

## Topic 17: Monotone mappings

### 17.1 The subdifferential correspondence

**17.1.1 Lemma** *The set of continuous subgradients of a convex function at a point is a (possibly empty) closed convex set.*

Move to a section  
the uhc of the  
subdifferential  
map

*Proof:* This is immediate since it is the set of solutions to a system of weak linear inequalities, one for each  $y$ . ■

### 17.2 Monotone and cyclically monotone mappings

Recall that a real function  $g: X \subset \mathbf{R} \rightarrow \mathbf{R}$  is increasing if  $x \geq y$  implies  $g(x) \geq g(y)$ . Another way to say this is  $[g(x) - g(y)](x - y) \geq 0$  for all  $x, y$ . Or equivalently,  $g$  is nondecreasing if

$$g(x)(y - x) + g(y)(x - y) \leq 0 \quad \text{for all } x, y.$$

More generally, a correspondence  $\varphi: X \subset \mathbf{R}^m \rightarrow \mathbf{R}^m$  is **monotone (increasing)** if

$$(p_x - p_y) \cdot (x - y) \geq 0 \quad \text{for all } x, y \in X, \text{ and all } p_x \in \varphi(x), p_y \in \varphi(y).$$

We could also write this as

$$p_x \cdot (y - x) + p_y \cdot (x - y) \leq 0.$$

A mapping  $\varphi$  is **monotone (decreasing)** if the reverse inequality always holds.

There is a natural generalization of these conditions. A **cycle** is a finite sequence  $x_0, x_1, \dots, x_n, x_{n+1}$  with  $x_{n+1} = x_0$ . A mapping  $g: U \subset \mathbf{R}^m \rightarrow \mathbf{R}^m$  is called **cyclically monotone (increasing)** if for every cycle  $x_0, x_1, \dots, x_n, x_{n+1} = x_0$  in  $U$ , we have

$$g(x_0) \cdot (x_1 - x_0) + g(x_1) \cdot (x_2 - x_1) + \dots + g(x_n) \cdot (x_0 - x_n) \leq 0.$$

If the same sum is always  $\geq 0$ , we shall say that  $g$  is **cyclically monotone (decreasing)**. Monotonicity is just cyclical monotonicity with  $n = 1$ .

More generally, a correspondence  $\varphi: U \subset \mathbf{R}^m \rightarrow \mathbf{R}^m$  is called **cyclically monotone (increasing)**<sup>1</sup> if for every cycle  $(x_0, p_0), (x_1, p_1), \dots, (x_{n+1}, p_{n+1}) = (x_0, p_0)$  in the graph of  $\varphi$ , that is, with  $p_i \in \varphi(x_i)$  for all  $i$ , we have

$$p_0 \cdot (x_1 - x_0) + p_1 \cdot (x_2 - x_1) + \dots + p_n \cdot (x_0 - x_n) \leq 0.$$

<sup>1</sup>Most authors define monotone and cyclically monotone correspondences to be increasing, and do not make a definition for decreasing monotonicity. This is because mathematicians

We mention that if  $m = 1$  ( $\mathbf{R}^m = \mathbf{R}$ ) then a function  $g$  is cyclically monotone if and only if it is monotone. For  $m \geq 2$ , there are monotone functions that are not cyclically monotone, see Rockafellar [5, p. 240].

### 17.3 Cyclic monotonicity characterizes subdifferentials

**17.3.1 Theorem (Cyclic monotonicity of the subdifferential)** *Let  $f$  be convex and subdifferentiable on a convex set  $C \subset \mathbf{R}^m$ . Then the subdifferential correspondence  $x \mapsto \partial f(x)$  is cyclically monotone (increasing). That is, for any cycle  $(x_0, p_0), (x_1, p_1), \dots, (x_{n+1}, p_{n+1}) = (x_0, p_0)$  in the graph of  $\partial f$ , we have*

