Optimization Theory and Methods

Prof. Xiaoe Ruan

Email: wruanxe@xjtu.edu.cn
Tel:13279321898

2019, Fall

Chapter1 Optimization Foundations

Introduction to optimization methods

1. Optimality Conditions

2. Introduction to Optimization Methods and Properties

3. Convergence Rate of Iterative Sequence

1. Optimality Conditions

Nonlinear Programming（NP)) problem:

$$(\mathrm{NP}):\begin{cases}\min\ f(x),\\ \text{s.t.}\quad c_i(x)=0, i\in E=\{1,2,\cdots m'\},\\ \qquad\quad c_i(x)\ge 0, i\in I=\{m'+1,\cdots m\}.\end{cases}$$

Denote $D=\{x \mid c_i(x)=0, i\in E, c_i(x)\ge 0, i\in I\}$

-constraint set or feasible domain

Def.1. **Let** $x^* \in D$. **For** $\forall x\in D$

(1) If $f(x)\ge f(x^*)$, **then** x^*

is a global optimizer (minimizer) of NP.

(2) If $f(x)> f(x^*)(x\ne x^*)$ **then** x^*

is a strictly global optimizer (minimizer) of NP.

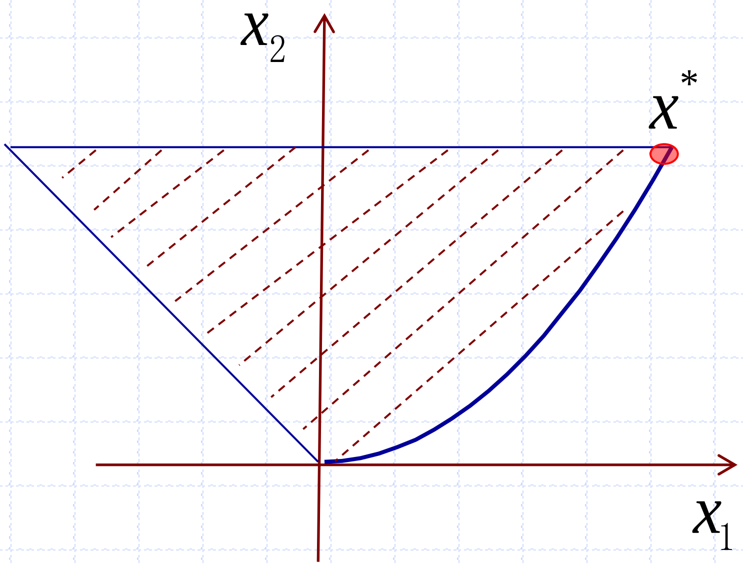
Ex1. Influence of constraints

$$\min\ f(x) = -x_1^3$$

$$s.t.\quad -x_1^2 + x_2 \geq 0 \qquad (1.1)$$

$$x_1 + x_2 \geq 0 \qquad (1.2)$$

$$1 - x_2 \geq 0 \qquad (1.3)$$



Extreme minimizer: $x^* = (1,\ 1)^T$

If constraint(1.2)has a slight perturbation,

$x^* = (1,\ 1)^T$ **is still the extreme minimizer.**

If constraint(1.1) or（1.3）is slightly perturbed,

the extreme minimizer would be changed.

Denote $N_\delta\left(x^*\right)=\left\{x \mid \left\|x-x^*\right\|<\delta\right\}$

Def.2. **Let** $x^* \in D$ **and** $\delta>0.$

(1) **If** $f(x) \geq f\left(x^*\right), \forall x \in D \cap N_\delta\left(x^*\right)$ **then** x^* **is said to be a local optimizer (minimizer)of NP problem.**

(2) **If** $f(x)>f\left(x^*\right), \forall x \in D \cap N_\delta\left(x^*\right), x \neq x^*$ **then** x^* **is said to be a strictly local optimizer (minimizer)of NP.**

Def.3. **Let** $f: R^n \rightarrow R^1$ **be differentiable at** $x \in R^n$.

If there exist a vector $0 \neq d \in R^n$ **and** $\bar{\alpha}>0$

s.t. $f(x+\alpha d)<f(x), \forall \alpha \in(0, \bar{\alpha}),$

then vector d **is named as a descent direction of function** $f(x)$ **at the point** x.

1.1. First-order Necessary Conditions

Th.1. Let $f: R^n \to R^1$ be differentiable at $x_k \in R^n$.

d is a descent direction $\Longleftrightarrow$ $\nabla f(x_k)^{\mathrm{T}} d < 0.$

Proof: $\Longleftarrow$ By means of Taylor's expansion, for $\alpha > 0$,

$$f(x_k + \alpha d) = f(x_k) + \alpha \nabla f(x_k)^{\mathrm{T}} d + o(\|\alpha d\|)$$

For $\varepsilon = -(\nabla f(x_k))^{\mathrm{T}} d$, $\exists \bar{\alpha} > 0$, s.t. for $\alpha < \bar{\alpha}$ we have

$$\frac{o(\|\alpha d\|)}{\alpha} < \frac{1}{2}\varepsilon \quad \text{that is} \quad o(\|\alpha d\|) < \frac{1}{2}\alpha\varepsilon.$$

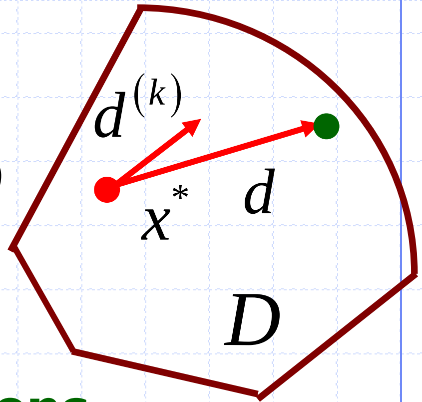
Thus
$$f(x_k + \alpha d) \le f(x_k) - \alpha\varepsilon + \frac{1}{2}\alpha\varepsilon < f(x_k).$$

$\Longrightarrow$ From sign reservation of continuous fcn, obvious.

