قرارداد نوشتار :

$a \rightarrow$ اسکالر a

$\underline{a} \rightarrow$ بردار ستونی $\longrightarrow \underline{a} = \begin{bmatrix} x \\ y \end{bmatrix} = (x, y)$

$\underline{A} \rightarrow$ ماتریس $\longrightarrow \underline{A} = \begin{bmatrix} a_{۱۱} & a_{۱۲} \\ a_{۲۱} & a_{۲۲} \end{bmatrix}$

معرفی بردار ها

- بردار : تعریفی است که اجزای غیر همجنس را در خود جای می دهد.

$$\underline{v} = \begin{bmatrix} v_۱ \\ v_۲ \end{bmatrix}$$

$v_۱ \rightarrow$ سیب

$v_۲ \rightarrow$ پرتقال

- جمع برداری :

$$\underline{v} = \begin{bmatrix} v_۱ \\ v_۲ \end{bmatrix} \quad \underline{w} = \begin{bmatrix} w_۱ \\ w_۲ \end{bmatrix} \Rightarrow \underline{v} + \underline{w} = \begin{bmatrix} v_۱ + w_۱ \\ v_۲ + w_۲ \end{bmatrix}$$

- تفاضل برداری :

$$\underline{v} - \underline{w} = \begin{bmatrix} v_۱ - w_۱ \\ v_۲ - w_۲ \end{bmatrix}$$

- ضرب عددی :

$$c\underline{v} = \begin{bmatrix} cv_۱ \\ cv_۲ \end{bmatrix}$$

- ترکیب خطی :

ترکیب خطی دو بردار $\underline{v}$ و $\underline{w}$ $= c\underline{v} + d\underline{w}$

حالت های خاص ترکیب خطی :

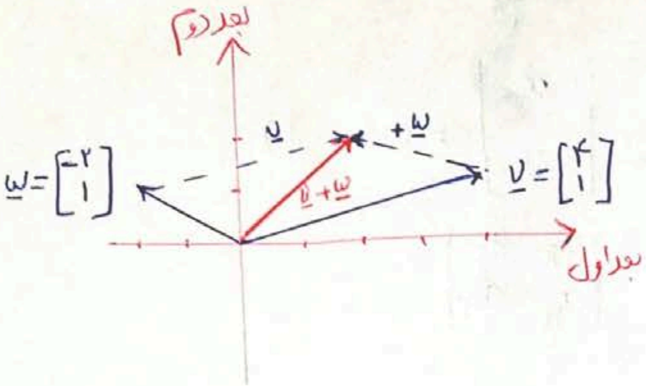
→ جمع برداری : $c = ۱ , d = ۱ \Rightarrow$ ترکیب خطی $= \underline{v} + \underline{w}$

→ تفاضل برداری : $c = ۱ , d = -۱ \Rightarrow$ ترکیب خطی $= \underline{v} - \underline{w}$

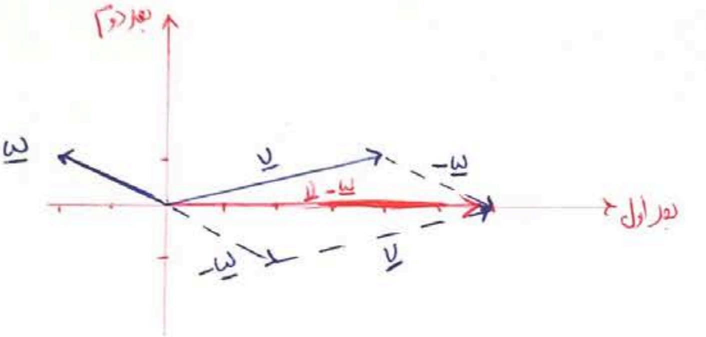
→ بردار صفر : $c = ۰ , d = ۰ \Rightarrow$ ترکیب خطی $= \underline{0}$

→ ضرب عددی : $d = ۰ \Rightarrow$ ترکیب خطی $= c\underline{v}$

- نمایش بردارها :

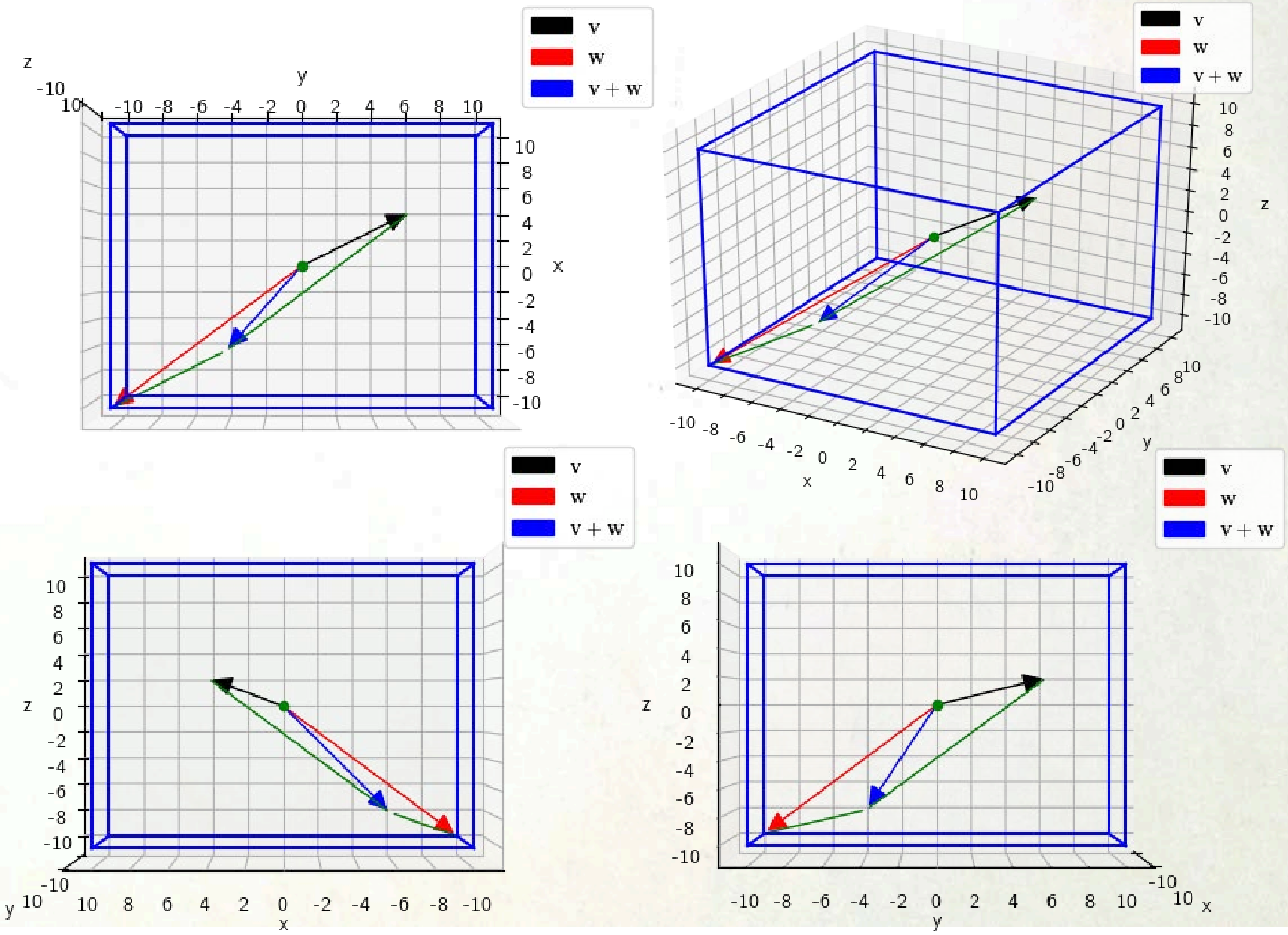


مجموع برداری



تفاضل برداری

- تعمیم به ابعاد بالاتر: با همان توجیهی که برای شکل دادن بردارهای دو بعدی داشتیم (وجود اعضای غیرهمجنس)، در برخی مواقع نیاز به ابعاد بالاتر نیز داریم.



- ترکیب‌های خطی در فضای سه بعدی

۱) یک بردار ← $c\underline{u}$ ← خط گذرنده از $\underline{0}$

۲) دو بردار ← $c\underline{u}+d\underline{v}$ ← صفحه شامل $\underline{0}$

۳) سه بردار ← $c\underline{u}+d\underline{v}+e\underline{w}$ ← کل فضای سه بعدی

حالت خاص یک: دو بردار - $\underline{v}$, $\underline{u}$ هم جهت ← خط گذرنده از $\underline{0}$

حالت خاص دو: سه بردار - $\underline{w}$ در صفحه $c\underline{u}+d\underline{v}$ ← صفحه شامل $\underline{0}$

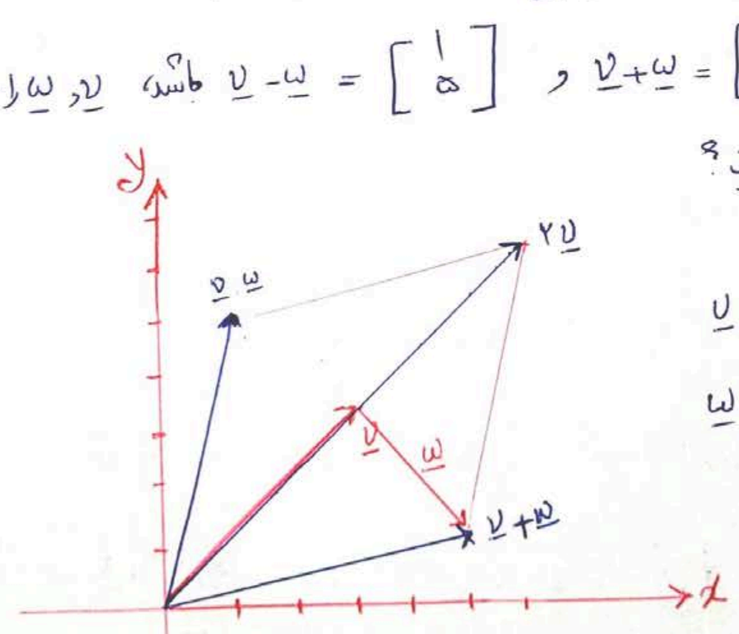
- مثال ۱: هندسه ترکیب خطی بردارهای زیر را توصیف کنید:

۱) $\begin{bmatrix}۱\\۲\\۳\end{bmatrix}$ و $\begin{bmatrix}۳\\۶\\۹\end{bmatrix}$ ← یک خط در $\mathbb{R}^۳$

۲) $\begin{bmatrix}۱\\۰\\۰\end{bmatrix}$ و $\begin{bmatrix}۰\\۲\\۳\end{bmatrix}$ ← یک صفحه در $\mathbb{R}^۳$

۳) $\begin{bmatrix}۲\\۰\\۰\end{bmatrix}$ و $\begin{bmatrix}۰\\۲\\۲\end{bmatrix}$ و $\begin{bmatrix}۲\\۲\\۳\end{bmatrix}$ ← تمام فضای $\mathbb{R}^۳$

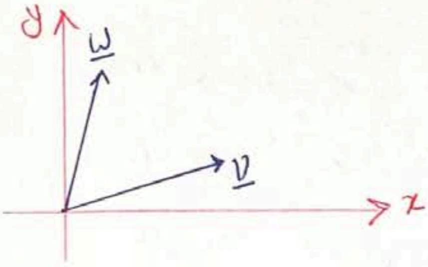
- مثال ۲: اگر $\underline{v}+\underline{w} = \begin{bmatrix}۵\\۱\end{bmatrix}$ و $\underline{v}-\underline{w} = \begin{bmatrix}۱\\۵\end{bmatrix}$ باشد، $\underline{v}$ و $\underline{w}$ را به صورت هندسی به دست آورید؟



$\underline{v} = \begin{bmatrix}۳\\۳\end{bmatrix}$

$\underline{w} = \begin{bmatrix}+۲\\-۲\end{bmatrix}$

- **مثال ۳**: بردارهای $\underline{v}$ و $\underline{w}$ را در نظر بگیرید. تمام ترکیب‌های خطی $c\underline{v} + d\underline{w}$ را برای $c+d=1$ توصیف نمایید.



خط گذرنده از $\underline{v}$ و $\underline{w}$

در مثال قبل ترکیب‌های خطی $c\underline{v} + c\underline{w}$ را توصیف نمایید.
خطی است که قطر متوازی الاضلاع بخشی از آن است.

- **مثال ۴**: دو ترکیب خطی متفاوت از بردارهای $\underline{u}=(۱,۳)$ ، $\underline{v}=(۲,۷)$ و $\underline{w}=(۱,۵)$ را بیابید که برابر $\underline{b}=(۰,۱)$ شود.

چون بردارهای $\underline{u}$، $\underline{v}$ و $\underline{w}$ در یک صفحه قرار دارند، در نتیجه یک ترکیب خطی $c\underline{u}+d\underline{v}+e\underline{w}$ وجود دارد که برابر صفر می‌شود (به جز جواب بدیهی $c=d=e=0$). بنابراین اگر یک ترکیب خطی $C\underline{u}+D\underline{v}+E\underline{w}$ داشته باشیم که برابر $\underline{b}$ گردد، با استفاده از ضرایب c، d و e می‌توان بی‌نهایت ترکیب خطی دیگر نیز ساخت.

- آیا برای هر سه بردار دلخواه دو بعدی $\underline{u}$، $\underline{v}$ و $\underline{w}$ می‌توان دو ترکیب مختلف ساخت که برابر $\underline{b}$ شود.

اگر بردارهای $\underline{u}$، $\underline{v}$ و $\underline{w}$ هم‌راستا با یکدیگر بوده و با $\underline{b}$ هم‌راستا نباشند، نمی‌توان دو ترکیب مختلف را ساخت.

- **ضرب داخلی (ضرب نقطه‌ای)**: ضرب داخلی دو بردار $\underline{v}$ و $\underline{w}$، عدد $\underline{w}.\underline{v}$ است که برابر است با:

$$\begin{array}{l}\underline{v}=(v_1,v_2)\\ \underline{w}=(w_1,w_2)\end{array} \Rightarrow \underline{v}.\underline{w}=v_1w_1+v_2w_2$$

- **بردارهای متعامد**: دو بردار $\underline{v}$ و $\underline{w}$ متعامدند اگر ضرب داخلی آن‌ها برابر صفر باشد.

- **بردارهای یکه استاندارد**:

$$\begin{array}{l}\underline{i}=(1,0)\\ \underline{j}=(0,1)\end{array} \Rightarrow \underline{i}.\underline{j}=0 \Rightarrow$$

بردار یکه استاندارد در راستای محور x بر بردار یکه استاندارد در راستای محور y متعامد است.

- **تقارن ضرب داخلی**: ضرب داخلی یک عملگر متقارن است به عبارت دیگر $\underline{v}.\underline{w}=\underline{w}.\underline{v}$

- طول بردار: طول بردار $\underline{v}$ با $\|\underline{v}\|$ نمایش داده می شود و برابر با ریشه دوم ضرب داخلی $\underline{v}\cdot\underline{v}$ است. یعنی:

$$\|\underline{v}\| = \sqrt{\underline{v}\cdot\underline{v}} = (v_1^2+v_2^2+\cdots+v_n^2)^{\frac{1}{2}}$$

- بردار یکه: بردار $\underline{u}$ را یکه گوییم اگر اندازه آن برابر واحد باشد. به عبارت دیگر: $\underline{u}\cdot\underline{u}=1$

- بردار یکه با زاویه θ با محور x: برداریکه‌ای که با محور x زاویه θ می سازد برابر است با:

$$\underline{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \longrightarrow \begin{cases} \theta = 0 \Rightarrow \underline{u} = \underline{i} \\ \theta = 90^\circ \Rightarrow \underline{u} = \underline{j} \end{cases}$$

- بردار یکه همراستا: برای بردار دلخواه و غیرصفر $\underline{v}$، بردار یکه همراستا برابر است با:

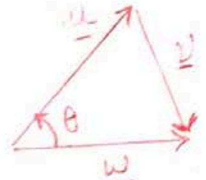
$$\underline{u} = \frac{\underline{v}}{\|\underline{v}\|}$$

- زاویه قائمه: اگر دو بردار $\underline{u}$ و $\underline{w}$ تشکیل یک زاویه قائمه دهند آنگاه $\underline{u}\cdot\underline{w}=0$ خواهد بود.

- ضرب داخلی با بردار صفر: ضرب داخلی بردار دلخواه $\underline{u}$ و بردار صفر $\underline{v}=\underline{0}$ برابر صفر است.

- ضرب داخلی بردارهای یکه: ضرب داخلی بردارهای یکه $\underline{u}$ و $\underline{w}$ که با یکدیگر زاویه θ را می سازند برابر است با:

$$\underline{u}\cdot\underline{w} = \cos\theta$$



$$\underline{u} + \underline{v} = \underline{w} \Rightarrow \underline{v} = \underline{w} - \underline{u}$$

$$\begin{cases} \text{قانون کسینوس} \Rightarrow |\underline{v}|^2 = \|\underline{u}\|^2 + \|\underline{w}\|^2 - 2\|\underline{w}\|\|\underline{u}\|\cos\theta \\ |v|^2 = (\underline{w}-\underline{u})\cdot(\underline{w}-\underline{u}) = \|w\|^2+\|\underline{u}\|^2 - 2w\cdot u \end{cases}$$

$$\Rightarrow \|w\|^2+\|u\|^2 - 2\,w\cdot u = \|u\|^2+\|w\|^2 - 2\|w\|\|\underline{u}\|\cos\theta \Rightarrow w\cdot u = \cos\theta$$

- ضرب داخلی بردارهای غیریکه: اگر $\underline{v}$ و $\underline{w}$ دو بردار غیرصفر باشند آنگاه (فرمول کسینوس):

$$\cos\theta = \frac{\underline{v}\cdot\underline{w}}{\|\underline{v}\|\|\underline{w}\|}$$

- نامساوی شوارتز: برای دو بردار دلخواه $\underline{v}$ و $\underline{w}$ نامساوی زیر برقرار است:

$$|\underline{v}\cdot\underline{w}| \leqslant \|v\|\,\|w\|$$

اثبات: فرمول کسینوس $\Rightarrow |\cos\theta| \leqslant 1 \Rightarrow \frac{|\underline{v}\cdot\underline{w}|}{\|\underline{v}\|\,\|\underline{w}\|} \leqslant 1$

$$\Rightarrow |\underline{v}\cdot\underline{w}| \leqslant \|\underline{v}\|\,\|\underline{w}\|$$

- نامساوی مثلثی: برای دو بردار دلخواه $\underline{v}$ و $\underline{w}$ نامساوی زیر برقرار است:

$$\|\underline{v}+\underline{w}\| \leqslant \|\underline{w}+\underline{v}\| \leqslant \|\underline{v}\| + \|\underline{w}\|$$

اثبات: فرمول کسینوس $\Rightarrow \cos\theta \leqslant 1 \Rightarrow \underline{v}\cdot\underline{w} \leqslant \|\underline{v}\|\,\|\underline{w}\|$

$$\xrightarrow[\text{جمع با } \|w\|^2+\|A\|^2]{\text{ضرب در } 2} \|w\|^2+\|\underline{v}\|^2+2\underline{v}\cdot\underline{w} \leqslant \|\underline{w}\|^2+\|\underline{v}\|^2+2\|\underline{v}\|\,\|\underline{w}\|$$

$$\Rightarrow \|\underline{v}+\underline{w}\|^2 \leqslant (\|\underline{v}\|+\|\underline{w}\|)^2 \Longrightarrow \|\underline{v}+\underline{w}\| \leqslant \|\underline{v}\|+\|\underline{w}\|$$

مثال ۱: آیا سه بردار $\underline{u}$ و $\underline{v}$ و $\underline{w}$ را می‌توان در فضای دو بعدی یافت به گونه‌ای که $\underline{u}\cdot\underline{v}<0$ و $\underline{v}\cdot\underline{w}<0$ و $\underline{u}\cdot\underline{w}<0$ برقرار باشد.

باتوجه به اینکه کل زاویه برابر 360° است بنابراین می‌توان سه بردار با زاویه بیش از 90° پیدا کرد که در مورد آنها هر سه رابطه برقرار است.

مثال ۲: سه عدد x و y و z را به گونه‌ای درنظر بگیرید که $x+y+z=0$ گردد. نشان دهید برای دو بردار $\underline{v}=(x,y,z)$ و $\underline{w}=(z,x,y)$ همواره رابطه $\frac{\underline{v}\cdot\underline{w}}{\|\underline{v}\|\,\|\underline{w}\|}=-\frac{1}{2}$ برقرار است.

$$\underline{v}\cdot\underline{w} = xz+xy+yz = \frac{1}{2}(x+y+z)^2 - \frac{1}{2}(x^2+y^2+z^2)$$

$$= -\frac{1}{2}\|\underline{v}\|\,\|\underline{w}\| \Longrightarrow \frac{\underline{v}\cdot\underline{w}}{\|\underline{v}\|\,\|\underline{w}\|} = -\frac{1}{2}$$

- همانطور که دیدیم ترکیب خطی بردارهای ستونی یکی از عملیات مهم در ارتباط با بردارها است. ماتریس‌ها روشی برای انجام ترکیب خطی ارائه می‌دهند.

$$x_1\underline{u} + x_2\underline{v} + x_3\underline{w} = \begin{bmatrix}\underline{u} & \underline{v} & \underline{w}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \underline{A}\,\underline{x} = \underline{b}$$

نگاه ستونی: ماتریس $\underline{A}$ بر روی بردار $\underline{x}$ عمل می‌کند. ترکیب خطی ستون‌های ماتریس $\underline{A}$ است.

چند مثال:

ماتریس تفاضل: $\underline{A} = \begin{bmatrix}1 & 0 & 0\\-1 & 1 & 0\\0 & -1 & 1\end{bmatrix} \Rightarrow \underline{A}\underline{x} = \begin{bmatrix}x_1\\x_2-x_1\\x_3-x_2\end{bmatrix} = \underline{b}$

ماتریس جمع تجمعی: $A = \begin{bmatrix}1 & 0 & 0\\1 & 1 & 0\\1 & 1 & 1\end{bmatrix} \Rightarrow \underline{A}\,\underline{x} = \begin{bmatrix}x_1\\x_1+x_2\\x_1+x_2+x_3\end{bmatrix} = \underline{b}$

ماتریس متوسط‌گیر: $A = \begin{bmatrix}1 & 0 & 0\\0.5 & 0.5 & 0\\0 & 0.5 & 0.5\end{bmatrix} \Rightarrow \underline{A}\,\underline{x} = \begin{bmatrix}x_1\\ \frac{x_1+x_2}{2}\\ \frac{x_2+x_3}{2}\end{bmatrix} = \underline{b}$

نگاه سطری: ضرب ماتریس $\underline{A}$ در بردار $\underline{x}$ را می‌توان به صورت ضرب داخلی سطرهای A در بردار $\underline{x}$ نیز تفسیر نمود.

$$\underline{A} = \begin{bmatrix}a_{11} & a_{12} & a_{13}\\a_{21} & a_{22} & a_{23}\\a_{31} & a_{32} & a_{33}\end{bmatrix} \Rightarrow \underline{A}\underline{x} = \begin{bmatrix}(a_{11},a_{12},a_{13})\cdot\underline{x}\\(a_{21},a_{22},a_{23})\cdot\underline{x}\\(a_{31},a_{32},a_{33})\cdot\underline{x}\end{bmatrix}$$

- معادلات خطی: رابطه $\underline{A}\,\underline{x} = \underline{b}$ یک دستگاه معادلات خطی را تشکیل می‌دهد. در مورد این دستگاه، دو نکته باید مورد توجه قرار گیرد.

۱) تعبیر ستونی دستگاه: $\underline{A}\,\underline{x}$ به معنی ترکیب خطی ستون‌های ماتریس $\underline{A}$ با استفاده از عناصر بردار $\underline{x}$ است.

۲) $\underline{b}$ در رابطه فوق معلوم بوده و ما به دنبال $\underline{x}$ هستیم (مسأله معکوس).

مثال: دستگاه معادلات خطی را برای ماتریس جمع تجمعی درنظر بگیرید.

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{\underline{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\underline{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\underline{b}} \Rightarrow \begin{cases} x_1 = b_1 \\ x_1 + x_2 = b_2 \\ x_1 + x_2 + x_3 = b_3 \end{cases} \Rightarrow x_2 = b_2 - b_1 \Rightarrow x_3 = b_3 - b_2$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}}_{\underline{A}^{-1}} \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\underline{b}} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\underline{x}} \Rightarrow \underline{A}^{-1}\ \underline{b} = \underline{x} *$$

ماتریس تفاضل $\leftarrow \underline{A}^{-1}$

- هر دستگاه معادلات خطی لزوماً جواب ندارد.
- هر دستگاه معادلات خطی لزوماً جواب منحصر به فرد ندارد.

- **ماتریس معکوس**: ماتریس تفاضل را معکوس ماتریس جمع تجمعی می‌نامیم. در حالت کلی‌تر اگر برای دستگاه معادلات $\underline{A}\,\underline{x} = \underline{b}$ بتوانیم ماتریس $\underline{A}^{-1}$ را بیابیم به گونه‌ای که $\underline{x} = \underline{A}^{-1}\underline{b}$، در اینصورت $\underline{A}^{-1}$ را ماتریس معکوس $\underline{A}$ می‌نامیم.

- **نتایج وجود ماتریس معکوس**:

۱- برای هر بردار $\underline{b}$، یک پاسخ به دستگاه معادلات $\underline{A}\,\underline{x} = \underline{b}$ وجود دارد.

۲- با استفاده از ماتریس معکوس این پاسخ به صورت $\underline{x} = \underline{A}^{-1}\underline{b}$ به دست می‌آید.

- **ماتریس تفاضل گردشی**:

$$\text{تفاضل گردشی}: \quad \underline{C}\,\underline{x} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

* در این حالت نمی‌توان ماتریسی با نام $\underline{C}^{-1}$ به دست آورد.

- **بررسی جواب‌ها**:

$$\begin{bmatrix} x_1 - x_3 \\ x_2 - x_1 \\ x_3 - x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \Rightarrow \begin{cases} x_1 = b_1 + c \\ x_2 = b_1 + b_2 + c \\ x_3 = c \end{cases}, \forall c$$

حالت اول: $b_1 + b_2 + b_3 = 0 \Leftarrow$

حالت دوم: $b_1 + b_2 + b_3 \neq 0 \Leftarrow$ دستگاه جواب ندارد.

* معادله صفحه حاصل از دو ستون اول:

$\underline{u}=(1,-1,0)$, $\underline{v}=(0,1,-1) \Rightarrow \underline{n}=(1,1,1) \Rightarrow$ معادله صفحه $\Leftarrow$ $x+y+z=0$

ستون سوم روی صفحه قرار دارد ← فضای ستون‌ها همان صفحه فوق است.

$\underline{b}$ روی صفحه است ← $b_1+b_2+b_3=0$ ← با بینهایت روش می‌توان $\underline{b}$ را ساخت.

$\underline{b}$ روی صفحه نیست ← $b_1+b_2+b_3\neq 0$ ← دستگاه جواب ندارد.

- استقلال و وابستگی:

$\underline{w}$ روی صفحه متشکل از $\underline{u}$ و $\underline{v}$ نیست ← استقلال

$$[\underline{u}\ \ \underline{v}\ \ \underline{w}]=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}=\text{ماتریس تفاضل}$$

$\underline{w}^*$ روی صفحه متشکل از $\underline{u}$ و $\underline{w}$ است. ← وابستگی

$$[\underline{u}\ \ \underline{v}\ \ \underline{w}^*]=\begin{bmatrix}1&0&-1\\-1&1&0\\0&-1&1\end{bmatrix}=\text{ماتریس تفاضل گردشی}$$

w^* آزادی عمل بیشتری در فضای ستونی ماتریس ایجاد نمی‌کند.

- تعاریف غیر رسمی:

$\underline{u}$ و $\underline{v}$ و $\underline{w}$ مستقلند زیرا هیچ ترکیب خطی جز $0\underline{u}+0\underline{v}+0\underline{w}=\underline{0}$ منجر به $\underline{b}=\underline{0}$ نمی‌گردد.

$\underline{u}$ و $\underline{v}$ و $\underline{w}^*$ وابسته‌اند زیرا $\underline{u}+\underline{v}+\underline{w}^*=\underline{0}$ منجر به $\underline{b}=\underline{0}$ می‌گردد.

- تعمیم (ماتریس $n\times n$):

اگر ستون‌های ماتریس $\underline{A}$ مستقل باشند آنگاه $\underline{A}\underline{x}=\underline{0}$ تنها یک جواب دارد و ماتریس $\underline{A}$ معکوس‌پذیر است.

اگر ستون‌های ماتریس $\underline{C}$ وابسته باشند آنگاه $\underline{C}\underline{x}=0$ بینهایت جواب دارد و ماتریس $\underline{C}$ تکین است.

- مثال 1: اگر (a,b) ضریبی از (c,d) باشد و $abcd\neq 0$ باشد، نشان دهید (a,c) ضریب (b,d) است. به عبارت دیگر اگر $\begin{bmatrix}a&b\\c&d\end{bmatrix}$ سطرهای وابسته داشته باشد آنگاه ستون‌های ماتریس نیز وابسته‌اند.

$$(a,b)=k(c,d)\Rightarrow\begin{cases}a=kc\\b=kd\end{cases}\Rightarrow a/c=b/d\Rightarrow ad=bc$$

$$\xRightarrow{\div bd}\frac{a}{b}=\frac{c}{d}=k'\Rightarrow\begin{cases}a=k'b\\c=k'd\end{cases}\Rightarrow(a,c)=k'(b,d)$$

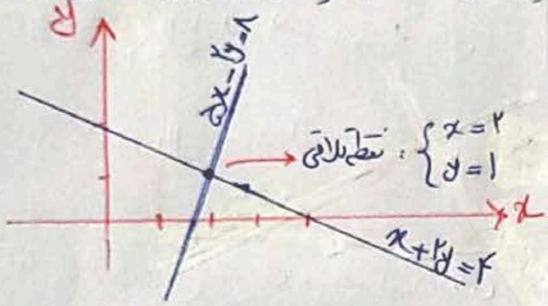
حل دستگاه معادلات خطی (n معادله n مجهول)

- مسأله اصلی در جبر خطی حل دستگاه معادلات است. دستگاه های مورد نظر خطی هستند به عبارت دیگر از ضرب کردن اعداد در متغیرها به دست می آیند (ضرب متغیرها وجود ندارد).

- نمونه دستگاه معادلات خطی (۲ معادله ۲ مجهول):

$$x + ۲y = ۴$$
$$۵x - ۲y = ۸$$

- تعبیر سطری: هر سطر یک خط را مشخص می کند که این دو خط در یک نقطه (جواب) تلاقی می کنند.



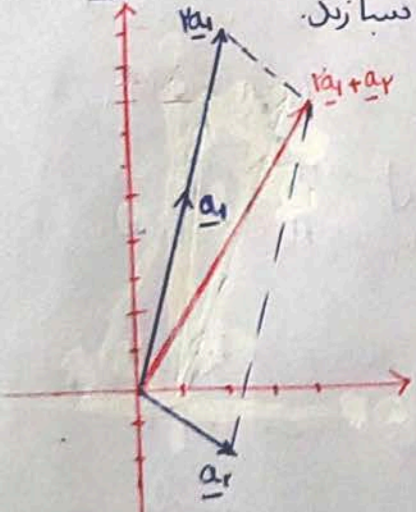
خط اول: $(۱,۲).(x,y) = ۴$

خط دوم: $(۵,-۲)(x,y) = ۸$

- تعبیر ستونی:

$$x\underbrace{\begin{bmatrix} ۱ \\ ۵ \end{bmatrix}}_{a_1} + y\underbrace{\begin{bmatrix} ۲ \\ -۲ \end{bmatrix}}_{a_2} = \begin{bmatrix} ۴ \\ ۸ \end{bmatrix} = \underline{b}$$

سؤال: دو بردار $a_1=(۱,۵)$، $a_2=(۲,-۲)$ چگونه ترکیب شوند که بردار $\underline{b}$ را بسازند.



$$\begin{matrix} x= \to xa_1 = \begin{bmatrix} ۲ \\ ۱۰ \end{bmatrix} \\ y=۱ \to ya_2 = \begin{bmatrix} ۲ \\ -۲ \end{bmatrix} \end{matrix} \Rightarrow xa_1 + ya_2 = \begin{bmatrix} ۴ \\ ۸ \end{bmatrix} = \underline{b}$$

- تعبیر سطری ← جواب تقاطع n ابرصفحه
تعبیر ستونی ← جواب ترکیب خطی n بردار n بعدی

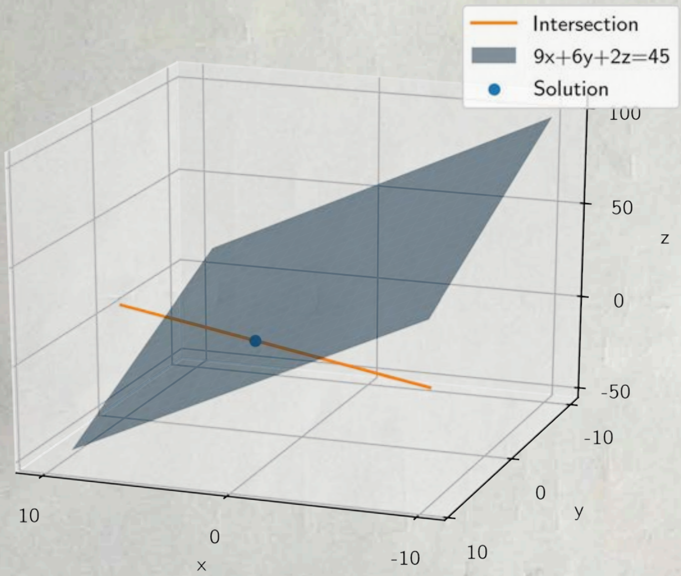
- نمایش ماتریسی:

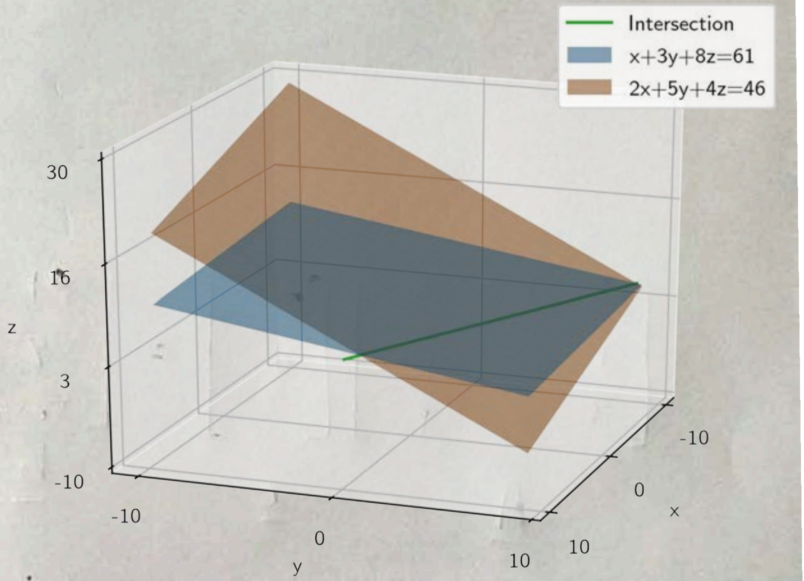
$$\underset{\underline{A}}{\begin{bmatrix} ۱ & ۲ \\ ۵ & -۲ \end{bmatrix}} \underset{\underline{x}}{\begin{bmatrix} x \\ y \end{bmatrix}} = \underset{\underline{b}}{\begin{bmatrix} ۴ \\ ۸ \end{bmatrix}}$$

- ۳ معادله، ۳ مجهول

$$x + ۳y + ۸z = ۶۱$$
$$۲x + ۵y + ۴z = ۴۶$$
$$۹x + ۶y + ۲z = ۴۵$$

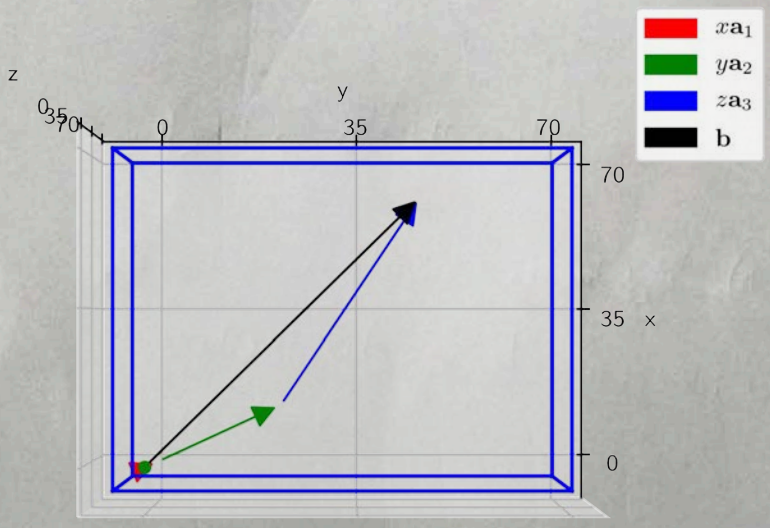
- تفسیر سطری: هر سطر یک صفحه را در فضای ۳ بعدی مشخص می‌کند. جواب دستگاه محل تقاطع این صفحات است.

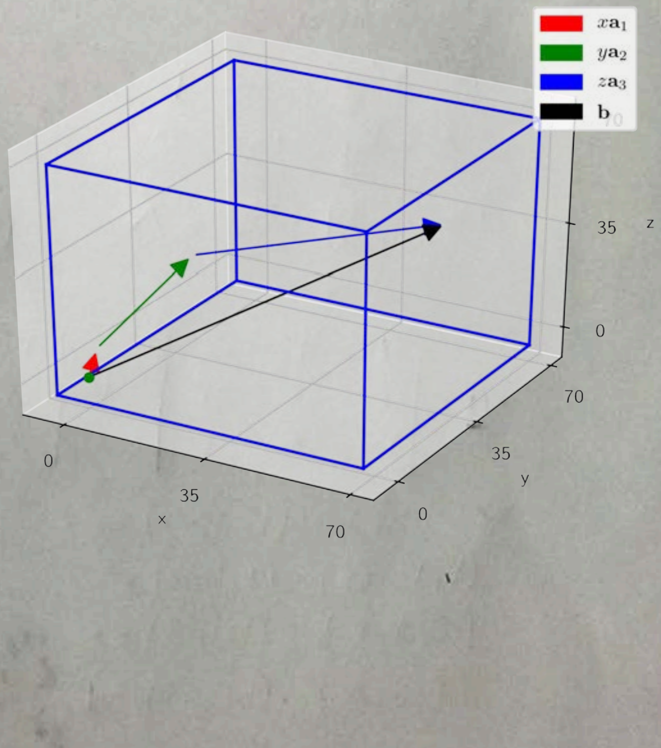


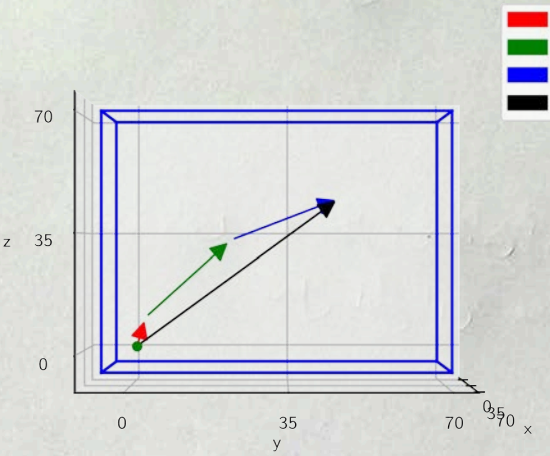


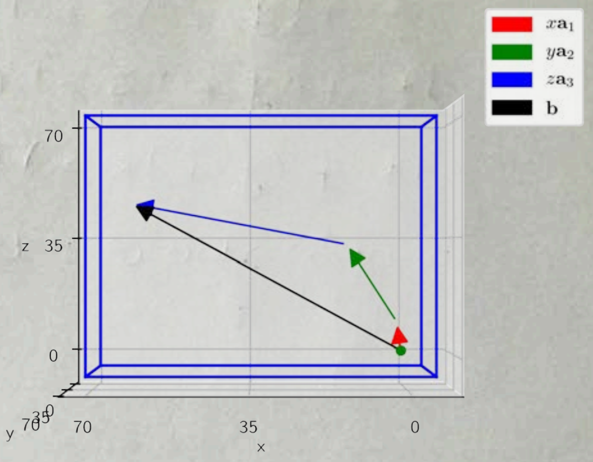
$$x\underset{\underline{a_1}}{\begin{bmatrix}۱\\۲\\۹\end{bmatrix}} + y\underset{\underline{a_2}}{\begin{bmatrix}۳\\۵\\۶\end{bmatrix}} + z\underset{\underline{a_3}}{\begin{bmatrix}۸\\۴\\۲\end{bmatrix}} = \begin{bmatrix}۶۱\\۴۶\\۴۵\end{bmatrix} = \underline{b}$$

- تفسیر ستونی: سه بردار مقابل باید به گونه‌ای ترکیب بشوند که بردار $\underline{b}$ را بسازند.









$$\underset{\underline{A}}{\begin{bmatrix} 1 & 3 & 8 \\ 2 & 5 & 4 \\ 9 & 6 & 2 \end{bmatrix}} \underset{\underline{x}}{\begin{bmatrix} x \\ y \\ x \end{bmatrix}} = \underset{\underline{b}}{\begin{bmatrix} 61 \\ 46 \\ 45 \end{bmatrix}}$$

- ضرب ماتریسی :

ضرب با استفاده از سطرها $\Rightarrow \underline{A}\,\underline{x} = \begin{bmatrix} (\underline{A}\ \text{سطر اول}) \cdot \underline{x} \\ (\underline{A}\ \text{سطر دوم}) \cdot \underline{x} \\ (\underline{A}\ \text{سطر سوم}) \cdot \underline{x} \end{bmatrix}$

ضرب با استفاده از ستون‌ها $\Rightarrow \underline{A}\,\underline{x} = x\,(\underline{A}\ \text{ستون اول}) + y\,(\underline{A}\ \text{ستون دوم}) + z\,(\underline{A}\ \text{ستون سوم})$

- مثال ۱ : دستگاه معادلات مقابل را در نظر بگیرید.

$$x + 3y + 5z = 4$$
$$x + 2y - 3z = 5$$
$$2x + 5y + 2z = 8$$

* از دیدگاه ستونی و با استفاده از ضرب داخلی بردار (۱، ۱، ۱-) با ستون‌های ماتریس $\underline{A}$ و بردار $\underline{b}$ ، نشان دهید که این دستگاه جواب ندارد.

$$\underline{A} = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 2 & -3 \\ 2 & 5 & 2 \end{bmatrix} \quad \underline{b} = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$$

$$\left.\begin{aligned} (1, 1, 2) \cdot (1, 1, -1) &= 0 \\ (3, 2, 5) \cdot (1, 1, -1) &= 0 \\ (5, -3, 2) \cdot (1, 1, -1) &= 0 \\ (4, 5, 8) \cdot (1, 1, -1) &= 1 \end{aligned}\right\} \Rightarrow$$

بردار (۱، ۱، ۱-) بر ستون‌های $\underline{A}$ عمود است اما بر $\underline{b}$ عمود نیست. بنابراین نمی‌توان بردار $\underline{b}$ را از ترکیب خطی ستون‌های $\underline{A}$ ساخت.

