Differentiable convex functions

$S \subseteq \mathbb{R}^n, \quad f: S \to \mathbb{R}$.

f is said to be *differentiable* at $\bar{x} \in \operatorname{int} S$ if there exists a vector $\nabla f(\bar{x})$, called the *gradient vector*, and a function $\alpha: \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \|x - \bar{x}\| \, \alpha(\bar{x}; x - \bar{x})$$

~~for each $x \in S$~~ $\quad \forall x \in S$,

where $\lim_{x \to \bar{x}} \alpha(\bar{x}, x - \bar{x}) = 0$.

f is *differentiable* on the open set $S' \subset S$ if it is differentiable at each point in S'.

If f is differentiable,

$$\nabla f(\bar{x}) = \left(\frac{\partial f(\bar{x})}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}(\bar{x}) \right)^T.$$

Theorem Let ~~S~~ C be a nonempty open convex set in $\mathbb{R}^n$.

Let $f: C \to \mathbb{R}$ be diff. on C.

Then f is convex if and only if for any $\bar{x} \in$ ~~S~~ C,

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}).$$

Proof Follows from Lemma and theorems.

[NB: Useful in, eg, Goffin & Vial approaches to convex optimization.]

Theorem $C \subseteq \mathbb{R}^n$ nonempty, convex, open. $f: C \to \mathbb{R}$ diff on C.

Then f is convex if and only if

~~$x_1, x_2 \in S \implies$~~ $\left[\nabla f(x_2) - \nabla f(x_1)\right]^T (x_2 - x_1) \geq 0$. $\forall x_1, x_2 \in S$

Proof. $\Rightarrow$: Assume f convex. Let $x_1, x_2 \in S$

Skip this, if doing proof on p 41a of previous theorem

So $$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1)$$

$$f(x_1) \geq f(x_2) + \nabla f(x_2)^T (x_1 - x_2)$$

Adding $$\left[\nabla f(x_2) - \nabla f(x_1)\right]^T (x_2 - x_1) \geq 0$$

$\Leftarrow$: Assume ~~$x_1, x_2 \in S \implies$~~ $\left[\nabla f(a) - \nabla f(b)\right]^T (a - b) \geq 0 \quad \forall a, b \in S$

Let $x_1, x_2 \in S$.

By MVT, there is some $x = \lambda x_1 + (1-\lambda) x_2 \quad (0 \leq \lambda \leq 1)$ s.t.

$$\nabla f(x)^T (x_2 - x_1) = f(x_2) - f(x_1).$$

[redacted]

Proof of first theorem without using subgradients:

⇒ Assume f is convex. Thus, the expression $\frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda}$ is monotonically increasing in λ, and

$$\frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda} \geq \lim_{\lambda \to 0+} \frac{f(\bar{x}+\lambda d)-f(\bar{x})}{\lambda} = \nabla f(\bar{x})^T d$$

⇐ Assume f is not convex:

So $\exists\, x_1, x_2, \lambda,\ 0<\lambda<1$ with $f(\lambda x_1+(1-\lambda)x_2) > \lambda f(x_1)+(1-\lambda)f($

Consider an interval ~~$[x_3, x_4] \subseteq [x_1, x_2]$ with~~

~~$f(x_3)$~~ $[\mu_1, \mu_2] \subseteq [0,1]$, with

(i) $\mu_1 \neq \mu_2$

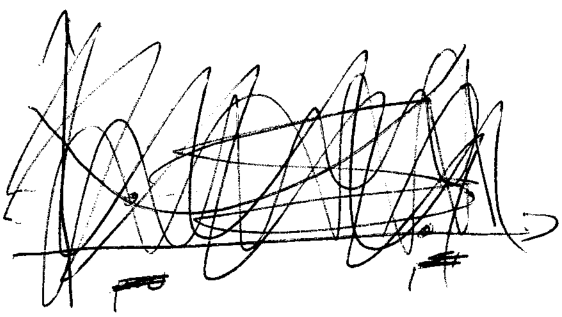
(ii) $f(\mu_1 x_1 + (1-\mu_1)x_2) = \mu_1 f(x_1) + (1-\mu_1) f(x_2)$

