NONLINEAR PROGRAMMING ~~(Assume all fns are smooth)~~

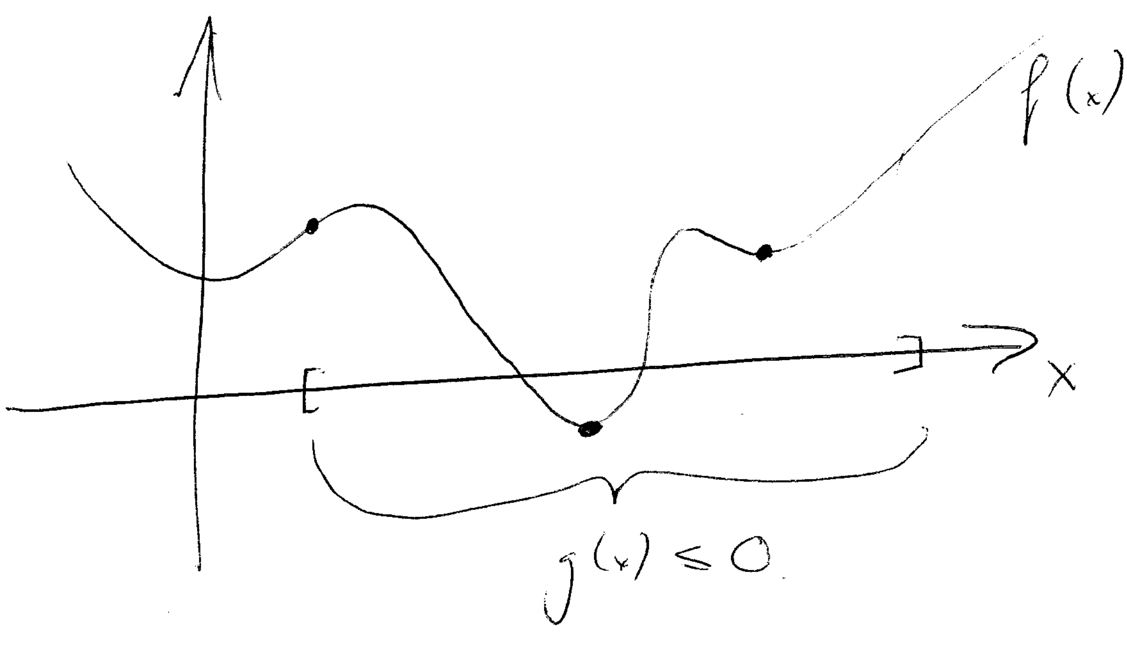
Consider two problems:

Unconstrained $\quad \min_{x \in \mathbb{R}^n} f(x) \qquad f: \mathbb{R}^n \to \mathbb{R}$

Constrained

$$\min f(x)$$
$$\text{s.t. } g_i(x) \leq 0 \quad i = 1, \dots, m$$
$$h_j(x) = 0 \quad j = 1, \dots, p$$
$$f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}.$$

Assume all fns are smooth.



Unconstrained case

So $\quad \min\limits_{x \in \mathbb{R}^n} f(x) \quad$ f smooth. $\quad$ (NLP)

Necessary condns:

Theorem If $\bar{x}$ is a local minimum for (NLP) then $\nabla f(\bar{x}) = 0$

Proof ~~As~~ By contradiction.

Assume $\nabla f(\bar{x}) \neq 0$.

Let $d = -\nabla f(\bar{x})$. (or $-H^{-1}\nabla f(\bar{x})$ for any symm p.d. H).

Define $\phi(\lambda) = f(\bar{x} + \lambda d) \qquad (\lambda \geq 0)$

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

$$< f(\bar{x}) \text{ for } \lambda \text{ small enough.}$$

//

Second order conditions:

<u>Theorem</u> If $\bar{x}$ is a local min of (NLP) then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is psd.

<u>Proof</u> Already shown $\nabla f(\bar{x}) = 0$.

Assume $\nabla^2 f(x)$ not psd.