* از دیدگاه سطری نشان دهید که دستگاه جواب ندارد.

ابتدا خط مشترک در دو صفحه اول را به دست می آوریم.

بردار عمود بر صفحه اول $= \underline{n}_۱ = (۱,۳,۵)$

بردار عمود بر صفحه دوم $= \underline{n}_۲ = (۱,۲,-۳)$

$\Rightarrow$ جهت خط مشترک $= \begin{vmatrix} i & j & k \\ ۱ & ۳ & ۵ \\ ۱ & ۲ & -۳ \end{vmatrix} = i(-۱۹)-j(-۸)+k(-۱)$

$= (-۱۹, ۸, -۱) = \underline{s}$

بردار عمود بر صفحه سوم $= \underline{n}_۳ = (۲,۵,۲)$

$\underline{s} = (-۱۹,۸,-۱)$

$\Rightarrow n_۳ \cdot \underline{s} = ۹۰ \longrightarrow$ خط مشترک یا تماماً در صفحه سوم است یا هیچ نقطه اشتراکی با آن ندارد.

در معادله دو صفحه اول $z=۰$ قرار می دهیم:

$$\begin{cases} x+۳y=۴ \\ x+۲y=۵ \end{cases}$$

$y=-۱ \Rightarrow x = ۵-۲y = ۷ \Rightarrow$

نقطه $(۷,-۱,۰)$ روی خط مشترک است.

چون نقطه $(۷,-۱,۰)$ روی صفحه سوم قرار ندارد پس خط اشتراک دو صفحه اول هیچ نقطه اشتراکی با صفحه سوم ندارد. در نتیجه دستگاه جواب ندارد.

– حل دستگاه معادلات با استفاده از روش حذف:

$$\begin{aligned} x-y &= -۳ \\ ۲x+۳y &= ۱۴ \end{aligned} \xrightarrow[\text{کم کردن حاصل از معادله دوم}]{\text{ضرب معادله اول در ۲ +}} \begin{aligned} x-y &= -۳ \\ +y &= ۲۰ \end{aligned}$$

حال از انتها شروع به مشخص کردن متغیرها می نماییم و مقادیر به دست آمده را در رابطه های بالاتر جایگذاری می کنیم (back substitution).

معادله دوم $\Rightarrow y=۴$

معادله اول $\Rightarrow x=-۳+y=۱$

– گام های روش حذف (نگاه کلی):

۱- تولید یک دستگاه معادلات بالامثلثی

۲- بدست آوردن متغیرها از آخرین معادله

$$\begin{aligned} &d_{۱۱}x_۱ + d_{۱۲}x_۲ + \cdots + d_{۱n}x_n = \\ &۰ + d_{۲۲}x_۲ + \cdots + d_{۲n}x_n = \\ &\vdots \\ &۰ + ۰ + \cdots + d_{nn}x_n = \end{aligned}$$

- گام های الگوریتم حذف: دستگاه معادلات مقابل را درنظر بگیرید.

$$a_{۱۱}x_۱ + a_{۱۲}x_۲ + \cdots + a_{۱(n-۱)}x_{n-۱} + a_{۱n}x_n = b_۱$$

$$a_{۲۱}x_۱ + a_{۲۲}x_۲ + \cdots + a_{۲(n-۱)}x_{n-۱} + a_{۲n}x_n = b_۲$$

$$\vdots$$

$$a_{n۱}x_۱ + a_{n۲}x_۲ + \cdots + a_{n(n-۱)}x_{n-۱} + a_{nn}x_n = b_n$$

۱- اولین معادله را درنظر بگیرید. پاشنه (pivot) معادله را تعیین نمایید، این مقدار در رابطه بالا برابر با $a_{۱۱}$ است (پاشنه اولین ضریب غیرصفر در معادله مورد بررسی است). متغیری که پاشنه در آن ضرب می‌شود در ادامه از معادلات بعدی حذف می‌گردد.

۲- برای $n \geqslant i \geqslant ۲$، ضریب (Multiplier) را به صورت $\frac{\text{ضریب متغیر پاشنه در رابطه } i}{\text{Pivot}}$ به دست آورد. و از معادله i-ام کم نمایید. این کار باعث حذف متغیر پاشنه در رابطه می‌گردد.

۳- به معادله دوم رفته و پاشنه را در آن تعیین می‌نماییم. مطابق روال فوق متغیر پاشنه را از معادلات بعدی حذف می‌کنیم.

۴- روال فوق را تا زمان دستیابی به فرم بالا مثلثی ادامه می‌دهیم.

۵- متغیرها را از رابطه انتهایی شروع به تعیین نمایید.

مثال ۱.

$$\begin{cases} x + ۳y + ۸z = ۶۱ & \rightarrow \text{Pivot} = ۱ \\ ۲x + ۵y + ۴z = ۴۶ & \rightarrow \text{Multiplier} = ۲ \\ ۶x + ۶y + ۲z = ۴۵ & \rightarrow \text{Mutiplier} = ۶ \end{cases}$$

$$\begin{cases} x + ۳y + ۸z = ۶۱ \\ ۰x - y - ۱۲z = -۷۶ & \rightarrow \text{Pivot} = -۱ \\ ۰x - ۲۱y - ۷۰z = -۵۰۴ & \rightarrow \text{Multiplier} = +۲۱ \end{cases}$$

$$x + ۳y + ۸z = ۶۱ \Rightarrow x = ۶۱ - ۳(۴) - ۸(۶) = ۱ \quad ③$$

$$۰x - y - ۱۲z = -۷۶ \Rightarrow y = ۷۶ - ۱۲z = ۷۶ - ۷۲ = ۴ \quad ②$$

$$۰x + ۰y + ۱۸۲z = ۱۰۹۲ \Rightarrow z = \frac{۱۰۹۲}{۱۸۲} = ۶ \quad ①$$

$\uparrow$ مسیر تعیین متغیرها

مثال ۲: در این مثال با شرایطی مواجه می‌شویم که نیاز به تغییر ترتیب معادلات داریم.

$$\begin{cases} x + ۳y + ۸z = ۶۱ & \rightarrow \text{pivot} = ۱ \\ ۲x + ۶y + ۱۰z = ۴۶ & \rightarrow \text{Multiplier} = ۲ \\ ۹x + ۶y + ۲z = ۴۵ & \rightarrow \text{Multiplier} = ۹ \end{cases}$$

$$\begin{cases} x + ۳y + ۸z = ۹۱ \\ ۰x + ۰y - ۶z = -۷۶ & \rightarrow \\ ۰x - ۲۱y - ۷۰z = -۵۰۴ \end{cases}$$

متغیر y ساختار بالا مثلثی را برهم زده است. Pivot برای معادله دوم ۶- و برای معادله سوم ۲۱- متناظر با y است. به منظور رفع این مشکل معادله دوم و سوم را جابه‌جا می‌کنیم.

$$\begin{cases} x + ۳y + ۸z = ۶۱ \\ ۰x - ۲۱y - ۷۰z = -۵۰۴ & \rightarrow \\ ۰x + ۰y - ۶z = -۷۶ \end{cases}$$

معادلات مقابل ساختار بالا مثلثی داشته و می‌توانیم به سادگی از انتها متغیرها را به دست آوریم.

مثال ۳: در این مثال شرایطی را می‌بینیم که دستگاه جواب ندارد.

$$\begin{cases} x + ۳y + ۵z = ۴ & \rightarrow \text{Pivot} = ۱ \\ x + ۲y - ۳z = ۵ & \rightarrow \text{Multiplier} = ۱ \\ ۲x + ۵y + ۲z = ۸ & \rightarrow \text{Multiplier} = ۲ \end{cases}$$

$$\begin{cases} x + ۳y + ۵z = ۴ \\ ۰x - y - ۸z = ۵ & \rightarrow \text{Pivot} = -۱ \\ ۰x - y - ۸z = ۰ & \rightarrow \text{Multiplier} = ۱ \end{cases}$$

$$\begin{cases} x + ۳y + ۵z = ۴ \\ ۰x - y - ۸z = ۵ \\ ۰x + ۰y + ۰z = -۵ & \rightarrow ۰ = -۵ \rightarrow \text{دستگاه جواب ندارد.} \end{cases}$$

توجه: اگر دستگاه بی‌نهایت جواب داشته باشد در سطر آخر معادلات پس از اعمال روش حذف با معادله‌ای به صورت $۰z = ۰$ خواهیم رسید که همواره برقرار است.

- **دستگاه معادلات تکین** : اگر پس از حذف در دستگاه معادلات حداقل یک متغیر وجود داشته باشد که متناظر با آن پاشنه‌ای نباشد، دستگاه تکین است (یا جواب ندارد و یا بی‌نهایت جواب دارد).

- **دستگاه معادلات غیرتکین** : اگر پس از حذف در دستگاه معادلات متناظر با هر متغیر یک پاشنه وجود داشته باشد دستگاه غیرتکین خواهد بود.

- **نمایش ماتریسی عملیات حذف** :

- **نمایش ماتریس** :

$$\underline{A} = \begin{bmatrix} a_{11} & a_{۱۲} & \cdots & a_{۱n} \\ a_{۲۱} & a_{۲۲} & \cdots & a_{۲n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n۲} & \cdots & a_{nn} \end{bmatrix} \quad , \quad a_{ij} = A(i,j)$$

- **بیان فرآیند حذف با استفاده از ضرب ماتریسی** :

$$\underset{\underline{A}}{\begin{bmatrix} 1 & ۳ & ۸ \\ ۲ & ۵ & ۴ \\ ۹ & ۶ & ۲ \end{bmatrix}} \underset{\underline{x}}{\begin{bmatrix} x \\ y \\ z \end{bmatrix}} = \underset{\underline{b}}{\begin{bmatrix} ۹۱ \\ ۴۶ \\ ۴۵ \end{bmatrix}} \begin{matrix} \rightarrow \text{Pivot} = ۱ \\ \rightarrow \text{Multiplier} = ۲ \\ \end{matrix}$$

$$E = \begin{bmatrix} 1 & ۰ & ۰ \\ -۲ & 1 & ۰ \\ ۰ & ۰ & 1 \end{bmatrix}$$

در این گام می‌بایست دو برابر سطر اول را از سطر دوم کم کنیم. ماتریس حذف متناظر مثل ماتریس همانی است که عنصر $E(۲و۱)$ آن برابر Multiplier_ قرار داده شده است.

بررسی عملکرد روی $\underline{b}$:

$$E\underline{b} = \begin{bmatrix} 1 & ۰ & ۰ \\ -۲ & 1 & ۰ \\ ۰ & ۰ & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_۲ \\ b_۳ \end{bmatrix} = \begin{bmatrix} b_1 \\ b_۲ - ۲b_1 \\ b_۳ \end{bmatrix}$$

سمت چپ معادله :

$$\underline{E}\,\underline{A}x = (\underline{E}\,\underline{A})x = \underline{E}(\underline{A}\,\underline{x})$$

توجه : برای رابطه فوق از قاعده شرکت‌پذیری ضرب ماتریسی استفاده کرده‌ایم به عبارت دیگر :

$$\underline{A}(\underline{B}\,\underline{C}) = (\underline{A}\,\underline{B})\underline{C}$$

برای بررسی سمت چپ معادله از ویژگی زیر استفاده می‌کنیم :

$$\underline{A}\underline{B} = \underline{A}[\underline{b}_1, \underline{b}_۲, \ldots, \underline{b}_n]$$
$$= [\underline{A}\underline{b}_1, \underline{A}\underline{b}_۲, \ldots, \underline{A}\underline{b}_n]$$

بنابراین مشابه توجیه $\underline{E}\underline{b}$ می‌توان برای توجیه $\underline{E}\,\underline{A}$ نیز استفاده نمود.

- عملیات تغییر سطر با استفاده از ماتریس:

$$P_{۲۳} = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۰ & ۰ & ۱ \\ ۰ & ۱ & ۰ \end{bmatrix}$$

ماتریس مقابل سطر دوم و سوم را جابه‌جا می‌کند.

- ماتریس افزوده: از آنجایی که عملیات صورت گرفته در روش حذف برای ستون‌های ماتریس $\underline{A}$ مشابه بردار ستونی $\underline{b}$ است برای یکپارچه کردن عملیات می‌توانیم ماتریس افزوده را به صورت زیر تعریف کنیم:

$$\text{ماتریس افزوده} = \begin{bmatrix} \underline{A} & \underline{b} \end{bmatrix} = \underline{\hat{A}}$$

در این صورت اعمال سطرهای ماتریس E بر ماتریس $[\underline{A} \;\; \underline{b}]$ سطرهای ماتریس $[E\underline{A} \;\; E\underline{b}]$ را می‌دهد.

به طور معادل ماتریس E بر روی ستون‌های ماتریس $[\underline{A} \;\; \underline{b}]$ عمل می‌کند و ستون‌های ماتریس $[\underline{E}\underline{A} \;\; \underline{E}\underline{b}]$ را می‌دهد.

- مثال:

$$\begin{cases} x + ۳y + ۸z = ۴۱ \\ ۲x + ۴y + ۱۰z = ۴۶ \\ ۹x + ۴y + ۲z = ۴۵ \end{cases} \xrightarrow{\text{ماتریس افزوده}} \underline{\hat{A}} = \begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۲ & ۶ & ۱۰ & ۴۶ \\ ۹ & ۶ & ۲ & ۴۵ \end{bmatrix}$$

$$\begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۲ & ۶ & ۱۰ & ۴۶ \\ ۹ & ۶ & ۲ & ۴۵ \end{bmatrix} \begin{matrix} \rightarrow \text{Pivot} = ۱ \\ \rightarrow \text{Multiplier} = ۲ \end{matrix} \Rightarrow E_{۲۱} = \begin{bmatrix} ۱ & ۰ & ۰ \\ -۲ & ۱ & ۰ \\ ۰ & ۰ & ۱ \end{bmatrix} \Rightarrow E_{۲۱}\hat{A} = \begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & ۰ & -۶ & -۷۶ \\ ۹ & ۶ & ۲ & ۴۵ \end{bmatrix}$$

$$\begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & ۰ & -۶ & -۷۶ \\ ۹ & ۶ & ۲ & ۴۵ \end{bmatrix} \begin{matrix} \rightarrow \text{Pivot} = ۱ \\ \rightarrow \text{Multiplier} = ۹ \end{matrix} \Rightarrow E_{۳۱} = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۰ & ۱ & ۰ \\ -۹ & ۰ & ۱ \end{bmatrix} \Rightarrow E_{۳۱}E_{۲۱}\hat{A} = \begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & ۰ & -۶ & -۷۶ \\ ۰ & -۲۱ & -۷۰ & -۵۰۴ \end{bmatrix}$$

$$\begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & ۰ & -۶ & -۷۶ \\ ۰ & -۲۱ & -۷۰ & -۵۰۴ \end{bmatrix} \longrightarrow P_{۲۳} = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۰ & ۰ & ۱ \\ ۰ & ۱ & ۰ \end{bmatrix} \Rightarrow P_{۲۳}E_{۳۱}E_{۲۱}\hat{A} = \begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & -۲۱ & -۷۰ & -۵۰۴ \\ ۰ & ۰ & -۶ & -۷۶ \end{bmatrix}$$

$$\begin{bmatrix} ۱ & ۳ & ۸ & ۴۱ \\ ۰ & -۲۱ & -۷۰ & -۵۰۴ \\ ۰ & ۰ & -۶ & -۷۶ \end{bmatrix} \xrightarrow{\text{فرم ماتریسی بالامثلثی است.}} \begin{bmatrix} ۱ & ۳ & ۸ \\ ۰ & -۲۱ & -۷۰ \\ ۰ & ۰ & -۶ \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ۴۱ \\ -۵۰۴ \\ -۷۶ \end{bmatrix} \rightarrow \begin{cases} x = ۱۴,۳۳ \\ y = -۱۸,۲۲ \\ z = ۱۲,۹۷ \end{cases}$$

- قواعد ضرب ماتریسی :

- ماتریس : یک آرایه مستطیلی از اعداد یا ورودی‌هاست. هنگامی که ماتریس $\underline{A}$، m سطر و n ستون دارد آن را یک ماتریس "m در n"

- جمع ماتریس‌ها: ماتریس‌های هم شکل با یکدیگر جمع می‌شوند. این جمع به صورت المان به المان انجام می‌شود

- ضرب عددی : یک ماتریس به روشی کاملاً مشابه یک بردار در یک عدد ضرب می‌شود

- بردار ستونی : بردار ستونی حالت خاص یک ماتریس با $n=1$ است.

- بردار سطری : بردار سطری حالت خاص یک ماتریس با $m=1$ است.

- ماتریس $-\underline{A}$ از ضرب عددی ماتریس $\underline{A}$ در عدد $c=-1$ به دست می‌آید.

- مجموع دو ماتریس $\underline{A}$ و $-\underline{A}$ ماتریس صفر ($\underline{0}$) است.

- عنصر ماتریس $\underline{A}$ در سطر i-ام و ستون j-ام را با $A(i,j)$ یا a_{ij} نمایش می‌دهیم.

- ضرب ماتریسی :

- امکان‌پذیری : ضرب ماتریسی ماتریس $\underline{B}$ در ماتریس $\underline{A}$ را که به صورت $\underline{A}\,\underline{B}$ نمایش می‌دهیم در صورتی امکان‌پذیر است که تعداد ستون‌های $\underline{A}$ با سطرهای $\underline{B}$ برابر باشد.

- اگر $\underline{A}$ یک ماتریس $m \times n$ و $\underline{B}$ یک ماتریس $n \times p$ باشد در اینصورت یک ماتریس با ابعاد $m \times p$ خواهد بود.

- روش اول محاسبه ضرب ماتریسی :

$$\left.\begin{array}{l} \underline{A} : m \times n \\ \underline{B} : n \times p \\ \underline{C} = \underline{A}\,\underline{B} : m \times p \end{array}\right\} \Rightarrow C(i,j) = C_{ij} = \left(\underline{A} \ \text{ردیف } i\text{-ام}\right) . \left(\underline{B} \ \text{ستون } j\text{-ام}\right)$$

- ضرب داخلی: ضرب ماتریسی یک بردار سطری در یک بردار ستونی را ضرب داخلی می‌نامیم.

- ضرب خارجی: ضرب ماتریسی یک بردار ستونی در یک بردار سطری را ضرب خارجی می‌نامیم.

- روش دوم محاسبه ضرب ماتریسی:

$$\underline{A}\,\underline{B} = \underline{A}\left[\underline{b}_1\ \underline{b}_2 \cdots \underline{b}_p\right] = \left[\underline{A}\underline{b}_1\ \ \underline{A}\underline{b}_2 \cdots \underline{A}\underline{b}_p\right]$$

- روش سوم محاسبه ضرب ماتریسی:

$$\underline{A}\,\underline{B} = \begin{bmatrix} \underline{A}\ \text{سطر اول} \\ \vdots \\ \underline{A}\ \text{سطر n ام} \end{bmatrix} B = \begin{bmatrix} \underline{A}\underline{B}\ \text{سطر اول} \\ \vdots \\ \underline{A}\underline{B}\ \text{سطر n ام} \end{bmatrix}$$

- روش چهارم محاسبه ضرب ماتریسی:

$$\underline{A}\,\underline{B} = \left[a_1\ a_2 \cdots a_m\right] \begin{bmatrix} B\ \text{سطر اول} \\ \\ \underline{B}\ \text{سطر m ام} \end{bmatrix}$$

$$= a_1\left[B\ \text{سطر اول}\right] + \cdots + a_m\left[\underline{B}\ \text{سطر m ام}\right]$$

- قوانین عملیات ماتریسی:

- $\underline{A}+\underline{B} = \underline{B}+\underline{A}$ خاصیت جابه‌جایی پذیری جمع ماتریسی
- $c(\underline{A}+\underline{B}) = c\underline{A}+c\underline{B}$ توزیع پذیری ضرب عددی
- $\underline{A}+(\underline{B}+\underline{C}) = (\underline{A}+\underline{B})+\underline{C}$ شرکت پذیری جمع ماتریسی
- $\underline{A}\underline{B} \neq \underline{B}\underline{A}$ ضرب ماتریسی خاصیت جابه‌جایی پذیر نیست.
- $\underline{A}(\underline{B}+\underline{C}) = \underline{A}\underline{B}+\underline{A}\underline{C}$ توزیع پذیری ضرب ماتریسی از چپ
- $(\underline{A}+\underline{B})\underline{C} = \underline{A}\underline{C}+\underline{B}\underline{C}$ توزیع پذیری ضرب ماتریسی از راست
- $\underline{A}(\underline{B}\underline{C}) = (\underline{A}\,\underline{B})\underline{C} = \underline{A}\underline{B}\underline{C}$ شرکت پذیری ضرب ماتریسی

- توان های یک ماتریس مربعی :

$$\left.\begin{array}{l}\underline{A}(\underline{B}\,\underline{C}) = (\underline{A}\,\underline{B})\underline{C} \\ \underline{A} = \underline{B} = \underline{C} : n \times n\end{array}\right\} \Rightarrow \left.\begin{array}{l}\underline{A}(\underline{A}\,\underline{A}) = (\underline{A}\,\underline{A})\underline{A} \\ \underline{A}\,\underline{A} = \underline{A}^2\end{array}\right\} \Rightarrow \underline{A}^2\underline{A} = \underline{A}\,\underline{A}^2 = \underline{A}^3$$

$$* \ \underline{A}^P = \underline{A}\,\underline{A} \dots \underline{A} \ (\text{مرتبه } P)$$

$$* \ \underline{A}^P\,\underline{A}^q = \underline{A}^{P+q}$$

$$(\underline{A}^P)^q = A^{Pq}$$

درصورتی که A^{-1} موجود باشد در سه رابطه مقابل P و q می توانند اعداد صحیح منفی شوند و طبق قرار داد $A^0 = I$

- ضرب بلوکی :

اگر بتوانیم ماتریس های $\underline{A}$ و $\underline{B}$ را به گونه ای بلوکی کنیم که بلوک ها در یکدیگر ضرب شونده باشند آنگاه می توانیم از ضرب بلوکی، ضرب ماتریس ها را به دست آوریم

$$\underline{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \Rightarrow \underline{A}\,\underline{B} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix}$$

$$\underline{B} = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

- بلوکی کردن یک ماتریس برخی مواقع موجب در درک عملکرد آن می شود.

به عنوان مثال ماتریس $\underline{A}$ را به صورت زیر در نظر بگیرید:

$$\underline{A} = \left[\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{A_{11}} & \multicolumn{2}{c|}{A_{12}} & \multicolumn{2}{c}{A_{13}} \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \multicolumn{2}{c|}{A_{21}} & \multicolumn{2}{c|}{A_{22}} & \multicolumn{2}{c}{A_{23}} \end{array}\right] \Rightarrow \underline{A}.\underline{B} = \begin{bmatrix} B_{11} + B_{21} + B_{31} \\ B_{11} + B_{21} + B_{31} \end{bmatrix}$$

$$B = \left[\begin{array}{cc} b_1 & B_{11} \\ b_2 & \\ \hline b_3 & B_{21} \\ b_4 & \\ \hline b_5 & B_{31} \\ b_6 & \end{array}\right]$$

- حالت خاص ضرب بلوکی:

$$A=\begin{bmatrix} | & & | \\ a_1 & \cdots & a_m \\ | & & | \end{bmatrix} \qquad \underline{B}=\begin{bmatrix} - & b_1 & - \\ & | & \\ - & b_m & - \end{bmatrix} \Rightarrow \underline{A}\cdot\underline{B}=\left[a_1b_1+\cdots+a_mb_m\right] \rightarrow$$

روش چهارم محاسبه ضرب ماتریسی

- حذف بلوکی (مکمل شور Schur):

$$A=\left[\begin{array}{c|cc} \overset{A}{1} & a & \overset{B}{} b \\ ۳\,C & c & \overset{D}{} d \\ ۴ & e & f \end{array}\right]$$

$$\left[\begin{array}{c|c} I & \underline{0} \\ \hline -CA^{-1} & I \end{array}\right]\left[\begin{array}{c|c} \underline{A} & \underline{B} \\ \hline \underline{C} & \underline{D} \end{array}\right]=\left[\begin{array}{c|c} A & B \\ \hline \underline{0} & \underline{D}-\underline{C}\underline{A}^{-1}\underline{B} \end{array}\right]$$

سطر اول در CA^{-1} ضرب می‌شود (مقایسه کنید با ضرب در c/a) و از $\underline{C}$ کم می‌شود تا بلوک صفر در نتیجه حاصل شود.

بلوک نهایی $\underline{D}-\underline{C}\underline{A}^{-1}\underline{B}$ را نیز مشابه $d-cb/a$ در نظر بگیرید.

مثال ۱: فرض کنید رابطه $\underline{A}\underline{x}=\underline{b}$ را برای سه مقدار خاص $\underline{b}$ حل کرده‌اید و به نتایج زیر رسیده‌اید:

$$\underline{A}\underline{x}_۱=\begin{bmatrix}1\\0\\0\end{bmatrix} \qquad \underline{A}\underline{x}_۲=\begin{bmatrix}0\\1\\0\end{bmatrix} \qquad Ax_۳=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

فرض کنید $X=[x_۱\ x_۲\ x_۳]$، شم

اگر $x_۱=(۱,۱,۱)$ و $x_۲=(۰,۱,۱)$ و $x_۳=(۰,۰,۱)$ باشند، $Ax=b$ را برای $b=(۳,۵,۸)$ حل کنید و A را به دست آورید.

$$AX=I$$

$$b=(۳,۵,۸)\Rightarrow x=۳\underline{x}_۱+۵\underline{x}_۲+۸x_۳=\begin{bmatrix}۳\\۸\\۱۴\end{bmatrix} \qquad A=\begin{bmatrix}1&0&0\\-1&1&0\\0&-1&1\end{bmatrix}$$

مثال ۲: فرض کنید سطرهای ماتریس $\underline{A}$ برابر با a_1^T، ...، a_m^T باشند. حاصل ضرب $\underline{A}^T\underline{A}$ را به دست آورید. اگر $\underline{C}$ یک ماتریس قطری با مؤلفه‌های روی قطر اصلی برابر با c_1، ...، c_m باشد، آنگاه مقدار $\underline{A}^T\underline{C}\underline{A}$ را به دست آورید.

$$A^TA = a_1a_1^T + \dots + a_ma_m^T$$

$$A^TA = c_1a_1a_1^T + \dots + c_ma_ma_m^T$$

ماتریس‌های معکوس

- تعریف: ماتریس $\underline{A}$ معکوس‌پذیر است اگر ماتریس $\underline{A}^{-1}$ وجود داشته باشد به گونه‌ای که:

$$\underline{A}\underline{A}^{-1} = I \qquad \underline{A}^{-1}\underline{A} = I$$

- چند نکته:

۱- ماتریس معکوس وجود دارد اگر و فقط اگر روش حذف n پاشنه تولید کند.

۲- اگر برای ماتریس مربعی $\underline{A}$، $\underline{A}\underline{C} = \underline{I}$, $\underline{B}\underline{A} = \underline{I}$ باشد آنگاه $B = C$ است.

$$\underline{B} = \underline{B}(\underline{A}\underline{C}) = (\underline{B}\underline{A})\underline{C} = \underline{C}$$

بنابراین معکوس سمت چپ $\underline{B}$ (ماتریسی که از سمت چپ در A ضرب می‌شود و نتیجه برابر I می‌گردد) با معکوس سمت راست $\underline{C}$ (ماتریسی که از سمت راست در A ضرب می‌شود و نتیجه برابر I می‌گردد) باید برابر باشند.

۳- اگر $\underline{A}$ معکوس‌پذیر باشد تنها جواب دستگاه معادلات $\underline{A}\underline{x} = \underline{b}$ برابر با $\underline{x} = \underline{A}^{-1}\underline{b}$ است.

۴- فرض کنید بردار غیرصفر $\underline{x}$ وجود دارد به گونه‌ای که $\underline{A}\underline{x} = \underline{0}$ در این صورت $\underline{A}$ معکوس‌پذیر نیست.

۵- ماتریس ۲×۲ به صورت $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ را در نظر بگیرید. دترمینان ماتریس به صورت $ad-bc$ تعریف می‌شود. این ماتریس معکوس‌پذیر است اگر و فقط اگر دترمینان آن غیرصفر باشد. معکوس این ماتریس به صورت زیر تعریف می‌شود:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

۶- ماتریس قطری $\underline{A} = \begin{bmatrix} d_1 & & \underline{0} \\ & \ddots & \\ \underline{0} & & d_n \end{bmatrix}$ معکوس‌پذیر است اگر عناصر روی قطر اصلی آن غیرصفر باشند.

در این صورت:

$$A^{-1} = \begin{bmatrix} 1/d_1 & & 0 \\ & \ddots & \\ 0 & & 1/d_n \end{bmatrix}$$

- معکوس ضرب ماتریس‌ها: اگر A و B معکوس‌پذیر باشند در اینصورت AB نیز معکوس‌پذیر است و معکوس آن برابر است با:

$$(AB)^{-1} = B^{-1}A^{-1}$$

- مثال: فرض کنید E ماتریس حذفی باشد که ۵ برابر سطر اول را از سطر دوم کم می‌کند. در اینصورت:

$$E = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

به صورت معکوس ماتریس E^{-1} باید ۵ برابر سطر اول را با سطر دوم جمع کند. پس:

$$E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

حال فرض کنید F، ۴ برابر سطر دوم را از سطر سوم کم می‌کند و به طور معادل F^{-1} ۴ برابر سطر دوم را به سطر سوم اضافه می‌کند. به عبارت دیگر:

$$F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix} \qquad F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

عملکرد ماتریس FE را توصیف کنید.

$$FE = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 20 & -4 & 1 \end{bmatrix}$$

عملکرد ماتریس $(FE)^{-1}$ را توصیف کنید.

$$E^{-1}F^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

– روش حذفی گاوس-جردن برای محاسبه A^{-1} :

$$\underline{A}\,\underline{A}^{-1} = \underline{I} \Rightarrow \underline{A}\left[\underline{x}_۱ \;\; \underline{x}_۲ \;\; \underline{x}_۳\right] = \left[e_۱ \;\; e_۲ \;\; e_۳\right] \Rightarrow Ax_i = e_i \,, \; i=۱,۲,۳$$

برای حل دستگاه معادلات $Ax=b$ ابتدا ماتریس افزوده $[\underline{A}\;\underline{b}]$ را تشکیل می‌دادیم و سپس با روش حذفی ماتریس A را به یک ماتریس بالامثلثی تبدیل می‌کردیم.

در روش گاوس-جردن هدف حل همزمان سه دستگاه حاصل فوق است. برای این منظور گام‌های زیر را طی می‌کنیم.

۱– ماتریس افزوده برای سه دستگاه را تشکیل می‌دهیم (فرض کنیم $A = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۱ & ۲ & ۱ \\ ۰ & ۱ & ۲ \end{bmatrix}$ است).

$$\text{ماتریس افزوده} = \begin{bmatrix} ۱ & ۰ & ۰ & ۱ & ۰ & ۰ \\ ۱ & ۲ & ۱ & ۰ & ۱ & ۰ \\ ۰ & ۱ & ۲ & ۰ & ۰ & ۱ \end{bmatrix}$$

۲– با روش حذفی ماتریس A را بالامثلثی می‌کنیم.

$$\begin{bmatrix} ۱ & ۰ & ۰ & ۱ & ۰ & ۰ \\ ۱ & ۲ & ۱ & ۰ & ۱ & ۰ \\ ۰ & ۱ & ۲ & ۰ & ۰ & ۱ \end{bmatrix} \begin{array}{l} \to \text{Pivot} = ۱ \\ \to \text{Multiplier} = ۱ \\ \to \text{Multiplier} = ۰ \end{array}$$

$$\begin{bmatrix} ۱ & ۰ & ۰ & ۱ & ۰ & ۰ \\ ۰ & ۲ & ۱ & -۱ & ۱ & ۰ \\ ۰ & ۱ & ۲ & ۰ & ۰ & ۱ \end{bmatrix} \begin{array}{l} \to \text{Pivot} = ۲ \\ \to \text{Multiplier} = ۱/۲ \end{array}$$

$$\begin{bmatrix} ۱ & ۰ & ۰ & ۱ & ۰ & ۰ \\ ۰ & ۲ & ۱ & -۱ & ۱ & ۰ \\ ۰ & ۰ & ۳/۲ & ۱/۲ & -۱/۲ & ۱ \end{bmatrix}$$

روش حذفی که با نام روش گاوس نیز شناخته می‌شود در همین مرحله متوقف شده و از انتها شروع به تعیین متغیرها می‌نماید. اما در روش گاوس-جردن فرآیند ادامه می‌یابد تا زمانی که بخش مربوط به ماتریس $\underline{A}$ برابر با ماتریس همانی $\underline{I}$ گردد. برای این کار قدم بعدی حذف عناصر غیرصفر بالای قطرها است.

۳- حذف عناصر غیر صفر بالای پاشنه‌ها:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & -1 & 1 & 0 \\ 0 & 0 & 3/2 & 1/2 & -1/2 & 1 \end{bmatrix} \begin{matrix} \to \text{Multiplier} = 2/3 \\ \to \text{Pivot} = 3/2 \end{matrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & -4/3 & 4/3 & -2/3 \\ 0 & 0 & 3/2 & 1/2 & -1/2 & 1 \end{bmatrix}$$

۴- هر سطر را به پاشنه تقسیم کنید تا تمام پاشنه‌ها برابر با واحد گردد.

$$\left[\underbrace{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}}_{A \to I} \quad \underbrace{\begin{matrix} 1 & 0 & 0 \\ -2/3 & 2/3 & -1/3 \\ 1/3 & -1/3 & 2/3 \end{matrix}}_{I \to A^{-1}} \right]$$

خلاصه روش گاوس-جردن: ماتریس افزوده $[A \ I]$ را در A^{-1} ضرب کنید تا به ماتریس $[I \ A^{-1}]$ برسید.

قضیه: ماتریس مربعی A با ابعاد $n \times n$ معکوس‌پذیر است اگر و تنها اگر n پاشنه داشته باشد.

اثبات: با داشتن n پاشنه می‌توان دستگاه معادلات $Ax_i = e_i$ را حل کرد و ستون i-ام
مستقیم ماتریس معکوس یعنی x_i را به دست آورد. بنابراین A حداقل یک معکوس راست B دارد که

$$AB = I$$

از طرف دیگر با انجام سطری به دستگاه. معادلات خطی این شیوه را داریم که برای به دست آوردن پاشنه‌ها از روش حذفی گاوس-جردن می‌توانیم از روابط ماتریسی زیر استفاده نماییم:

$$CA = (D^{-1} \dots E \dots P \dots E)A = I$$

که در رابطه فوق E ماتریس حذفی، P ماتریس تعویض سطرها و D^{-1} ماتریس تقسیم هر سطر بر پاشنه آن سطر است ← C معکوس چپ ماتریس A است.

بنابراین معکوس‌های چپ و راست وجود داشته و می‌دانیم که باید برابر باشند. در نتیجه اگر ماتریس $n\times n$: A دقیقاً n پاشنه داشته باشد آنگاه معکوس‌پذیر است.

اثبات معکوس: اگر $AC=I$ باشد آنگاه A باید n پاشنه داشته باشد ← برهان خلف :

۱- اگر A ، n پاشنه نداشته باشد ← در نتیجه روش حذف گاوس-جردن به یک سطر تمام صفر در محدود. ماتریس می‌رسیم.

۲- فرض کنیم عملیات انجام شده برای حذف همگی در ماتریس معکوس‌پذیر M باشد. در این صورت یک سطر ماتریس MA برابر صفر است.

۳- اگر $AC=I$ باشد در این صورت $MAC=M$ است. چون ماتریس MA سطر صفر دارد، ضرب آن در C یک سطر صفر در ماتریس MAC یا به طور معادل M ایجاد می‌کند.

۴- یک ماتریس معکوس‌پذیر مانند M نمی‌تواند سطر صفر داشته باشد ← فرض خلف باطل و A دقیقاً n پاشنه دارد.

- تشخیص ماتریس معکوس‌پذیر:

* مسیر رایج برای تشخیص معکوس‌پذیری ماتریس تعیین پاشنه‌ها است.

- ماتریس‌های برجسته قطری معکوس‌پذیر هستند.

تعریف: ماتریس $\underline{A}$ برجسته قطری است اگر برای هر سطر i، اندازه عنصر a_{ii} روی قطر اصلی از مجموع اندازه سایر عناصر روی همان قطر بیشتر باشد ← $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$

اثبات: فرض کنیم $\underline{x}$ یک بردار غیرصفر باشد و بزرگترین عنصر آن از نظر اندازه $|x_i|$ باشد. در اینصورت $\underline{A}\underline{x}=\underline{0}$ غیرممکن زیرا در سطر i-ام دستگاه معادلات داریم:

$$a_{i1}x_1+\cdots+a_{ii}x_i+\cdots+a_{in}x_n=0 \Rightarrow a_{ii}x_i=-\sum_{j\neq i}a_{ij}x_j \Rightarrow |a_{ii}x_i|=\left|\sum_{j\neq i}a_{ij}x_j\right|$$

چون A برجسته قطری است در نتیجه رابطه زیر را داریم:

$$\left|\sum_{j\neq i}a_{ij}x_j\right| \leqslant \sum_{j\neq i}|a_{ij}|\,|x_j| \leqslant |x_i|\sum_{j\neq i}|a_{ij}| < |x_i|\,|a_{ii}|=|x_i a_{ii}|$$

بنابراین تنها جواب دستگاه $\underline{A}\underline{x}=\underline{b}$ برابر با $\underline{x}=0$ است. در نتیجه $\underline{A}$ معکوس‌پذیر است.

تجزیه LU ماتریس مربعی A

- فرض کنید ماتریس A با ابعاد n×n، n پاشنه دارد و در مسیر رسیدن به پاشنه‌ها تغییر دادن سطرها مورد نیاز نیست. در این صورت رابطه زیر را داریم:

$$E_{n(n-1)}E_{n(n-۲)}\cdots E_{۲۱}A = U$$

U ماتریس بالا مثلثی است که پاشنه‌ها بر روی قطر اصلی آن قرار گرفته‌اند.

می‌توان مسیر معکوس از U به A را به صورت زیر پیمود:

$$A = E_{۲۱}^{-1}E_{۳۱}^{-1}\cdots E_{n(n-1)}^{-1}U$$

همانطور که قبلاً دیدیم در ماتریس E_{ij} عناصر روی قطر اصلی یک و تنها عنصر (i,j)-ام برای $-l_{ij}$ (l_{ij} multiplier است) و در ماتریس E_{ij}^{-1} نیز عناصر روی قطر اصلی یک و عنصر (i,j)-ام برابر $l_{ij}+$ است. بنابراین هر دو ماتریس E_{ij} و E_{ij}^{-1} پایین مثلثی هستند.

لم: ضرب ماتریس‌های پایین مثلثی خود یک ماتریس پایین مثلثی است

حال اگر تعریف مقابل را داشته باشیم:

$$L = E_{۲۱}^{-1}E_{۳۱}^{-1}\cdots E_{n(n-1)}^{-1}$$

به تجزیه LU ماتریس A می‌رسیم به صورت مقابل می‌رسیم:

$$A = LU$$

که L یک ماتریس پایین مثلثی و U یک ماتریس بالا مثلثی است. روی قطر اصلی ماتریس U پاشنه‌ها قرار دارند و روی قطر اصلی ماتریس L نیز همگی ۱ است.

- خاصیت ۱: اگر سطر i-ام ماتریس A با k صفر شروع شود سطر i-ام ماتریس L نیز با k صفر شروع خواهد شد.

- خاصیت ۲: اگر ستون i-ام ماتریس A با k صفر شروع شود i-ام ماتریس U نیز با k صفر شروع می‌شود.

