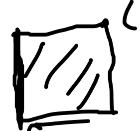
29 Lecture 29

abs max



[29.1] **Application of FONC** Optimize $f(x, y) = x^2y$ in $T = \mathsf{conv}(0, e_1, e_2, e_1 + e_2)$.

Answer. a) Find critical points.

$$\begin{bmatrix} D_x f \\ D_y f \end{bmatrix} \to \frac{\partial f}{\partial x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2xy \\ x^2 \end{bmatrix} = \Big[\; x = 0$$

$(0,t)$ $(t,0)$

From $D_x f = 2xy = 0$, $D_y f = 2x^2 = 0$, we get $x = 0$. No critical points in T°. abs min

b) So, we should check f on the boundary. (Not just on the segment 0 to e_2, but on all of ∂T.)

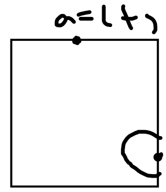
c) Conclude that $(0, t)$ and $(t, 0)$, $0 \le t \le 1$, are absolute minimums.

Notice that $f \ge 0$ on T. At points $(0, t)$ and $(t, 0)$, $0 \le t \le 1$, the f-value is 0. So, these are absolute minimums.

d) Conclude that point $(1, 1)$ is an absolute maximum.

See, the absolute maximum has to occur. It does not occur inside. It has to occur at the boundary. But among the boundary points, $(1, 1)$ has the largest f-value.

e) Can there be some other local optimums? For this, we can use FONC. We show that below. Draw picture.



f) Let $a = (1, t)$, $0 < t < 1$. Compute $f'(a)$, $D(a)$.

Can we find a d such that $D_d f(a) < 0$? What do we conclude?

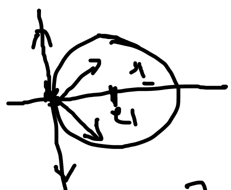
Can we find a d such that $D_d f(a) > 0$? What do we conclude?

$$\nabla f|_a = \begin{bmatrix} 2xy \\ x^2 \end{bmatrix} = \begin{bmatrix} 2t \\ 1 \end{bmatrix}, \quad D(a) = \left\{ \begin{bmatrix} \le 0 \\ * \end{bmatrix} \right\} = \left\{ d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} \;\middle|\; d_1 \le 0, \; d_2 \in \mathbb{R} \right\}$$

$\nabla f|_a^t d < 0$ for any $d \in D(a)$ never happen $\left\langle \begin{bmatrix} 2t \\ 1 \end{bmatrix}, \begin{bmatrix} \le 0 \\ * \end{bmatrix} \right\rangle$

We have $\nabla f(a) = \begin{bmatrix} 2t \\ 1 \end{bmatrix}$ and $D(a) = \{\begin{bmatrix} \le 0 \\ * \end{bmatrix}\}$. As $\nabla f(a)^t \begin{bmatrix} 0 \\ 1 \end{bmatrix} > 0$, a is not a local maximum by FONC. As $\nabla f(a)^t \begin{bmatrix} 0 \\ -1 \end{bmatrix} < 0$, it is not a local minimum by FONC.

g) Let $a = (t, 1)$, $0 < t < 1$. Proceed similarly.



$D(a) = \{ x \mid x_1 > 0, \; x_2 \in \mathbb{R} \} \cup \{(0,0)\}$ $\quad a = (0,0)$

$D(a) = \left\{ \begin{bmatrix} > 0 \\ * \end{bmatrix} \right\}$

$\bar{D} = \{ x \mid x_1 \geq 0, x_2 \geq 0 \}$

$\begin{bmatrix} 0 \\ d_2 \end{bmatrix}$

We have $\nabla f(a) = \begin{bmatrix} 2t \\ t^2 \end{bmatrix}$ and $D(a) = \{ \begin{bmatrix} * \\ \leq 0 \end{bmatrix} \}$.

As $\nabla f(a)^t e_1 > 0$ and $\nabla f(a)^t(-e_1) < 0$, it does not satisfy FONC for being a local optimum.

h) Write the final conclusion.