$\therefore \exists\ d$ s.t. $d^T \nabla^2 f(\bar{x}) d < 0$

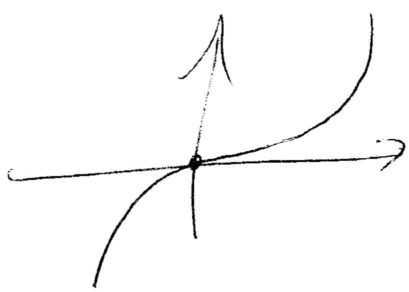
Now $$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda \underbrace{d^T \nabla f(\bar{x})}_{0} + \tfrac{1}{2}\lambda^2 \underbrace{d^T \nabla^2 f(\bar{x}) d}_{<0} + \lambda^2 \alpha(\bar{x}; \lambda d)$$

$$< f(\bar{x}) \text{ for small enough } \lambda.$$

//.

Condition not sufficient:

Eg: i) $f(x) = x^3$



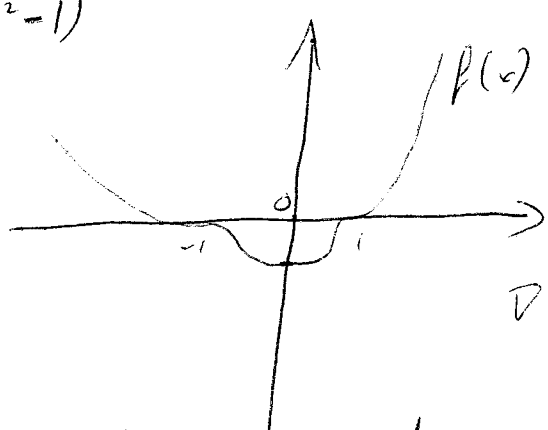
$\nabla f(x) = 3x^2$

$\nabla^2 f(x) = 6x$

At 0, $\nabla f(x) = \nabla^2 f(x) = 0$

Point of inflexion.

ii) $f(x) = (x^2 - 1)^3$



$\nabla f(x) = 6x(x^2-1)^2$

$\nabla^2 f(x) = 6(x^2-1)^2 + 24x^2(x^2-1$

$\nabla f(-1) = \nabla f(1) = \nabla f(0) = 0$

$\nabla^2 f(0) = 6,\ \nabla^2 f(1) = 0,\ \nabla^2 f(-1) =$

Cubed "confuses" things, obscures true nature of problem.

Not necessary

eg $f(x) = x_1^2(1-x_2^2)$

$(0,0)$ is local min

$\nabla f = \begin{bmatrix} 2x_1(1-x_2^2) \\ -2x_2 x_1^2 \end{bmatrix}$ $\nabla^2 f = \begin{bmatrix} 2(1-x_2^2) & -4 \\ -4x_2 & -2x \end{bmatrix}$

Sufficient conditions.

Theorem If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(x)$ is psd in a nbhd $N_\varepsilon(\bar{x})$ then $\bar{x}$ is a local minimum of (NLP).

Not psd at $(\varepsilon, \varepsilon)$

Proof Let $\omega \in N_\varepsilon(\bar{x})$

Then $\omega = \bar{x} + \lambda d$ for some d, $\|d\| = 1$, $\lambda > 0$.

So $f(\omega) = f(\bar{x}) + \lambda d^T \underbrace{\nabla f(\bar{x})}_{=0} + \underbrace{\tfrac{1}{2}\lambda^2 d^T(\nabla^2 f(\bar{x} + \theta\lambda d))d}_{\geq 0}$ for some $\theta \in [0,1]$

$\geq f(\bar{x})$

$\therefore$ $\bar{x}$ is a local minimum of (NLP). //.

Theorem If $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ is pos. def then $\bar{x}$ is a local minimum of (NLP).

Proof $\nabla^2 f(\bar{x})$ is p.d.

$\exists$: $\exists$ $a > 0$ (min eval of $\nabla^2 f(\bar{x})$) such that

$$x^T \nabla^2 f(\bar{x}) x \geq a\, x^T x \quad \forall x \in \mathbb{R}^n.$$

Take $\|d\| = 1$

$$f(\bar{x} + \lambda d) = f(\bar{x}) + \lambda d^T \nabla f(\bar{x}) + \tfrac{1}{2}\lambda^2 d^T \nabla^2 f(\bar{x}) d + \lambda^2 \alpha(\bar{x}, \lambda d)$$

$$\geq f(\bar{x}) + \lambda^2 a + \lambda^2 \alpha(\bar{x}; \lambda d)$$

So $\exists$ $\lambda(d) > 0$ s.t. $f(\bar{x} + \lambda d) \geq f(\bar{x})$ for $0 \leq \lambda \leq \lambda(d)$.

