CONCAVIFYING THE QUASICONCAVE

CHRIS CONNELL[†] AND ERIC B. RASMUSEN[‡]

ABSTRACT. We revisit a classic question of Fenchel from 1953: Which quasiconcave functions can be concavified by post-composition with a monotonic transformation? This question has a long history in economics utility theory. While many authors have given partial answers under various assumptions, we offer a complete characterization for quasiconcave functions without a priori assumptions on regularity. The answer hinges on the local regularity class of the function.

We establish this characterization of concavifiability for continuous functions whose domain is any arbitrary geodesic metric space. Under the additional assumption of twice differentiability, we also provide simpler necessary and sufficient conditions for concavifiability on Riemannian manifolds which essentially generalize those given by Kannai for the Euclidean case.

1. INTRODUCTION

Quasiconcavity is a property of a function which, if strict, guarantees a unique global maximum on any compact convex domain. As the name suggests, it weakens the property of concavity, and has found wide application in economics and other fields. The purpose of this paper is to further analyze a natural way to link quasiconcavity to concavity: by means of a monotonically increasing transformation. We will show that to say a continuous function is strictly quasiconcave is close, in a specific sense, to saying it can be made concave by a post-composing with a strictly increasing function. Almost any sufficiently regular quasiconcave function can be concavified this way. Any function that is not quasiconcave cannot be concavified by post-composition.

The "almost" above refers to the two additional qualifications beyond quasiconcavity that must be required of the function after concavifying one monotonic slice— which can always be done even if the function is not continuous. (It turns out to be easy to reduce the question of concavifiability of a monotonic discontinuous function f to that of an associated function which must be continuous to be concavifiable.) These qualifications, which appear as regularity assumptions involving upper derivatives (see Section 3.2), are most easily stated for the case when the domain of the objective function f is an interval $(a, b) \subset \mathbb{R}$. This leads to our first main result.

Theorem 1. *For any continuous nonmonotonic function $f\colon (a,b) \to \mathbb{R}$, there is a function $g\colon \mathrm{Range(f)} \to \mathbb{R}$ such that $g \circ f$ is strictly concave if and only if*

(i) f or $-f$ is strictly quasiconcave achieving it's maximum at $m \in (a, b)$.
(ii) for $h = f^{-1}_{\downharpoonright(a,m]}$, the function $h \circ f_{\downharpoonright(m,b)}$ and its inverse are locally Lipschitz.

[†]Supported in part by Simons Foundation Grant #210442.
[‡]Supported in part by Olin Center, Harvard Law School.

(iii) $\log\left|\overline{D}(h\circ f)\right| \in \mathrm{BV}_{loc}((m,b))$.

Moreover, when g exists it is strictly monotone.

The general case for arbitrary (complete) geodesic metric space domains is similar:

Theorem 2. *Let X be any geodesic metric space. For any continuous function $f\colon X\to\mathbb{R}$ there is a function $g\colon \mathrm{Range(f)}\to\mathbb{R}$ such that $g\circ f$ is strictly concave if and only if:*

(i) Either f or $-f$ is strictly quasiconcave with maximum
(ii) $h\circ f$ is locally Lipschitz on $X-\{m\}$ where $h=(f\circ\gamma_o)^{-1}$ for some geodesic γ_o in X ending at m such that $\mathrm{Range}(f\circ\gamma_o)=\mathrm{Range}(f)$. Moreover $\overline{D}(h\circ f\circ\gamma)$ does not vanish on any geodesic segment $\gamma\colon(0,1)\to X$ for which $h\circ f\circ\gamma$ is strictly increasing.
(iii) The total variation of $\log\overline{D}(h\circ f)$ along all geodesics $\gamma\colon[0,1]\to X$ for which $h\circ f\circ\gamma$ is strictly increasing is uniformly bounded away from the extrema of $h\circ f$, or in other words,

$$\inf_{\gamma}\left\{[\overline{D}(\log\overline{D}(h\circ f\circ\gamma)^{-1})]^{-}\right\}\in L^1_{loc}(R).$$

(Here R is the interior of the range of f and the infimum is taken over all geodesic segments $\gamma\colon[0,1]\to X$ for which $h\circ f\circ\gamma$ is strictly increasing.)

Our topic goes back to the origins of the study of quasiconcavity starting in the 1940's found in papers by DeFinetti [4] and Fenchel [5] (who invented the name, as Guerraggio & Molho explain in their history [8]). Both DeFinetti (in his "second problem") and Fenchel investigated whether any quasiconcave function could be transformed into a concave function, which with related problems is surveyed in [13] (see also Section 9 of [2]). Kannai [9] (and more elaborately in [10]) treats the question in depth in the context of utility functions, giving conditions under which, in the language of economics, continuous convex preference orderings can be represented by concave utility functions. Richter & Wong [14] and Kannai [11] similarly address preferences over discrete sets. We will discuss some of Kannai's conditions in greater detail in Section 5 below where we prove the following generalization for Riemannian manifolds. (See that section for definitions of f_{ij} and λ_i.)

Theorem 3. *Let M^n be a C^2 Riemannian manifold, possibly with boundary. For any twice-differentiable function $f: M\to\mathbb{R}$ there is a strictly increasing twice-differentiable $g\colon\mathbb{R}\to\mathbb{R}$ such that $g\circ f$ is strictly concave if and only if:*

(i) f is strictly quasiconcave;
(ii) ∇f does not vanish on the interior of M except possibly at the maximum point of f,
(iii) The function

$$q(t)=\inf_{x\in f^{-1}(t)}\frac{1}{\left\|\nabla f(x)\right\|^2}\left(-f_{11}(x)-\sum_{j=2}^{n}\frac{f_{1j}^2(x)}{\lambda_j(x)\left\|\nabla f(x)\right\|}\right)$$

has negative part q^- belonging to $L^1_{loc}(R)$, where R is the interior of the range of f.

A suitable g is

$$g(z) = \int_{f(m)}^{z} e^{\int_{f(m)}^{s} (-1+q^{-}(t))\, dt}\, ds.$$

Acknowledgements. We would like to thank Per Hjertstrand and Dan Zhao for helpful comments.

2. Preliminaries

Definition 1. (quasiconcavity) A function f defined on a subset $D \subset \mathbb{R}^n$ is *weakly quasiconcave* if for any two distinct points $x', x'' \in D$ and any number $t \in (0, 1)$ with $tx' + (1-t)x'' \in D$, we have

$$f(tx' + (1-t)x'') \geq \min\{f(x'), f(x'')\}\,. \tag{2.1}$$

We say that f is *strictly quasiconcave* if the inequality is strict whenever $x' \neq x''$, .

If $-f$ is strictly quasiconcave then f is *strictly quasiconvex.* If $-f$ is weakly quasiconcave then f is *weakly quasiconvex.*

Compare the right-hand side of (2.1) to the condition for concavity,

$$f(tx' + (1-t)x'') \geq tf(x') + (1-t)f(x''). \tag{2.2}$$

We will be looking at whether given a quasiconcave f we can always find a strictly monotonic function g that will transform f to a strictly concave $g \circ f$.

Many of the difficulties in concavifying quasiconcave functions already arise when the function's domain is just $\mathbb{R}$, so this will be the focus of the rest of the current section. As the complexity of the domain increases, these issues become harder to detect, even in the case of $\mathbb{R}^2$, as may be seen from Figure 1's example:

Example 1. Fenchel's example. The function $f(x, y) = y + \sqrt{x + y^2}$, graphed in Figure 1, cannot be concavified. This was first proposed by [5] and features in [2]. The function is strictly increasing in both x and y, and strictly concave in each variable separately. It is only weakly quasiconcave, however, because its level sets are straight lines, as shown in the right-hand side of Figure 1. To see that it

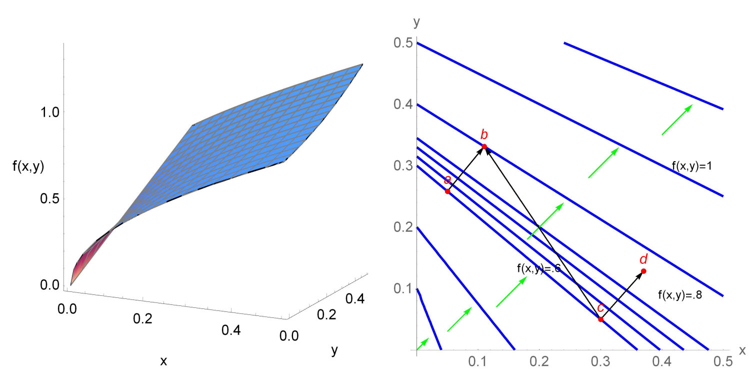


Figure 1. Fenchel's Example: Nonconcavifiable with Linear Level Sets.

cannot be weakly concavified, we note that for any postcomposition by a strictly increasing function g, the rate of increase of the function $g \circ f$ from point a to point b is greater than that of the same length segment from d to e. More precisely, the gradient at a is larger than the gradient at c. Hence, if we choose the point b sufficiently close to a, the level sets with values slightly larger than $g(f(a))$, but less than $g(f(b))$, must lie under the segment connecting $(c, g(f(c)))$ to $(b, g(f(b))$ near the point $(c, g(f(c))$. Hence part of the graph of $g \circ f$ lies over the line segment from $(c, g(f(c)))$ to $(b, g(f(b))$. Consequently, $g \circ f$ could not have been weakly concave.

