14 local extrema

Wednesday, October 21, 2020 2:36 AM

We're now going to switch over to the optimization half of the course, and get started with some basics.

Switching to Vol 2 - Optimization

Def. 4.1 If $J: E \to \mathbb{R}$ is a real-valued function defined on a normed vector space E, we say that J has a local minimum (or relative min) at the point $u \in E$ if there is some open subset $W \subseteq E$ containing u s.t. $J(u) \leq J(w)$ $\forall w \in W$.

Similarly, a local maximum at u if $J(u) \geq J(w)$ $\forall w \in W$.

Local extremum if either local minimum or maximum.

Strict if $J(u) < J(w)$ for minimum or $J(u) > J(w)$ for maximum.

Prop. 4.1 Let $\Omega \subseteq E$ be an open subset. Let $J: \Omega \to \mathbb{R}$ be a function. If J has a local extremum at $u \in \Omega$, and if J is differentiable at u, then $\underbrace{dJ_u}_{\text{notation for total derivative at } u} = J'(u) = 0$.

proof. Pick arbitrary $v \in E$. Since Ω is open, for t small enough, $u + tv \in \Omega$, where $t \in I \subseteq \mathbb{R}$ an open interval.

Let $\varphi(t) = J(u + tv)$, a well defined function $\varphi: I \to \mathbb{R}$.

Then $\varphi'(0) = dJ_u(v)$ (derivative of J at u going in the v direction)

WLOG, assume u is a local minimum. Then

$$\varphi'(0) = \lim_{t \to 0_-} \frac{\varphi(t) - \varphi(0)}{t} \leq 0$$

t approaching 0 from the negative dir.

$$\text{and} \quad \varphi'(0) = \lim_{t \to 0_+} \frac{\varphi(t) - \varphi(0)}{t} \geq 0.$$

t approaching 0 from the positive direction

$\Rightarrow \varphi'(0) = dJ_u(v) = 0$. But $v \in E$ is arbitrary, so $dJ_u = 0$. $\boxtimes$

Def. 4.2 A pt $u \in \Omega$ s.t. $J'(u) = 0$ is called a critical pt of J.

Def. 4.2 A pt $u \in \Omega$ s.t. $J'(u) = 0$ is called a critical pt of J.

Aside: If $E = \mathbb{R}^n$, $dJ_u = 0$ is equivalent to
$$\begin{cases} \frac{\partial J}{\partial x_1}(u_1, \dots, u_n) = 0 \\ \vdots \\ \frac{\partial u}{\partial x_n}(u_1, \dots, u_n) = 0 \end{cases}$$

Def. 4.3 If $J : \Omega \to \mathbb{R}$, where Ω is an open subset of a normed vector space E, and if $U \subseteq \Omega$ is a subset, we say J has a local minimum (resp. local maximum) at the pt $u \in U$ with respect to U if $\exists$ an open subset $W \subseteq \Omega$ containing u s.t.

$$J(u) \leq J(w)$$
resp $\quad J(u) \geq J(w) \quad$ for all $w \in U \cap W$.

Also called local extremum at u w.r.t. U.

These are known as constrained local extrema. The constraints defining $U \subseteq \Omega$ are generally defined by a set of either equality or inequality constraints.

equality contraints

(1) $U = \{x \in \Omega \mid \varphi_i(x) = 0, \ 1 \leq i \leq m\}$ where $\varphi_i : \Omega \to \mathbb{R}$ are continuous (and usually differentiable)

(2) $U = \{x \in \Omega \mid \varphi_i(x) \leq 0, \ 1 \leq i \leq m\}$ where $\varphi_i : \Omega \to \mathbb{R}$ are continuous (and usually differentiable). inequality constraints

Clearly, can write $\varphi_i(x) = 0$ as combination of $\varphi_i(x) \leq 0$, $-\varphi_i(x) \leq 0$, so the inequality constraints are more general. But they are harder to deal with, so we start our analysis with equality constraints

Thm 4.1/4.3 (Necessary condition for constrained extremum in terms of Lagrange multipliers)

Let Ω be an open subset of $\mathbb{R}^n$.

Consider m C^1-functions $\varphi_i : \Omega \to \mathbb{R}$ (C^1 means has continuous 1st derivative)

Let $U = \{v \in \Omega \mid \varphi_i(v) = 0, \ 1 \leq i \leq m\}$. linear maps from $\mathbb{R}^n \to \mathbb{R}$.

