_z^-([0,x]) \, dz = \mu_y^-([0,x]),$$

for almost all y and all rational x in $[0,1]$, and hence, by monotonicity, for all x such that $|\mu|_y\{x\}=0$. If $|\mu|_y\{x\}>0$ then conceivably $D_v \int_0^y |\mu|_z([0,x]) \, dz$ does not exist, but in any case monotonicity implies

$$|\mu|_y([0,x)) \leqq \liminf_{h \downarrow 0} h^{-1} \int_y^{y+h} |\mu|_z([0,x]) \, dz$$

$$\leqq \limsup_{h \downarrow 0} h^{-1} \int_y^{y+h} |\mu|_z([0,x]) \, dz \leqq |\mu|_y([0,x]).$$

The same inequalities hold for μ^+ and μ^- in place of $|\mu|$. By choice of μ_y^+ and μ_y^-, they never have an atom at the same point. Then, for almost all y and all x, $D_v \int_0^y \mu_z([0,x])\,dz$ lies in the interval between $\mu_y([0,x))$ and $\mu_y([0,x])$. For such y, then, $x \to D_v f(x,y)$ is of bounded variation.

PROPOSITION 10.2. *There is a convex function g on* $\mathbf{R}^2$ *such that* $\partial g(x,y)/\partial y\,|_{y=0}$ *is nowhere differentiable with respect to x.*

PROOF. The function $f(x) := \sum_{k \geqq 1} 2^{-k}\cos(2^k x)$ is continuous, bounded and nowhere differentiable on $\mathbf{R}^1$ (Hardy [10]). Let $h \in \mathscr{D}(\mathbf{R}^1)$ satisfy $h(0)=0$, $h'(0) \neq 0$, $\operatorname{spt} h \subset [-1,1]$, and

$$\sup_y \left(|h(y)|, |h'(y)|, |h''(y)|\right) \leqq 1 .$$

For each $n = 1, 2, \ldots$, let

$$g_n(x,y) := x^2 + 8|y|^{\frac{3}{2}}/3 + n^{-2}\cos(nx)h(n^{\frac{4}{3}}y) .$$

Then g_n is C^∞ for $y \neq 0$. *For* $|y| \geqq n^{-\frac{4}{3}}$,

$$g_n(x,y) = x^2 + 8|y|^{\frac{3}{2}}/3 .$$

which is convex. For $0 < |y| \leqq n^{-\frac{4}{3}}$, the second derivatives satisfy

$$\partial^2 g_n/\partial x^2 \geqq 2-1 \geqq 1, \qquad |\partial^2 g_n/\geqq x \geqq y| \leqq n^{\frac{4}{3}} ,$$

and

$$\partial^2 g_n/\geqq y^2 \geqq 2|y|^{-\frac{1}{2}} - n^{\frac{2}{3}} \geqq n^{\frac{2}{3}} .$$

Thus the Hessian is nonnegative definite. Taking derivatives in the distribution sense, we see that $D^2[g_n]$ is a nonnegative matrix-valued measure, having a density with respect to Lebesgue measure given by its Hessian, so by Theorem 3.1 g_n is convex.

Now let

$$n(k) := 2^k \quad \text{and} \quad g(x,y) := \sum_{k \geqq 1} 2^{-k/3} g_{n(k)}(x,y) .$$

The series is convergent, uniformly and absolutely for x, y bounded, so g is a continuous convex function. Differentiating the series termwise with respect to y gives

$$\sum_{k \geqq 1} 2^{-k/3} 4|y|^{\frac{3}{2}}/y + \sum_{k \geqq 1} 2^{-k}\cos(2^k x)h'(2^{4k/3}y) ,$$

which converges uniformly and absolutely, so that termwise differentiation is justified. For $y = 0$, we obtain $h'(0)$ times a nowhere differentiable function.

11. Cross derivatives.

For any distribution T, the distribution derivatives $\partial^2 T/\partial x_i \partial x_j$ and $\partial^2 T/\partial x_j \partial x_i$ are always equal as distributions for each i and j. For pointwise derivatives it may fail at some points, even for a convex function (Stoer and Witzgall [16, p. 152]). The same construction with different parameters proves the following (cf. [13, p. 119 problem I(2)]).