Consider a function $f \colon I \to \mathbb{R}$ defined on an interval $I \subset \mathbb{R}$ which is either open (in which case it might be unbounded), half open, or closed. For the rest of the paper we will define $a = \inf(I)$ and $b = \sup(I)$.

The following lemma explains why we may restrict our attention to the case when the postcomposing function g is strictly increasing.

Lemma 1. *Given continuous functions $f \colon D \to \mathbb{R}$ with $D \subset \mathbb{R}$ connected and $g \colon \mathrm{Range(f)} \to \mathbb{R}$, for $g \circ f$ to be strictly quasiconcave it is necessary that f and g fall into precisely one of these four cases:*

(i) f is strictly increasing and g is strictly quasiconcave,
(ii) $-f$ is strictly increasing and $-g$ is strictly quasiconcave,
(iii) f is strictly quasiconcave but not monotone and g is strictly increasing, or
(iv) $-f$ is strictly quasiconcave but not monotone and $-g$ is strictly increasing.

Proof. The cardinality $\mathrm{card}((g \circ f)^{-1}(t))$ of the level set of value t for $g \circ f$ is $\sum_{s \in g^{-1}(t)} \mathrm{card}(f^{-1}(s))$. In particular, postcomposition by a function never reduces a level set's cardinality. The maximum cardinality of a level set of the strictly quasiconcave function $g \circ f$ is two. Hence on values where the level sets of f have cardinality two, g must be one-to-one, and g can only be two-to-one on values where f has cardinality one. Such continuous functions are fairly simple to analyze.

If all the level sets of f have cardinality one everywhere then f is monotone by continuity and we are in case (i) or (ii). Then g or $-g$ must be quasiconcave since the composition of the strictly increasing function f^{-1} (or $(-f)^{-1}$) with the quasiconcave function $g \circ f$ is again quasiconcave.

If the cardinality of the level sets of f is two on a subset $L \subset \mathrm{Range}(f)$, then consider any $p \in L$. Let $\{x, y\} = f^{-1}(p)$, with $x < y$, and let $\{U_i\}$ be any sequence of connected open intervals centered at p and strictly decreasing to p. By continuity of f, the preimages of each U_i under f are open and contain both x and y. Since each point has at most two preimages, for some sufficiently large i, the open subintervals A containing x and B containing y in $f^{-1}(U_i)$ must be distinct, and hence disjoint. If $C = f(A) \cap f(B)$, then f must be monotone on $f^{-1}(C) \cap A$ and $f^{-1}(C) \cap B$. Now C, being connected, either has interior, or else is just p. If it is just p, some points are to the right of x and have values less than p and some points are to the left of y and have values greater than p. Hence, the intermediate value theorem guarantees, since D is connected, another point $z \in (x, y) \subset D$ for which $f(z) = f(x) = f(y) = p$, which violates the cardinality restriction on the level sets. Therefore, we conclude that each point $p \in L$ contains an interval about it, a priori not necessarily in L, on which f is monotone. Since f is one-to-one off of L, f is locally monotone everywhere except for local extrema. The cardinality rule then implies that f has at most two extrema and is monotone on each connected segment after removing these points.

Now g preserves the local extreme points of f. So if f has more than one local extreme point, then the values must coincide under g, but then there are at least four preimages of some value near this extreme value for $g \circ f$, violating strict quasiconvexity. So f can have at most one local extremum. It has exactly one because f is not monotone and D is connected. In particular, L is connected.

Since L is connected, g is one-to-one on the closure of L, not just on L. This contains all of the local extrema of f. Now assume g is strictly increasing on this closure. If the graph of g changes direction elsewhere, then $g \circ f$ has at least two local extrema, violating quasiconcavity. Hence, g is strictly increasing everywhere. This then implies that f was strictly quasiconcave since postcomposition by a strictly increasing function g^{-1} preserves quasiconcavity.

If g was strictly decreasing on the closure of L then similarly it is strictly decreasing everywhere and $-f$ is strictly quasiconcave. □

Observe that if $f : (a, b) \to \mathbb{R}$ is not continuous, but quasiconcave with maximum at $m \in (a, b)$, then any concavifier g must be constant on any intervals that correspond to the (at most countable) jump discontinuities in the range of f on $(a, m]$. Let $g_1 : \mathbb{R} \to \mathbb{R}$ be the continuous function that is constant on these segments and is linear with slope 1 on all other intervals. Note that $g_1 \circ f$ must be continuous for f to be strictly concavifiable, since collapsing any further interval would yield a constant segment in $g \circ f$. (To achieve weak concavifiability g must be constant on $[f(y), f(m)]$ for any discontinuity y of $g_1 \circ f$.) Hence, it is no loss of generality to consider only continuous f.

3. Functions on $\mathbb{R}$

In this section we prove the classification theorem for the case of real valued functions, which will be generalized later. We treat this case by progessively removing regularity conditions. The reason for this being that the general case and final conditions will make sense from the constructions we develop en-route.

3.1. **The Case Where the Objective Function Is Twice-Differentiable and Strictly Monotone.**

If the continuous function f is strictly increasing or decreasing, then $f \colon I \to \mathbb{R}$ is invertible. Hence, we can easily solve the problem of concavifying f by choosing $g = h \circ f^{-1}$ where h is a concave function and hence $g(f(x)) = h(x)$ is concave. Here, however, we will treat the twice-differentiable case more intrinsically and connect the definition of concavity more viscerally with the properties of f. This will build a foundation for the next section, where we treat the noninvertible case.

Thus, let us move to the special case where f is twice differentiable and strictly increasing. We will also for now assume g is twice differentiable and strictly increasing. Recall that we are searching for a strictly concave function $g \circ f$. The twice differentiable function $g \circ f$ is strictly concave if the expression

$$(g \circ f)''(x) = g''(f(x)) \cdot f'(x)^2 + g'(f(x)) \cdot f''(x), \tag{3.1}$$

is nonpositive and never vanishes on an interval. Equivalently, this occurs if

$$\frac{g''(f(x))}{g'(f(x))} \le -\frac{f''(x)}{f'(x)^2}, \tag{3.2}$$

with equality never holding on any interval. (There are examples of strictly convex functions where equality holds on a full measure set, however.)

Since we assume that f is strictly increasing, it is invertible. Define $z = f(x)$, so $x = f^{-1}(z)$. Note that $\frac{1}{f'(x)} = \frac{1}{f'(f^{-1}(z))} = \frac{\partial}{\partial z} f^{-1}(z)$, so the right-hand side of inequality (3.2) is

(3.3)
$$-\frac{f''(x)}{f'(x)^2} = \frac{\partial}{\partial x}\left(\frac{1}{f'(x)}\right) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial(z = f(x))} f^{-1}(f(x))\right) = f^{-1\prime\prime}(f(x)) \cdot f'(x).$$

Taking another step:

$$-\frac{f''(x)}{f'(x)^2} = f^{-1\prime\prime}(f(x)) \cdot f'(x) = \frac{\partial}{\partial x} \log f^{-1\prime}(f(x)). \tag{3.4}$$

Similarly (though since we are constructing g we need not invert), the left hand side of inequality (3.2) is $\frac{\partial}{\partial f(x)} \log g'(f(x))$. Hence a sufficient criterion for inequality (3.1) to be true is that

$$\frac{\partial}{\partial z} \log g'(z) < \frac{\partial}{\partial z} \log f^{-1\prime}(z) \tag{3.5}$$

for all z in the range of f, provided both sides are well defined.

