Lecture 14

Thursday, 17 February 2022 1:51 PM

Problem (Tut. Sheet 5) x^* is a minimizer of $f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} - \underline{b}^T\underline{x} + \underline{c}$, $Q > 0$ (SPD). If $(\underline{x}^{(0)} - x^*)$ is an eigenvector of Q, then the steepest descent converges in one step.

Pf. $\nabla f(\underline{x}) = Q\underline{x} - \underline{b}$

$\nabla f(\underline{x}^*) = Q\underline{x}^* - \underline{b} = 0$

$x^{(0)} - x^* = e^{(0)}$

$$\boxed{g^{(0)} = Q e^{(0)} = \lambda e^{(0)}}$$

$\nabla f(x^{(0)}) = Q x^{(0)} - \underline{b}$

$= Q(x^{(0)} - x^*)$

$\nabla f(x^{0}) = Q e^{(0)} = g^{(0)}$

$$\alpha_k = \frac{(g^{(k)})^T g^{(k)}}{(g^{(k)})^T Q g^{(k)}}$$

$$x^{(1)} = x^{(0)} - \alpha_0 g^{(0)}$$

$$= x^{(0)} - \frac{(g^{(0)})^T g^{(0)}}{(g^{(0)})^T Q \underbrace{g^{(0)}}_{\lambda e^{(0)}}} \overbrace{g^{(0)}}^{\lambda e^{(0)}}$$

$$= x^{(0)} - \frac{(g^{(0)})^T g^{(0)}}{(g^{(0)})^T \underbrace{Q e^{(0)}}_{(g^{(0)})}} e^{(0)} = x^{(0)} - (x^{(0)} - x^{(*)}) = \underline{\underline{x^*}}.$$

Steepest descent

[Conjugate directions Newton's

[Conjugated gradient.

Drawback of Steepest descent

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \nabla f(x^{(k)})$$

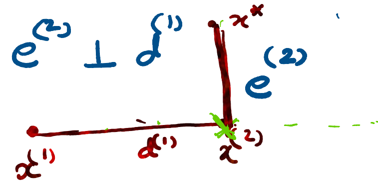
Takes same steps as earlier steps.

I. Idea : " Every time you take a step, take it right."

Pick up a set of mutually orthogonal search directions $\{d^{(1)}, d^{(2)}, \dots, d^{(n)}\}$. In each direction, take exactly ... After 'n' steps, we

Pick up ... $\{d^{(1)}, d^{(2)}, \ldots, d^{(n)}\}$. In each direction, take exactly one step with the right length. After 'n' steps, we will be done.

$e^{(2)} \perp d^{(1)}$



$$\underline{x}^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$$

$$\underline{e}^{(k+1)} = e^{(k)} + \alpha_k d^{(k)}.$$

$$e^{(k)} = x^{(k)} - x^{(*)}$$

Wish: $e^{(k+1)} \perp d^{(k)}$

$$Q e^{(k)} = Q x^{(k)} - \underline{b}$$
$$= \nabla f(x^{(k)}) = g^{(k)}$$

$$0 = (d^{(k)})^T e^{(k+1)} = (d^{(k)})^T e^{(k)} + \alpha_k (d^{(k)})^T \underline{d}^{(k)}$$

$$\Rightarrow \boxed{\alpha_k = -\frac{(d^{(k)})^T e^{(k)}}{(d^{(k)})^T d^{(k)}}}$$

We don't know $e^{(k)}$!

Residual equation $\leftarrow$ $Q e^{(k)} = r^{(k)}$

$Q \underline{x}^* = \underline{b}$

II. Rather than choosing orthogonal search directions $\{\underline{d}^{(1)}, \ldots, \underline{d}^{(n)}\}$, can we choose '$Q$' orthogonal search directions?

(i) What does This (Q orthogonality) mean?

(ii) Can we compute α_k easily?

$$(i) \quad (d^{(i)})^T Q d^{(j)} = 0 \qquad i \neq j. \quad [Q \text{ orthogonality}]$$

$$(d^{(k)})^T Q \underline{e}^{(k+1)} = 0 \qquad \underline{e}^{k+1} \overset{Q}{\perp} \underline{d}^{(k)}$$

$$\Rightarrow (d^{(k)})^T r^{(k+1)} = 0 \qquad \underline{x}^{k+1} = x^{(k)} + \alpha \underline{\underline{d}}^{(k)}$$

$$\Rightarrow \quad (\nabla f(x^{k+1}))^T \underline{d}^{(k)} = 0$$

$$\Rightarrow \quad \nabla f(x^{(k+1)})^T \frac{d}{d\alpha}(x^{(k+1)}) = 0$$

$\Rightarrow \quad \nabla f(x^{\cdots}) \quad \frac{d}{d\alpha}(\ldots)$

$$\Rightarrow \quad \frac{d}{d\alpha}\left(f(x^{(k+1)}) \right) = 0$$

Q-orthogonality is equivalent to finding a minimum in the direction $d^{(k)}$. [as in steepest descent, instead of finding min. $g^{(k)}$, we find in the direction of $d^{(k)}$]

(iii) <u>How to compute α_k ?</u>

$$(d^{(k)})^T Q e^{(k+1)} = 0$$

$$(d^{(k)})^T Q \left(e^{(k)} + \alpha_k d^{(k)} \right) = 0$$

$$\Rightarrow \quad \alpha_k = \frac{-(d^{(k)})^T \overbrace{Q e^{(k)}}^{g^{(k)}}}{(d^{(k)})^T Q d^{(k)}} = -\frac{(d^{(k)})^T g^{(k)}}{(d^{(k)})^T Q d^{(k)}}.$$

$$\boxed{\begin{aligned} x^{(k+1)} &= x^{(k)} + \alpha_k d^{(k)} \\ \alpha_k &= -\frac{(d^{(k)})^T g^{(k)}}{(d^{(k)})^T Q d^{(k)}} \end{aligned}}$$

$$g^{(k)} = \nabla f(x^{(k)})$$

III. If we choose 'Q' orthogonal directions can we have convergence after $(n+1)$ steps? $\boxed{\text{YES}}$

IV. How to construct 'Q' orthogonal search directions?

✓ Gram-Schmidt $\boxed{\text{Drawback}}$ Storage & computation.

→ Residuals $\{g^{(k)}\}$.

Choose ... such a way that most of

V. Choose Smart directions (→ Residuals $\{r^{(k)}\}$.) in such a way that most of the terms in Gram-Schmidt process vanish.

↳ Conjugate'd' Gradient method.

CG method:

Sketch.

- $d^{(1)} = \underline{\underline{r}}^{(1)} = Qx^{(1)} - \underline{b}$
- $\alpha_k = \cdots\cdots$
- $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$
- $r^{(k+1)} = r^{(k)} + \alpha_k A d^{(k)}$.
- $\beta_{k+1} = \cdots\cdots$ [Gram-Schmidt]
- $\underline{d}^{(k+1)} = \underline{r}^{(k+1)} + \underline{\beta_{k+1} d^{(k)}}$

Lecture 16

Thursday, 3 March 2022 1:33 PM

CG method

I. $\{ \underline{d}^{(1)}, \underline{d}^{(2)}, \ldots, \underline{d}^{(n)} \}$ mutually orthogonal; one step in one direction.

$$\boxed{\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \underline{d}^{(k)};}$$

$$\boxed{\alpha_k = - \frac{(\underline{d}^{(k)})^T \underline{e}^{(k)}}{(\underline{d}^{(k)})^T \underline{d}^{(k)}}} \; ??$$

Useless.

II. Change to Q orthogonal search directions.

$$(\underline{d}^{(i)})^T Q \, \underline{d}^{(k)} = 0$$

$$\boxed{\alpha_k = - \frac{(\underline{d}^{(k)})^T \underline{g}^{(k)}}{(\underline{d}^{(k)})^T Q \, \underline{d}^{(k)}}}$$

Can be computed.

III. 'Q' orthogonal directions $\{\underline{d}^{(j)}\}_{j=1}^{n}$ ⇒ Convergence after '(n+1)' steps.

→ Q orthogonal directions are l.i. if Q is SPD. ↳ (Proved in last class)

$$\underline{e}^{(1)} = \sum_{j=1}^{n} \delta_j \, \underline{d}^{(j)}$$

× by $(\underline{d}^{(k)})^T Q$. →

$$(\underline{d}^{(k)})^T Q \, \underline{e}^{(1)} = \sum_{j=1}^{n} \delta_j (\underline{d}^{(k)})^T Q \, \underline{d}^{(j)}$$

$$= \delta_k (\underline{d}^{(k)})^T Q \, \underline{d}^{(k)}$$

$$\delta_k = \frac{(\underline{d}^{(k)})^T Q \, \underline{e}^{(1)}}{(\underline{d}^{(k)})^T Q \, \underline{d}^{(k)}} \checkmark$$

$$= \frac{(\underline{d}^{(k)})^T Q \overbrace{\Big[\underline{e}^{(1)} + \underbrace{\sum_{j=1}^{k-1} \alpha_j \underline{d}^{(j)}}_{\text{"}0\text{"}} \Big]}^{\underline{e}^{(k)}}}{(\underline{d}^{(k)})^T Q \, \underline{d}^{(k)}}$$

Since $(\underline{d}^{(k)})^T Q \, \underline{d}^{(j)} = 0$.

$x^{(k)}$ $x^{(*)}$ $x^{(1)}$ $x^{(*)}$ + ...

$$= \frac{\cdots}{(d^{(k)})^T Q d^{(k)}}$$ Since $(d^{(k)})^T Q d^{(j)} = 0$.

$$x^{(k)} - x^{(*)} = x^{(1)} - x^{(*)} + \cdots$$

$$e^{(k)} = e^{(1)} + \sum_{j=1}^{k-1} \alpha_j d^{(j)}$$

$$x^{(k)} = x^{(k-1)} + \alpha_{k-1} d^{(k-1)}$$

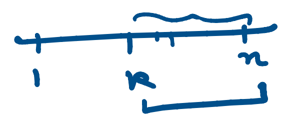
$$r^{(k)} = Q e^{(k)}$$

$$= \frac{(d^{(k)})^T Q e^{(k)}}{(d^{(k)})^T Q d^{(k)}} = \frac{(d^{(k)})^T r^{(k)}}{(d^{(k)})^T Q d^{(k)}}$$

$$\delta_k = -\alpha_k.$$

$$e^{(k)} = e^{(1)} + \sum_{j=1}^{k-1} \alpha_j d^{(j)}$$

$$= \sum_{j=1}^{n} \delta_j d^{(j)} - \sum_{j=1}^{k-1} \delta_j d^{(j)}$$



$$= \sum_{j=k}^{n} \delta_j d^{(j)}$$

$$\vdots$$

$$e^{(n+1)} = 0.$$

IV. How to construct Q orthogonal search directions? [Gram-Schmidt process].

$\{d^{(1)}, \ldots, d^{(n)}\} \to$ l.i.

Recall Gram-Schmidt: $\{v_1, \ldots, v_n\}$ l.i.

$\{e_1, \ldots, e_n\}$ ons

$u_1 = v_1$, $\quad e_1 = \frac{u_1}{\|u_1\|}$

$u_2 = v_2 - \langle v_2, u_1 \rangle u_1 \quad e_2 = \frac{u_2}{\|u_2\|}$

not needed

$$\underline{u}_2 = \underline{v}_2 - \frac{\langle v_2, \underline{u}_1 \rangle}{\langle \underline{u}_1, \underline{u}_1 \rangle} \underline{u}_1 \qquad e_2 = \frac{u_2}{\|u_2\|}$$

$$\vdots$$

$$\underline{u}_n = \underline{v}_n - \frac{\langle \underline{v}_n, u_1 \rangle}{\langle \underline{u}_1, u_1 \rangle} \underline{u}_1 - \cdots \frac{\langle \underline{v}_n, \underline{v}_{n-1} \rangle}{\langle \underline{u}_{n-1}, \underline{u}_{n-1} \rangle} \underline{u}_{n-1}; \qquad e_n = \frac{u_n}{\|u_n\|}$$

Modify Gram-Schmidt for Q-orthogonality

Let $\{\underline{u}^{(i)}\}_{i=1}^{n}$ be n- l.i. directions.

Aim: To construct $\{\underline{d}^{(i)}\}_{i=1}^{n}$ that are 'Q' orthogonal.

in such a way that it is

$\Big[$ $d^{(k)}$ is constructed Q orthogonal to $(\underline{d}^{j})_{j=1}^{k-1}$

$$\underline{d}^{(1)} = u^{(1)}$$

$$\underline{d}^{(k)} = \underline{u}^{(k)} + \sum_{j=1}^{k-1} \underline{\underline{\beta_{kj}}}\, \underline{d}^{(j)}$$

β_{kj}'s are to be chosen such that $\{\underline{d}^{(j)}\}_{j=1}^{k-1}$ are Q-orthogonal to $\underline{d}^{(k)}$.

[Try with a single value of j; $j=1$]

For $\boxed{j = 1, \ldots, k-1}$;

$$(Q\,\underline{d}^{(j)})^T \underline{d}^{(k)} = (\underline{d}^{(j)})^T Q \left(\underline{u}^{(k)} + \sum_{j=1}^{k-1} \beta_{kj}\, \underline{d}^{(j)} \right)$$

$$0 = (\underline{d}^{(j)})^T Q\, \underline{u}^{(k)} + \beta_{kj} (\underline{d}^{(j)})^T Q\, \underline{d}^{(j)}$$

$$\Rightarrow \boxed{\beta_{kj} = - \frac{(\underline{u}^{(k)})^T Q\, \underline{d}^{(j)}}{(\underline{d}^{j})^T Q\, \underline{d}^{(j)}}}$$

$$\boxed{\underline{d}^{(k)} = \underline{u}^{(k)} + \sum_{j=1}^{k-1} \beta_{kj}\, \underline{d}^{(j)}}$$

Drawbacks.

- Search directions need to be stored
- Linear combinations.

V. [Choice of $\{u^{(i)}\}_{i=1}^{n}$ (l.i) such that the drawbacks above are sorted out.

Choose $\underline{u}^{(i)} = \underline{r}^{(i)}$ $\rightarrow$ • Can we do this?

Qus: { ~~Are~~ $\{\underline{r}^{(k)}\}_{k=1}^{n}$ l.i.?

Do they simplify computation of δ_k $\underline{d}^{(k)}$?

Properties of residuals:

$$\boxed{\underline{r}^{(k)} = Q e^{(k)}}$$

(P1) $\underline{r}^{(k)}$ is orthogonal to $\boxed{\{\underline{d}^{(i)}\}_{i=1}^{k-1}}$

Pf: $\underline{e}^{(k)} = \sum_{j=k}^{n} \delta_j\, \underline{d}^{(j)}$.

$i = 1, \dots, k-1$ $\quad (d^{(i)})^T Q\, \underline{e}^{(k)} = \sum_{j=k}^{n} \delta_j\, (d^{(i)})^T Q\, d^{(j)}$

$$(d^{(i)})^T \underline{r}^{(k)} = 0 \qquad i = 1, \dots, k-1.$$

(P2) Claim $\underline{r}^{(k+1)} = \underline{r}^{(k)} + \alpha_k\, Q\, \underline{d}^{(k)}$.

$$\boxed{x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}}$$

Pf. $\underline{r}^{(k+1)} = Q\, \underline{e}^{(k+1)}$

Pf. $\quad \underline{r}^{(k+1)} = Q\, \underline{e}$

$$= Q\left(\underline{e}^{(k)} + \alpha_k\, \underline{d}^{(k)}\right)$$

$$\boxed{r^{(k+1)} = \underline{r}^{(k)} + \alpha_k\, Q\, \underline{d}^{(k)}}$$

(P3) $\quad \beta_{kj} = -\dfrac{(r^{(k)})^T Q\, d^{(j)}}{(\underline{d}^{(j)})^T Q\, d^{(j)}}$ [from derivation of β_{kj}'s with $\underline{u}^{(k)} = r^{(k)}$].

(P4) $\boxed{\{r^{(j)}\}_{j=1}^{n} \text{ are orthogonal}}$: $\rightarrow (\underline{r}^{(k)})^T \underline{r}^{(j)} = 0$, $k \neq j$.

HW

Hint: use (P1)

$$\underline{d}^{(k)} = \underline{u}^{(k)} + \sum_{j=1}^{k-1} \beta_{kj}\, d^{(j)}. \quad \| \quad \times (r^{(j)})^T$$

(P5) Most β_{kj}'s are zeroes when $\underline{u}^{(j)} = \underline{r}^{(j)}$.

$$\underline{d}^{(k)} = \underline{r}^{(k)} + \sum_{j=1}^{k-1} \beta_{k,j}\, d^{(j)}$$

$$= \underline{r}^{(k)} + \beta_{k,1}\, d^{(1)} + \dots + \beta_{k,k-1}\, \underline{d}^{(k-1)}$$

$$\boxed{\beta_{k,j} = \frac{-(r^{(k)})^T Q\, d^{(j)}}{(d^{(j)})^T Q\, d^{(j)}}}$$

↳ (P3)

$$\underline{r}^{(j+1)} = \underline{r}^{(j)} + \alpha_j\, Q\, \underline{d}^{(j)} \qquad \boxed{\text{from (P2)}}$$

$\times$ by $(r^{(k)})^T$

$$(r^{(k)})^T r^{(j+1)} = (r^{(k)})^T r^{(j)} + \alpha_j \cdot (r^{(k)})^T Q\, d^{(j)}$$

$(r^{(k)})^T r^{(j)} \quad (r^{(k)})^T \quad (j+1) \quad (r^{(k)})^T r^{(j)}$

$$\alpha_j (r^{(k)})^T Q d^{(j)} = (r^{(k)})^T r^{(j+1)} - (r^{(k)})^T r^{(j)}$$

$$\boxed{(r^{(k)})^T Q d^{(j)} = \begin{cases} -\frac{1}{\alpha_k} (r^{(k)})^T r^{(k)} & j = k \\ \frac{1}{\alpha_{k-1}} (r^{(k)})^T r^{(k)} & j+1 = k \\ 0 & \text{otherwise} \end{cases}}$$

$$\beta_{kj} = -\frac{(r^{(k)})^T Q d^{(j)}}{(d^{(j)})^T Q d^{(j)}}$$

$$\beta_k = \beta_{k,k-1} = -\frac{1}{\alpha_{k-1}} \frac{(r^{(k)})^T r^{(k)}}{\underbrace{(d^{(k-1)})^T Q d^{(k-1)}}}$$

$$\boxed{\beta_{k1}, \beta_{k2} \cdots, \beta_{k,k-2} = 0}$$

$$\alpha_{k-1} = -\frac{(d^{(k-1)})^T r^{(k-1)}}{\underbrace{(d^{(k-1)})^T Q d^{(k-1)}}}$$

$$\boxed{\beta_k = \frac{(r^{(k)})^T r^{(k)}}{(d^{(k-1)})^T r^{(k-1)}}}$$

HW $(d^{(k-1)})^T r^{(k-1)} = (r^{(k-1)})^T r^{(k-1)}$.

Hint. $d^{(k-1)} = r^{(k-1)} + \sum_{j=1}^{k-2} \beta_{k-1,j} d^{(j)}$

$$\boxed{\beta_k = \frac{(r^{(k)})^T r^{(k)}}{(r^{(k-1)})^T r^{(k-1)}}}$$

$$\boxed{\beta_k = \frac{(\underline{r}^{(k)})^T \, \underline{r}^{(k)}}{(\underline{r}^{(k-1)})^T \, \underline{r}^{(k-1)}}}$$

eg.

Lecture 17.

Monday, 7 March 2022 1:47 PM

$$f(x) := \frac{1}{2} x^T Q x - b^T x + c, \quad \boxed{\begin{array}{l} Q = Q^T \\ Q > 0 \end{array}}$$

CG Algorithm

1. $d^{(0)} = r^{(0)} \left(= g^{(0)}\right) = Q x^{(0)} - b$

$k = 0$

2. $\alpha_k = - \dfrac{(g^{(k)})^T r^{(k)}}{(d^{(k)})^T Q d^{(k)}}$

3. $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$

4. $r^{(k+1)} = r^{(k)} + \alpha_k Q d^{(k)}$

5. $\beta_{k+1} = \dfrac{(r^{(k+1)})^T r^{(k+1)}}{(r^{(k)})^T r^{(k)}}$

6. $d^{(k+1)} = r^{(k+1)} + \beta_{k+1} d^{(k)}$

→ CG for [non linear / non-quadratic] $f(x)$

Second order Taylor Series approximation of objective function.

Quasi-Newton Methods

Newton's method → $x^{(k+1)} = x^{(k)} - \alpha_k F(x^{(k)})^{-1} g^{(k)}$

α_k is s.t. $f(x^{(k+1)}) < f(x^{(k)})$

$\boxed{\text{Thm } 9.2}$ $\alpha_k = \arg\min_{\alpha \geq 0} f\left(x^{(k)} - \alpha F(x^{(k)})^{-1} g^{(k)}\right)$

→ Use an approximate value of $(F(x^k))^{-1}$ in place of

↳ Use an approximate value of $(F(x^{-}))$ in place of the true inverse.

That is, $\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H_k \underline{g}^{(k)}$.

$H_k \to n \times n$ real matrix, α_k is the search parameter.

$$f(\underline{x}^{k+1}) = f(\underline{x}^{(k)}) + (\underline{g}^{(k)})^T (\underline{x}^{(k+1)} - \underline{x}^{(k)}) + o(\| \underline{x}^{(k+1)} - x^{(k)} \|)$$

$$= f(\underline{x}^{(k)}) - \alpha_k (\underline{g}^{(k)})^T H_k \underline{g}^{(k)} + o(\alpha_k \| H_k \underline{g}^{(k)} \|).$$

As $\alpha_k \to 0$, the second term dominates the third.

$f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)})$ is guaranteed if

$$(\underline{g}^{(k)})^T H_k \underline{g}^{(k)} > 0$$

H_k is positive-definite guarantees this.

Summarize: $f \in C^{(1)}$

$\underline{x}^{(k)} \in \mathbb{R}^{(n)}$

$g^{(k)} = \nabla f(\underline{x}^{(k)}) \neq 0$ [Why?]

H_k is $n \times n$ real-symmetric SPD.

$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H_k \underline{g}^{(k)}$

$\alpha_k = \arg\min_{\alpha \geq 0} f(\underline{x}^{(k)} - \alpha H_k \underline{g}^{(k)}) > 0$

$f(\underline{x}^{(k+1)}) < f(\underline{x}^{(k)})$.

Hessian approximation using [objective fn

Hessian approximation using $\begin{bmatrix}\text{objective fn} \\ \text{gradient.}\end{bmatrix}$

Let us understand for quadratic function.

$$f(x) = \frac{1}{2} \underline{x}^T Q \underline{x} - \underline{x}^T \underline{b} + \underline{c}, \quad Q = Q^T > 0$$

Let $H_0, H_1, H_2, \ldots$ be successive approximation of $F(\underline{x}^{(k)})^{-1}$.

$$\underline{g}^{(k+1)} = \left\{\begin{aligned} g^{(k+1)} &= Q x^{(k+1)} - \underline{b} \\ g^{(k)} &= Q x^{(k)} - \underline{b} \end{aligned}\right\}$$

$$\Delta g^{(k)} = \underline{g}^{(k+1)} - \underline{g}^{(k)}$$

$$\Delta x^{(k)} = \underline{x}^{(k+1)} - \underline{x}^{(k)}.$$

$$\boxed{\Delta \underline{g}^{(k)} = Q \, \Delta \underline{x}^{(k)}}$$

$$Q^{-1} \Delta g^{(i)} = \Delta \underline{x}^{(i)} \qquad 0 \le i \le k$$

Impose the requirement $\rightarrow$ $\boxed{H_{k+1} \, \Delta g^{(i)} = \Delta x^{(i)}}$ $\rightarrow$ $\quad 0 \le i \le k.$

$\boxed{k=1}$ $\qquad H_2 \, \Delta g^{(i)} = \Delta x^{(i)} \qquad i = 0, 1$

$$[H_2]_{n\times n} \qquad \left.\begin{aligned} H_2 \, \Delta g^{(0)} &= \Delta x^{(0)} \\ H_2 \, \Delta g^{(1)} &= \Delta x^{(1)}. \end{aligned}\right]$$

$$[H_2]_{n\times n} \left[\Delta g^{(0)} \;\; \Delta g^{(1)}\right]_{n\times 2} = \left[\Delta x^{(0)} \;\; \Delta x^{(1)}\right]_{n\times 2}$$

$$H_{(n)} \left[\Delta g^{0} \; \Delta g^{(1)} \; \ldots \; \Delta g^{(n-1)}\right] = \left[\Delta x^{(0)}, \; \Delta x^{(1)}, \; \ldots \; \Delta x^{(n-1)}\right].$$

Q satisfies

$$Q^{-1} \underbrace{\left[\Delta g^{(0)} \; \ldots \; \Delta g^{(n-1)}\right]} = \left[\Delta x^{(0)}, \; \ldots \; \Delta x^{(n-1)}\right]$$

If $\left[\Delta g^{(0)} \; \ldots \; \Delta g^{(n-1)}\right]$ is non-singular,

$$Q^{-1} = H_n = \left[\Delta x^{(0)} \; \ldots \; \Delta x^{(n-1)}\right]\left[\Delta g^{(0)} \ldots \Delta g^{(n-1)}\right]^{-1}. \checkmark$$

[if $[\Delta g^{0} \; \ldots \; \Delta g^{(n-1)}]$ is

$$Q^{-1} = H_n = [\Delta x^{(0)} \dots \Delta x^{(n-1)}][\Delta g \dots \Delta g]^{-1}$$

[if $[\Delta g^{0}, \dots \Delta g^{(n-1)}]$ is invertible.]

At 'n' th iteration, H_n satisfies

$$\left[\begin{array}{l} H_n \Delta g^{(i)} = \Delta x^{(i)} \qquad 0 \le i \le n-1. \\ \underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H^{(k)} g^{(k)}. \end{array}\right.$$

$$\left[\begin{array}{l} \Delta x^{(i)} = \underline{x}^{(i+1)} - x^{(i)} \\ \Delta g^{(i)} = g^{(i+1)} - g^{(i)} = Q \Delta x. \end{array}\right.$$

Result: Consider a quasi-Newton method applied to a quadratic function with Hessian $Q = Q^T$ s.t.

$0 \le k < n-1$,

$$\boxed{H_{k+1} \Delta g^{(i)} = \Delta x^{(i)}} \quad \text{— (1)} \quad 0 \le i \le k,$$

where $H_{k+1} = H_{k+1}^T$.

If $\alpha_i \neq 0$, $0 \le i \le k$, then $\left(\underline{d}^{(k)} = -H_k g^{(k)}\right)$

$\underline{d}^{(0)}, \dots, \underline{d}^{(k+1)}$ are Q-conjugate (orthogonal).

[Conclude from III of Lecture 16 that convergence happens in finite steps.]

Pf. Induction $k = 0$.

$d^{(0)}$ and $d^{(1)}$ are Q orthogonal. $[d^{(k)} = -H_k g^{(k)}]$

$\underline{d}^{(0)}$ and $\underline{d}^{(1)}$ are Q orthogonal.

$$\boxed{\underline{d}^{0} = \frac{\Delta x^{(0)}}{\alpha_0}}$$

$$\underline{d}^{(k)} = -H_k \underline{g}^{(k)}$$
$$\Delta x^{(k)} = \underline{x}^{(k+1)} - \underline{x}^{(k)}$$
$$\Delta \underline{g}^{(k)} = \underline{g}^{(k+1)} - \underline{g}^{(k)}$$
$$\Delta \underline{g}^{(k)} = Q \Delta x^{(k)}$$

$$\underline{x}^{(1)} = x^{(0)} - \alpha_0 H_0 \underline{g}^{(0)}$$
$$\underline{x}^{(1)} = \underline{x}^{(0)} + \alpha_0 \underline{d}^{(0)}$$
$$\Rightarrow \Delta \underline{x}^{(0)} = \alpha_0 \underline{d}^{(0)}$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k H_k \underline{g}^{(k)}$$
$$\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha_k \underline{d}^{(k)}$$
$$\Delta x^{(k)} = \alpha_k d^{(k)}$$

$$(\underline{d}^{(1)})^T Q \underline{d}^{(0)} = -(H_1 \underline{g}^{(1)})^T Q \underline{d}^{(0)}$$
$$= -(\underline{g}^{(1)})^T H_1 Q \frac{\Delta x^{(0)}}{\alpha_0}$$
$$= -(\underline{g}^{(1)})^T H_1 \frac{\Delta \underline{g}^{(0)}}{\alpha_0}$$
$$= -(\underline{g}^{(1)})^T \frac{\Delta \underline{x}^{(0)}}{\alpha_0}$$
$$= -(\underline{g}^{(1)})^T \underline{d}^{(0)}$$
$$= 0$$

$$H_1 \Delta \underline{g}^{(0)} = \Delta x^{(0)}$$
from ①

$(\underline{g}^{(1)} d^{(0)}$

Ex. 11.1 Tut Sheet 6 [from CG method proof.]

Assume for $\underline{k-1}$ $(k < n-1)$.

We prove that $d^{(0)}, d^{(1)}, \ldots, d^{(k+1)}$ are Q-conjugate.

$$(d^{(k+1)})^T Q \underline{d}^{(i)} = 0 \qquad i = 0, \ldots, k.$$

$$(\underline{d}^{(k+1)})^T Q d^{(i)} = -(H_{k+1} \underline{g}^{(k+1)})^T Q \underline{d}^{(i)}$$
$$= -(\underline{g}^{(k+1)})^T H_{k+1} Q \frac{\Delta x^{(i)}}{\alpha_i}$$
$$= -(\underline{g}^{(k+1)})^T H_{k+1} \frac{\Delta \underline{g}^{(i)}}{\alpha_i} \quad (\text{use ①})$$

$$= -(g \quad \cdots \quad \frac{}{\alpha_i}$$

$$= -\left(g^{(k+1)}\right)^T \frac{\Delta \underline{x}^{(i)}}{\alpha_i}$$

$f(x^{(k+1)}) < f(x^{(k)})$

[residuals are orthogonal to $d^{(i)}$]. $\Leftarrow$ $= -\left(g^{(k+1)}\right)^T \underline{d}^{(i)} = 0$

$i = 0, \ldots, k$

$$= 0$$

$\boxed{\text{Qn:}}$ How to update H_k's ?

