Spl lecture: Stochastic GD — Praneeth N (Google Rs). on Thu 11 Nov.

— x —

Second Order Conditions

Revision

1. Linearized feasible directions:

$$\mathcal{F}(x) = \left\{ d \;\middle|\; \begin{array}{ll} d^T \nabla c_i(x) = 0, & i \in \mathcal{E} \quad \text{equality} \\ d^T \nabla c_i(x) \geq 0, & i \in \mathcal{A}(x) \cap \underset{\text{ineq}}{\mathcal{I}} \end{array} \right.$$

2. Active set: $\mathcal{A}(x) = \mathcal{E} \cup \{ i \in \mathcal{I} \mid c_i(x) = 0 \}$

3. KKT condns for x^* to be a local soln to: $\min_{x \in \Omega} f(x)$.

$$\boxed{\rightarrow \quad \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \quad \Rightarrow \quad \nabla f(x) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i \nabla c_i(x^*)}$$

→ $c_i(x) = 0$, $i \in \mathcal{E}$; $c_i(x) \geq 0$ $i \in \mathcal{I}$ (feasibility)

→ $\lambda_i^*(x) \geq 0$ for $i \in \mathcal{I}$

→ $\lambda_i^*(x) \underbrace{c_i(x)}_{} = 0$, $i \in \mathcal{E} \cup \mathcal{I}$

↳ if $i \in \mathcal{I}$, $i \notin \mathcal{A}(x^*)$ i.e. inactive ineq.
then $c_i(x) > 0 \Rightarrow \lambda_i^* = 0$.

FONC

Say that we are at x^* s.t. KKT are satisfied.

$\Rightarrow$ x^* is a local soln to $\left(\min_{x \in \Omega} f(x). \right)$

$\rightarrow$ say move from x^* in a direction w, then

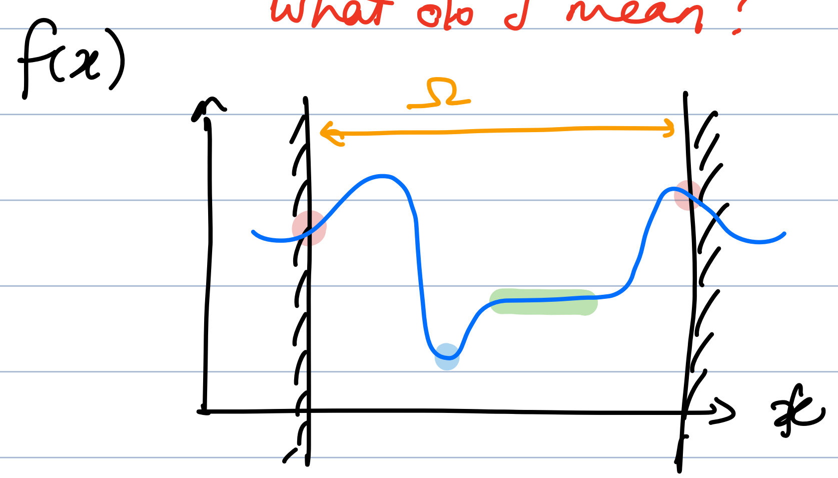
$w^T \nabla f(x^*) \geq 0$

obj. fn value increases $w^T \nabla f(x^*) > 0$ — ①

or stays the same $w^T \nabla f(x^*) = 0.$ — ②

In case ②, using only 1^{st} order info, I can't tell whether f increases/decreases/const along w. $\rightarrow$

What do I mean?



- Global minima

$f(x^*) \leq f(x) \;\; \forall x \in \Omega$

- strong local minima

there exists a nbd $\mathcal{N}$ of $x^* \in \Omega$ s.t.

$f(x^*) < f(x) \;\; \forall \; x \in \mathcal{N} \cap \Omega.$

- weak local minima

There exists a nbd N of $x^* \in \Omega$

s.t $f(x^*) \leq f(x) \;\; \forall x \in N \cap \Omega.$

KKT condns will be satisfied at all pts above

$\rightarrow$ Using 2^{nd} order info, resolve the difference between these pts $\rightarrow$ make a statement about strict local solns.

↪ So, motivation is to capture those directions (w) for which $w^T \nabla f(x^*) = 0$. What are the properties of such w?