Corollary1.

$-\nabla f(x)$ is a descent direction of $f(x)$ at x.

Def.4. (Feasible direction) Let $x^* \in D, 0 \neq d \in R^n$, d is called as a feasible direction at x^* w.r.t set D if $\exists \bar{\alpha} > 0$ s.t. $x^* + \alpha d \in D, \forall \alpha \in [0, \bar{\alpha}]$.



Denote $FD(x^*, D)$ as the set of all feasible directions at x^* w.r.t set D.

Def.5. (Sequential feasible direction) Let $x^* \in D, d \in R^n$, d is called a sequential feasible direction at x^* w.r.t the set D if there exist sequences $d^{(k)} \neq 0\ (k = 1, 2, \cdots)$ and $\alpha_k > 0\ (k = 1, 2, \cdots)$ s.t. $x^* + \alpha_k d^{(k)} \in D, k = 1, 2, \cdots$, $\lim\limits_{k\to\infty} d^{(k)} = d$, and $\lim\limits_{k\to\infty} \alpha_k = 0$.

Denote $SFD(x^*, D)$ as the set of all sequential feasible directions at x^* w.r.t the set D.

Def.6 . Effective constraint index set

For NP:
$$\begin{cases} \min\ f(x), \\ \text{s. t.}\quad c_i(x)=0, i\in E=\{1,2,\cdots m'\}, \\ \qquad\quad c_i(x)\ge 0, i\in I=\{m'+1,\cdots m\}. \end{cases}$$

Index set $I^*=\{i \mid i\in I=\{m'+1,\cdots m\}, c_i(x^*)=0\}$ **is called an active constraint index set at** x^*.

Def.7. Linearized feasible direction

Let $x^*\in D$. d **is called a linearized feasible direction at** x^*

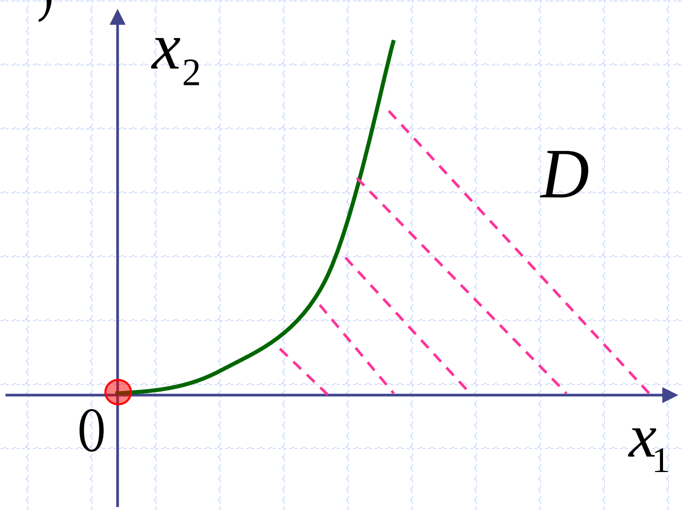
if $d^{\mathrm{T}}\nabla c_i(x^*)=0, i\in E$ **and** $d^{\mathrm{T}}\nabla c_i(x^*)\ge 0, i\in I^*$.

Denote $LFD(x^*,D)$ **as the set of all sequential feasible directions at** x^* **w.r.t the set** D.

Ex2. **Let** $D=\{(x_1,x_2)\mid x_1^3-x_2\geq 0, x_2\geq 0\},\ x^*=[0,0]^{\mathrm{T}}.$

Solve $SFD(x^*,D),\quad LFD(x^*,D).$

Solution: $SFD(x^*,D)=\left\{[a,0]^{\mathrm{T}},a>0\right\}$



Because $\nabla c_1(x^*)=[0,-1]^{\mathrm{T}},$

$\nabla c_2(x^*)=[0,1]^{\mathrm{T}}.$

Therefore

$LFD(x^*,D)=\left\{[b,0]^{\mathrm{T}},b\neq 0\right\}$

Obviously $SFD(x^*,D)\neq LFD(x^*,D),$

But $SFD(x^*,D)\subset LFD(x^*,D)$

Lemma1. Let all constraints fcns $c_i(x)(i \in E \bigcup I)$ be continuously differentiable at $x^* \in D$.

Then $FD(x^*,D) \subset SFD(x^*,D) \subset LFD(x^*,D)$.

Denote $D^* = D(x^*) = \{d \mid d^{\mathrm{T}} \nabla f(x^*) < 0, d \in R^n\}$

Th.2.（Optimality condition）

Let $x^* \in D$ be a local minimizer of NP and fcns $f(x)$ and $c_i(x)(i \in E \bigcup I)$ be continuously differentiable at x^*. Then

$$SFD(x^*,D) \bigcap D^* = \varnothing.$$

Proof:

$\forall d \in SFD(x^*,D), \exists x^{(k)} = x^* + \alpha_k d^{(k)}, s.t. \alpha_k \to 0, d^{(k)} \to d$ and

$$f(x^{(k)}) = f(x^* + \alpha_k d^{(k)}) = f(x^*) + \alpha_k d^{(k)\mathrm{T}} \nabla f(x^*) + o(\|\alpha_k d^{(k)}\|) \geq f(x^*)$$

Then $d^{\mathrm{T}} \nabla f(x^*) \geq 0$. That is $d \notin D(x^*)$.