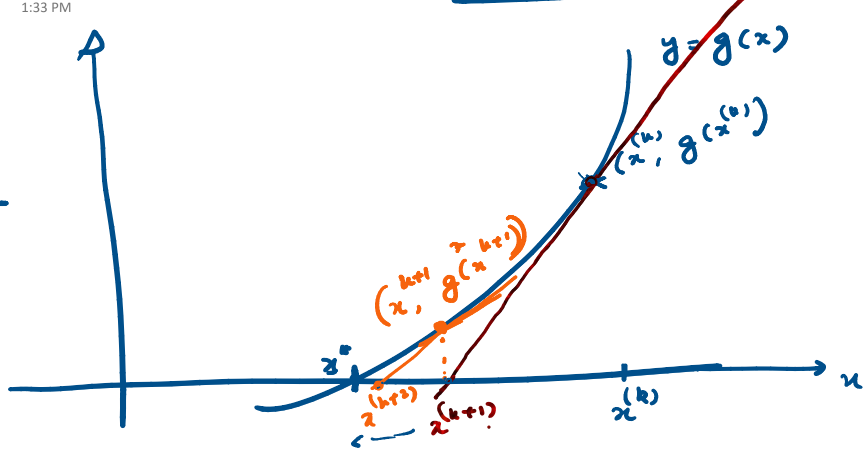
$$\boxed{H_{k+1} \Delta g^{(i)} = \Delta \underline{x}^{(i)}}$$

Lecture 18

Thursday, 10 March 2022 1:33 PM

(1D case)

Newton's



min $f(x)$

$f'(x^*) = 0$

$\Downarrow$

$g(x^*) = 0$

$$y - g(x^{(k)}) = g'(x^{(k)})\,(x - x^{(k)})$$

$x = x^{(k+1)}$

$$0 = y = g(x^k) + g'(x^{(k)})\,(x - x^{(k)})$$

$$\boxed{x^{(k+1)} = x^{(k)} - \left(g'(x^{(k)})\right)^{-1} g(x^{(k)})}$$

$$x^{(k+1)} = x^{(k)} - \left(f''(x^{(k)})\right)^{-1} f'(x^{(k)})$$

multivariable

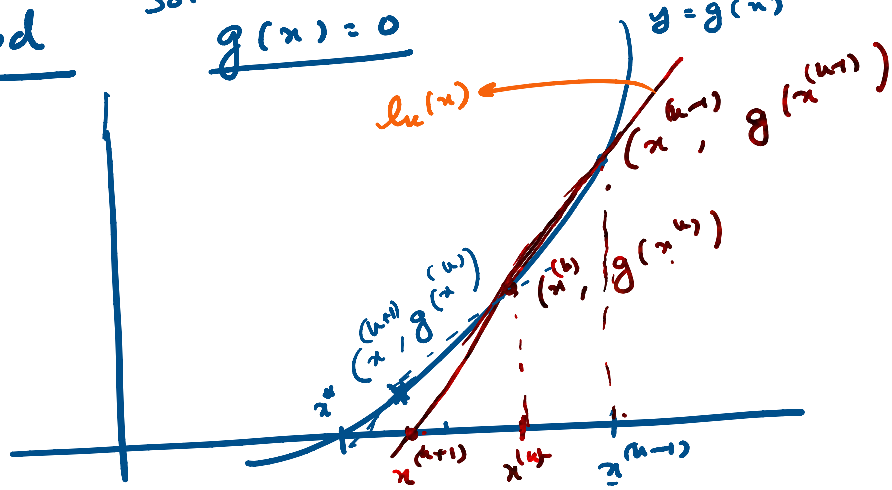
$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \left[\nabla g(\underline{x}^{(k)})\right]^{-1} g(\underline{x}^{(k)})$$

$$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \left[D^2 f(\underline{x}^{(k)})\right]^{-1} \nabla f(x^{(k)})$$

Secant' method

Solve.

$g(x) = 0$



$$x^{(k+1)} = x^{(k)} - \left[\frac{x^{(k)} - x^{(k-1)}}{g(x^{(k)}) - g(x^{(k-1)})}\right] g(x^{(k)}).$$

$l_k(x)$ passes

$\left[g(x^{(k)}) - g(x \quad) \right]$

$$\ell_k(x) = g(x^{(k)}) + B_k (x - x^k)$$

$\ell_k(x)$ passes thro' $(x^{(k+1)}, g(x^{(k+1)}))$ and $(x^{(k)}, g(x^{(k)}))$

$$\rightarrow \ell_{k+1}(x) = g(x^{(k+1)}) + B_{k+1}(x - x^{(k+1)})$$

$$\ell_{k+1}(x^{(k)}) = g(x^{(k+1)}) + B_{k+1}(x^{(k)} - x^{(k+1)})$$

$$g(x^{(k)}) - g(x^{(k+1)}) = B_{k+1}(x^{(k)} - x^{(k+1)})$$

$$\Delta g(x^k) = B_{k+1} \Delta x^{(k)}.$$

$\ell_{k+1}(x)$ passes thro' $(x^{(k+1)}, g(x^{(k+1)}))$ and $(x^{(k)}, g(x^{(k)}))$

Undetermined system $\leftarrow$ $\boxed{(B_{k+1})_{n\times n} \Delta \underline{x}^{(k)} = \Delta g(\underline{x}^{(k)})}$ Secant condition

Lecture 17 $\rightarrow$ $(B_{k+1} = H^{-1}_{k+1})$.

$$\nabla f(\underline{x}^{(k+1)}) - \nabla f(\underline{x}^{(k)}) = \boxed{B_{k+1}} (\underline{x}^{(k+1)} - \underline{x}^{(k)}).$$

$\Delta x^{(k)}$

$$[\;]_{n\times 1} = [\;]_{n\times n} [\;]_{n\times 1}$$

$n \times n$

$1 + 2 + \dots + n = \frac{n(n+1)}{2}$.

We want

$$\boxed{\begin{aligned} \min \quad & \| B_{k+1} - B_k \| \\ \text{s.t.} \quad & B_{k+1}^T = B_{k+1} \\ & B_{k+1} \Delta x^{(k)} = \Delta g^{(k)}. \end{aligned}}$$

$$\boxed{\begin{aligned} \min \quad & \| B_{k+1}^{-1} - B_k^{-1} \| \\ \text{s.t.} \quad & (B_{k+1}^{-1})^T = B_{k+1}^{-1} \\ & \Delta x^{(k)} = B_{k+1}^{-1} \Delta g^{(k)} \end{aligned}}$$

$\rightarrow$ (Motivation for Lagrange Method)

Quasi-Newton

- $\underline{d}^{(k)} = - H_k \, g^{(k)}$ (Hessian inverse in Newton)
- $\alpha^k = \arg\min_{\alpha \geq 0} f(\underline{x}^{(k)} + \alpha \underline{d}^{(k)})$
- $\underline{x}^{(k+1)} = \underline{x}^{(k)} + \alpha^{(k)} \underline{d}^{(k)}$

(Many options)

constants

$\dots u u^T + \beta v v^T$

(Many options)

Construct updates

(BFGS Method)

Broyden–Fletcher–Goldfarb–Shanno (1970)

Rank 1

$$B_{k+1} = B_k + \alpha\, \underline{u}\,\underline{u}^T + \beta\, \underline{v}\,\underline{v}^T$$

constants

(Rank-two update)

$$\underline{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$\underline{u}, \underline{v}$ l.i.

$$\underline{u}\,\underline{u}^T = \begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_n \\ u_nu_1 & u_nu_2 & \cdots & u_n^2 \end{bmatrix}$$

$$\boxed{B_{k+1}\, \Delta \underline{x}^{(k)} = \Delta g^{(k)}.}$$

$$B_{k+1}^T = B_{k+1}$$

$$B_{k+1}\, \Delta \underline{x}^{(k)} = B_k\, \Delta \underline{x}^{(k)} + \alpha\, \underline{u}\,\underline{u}^T\, \Delta \underline{x}^{(k)} + \beta\, \underline{v}\,\underline{v}^T \Delta x^k = \Delta g^{(k)}$$

Choose $\boxed{\begin{aligned} \underline{u} &= \Delta g^{(k)} \\ \underline{v} &= B_k\, \Delta \underline{x}^{(k)} \end{aligned}}$

[Qn: Are they l.i? Exercise. (Prove))

$$B_k\, \Delta \underline{x}^{(k)} + \alpha\, \underbrace{\Delta g^{(k)} (\Delta g^{(k)})^T \Delta \underline{x}^{(k)}} + \beta\, B_k\, \Delta \underline{x}^{(k)} (\Delta \underline{x}^{(k)})^T B_k\, \Delta x^{(k)} = \Delta g^{(k)}$$

$$\Rightarrow \underbrace{\Delta g^{(k)}}_{u} \Big[1 - \underbrace{\alpha\, (\Delta g^{(k)})^T \Delta x^{(k)}}_{\text{Constant}} \Big] = \underbrace{B_k\, \Delta \underline{x}^{(k)}}_{v} \Big[1 + \overbrace{\beta\, (\Delta x^{(k)})^T B_k\, \Delta x^{(k)}}^{\text{Constant}} \Big]$$

$$\Rightarrow \boxed{\alpha = \frac{1}{(\Delta g^{(k)})^T \Delta x^{(k)}} \qquad \beta = -\frac{1}{(\Delta x^{(k)})^T B_k\, (\Delta \underline{x}^{(k)})}}$$

$$\boxed{B_{k+1} = B_k + \frac{(\Delta g^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T \Delta x^{(k)}} - \frac{B_k\, \Delta x^{(k)} (\Delta x^{(k)})^T B_k}{(\Delta x^{(k)})^T B_k\, (\Delta x^{(k)})}}$$

Rank-two update

Lecture 19

Saturday, 12 March 2022 9:35 AM

Recall

$$\left[B_{k+1} = B_k - \underbrace{\frac{B_k (\Delta x^{(k)})(\Delta x^{(k)})^T B_k}{(\Delta x^{(k)})^T B_k (\Delta x^{(k)})} + \frac{(\Delta g^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T \Delta x^{(k)}}} \right.$$

$$H_{k+1} = B_{k+1}^{-1}.$$

[Verify $B B^{-1} = I$]

Woodbury formula

$$(\underbrace{A + UCV}_{B})^{-1} = \underbrace{A^{-1} - A^{-1} U (C^{-1} + VA^{-1}U)^{-1} VA^{-1}}_{B^{-1}}$$

Remove 'k' for notational convenience

$$B_+ = B - \frac{B \overset{a}{(\Delta x)} \overset{a^T}{(\Delta x)^T} B}{\underbrace{(\Delta x)^T}_{a} B \overset{}{\underset{a}{(\Delta x)}}} + \frac{\overset{b}{(\Delta g)} \overset{b^T}{(\Delta g)^T}}{\underbrace{(\Delta g)^T}_{b^T} \underset{a}{(\Delta x)}}$$

$$\underbrace{\begin{bmatrix} B\Delta x & \Delta g \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \frac{-1}{(\Delta x)^T B (\Delta x)} & 0 \\ 0 & \frac{1}{(\Delta g)^T (\Delta x)} \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} (\Delta x)^T B \\ (\Delta g)^T \end{bmatrix}}_{V}$$

$$\left\{ \begin{array}{l} A = B \\ U = [B\Delta x \quad \Delta g] \\ C = \begin{bmatrix} -\frac{1}{(\Delta x)^T B (\Delta x)} & 0 \\ 0 & \frac{1}{(\Delta g)^T \Delta x} \end{bmatrix} \\ V = \begin{bmatrix} (\Delta x)^T B \\ (\Delta g)^T \end{bmatrix} \end{array} \right.$$

$$C = \begin{bmatrix} d & 0 \\ 0 & e \end{bmatrix} \qquad C^{-1} = \frac{1}{de}\begin{bmatrix} e & 0 \\ 0 & d \end{bmatrix} = \begin{bmatrix} d^{-1} & 0 \\ 0 & e^{-1} \end{bmatrix}$$

$$\hookrightarrow C^{-1} = \begin{bmatrix} -(\Delta x)^T B (\Delta x) & 0 \\ 0 & (\Delta g)^T (\Delta x) \end{bmatrix}$$

Plug in Woodbury formula:

$$(B_+) \quad B_{k+1} = \Big\{ \underbrace{\left(I - \frac{(\Delta x^{(k)})(\Delta g^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right) B_k^{-1} \left(I - \frac{(\Delta g^{(k)})(\Delta x^{(k)})^T}{(\Delta g^{(k)})^T \Delta x^{(k)}} \right)}_{P}$$

$$+ \underbrace{\frac{\overset{a}{(\Delta x^{(k)})} \overset{a^T}{(\Delta x^{(k)})^T}}{(\Delta g^{(k)})^T (\Delta x^{(k)})}}_{Q} \Big\}$$

$\boxed{\text{Tut 7. (1)}}$.

$\boxed{\text{Tut 7 (2)}}$ - Rank two-update preserve positive-definiteness.

$\boxed{\text{Tut } 7\ (2)}$ → Rank two-update preserve positive-definiteness.

Hint: $\underline{z}^T B_{k+1} \underline{z} = \underline{z}^T P \underline{z} + z^T Q \underline{z}$

$$\left(\underline{z}^T - \frac{z^T (\Delta x^{(k)}) (\Delta g^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right) B_k^{-1} \left(\underline{z} - \frac{z\, \Delta g^{(k)} (\Delta x^{(k)})^T}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right)$$

$$= \underbrace{\left(\underline{z} - \frac{\Delta g^{(k)}\ \Delta x^{(k)}\ z}{(\Delta g^{(k)})^T (\Delta x^{(k)})} \right)^T}_{\underline{r}} B_k^{-1} \underbrace{\left(\quad \right)}_{\underline{r}}$$

$\underline{r}^T B_k^{-1} \underline{r} > 0$ [Induction]

Use B_k is positive-definite.

$$z^T \frac{\underline{a}\, a^T}{\underline{b}\, \underline{a}^T} z = \frac{(a^T \underline{z})^2}{\underline{b}\, \underline{a}} > 0$$

→ $\boxed{(\Delta g^{(k)})^T (\Delta x^{(k)})} \geq 0$?

Hint $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$
$= \alpha\, \underline{d}^{(k)}$

$$\boxed{\min_{B_{k+1}} \| B_{k+1}^{-1} - B_k^{-1} \| \quad \text{s.t. symmetry}, \quad B_{k+1}^{-1} \Delta g^{(i)} = \Delta x^{(i)}}$$ ✓

Tut. Sheet 7 (3) $\left[H_{k+1}\ \Delta g^{(i)} = \Delta \underline{x}^{(i)} \quad 0 \leq i \leq k. \right.$

↳ $\boxed{\text{Induction.}}$

$\boxed{\text{BFGS}}$ → Quasi-Newton method

→ positive-definiteness
→ $H_{k+1}\ \Delta g^{(i)} = \Delta x^{(i)}$, $0 \leq i \leq k$.
→ [Conjugate direction property.

Lagrange / KKT Conditions

Optimization problems with Constraint ✓

Equality constraint (Lagrange)
↓

(Karush-Kuhn-Tucker) (KKT) Inequality Constraint ✓

$\min\ f(x)$

$$\min \quad f(\underline{x})$$
$$\text{s.t.} \quad \underline{h}(\underline{x}) = 0$$

$\underline{x} \in \mathbb{R}^n, \quad f: \mathbb{R}^n \to \mathbb{R}$

$\underline{h}: \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad (m \le n)$

$$\underline{h} = \begin{bmatrix} h_1(\underline{x}) \\ h_2(\underline{x}) \\ \vdots \\ h_m(\underline{x}) \end{bmatrix} \qquad \boxed{h \in C^{(1)}}$$

$$\min \quad f(x)$$
$$\text{s.t.} \quad \underline{h}(x) = 0$$
$$\underline{g}(x) \le 0$$

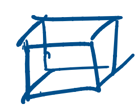
$f: \mathbb{R}^n \to \mathbb{R}$

$\underline{h}: \mathbb{R}^n \to \mathbb{R}^m \qquad (m \le n)$

$\underline{g}: \mathbb{R}^n \to \mathbb{R}^p, \quad \boxed{h, g \in C^{(1)}}$

Revise: Surface, Curve on S, Tangent space, Normal space.
[Section 20.3].

Examples ① Given a fixed area of a cardboard, construct a closed cardboard box with max. volume.



$\begin{bmatrix} m=1 \\ n=3 \end{bmatrix}$

$$\left.\begin{array}{ll} \max & x_1 x_2 x_3 \quad \to \boxed{f(x)} = -x_1x_2x_3 \\ \text{s.t.} & x_1x_2 + x_2x_3 + x_1x_3 = A/2 \end{array}\right] \qquad \boxed{x_1, x_2, x_3 > 0}$$

$$\boxed{\begin{array}{l} f: \mathbb{R}^3 \to \mathbb{R}. \\ h: \mathbb{R}^3 \to \mathbb{R}. \end{array}}$$

$\min \quad f(\underline{x}) \quad \text{s.t.} \quad h(\underline{x}) = (x_1x_2 + x_2x_3 + x_1x_3 - A/2) = 0$

$$\boxed{\nabla f(\underline{x}) + \lambda \nabla h(\underline{x}) = 0} \longrightarrow \text{Lagrange condition.}$$

$$\boxed{G = f + \lambda h = 0}$$

$$\left\{\begin{array}{l} -\begin{bmatrix} x_2x_3 \\ x_1x_3 \\ x_1x_2 \end{bmatrix} + \lambda \begin{bmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ \\ x_1x_2 + x_2x_3 + x_1x_3 = A/2. \\ \lambda \in \mathbb{R}. \end{array}\right.$$

$x_1 = 0$

$\lambda \ne 0 \qquad \lambda = 0$

Contradicts Constraints

$\ldots (x_1 + \ldots) \quad x_2 x_3 = A/2.$

$\boxed{x_1, x_2, x_3, \lambda \neq 0}$

$\underline{x_1 = 0}$ | $(\lambda \neq 0)$ $\underline{x_2 x_3 = A/2}$.

$\lambda \neq 0$ / $\lambda = 0$ → Contradicts Constraints

$\lambda x_3 = 0$ || $\underline{x_2 x_3 = 0}$ ↔ Contradicts Constraint $x_2 x_3 = A/2$.

$\lambda x_2 = 0$

$x_1 \neq 0$.

$\boxed{\lambda = 0}$ $x_2 x_3 = 0$ Contradicts the Constraint $x_2 x_3 = A/2$.

$\lambda = 0$ $\Rightarrow \underline{x_1 x_2 + x_2 x_3 + x_1 x_3 = 0}$ → $\underline{\text{Contradicts the Constraint.}}$

$$\left.\begin{array}{l} x_2 x_3 = 0 \\ x_1 x_3 = 0 \\ x_1 x_2 = 0 \end{array}\right]$$

$$\begin{array}{ll} [\, x_2 x_3 - \lambda (x_2 + x_3) = 0 \,] & x_1 \\ [\, x_1 x_3 - \lambda (x_1 + x_3) = 0 \,] & x_2 \\ [\, x_1 x_2 - \lambda (x_1 + x_2) = 0 \,] & x_3 \end{array}$$

$$x_1 x_2 + x_2 x_3 + x_3 x_1 = A/2$$

$x_1 x_2 x_3 - \lambda (x_1 x_2 + x_1 x_3) = 0$

$x_1 x_2 x_3 - \lambda (x_1 x_2 + x_2 x_3) = 0$

$\lambda (x_1 - x_2)\, x_3 = 0$

$\Downarrow$

$\boxed{x_1 = x_2}$

$|||^{ly}$ $x_2 = x_3$

$3 x_1^2 = A/2 \Rightarrow x_1^2 = A/6 \Rightarrow \boxed{x_1 = \sqrt{A/6} = x_2 = x_3}$

[$\underline{x_1, x_2, x_3 > 0 \text{ is ignored}}$; however, this is taken taken care of !]

Recall:

$\underline{\text{Second order "necessary"}}$ vs $\underline{\text{Second order "sufficient."}}$

$\boxed{\text{1 variable}}$ $f'(x^*) = 0$ [first order necessary] || [x^* can be a minimum, maximum, } "critical point"]

x^* is $\underline{\text{minimum}}$ $\underline{f''(x^*) \geq 0}$ | $\underline{f''(x^*) > 0.}$

second-order necessary ||

↓ guarantee's x^* is a

$f(\ldots x^*)$ [doesn't guarantee]

Second-order necessary || guarantee's x^* is a minimum.

Counter-example $f(x) = x^3$, $f''(0) = 0$ — doesn't guarantee that x^* is a minimum

We want to do this in the context of "Lagrangian".

$\boxed{\text{Ex. 2}}$ $\min f(x) := x_1^2 + x_2^2$

s.t. $\left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \underbrace{x_1^2 + 2x_2^2 - 1}_{h(x)} = 0 \right\} \longrightarrow$ (ellipse).

$$\left[\begin{array}{l} \nabla f(x) + \lambda h(x) = 0 \\ x_1^2 + 2x_2^2 - 1 = 0 \end{array} \right]$$

$$\left[\begin{array}{ll} x_1 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \end{pmatrix} & x_2 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \end{pmatrix} \\ x_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} & x_4 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \end{array} \right\}$$

Justification of $\rightarrow \left[\begin{array}{l} \nabla f(x) + \lambda \nabla h(x) = 0 \\ h(x) = 0 \end{array} \right]$

$\rightarrow$ Do we have second-order conditions?

[Example where $f \notin C^1$, $h \notin C^1$ but still we can consider min/max.]

Lecture 20 [Lagrange Method]

Monday, 14 March 2022 1:38 PM

$$\begin{cases} \min \ f(\underline{x}) \\ \text{s.t.} \quad \underline{h}(\underline{x}) = \underline{0} \\ \underline{h}(\underline{x}) = \begin{bmatrix} h_1(\underline{x}) \\ h_2(\underline{x}) \\ \vdots \\ h_m(\underline{x}) \end{bmatrix} \end{cases}$$

$\underline{x} \in \mathbb{R}^n$

$f : \mathbb{R}^n \to \mathbb{R}$

$\underline{h} : \mathbb{R}^n \to \mathbb{R}^m \quad (m \le n)$

$\underline{h}, \underline{f} \in \mathcal{C}^{(1)}$

Defn: [Regular point] A point $\underline{x}^*$ satisfying $h_1(\underline{x}^*) = h_2(\underline{x}^*) = \dots h_m(\underline{x}^*) = 0$ is said to be a regular point of the constraints, if the gradient vectors $\nabla h_1(\underline{x}^*), \nabla h_2(\underline{x}^*), \dots, \nabla h_m(\underline{x}^*)$ are linearly independent.

Jacobian matrix

$$D\underline{h}(\underline{x}^*) = \begin{bmatrix} Dh_1(\underline{x}^*) \\ \vdots \\ Dh_m(\underline{x}^*) \end{bmatrix} = \begin{bmatrix} (\nabla h_1(\underline{x}^*))^T \\ \cdots\cdots \\ (\nabla h_m(\underline{x}^*))^T \end{bmatrix}$$

$$\underline{x}^* \text{ is regular} \iff \operatorname{rank}(D\underline{h}(\underline{x}^*)) = m \quad [m \le n].$$

Ex. $\begin{bmatrix} h_1(\underline{x}) = x_1 \\ h_2(\underline{x}) = x_2^2 - x_3^2 \end{bmatrix}$

$\nabla h_1(\underline{x}) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$\nabla h_2(\underline{x}) = \begin{bmatrix} 0 \\ 2x_2 \\ -2x_3 \end{bmatrix}$

$$D\underline{h}(\underline{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 2x_2 \\ 0 & -2x_3 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2x_2 & -2x_3 \end{bmatrix}$$

$$Dh(\underline{x}) = \begin{bmatrix} 0 & 2x_2 \\ 0 & -2x_3 \end{bmatrix} \qquad \begin{bmatrix} 0 & 2x_2 & -2x_3 \end{bmatrix}$$

$\underline{x}^* = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ is a regular pt because (i) $h_1(x^*) = 0$, $h_2(x^*) = 0$.

(ii) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -2 \end{bmatrix}$ l.i

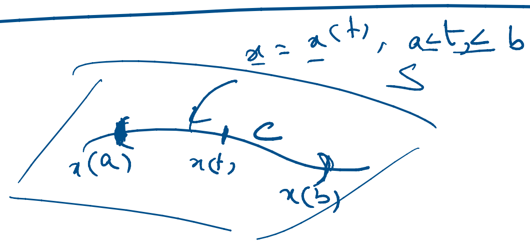
$x^* = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ not a regular point because (ii) is not satisfied.

The set $h_1(\underline{x}) = h_2(\underline{x}) = \dots = h_m(\underline{x}) = 0$ $\quad h_i : \mathbb{R}^n \to \mathbb{R}$

describes the surface

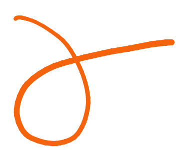
$$S = \{ \underline{x} \in \mathbb{R}^n : \ h_1(\underline{x}) = \dots = h_m(\underline{x}) = 0 \}.$$

$\boxed{\text{Exercise}}$ If the points are regular, then $\dim(S) = n-m$.



Def [Curve]

A curve on S is a set of points $\{ x(t) \in S : t \in (a,b) \}$ continuously parametrized by $t \in (a,b)$; that $x : (a,b) \to S$ is a continuous function.



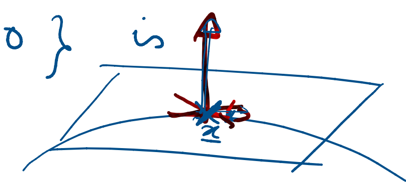
The curve $C = \{ \underline{x}(t) : t \in (a,b) \}$ is differentiable if

$$\dot{\underline{x}}(t) = \frac{d\underline{x}}{dt} = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}$$ exist for all $t \in (a,b)$.

The curve $C = \{ x(t) : t \in (a,b) \}$ is twice differentiable, The curve $C := \{ \underline{x}(t) : t \in (a,b) \}$ is twice differentiable,

if $\ddot{\underline{x}}(t) = \dfrac{d^2\underline{x}}{dt^2} = \begin{bmatrix} \ddot{x}_1(t) \\ \vdots \\ \ddot{x}_n(t) \end{bmatrix}$ exists for all $t \in (a,b)$.

Defn. [Tangent space] at a point $\underline{x}^*$ on the surface $S := \{ \underline{x} \in \mathbb{R}^n : \underline{h}(\underline{x}) = 0 \}$ is



$$\boxed{T(\underline{x}^*) = \{ \underline{y} : \overset{\text{normal}}{D\underline{h}}(\underline{x}^*) \cdot \underline{y} = 0 \}}$$

$$= \mathcal{N}\left(Dh(\underline{x}^*) \right) \subset \mathbb{R}^n.$$

Example: Let $S = \{ \underline{x} \in \mathbb{R}^3, \; h_1(x) = x_1 = 0, \; h_2(\underline{x}) = x_1 - x_2 = 0 \}$

Find the tangent space of S.

$$D\underline{h}(\underline{x}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

$$Dh(\underline{x}) \cdot \underline{y} = 0 \quad \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$R_2 \to R_2 - R_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$0 = y_1 \to$ basic

$0 = y_2 \to$ basic.

$y_3 \to$ free variable

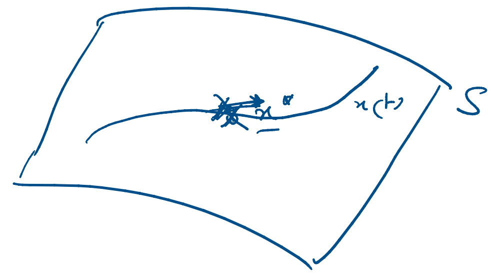
(Gauss-elimination)

$$\underline{y} = \left\{ \begin{bmatrix} 0 \\ 0 \\ \alpha \end{bmatrix} : \alpha \in \mathbb{R} \right\}$$

$\to$ $\boxed{\text{Z-axis}}$

Thm [20.1 in Chong | Zak]
(without proof).

Let $\underline{x}^* \in S$ be a regular point and $T(\underline{x}^*)$ be the tangent space at $\underline{x}^*$. Then $\underline{y} \in T(\underline{x}^*) \iff$ $\exists$ a differentiable curve in $\underline{S}$ passing through $\underline{x}^*$ with derivative $\underline{y}$ at $\underline{x}^*$.



Defn. [Normal space] $N(\underline{x}^*)$ at a point $\underline{x}^*$ on the surface $S = \{\underline{x} \in \mathbb{R}^n : h(\underline{x}) = 0\}$ is the set

$$N(\underline{x}^*) = \{\underline{x} \in \mathbb{R}^n : \underline{x} = \underbrace{Dh(\underline{x}^*)^T}_{n \times m} \underbrace{\underline{z}}_{m \times 1}, \quad \underline{z} \in \mathbb{R}^m\}$$

$$N(\underline{x}^*) = \mathcal{R}\left(D\underline{h}(\underline{x}^*)^T\right)$$

$$= \text{span}\{\nabla h_1(\underline{x}^*), \nabla h_2(\underline{x}^*), \ldots \nabla h_m(\underline{x}^*)\}$$

$$= \{\underline{x} \in \mathbb{R}^n : \underline{x} = z_1 \nabla h_1(\underline{x}^*) + \cdots + z_m \nabla h_m(\underline{x}^*), \; z_1, \ldots, z_m \in \mathbb{R}\}$$

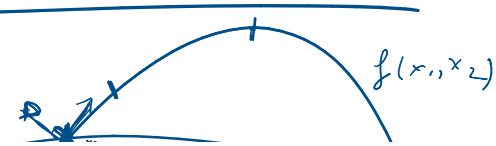
$$\mathbb{R}^n = N(\underline{x}^*) \oplus T(\underline{x}^*)$$

$$T(\underline{x}^*) = N(\underline{x}^*)^\perp$$

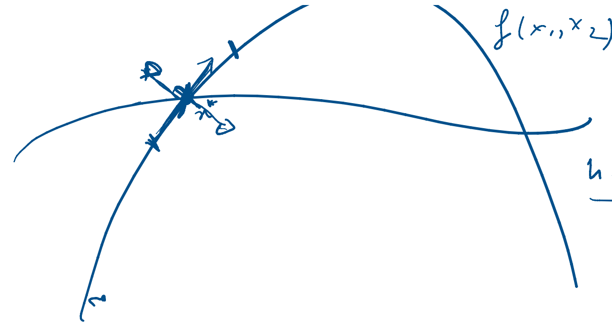
$$N(\underline{x}^*) = T(\underline{x}^*)^\perp$$

(Exercise)

to revise.



