15 local extrema part 2

Thursday, October 22, 2020 12:34 PM

Let's consider only minima because the situation for maxima is analogous.

Prop 4.3/4.4 Let E be a normed vector space and let $J:\Omega \to \mathbb{R}$ be a function, with Ω some open subset of E. If J is differentiable in Ω, if J has a second derivative $D^2J(u)$ at some pt $u \in \Omega$, and if J has a local minimum at u, then

$$D^2J(u)(w,w) \geq 0 \quad \forall w \in E.$$

proof. Let $w \in E$, $w \neq 0$. For t small enough, $u + tw \in \Omega$ and $J(u+tw) \geq J(u)$,

so there is an open interval $I \subseteq \mathbb{R}$ s.t.

$$u + tw \in \Omega \quad \text{and} \quad J(u+tw) \geq J(w). \quad \forall t \in I.$$

Using the Taylor-Young formula, and that $dJ(u) = 0$,

$$0 \leq J(u+tw) - J(u) = \frac{t^2}{2} D^2J(u)(w,w) + t^2 \|w\|^2 \epsilon(tw),$$

with $\lim_{t \mapsto 0} \epsilon(tw) = 0$

$$\Longrightarrow D^2J(u)(w,w) \geq 0.$$

But this is true $\forall w \in E$, proving the claim. $\square$

Note of course that this theorem only goes one way. (e.g. $f(x) = x^3$)

Aside: If $E = \mathbb{R}^n$, this means that a necessary condition for having a local minimum is that the Hessian $\nabla^2 J(u)$ is SPD.

We've only talked about necessary conditions so far. Let's give some sufficient conditions.

Thm 4.3/4.5 Let E be a normed vector space, let $J:\Omega \to \mathbb{R}$ be a function with $\Omega \subseteq E$ open, and assume that J is differentiable

Thm [illegible] Let E be a normed vector space, [illegible] a function with $\Omega \subseteq E$ open, and assume that J is differentiable in Ω and $dJ(u) = 0$ at some pt $u \in \Omega$. Then

(1) If $D^2J(u)$ exists and if there is some $\alpha \in \mathbb{R}$ s.t. $\alpha > 0$ and $D^2J(u)(w,w) \geq \alpha \|w\|^2$ for all $w \in E$, then J has a strict local minimum at u.

(2) If $D^2J(v)$ exists for all $v \in \Omega$ and if $\exists$ ball $B \subseteq \Omega$ centered at u s.t. $D^2J(v)(w,w) \geq 0$ $\forall$ $v \in B$ and $w \in E$, then J has a local minimum at u.

proof. (1) By the Taylor-Young formula, (Thm 3.6)

$$J(u+w) - J(u) = \frac{1}{2} D^2J(u)(w,w) + \|w\|^2 \varepsilon(w)$$

$$\geq \left(\frac{1}{2}\alpha + \varepsilon(w)\right)\|w\|^2 \quad , \text{ with } \lim_{w \to 0} \varepsilon(w) = 0.$$

Thus, for small enough $r > 0$, $|\varepsilon(w)| < \frac{\alpha}{2}$ $\forall$ w s.t. $\|w\| < r$.

Then $J(u+w) > J(u)$ $\forall u+w \in B$, so J has a strict local min at u.

(2) By the Taylor-Maclaurin formula, (Thm 3.8) (generalized mean-value-theorem)

$\forall u+w \in B$, $J(u+w) = J(u) + \frac{1}{2} D^2J(v)(w,w) \geq J(u)$ for some $v \in (u, u+w)$. ▱

Intuitively, if the 2nd derivative exists at a critical pt, then it's a minimum if $D^2J(v) \geq 0$

Prop. 4.4/4.6 For any symmetric matrix A, if A is positive definite, then, $\exists$ $\alpha > 0$ s.t. $x^TAx \geq \alpha \|x\|^2$ $\forall x \in \mathbb{R}^n$.

proof. Pick any norm in $\mathbb{R}^n$.

Since the unit sphere $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is compact, and since the function $f(x) = x^TAx \neq 0$ on S^{n-1}, f has a minimum $\alpha > 0$.

Thus, $x^TAx \geq \alpha$ $\forall$ $\|x\| = 1$ (because $x = \|x\| \cdot \frac{x}{\|x\|}$)

$\Rightarrow x^TAx \geq \alpha \|x\|^2$ $\forall x \in \mathbb{R}^n$. ▱

Thus, this shows that for quadratic forms that are symmetric pos. def., Thus, this shows that for quadratic forms that are symmetric pos. def., there exists a strict local minimum.