Let $u \in U$ be a pt s.t. the derivatives $d\varphi_i(u) \in \mathcal{L}(\mathbb{R}^n; \mathbb{R})$ are linearly independent.

↳ Equivalently the $m \times n$ matrix $\left(\frac{\partial \varphi_i}{\partial x_j}\right)(u)$ has rank m.

If $J : \Omega \to \mathbb{R}$ is a function which is differentiable at $u \in U$ and if J has a local constrained extremum at u, then $\exists$ m numbers $\lambda_i(u) \in \mathbb{R}$, uniquely defined, s.t.

$$dJ(u) + \lambda_1(u)\, d\varphi_1(u) + \dots + \lambda_m(u)\, d\varphi_m(u) = 0.$$

Aside: gradient here is a vector, whereas the total differential is a map $\mathbb{R}^n \to \mathbb{R}$

Equivalently

$$\nabla J(u) + \lambda_1(u) \nabla \varphi_1(u) + \dots + \lambda_m(u) \nabla \varphi_m(u) = 0.$$

← gradient

proof. Corollary of later theorem, so proof skipped for now.
Later theorem not restricted to $\mathbb{R}^n$.

Def. 4.4 The numbers $\lambda_i(u)$ are called the *Lagrange multipliers* associated with the constrained extremum u.

Def. 4.5 The *Lagrangian* associated with our constrained extremum problem is the function $L : \Omega \times \mathbb{R}^m \to \mathbb{R}$ given by

$$L(v, \lambda) = J(v) + \lambda_1 \varphi_1(v) + \dots + \lambda_m \varphi_m(v),$$

with $\lambda = (\lambda_1, \dots, \lambda_m)$.

Prop 4.2 There exists some $\mu = (\mu_1, \dots, \mu_m)$ and some $u \in U$ s.t.

$$dJ(u) + \mu_1\, d\varphi_1(u) + \dots + \mu_m\, d\varphi_m(u) = 0 \quad \text{iff}$$

$$dL(u, \mu) = 0, \qquad \text{(or equivalently } \nabla L(u,\mu) = 0)$$

that is, iff (u, λ) is a critical pt of the Lagrangian L.

proof. $dL(u,\mu) = 0$

$$\Longleftrightarrow \quad \frac{\partial L}{\partial v}(u,\mu) = 0, \quad \frac{\partial L}{\partial \lambda}(u,\mu) = 0$$

Proof: $\partial L(u,\mu) = 0$

$$\iff \quad \begin{aligned} &\frac{\partial L}{\partial v}(u,\mu) = 0 \\ &\frac{\partial L}{\partial \lambda_1}(u,\mu) = 0 \\ &\quad \vdots \\ &\frac{\partial L}{\partial \lambda_m}(u,\mu) = 0 \end{aligned}$$

by definition.

So $$\frac{\partial L}{\partial v}(u,\mu) = dJ(u) + \mu_1\, d\varphi_1(u) + \cdots + \mu_m\, d\varphi_m(u) = 0$$

$$\frac{\partial L}{\partial \lambda_i}(u,\mu) = \varphi_i(u) = 0 \iff u \in U.$$

And of course the proof here goes both ways $\boxtimes$

Notational aside: $$dJ(u) + \lambda_1\, d\varphi_1(u) + \cdots + \lambda_m\, d\varphi_m(u) = 0$$

$$\Updownarrow$$

$$\frac{\partial J}{\partial x_1}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_1}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_1}(u) = 0$$

$$\vdots$$

$$\frac{\partial J}{\partial x_n}(u) + \lambda_1 \frac{\partial \varphi_1}{\partial x_n}(u) + \cdots + \lambda_m \frac{\partial \varphi_m}{\partial x_n}(u) = 0$$

$$\Updownarrow$$

$$\begin{bmatrix} \frac{\partial J}{\partial x_1}(u) \\ \vdots \\ \frac{\partial J}{\partial x_n}(u) \end{bmatrix} + \begin{bmatrix} \frac{\partial \varphi_1}{\partial x_1}(u) & \cdots & \frac{\partial \varphi_m}{\partial x_1}(u) \\ \vdots & \ddots & \vdots \\ \frac{\partial \varphi_1}{\partial x_n}(u) & \cdots & \frac{\partial \varphi_m}{\partial x_n}(u) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} = \vec{0}$$

$$\Updownarrow$$

$$\underbrace{\nabla J(u)}_{\text{(grad)}} + \underbrace{(\mathrm{Jac}(\varphi))^T}_{\text{(Jacobian)}} \vec{\lambda} = \vec{0}$$

Also, $L(u,\lambda) = J(u) + (\varphi_1(u), \ldots, \varphi_m(u))\vec{\lambda}$ in vector form.