The points $(0, t)$ and $(t, 0)$, $0 \leq t \leq 1$, are the absolute minimums, $(1, 1)$ is the absolute maximum. We do not have any other local optimums. ┘

✓[29.2] **Practice** Consider $f(x, y) = x^2 y$ in $T = \mathsf{conv}(0, e_1, e_2, e_1 + e_2)$. Apply FONC at points $a = (0, t), b = (t, 0)$, $0 < t < 1$ and at $(1, 1)$.

[29.3] **NoPen** In general, can we conclude that a is a point of local minimum using FONC at a?

Second order conditions

[29.4] **Second order necessary condition** (SONC) Let $T \subseteq E \subseteq \mathbb{R}^n$, $f \in \mathcal{C}^2(E)$ and $a \in T$ be a local minimum for f on T.

$\nabla f(a)$ must

$a \in T^\circ \Rightarrow D(a) = \mathbb{R}^n$

✓ i) Suppose that $d \in D(a)$ with $\nabla f(a)^t d = 0$. Then we have $d^t H(a) d \geq 0$.

ii) In particular, if $a \in T^\circ$ then $H(a)$ must be a psd matrix.

Proof.

$[a, a + \delta d] \subseteq T$.

a $a + \delta d$

Taylor – I

$0 < \gamma \leq \delta$

$$f(a + \gamma d) = f(a) + \underbrace{\nabla f(a)^t \gamma d}_{\substack{= 0 \\ \text{given}}} + \frac{(\gamma d)^t H(a)(\gamma d)}{2!} + R(\gamma d)$$

$$\lim_{\gamma \to 0} \frac{R(\gamma d)}{\|\gamma d\|^2} = 0$$

$$\frac{(\gamma d)^t H(a)(\gamma d)}{\gamma^2 \, 2!} + \frac{R(\gamma d)}{\gamma^2} \geq 0$$

$$\lim_{\gamma \to 0} \frac{d^t H(a) d}{2} + \frac{R(\gamma d)}{\gamma^2} \geq 0$$

$$\frac{d^t H(a) d}{2} \geq 0$$

i) Let $d \in D(a)$ such that $\bigtriangledown f(a)^t d = 0$. If $d = 0$, then $d^t H(a) d = 0$. We don't have to show anything in this case.

So assume that $d \neq 0$. As $d \in D(a)$, by definition, $\exists \delta > 0$ such that $[a, a + \delta d] \subseteq T$. By Taylor-I, $\forall \theta \in (0, \delta)$, we have

$$0 \leq f(a + \theta d) - f(a) = \frac{1}{2}\theta^2 d^t H(a) d + r(\theta d),$$

where $\frac{r(\theta d)}{\|\theta d\|^2} \to 0$ as $\theta \to 0$. Dividing it by θ^2 and letting $\theta \to 0$, we get $d^t H(a) d \geq 0$.

ii) If $a \in T^\circ$, we already know that $D(a) = \mathbb{R}^n$. In that case, we also know by FONC that $\bigtriangledown f(a) = 0$. Thus, by i), we get that $d^t H(a) d \geq 0$, for each $d \in \mathbb{R}^n$. Since $H(a)$ is a symmetric matrix (as $f \in \mathcal{C}^2(T)$), we see that $H(a)$ is psd. ■

[29.5] **Necessary conditions are not sufficient enough** Note that the necessary conditions only help in "short-listing the points for local optimum". In general, we shall require some sufficient conditions to conclude whether some of the short-listed point is a local optimum.

[29.6] **Second order sufficient condition : interior case** (SOSC) Let $T \subseteq \mathbb{R}^n$, $f \in \mathcal{C}^2(T)$ and $a \in T^\circ$ be a critical point. TFAT.

a) If $H(a)$ is pd, then a is a strict local minimum.

b) If $H(x)$ is psd in some $B_\delta(a)$, then a is a local minimum.

c) If $H(x)$ is pd in a neighborhood of a except a, then a is a strict local minimum.

