Optimization with inequality constraints

3 Comments to KKT and CQ

$$
\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in S = \left\{ \boldsymbol{x} \in X \subseteq \mathbb{R}^n : g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m \right\} \end{array}
$$

A point $\boldsymbol{x}$ is *feasible* iff $\boldsymbol{x} \in S$. The set of active constraints at a point $\boldsymbol{x}$ is denoted

$$
I(\boldsymbol{x}) := \{ i : g_i(\boldsymbol{x}) = 0 \}.
$$

The condition *constraint qualification* at a feasible point is

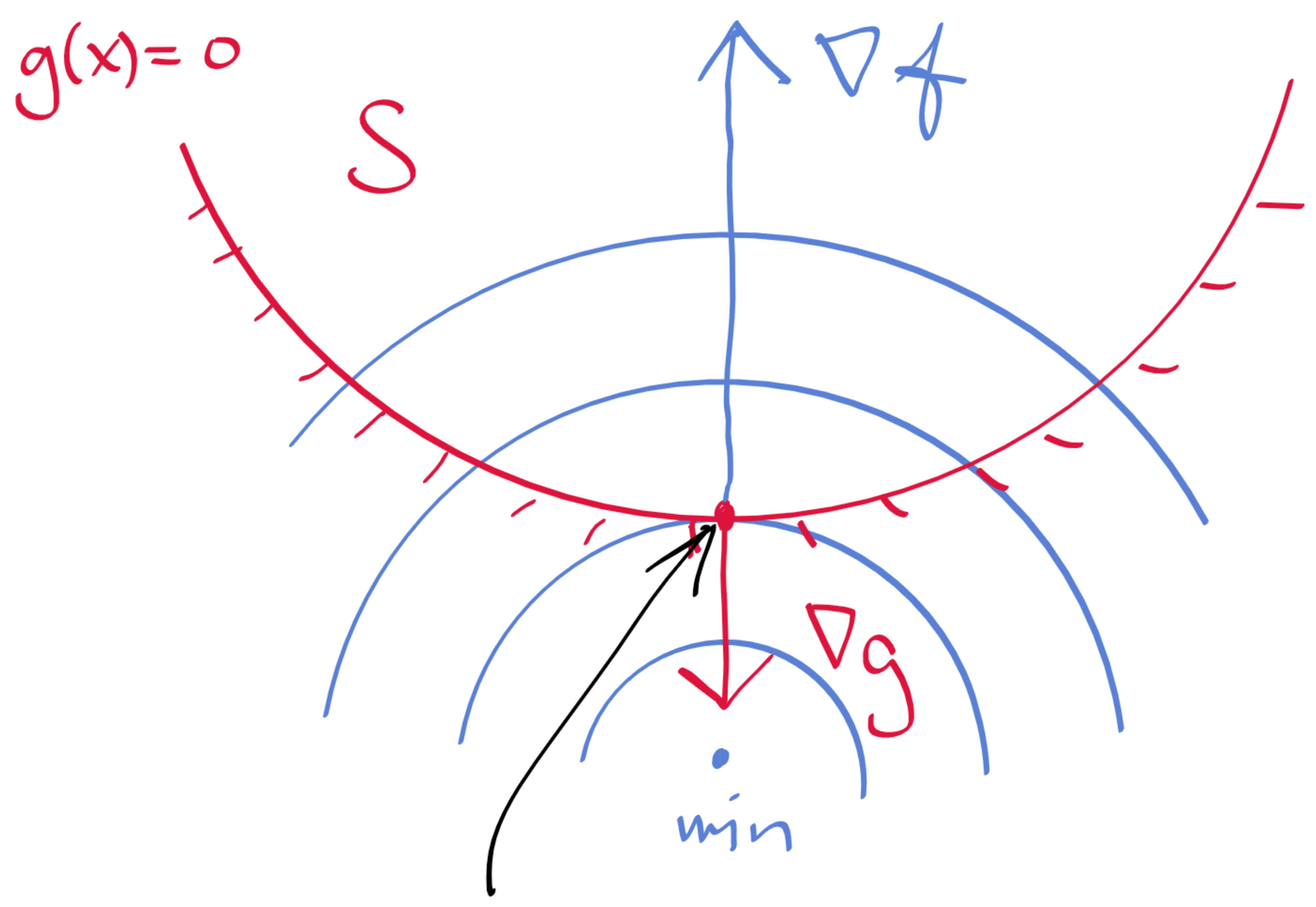
$$
\text{(CQ)} \qquad \begin{cases} \sum_{i \in I(\bar{\boldsymbol{x}})} \lambda_i \nabla g_i(\bar{\boldsymbol{x}}) = \boldsymbol{0} \\ \lambda_i \ge 0, \quad i \in I(\bar{\boldsymbol{x}}) \end{cases} \implies \lambda_i = 0, \quad i \in I(\bar{\boldsymbol{x}}).
$$

THEOREM 3. *$\bar{\boldsymbol{x}}$ satisfies (CQ) and solves (P)* $\implies$

$$
\text{(KKT)} \qquad \exists \boldsymbol{u} : \begin{cases} \nabla f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\boldsymbol{x}}) = \boldsymbol{0}, \\ u_i g_i(\bar{\boldsymbol{x}}) = 0, \quad \forall i, \\ u_i \ge 0, \quad \forall i. \end{cases}
$$