Th.3.(1st-order Kuhn-Tucker Necessity conditions)

Let x^* **be a local minimizer of NP.** $f(x)$ **and** $c_i(x)(i \in E \cup I)$ **are continuously differentiable at** x^*. **If** $SFD(x^*, D) = LFD(x^*, D)$, **then there exists such a vector** $\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^{\mathrm{T}}$ **that**

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, i \in E \cup I, \quad \lambda_i^* \geq 0, i \in I, \quad \lambda_i^* c_i(x^*) = 0, i \in I.$$

Proof: $d^{\mathrm{T}} \nabla f(x^*) \geq 0, \forall d \in SFD(x^*, D) = LFD(x^*, D).$

Let $A_1 = \left[\nabla c_1(x^*), \cdots, \nabla c_{m'}(x^*)\right]$, $A_2 = \left[-\nabla c_1(x^*), \cdots, -\nabla c_{m'}(x^*)\right]$,

$A_3 = \left[-\nabla c_{i_1}(x^*) \cdots, -\nabla c_{i_s}(x^*)\right]$, $i_1, \cdots, i_s \in I^*$, $b = -\nabla f(x^*)$. **Then**

$$\begin{cases} [A_1, A_2, A_3]^{\mathrm{T}} d \leq 0, \\ b^T d > 0. \end{cases} \text{ no solution } \Longleftrightarrow \begin{cases} \nabla c_i(x^*)^{\mathrm{T}} d = 0, i \in E, \\ \nabla c_i(x^*)^{\mathrm{T}} d \geq 0, i \in I^*, \\ b^T d > 0. \end{cases} \text{ no solution}$$

From Farkas Lemma, $\exists y = \begin{bmatrix} \mu^{*-} \\ \mu^{*+} \\ \omega^* \end{bmatrix} \geq 0$ so that

$$\left[A_1, A_2, A_3\right] y = b = -\nabla f\left(x^*\right).$$

$\left(x^{*T}, \lambda^{*T}\right)$

Kuhn-Tucker pair

That is $-\nabla f\left(x^*\right) = A_1\mu^{*-} + A_2\mu^{*+} + A_3\omega^*$

Kuhn-Tucker point

$$= \sum_{i\in E} \mu_i^{*-}\nabla c_i\left(x^*\right) - \sum_{i\in E} \mu_i^{*+}\nabla c_i\left(x^*\right) - \sum_{i\in I^*} \omega_i^*\nabla c_i\left(x^*\right)$$

i.e. $\nabla f\left(x^*\right) = \sum_{i\in E}\left(\mu_i^{*+} - \mu_i^{*-}\right)\nabla c_i\left(x^*\right) + \sum_{i\in I^*} \omega_i^*\nabla c_i\left(x^*\right)$

K-T conditions

Let $\lambda_i^* = \mu_i^{*+} - \mu_i^{*-}, i \in E;\quad \lambda_i^* = \omega_i^*, i \in I^*;\quad \lambda_i^* = 0, i \in I \setminus I^*.$

Then

$$\nabla f\left(\boxed{x^*}\right) - \sum_{i=1}^{m} \lambda_i^*\nabla c_i\left(x^*\right) = 0, i \in E \cup I,$$

$$\lambda_i^* \geq 0, i \in I, \quad \boxed{\lambda_i^* c_i\left(x^*\right) = 0, i \in I.}$$

Complementary relaxed conds.

Remarks:

1. If $SFD(x^*, D) = LFD(x^*, D)$ then x^* is Kuhn-Tucker pnt.

But not easy to testify. Can be replaced by simple condition later.

If $SFD(x^*, D) \neq LFD(x^*, D)$ then x^* is not definitely a K-T pnt.

2. Define $n+m$ -variable fcn $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$

Then $L(x, \lambda)$ is called Lagrangian fcn of NP.

Thus, Eq in K-T cond $\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, i \in E \cup I,$

is expressed as $\nabla_x L(x^*, \lambda^*) = 0.$ λ^* is Lagrangian multiplier

3. Let $f(x), x \in N(x^*)$ be continuously differentiable.

If $f(x^*) = \min_{x \in R^n} f(x)$ then $\nabla f(x^*) = 0.$

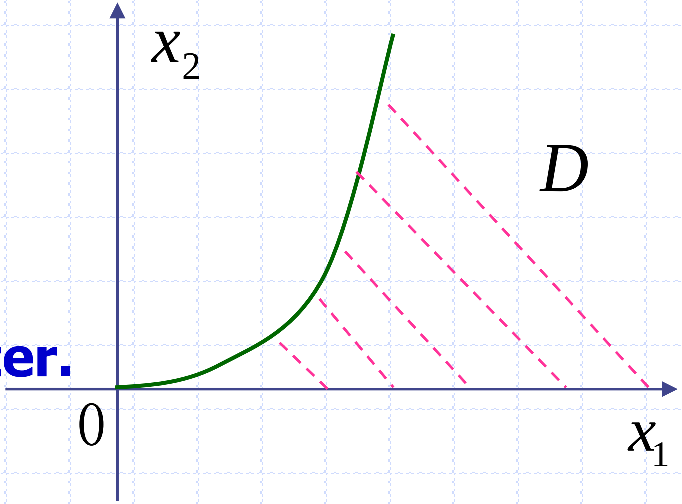
The constraint conditions are sufficient but not necessary.