If we choose a number $c > 0$ and a function g so that

$$g'(z) = e^{-cz} \cdot f^{-1\prime}(z) \tag{3.6}$$

then

$$\log g'(z) = -cz + \log f^{-1\prime}(z) \tag{3.7}$$

and

$$\frac{\partial}{\partial z} \log g'(z) = -c + \frac{\partial}{\partial z} \log f^{-1\prime}(z) < \frac{\partial}{\partial z} \log f^{-1\prime}(z) \tag{3.8}$$

Integrating g' from equation (3.6) will produce the desired function.

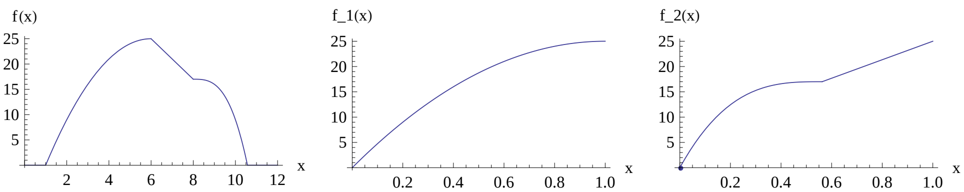
This approach, using Equation (3.6), has the advantage of relying only on first derivative data of f in its construction (although the properties requisite for its application do rely on the second derivatives of f.) Later we will make use of a second approach. To construct g, choose any function $u(z) < f'(f^{-1}(z)) \cdot f^{-1\prime\prime}(z)$. Then set

$$g(z) = \int_0^z \left(e^{\int_a^s u(t)dt}\right) ds \tag{3.9}$$

for any $a \in \mathbb{R}$ chosen at the top of the domain of $u(z)$. This yields a function with $g \circ f$ concave on $(a, d]$. (Note that shifting the function g by adding a constant is immaterial to its effectiveness.)

3.2. **The Case When the Objective Function Is Nonmonotonic but Twice Differentiable.** Now suppose the function f is twice differentiable and strictly quasiconcave but not monotone. In that case it achieves its maximum at a unique internal point $m \in (a, b)$, so that f is rising on $(a, m]$ and falling on $[m, b)$ as in Figure 2.

For now, we will also require that f have a non-vanishing derivative except at an internal maximum or endpoints.



(a) The Original f(x) (b) The Left−Side Function f_1(x) (b) The Right−Side Function f_2(x)

FIGURE 2. THE CONSTRUCTION OF f_1 AND f_2

Denote by f_1: $[0,1) \to \mathbb{R}$ the strictly increasing function

$$f_1(x) = f(a(1-x) + xm)$$

and by f_2: $[0,1) \to \mathbb{R}$ the strictly increasing function

$$f_2(x) = f(b(1-x) + xm).$$

Figure 2 illustrates this construction, which splits $f(x)$ into two strictly increasing functions on $[0,1]$ to save the bother of using negative signs and absolute values of slopes in our analysis. (In Figure 2 the $f(x)$ is not twice differentiable; it is drawn with a kink to illustrate the f_1, f_2 construction clearly.)

The new functions f_1 and f_2 are homeomorphisms onto their images, so they have inverses f_1^{-1} and f_2^{-1}. Hence, by post-composition we can easily choose a g such that either $g \circ f_1$ or $g \circ f_2$ is strictly concave and smooth. The difficulty is in making $g \circ f$ concave on its entire domain— that is, to use the same function to concavify both f_1 and f_2— especially when f is nondifferentiable or not defined over a compact set. We will treat this general problem in the next section.

Since f is twice differentiable with first derivative bounded away from 0, and I is compact, the problem becomes easy in light of what we discovered in the previous section. Simply set

$$\text{(Concavifying function)} \quad g'(z) = e^{\int_0^z u(t)dt} \cdot {f_1^{-1}}'(z) \cdot {f_2^{-1}}'(z) \tag{3.10}$$

for any continuous function u: $\mathbb{R} \to \mathbb{R}$ satisfying

$$u(z) < \min\left\{0, -\frac{\partial}{\partial z} \log {f_1^{-1}}'(z), -\frac{\partial}{\partial z} \log {f_2^{-1}}'(z)\right\}. \tag{3.11}$$

so that

(3.12)

$$\frac{\partial}{\partial z} \log g'(z) = \frac{\partial}{\partial z} \log {f_1^{-1}}'(z) + \frac{\partial}{\partial z} \log {f_2^{-1}}'(z) + u(z) < \min\left\{\frac{\partial}{\partial z} \log {f_1^{-1}}'(z), \frac{\partial}{\partial z} \log {f_2^{-1}}'(z)\right\}.$$

Note that we used the nonvanishing derivative condition on f simply to guarantee the existence of u in that the right-hand side of (3.11) is bounded from below. Thus we can solve for g, yielding $g \circ f$ concave.

3.3. **The Case When the Function Is Nondifferentiable and Nonmonotonic but Continuous.** In the case of a nondifferentiable f, we would like, following equation (3.10), to form

$$g'(z) = e^{\int_0^z u(t)dt} {f_1^{-1}}'(z) \cdot {f_2^{-1}}'(z), \tag{3.13}$$

but in a distributional sense, since we can only rely on weak derivatives. For this we consider the Sobolev space $W^{k,p}$, the space of functions whose weak k-th derivatives belong to L^p. Since f_1^{-1} and f_2^{-1} are strictly increasing, they are absolutely continuous and live in $W^{1,1}$. However, $W^{1,1}$ does not form an algebra, since it is not closed under multiplication of functions. This creates a problem as demonstrated in the following example.

Example 2. Suppose our quasiconcave function $f(x)$ was such that $f_1(x) = f_2(x) = x^3$. The strictly increasing function $f_1^{-1}(x) = f_2^{-1}(x) = x^{\frac{1}{3}}$ belong to $W^{1,1}$ on $[-1, 1]$. This has derivative ${f_1^{-1}}'(x) = \frac{1}{3}x^{-\frac{2}{3}} \in L^1$, but the product we would have for our construction in equation (3.13) is ${f_1^{-1}}'(z) \cdot {f_2^{-1}}'(z) = \frac{1}{9}x^{-\frac{4}{3}}$, which is not in L^1, and integrating it to get g yields $-\frac{1}{3}x^{-\frac{1}{3}}$, which is not even increasing on $[-1, 1]$. On the other hand, this f is easily concavified by $g(y) = -y^{2/3}$.

Hence, simply taking the product ${f_1^{-1}}'(z) \cdot {f_2^{-1}}'(z)$ for $g'(z)$ will not always work. If we assumed that each f_i^{-1} was in $W^{1,p}$ for $p \geq 2$, then the product would be in $W^{1,1}$. Though, if we allow one to be in $W^{1,1}$, then the other would necessarily have to be in $W^{1,\infty}$, and this is a stronger assumption than we need, since $W^{1,\infty}$ coincides with the space of Lipschitz functions and we know that there are non-Lipschitz quasiconcave functions that can be concavified (e.g. $f(x) = x^{\frac{1}{3}}$ on $[-1, 1]$).

Conversely, if we wanted to work directly on weak second derivatives to guarantee that $\frac{\partial}{\partial z} \log g'(z) < \min\left\{\frac{\partial}{\partial z} \log {f_1^{-1}}'(z), \frac{\partial}{\partial z} \log {f_2^{-1}}'(z)\right\}$ for equation (3.12), we would need to work with f_i^{-1} in $W^{2,1}$. By the Sobolev embedding theorem, however, $W^{2,1} \subset W^{1,\infty}$ for one-dimensional functions.[1] Thus we gain nothing over working with Lipschitz functions.

In view of these problems, instead of weak derivatives, we will work with the following upper and lower derivatives $\overline{D}f, \underline{D}f \colon \mathbb{R} \to [-\infty, \infty]$ defined for a function f by,

$$\overline{D}f(x) = \limsup_{h \to 0} \frac{f(x+h) - f(x)}{h} \quad \text{and} \quad \underline{D}f(x) = \liminf_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{3.14}$$

These quantities always exist, if we allow for values of $-\infty$ and ∞, and $\overline{D}f(x) \geq \underline{D}f(x)$ with equality occurring if and only if the derivative of f exists at x, in which case both quantities coincide with $f'(x)$.