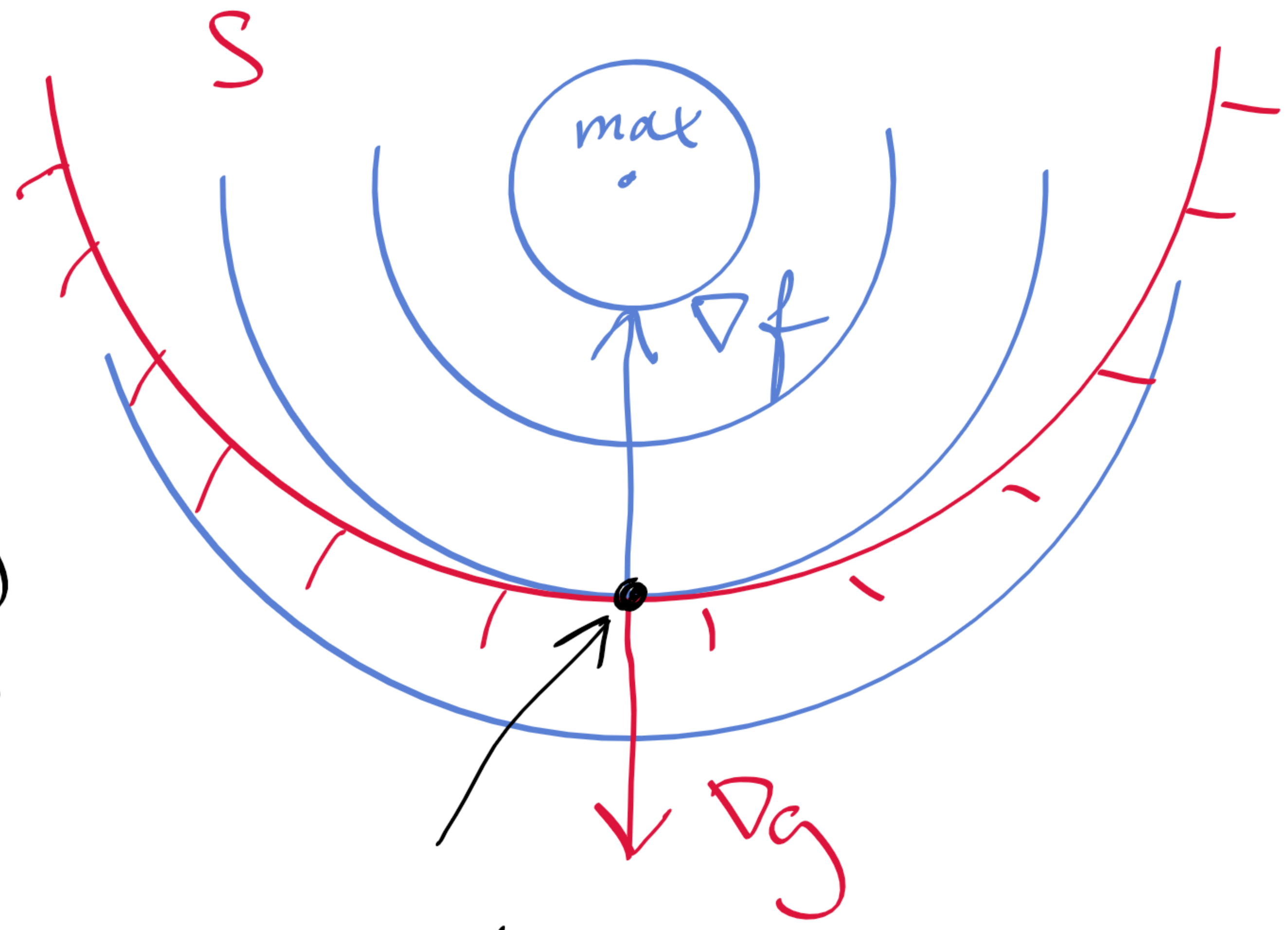
Comments

- These conditions (KKT) are only necessary for a local minimizer. Consider (P) with only one constraint $g(x) \le 0$ in $\mathbb{R}^2$:



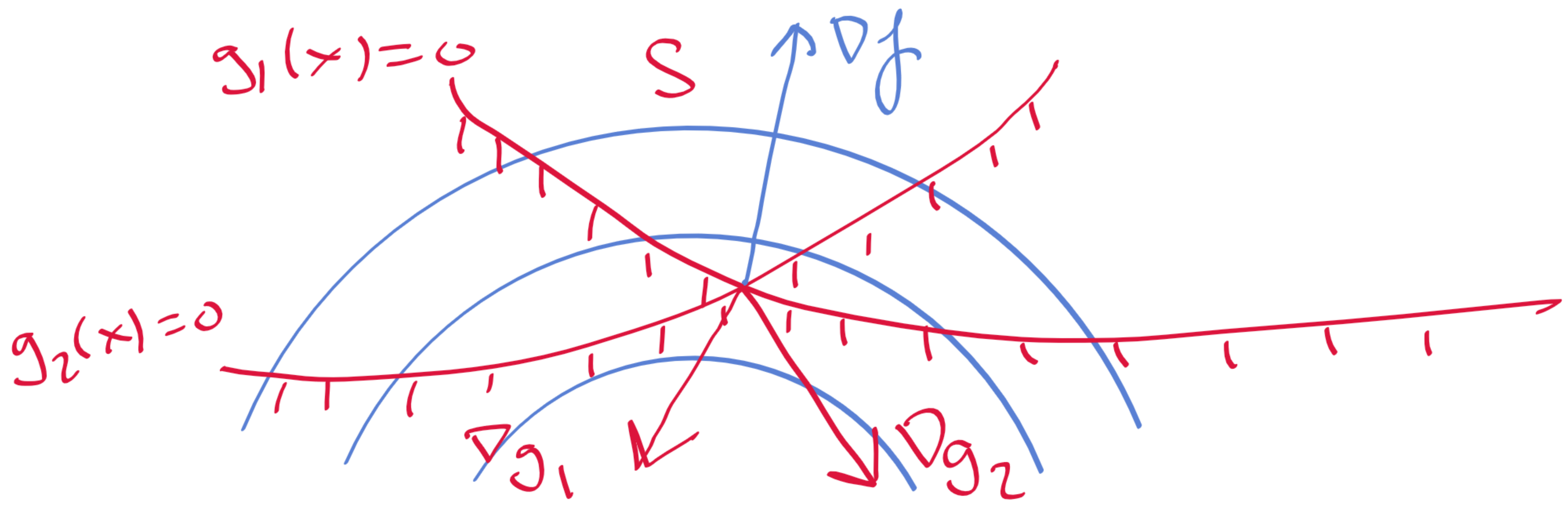
(KKT) and (CQ) satisfied

local min.



No local minimizer

An ex. with two constraints:



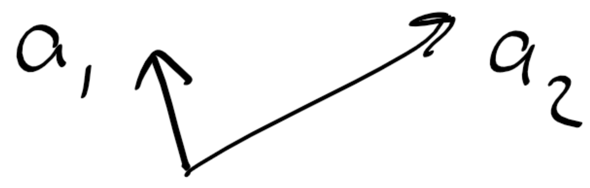
∇g_1 and ∇g_2 are PLI $\Rightarrow$ (CQ) is satisfied

(KKT) is satisfied

- LI: $\sum \lambda_i a_i = 0 \Rightarrow \text{all } \lambda_i = 0$

$\Rightarrow$ PLI: $\begin{cases} \sum \lambda_i a_i = 0 \\ \lambda_i \geq 0 \end{cases} \Rightarrow \text{all } \lambda_i = 0$

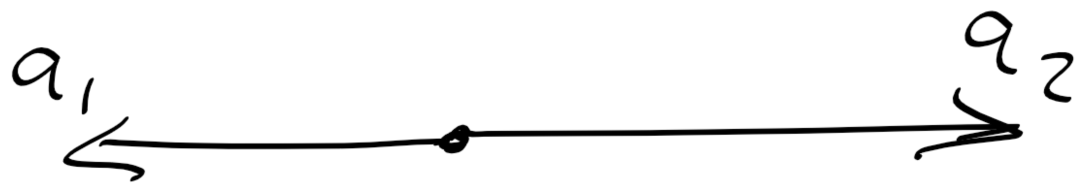
Ex. in $\mathbb{R}^2$:



Equivalently:

- PLD $\Rightarrow$ LD

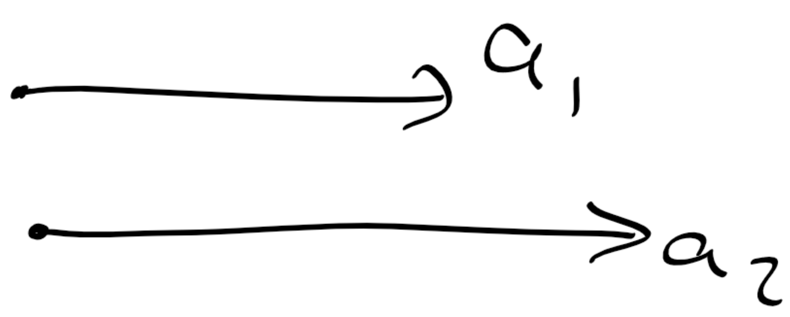
Ex. in $\mathbb{R}^2$



$2a_1 + a_2 = 0$

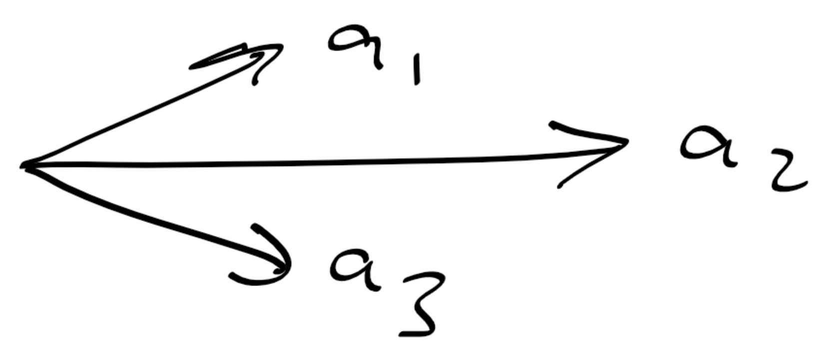
- The converse is not true

Ex. in $\mathbb{R}^2$



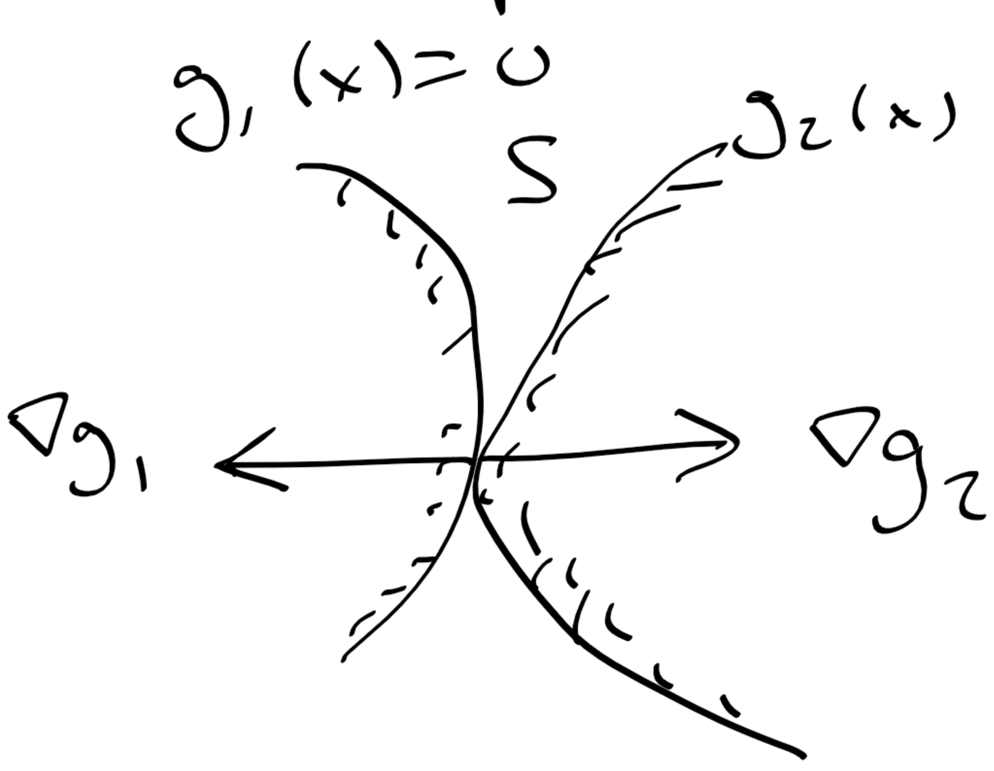
LD but PLI

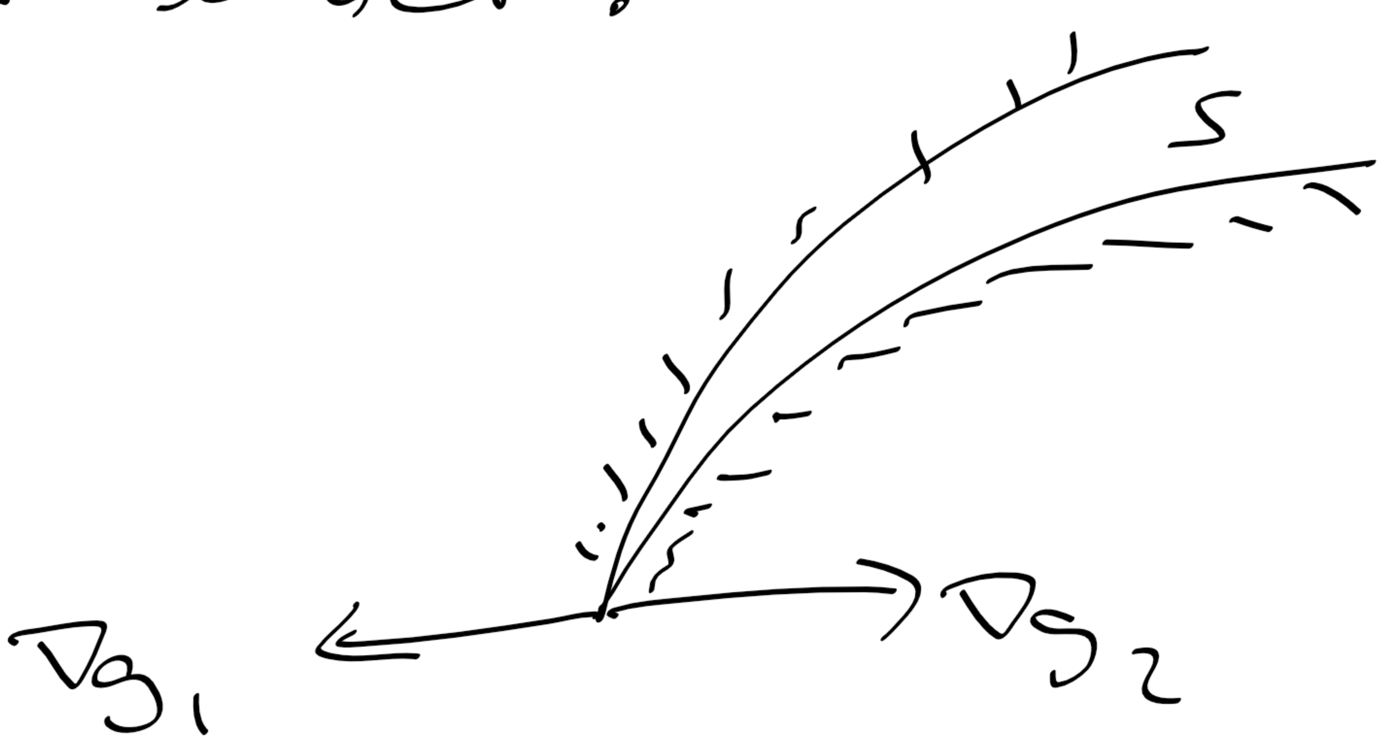
another example



LD but PLI

- Strategy to use Thm 3: candidates for local minimizers are solutions of (KKT) and points that do not satisfy (CQ). Examples of the latter:





- The complementary slackness condition

$u_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m \iff$

$(u_1 \ \dots \ u_m) \begin{pmatrix} g_1(\bar{x}) \\ \vdots \\ g_m(\bar{x}) \end{pmatrix} = 0 \iff u^T g(\bar{x}) = 0$

Proof: $\boxed{\Rightarrow}$ sum up zeroes. $\boxed{\Leftarrow}$ $\sum \underbrace{u_i g_i(\bar{x})}_{\leq 0} = 0 \Rightarrow u_1 g_1(\bar{x}) = -\sum_{i=2}^{m} u_i g_i \geq 0$

- Def. Lagrange function

$$\mathcal{L}(x;u) = f(x) + u^T g(x)$$

Then

$$(KKT) \quad \begin{cases} \nabla_x \mathcal{L}(\bar{x};u) = 0 \\ u^T g(\bar{x}) = 0 \\ u \geq 0 \\ g(\bar{x}) \leq 0 \end{cases}$$

- $\bar{x}$ local maximizer $\Rightarrow$ same except for $\underline{\underline{u \leq 0}}$

4. Example: $f(x) = x_1^2 + (x_2-1)^2$

$$g_1(x) = x_2 - x_1^2 \le 0$$

$$g_2(x) = -x_2 \le 0$$

$$\nabla f = \begin{pmatrix} 2x_1 \\ 2(x_2-1) \end{pmatrix} \qquad \nabla g_1 = \begin{pmatrix} -2x_1 \\ 1 \end{pmatrix} \qquad \nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

(KKT)
$$\begin{cases} 2x_1 - u_1 2x_1 = 0 \\ 2(x_2-1) + u_1 - u_2 = 0 \\ u_1(x_2 - x_1^2) = 0 \\ u_2 x_2 = 0 \\ u_1 \ge 0 \\ u_2 \ge 0 \end{cases}$$

I $u_1 = u_2 = 0$: $\begin{cases} 2x_1 = 0 \\ x_2 = 1 \end{cases}$ outside region (stationary point of f)

II $u_1 = 0, u_2 > 0$: $\begin{cases} x_1 = 0 \\ 2(x_2-1) - u_2 = 0 \\ x_2 = 0 \end{cases} \Longleftrightarrow \begin{array}{l} x_1 = 0 \\ x_2 = 0 \\ u_2 = -2 < 0 \end{array}$ not allowed

III $u_1 > 0, u_2 = 0$: $\begin{cases} x_1 - u_1 x_1 = 0 \\ 2(x_2-1) + u_1 = 0 \\ x_2 - x_1^2 = 0 \end{cases} \Longleftrightarrow \begin{cases} x_1(1-u_1) = 0 \\ 2x_2 + u_1 = 2 \\ x_2 = x_1^2 \end{cases}$

- $x_1 = 0 \Rightarrow x_2 = 0 \Rightarrow u_1 = 2 > 0$ OK
- $u_1 = 1 \Rightarrow 2x_2 + 1 = 2 \Rightarrow x_2 = 1/2 \Rightarrow x_1 = \pm \frac{1}{\sqrt{2}}$ OK

IV $u_1 > 0, u_2 > 0$:

$$\begin{cases} x_1(1-u_1) = 0 \\ 2x_2 + u_1 - u_2 = 2 \\ x_2 = x_1^2 \\ x_2 = 0 \end{cases} \Longleftrightarrow \begin{cases} x_2 = 0 \\ x_1 = 0 \\ u_1 = u_2 + 2 \end{cases}$$

Investigation of (CQ): $\nabla g_1 = \begin{pmatrix} -2x_1 \\ 1 \end{pmatrix}$, $\nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$

If both are active:

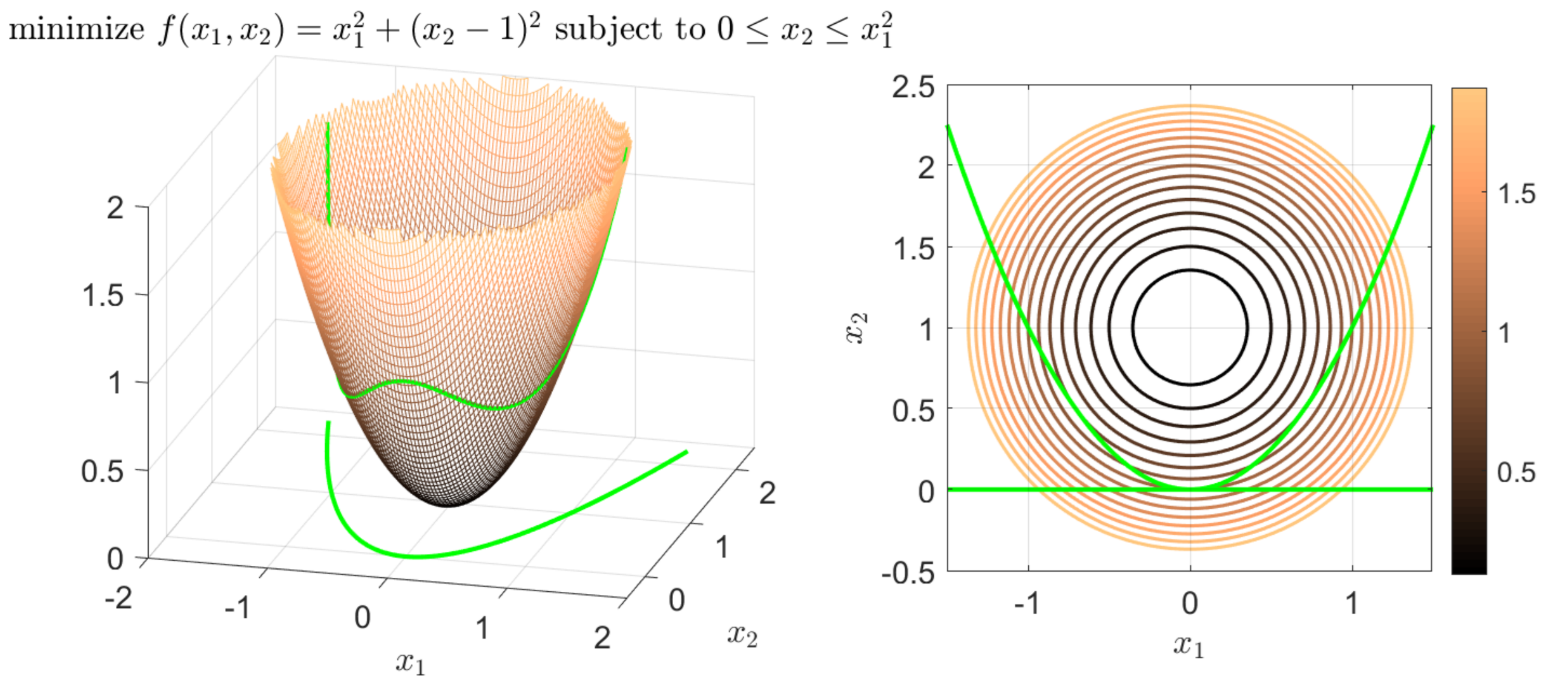
$$\lambda_1 \nabla g_1(x) + \lambda_2 \nabla g_2(x) = 0 \iff \begin{cases} -2x_1\lambda_1 = 0 \\ \lambda_1 - \lambda_2 = 0 \\ \lambda_1 \geq 0, \lambda_2 \geq 0 \end{cases}$$

(CQ) satisfied if $\lambda_1 = \lambda_2 = 0$ is the only solution. This is the case if $x_1 \neq 0$. (CQ) is not satisfied when $x_1 = 0$: constraints are $\begin{cases} x_2 \leq 0 \\ -x_2 \leq 0 \end{cases} \iff x_2 = 0$

$f(0,0) = 1$, $f\left(\pm\frac{1}{\sqrt{2}}, \frac{1}{2}\right) = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{2}\right)^2 = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

character? local min.