Ex3.

$$\min\ x_1$$

$$s.t.\ x_1^3 - x_2 \geq 0$$

$$x_2 \geq 0$$



Obviously, $x^* = [0,0]^{\mathrm{T}}$ **is the minimizer.**

As $\nabla f(x^*) = [1,0]^{\mathrm{T}}$,

$$\nabla c_1(x^*) = [0,-1]^{\mathrm{T}}, \quad \nabla c_2(x^*) = [0,1]^{\mathrm{T}}.$$

Therefore,

no matter what values $\lambda_1^*,\ \lambda_2^*$ **are chosen, we have**

$$\nabla f(x^*) \neq \lambda_1^* \nabla c_1(x^*) + \lambda_2^* \nabla c_2(x^*)$$

Th.4. **At the local minimizer x^* if any of the following cond. holds**

(1) $c_i(x)\left(i\in E\bigcup I^*\right)$ **are linear.**

(2) $\nabla c_i\left(x^*\right)\left(i\in E\bigcup I^*\right)$ **are linearly independent.**

Then $SFD\left(x^*,D\right)=LFD\left(x^*,D\right)$

By Lemma1, $SFD\left(x^*,D\right)\subset LFD\left(x^*,D\right).$

Proof: **(1)** $\forall d\in LFD\left(x^*,D\right)$ **then** $d\neq 0$

and $\nabla c_i\left(x^*\right)^{\mathrm{T}}d=0,i\in E,\quad \nabla c_i\left(x^*\right)^{\mathrm{T}}d\geq 0,i\in I^*.$ **Let** $x^{(k)}=x^*+\alpha_k d^{(k)}$

where $d^{(k)}=d\to d\left(k\to\infty\right),\quad \forall 0<\alpha_k\to 0\left(k\to\infty\right).$

Then $c_i\left(x^{(k)}\right)=c_i\left(x^*+\alpha_k d\right)=c_i\left(x^*\right)+\alpha_k d^{\mathrm{T}}\nabla c_i\left(x^*\right).$ **That is**

$$c_i\left(x^{(k)}\right)=\alpha_k d^{\mathrm{T}}\nabla c_i\left(x^*\right)=0,i\in E,\qquad c_i\left(x^{(k)}\right)=\alpha_k d^{\mathrm{T}}\nabla c_i\left(x^*\right)\geq 0,i\in I^*,$$

$$c_i\left(x^{(k)}\right)=c_i\left(x^*\right)+\alpha_k d^{\mathrm{T}}\nabla c_i\left(x^*+\theta\alpha_k d\right)\geq 0,\left(0<\theta<1,\alpha_k<<1\right),\ i\in I\setminus I^*.$$

Thus $d\in SFD\left(x^*,D\right)$ **Therefore** $LFD\left(x^*,D\right)\subset SFD\left(x^*,D\right).$

(2) $\nabla c_i\left(x^*\right)\left(i \in E \bigcup I^*\right)$ **are linearly independent.**

Proof: $\forall d \in LFD\left(x^*, D\right)$ **Then** $d \neq 0$ **and** $\nabla c_i\left(x^*\right)^{\mathrm{T}} d = 0, i \in E,$

$\nabla c_i\left(x^*\right)^{\mathrm{T}} d \geq 0, i \in I^*.$ **Denote** $E \bigcup I^* = \{1, 2, \cdots, l\}, m' \leq l \leq n.$

From $\nabla c_i\left(x^*\right)\left(i \in E \bigcup I^*\right)$ **independence,** $\exists b_{l+1}, \cdots, b_n \in R^n$ **s.t.**

$\nabla c_1\left(x^*\right), \cdots, \nabla c_l\left(x^*\right), b_{l+1}, \cdots, b_n$ **are independent.**

Then the Eq

$$\begin{cases} c_i\left(x\right) - d^{\mathrm{T}} \nabla c_i\left(x^*\right) \theta = 0, i = 1, 2, \cdots, l; \\ \left(x - x^*\right)^{\mathrm{T}} b_i - d^{\mathrm{T}} b_i \theta = 0, i = l+1, \cdots, n. \end{cases} \qquad (**)$$

has a solution $\begin{bmatrix} x^* \\ 0 \end{bmatrix}$ **w.r.t variables** $\left[x^{\mathrm{T}}, \theta\right]^{\mathrm{T}} \in R^{n+1}$

and Jacobian matrix is nonsingular at x^*. **Thus Eq (**) can**

determine fcn $x = x(\theta), \theta \in N(0)$ **at neighborhood of** $\begin{bmatrix} x^* \\ 0 \end{bmatrix}$

and $x^* = x(0).$

Differentiating both sides of Eq (**) w.r.t θ yields

$$\begin{cases} \nabla c_i\left(x^*\right)^{\mathrm{T}}\left(\left.\dfrac{dx}{d\theta}\right|_{\theta=0}-d\right)=0, i=1,2,\cdots,l; \\ b_i^{\mathrm{T}}\left(\left.\dfrac{dx}{d\theta}\right|_{\theta=0}-d\right)=0, i=l+1,\cdots,n. \end{cases}$$

has a unique solution $\left.\dfrac{dx}{d\theta}\right|_{\theta=0}=d.$

By mean theorem $x=x(\theta)=x(0)+\left.\dfrac{\mathrm{d}x}{\mathrm{d}\theta}\right|_{\theta=0+a(\theta)\theta}\theta=x^*+\theta d(\theta),\quad 0<a(\theta)<1.$

Specifically, choose

$$0<\theta_k\to 0(k\to\infty),\quad x^{(k)}=x^*+\theta_k d(\theta_k)=x^*+\theta_k d_k$$ **where**

$$d_k=d(\theta_k)=\left.\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta}\right|_{\theta=a(\theta_k)\theta_k}\to\left.\frac{\mathrm{d}x(\theta)}{\mathrm{d}\theta}\right|_{\theta=0}=d.$$ **Substituting** $\left[x^{(k)\mathrm{T}},\theta_k\right]^{\mathrm{T}}$ **into Eq makes**

$$c_i\left(x^{(k)}\right)=\theta_k d_k^{\mathrm{T}}\nabla c_i\left(x^*\right)=0, i\in E,\ c_i\left(x^{(k)}\right)=\theta_k d_k^{\mathrm{T}}\nabla c_i\left(x^*\right)\geq 0, i\in I^*,$$

$$c_i\left(x^{(k)}\right)=c_i\left(x^*\right)+\theta_k d_k^{\mathrm{T}}\nabla c_i\left(x^*\right)+o\left(\left\|\theta_k d_k\right\|\right)\geq 0,\left(i\in I\setminus I^*\right)$$

Thus $d\in SFD\left(x^*,D\right)$ **Therefore** $LFD\left(x^*,D\right)\subset SFD\left(x^*,D\right).$

Th.5. At the local minimizer x^* if any of the following cond. holds

(1) $c_i(x)\left(i \in E \cup I^*\right)$ are linear.