... condition

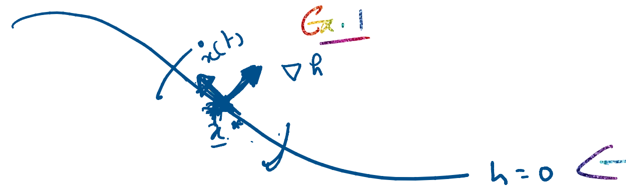


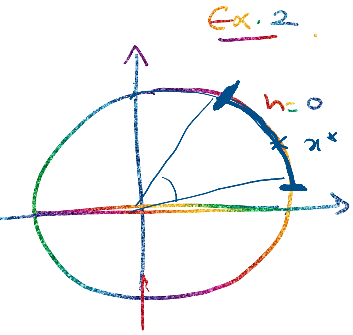
Lagrange condition

$\left[\begin{array}{l} f \quad \text{fn of two variables} \\ h : \mathbb{R}^2 \to \mathbb{R} \quad \to \text{one constraint.} \end{array}\right.$

$h=0$

$x_1^2+x_2^2=1$





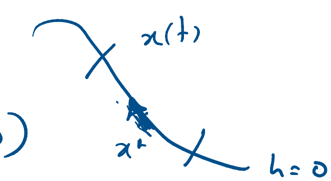
Parametrize the level set in the neighbourhood of $\underline{x}^*$ by $\{\underline{x}(t)\}$ that is a continuous differentiable function $\underline{x} : \mathbb{R}^2 \to \mathbb{R}$

s.t. $\underline{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$, $t \in (a,b)$

$\underline{x}^* = x(t^*)$, $\dot{x}(t^*) \neq 0$, $t^* \in (a,b)$.

Implicit fn theorem.

$\boxed{\nabla h(\underline{x}^*) \perp \dot{\underline{x}}(t^*)}$ Claim



$h=0 \Rightarrow h(x(t))=0 \quad t \in (a,b)$

$\nabla h \cdot \dot{x}(t) = 0$

$\boxed{\nabla h(\underline{x}^*) \perp \dot{\underline{x}}(t^*)} \to$

Let $\underline{x}^*$ be a minimizer of $f : \mathbb{R}^2 \to \mathbb{R}$ on the set $\{\underline{x} : h(\underline{x})=0\}$.

$x^* = x(t^*)$

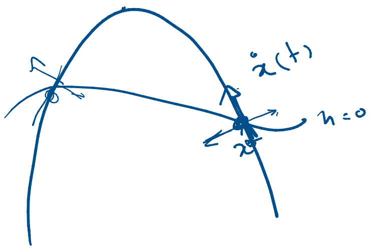
... is orthogonal to $\dot{\underline{x}}(t^*)$

the Set $\{ \underline{x} : h(\underline{x}) \ldots$

$x^* = x(t^*)$

Claim: $\boxed{\nabla f(x^*) \text{ is orthogonal to } \dot{\underline{x}}(t^*)}$

$\phi(t) = f(x(t))$ attains a minimum at t^*.



$\phi'(t^*) = 0$

$$\boxed{\nabla f(x^*) \cdot \dot{x}(t^*) = 0}$$

$$\nabla f(x^*) \perp \dot{x}(t^*)$$

$$\boxed{\nabla f(x^*) + \lambda \nabla h(x^*) = 0}$$

Lagrange ×

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

$$\ell(\underline{x}, \lambda) = f(\underline{x}) + \lambda h(\underline{x})$$

$$\ell_{\underline{x}}(\underline{x}^*, \lambda^*) = 0$$

$$\ell_\lambda(\underline{x}^*, \lambda^*) = 0 \longrightarrow \underline{h}(\underline{x}^*) = 0.$$

f

$F(x^*) > 0$

Recall Hessian

$F > 0$ [minimizer]

$F < 0$ [maximizer]

$F = 0] \longrightarrow$ Inconclusive.

$$y^T \mathcal{L}(\underline{x}^*) \underline{y}$$

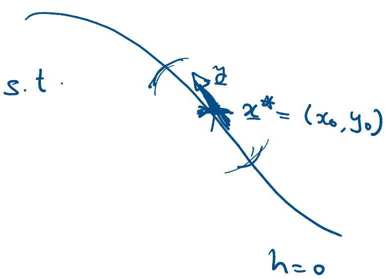
$T(x^*)$

Lecture 21

Thursday, 17 March 2022 1:40 PM

[Revisit - Implicit function Thm and Lagrange Thm].

Implicit fn Thm $\quad (h : \mathbb{R}^2 \to \mathbb{R})$



Assume
- (i) $h(x,y) \in \mathscr{C}^1$ near (x_0, y_0) s.t.
- (ii) $h(x_0, y_0) = 0$
- (iii) $\dfrac{\partial h}{\partial y}(x_0, y_0) \neq 0$

Then, $\exists$ a unique function $y = f(x) \in \mathscr{C}^1$ in the neighbourhood of x_0 s.t.

$h(x,y) = 0$

$h(x, f(x)) = 0$

(a) $y_0 = f(x_0)$

(b) $h(x, f(x)) = 0 \quad \forall x$ near x_0

(c) $f'(x) = -\dfrac{h_x(x, f(x))}{h_y(x, f(x))}$

$$\left[\begin{array}{l} x_1(t) = t \\ x_2(t) = f(t) \end{array}\right] \qquad \dot{x}(t) = \begin{bmatrix} 1 \\ f'(t) \end{bmatrix} \neq \underline{0}$$

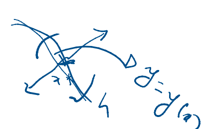
Lagrange multiplier theorem

$\left[\cdot \min f(\underline{x}) \text{ s.t. } h(\underline{x}) = 0 \right.$

$\left. f : \mathbb{R}^2 \to \mathbb{R}. \right]$

- $f \in \mathscr{C}^{(1)}$, $h \in \mathscr{C}^{(1)}$, $\nabla h \neq 0$
- $\underline{x}^* = (x_0, y_0)$ is an extremum such that $h(x_0, y_0) = 0$.

Conclusion: $\exists \lambda \in \mathbb{R}$ s.t. $\nabla (f + \lambda h)(x_0, y_0) = 0$.



Proof: Note that the conditions of implicit fn thm are satisfied.

$\Rightarrow \exists$ a function $y = y(x)$ that satisfies

$h(x, y(x)) = 0 \quad$ [in nbd of $\underline{x}^*$]

$h_x + h_y \, y'(x) = 0$

$$h_x + h_y \, y'(x) = 0$$

$$\Rightarrow \boxed{y'(x) = -\frac{h_x}{h_y}}$$

Given $f(x, y(x))$ has a local extrema at (x_0, y_0).

$$f_x + f_y \, y'(x) = 0 \rightarrow [\text{At } (x_0, y_0)]$$

$$f_x(x_0, y_0) - f_y(x_0, y_0) \frac{h_x(x_0, y_0)}{h_y(x_0, y_0)} = 0$$

$$\left[\frac{f_y(x_0, y_0)}{h_y(x_0, y_0)} = -\lambda \right]$$

$$\begin{cases} f_x(x_0, y_0) + \lambda h_x(x_0, y_0) = 0 \\ f_y(x_0, y_0) + \lambda h_y(x_0, y_0) = 0 \\ h(x_0, y_0) = 0 \end{cases}$$

$$\boxed{\nabla(f + \lambda h)(x_0, y_0) = 0}$$

Lagrange Theorem: $f : \mathbb{R}^n \to \mathbb{R}$,

$\underline{h} : \mathbb{R}^n \to \mathbb{R}^m \qquad m \leq n.$

Necessary condtn

(i) Let $\underline{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ s.t.

$\underline{h}(\underline{x}) = 0$, $\underline{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$.

(ii) Let $\underline{x}^*$ be a regular point.

(iii) $f, h \in C^{(1)}$

Then $\exists \; \underline{\lambda} \in \mathbb{R}^m$ s.t.

$$\boxed{Df(\underline{x}^*) + \underline{\lambda}^T D\underline{h}(\underline{x}^*) = \underline{0}^T} \quad (A)$$

$\nabla f = \begin{bmatrix} \; \end{bmatrix}_{n \times 1}$

$Df = [\quad]_{1 \times n}$

$\underline{h} = (h_1, \dots h_m)$

$D\underline{h} = [\quad]_{m \times n}$

$\underline{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}_{m \times 1}$

Second Order necessary condition

Let $\underline{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$

$C^{(2)}$ [with all other assumptions

Let $\underline{x}^*$ be a local minimizer of $f: \mathbb{R}^n \to \mathbb{R}$ s.t. $\underline{h(\underline{x})=0}$, $f, h \in C^{(2)}$ [with all other assumptions as above.]

Then $\exists\ \underline{\lambda}^* \in \mathbb{R}^m$ s.t.

(i) (A) holds [from I order necessary cond^n]

(ii) $\forall\ \underline{y} \in T(\underline{x}^*)$, $\underline{\underline{y}^T\ \mathcal{L}(\underline{x}^*, \underline{\lambda}^*)\ \underline{y} \geq 0.}$

$$\mathcal{L}(\underline{x}^*, \underline{\lambda}^*) = F(\underline{x}^*, \underline{\lambda}^*) + \lambda_1^* H_1(\underline{x}^*, \underline{\lambda}^*) + \cdots + \lambda_m^* H_m(\underline{x}^*, \lambda^*)$$

Recall

$$\left[\ \begin{aligned} l(\underline{x}, \underline{\lambda}) &= f(\underline{x}) + \underline{\lambda}^T \underline{h}(\underline{x}) \\ &= f(\underline{x}) + \lambda_1 h_1(\underline{x}) + \cdots + \lambda_m h_m(\underline{x}). \end{aligned}\right.$$

Example Min $x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3$

s.t. $\begin{bmatrix} x_1 + 2x_2 = 3 \\ 4x_1 + 5x_3 = 6 \end{bmatrix}$ $\boxed{\underline{h}(\underline{x}) = 0}$

$$l(\underset{x_1,x_2,x_3}{\underline{x}}, \underset{\lambda_1,\lambda_2}{\underline{\lambda}}) = \overbrace{x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3}^{f} + \lambda_1 (\underbrace{x_1 + 2x_2 - 3}_{h_1}) + \lambda_2 (\underbrace{4x_1 + 5x_3 - 6}_{h_2}).$$

$\underline{x}^* = \underline{x}$

$$\nabla(\underline{f} + \lambda \underline{h})(x^*) = 0$$

$$\begin{bmatrix} 2x_1 + 2x_2 + 4 \\ 2x_1 + 6x_2 + 5 \\ 6 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 4 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Constraints.
$$\begin{cases} x_1 + 2x_2 - 3 = 0 \\ 4x_1 + 5x_3 - 6 = 0. \end{cases}$$

Constraints. $4x_1 + 5x_3 - 6 = 0.$

$$\begin{bmatrix} 2 & 2 & 0 & 1 & 4 \\ 2 & 6 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 1 & 2 & 0 & 0 & 0 \\ 4 & 0 & 5 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} -4 \\ -5 \\ -6 \\ 3 \\ 6 \end{bmatrix}$$

$$\boxed{\underline{x}^* = \begin{bmatrix} 16/5 \\ -1/10 \\ -34/25 \end{bmatrix}} \qquad \underline{\lambda}^* = \begin{bmatrix} -27/5 \\ -6/5 \end{bmatrix} \rightarrow \boxed{\text{check}}$$

II order

$$\underline{y}^T \mathcal{L}(\underline{x}^*, \underline{\lambda}^*)\, \underline{y} \geq 0$$
$$\forall\, \underline{y} \in T(\underline{x}^*).$$

$$T(x^*) = \left\{ \underline{y} : \begin{bmatrix} 1 & 2 & 0 \\ 4 & 0 & 5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underline{0} \right\}.$$

$$\begin{bmatrix} 1 & 2 & 0 \\ 4 & 0 & 5 \end{bmatrix} \xrightarrow{R_2 \to R_2/4} \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 5/4 \end{bmatrix}$$

$$\xrightarrow{R_2 \to R_2 - R_1} \begin{bmatrix} 1 & 2 & 0 \\ 0 & \boxed{-2} & 5/4 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} =$$

y_2 is basic.

$y_1 \to$ basic variable.
$y_2 \to$ basic.
y_3 = free.

$$-2y_2 + \frac{5}{4} y_3 = 0 \Rightarrow \underline{y_2 = \frac{5}{8} y_3}$$

$$y_1 + 2y_2 = 0$$
$$y_1 = -2 \times y_2$$
$$= -2 \times \frac{5}{8} y_3$$
$$= -5/4\, y_3.$$

$$T(x^*) = \left\{ \underline{a} \begin{bmatrix} -5/4 \\ 5/8 \\ 1 \end{bmatrix} : a \in \mathbb{R} \right\}$$

$T(x^*)$

Check: $\underline{y}^T \mathcal{L}(\underline{x}^*, \lambda^*)\, \underline{y} = \frac{75}{32} a^2 > 0$ ✓

$a = 0$

$\Rightarrow$ $\boxed{\underline{x}^* \text{ is a local minimizer}}$

Proof of Lagrange Theorem

$f(x) = x^3$

$f'(0) = 0$

$f''(0) = 0$

$x = 0$

$$Df(\underline{x}^*) + \underline{\lambda}^{*T} Dh(\underline{x}^*) = 0$$

for some $\underline{\lambda}^* \in \mathbb{R}^m$.

$$\underline{\nabla f(\underline{x}^*)} = -\left(D\underline{h}(\underline{x}^*)^T \underline{\lambda}^*\right)$$

normal space

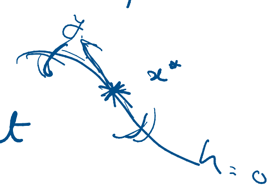
$$\underline{\nabla f(\underline{x}^*)} \in \mathcal{R}\left(D\underline{h}(\underline{x}^*)^T\right) = \underline{N(\underline{x}^*)}$$

$$= \left\{ \underline{x} \in \mathbb{R}^n : \underline{x} = Dh(\underline{x}^*)^T \underline{z},\ \underline{z} \in \mathbb{R}^m \right\}$$

S.T. $\nabla f(\underline{x}^*) \in T(\underline{x}^*)^\perp \Big] = T(\underline{x}^*)^\perp$ (Tangent space)

(Thm 20.1)

Let $\underline{y} \in T(\underline{x}^*) \Rightarrow \exists$ a differentiable curve $\{\underline{x}(t) : t \in (a,b)\}$ such that



$$\forall t \in (a,b), \quad \underline{h}(\underline{x}(t)) = 0$$

$\underline{x}(t^*) = x^*$

$\dot{\underline{x}}(t^*) = \underline{y}$

in tangent space

Let $\phi(t) = f(\underline{x}(t))$

f is attaining an extrema at t^*.

$\phi'(t) = 0$.

$$\phi'(t) = 0$$

$$\Rightarrow \quad Df(\underline{x}^*) \cdot \underline{\dot{x}}(t^*) = 0 \qquad \text{[Chain Rule]}$$

$$\Rightarrow \quad \underbrace{Df(\underline{x}^*)}_{\text{row}} \, \underline{y} = 0 \qquad \forall \underline{y} \in T(\underline{x}^*)$$

$$\Rightarrow \quad \boxed{Df(\underline{x}^*) \in T(\underline{x}^*)^{\perp}}$$

Concludes the proof.

VS ZSHR

Lecture 22

Monday, 21 March 2022 1:35 PM

Recall SONC :

A:
- $\underline{x}^*$ is a local minimizer of $f: \mathbb{R}^n \to \mathbb{R}$ s.t. $\underline{h}(\underline{x}) = \underline{0}$; $\underline{h}: \mathbb{R}^n \to \mathbb{R}^m$ $(m \le n)$,
- $f, h \in C^2$
- $\underline{x}^*$ is regular $\operatorname{rank}[Dh_1(x^*) \dots Dh_m(x^*)] = m$

Then

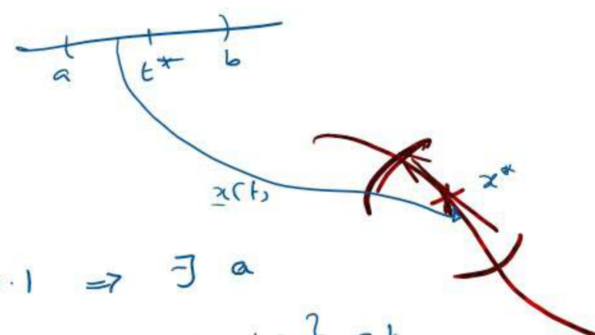
B:
$\exists\ \underline{\lambda}^* \in \mathbb{R}^m$ s.t.
(i) $Df(\underline{x}^*) + \underline{\lambda}^{*T} Dh(\underline{x}^*) = \underline{0}^T$
(ii) for all $\underline{y} \in T(\underline{x}^*)$, $\underline{y}^T \mathcal{L}(\underline{x}^*, \underline{\lambda}^*)\,\underline{y} \ge 0$.

$\boxed{\text{SOSC}}$: Let (A) hold, B(i) hold.

For all $\underline{y} \in T(\underline{x}^*)$, $\underline{y} \ne \underline{0}$

$$\underline{y}^T \mathcal{L}(\underline{x}^*, \lambda^*)\,\underline{y} > 0$$

$\Rightarrow$ $\underline{x}^*$ is a strict local minimizer of f s.t. $\underline{h}(\underline{x}) = \underline{0}$.



Proof of SONC.

(i) is I order condition.

(ii)
Let $y \in T(\underline{x}^*)$, Theorem 20.1 $\Rightarrow$ $\exists$ a twice differentiable curve $\{x(t),\ t \in (a,b)\}$ s.t.

$$\begin{cases} \underline{x}(t^*) = x^* \\ \dot{\underline{x}}(t^*) = y \end{cases} \quad \text{for some } t^* \in (a,b)$$

t^* is a local minimizer for $\phi(t) = f(\underline{x}(t))$

$$\Leftarrow \begin{cases} \phi'(t^*) = 0 \\ \phi''(t^*) \ge 0 \end{cases}$$

$\phi''(t^*) \geq 0$

$$\begin{cases} h(x(t)) = 0 \\ 0 = \frac{d^2}{dt^2}\left[\lambda_1^* h_1 + \dots + \lambda_m^* h_m\right] \end{cases}$$

HW

$$\frac{d}{dt}\left[\underline{y}(t)^T \underline{z}(t)\right] = \underline{z}(t)^T \frac{d\underline{y}}{dt}(t) + \underline{y}(t)^T \frac{d\underline{z}}{dt}(t)$$

$\boxed{\phi''(t^*) \geq 0}$

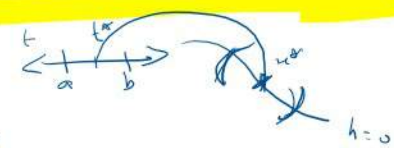
$\phi(t) = f(\underline{x}(t))$

$$\frac{d^2\phi}{dt^2}(t^*) = \frac{d}{dt}\left[\frac{d\phi}{dt}(t^*)\right]$$

$$= \frac{d}{dt}\left[\underbrace{Df}_{(\nabla f)^T}(\underline{x}(t^*))\ \dot{\underline{x}}(t^*)\right]$$

$$= \dot{\underline{x}}(t^*)^T \frac{d}{dt}\left[Df(\underline{x}(t^*))^T\right] + Df(\underline{x}(t^*))\ \ddot{\underline{x}}(t^*)$$

$$\phi''(t^*) = \dot{\underline{x}}(t^*)^T F(\underline{x}(t^*))\ \dot{\underline{x}}(t^*) + Df(\underline{x}(t^*))\ \ddot{\underline{x}}(t^*) \geq 0 \quad \text{--- (1)}$$



$\underline{h}(\underline{x}(t)) = 0 \quad \forall\, t \in (a,b)$

$$\Rightarrow \frac{d^2}{dt^2}\left[\lambda_1^* h_1(\underline{x}(t)) + \dots\dots \lambda_m^* h_m(\underline{x}(t))\right] = 0$$

$$\Rightarrow \frac{d^2}{dt^2}\left[\sum_{k=1}^{m} \lambda_k^* h_k(\underline{x}(t))\right] = 0$$

$$\Rightarrow \frac{d}{dt}\left[\sum_{k=1}^{m} \lambda_k^* \frac{d}{dt}\left(h_k(\underline{x}(t))\right)\right] = 0$$

$$\Rightarrow \frac{d}{dt}\left[\sum_{k=1}^{m} \lambda_k^* Dh_k(\underline{x}(t))\ \dot{\underline{x}}(t)\right] = 0$$

$$\Rightarrow \sum_{k=1}^{m} \lambda_k^* \frac{d}{dt}\left[Dh_k(\underline{x}(t))\ \dot{\underline{x}}(t)\right] = 0$$

$\Rightarrow \sum_{k=1} \lambda_k \frac{d}{dt}\left[\text{———}\right]$

$$\boxed{\begin{aligned}&\frac{d}{dt}\left[y(t)^T z(t)\right] \checkmark\\ &= z(t)^T \frac{dy}{dt} + y(t)^T \frac{dz}{dt}\end{aligned}}$$

$$= \sum_{k=1}^{m} \lambda_k^* \left((\dot{x}(t))^T H_k(x(t))\, \dot{x}(t) + Dh_k(x(t))\, \ddot{x}(t) \right) = 0$$

$$\Rightarrow \boxed{\underbrace{\sum_{k=1}^{m} \lambda_k^* (\dot{x}(t^*))^T H_k(x(t^*))\, \dot{x}(t^*)} + \sum_{k=1}^{m} \lambda_k^* Dh_k(x(t^*))\, \ddot{x}(t^*) = 0} \quad -③$$

① + ③ ≥ 0

$$\boxed{y^T F(x^*)\, y + y^T \left[\lambda^* H(x^*)\right] y} \quad \lambda^* \to \geq 0$$

$$+ \left[\underbrace{Df(x^*) + \lambda^* Dh(x^*)}_{=0}\right] \ddot{x}(t^*) \geq 0$$

$=0$ (I order condition).

$\boxed{y = \dot{x}(t^*)}$

SONC: $\boxed{y^T \mathcal{L}(x^*, \lambda^*)\, y \geq 0 \quad \forall y \in T(x^*)}$

$l(x, \lambda) \to$ Lagrangian function.

Example

$$\boxed{\max \left(\frac{x^T Q x}{x^T P x} \right)} \qquad \left.\begin{aligned} Q = Q^T > 0 \\ P = P^T > 0 \end{aligned}\right]$$

$(x \in \mathbb{R}^n)$

→ If x is a solution, $t\,x$ for $t \neq 0$ is also a solution.

→ To avoid multiplicity impose $\boxed{x^T P x = 1}$ ✓

$$\max \overbrace{x^T Q x}^{f} \quad \text{s.t.} \quad \overbrace{x^T P x}^{h(x)} = 1 \qquad m = 1$$

$x^T Q x$

$$\max \sqrt{\underline{x}^T Q \underline{x}} \quad \text{s.t.} \quad \underline{x}^T P \underline{x} = 1$$

$$\ell(\underline{x}, \lambda) = \underbrace{\sqrt{\underline{x}^T Q \underline{x}}}_{f} + \lambda \big(\underbrace{1 - \underline{x}^T P \underline{x}}_{h}\big)$$

$\underline{x}^T Q \underline{x}$
$= a_{11} x_1^2 + 2a_{12} x_1 x_2 + \dots + 2a_{1n} x_1 x_n$
$+ a_{22} x_2^2 + \dots + a_{nn} x_n^2$

I order condition

$$\cancel{2}\, \underline{x}^T Q - \cancel{2} \lambda \underline{x}^T P = \underline{0}^T$$

$$1 - \underline{x}^T P \underline{x} = 0$$

$$Q \underline{x} - \lambda P \underline{x} = 0 \Rightarrow (\lambda P - Q) \underline{x} = 0$$

$$\Rightarrow \boxed{P^{-1} Q \underline{x} = \lambda \underline{x}}$$

Stationary points are eigenvectors of $P^{-1}Q$; Eigenvalues are Lagrange x.

$(\underline{\lambda}^*, \underline{x}^*)$ is optimal $\boxed{\underline{x}^{*T} P \underline{x}^* = 1}$

$$Q = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \qquad P = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify:

$$P^{-1} Q = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$(P^{-1}Q)\, \underline{x} = \lambda \underline{x}$$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} = 0 \Rightarrow \boxed{\lambda_1 = 2, \quad \lambda_2 = 1}$$

$\boxed{\lambda_1 = 2}$ Eigenvectors $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^* = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}$$

$$(\alpha \quad 0) \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ 0 \end{bmatrix} = 1$$

$$(\alpha \quad 0) \begin{pmatrix} 2\alpha \\ 0 \end{pmatrix} = 1$$

$$2\alpha^2 = 1 \qquad \alpha = \pm \frac{1}{\sqrt{2}}$$

$\boxed{\lambda_1 = 2}$

Consider e.v. $\begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 0 \end{pmatrix}$ that satisfy $\underline{x}^* P \underline{x} = 1$

$\boxed{\lambda_2 = 1}$

e.v. $\rightarrow \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$

$$\underline{x}^* P \underline{x} = 1$$

$$[0 \quad \alpha] \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha \end{bmatrix} = 1$$

$$\alpha^2 = 1 \qquad \alpha = \pm 1$$

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\left(2 ; \overset{x_1^*}{\begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}}\right) ; \left(2, \overset{x_2^*}{\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}}\right) ; \left(1, \overset{x_3^*}{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}\right) ; \left(1, \overset{x_4^*}{\begin{pmatrix} 0 \\ -1 \end{pmatrix}}\right)$$

Tangent space for $1 - \underline{x}^T P \underline{x} = 0$ at x_1^*.

$$T(\underline{x}_1^*) = \{ \underline{y} \in \mathbb{R}^2 : Dh(x^*) \, \underline{y} = 0 \}$$

$$= \{ \underline{y} \in \mathbb{R}^2 : \underline{x}^{*T} \underline{P} \, \underline{y} = 0 \}$$

$$= \{ \underline{y} \in \mathbb{R}^2 : \left[\tfrac{1}{\sqrt{2}} \quad 0\right] \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \underline{y} = 0 \}$$

$$= \{ \underline{y} \in \mathbb{R}^2 : [\sqrt{2} \quad 0] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \}$$

$$\sqrt{2} y_1 = 0 \Rightarrow y_1 = 0 \qquad y_2 = \text{free}.$$

$\sqrt{2}\, y_1 = 0 \Rightarrow y_1 = 0 \qquad y_2 = \text{free}.$

$$T(\underline{x}_1^*) = \left\{ \begin{pmatrix} 0 \\ \alpha \end{pmatrix} : \alpha \in \mathbb{R} \right\}$$

$$\ell(x, \lambda) = \underline{x}^T Q x + \lambda (1 - x^T P x)$$

$$\mathcal{L}(\underline{x}_1^*, \lambda^*) = \underline{2(Q - \lambda P)}$$

$$= 2Q - 4P$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$$

$$\underline{y}^T \mathcal{L}\, \underline{y}$$

$$= \begin{bmatrix} 0 & \alpha \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha \end{bmatrix}$$

$$= -2\alpha^2 < 0$$

$\underline{x}_1^* = \begin{pmatrix} 1/\sqrt{2} \\ 0 \end{pmatrix}$ is a strict local maxima.

Problems with inequality constraints

$$\left\{ \begin{array}{ll} \min & f(\underline{x}) \\ \text{s.t} & \underline{h}(x) = 0 \\ & \underline{g}(\underline{x}) \leq 0 \end{array} \right.$$

$f : \mathbb{R}^n \to \mathbb{R}$

$\underline{h} : \mathbb{R}^n \to \mathbb{R}^m \;(m \leq n)$

$\underline{g} : \mathbb{R}^n \to \mathbb{R}^p$

$\boxed{\text{KKT}}$

Lecture 23 [Problems with inequality constraints]

Thursday, 24 March 2022 1:46 PM

$$\left.\begin{array}{ll} \text{minimize} & f(\underline{x}) \\ \text{s.t.} & \begin{cases} \underline{h}(\underline{x}) = 0 \\ \underline{g}(\underline{x}) \le 0 \end{cases} \end{array}\right] \qquad \begin{array}{l} f : \mathbb{R}^n \to \mathbb{R} \\ \underline{h} : \mathbb{R}^n \to \mathbb{R}^m \\ \underline{g} : \mathbb{R}^n \to \mathbb{R}^p \end{array}$$

$\boxed{\text{Defn}}$ An inequality constraint $g(x) \le 0$ is said to be active at $\underline{x}^*$ if $g_j(\underline{x}^*) = 0$. It is said to be inactive at $\underline{x}^*$ if $g_j(\underline{x}^*) < 0$.

---

min $f(x) := (x+5)^2$ s.t. $x \le 0$

$g_1(x) = x$.