Also, $L(u, \lambda) = J(u) + (\varphi_1(u), \ldots, \varphi_m(u))^T \lambda$ in vector form.

The Lagrangian encodes the constraints into the function, so an unconstrained critical pt of L is needed to have a constrained local extremum of J. Note that this is not a sufficient condition.

Ex. 4.1 Let $J : \mathbb{R}^3 \to \mathbb{R}$, $J(x, y, z) = x + y + z^2$

and $g : \mathbb{R}^3 \to \mathbb{R}$, $g(x, y, z) = x^2 + y^2$.

Let $U = \{(x, y, z) \mid g(x, y, z) = 0\} = \{(0, 0, z) \mid z \in \mathbb{R}\}$.

Clearly, $J(0, 0, z) = z^2$, so $\min_{u \in U} J(u) = 0$. $\quad u \in \mathbb{R}^3$

But suppose we use Lagrange multipliers "blindly" and assume an extrema at some $u = \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix}$.

$$dJ(u) + \lambda \, dg(u) = 0$$

$$\Rightarrow \frac{\partial J}{\partial x}(0, 0, z) = -\lambda \frac{\partial g}{\partial x}(0, 0, z) \qquad \text{And} \qquad \frac{\partial g}{\partial x}(x, y, z) = 2x$$

$$\frac{\partial J}{\partial y}(0, 0, z) = -\lambda \frac{\partial g}{\partial y}(0, 0, z) \qquad \frac{\partial g}{\partial y}(x, y, z) = 2y$$

$$\frac{\partial J}{\partial z}(0, 0, z) = -\lambda \frac{\partial g}{\partial z}(0, 0, z) \qquad \frac{\partial g}{\partial z}(x, y, z) = 0$$

$$\Rightarrow \frac{\partial J}{\partial x}(0, 0, z) = -\lambda \cdot 2x = 0 \qquad \text{However} \qquad \frac{\partial J}{\partial x}(x, y, z) = 1$$

$$\frac{\partial J}{\partial y}(0, 0, z) = -\lambda \cdot 2y = 0 \qquad \frac{\partial J}{\partial y}(x, y, z) = 1$$

$$\frac{\partial J}{\partial z}(0, 0, z) = -\lambda \cdot 0 = 0. \qquad \frac{\partial J}{\partial y}(x, y, z) = 2z \quad \text{by direct computation.}$$

Thus, we get a contradiction. What went wrong?

Note $\left[\frac{\partial g}{\partial x}(0, 0, z) \quad \frac{\partial g}{\partial y}(0, 0, z) \quad \frac{\partial g}{\partial z}(0, 0, z) \right] = [0 \quad 0 \quad 0]$

Thus, we ...

Note $\left[\frac{\partial g}{\partial x}(0,0,z) \quad \frac{\partial g}{\partial y}(0,0,z) \quad \frac{\partial g}{\partial z}(0,0,z)\right] = \left[0 \quad 0 \quad 0\right]$

Ex. 4.2 Let $E_1 = \mathbb{R}$, $E_2 = \mathbb{R}$, $\Omega = \mathbb{R}^2$, and

$$J(x_1, x_2) = -x_2$$

$$\varphi(x_1, x_2) = x_1^2 + x_2^2 - 1$$

Then $U = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}$ is the unit circle.

$$\nabla \varphi(x_1, x_2) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}, \quad \text{so} \quad \nabla\varphi(x_1, x_2) \neq 0 \quad \forall \; x_1^2 + x_2^2 = 1.$$

Then the Lagrangian $L(x_1, x_2, \lambda) = -x_2 + \lambda(x_1^2 + x_2^2 - 1)$.