$$\sum_{k=0}^n p_k \cdot (x_{k+1} - x_k) \leq 0.$$

*Proof:* By the subgradient inequality, for each  $k$ ,

$$f(x_{k+1}) \geq f(x_k) + p_k \cdot (x_{k+1} - x_k)$$

or

$$p_k \cdot (x_{k+1} - x_k) \leq f(x_{k+1}) - f(x_k)$$

Summing both sides gives

$$\sum_{k=0}^n p_k \cdot (x_{k+1} - x_k) \leq \sum_{k=0}^n [f(x_{k+1}) - f(x_k)] = 0,$$

where the last equality follows from the fact that  $x_{n+1} = x_0$ . ■

**17.3.2 Corollary (Cyclic monotonicity of the derivative)** *Let  $U \subset \mathbf{R}^m$  be a nonempty open convex set, and let  $f$  be convex and differentiable on  $U$ . Then the gradient mapping  $x \mapsto f'(x)$  is cyclically monotone (increasing). That is, for any cycle  $x_0, x_1, \dots, x_n, x_{n+1}$  in  $U$  with  $x_{n+1} = x_0$ , we have*

$$\sum_{k=0}^n f'(x_k) \cdot (x_{k+1} - x_k) \leq 0.$$

Note that the gradient of a concave function is cyclically monotone (decreasing).

Remarkably, cyclic monotonicity characterizes the subdifferential mapping. This result is due to Rockafellar, and may be found in his book [5, Theorem 24.8, p. 238].

---

find convex functions (such as norms) to be natural, and as we shall see below there is an important relationship between convex functions and (cyclically) monotone increasing mappings. Economists however find concave functions to be naturally occurring (as in production functions) so it seems natural to introduce a term for (cyclically) monotone decreasing mappings. Just keep in mind that for every statement about convex functions, there is a corresponding one for concave functions derived by replacing  $f$  by  $-f$ .

**17.3.3 Theorem (Rockafellar)** *Let  $C \subset \mathbf{R}^m$  be a nonempty convex set and let  $\varphi: C \rightarrow \mathbf{R}^m$  be a correspondence with nonempty values. Then  $\varphi$  is cyclically monotone (increasing) if and only if there is a lower semicontinuous convex function  $f: C \rightarrow \mathbf{R}$  satisfying*

$$\varphi(x) \subset \partial f(x) \quad \text{for every } x \in C.$$

The proof uses a construction that may seem a bit unusual at first, but is actually rather natural. Consider the problem of recovering a continuously differentiable function  $f$  from its derivative  $f'$ . Since for any constant  $c$ ,  $f$  and  $f + c$  have the same derivative, the best we can hope to do is recover  $f$  up to some constant. One way to do this is to fix  $f$  at some point  $x_0$  and use the Fundamental Theorem of Calculus to write  $f(y) = \int_{x_0}^y f'(x) dx$ . Now the integral  $\int_{x_0}^y f'(x) dx$  can be approximated by using points  $x_1, \dots, x_n$  to partition the interval from  $x_0$  to  $y$  and computing the sum

$$S = f'(x_0)(x_1 - x_0) + f'(x_1)(x_2 - x_1) + \cdots + f'(x_n)(y - x_n).$$

Since  $f'$  is integrable, as the mesh of this partition goes to zero, the sums converge to  $f(y) - f(x_0)$ . The proof below will use a similar idea, but we will work in  $\mathbf{R}^m$ , and we won't require the  $x_1, \dots, x_n$  to partition the segment between  $x_0$  and  $y$ . Nevertheless, the subgradient inequality will assure us that the sums we use will always underestimate  $f(y) - f(x_0)$ , so we can take a supremum rather than a more general limit.

*Proof of Theorem 17.3.3:* (  $\Leftarrow$  ) If  $\varphi(x) \subset \partial f(x)$  for a convex  $f$ , Theorem 17.3.1 shows that  $\varphi$  is cyclically monotone (increasing).

(  $\Rightarrow$  ) For the converse, assume  $\varphi$  is cyclically monotone (increasing) and fix some point  $x_0$  in  $C$  and fix  $p_0 \in \varphi(x_0)$ .