- If I go along w, then:

1) equality constraints:

$$c_i(x^*) = 0 \rightarrow c_i(x^*+s) \approx c_i(x^*) + \nabla c_i(x^*)^T s = 0$$

$$\Rightarrow \quad \nabla c_i(x^*)^T w = 0 \quad , i \in \mathcal{E}$$

2) Deal with ineq. next.

Recall: $\min_{x \in \mathbb{R}} f(x) \quad s.t. \; c(x) \geq 0$

we were at x, moving to $x+s$: $c(x+s) \geq 0$

1st order Taylor : $c(x+s) = c(x) + \nabla c^T(x) s \geq 0$

Case 1 : $c(x) > 0 \Rightarrow$ Any small enough s ✓

Case 2 : $c(x) = 0 \Rightarrow \nabla c^T(x) d \geq 0$

KKT condns : $[\lambda_i^* c_i(x) = 0]$

Case 1 : Here $\lambda_i^* = 0$

Case 2 : Here $c(x) = 0 \Rightarrow \lambda_i > 0$

$\lambda_i = 0$

$\boxed{\text{claim:}}$

Choose w, s.t. $\left[\begin{array}{ll} \nabla c_i(x^*)^T w = 0 & \text{for } \lambda_i > 0, \\ \nabla c_i(x^*)^T w \geq 0 & \text{for } \lambda_i = 0. \end{array} \right]$

How does it help? What is $\boxed{\lambda_i^* \nabla c_i(x^*)^T w}$ for different types of i?

$i \in \mathcal{E} \rightarrow \lambda_i^* \underbrace{\nabla c_i(x^*)^T w}_{=0} = 0$

$i \in \mathcal{I} \setminus \mathcal{A}(x^*) \rightarrow \lambda_i^* \nabla c_i(x^*)^T w = 0 \quad (\because \lambda_i^* = 0)$

$$i \in \mathcal{I} \cap \mathcal{A}(x^*) \rightarrow \lambda_i^* \underline{\nabla c_i(x^*)^T w} = \begin{cases} 0 & , \lambda_i > 0 \\ 0 & , \lambda_i = 0 \end{cases}$$

Combine: $\lambda_i^* \nabla c_i(x^*)^T w = 0 \quad \forall\, i \in \mathcal{E} \cup \mathcal{I}$

$\Rightarrow$ What did KKT say?

$$\nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* \nabla c_i(x^*)$$

Take i/p with w

$$w^T \nabla f(x^*) = \sum_i \lambda_i^* w^T \nabla c_i(x^*) = 0$$

These w's are spl because if I go along these w's, $w^T \nabla f(x^*) = 0$.

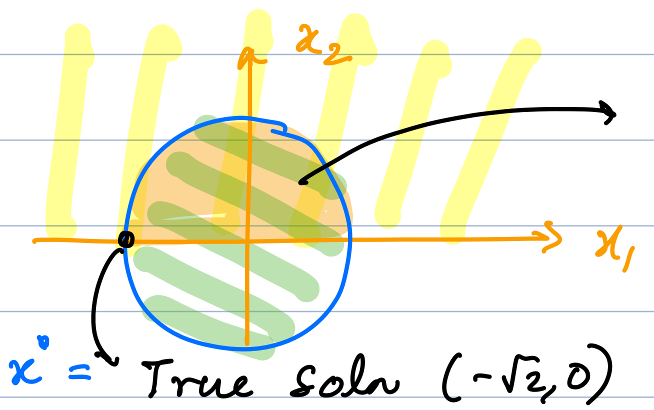
They define a <u>critical cone $C(x^*, \lambda^*)$</u>

$$w \in C(x^*, \lambda^*) \Leftrightarrow \begin{cases} \nabla c_i^T(x^*)\, w = 0 & i \in \mathcal{E} \\ \nabla c_i^T(x^*)\, w = 0, & i \in A(x) \cap \mathcal{I},\ \lambda_i > 0 \\ \nabla c_i^T(x^*)\, w \geq 0, & i \in A(x) \cap \mathcal{I},\ \underline{\lambda_i = 0} \end{cases}$$

→ If I move in this cone, obj fn value unchanged (to first order).

e.g. to show diff $\mathcal{F}(x^*)$, $C(x^*, \lambda^*)$

$$\min f(x) = x_1 \quad \text{s.t. } x_2 \geq 0,\ 2 - x_1^2 - x_2^2 \geq 0$$



Active set $= A(x^*) = \{1, 2\}$

$\nabla f = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\nabla c_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$,

$\nabla c_2 = -2 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2\sqrt{2} \\ 0 \end{pmatrix}$

At x^*, LICQ? Yes $\nabla c_1, \nabla c_2$ are lin indepn.