- تجزیه $LD\bar{U}$: از آنجایی که روی قطر اصلی ماتریس U، پاشنه‌ها قرار گرفته‌اند، این ماتریس متقارن نیست. اگر ماتریس U را به ضرب یک ماتریس قطری که پاشنه‌ها روی قطر اصلی آن قرار گرفته‌اند و یک ماتریس $\bar{U}$ نرمالیزه که هر سطر بر پاشنه همان سطر تقسیم شده است تجزیه کنیم به تجزیه $LD\bar{U}$ می رسیم که در آن:

۱) $\underline{L}$ پایین‌مثلثی است و روی قطر اصلی آن همگی عناصر $\underline{1}$ هستند.

۲) $\bar{U}$ بالامثلثی است و روی قطر اصلی آن همگی عناصر $\underline{1}$ هستند.

۳) D قطری است و پاشنه‌ها بر روی قطر اصلی آن قرار گرفته‌اند.

- حل دستگاه $\underline{A}\underline{x}=\underline{b}$ با استفاده از تجزیه LU :

گام اول: تجزیه $\underline{L}\underline{U}$ ماتریس $\underline{A}$ را به دست آورید.

گام دوم: ابتدا دستگاه معادله پایین‌مثلثی $\underline{L}\underline{c}=\underline{b}$ را حل کرده و $\underline{c}$ را به دست آورید. سپس دستگاه معادله $\underline{U}\underline{x}=\underline{c}$ را حل کرده و $\underline{x}$ را به دست آورید.

ترانهاده و جایگشت

- ماتریس ترانهاده: اگر A یک ماتریس دلخواه $n\times m$ باشد در اینصورت ماتریس ترانهاده را با A^T نشان می‌دهیم ماتریس $\underline{A}$، ستون‌های ماتریس $\underline{A}^T$ هستند.

به عبارت دیگر:

$$(A^T)_{ij} = A_{ji}$$

- خواص ترانهاده:

- جمع: $(\underline{A}+\underline{B})^T=\underline{A}^T+\underline{B}^T$

- ضرب: $(AB)^T=B^TA^T$

- معکوس: $(A^{-1})^T=(A^T)^{-1}$

- خاصیت ضرب:

$$\left.\begin{matrix}\underline{B}=[\underline{x}_1,\dots,\underline{x}_m]\\ \underline{A}\end{matrix}\right\}\Rightarrow \underline{A}\,\underline{B}=[\underline{A}\underline{x}_1\ \ \underline{A}\underline{x}_2\ \dots\ \underline{A}\underline{x}_m]$$

$$(\underline{A}\,\underline{B})^T=\begin{bmatrix}x_1^TA^T\\ \vdots\\ x_m^T\underline{A}^T\end{bmatrix}=\begin{bmatrix}x_1^T\\ \vdots\\ x_m^T\end{bmatrix}A^T=B^TA^T$$

- نمایش ضرب داخلی با استفاده از ترانهاده:

$\underline{x} \cdot \underline{y} = x^T y$ ضرب داخلی

$x y^T \rightarrow$ نمایش ضرب خارجی

- ماتریس متقارن: ماتریس S را متقارن گوییم اگر رابطه $S = S^T$ برقرار باشد.

توجه: معکوس یک ماتریس متقارن خود یک ماتریس متقارن است.

توجه: برای ماتریس دلخواه A، ماتریس A^TA مربعی و متقارن است.

قضیه: اگر $S = S^T$ باشد و بتوان بدون تغییر سطر تجزیه $S = LDU$ را انجام داد در اینصورت $U = L^T$.

اثبات: ابتدا باید نشان دهیم تجزیه LU منحصر به فرد است. برای اینکار فرض کنید $S = L_1U_1 = L_2U_2$ باشد. با توجه به معکوس پذیر بودن ماتریس‌های تجزیه LU رابطه زیر را داریم.

$$L_2^{-1}L_1 = U_2U_1^{-1}$$

چون سمت راست بالامثلثی و سمت چپ پایین مثلثی است پس هر دو سمت می‌بایست قطری باشند. برای ماتریس سمت چپ می‌دانیم که عنصر روی قطر اصلی همگی 1 هستند.

$$L_2^{-1}L_1 = U_2U_1^{-1} = I \Rightarrow \begin{cases} L_1 = L_2 \\ U_2 = U_1 \end{cases}$$ تجزیه منحصر ⇒ به فرد است.

$$\left.\begin{array}{l} S = L_1 D_1 U_1 \\ S = L_2 D_2 U_2 \end{array}\right\} \Rightarrow \begin{cases} L_1 = L_2 \\ D_1 = D_2 \\ U_1 = U_2 \end{cases}$$

حال فرض می‌کنیم S دارای تجزیه تقابل باشد.

$$\left.\begin{array}{l} S = LDU \\ S^T = S = U^TD^TL^T = U^TDL^T \\ \text{LDU منحصر به فرد بودن تجزیه} \end{array}\right\} \Rightarrow U = L^T$$

بنابراین برای ماتریس متقارن S رابطه مقابل را داریم:

$$S = LDL^T$$

- ماتریس‌های جایگشت : ماتریسی است که در هر سطر و هر ستون آن دقیقاً یک عنصر غیرصفر وجود دارد که برابر با ۱ است. به عبارت دیگر ماتریس جایگشت P تمام سطرهای ماتریس همانی I را با یک ترتیب دلخواه دارد.

- توجه : اگر P یک ماتریس جایگشت باشد آنگاه P^T نیز یک ماتریس جایگشت است.

- توجه : ضرب ماتریسی دو ماتریس جایگشت خود یک ماتریس جایگشت است.

- توجه : برای یک n ، دقیقاً $n!$ ماتریس جایگشت منحصربه‌فرد وجود دارد.

- توجه : اگر P یک ماتریس جایگشت باشد آنگاه ، P^{-1} نیز یک ماتریس جایگشت است.

- قضیه : برای یک ماتریس جایگشت P ، رابطه مقابل برقرار است :

$$P^{-1} = P^T$$

اثبات :

$$\left.\begin{aligned}(PP^T)_{ij} &= \sum_{k=1}^{n} P_{ik} P^T_{kj} = \sum_{k=1}^{n} P_{ik} P_{jk}\\ \sum_{k=1}^{n} P_{ik} P_{jk} &= \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}\end{aligned}\right\} \Rightarrow PP^T = I \Rightarrow P^T = P^{-1}$$

- تجزیه LU با جابه‌جایی ردیف :

همانطور که دیدیم اگر در مسیر حذف گاوسی ماتریس A نیاز به جابه‌جایی ردیف‌ها باشد نمی‌توان تجزیه $A = LU$ را انجام داد. در این حالت دو راه حل وجود دارد :

۱- تغییر سطرهای مورد نیاز را در ابتدا روی A با استفاده از ماتریس جایگشت P انجام دهیم و سپس تجزیه $PA = LU$ را خواهیم داشت.

۲- جابه‌جایی را تا اتمام روش حذف به تأخیر بیندازیم. در این حالت ردیف‌های پاشنه در شرایط نامرتبی خواهند بود. سپس با استفاده از ماتریس P_1 آنها را به فرم مثلثی مناسب در بیاوریم تا رابطه زیر را داشته باشیم :

$$A = L_1 P_1 U_1$$

فضاهای برداری و زیرِ فضاها

فضاهای برداری

- میدان: مجموعه F شامل اعضای x، y، z و ... را به همراه دو عملگر جمع و ضرب درنظر بگیرید. عملگر جمع به هر جفت عنصر x و y در F یک عنصر $x+y$ در F را نسبت می دهد. عملگر ضرب به هرجفت عنصر x و y در F یک عنصر xy در F رانسبت می دهد. مجموعه F به همراه دو عملگر جمع و ضرب را یک میدان می نامیم اگر این دو عملگر ۹ رابطه زیر را ارضا نمایند:

۱- عملگر جمع جابه جایی پذیر است. $\forall x,y \in F: x+y=y+x$

۲- عملگر جمع شرکت پذیر است. $\forall x,y,z \in F: x+(y+z)=(x+y)+z$

۳- یک عنصر یکتای 0 (صفر) در F وجود دارد به گونه‌ای که: $\forall x \in F: x+0=x$

۴- برای هر عنصر x در مجموعه F یک عنصر منحصربه فرد $-x$ در F وجود دارد که $x+(-x)=0$.

۵- عملگر ضرب جابه‌جایی پذیر است. $\forall x,y \in F: xy=yx$

۶- عملگر ضرب شرکت پذیر است. $\forall x,y,z \in F: x(yz)=(xy)z$

۷- یک عنصر یکتای غیرصفر 1 (یک) در F وجود دارد به گونه‌ای که: $\forall x \in F: x1=x$

۸- برای هر عنصر غیر صفر x در مجموعه F یک عنصر منحصربه‌فرد x^{-1} (یا $1/x$) در F وجود دارد به‌گونه‌ای که $xx^{-1}=1$

۹- ضرب بر روی جمع توزیع پذیر است یعنی: $\forall x,y,z \in F: x(y+z)=xy+xz$

- مثال: مجموعه اعداد مختلط بهمراه تعاریف ضرب و جمع مختلط یک میدان است.

- زیرمیدان: فرض کنیم F به‌همراه عملیات جمع و ضرب مشخص یک میدان باشد. دراینصورت زیرمجموعه E از F یک زیر میدان است اگر E به همراه عملیات ضرب و جمع متناظر با میدان F، خود یک میدان باشد.

مثال: مجموعه اعداد صحیح مثبت به همراه عملیات جمع و ضرب معمولی یک زیرمیدان از اعداد حقیقی نیست زیرا صفر ندارد.

مجموعه اعداد صحیح یک زیرمیدان از اعداد حقیقی نیست چون x^{-1} ندارد.

مجموعه اعداد گویا یک زیرمیدان از اعداد مختلط است.

- فضای برداری: یک فضای برداری (فضای خطی) شامل اعضای زیر می‌گردد:

۱- میدان F شامل اعداد.

۲- مجموعه V از اشیاء که به آنها بردار نیز گفته می‌شود.

۳- عملگر جمع که به جفت بردار α و β در V یک بردار $\alpha+\beta$ در V را نسبت می‌دهد و بردار $\alpha+\beta$ را مجموع بردارهای α و β می‌نامیم به گونه‌ای که:

الف- عملگر جمع جابجایی پذیر است یعنی $\forall \alpha, \beta \in V: \alpha+\beta=\beta+\alpha$

ب- عملگر جمع شرکت پذیر است یعنی $\forall \alpha, \beta, \gamma \in V: (\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$

ج- یک بردار منحصر به فرد 0 در V وجود دارد. که بردار صفر نامیده می‌شود به گونه‌ای که: $\forall \alpha \in V: \alpha+0=\alpha$

د- برای هر بردار α در V یک بردار منحصر به فرد $-\alpha$ در V وجود دارد به گونه‌ای که: $\alpha+(-\alpha)=0$

۴- عملگر ضرب عددی که به هر عدد $c \in F$ و بردار $\alpha \in V$ یک بردار $c\alpha \in V$ را نسبت می‌دهد که $c\alpha$ را ضرب c و α می‌نامیم به گونه‌ای که:

الف- $\forall \alpha \in V: 1\alpha=\alpha$

ب- $(c_1c_2)\alpha=c_1(c_2\alpha)$

ج- $c(\alpha+\beta)=c\alpha+c\beta$

د- $(c_1+c_2)\alpha=c_1\alpha+c_2\alpha$

- مثال: فضای n تایی F^n: فرض F یک میدان بوده و V مجموعه تمام n تایی‌هایی به صورت $\alpha=(x_1, x_2, \dots, x_n)$ است که $x_i \in F$ است. اگر $\beta=(y_1, y_2, \dots, y_n)$ باشد به گونه‌ای که $y_i \in F$ باشد در اینصورت عملیات جمع و ضرب عددی به صورت زیر تعریف می‌شوند:

$$\alpha+\beta=(x_1+y_1, x_2+y_2, \dots, x_n+y_n)$$

$$c\alpha=(cx_1, cx_2, \dots, cx_n)$$

- فضای ماتریس‌های $m\times n$ ، $F^{m\times n}$: فرض کنید F یک میدان است و m ، n دو عدد صحیح مثبت باشند. فرض کنید $F^{m\times n}$ مجموعه تمام ماتریس‌های $m\times n$ بر روی میدان F باشد. مجموع دو بردار A و B در $F^{m\times n}$ به صورت زیر تعریف می‌شود:

$$(A+B)_{ij} = A_{ij} + B_{ij}$$

ضرب عددی ماتریس A در عدد c به صورت زیر تعریف می‌شود:

$$(cA)_{ij} = cA_{ij}$$

- فضای توابع از یک مجموعه به یک میدان: فرض کنید F یک میدان و S یک مجموعه غیرتهی باشد. V را مجموعه تمام توابع از مجموعه S به میدان F در نظر بگیرید. مجموع دو بردار f و g در V بردار $f+g$ است یعنی تابعی از S به F که به صورت مقابل تعریف می‌شود:

$$(f+g)(s) = f(s) + g(s)$$

ضرب عددی تابع f یک تابع cf است که به صورت زیر تعریف می‌شود:

$$(cf)(s) = cf(s)$$

- در تعریف فضای n تایی F^n اگر میدان F را برابر با $\mathbb{R}$ در نظر بگیریم به تعریف فضای برداری $\mathbb{R}^n$ می‌رسیم که تمام بردارهای ستونی $\underline{v}$ با n عنصر را در بر می‌گیرد.

- فضای توابع چندجمله‌ای روی میدان F : فرض کنید F یک میدان باشد و V مجموعه تمام توابع f از F به F باشد به گونه‌ای که:

$$f(x) = c_0 + c_1 x + \dots + c_n x^n$$

که c_0 ، c_1 ، ... و c_n اعداد ثابتی در F هستند. همچنین فرض کنید تابع دلخواه $g(x)$ به صورت $g(x) = d_0 + d_1 x + \dots + d_n x^n$ تعریف گردد و روابط زیر را نیز داشته باشیم:

$$(f+g)(x) = (c_0+d_0) + (c_1+d_1)x + \dots + (c_n+d_n)x^n = f(x)+g(x)$$

$$(cf)(x) = c(c_0 + c_1 x + \dots + c_n x^n) = cf(x)$$

- زیرفضا: فرض کنید V یک فضای برداری روی F باشد. یک زیرفضای V، یک زیرمجموعه از V مانند W است که خود به همراه عملیات جمع و ضرب عددی تعریف شده برای V، یک فضای برداری روی F است.

- یک زیرفضای دلخواه در $\mathbb{R}^n$ در نظر بگیرید. در اینصورت اگر بردارهای $\underline{u}$ و $\underline{v}$ در این زیرفضا قرار بگیرند، تمام ترکیب‌های خطی این دو بردار نیز باید در زیرفضا قرار گیرد.

- هر زیرفضای دلخواه در $\mathbb{R}^n$ می‌بایست بردار صفر را شامل شود.

- خط‌هایی که از مبدأ مختصات می‌گذرند یک زیرفضا در $\mathbb{R}^n$ هستند.

- ابرصفحه‌هایی که از مبدأ مختصات می‌گذرند یک زیرفضا در $\mathbb{R}^n$ هستند.

فضای ستونی ماتریس A

- تعریف: فضای ستونی ماتریس $\underline{A}$ شامل تمام ترکیب‌های خطی ستون‌های A می‌شود. این ترکیب‌های خطی تمامی بردارهای ممکن به صورت $\underline{A}\,\underline{x}$ هستند و این بردارها فضای ستونی ماتریس $\underline{A}$ را تشکیل می‌دهند. این فضا را با $C(A)$ نمایش می‌دهیم.

- توجه: دستگاه معادلات خطی $\underline{A}\,\underline{x}=\underline{b}$ قابل حل است اگر و تنها اگر $\underline{b}$ در فضای ستونی ماتریس A باشد یعنی $\underline{b} \in C(\underline{A})$

- مثال:

$$A=\begin{bmatrix} ۱ & ۰ \\ ۴ & ۳ \\ ۲ & ۳ \end{bmatrix} \quad x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow A \text{ فضای ستونی } = x_1\begin{bmatrix} ۱ \\ ۴ \\ ۲ \end{bmatrix} + x_2\begin{bmatrix} ۰ \\ ۳ \\ ۳ \end{bmatrix}$$

* فضای ستونی A صفحه ای است که دو ستون بردار A در آن قرار گرفته‌اند.
* اگر $\underline{b}$ در این فضا قرار گرفته باشد دستگاه فوق جواب دارد.
* بردار صفر حتماً در فضای ستونی ماتریس A قرار دارد و این بردار از ترکیب خطی $x_1=x_2=0$ به دست می‌آید.

- فضای برداری $\mathbb{R}^n$ را درنظر بگیرید. اگر S مجموعه‌ای از بردارهای موجود در این زیرفضای برداری باشد در این صورت زیرفضای V پوشش داده شده توسط مجموعه S را به صورت زیرفضای حاصل از ترکیب خطی تمام بردارهای داخل مجموعه S تعریف می نماییم.

$S = \{\underline{v}_1, \dots, \underline{v}_N\} \Rightarrow$ زیرفضای V پوشش داده شده توسط S = تمام بردارهای $c_1\underline{v}_1 + \dots + c_N\underline{v}_N$

- توجه: اگر S ستون های ماتریس A باشد در اینصورت $SS = C(A)$ است.

فضای پوچی ماتریس A

- تعریف: فضای پوچی ماتریس A که با $N(A)$ نشان می دهیم شامل تمام بردارهای $\underline{x}$ می گردد که در رابطه $\underline{A}\underline{x} = \underline{0}$ صادق هستند (A می تواند در حالت کلی یک ماتریس غیرمربعی باشد).

- تمرین: نشان دهید که بردارهایی که در رابطه $\underline{A}\underline{x} = \underline{0}$ صادق هستند تشکیل یک زیرفضای برداری می دهند.

می دانیم بردار $\underline{x} = \underline{0}$ یک جواب بدیهی دستگاه معادلات $\underline{A}\underline{x} = \underline{0}$ است

$$\left.\begin{array}{l}\underline{A}\underline{x}_1 = \underline{0}\\ \underline{A}\underline{y}_1 = \underline{0}\end{array}\right\} \Rightarrow \underline{A}\underline{x}_1 + \underline{A}\underline{y}_1 = 0 \Rightarrow \underline{A}(\underline{x}_1 + \underline{y}_1) = \underline{0} \Rightarrow$$

اگر $\underline{x}_1$ و $\underline{y}_1$ پاسخ دستگاه $\underline{A}\underline{x} = b$ باشد آنگاه $\underline{x}_1 + \underline{y}_1$ نیز پاسخ این دستگاه است.

$$\underline{A}\underline{x}_1 = 0 \Rightarrow \underline{A}(c\underline{x}_1) = 0 \Rightarrow$$

اگر $\underline{x}_1$ پاسخ دستگاه معادلات باشد $c\underline{x}_1$ نیز پاسخ این دستگاه است.

بنابراین مجموعه پاسخ های دستگاه $\underline{A}\underline{x} = \underline{0}$ تشکیل یک فضای برداری می دهد که آن را با $N(\underline{A})$ نمایش می دهیم

- توجه: برای ماتریس $\underline{A}$ با ابعاد $m \times n$، $C(\underline{A})$ یک زیرفضای $\mathbb{R}^n$، $N(\underline{A})$ یک زیرفضای $\mathbb{R}^m$ است.

- مثال: فضای پوچی ماتریس $A = \begin{bmatrix} ۱ & ۲ \\ ۳ & ۴ \end{bmatrix}$ را توصیف کنید

$$\begin{cases} x_۱ + ۲x_۲ = ۰ \\ ۳x_۱ + ۴x_۲ = ۰ \end{cases} \xrightarrow[\text{از سطر دوم کم می‌کنیم}]{\text{۳ برابر سطر اول را}} \begin{cases} x_۱ + ۲x_۲ = ۰ \\ ۰ = ۰ \end{cases}$$

بنابراین فضای پوچی ماتریس $\underline{A}$، خط $۳x_۱ + ۴x_۲ = ۰$ است.

$$\underset{\underline{A}}{\begin{bmatrix} ۱ & ۲ \\ ۳ & ۴ \end{bmatrix}} \begin{bmatrix} x_۱ \\ x_۲ \end{bmatrix} = \begin{bmatrix} ۰ \\ ۰ \end{bmatrix} \longrightarrow \underset{U}{\begin{bmatrix} ۱ & ۲ \\ ۰ & ۰ \end{bmatrix}} \begin{bmatrix} x_۱ \\ x_۲ \end{bmatrix} = \begin{bmatrix} ۰ \\ ۰ \end{bmatrix}$$

فضای پوچی ماتریس A و U یکسان است.

$$\rightarrow N(A) = N(U)$$

در رابطه $\underline{U}\underline{x} = \underline{0}$، $x_۲$ متغیر دلخواه (آزاد) است. یک روش برای به دست آوردن و توصیف فضای پوچی این است که متغیر آزاد را برابر یک مقدار دلخواه قرار دهیم و بقیه متغیرها را به دست آوریم. به عنوان مثال در رابطه فوق $x_۲ = ۱$ قرار می‌دهیم. در اینصورت $x_۱ = -۲$ و در نتیجه فضای پوچی تمام مضارب بردار $(-۲, ۱)$ را شامل می‌شود.

- مثال: برای $A = \begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix}$ فضای پوچی را به دست آورید.

اول: معادله یک صفحه گذرنده از مبدأ را توصیف می‌کند که بردار نرمال بر آن $(۱, ۲, ۳)$ است.

$$\underline{A}\,\underline{x} = \underline{0} \Rightarrow x_۱ + ۲x_۲ + ۳x_۳ = ۰ \rightarrow$$

راه حل دوم: برای ماتریس A فوق دو متغیر $x_۲$, $x_۳$ آزاد هستند. پس می‌توانیم دو جواب خاص برای معادله $\underline{A}\underline{x} = \underline{0}$ به صورت زیر ارائه دهیم.

$$\begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \end{bmatrix} = ۰ \Rightarrow \begin{cases} s_۱ = \begin{bmatrix} -۲ \\ ۱ \\ ۰ \end{bmatrix} \\ s_۲ = \begin{bmatrix} -۳ \\ ۰ \\ ۱ \end{bmatrix} \end{cases}$$

فضای پوچی تمام بردارهایی هستند که ترکیب خطی دو بردار $s_۱$, $s_۲$ است که برابر همان صفحه حاصل در روش اول است.

به عبارت دیگر فرض کنید از شما خواسته شود فرآیند حذف گاوس را بر روی ماتریس A پیاده نمایید. در این صورت چون A تنها یک سطر دارد خود آن نتیجه فرآیند خواهد بود.

$$A = \begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix}$$

بنابراین در ماتریس مستطیلی A فقط یک پاشنه متناظر با متغیر x_1 داریم. متغیرهایی که پاشنه ندارند را متغیر آزاد می نامیم. می توان به صورت زیر جواب های ویژه متناظر با هر یک از دو متغیر را به دست آورد:

$$\begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} -۲ \\ ۱ \\ ۰ \end{bmatrix} = ۰ \Rightarrow \text{بردار متناظر با } x_2$$

$$\begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} -۳ \\ ۰ \\ ۱ \end{bmatrix} = ۰ \Rightarrow \text{بردار متناظر با } x_3$$

- به دست آوردن جواب های ویژه معادل به دست آوردن فضای پوچی ماتریس A است زیرا فضای پوچی شامل تمام ترکیب های خطی جواب های ویژه است.

- تعداد جواب های ویژه دستگاه معادلات $\underline{A}\underline{x}=\underline{0}$ برابر تعداد متغیرهای آزاد در بردار $\underline{x}$ است. در ادامه روشی را برای تعیین متغیرهای آزاد و جواب های ویژه مبتنی بر روش حذف گاوسی و گاوس جردن ارائه می دهیم.

- تعیین فضای پوچی با استفاده از روش حذف گاوسی:

۱- عملیات مربوط به حذف گاوسی را روی ماتریس $\underline{A}$ انجام دهید تا ماتریس $\underline{U}$ به دست آید. این ماتریس را فرم سطری پلکانی متناظر با $\underline{A}$ می نامیم. همانطور که قبلاً نیز گفتیم چون بردار $\underline{b}=\underline{0}$ است، فضای پوچی ماتریس $\underline{A}$ و $\underline{U}$ یکسان است

۲- در ماتریس $\underline{U}$، ستون هایی که در آنها پاشنه قرار دارد آزاد نیستند و ستون هایی که پاشنه ندارند آزادند. بنابراین تعداد جواب های ویژه دستگاه $\underline{A}\underline{x}=\underline{0}$ برابر با تعداد ستون هایی از $\underline{U}$ است که پاشنه ندارد.

۳- فرض کنید متغیرهای $\{x_i, x_j, x_k\}$ آزاد هستند. حال سه جواب ویژه را به صورت زیر به دست آورید. یکی از متغیرهای آزاد را برابر $\underline{1}$ و سایر آنها را برابر صفر قرار دهید. دستگاه حاصل را برای متغیرهای غیرآزاد حل کنید. حاصل یکی از جواب های ویژه است. این کار را برای سایر متغیرهای آزاد ادامه دهید.

مثال: فرض کنید ماتریس A به صورت مقابل باشد تمام بردارهای خاص متصل برحی آن را به دست آورید.

$$\underline{A} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \begin{array}{l} \rightarrow \text{۱ = پاشنه} \\ \rightarrow \text{۲ = ضریب} \\ \rightarrow \text{۳ = ضریب} \end{array}$$

$$\underline{U}_1 = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix} \begin{array}{l} \\ \rightarrow \text{۲ = پاشنه} \\ \rightarrow \text{۱ = ضریب} \end{array}$$

$$\underline{U} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \text{فرم سطری پلکانی}$$

↑	↑	↑	↑
پاشنه دارد.	پاشنه ندارد.	پاشنه دارد.	پاشنه ندارد.
x_1 غیرآزاد است.	x_2 آزاد است.	x_3 غیرآزاد است.	x_4 آزاد است.

۲ متغیر آزاد و در نتیجه ۲ جواب خاص داریم.

جواب اول: $x_2=1$, $x_4=0$ ← x_1, x_3 را به دست می‌آوریم.

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ 1 \\ x_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2 + 2x_3 = 0 \\ 2x_3 = 0 \end{cases} \Rightarrow \begin{cases} x_3 = 0 \\ x_1 = -2 \end{cases}$$

$$\Longrightarrow \text{جواب خاص اول} : \underline{x}_1^{(P)} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

جواب دوم: $x_2=0$, $x_4=1$ ← x_1, x_3 را به دست می‌آوریم.

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2x_3 + 2 = 0 \\ 2x_3 + 4 = 0 \end{cases} \Rightarrow \begin{cases} x_3 = -2 \\ x_1 = 2 \end{cases}$$

$$\Longrightarrow \text{جواب خاص دوم} : x_2^{(P)} = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$$

$$N(\underline{U}) = N(\underline{A}) = \text{span}\left\{ \underline{x}_1^{(P)}, \underline{x}_2^{(P)} \right\} \rightarrow$$

$\text{span}\{\underline{x}_1, \dots, \underline{x}_N\}$ زیرفضای پوشش داده شده در $\mathbb{R}^n$ به وسیله بردارهای x_1، ...، x_N را نشان می‌دهد که n بعد بردارها است.

- ستون پاشنه: ستونی است که در آن یک پاشنه قرار دارد.
- ستون آزاد: ستونی است که در آن پاشنه وجود ندارد.

- فضای پوچی Z: اگر ماتریس A ستون آزاد نداشته باشد در اینصورت تنها جواب دستگاه $\underline{A}\underline{x}=\underline{0}$ بردار $\underline{x}=\underline{0}$ است. در این حالت فضای پوچی فضای پوشش داده شده توسط بردار صفر بود. و آن را با Z نمایش می‌دهیم.

- تعیین فضای پوچی با استفاده از روش گاوس-جردن:

۱- عملیات حذف گاوس-جردن را روی $\underline{A}$ انجام دهید تا ماتریس $\underline{R}$ به دست آید. (در مقایسه با $\underline{U}$، در ماتریس $\underline{R}$ درایه‌های بالای پاشنه برابر صفر شده است و خود پاشنه نیز برابر با یک است). به سادگی می‌توان نشان داد: $N(\underline{A})=N(\underline{U})=N(\underline{R})$

۲- ستون‌های پاشنه و آزاد را در ماتریس $\underline{R}$ به دست آورید و متغیرهای متناظر با آنها را مشخص کنید.

۳- مشابه روش قبل اگر متغیرهای آزاد $\{x_i, x_j, x_k\}$ باشند، جواب ویژه متناظر با هر یک از متغیرهای آزاد با برابر یک قرار دادن آن متغیر، برابر صفر قرار دادن سایر متغیرهای آزاد و حل دستگاه، حاصل برای متغیرهای غیرآزاد به دست می‌آید.

توجه: فرم ایجاد شده در $\underline{R}$ را فرم کاهش یافته سطری پلکانی می‌نامیم و با $R=\text{rref}(A)$ نمایش می‌دهیم.

توجه: جواب‌های ویژه به سادگی از دستگاه $\underline{R}\,\underline{x}=\underline{0}$ به دست می‌آیند و از این نظر این فرم بر فرم سطری پلکانی ارجحیت دارد.

توجه: اگر $N(A)=Z$ باشد در اینصورت معادله $\sum_{i=1}^{n} x_i \underline{a}_i=0$ تنها جواب بدیهی $\underline{x}=\underline{0}$ دارد. به عبارت دیگر ستون‌های $\underline{A}$ مستقل خطی هستند.

توجه: اگر در $A : m \times n$ تعداد ستون‌ها بیشتر باشد $(n > m)$ در اینصورت حداقل یک ستون آزاد در $\underline{R}$ یا $\underline{U}$ وجود خواهد داشت و در نتیجه دستگاه معادلات $\underline{A}\underline{x} = \underline{0}$ جواب غیرصفر خواهد داشت ← فضای پوچی A بردارهای غیرصفر خواهد داشت ← اگر $d = n - m$ قرار دهیم در اینصورت حداقل d متغیر آزاد داریم.

- **بعد فضای پوچی:** بعد فضای پوچی برابر با تعداد متغیرهای آزاد تعریف می‌شود.

- **رتبه ماتریس:** اگرچه یک ماتریس با ابعاد ظاهری آن مشخص می‌شود (مثلاً $m \times n$) اما از منظر دستگاه معادلات خطی این ابعاد ظاهری، ابعاد واقعی دستگاه. را نشان نمی‌دهد. به عنوان مثال سطری که مجموع دو سطر دیگر ماتریس است هنگام تشکیل ماتریس‌های $\underline{U}$ و $\underline{R}$ حذف می‌شود و نباید در شمارش ابعاد دستگاه از منظر دستگاه. معادلات خطی شمرده شود. رتبه ماتریس بعد واقعی یک ماتریس را نشان می‌دهد.

تعریف: رتبه ماتریس A برابر با تعداد پاشنه‌های ماتریس بوده و با r نشان داده می‌شود. به عبارت دیگر:

$$r = \text{rank}(\underline{A})$$

- ماتریس به فرم کاهش یافته سطری پلکانی $\underline{R}$ متناظر با $\underline{A}$ دقیقاً r سطر دارد که تمام غیر صفر هستند.

- **ارتباط تفسیر سطری و ستونی:** هر ستون آزاد ترکیب خطی ستون‌های پاشنه است که این ترکیب خطی در جواب‌های خاص به دست آمده برای تعیین فضای پوچی تعیین شده است.

- **حالت خاص: ماتریس رتبه اول:** ماتریسی است که فقط یک پاشنه دارد ← در ماتریس $\text{ref}(A)$ فقط سطر اول عنصر غیرصفر دارد ← تمام سطرهای ماتریس A ضریبی از سطر اول هستند و صفر کردن عنصر زیر پاشنه در سطر اول منجر به صفر شدن تمام سطرها می‌گردد.

همزمان چون فقط ستون اول پاشنه است، سایر ستون‌ها آزاد هستند در نتیجه تمام ستون‌ها نیز مضربی از ستون اول هستند.

- تعریف دوم رتبه ماتریس: r برابر تعداد سطرهای مستقل و تعداد ستون های مستقل در ماتریس‌های A، $ref(A)$, $rref(A)$ است.

- تعریف سوم رتبه ماتریس: r برابر بعد فضای ستون‌ها و بعد فضای سطرهای ماتریس است.

- طبق تعریف دوم r تعداد ستون‌های پاشنه (مستقل) ماتریس است. اگر تعداد کل ستون‌های ماتریس را با n نشان دهیم درنتیجه بعد فضای پوچی ماتریس برابر با $n-r$ خواهد بود.

- چند نکته مهم در مورد حذف:

۱- اگر یک ستون طی فرآیند حذف پاشنه نداشته باشد در اینصورت آن ستون ترکیب خطی ستون‌های قبلی است در غیراینصورت آن ستون از ستون‌های قبلی مستقل خطی است.

۲- اگر یک سطر پاشنه نداشته باشد در طول فرآیند حذف تماماً صفر خواهد شد و در انتهای ماتریس $\underline{R}$ قرار خواهد گرفت. در غیراینصورت آن سطر از سطرهای قبلی مستقل خطی است.

۳- اگر در فرآیند کاهش ماتریس $[A \quad I]$ را کاهش دهید و به $[R \quad E]$ دست یابید در اینصورت ماتریس E کل تاریخچه عملیات کاهش را به شما ارائه می‌دهد. به عبارت دیگر $EA=R$ خواهد بود. درحالتی که A مربعی و معکوس‌پذیر باشد E برابر A^{-1} و $R=I$ خواهد بود.

پاسخ کامل به دستگاه معادلات $\underline{A}\underline{x}=\underline{b}$

- در بخش قبل دستگاه معادلات $\underline{A}\underline{x}=\underline{b}$ را برای $\underline{b}=\underline{0}$ به‌طور کامل مورد بررسی قرار دادیم و مشاهده کردیم که حاصل برابر تمام بردارها در زیرفضای پوچی ماتریس A است و نحوه به دست آوردن این زیرفضا را نیز دیدیم. در این بخش به سراغ دستگاه معادلات $\underline{A}\underline{x}=\underline{b}$ برای حالت کلی $\underline{b}$ دلخواه می‌رویم.

- تفاوت با فضای پوچی: برای به دست آوردن جواب دستگاه $\underline{A}\underline{x}=\underline{0}$ نیازی به اعمال عملیات حذف بر روی بردار $\underline{0}$ نیست زیرا نتیجه برای هرگونه عملیات حذف خود بردار $\underline{0}$ است اما برای دستگاه $\underline{A}\underline{x}=\underline{b}$ باید عملیات بر روی $\underline{b}$ نیز انجام شود که در این شرایط از ماتریس $[A \quad b]$ استفاده می‌کنیم.

- جواب خصوصی دستگاه $\underline{A}\,\underline{x}_p = \underline{b}$:

۱- ماتریس افزوده $[A \;\; b]$ را تشکیل دهید.

۲- فرآیند حذف گاوس-جردن را بر روی آن اعمال کنید تا به ماتریس $[R \;\; d]$ برسید.

۳- دستگاه در صورتی جواب دارد که اگر سطری در $\underline{R}$ تماماً صفر گردید، سطر متناظر در d نیز صفر باشد (در این صورت بردار d در فضای ستونی ماتریس $\underline{R}$ است). یک جواب خصوصی دستگاه را می‌توان با برابر صفر قرار دادن متغیرهای آزاد و حل دستگاه حاصل برای متغیرهای غیرآزاد (متغیرهای پاشنه) به دست آورد.

مثال :

$$\begin{bmatrix} ۱ & ۳ & ۰ & ۲ \\ ۰ & ۰ & ۱ & ۴ \\ ۱ & ۳ & ۱ & ۶ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \\ x_۴ \end{bmatrix} = \begin{bmatrix} ۱ \\ ۶ \\ ۷ \end{bmatrix}$$

$$[A \;\; b] = \begin{bmatrix} ۱ & ۳ & ۰ & ۲ & ۱ \\ ۰ & ۰ & ۱ & ۴ & ۶ \\ ۱ & ۳ & ۱ & ۶ & \underset{\underset{\underline{b}}{\uparrow}}{۷} \end{bmatrix} \begin{array}{l} \rightarrow ۱ = \text{پاشنه} \\ \\ \rightarrow ۱ = \text{ضرب} \end{array}$$

$$\begin{bmatrix} ۱ & ۳ & ۰ & ۲ & ۱ \\ ۰ & ۰ & ۱ & ۴ & ۶ \\ ۰ & ۰ & ۱ & ۴ & ۶ \end{bmatrix} \begin{array}{l} \\ \rightarrow ۱ = \text{پاشنه} \\ \rightarrow ۱ = \text{ضرب} \end{array}$$

$$\begin{bmatrix} ۱ & ۳ & ۰ & ۲ & ۱ \\ ۰ & ۰ & ۱ & ۴ & ۶ \\ ۰ & ۰ & ۰ & ۰ & ۰ \end{bmatrix} = [R \;\; d]$$

$$\begin{array}{cccc} ۱ & ۳ & ۰ & ۲ \\ ۰ & ۰ & ۱ & ۴ \\ ۰ & ۰ & ۰ & ۰ \\ \uparrow & \uparrow & \uparrow & \uparrow \\ \text{متغیر پاشنه } x_۱ & \text{متغیر آزاد } x_۲ & \text{متغیر پاشنه } x_۳ & \text{متغیر آزاد } x_۴ \end{array} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \\ x_۴ \end{bmatrix} = \begin{bmatrix} ۱ \\ ۶ \\ ۰ \end{bmatrix}$$

$x_۲ = x_۴ = ۰$ قرار می‌دهیم و دستگاه را برای $x_۱$ ، $x_۳$ حل می‌کنیم.

$$\begin{cases} x_۱ + ۳x_۲ + ۲x_۴ = ۱ \Rightarrow x_۱ = ۱ \\ x_۳ + ۴x_۴ = ۶ \Rightarrow x_۳ = ۶ \\ ۰ = ۰ \Rightarrow \text{این رابطه نیز نشان می‌دهد که دستگاه جواب دارد.} \end{cases} \Rightarrow \underline{x}_p = \begin{bmatrix} ۱ \\ ۰ \\ ۶ \\ ۰ \end{bmatrix}$$

- پاسخ کامل دستگاه معادلات $\underline{A}\underline{x} = \underline{b}$:

$$\underline{x} = \underline{x}_p + \underline{x}_n$$

$\underline{x}_n$ ← کلیه بردارها در فضای پوچی ماتریس $\underline{A}$

$\underline{x}_p$ ← یک جواب خصوصی به دست آمده

- مثال: اگر A معکوس پذیر باشد، پاسخ کامل دستگاه $\underline{A}\underline{x}=\underline{b}$ به چه صورتی است.

$A \longrightarrow N(A) = Z$ معکوس پذیر است.

$A \longrightarrow \underline{x}_p = A^{-1}b$ معکوس پذیر است.

$$\Rightarrow \underline{x} = \underline{x}_p$$

- توجه: در حل دستگاه معادلات در مورد فضای پوچی ماتریس A صحبتی نکردیم زیرا فرض براین بود که A معکوس پذیر است و در نتیجه $N(A) = Z$ بود.

- مثال: شرایط مورد نیاز بردار $\underline{b}$ را به گونه ای به دست آورید که دستگاه معادلات زیر جواب داشته باشد.

$$\begin{bmatrix}1 & 1\\ 2 & 4\\ 3 & 6\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}b_1\\ b_2\\ b_3\end{bmatrix} = \underline{b}$$

$$[A \quad b] = \begin{bmatrix}1 & 1 & b_1\\ 2 & 4 & b_2\\ 3 & 6 & b_3\end{bmatrix}\begin{array}{l}\rightarrow \text{پاشنه} = 1\\ \rightarrow \text{ضرب} = 2\\ \rightarrow \text{ضرب} = 3\end{array} \Longrightarrow \begin{bmatrix}1 & 1 & b_1\\ 0 & 2 & b_2-2b_1\\ 0 & 3 & b_3-3b_1\end{bmatrix}$$

پاشنه = ۲

ضرب = ۳/۲

$$\Rightarrow \begin{bmatrix}1 & 1 & b_1\\ 0 & 2 & b_2-2b_1\\ 0 & 0 & (b_3-3b_1)-\frac{3}{2}(b_2-2b_1)\end{bmatrix}\begin{array}{l}\rightarrow \text{ضرب} = \frac{1}{2}\\ \rightarrow \text{پاشنه} = 2\\ \\ \end{array}$$

$$\Rightarrow \begin{bmatrix}1 & 0 & b_1-\frac{1}{2}(b_2-2b_1)\\ 0 & 2 & b_2-2b_1\\ 0 & 0 & b_3-\frac{3}{2}b_2\end{bmatrix}$$

دستگاه در صورتی جواب دارد که $b_3-\frac{3}{2}b_2=0$ گردد.

در این حالت جواب دستگاه برابر زیر است:

$$\xrightarrow[\text{پاشنه سطر دوم}]{\text{واحد کردن}} \left[\underbrace{\begin{matrix}1 & 0\\ 0 & 1\\ 0 & 0\end{matrix}}_{\underline{R}} \quad \underbrace{\begin{matrix}2b_1-\frac{1}{2}b_2\\ \frac{1}{2}b_2-b_1\\ b_3-\frac{3}{2}b_1\end{matrix}}_{\underline{d}}\right] \Rightarrow \underline{x}_p = \begin{bmatrix}2b_1-\frac{1}{2}b_2\\ \frac{1}{2}b_2-b_1\end{bmatrix}$$

در این مثال $N(\underline{A}) = Z$ است در نتیجه: $\underline{x} = \underline{x}_p$

- تعریف : ماتریس با رتبه کامل ستونی : ماتریس $A: m \times n$ رتبه کامل ستونی دارد اگر $r=n$ باشد. در این نوع از ماتریس ها $m \geq n$ است و ماتریس R متناظر آن به صورت زیر خواهد بود :

$$\underline{R} = \begin{bmatrix} \underline{I} \\ \hline \underline{0} \end{bmatrix} = \begin{bmatrix} \text{ماتریس همانی } n \times n \\ \text{سطر صفر } m-n \end{bmatrix}$$

ویژگی ها :

۱- تمام ستون های A ستون پاشنه هستند.