Given any finite sequence  $(x_1, p_1), \dots, (x_n, p_n)$  in  $\mathbf{R}^m \times \mathbf{R}^m$ , define the affine function  $S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}$  by

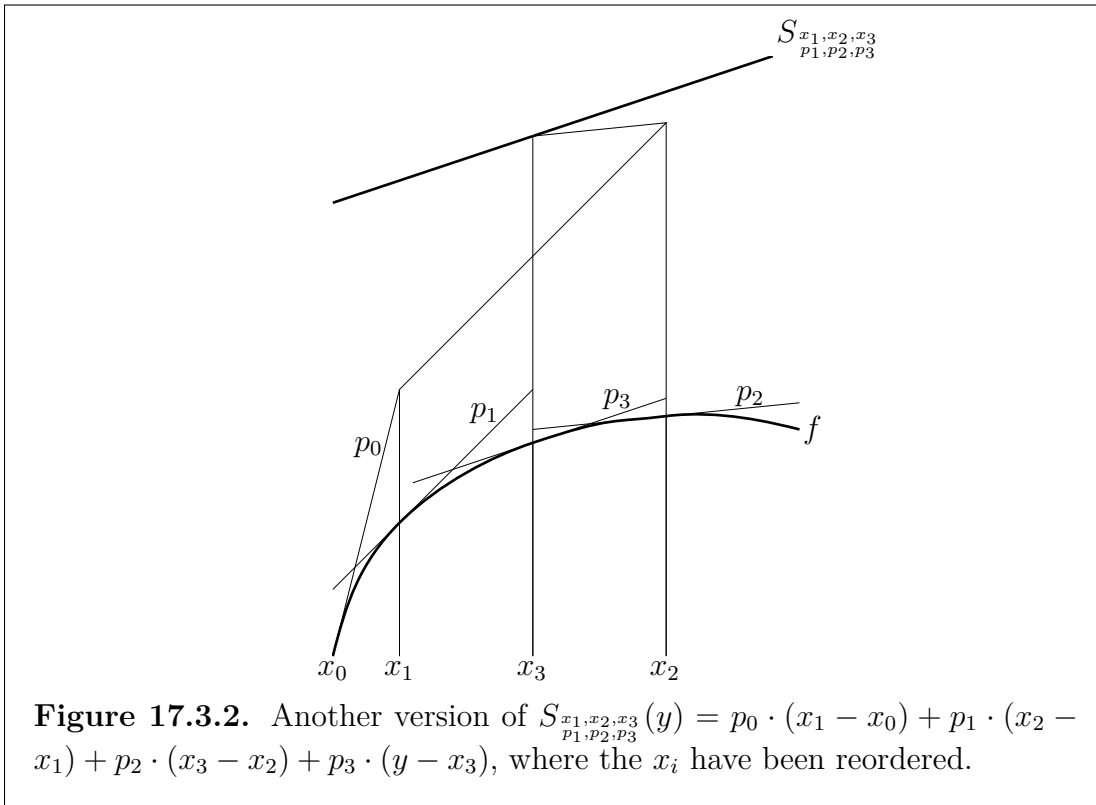
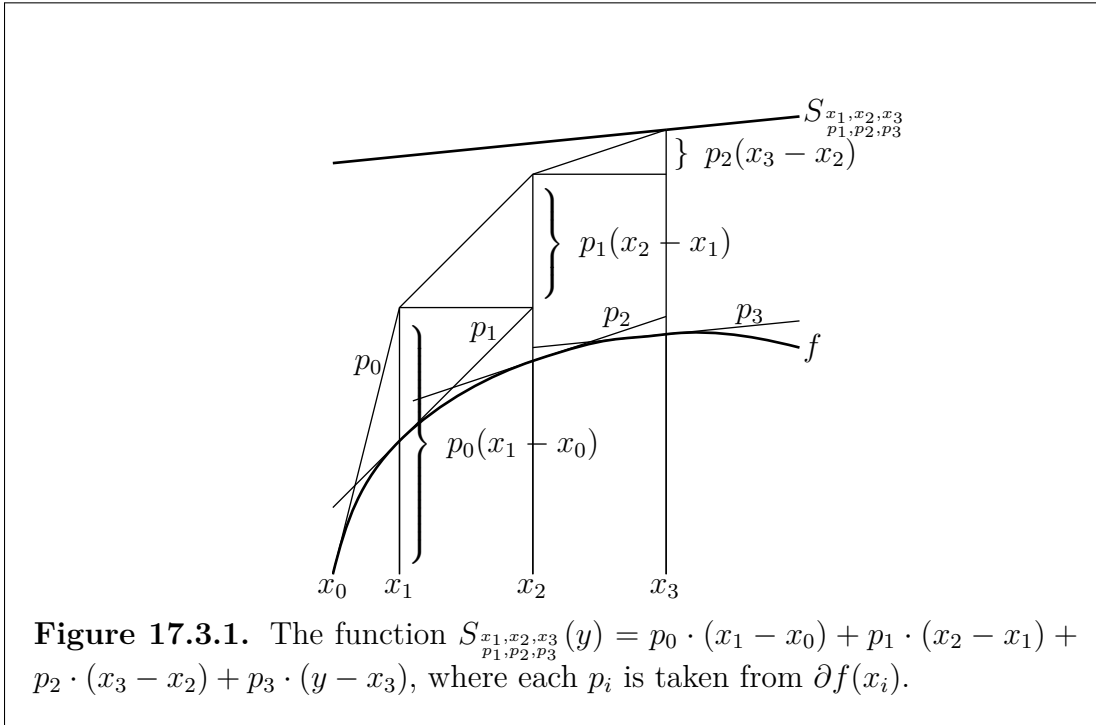
$$S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(y) = p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (y - x_n).$$