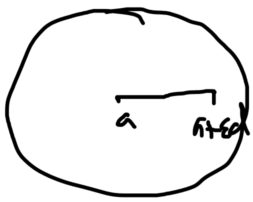
d) If there exist x, y such that $x^t H(a) x < 0 < y^t H(a) y$, then a is a saddle point.

e) If $H(a)$ has a positive and a negative eigenvalue, then a is a saddle point.

$H(a+\theta d)$

Proof.

a) $H(a)$ pd, $a \in T^\circ$, $\nabla f(a) = 0$ $\Rightarrow$ a is a st loc min



$\lim_{r \to 0} \frac{R(rd)}{r^2} = 0$

$\left|\frac{R(rd)}{r^2}\right| < \frac{\lambda_1}{4}$

Taylor-I

$f(a+rd) = f(a) + \frac{(rd)^t H(a) rd}{2!} + R(rd)$

$\lambda_1, \lambda_2, \ldots, \lambda_n$

$x_1 \quad x_2 \quad x_n$

orthonormal basis of $\mathbb{R}^n$

$d = d_1 x_1 + \cdots + d_n x_n$

$d^t d = d_1^2 + \cdots + d_n^2$

$H(a) d = H(a)(d_1 x_1 + \cdots + d_n x_n) = d_1 \lambda_1 x_1 + \cdots + d_n \lambda_n x_n$

$d^t H(a) d = d_1^2 \lambda_1 + \cdots + d_n^2 \lambda_n \geq d_1^2 \lambda_1 + \cdots + d_n^2 \lambda_1 = \lambda_1 d^t d$

$\forall \gamma$, with $|\gamma|<\delta$

$$\frac{f(a+\gamma d)-f(a)}{\|\gamma d\|^2} = \frac{\gamma^2}{2!} d^t H(a) d + R(\gamma d) \geq \frac{\gamma^2}{2!} \lambda d^t d + R(\gamma d)$$

$$\frac{\lambda}{2} + \frac{R}{4} > 0$$

a) Since we know only about $H(a)$, we have to use Taylor-I (because Taylor-II, uses the Hessian matrix at nearby points). But in Taylor-I, apart from the Hessian term, we also have the remainder term $r(x)$ and that needs to be managed carefully.

Let $H(a)$ be pd. So $H(a)$ (is Hermitian and) has positive eigenvalues. Let $\lambda > 0$ be the smallest eigenvalue. By [27.11], we see that

$$d^t H d \geq \lambda \|d\|^2 \qquad \forall d \in \mathbb{R}^n.$$

As $a \in T^\circ$, we have $B_\delta(a) \subseteq T$ for some $\delta > 0$. For each d with $a + d \in B_\delta(a)$, by Taylor-I, we have

$$f(a+d) \;-\; f(a) \quad = \quad \frac{1}{2} d^t H(a) d \;+\; r(d),$$

where $\lim\limits_{\|d\|\to 0} \frac{r(d)}{\|d\|^2} \to 0$. Thus, $\exists \epsilon > 0$, $\epsilon < \delta$, such that for each $\|d\| \leq \epsilon$ we have $|\frac{r(d)}{\|d\|^2}| < \frac{\lambda}{4}$. Hence

$$\frac{f(a+d)-f(a)}{\|d\|^2} \geq \frac{\lambda}{2} + \frac{r(d)}{\|d\|^2} > \frac{\lambda}{4} > 0,$$

for all d with $\|d\| \leq \epsilon$. Hence a is a strict local minimum.

b) This one is a direct application of Taylor-II.