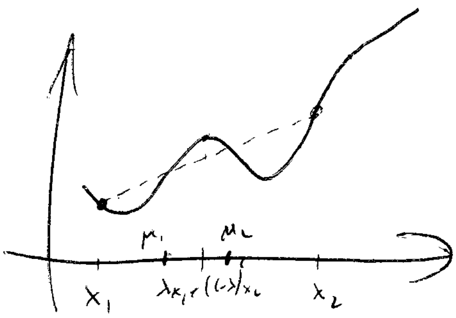
(iii) $f(\mu_2 x_1 + (1-\mu_2)x_2) = \mu_2 f(x_1) + (1-\mu_2) f(x_2)$

(iv) $f(\mu x_1 + (1-\mu)x_2) \geq \mu f(x_1) + (1-\mu) f(x_2), \quad \mu_1 \leq \mu \leq \mu_2$

Because λ exists, such an interval in μ must exist.

Don't need μ. Could jump straight to X_3.





Then $\exists\, \hat{\mu} \in [\mu_1, \mu_2]$ with $\nabla f(\underbrace{\hat{\mu} x_1 + (1-\hat{\mu})x_2}_{=:x_3})^T (x_2 - x_1) = f(x_2) - f(x_1)$

x_3 with $\nabla f(x_3) = \frac{f(x_2)-f(x_1)}{x_2-x_1}$ in 1D,

Then $\nabla f(x_3)^T (x_3 - x_1) < f(x_3) - f(x_1)$, or $f(x_1) < f(x_3) + \nabla f(x_3)^T (x_1 - x_3)$ ☐

$\nabla f(x_3)^T (x_2-x_1) = f(x_2) - f(x_1)$

By assumption, $[\nabla f(x) - \nabla f(x_1)]^T (x - x_1) \geq 0$

$$\Rightarrow \quad [\nabla f(x) - \nabla f(x_1)]^T (x_2 - x_1) \geq 0$$

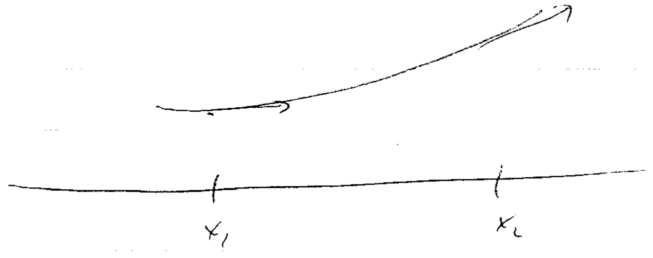
since $x_2 - x_1$ is parallel to $x - x_1$.

$$\therefore \quad \nabla f(x)^T (x_2 - x_1) \geq \nabla f(x_1)^T (x_2 - x_1)$$

$$\therefore \quad f(x_2) - f(x_1) \geq \nabla f(x_1)^T (x_2 - x_1)$$

$$\therefore \quad f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \quad //$$

This last theorem looks sort of second order, with differences in derivatives: as you move from x_1 to x_2, the component of the derivative in that direction increases.



Twice differentiable convex fcns

Defn Let $S \subseteq \mathbb{R}^n$, nonempty. $f: S \to \mathbb{R}$

f is twice differentiable ~~of~~ ~~[illegible]~~ at $\bar{x} \in \text{int } S$ if there exists a vector $\nabla f(\bar{x})$, the gradient vector, ~~and~~ an $n \times n$ matrix $\nabla^2 f(\bar{x})$ (or $H(\bar{x})$), the Hessian matrix, and a function $\alpha(\bar{x}; x - \bar{x})$ such that

$$f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T \nabla^2 f(\bar{x})(x - \bar{x})$$
$$+ \|x - \bar{x}\|^2 \alpha(\bar{x}; x - \bar{x}), \quad \forall x \in S$$

where $\lim_{x \to \bar{x}} \alpha(\bar{x}; x - \bar{x}) = 0$.

f is twice differentiable on open set $S' \subseteq S$ if it is twice differentiable at each point in S'. (also called smooth, C^2)

(i,j)th component of $\nabla^2 f(\bar{x})$ is $\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x})$.