۲- متغیر آزاد و جواب ویژه ندارد.

۳- $N(\underline{A}) = Z$

۴- اگر $\underline{A}\,\underline{x} = \underline{b}$ جواب داشته باشد (سی ترازند نداشته باشد) آنگاه این جواب منحصر به فرد است.

۵- ستون های ماتریس A مستقل خطی هستند.

۶- $m-n$ شرط باید روی $\underline{b}$ وجود داشته باشد که دستگاه $\underline{A}\,\underline{x} = \underline{b}$ جواب منحصر به فرد داشته باشد.

- تعریف : ماتریس با رتبه کامل سطری : ماتریس $A: m \times n$ رتبه کامل سطری دارد اگر $r=m$ باشد.

ویژگی ها :

۱- تمام سطرهای A پاشنه دارند و A سطر ندارد.

۲- برای هر $\underline{b}$، دستگاه $\underline{A}\,\underline{x} = \underline{b}$ جواب دارد.

۳- فضای ستونی ستونی کل $\mathbb{R}^m$ است.

۴- بعد فضای پوچی ماتریس A برابر با $n-m$ است.

۵- سطرهای ماتریس A مستقل خطی هستند.

– مثال: جواب کامل دستگاه مقابل را به دست آورید.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \underline{b}$$

$$[A \quad b] = \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 1 & 4 & 6 & b_2 \end{bmatrix} \begin{matrix} \rightarrow \text{پاشنه} = 1 \\ \rightarrow \text{ضریب} = 1 \end{matrix} \Rightarrow \begin{bmatrix} 1 & 2 & 3 & b_1 \\ 0 & 2 & 3 & b_2 - b_1 \end{bmatrix} \begin{matrix} \rightarrow \text{ضریب} = 1 \\ \rightarrow \text{پاشنه} = 2 \end{matrix}$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 & 2b_1 - b_2 \\ 0 & 2 & 3 & b_2 - b_1 \end{bmatrix} \xrightarrow{\text{نرمالیزه کردن}} \begin{bmatrix} 1 & 0 & 0 & 2b_1 - b_2 \\ 0 & 1 & \underset{\uparrow}{3/2} & \frac{b_2 - b_1}{2} \end{bmatrix}$$

x_3 متغیر آزاد است

جواب خصوصی x_p :

$x_3 = 0$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2b_1 - b_2 \\ \frac{b_2 - b_1}{2} \end{bmatrix} \Rightarrow \begin{cases} x_1 = 2b_1 - b_2 \\ x_2 = \frac{b_2 - b_1}{2} \end{cases}$$

$$\Rightarrow \underline{x}_p = \begin{bmatrix} 2b_1 - b_2 \\ \frac{b_2 - b_1}{2} \\ 0 \end{bmatrix}$$

جواب ویژه $\underline{x}_n$:

$x_3 = 1$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = -3/2 \end{cases}$$

$$\Rightarrow \underline{x}_n = \alpha \begin{bmatrix} 0 \\ -3/2 \\ 1 \end{bmatrix}, \forall \alpha \in \mathbb{R}$$

$$\underline{x} = \begin{bmatrix} 2b_1 - b_2 \\ \frac{b_2 - b_1}{2} \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ -3/2 \\ 1 \end{bmatrix}$$

- مثال : جواب کامل دستگاه زیر را به دست آورید و با رسم شکل آن را نشان دهید.

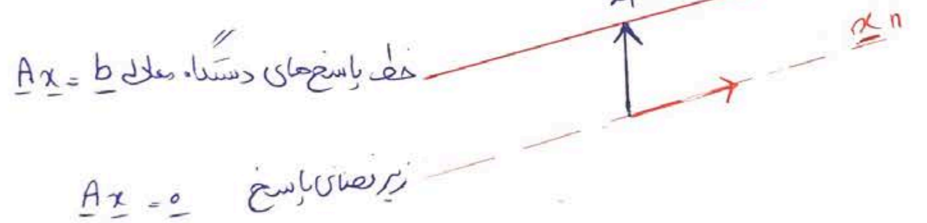
$$\begin{bmatrix}۱&۱&۱\\۱&۲&-۱\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}۳\\۴\end{bmatrix}$$

$$[\underline{A}\ \ \underline{b}]=\begin{bmatrix}۱&۱&۱&۳\\۱&۲&-۱&۴\end{bmatrix}\begin{matrix}\rightarrow \text{۱ = پاشنه}\\ \rightarrow \text{۱ = ضرب}\end{matrix}\Rightarrow\begin{bmatrix}۱&۱&۱&۳\\۰&۱&-۲&۱\end{bmatrix}\begin{matrix}\rightarrow \text{۱ = ضرب}\\ \rightarrow \text{۱ = پاشنه}\end{matrix}$$

$$\Rightarrow\begin{bmatrix}۱&۰&۳&۲\\۰&۱&-۲&۱\end{bmatrix}\qquad \underline{x}_p=\begin{bmatrix}۲\\۱\\۰\end{bmatrix}$$

$\uparrow$ متغیر آزاد

$$\underline{x}_n=\alpha\begin{bmatrix}-۱\\۳\\۱\end{bmatrix}\qquad \underline{x}=\underline{x}_p+x_n$$



- جمع بندی :

$۱\rightarrow \begin{cases}r=m\\ r=n\end{cases}\rightarrow$ A مربعی و معکوس‌پذیر $\rightarrow$ $\underline{A}\underline{x}=b$ دقیقاً یک جواب دارد. $\rightarrow \underline{R}=[\underline{I}]$

$۲\rightarrow \begin{cases}r=m\\ m<n\end{cases}\rightarrow$ A کشیده افقی $\rightarrow$ $\underline{A}\underline{x}=\underline{b}$ بی‌نهایت جواب دارد. $\rightarrow \underline{R}=[\underline{I}\ \ F]$

$۳\rightarrow \begin{cases}r<m\\ r=n\end{cases}\rightarrow$ A کشیده عمودی $\rightarrow$ $A\underline{x}=\underline{b}$ صفر یا یک جواب دارد. $\rightarrow \underline{R}=\begin{bmatrix}I\\ \underline{0}\end{bmatrix}$

$۴\rightarrow \begin{cases}r<m\\ r<n\end{cases}\rightarrow$ A رتبه کامل نیست. $\rightarrow$ $\underline{A}\underline{x}=\underline{b}$ صفر یا بی‌نهایت جواب دارد. $\rightarrow R=\begin{bmatrix}I&F\\0&0\end{bmatrix}$

استقلال، پایه و بعد

- تعریف: استقلال و وابستگی: یک زیرمجموعه S از فضای برداری V وابسته خطی (وابسته) گوییم اگر بردارهای متمایز α_1 و α_2 و ... و α_n در S و اعداد c_1، c_2، ...، c_n در F (که حداقل یکی از آنها غیرصفر است) وجود داشته باشند به گونه‌ای که:

$$c_1\alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n = \underline{0}$$

مجموعه‌ای که وابسته خطی نباشد را مستقل خطی می‌نامند.

- تعریف: زیرفضای پوشش داده شده توسط مجموعه S: فرض کنید S یک زیر مجموعه از بردارها در فضای برداری V باشد. زیرفضای پوشش داده شده توسط S به صورت اشتراک W از تمام زیر فضاهای V که S را شامل می‌شوند تعریف می‌گردد. هنگامی که S یک مجموعه محدود از بردارها به صورت $S = \{\alpha_1, \ldots, \alpha_n\}$ است، W را زیرفضای پوشش داده شده توسط بردارهای α_1، α_2 و ...، α_n می‌نامیم.

- قضیه: زیرفضای پوشش داده شده توسط زیرمجموعه غیرتهی S از فضای برداری V مجموعه تمام ترکیب‌های خطی بردارهای موجود در S است.

- تعریف: پایه: فرض کنید V یک فضای برداری است. یک پایه برای V مجموعه‌ای از بردارهای مستقل خطی در V است که فضای V را پوشش می‌دهد. فضای V بعد محدود است اگر پایه آن محدود باشد (تعداد بردارها در پایه محدود باشد).

- قضیه: فرض کنید V فضای برداری است که با مجموعه محدودی از بردارهای $\underline{\beta}_1$، ... و $\underline{\beta}_m$ پوشش داده می‌شود. در اینصورت هر مجموعه از بردارهای مستقل در V محدود بوده و نمی‌تواند بیشتر از m عضو داشته باشد.

اثبات: فرض کنیم S زیرمجموعه از V با n عضو به صورت $\underline{\alpha}_1$، $\underline{\alpha}_2$، ...، $\underline{\alpha}_n$ است که $n > m$. چون $\underline{\beta}_1$، $\underline{\beta}_2$، ...، $\underline{\beta}_m$ فضای V را پوشش می‌دهند ضرایب A_{ij} در F وجود دارد به گونه‌ای که:

$$\underline{\alpha}_j = \sum_{i=1}^{m} A_{ij}\,\underline{\beta}_i$$

برای هر n عدد x_1، x_2، … ، x_n رابطه زیر را داریم:

$$x_1\underline{\alpha}_1+\cdots+x_n\underline{\alpha}_n=\sum_{j=1}^{n}x_j\alpha_j=\sum_{j=1}^{n}x_j\sum_{i=1}^{m}A_{ij}\underline{\beta}_i$$

$$=\sum_{i=1}^{m}\underbrace{\left(\sum_{j=1}^{n}A_{ij}\cdot x_j\right)}_{y_i}\beta_i$$

$$\begin{bmatrix}A_{11} & \cdots & A_{1n}\\ \vdots & \ddots & \vdots\\ A_{m1} & \cdots & A_{mn}\end{bmatrix}_{m\times n}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}\rightarrow$$

چون $m<n$ است، حتماً ستون آزاد داریم و دستگاه، مقابل جواب غیربدیهی دارد. درنتیجه:

مجموعه S وابسته خطی است

$x_1\alpha_1+\cdots+x_n\alpha_n=0\Longrightarrow$

نتیجه: اگر V یک فضای برداری با بعد محدود باشد درانصورت هردو پایه V تعداد اعضای برابری دارند ← تعداد اعضای پایه مستقل از انتخاب پایه بوده و یک ویژگی V است.

تعریف بعد: تعداد بردارهای موجود در هر پایه برای فضای برداری V را بعد فضای برداری V می‌نامیم و با dim(V) نشان می‌دهیم.

تعریف: ستون‌های ماتریس A مستقل خطی هستند اگر تنها جواب به دستگاه $A\underline{x}=\underline{0}$، جواب بدیهی $\underline{x}=\underline{0}$ باشد. به طور معادل ستون‌ها مستقل خطی هستند اگر فضای پوچی ماتریس A برابر Z باشد.

- ماتریس $A: m\times n$ با رتبه کامل ستونی درنظر بگیرید $(r=n)$. درانصورت ستون آزاد نداشته و تمامی ستون‌ها مستقل خطی هستند.

- فضای ستونی: فضای ستونی ماتریس A فضایی است که توسط ستون های ماتریس پوشش داده می‌شود.

- فضای سطری: فضای سطری ماتریس A فضایی است که توسط سطرهای ماتریس پوشش داده می‌شود.

- فضای سطری ماتریس A برابر فضای ستونی ماتریس A^T است، یعنی:

A فضای سطری ماتریس $= C(A^T)$

- قضیه: نمایش یک بردار $\underline{v}$ برحسب پایه $\underline{v}_1$، v_2 و ...، v_n منحصر به فرد است.

اثبات:

$$\left.\begin{array}{l}\underline{v} = a_1\underline{v}_1 + a_2\underline{v}_2 + \cdots + a_n v_n \\ \underline{v} = b_1\underline{v}_1 + b_2\underline{v}_2 + \cdots + b_n v_n\end{array}\right\} \Rightarrow \underline{0} = (a_1-b_1)\underline{v}_1 + \cdots + (a_n-b_n)\underline{v}_n$$

رابطه فوق تناقض است زیرا بردارهای v_1، v_2 و ...، v_n مستقل خطی هستند. پس نمایش یک بردار دلخواه برحسب پایه منحصر به فرد است.

- پایه استاندارد: ستون‌های ماتریس $\underline{I}$ پایه استاندارد برای فضای برداری $\mathbb{R}^n$ هستند.

- پایه فضای ستونی: در ماتریس A، ستون‌های پاشنه پایه‌ای برای فضای ستونی هستند.

- پایه فضای سطری: در ماتریس A، سطرهای پاشنه پایه‌ای برای فضای سطری هستند.

* فضای برداری ستون‌های A و R متفاوت است. پایه‌های این فضاها متفاوت است اما بعد آنها برابر است.

* فضای سطری A و R یکسان است.

- سؤال: پنج بردار در فضای $\mathbb{R}^7$ در اختیار شما قرار می‌گیرد. چگونه یک پایه برای این بردارها به دست می‌آورید

پاسخ: بردارها را به صورت ستونی در یک ماتریس قرار می‌دهید (ماتریس A) با استفاده از روش حذفی ماتریس R را تعیین می‌نمایید و ستون‌های پاشنه را مشخص می‌کنید. ستون‌های پاشنه در ماتریس A پایه‌ای برای بردارها هستند.

- بعد فضای برداری Z برابر صفر تعریف می‌شود و مجموعه تهی یک پایه برای این فضای برداری است.

ابعاد چهار زیرفضای اساسی

- چهار زیرفضای اساسی ماتریس $A: m\times n$ عبارتند از :

۱) فضای سطری ماتریس A که با $C(A^T)$ نشان می‌دهیم و زیرفضایی از $\mathbb{R}^n$ است .

۲) فضای ستونی ماتریس A که با $C(A)$ نشان می‌دهیم و زیرفضایی از $\mathbb{R}^m$ است .

۳) فضای پوچی ماتریس A که با $N(A)$ نشان می‌دهیم و زیرفضایی از $\mathbb{R}^n$ است .

۴) فضای پوچی چپ ماتریس A که با $N(A^T)$ نشان می‌دهیم و زیرفضایی از $\mathbb{R}^m$ است .

- برای به دست آوردن $N(A^T)$ دستگاه معادلات $\underline{A}^T\underline{y}=\underline{0}$ را حل می‌کنیم .
این دستگاه، معادل دستگاه $y^TA=\underline{0}^T$ است . به همین دلیل $N(A^T)$ را فضای پوچی چپ ماتریس A نیز می‌نامیم .

- روابط مقابل را برای ابعاد زیرفضاهای اساسی ماتریس $A: m\times n$ داریم :

$$\dim\big(C(A)\big)=r=\dim\big(C(A^T)\big)$$
$$\dim\big(N(A)\big)=n-r$$
$$\dim\big(N(A^T)\big)=m-r$$

- زیرفضاها در R $(R=rref(A))$:

$$R=\begin{array}{c|cc} & \overleftrightarrow{r} & \overleftrightarrow{n-r} \\ \hline r\updownarrow & \underline{I} & \underline{F} \\ m-r & \underline{0} & \underline{0} \end{array}$$

همانطور که می‌دانیم R در حالت کلی به فرم زیر قابل بیان است :
براساس این فرم نتایج زیر به سادگی قابل استنتاج هستند :

۱- بعد فضای سطری R برابر r است و سطرهای غیرصفر یک پایه این فضا هستند .

۲- بعد فضای ستونی R برابر r است و ستون‌های پاشنه یک پایه برای این فضا هستند .

۳- بعد فضای پوچی R برابر $n-r$ است و جواب‌های ویژه یک پایه برای این فضا هستند .

۴- بعد فضای پوچی چپ R برابر $m-r$ است و آن به صورت زیر به دست می‌آید :

$$\underline{R}^T\underline{y}=\underline{0}\Rightarrow \underline{y}^T\underline{R}=\underline{0}^T\Rightarrow \underline{y}^T\begin{bmatrix}-\underline{r}_1^T-\\ \vdots\\ -\underline{r}_m^T-\end{bmatrix}=0$$

$$\Rightarrow \left.\begin{array}{l} y_1 \underline{r}_1^T \\ \vdots \\ + y_r \underline{r}_r^T \end{array}\right\} \rightarrow$$

در این بخش به علت حضور بردار I در فرم کلی R ضرایب می بایست صفر باشند

$$\left.\begin{array}{l} y_{r+1} r_{r+1}^T \\ \vdots \\ + y_m r_m^T \end{array}\right\} \rightarrow$$

$$\overline{\qquad \underline{o}^T \qquad}$$

در این قسمت ما برابر یک قرار دادن یکی از ضرایب و صفر کردن بقیه یک پایه از فضای پوچی چپ به دست می آید. در نتیجه در کل $m-r$ پایه خواهیم داشت.

مثال:

$$\underline{R} = \begin{bmatrix} \boxed{۱} & ۳ & ۵ & ۰ & ۷ \\ ۰ & ۰ & ۰ & \boxed{۱} & ۲ \\ ۰ & ۰ & ۰ & ۰ & ۰ \end{bmatrix} \rightarrow$$

چون دو پاشنه داریم رتبه برابر ۲ است.
بعد فضای سطری و ستونی برابر ۲ است.
بعد فضای پوچی برابر $n-r = ۵-۲ = ۳$ است.
بعد فضای پوچی چپ برابر $m-r = ۳-۲ = ۱$ است.

پایه های فضای ستونی، ستون هایی هستند که در آن ها پاشنه داریم، یعنی: $\begin{bmatrix} ۰ \\ ۱ \\ ۰ \end{bmatrix}$, $\begin{bmatrix} ۱ \\ ۰ \\ ۰ \end{bmatrix}$

پایه های فضای سطری، سطرهایی هستند که در آن ها پاشنه داریم، یعنی:

$$\begin{bmatrix} ۱ \\ ۳ \\ ۵ \\ ۰ \\ ۷ \end{bmatrix} \quad \begin{bmatrix} ۰ \\ ۰ \\ ۰ \\ ۱ \\ ۲ \end{bmatrix}$$

پایه های فضای پوچی برابر جواب های ویژه هستند، یعنی:

$$s_۲ = \begin{bmatrix} -۳ \\ ۱ \\ ۰ \\ ۰ \\ ۰ \end{bmatrix} \quad s_۳ = \begin{bmatrix} -۵ \\ ۰ \\ ۱ \\ ۰ \\ ۰ \end{bmatrix} \quad s_۵ = \begin{bmatrix} -۷ \\ ۰ \\ ۰ \\ -۲ \\ ۱ \end{bmatrix}$$

پایه های فضای پوچی چپ برابر است با:

$$\begin{bmatrix} ۰ \\ ۰ \\ ۱ \end{bmatrix}$$

- زیرفضاهای اساسی در A:

$$A = \begin{bmatrix} ۱ & ۳ & ۵ & ۰ & ۷ \\ ۰ & ۰ & ۰ & ۱ & ۲ \\ ۱ & ۳ & ۵ & ۱ & ۹ \end{bmatrix} \xrightarrow{\text{حذف گاوس جردن}} R = \begin{bmatrix} ۱ & ۳ & ۵ & ۰ & ۷ \\ ۰ & ۰ & ۰ & ۱ & ۲ \\ ۰ & ۰ & ۰ & ۰ & ۰ \end{bmatrix}$$

۱) فضای سطری A: فضای سطری A و R یکسان است. بعد این فضا r و پایه های آن نیز یکسان هستند.
علت: سطرهای A ترکیب خطی سطرهای R است بنابراین فضای سطری این دو ماتریس یکسان است.
به عبارت دیگر عملیات حذف سطرها را تغییر می دهد اما فضای سطری را تغییر نمی دهد.
پایه های فضای سطری A می تواند برابر با r سطر اول ماتریس R انتخاب شود. به طور معادل می توان سطرهای پاشنه از ماتریس A را نیز به عنوان پایه فضای سطری A در نظر گرفت.

۲) فضای ستونی A : بعد فضای ستونی A برابر رتبه ماتریس r است (قضیه بعد).
بنابراین بعد فضای ستونی A و R برابر است. اما به سادگی از فرم R مشخص است که خود این فضاها متفاوت هستند. هر ستون در ماتریس R به m-r صفر ختم می‌شود و در نتیجه در فضای ستونی ماتریس نیز این شرایط را داریم در حالیکه این شرایط لزوماً برای A وجود ندارد.

توجه: هر ترکیب خطی از ستون های A که منجر به بردار صفر شود همان ترکیب خطی از ستون های R نیز صفر می‌شود.

علت: $A\underline{x}=0$ است دقیقاً زمانی که $R\underline{x}=0$ باشد $(N(A)=N(R))$

بنابراین ستون های پاشنه پایه‌ای برای فضای ستونی A هستند.

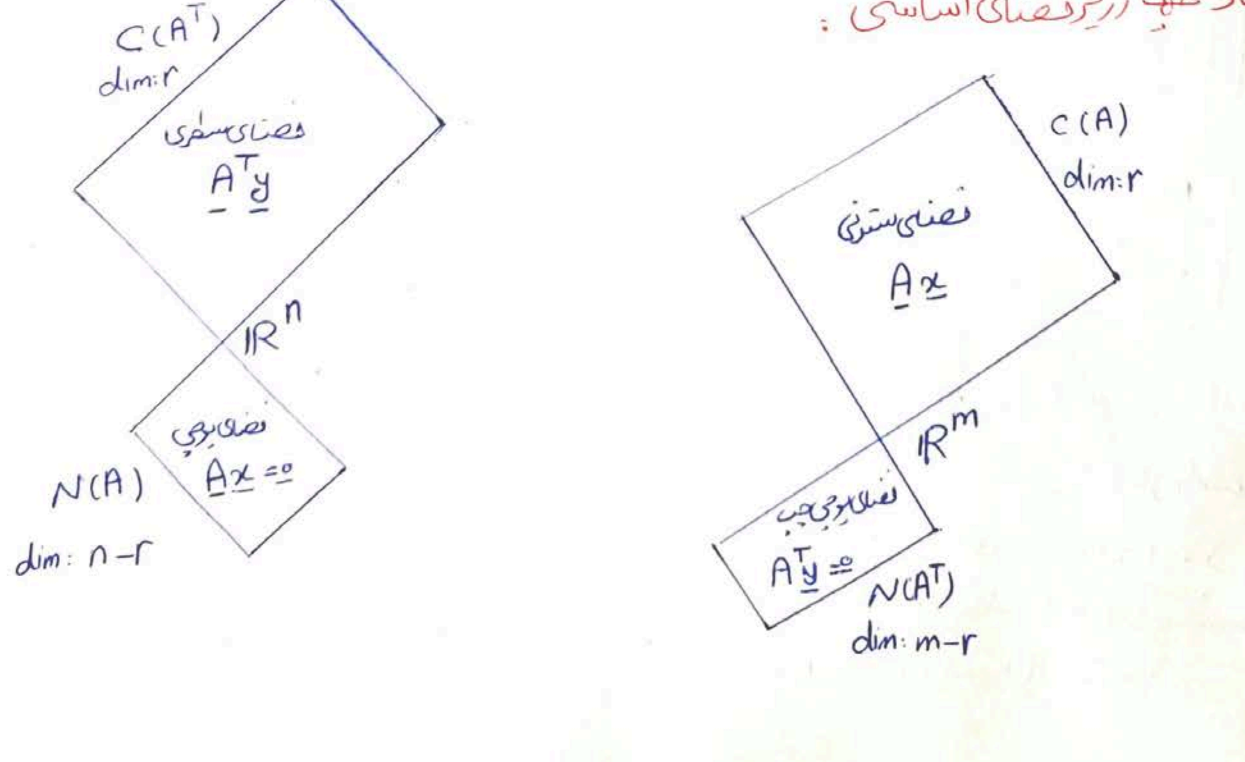
۳) فضای پوچی A : فضای پوچی ماتریس A و R یکسان است و بعد آن n-r است.

قضیه شمارش: مجموع بعد فضای ستونی و فضای پوچی در ماتریس $A: m\times n$ برابر با n است.

۴) فضای پوچی چپ ماتریس A : بعد فضای پوچی چپ ماتریس A برابر m-r است.

علت: در ماتریس A^T بعد فضای ستونی r است.
در ماتریس A^T طبق قضیه شمارش مجموع ابعاد فضای ستونی و پوچی برابر m است.
⇐ بعد فضای پوچی چپ (فضای پوچی A^T) برابر m-r است.

- ابعاد چهار زیرفضای اساسی :



- مثال:

$$A = \begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix} \Rightarrow r = ۱ \Rightarrow \begin{cases} \text{پایه فضای ستونی} = \{ [۱] \} \\ \text{پایه فضای سطری} = \left\{ \begin{bmatrix} ۱ \\ ۲ \\ ۳ \end{bmatrix} \right\} \end{cases}$$

فضای پوچی: $\underline{A}\,\underline{x} = ۰ \Rightarrow \begin{bmatrix} ۱ & ۲ & ۳ \end{bmatrix} \begin{bmatrix} x_۱ \\ x_۲ \\ x_۳ \end{bmatrix} = ۰ \Rightarrow$ $x_۲$ و $x_۳$ متغیر آزادند

$$\Rightarrow \text{جواب‌های ویژه:} \begin{cases} x_۲ = ۱ ,\ x_۳ = ۰ \Rightarrow s_۱ = \begin{bmatrix} -۲ \\ ۱ \\ ۰ \end{bmatrix} \\ x_۲ = ۰ ,\ x_۳ = ۱ \Rightarrow s_۲ = \begin{bmatrix} -۳ \\ ۰ \\ ۱ \end{bmatrix} \end{cases}$$

$$\text{پایه‌های فضای پوچی} = \left\{ \begin{bmatrix} -۲ \\ ۱ \\ ۰ \end{bmatrix} , \begin{bmatrix} -۳ \\ ۰ \\ ۱ \end{bmatrix} \right\}$$

فضای پوچی چپ: بعد فضای پوچی چپ $= m - r = ۰ \Rightarrow N(A^T) = Z$

- ماتریس‌های رتبه r (تعبیر جدید): فرض کنید ماتریس $A: m \times n$ از رتبه r باشد. در این صورت طی عملیات حذف گاوس-جردن می‌توانیم ماتریس E را به گونه‌ای به دست آوریم که $\underline{E}\underline{A} = \underline{R}$ گردد. دیدیم که این ماتریس معکوس‌پذیر است. فرض کنید $C = E^{-1}$ باشد.

در اینصورت $\underline{A} = \underline{C}\underline{R}$. حال فرم‌های مقابل را درنظر بگیرید.

$$\underline{A} = \begin{bmatrix} \underline{a}_۱ & \underline{a}_۲ & \cdots & \underline{a}_n \end{bmatrix}$$

$$\underline{C} = \begin{bmatrix} \underline{c}_۱ & \underline{c}_۲ & -- & \underline{c}_m \end{bmatrix}$$

$$R = \begin{bmatrix} - r_۱^T - \\ \vdots \\ - r_m^T - \end{bmatrix} = r\updownarrow \begin{bmatrix} \overset{\overset{r}{\leftrightarrow}}{I} & F \\ ۰ & ۰ \end{bmatrix} = \begin{bmatrix} \tilde{r}_۱ , & — & \tilde{r}_n \end{bmatrix}$$

$$\left.\begin{aligned} \underline{a}_i &= \underline{C}\ \tilde{r}_i \\ ۱ \leq i \leq r \Rightarrow \tilde{r}_i &= \underline{e}_i \end{aligned}\right\} \Rightarrow \underline{a}_i = \underline{c}_i ,\ ۱ \leq i \leq r$$

ستون‌های اول تا r-ام ماتریس $\underline{A}$، $\underline{C}$ برابر هستند.

$$\left.\begin{aligned} \underline{A} = \underline{C}\,\underline{R} &= \sum_i c_i r_i^T \\ r_i^T = \underline{0}^T \qquad & i \geq r \end{aligned}\right\} \Rightarrow \underline{A} = \sum_{i=1}^{r} c_i\ r_i^T \Rightarrow$$

ماتریس $\underline{A}$ با رتبه $\underline{r}$ را می‌توان به صورت مجموع $\underline{r}$ ماتریس رتبه یک نوشت

بردارهای مختلط

- اعداد مختلط :

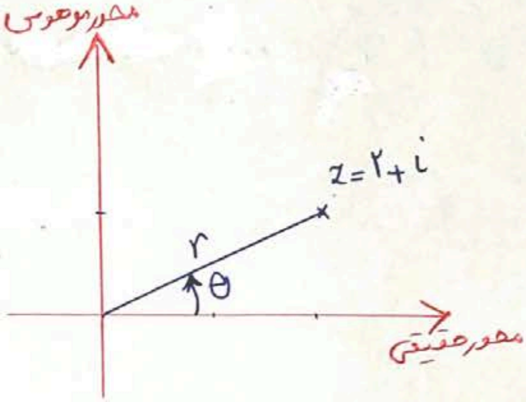
$$z = x + jy = re^{j\theta} \qquad e^{j\theta} = \cos\theta + j\sin\theta$$

$$z_1 = x_1 + j\,y_1 = r_1 e^{j\theta_1} \qquad j \triangleq \sqrt{-1}$$

$$z_۲ = x_۲ + j\,y_۲ = r_۲ e^{j\theta_۲}$$

$$r = \sqrt{x^۲ + y^۲}$$

$$\theta = \tan^{-1} y/x$$



$$Re\{z\} = x$$
$$Im\{z\} = y$$

- اعداد مختلط متناظر با یک صفحه در دستگاه با دو محور اعداد حقیقی و موهومی هستند. این صفحه با $\mathbb{C}$ نمایش داده می‌شود و رفتاری مشابه $\mathbb{R}^۲$ دارد.

- مزدوج مختلط :

$$\bar{z} = x - jy = re^{j(-\theta)}$$

- خواص :

$$z_1 + z_۲ = (x_1 + x_۲) + j(y_1 + y_۲)$$

$$z_1 z_۲ = (x_1 x_۲ - y_1 y_۲) + j\,(x_1 y_۲ + x_۲ y_1) = r_1 r_۲ e^{j(\theta_1 + \theta_۲)}$$

$$\overline{z_1 z_۲} = \bar{z}_1\,\bar{z}_۲$$

$$\overline{z_1 + z_۲} = \bar{z}_1 + \bar{z}_۲$$

$$z\bar{z} = r^۲$$

$$z^n = r^n e^{j(n\theta)} = r^n(\cos n\theta + j\sin n\theta)$$

- ماتریس‌های هرمیتی (ترانهاده مزدوج) . برای ماتریس $A = [a_{ij}]_{m\times n}$ ماتریس A^H را ترانهاده مزدوج می‌نامیم و به صورت زیر تعریف می‌شود.

$$A^H = [\bar{a}_{ji}]_{n\times m}$$

- طول بردار (حالت کلی) : $\|z\| = \sqrt{z^H z} = \sqrt{\sum_i |z_i|^2}$

- ضرب داخلی (حالت کلی) : $\underline{u} \cdot \underline{v} = \underline{u}^H \underline{v} = \sum_i \bar{u}_i v_i$

توجه : $\underline{u} \cdot \underline{v} = \overline{\underline{v} \cdot \underline{u}}$

- مزدوج مختلط ضرب ماتریس‌ها : $(AB)^H = B^H A^H$

- تعریف : ماتریس هرمیتی : ماتریس $\underline{S}$ را هرمیتی گوییم اگر : $\underline{S} = \underline{S}^H$

- خاصیت : برای ماتریس هرمیتی $\underline{S}$ و بردار ستونی دلخواه $\underline{z}$ ، عدد $\underline{z}^H \underline{S}\, \underline{z}$ حقیقی است.

تعامد

- بردارهای متعامد : دو بردار $\underline{v}$ و $\underline{w}$ متعامدند اگر $\underline{v}^T \cdot \underline{w} = 0 \quad (\underline{v}^H \underline{w} = 0)$

- زیرفضاهای متعامد : دو زیرفضای W, V از یک فضای برداری متعامدند اگر به ازای هر بردار $\underline{v}$ در V و هر بردار $\underline{w}$ در W رابطه زیر برقرار باشد :

$$\underline{v}^T \cdot \underline{w} = 0 \quad (\underline{v}^H \underline{w} = 0)$$

- هر $\underline{x}$ در فضای پوچی ماتریس A بر تمام سطرهای ماتریس A عمود است. درنتیجه فضای پوچی ماتریس A بر فضای سطری این ماتریس متعامد است.

$$\underline{A}\, \underline{x} = \begin{bmatrix} -a_1^T- \\ \vdots \\ -a_m^T- \end{bmatrix} \underline{x} \Rightarrow \underline{a}_i^T \underline{x} = 0 \Rightarrow$$

$\underline{x}$ بر بردارهای پایه فضای سطری متعامد است

$\underline{A}^T \underline{y} \rightarrow$ برداری دلخواه در

$\underline{x} \rightarrow$ برداری دلخواه در فضای پوچی

$$\left.\right\} \; x^T (A^T y) = (Ax)^T y = \underline{0}^T \underline{y} = 0$$

- هر بردار $\underline{y}$ در فضای پوچی A^T (فضای پوچی چپ ماتریس A) به تمام ستون‌های ماتریس $\underline{A}$ عمود است. در نتیجه فضای پوچی $N(A^T)$ بر فضای ستونی ماتریس A است.

$$\underline{A}^T\underline{y} = \begin{bmatrix} (\underline{A} \ \text{ستون اول})^T \\ \vdots \\ (\underline{A} \ \text{ستون n-ام})^T \end{bmatrix} \underline{y} = \underline{0} \Rightarrow$$

$\underline{y}$ بر تمام ستون‌های $\underline{A}$ عمود است.

- **مکمل‌های متعامد**: مکمل متعامد یک زیرفضای برداری V هر برداری که بر V عمود است را شامل می‌شود. این زیرفضای متعامد با $V^{\perp}$ نمایش داده می‌شود.

- $N(A)$ مکمل متعامد فضای سطری $C(A^T)$ در $\mathbb{R}^n$ است.

اثبات: اگر $\underline{x}$ بر فضای سطری عمود باشد باید رابطه $\underline{A}\,\underline{x} = \underline{0}$ را برآورده سازد. پس $\underline{x}$ در فضای پوچی قرار می‌گیرد.

اگر $\underline{x}$ بر فضای پوچی $N(A)$ عمود باشد و فرض کنیم در فضای سطری $C(A^T)$ قرار نگیرد. در اینصورت می‌توانیم $\underline{x}$ را به عنوان یک سطر جدید به ماتریس A اضافه کنیم. چون $\underline{x}$ بر فضای پوچی عمود است، فضای پوچی تغییری نکند. از طرف دیگر چون سطر جدید در فضای سطری نیست بعد این فضا افزایش می‌یابد. در نتیجه مجموع بعد فضای پوچی و فضای سطری از n بیشتر می‌شود که با قضیه شمارش در تناقض است.

- $N(A^T)$ مکمل متعامد فضای ستونی $C(A)$ در $\mathbb{R}^m$ است.

ـ خلاصه :

تئوری اساسی جبرخطی - قسمت اول : فضای سطری و ستونی هر دو دارای بعد r هستند .

فضای پوچی دارای بعد $n-r$ است .

فضای پوچی چپ دارای بعد $m-r$ است .

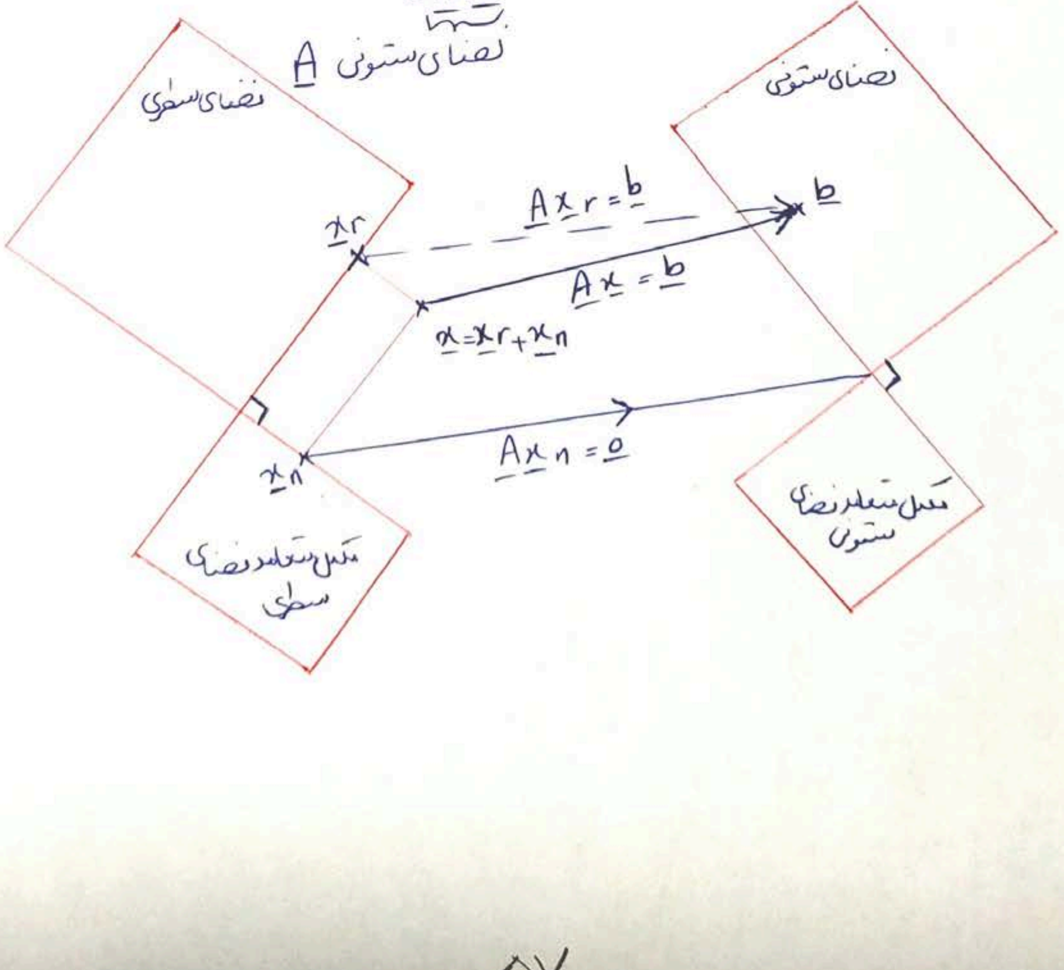
تئوری اساسی جبرخطی - قسمت دوم : $N(A)$ مکمل متعامد $C(A^T)$ در $\mathbb{R}^n$ است .

$N(A^T)$ مکمل متعامد $C(A)$ در $\mathbb{R}^m$ است .

توجه : قسمت اول بعد زیرفضاها ، قسمت دوم متعامد بودن آنها را نشان داد .

- با توجه به تئوری مکمل های متعامد می توان نشان داد که هر بردار $\underline{x}$ می تواند به دو بخش فضای سطری $\underline{x}_r$ ، فضای پوچی $\underline{x}_n$ شکسته شود . در اینصورت برای $\underline{A}\,\underline{x}$ روابط زیر را داریم :

$$\underline{A}\,\underline{x} = \underline{A}(\underline{x}_r + \underline{x}_n) = \underbrace{\underline{A}\,\underline{x}_r}_{\underbrace{=\underline{A}\,\underline{x}}_{\text{فضای ستونی } \underline{A}}} + \underbrace{\underline{A}\,\underline{x}_n}_{=\underline{0}}$$



- قضیه: هر بردار $\underline{b}$ در فضای ستونی ماتریس $\underline{A}$ متناظر با یک و فقط یک بردار $\underline{x}_r$ از فضای سطری است.

اثبات: فرض کنیم $\underline{A}\underline{x}_r = \underline{A}\underline{x}_r'$ در اینصورت $\underline{x}_r - \underline{x}_r'$ در فضای پوچی است و این با فرض متعامد بودن فضای سطری بر فضای پوچی در تناقض است مگر اینکه $\underline{x}_r - \underline{x}_r' = \underline{0}$ باشد. یعنی $\underline{x}_r = \underline{x}_r'$

نگاشت

- نگاشت به محورها و صفحه‌های استاندارد:

- نگاشت به محور z: بردار ستونی $b=(b_x,b_y,b_z)$ را می‌خواهیم روی محور z نگاشت دهیم. همانطور که مشخص است تنها مؤلفه سوم روی محور z قابل نمایش است و محورهای x و y بر این محور متعامدند. پس نگاشت b روی محور z که با $\underline{P}_1$ نمایش می‌دهیم برابر $(0,0,b_z)$ است. این نتیجه با استفاده از ضرب ماتریسی ماتریس $\underline{P}_1$ در بردار $\underline{b}$ نیز بصورت زیر حاصل می‌شود.

$$\underline{P}_1 = \underbrace{\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}}_{\underline{P}_1}\underline{b}$$

اگر بخواهیم این بردار را روی صفحه xy نگاشت دهیم حاصل $\underline{P}_2=(b_x,b_y,0)$ خواهد بود که از ضرب ماتریسی زیر به دست می‌آید.

$$\underline{P}_2 = \underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&0\end{bmatrix}}_{\underline{P}_2}\underline{b}$$

توجه: محور z و صفحه xy تشکیل دو زیرفضای مکمل متعامد در فضای $\mathbb{R}^3$ می‌دهند. هر بردار $\underline{b}$ در فضای $\mathbb{R}^3$ را می‌توان بصورت ترکیب اجزای آن در این دو زیرفضا نوشت. برای بردار $\underline{b}$، $\underline{P}_1$ و $\underline{P}_2$ اجزای آن در دو زیرفضای محور z و صفحه xy است.