The points $\left(\pm\frac{1}{\sqrt{2}}, \frac{1}{2}\right)$ are global minimizers, because if we add the constraint $x_1^2 + (x_2-1)^2 \leq R^2$ with a large R, we get a compact set on which boundary $f = R^2$.



Optimization with mixed constraints

5 KKT Theorem

Let $X \subseteq \mathbb{R}^n$ and assume that $f, \boldsymbol{g}, \boldsymbol{h} \in \mathcal{C}^1(X)$ if X is open, otherwise $\mathcal{C}^1$ in a neighbourhood of X. The nonlinear optimization problem is

$$\text{(P)} \qquad \begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in X, \quad \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}, \quad \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}. \end{array}$$

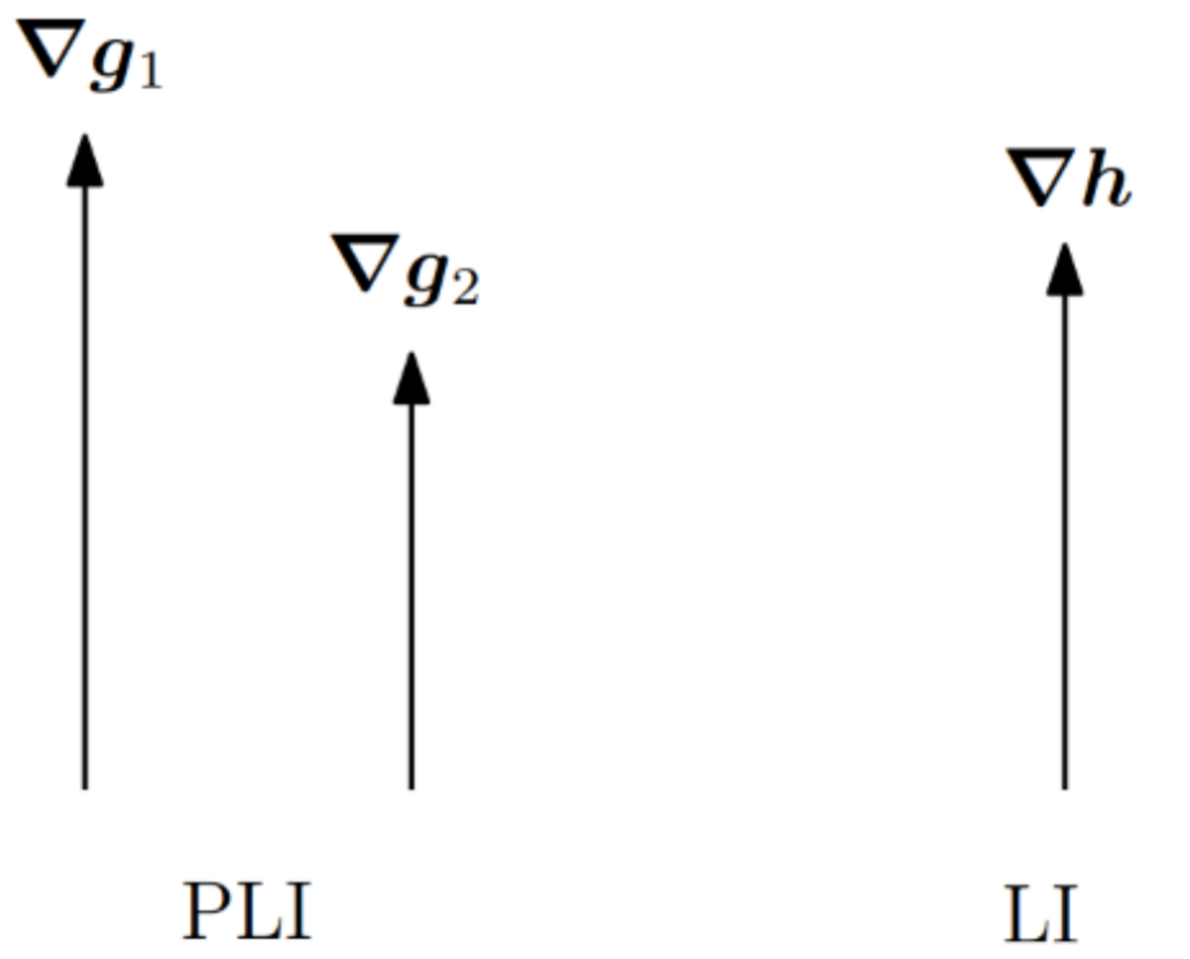
The set of active constraints at a point $\boldsymbol{x}$ (of the inequality constraints) is denoted

$$\mathcal{I}(\boldsymbol{x}) := \{i : g_i(\boldsymbol{x}) = 0\}.$$

The condition *constraint qualification* at a feasible point $\bar{\boldsymbol{x}}$ is

$$\text{(CQ)} \qquad \begin{cases} \sum_{i \in \mathcal{I}(\bar{\boldsymbol{x}})} \lambda_i \nabla g_i(\bar{\boldsymbol{x}}) + \sum_j \mu_i \nabla h_j(\bar{\boldsymbol{x}}) = \boldsymbol{0} \\ \lambda_i \ge 0, \quad i \in \mathcal{I}(\bar{\boldsymbol{x}}) \end{cases} \implies \begin{cases} \lambda_i = 0 & i \in \mathcal{I}(\bar{\boldsymbol{x}}) \\ \mu_j = 0 & \forall j. \end{cases}$$

We note that (CQ) implies that the gradients $\nabla g_i(\bar{\boldsymbol{x}})$ are positively linearly independent (set all $\mu_i = 0$) and that $\nabla h_j(\bar{\boldsymbol{x}})$ are linearly independent (set all $\lambda_i = 0$), but not the converse; for example,



CQ is not satisfied

$$\nabla h = \tfrac{1}{2}(\nabla g_1 + \nabla g_2)$$

THEOREM 3. *Assume that* $\boldsymbol{h} \equiv \boldsymbol{0}$.

$$\bar{\boldsymbol{x}} \textit{ satisfies (CQ) and solves (P)} \implies \exists \boldsymbol{u} : \begin{cases} \nabla f(\bar{\boldsymbol{x}}) + \boldsymbol{u}^{\mathsf{T}} \nabla \boldsymbol{g}(\bar{\boldsymbol{x}}) = \boldsymbol{0}, \\ \boldsymbol{u}^{\mathsf{T}} \boldsymbol{g}(\bar{\boldsymbol{x}}) = 0, \\ \boldsymbol{u} \ge \boldsymbol{0}. \end{cases}$$

PROOF: By Farkas' theorem, see previous notes.