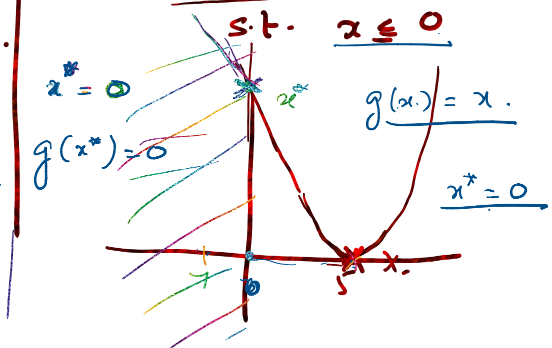
$\boxed{\text{inactive constraint}}$

$\downarrow$

$x^* = -5. < 0$

min $f(x) = (x-5)^2 \to \boxed{x=5}$

s.t. $x \le 0$



Defn $\boxed{\text{Regular point}}$ Let $\underline{x}^*$ satisfy $\begin{cases} \underline{h}(\underline{x}^*) = 0 \\ \underline{g}(\underline{x}^*) \le 0 \end{cases}$

and let $J(\underline{x}^*)$ be the index set of active inequality constraints

$$J(\underline{x}^*) = \{ j : g_j(\underline{x}^*) = 0 \}.$$

Then $\underline{x}^*$ is a regular point if the vectors

$$\underset{(i=1,\dots,m)}{\nabla h_i(x^*)} \quad \text{and} \quad \underset{j \in J(\underline{x}^*)}{\nabla g_j(x^*)} \quad \text{are linearly independent.}$$

Karush Kuhn Tucker Theorem [KKT] FONC [KT condn].

Karush-Kuhn-Tucker Theorem [KKT] (FONC) [KT condtn].

- Let $f, h, g \in \mathcal{C}^{(1)}$
- Let $\underline{x}^*$ be a regular point and a local minimizer for f s.t. $\underline{h}(\underline{x}) = 0$, $g(x) \le 0$.

Then $\exists\ \underline{\lambda}^* \in \mathbb{R}^m$ and $\underline{\mu}^* \in \mathbb{R}^p$ such that

(i) $\underline{\mu}^* \ge 0$ → (dual feasibility)

(ii) $Df(x^*) + \underline{\lambda}^{*T} D\underline{h}(\underline{x}^*) + \underbrace{\underline{\mu}^{*T}}_{\text{KKT multiplier}} Dg(\underline{x}^*) = 0$ → (Stationarity condition)

Lagrange multiplier

(iii) $\underline{\mu}^{*T} \underline{g}(\underline{x}^*) = 0$ → (Complementary Slack Condition).

(iv) $\underline{h}(\underline{x}^*) = 0$

(v) $\underline{g}(\underline{x}^*) \le 0$

(iv), (v) → primal feasibity condtns.

[1] $0 = \underbrace{\mu_1^*}_{\ge 0 \text{ from (i)}} \underbrace{g_1(x^*)}_{\le 0 \text{ (from (v))}} + \mu_2^* g_2(x^*) + \cdots + \mu_p^* g_p(x^*)$

(iii) is ⟸ $\left\{ \mu_j^* g_j(x^*) = 0 \quad j = 1, \cdots p. \right.$ →

[2]

	Inactive	Active
Constraints	$g_j(x^*) < 0$	$g_j(x^*) = 0$
KKT multiplies	$\mu_j^* = 0$	$\mu_j(x^*) \ge 0$
		$\mu_j > 0$ | $\mu_j = 0$

Second order Necessary Condition: $h_1, \ldots, h_m$

$$l(\underline{x}, \underline{\lambda}, \underline{\mu}) = f(\underline{x}) + \underline{\lambda}^T \underline{h}(\underline{x}) + \underline{\mu}^T \underline{g}(\underline{x})$$

Define $\mathcal{L}(\underline{x}, \underline{\lambda}, \underline{\mu}) = F(\underline{x}) + [\underline{\lambda}\ \ H(\underline{x})] + [\underline{\mu}\ \ G(x)]$

Hessian μ_1(Hessian of g_1)

Hessian of f $\quad \lambda_1$ (Hessian of h_1) $+\cdots+ \lambda_m$ (Hessian of h_m) $\quad \mu_1$(Hessian of g_1) $+\cdots+ \mu_p$ (Hessian of g_p)

$$T(\underline{x}^*) = \{ \underline{y} \in \mathbb{R}^n : D\underline{h}(\underline{x}^*)\,\underline{y} = 0,\ Dg_j(\underline{x}^*)\,\underline{y} = 0;\ j \in \underbrace{J(\underline{x}^*)}_{\text{index set of active constraints.}} \}$$

$\boxed{\text{SONC}}$

Thm: Let $\underline{x}^*$ be a local minimizer of $f: \mathbb{R}^n \to \mathbb{R}$, s.t. $\boxed{\underline{h}(\underline{x}) = 0,\ \underline{g}(\underline{x}) \le 0,}$ $\underline{h}: \mathbb{R}^n \to \mathbb{R}^m$, $\underline{g}: \mathbb{R}^n \to \mathbb{R}^p$, $f, \underline{h}, \underline{g} \in \mathcal{C}^2$. Let $\underline{x}^*$ be regular. Then

$\exists\ \underline{\lambda}^* \in \mathbb{R}^m$, $\underline{\mu}^* \in \mathbb{R}^p$ s.t.

(i) $\underline{\mu}^* \geq 0$ (ii) $Df(\underline{x}^*) + \underline{\lambda}^{*T} D\underline{h}(\underline{x}^*) + \underline{\mu}^{*T} D\underline{g}(\underline{x}^*) = \underline{0}^T$

(iii) $\underline{\mu}^{*T} \underline{g}(\underline{x}^*) = 0$

(iv) $\forall\ \underline{y} \in T(\underline{x}^*)$, $\underline{y}^T \mathcal{L}(\underline{x}^*, \underline{\lambda}^*, \underline{\mu}^*)\,\underline{y} \geq 0$.

$\boxed{\text{II order Suff. conditions}}$ (SOSC).

- $f, g, h \in \mathcal{C}^2$
- $\exists$ a feasible point $\boxed{\underline{h}(\underline{x}^*) = 0,\ \underline{g}(\underline{x}^*) \le 0.}$ $\underline{x}^* \in \mathbb{R}^n$ and vectors $\underline{\lambda}^* \in \mathbb{R}^m$ and $\underline{\mu}^* \in \mathbb{R}^p$ s.t.

(i) $\underline{\mu}^* \geq 0$ (ii) $Df(\underline{x}^*) + \underline{\lambda}^{*T} Dh(\underline{x}^*) + \underline{\mu}^{*T} D\underline{g}(\underline{x}^*) = \underline{0}^T$

(iii) $\underline{\mu}^* \underline{g}(\underline{x}^*) = 0$ (iv) $\forall\ y \in \underline{\tilde{T}(\underline{x}^*, \underline{\mu}^*)}$; $\underline{y}^T \mathcal{L}(\underline{x}^*, \underline{\lambda}^*, \underline{\mu}^*)\,\underline{y} > 0$ $\underline{y} \neq 0$

$\tilde{T}(\underline{x}^*, \underline{\mu}^*) = \{ \underline{y} : Dh(\underline{x}^*)\,\underline{y} = 0;\ Dg_j(\underline{x}^*)\,\underline{y} = 0\ \ \forall j \in \tilde{J}(\underline{x}^*, \underline{\mu}^*) \}$

$\tilde{T}(\underline{x}^*, \underline{\mu}^*) = \{\underline{y} : Dh(\underline{x}^*)\underline{y} = 0;\ Dg_j(\underline{x}^*)\underline{y} = 0 \quad \forall j \in \tilde{J}(\underline{x}, \underline{\mu})\}$

Then $\underline{x}^*$ is a strict local minimizer of f : s.t. $\begin{bmatrix} h(\underline{x}) = 0 \\ g(\underline{x}) \leq 0 \end{bmatrix}$

with $\tilde{J}(\underline{x}^*, \mu^*) = \{ j : g_j(\underline{x}^*) = 0,\ \mu_j^* > 0 \} \subset \underline{J(\underline{x}^*)}$

$= \{ j : \underline{g_j(x^*) = 0} \}$

<u>Example</u>

$$\min \quad x_1 x_2$$
$$x_1 + x_2 \geq 2$$
$$x_2 \geq x_1$$

$f(\underline{x}) = f(x_1, x_2) = x_1 x_2$

$g_1(\underline{x}) = 2 - x_1 - x_2 \leq 0$

$g_2(\underline{x}) = x_1 - x_2 \leq 0$

$\boxed{\text{KKT conditions}}$

$$[x_2 \quad x_1] + \mu_1 [-1 \quad -1] + \mu_2 [1 \quad -1] = 0$$

$x_1 - \mu_1 - \mu_2 = 0$

$x_2 - \mu_1 + \mu_2 = 0$

$\mu_1 (2 - x_1 - x_2) + \mu_2 (x_1 - x_2) = 0$

$\mu_1, \mu_2 \geq 0$

$2 - x_1 - x_2 \leq 0$

$x_1 - x_2 \leq 0$

$x_1 = \mu_1 + \mu_2$

$x_2 = \mu_1 - \mu_2$

$\mu_1 = \frac{x_1 + x_2}{2}$

$\mu_2 = \frac{x_1 - x_2}{2}$

$\longrightarrow \mu_1 (2 - 2\mu_1) + \mu_2 (2\mu_2) = 0$

$\mu_1 (1 - \mu_1) = 0 \qquad \underline{\mu_2 = 0}$

$\mu_1 = 0$ or $\mu_1 = 1$

μ_1^*	~~0~~	1
μ_2^*	0	
x_1	1	
x_2	1	

$x_1 + x_2 \geq 2$

$\mu_1 \geq 1$

1 $\boxed{\begin{matrix} x_1^* = 1 & \mu_1^* = 1 \\ x_2^* = 1 & \mu_2^* = 0 \end{matrix}}$ <u>feasible pt.</u>

① $\boxed{x_2^* = 1 \qquad \mu_2^* = 0}$ $\underline{\text{feasible pt.}}$

③ Is $\underline{x}^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ a $\underline{\text{regular point}}$? Yes. [Check].

③ $$\mathcal{L}(\underline{x}^*, \underline{\lambda}^*, \mu^*) = F(\underline{x}^*) + \mu_1 (\text{Hessian of } g_1) + \mu_2 (\text{Hessian of } g_2).$$

$f(x_1, x_2) = x_1 x_2$

$\nabla f = \begin{pmatrix} \underline{x_2} \\ \underline{x_1} \end{pmatrix}$ $\qquad F = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $\qquad \nabla g_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$, $\nabla g_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ Hessian are 0

$$\mathcal{L}(\underline{x}^*, \lambda^*, \mu^*) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\tilde{T}(\underline{x}^*, \mu^*) = \left\{ \underline{y} \in \mathbb{R}^2 : \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \right\}$$

$y_1 + y_2 = 0.$

$\alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$

$$\underline{y}^T \mathcal{L}(\underline{x}^*, \underline{\lambda}^*, \underline{\mu}^*) \underline{y} = \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \alpha^2 \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \alpha^2 \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -2\alpha^2 < 0.$$

$\boxed{\text{Local maximizer}}$

Lecture 24

Monday, 28 March 2022 1:38 PM

Problem: [Tut Sheet 8] Let $f: \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, f is a convex function on set of feasible points

$$\Omega = \{\underline{x} \in \mathbb{R}^n : \underline{h}(\underline{x}) = 0\}$$

$\left(\begin{array}{l}\underline{h}: \mathbb{R}^n \to \mathbb{R}^m \\ \Omega \text{ is convex}\end{array}\right)$.

Let $\exists\ \underline{x}^* \in \Omega$ & $\underline{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\underline{x}^*) + \underline{\lambda}^{*T} D\underline{h}(\underline{x}^*) = \underline{0}^T.$$

Then $\underline{x}^*$ is a global minimizer of f.

Proof: (Thm 22.4) Convexity

f is convex on Ω

$\iff \forall\ x, y \in \Omega,$

$$f(y) \geq f(x) + Df(\underline{x})(\underline{y} - x)$$

Let $\underline{x}, \underline{x}^* \in \Omega$. Since f is convex and $f \in C^{(1)}$,

$$f(\underline{x}) \geq f(\underline{x}^*) + Df(\underline{x}^*)(\underline{x} - \underline{x}^*)$$

Lagrange condtn

$$\Rightarrow f(\underline{x}) \geq f(\underline{x}^*) - \underbrace{\underline{\lambda}^{*T} D\underline{h}(\underline{x}^*)(\underline{x} - \underline{x}^*)}$$

$\underline{x}^*, \underline{x} \in \Omega$; Ω is convex

$$\Rightarrow (1-\alpha)\,\underline{x}^* + \alpha \underline{x} \in \Omega \qquad \alpha \in [0,1].$$

$$\Rightarrow h\big((1-\alpha)\underline{x}^* + \alpha \underline{x}\big) = 0$$

$$\Rightarrow \underline{h}\big(\underline{x}^* + \alpha(\underline{x} - \underline{x}^*)\big) = 0$$

$$\Rightarrow \quad \underline{h}(\underline{x}^* + \alpha(\underline{x} - x^*)) = 0$$

$$\Rightarrow \quad \frac{\underline{\lambda}^{*T}\, \underline{h}\,(\underline{x}^* + \overbrace{\alpha\,(x - x^*)})) - \overset{0}{\overset{\shortparallel}{\underline{\lambda}^{*T}\, \underline{h}\,(\underline{x}^*)}}}{\alpha} = 0$$

Take limit as $\alpha \to 0$, $\underline{\lambda}^{*T} D\, \underline{h}\,(x^*)\,(\underline{x} - \underline{x}^*) = 0$.

Hence $\quad f(\underline{x}) \geq f(x^*) \qquad \forall\, \underline{x} \in \Omega$

$\Rightarrow$ $\underline{x}^*$ is a global min.

Problem 11 Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$ is convex on

$\Omega = \{ x \in \mathbb{R}^n : \underline{h}(\underline{x}) = 0,\ g(\underline{x}) \leq 0 \}$; $h : \mathbb{R}^n \to \mathbb{R}^m$,

$\underline{g} : \mathbb{R}^n \to \mathbb{R}^p$, $h, g \in C^1$; Ω is convex.

Suppose $\exists\ \underline{x}^* \in \Omega$, $\underline{\lambda}^* \in \mathbb{R}^m$ and $\underline{\mu}^* \in \mathbb{R}^p$ s.t.

(i) $\underline{\mu}^* \geq 0$

(ii) $Df(\underline{x}^*) + \underline{\lambda}^{*T} D\underline{h}(\underline{x}^*) + \underline{\mu}^{*T} D\underline{g}(\underline{x}^*) = 0$

(iii) $\underline{\mu}^{*T}\, \underline{g}\,(\underline{x}^*) = 0$.

Then $\underline{x}^*$ is a global minimizer of f over Ω.

Proof: Proceed as in the previous result.

$$Df(\underline{x}^*) = -\underline{\lambda}^{*T} Dh(\underline{x}^*) - \underline{\mu}^{*T} Dg(x^*).$$

$$f(\underline{x}) \geq f(\underline{x}^*) - \underbrace{\underline{\lambda}^{*T} D\underline{h}(x^*)(\underline{x} - \underline{x}^*)}_{\overset{\shortparallel}{0}} - \overbrace{\underline{\mu}^{*T} Dg(\underline{x}^*)\underbrace{(\underline{x} - \underline{x}^*)}}$$

(as in the last result).

(as in ^{0} the last result).

$f(x) \geq f(x^*) + (+ve)$

$$\overline{(1-\alpha)\,\underline{x}^* + \alpha\,\underline{x}} \in \Omega \quad \to \Omega \text{ is convex.}$$

$$g\left((1-\alpha)\,\underline{x}^* + \alpha\,\underline{x}\right) \leq 0$$

$$g\left(\underline{x}^* + \alpha(\underline{x} - \underline{x}^*)\right) \leq 0$$

≥ 0 (by condtn (ii))

$= 0$ by condtn (iii)

$$\lim_{\alpha \to 0} \frac{\underline{\mu}^{*T}\, g\left(\underline{x}^* + \alpha(x - x^*)\right) - \overbrace{\underline{\mu}^{*}\, g(\underline{x}^*)}^{=0}}{\alpha} \leq 0$$

$$\underline{\mu}^{*T}\, Dg(\underline{x}^*)\,(\underline{x} - \underline{x}^*) \leq 0.$$

$$\Rightarrow f(\underline{x}) \geq f(\underline{x}^*) \quad \forall\, x \in \Omega.$$

Example: Linear programming problem in Standard form.

I. maximize $\underline{c}^T \underline{x}_{n\times 1}$

s.t. $\underset{m\times n}{A}\,x = \underline{b}_{m\times 1}$

$\underline{x} \geq 0$

II. maximize $\underline{c}^T \underline{x}$

s.t. $\boxed{\begin{array}{l} A\underline{x} \leq \underline{b} \\ \underline{x} \geq 0 \end{array}}$

Write KKT Conditions.

[$m \leq n$ may not be true]

I. min $-\underline{c}^T \underline{x}$ $\longrightarrow$ $f(\underline{x}) = -\underline{c}^T \underline{x} = -c_1x_1 - c_2x_2 \cdots - c_nx_n$

s.t. $A\underline{x} - \underline{b} = 0$ $\longrightarrow$ $h(\underline{x}) = A\underline{x} - \underline{b} = 0$ [m condtn].

$-\underline{x} \leq 0$ $\quad$ $g(\underline{x}) = -\underline{x} \leq 0.$ [n condtns]

KKT conditions

[$\underline{\lambda}^* \to$ no restriction on sign

KKT conditions

$$\left[\begin{array}{l} \underline{\lambda}^* \to \text{no restriction on sign} \\ \underline{\mu}^* \geq 0 \end{array}\right.$$

Primal Feasibility $\quad \left.\begin{array}{l} A\underline{x}^* = \underline{b} \\ \underline{x}^* \geq 0 \end{array}\right]$

$g_1(x) = -x_1 \quad \nabla g_1 = \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

$\vdots$

$g_n(x) = -x_n \quad \nabla g_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix}$

Stationarity $\quad Df(\underline{x}^*) + \underline{\lambda}^{*T} Dh(\underline{x}^*) + \underline{\mu}^{*T} Dg(\underline{x}^*) = \underline{0}$

$$\boxed{\underline{\lambda}^{*T} A - \underline{\mu}^{*T} = \underline{c}^T}$$

$\underline{\mu}^{*T} = [\mu_1^*, \dots, \mu_n^*]$

$Dg(\underline{x}^*) = \begin{bmatrix} -1 & 0 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ 0 & \dots & \dots & -1 \end{bmatrix}$

Complementarity $\quad \underline{\mu}^{*T} \underline{x} = \underline{0}.$

II.

$$\left.\begin{array}{ll} \max & \underline{c}^T \underline{x} \\ \text{s.t.} & A\underline{x} \leq \underline{b} \\ & \underline{x} \geq 0 \end{array}\right] \quad m + n$$

		Dual variables
min	$-\underline{c}^T \underline{x}$	
s.t.	$\underline{g}_1(\underline{x}) = A\underline{x} - \underline{b} \leq 0$	$\underline{w}^*$
	$\underline{g}_2(x) = -\underline{x} \leq 0$	$\underline{v}^*$

$\underline{\mu}^* = \begin{bmatrix} w^* \\ v^* \end{bmatrix}.$

- $\underline{w}^*, \underline{v}^* \geq 0.$
- $A\underline{x}^* \leq \underline{b}, \quad \underline{x}^* \geq 0 \quad$ [Primal feasibility].
- ✓ $\underline{w}^{*T} A - \underline{v}^{*T} = \underline{c}^T \quad$ [Stationarity] $\quad Df(\underline{x}^*) + \underline{w}^* Dg_1(x^*) + \underline{v}^* Dg_2(x^*) = 0$
- $\underline{w}^{*T}(A\underline{x}^* - \underline{b}) = 0; \quad \underline{v}^{*T}\underline{x}^* = 0.$

Example.

$$\max \quad f(x_1, x_2) = 7x_1 + 6x_2$$

s.t.

	Dual variables
$3x_1 + x_2 \leq 120$	$\to w_1$
$x_1 + 2x_2 \leq 160$	$\to w_2$
$x_1 \leq 35$	$\to w_3$
$x_1 \geq 0,$	$\to v_1$

$x_1 \leq 35 \longrightarrow w_3$

$x_1 \geq 0, \longrightarrow v_1$

$x_2 \geq 0 \longrightarrow v_2$

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \quad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \underline{b} = \begin{bmatrix} 120 \\ 160 \\ 35 \end{bmatrix}$$

$$\underline{c} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

Primal feasibility

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} \leq \begin{bmatrix} 120 \\ 160 \\ 35 \end{bmatrix}$$

$$\begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Dual feasibility

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \geq \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} \geq 0.$$

Stationarity

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix} 7 & 6 \end{bmatrix}.$$

$\boxed{\text{Complementary Slack}}$

$$\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \left(\begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 120 \\ 160 \\ 35 \end{bmatrix} \right) = \underline{0}$$

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underline{0}$$

A Q M A V F

Tutorial Sheet 7:

Q5. $f(x) = \frac{1}{2}x^T Q x - b^T x + c,$

$$Q = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad c = \pi^2$$

Take: $H_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$d^{(0)} = -g^{(0)} = -\left(Q x^{(0)} - b\right) = b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

$$x^{(1)} = x^{(0)} + \alpha_0 d^{(0)}, \quad \alpha_0 = \frac{-g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} = \frac{-\begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}$$

$$= \frac{1}{2}$$

Next: Compute B_1^{-1} from the formula in Lecture Slides 19.

Then update $d^{(1)}$. Compute $\alpha^{(1)}$.

Repeat it once more for getting B_2^{-1} etc...

Q6.b)

Minimize $-4x_1 - x_2^2$

subject to $x_1^2 + x_2^2 = 9$

(Let's do this using Lagrange multipliers).

$$f(\underset{\sim}{x}) = -4x_1 - x_2^2$$

$$h(\underset{\sim}{x}) = x_1^2 + x_2^2 - 9$$

$$\left.\begin{aligned} \nabla f(x^*) + \lambda^* \nabla h(\underset{\sim}{x}^*) &= \underset{\sim}{0} \\ h(x^*) &= 0 \end{aligned}\right\} \longrightarrow \begin{aligned} &\begin{bmatrix} -4 \\ -2x_2 \end{bmatrix} + \begin{bmatrix} 2\lambda x_1 \\ 2\lambda x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ &x_1^2 + x_2^2 - 9 = 0 \end{aligned}$$

$$-4 + 2\lambda^* x_1 = 0$$

$$-2x_2 + 2\lambda^* x_2 = 0 \longrightarrow 2x_2(-1+\lambda^*) = 0$$

$$x_1^2 + x_2^2 = 9$$

Case I: $x_2 = 0$

Then $x_1 = \pm 3$, $\lambda^* = \pm \frac{2}{3}$, $x^* = \begin{bmatrix} \pm 3 \\ 0 \end{bmatrix}$

Case II: $\lambda^* = 1$

$x_1 = 2$, $x_2 = \pm\sqrt{5}$, $x^* = \begin{bmatrix} 2 \\ \pm\sqrt{5} \end{bmatrix}$

If x^* is a minimizer, then by SONC,

$$y^T L(x^*, \lambda^*) y \geq 0 \quad \text{for all} \quad y \in T(x^*),$$

where $L(x^*, \lambda^*) = F(x^*) + \lambda_1 H_1(x^*, \lambda^*) + \cdots + \lambda_k H_k(x^*, \lambda^*)$

In our case:

$$F(x) = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}$$

$$H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$\therefore$ SONC is satisfied at x^* whenever

$$y^T \begin{bmatrix} 2\lambda^* & 0 \\ 0 & -2+2\lambda^* \end{bmatrix} y \geq 0$$

for all $y \in T(x^*)$.

Consider the pair $(x^*, \lambda^*) = \left([3, 0]^T, \frac{2}{3}\right)$

$$L(x^*, \lambda^*) = \begin{bmatrix} 4/3 & 0 \\ 0 & -2/3 \end{bmatrix}$$

$$T(x^*) = \left\{ y : Dh(x^*) \cdot y = 0 \right\}.$$

$$= \left\{ y : \begin{bmatrix} 2\cdot 3 & 2\cdot 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \right\}$$

$$= \left\{ y : \begin{bmatrix} 6 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 0 \right\}$$

$$= \left\{ y : y_1 = 0 \right\}.$$

So, an arbitrary vector $y \in T(x^*)$ ~~that~~ is of the form: $y = [0, y_2]^T$.

$\therefore$ $y^T L(x^*, \lambda^*)\, y = -\frac{2}{3} y_2^2 \le 0$

In particular, $y = [0, 1]^T \in T(x^*)$ and we have $y^T L(x^*, \lambda^*)\, y < 0$.

$\therefore$ x^* is <u>not</u> a local minimizer.

Next, check for $(x^*, \lambda^*) = ([-3, 0]^T, -\frac{2}{3})$

$$\mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} -4/3 & 0 \\ 0 & -10/3 \end{bmatrix} < 0$$

$\therefore$ $x^* = [-3, 0]^T$ is NOT a local minimizer.

Next, check for $\lambda^* = 1$: $\mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \geq 0$

$\therefore$ Potential minimizers are $x^* = \begin{bmatrix} 2 \\ \pm\sqrt{5} \end{bmatrix}$.

If x^* is a feasible point with Lagrange multiplier λ^*, such that

$$y^T L(x^*, \lambda^*) y > 0$$

for all $y \neq 0$ in $T(x^*)$, then by SOSC, x^* is a strict local minimizer for f.

$$T([2 \ \sqrt{5}]^T) = \{ y : Dh([2 \ \sqrt{5}]^T) \cdot y = 0 \}$$

$$= \{ y : [4 \ \ 2\sqrt{5}] \, y = 0 \}$$

$$= \{ y : 4y_1 + 2\sqrt{5} y_2 = 0 \}$$

$y^T \mathcal{L}(x^*, \lambda^*) y^* = 2y_1^2$ where $y = [y_1, y_2]^T$.

$x^* = [2 \ \sqrt{5}]^T$

Now, this is zero $\Leftrightarrow y_1 = 0$.

But if $y \in T(x^*)$ with $y_1 = 0$,

then we also must have that

$$4 \cdot 0 + 2\sqrt{5} \cdot y_2 = 0 \Rightarrow y_2 = 0.$$

$\therefore \ y^T \mathcal{L}(x^*, \lambda^*) y = 0$ for $y \in T(x^*) \Leftrightarrow y = 0$.

Else it is > 0.

SOSC is satisfied by $[2 \quad \pm\sqrt{5}]$.

So, both are strict local minimizers.

Q8. a)

minimize $2x_1x_2 + 2x_1x_3 + 2x_2x_3$

subject to $x_1x_2x_3 - V = 0$

$$\underset{\sim}{h}(\underset{\sim}{x}) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix}$$

$$\nabla f + \lambda \nabla h = 0$$

$$\begin{bmatrix} 2x_2 + 2x_3 \\ 2x_1 + 2x_3 \\ 2x_1 + 2x_2 \end{bmatrix} + \lambda \begin{bmatrix} x_2x_3 \\ x_1x_3 \\ x_1x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

b) $h(\underset{\sim}{x}) = x_1x_2x_3 - V$

$\{\nabla h(x^*)\}$ is L.I. $\Longleftrightarrow$ $\nabla h(x^*) \neq 0$.

Observe that in our case:

$$\left.\begin{array}{l} x_1 x_2 = 0 \\ x_2 x_3 = 0 \\ x_1 x_3 = 0 \end{array}\right\} \iff x_i = x_j = 0 \text{ for some } i \neq j$$

x satisfying

However any such point $[x_1, x_2, x_3]$ does not satisfy the constraint $h(x) = x_1 x_2 x_3 - V$

$\therefore$ All feasible $\underset{\sim}{x}$ in this case are regular points.

What is a regular point?

A point $\underset{\sim}{x}^*$ satisfying $\underset{\sim}{h}(x^*) = 0$ is regular if $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are m linearly independent vectors in $\mathbb{R}^n$. [Here $h : \mathbb{R}^n \to \mathbb{R}^m$,

$$h(\underset{\sim}{x}) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(\underset{\sim}{x}) \end{bmatrix}$$

Linear Programming Problems.

Monday, 4 April 2022 1:05 PM

Standard form

I. $\min \; \underline{c}^T \underline{x}$ s.t. $A\underline{x} = \underline{b}$, $\underline{x} \geq 0$

$\left(A_{m\times n} \; \underline{x}_{n\times 1} = \underline{b}_{m\times 1}\right)$

$m < n$

rank $A = m$.

$[\underline{b} \geq 0]$

II. $\min \; \underline{c}^T \underline{x}$

s.t. $A\underline{x} \geq \underline{b}$

$\underline{x} \geq 0$

III. $\min \; \underline{c}^T \underline{x}$

s.t. $A\underline{x} \leq \underline{b}$

$\underline{x} \geq 0$

$\downarrow$

$A\underline{x} + I_m \underline{y} = \underline{b}$

(Slack variables)

Standard form for II. & III.