(2) $\nabla c_i\left(x^*\right)\left(i \in E \cup I^*\right)$ are linearly independent.

Then there exists such a vector $\lambda^* = \left(\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*\right)^{\mathrm{T}}$ that

$$\nabla f\left(x^*\right) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i\left(x^*\right) = 0, i \in E \cup I,$$

$$\lambda_i^* \geq 0, i \in I, \quad \lambda_i^* c_i\left(x^*\right) = 0, i \in I.$$

Th.6.（1st-order sufficiency cond. of an optimizer）

$x^* \in D$ is a strictly local minimizer if $f(x)$ and $c_i(x)(i \in E \cup I)$ are 1st-order continuously differentiable at x^* and

$$d^{\mathrm{T}} \nabla f\left(x^*\right) > 0, \forall d \in SFD\left(x^*, D\right) \subset LFD\left(x^*, D\right)$$

1.2. 2nd-order Optimization Conditions

The sufficiency requires the 2-nd-order derivatives of the objective fcn and the constraints fcns.

Reminder: Let x^* be K--T point. Then Lagrangian multiplier

satisfying $\nabla f\left(x^*\right)-\sum_{i=1}^{m}\boxed{\lambda_i^*}\nabla c_i\left(x^*\right)=0, i\in E\bigcup I,\quad \lambda_i^*\geq 0, i\in I,$

and $\boxed{\lambda_i^*}c_i\left(x^*\right)=0, i\in I.$ Then $\lambda_i^*\geq 0, i\in I^*.$

Thus for some index *i* $\boxed{\lambda_i^*=0}$ or $\boxed{\lambda_i^*>0}$ $\Rightarrow$ effective on cond.

$\Downarrow$

no effect on cond.

Denote $I_+^*=I_+\left(x^*\right)=\left\{i\middle| i\in I^*,\lambda_i^*>0\right\}$ Then $A_+(x^*)=E\bigcup I_+(x^*)$

is called a strongly effective constraints index set at x^*.

or a strongly effective set.

Def.8. Sequential nullity constraints direction

$$G\left(x^*,\lambda^*\right)\subseteq LFD\left(x^*,D\right)$$
$$S\left(x^*,\lambda^*\right)\subseteq SFD\left(x^*,D\right)$$
$$S\left(x^*,\lambda^*\right)\subseteq G\left(x^*,\lambda^*\right)$$

Let $\left(x^{*\mathrm{T}},\lambda^{*\mathrm{T}}\right)$ **be Kuhn-Tucker pair of NP.**

d **is a sequential nullity constraints direction at** x^* **if there exist sequences**

$$d^{(k)}\neq 0\left(k=1,2,\cdots\right) \quad \textbf{and} \quad \alpha_k>0\left(k=1,2,\cdots\right).$$

so that $x^{(k)}=x^*+\alpha_k d^{(k)}\in D, k=1,2,\cdots$ **satisfying**

$$c_i\left(x^{(k)}\right)=0, i\in E\bigcup I_+^* \quad \textbf{and} \quad c_i\left(x^{(k)}\right)\geq 0, i\in I\setminus I_+^*$$

together with $\lim\limits_{k\to\infty} d^{(k)}=d$ **and** $\lim\limits_{k\to\infty}\alpha_k=0.$

Denote $S\left(x^*,\lambda^*\right)$ **as a set of all SNCDs at** x^* **and**

$$G\left(x^*,\lambda^*\right)=\left\{d \,\middle|\, d\neq 0, d^{\mathrm{T}}\nabla c_i\left(x^*\right)=0\left(i\in E\bigcup I_+^*\right), d^{\mathrm{T}}\nabla c_i\left(x^*\right)\geq 0\left(i\in I^*\setminus I_+^*\right)\right\}$$

as a set of all linearized nullity constraints directions at x^*

Th.7. (2nd-order Kuhn-Tucker Necessity Conditions)

Let x^* **be a local minimizer of NP.** $f(x)$ **and** $c_i(x)(i \in E \bigcup I)$ **are 2nd-order continuously differentiable at** x^*.

If $SFD(x^*, D) = LFD(x^*, D)$, **then there exists such a vector** $\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^{\mathrm{T}}$ **that**

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, i \in E \bigcup I,$$

$$\lambda_i^* \geq 0, i \in I, \qquad \lambda_i^* c_i(x^*) = 0, i \in I.$$

$$d^{\mathrm{T}} \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, \forall d \in S(x^*, \lambda^*).$$

where $L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$

$$\nabla_{xx}^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 c_i(x^*)$$

Lemma2:

At a local optimizer x^* if any of the following holds

(1) $c_i(x)\left(i \in E \cup I^*\right)$ **are linear.**

(2) $\nabla c_i\left(x^*\right)\left(i \in E \cup I^*\right)$ **are linearly independent.**

Then $G\left(x^*, \lambda^*\right) = S\left(x^*, \lambda^*\right).$

Th.8. (2nd-order Necessity Conditions)

Let x^* be a local minimizer of NP. $f(x)$ **and** $c_i(x)(i \in E \cup I)$ **are 2nd-order continuously differentiable at** x^*. **If**

(1) $c_i(x)(i \in E \cup I^*)$ **are linear. or**

(2) $\nabla c_i(x^*)(i \in E \cup I^*)$ **are linearly independent.**

Then there exists a vector $\lambda^* = (\lambda_1^*, \lambda_2^*, \cdots, \lambda_m^*)^{\mathrm{T}}$ **such that**

$$\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, i \in E \cup I,$$

$$\lambda_i^* \geq 0, i \in I, \qquad \lambda_i^* c_i(x^*) = 0, i \in I.$$

$$d^{\mathrm{T}} \nabla_{xx}^2 L(x^*, \lambda^*) d \geq 0, \forall d \in S(x^*, \lambda^*).$$

where

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x)$$

$$\nabla_{xx}^2 L(x^*, \lambda^*) = \nabla^2 f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla^2 c_i(x^*)$$

Th.9. (2nd-order Sufficiency Conditions)

Let $\left(x^{*\mathrm{T}}, \lambda^{*\mathrm{T}}\right)$ be Kuhn-Tucker pair of NP.