THEOREM 4 (KKT, FIRST-ORDER NECESSARY CONDITIONS).

$$\bar{\boldsymbol{x}} \textit{ satisfies (CQ) and solves (P)} \implies \exists \boldsymbol{u}, \boldsymbol{v} : \begin{cases} \nabla f(\bar{\boldsymbol{x}}) + \boldsymbol{u}^{\mathsf{T}} \nabla \boldsymbol{g}(\bar{\boldsymbol{x}}) + \boldsymbol{v}^{\mathsf{T}} \nabla \boldsymbol{h}(\bar{\boldsymbol{x}}) = \boldsymbol{0}, \\ \boldsymbol{u}^{\mathsf{T}} \boldsymbol{g}(\bar{\boldsymbol{x}}) = 0, \\ \boldsymbol{u} \ge \boldsymbol{0}. \end{cases}$$

PROOF: We do the proof for the case of a single equality constraint $h(\boldsymbol{x}) = 0$ and comment on the general case at the end. Condition (CQ) implies $\nabla h(\bar{\boldsymbol{x}}) \neq \boldsymbol{0}$ and we assume that $\partial h(\bar{\boldsymbol{x}})/\partial x_n \neq 0$. By the implicit function theorem, there is a neighbourhood of $\bar{\boldsymbol{x}}$ in which we can solve $h(\boldsymbol{x}) = 0$ for x_n as a $\mathcal{C}^1$ function $\tilde{h}$ of the other variables:

$$h(\boldsymbol{x}) = 0 \quad \iff \quad x_n = \tilde{h}(\boldsymbol{x}'), \quad \text{where } \boldsymbol{x}' = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \end{pmatrix}.$$

We now substitute this into (P), which becomes a problem without equality constraints, and use Theorem 3 on the objective function $F(\boldsymbol{x}') := f(\boldsymbol{x}', \tilde{h}(\boldsymbol{x}'))$ and the inequality constraints $\boldsymbol{G}(\boldsymbol{x}') := \boldsymbol{g}(\boldsymbol{x}', \tilde{h}(\boldsymbol{x}')) \leq \boldsymbol{0}$. The chain rule gives

$$\nabla_{x'} F(\boldsymbol{x}') = \nabla_{x'} f(\boldsymbol{x}', \tilde{h}(\boldsymbol{x}')) + \frac{\partial f}{\partial x_n}(\boldsymbol{x}', \tilde{h}(\boldsymbol{x}'))\nabla_{x'}\tilde{h}(\boldsymbol{x}'),$$

and similarly for each component G_i and h, so that we have

$$\begin{aligned} \nabla_{x'} F(\bar{\boldsymbol{x}}') &= \nabla_{x'} f(\bar{\boldsymbol{x}}) + \frac{\partial f}{\partial x_n}(\bar{\boldsymbol{x}})\nabla_{x'}\tilde{h}(\bar{\boldsymbol{x}}'), \\ \nabla_{x'} G_i(\bar{\boldsymbol{x}}') &= \nabla_{x'} g_i(\bar{\boldsymbol{x}}) + \frac{\partial g_i}{\partial x_n}(\bar{\boldsymbol{x}})\nabla_{x'}\tilde{h}(\bar{\boldsymbol{x}}'), \\ 0 &= \nabla_{x'} h(\bar{\boldsymbol{x}}) + \frac{\partial \tilde{h}}{\partial x_n}(\bar{\boldsymbol{x}})\nabla_{x'}\tilde{h}(\bar{\boldsymbol{x}}'). \end{aligned}$$

Then we can express the gradient of h as

$$\nabla h(\bar{\boldsymbol{x}}) = \begin{pmatrix} \nabla_{x'} h(\bar{\boldsymbol{x}}') \\ \dfrac{\partial h}{\partial x_n}(\bar{\boldsymbol{x}}) \end{pmatrix} = \frac{\partial h}{\partial x_n}(\bar{\boldsymbol{x}}) \begin{pmatrix} -\nabla_{x'}\tilde{h}(\bar{\boldsymbol{x}}') \\ 1 \end{pmatrix}$$

and condition (CQ) as

$$(1) \quad \begin{cases} \displaystyle\sum_{i \in I(\bar{\boldsymbol{x}})} \lambda_i \begin{pmatrix} \nabla_{x'} g_i(\bar{\boldsymbol{x}}) \\ \dfrac{\partial g_i}{\partial x_n}(\bar{\boldsymbol{x}}) \end{pmatrix} + \mu \frac{\partial h}{\partial x_n}(\bar{\boldsymbol{x}}) \begin{pmatrix} -\nabla_{x'}\tilde{h}(\bar{\boldsymbol{x}}') \\ 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{0} \\ 0 \end{pmatrix} \\ \lambda_i \geq 0, \quad i \in I(\bar{\boldsymbol{x}}) \end{cases} \implies \begin{cases} \lambda_i = 0 \quad \forall i \\ \mu = 0. \end{cases}$$

Equality no. n of (1) gives

$$\mu \frac{\partial h}{\partial x_n}(\bar{\boldsymbol{x}}) = -\lambda_i \frac{\partial g_i}{\partial x_n}(\bar{\boldsymbol{x}}),$$

which implies that the first $n-1$ equalities in (1) can be written

$$\begin{cases} \displaystyle\sum_{i \in I(\bar{\boldsymbol{x}})} \lambda_i \nabla_{x'} G_i(\bar{\boldsymbol{x}}') = \boldsymbol{0} \\ \lambda_i \geq 0, \quad i \in I(\bar{\boldsymbol{x}}') \end{cases} \implies \lambda_i = 0 \quad \forall i.$$

This is exactly (CQ) of Theorem 3 (where $\boldsymbol{h} \equiv \boldsymbol{0}$), which states the existence of $\boldsymbol{u} \geq \boldsymbol{0}$ such that $0 = u_i G_i(\bar{\boldsymbol{x}}') = u_i g_i(\bar{\boldsymbol{x}})$ for all i and

$$\nabla_{x'} F(\bar{\boldsymbol{x}}') + \sum_{i=1}^{m} u_i \nabla_{x'} G_i(\bar{\boldsymbol{x}}') = \boldsymbol{0} \quad \iff$$

$$
\nabla_{x'} f(\bar{\boldsymbol{x}}) + \frac{\partial f}{\partial x_n}(\bar{\boldsymbol{x}}) \nabla_{x'} \tilde{h}(\bar{\boldsymbol{x}}') + \sum_{i=1}^{m} u_i \left(\nabla_{x'} g_i(\bar{\boldsymbol{x}}) + \frac{\partial g_i}{\partial x_n}(\bar{\boldsymbol{x}}) \nabla_{x'} \tilde{h}(\bar{\boldsymbol{x}}') \right) = \mathbf{0} \quad \Longleftrightarrow
$$