Def. 4.6/4.3 Given a function $J : \Omega \to \mathbb{R}$ as before, a point $u \in \Omega$ is a nondegenerate critical pt if $dJ(u) = 0$ and if the Hessian matrix

$$\nabla^2 J(u) = \begin{pmatrix} \frac{\partial^2 J}{\partial x_1^2}(u) & \frac{\partial^2 J}{\partial x_1 \partial x_2}(u) & \cdots & \frac{\partial^2 J}{\partial x_1 \partial x_n}(u) \\ \frac{\partial^2 J}{\partial x_2 \partial x_1}(u) & \frac{\partial^2 J}{\partial x_2^2}(u) & \cdots & \frac{\partial^2 J}{\partial x_2 \partial x_n}(u) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J}{\partial x_n \partial x_1}(u) & \frac{\partial^2 J}{\partial x_n \partial x_2}(u) & \cdots & \frac{\partial^2 J}{\partial x_n^2}(u) \end{pmatrix}$$

is invertible.

(note that the Hessian is always symmetric)

Prop. 4.5/4.7 Let $J : \Omega \to \mathbb{R}$ be defined on an open subset $\Omega \subseteq \mathbb{R}^n$. If J is differentiable in Ω and if $u \in \Omega$ is a nondegenerate critical pt. s.t. $\nabla^2 J(u)$ is positive definite, then J has a strict local min at u.

Remark: Can generalize to infinite-dimensional spaces. Let E be a Banach space, and let $E' = \mathcal{L}(E; \mathbb{R})$ (continuous linear forms).

Let $\varphi : E \times E \to \mathbb{R}$ be $\varphi \in \mathcal{L}(E, E; \mathbb{R})$ (continuous bilinear map),

giving us a map $\Phi : E \to E'$ by $\Phi(u) = \varphi_u$, where $\varphi_u(v) = \varphi(u, v)$.

φ is nondegenerate iff $\Phi : E \to E'$ is an isomorphism of Banach spaces.

Given $J : \Omega \to \mathbb{R}$ differentiable on Ω, if $D^2 J(u)$ exists for some $u \in \Omega$, then u is a nondegenerate crit. pt. if $dJ(u) = 0$ and $D^2 J(u)$ is nondegenerate.

$D^2 J(u)$ is pos def. if $D^2 J(u)(w, w) > 0$ $\forall w \in E - \{0\}$.

Def. 4.7/4.4 Given any real vector space E, a subset $C \subseteq E$ is convex if either $C = \emptyset$ or if $\forall u, v \in C$, $(1-\lambda)u + \lambda v \in C$ for $0 \le \lambda \le 1$.

Notation: Line segment $[u, v] = \{(1-\lambda)u + \lambda v \in E \mid \lambda \in \mathbb{R},\ 0 \le \lambda \le 1\}$.

Def. 4.8/4.5 If $C \subseteq E$ is nonempty and convex, a function $f : C \to \mathbb{R}$ is

Def. 4.8/4.5 If $C \subseteq E$ is nonempty and convex, a function $f: C \to \mathbb{R}$ is

convex if $\forall u, v \in C$, $f((1-\lambda)u + \lambda v) \leq (1-\lambda) f(u) + \lambda f(v)$, $0 \leq \lambda \leq 1$.

strictly convex if $\forall u, v \in C$, $u \neq v$, $f((1-\lambda)u + \lambda v) < (1-\lambda) f(u) + \lambda f(v)$.

Define: The epigraph $epi(f)$ of a function $f: A \to \mathbb{R}$, $A \subseteq \mathbb{R}^n$, is a subset of $\mathbb{R}^{n+1}$ defined as

$$epi(f) = \{(x, y) \in \mathbb{R}^{n+1} \mid f(x) \leq y,\ x \in A\}.$$

A function $f: C \to \mathbb{R}$ is convex iff $epi(f)$ is convex.

We call a function $f: C \to \mathbb{R}$ concave if $-f$ is convex.

Thm 4.4/4.8 (Necessary cond. for local min on convex subset)

Let $J: \Omega \to \mathbb{R}$, $\Omega \subseteq E$ an open subset of a normed vector space.

Let $U \subseteq \Omega$ be a nonempty convex subset.

Given any $u \in U$, if $dJ(u)$ exists and if J has a local minimum in u w.r.t. U, then

$$dJ(u)(v-u) \geq 0 \quad \forall v \in U.$$

proof. Let $v = u + w$ be an arbitrary pt. in U.