If $f(x)$ and $c_i(x)(i \in E \bigcup I)$ are 2nd-order continuously differentiable at x^*

and $d^{\mathrm{T}} \nabla_{xx}^2 L\left(x^*, \lambda^*\right) d > 0, \forall d \in G\left(x^*, \lambda^*\right)$.

Then x^* is a strictly local optimizer.

Corollary2.(2nd-order unconstraint sufficiency cond.)

If $f(x)$ is 2nd-order continuously differentiable in nbhd $N\left(x^*\right)$ along with $\nabla f\left(x^*\right) = 0$ and $\nabla^2 f\left(x^*\right)$ positive definite.

Then x^* is a strictly local optimizer of $\min\limits_{x \in R^n.} f(x)$.

1.3. Convex Programming (CP)

For Nonlinear Programming（NP)):

$$\min f(x), \text{s. t.} \begin{cases} c_i(x)=0, i\in E=\{1,2,\cdots m'\}, \\ c_i(x)\ge 0, i\in I=\{m'+1,\cdots m\}. \end{cases}$$

$$D=\{x \mid c_i(x)=0, i\in E, c_i(x)\ge 0, i\in I\}$$

If $f(x)$ **and** D **are convex, then NP is called CP.**

D **is convex if** $c_i(x)(i\in E)$ **linear and** $c_i(x)(i\in I)$ **concave.**

The course addresses CP on Convex Domain.

Ex4. **（NP) :**

$$\begin{cases} \min f(x,y,z)=2x^2+3y^2+4z^2, \\ \text{s. t.} \quad c_1(x,y,z)=x+y-z+\frac{2}{3}=0, \\ \qquad c_2(x,y,z)=2-(x^2+y^2+z^2)\ge 0. \end{cases}$$

Th.10. A local minimizer of CP is a global minimizer.

Proof: **Let** x^* **be a local minimizer but not a global one.**

Then there exists a neighborhood $N_\delta(x^*) = \{x \mid \|x - x^*\| < \delta\}$

and a point $\bar{x} \neq x^*, \bar{x} \in D$ **but** $\bar{x} \notin N_\delta(x^*)$

such that $f(x^*) \leq f(x), \forall x \in N_\delta(x^*)$ **but** $f(\bar{x}) < f(x^*)$.

Choose $0 < \alpha < \dfrac{\delta}{\|\bar{x} - x^*\|}$ **and** $\tilde{x} = \alpha \bar{x} + (1-\alpha) x^* \in D,$

Then $\|\tilde{x} - x^*\| = \alpha \|\bar{x} - x^*\| < \delta.$ **Thus** $\tilde{x} \in N_\delta(x^*).$

From convexity

$$f(\tilde{x}) = f\left(\alpha\bar{x} + (1-\alpha)x^*\right) \leq \alpha f(\bar{x}) + (1-\alpha) f(x^*) \leq f(x^*).$$

Th.11. Let $f(x)$ and $c_i(x)(i \in E \cup I)$ of CP be continuously differentiable and x^* be K-T point. Then x^* must be a global minimizer.

Proof: Let $\left(x^{*\mathrm{T}}, \lambda^{*\mathrm{T}}\right)$ be Kuhn-Tucker pair. Then

$$f(x) \geq f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*), \quad c_i(x) = c_i(x^*) + \nabla c_i(x^*)^{\mathrm{T}}(x - x^*)(i \in E)$$

$$-c_i(x) \geq -c_i(x^*) - \nabla c_i(x^*)^{\mathrm{T}}(x - x^*)(i \in I)$$

Thus

$$\begin{aligned} f(x) &\geq f(x) - \sum_{i=1}^{m} \lambda_i^* c_i(x) \\ &\geq f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*) - \sum \lambda_i^* \left(c_i(x^*) + \nabla c_i(x^*)(x - x^*)\right) \\ &= f(x^*) - \sum_{i=1}^{m} \lambda_i^* c_i(x^*) + \left[\nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*)\right]^{\mathrm{T}}(x - x^*) \\ &= f(x^*) - \sum_{i=1}^{m} \lambda_i^* c_i(x^*) = f(x^*) \end{aligned}$$

Corollary3.

If $f(x)$ of CP is strictly convex, the minimizer is unique.

Slater condition: There exists $\bar{x} \in R^n$ such that

$$c_i(\bar{x}) = 0, i \in E; \text{ and } c_i(\bar{x}) > 0, i \in I.$$

Th.12. Let $f(x)$ and $c_i(x)(i \in E \cup I)$ of CP is continuously differentiable and satisfy Slater condition.

Then

A feasible point x^* is a minimizer

iff x^* is a Kuhn-Tucker point.

Ex5.

NP:

$$\min \; f(x) = f(x_1, x_2) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}x_2^{\,2}$$

$$s.t. \quad c(x) = -x_1 + bx_2^{\,2} \geq 0$$ where ***b*** is const.