Need to show $\{\lambda(d)\}$ is bounded away from zero. Follows since set of dirns is closed & bounded so $\lambda(d)$ achieves its minimum //.

Constrained case

Consider first case where all constraints are inequalities:

$$\min f(x)$$
$$g_i(x) \leq 0 \quad i=1,\dots,m \qquad \text{(NLP)}$$

$f, g_i : \mathbb{R}^n \to \mathbb{R}$, $\quad f, g_i$ are smooth.

Defn Let $\bar{x}$ be feasible for (NLP). Then $d \in \mathbb{R}^n, d \neq 0$ is a feasible direction at $\bar{x}$ if

$$\bar{x} + \lambda d \text{ is feasible } \quad \forall \lambda \in (0, \delta) \text{ for some } \delta > 0$$

Cone of feas dirns $:= \{d : d \text{ is a feas dirn at } \bar{x}\} =: D$

Defn A vector $d \in \mathbb{R}^n$ is called a direction of descent of f at $\bar{x}$ if there exists $\delta > 0$ such that

$$f(\bar{x} + \lambda d) < f(\bar{x}) \text{ for each } \lambda \in (0, \delta).$$

Note that

$$F_0 := \{d : \nabla f(\bar{x})^T d < 0\} \subseteq \{d : d \text{ is a descent dirn at } \bar{x}\}$$
$$\subseteq \{d : \nabla f(\bar{x})^T d \leq 0\}.$$

<u>Theorem</u> Let $\bar{x}$ be a Local minimum for (NLP).
Then $F_0 \cap D = \phi$.

<u>Proof</u> By contradiction.

Let $d \in F_0 \cap D$

Then $\exists\ \delta_1 > 0$ s.t. $f(\bar{x} + \lambda d) < f(\bar{x}) \quad \forall \lambda \in (0, \delta_1)$

$\exists\ \delta_2 > 0$ s.t. $\bar{x} + \lambda d$ is feasible $\quad \forall \lambda \in (0, \delta_2)$ //.

Picture now.

Consider $\{i : g_i(\bar{x}) = 0\} =: I$ Set of ACTIVE constraints

Define $G_0 := \{ d \in \mathbb{R}^n : \nabla g_i(\bar{x})^T d < 0 \quad \forall i \in I \}$.

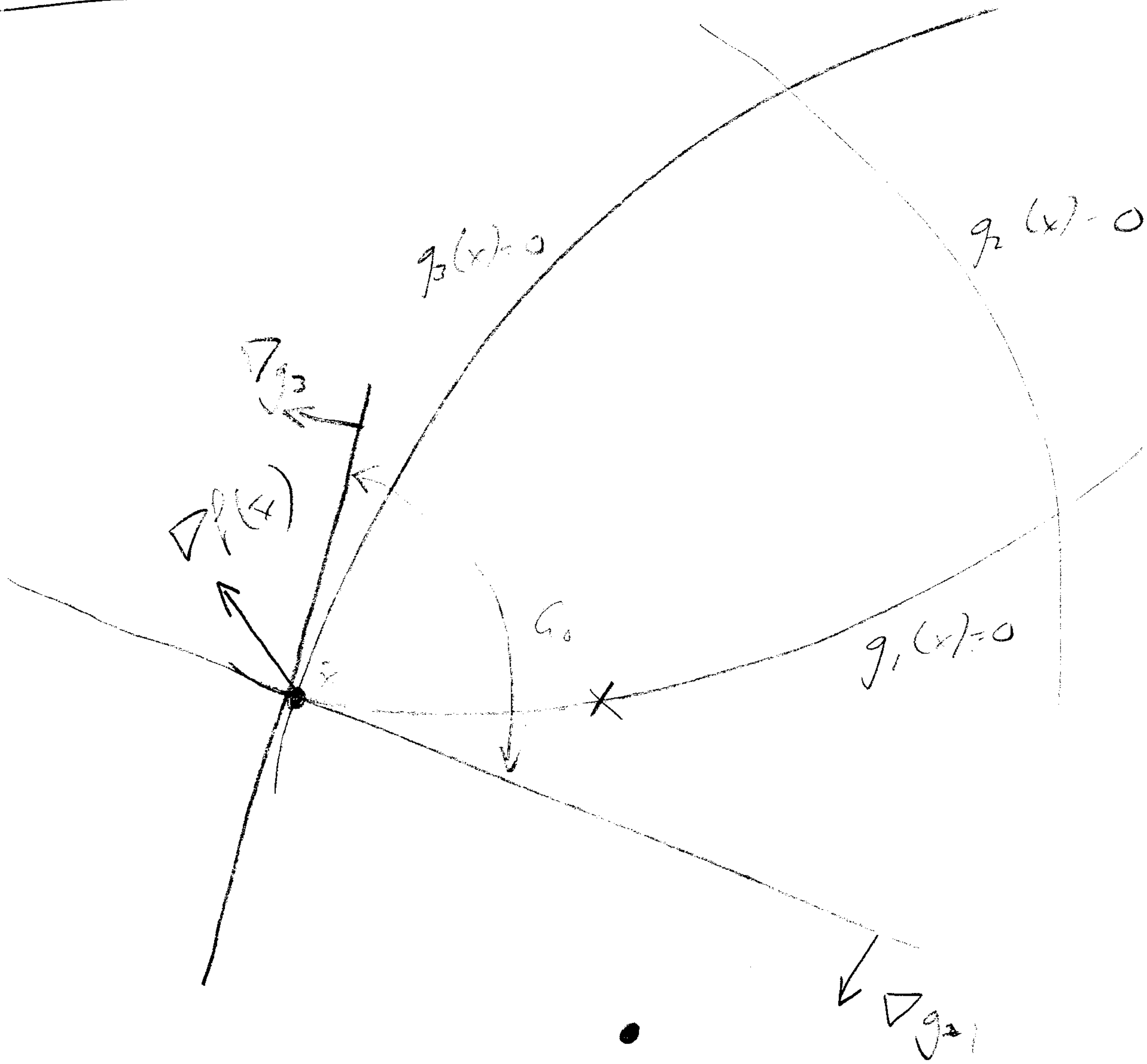
<u>Theorem</u> Let $\bar{x}$ be a Local minimum for (NLP). Then $F_0 \cap G_0 = \phi$.