If there exists a constrained local extremum, then

$$\nabla L(x_1, x_2, \lambda) = 0 \iff \frac{\partial L}{\partial x_1} = 2\lambda x_1 = 0$$

$$\frac{\partial L}{\partial x_2} = -1 + 2\lambda x_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0$$

$\Rightarrow \lambda \neq 0$

$\Rightarrow x_1 = 0$

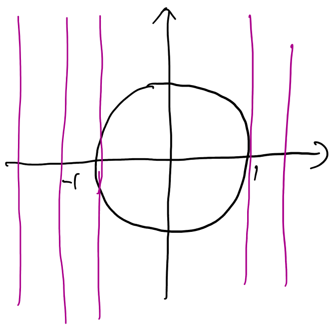
$\Rightarrow x_2 = \pm 1$

$\Rightarrow \lambda = \frac{1}{2}, \quad (x_1, x_2) = (0, 1)$

or $\lambda = -\frac{1}{2}, \quad (x_1, x_2) = (0, -1)$

Clearly, $(0,1)$ is a minimum and $(0,-1)$ is a maximum to $J(x_1, x_2) = -x_2$.

Geometrically,



level curves of J

In order to be a constrained extrema, the constraint must be tangent to a level curve of J. If not, then we could move infinitesimally along the tangent, and change level curves, increasing or decreasing J. For this easier to think of the equivalent $L = J - \lambda \varphi$

$$\nabla L = 0 \Rightarrow \nabla J = \lambda \nabla \varphi.$$

Ex. 4.3 Let J be a quadratic function (symmetric)

$$J(v) = \frac{1}{2} v^T A v - v^T b, \quad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n$$

with constraints $Cv = d$, $C \in \mathbb{R}^{m \times n}$, $m < n$, $d \in \mathbb{R}^n$, $\mathrm{rank}(C) = m$.

Then $\varphi(v) = (Cv - d)^T$ (viewing $\varphi(v)$ as a row vector, v as a col vector)

$$\Rightarrow d\varphi(v)(w) = C^T w,$$

$$\mathrm{rank}\,(J(\varphi(u))) = m.$$

The Lagrangian
$$L(v, \lambda) = \frac{1}{2} v^T A v - v^T b + (Cv - d)^T \lambda$$
$$= \frac{1}{2} v^T A v - v^T b + \lambda^T (Cv - d)$$

Then the gradient
$$\nabla L(v, \lambda) = \begin{pmatrix} Av - b + C^T \lambda \\ Cv - d \end{pmatrix}$$

Recall $\dfrac{\partial (x^T y)}{\partial x} = y$

$$\frac{df(g,h)}{dx} = \frac{d(g(x)^T)}{dx} \cdot \frac{\partial f(g,h)}{\partial g} + \frac{d(h(x)^T)}{dx} \cdot \frac{\partial f(g,h)}{\partial h}$$ Chain rule

$\dfrac{d(x^T A x)}{dx}$ let $g(x) = x^T$, $h(x) = Ax$, $f(g,h) = gh$.

$$= \frac{dx}{dx} \cdot h + \frac{d(x^T A^T)}{dx} \cdot g^T$$
$$= Ax + A^T x$$
$$= (A + A^T) x$$

Thus, a necessary condition for a constrained local extremum is

$$Av + C^T \lambda = b$$
$$Cv = d$$

$$\Longleftrightarrow \begin{pmatrix} A & C^T \end{pmatrix} \begin{pmatrix} v \end{pmatrix} = \begin{pmatrix} b \end{pmatrix}$$

$$\iff \begin{pmatrix} A & C^T \\ C & 0 \end{pmatrix} \begin{pmatrix} v \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}$$

Note that this is a symmetric matrix. Later, we will show that J has a minimum iff A is pos. def. (not true for general symmetric A).

Thm 4.2/4.2 (Necessary condition for a constrained extremum).
Let $\Omega \subseteq E_1 \times E_2$ be an open subset of a product of normed vector spaces, with E_1 a Banach space (E_1 is complete), let $\varphi : \Omega \to E_2$ be a C^1 function (so $d\varphi(w)$ exists and is continuous for all $w \in \Omega$), and let

$$U = \{(u_1, u_2) \in \Omega \mid \varphi(u_1, u_2) = 0\}.$$

Let $u = (u_1, u_2) \in U$ be a point s.t.

$$\frac{\partial \varphi}{\partial x_2}(u_1, u_2) \in \mathcal{L}(E_2; E_2) \quad \text{and} \quad \left(\frac{\partial \varphi}{\partial x_2}(u_1, u_2)\right)^{-1} \in \mathcal{L}(E_2; E_2),$$

and let $J : \Omega \to \mathbb{R}$ be a function which is differentiable at u.
If J has a constrained local extremum at u, then there is a continuous linear form $\Lambda(u) \in \mathcal{L}(E_2; \mathbb{R})$ s.t.