Note that if $w < x < y$ then the slope of the secant line between $(w, f(w))$ and $(y, f(y))$ lies between the values of the slopes of the secant lines from $(w, f(w)))$ to $(x, f(x))$ and $(x, f(x))$ to $(y, f(y)))$. Hence we have,

$$\overline{D}f(x) = \limsup_{\substack{|y-w| \to 0 \\ w \leq x < y}} \frac{f(y) - f(w)}{y - w} \quad \text{and} \quad \underline{D}f(x) = \liminf_{\substack{|y-w| \to 0 \\ w \leq x < y}} \frac{f(y) - f(w)}{y - w}. \tag{3.15}$$

[1] For the general Sobolev Embedding Theorem, see Chapter 2 of [1]. For $W^{2,1}(\mathbb{R}) \subset W^{1,\infty}(\mathbb{R})$, see Exercise 2.18 in [15].

In other words, these quantities reflect the lower and upper limit of the slopes of all secant lines between points before and after x, not just those with an endpoint at x. Also, we cannot dispense with the ordering $w \leq x < y$ in the above limits; for example, the continuous extension of the function $x^2 \sin(\frac{1}{x^2})$ has derivative 0 at 0 and yet admits secant lines of unboundedly positive and negative slope whose endpoints are arbitrarily close to 0.

We now begin to explore analogues of conditions for concavity of C^2 functions using the above objects that are available to us for arbitrary continuous functions. For what follows let $\ell(x, y)$ represent the secant line between $(x, f(x))$ and $(y, f(y))$, and let $s(x, y)$ represent the slope of $\ell(x, y)$. In the next three lemmas, we will explore the relationship between concavity and conditions on $\overline{D}f$. (Analogous statements invoking $\underline{D}(f)$ could also naturally be formulated.)

Lemma 2. *A continuous function $f : I \to \mathbb{R}$ is strictly concave if and only if for all $x \in I$, and all $w < x$, $\overline{D}f(x) < s(w, x)$.*

Proof. If f is concave, then for any $y > w$ in I we have $s(x, y) < s(x, w)$ and hence taking the lim sup as $y \to x$ we obtain the forward implication. Conversely, if f is not convex then there exist points $r < s < t$ in I such that $f(s)$ lies below the secant line $\ell(r, t)$. By continuity, one may trace the graph in both directions from $(s, f(s))$ until it runs into the segment $\ell(r, t)$ showing that there is some open interval $(w, x) \subset (s, t)$ such that the entire graph of f over (w, x) lies strictly below the secant line $\ell(w, x)$. For all $z \in (w, x)$, we then have $s(z, x) > s(w, x)$ and therefore $\overline{D}f(x) \geq s(w, x)$ contradicting our hypothesis. □

Lemma 3. *A continuous function $f : I \to \mathbb{R}$ is strictly concave if and only if $\overline{D}f$ is a strictly decreasing function.*

Proof. If f is concave, then for any three points $w < x < y$ in I we have $s(w, x) > s(w, y)$ and $s(w, y) > \overline{D}f(y)$ by Lemma 2. Taking the lim sup as w approaches x from below we see that $\overline{D}f(x) \geq s(w, y) > \overline{D}f(y)$, as desired.

Conversely, if f is not concave we can find, as in the proof of Lemma 2, points $w < x$ for which the entire graph of f over (w, x) lies strictly below the secant line $\ell(w, x)$. After possibly shrinking this neighborhood we may assume the graph of f changes sides of the secant line $\ell(x, w)$ at both w and x. Then for any point z sufficiently near w, and any point y sufficiently near x, we have $s(z, w) < s(w, x) < s(x, y)$. Taking limsup's as z approaches w and y approaches x, we obtain $\overline{D}f(w) \leq \overline{D}f(x)$, contradicting our hypothesis. □

We now combine Lemma 3 with the fact that one can always control, after post-composition, the Lipschitz constant of f over any interval where f is monotone. This yields the following necessary criterion for strict quasiconcavity of f. First recall that a was defined as the lower bound of the support of f and m as its argmax in Section 3.2. Also, for the remainder of the paper, let $f\rfloor_{(a,m]}$ denote the restriction of the function f to the interval $(a, m]$.

Lemma 4. *Given a strictly quasiconcave function $f\colon (a, b) \to \mathbb{R}$, there is a $g\colon \mathbb{R} \to \mathbb{R}$ such that $g \circ f$ is strictly concave only if there is a function $h\colon \mathbb{R} \to \mathbb{R}$ such that $h \circ f$*

satisfies,

$$
\begin{cases} 0 < \underline{D}(h\circ f)(x) \le \overline{D}(h\circ f)(x) < \infty & \text{for } x\in(a,m) \\ \\ -\infty < \underline{D}(h\circ f)(x) \le \overline{D}(h\circ f)(x) < 0 & \text{for } x\in(m,b). \end{cases} \tag{3.16}
$$

In particular, $h\circ f$ must be (locally) Lipschitz except at m.

Proof. Suppose there is no such h. We can compose f by the function $h = f^{-1}_{\downarrow_{(a,m]}}$ so that $h\circ f$ is still strictly quasiconcave, but $(h\circ f)_1$ is linear. By hypothesis, $(h\circ f)_2$ either admits a vertical tangency on the pre-image of the range of $(h\circ f)_1$, or else $\underline{D}(h\circ f)_2(x) = 0$ for some $x\in(0,1)$. In the latter case, if we had instead chosen $h = -f^{-1}_{\downarrow_{[m,b)}}$ then $(h\circ f)_2$ would be linear and $(h\circ f)_1$ would admit a vertical tangency, and so we are back in the first case after switching "1"and "2". Hence, without loss of generality we may assume that there is a point $x\in(0,1)$ with $\overline{D}(h\circ f)_2(x) = \infty$. (Recall here that $(h\circ f)_2$ is increasing, see Figure 2.)

Since $(h\circ f)_1$ is the identity, any strictly concavifying g for $h\circ f$ must be concave and strictly increasing, and hence with $\underline{D}(g)(z) > 0$ for any z in the interior of the range of $(h\circ f)_1$. Then, however, it could not have concavified $(h\circ f)_2$. □

The function h in Lemma 4 can also be taken to be the inverse of the restriction to the strictly increasing side, so $h = f^{-1}_{\downarrow_{[m,b)}}$.

We shall see that the necessary conditions in Lemma 4 turn out to not be sufficient for concavifiability. A significantly more subtle problem arises. Consider the following example.

Example 3. A positive log-derivative with unbounded variation. Consider the strictly quasiconcave function $f(x)$ on $[-1,4]$ shown in Figure 3, which is defined as follows.

$$
f(x) = \begin{cases} q(x) & -1\le x<1 \\ q(1) - \frac{1}{2}(x-3)(x-1)q'(1) & 1\le x\le 4 \end{cases}
$$

where

$$
q(x) = \int_{-1}^{x} e^{t\sin\left(\frac{1}{t}\right)+1}\,dt
$$

From the formula we can readily verify that the first derivative,

$$
f'(x) = \begin{cases} e^{x\sin(1/x)+1} & -1\le x<1 \text{ and } x\ne 0 \\ e & x=0 \\ e^{1+\sin(1)}(2-x) & 1\le x\le 4 \end{cases},
$$

is a C^1 function with derivative bounded away from 0, except at the peak of f at $x=2$, and is strictly concave on $[1,4]$. Nevertheless, on the interval $(-1,1)$, we have $\log f'(x) = x\sin(\frac{1}{x}) + 1$ which is a classic example of a function with unbounded variation. While not obvious, Theorem 1 will show that such an f cannot be concavified by any postcomposition.

Recall that the variation of a function $f\colon [a,b]\to\mathbb{R}$ is defined as

$$
\mathrm{Var}(f) = \sup\left\{ \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \colon\ n\in\mathbb{N} \text{ and } a\le x_0 < x_1 < \cdots < x_n \le b \right\}. \tag{3.17}
$$

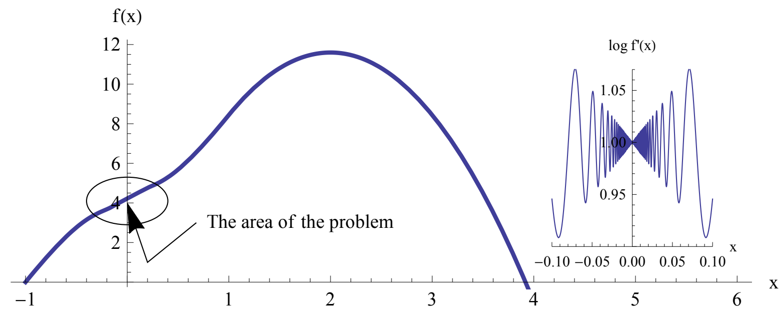


FIGURE 3. A Nonconcavifiable Strictly Quasiconcave Function with Strictly Positive Derivatives but Unbounded Variation

Denote the functions of bounded variation on the closed interval $[a, b]$ by

$$\mathrm{BV}([a,b]) = \{f\colon [a,b] \to \mathbb{R}\colon \mathrm{Var}(f) < \infty\}\,. \tag{3.18}$$

For a general interval $I \subset \mathbb{R}$, denote by $\mathrm{BV}_{loc}(I)$ the set of *locally* bounded variation functions, i.e. those which belong to $\mathrm{BV}([a,b])$ for every compact interval $[a,b] \subset I$.