By convexity of U, $u + tw \in U$ $\forall$ $0 \leq t \leq 1$.

$\lambda u + (1-\lambda)(u+w) \in U$

Since $dJ(u)$ exists, $J(u+tw) - J(u) = dJ(u)(tw) + \|tw\| \varepsilon(tw)$, $\lim_{t \to 0} \varepsilon(tw) = 0$.

$(0 \leq t \leq 1)$ $\Rightarrow J(u+tw) - J(u) = t\left[dJ(u)(w) + \|w\| \varepsilon(tw)\right]$.

Because u is a local min. w.r.t. U, $J(u+tw) - J(u) \geq 0$

$$\Rightarrow t\left[dJ(u)(w) + \|w\| \varepsilon(tw)\right] \geq 0.$$

Suppose $dJ(u)(w) < 0$. Then choose small enough $t > 0$ s.t.

$t\left[dJ(u)(w) + \|w\| \varepsilon(tw)\right] < 0$, a contradiction.

Thus, $dJ(u)(w) \geq 0 \Rightarrow dJ(u)(v-u) \geq 0$ $\forall v \in U$.

Thus, we can use convexity of U as a substitute for using Lagrange multipliers, though we now have an inequality instead of an equality.

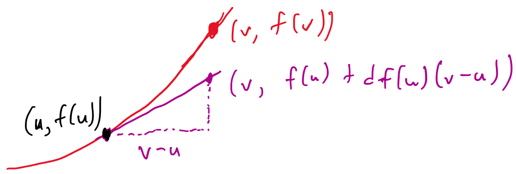
Corollary 4.1/ If $U \subseteq E$ is a subspace, if $dJ(u)$ exists and J has a local minimum w.r.t. U, then $dJ(u)(w) = 0 \quad \forall\, w \in U$.

Prop. 4.6/4.9 (Convexity and first derivative)

Let $f : \Omega \to \mathbb{R}$ be a function differentiable on an open subset $\Omega \subseteq E$, where E is a normed vector space, and let $U \subseteq \Omega$ be a convex subset.

(1) f is convex on U iff $f(v) \geq f(u) + df(u)(v-u)$ for all $u, v \in U$.

(2) f is strictly convex on U iff $f(v) > f(u) + df(u)(v-u) \quad \forall\, u, v \in U,\ u \neq v$.



proof. (1) Let $u, v \in U$ be distinct, and let $\lambda \in \mathbb{R}$, $0 < \lambda < 1$.

If f is convex, then

$$f((1-\lambda)u + \lambda v) \leq (1-\lambda) f(u) + \lambda f(v).$$

$$\Rightarrow \quad f((1-\lambda)u + \lambda v) - f(u) \leq \lambda f(v) - \lambda f(u)$$

$$\Rightarrow \quad \frac{f((1-\lambda)u + \lambda v) - f(u)}{\lambda} \leq f(v) - f(u)$$

$$\Rightarrow \quad df(u)(v-u) = \lim_{\lambda \to 0} \frac{f((1-\lambda)u + \lambda v) - f(u)}{\lambda} \leq f(v) - f(u).$$

Consider $f(v) \geq f(u) + df(u)(v-u) \quad \forall\, u, v \in U$

Then $\forall\, u \neq v$, $0 < \lambda < 1$, we have

(1) $\quad f(v) \geq f(v + \lambda(u-v)) - \lambda\, df(v + \lambda(u-v))(u-v) \qquad$ (let $v = v$, $u = v + \lambda(u-v)$)

(2) $\quad f(u) \geq f(v + \lambda(u-v)) + (1-\lambda)\, df(v + \lambda(u-v))(u-v) \qquad$ (let $v = u$, $u = v + \lambda(u-v)$)

Then $(1-\lambda)(1) + \lambda(2)$ gives

$$(1-\lambda) f(v) + \lambda f(u) \geq f(v + \lambda(u-v)) = f((1-\lambda)v + \lambda u).$$

$$(1-\lambda)f(v)+\lambda f(u) \geq f(v+\lambda(u-v)) = f((1-\lambda)v+\lambda u).$$

$\Rightarrow f$ is convex.

(2) sketch: passing to the limit doesn't preserve strict equality, so have to use additional technical tricks. Otherwise similar argument.