<u>Proof</u> $d \in G_0 \Rightarrow d \in D$ so $G_0 \subseteq D$

By previous theorem, $F_0 \cap D = \phi$

$\therefore F_0 \cap G_0 = \phi$ //.

Example



Use this to argue that if have local minimizer then $-\nabla f(\bar{x})$ is in cone generated by ∇g_1 and ∇g_3. (Maybe switch to different vertex $\bar{x}$, or place • more carefully — f • is distance from •).

Theorem (Gordon's Theorem).

(I) $Ad < 0$

(II) $p \geq 0,\ p \neq 0 \qquad A^T p = 0$

Proof

$$\max \quad 0 \qquad \text{(P)}$$

$$Ad \leq -e$$

lin var

$$\min \quad -e^T p \qquad \text{(D)}.$$

$$A^T p = 0$$

$$p \geq 0$$

//.

<u>Theorem</u> (Fritz-John conditions)

Let $\bar{x}$ be a local minimum for (NLP). Then there exist scalars u_0 and u_i, $i \in I$, such that

$$u_0 Df(\bar{x}) + \sum_{i \in I} u_i D g_i(\bar{x}) = 0$$

$$u_0, u_i \geq 0 \quad \text{for } i \in I.$$

$$(u_0, u_I) \neq (0, 0).$$

u_I = vector whose components are u_i, $i \in I$.

Equivalent to

$$u_0 Df(\bar{x}) + \sum_{i=1}^{m} u_i Dg_i(\bar{x}) = 0$$

$$u_0, u_i \geq 0$$

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

$$(u_0, u) \neq (0, 0).$$

Proof

There exists no vector d such that

$$Df(\bar{x})^T d < 0 \quad \text{and} \quad Dg_i(\bar{x})^T d < 0 \quad \forall i \in I.$$

Let A be the matrix whose rows are $Df(\bar{x})^T$ and $Dg_i(\bar{x})^T$ for $i \in I$.

So $Ad < 0$ is inconsistent.

Then $\exists^{\text{nonzero}}\, p \geq 0$ s.t. $A^T p = 0$

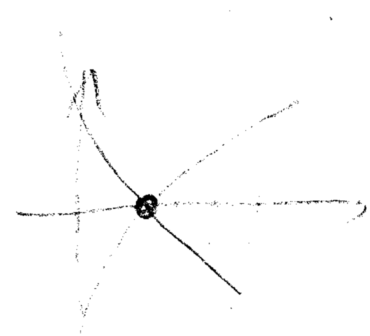
Setting u_0 and u_i equal to appropriate components of p gives result.