PROPOSITION 11.1. *There exist convex functions g on $\mathbf{R}^2$, C^∞ except at 0, with a Hessian at 0 which may be symmetric or not, such that $g''(0)$ does not exist.*

PROOF. We first consider (non-convex) functions of the form

$$f(x,y) = \sum_{0 \leq j \leq 4} a_j x^j y^{4-j}/r^2, \qquad \langle x,y\rangle \neq \langle 0,0\rangle\,,$$

$$f(0,0) = 0, \qquad \text{where } r^2 := x^2+y^2\,.$$

Any such function is C^1 and C^∞ except at the origin. The Hessian matrix exists everywhere, with

$$(\partial^2 f/\partial x \partial y)(x,0) = a_3 \quad \text{for all } x \text{ and}$$

$$(\partial^2 f/\partial y \partial x)(0,y) = a_1 \quad \text{for all } y\,.$$

Thus the Hessian is symmetric at 0 iff $a_1 = a_3$. Calculation shows that the second Fréchet derivative $f''(x)$ exists at the origin iff both $a_1 = a_3$ and $a_0 + a_4 = a_2$. The entries in the Hessian are homogeneous of degree 0 (quotients of 6th degree homogeneous polynomials) and are uniformly bounded. A continuous function, convex on all lines not passing through 0, is actually convex. Thus

$$g(x,y) := f(x,y) + A(x^2+y^2)$$

is convex for A large enough. We can take $a_0 = 1$, $a_2 = a_4 = 0$, and either $a_1 = a_3 = 0$ or $a_1 = 1$, $a_3 = 0$ to obtain the desired examples.

Stoer and Witzgall [16, p. 152] take $a_0 = a_2 = a_4 = 0$, $a_1 = 1 = -a_3$, $A = 13$ to obtain a convex function with a non-symmetric Hessian at 0. ($A \geqq 3$ will suffice.)

REFERENCES

1. S. Alexander and R. L. Bishop, *Convex-supporting domains on spheres*, Ill. J. Math. 18 (1974), 37–47.

2. A. D. Alexandrov, *The existence almost everywhere of the second differential of a convex function and some associated properties of convex surfaces* (in Russian), Učenye Zapiski Leningr. Gos. Univ. Ser. Mat. 37 no. 6 (1939), 3–35.
3. R. D. Anderson and V. L. Klee, *Convex functions and upper semi-continuous decompositions*, Duke Math. J. 19 (1952), 349–357.
4. E. Asplund, *Fréchet differentiability of convex functions*, Acta. Math. 121 (1968), 31–48.
5. W. Blaschke and G. Pick, *Distanzschätzungen in Funktionenraümen* II, Math. Ann. 77 (1916), 277–302.
6. N Bourbaki, *Éléments de Mathématique*, Livre 6, *Intégration*, Chapitre 6 (Act. Sci. Ind. 1281) Hermann, Paris, 1959–1969.
7. H. Busemann and W. Feller, *Krümmungseigenschaften konvexer Flächen*, Acta. Math. 66 (1935), 1–47.
8. R. M. Dudley, *Convergence of sequences of distributions*, Proc. Amer. Math. Soc. 27 (1971), 531–534.
9. H. Federer, *Geometric Measure Theory* (Grundlehren Math. Wissensch. 153), Springer-Verlag, Berlin · Heidelberg · New York.
10. G. H. Hardy, *Weierstrass's non-differentiable function*, Trans. Amer. Math. Soc. 17 (1916), 301–325.
11. C. Ionescu Tulcea, *Two theorems concerning the disintegration of a measure*, J. Math. Anal. Appl. 26 (1969), 376–380.
12. S. Johansen, *The extremal convex functions*, Math. Scand. 34 (1974), 61–68.
13. A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York and London, 1973.
14. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton 1970.
15. L. Schwartz, *Théorie des Distributions* (Publications de l'Institut de mathematique de l'Université de Strasbourg 9), 2e ed., Hermann, Paris, 1966.
16. J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions* (Grundlehren Math. Wissensch. 163), Springer-Verlag, Berlin · Heidelberg · New York, 1970.
17. J. H. Webb, *Sequential convergence in locally convex spaces*, Proc. Cambridge Philos. Soc. 64 (1968), 341–364.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
U.S.A.

AND

AARHUS UNIVERSITET
DENMARK