(b) The conjecture is true when v and w are scaled versions of each other. Show that $\mathbf{E}\, f(x_0 + tv)$ is monotone increasing in $t \geq 0$, when f is convex and v is zero mean.

3.11 *Monotone mappings.* A function $\psi : \mathbf{R}^n \to \mathbf{R}^n$ is called *monotone* if for all $x,\ y \in \mathbf{dom}\, \psi$,

$$(\psi(x) - \psi(y))^T (x - y) \geq 0.$$

(Note that 'monotone' as defined here is not the same as the definition given in §3.6.1. Both definitions are widely used.) Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is a differentiable convex function. Show that its gradient ∇f is monotone. Is the converse true, *i.e.*, is every monotone mapping the gradient of a convex function?

3.12 Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is convex, $g : \mathbf{R}^n \to \mathbf{R}$ is concave, $\mathbf{dom}\, f = \mathbf{dom}\, g = \mathbf{R}^n$, and for all x, $g(x) \leq f(x)$. Show that there exists an affine function h such that for all x, $g(x) \leq h(x) \leq f(x)$. In other words, if a concave function g is an underestimator of a convex function f, then we can fit an affine function between f and g.

3.13 *Kullback-Leibler divergence and the information inequality.* Let D_{kl} be the Kullback-Leibler divergence, as defined in (3.17). Prove the *information inequality*: $D_{\mathrm{kl}}(u, v) \geq 0$ for all $u,\ v \in \mathbf{R}^n_{++}$. Also show that $D_{\mathrm{kl}}(u, v) = 0$ if and only if $u = v$.

Hint. The Kullback-Leibler divergence can be expressed as

$$D_{\mathrm{kl}}(u, v) = f(u) - f(v) - \nabla f(v)^T (u - v),$$

where $f(v) = \sum_{i=1}^n v_i \log v_i$ is the negative entropy of v.

3.14 *Convex-concave functions and saddle-points.* We say the function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is *convex-concave* if $f(x, z)$ is a concave function of z, for each fixed x, and a convex function of x, for each fixed z. We also require its domain to have the product form $\mathbf{dom}\, f = A \times B$, where $A \subseteq \mathbf{R}^n$ and $B \subseteq \mathbf{R}^m$ are convex.

(a) Give a second-order condition for a twice differentiable function $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ to be convex-concave, in terms of its Hessian $\nabla^2 f(x, z)$.

(b) Suppose that $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is convex-concave and differentiable, with $\nabla f(\tilde{x}, \tilde{z}) = 0$. Show that the *saddle-point property* holds: for all x, z, we have

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z}).$$

Show that this implies that f satisfies the *strong max-min property*:

$$\sup_z \inf_x f(x, z) = \inf_x \sup_z f(x, z)$$

(and their common value is $f(\tilde{x}, \tilde{z})$).

(c) Now suppose that $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is differentiable, but not necessarily convex-concave, and the saddle-point property holds at $\tilde{x}, \tilde{z}$:

$$f(\tilde{x}, z) \leq f(\tilde{x}, \tilde{z}) \leq f(x, \tilde{z})$$

for all x, z. Show that $\nabla f(\tilde{x}, \tilde{z}) = 0$.

Examples

3.15 *A family of concave utility functions.* For $0 < \alpha \leq 1$ let

$$u_\alpha(x) = \frac{x^\alpha - 1}{\alpha},$$

with $\mathbf{dom}\, u_\alpha = \mathbf{R}_+$. We also define $u_0(x) = \log x$ (with $\mathbf{dom}\, u_0 = \mathbf{R}_{++}$).

(a) Show that for $x > 0$, $u_0(x) = \lim_{\alpha \to 0} u_\alpha(x)$.