Equivalent form obtained by setting $u_i = 0$ for $i \notin I$ //

eg: min $x_1^2 + x_2^2$

s.t. $(x_1 - 2)^2 + (x_2 - 1)^2 \leq 2$

$(x_1 - 2)^2 + (x_2 + 1)^2 \leq 2$



$\nabla f_0 = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \qquad \nabla f_1 = \begin{bmatrix} 2(x_1 - 2) \\ 2(x_2 - 1) \end{bmatrix} \qquad \nabla f_2 = \begin{bmatrix} 2(x_1 - 2) \\ 2(x_2 + 1) \end{bmatrix}$

$x = (1, 0)$: $\quad u_0 \begin{bmatrix} 2 \\ 0 \end{bmatrix} + u_1 \begin{bmatrix} -2 \\ -2 \end{bmatrix} + u_2 \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$u_0 = 2,\ u_1 = 1,\ u_2 = 1.$

May have $u_0 = 0$. Seems counterintuitive, but may be necessary:

Example

Equivalent to

$$\min \quad -x_1$$

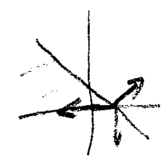
$$x_2 - (1-x_1)^3 \le 0 \qquad - g_1$$

$$-x_2 \le 0 \qquad - g_2$$

$\min \; -x$

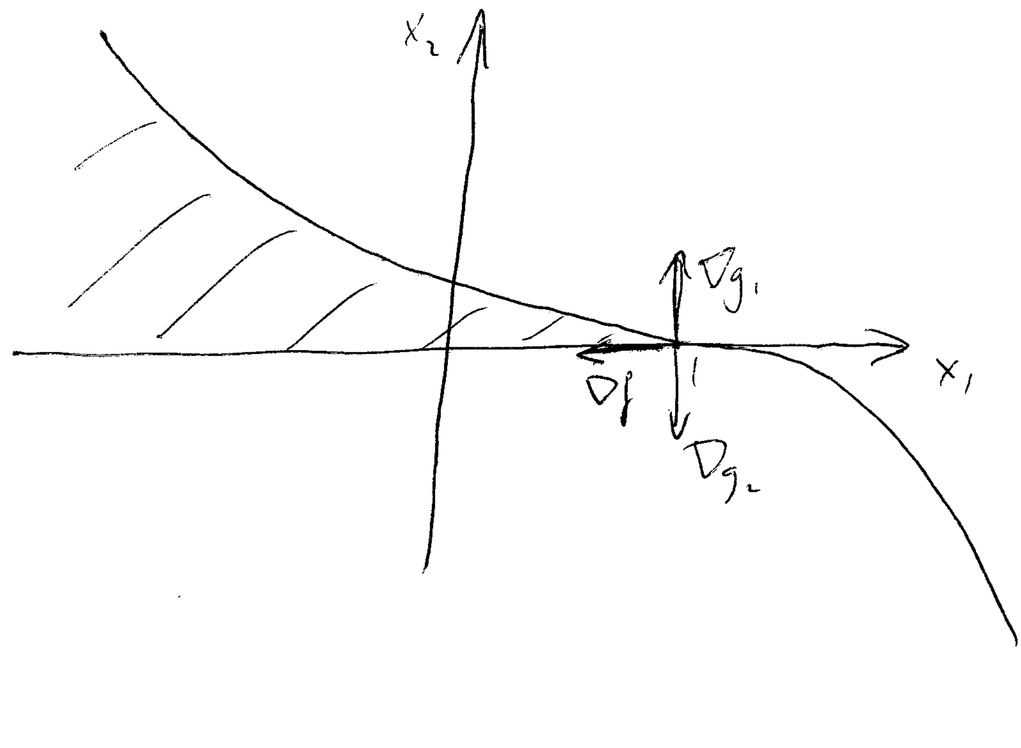
$x_2 - (1-x_1) \le 0$

$-x_2 \le 0$



at optimum.

Picture:



$$\nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \qquad \nabla g_1 = \begin{pmatrix} 3(1-x_1)^2 \\ 1 \end{pmatrix}$$

$$\bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} : \quad \nabla g_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \nabla g_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Another example: $\min \; -x_1 - x_2$

$\cancel{(1-x_1) \le 0}$ $\quad (x_1 - 1)^3 \le 0$

$(x_2 - 1)^3 \le 0$