THM $C \subseteq \mathbb{R}^n$ open, convex, nonempty

$f: C \to \mathbb{R}$, twice differentiable.

f is convex on $C \iff H$ is p.s.d. on C.

Pf $\Rightarrow$ f convex:

Take $\bar{x} \in S$

Need to show $d^T H(\bar{x}) d \geq 0 \quad \forall d \in \mathbb{R}^n$.

$\bar{x} + \lambda d \in S$ for λ small enough for any d.

$\therefore f(\bar{x} + \lambda d) \geq f(\bar{x}) + \lambda \nabla f(\bar{x})^T d$, since f convex.

Also $f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \nabla f(\bar{x})^T d + \lambda^2 d^T \nabla^2 f(\bar{x})^T d$
$+ \lambda^2 \|d\|^2 \alpha(\bar{x}; \lambda d)$

So $\lambda^2 d^T \nabla^2 f(\bar{x})^T d + \lambda^2 \|d\|^2 \alpha(\bar{x}; \lambda d) \geq 0$

Cancel λ^2.
Take limit as $\lambda \to 0$:
$d^T \nabla^2 f(\bar{x})^T d \geq 0$.

$\Leftarrow$ H is psd on C.

Given $x, \bar{x}$, ~~Taylor's theorem~~

By ~~[illegible]~~ MVT, we have ~~some~~ $\hat{x} = \lambda \bar{x} + (1-\lambda)x$ for some $\lambda \in (0,1)$

satisfying $f(x) = f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H(\hat{x}) (x - \bar{x})$.

~~for some $\hat{x} \in (x, \bar{x})$~~.

Characterizations of positive definiteness

Consider matrix M

By defn., M is positive definite if $x^T M x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0$

Equivalent to saying all eigenvalues of M are positive.

Also:

- M p.d $\Rightarrow$ All principal subdeterminants are positive
- M p.d. $\Rightarrow$ $\exists$ factorization LDU where L lower tri, U upper tri, diagonal entries 1, D positive diag matrix.

If M is symmetric:

M p.d $\Leftrightarrow$ $\exists$ Cholesky factorization $M = LL^T$, where all diag entries of L are +ve.

M p.d. $\Leftrightarrow$ All leading principal subdeterminants are positive.

M psd $\Leftrightarrow$ All principal subdeterminants are nonnegative.

$\left| \begin{matrix} 1 & 5 \\ -25 & 1 \end{matrix} \right| = 126$

... in GoVL, p 86, ...

$\hat{x} \in S$, so $H(\hat{x})$ is psd, so $(x-\hat{x})^T H(\hat{x})(x-\hat{x}) \geq 0$

$\therefore f(x) \geq f(\bar{x}) + Df(\bar{x})^T(x-\bar{x})$ //.

Useful for checking ~~concavity~~ convexity of a C^2 function.

If f quadratic, $H(x) \equiv H$, so only need to check if one matrix is psd.

eg: $f(x) = 3x_1^2 + x_2^2 + 2x_1x_2$:

$$= \frac{1}{2}(x_1, x_2)\begin{pmatrix} 6 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$H(x) = D^2 f(x) = \begin{pmatrix} 6 & 1 \\ 1 & 2 \end{pmatrix} \quad \forall x \in \mathbb{R}^2.$$

Eigenvalues of H are $4 \pm \sqrt{5} > 0$, so p.d.

Defn: ~~f is strictly convex if f(x~~

$C \subseteq \mathbb{R}^n$ convex, nonempty. $f: C \to \mathbb{R}$.

f is strictly convex if ~~$x_1, x_2 \in C$ $x_1 \neq x_2$,~~

~~$f(x_2) > f(x_1) +$~~

$$\left.\begin{array}{l} x_1 \in C \\ x_2 \in C \\ \lambda \in (0,1) \\ x_1 \neq x_2 \end{array}\right\} \Rightarrow f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2).$$