How to choose constant *b* so that $x^* = [0,0]^{\mathrm{T}}$ **is a local minimizer.**

Solution:

$$D = \{(x_1, x_2)^{\mathrm{T}} \mid c(x) = -x_1 + bx_2^{\,2} \geq 0\}$$

$$c(x^*) = 0, \qquad \nabla f(x^*) = [-1, 0]^{\mathrm{T}}, \qquad \nabla c(x^*) = [-1, 0]^{\mathrm{T}}.$$

$$L(x, \lambda) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}x_2^{\,2} - \lambda(-x_1 + bx_2^{\,2})$$

K-T conds:

$$\nabla_x L(x^*, \lambda) = [-1, 0]^{\mathrm{T}} - \lambda[-1, 0]^{\mathrm{T}} = 0$$

$$c(x^*) = 0$$

$$\lambda \geq 0$$

Solving yields $\lambda^* = 1 > 0$. **Thus** $x^* = [0,0]^{\mathrm{T}}$ **is a K-T point.**

Hessian matrix of Lagrangian fcn at *x is** $W^* = \begin{bmatrix} 1 & 0 \\ 0 & 1-2b \end{bmatrix}$

Kinds of feasible directions sets are

$$S(x^*,\lambda^*) = G(x^*,\lambda^*) = \{d = [0,a]^{\mathrm{T}}, a \neq 0\}$$

Note $SFD(x^*,D) = \{d \mid [-1,0]^{\mathrm{T}} d \geq 0, d \neq 0\} = \{[a_1,a_2]^{\mathrm{T}}, a_1 < 0\}$

Therefore

$$[0,a]W^*\begin{bmatrix} 0 \\ a \end{bmatrix} = [0,a]\begin{bmatrix} 1 & 0 \\ 0 & 1-2b \end{bmatrix}\begin{bmatrix} 0 \\ a \end{bmatrix} = a^2(1-2b)$$

Thus

If $b > 0.5$ x^* **is not a local minimizer from 2nd-order necessity.**

If $b < 0.5$ x^* **is a local minimizer from 2nd-order sufficiency.**

If $b = 0.5$

As $f(x) = (1 + \frac{1}{2}x_1^{\ 2}) - x_1 + \frac{1}{2}x_2^{\ 2}$

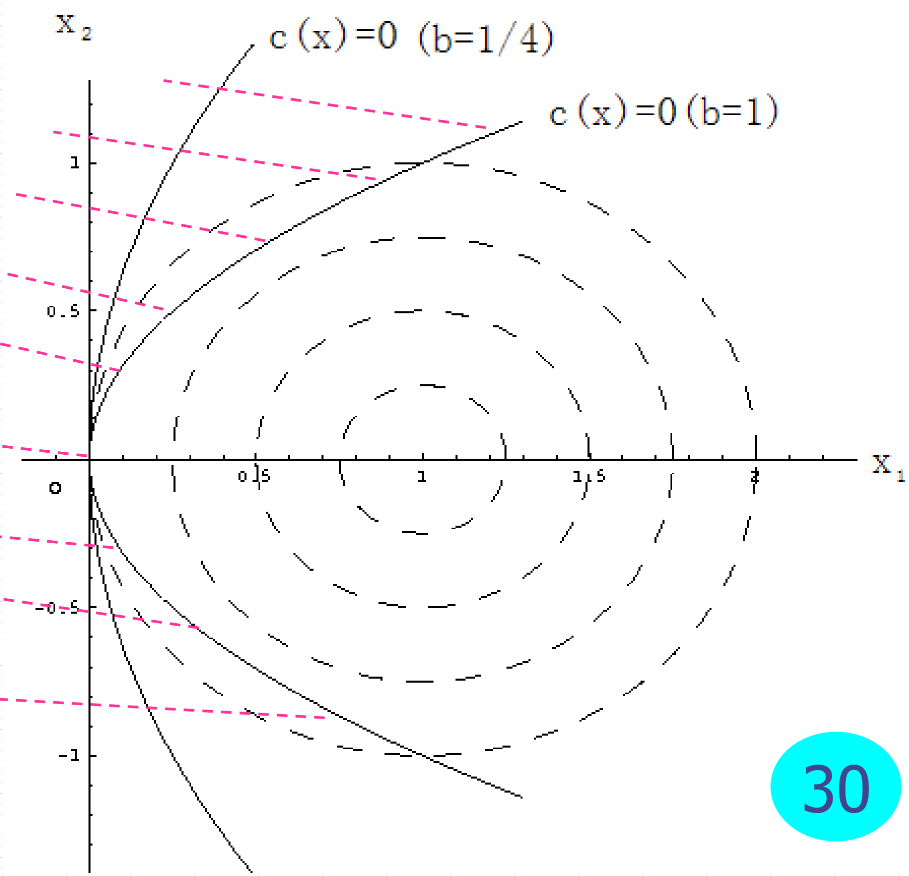
s.t. $c(x) = -x_1 + \frac{1}{2}x_2^{\ 2} \geq 0$

i.e. $f(x) \geq (1 + \frac{1}{2}x_1^{\ 2})$

Then

x^* **is a local minimizer.**

$b - x^*$ **relationship as figured.**



2. Brief Introduction to Optimization Methods and Properties

For Nonlinear Programming（NP）:

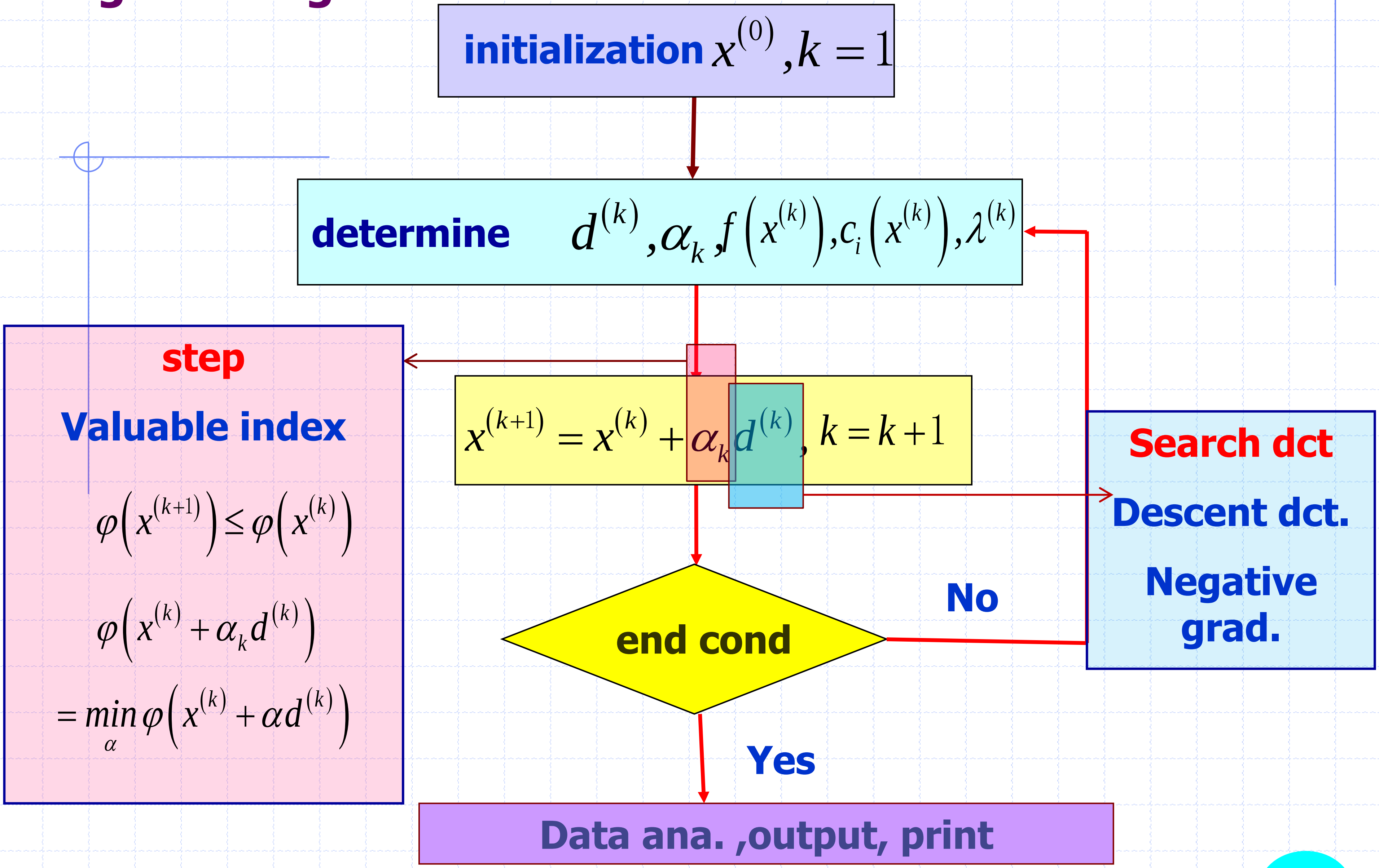
$$(\text{NP}):\quad \begin{cases} \min\ f(x), \\ \text{s. t.}\quad c_i(x)=0, i\in E=\{1,2,\cdots m'\}, \\ \qquad\quad c_i(x)\ge 0, i\in I=\{m'+1,\cdots m\}. \end{cases}$$

Usual method is to solve a Kuhn-Tucker point of following Eqs. by iterative algorithm

$$\nabla f(x^*)-\sum_{i=1}^{m}\lambda_i^*\nabla c_i(x^*)=0, i\in E\cup I,$$

$$\lambda_i^*\ge 0, i\in I, \qquad \lambda_i^* c_i(x^*)=0, i\in I.$$

Programming flow



3. Convergence Rate of Iterative Sequence

Let $\left\{x^{(k)}\right\}, e\left(x^{(k)}\right) = \left\| x^{(k)} - x^{*} \right\| \to 0 \left(k \to \infty\right)$

$$Q_p = \limsup_{k\to\infty} \frac{e\left(x^{(k+1)}\right)}{e\left(x^{(k)}\right)^p}, \quad p \in [1, +\infty) \qquad Q-\textbf{factor}$$

is called Quotient convergence factor of sequence $\left\{x^{(k)}\right\}$.

If $Q_1 = 0, \left\{x^{(k)}\right\}$ **is said to Q super-linearly converges to** x^{*}

If $0 < Q_1 < 1, \left\{x^{(k)}\right\}$ **is said to Q linearly converges to** x^{*}

If $Q_1 = 1, \left\{x^{(k)}\right\}$ **is said to Q sub-linearly converges to** x^{*}

Analogously $Q_2 = 0, 0 < Q_2 < 1, Q_2 = 1$

are respectively Q super-square, square and sub-square convergent.

$$R_p = \begin{cases} \limsup\limits_{k\to\infty} e\left(x^{(k)}\right)^{\frac{1}{k}}, p = 1, \\ \limsup\limits_{k\to\infty} e\left(x^{(k)}\right)^{\frac{1}{p^k}}, p > 1. \end{cases}$$

$R-$ **factor**

is called Root convergence factor of sequence $\left\{x^{(k)}\right\}$.

If $R_1 = 0$, $\left\{x^{(k)}\right\}$ **is said to R superlinearly converges to** x^*

If $0 < R_1 < 1$, $\left\{x^{(k)}\right\}$ **is said to R linearly converges to** x^*

If $R_1 = 1$, $\left\{x^{(k)}\right\}$ **is said to R sublinearly converges to** x^*

Analogously **If** $R_2 = 0, 0 < R_2 < 1, R_2 = 1,$ **sequence** $\left\{x^{(k)}\right\}$

are respectively R super-quadratically, quadratically and sub-quadratic ally convergent.

THANK YOU FOR ATTENDING