- توجه:

$\underline{P}_1 + \underline{P}_2 = \underline{b}$

$\underline{P}_1 + \underline{P}_2 = \underline{I}$

- نگاشت به $C(A)$:

نگاشت به $C(A)$ همان $\underline{P}_1$ است. ← فضای ستونی A محور z است. $\leftarrow$ $A_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

نگاشت به $C(A_2)$ همان $\underline{P}_2$ است. ← فضای ستون A صفحه xy است. $\leftarrow$ $A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

نگاشت به $C(A_3)$ همان $\underline{P}_2$ است. ← فضای ستون A صفحه xy است. $\leftarrow$ $A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}$

- مشخصه اصلی در نگاشت: در تمامی موارد فوق، می‌توان نشان داد که تفاضل دو بردار $\underline{b}$ و $\underline{P}$ ($\underline{b} - \underline{P}$) بر بردار $\underline{P}$ عمود است. این تعامد مشخصه اصلی در نگاشت است و در ادامه نیز از آن استفاده خواهیم نمود. تفاضل بردارهای $\underline{P}$ و $\underline{b}$ را بردار خطا می‌نامیم.

- نگاشت به زیرفضای خط با جهت $\underline{a} = (a_1, \dots, a_m)$:

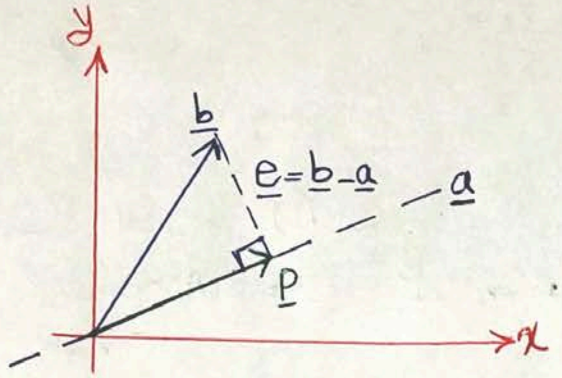
$$\left.\begin{array}{l} \underline{b} \\ \underline{P} = \hat{x}\,\underline{a} \end{array}\right\} \Rightarrow \left.\begin{array}{c} \underline{e} = \underline{b} - \hat{x}\underline{a} \\ \text{تعامد} \end{array}\right\} \Rightarrow \underline{a} \cdot \underline{e} = 0 \Rightarrow \underline{a} \cdot \underline{b} - \hat{x}\, a \cdot a = 0$$

$$\Rightarrow \hat{x} = \frac{\underline{a}^T \underline{b}}{\underline{a}^T \underline{a}} \Rightarrow \underline{P} = \frac{\underline{a}^T \underline{b}}{\underline{a}^T \underline{a}}\, \underline{a}$$

بنابراین نگاشت بردار $\underline{b}$ روی راستای $\underline{a}$ به صورت $\frac{\underline{a}^T \underline{b}}{\underline{a}^T \underline{a}}\, \underline{a}$ خواهد بود.

حالت ۱: $\underline{a} = \underline{b}$ ⇐ در این حالت $\underline{P} = \underline{b}$

حالت ۲: $\underline{a}$ متعامد بر $\underline{b}$ ⇐ $\underline{P} = \underline{0}$



تعبیر هندسی:

اگر به مثلث متشکل از $\underline{p}$، $\underline{e}$ و $\underline{b}$ توجه کنیم می‌بینیم که بردار $\underline{b}$ به دو مؤلفه تقسیم شده است. یک مؤلفه در راستای $\underline{a}$ و مؤلفه دیگر متعامد بر آن (مکمل‌های متعامد). براساس روابط مثلثی داریم:

$$\underline{P} = \frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}} \Rightarrow \|P\| = \frac{\|a\|\,\|b\|\,|\cos\theta|}{\|a\|^{۲}} = \|b\|\;|\cos\theta|$$

$$\text{رابطه فیثاغورس} \Rightarrow \|e\| = \|b\|\;|\sin\theta|$$

ماتریس نگاشت:

$$\underline{P} = \underline{a}\hat{x} = \underline{a}\;\frac{a^Tb}{a^Ta} = \frac{aa^T}{a^Ta}\;\underline{b} \Rightarrow \underline{P} = \frac{aa^T}{a^Ta}$$

ماتریس نگاشت یک ماتریس $m \times m$ است.

رتبه ماتریس نگاشت یک است.

راستای $\underline{a}$ فضای ستونی ماتریس $\underline{P}$ است.

مثال: ماتریس نگاشت به راستای $\underline{a} = (۲, ۰, ۰)$ را به دست آورید.

$$\underline{P} = \frac{aa^T}{a^Ta} = \frac{\begin{bmatrix} ۲ \\ ۰ \\ ۰ \end{bmatrix}\begin{bmatrix} ۲ & ۰ & ۰ \end{bmatrix}}{۴} = \begin{bmatrix} ۱ & ۰ & ۰ \\ ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ \end{bmatrix}$$

- با توجه به تعریف نگاشت براساس تعامد انتظار داریم $\underline{P}^{۲} = \underline{P}$. با توجه به فرم ماتریس $\underline{P}$ می‌توان به سادگی نشان داد که این شرایط وجود دارد.
- با توجه به رابطه $\underline{P} = \frac{aa^T}{a^Ta}$ می‌توان به سادگی نشان داد که مجموع عناصر روی قطر اصلی $\underline{P}$ برابر واحد است.

- ماتریس $\underline{I}-\underline{P}$ نیز نگاشت به زیرفضای متعامد بر راستای $\underline{a}$ است. توجه کنید که $(\underline{I}-\underline{P})\underline{b}=\underline{e}$.

- $\underline{e}$ در زیرفضای متعامد به فضای ستون $\underline{P}$ است. بنابراین $\underline{e}$ در فضای پوچ چپ ماتریس $\underline{P}$ است. $\Leftarrow$ $\underline{P}\,\underline{e}=\underline{P}(\underline{b}-\underline{P})=P-\underline{P}=0$

- نگاشت به زیرفضا: فرض کنید می‌خواهیم به فضای ستونی ماتریس $\underline{A}$ نگاشت دهیم و ستون‌های ماتریس $\underline{A}$ نیز مستقل خطی هستند. بنابراین: $A=[\underline{a}_1,\dots,\underline{a}_n]$

به طور معادل می‌خواهیم برای بردار $\underline{b}$، ضرایب $\hat{x}_1,\dots,\hat{x}_n$ را به گونه‌ای بیابیم که بردار $\underline{p}=\hat{x}_1\underline{a}_1+\dots+\hat{x}_n\underline{a}_n$ نگاشت بردار $\underline{b}$ روی فضای تولیدشده توسط بردارهای $\underline{a}_1,\dots,\underline{a}_n$ گردد. به عبارت دیگر بردار خطای $\underline{e}=\underline{b}-\underline{p}$ بر بردار نگاشت $\underline{p}$ عمود باشد.

نکته کلیدی: بردار خطای $\underline{e}$ بر زیرفضای تولیدی توسط ستون‌های ماتریس $\underline{A}$ عمود است. به طور معادل بردار $\underline{e}$ بر ستون‌های ماتریس $\underline{A}$ عمود است.

$$\left.\begin{array}{l}\underline{a}_1^T\underline{e}=0\\ \vdots\\ \underline{a}_n^T\underline{e}=0\end{array}\right\}\xrightarrow[\text{ماتریسی}]{\text{فرم}}\left.\begin{array}{l}A^Te=\underline{0}\\ e=\underline{b}-\underline{p}=\underline{b}-\underline{A}\hat{\underline{x}}\end{array}\right\}\Rightarrow\underline{A}^T(\underline{b}-\underline{A}\hat{\underline{x}})=\underline{0}$$

$$\Rightarrow\underline{A}^T\underline{A}\hat{\underline{x}}=\underline{A}^T\underline{b}\longrightarrow$$ بردار $\hat{\underline{x}}$ از دستگاه مقابل به دست می‌آید.

- مهم!: ماتریس A^TA دارای ابعاد $n\times n$ بوده و چون ستون‌های $\underline{A}$ مستقل خطی هستند، این ماتریس معکوس پذیر است.

- فرم بسته بردار $\hat{\underline{x}}$ به صورت مقابل است:

و در نتیجه روابط مقابل را داریم:

$$\hat{\underline{x}}=(A^TA)^{-1}A^Tb\ :n\times1$$
$$\underline{P}=\underline{A}\hat{\underline{x}}=\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{b}\ :m\times1$$
$$\underline{P}=A(A^TA)^{-1}A^T\ :m\times m$$

- توجه: بردار $\underline{e}$ به ستون‌های ماتریس $\underline{A}$ عمود است. یعنی $A^T\underline{e} = 0$
عبارت تقابل به این معنی است که $\underline{e}$ در فضای پوچی چپ A یا فضای پوچی A^T قرار دارد.

- توجه: براساس روابط نگاشت به فضای ستونی یک ماتریس به سادگی می‌توان برای بردار دلخواه $\underline{P}$، دو مؤلفه $\underline{P} = \underline{b}_c$ (مؤلفه فضای ستونی) و $\underline{e} = \underline{b}_n$ (مؤلفه فضای پوچی چپ) را به دست آورد.

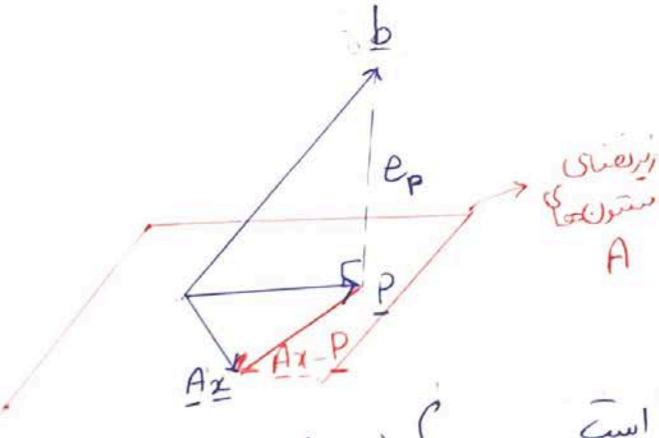
- توجه: ماتریس $\underline{P}$ به دست آمده رابطه $\underline{P}^2 = \underline{P}$ را برآورده می‌سازد.

تقریب حداقل مربعات

- حالتی را در نظر بگیرید که از نگاه سطری تعداد معادلات از تعداد متغیرها بیشتر است.
به طور معادل از نگاه ستونی در فضای $\mathbb{R}^m$، تنها n بردار داریم که کل فضا را نمی‌توانند پوشش دهند.

در شرایط فوق روش حذف گاوس - جردن به معادلات غیر ممکن از نوع $1 = 0$ می‌رسد. اگر $\underline{e} = \underline{b} - \underline{A}\underline{x}$ در نظر بگیریم، در این شرایط نمی‌توان $\underline{e} = 0$ قرار داد. در تقریب حداقل مربعات $\underline{x}$ به گونه‌ای تعیین می‌شود که $\| e \|^2$ حداقل گردد و جواب حاصل را $\hat{\underline{x}}$ می‌نامیم.

پاسخ حداقل مربعات:



$\underline{A}\,\underline{x} - \underline{P}$ در فضای ستون‌های $\underline{A}$ است.
e_P در فضای تکمیل متعامد ستون‌های $\underline{A}$ است.
$\Big\}$ $(\underline{A}\underline{x} - \underline{P})$ بر e_p عمود است.

$$\underline{A}\underline{x} - \underline{b} = (\underline{A}\underline{x} - \underline{P}) - \underbrace{(\underline{P} - \underline{b})}_{= e_P}$$

بنابراین : $\|\underline{A}\,\underline{x} - \underline{b}\|^2 = \|\underline{A}\,\underline{x} - p\|^2 + \|e_p\|^2 = \|e\|^2$

حداقل $\|e\|^2$ زمانی حاصل می‌شود که $\|Ax - p\|^2 = 0$ گردد، در این شرایط یعنی $\underline{x} = \underline{\hat{x}}$.

- توضیح : با استفاده از مشتق‌گیری از $\|e\|^2 = \|Ax - \underline{b}\|$ نیز می‌توان نشان داد که جواب تقریب حداقل مربعات برابر $\underline{x} = \underline{\hat{x}}$ است.

- جمع‌بندی : برای دستگاه معادلات $\underline{A}\,\underline{x} = \underline{b}$ که ستون‌های $\underline{A}$ مستقل خطی هستند و دستگاه معادلات جواب ندارد، پاسخ تقریب حداقل مربعات پاسخی است که $\|Ax - b\|^2$ را حداقل می‌نماید و برابر تصویر بردار $\underline{b}$ روی فضای ستونی ماتریس $\underline{A}$ است که به صورت زیر به دست می‌آید:

$$\underline{x}_{ls} = \underline{\hat{x}} = (A^TA)^{-1}A^Tb$$

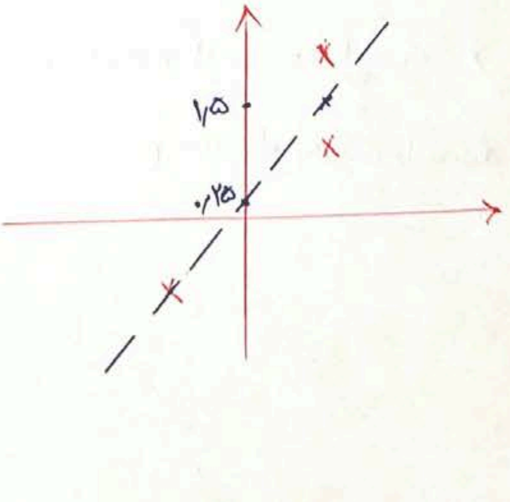
- مثال : برازش منحنی : نزدیکترین خط به نقاط (۱و۱)، (۱-و۱-) و (۲و۱) را به دست آورید.

همانطور که مشخص است خطی که بتواند از تمام نقاط عبور کند وجود ندارد. فرض کنیم معادله خط را به صورت $y = a + bx$ در نظر بگیریم. در این صورت a و b پاسخ دستگاه زیر خواهند بود:

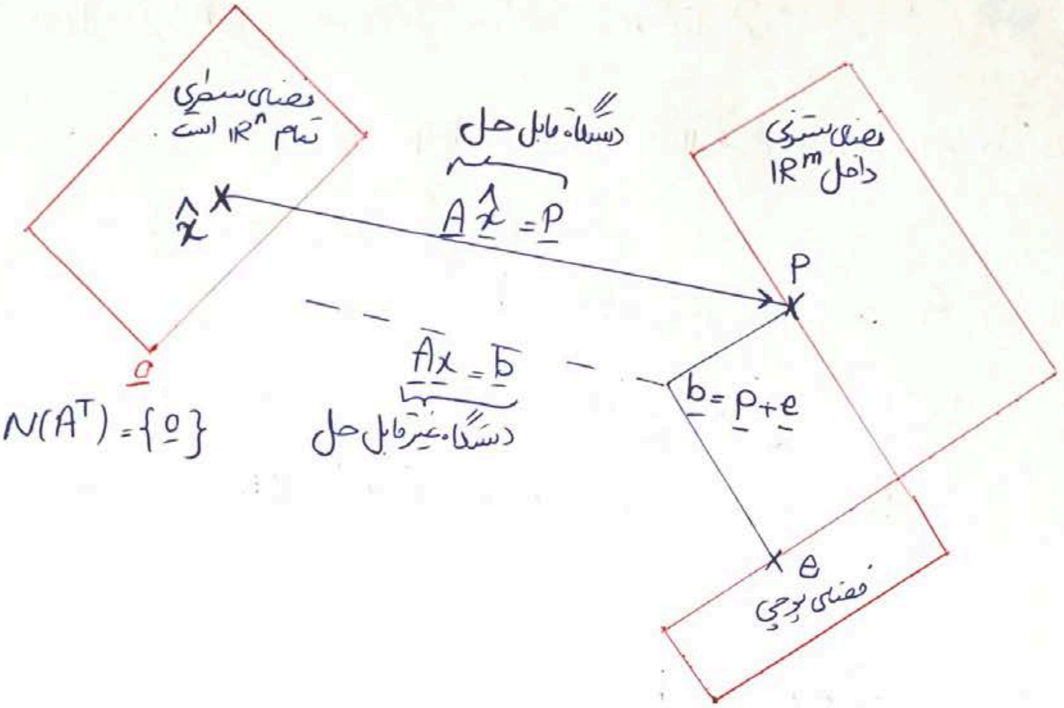
$$\underset{\underline{A}}{\begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & 1\end{bmatrix}}\ \underset{\underline{x}}{\begin{bmatrix}a\\ b\end{bmatrix}} = \underset{\underline{b}}{\begin{bmatrix}1\\ -1\\ 2\end{bmatrix}}$$

دستگاه مقابل جواب ندارد اما می‌توان جواب تقریب حداقل مربعات آن را محاسبه نمود.

$$x_{ls} = \hat{x} = (A^TA)^{-1}A^Tb = \begin{bmatrix}\text{۰/۲۵}\\ \text{۱/۲۵}\end{bmatrix}$$



- تصویر کلی دستگاه $\underline{A}.\underline{x} = \underline{b}$ برای حالتی که ستون‌های $\underline{A}$ مستقل خطی هستند



- ماتریس با ستون‌های وابسته:

- ابتدا ستون‌های وابسته را مشخص می‌کنیم. سپس ماتریس $\underline{A}_{ind}$ را که شامل تمام ستون‌های مستقل می‌گردد (ستون‌های پاشنه) تشکیل می‌دهیم.

- نگاشت بردار $\underline{b}$ بر روی فضای ستونی ماتریس $\underline{A}_{ind}$ را به دست می‌آوریم:

$$\underline{p} = A_{ind}(A_{ind}^T A_{ind})^{-1} A_{ind}^T\ b$$

- دستگاه $\underline{A}\underline{x} = \underline{p}$ بینهایت جواب . همگی $\underline{e}$ یکسان و $\|e\|^2$ برابری دارند.

- یکی از جواب‌های حداقل مربعات ویژگی حداقل طول را دارد. یعنی همزمان $\|e\|^2$ حداقل می‌گردد و $\|\underline{x}\|^2$ نیز در میان جواب‌های موجود حداقل می‌گردد. در ادامه خواهیم دید که این جواب با استفاده از شبه‌معکوس به دست می‌آید.

مثال: فرض کنید از شما خواسته شود یک خط به نقاط (۱,۱) و (۱,۳) برازش نمایید. جواب حداقل مربعات را به دست آورید.

پاسخ: اگر بخواهیم خط $y=a+bx$ را به داده‌ها برازش نماییم در اینصورت دستگاه زیر را خواهیم داشت:

$$\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}3\\1\end{bmatrix}$$

همانطورکه مشخص است ستون‌ها وابسته خطی هستند و یک ستون مستقل (۱,۱) خواهد بود. با استفاده از ستون‌های مستقل $\underline{p}$ را به صورت زیر به دست می‌آوریم:

$$\underline{A}_{ind}=\begin{bmatrix}1\\1\end{bmatrix}\Rightarrow \underline{p}=\underline{A}_{ind}\left(\underline{A}_{ind}^{T}\underline{A}_{ind}\right)^{-1}\underline{A}_{ind}^{T}\underline{b}$$

$$=\begin{bmatrix}1\\1\end{bmatrix}\left(\begin{bmatrix}1&1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}\right)^{-1}\begin{bmatrix}1&1\end{bmatrix}\begin{bmatrix}3\\1\end{bmatrix}=\begin{bmatrix}2\\2\end{bmatrix}$$

دستگاه $\underline{A}\underline{x}=\underline{p}$ بینهایت جواب دارد زیرا فضای پوچ آن Z نیست. تمامی این جواب‌ها حداقل مربعات خطا را دارند. جواب کامل این دستگاه به صورت زیر است:

$$\underline{A}\underline{x}_p=\underline{p}\Rightarrow\begin{bmatrix}1&1&2\\1&1&2\end{bmatrix}\begin{array}{l}\to \text{پاشنه}=1\\ \to \text{ضریب}=1\end{array}\longrightarrow\begin{bmatrix}1&1&2\\0&0&0\end{bmatrix}$$

$$\Rightarrow\begin{bmatrix}1&1\\0&0\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}2\\0\end{bmatrix}$$

$$\begin{array}{cc}\uparrow&\uparrow\\ \text{a پاشنه است.}&\text{b آزاد است.}\end{array}$$

$$\left.\begin{array}{l}b=1\Rightarrow a=1\Rightarrow \underline{x}_p=\begin{bmatrix}1\\1\end{bmatrix}\\ \underline{A}\underline{x}_n=\underline{0}\Rightarrow x_n=\alpha\begin{bmatrix}1\\-1\end{bmatrix}\end{array}\right\}\Rightarrow\underline{x}=\underline{x}_p+\underline{x}_n=\begin{bmatrix}1\\1\end{bmatrix}+\alpha\begin{bmatrix}1\\-1\end{bmatrix}$$

از بین جواب‌های فوق پاسخ $\begin{bmatrix}1\\1\end{bmatrix}$ یک ویژگی خاص دارد. این جواب حداقل طول را دارد و در فصل‌های آینده خواهیم دید که می‌توان این جواب را با استفاده از شبه‌معکوس به دست آورد.

پایه های متعامد و گرام-اشمیت

- علت استفاده از پایه های متعامد ← درشرایطی که پایه ها متعامد است A^TA قطری است و به سادگی می توان از رابطه $A\hat{x}=p$ ، $\hat{x}$ را به دست آورد.

- تعریف: بردارهای q_1 ، ... ، q_n متعامد یکه هستند اگر:

$$q_i^T q_j = \begin{cases} 0 & i\neq j \\ 1 & i=j \end{cases}$$

ماتریسی که ستون های آن بردارهای متعامد یکه است را با Q نشان می دهیم.

- توجه: ماتریس Q در رابطه $Q^TQ=I$ صادق است. درشرایطی که Q مربعی است درنتیجه این رابطه داریم:

$$Q^T = Q^{-1}$$

- توجه: در بردارهای متعامد یکه طول هر بردار برابر واحد است. اگر این طول واحد نباشد، حاصل Q^TQ یک ماتریس قطری خواهد بود. برای چنین حالتی به سادگی می توان با استفاده از یک ماتریس وزن ستون ها را به گونه ای نرمالیزه نمود که طول آنها واحد گردد.

- مثال:

$$Q=\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} \longrightarrow QQ^T=I \rightarrow$$ ماتریس چرخش با زاویه θ

Q ← چرخش با زاویه θ درجه

$Q^T=Q^{-1}$ ← چرخش با زاویه $-\theta$ درجه

$Q=P \longrightarrow$ هر ماتریس جایگشتی یک ماتریس متعامد یکه است.

- خاصیت: اگر Q ماتریسی با ستون های متعامد یکه باشد آنگاه، طول بردار Qx ، x برای هر بردار دلخواه x برابر خواهد بود.

اثبات:

$$\|Qx\|^2=(Qx)^TQx=x^TQ^TQx=x^TIx=x^Tx=\|x\|^2$$

- خاصیت: ماتریس Q با ستون‌های متعامد یکه مقدار ضرب داخلی را حفظ می‌کند، یعنی:

$$(\underline{Q}\underline{x})^T \underline{Q}\,\underline{y} = \underline{x}^T \underline{Q}^T \underline{Q}\,\underline{y} = \underline{x}^T \underline{y}$$

- نگاشت با استفاده از پایه‌های متعامد یکه:

$$\hat{x} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T b \xrightarrow{A \to Q} \hat{x} = (\underline{Q}^T\underline{Q})^{-1}\underline{Q}^T\underline{b} = \underline{Q}^T\underline{b}$$

$$\underline{P} = A\hat{x} = A(A^TA)^{-1}A^T b \xrightarrow{A \to Q} \underline{P} = \underline{Q}\,\underline{Q}^T\underline{b}$$

$$\underline{P} = A(A^TA)^{-1}A^T \xrightarrow{A \to Q} \underline{P} = \underline{Q}\,\underline{Q}^T$$

بنابراین اگر A یک ماتریس با ستون‌های متعامد یکه باشد نگاشت هر بردار دلخواه $\underline{b}$ به سادگی قابل انجام خواهد بود. فرض کنید $\underline{Q} = [\underline{q}_1, \dots, \underline{q}_n]$ باشد در اینصورت:

$$\underline{P} = \underline{Q}\,\underline{Q}^T\underline{b} = [\underline{q}_1, \dots, \underline{q}_n] \begin{bmatrix} -q_1^T- \\ \vdots \\ -q_n^T- \end{bmatrix} \underline{b}$$

$$= [q_1, \dots, q_n] \begin{bmatrix} q_1^T b \\ \vdots \\ q_n^T \underline{b} \end{bmatrix} = q_1(q_1^T b) + \dots + q_n(q_n^T b) \; *$$

حالت خاص ماتریس مربعی: اگر Q مربعی باشد در اینصورت:

۱) $Q^T = Q^{-1}$ ، $\hat{\underline{x}} = \underline{Q}^T\underline{b}$ همان $\hat{\underline{x}} = \underline{Q}^{-1}\underline{b}$ خواهد بود.

۲) زیرفضای تولید شده توسط ستون های Q کل فضای $\mathbb{R}^m$ است.

۳) در این حالت $\underline{P} = \underline{b}$ ، $\underline{P} = \underline{Q}\underline{Q}^T = \underline{I}$.

۴) در این حالت فرمول * $\underline{b}$ را به صورت مجموع نگاشت بر زیرفضاهای یک بعدی بازیابی می‌کند.

$$\underline{b} = q_1(\underline{q}_1^T\underline{b}) + \dots + \underline{q}_n(\underline{q}_n^T\underline{b})$$

قطعات فوق بر یکدیگر عمود هستند و این اساس بسیاری از تبدیل ها است.

مثال: ماتریس متعامد یکه مربعی $\underline{Q}$ به فرم مقابل را در نظر بگیرید:

$$Q = \frac{۱}{۳}\begin{bmatrix} -۱ & ۲ & ۲ \\ ۲ & -۱ & ۲ \\ ۲ & ۲ & -۱ \end{bmatrix}$$

فرض کنید می‌خواهیم نگاشت $\underline{b} = (۰,۰,۱)$ بر روی ستون‌های $\underline{q}_۱$، $q_۲$ و $q_۳$ را که به ترتیب $P_۱$، $P_۲$ و $P_۳$ می‌نامیم به دست آوریم. این نگاشت‌ها برابرند با:

$$\left.\begin{aligned} P_۱ &= q_۱(q_۱^T b) = ۲/۳\, q_۱ \\ P_۲ &= q_۲(q_۲^T b) = ۲/۳\, q_۲ \end{aligned}\right\} \Rightarrow P_۱ + P_۲ \rightarrow$$

نگاشت بردار $\underline{b}$ روی زیرفضای تولیدی توسط دو بردار $\underline{q}_۱$ و $q_۲$

$$P_۳ = q_۳(q_۳^T b) = -\frac{۱}{۳} q_۳$$

$$\underline{b} = P_۱ + P_۲ + P_۳ = ۲/۳\, \underline{q}_۱ + ۲/۳\, q_۲ - ۱/۳\, q_۳$$

فرآیند گرام اشمیت: در این فرآیند می‌خواهیم نحوه به دست آوردن پایه‌های متعامد یکه برای ستون‌های ماتریس $\underline{A}$ را به دست آوریم.

مثال: فرض کنید ستون‌های ماتریس $A = [a_۱ \;--\; a_n]$ مستقل خطی باشند. می‌خواهیم بردارهای ستونی جدید $\tilde{a}_۱$، ...، $\tilde{a}_n$ را به گونه ای به دست آوریم که متعامد باشند. روال گرام-اشمیت این کار را در گام‌های زیر انجام می‌دهد:

جهت اول هم‌راستا با ستون اول $\underline{A}$ در نظر گرفته می‌شود

گام اول: $\tilde{a}_۱ = a_۱ \longrightarrow$ در نظر گرفته می‌شود

گام دوم: $$\tilde{a}_۲ = a_۲ - \underbrace{\frac{\tilde{a}_۱^T a_۲}{\tilde{a}_۱^T \tilde{a}_۱}\tilde{a}_۱}_{\text{نگاشت } a_۲ \text{ در راستای } \tilde{a}_۱} \rightarrow$$

بردار متقابل در واقع خطای است هنگامی که بخواهیم بردار $a_۲$ را با نگاشت آن روی $\tilde{a}_۱$ نشان دهیم.

گام سوم: $$\tilde{a}_۳ = a_۳ - \underbrace{\left(\frac{\tilde{a}_۱^T a_۳}{\tilde{a}_۱^T \tilde{a}_۱}\tilde{a}_۱ + \frac{\tilde{a}_۲^T a_۳}{\tilde{a}_۲^T \tilde{a}_۲}\tilde{a}_۲\right)}$$

نگاشت بردار $a_۳$ روی صفحه حاصل از $\tilde{a}_۱$، $\tilde{a}_۲$ (می‌دانیم این نگاشت مجموع نگاشت روی هر خط در راستای $\tilde{a}_۱$ و $\tilde{a}_۲$ است.)

گام n-۲ :

$$\tilde{a}_n = a_n - \underbrace{\left(\sum_{i=1}^{n-1} \frac{\tilde{a}_i^T a_n}{\tilde{a}_i^T \tilde{a}_i} \tilde{a}_i \right)}$$

نگاشت روی ابرصفحه حاصل از بردارهای $\tilde{a}_1$ و ... و $\tilde{a}_{n-1}$

می دانیم این نگاشت برابر مجموع نگاشت یک بعدی روی راستاهای $\tilde{a}_1$ و ... و $\tilde{a}_n$ است.

توجه: با توجه به اینکه $\tilde{a}_i$، برابر خطا نگاشت a_i روی زیرفضای پوشش داده شده بردارهای $\tilde{a}_1$ و ... و $\tilde{a}_{i-1}$ است، بنابراین $\tilde{a}_i$ بر تمامی بردارهای این زیرفضا از جمله $\tilde{a}_1$ و ... و $\tilde{a}_{i-1}$ متعامدند.

- نگاه کردن به روال گرام-اشمیت به فرم عامل بندی:

شبه کد زیر روال گرام اشمیت اصلاح شده را نشان می دهد. در این کد q_i ها همان $\tilde{a}_i$ ها هستند که یکه شده اند.

```
for  j = 1:n
    $v = a_j$
    for  i = 1 : j-1
        $R(i,j) = q_i^T a_j$
        $v = v - R(i,j) q_i$
    end
    $R(j,j) = \|v\|$
    $q_j = \frac{v}{R(j,j)}$
end
```

همانطور که در شبه کد فوق مشخص است رابطه زیر برقرار است:

$$R(j,j) q_j = a_j - \sum_{i=1}^{j-1} R(i,j) q_i$$

از رابطه بالا می توان ستون های ماتریس A را به فرم زیر بازیابی نمود:

$$A = QR$$

که R ماتریسی است که طی فرآیند گرام اشمیت اصلاح شده، عناصر آن به دست آمده اند. همانطور که در شبه کد مشخص است این ماتریس بالامثلثی است و فرم آن به صورت زیر است:

$$R = \begin{bmatrix} q_1^T a_1 & q_1^T a_2 & \cdots & q_1^T a_n \\ 0 & q_2^T a_2 & \cdots & q_2^T a_n \\ 0 & 0 & - & q_n^T a_n \end{bmatrix}$$

برای بردارهای a_1 ، ... ، a_n ، گرام-اشمیت بردارهای متعامد یکه q_1 ، ... ، q_n را می‌سازد. ماتریس‌هایی که این ها را در ستون‌های خود دارند رابطه $A = QR$ را برآورده می‌سازند. به عبارت دیگر $R = Q^T A$ خواهد بود. این ماتریس بالامثلثی است زیرا q_i بر تمامی ستون‌های ماتریس A بعد از اندیس i عمود است.

محاسبه نگاشت: می‌دانیم نگاشت معادل حل دستگاه حداقل مربعات زیر است:

$$\left.\begin{array}{l} \underline{A}^T \underline{A}\, \hat{\underline{x}} = \underline{A}^T \underline{b} \\ A^T A = (QR)^T QR = R^T R \end{array}\right\} \Rightarrow R^T R \hat{x} = R^T Q^T b \Rightarrow R \hat{x} = Q^T b$$

حل دستگاه حاصل به سادگی با روش جایگذاری انجام می‌شود.

دترمینان

- **تعریف:** دترمینان یک عدد است که برای هر ماتریس مربعی قابل تعریف است و اطلاعات مختلفی از آن ماتریس ارائه می‌نماید. در این فصل ابتدا با برخی از ویژگی‌های دترمینان آشنا می‌شویم و سپس آن را به صورت رسمی تعریف خواهیم کرد. دترمینان ماتریس A را با $\det A$ یا $|A|$ نمایش می‌دهیم.

- **برخی ویژگی‌های مهم دترمینان:**

۱- با استفاده از دترمینان ماتریس A می‌توان A^{-1} و $A^{-1}b$ را محاسبه نمود.

۲- اگر سطرهای ماتریس A، موقعیت لبه‌های یک جعبه در فضای n بعدی باشند در این صورت حجم جعبه برابر $|\det A|$ است.

۳- برای ماتریس $A: n\times n$، اعداد خاص $\{\lambda_i\}_{i=1}^{n}$ وجود دارند به گونه‌ای که دترمینان $A-\lambda_i I$ برابر صفر می‌گردد.

- **حالت خاص:** برای ماتریس دو بعدی $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$، دترمینان برابر $ad-bc$ است، ماتریس معکوس به صورت زیر محاسبه می‌شود:

$$A^{-1}=\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

* اگر $ad-bc=0$ باشد در اینصورت $a/c=b/d$ خواهد بود و سطرها موازی هستند.

* اگر از روش حذف برای ماتریس $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ استفاده کنیم به ماتریس زیر می‌رسیم.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{matrix}\rightarrow \text{پاشنه}: a \\ \rightarrow \text{ضریب}: c/a\end{matrix}\longrightarrow\begin{bmatrix} \overset{P_1}{\boxed{a}} & b \\ 0 & \overset{P_2}{\boxed{d-\frac{c}{a}b}} \end{bmatrix}$$

$$\det A=\prod_{i=1}^{n} P_i \rightarrow$$ دترمینان برابر حاصل ضرب پاشنه‌ها است.

* اگر جای سطرهای ماتریس را عوض کنیم حاصلضرب پاشنه‌ها برابر خواهد بود با:

$$\begin{bmatrix} c & d \\ a & b \end{bmatrix}\begin{matrix}\rightarrow \text{پاشنه}: c \\ \rightarrow \text{ضریب}: a/c\end{matrix}\longrightarrow\begin{bmatrix} c & d \\ 0 & b-\frac{a}{c}d \end{bmatrix}\rightarrow \prod P_i = bc-ad=-\det A$$

جابه‌جایی دو سطر از ماتریس علامت دترمینان را عوض می‌کند. این قاعده به طور معادل با جابه‌جایی دو ستون نیز برقرار است.

- ویژگی های دترمینان :

۱- دترمینان ماتریس همانی $n \times n$ برابر یک است.

$$\begin{vmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{vmatrix} = 1$$

۲- با جابجایی هر دو سطر ماتریس، دترمینان قرینه می‌شود.

* با استفاده از این خاصیت می‌توان دترمینان هر ماتریس جایگشت را محاسبه نمود:

$\det P = \begin{cases} +1 \rightarrow \\ -1 \rightarrow \end{cases}$ تعداد جایگزینی‌های مورد نیاز برای رسیدن به P از ماتریس همانی زوج باشد.

〃 〃 〃 〃 فرد باشد.

۳- دترمینان یک تابع خطی از سطر به صورت جداگانه است.

خاصیت همگنی :

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ ta_{i1} & \cdots & ta_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = t \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

خاصیت جمع پذیری :

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a'_{i1} & \cdots & a'_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a''_{i1} & \cdots & a''_{in} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ (a'_{i1}+a''_{i1}) & \cdots & (a'_{in}+a''_{in}) \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

* براساس خاصیت ۳، $\det tA = t^n \det A$

۴- اگر دو سطر A برابر باشند در اینصورت $\det A = 0$ می‌گردد.

علت : با جابجایی دو سطر برابر، دترمینان عوض نمی‌شود. براساس خاصیت ۲ باید دترمینان قرینه شود. تنها عددی که با قرینه شدن عوض نمی‌شود صفر است.

* اگر ماتریسی دو سطر برابر داشته باشد، معکوس پذیر نیست.

۵- کم کردن مضربی از یک سطر از سطر دیگر ماتریس دترمینان را عوض نمی‌کند.

علت:

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & & \vdots \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \det A$$

l برابر سطر i-ام را از سطر j-ام کم کنیم.

$$\begin{vmatrix} a_{11} & & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & & a_{in} \\ a_{j1} - l a_{i1} & & a_{jn} - l a_{in} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{vmatrix}$$

$$\overset{\text{ویژگی ۳}}{=} \underbrace{\begin{vmatrix} a_{11} & & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & & a_{in} \\ \vdots & & \vdots \\ a_{j1} & & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{vmatrix}}_{= \det A} + l \underbrace{\begin{vmatrix} a_{11} & - & a_{1n} \\ a_{i1} & - & a_{in} \\ \vdots & & \vdots \\ a_{i1} & - & a_{in} \\ \vdots & & \vdots \\ a_{n1} & - & a_{nn} \end{vmatrix}}_{\overset{\text{ویژگی ۴}}{=} ۰}$$

$= \det A$

نتیجه: طی عملیات حذف برای رسیدن از A به U، دترمینان عوض نمی‌شود و دترمینان U برابر دترمینان A است. اگر برای حذف نیاز به تغییر سطر داشته باشیم در اینصورت به ازای هر تغییر سطر، علامت دترمینان قرینه می‌شود. بنابراین در حالت کلی

$\det A = \pm \det U$.

۶- اگر A یک سطر تماماً صفر داشته باشد دراینصورت $\det A = 0$ است.

علت: از ماتریس A ماتریس B را می‌سازیم که در آن یکی از سطرهای غیرصفر A به سطر صفر آن جمع شده است. دراینصورت:

$$\left.\begin{array}{l} \text{ویژگی ۵} \Rightarrow \det A = \det B \\ \text{ویژگی ۴} \Rightarrow \det B = 0 \end{array}\right\} \Rightarrow \det A = 0$$

۷- اگر A ماتریسی مثلثی باشد دراینصورت: $\det A = \prod_{i=1}^{n} a_{ii}$

علت: فرض کنید تمام عناصر روی قطر اصلی غیرصفر هستند. دراینصورت با عملیات حذف و بدون تغییر دترمینان به ماتریس قطری به صورت زیر رسید.

$$D = \begin{bmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots \\ & & & a_{nn} \end{bmatrix}$$

که $\det A = \det D$.

حال با استفاده از ویژگی ۳، در سطر اول ماتریس D از a_{11} و ... از سطر n-ام ماتریس D از a_{nn} فاکتورگیری کنیم. دراینصورت:

$$\left.\begin{array}{l} \det D = \prod_{i=1}^{n} a_{ii} \det I \\ \det I = 1 \quad (\text{ویژگی ۱}) \end{array}\right\} \Rightarrow \det A = \det D = \prod_{i=1}^{n} a_{ii}$$

اگر حداقل یکی از عناصر روی قطر اصلی صفر باشد، با عملیات حذف می‌توان یک سطر صفر در ماتریس تولید کرد که دراینصورت دترمینان برابر صفر خواهد بود. از طرف دیگر می‌دانیم عملیات حذف دترمینان را عوض نمی‌کند. بنابراین اگر حداقل یکی از عناصر قطر اصلی صفر باشد $\det A = 0$ است.

۸ - اگر $\underline{A}$ تکین باشد درانصورت $\det A = ۰$ است. اگر $\underline{A}$ معکوس پذیر باشد در انصورت $\det A \neq ۰$ است.

علت: با استفاده از فرآیند حذف گاوس می‌توانیم از ماتریس $\underline{A}$ به ماتریس $\underline{U}$ برسیم و در ماتریس $\underline{U}$ حتماً یک سطر صفر خواهد بود (در غیرانصورت $\underline{U}$ پاشنه کامل بود. و معکوس پذیر خواهد بود و درنتیجه $\underline{A}$ نیز معکوس پذیر است). چون در ماتریس U یک سطر تماما صفر است در نتیجه $\det U = ۰$ است. از طرفی می‌دانیم $\det A = \pm \det U$ است و درنتیجه $\det A = ۰$.

اگر $\underline{A}$ معکوس پذیر باشد $\underline{U}$ پاشنه کامل است و دترمینان آن از ویژگی ۷ به صورت زیر قابل محاسبه است.

$\det U = \prod_{i=1}^{n} u_{ii} =$ حاصلضرب پاشنه‌ها

چون پاشنه‌ها غیرصفر هستند درنتیجه $\det U \neq ۰$ است.

از طرف دیگر $\det A = \pm \det U$. بنابراین $\det A \neq ۰$ است.

۹ - دترمینان ضرب دو ماتریس $\underline{A}$ و $\underline{B}$ برابر با حاصلضرب دترمینان آن دو است.
یعنی: $|AB| = |A||B|$

علت: فرض کنید A معکوس پذیر است. درانصورت می‌توان با استفاده از عملیات حذف گاوس جردن این ماتریس را به صورت حاصلضرب زیر نمایش داد:

$$A = E_۱ E_۲ \dots E_k I = E_۱ \dots E_k$$

که ماتریس‌ها E_i یکی از انواع زیر هستند:

۱- ماتریس جابگشت.