We will use these sums  $S$  to approximate  $f(y) - f(x_0)$ . The construction of such functions  $S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}$  is illustrated in Figures 17.3.1 and 17.3.2, unfortunately for the case of a concave function rather than a convex function.

So define the function  $f: C \rightarrow \mathbf{R}$  to be the pointwise supremum of the functions  $S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}$  as  $(x_1, p_1), \dots, (x_n, p_n)$  ranges over all finite sequences in the graph of  $\varphi$ . That is,

$$f(y) = \sup \left\{ S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(y) : (\forall i) [x_i \in C \ \& \ p_i \in \varphi(x_i)] \right\}.$$

We now show that  $f$  has the desired properties.



- $f$  is convex and lower semicontinuous.

Since  $f$  is the pointwise supremum of a collection of continuous affine functions, it is convex by part 4 of Exercise 1.3.3, and lower semicontinuous by Proposition 13.4.5.

- $f$  is proper. Indeed  $f(y) < \infty$  for all  $y \in C$ .

To see this, pick any  $y$  and fix some  $p_y$  in  $\varphi(y)$ . Let  $(x_1, p_1), \dots, (x_n, p_n)$  be a finite sequence in the graph of  $\varphi$ . Consider the affine function

$$p_y \cdot (x_0 - y) + \underbrace{p_0 \cdot (x_1 - x_0) + \dots + p_n \cdot (y - x_n)}_{S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(y)}.$$

But  $y, x_0, x_1, \dots, x_n, y$  constitutes a cycle, so by cyclic monotonicity,

$$p_y \cdot (x_0 - y) + S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(y) \leq 0.$$

Rearranging gives

$$S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(y) \leq p_y \cdot (y - x_0).$$

Note that the right-hand side is independent of  $(x_1, p_1), \dots, (x_n, p_n)$ , so taking the supremum on the left-hand side gives

$$f(y) \leq p_y \cdot (y - x_0) < \infty.$$

- For every  $x \in C$ ,  $\varphi(x) \subset \partial f(x)$ . That is, for any  $x, y$  in  $C$  and any  $p_x \in \varphi(x)$  the subgradient inequality

$$f(y) \geq f(x) + p_x \cdot (y - x)$$

is satisfied.