In Theorem 1 below, without loss of generality, we will assume that our function $f\colon (a,b) \to \mathbb{R}$ has the property that if either f or $-f$ is quasiconcave, then $\mathrm{Range}(f\rfloor_{(a,m]}) = \mathrm{Range}(f)$ where m represents the unique extremal point (the argmax or argmin). (Otherwise, just replace $f(x)$ by its reflection about $\frac{b+a}{2}$, namely $f(b+a-x)$. The resulting g will concavify the original f.)

Theorem 1. *For any continuous nonmonotonic function $f\colon (a,b) \to \mathbb{R}$, there is a function $g\colon \mathrm{Range(f)} \to \mathbb{R}$ such that $g \circ f$ is strictly concave if and only if*

(i) f or $-f$ is strictly quasiconcave.
(ii) for $h = f^{-1}_{\rfloor_{(a,m]}}$, the function $h \circ f\rfloor_{(m,b)}$ and its inverse are locally Lipschitz.
(iii) $\log\left|\overline{D}(h \circ f)\right| \in \mathrm{BV}_{loc}((m,b))$.

Moreover, when g exists it is strictly monotone.

Proof. We will first explain the necessity of (i). If $f(x)$ is not strictly quasiconcave, then there exist points w, y, z such that $w < y < z$ and one of the following three conditions holds:

(a) $f(y) < \min(f(w), f(z))$ (if $f(x)$ is not even weakly quasiconcave)
(b) $f(y) = \min(f(w), f(z))$ and $f(w) \neq f(z)$
(c) $f(y) = f(w) = f(z)$.

In any of these cases, for any strictly monotonic function g, $g(f(y))$ will be below a straight line connecting $g(w)$ and $g(z)$ and hence by definition is not concave, because:

$$g(f(y)) \leq \left|\frac{w-y}{w-z}\right| g(f(w)) + \left|\frac{y-z}{w-z}\right| g(f(z)) \tag{3.19}$$

In case (a), this is because $g(f(y))$ is less than either $g(f(w))$ or $g(f(z))$, so the inequality in (3.19) is strict. In case (b), $g(f(y))$ is equal to one of the other two g's and less than the other, so inequality (3.19) is again strict. In case (c), $g(f(y))$ is equal to both the other two g's, so the inequality becomes an equality.

Lemma 4 demonstrates the necessity of condition (ii) on the upper derivatives. Necessity of condition (iii) will be deferred until the end of the proof.

To prove sufficiency without loss of generality we may assume f, as opposed to $-f$, is strictly quasiconcave. The existence of g implies that it is continuous, since f and $g \circ f$ are. Moreover, $\overline{D}(h \circ f)$ is nonpositive on (m, b) since f is strictly decreasing there and h is strictly increasing on its domain. (This accounts for taking the absolute value in condition (iii), which is unnecessary in the case when $-f$ is quasiconcave.) Since $\log \left|\overline{D}(h \circ f)\right| \in \mathrm{BV}_{loc}((m, b))$, it is a standard fact, e.g. see [6], that $\log \left|\overline{D}(h \circ f)\right|$ is the difference of two strictly increasing functions. Also, $\log \left|\overline{D}(h \circ f)\right|$ is continuous except at a countable set of points such that the sizes of the jumps at the discontinuities on any compact interval are summable. Now $\log |(h \circ f)'|$ agrees with this function wherever it is defined, which is almost everywhere since $h \circ f$ is monotone.

By Darboux's Theorem, and since a full measure set is dense, $\log |(h \circ f)'|$ then agrees with $\log \left|\overline{D}(h \circ f)\right|$ at each point where $\log \left|\overline{D}(h \circ f)\right|$ is continuous. Since $-\log |(h \circ f)'(x)| = \log |((h \circ f)^{-1})'(h \circ f(x))|$, it also agrees with $\log \left|\overline{D}(h \circ f)^{-1} \circ h \circ f\right|$ except at a countable number of points where $\log \left|\overline{D}(h \circ f)^{-1} \circ h \circ f\right|$ has discontinuities with summable gaps. Thus, since precomposition does not affect the BV property, except for the domain over which it applies, $\log \left|\overline{D}(h \circ f)^{-1}\right| \in \mathrm{BV}_{loc}\left((h(f(m)), h(f(b)))\right)$.

Let $h_1 = h_0 \circ h$, where h_0 is a smooth strictly increasing concave function on $(-\infty, f(m)]$ with $\lim_{x \to m} \overline{D}(h_0 \circ f)(x) = 0$. This can always be done by using an h_0 that increases sufficiently slowly near $f(m)$.

From now on in the proof, we will write f for $f\rfloor_{[m,b)}$ to avoid distraction from the subscript. Since the derivative of h_0 is bounded away from 0 and ∞ and is strictly decreasing on any compact subinterval of $(f(b), f(m))$, the function $\log \left|\overline{D}((h_1 \circ f)^{-1})\right|$ still lies in $\mathrm{BV}_{loc}((h_1(f(b)), h_1(f(m))))$ and $h_1 \circ f$ is concave on $(a, m]$.

Now choose $z_0 \in (h_1(f(b)), h_1(f(m)))$. Since

$$\log \left|\overline{D}((h_1 \circ f)^{-1})\right| \in \mathrm{BV}_{loc}\left((h_1(f(b)), h_1(f(m)))\right),$$

there is a representative

$$q \in L^1_{loc}\left((h_1(f(b)), h_1(f(m)))\right) \tag{3.20}$$

of the almost everywhere defined function $\left(\log \left|\overline{D}((h_1 \circ f)^{-1})\right|\right)'$ such that

$$\log \left|\overline{D}((h_1 \circ f)^{-1})(z)\right| = \log \left|\overline{D}((h_1 \circ f)^{-1})(z_0)\right| + \int_{z_0}^{z} q(t)\, dt.$$

Since $\lim_{z \to m} \overline{D}(h_0 \circ f)(z) = 0$, the negative part of q, namely

$$q^-(x) = \begin{cases} q(x) & q(x) < 0 \\ 0 & q(x) \geq 0 \end{cases}, \tag{3.21}$$

belongs to $L^1([h_1(f(c)), h_1(f(m))])$ for any $c < b$. By integrating, we can find a twice-differentiable (though not necessarily in C^2) function g_0: $(h_1(f(b)), h_1(f(m))) \to \mathbb{R}$ such that

$$g_0'(z) = e^{\int_{h_1(f(m))}^{z} (-1 + q^-(t))\, dt}. \tag{3.22}$$

Consequently, $g_0'(z) > 0$, $g_0''(z) < 0$ and $(\log g_0')' < q^-(z)$ for each $z \in (h_1(f(b)), h_1(f(m)))$. Since by construction $(\log g_0')'(z) < (\log |((h_1 \circ f)^{-1})'|)'(z)$ for almost every $z \in$

$(h_1(f(b)), h_1(f(m)))$, it follows that $\overline{D}(g_0 \circ h_1 \circ f)$ is strictly decreasing on (a, b) and so $g_0 \circ h_1 \circ f$ is concave, which is what we needed to prove. Lastly note that $g = g_0 \circ h_1$ is strictly monotone since h_1 is and $g_0' > 0$.

All that remains to be proved is the necessity of condition (iii). If $\log \overline{D}(h \circ f) \notin \mathrm{BV}_{loc}((m, b))$, then no such function $q \in L^1$ can be found: there exists no g_0 for which $\log g_0'$ grows slower than $\log(h_1 \circ f)'$ since $\log g_0'(z)$ would necessarily become unbounded before z reached $h_1(f(b))$. □

Remark 1. Theorem 1 shows that quasiconcavity is not quite equivalent to concavifiability. In addition, we require condition (ii), which roughly says that after straightening out one side, the other side has no horizontal or vertical tangencies. And beyond that, one still needs the yet more subtle condition (iii) governing the oscillation of the derivative on the unstraightened side.