$$dJ(u) + \Lambda(u) \circ d\varphi(u) = 0.$$

proof. We will use the Implicit Function Thm (Thm 3.3/3.14), which shows that $\exists$ open subsets $U_1 \subseteq E_1$, $U_2 \subseteq E_2$ and a continuous function $g : U_1 \to U_2$ with $(u_1, u_2) \in U_1 \times U_2 \subseteq \Omega$ and such that

$$\varphi(v_1, g(v_1)) = 0. \quad \forall v_1 \in U_1.$$

Moreover, g is differentiable at $u_1 \in U_1$, and

$$dg(u_1) = -\left(\frac{\partial \varphi}{\partial x_2}(u)\right)^{-1} \circ \frac{\partial \varphi}{\partial x_1}(u)$$

Then the restriction of J to $(U_1 \times U_2) \cap U$ yields a function G

Then the restriction of J to $(U_1 \times U_2) \cap U$ yields a function G of a single variable, with $G(v_1) = J(v_1, g(v_1))$ $\forall\, v_1 \in U_1$.

G is differentiable at u_1 and has a local extremum at u_1 on U_1, so

$$dG(u_1) = 0.$$

By the chain rule,

$$dG(u_1) = \frac{\partial J}{\partial x_1}(u) + \frac{\partial J}{\partial x_2}(u) \circ dg(u_1)$$

$$= \frac{\partial J}{\partial x_1}(u) - \frac{\partial J}{\partial x_2}(u) \circ \left(\frac{\partial \varphi}{\partial x_2}(u)\right)^{-1} \circ \frac{\partial \varphi}{\partial x_1}(u) = 0$$

$$\Longrightarrow \quad \frac{\partial J}{\partial x_1}(u) = \frac{\partial J}{\partial x_2}(u) \circ \left(\frac{\partial \varphi}{\partial x_2}(u)\right)^{-1} \circ \frac{\partial \varphi}{\partial x_1}(u).$$

Let $\Lambda(u) = -\dfrac{\partial J}{\partial x_2}(u) \circ \left(\dfrac{\partial \varphi}{\partial x_2}(u)\right)^{-1}$.

Then
$$dJ(u) = \frac{\partial J(u)}{\partial x_1} + \frac{\partial J(u)}{\partial x_2}$$

$$= \frac{\partial J(u)}{\partial x_2} \circ \left(\frac{\partial \varphi}{\partial x_2}(u)\right)^{-1} \circ \left(\frac{\partial \varphi}{\partial x_1}(u) + \frac{\partial \varphi}{\partial x_2}(u)\right)$$

$$= -\Lambda(u) \circ d\varphi(u).$$

$$\Rightarrow \quad dJ(u) + \Lambda(u) \circ d\varphi(u) = 0.$$

Corollary: (proof of Thm 4.1/4.2)

Linear independence of m linear forms $d\varphi_i(u)$

Linear independence of m linear forms $d\varphi_i(u)$

$\iff$ $m \times n$ matrix $A = \left(\left(\partial \varphi_i / \partial \varphi_{x_j} \right)(u) \right)$ has rank m.

WLOG, can assume 1st m cols of A are lin. ind.

Let $\varphi : \Omega \to \mathbb{R}^m$ be $\varphi(v) = (\varphi_1(v) \dots \varphi_m(v))$.

Then $\dfrac{\partial \varphi}{\partial x_2}(u)$ is invertible and $\dfrac{\partial \varphi}{\partial x_2}(u)$, $\left(\dfrac{\partial \varphi}{\partial x_2}(u) \right)^{-1}$ are continuous,

so the last Thm applies

$\Rightarrow$ $\exists$ continuous linear form $\Lambda(u) \in \mathcal{L}(\mathbb{R}^m ; \mathbb{R})$ s.t.

$$dJ(u) + \Lambda(u) \circ d\varphi(u) = 0$$

But $\Lambda(u)$ is defined by a tuple $(\lambda_1(u) \dots \lambda_m(u)) \in \mathbb{R}^m$, so

$$\Rightarrow dJ(u) + \lambda_1(u) d\varphi_1(u) + \dots + \lambda_m(u) d\varphi_m(u) = 0$$

Uniqueness of $\lambda_i(u)$'s comes from linear ind. of $d\varphi_i(u)$'s.

Next time we will see how to use second derivatives and convexity to find extrema.