To see this, let  $\varepsilon > 0$  be given. Then by the definition of  $f$ , since  $f(x)$  is finite, there is a finite sequence  $(x_1, p_1), \dots, (x_n, p_n)$  in the graph of  $\varphi$  with

$$f(x) - \varepsilon < S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(x) \leq f(x). \quad (1)$$

Extend this sequence by appending  $(x, p_x)$ . Again by the definition of  $f$  as the supremum of these  $S$  functions, for all  $y$ ,

$$S_{\substack{x_1, \dots, x_n, x \\ p_1, \dots, p_n, p_x}}(y) \leq f(y). \quad (2)$$

But

$$\begin{aligned} S_{\substack{x_1, \dots, x_n, x \\ p_1, \dots, p_n, p_x}}(y) &= \underbrace{p_0 \cdot (x_1 - x_0) + \dots + p_n \cdot (x - x_n)}_{S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(x)} + p_x \cdot (y - x) \\ &= S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(x) + p_x \cdot (y - x). \end{aligned} \quad (3)$$

Adding  $p_x \cdot (y - x)$  to both sides of the first inequality in (1) and combining with (3) and (2) gives

$$f(x) - \varepsilon + p_x \cdot (y - x) < S_{\substack{x_1, \dots, x_n \\ p_1, \dots, p_n}}(x) + p_x \cdot (y - x) = S_{\substack{x_1, \dots, x_n, x \\ p_1, \dots, p_n, p_x}}(y) \leq f(y).$$

Since  $\varepsilon > 0$  is arbitrary, we conclude that  $f(y) \geq f(x) + p_x \cdot (y - x)$ , so indeed  $\varphi(x) \subset \partial f(x)$ . ■

## 17.4 Monotonicity vs. cyclic monotonicity

We mention that if  $m = 1$  ( $\mathbf{R}^m = \mathbf{R}$ ) then a function  $g$  is cyclically monotone if and only if it is monotone. For  $m \geq 2$ , there are monotone functions that are not cyclically monotone.

**17.4.1 Example (Monotonicity vs. cyclic monotonicity)** This example is based on a remark of Rockafellar [5, p. 240]. Define the function  $g: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  by

$$g(x, y) = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2y - x, -y).$$

Then  $g$  is monotone (decreasing):

$$\begin{aligned} g(x_0, y_0) \cdot (x_1 - x_0, y_1 - y_0) + g(x_1, y_1) \cdot (x_0 - x_1, y_0 - y_1) \\ &= (2y_0 - x_0, -y_0) \cdot (x_1 - x_0, y_1 - y_0) + (2y_1 - x_1, -y_1) \cdot (x_0 - x_1, y_0 - y_1) \\ &= (2y_0 - x_0, -y_0) \cdot (x_1 - x_0, y_1 - y_0) - (2y_1 - x_1, -y_1) \cdot (x_1 - x_0, y_1 - y_0) \\ &= (2y_0 - x_0 - 2y_1 + x_1, y_1 - y_0) \cdot (x_1 - x_0, y_1 - y_0) \\ &= (x_1 - x_0)^2 - 2(y_1 - y_0)(x_1 - x_0) + (y_1 - y_0)^2 \\ &= \left( (x_1 - x_0) - (y_1 - y_0) \right)^2 \\ &\geq 0. \end{aligned}$$

But  $g$  is not cyclically monotone (decreasing): Consider the cycle  $(0, -2), (2, -2), (3, 0), (0, -2)$ . Then

$$\begin{aligned} g(0, -2) \cdot \left( (2, -2) - (0, -2) \right) + g(2, -2) \cdot \left( (3, 0) - (2, -2) \right) \\ + g(3, 0) \cdot \left( (0, -2) - (3, 0) \right) \\ &= (-4, 2) \cdot (2, 0) + (-6, 2) \cdot (1, 2) + (-3, 0) \cdot (-3, -2) \\ &= -8 - 2 + 9 \\ &= -1. \end{aligned}$$

Can anyone prove this???

In fact, Rockafellar [5, p. 240] asserts the following. Let  $g: \mathbf{R}^n \rightarrow \mathbf{R}^n$  be linear, that is,  $g(x) = Ax$ , where  $A$  is an  $n \times n$  matrix. If  $A$  is negative quasi-semidefinite but not symmetric, then  $g$  is monotone decreasing, but not cyclically monotone decreasing. □

## 17.5 Monotonicity and second derivatives

From Corollary 17.3.2 we know that the gradient of a convex function  $f: C \rightarrow \mathbf{R}$ , where  $C$  is an open convex set in  $\mathbf{R}^n$ , is monotone (increasing). That is, it satisfies

$$f'(x_0) \cdot (x_1 - x_0) + f'(x_1) \cdot (x_0 - x_1) \leq 0,$$

which can be rearranged as

$$(f'(x_1) - f'(x_0)) \cdot (x_1 - x_0) \geq 0.$$

This is enough to show that the second differential (if it exists) is positive semidefinite.