Prop. 4.7/4.10 (convexity and 2nd derivative)

Let $f: \Omega \to \mathbb{R}$ be twice differentiable, Ω an open subset, and let $U \subseteq \Omega$ be a nonempty convex subset.

(1) f is convex on U iff $D^2 f(u)(v-u, v-u) \geq 0 \quad \forall u, v \in U$.

(2) If $D^2 f(u)(v-u, v-u) > 0 \quad \forall u, v \in U, u \neq v$, then f is strictly convex.

proof sketch: Use various forms of Taylor expansions.

Aside: $f(x) = x^4$ is strictly convex but $f''(0) = 0$.

Ex. 4.6/4.1 Let $f(u) = \frac{1}{2}u^T A u - u^T b$, where A is a symmetric matrix.

Then
$$\begin{aligned} f(v) - f(u) - df(u)(v-u) &= \tfrac{1}{2}v^T A v - v^T b - \tfrac{1}{2}u^T A u + u^T b - (v-u)^T(Au-b) \\ &= \tfrac{1}{2}v^T A v - \tfrac{1}{2}u^T A u - (v-u)^T A u \\ &= \tfrac{1}{2}v^T A v + \tfrac{1}{2}u^T A u - v^T A u \\ &= \tfrac{1}{2}(v-u)^T A (v-u). \end{aligned}$$

So if A is pos. semidefinite, then f is convex.

Can show the converse by the 2nd derivative test.

$$f(u) = \tfrac{1}{2}u^T A u - u^T b$$

$$Df(u) = u^T A - b^T \qquad D_v f(u) = u^T A v - u^T v$$

$$D^2 f(u) = A \qquad D^2 f(u)(v_1, v_2) = v_1^T A v_2$$

So $D^2 f(u)(v, v) = v^T A v \geq 0$.

Def. 4.9/4.6 f has a minimum at $u \in E$ if $f(u) \leq f(v)$ $\forall v \in E$.
strict minimum at $u \in E$ if $f(u) < f(v)$ $\forall v \in E - \{u\}$.

f has a minimum at $u \in U \subseteq E$ w.r.t. U if $f(u) \leq f(v)$ $\forall v \in U$.
strict minimum at $u \in U \subseteq E$ w.r.t. U if $f(u) < f(v)$ $\forall v \in U - \{u\}$.

We sometimes stress that these are global minimums (vs. local)

Thm 4.5/4.11 Given any normed vector space E, let $U \subseteq E$ be nonempty and convex.

(1) For any convex function $J: U \to \mathbb{R}$, $\forall u \in U$, if J has a local minimum at u in U, then J has a (global) minimum at u in U.

(2) Any strict convex function $J: U \to \mathbb{R}$ has at most one minimum (in U). If it does have a minimum, then it is a strict minimum (in U).

(3) Let $J: \Omega \to \mathbb{R}$ be any function defined on an open subset $\Omega \subseteq E$ with $U \subseteq \Omega$ convex and assume J is convex on U. $\forall u \in U$, if $dJ(u)$ exists, then J has a minimum in u w.r.t. U iff

$$dJ(u)(v-u) \geq 0 \quad \forall v \in U.$$

(4) If the convex subset U in (3) is open, then the above condition is equivalent to $dJ(u) = 0$.

proof sketch: Repeated application of definitions and convexity.

Punchline is that for convex functions, local and global minima correspond.

Ex. 4.7/— Consider least squares solution to $Ax = b$.

i.e. Find $\underset{v}{\arg\min} \|Av - b\|_2^2$.

Consider $J(v) = \frac{1}{2}\|Av - b\|_2^2 - \frac{1}{2}\|b\|_2^2$, a quadratic function.

$$= \frac{1}{2}(Av-b)^T(Av-b) - \frac{1}{2}b^Tb$$

$$= \frac{1}{2}(v^TA^T - b^T)(Av-b) - \frac{1}{2}b^Tb$$

$$= \frac{1}{2}v^TA^TAv - v^TA^Tb$$

$\Rightarrow \quad dJ(u) = A^T A u - A^T b.$

Since $A^T A$ is pos. semidefinite, J is convex, then the last theorem implies that the (global) minima of J are solutions to $A^T A u - A^T b = 0$

Recall SVD $A = VDU^T$ and $A^+ = UD^+V^T$. Let $u = A^+ b$.

Then $UDV^TVDU^TUD^+V^Tb = UDV^Tb = A^Tb.$

Conclusion: We want to find zeros of $dJ : \Omega \to E'$, as that tells us a lot about minima. Next time, generalizations of Newton's method.