$H(x)$ is psd in $B_\delta(a)$, then a is a loc min.

$$f(a+x) - f(a) = \frac{x^t H(a+\theta x) x}{2!} \geq 0. \quad \text{for some } 0<\theta<1$$

As $a \in T^\circ$, we have $B_\epsilon(a) \subseteq T$ for some $\epsilon > 0$ and $\epsilon < \delta$. For each d with $a + d \in B_\epsilon(a)$, by Taylor-II,

$$f(a+d) \;-\; f(a) \quad = \quad \frac{1}{2} d^t H(a+td) d,$$

for some $t \in (0,1)$. But the rhs is nonnegative, as $H(x)$ is psd for each $x \in B_\epsilon(a) \subseteq B_\delta(a)$. Hence a is a local minimum.

c) This one is a direct application of Taylor-II.

d) We apply Taylor-I here.

Consider tx where $t > 0$ is so small that $[a, a + tx] \subseteq T$. (This is possible, as $a \in T^\circ$.) By Taylor-I, we have for all $p \le t$,

$$f(a + px) \; - \; f(a) \quad = \quad \frac{1}{2}(px)^t H(a)(px) \; + \; r(px),$$

where $\lim\limits_{p \to 0} \frac{r(px)}{\|px\|^2} \to 0$. Thus

$$\frac{f(a + px) - f(a)}{\|px\|^2} = \frac{1}{2}\frac{x^t H(a) x}{\|x\|^2} + \frac{r(px)}{\|px\|^2} \le \frac{1}{4}\frac{x^t H(a) x}{\|x\|^2} < 0,$$

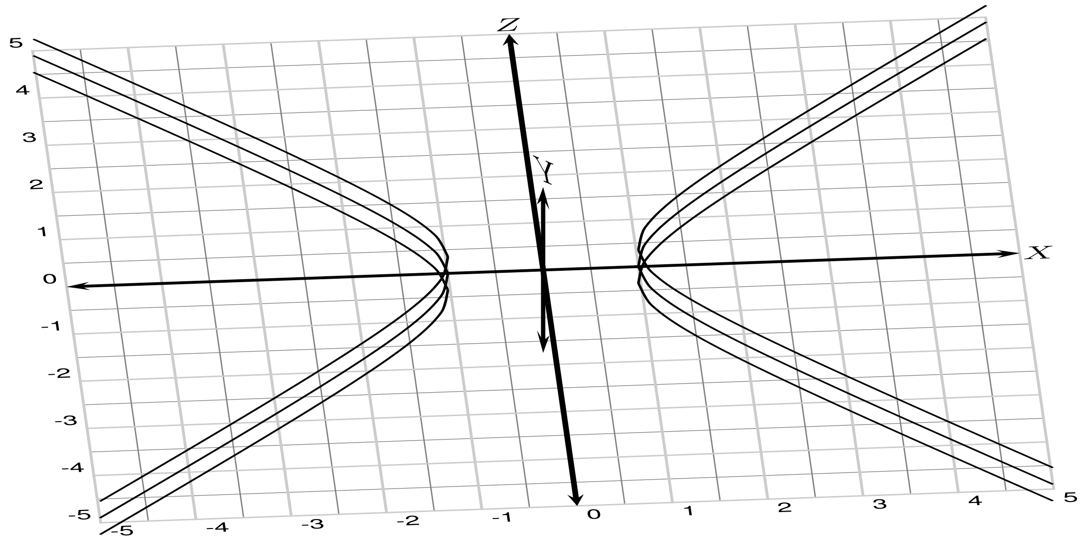
for all small p.

Thus, a cannot be a point of local minimum. Similarly, we can show that a cannot be a local maximum.

e) Apply d). ■

[29.7] **Application of SOSC** Find a point (x, y, z) on the surface $x^2 - z^2 = 1$ which is nearest to the origin.