$\downarrow$

$\min \; \underline{c}^T \underline{x}$

(Surplus variables)

s.t. $\begin{cases} a_{i1}x_1 + \cdots + a_{in}x_n - y_i = b_i \\ \quad i = 1, \ldots, m \\ x_1, x_2, \ldots, x_n \geq 0 \\ y_1, y_2, \ldots y_m \geq 0 \end{cases}$

$A\underline{x} - I_m \underline{y} = \underline{b}$

$A_{m\times n}$ $\quad m < n$

$A = [\,\underbrace{B}\,|\; D\,]$ (reordering)

s.t. $B_{m\times m}$ is non-singular

$D_{m\times(n-m)}$

$A\underline{x} = \underline{b}_{m\times 1}$

$\begin{bmatrix} x_1 \end{bmatrix}$

$m \times (n-m)$

$$A_{m\times n}\,\underline{x} = \underline{b}_{m\times 1}$$

$$B\,\underline{x}_B = \underline{b}$$

$$\begin{array}{c} x_B \\ \\ x_D \end{array} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \hline x_{m+1} \\ \vdots \\ x_n \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{matrix}}\right\} m \\ \left.\vphantom{\begin{matrix} x_{m+1} \\ \vdots \\ x_n \end{matrix}}\right\} (n-m) \end{array}$$

$$\left[\, B_{m\times m} \;\middle|\; D_{m\times(n-m)} \,\right] \begin{bmatrix} \underline{x}_B \\ \hline \underline{\underline{x}}_D \end{bmatrix} = \underline{b}$$

$$\underline{x}_B = B^{-1}\underline{b}$$

$$\underbrace{B\,\underline{x}_B}_{\text{basic variables}} + \underbrace{D\,\underline{x}_D}_{\text{free variables.}} = \underline{b}$$

$$\boxed{\underline{x} = \left[\underline{\underline{x}}_B^T,\ \underline{0}^T\right]^T}$$

Ex.

$$A = [\underline{a}_1 \quad \underline{a}_2 \quad \underline{a}_3 \quad \underline{a}_4] = \begin{bmatrix} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}.$$

$m = 2$

Aim: To find basic solutions

$$B = [\underline{a}_1 \quad \underline{a}_2] \qquad D = [\underline{a}_3 \quad \underline{a}_4]$$

$$\left[\begin{array}{cc|cc} 1 & 1 & -1 & 4 \\ 1 & -2 & -1 & 1 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}.$$

$R_2 \to R_2 - R_1$

$$\left[\begin{array}{cc|cc} 1 & 1 & -1 & 4 \\ 0 & -3 & 0 & -3 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ \hline x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ -6 \end{bmatrix}$$

$0 \quad -3 \quad 0 \quad x_4$

$x_3 = x_4 = 0 \qquad -3x_2 = -6 \qquad x_2 = 2.$

$x_1 + x_2 = 8 \qquad x_1 = 6$

$$\underline{x}_B = \begin{bmatrix} 6 \\ 2 \end{bmatrix} \qquad \boxed{\underline{x} = \begin{bmatrix} 6 \\ 2 \\ 0 \\ 0 \end{bmatrix}} \to \boxed{\text{basic solution}}$$

$\underline{a}_1 x_1 + \underline{a}_2 x_2 + \underline{a}_3 x_3 + \underline{a}_4 x_4 = \underline{b}$

$$\begin{cases} x_1 + x_2 - x_3 + 4x_4 = b_1 \\ x_1 + 4x_4 + x_2 - x_3 = b_1 \end{cases}$$

<u>Another choice</u>:

$$B = [\underline{a}_1 \quad \underline{a}_4] \qquad \underline{D} = [\underline{a}_2 \quad \underline{a}_3]$$

$$\left[\begin{array}{cc|cc} 1 & 4 & 1 & -1 \\ 1 & 1 & -2 & -1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_4 \\ \hline x_2 \\ x_3 \end{array}\right] = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$

$\boxed{R_2 \to R_2 - R_1}$

$$\left[\begin{array}{cc|cc} 1 & 4 & 1 & -1 \\ 0 & -3 & -3 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_4 \\ \hline x_2 \\ x_3 \end{array}\right] = \begin{bmatrix} 8 \\ -6 \end{bmatrix}$$

$x_2 = x_3 = 0 \qquad -3x_4 = -6 \qquad x_4 = 2$

$x_1 + 4x_4 = 8 \qquad x_1 + 8 = 8 \qquad \boxed{x_1 = 0}$

<u>Basic soln.</u> $\quad \underline{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix}$ is the basic soln corresponding $\underline{a}_1$ and $\underline{a}_4$

(<u>degenerate basic solution</u>)

III choice

$B = [\underline{a}_2 \quad \underline{a}_3]$

$$\begin{array}{cccc} a_2 & a_3 & a_1 & a_4 \end{array}$$
$$\left[\begin{array}{cc|cc} 1 & -1 & 1 & 4 \\ -2 & -1 & 1 & 1 \end{array}\right] \begin{bmatrix} x_2 \\ x_3 \\ x_1 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$

$R_2 \to R_2 + 2R_1$

$$\left[\begin{array}{cc|cc} 1 & -1 & 1 & 4 \\ 0 & -3 & \cdots & \cdots \end{array}\right] \begin{bmatrix} x_2 \\ x_3 \\ x_1 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$$

$x_1 = x_4 = 0$ $\qquad -3x_3 = 18 \qquad \boxed{x_3 = -6}$

$x_2 - x_3 = 8$

$x_2 = 8 + -6 = 2.$

$$\boxed{\underline{x} = \begin{bmatrix} 0 \\ 2 \\ -6 \\ 0 \end{bmatrix}} \longrightarrow \underline{x} \not\geq 0 \quad \text{[Basic but not feasible].}$$

Defins:

- If some of the basic variables of the basic soln are zeroes, then the basic soln is called a degenerate basic soln.

- A vector $\underline{x}$ Satisfying $A\underline{x} = \underline{b}$, $\underline{x} \geq 0$ is called a feasible soln.

- A feasible soln that is also basic is called a a feasible soln.

- A feasible soln. that is also basic is called a basic feasible soln.

max # of basic solns $= {}^{n}C_{m}$.

Ex. 2 $A = \begin{bmatrix} 2 & 3 & -1 & -1 \\ 4 & 1 & 1 & -2 \end{bmatrix}$ $\underline{b} = \begin{bmatrix} -1 \\ 9 \end{bmatrix}$

Gauss Jordan Elimination (Augmented matrix)

$A\underline{x} = \underline{b}$

$\underline{x} = \underline{v} + \underline{h}$

g.soln. $A\underline{h} = 0$

$A\underline{v} = \underline{b}$

$\underline{v}$ particular soln

$$\left[\begin{array}{cccc|c} 2 & 3 & -1 & -1 & -1 \\ 4 & 1 & 1 & -2 & 9 \end{array}\right]$$

$\boxed{R_1 \to \frac{R_1}{2}}$ $\sim \left[\begin{array}{cccc|c} 1 & 3/2 & -1/2 & -1/2 & -1/2 \\ 4 & 1 & 1 & -2 & 9 \end{array}\right]$

$\boxed{R_2 \to R_2 - 4R_1}$ $\sim \left[\begin{array}{cccc|c} 1 & 3/2 & -1/2 & -1/2 & -1/2 \\ 0 & \boxed{-5} & 3 & 0 & 11 \end{array}\right]$

$\boxed{R_2 \to \frac{R_2}{-5}}$ $\sim \left[\begin{array}{cccc|c} 1 & 3/2 & -1/2 & -1/2 & -1/2 \\ 0 & 1 & -3/5 & 0 & -11/5 \end{array}\right]$

$$\begin{array}{ccccc} x_1 & x_2 & & & \\ \end{array}$$
$\left[\begin{array}{cccc|c} 1 & 0 & 2/5 & -1/2 & 14/5 \end{array}\right.$

$$\boxed{R_1 \to R_1 - 3/2\, R_2}$$

$$\sim \left[\begin{array}{cccc|c} 1 & 0 & 2/5 & -1/2 & 14/5 \\ 0 & 1 & -3/5 & 0 & -11/5 \end{array}\right]$$

$$x_1 + \underline{2/5\, x_3 - 1/2\, x_4} = 14/5$$

$$x_2 - 3/5\, x_3 = -11/5$$

$x_3 = s$

$x_4 = t$

$$x_1 = \frac{14}{5} - \frac{2}{5} s + \frac{1}{2} t$$

$$x_2 = -11/5 + 3/5\, s$$

$$x_3 = s$$

$$x_4 = t$$

$x_2 = x_4 = 0$

$\Rightarrow \boxed{s = 11/3} = x_3$

$x_1 = \frac{14}{5} - \frac{2}{5} \times \frac{11}{3}$

$= 4/3$

$$\begin{bmatrix} 4/3 \\ 0 \\ 11/3 \\ 0 \end{bmatrix}$$

$$\boxed{\underline{x} = \underbrace{\begin{pmatrix} 14/5 \\ -11/5 \\ 0 \\ 0 \end{pmatrix}}_{\underline{v}} + \underbrace{s \begin{pmatrix} -2/5 \\ 3/5 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1 \end{pmatrix}}_{\underline{h}}}$$

$$\underline{h} = s \begin{pmatrix} -2/5 \\ 3/5 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\underline{v} = \begin{pmatrix} 14/5 \\ -11/5 \\ 0 \\ 0 \end{pmatrix}$$

$\underline{x} = \underline{v} + \underline{h}$

(general soln).

$\boxed{\text{Basic Solutions}}$?

$x_3 = x_4 = 0 \qquad x_B = \begin{pmatrix} 14/5 \\ -11/5 \end{pmatrix}$

$\left[\frac{14}{5} \quad \boxed{\frac{-11}{5}} \quad 0 \quad 0\right]^T$ solves $A\underline{x} = \underline{b}$

but not <u>feasible</u>.

$\boxed{x_2 = x_4 = 0}$

<u>Importance of basic feasible solutions</u>

While solving LPP, we need to consider only basic feasible solutions. The optimal solution (if it exists) is always achieved at the <u>basic feasible soln</u>.

$\boxed{\text{Convexity}}$ The set of points that satisfy [feasible solns]

$A\underline{x} = \underline{b}, \quad \underline{x} \geq 0$

form a <u>convex set</u>.

$\Omega = \{\underline{x} : A\underline{x} = \underline{b}, \ \underline{x} \geq 0\}$

$A\underline{x}_1 = \underline{b} \qquad x_1 \geq 0$

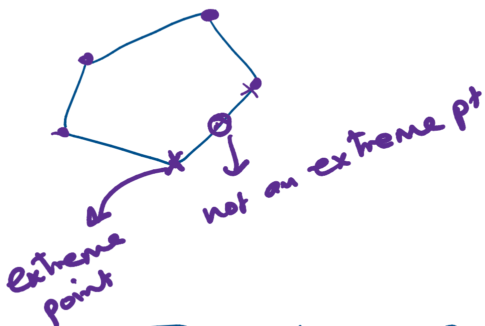
$A\underline{x}_2 = \underline{b} \qquad x_2 \geq 0$

$\underbrace{\alpha x_1 + (1-\alpha) x_2} \geq 0$

$A(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2)$

$= \alpha A x_1 + (1-\alpha) A x_2 = \underline{b}$

Extreme point $\underline{x}$ is an extreme point of Ω if there are no two distinct points $\underline{x}_1$ and $\underline{x}_2$ in Ω s.t. $\underline{x} = \alpha \underline{x}_1 + (1-\alpha)\underline{x}_2$, for $\alpha \in (0,1)$.



$\boxed{\text{Thm}}$ Let Ω be a convex set of all feasible solutions, that is, all the vectors satisfy

$$A\underline{x} = \underline{b}, \quad \underline{x} \geq 0. \qquad [A \in \mathbb{R}^{m\times n}, m < n]$$

Then $\underline{x}$ is an extreme point of Ω

$\iff$ $\underline{x}$ is a basic feasible solution of $A\underline{x} = \underline{b}$, $\underline{x} \geq 0$

Simplex Method for LPP.

Thursday, 7 April 2022 1:33 PM

1. Start with an initial BFS.
2. Is the current BFS optimal?
3. If yes, Stop, else; if no move to a new and improved BFS and return to step 2.

Ex. maximise $x_1 + x_2$

s.t. $x_1 + 3x_2 \le 9$

$2x_1 + x_2 \le 8$

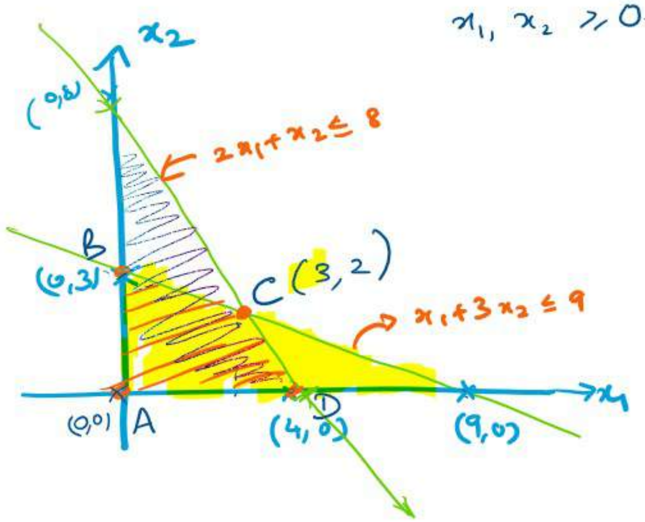
$x_1, x_2 \geq 0$

$f(x) = \boxed{x_1 + x_2}$ — (3) → slack variables.

$x_1 + 3x_2 + \boxed{x_3} = 9$ — (1)

$2x_1 + x_2 + \boxed{x_4} = 8$ — (2)

$(x_3, x_4 \geq 0)$



Step I

$$\begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} 1 & 3 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix} \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$$

x_3, x_4 basic

x_1, x_2 are free Variables $x_1 = x_2 = 0$.

$BFS = [0, 0, 9, 8]^t$ $\quad \underline{f = 0}$ $\quad (f = x_1 + x_2)$

At vertex A, we have $f = 0$.

From (3), increasing x_1 or x_2 will increase 'f'.

Let us increase x_1:

(i) From (1), increase x_1 to 9, decrease x_3 to 0

(ii) From (2), increase x_1 to 4, decrease x_4 to 0

(ii) From ③, increase x_1 to 4, decrease x_4 to 0

Choose the stricter restrictions, so that all variables remain positive.

$\uparrow x_1$ to 4, $x_2 = 0$, $\downarrow x_3 = 5$; $\downarrow x_4 = 0$

$[4, 0, 5, 0]$ $f = 4$.

free variables $\rightarrow x_2, x_4$

basic variables $\rightarrow x_1, x_3$

Write basic variables and f in terms of the free variables

$$\begin{bmatrix} 1 & 3 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \\ f \end{bmatrix}$$

$x_1 + 3x_2 + x_3 = 9$

$$\begin{bmatrix} 1 & 1 & 3 & 0 \\ 2 & 0 & 1 & 1 \\ \hline 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \\ f \end{bmatrix}$$

$x_1 + x_3 + 3x_2 = 9$

$R_1 \rightarrow R_1 - \frac{1}{2} R_2$

$$\begin{bmatrix} 0 & 1 & 5/2 & -1/2 \\ 2 & 0 & 1 & 1 \\ \hline 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ f \end{bmatrix}$$

$R_2 \rightarrow \frac{R_2}{2}$

$$\begin{bmatrix} 0 & 1 & 5/2 & -1/2 \\ 1 & 0 & 1/2 & 1/2 \\ \hline 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ f \end{bmatrix}$$

$R_3 \to R_3 - R_2$

$$\begin{bmatrix} 0 & 1 & 5/2 & -1/2 \\ 1 & 0 & 1/2 & 1/2 \\ \hline 0 & 0 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ \hline f-4 \end{bmatrix}$$

$$f = 4 + \tfrac{1}{2} x_2 - \tfrac{1}{2} x_4 \qquad \boxed{f = 4}$$

$$x_3 + \tfrac{5}{2} x_2 - \tfrac{1}{2} x_4 = 5 \qquad x_2 = 0$$

$$x_1 + \tfrac{1}{2} x_2 + \tfrac{1}{2} x_4 = 4 \qquad x_4 = 0$$

$$\boxed{[4, 0, 5, 0]^t \qquad f = 4.}$$

$$\tfrac{5}{2} x_2 + x_3 - \tfrac{1}{2} x_4 = 5 \quad \text{— (4)} \qquad f = x_1 + x_2$$

$$x_1 + \tfrac{1}{2} x_2 + \tfrac{1}{2} x_4 = 4 \quad \text{— (5)}$$

$$\boxed{\tfrac{1}{2} x_2 - \tfrac{1}{2} x_4 + 4 = f \quad \text{— (6)}}$$

Increase x_2 to maximise f.

Go to (4), (5) $\left\{ \begin{array}{ll} x_2 = 2 & \text{from (4)} \\ x_2 = 8 & \text{from (5)} \end{array} \right.$ $\quad \min\left\{ \frac{5}{5/2}, \frac{4}{(1/2)} \right\}$.

$x_2 = 2, \quad x_3 = 0, \quad x_4 = 0$ $\qquad \uparrow \underset{x_1}{x_2}$ to 2, decrease $x_3 = 0$

$x_1 = 3$

$$[3, 2, 0, 0]^t \qquad f = 5$$

$5/2 \quad -1/2 \quad x_1$ ...

$$\left[\begin{array}{cccc} 0 & 1 & 5/2 & -1/2 \\ 1 & 0 & 1/2 & 1/2 \\ \hline 0 & 0 & 1/2 & -1/2 \end{array}\right] \begin{bmatrix} \left.\begin{matrix} x_1 \\ x_3 \end{matrix}\right\} \text{basic} \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ -4 \\ f-4 \end{bmatrix}$$

x_3 free. $\quad$ x_2 basic.

$$\left[\begin{array}{cccc} 0 & 5/2 & 1 & -1/2 \\ 1 & 1/2 & 0 & 1/2 \\ \hline 0 & 1/2 & 0 & -1/2 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \left[\begin{array}{c} 5 \\ 4 \\ \hline f-4 \end{array}\right]$$

$R_1 \to \frac{2}{5} R_1$

$$\left[\begin{array}{cccc} 0 & 1 & 2/5 & -1/5 \\ 1 & 1/2 & 0 & 1/2 \\ \hline 0 & 1/2 & 0 & -1/2 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ f-4 \end{bmatrix}$$

$R_2 \to R_2 - \frac{1}{2} R_1 \quad \sim$

$$\left[\begin{array}{cccc} 0 & 1 & 2/5 & -1/5 \\ 1 & 0 & -1/5 & 3/5 \\ \hline 0 & \underline{\underline{1/2}} & 0 & -1/2 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \left[\begin{array}{c} 2 \\ 3 \\ \hline f-4 \end{array}\right]$$

$R_3 \to R_3 - \frac{1}{2} R_1$

$$\sim \left[\begin{array}{cccc} 0 & 1 & 2/5 & -1/5 \\ 1 & 0 & -1/5 & 3/5 \\ \hline 0 & 0 & -1/5 & -2/5 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \left[\begin{array}{c} 2 \\ 3 \\ \hline f-5 \end{array}\right]$$

$x_3 = x_4 = 0$

$[3, 2, 0, 0] \qquad f = 5$

$x_2 + \boxed{\frac{2}{5} x_3 - \frac{1}{5} x_4} = 2$

$x_1 - \underbrace{\frac{1}{5} x_3 + \frac{3}{5} x_4} = 3$

$f - 5 = -\frac{1}{5} x_3 - \frac{2}{5} x_4.$

$$f = 5 - \frac{1}{5} x_3 - \frac{2}{5} x_4 \qquad f \le 5.$$

- $B \to$ basic variables
- express x_i, $i \in B$ and f in terms of the free variables, x_i, $i \notin B$
- Set $x_i = 0$, $i \notin B$; and calculate f, & x_i ($i \in B$)

$$
\begin{array}{c|cccc|c|c}
 & \multicolumn{2}{c}{\text{(free = 0)}} & \multicolumn{2}{c|}{\text{basic}} & & \\
 & x_1 & x_2 & x_3 & x_4 & & \\
\hline
0 \leftarrow & 1 & 3 & 1 & 0 & 9 & \rho_1 \\
1 \leftarrow & \boxed{2} & 1 & 0 & 1 & 8 & \rho_2 \\
\hline
0 \leftarrow & 1 & 1 & 0 & 0 & 0 & \rho_3
\end{array}
\qquad f = \underline{c}^T \underline{x}
$$

Initial

$$
\begin{array}{c|c}
\bar{a}_{ij} = \bar{A} & \bar{b} \\
\hline
\bar{\underline{c}}^T & \bar{f}
\end{array}
\qquad
\begin{aligned}
\bar{A} &= A \\
\bar{b} &= b \\
\bar{c}^T &= c^T \\
\bar{f} &= 0
\end{aligned}
$$

1. Choose a pivot column.
   Choose a j such that $\bar{c}_j > 0$ [Corresponds to the variable that we want to increase]
2. Choose a pivot row
   Among all i's with $\bar{a}_{ij} > 0$, choose i to minimize $\bar{b}_i / \bar{a}_{ij}$ [gives how much we can increase x_j].

1	3	1	0	9	ρ_1
2	1	0	1	8	ρ_2
1	1	0	0	0	ρ_3

0	$5/2 \to 1$	1	$-1/2$	5	$\rho_1' \to \rho_1 - \rho_2'$
1	$1/2 \to 0$	0	$1/2$	4	$\rho_2' \to \rho_2/2$
0	$1/2 \to 0$	0	$-1/2$	-4	$\rho_3' \to \rho_3 - \rho_2'$

$f - 4$

Pivot column = 2^{nd} column

Pivot row = $\min\left\{ 5/\tfrac{5}{2},\ 4/\tfrac{1}{2} \right\} = \min\{2, 8\}$

x_1	x_2	x_3	x_4	
0	1	$2/5$	$-1/5$	2
1	0	$-1/5$	$3/5$	3
0	0	$-1/5$	$-2/5$	-5

$\rho_1'' = \rho_1' \times \frac{2}{5}$

$\rho_2'' = \rho_2' - \frac{1}{2}\rho_1''$

$f - 5$

$f - 5 = -\frac{1}{5}x_3 - \frac{2}{5}x_4.$

Proof of KKT

Monday, 11 April 2022 2:03 PM

Let $f, h, g \in C^1$

Let $\underline{x}^*$ be a regular point and a local minimizer for the minimization of $f : \mathbb{R}^n \to \mathbb{R}$ s.t. $\overbrace{\underline{h}(x) = 0, \; g(x) \le 0}^{\text{Primal feasibility}}$

Then $\exists \; \underline{\lambda}^* \in \mathbb{R}^m$ and $\underline{\mu}^* \in \mathbb{R}^p$ such that
$(m \le n)$.

$\boxed{1}$ $\underline{\mu}^* \ge 0$ (dual feasibility) → KKT ×

$\boxed{2}$ $Df(\underline{x}^*) + \underbrace{\underline{\lambda}^{*T}}_{\text{Lagrange} \times} D\underline{h}(\underline{x}^*) + \underline{\mu}^{*T} D\underline{g}(\underline{x}^*) = \underline{0}^T$ [Stationarity]

$\boxed{3}$ $\underline{\mu}^{*T} \underline{g}(\underline{x}^*) = 0$ [Complementarity Slack].

Proof: Step 1

$S = \{ \underline{x} : \underline{h}(x) = 0, \; \underline{g}(x) \le 0 \}$

$S_1 = \{ \underline{x} : \underline{h}(x) = 0, \; g_j(\underline{x}) = 0, \; j \in \underbrace{J(x^*)}_{\text{active set.}} \}$

If $\underline{x}^*$ is a local minimizer of f on S, $\underline{x}^*$ is also a local minimizer of f on S_1. Ex. 21.16

[$\underline{x}^*$ is a regular local minimum on a constrained set ⇒ it is also a regular local minimum on a subset defined by setting the active constraints to 0].

Step 2. Lagrange Thm ⇒ $\exists \; \underline{\lambda}^* \in \mathbb{R}^m, \; \underline{\mu}^* \in \mathbb{R}^p$ s.t.

$$Df(\underline{x}^*) + \underline{\lambda}^{*T} D\underline{h}(\underline{x}^*) + \underline{\mu}^{*T} D\underline{g}(\underline{x}^*) = 0^T$$

[Choose $\mu_j^* = 0 \;\; \forall j \notin J(x^*)$]

	Active	Inactive
Constraints	$g_j(x^*) = 0$	$g_j(x^*) < 0$
Multipliers	μ_j^*	$\mu_j^* = 0$

Step gives us the condition 3

Step gives us the condition 3

Multipliers | μ_j^* | $\mu_j^* = 0$

(Lagrange Thm) (Choose).

Step 3:

$$\underline{\mu_1^* g_1(x^*) + \mu_2^* g_2(x^*) + \cdots + \mu_p g_p(x^*) = 0}$$

$\left\{ \begin{array}{l} j \notin J(x^*), \\ \underline{\mu_j^* = 0}; \ \mu_j^* g_j(x^*) = 0 \\ \left. \begin{array}{l} j \in J(x^*), \\ \underline{g_j(x^*) = 0} \end{array} \right\} \\ \mu_j g_j(x^*) = 0 \end{array} \right.$

Condtn 3 also holds.

Step 4: $\mu_j^* \geq 0 \quad j = 1, \cdots p.$

To S.T. $\mu_j^* \geq 0$ for $j \in J(x^*)$. [active sets; $g_j(x^*) = 0$].

Proof by contradiction: If possible, let $\exists\, j \in J(\underline{x}^*)$ for which $\mu_j^* < 0$.

Define $\hat{S} = \{ \underline{x} : \underline{h(\underline{x}) = 0; \ g_i(\underline{x}) = 0}, \ i \in J(x^*), \ i \neq j \}$.

$$\hat{T}(\underline{x}^*) = \{ \underline{y} : Dh(x^*)\, \underline{y} = 0, \ Dg_i(\underline{x}^*)\, y = 0, \ i \in J(x^*), \ i \neq j \}.$$

Step 4a

Claim: $\exists\ \underline{y} \in \hat{T}(\underline{x}^*)$ s.t. $Dg_j(x^*)\, \underline{y} \neq 0$. [$\underline{x}^*$ is a regular point].

Justify: If possible, $\nabla g_j(\underline{x}^*)^T \underline{y} = Dg_j(\underline{x}^*)\, \underline{y} = 0$

$$\nabla g_j(x^*) \in \hat{T}(\underline{x}^*)^{\perp} = \hat{N}(\underline{x}^*)$$
$$= \mathrm{Span}\{ \nabla h_1(x^*), \ldots \nabla h_m(x^*), \nabla g_i(x^*) \ \ i \neq j \}.$$

Contradiction to the fact that $\underline{x}^*$ is a regular point.
[Defn of regular point].

[Def'n of regular point].

Step 4b W.l.o.g. assume that $\boxed{Dg_j(x^*)\, \underline{y} < 0}$ $(\underline{y} \in \hat{T}(x^*))$

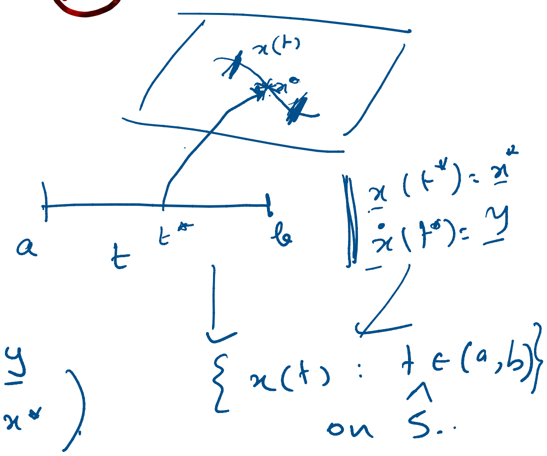
From Lagrange cond'n,

$$Df(x^*)\underline{y} + \underbrace{\lambda^{*T} Dh(x^*)\underline{y}}_{0} + \mu_j^* Dg_j(x^*)\underline{y} + \overbrace{\sum_{i \neq j} \mu_i^* Dg_i(x^*)\underline{y}}^{0} = 0$$

as $\underline{y} \in \hat{T}(x^*)$.

$$\boxed{Df(x^*)\,\underline{y}} = -\overbrace{\mu_j^*}^{<0} \overbrace{Dg_j(x^*)\,\underline{y}}^{<0} < 0 \quad \text{for } \underline{y} \in \hat{T}(x^*) \quad \text{Ⓐ}$$

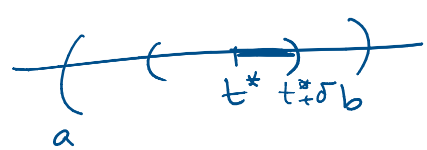
Thm 20.1 Let $x^* \in \hat{S}$ be a regular pt and $\hat{T}(x^*)$ is the tangent space at $\underline{x}^*$. Then $\underline{y} \in \hat{T}(\underline{x}^*) \iff \exists$ a diff. curve in $\hat{S}$ s.t. $\dot{x}(t^*) = \underline{y}$, $x(t^*) = x^*$)



$$\frac{d}{dt}\Big(f(\underbrace{x(t^*)}_{x(t)}) \Big) = \underbrace{Df(x(t^*))^T \overbrace{\dot{x}(t^*)}^{\underline{y}}}_{} < 0$$

$$= Df(x^*)\,\underline{y} < 0 \quad [\text{from Ⓐ}]$$

$\underline{x'(t^*) < 0}$



$\Rightarrow \exists\, \delta > 0$ s.t. $\forall\, t \in (t^*, t^*+\delta)$,

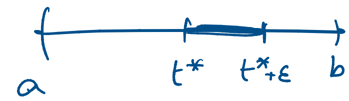
$$f(x(t)) < f(x(t^*)) = f(x^*).$$

$$\frac{d}{dt}\Big(g_j(x(t^*)) \Big) = Dg_j(x^*)\,\underline{y} < 0$$

$\frac{d}{dt}\left(\underline{g_j(x(t))} \right) = \ldots$

$\Rightarrow \exists\ \varepsilon > 0$ and $t \in [t^*, t^* + \varepsilon]$ s.t.

$$g_j(x(t)) < 0$$



$\forall\ t \in [t^*, t^* + \min\{\varepsilon, \delta\}]$,

$$\left.\begin{array}{l} g_j(x(t)) < 0 \\ f(\underline{x}(t)) < f(\underline{x}^*) \end{array}\right]$$

The point $x(t) \in \hat{S}$, $t \in [t^*, t^* + \min\{\varepsilon, \delta\}]$ are feasible points with lower values of the objective function f (than $f(x^*)$); this contradicts the assumption that $\underline{x}^*$ is a local minimizer.

$\Rightarrow$ This assumption $(\mu_j < 0)$ is wrong.