۲- ماتریس جمع کردن مضربی از یک سطر با سطر دیگر $\begin{bmatrix} ۱ & & ۰ \\ x & \ddots & \\ & & ۱ \end{bmatrix}$ یا $\begin{bmatrix} ۱ & & x \\ ۰ & \ddots & \\ & & ۱ \end{bmatrix}$

۳- ماتریس ضرب یک سطر در یک عدد ناصفر کلی: $\begin{bmatrix} ۱ & & \\ & x & ۰ \\ ۰ & & \ddots \; ۱ \end{bmatrix}$

حال برای هر یک از حالت‌های فوق برقرار ویژگی را بررسی می‌کنیم:

۱- اگر A ماتریس جابگشت باشد:

$$\left.\begin{array}{l} |AB| \overset{\text{ویژگی ۲}}{=\!=\!=\!=} -|B| \\ |A| \overset{\text{ویژگی ۱ و ۲}}{=\!=\!=\!=} -۱ \end{array}\right\} \Rightarrow |AB| = |A||B|$$

۲- اگر A مضربی از یک سطر را با سطر دیگر جمع کند:

$$\left.\begin{array}{l} |A| \xlongequal{\text{ویژگی ۷}} 1 \\ |AB| \xlongequal{\text{ویژگی ۵}} |B| \end{array}\right\} \Rightarrow |AB| = |A|\,|B|$$

۳- اگر A یک سطر را در یک عدد ضرب کند و به فرض کلی تبدیل باشد:

$$A = \begin{bmatrix} 1 & & 0 \\ & \ddots\; x \;\ddots & \\ 0 & & 1 \end{bmatrix}$$

$$\left.\begin{array}{l} |A| \xlongequal{\text{ویژگی ۷}} x \\ |AB| \xlongequal{\text{ویژگی ۳}} x\,|B| \end{array}\right\} \Rightarrow |AB| = |A|\,|B|$$

بنابراین رابطه مورد نظر برای تمام حالاتی که A یکی از سه نوع گفته شده باشد برقرار است. حال حالت کلی را در نظر بگیرید:

$$\begin{aligned} |AB| &= |E_1 E_2 \ldots E_{k-1} E_k B| \\ &= |E_1|\,|E_2 \ldots E_k B| \\ &\vdots \\ &= |E_1| \ldots |E_k|\,|B| \\ &= |E_1| \ldots |E_{k-1} E_k|\,|B| \\ &\vdots \\ &= |E_1 \ldots E_k|\,|B| \\ &= |A|\,|B| \end{aligned}$$

حال فرض کنید A تکین است. در این صورت سمت راست تساوی $|AB| = |A|\,|B|$ برابر صفر می‌شود. می‌توان نشان داد اگر A تکین باشد AB نیز تکین خواهد بود. برای این منظور دو حالت زیر را در نظر بگیرید:

الف) B تکین است ← در این شرایط بردار غیرصفر $\underline{x}$ وجود دارد به گونه‌ای که $B\underline{x} = \underline{0}$ است. در نتیجه طبق رابطه مقابل، AB نیز تکین است.

$$A(Bx) = (AB)\,x = 0$$

ب) $\underline{B}$ غیرتکین است ← در این شرایط چون A تکین است، می‌توان بردار را به گونه‌ای یافت که $\underline{A}\underline{y} = \underline{0}$ باشد. حال دستگاه $\underline{B}\underline{x} = \underline{y}$ به علت غیرتکین بودن $\underline{B}$ جواب منحصر به فرد $\underline{B}^{-1}\underline{y}$ را دارد. این جواب را می‌نامیم. در نتیجه:

$$ABz = \underline{A}\,\underline{y} = \underline{0}$$

رابطه فوق نشان می‌دهد در این شرایط نیز AB تکین است.

بنابراین اگر A تکین باشد، دو طرف تساوی $|AB| = |A||B|$ صفر خواهد بود و تساوی برقرار است.

$$\det A = \det A^T \qquad -10$$

حل: فرض کنید $\underline{A}$ تکین است. ابتدا نشان می‌دهیم در این شرایط A^T نیز تکین خواهد بود.

بر اساس برهان خلف فرض می‌کنیم A^T معکوس‌پذیر باشد یعنی ماتریسی مانند $\underline{B}$ وجود دارد به گونه‌ای که $BA^T = I$. در نتیجه:

$$BA^T = I \longrightarrow (BA^T)^T = I^T = I \Rightarrow AB^T = I$$

رابطه فوق نشان می‌دهد که B^T معکوس A است که با فرض تکین بودن A در تناقض است. بنابراین A^T تکین خواهد بود.

بنابراین در شرایطی که $\underline{A}$ تکین است، طرفین تساوی هر دو صفر شده و رابطه برقرار است.

حال فرض کنید A غیرتکین است. در این صورت با استفاده از روش حذفی می‌توانیم به عامل‌بندی $PA = LU$ برسیم. در نتیجه:

$$|P||A| = |L||U|$$

اگر از طرفین عامل‌بندی ترانهاده بگیریم به رابطه $A^TP^T = U^TL^T$ می‌رسیم و در نتیجه:

$$|P^T||A^T| = |L^T||U^T|$$

حال دو رابطه مقابل را در نظر بگیرید:

$$\begin{cases} |P|\ |A| = |L|\ |U| \\ |P^T|\ |A^T| = |L^T|\,|U^T| \end{cases}$$

روابط زیر را داریم:

۱- $|U| = |U^T|$ (براساس ویژگی ۷)

۲- $|L| = |L^T|$ (براساس ویژگی ۷)

۳- $|P| = |P^T|$

$$\left.\begin{array}{l} PP^T = I \Rightarrow |P|\ |P^T| = 1 \Rightarrow |P^T| = \dfrac{1}{|P|} \\ |P| = \pm 1 \end{array}\right\} \Rightarrow |P^T| = |P|$$

بنابراین لازم است $|A| = |A^T|$.

توجه: ویژگی ۱۰ سبب می‌شود بتوان تمام ویژگی‌هایی که در مورد سطرهای ماتریس بیان نمودیم به ستون‌های آن نیز قابل اعمال باشد.

← اگر دو ستون جابه‌جا شوند دترمینان قرینه می‌شود.

← ستون صفر یا ستون‌های برابر منجر به دترمینان صفر می‌شود.

← دترمینان تابع خطی از هر ستون به طور مجزا است.

روش‌های محاسبه دترمینان

۱- محاسبه با استفاده از پاشنه‌ها:

همانطور که دیدیم با استفاده از جابگشت‌های مورد نیاز می‌توان هر ماتریس مربعی A را با استفاده از روش حذف به صورت مقابل عامل‌بندی نمود: $PA = LU$

که در رابطه فوق L یک ماتریس پایین مثلثی بود. و عناصر روی قطر اصلی آن یک است.

U یک ماتریس بالا مثلثی است.

P یک ماتریس جابگشت است.

بنابراین روابط زیر را در ارتباط با دترمینان ها داریم:

$$|P|\;|A| = |L|\;|U|$$

با استفاده از ویژگی های ماتریس می توان ساده سازی های زیر را انجام داد:

$$|P| = \begin{cases} +1 \rightarrow & \text{اگر تعداد جابه جایی سطرها فرد باشد} \\ -1 \rightarrow & \text{اگر تعداد جابه جایی سطرها زوج باشد} \end{cases}$$

$$|L| = \prod_{i=1}^{n} l_{ii} = \prod_{i=1}^{n} 1 = 1$$

$$|U| = \prod_{i=1}^{n} u_{ii}$$

بنابراین رابطه اول برای به دست آوردن دترمینان ماتریس A بصورت زیر است:

$$|A| = \pm \prod_{i=1}^{n} u_{ii}$$

$$A = \left[\begin{array}{ccc|cc} \multicolumn{3}{c}{A_k} & & \\ \hline a_{11} & \text{---} & a_{1k} & \text{---} & a_{n} \\ | & & | & & \\ a_{k1} & \text{---} & a_{kk} & & a_{kn} \\ \hline | & & & & \\ a_{kn} & & a_{nk} & & a_{nn} \end{array}\right]$$

- توجه: ماتریس A و A_k را به صورت مقابل در نظر بگیرید.

اگر در فرآیند حذف در ماتریس A، جابجایی نیاز نباشد، k پاشنه اول تنها به ماتریس A_k وابسته است. دترمینان ماتریس A_k نیز برابر است با:

$$|A_k| = \prod_{i=1}^{k} u_{ii}$$

براین اساس می توانیم فرمولی برای محاسبه پاشنه ها از روی دترمینان ماتریس های A_k به صورت زیر ارائه نماییم:

$$u_{kk} = \frac{u_{11}\cdots u_{kk}}{u_{11}\cdots u_{(k-1)(k-1)}} = \frac{|A_k|}{|A_{k-1}|}$$

۲ـ محاسبه با استفاده از فرمول بزرگ:

هنگام استفاده از پاشنه‌ها برای محاسبه دترمینان، رابطه صریح میان عناصر ماتریس و دترمینان به دست نمی‌آید. در این بخش به دنبال این رابطه صریح هستیم.

فرمول بزرگ در واقع از رابطه خطی بودن دترمینان نسبت به یک سطر در هسته مرکزی خود استفاده می‌کند.

حالت خاص ماتریس ۲x۲ زیر را در نظر بگیرید. با استفاده از خاصیت خطی بودن داریم:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix}$$ * شکستن سطر اول:

$$= \underset{①}{\begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix}} + \underset{②}{\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix}}$$ * شکستن سطر دوم روی عبارت اول

$$+ \underset{③}{\begin{vmatrix} a & b \\ c & 0 \end{vmatrix}} + \underset{④}{\begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix}}$$ * شکستن سطر دوم روی عبارت دوم

با استفاده از روش فوق در حالت کلی برای یک ماتریس $A: n \times n$ به n^n دترمینان می‌رسیم که در هر سطر آن تنها یک مؤلفه غیرصفر حضور دارد.

اگر در هر یک از سطرها، شماره ستونی که مؤلفه غیرصفر در آن قرار دارد یکسان باشد، با استفاده از ویژگی ۵ می‌توان به یک سطر صفر در دترمینان رسید و در نتیجه دترمینان برابر صفر خواهد بود.

بنابراین تعداد دترمینان‌های غیرصفر از تعداد حالت‌های ممکن برای هر سطر به دست می‌آید.

$$\left.\begin{array}{ll} \text{تعداد حالت‌های ممکن برای سطر اول} & = n \\ \text{دوم} \quad '' \quad '' \quad '' & = n-1 \\ \vdots & \\ \text{n-ام} \quad '' \quad '' \quad '' & = 1 \end{array}\right\} \Rightarrow \text{کل تعداد دترمینان‌های غیرصفر} = n!$$

بنابراین از n^n جمله تنها $n!$ جمله می‌تواند غیرصفر باشد. به عنوان نمونه در مثال دترمینان ۲x۲ روابط زیر را داریم:

$$\begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} = 0$$

$$\begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} = 0$$

برای دترمینان‌هایی که می‌توانند غیرصفر باشند نیز با استفاده از ویژگی خطی بودن دترمینان محاسبه می‌شود. به عنوان نمونه برای ماتریس ۲×۲ داریم:

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} = ad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = ad$$

(براساس ویژگی ۱، $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$ است)

$$\begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = bc \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -bc$$

(براساس ویژگی ۱و۲، $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$ است)

- ماتریس ۳×۳ : در این شرایط ۶ = !۳ دترمینان خواهیم داشت که می‌توانند غیرصفر باشند

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \underbrace{\begin{vmatrix} a_{11} & & \\ & a_{12} & \\ & & a_{13} \end{vmatrix}}_{\text{متناظر با ترتیب ستونی } (1,2,3)} + \underbrace{\begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix}}_{\text{متناظر با ترتیب ستونی } (2,3,1)}$$

$$+ \underbrace{\begin{vmatrix} & & a_{13} \\ a_{21} & & \\ & a_{32} & \end{vmatrix}}_{(3,1,2)} + \underbrace{\begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix}}_{(1,3,2)}$$

$$+ \underbrace{\begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix}}_{(2,1,3)} + \underbrace{\begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}}_{(3,2,1)}$$

$$= a_{11}a_{12}a_{13} \begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31} \begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32} \begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix}$$

$$+ a_{11}a_{23}a_{32} \begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33} \begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31} \begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}$$

$$= a_{11}a_{12}a_{13}\overbrace{\underline{P}(1,2,3)}^{=1} + a_{12}a_{23}a_{31}\overbrace{\underline{P}(2,3,1)}^{=1} + a_{13}a_{21}a_{32}\overbrace{\underline{P}(3,1,2)}^{=1}$$

$$+ a_{11}a_{23}a_{32}\underbrace{\underline{P}(1,3,2)}_{=-1} + a_{12}a_{21}a_{33}\underbrace{\underline{P}(2,1,3)}_{=-1} + a_{13}a_{21}a_{32}\underbrace{\underline{P}(3,2,1)}_{=-1}$$

- درحالت کلی برای ماتریس $A: n \times n$ ، رابطه فرمول بزرگ برای محاسبه دترمینان به صورت زیر خواهد بود:

$$\det A = \sum_{\underline{P}} (\det \underline{P})\, a_{1P_1}\, a_{۲P_۲}\, a_{۳P_۳} \cdots a_{nP_n}$$

که

$$\underline{P} = (P_1, P_۲, \dots, P_n)$$

- مثال: اگر A یک ماتریس مثلثی باشد آنگاه تنها جایگشت که در مجموع فوق غیرصفر است به صورت $\underline{P} = (۱, ۲, \dots, n)$ خواهد بود. در نتیجه برای این نوع ماتریس رابطه زیر را داریم:

$$\det A = (+1)\, a_{۱۱}\, a_{۲۲} \cdots a_{nn}$$

که منطبق بر ویژگی ۷ است.

- مثال: با استفاده از فرمول بزرگ دترمینان مقابل را محاسبه نمایید.

$$\begin{vmatrix} 1 & 0 & a & 0 \\ 0 & 1 & b & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & d & 1 \end{vmatrix} = c \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = c$$

با استفاده از پاشنه‌ها:

حالت اول: $c = 0$ ← در دترمینان یک سطر صفر است ← دترمینان صفر است.

حالت دوم: $c \neq 0$ ← با روش حذف و بدون تغییر دترمینان می‌توان به یک دترمینان قطری رسید ← دترمینان برابر c می‌شود.

۳- محاسبه با استفاده از هم‌عامل‌ها (Cofactors):

در قسمت قبل دیدیم که با استفاده از خاصیت خطی بودن دترمینان نسبت به یک سطر مجزا می‌توان به جداسازی زیر دست پیدا کرد:

$$\begin{vmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{vmatrix} = \begin{vmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{vmatrix} + \begin{vmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{vmatrix} + \begin{vmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{vmatrix}$$

در هر سه دترمینان سمت راست مساوی فوق، مقدار المان‌های قرمزرنگ هیچ تأثیری در محاسبه دترمینان ندارد.

در دترمینان اول، چون عنصر $a_{۱۱}$ از سطر و ستون اول انتخاب شده است، امکان انتخاب مجدد این سطر و ستون برای رسیدن به یک دترمینان غیرصفر وجود ندارد. بنابراین جملاتی در دترمینان A که $a_{۱۱}$ در آنها حضور دارد تنها می‌توانند تابع $a_{۲۲}$، $a_{۲۳}$، $a_{۳۲}$ و $a_{۳۳}$ باشند. این موضوع برای عنصرهای $a_{۱۲}$ و $a_{۱۳}$ نیز برقرار است. رابطه به دست آمده از فرمول بزرگ برای دترمینان ماتریس ۳x۳ این موضوع را تأیید می‌نماید.

$$\det A = a_{۱۱}\underbrace{(a_{۲۲}a_{۳۳}-a_{۲۳}a_{۳۲})}_{=\begin{vmatrix} a_{۲۲} & a_{۲۳} \\ a_{۳۲} & a_{۳۳} \end{vmatrix}} + a_{۱۲}\underbrace{(a_{۲۳}a_{۳۱}-a_{۲۱}a_{۳۳})}_{-\begin{vmatrix} a_{۲۱} & a_{۲۳} \\ a_{۳۱} & a_{۳۳} \end{vmatrix}} + a_{۱۳}\underbrace{(a_{۲۱}a_{۳۲}-a_{۲۲}a_{۳۱})}_{\begin{vmatrix} a_{۲۱} & a_{۲۲} \\ a_{۳۱} & a_{۳۲} \end{vmatrix}}$$

بنابراین دترمینان ماتریس ۳x۳ : A به سه دترمینان با ابعاد ۲x۲ وابسته است. این رابطه در حالت کلی نیز برای ماتریس $A: n\times n$ به صورت زیر برقرار است:

$$\det A = a_{۱۱}C_{۱۱} + a_{۱۲}C_{۱۲} + \dots + a_{۱n}C_{۱n}$$

$$C_{۱j} = (-1)^{1+j} \det M_{۱j}$$

که در رابطه فوق $M_{۱j}$ کهاد ۱j-ام ماتریس A است و از حذف سطر اول و ستون j-ام ماتریس A به دست می‌آید. $C_{۱j}$ نیز هم‌عامل ۱j-ام ماتریس A است و از $M_{۱j}$ به دست می‌آید.

- در حالت کلی برای ماتریس $A: n\times n$، دترمینان برابر حاصلضرب سطر i-ام ماتریس A با هم عامل‌های آن سطر است یعنی:

$$\det = a_{i۱}C_{i۱} + a_{i۲}C_{i۲} + \dots + a_{in}C_{in}$$

که هم عامل C_{ij} از رابطه زیر به دست می‌آید:

$$C_{ij} = (-1)^{i+j} \det M_{ij}$$

که M_{ij} کهاد ij-ام ماتریس A بوده و از حذف سطر i-ام و ستون j-ام ماتریس A به دست می‌آید.

توجه: برای نشان دادن این موضوع که $\det A = \sum_{i=1}^{n} a_{ij} c_{ij}$ است می‌توان نشان داد که این تعریف ۳ ویژگی اول دترمینان را داراست.

توجه: با استفاده از ویژگی جابه‌جایی سطرها و تأثیر آن در دترمینان می‌توان نشان داد که رابطه $\det A = \sum_{i=1}^{n} a_{ij} c_{ij}$ برای هر سطر دلخواه i نیز به صورت زیر برقرار است:

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}$$

توجه: روش محاسبه دترمینان براساس هم‌عامل‌ها، دترمینان یک ماتریس $A: n \times n$ را به صورت بازگشتی به ترتیب براساس دترمینان‌های $(n-1)\times(n-1)$، ...، 1×1 محاسبه می‌نماید.

توجه: با استفاده از ویژگی ۱۰ به سهولت می‌توان دترمینان را برحسب بسط هم‌عامل‌ها روی یک ستون به صورت زیر نوشت:

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij}$$

مثال

$$\underbrace{\begin{vmatrix} ۲ & -۱ & ۰ & ۰ \\ -۱ & ۲ & -۱ & ۰ \\ ۰ & -۱ & ۲ & -۱ \\ ۰ & ۰ & -۱ & ۲ \end{vmatrix}}_{D_۴} = (۲)(-۱)^۲ \underbrace{\begin{vmatrix} ۲ & -۱ & ۰ \\ -۱ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{vmatrix}}_{D_۳} + (-۱)(-۱)^۳ \underbrace{\begin{vmatrix} -۱ & -۱ & ۰ \\ ۰ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{vmatrix}}$$

$$= -۱ \begin{vmatrix} ۲ & -۱ \\ -۱ & ۲ \end{vmatrix} = -D_۲$$

که در رابطه فوق D_n دترمینان ماتریسی است که روی قطر اصلی آن ۲ و روی عناصر بالا و پایین قطر اصلی ۱- قرار دارد و با نام ماتریس ۱-،۲،۱- شناخته می‌شود.

$$D_۳ = \begin{vmatrix} ۲ & -۱ & ۰ \\ -۱ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{vmatrix} = (۲)(-۱)^۲ \begin{vmatrix} ۲ & -۱ \\ -۱ & ۲ \end{vmatrix} + (-۱)(-۱)^۳ \begin{vmatrix} -۱ & -۱ \\ ۰ & ۲ \end{vmatrix}$$

$$= ۲(۳) - ۲ = ۳$$

$$D_۲ = ۳$$

بنابراین درحالت کلی داریم :

$$D_۴ = ۲D_۳ - D_۲ = ۲(۴) - ۳ = ۵$$

توجه : به طور کلی می‌توان نشان داد برای یک ماتریس D_n روابط زیر را داریم :

$$D_n = n + ۱$$

$$D_n = ۲D_{n-۱} - D_{n-۲}$$

قانون کرامر

- دستگاه معادلات $\underline{A}\,\underline{x} = \underline{b}$ را درنظر بگیرید. به طور معادل می‌توانیم رابطه زیر را داشته باشیم :

$$\begin{bmatrix} a_{۱۱} & a_{۱۲} & a_{۱۳} \\ a_{۲۱} & a_{۲۲} & a_{۲۳} \\ a_{۳۱} & a_{۳۲} & a_{۳۳} \end{bmatrix} \begin{bmatrix} x_۱ & ۰ & ۰ \\ x_۲ & ۱ & ۰ \\ x_۳ & ۰ & ۱ \end{bmatrix} = \begin{bmatrix} b_۱ & a_{۱۲} & a_{۱۳} \\ b_۲ & a_{۲۲} & a_{۲۳} \\ b_۳ & a_{۳۲} & a_{۳۳} \end{bmatrix} = B_۱$$

با استفاده از تعبیر ستونی می‌توان به سادگی معادل بودن دو رابطه را نشان داد. اگر از طرفین رابطه فوق دترمینان بگیریم می‌توانیم رابطه زیر را داشته باشیم :

$$(\det A)\, x_۱ = \det B_۱ \Rightarrow x = \frac{\det B_۱}{\det A}$$

که $B_۱$ ماتریس A است درحالیکه ستون اول آن با بردار $\underline{b}$ جایگزین شده است.

- قانون کرامر: اگر $\det A \neq ۰$ باشد درانصورت می‌توان دستگاه $\underline{A}\underline{x} = \underline{b}$ را به صورت زیر حل کرد:

$$x_i = \frac{\det B_i}{\det A} \quad , \quad i = ۱, ..., n$$

که در رابطه فوق B_i ماتریس A است که i-امین ستون آن برابر $\underline{b}$ قرار داده شده است.

- توجه: برای حل یک دستگاه با n مجهول با استفاده از روش کرامر لازم است $n+1$ دترمینان حساب شود که اگر هر دترمینان با استفاده از فرمول بزرگ طی $n!$ عبارت حساب شود $(n+1)!$ عبارت در محاسبات حضور خواهند داشت.

- محاسبه ماتریس معکوس: می‌دانیم ستون j-ام ماتریس معکوس از دستگاه معادلات زیر قابل محاسبه است:

$$\underline{A}\ \underline{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_j$$

که e_j برداری است که تمام مؤلفه‌های آن صفر بوده و فقط مؤلفه j-ام آن برابر یک.

برای به دست آوردن مؤلفه i-ام در بردار $\underline{x}$ فوق می‌توان به ترتیب زیر عمل کرد:

$$A\left[e_1\ e_2 \cdots e_{i-1}\ \overset{\text{ستون } i\text{-ام}}{\overset{\downarrow}{x}}\ e_{i+1} \cdots e_n\right] = \left[a_1 \cdots a_{i-1}\ \overset{\text{ستون } i\text{-ام}}{\overset{\downarrow}{e_j}}\ a_{i+1} \cdots a_n\right]$$
$$= B_i$$

حال با دترمینان‌گیری از طرفین رابطه زیر را داریم:

$$(\det A)\ x_i = \det B_i$$

اما در این شرایط فرم B_i به صورت زیر است:

$$B_i = \begin{bmatrix} a_{11} & \cdots & a_{1(i-1)} & 0 & a_{1(i+1)} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ & & & 0 & & & \\ & & & 1 & & & \\ & & & 0 & & & \\ & & & \vdots & & & \\ a_{n1} & & a_{n(i-1)} & 0 & a_{n(i+1)} & & a_{nn} \end{bmatrix}$$

که با بسط همه عامل‌ها روی ستون i-ام به سادگی می‌توان $\det B_i$ را به صورت زیر محاسبه کرد:

$$\det B_i = C_{ji}$$

بنابراین به طور کلی می‌توانیم روابط زیر را داشته باشیم:

$$(A^{-1})_{ij} = \frac{C_{ji}}{\det A} \Rightarrow \underline{A}^{-1} = \frac{\underline{C}^{T}}{\det \underline{A}}$$

که در رابطه فوق $\underline{C}$ ماتریس هم عامل‌ها است.

توجه: رابطه ماتریس معکوس به سادگی از بسط هم عامل‌های برای دترمینان در سطرهای مختلف به صورت زیر به دست می‌آید (برای سادگی روابط در یک ماتریس ۳x۳ نوشته شده است)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix}$$

سوال: چرا در رابطه ماتریسی فوق در ماتریس سمت راست تساوی عناصر غیرقطر اصلی صفر هستند.

پاسخ: عنصر ij $(i \neq j)$ $= \sum a_{ik} C_{jk}$ -

کمیت فوق معادل دترمینان برای ماتریسی است که سطر i-ام آن در سطر j-ام آن کپی شده است و می‌دانیم که این دترمینان برابر صفر است.

ضرب خارجی

- ضرب خارجی دو بردار $\underline{u} = (u_1, u_2, u_3)$ و $\underline{v} = (v_1, v_2, v_3)$ برداری است که از رابطه زیر به دست می‌آید:

$$\underline{u} \times \underline{v} = \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = (u_2 v_3 - u_3 v_2)\underline{i} + (u_3 v_1 - u_1 v_3)\underline{j} + (u_1 v_2 - u_2 v_1)k$$

این بردار بر هر دو بردار $\underline{u}$, $\underline{v}$ عمود است.

- توجه: نگارش ضرب خارجی براساس دترمینان تنها روشی است که برای به خاطرسپاری ضرب خارجی استفاده می‌شود. در این دترمینان سطر اول از جنس بردار و دو سطر دیگر از جنس عدد هستند.

- توجه: $\underline{u} \times \underline{v} = -\underline{v} \times \underline{u}$ زیرا جابه‌جایی دو سطر در دترمینان علامت آن را قرینه می‌کند.

- توجه: $\underline{u} \times \underline{u} = 0$ زیرا وجود دو سطر برابر دترمینان را صفر می‌کند.

- **تعریف**: ضرب خارجی دو بردار $\underline{u}$ و $\underline{v}$ برداری با طول $\|u\| \, \|v\| \, |\sin\theta|$ است. جهت این بردار بر $\underline{u}$ و $\underline{v}$ عمود بود. و از قاعده دست راست به دست می آید.

مقدارها و بردارهای ویژه

تعریف: λ را مقدار ویژه و $\underline{x}$ را بردار ویژه ماتریس $\underline{A}$ می‌نامیم اگر در رابطه زیر صادق باشند:

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

بنابراین بردار ویژه ماتریس $\underline{A}$ برداری است که وقتی ماتریس $\underline{A}$ روی آن عمل می‌کند، جهت آن را تغییر نمی‌دهد.

توجه: اگر $\lambda = ۰$ باشد در اینصورت بردار $\underline{x}$ در فضای پوچی ماتریس $\underline{A}$ است.

توجه: طبق تعریف برای $\underline{A} = \underline{I}$، همه بردارهای $\underline{x}$ بردار ویژه بوده و $\lambda = ۱$ مقدار ویژه متناظر با آن است.

نحوه محاسبه: برای محاسبه بردارها و مقدارهای ویژه باید گام‌های زیر را طی کنیم:

۱- دترمینان $\underline{A} - \lambda \underline{I}$ را محاسبه نمایید (از آنجایی که λ روی قطر اصلی است، این دترمینان یک چندجمله‌ای از λ با درجه n خواهد بود).

۲- ریشه‌های چندجمله‌ای حاصل از گام ۱ را محاسبه نمایید. این ریشه‌ها $\underline{A} - \lambda \underline{I}$ را تکین می‌کند.

۳- برای هرکدام از مقدارهای ویژه λ، دستگاه $(\underline{A} - \lambda \underline{I})\underline{x} = \underline{0}$ را حل کنید تا بردار ویژه متناظر را به دست آورید.

مثال: بردارها و مقدارهای ویژه ماتریس $A = \begin{bmatrix} ۰/۸ & ۰/۳ \\ ۰/۲ & ۰/۷ \end{bmatrix}$ را به دست آورید.

$$1 \rightarrow \det(\underline{A} - \lambda \underline{I}) = \det \begin{bmatrix} ۰/۸ - \lambda & ۰/۳ \\ ۰/۲ & ۰/۷ - \lambda \end{bmatrix} = \lambda^۲ - \frac{۳}{۲}\lambda + ۱/۲ = (\lambda - 1)(\lambda - ۱/۲)$$

$$۲ \rightarrow \det(\underline{A} - \lambda \underline{I}) = ۰ \Rightarrow \begin{cases} \lambda_1 = 1 \\ \lambda_۲ = ۱/۲ \end{cases}$$

$$۳ \rightarrow \underline{A} - \lambda_1 \underline{I} = \begin{bmatrix} -۰/۲ & ۰/۳ \\ ۰/۲ & -۰/۳ \end{bmatrix} \longrightarrow \text{پایه فضای پوچی} = \begin{bmatrix} ۱/۵ \\ ۱ \end{bmatrix} \rightarrow$$

این پایه بردار ویژه متناظر با $\lambda_1 = 1$ است.

$$\underline{A} - \lambda_۲ \underline{I} = \begin{bmatrix} ۰/۳ & ۰/۳ \\ ۰/۲ & ۰/۲ \end{bmatrix} \longrightarrow \text{پایه فضای پوچی} = \begin{bmatrix} ۱ \\ -۱ \end{bmatrix} \rightarrow$$

این پایه بردار ویژه متناظر با $\lambda = ۰/۵$ است.

- توجه: بردارها و مقدارهای ویژه ماتریس‌های $A^{۲}$، $A^{۳}$ و ... برحسب بردارها و مقدارهای ویژه ماتریس A، به صورت زیر به دست می‌آید:

$$\underline{A}\,\underline{x} = \lambda\,\underline{x} \xrightarrow{\times A} A^{۲}\,\underline{x} = \lambda\,(\underline{A}\underline{x}) = \lambda\,\lambda\,\underline{x} = \lambda^{۲}x$$

$\Rightarrow \underline{A}^{۲}\,\underline{x} = \lambda^{۲}\underline{x} \longrightarrow$ اگر x و λ به ترتیب بردار و مقدار ویژه ماتریس A باشند، x و $\lambda^{۲}$ به ترتیب بردار و مقدار ویژه ماتریس $A^{۲}$ هستند.

⋮

$A^{n}\,\underline{x} = \lambda^{n}\,\underline{x} \longrightarrow$ x و λ^{n} به ترتیب بردار و مقدار ویژه ماتریس A^{n} هستند.

- مثال: برای ماتریس $\underline{A} = \begin{bmatrix} ۰/۸ & ۰/۳ \\ ۰/۲ & ۰/۷ \end{bmatrix}$، ماتریس $A^{۱۰۰}$ را به دست آورید.

می‌دانیم بردارها و مقدارهای ویژه این ماتریس به صورت مقابل است.

$\lambda = ۱ \longrightarrow \underline{x}_{۱} = \begin{bmatrix} ۱/۵ \\ ۱ \end{bmatrix}$

$\lambda = ۰/۵ \longrightarrow x_{۲} = \begin{bmatrix} ۱ \\ -۱ \end{bmatrix}$

دو بردار فوق مستقل خطی هستند در نتیجه کل فضای $\mathbb{R}^{۲}$ را پوشش می‌دهند. پس می‌توان ستون‌های ماتریس A را برحسب این دو بردار نوشت. برای ستون اول داریم:

$a_{۱} = ۰/۴\,x_{۱} + ۰/۲\,x_{۲}$

$A^{۲}$ ستون اول $= A\,a_{۱} = A(۰/۴x_{۱} + ۰/۲x_{۲}) = ۰/۴(\lambda_{۱})x_{۱} + ۰/۲(\lambda_{۲})x_{۲}$

$A^{۱۰۰}$ ستون اول $= A^{۹۹}a_{۱} = A^{۹۹}(۰/۴x_{۱} + ۰/۲x_{۲}) = ۰/۴\,(۱)^{۹۹}x_{۱} + ۰/۲(۰/۵)^{۹۹}x_{۲}$

$$\simeq ۰/۴\,x_{۱} = \begin{bmatrix} ۰/۶ \\ ۰/۴ \end{bmatrix}$$

$a_{۲} = ۰/۴\,x_{۱} - ۰/۳\,x_{۲}$

$A^{۱۰۰}$ ستون دوم $= A^{۹۹}a_{۲} = A^{۹۹}(۰/۴x_{۱} - ۰/۳x_{۲}) = ۰/۴(۱)^{۹۹}x_{۱} - ۰/۳(۰/۵)^{۹۹}x_{۲} \simeq \begin{bmatrix} ۰/۶ \\ ۰/۴ \end{bmatrix}$

- **تعریف:** بردار ویژه $\underline{x}$ متناظر با مقدار ویژه $\lambda = 1$ را که با افزایش توان ماتریس از بین نمی‌رود (λ^n به سمت صفر میل نمی‌کند) که حالت دائمی است.

بردار ویژه $\underline{x}$ متناظر با مقدار ویژه $|\lambda| < 1$ که با افزایش توان ماتریس مقدار ویژه آن به صفر میل می‌کند یک مورد کاهشی است.

- **ماتریس نگاشت:** اگر $\underline{P}$ یک ماتریس نگاشت باشد آنگاه مقدارهای ویژه آن ۰ و ۱ خواهد بود.

علت: P ماتریس نگاشت است $\Rightarrow P^2 = P$ *

$$P\underline{v} = \lambda \underline{v} \xrightarrow{\times P} P^2 \underline{v} = \lambda P v \Rightarrow P\underline{v} = \lambda^2 \underline{v}$$

$$\Rightarrow \lambda \underline{v} = \lambda^2 \underline{v} \Rightarrow \lambda = \lambda^2 \Longrightarrow \begin{cases} \lambda = 0 \\ \lambda = 1 \end{cases}$$

- در یک ماتریس مثلثی، مقادیر ویژه عناصر روی قطر اصلی هستند.

علت: در ماتریس مثلثی $\underline{T}$، $\underline{T} - T_{ii}\underline{I}$ برابر ماتریس مثلثی خواهد بود که i-امین عنصر روی قطر اصلی آن صفر است و در نتیجه دترمینان آن صفر خواهد بود. پس عناصر روی قطر اصلی ماتریس مثلثی مقادیر ویژه آن هستند.

- **توجه:** عملیات حذف مقدارهای ویژه را عوض می‌کند. بنابراین مقدارهای ویژه ماتریس $\underline{A}$ و ماتریس $\underline{U}$ حاصل از عملیات حذف گاوس متفاوت خواهند بود.

- **تعریف:** اثر ماتریس (trace) : اثر یک ماتریس مربعی برابر مجموع عناصر روی قطر اصلی آن ماتریس است و با trace A نمایش داده می‌شود، یعنی:

$$\text{trace } A = \sum_{i=1}^{n} a_{ii}$$

- **توجه:** در یک ماتریس مربعی $A_{n\times n}$، حاصلضرب n مقدار ویژه برابر دترمینان ماتریس A، حاصل جمع n مقدار ویژه برابر اثر ماتریس A می‌باشد.

- **مقدار ویژه موهومی** : با توجه به تعریف مقدارهای ویژه، می توانند موهومی باشند. به عنوان مثال ماتریس چرخش به اندازه $\theta = 90°$ به صورت $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ را در نظر بگیرید.

می توان نشان داد که روابط مقابل برقرار است :

$$\left.\begin{array}{l} \lambda_1 + \lambda_2 = \text{trace } Q = 0 \\ \lambda_1\lambda_2 = \det Q = 1 \end{array}\right\} \Rightarrow \lambda_1 = -\lambda_2 = j$$

← براساس ماتریس Q می دانیم بردار $\underline{x}$ ، $Q\underline{x}$ $90°$ چرخش نسبت بردار $\underline{x}$ دارد. اما به طور شهودی دانیم اگر $\underline{x}$ حقیقی باشد نمی تواند با بردار چرخش یافته خود به میزان $90°$ هم راستا گردد. بنابراین بردارها و مقدارها ویژه ماتریس Q نمی توانند حقیقی باشند.

← **بردارهای ویژه متناظر:**

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 + jy_1 \\ x_2 + jy_2 \end{bmatrix} = \lambda_1 \begin{bmatrix} x_1 + jy_1 \\ x_2 + jy_2 \end{bmatrix} = \begin{bmatrix} -y_1 + jx_1 \\ -y_2 + jx_2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 0 & -1 & +1 & 0 \\ +1 & 0 & 0 & +1 \\ +1 & 0 & +1 & 0 \\ +1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \underline{x}_1 = \begin{bmatrix} j \\ 1 \end{bmatrix}$$

$$\lambda = -j \Rightarrow \underline{x}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$$

- **ماتریس متعامد یکه Q** : برای ماتریس متعامد یکه Q ، اندازه تمام مقدارهای ویژه برابر با واحد است.

علت :

$$\left.\begin{array}{l} \text{ماتریس Q اندازه بردار را حفظ می کند} \Rightarrow \|Qx\|^2 = \|x\|^2 \\ \text{تعریف بردار و مقدار ویژه} \Rightarrow \|Qx\|^2 = |\lambda|^2 \|x\|^2 \end{array}\right\} \Rightarrow |\lambda|^2 = 1 \Rightarrow |\lambda| = 1$$

- **ماتریس هرمیتی** : برای ماتریس هرمیتی A ($A^H = A$) تمام مقادیر ویژه حقیقی هستند.

$$\text{تعریف بردار و مقدار ویژه} \Rightarrow A\underline{x} = \lambda \underline{x} \xrightarrow{H} x^H A^H = \bar{\lambda} \underline{x}^H \Rightarrow x^H A = \bar{\lambda} x^H$$

$$\xrightarrow[\text{غیر صفر } \underline{x}]{\text{ضرب از راست در بردار}} x^H A x = \bar{\lambda} \underline{x}^H \underline{x}$$

$$\Rightarrow \underline{x}^H \lambda \underline{x} = \bar{\lambda} x^H \underline{x} \Rightarrow \lambda \|x\|^2 = \bar{\lambda} \|x\|^2$$

$$\Rightarrow \lambda = \bar{\lambda}$$

- ماتریس پادهرمیتی: برای ماتریس پاد هرمیتی A ($A^H=-A$)، تمام مقادیر ویژه موهومی خالص هستند.

- توجه: با توجه به اینکه بردارهای ویژه ماتریس دلخواه A و B لزوماً یکسان نیست، به طور عمومی نمی‌توان اظهار نظری در مورد بردارها و مقدارهای ویژه AB و A+B کرد.

- خاصیت: $AB = BA$ است اگر و فقط اگر A، B، n بردار ویژه مستقل یکسان داشته باشند.

قطری سازی

- قطری سازی: فرض کنید $A: n\times n$، n بردار ویژه مستقل خطی $x_1, \dots, x_n$ داشته باشد. اگر این بردارها را به عنوان ستون‌های ماتریس $\underline{X}$ در نظر بگیریم، رابطه بردار-مقدار ویژه به صورت زیر درخواهد آمد:

$$A \underset{\underset{X}{\uparrow}}{\left[\underline{x}_1, \dots, x_n\right]} = \underset{\underset{X}{\uparrow}}{\left[\underline{x}_1, \dots, \underline{x}_n\right]} \underset{\underset{\Lambda}{\uparrow}}{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}$$

$$\Rightarrow X^{-1} A X = \Lambda$$

بنابراین حاصلضرب $X^{-1}AX$ یک ماتریس قطری است که عناصر روی قطر اصلی، مقادیر ویژه به ترتیبی هستند که بردارهای ویژه متناظر آنها در ماتریس $\underline{X}$ ظاهر شده است. طی این حاصلضرب ماتریس A قطری می‌شود.

مثال: ماتریس $A=\begin{bmatrix}1 & 5\\ 0 & 4\end{bmatrix}$ را قطری کنید.

$$A=\begin{bmatrix}1 & 5\\ 0 & 4\end{bmatrix} \longrightarrow \begin{cases}\lambda_1=1\ ,\ x_1=\begin{bmatrix}1\\ 0\end{bmatrix}\\ \lambda_2=4\ ,\ x_2=\begin{bmatrix}1\\ 1\end{bmatrix}\end{cases}$$

$$X^{-1}AX=\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 5\\ 0 & 4\end{bmatrix}\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 4\end{bmatrix}$$

توجه: از رابطه $X^{-1}AX=\Lambda$ نتیجه می‌گیریم $A=X\Lambda X^{-1}$. به استفاده از این رابطه می‌توانیم A^2 را به صورت زیر محاسبه نماییم.

$$A^2=X\Lambda X^{-1}X\Lambda X^{-1}=X\Lambda^2X^{-1}$$

رابطه فوق مورد انتظار است زیرا می‌دانیم بردارهای ویژه ماتریس A و A^2 یکسان است و مقدارهای ویژه ماتریس A^2 مربع مقدارهای ویژه ماتریس A است.

به طور تعمیم یافته می‌توانیم نشان دهیم:

$$A^n=X\Lambda^nX^{-1}$$

قضیه: اگر $\underline{x}_1$ و $\underline{x}_2$ دو بردار ویژه متناظر با دو مقدار ویژه متمایز λ_1 و λ_2 ماتریس A باشند، آنگاه $\underline{x}_1$ و $\underline{x}_2$ مستقل خطی هستند.