**17.5.1 Theorem** *Let  $C$  be an open convex subset of  $\mathbf{R}^m$ , and let  $f$  be a twice differentiable function on  $C$ . If the derivative  $f'$  of  $f$  is monotone (increasing), then the Hessian of  $f$  is everywhere positive semidefinite*

*Proof:* Consider a point  $x$  in  $C$  and choose  $v$  so that  $x \pm v$  belong to  $C$ . Then by monotonicity with  $x_0 = x$  and  $x_1 = x + \lambda v$ ,

$$(f'(x + \lambda v) - f'(x)) \cdot (\lambda v) \geq 0.$$

Dividing by the positive quantity  $\lambda^2$  implies

$$v \cdot \frac{(f'(x + \lambda v) - f'(x))}{\lambda} \geq 0. \quad (4)$$

Define the function  $g: (-1, 1) \rightarrow \mathbf{R}$  by

$$g(\lambda) = v \cdot f'(x + \lambda v) = \sum_{i=1}^m D_i f(x + \lambda v) v_i.$$

In particular, if  $f$  is twice differentiable, then by the Chain Rule

$$g'(\lambda) = \sum_{i=1}^m \sum_{j=1}^m D_{ij} f(x + \lambda v) v_i v_j = f''(x + \lambda v)(v, v). \quad (5)$$

On the other hand,

$$g'(0) = \lim_{\lambda \rightarrow 0} v \cdot \frac{g(\lambda) - g(0)}{\lambda} = \lim_{\lambda \rightarrow 0} v \cdot \frac{(f'(x + \lambda v) - f'(x))}{\lambda} \geq 0, \quad (6)$$

where, the last inequality is (4). Evaluating (5) at  $\lambda = 0$  and substituting into (6) shows that the Hessian matrix  $f''(x)$  is positive semidefinite. ■

**17.5.2 Theorem (Second order characterization of convexity)** *Let  $C$  be an open convex subset of  $\mathbf{R}^m$ , and let  $f$  be a twice differentiable function on  $C$ .*

*Then  $f$  is convex if and only if the Hessian of  $f$  is everywhere positive semidefinite. And  $f$  is concave if and only if the Hessian of  $f$  is everywhere negative semidefinite.*

*Proof:* I'll prove the assertion for the positive definite case. If  $f$  is convex, then by Corollary 17.3.2, its derivative is monotone (increasing), so by Theorem 17.5.1, its Hessian is positive semidefinite.

For the converse, assume that the Hessian of  $f$  is everywhere positive semidefinite. Let  $x, y$  belong to  $C$ , and let  $0 \leq \lambda \leq 1$ . We wish to show that  $f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$ . Setting  $v = y - x$ , this is equivalent to

$$f(x + \lambda v) - f(x) \leq \lambda(f(x + v) - f(x)).$$

By the Chain Rule  $g(t) = f(x + tv)$  satisfies

$$\begin{aligned} g'(t) &= f'(x + tv) \cdot v \\ g''(t) &= \sum_{i=1}^n \sum_{j=1}^n D_{ij} f(x + tv) v_i v_j \geq 0. \end{aligned}$$

This implies that  $g'' \geq 0$ , so  $g'$  is nondecreasing. So by the Second Fundamental Theorem of Calculus,

$$\begin{aligned} f(x + \lambda v) - f(x) &= g(\lambda) - g(0) = \int_0^\lambda g'(t) dt \\ &\leq \int_0^\lambda g'(1) dt \\ &= \lambda \int_0^1 g'(1) dt \\ &= \lambda(g(1) - g(0)) = \lambda(f(x + v) - f(x)). \end{aligned}$$

■

**17.5.3 Remark** At this point I was a bit confused. If you are not confused, you may not wish to read this.

We have just shown that if a twice differentiable function has a monotone gradient, then it has positive semidefinite Hessian, so it is convex, and therefore its gradient is actually cyclically monotone. Thus every monotone gradient is cyclically monotone. Now Theorem 17.3.3 says that every cyclically monotone vector field is a selection from the subdifferential of a convex function. I am embarrassed to admit it, but I thought for a while therefore that the argument above allowed me to conclude that every monotone vector field is a selection from the subdifferential of a convex function, which would be a stronger result (except that it is not true).