SYMQBD

Tutorial Sheet 5

Q1. Compute two iterations for the minimization of

$$f(x_1, x_2) = x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1^2 + x_2^2 + 3$$

using (i) steepest descent method (ii) Newton's method with starting point $\boldsymbol{x}^{(0)} = \mathbf{0}$. Determine the optimal solution analytically. Compare the rates of convergences.

Q2. The fixed-step-size gradient algorithm defined by

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)})$$

is known to converge iff $0 < \alpha < 2/\lambda_{\max}(\boldsymbol{Q})$. Find the largest ranges of values of α for which the minimization algorithm is globally convergent if:

(i) $f(x_1, x_2) = 3(x_1^2 + x_2^2) + 4x_1x_2 + 5x_1 + 6x_2 + 7$;

(ii) $f(x_1, x_2) = 1 + 2x_1 + 3(x_1^2 + x_2^2) + 4x_1x_2$;

(iii) $f(x_1, x_2) = \boldsymbol{x}^T \begin{bmatrix} 3 & 3 \\ 1 & 3 \end{bmatrix} \boldsymbol{x} + [16, 23]\, \boldsymbol{x} + \pi^2$.

Q3. Let $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ where A is SPD. If $\boldsymbol{x}^{(0)}$ is such that $\boldsymbol{x}^{(0)} - \boldsymbol{x}^*$ is an eigenvector of $\boldsymbol{A}$, then show that the steepest descent method converges in one step.

Q4. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\boldsymbol{x}) = \frac{3}{2}(x_1^2 + x_2^2) + (1 + a)x_1x_2 - (x_1 + x_2) + b,$$

where a and b are some unknown real-valued parameters.

(a) Write the function f in the usual multivariable quadratic form.

(b) Find the largest set of values a and b such that the unique global minimizer of f exists, and write down the minimizer (in terms of the parameters a and b).

(c) Consider the following algorithm:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \frac{2}{5}\nabla f(\boldsymbol{x}^{(k)}).$$

Find the largest set of values of a and b for which this algorithm converges to the global minimizer of f for any initial point $\boldsymbol{x}^{(0)}$.

Q5. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3 - x$. Suppose that we use a fixed-step-size algorithm $x^{(k+1)} = x^{(k)} - \alpha f'(x^{(k)})$ to find a local minimizer of f. Find the largest range of values of α such that the algorithm is locally convergent (i.e., for all x_0 sufficiently close to a local minimizer x^*, we have $x^{(k)} \to x^*$).

Q6. Consider the optimization problem

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2,$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m \geq n$, and $\boldsymbol{b} \in \mathbb{R}^m$.

(a) Show that the objective function for this problem is a quadratic function, and write down the gradient and Hessian of this quadratic.

(b) Write down the fixed-step-size gradient algorithm for solving this optimization problem.

(c) Suppose that

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Find the largest range of values for α such that the algorithm in part **Q6.**b converges to the solution of the problem.

Q7. Consider a function $\boldsymbol{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ given by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$. Suppose that $\boldsymbol{A}$ is invertible and $\boldsymbol{x}^*$ is the zero of $\boldsymbol{f}$ [i.e., $\boldsymbol{f}(\boldsymbol{x}^*) = \boldsymbol{0}$]. We wish to compute $\boldsymbol{x}^*$ using the iterative algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha \boldsymbol{f}(\boldsymbol{x}^{(k)}),$$

where $\alpha \in \mathbb{R}$, $\alpha > 0$. We say that the algorithm is *globally monotone* if for any $\boldsymbol{x}^{(0)}$, $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^*\| \leq \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|$ for all k.

(a) Assume that all the eigenvalues of $\boldsymbol{A}$ are real. Show that a necessary condition for the algorithm above to be *globally monotone* is that all the eigenvalues of $\boldsymbol{A}$ are nonnegative.

Hint: Use contraposition.

(b) Suppose that

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

Find the largest range of values of α for which the algorithm is *globally convergent* (i.e., $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$ for all $\boldsymbol{x}^{(0)}$).

Q8. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be given by $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, where $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{Q}$ is a real symmetric positive definite $n \times n$ matrix. Suppose that we apply the steepest descent method to this function, with $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$. Show that the method converges in one step, that is $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b}$, if and only if $\boldsymbol{x}^{(0)}$ is chosen such that $\boldsymbol{g}^{(0)} = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ is an eigenvector of $\boldsymbol{Q}$.

Solution outlines

A1. Express f in the form $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ for a symmetric matrix $\boldsymbol{Q}$. Verify that $\boldsymbol{Q}$ turns out to be positive definite.

(i) Apply the formula for the method of steepest descent

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{g}^{(k)}$$

using the explicit formula for α_k given as

$$\alpha_k = \frac{\left(\boldsymbol{g}^{(k)}\right)^T \boldsymbol{g}^{(k)}}{\left(\boldsymbol{g}^{(k)}\right)^T \boldsymbol{Q}\boldsymbol{g}^{(k)}},$$

where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$.

You should get the step sizes as $\alpha_0 = \frac{5}{6}$ and $\alpha_1 = \frac{5}{9}$. The result of two iterations should be $\boldsymbol{x}^{(2)} = [-\frac{25}{27}, -\frac{25}{108}]^T$.

(ii) Use the formula for Newton's method

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)},$$

where $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$.

The result of two iterations should be $\boldsymbol{x}^{(2)} = [-1, -\frac{1}{4}]^T$.

The optimal solution can be found analytically by solving $\boldsymbol{Q}\boldsymbol{x}^* = \boldsymbol{b}$. (Justify!) Hence, $\boldsymbol{x}^* = [-1, -\frac{1}{4}]^T$.

A2. In each case, write f in the "standard form" $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$ for a symmetric matrix $\boldsymbol{Q}$, and compute the eigenvalues of $\boldsymbol{Q}$.

(i) $0 < \alpha < \frac{1}{5}$.

(ii) $0 < \alpha < \frac{1}{5}$.

(iii) $0 < \alpha < \frac{1}{5}$.

A3. $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x} + c$.

Given: $\boldsymbol{A}(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*) = \lambda(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*)$ for some $\lambda > 0$.

Steepest descent algorithm: $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})$.

To prove: $\boldsymbol{x}^{(1)} = \boldsymbol{x}^*$.

Proof: Note that $\boldsymbol{g}^{(0)} = \boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b}$ and $\boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}$. (Why? Justify!) So,

$$
\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{A}\boldsymbol{g}^{(0)}}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{A}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \quad \text{(subtract } \boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} \text{ in } \boldsymbol{g}^{(0)}\text{)} \\
&= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\lambda\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{A}[\boldsymbol{x}^{(0)} - \boldsymbol{x}^*]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \\
&= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\lambda\left(\boldsymbol{g}^{(0)}\right)^T [\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}]}[\boldsymbol{A}\boldsymbol{x}^{(0)} - \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b}] \quad \text{(add and subtract } \boldsymbol{b}\text{)} \\
&= \boldsymbol{x}^{(0)} - \frac{1}{\lambda}\boldsymbol{A}[\boldsymbol{x}^{(0)} - \boldsymbol{x}^*] \quad \text{(subtract } \boldsymbol{0} = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} \text{ in } \boldsymbol{g}^{(0)}\text{)} \\
&= \boldsymbol{x}^{(0)} - [\boldsymbol{x}^{(0)} - \boldsymbol{x}^*]
\end{aligned}
$$

A4. (a)

$$
f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 3 & 1+a \\ 1+a & 3 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b.
$$

(b) Unique global minimizer exists iff Hessian is positive definite. Use Sylvester's criterion for positive definiteness to find the required range of a and b. Final answer: $a \in (-4, 2)$ and $b \in (-\infty, \infty)$.

(c) This is a fixed-step-size gradient algorithm, so the algorithm is globally convergent iff $\frac{2}{5} < \frac{2}{\lambda_{\max}(\boldsymbol{Q})}$. Compute the eigenvalues of $\boldsymbol{Q}$ to find the required range of a and b. Final answer: $a \in (-3, 1)$ and $b \in (-\infty, \infty)$.

A5. Firstly, show that the only local minimizer of f is $x^* = 1/\sqrt{3}$. Then, note that to check for "local convergence", we may linearize the given fixed-step-size algorithm (since x_0 is assumed to be sufficiently close to x^*) as:

$$
x^{(k+1)} = x^{(k)} - \alpha f''(x^*)(x^{(k)} - x^*).
$$

Now, notice that this is nothing but the fixed-step-size algorithm applied to a quadratic with second derivative $f''(x^*)$. (Think about this!)

Since $f''(x^*) = 2\sqrt{3}$, the algorithm is locally convergent for $0 < \alpha < \frac{2}{2\sqrt{3}}$.

A6. (a) *Hint:* $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$. Final answers: $\nabla f(\boldsymbol{x}) = 2(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{x} - 2(\boldsymbol{A}^T\boldsymbol{b})$ and $\boldsymbol{F}(\boldsymbol{x}) = 2(\boldsymbol{A}^T\boldsymbol{A})$.

(b) $x^{(k+1)} = x^{(k)} - 2\alpha\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b})$.

(c) $0 < \alpha < \frac{2}{\lambda_{\max}(\boldsymbol{F})} = \frac{1}{4}$.

A7. (a) Suppose $\boldsymbol{Av} = \lambda \boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{R}^n$ and $\lambda < 0$. Choose $\boldsymbol{x}^{(0)} = \boldsymbol{x}^* + \boldsymbol{v}$. Then,

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^* + \boldsymbol{v} - \alpha(\boldsymbol{A}(\boldsymbol{x}^* + \boldsymbol{v}) + \boldsymbol{b}) \implies \boldsymbol{x}^1 - \boldsymbol{x}^* = (1 - \alpha\lambda)(\boldsymbol{x}^{(0)} - \boldsymbol{x}^*).$$

Since $1 - \alpha\lambda > 1$, the algorithm is not globally monotone.

(b) Observe that the given iterative algorithm is identical to a fixed-step-size algorithm for a quadratic whose Hessian is $\boldsymbol{A}$. Hence, the required range of α is $0 < \alpha < \frac{2}{\lambda_{\max}(\boldsymbol{A})} = \frac{2}{5}$.

A8. Given: $\boldsymbol{x}^{(0)} \neq \boldsymbol{Q}^{-1}\boldsymbol{b}$.

To prove: $\boldsymbol{x}^{(1)} = \boldsymbol{Q}^{-1}\boldsymbol{b} \iff \boldsymbol{Q}\boldsymbol{g}^{(0)} = \lambda \boldsymbol{g}^{(0)}$ for some $\lambda \in \mathbb{R}$.

Proof of ($\implies$):

$$\begin{aligned}
& \boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \alpha_0 \boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{Q}^{-1}\boldsymbol{b} = \boldsymbol{x}^{(0)} - \alpha_0 \boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{b} = \boldsymbol{Q}\boldsymbol{x}^{(\mathbf{0})} - \alpha_0 \boldsymbol{Q}\boldsymbol{g}^{(0)} \\
\implies \quad & \alpha_0 \boldsymbol{Q}\boldsymbol{g}^{(0)} = \boldsymbol{g}^{(0)} \\
\implies \quad & \boldsymbol{Q}\boldsymbol{g}^{(0)} = \frac{1}{\alpha_0}\boldsymbol{g}^{(0)}.
\end{aligned}$$

Call $\frac{1}{\alpha_0} = \lambda$. (Why is $\alpha_0 \neq 0$? Justify!)

Proof of ($\impliedby$):

$$\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{x}^{(0)} - \alpha_0 \boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{Q}\boldsymbol{g}^{(0)}} \boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(0)} - \frac{\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}}{\lambda\left(\boldsymbol{g}^{(0)}\right)^T \boldsymbol{g}^{(0)}} \boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(\mathbf{0})} - \frac{1}{\lambda}\boldsymbol{g}^{(0)} \\
&= \boldsymbol{x}^{(\mathbf{0})} - \boldsymbol{Q}^{-1}\boldsymbol{g}^{(0)}.
\end{aligned}$$

Tutorial Sheet 6

Q1. Let $\{\boldsymbol{x}^{(k)}\}$ be the sequence generated by Newton's method for minimizing a given objective function $f(\boldsymbol{x})$. Show that if the Hessian $\boldsymbol{F}(\boldsymbol{x}^{(k)}) > 0$ and $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)}) \neq \boldsymbol{0}$, then the search direction

$$\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$$

from $\boldsymbol{x}^{(k)}$ to $\boldsymbol{x}^{(k+1)}$ is a descent direction for f in the sense that there exists an $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}).$$

Q2. Show that in the conjugate gradient algorithm,

$$\boldsymbol{g}^{(k+1)^T} \boldsymbol{d}^{(i)} = 0$$

for all k, $0 \leq k \leq n-1$, and $0 \leq i \leq k$.

Q3. Represent the function

$$f(x_1, x_2) = \frac{5}{2}x_1^2 + x_2^2 - 3x_1x_2 - x_2 - 7$$

in the form $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} + c$. Then use the *conjugate gradient algorithm* to construct a vector $\boldsymbol{d}^{(1)}$ that is $\boldsymbol{Q}$-conjugate with $\boldsymbol{d}^{(0)} = \nabla f(\boldsymbol{x}^{(0)})$, where $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.

Q4. Let $f(\boldsymbol{x})$, $\boldsymbol{x} = [x_1, x_2]^T \in \mathbb{R}^2$, be given by

$$f(\boldsymbol{x}) = \frac{5}{2}x_1^2 + \frac{1}{2}x_2^2 + 2x_1x_2 - 3x_1 - x_2.$$

a. Express $f(\boldsymbol{x})$ in the form of $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$.

b. Find the minimizer of f using the conjugate gradient algorithm. Use a starting point of $\boldsymbol{x}^{(0)} = [0, 0]^T$.

c. Calculate the minimizer of f analytically from $\boldsymbol{Q}$ and $\boldsymbol{b}$, and check it with your answer in part b..

Q5. Given $f\colon \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, consider the algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)},$$

where $\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots$ are vectors in $\mathbb{R}^n$, and $\alpha_k \geq 0$ is chosen to minimize $f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)})$; that is,

$$\alpha_k = \arg\min_{\alpha \geq 0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}).$$

Note that the general algorithm encompasses almost all algorithms that we discussed in this part, including the steepest descent, Newton, conjugate gradient, and quasi-Newton algorithms.

Let $\boldsymbol{g}^{(k)} = \nabla f(\boldsymbol{x}^{(k)})$, and assume that $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$.

a. Show that $\boldsymbol{d}^{(k)}$ is a descent direction for f in the sense that there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$,

$$f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{d}^{(k)}) < f(\boldsymbol{x}^{(k)}).$$

b. Show that $\alpha_k > 0$.

c. Show that $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k+1)} = 0$.

d. Show that the following algorithms all satisfy the condition $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$:

1. Steepest descent algorithm.
2. Newton's method, assuming that the Hessian is positive definite.
3. Conjugate gradient algorithm.

4. Quasi-Newton algorithm, assuming that $\boldsymbol{H}_k > 0$.

e. For the case where $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b}$, with $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$, derive an expression for α_k in terms of $\boldsymbol{Q}$, $\boldsymbol{d}^{(k)}$, and $\boldsymbol{g}^{(k)}$.

Q6. Minimize the function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \boldsymbol{x} - \boldsymbol{x}^T \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 7$$

using a quasi-Newton method with the starting point $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.

Solution outlines

For **A1**–**A4**, see the handwritten slides titled "Solution outlines to Tutorial 6".

A5. (a) $\phi'(\alpha) = \boldsymbol{d}^{(k)^T}\nabla f(\boldsymbol{x}^{(k)} + \alpha\boldsymbol{d}^{(k)})$.
Hence, $\phi'(0) = \boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)}$.
Since ϕ' is continuous, if $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)} < 0$, then there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, $\phi(\alpha) < \phi(0)$.

(b) From part (a), $\phi(\alpha) < \phi(0)$ for all $\alpha \in (0, \bar{\alpha}]$.
Hence, $\alpha_k = \arg\min \phi(\alpha) \neq 0$, implying that $\alpha_k > 0$.

(c) $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k+1)} = \boldsymbol{d}^{(k)^T}\nabla f(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) = \phi'(\alpha_k)$. Since $\alpha_k > 0$, we have $\phi'(\alpha_k) = 0$.

(d) i) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)}$. Hence $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\|\boldsymbol{g}^{(k)}\|^2$. If $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\|\boldsymbol{g}^{(k)}\|^2 > 0$, and hence proved.

ii) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)}$. Since $\boldsymbol{F}(\boldsymbol{x}^{(k)}) > 0$, we have $\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1} > 0$. Therefore, $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\boldsymbol{g}^{(k)^T}\boldsymbol{F}(\boldsymbol{x}^{(k)})^{-1}\boldsymbol{g}^{(k)} < 0$ if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$.

iii) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{g}^{(k)} + \beta_{k-1}\boldsymbol{d}^{(k-1)}$.
Hence, $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k = -\|\boldsymbol{g}^{(k)}\|^2 + \beta_{k-1}\boldsymbol{d}^{(k-1)}\boldsymbol{g}^{(k)^T}$.
By part (c), $\boldsymbol{d}^{(k-1)^T}\boldsymbol{g}^{(k)} = 0$. Hence, if $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$ and $\|\boldsymbol{g}^{(k)}\| > 0$, then $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k < 0$.

iv) We have $\boldsymbol{d}^{(k)} = -\boldsymbol{H}_k\boldsymbol{g}^{(k)}$. Therefore, if $\boldsymbol{H} > 0$ and $\boldsymbol{g}^{(k)} \neq \boldsymbol{0}$, then $\boldsymbol{d}^{(k)^T}\boldsymbol{g}^k < 0$.

(e) Using the equation $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{b}$, we get

$$\begin{aligned}
\boldsymbol{d}^{(k)^T}\boldsymbol{g}^{k+1} &= \boldsymbol{d}^{(k)^T}(\boldsymbol{Q}\boldsymbol{x}^{(k+1)} - \boldsymbol{b}) \\
&= \boldsymbol{d}^{(k)^T}(\boldsymbol{Q}(\boldsymbol{x}^{(k)} + \alpha_k\boldsymbol{d}^{(k)}) - \boldsymbol{b}) \\
&= \alpha\boldsymbol{d}^{(k)^T}\boldsymbol{Q}\boldsymbol{d}^{(k)} + \boldsymbol{d}^{(k)^T}\boldsymbol{g}^{(k)}.
\end{aligned}$$

Now use part (c) to get the value of α.

A6. Using the rank-one correction method, we obtain the following steps:

$$\nabla f(\boldsymbol{x}^{(0)}) = \boldsymbol{Q}\boldsymbol{x}^{(0)} - \boldsymbol{b} = [-1, 1]^T$$
$$\boldsymbol{d}^{(0)} = -\boldsymbol{H}_0\boldsymbol{g}^{(0)} = [1, -1]^T$$
$$\alpha_0 = -\frac{\boldsymbol{g}^{(0)T}\boldsymbol{d}^{(0)}}{\boldsymbol{d}^{(0)T}\boldsymbol{Q}\boldsymbol{d}^{(0)}} = \frac{2}{3}$$
$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \alpha_0\boldsymbol{d}^{(0)} = [\tfrac{2}{3}, -\tfrac{2}{3}]^T$$
$$\nabla f(\boldsymbol{x}^{(1)} = [\tfrac{-1}{3}, \tfrac{-1}{3}]^T$$
$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$
$$\boldsymbol{d}^{(1)} = [\tfrac{1}{3}, \tfrac{1}{6}]^T$$
$$\alpha_1 = 1$$
$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^* = [1, \tfrac{-1}{2}]^T$$

(Try to also apply a rank-two correction method instead, and compare with the above.)

Tutorial Sheet 7

Q1. Show that in the BFGS method when one chooses $\boldsymbol{u} = \Delta \boldsymbol{g}^{(k)}$ and $\boldsymbol{v} = \boldsymbol{B}_k \Delta \boldsymbol{x}^{(k)}$, the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent.

Q2. Verify that

$$\boldsymbol{B}_{k+1}^{-1} = \left(\boldsymbol{I} - \frac{(\Delta \boldsymbol{x}^{(k)})(\Delta \boldsymbol{g}^{(k)})^T}{(\Delta \boldsymbol{g}^{(k)})^T(\Delta \boldsymbol{x}^{(k)})}\right) \boldsymbol{B}_k^{-1} \left(\boldsymbol{I} - \frac{(\Delta \boldsymbol{g}^{(k)})(\Delta \boldsymbol{x}^{(k)})^T}{(\Delta \boldsymbol{g}^{(k)})^T(\Delta \boldsymbol{x}^{(k)})}\right) + \frac{(\Delta \boldsymbol{x}^{(k)})(\Delta \boldsymbol{x}^{(k)})^T}{(\Delta \boldsymbol{g}^{(k)})^T(\Delta \boldsymbol{x}^{(k)})}$$

by plugging the values (here, the 'k' is removed for notational convenience)

$$\begin{aligned}
\boldsymbol{A} &= \boldsymbol{B} \\
\boldsymbol{U} &= \begin{bmatrix} \boldsymbol{B}\Delta \boldsymbol{x} & \Delta \boldsymbol{g} \end{bmatrix} \\
\boldsymbol{C} &= \begin{bmatrix} \frac{-1}{(\Delta \boldsymbol{x})^T \boldsymbol{B}(\Delta \boldsymbol{x})} & 0 \\ 0 & \frac{1}{(\Delta \boldsymbol{g})^T \Delta \boldsymbol{x}} \end{bmatrix} \\
\boldsymbol{V} &= \begin{bmatrix} (\Delta \boldsymbol{x})^T \boldsymbol{B} \\ (\Delta \boldsymbol{g})^T \end{bmatrix}
\end{aligned}$$

into the Woodbury formula

$$(\boldsymbol{A} + \boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1} + \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.$$

Q3. Show that rank-two updates preserve positive definiteness. *Hint:* Write the expression for $\boldsymbol{B}_{k+1}^{-1}$ in **Q2.** as $\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{P} + \boldsymbol{Q}$. Then, assuming that $\boldsymbol{B}_k^{-1} > 0$, one needs to show that $\boldsymbol{B}_{k+1}^{-1} > 0$. Use the definition of positive definiteness to analyze the expressions $\boldsymbol{z}^T\boldsymbol{P}\boldsymbol{z}$ and $\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z}$ separately, and put together all this information to deduce the positive definiteness of $\boldsymbol{B}_{k+1}^{-1}$.

Q4. Show that in the BFGS algorithm applied to a quadratic with Hessian $\boldsymbol{Q} = \boldsymbol{Q}^T$, we have $\boldsymbol{B}_{k+1}\Delta \boldsymbol{x}^{(i)} = \Delta \boldsymbol{g}^{(i)}$ for all $0 \leq i \leq k$. *Hint:* Use induction.

Q5. Use the BFGS method to minimize

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{b} + \pi^2,$$

where

$$\boldsymbol{Q} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Take $\boldsymbol{H}_0 = \boldsymbol{I}_2$ and $\boldsymbol{x}^{(0)} = [0, 0]^T$. Verify that $\boldsymbol{H}_2 = \boldsymbol{Q}^{-1}$.

Q6. Find local extremizers for the following optimization problems:

(a)

$$\begin{aligned}
\text{Minimize} \quad & x_1^2 + 2x_1x_2 + 3x_2^2 + 4x_1 + 5x_2 + 6x_3 \\
\text{subject to} \quad & x_1 + 2x_2 = 3 \\
& 4x_1 + 5x_3 = 6.
\end{aligned}$$

(b)

$$\begin{aligned}
\text{Maximize} \quad & 4x_1 + x_2^2 \\
\text{subject to} \quad & x_1^2 + x_2^2 = 9.
\end{aligned}$$

(c)

$$\begin{aligned} \text{Maximize} \quad & x_1x_2 \\ \text{subject to} \quad & x_1^2 + 4x_2^2 = 1. \end{aligned}$$

Q7. Find minimizers and maximizers of the function

$$f(\boldsymbol{x}) = (\boldsymbol{a}^T\boldsymbol{x})(\boldsymbol{b}^T\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^3,$$

subject to

$$\begin{aligned} x_1 + x_2 &= 0 \\ x_2 + x_3 &= 0, \end{aligned}$$

where

$$\boldsymbol{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Q8. We wish to construct a closed box with minimum surface area that encloses a volume V cubic feet, where $V > 0$.

(a) Let a, b, and c denote the dimensions of the box with minimum surface area (with volume V). Derive the Lagrangian condition that must be satisfied by a, b, and c.

(b) What does it mean for a point $\boldsymbol{x}^*$ to be a *regular* point in this problem?

(c) Find a, b, and c.

(d) Does the point $\boldsymbol{x}^* = [a, b, c]^T$ found in **Q8.**c satisfy the second-order sufficient condition?

Q9. Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_1x_2 - 2x_1, \qquad x_1, x_2 \in \mathbb{R} \\ \text{subject to} \quad & x_1^2 - x_2^2 = 0. \end{aligned}$$

(a) Apply Lagrange's theorem directly to the problem to show that if a solution exists, then it must be either $[1,1]^T$ or $[-1,1]^T$.

(b) Use the second-order necessary conditions to show that $[-1,1]^T$ cannot possibly be the solution.

(c) Use the second-order sufficient conditions to show that $[1,1]^T$ is a strict local minimizer.

Solution outlines

A4. Follow the proof of Theorem 11.3 in [CZ13] given on pages 203–204, taking $\boldsymbol{H}_{k+1} = \boldsymbol{B}_{k+1}^{-1}$.

A5. This is Example 11.4 in [CZ13] given on pages 209–211.

A6. (a) $l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T\boldsymbol{h}(\boldsymbol{x})$ is the lagrangian. We now find critical points by solving following equations:

$$\begin{aligned} \nabla f(\boldsymbol{x}) + \boldsymbol{\lambda}^T\nabla \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0} \\ \boldsymbol{h}(\boldsymbol{x}) &= \boldsymbol{0}^T. \end{aligned}$$

The unique solution is $\boldsymbol{x}^* = [\frac{16}{5}, \frac{-1}{10}, \frac{-34}{25}]^T$ and $\boldsymbol{\lambda}^* = [\frac{-27}{5}, \frac{-6}{5}]^T$.

The Hessian of the lagrangian is: $\begin{pmatrix} 2 & 2 & 0 \\ 2 & 6 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

$T(\boldsymbol{x^*}) = \{a[\frac{-5}{4}, \frac{5}{8}, 1]^T : a \in \mathbb{R}\}$. $\therefore$ $\boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda})\boldsymbol{y} = \frac{75}{32}a^2 > 0$ for all $a \neq 0$. Hence, $\boldsymbol{x}^*$ is a local minimizer.

(b) Proceed exactly as in the previous problem. The four points satisfying the Lagrange condition are:

$$\boldsymbol{x}^{(1)} = [3,0]^T, \lambda^{(1)} = \tfrac{-2}{3};$$
$$\boldsymbol{x}^{(2)} = [-3,0]^T, \lambda^{(2)} = \tfrac{2}{3};$$
$$\boldsymbol{x}^{(3)} = [2,\sqrt{5}]^T, \lambda^{(3)} = -1;$$
$$\boldsymbol{x}^{(4)} = [2,-\sqrt{5}]^T, \lambda^{(4)} = -1.$$

For the first point, we have $\boldsymbol{L}(\boldsymbol{x}^{(1)}, \lambda^{(1)}) = \begin{pmatrix} \frac{-4}{3} & 0 \\ 0 & \frac{2}{3} \end{pmatrix}$.
$T(\boldsymbol{x}^{(1)}) = \{a[0,1]^T : a \in \mathbb{R}\}$. $\therefore$ $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^{(1)}, \lambda^{(1)})\boldsymbol{y} > 0$ for all nonzero $\boldsymbol{y} \in T(\boldsymbol{x}^{(1)})$.
Similarly, do for the remaining 3 points:
$\boldsymbol{x}^{(2)}$ is a strict local minimizer;
$\boldsymbol{x}^{(3)}$ is a strict local maximizer;
$\boldsymbol{x}^{(4)}$ is a strict local maximizer.

(c) The four points satisfying the Lagrange condition are:

$$\boldsymbol{x}^{(1)} = [\tfrac{1}{\sqrt{2}}, -\tfrac{1}{2\sqrt{2}}]^T, \lambda^{(1)} = \tfrac{1}{4}$$
$$\boldsymbol{x}^{(2)} = [\tfrac{-1}{\sqrt{2}}, \tfrac{1}{2\sqrt{2}}]^T, \lambda^{(2)} = \tfrac{1}{4}$$
$$\boldsymbol{x}^{(3)} = [\tfrac{1}{\sqrt{2}}, \tfrac{1}{2\sqrt{2}}]^T, \lambda^{(3)} = \tfrac{-1}{4}$$
$$\boldsymbol{x}^{(4)} = [\tfrac{-1}{\sqrt{2}}, -\tfrac{1}{2\sqrt{2}}]^T, \lambda^{(4)} = \tfrac{-1}{4}$$

After calculations of SOSC we realize that first two points are strict local maximizers and last two are strict local minimizers.

A7. $l(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{a}^T\boldsymbol{x}\boldsymbol{b}^T\boldsymbol{x} + \lambda_1(x_1 + x_2) + \lambda_2(x_2 + x_3)$

$\nabla f(\boldsymbol{x}) + \boldsymbol{\lambda}^T\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} x_2 + \lambda_1 \\ x_1 + x_3 + \lambda_1 + \lambda_2 \\ x_2 + \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

$\boldsymbol{h}(\boldsymbol{x}) = \begin{pmatrix} x_1 + x_2 \\ x_2 + x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\boldsymbol{x}^* = 0$, $\boldsymbol{\lambda}^* = 0$ satisfy the Lagrange, FONC conditions.

The Hessian of the lagrangian is: $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

Tangent space: $\{\boldsymbol{y} : \boldsymbol{y} = a[1,-1,1]^T\}$.
$\boldsymbol{y}^T\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda})\boldsymbol{y} = -4a^2 < 0$. Hence, the critical point satisfies SOSC and is a strict local minimizer.