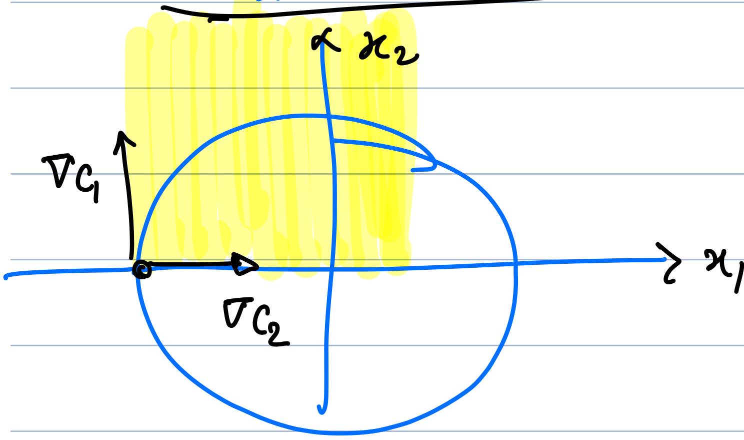
$$\nabla f = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2 \Leftrightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_1 \end{pmatrix} + \begin{pmatrix} \lambda_2 2\sqrt{2} \\ 0 \end{pmatrix}$$

$\Rightarrow$ $\underline{\lambda_1^* = 0,\ \lambda_2^* = 1/2\sqrt{2}}$. $\Rightarrow$ unique

$\Rightarrow$ KKT is satisfied at $(-\sqrt{2}, 0)$.

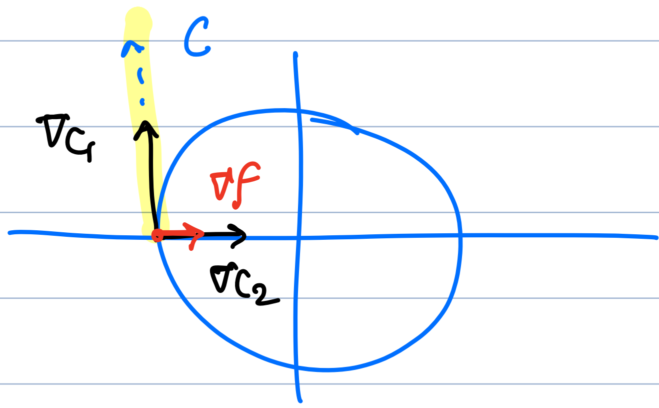
$\Rightarrow$ x^* is a local soln.

Feasible set



defn: $\mathcal{F}(x)$: $\nabla c_i^T(x^*) d \geq 0 \quad \forall i$

Critical Cone



$$\begin{cases} \nabla c_1(x^*)^T w \geq 0 & (\lambda_1^* = 0) \\ \nabla c_2(x^*)^T w = 0 & (\lambda_2^* > 0) \end{cases}$$

$$C(x^*, \lambda^*) = \left\{ \begin{pmatrix} 0 \\ w_2 \end{pmatrix} \middle| \; w_2 \geq 0 \right.$$

(check $w^T \nabla f = 0$)

Statements of 2nd order thms.

2nd order necessary condn.

If: ① x^* is a local soln to $\min_{x \in \Omega} f(x)$
② LICQ is satisfied
③ λ^* is the LM for which KKT is satisfied

Then: $\underbrace{w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)\, w \geq 0}$ $\forall\, w \in C(x^*, \lambda^*)$

2nd order sufficiency thm

If:
1. x^* is a feasible point $\in \Omega$
2. There exists a LM λ^* s.t KKT condn, are satisfied at x^*
3. $w^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)\, w \geq 0$ $\forall\, w \in C(x^*, \lambda^*)$

Then: x^* is a strict local soln of $\min_{x \in \Omega} f(x)$.