$$
\begin{cases} \nabla_{x'} f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} u_i \nabla_{x'} g_i(\bar{\boldsymbol{x}}) + \tilde{v} \nabla_{x'} \tilde{h}(\bar{\boldsymbol{x}}) = \mathbf{0} \\ \frac{\partial f}{\partial x_n}(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} u_i \frac{\partial g_i}{\partial x_n}(\bar{\boldsymbol{x}}) - \tilde{v} = 0 \end{cases} \quad \Longleftrightarrow
$$

$$
\nabla_x f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} u_i \nabla_x g_i(\bar{\boldsymbol{x}}) - \tilde{v} \begin{pmatrix} -\nabla_{x'} \tilde{h}(\bar{\boldsymbol{x}}') \\ 1 \end{pmatrix} = \mathbf{0} \quad \Longleftrightarrow
$$

$$
\nabla_x f(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} u_i \nabla_x g_i(\bar{\boldsymbol{x}}) + v \nabla h(\bar{\boldsymbol{x}}) = \mathbf{0}, \quad \text{where} \quad -\tilde{v} = v \frac{\partial h}{\partial x_n}(\bar{\boldsymbol{x}}),
$$

and the theorem is proved. In the general case of several equality constraints $h_j(\boldsymbol{x}) = 0$, $j = 1, \ldots, l$, the implicit function theorem is used to solve for ℓ variables as functions of the rest $n - \ell$, which is possible since the corresponding Jacobian having the gradients $\nabla h_j(\bar{\boldsymbol{x}})^{\mathsf{T}}$ as rows is non-singular by (CQ). $\square$

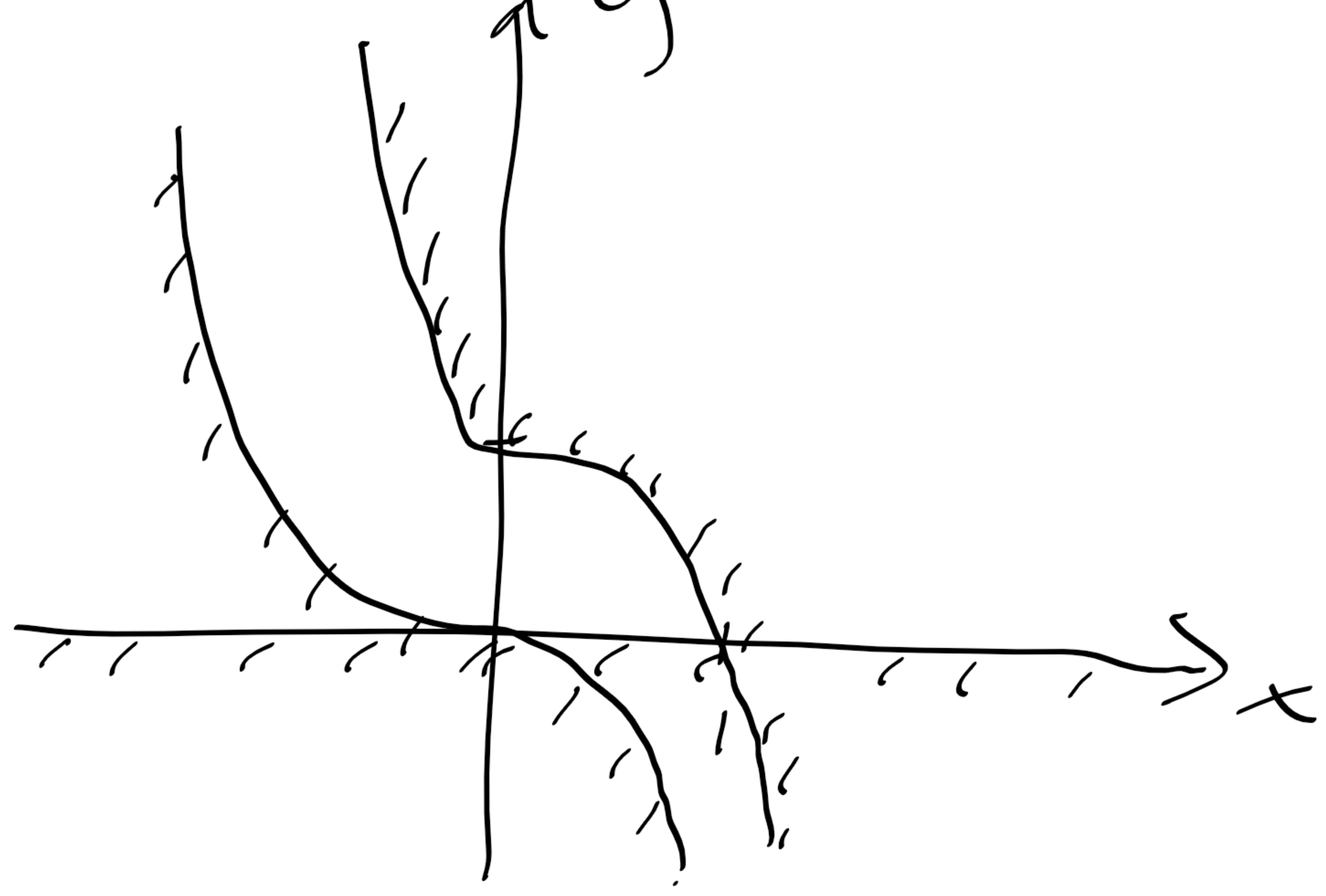
6. Examples

Ex. minimize $f(x,y) = x$

subject to $(x-1)^3 + y \le 0$

$-x^3 - y \le 0$

$-y \le 0$



The curve

$(x,y) = (t, -t^3), \quad t < 0$

lies in the region

and $f(t,-t^3) = t \to -\infty$

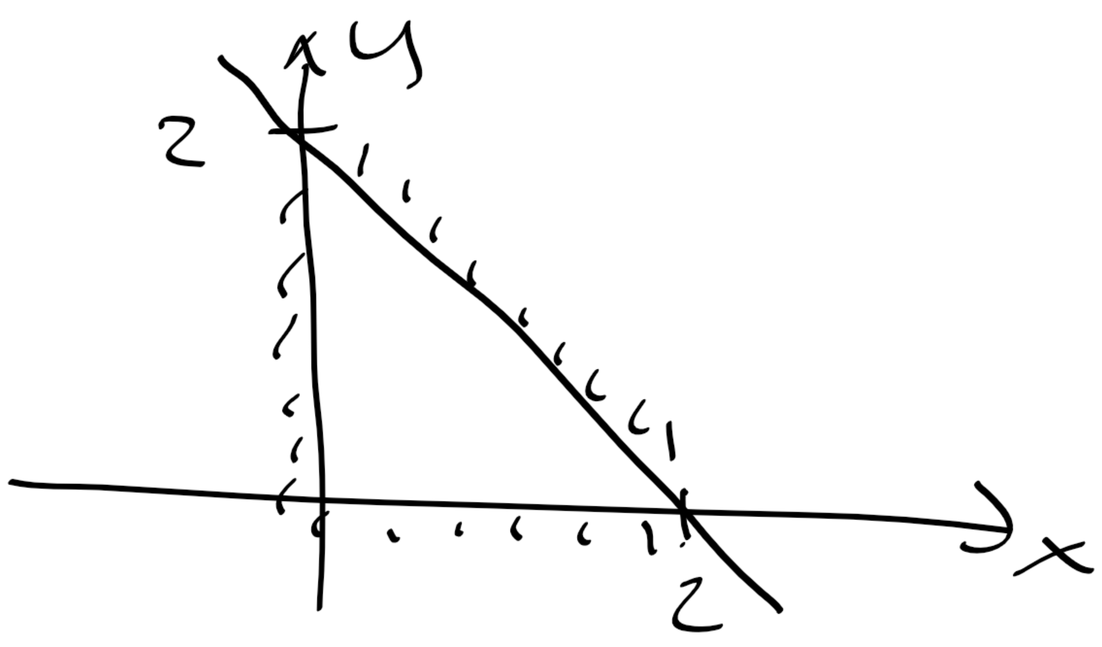
as $t \to -\infty$

so there is no solution.