A8. (a) Let x_1, x_2, x_3 be the dimensions of the box. The problem is :

$$\begin{aligned} \text{minimize} \quad & 2(x_1x_2 + x_2x_3 + x_1x_3 \\ \text{subject to} \quad & x_1x_2x_3 = V \end{aligned}$$

The dimensions of the box with minimum surface area say [a,b,c] satisfies:

$$\begin{aligned} 2(b+c) + \lambda bc &= 0 \\ 2(a+c) + \lambda ac &= 0 \\ 2(a+b) + \lambda ab &= 0 \\ abc - V &= 0 \end{aligned}$$

where $\lambda \in \mathbb{R}$.

(b) $a, b, c \neq 0$ (Why?)

(c) Multiply the first equation by a and second equation by b, and subtracting the first from the second, we obtain $c(a-b) = 0$ implying $a = b$. By a similar procedure we can conclude that $a = b = c = V^{\frac{1}{3}}$ and $\lambda = -4V^{\frac{-1}{3}}$.

(d) $\boldsymbol{L(x^*,}\lambda) = -2\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$.

$T(\boldsymbol{x}^*) = (\boldsymbol{y} : y_3 = -(y_1 + y_2))$

$\boldsymbol{y^T L(x^*, \lambda) y} > 0$. Hence SOSC is satisfied.

A9. (a) Denote the solution by $[x_1^*, x_2^*]$. The Lagrange condition for this problem has the form:

$$\begin{aligned} x_2^* - 2 + 2\lambda^* x_1^* &= 0 \\ x_1^* - 2\lambda^* x_2^* &= 0 \\ (x_1^*)^2 - (x_2^*) &= 0. \end{aligned}$$

Note that $x_1^*, x_2^* \neq 0$.

Combining the first and second equations we obtain $\lambda^* = \frac{2-x_2^*}{2x_1^*} = \frac{x_1^*}{2x_2^*}$.

Hence, $x_2^* = 1$, and hence $x_1^* = 1$. Thus, the only two points satisfying Lagrange condition are $[1, 1]^T$, $[-1, 1]^T$.

(b) $\boldsymbol{y^T} L(\boldsymbol{x^*, \lambda^*})\boldsymbol{y} = -2a^2 < 0$. Hence, not a local minimizer.

(c) $\boldsymbol{y^T} L(\boldsymbol{x^*, \lambda^*})\boldsymbol{y} = 2a^2 > 0$ Hence it is a strict local minimizer.

Tutorial Sheet 8

Q1. Consider the problem

$$\begin{aligned} \text{maximize} \quad & ax_1 + bx_2, \qquad x_1, x_2 \in \mathbb{R} \\ \text{subject to} \quad & x_1^2 + x_2^2 = 2, \end{aligned}$$

where $a, b \in \mathbb{R}$. Show that if $[1, 1]^T$ is a solution to the problem, then $a = b$.

Q2. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m \leq n$, $\text{rank}(\boldsymbol{A}) = m$, and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Let $\boldsymbol{x}^*$ be the point on the nullspace of $\boldsymbol{A}$ that is closest to $\boldsymbol{x}_0$ (in the sense of Euclidean norm).

(a) Show that $\boldsymbol{x}^*$ is orthogonal to $\boldsymbol{x}^* - \boldsymbol{x}_0$.

(b) Find a formula for $\boldsymbol{x}^*$ in terms of $\boldsymbol{A}$ and $\boldsymbol{x}_0$.

Q3. Consider the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{subject to} \quad & \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m > n$, $\boldsymbol{C} \in \mathbb{R}^{p \times n}$, $p < n$, and both $\boldsymbol{A}$ and $\boldsymbol{C}$ are of full rank. We wish to find an expression for the solution (in terms of $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{C}$, and $\boldsymbol{d}$).

Apply Lagrange's theorem to solve this problem.

Q4. Find local extremizers for:

(a) $x_1^2 + x_2^2 - 2x_1 - 10x_2 + 26$ subject to $\frac{1}{5}x_2 - x_1^2 \leq 0$, $5x_1 + \frac{1}{2}x_2 \leq 5$.

(b) $x_1^2 + x_2^2$ subject to $x_1 \geq 0$, $x_2 \geq 0$, $x_1 + x_2 \geq 5$.

(c) $x_1^2 + 6x_1x_2 - 4x_1 - 2x_2$ subject to $x_1^2 + 2x_2 \leq 1$, $2x_1 - 2x_2 \leq 1$.

Q5. Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_2 \\ \text{subject to} \quad & x_2 \geq -(x_1 - 1)^2 + 3. \end{aligned}$$

(a) Find all points satisfying the KKT condition for the problem.

(b) For each point $\boldsymbol{x}^*$ in **Q5.**a, find $T(\boldsymbol{x}^*)$, $N(\boldsymbol{x}^*)$, and $\tilde{T}(\boldsymbol{x}^*, \mu^*)$.

(c) Find the subset of points from **Q5.**a that satisfy the second-order necessary condition.

Q6. Consider the problem of optimizing (either minimizing or maximizing) $(x_1 - 2)^2 + (x_2 - 1)^2$ subject to

$$\begin{aligned} x_2 - x_1^2 &\geq 0 \\ 2 - x_1 - x_2 &\geq 0 \\ x_1 &\geq 0. \end{aligned}$$

The point $\boldsymbol{x}^* = \boldsymbol{0}$ satisfies the KKT conditions.

(a) Does $\boldsymbol{x}^*$ satisfy the FONC for minimization or maximization? What are the KKT multipliers?

(b) Does $\boldsymbol{x}^*$ satisfy the SOSC? Carefully justify your answer.

Q7. Let $g : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$ be given, where $g(\boldsymbol{x}_0) > 0$. Consider the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ \text{subject to} \quad & g(\boldsymbol{x}) \leq 0. \end{aligned}$$

Suppose that $\boldsymbol{x}^*$ is a solution to the problem and $g \in \mathcal{C}^1$. Use the KKT theorem to decide which of the following equations/inequalities hold:

(i) $g(\boldsymbol{x}^*) < 0$.

(ii) $g(\boldsymbol{x}^*) = 0$.

(iii) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) < 0$.

(iv) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) = 0$.

(v) $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) > 0$.

Q8. Consider the constraint set $S = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$. Let $\boldsymbol{x}^* \in S$ be a regular local minimizer of f over S and $J(\boldsymbol{x}^*)$ the index set of active inequality constraints. Show that $\boldsymbol{x}^*$ is also a regular local minimizer of f over the set $S' = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, g_j(\boldsymbol{x}) = 0, j \in J(\boldsymbol{x}^*)\}$.

Hints:

(1) If g_i is not an active inequality constraint, then what can you say about $g_i(\boldsymbol{x}^*)$? By the continuity of g_i, can you say something about $g_i(\boldsymbol{x})$ for $\boldsymbol{x}$ close enough to $\boldsymbol{x}^*$?

(2) Let B be a ball of radius $\epsilon > 0$ centered at $\boldsymbol{x}^*$, where the choice of ϵ comes from the previous hint. Consider the constraint set $S_1 = \{\boldsymbol{x} : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, g_j(\boldsymbol{x}) \leq 0, j \in J(\boldsymbol{x}^*)\}$. Show that $S \cap B = S_1 \cap B$. (To show this, you need to prove that $S \cap B \subseteq S_1 \cap B$ as well as $S \cap B \supseteq S_1 \cap B$.) This means that "locally" near $\boldsymbol{x}^*$ the constraint set S looks the same as the constraint set S_1.

(3) Is $\boldsymbol{x}^*$ a regular local minimizer of f on S_1? Argue carefully.

(4) Note that $S' \subseteq S_1$ and $\boldsymbol{x}^* \in S'$. Conclude.

Q9. Consider the problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^T \boldsymbol{x} + 8 \\ \text{subject to} & \frac{1}{2}\|\boldsymbol{x}\|^2 \leq 1, \end{array}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{c} \neq \boldsymbol{0}$. Suppose that $\boldsymbol{x}^* = \alpha \boldsymbol{e}$ is a solution to the problem, where $\alpha \in \mathbb{R}$ and $\boldsymbol{e} = [1, \ldots, 1]^T$, and the corresponding objective value is 4.

(a) Show that $\|\boldsymbol{x}^*\|^2 = 2$.

(b) Find α and $\boldsymbol{c}$ (they may depend on n).

Q10. Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exists $\boldsymbol{x}^* \in \Omega$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T.$$

Show that $\boldsymbol{x}^*$ is a global minimizer of f over Ω.

Hints:

(1) Recall that for a convex function $f : \Omega \to \mathbb{R}$, we have $f(\boldsymbol{x}) \geq f(\boldsymbol{y}) + Df(\boldsymbol{y})(\boldsymbol{x} - \boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Omega$.
Apply this to the above scenario when $\boldsymbol{y} = \boldsymbol{x}^*$, and express the Df term in terms of $D\boldsymbol{h}$.

(2) To examine the $D\boldsymbol{h}$ term more closely, you need to use the definition of directional derivative. So, start by writing down the expression $\boldsymbol{h}(\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*)) - \boldsymbol{h}(\boldsymbol{x}^*)$.
Use the convexity and definition of Ω to say something about this expression.
Then, left-multiply by $\boldsymbol{\lambda}^{*T}$, divide by α, and take the limit as $\alpha \to 0$. Conclude.

Q11. Let $f : \mathbb{R}^n \to \mathbb{R}$, $f \in \mathcal{C}^1$, be a convex function on the set of feasible points

$$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\},$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$, $\boldsymbol{h}, \boldsymbol{g} \in \mathcal{C}^1$, and Ω is convex. Suppose that there exist $\boldsymbol{x}^* \in \Omega$, $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, and $\boldsymbol{\mu}^* \in \mathbb{R}^p$, such that

1. $\boldsymbol{\mu}^* \geq \boldsymbol{0}$.
2. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) + \boldsymbol{\mu}^{*T} D\boldsymbol{g}(\boldsymbol{x}^*) = \boldsymbol{0}^T$.
3. $\boldsymbol{\mu}^{*T} \boldsymbol{g}(\boldsymbol{x}^*) = 0$.

Show that $\boldsymbol{x}^*$ is a global minimizer of f over Ω.

Hints:

(1) Use the convexity of f similarly to the previous exercise. This time, you should be able to rewrite the Df expression in terms of $D\boldsymbol{h}$ and $D\boldsymbol{g}$.

(2) Similar to the previous exercise, show that the expression involving $D\boldsymbol{h}$ evaluates to zero.

(3) Similar to the method for handling the $D\boldsymbol{h}$ expression, can you write down the relevant expression that needs to be examined in order to examine the $D\boldsymbol{g}$ expression? Use the convexity of Ω and the third condition given in the problem. Conclude.

Solution outlines

A1. Verify that $[1, 1]^T$ is a regular point, so that one can apply the Lagrange multiplier theorem at this point to deduce that $a = b$.

A2. Observe that $\boldsymbol{x}^*$ is a solution to the problem

$$\begin{aligned} \text{minimize} \quad & \|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}. \end{aligned}$$

(a) Since $\operatorname{rank}(\boldsymbol{A}) = m$, every feasible point is also regular. Write down the Lagrange multiplier condition for $\boldsymbol{x}^*$. Multiply by $\boldsymbol{x}^*$ appropriately. Conclude.

(b) The Lagrange multiplier condition implies that $\boldsymbol{x}^* - \boldsymbol{x}_0 = \boldsymbol{A}^T \boldsymbol{\lambda}^*$. Multiply on the left by $\boldsymbol{A}$, and conclude. You should get $\boldsymbol{\lambda}^* = -(\boldsymbol{A}\boldsymbol{A}^T)^{-1} \boldsymbol{A}\boldsymbol{x}_0$ and $\boldsymbol{x}^* = (\boldsymbol{I}_n - \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A})\boldsymbol{x}_0$.

A3. The Lagrange condition is:

$$\begin{aligned} (\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b})^T \boldsymbol{A} + \boldsymbol{\lambda}^{*T} \boldsymbol{C} &= \boldsymbol{0}^T \\ \boldsymbol{C}\boldsymbol{x}^* &= \boldsymbol{d}. \end{aligned}$$

The first equation gives

$$\boldsymbol{x}^* = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} - (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{C}^T\boldsymbol{\lambda}^*.$$

One needs to eliminate $\boldsymbol{\lambda}$. So, multiply on the left by $\boldsymbol{C}$ and use the second equation to get

$$\boldsymbol{d} = \boldsymbol{C}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} - \boldsymbol{C}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{C}^T\boldsymbol{\lambda}^*$$

Use this to find an expression for $\boldsymbol{\lambda}^*$ in terms of $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{C}$, and $\boldsymbol{d}$ only. Substitute that back into the expression for $\boldsymbol{x}^*$ to get your final answer.

A4. (a) The KKT conditions for the minimization problem are:

$$\begin{aligned} 2x_1 - 2 - 2\mu_1 x_1 + 5\mu_2 &= 0 \\ 2x_2 - 10 + \frac{1}{5}\mu_1 + \frac{1}{2}\mu_2 &= 0 \\ \mu_1\left(\frac{1}{5}x_2 - x_1^2\right) + \mu_2\left(5x_1 + \frac{1}{2}x_2 - 5\right) &= 0 \\ \boldsymbol{\mu} &\geq 0. \end{aligned}$$

We now have four cases for μ_1 and μ_2.

- Case I: $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{x}_{\mathrm{I}} = [1, 5]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = 0$. The complementarity slack condition implies that $x_2 = 5x_1^2$. Using this, the KKT conditions simplify to:

$$x_1(1 - \mu_1) = 1$$
$$50x_1^2 - 50 + \mu_1 = 0.$$

Solve for μ_1 and then x_1. You should get:

$$\boldsymbol{\mu}_{\mathrm{II}}^+ = [26 + 5\sqrt{23}, 0]^T, \ \boldsymbol{x}_{\mathrm{II}}^+ = [\tfrac{-5+\sqrt{23}}{10}, \tfrac{24-5\sqrt{23}}{10}]^T$$

and

$$\boldsymbol{\mu}_{\mathrm{II}}^- = [26 - 5\sqrt{23}, 0]^T, \ \boldsymbol{x}_{\mathbf{II}}^- = [\tfrac{-5-\sqrt{23}}{10}, \tfrac{24+5\sqrt{23}}{10}]^T.$$

Both points are feasible.
- Case III: $\mu_1 = 0$, $\mu_2 > 0$. The complementarity slack condition implies that $x_2 = 10 - 10x_1$. Using this, the KKT conditions simplify to:

$$2x_1 - 2 + 5\mu_2 = 0$$
$$-20x_1 + 10 + \frac{1}{5}\mu_2 = 0.$$

Solve for μ_2 and then x_1. You should get:

$$\boldsymbol{\mu}_{\mathrm{III}} = [0, \tfrac{50}{251}]^T, \ \boldsymbol{x}_{\mathrm{III}} = [\tfrac{126}{251}, \tfrac{1250}{251}]^T.$$

However, this point is not feasible.
- Case IV: $\mu_1, \mu_2 > 0$. The complementarity slack condition implies that:

$$\frac{1}{5}x_2 - x_1^2 = 0$$
$$5x_1 + \frac{1}{2}x_2 - 5 = 0.$$

Solving for x_1 and x_2, we get:

$$\boldsymbol{x}_{\mathrm{IV}}^+ = [-1 - \sqrt{3}, 20 + 10\sqrt{3}]^T, \ \boldsymbol{x}_{\mathrm{IV}}^- = [-1 + \sqrt{3}, 20 - 10\sqrt{3}]^T.$$

Using these, solve for μ_1 and μ_2. You should get:

$$\boldsymbol{\mu}_{\mathrm{IV}}^+ = [101 + \tfrac{152}{\sqrt{3}}, \tfrac{-502}{5} - \tfrac{904}{5\sqrt{3}}]^T, \ \boldsymbol{\mu}_{\mathrm{IV}}^- = [101 - \tfrac{152}{\sqrt{3}}, \tfrac{-502}{5} + \tfrac{904}{5\sqrt{3}}]^T.$$

Since $\boldsymbol{\mu}_{\mathrm{IV}}^+$ does not satisfy the dual feasibility condition, we eliminate $\boldsymbol{x}_{\mathrm{IV}}^+$ from consideration, and only retain $\boldsymbol{x}_{\mathrm{IV}}^-$, which is feasible.

Finally, check that the three candidate points $\boldsymbol{x}_{\mathrm{II}}^{\pm}$ and $\boldsymbol{x}_{\mathrm{IV}}^-$ are all regular.

What about for the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While three of the four cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multipliers found for the feasible points do not satisfy $\boldsymbol{\mu}^* \leq 0$, there are no new points to consider.

So, we now have to apply the second order Lagrange conditions. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2 - 2\mu_1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Since there are no immediate conclusions that can be drawn from this for any of the three points, let's compute the tangent spaces at each point to apply the SONC. We have

$$J(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{j : g_j(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = 0\} = \{1\},$$

so,

$$T(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_1(\boldsymbol{x}_{\mathrm{II}}^{\pm})\boldsymbol{y} = 0\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\mathrm{II}}^{\pm}) = \{a[1, -5 \pm \sqrt{23}]^T : a \in \mathbb{R}\}.$$

Similarly,

$$J(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{IV}}^{-}) = 0\} = \{1, 2\},$$

so,

$$T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{y} \in \mathbb{R}^2 : D\boldsymbol{g}(\boldsymbol{x}_{\mathrm{IV}}^{-})\boldsymbol{y} = \boldsymbol{0}\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{0}\}.$$

Now, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{II}}^{+})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{+}, \boldsymbol{\mu}_{\mathrm{II}}^{+})\boldsymbol{y} = a^2[2 - 2(26 + 5\sqrt{23}) + 2(-5 + \sqrt{23})^2] = -2a^2[15\sqrt{23} - 23],$$

which is < 0 when $a \neq 0$ (for instance, take $a = 1$). Hence, $\boldsymbol{x}_{\mathrm{II}}^{+}$ does not satisfy the SONC. Similarly, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{II}}^{-})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})\boldsymbol{y} = a^2[2 - 2(26 - 5\sqrt{23}) + 2(-5 - \sqrt{23})^2] = 2a^2[15\sqrt{23} + 23],$$

which is always ≥ 0. Hence, $\boldsymbol{x}_{\mathrm{II}}^{-}$ satisfies the SONC.
Lastly, notice that since $T(\boldsymbol{x}_{\mathrm{IV}}^{-}) = \{\boldsymbol{0}\}$, it is true that $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})\boldsymbol{y} \leq 0$ for all $\boldsymbol{y} \in T(\boldsymbol{x}_{\mathrm{IV}}^{-})$, and so $\boldsymbol{x}_{\mathrm{IV}}^{-}$ also satisfies the SONC.
Next, to check the SOSC condition for $\boldsymbol{x}_{\mathrm{II}}^{-}$ and $\boldsymbol{x}_{\mathrm{IV}}^{-}$, we need to compute $\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$. We have

$$\tilde{J}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{II}}^{-}) = 0, \mu_j > 0\} = \{1\} = J(\boldsymbol{x}_{\mathrm{II}}^{-}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-}) = T(\boldsymbol{x}_{\mathrm{II}}^{-}).$$

Hence, for all nonzero $y \in \tilde{T}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{II}}^{-}, \boldsymbol{\mu}_{\mathrm{II}}^{-})\boldsymbol{y} > 0.$$

Hence, $\boldsymbol{x}_{\mathrm{II}}^{-}$ satisfies the SOSC and is a strict local minimizer.
Similarly, we have

$$\tilde{J}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-}) = \{j : g_j(\boldsymbol{x}_{\mathrm{IV}}^{-}) = 0, \mu_j > 0\} = \{1, 2\} = J(\boldsymbol{x}_{\mathrm{IV}}^{-}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-}) = T(\boldsymbol{x}_{\mathrm{IV}}^{-}).$$

Since there is no nonzero $y \in \tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})$, it is vacuously true that

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})\boldsymbol{y} > 0$$

for all nonzero $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}_{\mathrm{IV}}^{-}, \boldsymbol{\mu}_{\mathrm{IV}}^{-})$. Hence, $\boldsymbol{x}_{\mathrm{IV}}^{-}$ satisfies the SOSC and is a strict local minimizer.

(b) The KKT conditions for the minimization problem are:

$$\begin{aligned} 2x_1 - \mu_1 - \mu_3 &= 0 \\ 2x_2 - \mu_2 - \mu_3 &= 0 \\ -x_1 &\leq 0 \\ -x_2 &\leq 0 \\ -x_1 - x_2 + 5 &\leq 0 \\ -\mu_1 x_1 - \mu_2 x_2 + \mu_3(-x_2 - x_2 + 5) &= 0 \\ \boldsymbol{\mu} &\geq 0. \end{aligned}$$

We now have eight cases for μ_1, μ_2, and μ_3.

- Case I: $\mu_1 = \mu_2 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{I}} = [0,0]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{II}} = [0,0]^T$ is the only critical point, but it is not feasible.
- Case III: $\mu_2 > 0$, $\mu_1 = \mu_3 = 0$. Then, $\boldsymbol{x}_{\text{III}} = [0,0]^T$ is the only critical point, but it is not feasible.
- Case IV: $\mu_3 > 0$, $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{\mu}_{\text{IV}} = [0,0,5]^T$ and $\boldsymbol{x}_{\text{IV}} = [\frac{5}{2}, \frac{5}{2}]^T$. This point is feasible.
- Case V: $\mu_1, \mu_2 > 0$, $\mu_3 = 0$. Then, $\boldsymbol{x}_{\text{V}} = [0,0]^T$ is the only critical point, but it is not feasible.
- Case VI: $\mu_1, \mu_3 > 0$, $\mu_2 = 0$. Then, $\boldsymbol{x}_{\text{VI}} = [0,5]^T$, $\boldsymbol{\mu}_{\text{VI}} = [-10,0,10]^T$. Since the dual feasibility condition is not satisfied, we eliminate $\boldsymbol{x}_{\text{VI}}$ from consideration.
- Case VII: $\mu_2, \mu_3 > 0$, $\mu_1 = 0$. Then, $\boldsymbol{x}_{\text{VII}} = [5,0]^T$, $\boldsymbol{\mu}_{\text{VII}} = [0,-10,10]^T$. Since the dual feasibility condition is not satisfied, we eliminate $\boldsymbol{x}_{\text{VII}}$ from consideration.
- Case VIII: $\mu_1, \mu_2, \mu_3 > 0$. Then, $\boldsymbol{x}_{\text{VIII}} = [0,0]^T$ is the only critical point, but it is not feasible.

Finally, check that the candidate point $\boldsymbol{x}_{\text{IV}}$ is regular.

What about for the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While seven of the eight cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multipliers found for the feasible points do not satisfy $\boldsymbol{\mu}^* \leq 0$, there are no new points to consider.

So, we now have to apply the second order Lagrange conditions. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0.$$

Hence, $\boldsymbol{x}_{\text{IV}}$ is a strict local minimizer.

(c) The KKT conditions for the minimization problem are:

$$\begin{aligned} 2x_1 + 6x_2 - 4 + 2\mu_1 x_1 + 2\mu_2 &= 0 \\ 6x_1 - 2 + 2\mu_1 - 2\mu_2 &= 0 \\ x_1^2 + 2x_2 - 1 &\leq 0 \\ 2x_1 - 2x_2 - 1 &\leq 0 \\ \mu_1(x_1^2 + 2x_2 - 1) + \mu_2(2x_1 - 2x_2 - 1) &= 0 \\ \boldsymbol{\mu}^* &\geq 0. \end{aligned}$$

We now have four cases for μ_1 and μ_2.

- Case I: $\mu_1 = \mu_2 = 0$. Then, $\boldsymbol{x}_{\text{I}} = [\frac{1}{3}, \frac{5}{9}]^T$ is the only critical point, but it is not feasible.
- Case II: $\mu_1 > 0$, $\mu_2 = 0$. The complementarity slack condition implies that $x_2 = \frac{1}{2}(1 - x_1^2)$. Using this, the KKT conditions simplify to:

  $$\begin{aligned} 1 - 3x_1 &= \mu_1 \\ 9x_1^2 - 4x_1 + 1 &= 0. \end{aligned}$$

  Since the quadratic in x_1 has no real roots, there are no critical points from this case.
- Case III: $\mu_1 = 0$, $\mu_2 > 0$. The complementarity slack condition implies that $x_2 = x_1 - \frac{1}{2}$. Using this, the KKT conditions simplify to:

  $$\begin{aligned} 3x_1 - 1 &= \mu_2 \\ 14x_1 - 9 &= 0. \end{aligned}$$

Solving, you should get:

$$\boldsymbol{x}_{\text{III}} = [\tfrac{9}{14}, \tfrac{1}{7}]^T, \ \boldsymbol{\mu}_{\text{III}} = [0, \tfrac{13}{14}]^T.$$

This point is feasible.

- Case IV: $\mu_1, \mu_2 > 0$. The complementarity slack condition implies that:

$$x_1^2 + 2x_2 - 1 = 0$$
$$2x_1 - 2x_2 - 1 = 0.$$

Solving for x_1 and x_2, we get:

$$\boldsymbol{x}_{\text{IV}}^+ = [-1+\sqrt{3}, \tfrac{-3}{2}+\sqrt{3}]^T, \ \boldsymbol{x}_{\text{IV}}^- = [-1-\sqrt{3}, \tfrac{-3}{2}-\sqrt{3}]^T.$$

Using these, solve for μ_1 and μ_2. You should get:

$$\boldsymbol{\mu}_{\text{IV}}^+ = [\tfrac{-14\sqrt{3}+23}{2\sqrt{3}}, \tfrac{-22\sqrt{3}+41}{2\sqrt{3}}]^T, \ \boldsymbol{\mu}_{\text{IV}}^- = [\tfrac{-14\sqrt{3}-23}{2\sqrt{3}}, \tfrac{-22\sqrt{3}-41}{2\sqrt{3}}]^T.$$

Since neither KKT multiplier satisfies the dual feasibility condition, we eliminate both points from consideration.

Finally, check that the candidate point $\boldsymbol{x}_{\text{III}}$ is regular.

What about the corresponding maximization problem? The only difference is that the dual feasibility condition changes to $\boldsymbol{\mu}^* \leq 0$. While three of the four cases will correspondingly change, this does not result in any change in the final calculations. Since the KKT multiplier $\boldsymbol{\mu}_{\text{IV}}^-$ satisfies $\boldsymbol{\mu}^* \leq 0$, the point $\boldsymbol{x}_{\text{IV}}^-$ is a candidate maximizer. We also check that the $\boldsymbol{x}_{\text{IV}}^-$ is regular.

So, we now have to apply the second order Lagrange condition. The Hessian of the Lagrangian here is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2+2\mu_1 & 6 \\ 6 & 0 \end{bmatrix}.$$

Since $\boldsymbol{L}(\boldsymbol{x}_{\text{IV}}^-, \boldsymbol{\mu}_{\text{IV}}^-) < 0$, the SOSC holds for $\boldsymbol{x}_{\text{IV}}^-$. Hence, $\boldsymbol{x}_{\text{IV}}^-$ is a strict local maximizer.

For the point $\boldsymbol{x}_{\text{III}}$, no immediate conclusions that can be drawn by looking at

$$\boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = \begin{bmatrix} 2 & 6 \\ 6 & 0 \end{bmatrix},$$

so let's compute the tangent space at $\boldsymbol{x}_{\text{III}}$ to apply the SONC. We have

$$J(\boldsymbol{x}_{\text{III}}) = \{j : g_j(\boldsymbol{x}_{\text{III}}) = 0\} = \{2\},$$

so,

$$T(\boldsymbol{x}_{\text{III}}) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_2(\boldsymbol{x}_{\text{III}})\boldsymbol{y} = 0\}.$$

Solving, we get

$$T(\boldsymbol{x}_{\text{III}}) = \{a[1,1]^T : a \in \mathbb{R}\}.$$

Now, check that for $\boldsymbol{y} \in T(\boldsymbol{x}_{\text{III}})$, we have

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})\boldsymbol{y} = 14a^2,$$

which is always ≥ 0. Hence, $\boldsymbol{x}_{\text{III}}$ satisfies the SONC.

Next, to check the SOSC condition for $\boldsymbol{x}_{\text{III}}$, we need to compute $\tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})$. We have

$$\tilde{J}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = \{j : g_j(\boldsymbol{x}_{\text{III}}) = 0, \mu_j > 0\} = \{2\} = J(\boldsymbol{x}_{\text{III}}),$$

so,

$$\tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}}) = T(\boldsymbol{x}_{\text{III}}).$$

Hence, for all nonzero $\boldsymbol{y} \in \tilde{T}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}_{\text{III}}, \boldsymbol{\mu}_{\text{III}})\boldsymbol{y} > 0.$$

Hence, $\boldsymbol{x}_{\text{III}}$ satisfies the SOSC and is a strict local minimizer.

A5. (a) The only point satisfying the KKT condition is $\boldsymbol{x}^* = [1,3]^T$ with KKT multiplier $\mu^* = 1$.

(b) The constraint $g(\boldsymbol{x})$ is active at $\boldsymbol{x}^*$, so

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg(\boldsymbol{x}^*)\boldsymbol{y} = 0\} = \{a[1,0]^T : a \in \mathbb{R}\},$$

and

$$N(\boldsymbol{x}^*) = \{\boldsymbol{y} : \boldsymbol{y} = Dg(\boldsymbol{x}^*)^T z,\ z \in \mathbb{R}\} = \{a[0,1]^T : a \in \mathbb{R}\}.$$

Since $\mu^* > 0$, $\tilde{J}(\boldsymbol{x}^*, \mu^*) = J(\boldsymbol{x}^*)$ in this case. So, $\tilde{T}(\boldsymbol{x}^*, \mu^*) = T(\boldsymbol{x}^*) = \{a[1,0]^T : a \in \mathbb{R}\}$.