Strictly speaking, condition (ii) is superfluous in that it only serves to establish the existence of the function in condition (iii), where its existence is implicit. In particular, we need $\log \left|\overline{D}(h \circ f)\right|$ to exist almost everywhere in order to make sense of it being in BV_{loc}. Once it belongs to BV_{loc} we can conclude that $h \circ f$ on (m, b), and its inverse, are locally Lipschitz.

4. Functions on an Arbitrary Geodesic Metric Space

We now extend our results to $\mathbb{R}^n$ and more general geodesic metric spaces.

Definition 2. A function $f\colon X \to \mathbb{R}$ on a geodesic metric space is *(strictly/weakly) quasiconcave* if and only if $f \circ \gamma\colon [0, 1] \to \mathbb{R}$ is (strictly/weakly) quasiconcave for every geodesic $\gamma\colon [0, 1] \to \mathbb{R}$, where we assume that γ is parameterized so that $d(\gamma(s), \gamma(t)) = |t - s|\, d(\gamma(0), \gamma(1))$ for all $s, t \in [0, 1]$.

Similarly, f is *(strictly/weakly) concave* if and only if for each geodesic γ: $[0, 1] \to X$, $f \circ \gamma$ is (strictly/weakly) concave as a function on $[0, 1]$. (Note that this definition generalizes standard concavity for $X = \mathbb{R}^n$ with the Euclidean metric.)

In what follows, we let $m \in X$ be the unique point maximizing f if f is quasiconcave or minimizing f if $-f$ is quasiconcave. We will denote the negative part of F by F^- as in (3.21).

We now state the complete criterion for concavification of quasiconcave functions, generalizing Theorem 1.

Theorem 2. *Let X be any geodesic metric space. For any continuous function $f\colon X \to \mathbb{R}$ there is a function $g\colon \mathrm{Range(f)} \to \mathbb{R}$ such that $g \circ f$ is strictly concave if and only if:*

(i) Either f or $-f$ is strictly quasiconcave;

(ii) $h \circ f$ is locally Lipschitz on $X - \{m\}$ where $h = (f \circ \gamma_o)^{-1}$ for some geodesic γ_o in X ending at m such that $\mathrm{Range}(f \circ \gamma_o) = \mathrm{Range}(f)$. Moreover $\overline{D}(h \circ f \circ \gamma)$ does not vanish on any geodesic segment $\gamma\colon (0, 1) \to X$ for which $h \circ f \circ \gamma$ is strictly increasing.

(iii) The total variation of $\log \overline{D}(h \circ f)$ along all geodesics $\gamma\colon [0, 1] \to X$ for which $h \circ f \circ \gamma$ is strictly increasing is uniformly bounded away from the extrema of $h \circ f$, or in other words,

$$\inf_{\gamma} \left\{ [\overline{D}(\log \overline{D}(h \circ f \circ \gamma)^{-1})]^- \right\} \in L^1_{loc}(R). \tag{4.1}$$

(Here R is the interior of the range of f and the infimum is taken over all geodesic segments $\gamma\colon [0,1] \to X$ for which $h \circ f \circ \gamma$ is strictly increasing.)

Proof. Suppose first that the conditions are met. Let q be the infimal function

$$q = \inf_{\gamma} \left\{ [\overline{D}(\log \overline{D}(h \circ f \circ \gamma)^{-1})]^{-} \right\}.$$

From the proof of Theorem 1, any function g_0 such that $(\log g_0')' < q$ (pointwise) will concavify $h \circ f \circ \gamma$ for each such γ. Choose g_0 such that $(\log g_0')' = -1 + q$. (Note that we can extend g_o to a function at the endpoints of R as well.)

Observe that every segment $\gamma\colon [0,1] \to X$ contains a subsegment $[s,1]$ where $h \circ f \circ \gamma$ is strictly increasing on $[s,1]$. Moreover, $h \circ f \circ \gamma(1-t)$ is also strictly increasing for $t \in [1-s,1]$. We now note that a strictly quasiconcave function that is concave on both its strictly increasing and strictly decreasing part separately is concave. Hence, the function g_0 concavifies $h \circ f \circ \gamma$ for every geodesic γ, and so $g_0 \circ h \circ f$ is concave. (This applies even to geodesics through m since $g_0 \circ h \circ f$ is concave on every subinterval on either side of m.) Taking $g = g_0 \circ h$ finishes this direction of the proof.

Conversely, suppose that there is a function g such that $g \circ f$ is concave. Since h is invertible, we may write $g \circ f = g \circ h^{-1} \circ h \circ f$, and set $g_o = g \circ h^{-1}$. By concavity of $g \circ f$ along γ_o, we have that $g_o = g \circ f \circ \gamma_o$ is convex. In particular it C^1 with $\log(g_o)'$ Lipschitz with derivative belonging to $L^1_{loc}(R)$. Moreover, for any geodesic $\gamma\colon [0,1] \to X$, with $f \circ \gamma$ strictly increasing, we have $(\log(g_o)')' \le q_\gamma$ where $q_\gamma = [\overline{D}(\log \overline{D}(h \circ f \circ \gamma)^{-1})]^{-}$, the comparison holding almost everywhere. Taking infima over all such γ implies $(\log(g_o)')' \le q$ as desired. □

Put crudely, Theorem 2 says that in any geodesic metric space, even nonseparable ones, the function f being strictly concavifiable by a strictly increasing g is equivalent to three conditions on f. First, f must be strictly quasiconcave. Second, after being straightened to linear along one geodesic spanning the whole range, the resulting function must not be too flat or too steep in any direction. Lastly, the total variation of the log of the derivative along all geodesic segments must be bounded uniformly, away from the endpoints. Note that since there are typically an infinite number of geodesics in X, these criteria become much more restrictive than in the setting of Theorem 1. Fortunately, for the special case of smooth functions on manifolds, these conditions are both more easily satisfied and more easily verified.

5. Twice Differentiable Functions on a Riemannian Manifold M

Here we consider a C^2 n-dimensional Riemannian manifold M with its Riemannian connection ∇. We will make the assumption that the strictly quasiconcave function $f\colon M \to \mathbb{R}$ is twice differentiable. Later, for our last theorem, we will weaken this to functions belonging only to the Sobolev space $W^{2,1}$ (i.e. possessing weak second derivatives).

Let $f\colon M \to \mathbb{R}$ be strictly quasiconcave. Suppose ∇f does not vanish at a point x. Then some contractible open neighborhood U of x within the level set $f^{-1}(f(x))$ is a hypersurface of M. On U, choose an orthonormal framing $\{e_i(x)\}$ of the tangent bundle TM of M restricted to U such that (i) $e_1 = \frac{\nabla f}{\|\nabla f\|}$, the unit normal to U, and (ii) $e_2, \dots, e_n$ forms a diagonal basis for the second fundamental form (i.e. the shape operator) of the hypersurface $U \subset M$.

The Hessian of a function f on an arbitrary Riemannian manifold is the $(0,2)$ tensor $\mathrm{Hess}(f) = \nabla df$. Given any basis, $v_1, \ldots, v_n$ of the tangent space T_pM at the point p, the corresponding matrix of the Hessian at p has entries

$$f_{ij} = \langle \nabla_{v_i}(\nabla f), v_j \rangle = \nabla_{v_i} \langle \nabla f, v_j \rangle - \langle \nabla f, \nabla_{v_i} v_j \rangle , \tag{5.1}$$

where on the right hand side we have extended the basis of vectors to local vector fields, though it is independent of the extension.

This matrix depends on the metric, and not just on the smooth structure (except at critical points of the function f, where $\nabla df = d^2 f$). Note, too, that $\mathrm{Hess}(f)$ is symmetric, which can be seen easily by extending the basis $\{v_i\}$ to a coordinate basis so that

$$f_{ij} - f_{ji} = v_i(df(v_j)) - v_j(df(v_i)) - df(\nabla_{v_i} v_j - \nabla_{v_j} v_i) = [v_i, v_j](f) - df([v_i, v_j]) = 0. \tag{5.2}$$

Thus equipped, we can present a necessary and sufficient condition for concavifiability of a twice-differentiable quasiconcave function f by postcomposing with a twice-differentiable function g. Our theorem will apply to C^2 Riemannian manifolds; that is, manifolds admitting C^2 charts for which the metric tensor coefficients are also C^2 functions. It will depend on the principal curvatures $\lambda_2(x), \ldots, \lambda_n(x)$ of the level sets of f, which are the eigenvalues of the second fundamental form at the point $x \in M$. These values (which are always positive for a strictly quasiconcave function) indicate the bending of the submanifold relative to the ambient manifold's curvature. For an $\mathbb{R}^2$ example to illustrate the theorem, look ahead to the example in Figure 4 after the proof.