Answer.



a) Note that, from the figure, it is visible that the points are $\pm e_1$. But let us argue it, assuming that we do not have the picture.

b) The problem is
$$\begin{array}{ll} \min & \underline{x^2 + y^2 + z^2} \\ \text{s.t.} & x^2 - z^2 = 1 \end{array} \quad \equiv \quad \begin{array}{ll} \min & \underline{y^2 + 2z^2 + 1} \\ \text{s.t.} & y, z \in \mathbb{R} \end{array}.$$

c) The rhs is an unconstrained problem in two variables. We solve that.

d) Critical points: We have only one critical point $\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

e) Time to use SOSC: The Hessian $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ is pd. So by SOSC, it is a strict local minimum.

f) As the function $y^2 + 2z^2 + 1$ is always ≥ 1, we see that it is a strict absolute minimum.

g) For the original problem: we have $x^2 = 1 + z^2 = 1$. Hence $x = \pm 1$, $y = 0$, $z = 0$ are strict absolute minimums.

h) Can we conclude them to be global minimum in some other way also?

Of course. Take a big r and consider the part of the surface lying inside $\overline{B_r(0)}$. That becomes a closed and bounded set, that is, a compact set. As $f(x, y, z) = \|(x, y, z)\|^2$ is a continuous function, it will attain its minimum. This point of minimum, must be a global minimum, as points outside the ball have larger distance from origin. It will also be a critical point. Since $f(e_1)$ is the same as $f(-e_1)$, they both must be global minimums. (Otherwise, we should have another critical point.)

Alternate. A) Alternately, we can use $z^2 = x^2 - 1$. Then <u>we must use</u> $x^2 - 1 \geq 0$. So we have

$$\begin{array}{ll} \min & \underline{2x^2 + y^2 - 1} \\ \text{s.t.} & x^2 \geq 1,\ y \in \mathbb{R}. \end{array}$$

B) FONC: We have no critical points in the interior.

C) FONC: For a point $a = (1, t)$, where $t \neq 0$, we have $\triangledown f(a) = \begin{bmatrix} 4 \\ 2t \end{bmatrix}$, feasible directions are $\begin{bmatrix} \geq 0 \\ * \end{bmatrix}$ and $\triangledown f(a)^t \begin{bmatrix} 0 \\ -1 \end{bmatrix} < 0$. So this point is not a local minimum. Similarly, $(-1, t)$, $t \neq 0$, is not a point of local maximum.

D) FONC: The point $a = (1, 0)$ satisfies FONC for being a local minimum as

$$\triangledown f(a)^t d = \begin{bmatrix} 4 & 0 \end{bmatrix} \begin{bmatrix} \geq 0 \\ * \end{bmatrix} \geq 0, \qquad \forall d \in D(a).$$

E) SONC: The directions $d \in D(a)$ such that $\bigtriangledown f(a)^t d = 0$ are $\alpha \begin{bmatrix} 0 \\ \pm 1 \end{bmatrix}$. As

$$e_2^t H(a) e_2 = e_2^t \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} e_2 = 2 > 0,$$

we see that the point a satisfies SONC for being a local minimum.

F) Similarly, $(-1, 0)$ satisfies FONC and SONC.

G) Since $(1, 0)$ is not a point of interior, we cannot use SOSC.

H) Can we conclude the minimality in this way? Sometimes.

Since the function is large outside a ball, we see that f must have an absolute minimum. At these points FONC and SONC must be satisfied. We have only two such points $(\pm 1, 0)$. The value of the function is the same at these points. So both these points are global minimums.

Some exercises

[29.8] **Exercises(E)** Let $f \in \mathcal{C}^2(T)$ and $a \in T$ be a point at which FONC and SONC holds. Suppose that a is not a point of interior but the Hessian $H(a)$ is positive definite. Show that a may not be a point of minimum.

[29.9] **Exercise(M)** `(Why is it happening?)`

a) I have a convex cube in $\mathbb{R}^3$. Suppose that I have a linear function which takes equal values at two diametrically opposite vertices. Must that function be a constant?

b) I have a convex cube in $\mathbb{R}^3$. Suppose that I have a linear function which is minimized at two diametrically opposite vertices. Must that function be a constant?