اثبات: برای اثبات قضیه باید نشان دهیم اگر $\alpha_1\underline{\ }_1+\alpha_2\underline{x}_2=\underline{0}$ باشد آنگاه $\alpha_1=\alpha_2=0$ است.

$$\alpha_1\underline{x}_1+\alpha_2\underline{x}_2=\underline{0}\Longrightarrow\begin{cases}\text{ضرب طرفین در }\lambda_1:\ \alpha_1\lambda_1x_1+\alpha_2\lambda_1x_2=\underline{0}\\ \text{ضرب از چپ در }A:\ \alpha_1\lambda_1x_1+\alpha_2\lambda_2x_2=0\end{cases}$$

$$\overline{\alpha_2(\lambda_1-\lambda_2)x_2=0\Rightarrow\alpha_2=0}$$

به طور مشابه می‌توان نشان داد $\alpha_1=0$ است و در نتیجه $\underline{x}_1$ و $\underline{x}_2$ مستقل خطی هستند.

- **ماتریس معکوس:** اگر در رابطه $A^n = X\Lambda^n X^{-1}$، $n = -1$ قرار دهیم به رابطه زیر می‌رسیم:

$$A^{-1} = X\Lambda^{-1}X^{-1}$$

رابطه فوق در صورتی برقرار است که تمامی مقدارهای ویژه غیر صفر باشند.

- **قطری شدنی:** ماتریس $A: n \times n$ قطری شدنی است اگر و فقط اگر n بردار ویژه مستقل خطی داشته باشد.

- **شرح:** اگر ریشه‌های معادله مشخصه $\det(A - \lambda I) = ۰$ منحصر به فرد باشند، ماتریس A قطری شدنی است.

- **مثال:** آیا ماتریس $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$ و $B = \begin{bmatrix} ۰ & 1 \\ ۰ & ۰ \end{bmatrix}$ قطری شدنی هستند.

A: معادله مشخصه: $\det\left(\begin{bmatrix} 1-\lambda & -1 \\ 1 & -1-\lambda \end{bmatrix}\right) = \lambda^{۲} = ۰ \Rightarrow \lambda_1 = \lambda_{۲} = ۰$

نقش بردار ویژه: $A - \lambda_1 I = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \rightarrow$ بردارهای ویژه، بردارهای پایه فضای پوچی ماتریس مقابل هستند که در اینجا بعد این فضا 1 است.

B: معادله مشخصه: $\det\left(\begin{bmatrix} -\lambda & 1 \\ ۰ & -\lambda \end{bmatrix}\right) = \lambda^{۲} = ۰ \Rightarrow \lambda_1 = \lambda_{۲} = ۰$

نقش بردار ویژه: $A - \lambda_1 I = \begin{bmatrix} ۰ & 1 \\ ۰ & ۰ \end{bmatrix} \rightarrow$ بعد فضای پوچی 1 است و در نتیجه به ازای دو مقدار ویژه، تک بردار ویژه حضور دارد.

دو مثال فوق شرایطی را نشان می‌دهد که مقدارهای ویژه تکراری در ماتریس حضور دارند اما بعد فضای پوچی ماتریس برای آن مقدار ویژه از تکرار مقدار ویژه کمتر است. این دسته از ماتریس‌ها با نام ماتریس‌های ناقص شناخته می‌شوند.

- **قطری نشدنی:** اگر ماتریس $A: n \times n$ ناقص باشد در اینصورت n بردار ویژه مستقل خطی نداشته و قطری نشدنی است.

- **توجه:** ارتباطی میان معکوس پذیری و قطری شدنی بودن وجود ندارد.

ماتریس A معکوس پذیر است اگر تمام مقادیر ویژه آن غیر صفر باشند.

ماتریس A قطری شدنی است اگر n بردار ویژه مستقل داشته باشد.

تعریف: ماتریس‌های مشابه: ماتریس ثابت C را در نظر بگیرید که لزوماً قطری نیست. با استفاده از ماتریس دلخواه معکوس پذیر B می‌توان یک خانواده از ماتریس ها به صورت $A = BCB^{-1}$ تشکیل داد. در این صورت ماتریس‌های A و C را مشابه می‌نامیم. می‌توان نشان داد مقادیر ویژه ماتریس های خانواده A مقادیر ویژه C است.

ماتریس های متقارن

همانطورکه در بخش قبل دیدیم مقدارهای ویژه ماتریس متقارن همگی حقیقی هستند. این ماتریس‌ها از منظر بردارها و مقدارهای ویژه شرایط منحصر به فردی دارند که این شرایط معمولاً در قضیه‌ای با عنوان قضیه طیفی بیان می‌شود.

قضیه طیفی: هر ماتریس حقیقی متقارن $\underline{S}$ عامل بندی به صورت $\underline{S} = Q\Lambda Q^{T}$ دارد که Λ ماتریس قطری است که مقدارهای ویژه حقیقی روی قطر آن قرار دارند و Q که ماتریس تشکیل شده توسط بردارهای ویژه است یک ماتریس متعامد یکه است. چون Q متعامد یکه است، عامل بندی فوق به صورت فرم عامل بندی قطری سازی زیر نیز قابل بیان است:

$$\underline{S} = Q\Lambda Q^{-1}$$

قضیه: بردارهای ویژه متناظر با مقدارهای ویژه متمایز ماتریس متقارن S، همواره متعامد هستند.

اثبات:

$$\text{فرض:}\begin{cases} S = S^{T} \\ S\underline{x}_1 = \lambda_1 \underline{x}_1 \\ S\underline{x}_2 = \lambda_2 \underline{x}_2 \end{cases}$$

$$\left.(\lambda_1 \underline{x}_1)^{T}\underline{x}_2 = (\underline{S}\underline{x}_1)^{T}\underline{x}_2 = x_1^{T} S \underline{x}_2 = \underline{x}_1^{T}\underline{\lambda}_2 \underline{x}_2 \right\}$$

$$\Rightarrow x_1^{T} x_2 = 0$$

- **مثال:** نشان دهید برای ماتریس متقارن $\underline{S}=\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ مقادیر ویژه حقیقی هستند.

$$\det\left(S-\lambda I\right)=۰ \Rightarrow \det\left(\begin{bmatrix} a-\lambda & b \\ b & c-\lambda \end{bmatrix}\right)=۰$$

$$\Rightarrow (a-\lambda)(c-\lambda)-b^{۲}=۰$$

$$\Delta=(a-c)^{۲}+۴b^{۲} \geqslant ۰ \Rightarrow$$ ریشه‌ها حقیقی هستند.

- **توجه:** برای ماتریس متقارن S، رابطه قطری‌سازی را می‌توان به صورت زیر نیز تعبیر نمود:

$$S=Q\Lambda Q^T=\left[\underline{q}_1,\dots,\underline{q}_n\right]\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$=\left[\lambda_1 q_1 \;\cdots\; \lambda_n q_n\right]\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$=\lambda_1 q_1 q_1^T+\dots+\lambda_n q_n q_n^T$$

با استفاده از رابطه فوق می‌توان به سهولت رابطه بردار و مقدار ویژه را نشان داد، یعنی:

$$\left.\begin{array}{l} \underline{S}\underline{q}_i=\sum_{j=1}^{n}\lambda_j q_j q_j^T \\ q_j^T q_i=\begin{cases} 0 & i\neq j \\ 1 & i=j \end{cases} \end{array}\right\} \Rightarrow S q_i=\lambda_i q_i \underbrace{q_i^T q_i}_{=1}=\lambda_i q_i$$

- **نکته:** اگر برای ماتریس مربعی $\underline{A}$، عدد مختلط λ و بردار مختلط $\underline{x}$ به ترتیب مقدار و بردار ویژه باشند در اینصورت $\bar{\lambda}$ و $\bar{x}$ نیز به ترتیب بردار و مقدار ویژه ماتریس A خواهند بود.

اثبات:

$$\left.\begin{array}{l} Ax=\lambda x \Rightarrow \bar{A}\bar{x}=\bar{\lambda}\bar{x} \\ \bar{A}=A : \text{ماتریس A حقیقی است} \end{array}\right\} \Rightarrow A\bar{x}=\bar{\lambda}\bar{x}$$

ماتریس های مثبت معین

- تعریف: ماتریس متقارن S را مثبت معین گوییم اگر تمام مقادیر ویژه آن مثبت باشند.

- کاربرد: در مسائل بهینه سازی، مثبت معین بودن کاربرد فراوانی دارد. اگر بخواهیم نشان دهیم که پاسخ حداقل مربعات با استفاده از مشتق گیری به چه صورت است می توانیم از تعریف مثبت معین بودن استفاده نماییم.

- تعریف: ماتریس متقارن S مثبت معین است اگر برای هر بردار غیرصفر $\underline{x}$ رابطه $\underline{x}^T\underline{S}\,\underline{x}>0$ برقرار باشد.

اثبات:

$$S=Q\Lambda Q^T \Rightarrow x^TSx = x^TQ\Lambda Q^Tx$$

$$\left.\begin{array}{l} =(Q^Tx)^T\Lambda Q^Tx \\ y \triangleq Q^Tx \end{array}\right\} \Rightarrow x^TSx = y^T\Lambda y = \sum_{i=1}^{n}\lambda_i y_i^2$$

با توجه به اینکه بردار $\underline{y}$ غیرصفر است مجموع فوق مثبت است اگر تمام بردارهای ویژه مثبت باشند.

- تست مثبت معین بودن براساس پاشنه ها: مقادیر ویژه ماتریس متقارن S مثبت هستند اگر و فقط اگر پاشنه ها مثبت باشند.

- تست براساس عامل بندی: اگر ستون های ماتریس $\underline{A}$ مستقل خطی باشند آنگاه $\underline{S}=\underline{A}^T\underline{A}$ مثبت معین است.

علت:

$$\left.\begin{array}{l} x^TSx = x^TA^TAx = (Ax)^TAx = \|Ax\|^2 \\ \text{ستون های } A \text{ مستقل خطی} \Rightarrow \underline{A}\,\underline{x} \neq \underline{0} \end{array}\right\} \Rightarrow x^TSx > 0$$

گزاره های معادل: برای ماتریس متقارن S ، برقراری یکی از گزاره های زیر معادل با برقراری همگی آنها است.

۱- تمام پاشنه‌های S مثبت است.

۲- تمام دترمینان‌های زیر ماتریس های S_k مثبت است.

۳- تمام مقدارهای ویژه ماتریس S مثبت است.

۴- برای تمام بردارهای غیرصفر $\underline{x}$ ، $x^TSx > ۰$ است.

۵- $S = A^TA$ که A یک ماتریس با ستون‌های مستقل خطی است.

مثال: برای ماتریس $S = \begin{bmatrix} ۲ & -۱ & ۰ \\ -۱ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{bmatrix}$ ، گزاره های فوق را تست کنید.

۱-

$$\begin{bmatrix} ۲ & -۱ & ۰ \\ -۱ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{bmatrix} \begin{array}{l} \rightarrow \text{پاشنه} = ۲ \\ \rightarrow \text{ضریب} = -\frac{۱}{۲} \\ \rightarrow \text{ضریب} = ۰ \end{array} \Rightarrow \begin{bmatrix} ۲ & -۱ & ۰ \\ ۰ & ۱٫۵ & -۱ \\ ۰ & -۱ & ۲ \end{bmatrix} \begin{array}{l} \rightarrow \text{پاشنه} = ۱٫۵ \\ \rightarrow \text{ضریب} = -\frac{۲}{۳} \end{array}$$

$$\Rightarrow \begin{bmatrix} ۲ & -۱ & ۰ \\ ۰ & ۱٫۵ & -۱ \\ ۰ & ۰ & \frac{۴}{۳} \end{bmatrix} \Rightarrow \begin{cases} P_۱ = ۲ > ۰ \\ P_۲ = ۱٫۵ > ۰ \\ P_۳ = ۴/۳ > ۰ \end{cases}$$

۲-

$$S_۱ = \begin{bmatrix} ۲ \end{bmatrix} \Rightarrow |S_۱| = ۲ > ۰$$

$$S_۲ = \begin{bmatrix} ۲ & -۱ \\ -۱ & ۲ \end{bmatrix} \Rightarrow |S_۲| = ۳ > ۰$$

$$S_۳ = \begin{bmatrix} ۲ & -۱ & ۰ \\ -۱ & ۲ & -۱ \\ ۰ & -۱ & ۲ \end{bmatrix} \Rightarrow |S_۳| = ۴ > ۰$$

۳-

$$\det(S - \lambda I) = ۰ \Rightarrow \begin{cases} \lambda_۱ = ۲ - \sqrt{۲} > ۰ \\ \lambda_۲ = ۲ > ۰ \\ \lambda_۳ = ۲ + \sqrt{۲} > ۰ \end{cases}$$

۴- می‌دانیم برای یک ماتریس متقارن، تجزیه LU به صورت $S = LDL^T$ قابل بیان است. براین اساس می‌توانیم ماتریس‌های L , D را به صورت زیر به دست آوریم.

$$LDL^T = \begin{bmatrix} ۱ & ۰ & ۰ \\ -۱/۲ & ۱ & ۰ \\ ۰ & -۲/۳ & ۱ \end{bmatrix} \begin{bmatrix} ۲ & & \\ & ۳/۲ & \\ & & ۴/۳ \end{bmatrix} \begin{bmatrix} ۱ & -۱/۲ & ۰ \\ ۰ & ۱ & -۲/۳ \\ ۰ & ۰ & ۱ \end{bmatrix}$$

$$= (L\sqrt{D})(L\sqrt{D})^T = A_۱^T A_۱$$

$$\Rightarrow A_۱^T = \begin{bmatrix} ۱ & & \\ -۱/۲ & ۱ & \\ ۰ & -۲/۳ & ۱ \end{bmatrix} \begin{bmatrix} \sqrt{۲} & & \\ & \sqrt{۳/۲} & \\ & & \sqrt{۴/۳} \end{bmatrix}$$

توجه: $A_۱$ با نام عامل چالسکی S نیز شناخته می‌شود.

با توجه به رابطه قطری‌سازی S می‌دانیم:

$$S = Q\Lambda Q^T$$

عامل‌بندی دوم را می‌توان به صورت زیر تعریف کرد:

$$A_۲ = Q\sqrt{\Lambda}Q^T \Rightarrow A_۳^T A_۳ = Q\Lambda Q^T = S$$

۵-

$$x^TSx = ۲x_۱^۲ - ۲x_۱x_۲ + ۲x_۲^۲ - ۲x_۲x_۳ + ۲x_۳^۲$$

از تعریف ماتریس‌های $A_۱$, $A_۲$ کمک می‌گیریم:

$$x^TSx = \|A_۱x\|^۲ = ۲(x_۱ - ۱/۲\, x_۲)^۲ + ۳/۲\,(x_۲ - ۲/۳\, x_۳)^۲ + ۴/۳\, x_۳^۲$$

$$x^TSx = \|A_۲x\|^۲ = \lambda_۱(q_۱^Tx)^۲ + \lambda_۲(q_۲^Tx)^۲ + \lambda_۳(q_۳^Tx)^۲$$

ماتریس‌های مثبت نیمه‌معین

- تعریف: ماتریس متقارن S مثبت نیمه‌معین است اگر کوچکترین مقدار ویژه آن صفر باشد

- توجه: برای ماتریس متقارن مثبت نیمه‌معین S، $\underline{x}^T S x \geq 0$ خواهد بود.

- توجه: در ماتریس مثبت نیمه معین، دترمینان یکی از زیرماتریس‌های S_k صفر خواهد بود

- توجه: اگر ماتریس متقارن S که مثبت نیمه‌معین است را به صورت $S = A^T A$ عامل‌بندی کنیم ستون‌های A وابسته خواهند بود.

فرم های درجه دوم

- فرم خطی: تابع $f: \mathbb{R}^n \to \mathbb{R}$ که تنها عبارت‌هایی به فرم $c_i x_i$ دارد را تابع خطی می‌نامیم. این تابع به فرم ماتریسی نیز قابل نگارش است:

$$f(x_1, \dots, x_n) = \sum_{i=1}^{n} c_i x_i = \underline{c}^T \underline{x}$$

$$\underline{c} \triangleq \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

$$\underline{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- فرم درجه دوم: تابع $f: \mathbb{R}^n \to \mathbb{R}$ که تنها عبارت‌هایی به فرم $s_{ij} x_i x_j$ دارد را تابع درجه دوم می‌نامیم. این تابع به فرم ماتریسی نیز قابل نگارش است:

$$f(x_1, \dots, x_n) = \sum_i \sum_j s_{ij} x_i x_j = \underline{x}^T S \underline{x}$$

$$\underline{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$S \triangleq \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ \vdots & \ddots & & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}$$

- معادله بیضی: می‌توان معادله بیضی را به صورت ماتریسی زیر نوشت:

$$\underline{x}^T \underline{S}\, \underline{x} = 1$$

که در آن ماتریس S، مثبت معین است. در اینصورت در فضای دو بعدی داریم:

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} S \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} Q \Lambda Q^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$= \left(Q \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)^T \Lambda \left(Q \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)$$

$$= \underline{y}^T \Lambda\, \underline{y}$$

در بیضی $x^T S x = 1$ ، بردارهای ویژه محورهای بیضی و مقادیر $\frac{1}{\sqrt{\lambda_1}}$ و ... و $\frac{1}{\sqrt{\lambda_n}}$ نیم طول‌های بیضی هستند.

تست حداقل تابع

- فرض کنید تابع $F(\underline{x})$ را داریم و در نقطه $\underline{x} = \hat{\underline{x}}$ ، مشتق تابع نسبت به تمام عناصر بردار $\underline{x}$ صفر است. یعنی:

$$\left. \frac{\partial F}{\partial x_i} \right|_{x_i = \hat{x}_i} = 0 \quad \forall i \in \{1, \dots, n\}$$

در اینصورت می‌توان از ماتریس مشتق دوم به منظور تعیین حداقل یا حداکثر بودن بردار $\hat{\underline{x}}$ استفاده کرد. اگر ماتریس S را به صورت زیر تعریف کنیم:

$$S = \begin{bmatrix} \left. \frac{\partial^2 F}{\partial x_1 \partial x_1} \right|_{x_1 = \hat{x}_1} & \cdots & \left. \frac{\partial^2 F}{\partial x_1 \partial x_n} \right|_{\substack{x_1 = \hat{x}_1 \\ x_n = \hat{x}_n}} \\ \vdots & \ddots & \\ \left. \frac{\partial^2 F}{\partial x_n \partial x_1} \right|_{\substack{x_1 = \hat{x}_1 \\ x_n = \hat{x}_n}} & \cdots & \left. \frac{\partial^2 F}{\partial x_n \partial x_n} \right|_{x_n = \hat{x}_n} \end{bmatrix} = \nabla^2 F \Big|_{\underline{x} = \hat{\underline{x}}}$$

می‌توان نشان داد $\hat{x}$ حداقل نسبی است اگر S مثبت معین باشد و حداکثر نسبی است اگر $-S$ مثبت معین باشد.

علت: بسط تیلور تابع F حول نقطه $\underline{x} = \underline{\hat{x}}$ به صورت زیر است:

$$F(\underline{x}) = F(\underline{\hat{x}}) + (\underline{x}-\underline{\hat{x}})^T \underbrace{\nabla F(\underline{\hat{x}})}_{=0} + (\underline{x}-\underline{\hat{x}})^T \nabla^2 F(\underline{\hat{x}}) (\underline{x}-\underline{\hat{x}}) + O(x^3)$$

حال اگر مقادیر $F(\underline{x})$ در همسایگی $\hat{x}$ را محاسبه کنیم به رابطه زیر می رسیم:

$$F(\underline{x}) = F(\hat{x}) + (x-\hat{x})^T \nabla^2 F(\hat{x}) (x-\hat{x})$$

اگر $\nabla F(\hat{x})$ مثبت معین باشد درنتیجه:

$$(x-\hat{x})^T \nabla^2 F(\hat{x})(x-\hat{x}) > 0 \Rightarrow F(\underline{x}) > F(\hat{x}) \Rightarrow$$ $\hat{x}$ حداقل نسبی است.

اگر $-\nabla F(\hat{x})$ مثبت معین باشد درنتیجه:

$$(x-\hat{x})^T \nabla^2 F(\hat{x})(x-\hat{x}) < 0 \Rightarrow F(\underline{x}) < F(\hat{x}) \Rightarrow$$ $\hat{x}$ حداکثر نسبی است.

تجزیه به مقادیر تکین

- در فصل قبل با فرآیند قطری‌سازی یک ماتریس دلخواه با استفاده از بردارها و مقدارهای ویژه آشنا شدیم. در فرآیند قطری‌سازی سه محدودیت اصلی وجود داشت (فرض کنید ماتریس A قطری‌شدنی بود و $\Lambda = X^{-1}AX$):

۱- بردارهای ویژه که در ستون‌های X قرار می‌گیرند لزوماً متعامد نیستند.
۲- همه ماتریس‌های A قطری شدنی نیستند.
۳- قطری‌سازی محدود به ماتریس‌های مربعی است.

- هدف: تجزیه به مقادیر تکین (Singalar Value Decomposition) یا به اختصار SVD هر ماتریس را به قطعات کوچکتری تجزیه می‌کند که هر قطعه ضرب یک بردار ستونی در یک بردار سطری است.

- شرح: اگر بتوان یک ماتریس $m \times n$ را به صورت ضرب یک برداری ستونی m تایی و یک بردار سطری n تایی نوشت، به جای $m \times n$ عنصر، تنها $m+n$ عنصر خواهیم داشت. این کار می‌تواند در فشرده‌سازی یک سیگنال بسیار مفید باشد.

- مثال فشرده‌سازی: یک تصویر را می‌توان به صورت یک ماتریس $m \times n$ درنظر گرفت که m تعداد سطرهای پیکسل و n تعداد ستون‌های آن است. فرض کنید تصویر مورد نظر به صورت زیر باشد:

$$I = \begin{bmatrix} a & b & c & d & e \\ a & b & c & d & e \\ a & b & c & d & e \\ a & b & c & d & e \end{bmatrix}_{۴\times۵}$$

اگر بخواهیم این ماتریس را به همین صورت فوق ذخیره کنیم و برای هر عنصر ۸ بیت درنظر بگیریم، در اینصورت حجم حافظه مورد نیاز برابر است با:

$۴\times۵\times۸ = ۱۶۰$ bits

اما تصویر فوق را می‌توان به سادگی به صورت ضرب یک بردار ستونی در یک بردار سطری نوشت:

$$I = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} a & b & c & d & e \end{bmatrix}$$

که طبق رابطه فوق حجم حافظه مورد نیاز برابر با (۴+۵) × ۸ = ۷۲ خواهد بود.

- SVD در مقابل تبدیل فوریه: همانطور که با استفاده از SVD می‌توان به صورت فوق به فشرده‌سازی دست یافت، با استفاده از روش تبدیل فوریه و حذف فرکانس‌های غیراثرگذار در فهم سیگنال نیز می‌توان به فشرده‌سازی دست یافت. این دو نگاه اما یک تفاوت جدی دارند.

پایه‌های ثابت: در فشرده‌سازی به روش تبدیل فوریه، شما سیگنال تصویر را روی یک سری پایه ثابت بسط می‌دهید که سیگنال صحبت هیچ ارتباطی ندارند.

پایه‌های وفقی: در فشرده‌سازی با استفاده از SVD، پایه‌های تجزیه براساس خود سیگنال تصویر به دست می‌آیند.

- توجه: تجزیه براساس پایه‌های وفقی معمولاً کارایی بالاتری دارد اما با پیچیدگی بیشتری حساب می‌شود.

نحوه به دست آوردن SVD

ابتدا تعاریف زیر را درنظر بگیرید:

$u_1, \dots, u_r \rightarrow$ پایه متعامد یکه برای فضای ستونی A

$u_{r+1}, \dots, u_m \rightarrow$ پایه متعامد یکه برای فضای پوچ چپ A

$v_1, \dots, v_r \rightarrow$ (فضای سطری A) پایه متعامد یکه برای فضای ستونی A^T

$v_{r+1}, \dots, v_n \rightarrow$ پایه متعامد یکه برای فضای پوچ A

لم ۱: فضای سطری ماتریس A و ماتریس A^TA یکسان است.

علت: اگر فضای سطری را با $R(\circ)$ نشان دهیم داریم:

ستون‌های D ترکیب خطی سطرهای A است پس:

$$D = A^TA \longrightarrow$$

$$C(A^TA) \subseteq R(A) \quad (۱)$$

از طرف دیگر A^TA ماتریسی متقارن است و:

$$C(A^TA) = R(A^TA) \quad (۲)$$

از (۱) و (۲) نتیجه زیر را داریم:

$$R(A^TA) \subseteq R(A)$$

از طرفی قبلاً نشان دادیم $R(A^TA) = R(A)$

بنابراین:

$$R(A^TA) = R(A)$$

لم ۲: بردارهای ویژه A^TA که متناظر با مقدارهای ویژه غیرصفر هستند در فضای سطری A قرار می‌گیرند و می‌توانند یک پایه برای این فضا باشند.

علت: فرض کنیم $\underline{v}$ بردار ویژه A^TA متناظر با مقدار ویژه σ^2 باشد در این صورت:

$$A^TAv = \sigma^2 v \Rightarrow$$

v در فضای ستونی یا سطری A^TA است.

نتیجه: بردارهای $v_1, \dots, v_n$ را برابر بردارهای ویژه ماتریس A^TA قرار می‌دهیم.

حال تعاریف زیر را در نظر بگیرید:

$$\underline{A}\,\underline{v}_1 \triangleq \underline{\hat{\sigma}}_1\,\underline{\hat{u}}_1$$

$$\vdots$$

$$\underline{A}\,\underline{v}_r \triangleq \underline{\hat{\sigma}}_r\,\hat{u}_r$$

اگر ضرب داخلی $\hat{u}_i$ و $\hat{u}_j$ را محاسبه کنیم داریم:

$$\hat{u}_i^T \hat{u}_j = \left(\frac{Av_i}{\hat{\sigma}_i}\right)^T \left(\frac{Av_j}{\hat{\sigma}_j}\right) = \frac{v_i^T A^T A v_j}{\hat{\sigma}_i \hat{\sigma}_j}$$

$$= \begin{cases} \dfrac{\sigma_j^2 \; v_i^T v_j}{\hat{\sigma}_i \hat{\sigma}_j} = 0 & i \neq j \\[2ex] \dfrac{\sigma_i^2}{\hat{\sigma}_i^2} \|v_i\|^2 = \dfrac{\sigma_i^2}{\hat{\sigma}_i^2} & i = j \end{cases}$$

بنابراین مجموعه $\hat{u}_1, \dots, \hat{u}_r$ مستقل خطی بود. و چون به صورت $Av = \hat{\sigma}\hat{u}$ تعریف شده است در فضای ستونی ماتریس A قرار دارد. بنابراین می‌توان آنها را به عنوان $u_1, \dots, u_r$ در نظر گرفت و تساوی مقابل را داشت: $\sigma_i = \hat{\sigma}_i \,, \; i = 1, \dots, r$

بنابراین رابطه ماتریسی زیر به دست می‌آید:

$$\left.\begin{aligned} Av_1 &= \sigma_1 u_1 \\ &\;\vdots \\ Av_r &= \sigma_r u_r \end{aligned}\right\} \Rightarrow A \underbrace{[v_1, \dots, v_r]}_{\triangleq V_r} = \underbrace{[u_1, \dots, u_r]}_{\triangleq U_r} \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}}_{\triangleq \Sigma_r}$$

برای دو ماتریس U_r و V_r روابط زیر را داریم:

$$V_r^T V_r = I$$

$$U_r^T U_r = I$$

در ماتریس Σ_r نیز عناصر روی قطر اصلی به ترتیب نزولی مرتب شده‌اند.

عبارت به دست آمد. هسته اصلی تجزیه به مقادیر تکین است اما ماتریس های U و V در این هسته مربعی نیستند. برای این منظور از فضاهای پوچی استفاده می کنیم.

ابتدا نشان می دهیم u_i ها برای $i=1,\dots,r$ بردارهای ویژه AA^T هستند.

$$AA^T u_i = AA^T \frac{Av_i}{\sigma_i} = \frac{A\sigma_i^2 v_i}{\sigma_i} = \sigma_i^2 u_i \Rightarrow$$ u_i بردار ویژه متناظر با AA^T و مقدار ویژه σ_i^2 است.

ماتریس A^TA متقارن است. ← r بردار ویژه $v_1,\dots,v_r$ ← پایه های فضای سطری A هستند.

← $n-r$ بردار ویژه متعامد بر بردارهای فوق ← پایه های فضای پوچی A هستند.

ماتریس AA^T متقارن است. ← r بردار ویژه $u_1,\dots,u_r$ ← پایه های فضای ستونی A هستند.

← $m-r$ بردار ویژه متعامد بر بردارهای فوق ← پایه های فضای پوچی چپ A هستند.

مقادیر ویژه AA^T : $$\overbrace{u_1,\dots,u_r}^{\text{پایه های فضای ستونی } A} \qquad \overbrace{u_{r+1},\dots,u_m}^{\text{پایه های فضای پوچی چپ } A}$$

مقادیر ویژه A^TA : $$\underbrace{v_1,\dots,v_r}_{\text{پایه های فضای سطری } A} \qquad \underbrace{v_{r+1},\dots,v_n}_{\text{پایه های فضای پوچی } A}$$

حال ماتریس های U_r و V_r را با استفاده از پایه های فضای پوچی، فضای پوچی چپ به صورت زیر تکمیل می کنیم. ستون های ماتریس V بردارهای تکین ر، ستون های ماتریس U بردارهای تکین چپ ماتریس A هستند.

$$V = [v_1 \ v_2 \ \dots \ v_r \ v_{r+1} \ \dots \ v_n]$$

$$U = [u_1 \ u_2 \ \dots \ u_r \ u_{r+1} \ \dots \ u_m]$$

با استفاده از ماتریس‌های جدید روابط زیر را خواهیم داشت:

$$AV = U\Sigma$$

در رابطه فوق Σ ماتریس $m\times n$ بوده که دارای $m-r$ سطر صفر و $n-r$ ستون صفر است. U و V نیز در رابطه فوق مربعی و متعامد یکه هستند.

توجه: روابط اضافه‌شده در تساوی ماتریسی فوق به‌صورت مقابل است:

$$\begin{aligned} AV &= \left[Av_1 \quad Av_2 \quad \cdots \quad Av_r \quad \boxed{Av_{r+1} \cdots Av_n}\right] \\ U\Sigma &= \left[\sigma_1 u_1 \quad \sigma_2 u_2 \quad \cdots \quad \sigma_r u_r \quad \boxed{0 \quad \cdots \quad 0}\right] \end{aligned}$$

بخش اضافه‌شده

چون v_{r+1} تا v_n پایه زیرفضای پوچی ماتریس A است پس: $Av_i = 0 \quad i=r+1,\dots,n$

بنابراین SVD به‌صورت زیر خواهد بود:

$$A = U\Sigma V^{-1} = U\Sigma V^T = \sum \sigma_i u_i v_i^T$$

مثال: تجزیه مقادیر تکین را برای ماتریس $A=\begin{bmatrix}1&1&0\\0&0&1\end{bmatrix}$, $B=\begin{bmatrix}1&1\\1&0\\0&1\end{bmatrix}$ محاسبه نمایید.

ماتریس A:

$$A^TA = \begin{bmatrix}1&1&0\\1&1&0\\0&0&1\end{bmatrix} \longrightarrow \begin{cases} \lambda_1 = 2 & v_1 = \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\\0\end{bmatrix} \\ \lambda_2 = 1 & v_2 = \begin{bmatrix}0\\0\\1\end{bmatrix} \\ \lambda_3 = 0 & v_3 = \begin{bmatrix}-1/\sqrt{2}\\1/\sqrt{2}\\0\end{bmatrix} \end{cases}$$

از محاسبه فوق مقادیر تکین ماتریس A نیز به‌صورت $\sigma_1 = \sqrt{\lambda_1} = \sqrt{2}$ و $\sigma_2 = \sqrt{\lambda_2} = 1$ به دست می‌آید. بنابراین:

$$V = \begin{bmatrix}1/\sqrt{2} & 0 & -1/\sqrt{2}\\ 1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 0 & 1 & 0\end{bmatrix} \qquad \Sigma = \begin{bmatrix}\sqrt{2} & 0 & 0\\ 0 & 1 & 0\end{bmatrix}$$

حال برای محاسبه u_1 و u_2 از رابطه $Av=\sigma u$ استفاده می‌کنیم:

$$u_1 = \frac{Av_1}{\sigma_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_2 = \frac{Av_2}{\sigma_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

بنابراین تجزیه مقادیر تکین ماتریس A به صورت زیر است:

$$A = U \Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$$

ماتریس B:

$$B^TB = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \longrightarrow \begin{cases} \lambda_1 = \sigma_1^2 = 3 & v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \\ \lambda_2 = \sigma_2^2 = 1 & v_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \end{cases}$$

بنابراین:

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$u_1 = \frac{1}{\sigma_1} Av_1 = \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} Av_2 = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

برای محاسبه u_3 نمی‌توانیم به صورت فوق عمل کنیم زیرا تنها دو بردار v داریم. برای این کار از ایده متعامدسازی گرام-اشمیت استفاده می‌کنیم. e_3 را فرض می‌کنیم و بخشی از این [illegible] که بر u_1 و u_2 متعامد است را به دست می‌آوریم (می‌توانید هر راستایی را به دلخواه انتخاب کنید اما این راستا نباید به طور کامل در صفحه متشکل از u_1 و u_2 قرار گیرد)

$$\hat{u}_3 = e_3 - \left(\frac{u_1^T e_3}{u_1^T u_1} u_1 + \frac{u_2^T e_3}{u_2^T u_2} u_2 \right) = \begin{bmatrix} -1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \xrightarrow{\text{نرمالیزه کردن}} u_3 = \begin{bmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

توجه کنید که به طور جایگزین می‌توانید از روش‌های زیر نیز برای به دست آوردن u_3 استفاده کنید.

۱- u_1 و u_2 را دو سطر یک ماتریس در نظر بگیرید و یک پایه برای فضای پوچی این ماتریس به دست آورید. این پایه در راستای u_3 است.

۲- بردارهای ویژه ماتریس BB^T را به دست آورید.

بنابراین SVD به صورت زیر خواهد بود:

$$B = U \Sigma V^T = \begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

توجه: برای ماتریس‌های متقارن و مثبت نیمه معین، تجزیه به مقادیر تکین معادل قطری‌سازی است.

- مثال: ماتریس A به صورت مقابل را در نظر بگیرید:
در مورد این ماتریس موارد زیر برقرار است:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

۱- این ماتریس چهار مقدار ویژه $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$ دارد.

۲- بردار ویژه متناظر با این چهار مقدار ویژه برابر (۱ و ۰ و ۰ و ۰) است.

۳- مقادیر ویژه برابر $\sigma_1 = 3, \sigma_2 = 2, \sigma_3 = 1$ هستند.

۴- بردارهای ویژه، ستون‌های ماتریس I هستند.

۵- ماتریس A قطری‌ناشدنی است.

۶- ماتریس‌های U، Σ و V عبارت از:

$$U = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 3 & & & \\ & 2 & & \\ & & 1 & \\ & & & 0 \end{bmatrix} \quad V = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$A = U\Sigma V^T = \underbrace{3u_1v_1^T}_{\text{ماتریسی که عنصر (۳,۴) آن یک است}} + \underbrace{2u_2v_2^T}_{\text{فقط عنصر (۲,۳) یک است}} + \underbrace{u_3v_3^T}_{\text{فقط عنصر (۲,۱) یک است}}$$

- بررسی پایداری مقدارهای ویژه و تکین :

فرض کنید در ماتریس A تغییر بسیار کوچکی به صورت مقابل رخ دهد:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ \frac{1}{60000} & 0 & 0 & 0 \end{bmatrix}$$

در اینصورت:

بردارهای ویژه برابر $\frac{1}{10}$ ، $\frac{i}{10}$ ، $-\frac{1}{10}$ و $-\frac{i}{10}$ می‌گردند (به پرش بردارها توجه کنید)

مقدارهای ویژه برابر $\frac{1}{60000}$, ۱ , ۲ , ۳ می‌گردند.

همانطور که مشاهده می‌کنید مقدارهای ویژه مقاومت بالاتری در مقابل تغییرات جزئی در عناصر ماتریس دارند.

نقش ترتیبی بردارها

نسبت $r(\underline{x}) = \frac{\underline{x}^T S \underline{x}}{\underline{x}^T \underline{x}}$ نسبت رایلی نامیده می‌شود. حال مسأله بیشینه سازی زیر را در نظر بگیرید (فرض کنید $S = A^TA$)

$$\max_{\underline{x}} r(\underline{x}) = \frac{\underline{x}^T S \underline{x}}{\underline{x}^T \underline{x}}$$

با استفاده از مشتق گیری می‌توان نشان داد که تمام بردارهای $\underline{x}$ که به صورت $S\underline{x} = r(\underline{x})\underline{x}$ هستند مشتق اول عبارت فوق را صفر و منفی ماتریس مشتق مرتبه دوم در آنها مثبت معین است. بنابراین بردارهای ویژه بیشینه نسبی برای نسبت رایلی هستند (تنها در بردارهای ویژه مشتق صفر و منفی مشتق دوم، مثبت معین است). مقدار $r(\underline{x})$ در هر بردار ویژه برابر مقدار ویژه به دست می‌آید و برابر با همین مقدار ویژه است. بنابراین بیشینه مطلق به ازای بردار ویژه متناظر با بزرگترین مقدار ویژه (λ_1) به دست می‌آید و برابر با همین مقدار ویژه است.

بنابراین اگر بردارها و مقدارهای ویژه ماتریس S را براساس ترتیب نزولی مقدارهای ویژه مرتب نماییم به رابطه زیر می‌رسیم:

$$\lambda_1 = \max_{\underline{x}} \frac{\underline{x}^T S \underline{x}}{\underline{x}^T \underline{x}} , \quad \underline{x}^* = q_1 , \ S q_1 = \lambda_1 q_1$$

حال مسئله زیر را درنظر بگیرید:

$$\max_{\underline{x}} \frac{\|Ax\|}{\|x\|}$$

می‌دانیم برداری که $\frac{\|Ax\|}{\|x\|}$ را بیشینه نماید توان دوم آن را نیز بیشینه می‌نماید.

از طرفی با توجه به رابطه $S = A^T A$ داریم:

$$\left(\frac{\|Ax\|}{\|x\|}\right)^{۲} = \frac{(Ax)^T Ax}{x^T x} = \frac{x^T A^T A x}{x^T x} = \frac{x^T S x}{x^T x}$$

بنابراین مقدار q_1 عبارت $\frac{\|Ax\|}{\|x\|}$ را نیز بیشینه می‌نماید. مقدار این بیشینه $\lambda_1^{۲}$ است که با σ_1 نمایش می‌دهیم. پس:

$$\sigma_1 = \max_{\underline{x}} \frac{\|Ax\|}{\|x\|} , \quad x^* = v_1 , \ A v_1 = \sigma_1 \underline{u}_1$$

حال مسئله زیر را درنظر بگیرید:

$$\max_{\underline{x}} \frac{\underline{x}^T S \underline{x}}{\underline{x}^T x} , \quad q_1^T x = ۰$$

به عبارت دیگر دنبال راستای متعامد بر بردار ویژه q_1 هستیم که مقدار نسبت رایلی را بیشینه می‌کند.

به این منظور ابتدا ماتریس متعامد Q_1 را به صورت مقابل درنظر بگیرید. توجه شود که q_1 درستون اول این ماتریس قرار گرفته است.

$$\underline{Q}_1 = [\underline{q}_1 , \dots , \underline{q}_n]$$

اگر از تساوی $Sq_1=\lambda_1 q_1$ استفاده نمائیم به رابطه ماتریسی زیر دست پیدا می‌کنیم :

$$SQ_1 = S[q_1\ q_2 \cdots q_n] = [q_1\ q_2 \cdots q_n]\begin{bmatrix}\lambda_1 & \omega^T \\ \underline{0} & \underline{S}_{n-1}\end{bmatrix}$$

اگر ستون اول طرفین تساوی را محاسبه نمائیم به رابطه $Sq_1=\lambda_1 q_1$ خواهیم رسید. حال اگر طرفین را از سمت راست در $\underline{Q}_1^T$ ضرب کنیم ، رابطه زیر را خواهیم داشت :

$$Q_1^T S Q_1 = \begin{bmatrix}\lambda_1 & \underline{\omega}^T \\ \underline{0} & \underline{S}_{n-1}\end{bmatrix}$$

چون سمت چپ تساوی فوق متقارن است بنابراین روابط زیر می‌بایست برقرار باشد:

$$\underline{\omega} = \underline{0}$$
$$S_{n-1} = S_{n-1}^T$$

بنابراین :

$$Q_1^T S Q_1 = \begin{bmatrix}\lambda_1 & \underline{0} \\ \underline{0} & S_{n-1}\end{bmatrix} \Rightarrow S = Q_1\begin{bmatrix}\lambda_1 & \underline{0} \\ \underline{0} & S_{n-1}\end{bmatrix}Q_1^T$$

حال مسأله بهینه‌سازی مقابل را در نظر بگیرید :

$$\max_{\underline{x}} \frac{\underline{x}^T \underline{S}\, \underline{x}}{\underline{x}^T \underline{x}}\ ,\quad q_1^T x = 0$$

اگر S را جایگزین نمائیم ، به مسأله زیر می‌رسیم :

$$\max_{\underline{x}} \frac{x^T Q_1 \begin{bmatrix}\lambda_1 & 0 \\ 0 & S_{n-1}\end{bmatrix} Q^T x}{x^T x}\ ,\quad q_1^T x = 0$$

$$\equiv \max_{\underline{x}} \frac{(Q_1^T x)^T \begin{bmatrix}\lambda_1 & 0 \\ 0 & S_{n-1}\end{bmatrix} (Q^T x)}{x^T x}\ ,\quad q_1^T x = 0$$

$$\overset{y \triangleq Q_1^T x}{\equiv} \max_{\underline{y}} \frac{\underline{y}^T \begin{bmatrix}\lambda_1 & 0 \\ 0 & S_{n-1}\end{bmatrix} \underline{y}}{y^T y}\ ,\quad y = (0, \underline{\hat{y}})$$

$$\equiv \max_{\underline{\hat{y}}} \frac{\left[0, \underline{\hat{y}}^T\right] \begin{bmatrix} \lambda_1 & 0 \\ 0 & S_{n-1} \end{bmatrix} \begin{bmatrix} 0 \\ \underline{\hat{y}} \end{bmatrix}}{\hat{y}^T \hat{y}}$$

$$\equiv \max_{\hat{y}} \frac{\hat{y}^T S_{n-1} \hat{y}}{\hat{y}^T \hat{y}}$$

مسأله فوق مشابه مسأله بدون قید اول است با این تفاوت که $\hat{y}$ از بعد $n-1$ است. بنابراین مقدار بیشینه نسبت رایلی فوق برابر مقدار ویژه بیشینه ماتریس S_{n-1} است.