What the argument above shows is this: Every monotone vector field that happens also to be a gradient of a twice differentiable function is indeed cyclically monotone. But, there are differentiable vector fields that are not gradients of twice differentiable functions. (A vector field



Cite symmetry of second differential.

is just a function from  $\mathbf{R}^m$  into  $\mathbf{R}^m$ . If it is the gradient of a real function  $f$ , then  $f$  is called the **potential** of the field.) Recall that second differentials are symmetric. So if  $x \mapsto g(x)$  is a gradient of a twice differentiable function  $f$ , then

$$D_j g_i(x) = D_i D_j f(x) = D_j D_i f(x) = D_i g_j(x).$$

Now consider the vector field of Example 17.4.1, namely  $g: \mathbf{R}^2 \rightarrow \mathbf{R}^2$  defined by

$$g(x, y) = (2y - x, -y).$$

This vector field is continuously differentiable, but

$$D_1 g_2 = 0, \quad D_2 g_1 = 2,$$

so  $g$  cannot be the gradient of any twice differentiable function. However, as we saw in Example 17.4.1,  $g$  is monotone (decreasing), but not cyclically monotone (decreasing).

By the way, this is analogous to the “integrability problem” in demand theory. The Weak Axiom of Revealed Preference can be used to show that the Slutsky matrix is quasi-negative semidefinite (negative semidefinite without necessarily being symmetric), see, e.g., Samuelson [6, pp. 109–111] or Kihlstrom, Mas-Colell, and Sonnenschein [3], but it takes the Strong Axiom to show symmetry: Gale [1], Houthakker [2], Uzawa [7].

Now let’s return to support functions.

**17.5.4 Lemma** *Suppose  $x(p)$  minimizes  $p \cdot x$  over the nonempty set  $A$ . Suppose further that it is the unique minimizer of  $p \cdot x$  over  $\overline{\text{co}} A$ . If  $\frac{\partial^2 \mu_C(p)}{\partial p_i^2}$  exists (or equivalently  $\frac{\partial x(p)}{\partial p_i}$  exists), then*

$$\frac{\partial x_i(p)}{\partial p_i} \leq 0.$$

*Proof:* This follows from Corollary 15.2.2 and the discussion above. ■

Do we need twice differentiability or just the existence of the second partial????

This, by the way, summarizes almost everything interesting we know about cost minimization.

## References

- [1] D. Gale. 1960. A note on revealed preference. *Economica N.S.* 27(108):348–354. <http://www.jstor.org/stable/2550547>
- [2] H. S. Houthakker. 1950. Revealed preference and the utility function. *Economica N.S.* 17:159–174. <http://www.jstor.org/stable/2549382>
- [3] R. Kihlstrom, A. Mas-Colell, and H. F. Sonnenschein. 1976. The demand theory of the weak axiom of revealed preference. *Econometrica* 44(5):971–978. <http://www.jstor.org/stable/1911539>

- [4] J.-C. Rochet. 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics* 16(2):191–200.  
[DOI: 10.1016/0304-4068\(87\)90007-3](https://doi.org/10.1016/0304-4068(87)90007-3)
- [5] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.
- [6] P. A. Samuelson. 1965. *Foundations of economic analysis*. New York: Athenaeum. Reprint of the 1947 edition published by Harvard University Press.
- [7] H. Uzawa. 1958. Preference and rational choice in the theory of consumption. In K. J. Arrow, L. Hurwicz, and H. Uzawa, eds., *Studies in Linear and Non-linear Programming*, number 2 in Stanford Mathematical Studies in the Social Sciences, chapter 9, pages 129–148. Stanford, California: Stanford University Press.