Ex. minimize $f(x,y) = \frac{1}{x} + \frac{4}{y}$

subject to $(x,y) \in S = \{(x,y) : x + y \le 2, \ x > 0, y > 0\}$

Solution: S convex and



$$\nabla f = \begin{pmatrix} -\frac{1}{x^2} \\ -\frac{4}{y^2} \end{pmatrix}, \quad \nabla^2 f = \begin{pmatrix} \frac{2}{x^3} & 0 \\ 0 & \frac{8}{y^3} \end{pmatrix}$$

pos. def. in S

$\Rightarrow$ f strictly convex on S, so $\bar{x}$ local min $\Rightarrow$ $\bar{x}$ global min (thm). Set

$X = \{(x,y) : x > 0, y > 0\}$ and we consider only

$g(x,y) = x + y - 2 \le 0$.

$\nabla g = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \neq 0$ (CQ) satisfied

KKT:

$$(KKT)\quad \begin{cases} -\frac{1}{x^2} + u = 0 \\ -\frac{4}{y^2} + u = 0 \\ u(x+y-2) = 0 \\ u \geq 0 \\ x+y \leq 2 \\ x,y \in X \end{cases}$$

$u=0$: impossible

$u>0$: $\begin{cases} \frac{1}{x^2} = \frac{4}{y^2} = u > 0 \\ x+y = 2 \end{cases}$

$$\Leftrightarrow \quad \begin{cases} y = \pm 2x \\ u = \frac{1}{x^2} > 0 \\ x+y = 2 \end{cases} \Leftrightarrow \quad \begin{cases} y = 2x \Rightarrow x = \frac{2}{3} \Rightarrow y = \frac{4}{3} \\ \text{or} \\ y = -2x \Rightarrow x = -2 < 0 \text{ impossible} \end{cases}$$

Answer: $(x,y) = \left(\frac{2}{3}, \frac{4}{3}\right)$ global min.

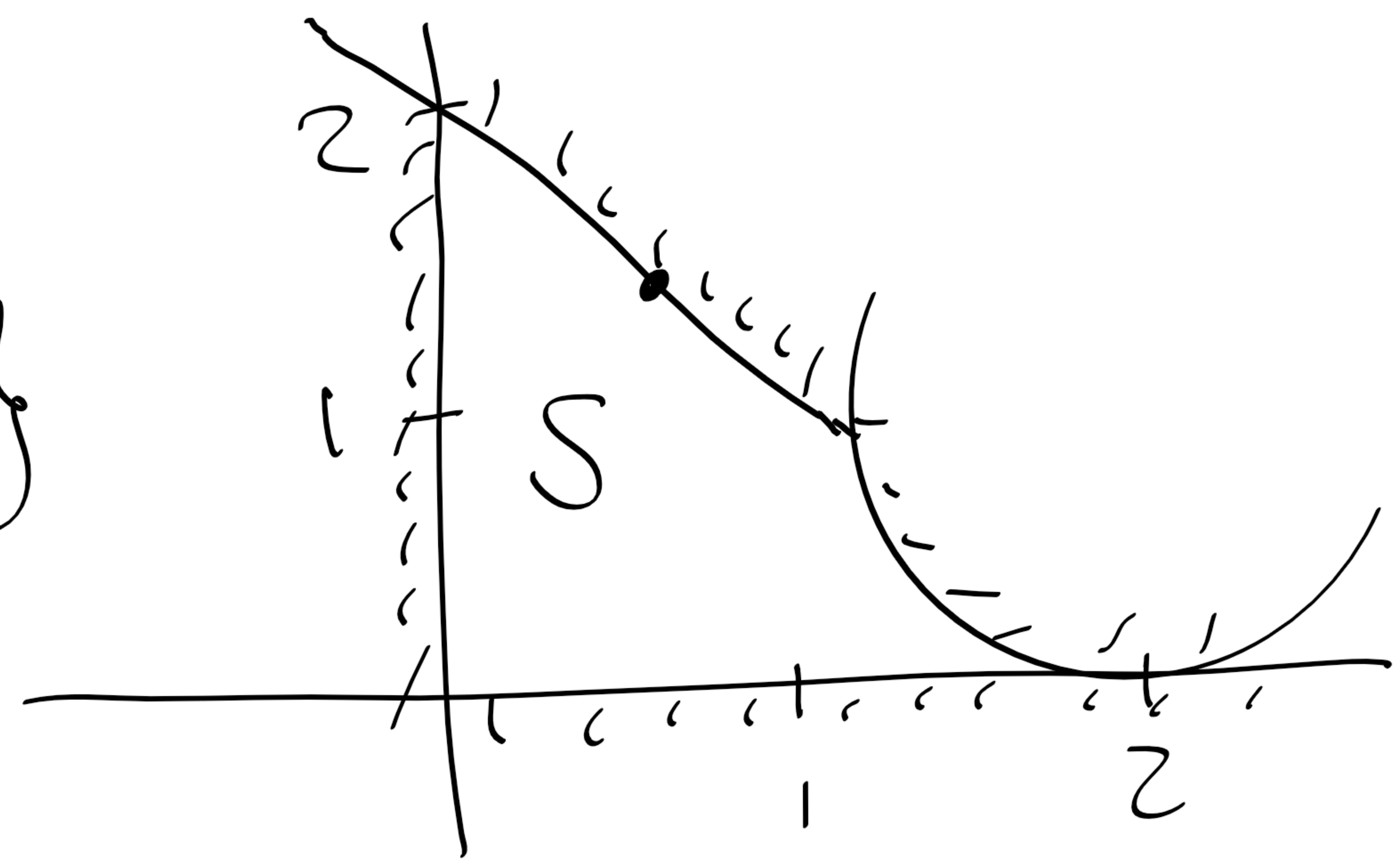
Ex. minimize $f(x,y) = \frac{1}{x} + \frac{4}{y}$, $X = \{x>0, y>0\}$

S. to $(x,y) \in S = \{ (x,y) \in X : x+y \leq 2,$
$(x-2)^2 + (y-1)^2 \geq 1 \}$

Sol.: S is not convex, but it is bounded. Let $\varepsilon > 0$ and set

$$S_\varepsilon = \{ (x,y) \in S,\ x \geq \varepsilon,\ y \geq \varepsilon \}$$

which is a compact set. Weierstraß' thm $\Rightarrow \exists$ global solution.



Of course the solution is $\left(\frac{2}{3}, \frac{4}{3}\right)$; see previous ex.