(c)

$$\boldsymbol{L}(\boldsymbol{x}^*, \mu^*) = \begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$$

which is negative semidefinite. So, we cannot conclude anything immediately, and need to check the sign of $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \mu^*)\boldsymbol{y}$ for an arbitrary $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ to see if $\boldsymbol{x}^*$ satisfies the SONC. In this case, we get for $\boldsymbol{y} = a[1,0]^T$,

$$\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \mu^*)\boldsymbol{y} = -2a^2$$

which is < 0 when $a \neq 0$ (for instance, when $a = 1$). Thus, $\boldsymbol{x}^*$ does not satisfy the SONC, and is not a local minimizer.

A6. (a) Write down the KKT conditions and find the KKT multipliers for the point $\boldsymbol{x}^* = \boldsymbol{0}$. You should get $\mu_1 = -2$, $\mu_2 = 0$, and $\mu_3 = -4$. Hence, $\boldsymbol{x}^*$ satisfies the FONC for maximization, but not for minimization.

(b) The Hessian of the Lagrangian in this case is:

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix},$$

so we cannot conclude anything immediately. To check whether $\boldsymbol{x}^*$ satisfies the SOSC, we need to find $\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$. Here, we have $\tilde{J}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{1, 3\}$, so

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{y} \in \mathbb{R}^2 : Dg_1(\boldsymbol{x}^*)\boldsymbol{y} = 0,\ Dg_3(\boldsymbol{x}^*)\boldsymbol{y} = 0\}.$$

Solving, we get

$$\tilde{T}(\boldsymbol{x}^*, \boldsymbol{\mu}^*) = \{\boldsymbol{0}\}.$$

Hence, the SOSC is vacuously satisfied by $\boldsymbol{x}^*$, and $\boldsymbol{x}^*$ is a strict local maximizer.

A7. The KKT conditions are:

$$\begin{aligned} (\boldsymbol{x}^* - \boldsymbol{x}_0) + \mu^* \nabla g(\boldsymbol{x}^*) &= 0 \\ \mu^* g(\boldsymbol{x}^*) &= 0. \end{aligned}$$

Multiply on the left by $(\boldsymbol{x}^* - \boldsymbol{x}_0)^T$ to get

$$\|\boldsymbol{x}^* - \boldsymbol{x}_0\|^2 + \mu^* (\boldsymbol{x}^* - \boldsymbol{x}_0)^T \nabla g(\boldsymbol{x}^*) = 0.$$

Since $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|^2 > 0$ (why?) and $\mu^* \geq 0$, condition (iii) must be true, with $\mu^* > 0$. Hence, condition (ii) is also true. These are the only ones that hold.

A8. (1) By definition of $J(\boldsymbol{x}^*)$, we have $g_i(\boldsymbol{x}^*) < 0$ for all $i \notin J(\boldsymbol{x}^*)$. Since by assumption g_i is continuous for all i, there exists $\epsilon > 0$ such that $g_i(\boldsymbol{x}) < 0$ for all $i \in J(\boldsymbol{x}^*)$ and all $\boldsymbol{x}$ such that $\|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon$.

(2) Clearly, $S \cap B \subseteq S_1 \cap B$. To show that $S_1 \cap B \subseteq S \cap B$, suppose that $\boldsymbol{x} \in S_1 \cap B$. Then, by definition of S_1 and B, we have $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $g_j(\boldsymbol{x}) \leq 0$ for all $j \in J(\boldsymbol{x}^*)$, and $g_i(\boldsymbol{x}^*) < 0$ for all $i \notin J(\boldsymbol{x}^*)$. Hence, $x \in S \cap B$.

(3) Since $\boldsymbol{x}^*$ is a local minimizer of f over S, and $S \cap B \subseteq S$, $\boldsymbol{x}^*$ is also a local minimizer of f over $S \cap B = S_1 \cap B$. Hence, we conclude that $\boldsymbol{x}^*$ is a regular local minimizer of f on S_1.

(4) Since $S' \subseteq S_1$ and $\boldsymbol{x}^* \in S'$, we conclude that $\boldsymbol{x}^*$ is a regular local minimizer of f on S'.

A9. (a) Here, $\nabla f(\boldsymbol{x}) = \boldsymbol{c}$ and $\nabla g(\boldsymbol{x}) - \boldsymbol{x}$. Note that $\boldsymbol{x}^* \neq \mathbf{0}$ (why?), and therefore it is regular. By the KKT Theorem, there exists $\mu^* \geq 0$ such that $\boldsymbol{c} = -\mu^* \boldsymbol{x}^*$ and $\mu^* g(\boldsymbol{x}^*) = 0$. Since $\boldsymbol{c} \neq \mathbf{0}$, we must have $\mu^* \neq 0$. Hence, $g(\boldsymbol{x}^*) = 0$, so $\frac{1}{2}\|\boldsymbol{x}^*\|^2 - 1 = 0$.

(b) Since $\|\boldsymbol{e}\|^2 = n$ and $\alpha^2 \|\boldsymbol{e}\|^2 = 2$ from the previous part, we have $\alpha = (2.n)^{1/2}$.

To find $\boldsymbol{c}$, use $f(\boldsymbol{x}^*) = 4$ and $\boldsymbol{c} = \mu^* \boldsymbol{x}^*$ to find μ^* and then $\boldsymbol{c}$. You should get $\mu^* = -2$ and $\boldsymbol{c} = (-2(2/n)^{1/2})\boldsymbol{e}$.

A10. See Theorem 22.8 in [CZ13]. Also see Lecture 24 slides.

A11. See Theorem 22.9 in [CZ13]. Also see Lecture 24 slides.

Tutorial Sheet 9

Q1. Consider the following standard form LP problem:

$$\begin{aligned} \text{minimize} \quad & 2x_1 - x_2 - x_3 \\ \text{subject to} \quad & 3x_1 + x_2 + x_4 = 4 \\ & 6x_1 + 2x_2 + x_3 + x_4 = 5 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

(a) Write down the $\boldsymbol{A}$, $\boldsymbol{b}$, and $\boldsymbol{c}$ matrices/vectors for the problem.

(b) Consider the basis consisting of the third and fourth columns of $\boldsymbol{A}$, ordered according to $[\boldsymbol{a}_4, \boldsymbol{a}_3]$. Compute the canonical tableau corresponding to this basis.

(c) Write down the basic feasible solution corresponding to the basis above, and its objective function value.

(d) Write down the values of the reduced cost coefficients (for all the variables) corresponding to the basis.

(e) Is the basic feasible solution in part (c) an optimal feasible solution? If yes, explain why. If not, determine which element of the canonical tableau to pivot about so that the new basic feasible solution will have a lower objective function value.

Q2. Use the simplex method to solve the following linear program:

$$\begin{aligned} \text{maximize} \quad & x_1 + x_2 + 3x_3 \\ \text{subject to} \quad & x_1 + x_3 = 1 \\ & x_2 + x_3 = 2 \\ & x_1, x_2, x_3 \geq 0. \end{aligned}$$

Q3. Consider the linear program

$$\begin{aligned} \text{maximize} \quad & 2x_1 + x_2 \\ \text{subject to} \quad & 0 \leq x_1 \leq 5 \\ & 0 \leq x_2 \leq 7 \\ & x_1 + x_2 \leq 9. \end{aligned}$$

Convert the problem to standard form and solve it using the simplex method.

Q4. Consider the problem

$$\begin{aligned} \text{maximize} \quad & -x_1 - 2x_2 \\ \text{subject to} \quad & x_1 \geq 0 \\ & x_2 \geq 1. \end{aligned}$$

(a) Convert the problem into a standard form linear programming problem.

(b) Use the simplex method to compute the solution to this problem and the value of the objective function at the optimal solution of the problem.

Solution outlines

A1. (a)

$$A = \begin{bmatrix} 3 & 1 & 0 & 1 \\ 6 & 2 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \quad c = [2, -1, -1, 0].$$

(b) Pivoting about $(1,4)$ and $(2,3)$, we obtain

$$\begin{array}{ccccc} 3 & 1 & 0 & 1 & 4 \\ 3 & 1 & 1 & 0 & 1 \\ 5 & 0 & 0 & 0 & 1 \end{array}$$

(c) $\boldsymbol{x} = [0,0,1,4]^T$, $\boldsymbol{x}^T\boldsymbol{c} = -1$.

(d) $[r_1, r_2, r_3, r_4] = [5,0,0,0]$

(e) Since reduced cost solution is ≥ 0, feasible solution in part (c) is optimal.

A2. Write down the problem in standard form and form its tableau. Performing the necessary row operations to bring it into canonical form, you can get:

$$\begin{array}{cccc} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 3 \end{array}$$

Pivot about the $(1,3)$th element, and complete the solution. The optimal solution is $[0,1,1]^T$ and optimal cost is 4.

A3. Again, bring the system into the standard form, as:

$$\begin{array}{ll} \text{minimize} & -2x_1 - x_2 \\ \text{subject to} & x_1 + x_3 = 5 \\ & x_2 + x_4 = 7 \\ & x_1 + x_2 + x_5 = 9 \\ & x_1, \ldots, x_5 \geq 0. \end{array}$$

Tableau:

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 1 & 0 & 7 \\ 1 & 1 & 0 & 0 & 1 & 9 \\ -2 & -1 & 0 & 0 & 0 & 0 \end{array}$$

Pivot about the $(1,1)$th element. After row operations,

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 1 & 0 & 7 \\ 0 & 1 & -1 & 0 & 1 & 4 \\ 0 & -1 & 2 & 0 & 0 & 10 \end{array}$$

Pivot about the $(3,2)$th element. After appropriate row operations:

$$\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 1 & 1 & -1 & 3 \\ 0 & 1 & -1 & 0 & 1 & 4 \\ 0 & 0 & 1 & 0 & 1 & 14 \end{array}$$

Therefore solution is $[5,4,0,3,0]^T$ and optimal cost is 14.

A4. (a) In standard form:

$$\begin{array}{ll} \text{minimize} & x_1 + 2x_2 \\ \text{subject to} & x_2 - x_3 = 1 \\ & x_1, x_2, x_3 \geq 1. \end{array}$$

(b) Final answer: The optimal solution to the original problem is $[0,1]^T$ with objective function value -2.

Q1. WTS: $f(x^{(k)} + \alpha d^{(k)}) < f(x^{(k)})$ for all $\alpha \in (0, \bar{\alpha}]$, for some $\bar{\alpha} > 0$.

$$\phi(\alpha) = f(x^{(k)} + \alpha d^{(k)})$$

$$\phi'(\alpha) = \nabla f(x^{(k)} + \alpha d^{(k)})^T d^{(k)} \quad \text{[Chain Rule]}$$

$$\therefore \phi'(0) = \nabla f(x^{(k)})^T d^{(k)} = -\underbrace{g^{(k)T} F(x^{(k)})^{-1} g^{(k)}} < 0$$

because $F(x^{(k)}) > 0$ and inverse of a p.d. matrix is p.d.

$\therefore \exists\ \bar{\alpha} > 0$ s.t. $\phi(\alpha) < \phi(0)$ for all $\alpha \in (0, \bar{\alpha}]$.

Q2. WTS: In the conjugate gradient algorithm,

$$g^{(k+1)^T} d^{(i)} = 0$$

for all $0 \leq k \leq n-1$, $0 \leq i \leq k$. Let's assume that Q is SPD.

Hint: Use induction.

Let's show this for $k=0$, $i=0$.

$$g^{(1)^T} d^{(0)} \stackrel{?}{=} 0$$

$$g^{(1)T} = (Qx^{(1)} - b)^T$$

$$f(x) = \frac{1}{2}x^T Q x - b^T x + c$$

$$Q = Q^T > 0$$

$$g^{(1)T} d^{(0)} = (Qx^{(1)} - b)^T d^{(0)}$$

$$= x^{(1)T} Q d^{(0)} - b^T d^{(0)}$$

$$\boxed{x^{(1)} = x^{(0)} + \alpha_0 d^{(0)}, \quad \text{where } \alpha_0 = -\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}}}$$

$$= \underbrace{x^{(0)T} Q d^{(0)}}$$

$$- \frac{g^{(0)T} d^{(0)}}{\cancel{d^{(0)T} Q d^{(0)}}} \cancel{d^{(0)T} Q d^{(0)}} \underbrace{- b^T d^{(0)}}$$

$$= (Qx^{(0)} - b)^T d^{(0)} - g^{(0)T} d^{(0)}$$

$$= g^{(0)T} d^{(0)} - g^{(0)T} d^{(0)} = 0.$$

Consider $Q(x^{(k+1)} - x^{(k)}) = Qx^{(k+1)} - b - (Qx^{(k)} - b)$

$$= g^{(k+1)} - g^{(k)}.$$

$$g^{(k+1)} = g^{(k)} + \alpha_k Q d^{(k)}.$$

$0 \le i < k$

$$g^{(k+1)T} d^{(k)} = (Qx^{(k+1)} - b)^T d^{(k)}$$

$$= \left(x^{(k)} - \left(\frac{g^{(k)T} d^{(k)}}{d^{(k)T} Q d^{(k)}}\right) d^{(k)}\right)^T Q d^{(k)}$$

$$- b^T d^{(k)}$$

$$= \underbrace{x^{(k)T} Q d^{(k)}} - g^{(k)T} d^{(k)} \underbrace{- b^T d^{(k)}}$$

$$= (Qx^{(k)} - b)^T d^{(k)} - g^{(k)T} d^{(k)} = 0$$

Q3. $f(x_1, x_2) = \frac{5}{2}x_1^2 + x_2^2 - 3x_1x_2 - x_2 - 7$

$= \frac{1}{2}\left[5x_1^2 + 2x_2^2 - 6x_1x_2\right] - x_2 - 7$

$$Q = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

$Q = Q^T > 0$ (the Sylvester's criterion)

$$b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$c = -7$$

$$\boxed{x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

$$\nabla f(x^{(0)}) = Qx^{(0)} - b = \begin{bmatrix} 0 \\ -1 \end{bmatrix} = d^{(0)}$$

$$\alpha_0 = -\frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}} = \frac{-1}{2}$$

$$x^{(1)} = x^{(0)} + \alpha_0 d^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \left(-\frac{1}{2}\right) \cdot \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}.$$

$$g^{(1)} = Q x^{(1)} - b = \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/2 \\ 0 \end{bmatrix}$$

$$\beta_1 = \frac{g^{(1)T} g^{(1)}}{g^{(0)T} g^{(0)}} = \frac{9/4}{1} = \frac{9}{4}$$

or β_0

$$d^{(1)} = g^{(1)} + \beta^{(1)} d^{(0)} = \begin{bmatrix} -3/2 \\ 0 \end{bmatrix} + \frac{9}{4} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -3/2 \\ -9/4 \end{bmatrix}$$

Verify:

$$d^{(0)T} Q\, d^{(1)} = \begin{bmatrix} 0 & -1 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} -3/2 \\ -9/4 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} -3/2 \\ -9/4 \end{bmatrix}$$

$$= \frac{-9}{2} + \frac{9}{2} = 0. \checkmark$$

Elementary_Matrices

Monday, 4 April 2022 1:04 PM

Elementar...

Elementary Matrices

Neela Nataraj

Department of Mathematics,
Indian Institute of Technology Bombay,
Powai, Mumbai 76
neela@math.iitb.ac.in

April 4, 2022

Supplementary Material

1. Elementary matrices and properties
2. Gauss Jordan method

Elementary Matrices

An $m \times m$ elementary matrix is a matrix obtained from the $m \times m$ identity matrix I_m by one of the elementary operations; namely,

1. interchange of two rows,
2. multiplying a row by a non-zero constant,
3. adding a constant multiple of a row to another row.

That is, an elementary matrix is a matrix which differs from the identity matrix by one single elementary row operation.

1. Row Switching Operation

Interchange of two rows - $\boxed{R_i \leftrightarrow R_j}$.

The elementary matrix P_{ij} corresponding to this operation on I_m is obtained by swapping row i and row j of the identity matrix.

$$P_{ij} = \begin{array}{c} \\ R_1 \\ \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{ccccccc} C_1 & \dots & C_i & \dots & C_j & \dots & C_m \end{array} \\ \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \dots & 1 & & \\ & & & \ddots & & & \\ & & 1 & \dots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array}$$

Example : Consider I_3. $R_1 \leftrightarrow R_2$ gives $P_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

2. Row Multiplying Transformation

Multiplying a row by a non-zero constant - $\boxed{R_i \longrightarrow kR_i}$. The elementary matrix $M_i(k)$ corresponding to this operation on I_m is obtained by multiplying row i of the identity matrix by a non-zero constant k.

$$M_i(k) = \begin{array}{c} \\ R_1 \\ R_2 \\ \vdots \\ R_i \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{cccccc} C_1 & C_2 & \dots & C_i & \dots & C_m \end{array} \\ \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & k & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \end{array}$$

Example : Consider I_3. $R_3 \rightarrow 7R_3$ gives $M_3(7) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}$.

3. Row Addition Transformation

$\boxed{R_i \longrightarrow R_i + kR_j}$: The elementary matrix $E_{ij}(k)$ corresponding to this operation on I_m is obtained by multiplying row j of the identity matrix by a non-zero constant k and adding with row i.

$$E_{ij}(k) = \begin{array}{c} \\ R_1 \\ \vdots \\ R_i \\ \vdots \\ R_j \\ \vdots \\ R_m \end{array} \begin{array}{c} \begin{array}{ccccccc} C_1 & \cdots & C_i & \cdots & C_j & \cdots & C_m \end{array} \\ \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & k & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array}$$

Example : $R_2 \to R_2 + (-3)R_1$ for I_3 gives $E_{21}(-3) = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Proposition 1

Let A be an $m \times n$ matrix. If $\widetilde{A}$ is obtained from A by an elementary row operation, and $\mathscr{E}$ is the corresponding $m \times m$ elementary matrix, then $\mathscr{E}A = \widetilde{A}$.

Proof :

Let $I_m = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_m^T \end{bmatrix}$

where $e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, with 1 is in the i^{th} position, $1 \le i \le m$.

Also, $e_i^T A = A_{(i)}$, the i^{th} row of A.

$\boxed{\text{Recall that row } i \text{ of } AB \text{ is row } i \text{ of } A \text{ times } B.}$

$\boxed{R_i \leftrightarrow R_j (i < j)}$

$$P_{ij}A = \begin{bmatrix} \vdots \\ e_j^T \\ \vdots \\ e_i^T \\ \vdots \end{bmatrix} A = \begin{bmatrix} \vdots \\ e_j^T A \\ \vdots \\ e_i^T A \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ A_{(j)} \\ \vdots \\ A_{(i)} \\ \vdots \end{bmatrix} = \widetilde{A}.$$

$\boxed{R_i \longrightarrow kR_i}$

$$M_i(k)A = \begin{bmatrix} \vdots \\ ke_i^T \\ \vdots \end{bmatrix} A = \begin{bmatrix} \vdots \\ kA_{(i)} \\ \vdots \end{bmatrix} = \widetilde{A}.$$

$\boxed{R_i \longrightarrow R_i + kR_j}$

$$E_{ij}(k)A = \begin{bmatrix} \vdots \\ e_i^T + ke_j^T \\ \vdots \end{bmatrix} A = \begin{bmatrix} \vdots \\ (e_i^T + ke_j^T)A \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ A_{(i)} + kA_{(j)} \\ \vdots \end{bmatrix} = \widetilde{A}.$$

Reduced Row Echelon form contd.

Exercise: Let A be an $m \times n$ matrix. There exist elementary matrices $E_1, E_2, ..., E_N$ of order m such that the product $\underline{E_N \cdots E_2 E_1 A}$ is a row echelon form of A.

Proposition 2 :

Elementary matrices are invertible and the inverses are also elementary matrices.

In fact,

• If $\mathscr{E}$ corresponds to $R_i \leftrightarrow R_j$, then $\mathscr{E}^{-1} = P_{ij}$;

• if $\mathscr{E}$ corresponds to $R_i \longrightarrow kR_i$, then $\mathscr{E}^{-1}$ is the elementary matrix corresponding to $R_i \to \frac{1}{k} R_i$; that is, $M_i(1/k)$.

• if $\mathscr{E}$ corresponds to $R_i \longrightarrow R_i + kR_j$, then $\mathscr{E}^{-1}$ is the elementary matrix corresponding to $R_i \to R_i - kR_j$, that is, $E_{ij}(-k)$.

Exercise : In each case, check $\mathscr{E}\mathscr{E}^{-1} = I$.

Since row operations are reversible, elementary matrices are invertible, for if $\mathscr{E}$ is produced by a row operation on I_m, there is another row operation of the same type, that changes $\mathscr{E}$ back to I_m. Hence, there is an elementary operation $\mathscr{F}$ such that $\mathscr{F}\mathscr{E} = I_m$. Since $\mathscr{E}$ and $\mathscr{F}$ correspond to reverse operations, $\mathscr{E}\mathscr{F} = I_m$ too. $\square$

Example

Find the inverses of

$$E_{31}(-4) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, M_2(7) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ \& \ P_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Soln: $E_{31}(-4)^{-1} = E_{31}(4) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}, M_2(7)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/7 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\& \ P_{13}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$

Proposition 3 :

> If A and B are row equivalent square matrices and A is invertible, then B is invertible.

Proof :

- If $A_1, \dots, A_k$ are invertible matrices of same size, then the product $A_1 \dots A_k$ is also invertible; and $(A_1 \dots A_k)^{-1} = A_k^{-1} \dots A_1^{-1}$.
  (Why?
  Hint : Use the associative law for multiplication to derive

$$(AB)(B^{-1}A^{-1}) = \cdots = A(BB^{-1})A^{-1} = \cdots = I.$$

  Generalize.)
- If $\mathscr{E}_1, \dots, \mathscr{E}_k$ are elementary matrices such that

$$\boxed{B = \mathscr{E}_k \dots \mathscr{E}_1 A,}$$

  then B is a product of invertible matrices; and hence is invertible.

Theorem

A square matrix A is invertible if and only if A is row equivalent to the identity matrix.

Proof :

$\boxed{(\Longleftarrow)}$ Let A be row equivalent to the identity matrix. From Proposition 3, as I is invertible, A is also invertible.

$\boxed{(\Longrightarrow)}$ If A is invertible, then its row echelon form

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ 0 & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{nn} \end{pmatrix}$$

is an upper triangular matrix with $b_{ii} \neq 0$ for all i. (Why?)
(B is obtained from A via $B = \mathscr{E}_k \dots \mathscr{E}_1 A$. Since A is invertible, B is invertible (Proposition 3). Now, if $b_{ii} = 0$ for any i, this would imply $det(B) = 0$ and this contradicts invertibility of B, as ($since\ |B||B^{-1}| = 1$)).

Proof continued...

Applying $R_i \to b_{ii}^{-1} R_i$ (for $i = 1, \ldots, n$) gives

$$B \sim \begin{pmatrix} 1 & b_{11}^{-1} b_{12} & \ldots & b_{11}^{-1} b_{1n} \\ 0 & 1 & \ldots & b_{22}^{-1} b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Now applying some more row operations of the type $R_i \to R_i - cR_j$, we get the required result, that is,

$$A \sim B \sim I.$$

Result: Let A be a square matrix, say of order $n \times n$. There exist elementary matrices $E_1, E_2, ..., E_N$ of order n such that the product $E_N \cdots E_2 E_1 A$ is either the $n \times n$ identity matrix I or its last row is 0.

Proof:

Consider the *reduced* row echelon form of A. Recall that there must be $p \leq n$ pivots in all. If there are $p = n$ pivots then the *reduced* REF must be I. If there are $p < n$ pivots,then the last $n - p$ rows must vanish.

Result: If A is an invertible matrix, then A can be written as a product of elementary matrices.

Proof : If A is invertible, then A is row equivalent to identity matrix, that is, there exists elementary matrices $\mathscr{E}_1, \ldots, \mathscr{E}_k$ such that $\mathscr{E}_k \ldots \mathscr{E}_1 A = I$.

This gives $A = (\mathscr{E}_k \ldots \mathscr{E}_1)^{-1} = \mathscr{E}_1^{-1} \ldots \mathscr{E}_k^{-1}$.

Gauss-Jordan method for finding A^{-1}

Let A be invertible, and $\mathscr{E}_k \ldots \mathscr{E}_1 A = I$.

Then, $\mathscr{E}_k \ldots \mathscr{E}_1 I = A^{-1}$.

That is, the elementary row operations performed on A to reduce it to identity matrix, performed in the same order on I reduces I to A^{-1}.

Inverse of A : an example

$$[A|I] \quad = \quad \begin{bmatrix} 2 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 4 & -6 & 0 & \vdots & 0 & 1 & 0 \\ -2 & 7 & 2 & \vdots & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow[R_3 \to R_3 + (1) R_1]{R_2 \to R_2 + (-2) R_1} \begin{bmatrix} 2 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & -8 & -2 & \vdots & -2 & 1 & 0 \\ 0 & 8 & 3 & \vdots & 1 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{R_3 \to R_3 + (1) R_1} \begin{bmatrix} 2 & 1 & 1 & \vdots & 1 & 0 & 0 \\ 0 & -8 & -2 & \vdots & -2 & 1 & 0 \\ 0 & 0 & 1 & \vdots & -1 & 1 & 1 \end{bmatrix} = [B|L]$$

This completes the forward elimination. The first half of elimination has taken A to echelon form B, and now the second half will take B to I. That is, we create $0's$ above the pivots in the last matrix.

Example continued...

$$[B|L] \xrightarrow[R_1 \to R_1 + (-1) R_3]{R_2 \to R_2 + (2) R_3} \begin{bmatrix} 2 & 1 & 0 & \vdots & 2 & -1 & -1 \\ 0 & -8 & 0 & \vdots & -4 & 3 & 2 \\ 0 & 0 & 1 & \vdots & -1 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow{R_1 \to R_1 + (1/8) R_3} \begin{bmatrix} 2 & 0 & 0 & \vdots & \frac{12}{8} & -\frac{5}{8} & -\frac{6}{8} \\ 0 & -8 & 0 & \vdots & -4 & 3 & 2 \\ 0 & 0 & 1 & \vdots & -1 & 1 & 1 \end{bmatrix}$$

$$\xrightarrow[(2) \times R_2 / -8]{(1) \times R_1 / 2} \begin{bmatrix} 1 & 0 & 0 & \vdots & \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ 0 & 1 & 0 & \vdots & \frac{4}{8} & -\frac{3}{8} & -\frac{2}{8} \\ 0 & 0 & 1 & \vdots & -1 & 1 & 1 \end{bmatrix} = [I|A^{-1}]$$

Example

Write $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ as a product of 4 elementary row matrices if $D = ad - bc \neq 0$.

Case 1 ($a \neq 0$): $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a & b \\ 0 & D/a \end{bmatrix} \mapsto \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix} \mapsto \mathrm{I}_2$.
where $D = ad - bc$ (see next slide for details).
Hence

$$A = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} 1/a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D/a \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}$$

$$= E_{21}(c/a) M_1(a) M_2(D/a) E_{12}(b/a).$$

Case 2: $(a = 0) \Longrightarrow -D = bc \neq 0$: Then

$$A = P_{12} M_2(b) M_1(c) E_{12}(d/c).$$

Detailing the previous example (Case 1)

Start with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, let $D = ad - bc$.

Operation	Matrix	Product with elementary matrix
$R_2 \to R_2 - (c/a)R_1$	$\begin{bmatrix} a & b \\ 0 & D/a \end{bmatrix}$	$E_{21}(-c/a)A$
$R_2 \to R_2/(D/a)$ $R_1 \to R_1/a$	$\begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix}$	$M_1(1/a)M_2(a/D)E_{21}(-c/a)A$
$R_1 \to R_1 - (b/a)R_2$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$E_{12}(-b/a)M_1(1/a)M_2(a/D)$ $\times E_{21}(-c/a)A$

Hence, I can be obtained from A by performing $E_{12}(-b/a)M_1(1/a)M_2(a/D)E_{21}(-c/a)$ on A; that is,

$\boxed{I = E_{12}(-b/a)M_1(1/a)M_2(a/D)E_{21}(-c/a)A.}$

Arguing in a similar manner,

$$A = E_{21}(c/a)M_2(D/a)M_1(a)E_{12}(b/a)I$$
$$= \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D/a \end{bmatrix} \begin{bmatrix} 1/a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix} (\textit{Verify}!)$$

Finding inverse by Gauss Jordan Method (Exercise)

Example: Find the inverse of $A = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}$.

Solution:

$$\begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{E_{21}(3), E_{31}(-1)} \begin{bmatrix} -1 & 1 & 2 \\ 0 & 2 & 7 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{E_{32}(-1)} \begin{bmatrix} -1 & 1 & 2 \\ 0 & 2 & 7 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & -1 & 1 \end{bmatrix}$$

Example contd.

$$\xrightarrow{M_1(-1), M_2(1/2), M_3(-1/5)} \begin{bmatrix} 1 & -1 & -2 \\ 0 & 1 & 7/2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 3/2 & 1/2 & 0 \\ 4/5 & 1/5 & -1/5 \end{bmatrix}$$

$$\xrightarrow{E_{13}(2), E_{23}(-7/2)} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3/5 & 2/5 & -2/5 \\ -13/10 & -1/5 & 7/10 \\ 4/5 & 1/5 & -1/5 \end{bmatrix}$$

$$\xrightarrow{E_{12}(1)} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -7/10 & 1/5 & 3/10 \\ -13/10 & -1/5 & 7/10 \\ 4/5 & 1/5 & -1/5 \end{bmatrix}$$

Example contd.

It follows that

$$A^{-1} = \frac{1}{10}\begin{bmatrix} -7 & 2 & 3 \\ -13 & -2 & 7 \\ 8 & 2 & -2 \end{bmatrix}.$$

Also putting all the row ops together,

$$\begin{aligned} A^{-1} = E_{12}(1)E_{13}(2)E_{23}(-7/2)M_1(-1) \\ \times M_2(1/2)M_3(-1/5)E_{32}(-1)E_{21}(3)E_{31}(-1) \end{aligned}$$

as a product of ERM's.