Theorem 3. *Let M^n be a C^2 Riemannian manifold, possibly with boundary. For any twice-differentiable function $f\colon M \to \mathbb{R}$ there is a strictly increasing twice-differentiable $g\colon \mathbb{R} \to \mathbb{R}$ such that $g \circ f$ is strictly concave if and only if:*

(i) f is strictly quasiconcave;
(ii) ∇f does not vanish on the interior of M except possibly at the maximum point of f,
(iii) The function

$$q(t) = \inf_{x \in f^{-1}(t)} \frac{1}{\|\nabla f(x)\|^2} \left(-f_{11}(x) - \sum_{j=2}^{n} \frac{f_{1j}^2(x)}{\lambda_j(x) \|\nabla f(x)\|} \right) \tag{5.3}$$

has negative part q^- belonging to $L^1_{loc}(R)$, where R is the interior of the range of f.

A suitable g is

$$g(z) = \int_{f(m)}^{z} e^{\int_{f(m)}^{s} (-1+q^-(t))\, dt}\, ds. \tag{5.4}$$

Proof. Consider f in a neighborhood of a point $p \in M$. We assumed $\nabla_p f \neq 0$ which allows us to choose an orthonormal basis of T_xM for x in a neighborhood U of p as before, so that $e_1(x) = \frac{\nabla_x f}{\|\nabla_x f\|}$ and $e_2(x), \ldots, e_n(x)$ is a basis of $\mathrm{span}\{e_1(x)\}^{\perp}$ which diagonalizes the second fundamental form of the level set $f^{-1}(f(x))$ at the point $x \in U$. By quasiconcavity of f (condition (i)), these are all strictly positive. In terms of our basis we have $\lambda_j = -\left\langle \nabla_{e_j} e_1, e_j \right\rangle$ (see e.g. [3]).

For any twice-differentiable function $g\colon \mathbb{R} \to \mathbb{R}$, the Hessian of $g \circ f$ is given by $\nabla^2(g \circ f) = (g'' \circ f) df \otimes df + g' \circ f \operatorname{Hess}(f)$. We need to show that this is negative definite under the hypotheses.

Recall that in the above frame we computed the (i,j)-entry of the Hessian to be $f_{ij} = \langle \nabla_{e_i}(\nabla f), e_j \rangle = \nabla_{e_i} \langle \nabla f, e_j \rangle - \langle \nabla f, \nabla_{e_i} e_j \rangle$. By our choice of frame, for $j > 1$ the term $\langle \nabla f, e_j \rangle$ identically vanishes, and so

$$f_{ij} = -\langle \nabla f, \nabla_{e_i} e_j \rangle = -\|\nabla f\| \langle e_1, \nabla_{e_i} e_j \rangle = -\|\nabla f\| (\nabla_{e_i} \langle e_1, e_j \rangle - \langle \nabla_{e_i} e_1, e_j \rangle) = \|\nabla f\| (\langle \nabla_{e_i} e_1, e_j \rangle).$$

In particular $f_{ii} = -\|\nabla f\| \lambda_i$ when $i > 1$. Putting this together we compute the Hessian of $g \circ f$ to be,

(5.5)
$$\nabla^2(g \circ f) = (g'' \circ f) \begin{pmatrix} \|\nabla f\|^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} + (g' \circ f) \begin{pmatrix} f_{11} & f_{12} & f_{13} & \cdots & f_{1n} \\ f_{21} & -\lambda_2 \|\nabla f\| & 0 & \cdots & 0 \\ f_{31} & 0 & -\lambda_3 \|\nabla f\| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ f_{n1} & 0 & \cdots & 0 & -\lambda_n \|\nabla f\| \end{pmatrix},$$

Note we have $f_{1j} = -\langle \nabla f, \nabla_{e_1} e_j \rangle$ for $j > 1$, and moreover,

$$f_{11} = \frac{1}{\|\nabla f\|^2} \langle \nabla_{\nabla f} \nabla f, \nabla f \rangle = \frac{\nabla_{\nabla f} \|\nabla f\|^2}{2\|\nabla f\|^2} = \frac{\nabla_{\nabla f} \|\nabla f\|}{\|\nabla f\|} = \nabla_{e_1} \|\nabla f\|,$$

or, in other words, f_{11} is the growth rate of $\|\nabla f\|$ in the ∇f direction.

Similarly, since $\langle e_1, e_1 \rangle = 1$ identically, $\langle e_1, \nabla_{e_j} e_1 \rangle = \frac{1}{2} e_j(\langle e_1, e_1 \rangle) = 0$. Therefore,

$$f_{1,j} = f_{j,1} = \nabla_{e_j} \langle \nabla f, e_1 \rangle - \langle \nabla f, \nabla_{e_j} e_1 \rangle = \nabla_{e_j} \|\nabla f\|. \tag{5.6}$$

Since the values λ_i are all positive, we see that the principal minors, starting from the lower right, alternate sign. Hence in order to show that the eigenvalues of $\operatorname{Hess}(g \circ f)$ are all negative it remains to show that the sign of the entire determinant is $(-1)^n$.

Observe that for $j > 1$, the $(1,j)$-minor matrix of the combined matrix, formed by removing the first row and j-th column, can be made lower triangular by the following operations. We move the $j-1$-th row of the minor matrix, whose entry begins with f_{j1}, to the first row and shift all of the $j-2$ rows above the $j-1$-th row down by one. These operations introduce a $(-1)^{j-2}$ factor to the value of the minor, the determinant of the minor matrix, which is then $(-1)^{j-2}(g' \circ f)^{n-1} f_{j1} \lambda_2 \dots \lambda_{j-1} \lambda_{j+1} \dots \lambda_n (-\|\nabla f\|)^{n-2}$. In particular the $(1,j)$-entry times the $(1,j)$-cofactor for $j > 1$, namely $(-1)^{j-1} f_{1j}(g' \circ f)$ times the $(1,j)$-minor, is simply

$$(-1)^{n-1}(g' \circ f)^n \|\nabla f\|^{n-2} \frac{f_{1j}^2}{\lambda_j} \prod_{i=2}^n \lambda_i.$$

Hence the sum of these expressions from $j = 1$ to n, corresponding to the determinant expansion across the first row, yields the entire determinant of $\operatorname{Hess}(g \circ f)$, namely

(5.7)
$$\det \operatorname{Hess}(g \circ f) = (-1)^{n-1} \left(\frac{g'' \circ f}{g' \circ f} \|\nabla f\|^2 + f_{11} + \sum_{j=2}^n \frac{f_{1j}^2}{\lambda_j \|\nabla f\|} \right) (\|\nabla f\|)^{n-1} (g' \circ f)^n \prod_{i=2}^n \lambda_i.$$

For $j < n$ the j-th (lower corner) principal minor, namely the determinant of the last j rows and columns, is $(-1)^j (g' \circ f)^j \|\nabla f\|^j \prod_{i=n-j+1}^n \lambda_i$. This sequence in j has alternating sign since $\|\nabla f\|$, g' and the λ_i are all positive. In order for $\mathrm{Hess}(g \circ f)$ to be negative definite, it remains to show that expression (5.7) has sign $(-1)^n$. Since f_{1j}^2 is also positive, this happens if and only if

$$\frac{(g'' \circ f)}{(g' \circ f)} < \frac{1}{\|\nabla f\|^2} \left(-f_{11} - \sum_{j=2}^{n} \frac{f_{1j}^2}{\lambda_j \|\nabla f\|} \right), \tag{5.8}$$

which can be satisfied for a given f by a properly chosen g with $g' > 0$ and $g'' < 0$, provided that for almost every value t in the range of f, the quantity

$$\frac{1}{\|\nabla f\|^2} \left(-f_{11} - \sum_{j=2}^{n} \frac{f_{1j}^2}{\lambda_j \|\nabla f\|} \right)$$

is bounded below by a value $q(t) > -\infty$ on the level set $f^{-1}(t)$, and where $q^- \in L^1_{loc}$. By the theorem's assumption we have such a bound. The g function in the statement of the theorem then satisfies condition (5.8).