[29.10] **NoPen**

a) Let $f \in \mathcal{C}^2(T)$ and $a \in T^o$ be a critical point. In SOSC item b), it says that if $H(x)$ is psd in a neighborhood $B_\delta(a)$, then a is local minimum. Can we go with just that $H(a)$ is psd?

b) Give an example of a closed convex feasible set and a point for which $D(a)$ is not closed.

c) Consider minimizing $c^t x$, $c \neq 0$ over a set T. Can we have an optimum at an interior point?

d) Let $A \in M_n(\mathbb{R})$ and $f : \mathbb{R}^n \to \mathbb{R}$ be defined as $f(x) = x^t A x$. What is $\triangledown f(x)$? What is $H(x)$?

e) Let $A \in M_3(\mathbb{R})$ and $b \in \mathbb{R}^3$. Must the system $Ax = b$ have at least one solution? What if A is psd? What if A is pd?

[29.11] <u>**Exercise(E)**</u> Let T be convex and $a \in T$. Is $D(a)$ necessarily a convex cone?

[29.12] <u>**Exercise(E)**</u> Let $A \in M_n(\mathbb{R})$, $b \in \mathbb{R}^n$, and consider $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(x) = x^t A x + b^t x$.

a) What is $\triangledown f(x)$? What is $H(x)$?

b) Suppose that A is psd. Is it necessary that we should have at least one critical point?

c) Suppose that A is psd. Suppose that a is a critical point. Can it be a saddle point/ local maximum/ local minimum?

d) Let A be positive definite. Is it necessary that we should have at least one critical point? Check whether they are minimums or maximums or saddle points.

[29.13] <u>**Practice**</u> Find local optimums and saddle points of $f = 2x_1x_2x_3 - 4x_1x_3 - 2x_2x_3 + x_1^2 + x_2^2 + x_3^2 - 2x_1 - 4x_2 + 4x_3$.

[29.14] <u>**Practice**</u> Find the local optimums and saddle points of $f(x, y) = 5x^3 + 4xy + x + y^2$.

[29.15] <u>**Exercise(E)**</u> Let $v_1, \ldots, v_p \in \mathbb{R}^n$ be distinct and define $f : \mathbb{R}^n \to \mathbb{R}$ as $f(x) = \sum_{i=1}^{p} \|x - v_i\|^2$. Optimize it.

[29.16] <u>**Exercise(E)**</u> Let $v_1 < \cdots < v_p$ be some real numbers and define $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = \sum_{i=1}^{p} |x - v_i|$. Optimize it.

[29.17] <u>**Exercise(E)**</u> Consider $\begin{array}{ll} \max & cx + dy \\ \hline \text{s.t.} & x + y \leq 1,\ x, y \geq 0, \end{array}$ where $c > d \geq 0$. Use FONC to show that the unique maximum solution is $(1, 0)^t$. Use graphical method to give an alternate solution.

30 Lecture 30

Constrained optimization

[30.1] **Definitions** `(Feasible directions and linearizing cones)`

a) Suppose that the set T is defined using some $\mathcal{C}^1$ functions,

$$T = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, i = 1, \ldots, m, \quad h_j(x) = 0, j = 1, \ldots, p\}.$$

Notice here 'the inequality constraints are in '$g_i(x) \geq 0$' form.

b) A constraint is said to be an ACTIVE CONSTRAINT at a feasible point $a \in T$, if it satisfies equality in the constraint.

c) Since all h_j are active at each feasible points, let us denote by $\boldsymbol{A(a)}$ the set

$$A(a) := \{i \mid g_i \text{ is active at } a\}.$$

d) The LINEARIZING CONE $\boldsymbol{\mathcal{D}(a)}$ of T at a is defined as

$$\mathcal{D}(a) := \Big\{d \in \mathbb{R}^n \mid \langle \nabla g_i(a), d\rangle \geq 0, \forall i \in A(a), \quad \langle \nabla h_j(a), d\rangle = 0, \forall j\Big\}.$$

It is a nonempty closed convex cone. Loosely saying, it gives us the directions along which we can move a little bit and still stay inside the feasible set, that is, $\overline{D}(a)$. (We will show this later.)