Conversely, ∇f must be bounded, away from the maximum point m, by the condition in Theorem 2. If we cannot find such a function q then we cannot obtain a g which everywhere satisfies the needed inequality, for the same reason as for the analogous result in the proof of Theorem 1.

□

Theorem 3 generalizes the "one-point" conditions of Fenchel [5] for $\mathbb{R}^n$ (as reformulated in Section 4 of [9]) to the Riemannian setting. Kannai's condition (I) on utility v corresponds to our condition (ii) on f. However he is allowing for weak concavifiability, which accounts for his necessary conditions (II) and (III) differing from our condition (i) when the sublevel sets of v are not strictly convex. Otherwise, these conditions are equivalent to our condition (i) and his conditions (IV) and (V) are folded into our condition (iii). This is best seen through the rephrasing of Kannai's condition (IV) as (IV′) and noting that his quantity "k" equals our $\|\nabla f\|$ and that under our assumptions in his setting when $M = \mathbb{R}^n$, we have $-\lambda_j \|\nabla f\| = f_{jj}$. Note also that Kannai's perspective is that of constructing a concave utility function based on weakly convex preference relations, whereas we start with an arbitrary function and see if it can be concavified.

Example 4. What condition (iii) excludes. Condition (iii) can be easily violated by a C^2 function f satisfying conditions (i) and (ii) by allowing for non-compact level sets which become asymptotically flat sufficiently quickly as points tend to infinity. A simple example is the quasiconcave function $f(x,y) = e^{e^x} y$ defined in the open positive quadrant of $\mathbb{R}^2$, shown in Figure 4. (While this is not a C^2 manifold with boundary, we could smooth the corner to make it so.) Its gradient, $\nabla f = \left(e^{x+e^x} y, e^{e^x}\right)$, is nonvanishing and its Hessian restricted to the level set of value t as a function of the x coordinate is $f_{22} = -\lambda_2 \|\nabla f\| = -\frac{te^x(e^x-1)}{t^2 e^{2x-2e^x}+1}$. Similarly, $f_{11} = \frac{te^{2x}\left(t^2 e^{x-2e^x}(e^x+1)+2\right)}{t^2 e^{2x-2e^x}+1}$ and $f_{12} = -\frac{e^{x+e^x}\left(t^2 e^x + e^{2e^x}\right)}{t^2 e^{2x}+e^{2e^x}}$. The negative definiteness shows that f is strictly quasiconcave. The quantity in condition (iii) on the level set of t works out to be $\frac{1}{t(e^{-x}-1)}$, whose infimum over $x > 0$ is always $-\infty$ for each t, and thus f is not concavifiable. Intuitively, a concavifying g must raise the values of the level sets along the y-axis, whereas much further along the x-axis on the same level sets g must squash the function f.

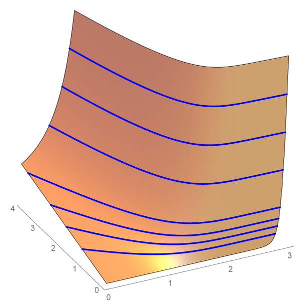


Figure 4. A function violating condition (iii) of Theorem 3 and some level sets

Remark 2. Since $f_{1j} = -\langle \nabla f, \nabla_{e_1} e_j \rangle$, in the special case that the integral curves of the vector field ∇f lie along geodesics of M, then $f_{1j} = 0$ for all $j > 1$. This occurs, for instance, when f is constant on distance spheres about a fixed point. In this case condition (iii) in Theorem 3 becomes vacuous.

Remark 3. Some twice-differentiable functions f with ∇f vanishing at points other than the maximum can also be concavified, provided we are willing to concavify using a g which is not twice differentiable. The more general condition is that after an initial postcomposition by a non-twice differentiable function g_o the resulting $g_o \circ f$ must satisfy conditions (ii) and (iii). In particular, when ∇f vanishes at a point, it must do so on the entire level set, though this alone is not sufficient.

Remark 4. In contrast to Theorems 1 and 2, here f is Lipschitz from the beginning, by virtue of being twice differentiable, and moreover $\log(h \circ f \circ \gamma)'$ automatically belongs to BV_{loc} for any twice differentiable increasing function h and geodesic γ under the assumption of condition (ii). Also, condition (iii) of Theorem 3 is vacuous for one-dimensional M and C^2 function f when condition (ii) holds. So applying the theorem to one-dimensional examples is pointless.

If f is C^2 with nonvanishing gradient, then the quantity (5.3) in the infimum of the definition of $q(t)$ in condition (iii) of Theorem 3 is uniformly bounded and continuous on compact sets. (Recall that the λ_j are uniformly positive on the convex compact level sets.) Moreover, the infimum of any compact family of continuous functions is always continuous. Hence, we immediately obtain that the variation function q from (5.3) is continuous if f is C^2 with compact level sets. We express this as the following especially simple corollary.

Corollary 1. *If $f\colon M \to \mathbb{R}$ is strictly quasiconcave and C^2, with compact level sets, then there is a C^2 strictly concavifying g if and only if ∇f does not vanish except possibly at f's global maximum and minimum points, if any.*

Remark 5. Fenchel's Example from Figure 1 does not satisfy the conditions of Corollary 1, because it is not strictly quasiconcave. In fact, for any function not strictly quasiconcave, at least one of the principal curvatures λ_i vanishes somewhere and thus quantity (5.3) becomes unbounded.

References

[1] Aubin, Thierry, *Nonlinear Analysis on Manifolds. Monge-Ampre equations.* Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 252, Berlin, New York: Springer-Verlag (1982).

[2] Aumann, Robert, "Values of Markets with a Continuum of Traders," *Econometrica*, 43:(1975) 611–646.

[3] Chavel, Isaac (1993) *Riemannian Geometry— A Modern Introduction*, Cambridge Tracts in Mathematics, vol. 108, Cambridge University Press, Cambridge (1993).

[4] DeFinetti, B., "Sulle Stratificazioni Convesse," *Annali di Matematica*, 30:(1949) 173–183.

[5] Fenchel, W., "Convex Cones, Sets, and Functions," unpublished notes, Princeton Mathematics Department, Princeton, New Jersey. (1953) http://rasmusen.org/x/abros/fenchel.pdf.

[6] Folland, G. B., *Real Analysis. Modern Techniques and Their Applications*, Pure and Applied Mathematics Series, New York: John Wiley & Sons Inc. (1984).

[7] Ginsberg, W., "Concavity and Quasiconcavity in Economics," *Journal of Economic Theory*, 6:(1973) 596–605.

[8] Guerraggio, Angelo & Elena Molho, "The Origins of Quasiconcavity: A Development between Mathematics and Economics," *Historia Mathematica*, 31:(2004) 62–75.

[9] Kannai, Yakar, "Concavifiability and Constructions of Concave Utility Functions," *Journal of Mathematical Economics*, 4:(1977) l–56.

[10] Kannai, Yakar, "Concave Utility Functions— Existence, Construction, and Cardinality," pp. 543-612 of *Generalized Concavity in Optimization and Economics,* ed. by Siegried Schaible and William T. Ziemba, New York: Academic Press, (1981).

[11] Kannai, Yakar, "Remarks Concerning Concave Utility Functions on Finite Sets," *Journal of Economic Theory* 26:(2005) 333–344.

[12] Papadopoulos, Athanase, *Metric Spaces, Convexity and Nonpositive Curvature*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 6, European Mathematical Society (EMS), Zürich (2005).

[13] Rapcsak, T., "Survey of the Fenchel Problem of Level Sets," in *Variational Analysis and Applications,* F. Giannessi and A. Maugeri (eds), pp. 935–956, Dordrecht, Holland: Kluwer Academic Publishers (2005).

[14] Richter, Marcel K. & K.-C. Kam-Chau Wong, "Concave Utility on Finite Sets," *Journal of Economic Theory,* 115:(2004) 341–357.

[15] Ziemer, William, *Weakly Differentiable Functions.* Graduate Texts in Mathematics, New York: Springer-Verlag (1989).

Department of Mathematics, Indiana University
E-mail address: `connell@indiana.edu`

Department of Business Economics and Public Policy, Kelley School of Business, Indiana University
E-mail address: `erasmuse@indiana.edu`