لم: اگر مقادیر ویژه S را با $\lambda_1 \geqslant \lambda_2 \geqslant \dots \geqslant \lambda_n$ نشان دهیم و مقادیر ویژه S_{n-1} را با $\lambda'_1 \geqslant \lambda'_2 \geqslant \dots \geqslant \lambda'_{n-1}$ نشان دهیم در اینصورت تساوی زیر برقرار است:

$$\lambda_{i+1} = \lambda'_i \quad , i = 1, \dots, n-1$$

با استفاده از لم فوق مقدار بهینه برای بردار $\underline{x}$ در مسأله زیر برابر $\underline{q}_2$ و بیشینه نسبت رایلی برابر با λ_2 می‌گردد.

$$\max_{\underline{x}} \frac{\underline{x}^T S \underline{x}}{\underline{x}^T \underline{x}} \quad , \underline{q}_1^T \underline{x} = 0$$

به طور مشابه می‌توان نشان داد برای مسأله $\max_{\underline{x}} \frac{\|Ax\|}{\|x\|} , \underline{v}_1^T \underline{x} = 0$ ، مقدار بیشینه برابر σ_2 و بردار سازنده این بیشینه نیز v_2 است.

- جمع‌بندی: ماتریس دلخواه A را درنظر بگیرید. بردارهای تکین این ماتریس بردارهای ویژه ماتریس مربعی و متقارن $S=A^TA$ است. بردارهای تکین ماتریس A را با v_i و بردارهای ویژه ماتریس S را با q_i نشان می‌دهیم. مقادیر ویژه λ_i متناظر با ماتریس S نیز مربع مقادیر تکین σ_i^2 متناظر با ماتریس A است.

S را با قطری سازی می توان به صورت زیر تجزیه کرد:

$$S = Q\Lambda Q^T = \sum_{i=1}^{r} \lambda_i q_i q_i^T$$

A را با SVD می توان به صورت زیر تجزیه کرد:

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

اگر فرض کنیم اندیس‌گذاری مقادیر ویژه و تکین λ_i و σ_i به صورت نزولی است در اینصورت بردارها و مقداری ویژه q_i و λ_i و بردارها و مقدارهای تکین v_i و σ_i را می توان به صورت ترتیبی زیر به دست آورد:

$$\lambda_1 = \max_x \frac{x^T S x}{x^T x} \qquad \underline{x}^* = q_1$$

$$\vdots$$

$$\lambda_i = \max_x \frac{x^T S x}{x^T x}, \quad q_j^T x = 0, \; j = 1, \dots, i-1 \qquad \underline{x}^* = q_i$$

$$\sigma_1 = \max_x \frac{\|Ax\|}{\|x\|} \qquad x^* = v_1$$

$$\vdots$$

$$\sigma_i = \max_{\underline{x}} \frac{\|Ax\|}{\|x\|}, \quad v_j^T x = 0, \; j = 1, \dots, i-1 \qquad x^* = v_i$$

تجزیه به مؤلفه‌های اصلی

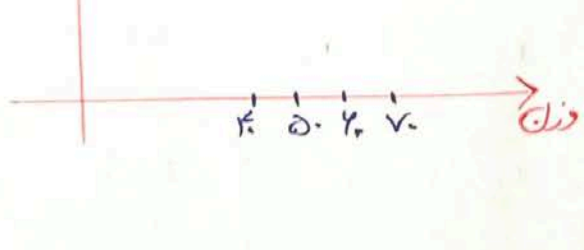
ماتریس داده‌های مربوط به ویژگی های شماره ۱و۲و... و m را در جامعه افراد ۱و۲و... و n در نظر بگیرید. فرض کنید این داده‌ها را در ماتریس A_0 به صورت زیر قرار دهیم:

$$A_0 = \begin{array}{cc} \left[\begin{array}{ccccc} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & & \vdots \\ a_{m1} & a_{m2} & \cdots & \cdots & a_{mn} \end{array}\right] & \begin{array}{l} \rightarrow \text{ویژگی اول} \\ \rightarrow \text{ویژگی دوم} \\ \\ \rightarrow \text{ویژگی m-ام} \end{array} \\ \begin{array}{ccccc} \uparrow & \uparrow & & & \uparrow \\ \text{فرد اول} & \text{فرد دوم} & & & \text{فرد n-ام} \end{array} & \end{array}$$

یک نمونه از این ماتریس می‌تواند قد و وزن دانش‌آموزان یک کلاس ۴ نفره باشد که به صورت زیر است.

$$B_0 = \begin{array}{cc} \left[\begin{array}{cccc} ۱۳۰ & ۱۳۵ & ۱۶۰ & ۱۵۰ \\ ۴۰ & ۵۰ & ۶۰ & ۷۰ \end{array}\right] & \begin{array}{l} \rightarrow \text{قد} \\ \rightarrow \text{وزن} \end{array} \\ \begin{array}{cccc} \uparrow & \uparrow & \uparrow & \uparrow \\ \text{نفر اول} & \text{نفر دوم} & \text{نفر سوم} & \text{نفر چهارم} \end{array} & \end{array}$$

اگر در صفحه قد برحسب وزن به این داده‌ها نگاه کنیم نمودار پراکندگی زیر را خواهیم داشت.



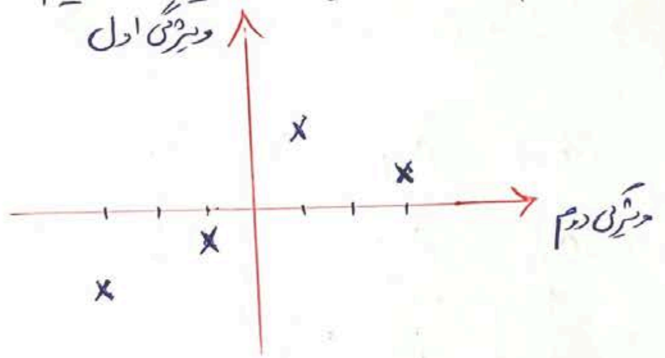
می توان برای هر سطر (هر ویژگی) در ماتریس B_0، متوسط ویژگی را حساب کرد و از داده های آن ویژگی کم کرد تا متوسط آن صفر گردد.

$$B_0 = \begin{bmatrix} ۱۳۰ & ۱۳۵ & ۱۴۰ & ۱۵۰ \\ ۴۰ & ۵۰ & ۶۰ & ۷۰ \end{bmatrix} \begin{matrix} \rightarrow \text{متوسط قد} = ۱۴۳,۷۵ \\ \rightarrow \text{متوسط وزن} = ۵۵ \end{matrix}$$

ماتریس جدید پس از صفر کردن میانگین سطرها (ویژگی ها) که با B نمایش می دهیم به صورت زیر است.

$$B = \begin{bmatrix} -۱۳,۷۵ & -۸,۷۵ & ۱۶,۲۵ & ۶,۲۵ \\ -۱۵ & -۵ & ۵ & ۱۵ \end{bmatrix} \begin{matrix} \rightarrow \text{متوسط ویژگی اول} = ۰ \\ \rightarrow \text{متوسط ویژگی دوم} = ۰ \end{matrix}$$

حال اگر به نقاط (هر ستون) در ماتریس B در فضای ویژگی اول (قد مرکزی شده) برحسب ویژگی دوم (وزن مرکزی شده) نگاه کنیم، نمودار پراکندگی زیر را خواهیم داشت.



ماتریس کوواریانس نمونه‌ای داده ها در ماتریس B به صورت $S = \frac{BB^T}{n-1}$ تعریف می شود. عناصر این ماتریس واریانس ویژگی ها و کوواریانس بین دو ویژگی را نشان می دهد. به عبارت دیگر:

$$[S]_{ij} = \frac{\sum_{k=1}^{n} b_{ik} b_{jk}}{n-1} = \frac{\sum_{k=1}^{n} (b_{ik}-\mu_i)(b_{jk}-\mu_j)}{n-1}.$$

$$= \begin{cases} \text{واریانس نمونه‌ای ویژگی } i\text{-ام} & i=j \\ \text{کوواریانس نمونه‌ای ویژگی } i\text{-ام و } j\text{-ام} & i \neq j \end{cases}$$

که در رابطه بالا μ_i متوسط ویژگی i-ام است و همانطور که می دانیم متوسط تمام ویژگی ها در ماتریس B صفر است.

با استفاده از مقادیر ویژه ماتریس S می‌توان مقادیر تکین ماتریس B را به دست آورد که عبارتند از:

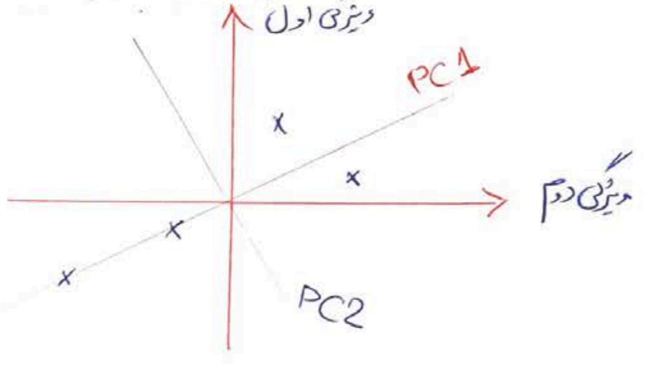
σ_1 = ۳۰٫۹۹۶

σ_2 = ۱۰٫۳۹۲

همانطور که می‌بینیم یکی از مقادیر ویژه اندازه بزرگتری نسبت به دیگری دارد.
اگر تقطیع رتبه اول ماتریس B را بازنویسی کنیم رابطه زیر را خواهیم داشت:

B = ۳۰٫۹۹۶ $u_1 v_1^T$ + ۱۰٫۳۹۲ $u_2 v_2^T$

جهت‌های u_1 و u_2 اطلاعات اصلی در مورد جهت‌های پراکندگی داده‌ها را ارائه می‌نمایند. مؤلفه u_1 جهتی را نشان می‌دهد که واریانس داده در آن جهت بیشینه است (این موضوع به سادگی قابل اثبات است). این مؤلفه با عنوان مؤلفه اصلی اول (PC 1) شناخته می‌شود. حال اگر جهتی متعامد بر u_1 را بخواهیم که پس از نگاشت داده‌ها در آن جهت بیشترین واریانس را داشته باشد پاسخ u_2 خواهد بود. این مؤلفه با عنوان مؤلفه اصلی دوم (PC2) شناخته می‌شود. شکل زیر این مؤلفه‌ها را نشان می‌دهد.



- استفاده از تجزیه به مؤلفه‌های اصلی: ماتریس $A: m \times n$ را در نظر بگیرید که داده‌های m ویژگی از n مشاهده را نشان می‌دهد. در اینصورت ماتریس کوواریانس به صورت $S = \frac{AA^T}{n-1}$ خواهد بود. گام‌های زیر نحوه رسیدن به استفاده عملیاتی از تجزیه به مقادیر اصلی را نشان می‌دهد.

۱- مجموع واریانس ویژگی‌ها ($\sum_{i=1}^{m} s_{ii}$) برابر مجموع مقادیر تکین ماتریس A است. یعنی:

$$\text{Trace } S = \sum_{i=1}^{m} \sigma_i^2 = T_s$$

۲- اگر میزان اهمیت یک جهت را برابر با میزان واریانس نمایش داده‌ها در آن جهت تعریف کنیم، یعنی:

$$\left.\begin{array}{l} A = [a_1, \dots, a_n] \\ \text{جهت دلخواه} = d \end{array}\right\} r_i = a_i^T d$$

r_i: نمایش داده i-ام در جهت d

$$\text{میزان اهمیت } d = \operatorname{var}(r_i) = \sum_{i=1} r_i^2$$

جهتی که بیشترین اهمیت را دارد u_1 خواهد بود. میزان واریانس در این جهت σ_1^2 است که برابر با $\frac{\sigma_1^2}{T_s}$ از مجموع واریانس ویژگی‌ها است.

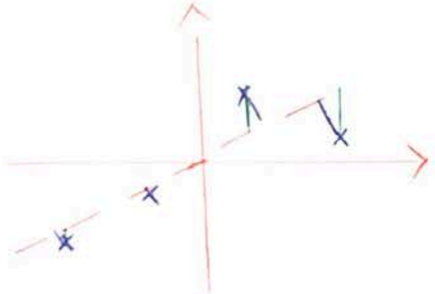
۳- حال اگر در میان جهت‌های عمود بر u_1، دنبال جهتی با بیشترین اهمیت باشیم پاسخ u_2 است که $\frac{\sigma_2^2}{T_s}$ از مجموع واریانس را دارد.

۴- اگر همین فرآیند را تا گام R ادامه دهیم در مجموع $\sum_{i=1}^{R} \frac{\sigma_i^2}{T_s}$ از مجموع واریانس را در نمایش داده توسط جهت‌های $u_1, \dots, u_r$ خواهیم داشت. نسبت $\sum_{i=1}^{R} \frac{\sigma_i^2}{T_s}$ به گونه ای است که قابل توجهی از مجموع واریانس را نشان می‌دهد.

۵- در این شرایط R را رتبه مؤثر ماتریس A می‌نامیم. در خود ماتریس A، بعد هر بردار مربوط به یک مشاهده، m است اما اگر آن را در فضای ایجاد. به u_R تصویر نماییم، این عدد به R می‌رسد که معمولاً خیلی کوچکتر از m و یا حتی ۲ است.

تجزیه به مؤلفه‌های اساسی (نگاه حداقل مربعات متعامد)

ماتریس B که در آن اطلاعات قد و وزن مرکزی شده (منظور از مرکزی شدن شرایطی است که متوسط ویژگی‌ها صفر شده است) را در نظر بگیرید. اگر بخواهیم با استفاده از حداقل مربعات که قبلاً با آن آشنا شدیم پارامترهای خط $y=a+bx$ را به دست آوریم در واقع فواصل نشان داده شده با خط سبز رنگ را حداقل می‌نماییم.



اما جهت u_1 که با استفاده از SVD به دست می‌آوریم متفاوت از جهت فوق است. برای نشان دادن این تفاوت تساوی زیر را در نظر بگیرید:

$$\sum_{j=1}^{4} \|b_j\|^2 = \sum_{j=1}^{4} |b_j^T u_1|^2 + \sum_{j=1}^{n} |b_j^T u_2|$$

سمت چپ تساوی فوق ثابت است. جهت u_1 به گونه‌ای انتخاب می‌گردد که عبارت اول در سمت راست تساوی بیشینه گردد یا به طور معادل عبارت دوم کمینه شود. عبارت دوم مربع فاصله نقاط از راستای u_1 است. بنابراین پاسخ PCA (PC1) در واقع حداقل مربعات متعامد خواهد بود.

بهترین تقریب رتبه k از ماتریس

در این بخش می‌خواهیم به کاربرد SVD در تعیین بهترین تقریب رتبه k از ماتریس دلخواه A بپردازیم. به این منظور ابتدا لازم است معیاری برای کیفیت تقریب در نظر بگیریم. در این بخش ابتدا نرم فروبنیوس را به این منظور معرفی می‌نماییم.

تعریف: نرم فروبنیوس ماتریس A با ابعاد m×n و عناصر a_{ij}، به صورت $\|A\|_F$ نمایش داده شده و برابر است با:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}}$$

توضیح: مسأله زیر را در نظر بگیرید:

$$\min_{B} \ \|A-B\|_F \quad \text{subject to} \quad r(B) \leqslant k$$

پاسخ مسأله فوق برابر است با:

$$B^* = A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

که σ_i، u_i و v_iها به ترتیب مقدار تکین، بردار تکین چپ و بردار تکین iـام ماتریس A هستند و فرض شده است که تکین به صورت نزولی مرتب شده اند.

اثبات:

لم ۱: می‌توان نشان داد برای ماتریس دلخواه A، $\|A\|_F^2$ برابر است با:

$$\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2$$

با استفاده از لم ۱ می‌توان نشان داد:

$$\|A-A_k\|_F^2 = \sum_{i=k+1}^{n} \sigma_i^2$$

حال فرض می‌کنیم B_k ماتریس دلخواهی است که $B_k = XY^T$ و $X: m\times k$ و $Y: n\times k$ است.

باید نشان دهیم:

$$\|A-B_k\|_F^2 \geqslant \sum_{i=k+1}^{n} \sigma_i^2$$

لم ۲: اگر $A = A' + A''$ در اینصورت $\sigma_1(A) \leqslant \sigma_1(A') + \sigma_1(A'')$

که $\sigma_1(\cdot)$ بزرگترین مقدار تکین ماتریس ورودی را نمایش می دهد. نامساوی فوق به نامساوی مثلثی نرم طیفی مشهور است.

فرض کنیم $A = A' + A''$ و A'_k ، A''_k تقریب رتبه k از A' ، A'' هستند که با استفاده از رابطه SVD به دست آمده‌اند. در اینصورت به ازای $i,j \geqslant 1$ داریم:

$$\sigma_i(A') + \sigma_j(A'') = \sigma_1(A' - A'_{i-1}) + \sigma_1(A'' - A''_{j-1})$$

$$\geqslant \sigma_1(A - A'_{i-1} - A''_{j-1})$$

$$\overset{*}{\geqslant} \sigma_1(A - A_{i+j-2})$$

$$= \sigma_{i+j-1}(A)$$

چون $\sigma_{k+1}(B_k) = 0$ است اگر $A' = A - B_k$ و $A'' = B_k$ قرار دهیم در اینصورت برای $i \geqslant 1$ و $j = k+1$ رابطه زیر را داریم: $\sigma_i(A - B_k) \geqslant \sigma_{k+i}(A)$

درنتیجه:

$$\|A - B_k\|_F^2 = \sum_{i=1}^{n} \sigma_i(A - B_k)^2 \geqslant \sum_{i=k+1}^{n} \sigma_i(A)^2 = \|A - A_k\|_F^2$$

توجه: نامساوی * از اینکه $r(A'_{i-1} + A''_{j-1}) \leqslant r(A_{i+j-2})$ است ناشی می‌شود (این نامساوی را نشان دهید برقرار است)

کار با ماتریس همبستگی: همانطور که دیدیم در ماتریس A_0 سطرهای غیرمرکزی بودند و متوسط آنها معمولاً غیرصفر بود. دیدیم که با داده‌های مرکزی شده حول مبدأ مختصات می‌توان متوسط سطرها را صفر کرد. پس از مرکزی کردن چون ویژگی‌های مختلف معمولاً مقیاس‌های متفاوتی دارند (قد و وزن)، می‌توان از یک ماتریس قطری D به گونه‌ای استفاده کرد که طول هر سطر در ماتریس DA برابر $\sqrt{n-1}$ گردد. با این کار در ماتریس کوواریانس برای ماتریس داده جدید $A' = DA$ روی قطری اصلی یک خواهد بود. این ماتریس را ماتریس همبستگی نمونه‌ای A می‌نامیم و معمولاً در پردازش‌ها از این ماتریس استفاده می‌شود.

نرم ماتریس: برای ماتریس دلخواه $A: m \times n$، نرم ماتریس با $\|A\|$ نمایش داده شده و از رابطه زیر به دست می‌آید:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1$$

خواص نرم ماتریس: برای دو ماتریس دلخواه دو خاصیت زیر برقرار است:

۱ـ نامساوی مثلثی:

$$\|A+B\| \leqslant \|A\| + \|B\|$$

۲ـ نامساوی ضرب.

$$\|AB\| \leqslant \|A\| \, \|B\|$$

تجزیه QS: هر ماتریس مربعی حقیقی یک عامل‌بندی به صورت $A = QS$ دارد که Q یک ماتریس متعامد یکه و S یک ماتریس مثبت نیمه‌معین است (اگر A معکوس‌پذیر باشد S مثبت معین خواهد بود). ماتریس‌های Q و S نیز به صورت زیر به دست می‌آیند:

$$A = U\Sigma V^T = U(V^T V)\Sigma V^T$$

$$= \underbrace{(UV^T)}_{Q}\underbrace{(V\Sigma V^T)}_{S}$$

- نزدیکترین ماتریس متعامد یکه: اگر $A=$ باشد، در اینصورت نزدیکترین ماتریس متعامد یکه به A در مفهوم نرم ماتریسی برابر $\overset{*}{Q}=UV^T$ است. به عبارت دیگر این ماتریس جواب مسأله زیر است:

$$\min_{Q \text{ متعامد یکه}} \| A-Q \|$$

- نزدیکترین ماتریس تکین: نزدیکترین ماتریس تکین به ماتریس A در معیار نرم ماتریسی را با A_0 نمایش می‌دهیم و برابر است با:

$$A_0=\sum_{i=1}^{r-1} \sigma_i u_i v_i^T$$

شبه معکوس

همانطور که دیدیم ماتریس A یک بردار v_i در فضای سطری را به بردار $\sigma_i u_i$ در فضای ستونی تبدیل می‌کرد. ماتریس معکوس در حالت کلی این کار را در جهت عکس انجام می‌دهد یعنی $A^{-1}u_i=\frac{v_i}{\sigma_i}$. بنابراین در ماتریس معکوس مقادیر تکین قرینه شده و u_i ها و v_i ها به ترتیب پایه‌های فضای سطری و ستونی خواهند بود.

A^{-1} در حالت کلی همواره وجود ندارد. شبه معکوس در واقع عملکردی مشابه معکوس برای حالت کلی‌تر ماتریس‌های تکین و غیرمربعی دارد. شبه معکوس با ضرب شدن در u_i، بردار $\frac{v_i}{\sigma_i}$ را تولید می‌نماید و به صورت زیر تعریف می‌شود:

$$A^+=V\Sigma^+U^T=[v_1,\dots,v_r,\dots,v_n]\begin{bmatrix}\sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \\ & & \end{bmatrix}[u_1\dots u_r\dots u_n]^T$$

توجه: با استفاده از تعریف می‌توان نشان داد:

$$A^+u_i=\frac{1}{\sigma_i}v_i \quad , \quad i\leq r$$

$$A^+u_i=0 \qquad i>r$$

توجه: اگر A^{-1} وجود داشته باشد $A^{+} = A^{-1}$ خواهد بود.

توجه: برای ماتریس Σ، شبه معکوس از معکوس کردن عناصر غیر صفر به دست می‌آید. در این صورت $\Sigma^{+}\Sigma$ یک ماتریس نزدیک به همانی خواهد بود. مثلاً:

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \Sigma^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\Sigma^{+}\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & \underline{0} \\ \underline{0} & \underline{0} \end{bmatrix}$$

توجه: ماتریس AA^{+} برابر ماتریس نگاشت به فضای ستونی A است.
ماتریس $A^{+}A$ برابر ماتریس نگاشت به فضای سطری A است.

توجه: همانطور که قبلاً دیدیم، شبه معکوس برای یافتن جواب حداقل مربعات با ویژگی حداقل طول در شرایطی که ستون‌های ماتریس A وابسته است کاربرد دارد. به این منظور فرض کنید برای حداقل مربعات به دستگاه $A^{T}A = A^{T}b$ رسید. دیدیم که چون ستون‌های A وابسته است بی‌نهایت جواب دارد. از میان این جواب‌ها، جوابی که حداقل طول را دارد برابر است با:
$x^{+} = A^{+}b$

- استفاده از شبه معکوس در حل دستگاه $Ax = b$:

پایه‌های متعامد یکه فضای ستونی $u_1, \dots, u_r$

پایه‌های متعامد یکه فضای پوچی چپ $u_{r+1}, \dots, u_m$

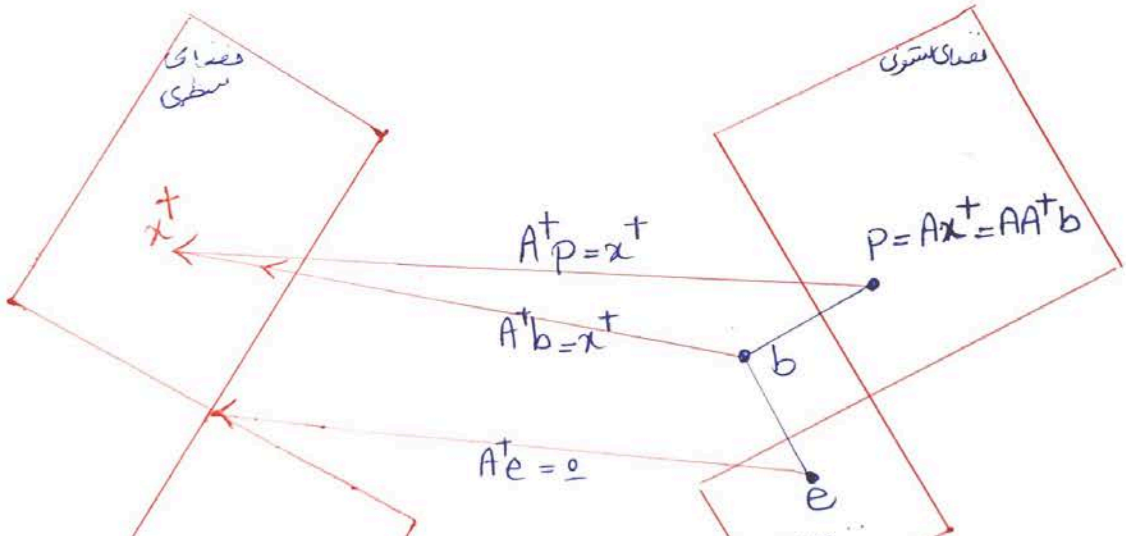
$$b = p + e = \sum_{i=1}^{r} c_i u_i + \sum_{j=1}^{m-r} d_j u_{r+j}$$

$$A^+ b = \sum_{i=1}^{r} c_i \underbrace{(A^+ u_i)}_{\frac{1}{\sigma_i} v_i} + \sum_{j=1}^{m-r} d_j \underbrace{(A^+ u_{r+j})}_{= \underline{0}}$$

$$= \sum_{i=1}^{r} \frac{c_i}{\sigma_i} v_i = x^+ \longrightarrow$$

این بردار در فضای سطری ماتریس A قرار دارد زیرا $v_i, i=1, \dots, r$، پایه‌های فضای سطری هستند.

$$Ax^+ = \sum_{i=1}^{r} \frac{c_i}{\sigma_i} A v_i = \sum_{i=1}^{r} \frac{c_i}{\sigma_i} \sigma_i u_i = \sum_{i=1}^{r} c_i u_i = P$$



$AA^+b \leftarrow$

$x^+ = A^+b \leftarrow$ حذف مؤلفه فضای پوچی چپ و به دست آوردن بردار متناظر در فضای سطری

$p = Ax^+ \leftarrow$ بازگشت از فضای سطری به فضای ستونی

$A^+Ax \leftarrow$

$b^+ = Ax \leftarrow$ حذف مؤلفه فضای پوچی و به دست آوردن بردار متناظر در فضای ستونی

$x_p = A^+b^+ \leftarrow$ بازگشت از فضای ستونی به فضای سطری

مثال: شبه معکوس ماتریس $A=\begin{bmatrix}1&1\\1&1\end{bmatrix}$ را به دست آورید.

$$\left.\begin{aligned}U&=\frac{1}{\sqrt{۲}}\begin{bmatrix}+1&+1\\+1&-1\end{bmatrix}\\V&=\frac{1}{\sqrt{۲}}\begin{bmatrix}+1&+1\\+1&-1\end{bmatrix}\\\Sigma&=\begin{bmatrix}۲&۰\\۰&۰\end{bmatrix}\end{aligned}\right\}\Rightarrow A^{+}=V\Sigma^{+}U^{T}=\frac{1}{\sqrt{۲}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}\begin{bmatrix}1/۲&۰\\۰&۰\end{bmatrix}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$$

$$=\frac{1}{۴}\begin{bmatrix}1&1\\1&1\end{bmatrix}$$

تبدیل‌های خطی

- تا این قسمت از درس با مفهوم ضرب ماتریسی به خوبی آشنا شدیم. ماتریس A و بردار v را درنظر بگیرید. ضرب ماتریسی Av را می‌توان به عنوان یک تبدیل درنظر گرفت که در آن ماتریس A، بردار v را به عنوان ورودی گرفته به بردار Av تبدیل می‌کند.

- تعریف: تبدیل خطی: فرض کنید V و W دو فضای برداری هستند که روی میدان F تعریف شده‌اند. یک تبدیل خطی از V به W یک تابع از V به W است که دو خاصیت زیر را برآورده می‌سازد:

۱) $T(\underline{v}+\underline{\ }) = T(\underline{v}) + T(\underline{\ }) \qquad \underline{u} \in V$

۲) $T(c\underline{v}) = cT(\underline{v}) \qquad \forall v \in V, \forall c \in F$

- توجه: براساس دو ویژگی تبدیل خطی می‌توان نشان داد هر تبدیل خطی است اگر و فقط اگر خاصیت زیر را داشته باشد:

$$T(c_1v_1 + c_2v_2) = c_1T(v_1) + c_2T(v_2)$$

- مثال: ۱- تبدیل $T(\underline{v}) = \underline{v} + \underline{u_0}$ را درنظر بگیرید. این تبدیل خطی نیست زیرا:

$$\left.\begin{aligned} T(v_1) &= v_1 + u_0 \\ T(v_2) &= v_2 + u_0 \\ T(v_1+v_2) &= v_1+v_2+u_0 \end{aligned}\right\} \Rightarrow T(v_1+v_2) \neq T(v_1) + T(v_2)$$

درحالت خاص $\underline{u_0} = \underline{0}$ تبدیل فوق خطی خواهد بود.

توجه: تبدیل فوق با نام تبدیل انتقال شناخته می‌شود زیرا هر بردار ورودی را به اندازه $\underline{u_0}$ انتقال می‌دهد و یک تبدیل غیرخطی است. تبدیل‌های افاین (Affine transformation) حالت کلی‌تری از تبدیل‌های خطی هستند که به صورت $T_{Aff}(v) = T_{Lin}(v) + u_0$ تعریف می‌شوند که $T_{Lin}(v)$ یک تبدیل خطی است. با توجه به اینکه $T(v) = v$ یک تبدیل خطی است (تبدیل همانی) بنابراین $T(v) = v + u_0$ یک تبدیل افاین است.

۲- $T(v) = a \cdot v$ که a یک بردار ثابت دلخواه است.

$$T(c_1v_1 + c_2v_2) = a\cdot(c_1v_1 + c_2v_2) = c_1 a\cdot v_1 + c_2 a\cdot v_2 = c_1T(v_1) + c_2T(v_2)$$

تبدیل فوق خطی است و در آن $V = \mathbb{R}^3$, $W = \mathbb{R}$ است.

۳- طول بردار $T(v) = \|v\|$ یک تبدیل غیرخطی است زیرا:

$$T(v_1+v_2) = \|v_1+v_2\| \leqslant \|v_1\| + \|v_2\| = T(v_1) + T(v_2)$$

$$T(-v_1) = \|-v_1\| = \|v_1\| \neq -\|v_1\|$$

خواص تبدیل خطی

- اگر بردار v_2 از دو بردار v_1 و v_2 دارای فاصله برابر باشد و T یک تبدیل خطی باشد آنگاه فاصله بردار $T(v_2)$ از بردارهای $T(v_1)$, $T(v_3)$ نیز برابر خواهد بود.

- توسعه ویژگی خطی بودن: ویژگی خطی بودن را به سادگی می‌توان برای حالتی که n بردار داریم، توسعه دهیم. فرض کنید $v = c_1v_1 + c_2v_2 + \dots + c_nv_n$ باشد. در اینصورت:

$$T(v) = c_1T(v_1) + c_2T(v_2) + \dots + c_nT(v_n)$$

- فرض کنید $v_1, v_2, \dots, v_n$ یک پایه برای فضای برداری V باشند و $T(v)$ را برای تمام بردارهای این پایه بدانیم. در اینصورت با استفاده از ویژگی توسعه یافته خطی می‌توانیم $T(u)$ را برای هر بردار دلخواه u در فضای برداری V به دست آوریم.

تعمیم زیرفضاهای اساسی به تبدیل خطی

- زیرفضاهایی که با آنها آشنا شدیم قابلیت توسعه به تبدیل‌های خطی را دارند. به عنوان نمونه تبدیل خطی $T(u) = \frac{du}{dx}$ را درنظر بگیرید که روی فضای چندجمله ای های درجه دوم تعریف شده است. می دانیم برای این تبدیل شرایط زیر را داریم:

$V \leftarrow$ زیرفضای چندجمله‌ای‌های درجه دوم (بعد ۳)

$W \leftarrow$ زیرفضای چندجمله‌ای‌های درجه اول (بعد ۲)

- **فضای پوچی** : فضای پوچی تبدیل T تمام بردارهای $\underline{u}$ است که در رابطه $T(u) = ۰$ صادق هستند. می دانیم که برای این تبدیل فضای پوچی تمام چندجمله ای های ثابت (درجه صفر) است. بعد این فضا نیز برابر یک است.

- **فضای ستونی** : می دانیم فضای ورودی شامل تمام چند جمله‌ای ها از درجه ۲ است. بنابراین فضای ستونی شامل تمام چندجمله ای ها از درجه ۱ است و بنابراین بعد آن برابر ۲ است.

- **قضیه شمارش** : بعد فضای پوچی (۱) + بعد فضای ستونی (۲) = بعد فضای ورودی (۳)

- **پایه های فضای ورودی** : می دانیم ورودی این تبدیل چندجمله‌ای های با حداکثر درجه ۲ هستند پس پایه های فضای ورودی عبارتست از :

$v_1 = 1$

$v_{۲} = x$

$v_{۲} = x^{۲}$

اگر تبدیل این پایه ها را داشته باشیم می توانیم تبدیل هر چندجمله درفضای ورودی را به دست آوریم :

$T(v_1) = ۰$

$T(v_{۲}) = 1$

$T(v_{۳}) = ۲x$

برای این پایه‌های خاص انتخابی می‌توانیم ماتریس متناظر با تبدیل T را نیز به دست آوریم.

$$\left.\begin{array}{l} u_1 = T(v_1) \Rightarrow v_1 = \begin{bmatrix} 1 \\ ۰ \\ ۰ \end{bmatrix} \Rightarrow Av_1 = \begin{bmatrix} a_{11} \\ a_{۲1} \end{bmatrix} \\ \qquad\qquad\qquad u_1 = \begin{bmatrix} ۰ \\ ۰ \end{bmatrix} \end{array}\right\} \Rightarrow \begin{bmatrix} a_{11} \\ a_{۲1} \end{bmatrix} = \begin{bmatrix} ۰ \\ ۰ \end{bmatrix}$$

$$u_{۲} = T(v_{۲}) \Rightarrow \begin{bmatrix} a_{۲۱} \\ a_{۲۲} \end{bmatrix} = \begin{bmatrix} 1 \\ ۰ \end{bmatrix}$$

$$u_{۳} = T(v_{۳}) \Rightarrow \begin{bmatrix} a_{۱۳} \\ a_{۲۳} \end{bmatrix} = \begin{bmatrix} ۰ \\ ۲ \end{bmatrix}$$

بنابراین اگر پایه‌های فضای ورودی V را برابر ۱، x، x^2 و پایه‌های فضای خروجی W را برابر با ۱ و x در نظر بگیریم، ماتریس متناظر با تبدیل خطی $T(u) = \frac{du}{dx}$ برابر است با:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

- شبه معکوس: چون در تبدیل T فضای پوچی Z نبود. و دارای بعد ۱ است پس این تبدیل معکوس ندارد اما شبه معکوس را می‌توان برای آن به صورت زیر به دست آورد:

$$T^{+}_{(u)} = \int_{0}^{x} u \quad x$$

برای تبدیل T^{+} فضای ورودی چند جمله‌ای‌های با حداکثر درجه ۱ است پس بردارهای پایه برابر است با:

$$u_1 = 1 \Longrightarrow A^{+} u_1 = v_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \underline{a}_1$$

$$u_2 = x \Longrightarrow A^{+} u_2 = v_2 = \begin{bmatrix} 0 \\ 0 \\ 1/2 \end{bmatrix} = a_2$$

$$\Rightarrow A^{+} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

توجه شود که A^{+} را می‌توان با استفاده از شبه معکوس گیری از A نیز به دست آورد.

- تعریف: گستره تبدیل (Range T): مجموعه تمام خروجی‌های ممکن $T(v)$ را گستره تبدیل T می‌نامند (گستره متناظر با فضای ستونی است)

- تعریف: کرنل تبدیل (kernel T): مجموعه تمام ورودی‌های v که $T(v)=0$ است را کرنل تبدیل T می‌نامند (کرنل متناظر با فضای پوچی است)

- توجه: گستره در فضای خروجی W و کرنل در فضای ورودی V است.

ماتریس یک تبدیل خطی

تبدیل خطی $T(\underline{v})$ که از فضای برداری $V=\mathbb{R}^n$ به فضای برداری $W\in\mathbb{R}^m$ را در نظر بگیرید. با توجه به اینکه برای فضای برداری V ، W چه پایه‌ای انتخاب شود، ماتریس تبدیل A_T متفاوت خواهد بود. به عنوان نمونه اگر ستون‌های ماتریس همانی $\underline{I}$ به عنوان پایه انتخاب نماییم، در اینصورت تبدیل با استناد از ماتریس A خواهد بود. با تغییر پایه‌های انتخابی، ماتریس A_T نیز متفاوت خواهد بود.

به عنوان نمونه اگر پایه فضای برداری V ، W را استاندارد (ستون‌های ماتریس همانی) انتخاب کنیم، ماتریس تبدیل همانی $T(v)=v$ ، ماتریس همانی خواهد بود اما اگر پایه‌های انتخابی برای V و W متفاوت باشد در اینصورت ماتریس این تبدیل همانی نخواهد بود.

نحوه تعیین ماتریس تبدیل : فرض کنید $T(\underline{v})$ را برای بردارهای پایه ورودی $\underline{v}_1$ تا $\underline{v}_n$ می دانیم. در اینصورت ستون اول تا n-ام ماتریس تبدیل برابر خروجی‌های $T(v_1)$ تا $T(v_n)$ خواهد بود. (در این حالت فرض کرده‌ایم که پایه‌های استاندارد برای زیر فضای W در نظر گرفته شده است).

توجه : بردار دلخواه v را در نظر بگیرید. این بردار را می‌توان به صورت $\underline{v}=\sum_{i=1}^{n} c_i v_i$ بسط داد. در اینصورت حاصلضرب ماتریس تبدیل A_T در بردار نمایش $c=(c_1, \dots, c_n)$ برابر است با:

$$\begin{bmatrix} T(v_1), \dots, T(v_n) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \sum_{i=1}^{n} c_i T(v_i) \overset{\text{خطی بودن}}{=} T\left(\sum_{i=1}^{n} c_i v_i\right)$$

$$= T(v)$$

بنابراین با داشتن یک پایه در فضای ورودی V می‌توانیم ماتریس تبدیل A_T را محاسبه نماییم.

مثال: تبدیل T را درنظر بگیرید که از فضای $V=\mathbb{R}^2$ به فضای $W=\mathbb{R}^3$ تعریف شده است. همچنین فرض کنید به ازای $v_1=(1,0)$ و $v_2=(0,1)$ روابط زیر را داریم:

$T(v_1)=(2,3,4)$

$T(v_2)=(5,5,5)$

اگر T یک تبدیل خطی باشد، ماتریس آن برابر است با (ماتریس استاندارد ماتریس تبدیل است هنگامی که پایه‌های فضاهای V, W استاندارد هستند):

$$A_T=\begin{bmatrix} 2 & 5 \\ 3 & 5 \\ 4 & 5 \end{bmatrix}$$

تغییر پایه

فرض کنید $V=\mathbb{R}^2$, $W=\mathbb{R}^2$ است و تبدیل $T(v)=v$ روی این زیر فضاها تعریف شده است. اگر پایه‌های درنظر گرفته شده برای V و W یکسان باشد، $A_T=I$ خواهد بود. در تغییر پایه به دنبال آن هستیم که از نمایش در پایه به نمایش در پایه برویم. اگر بتوانیم ماتریس تبدیل را برای این حالت به دست آوریم، در اینصورت با استفاده از این ماتریس می‌توانیم به سادگی تغییر پایه را انجام دهیم.

پایه ورودی v_1 و v_2 را به صورت مقابل درنظر بگیرید: $\begin{bmatrix} v_1 & v_2 \end{bmatrix}$

می‌دانیم نتیجه اثر تبدیل T بر روی این دو پایه به صورت زیر است:

$T(v_1)=v_1$

$T(v_2)=v_2$

اما در نمایش فوق در پایه V است و باید به پایه خروجی $\begin{bmatrix} w_1 & w_2 \end{bmatrix}$ انتقال یابد. برای این منظور باید v_1 و v_2 برحسب w_1 و w_2 بازنویسی شوند.

$v_1=1w_1+1w_2$

$v_2=2w_1+3w_2$

بنابراین ماتریس تبدیل برای حالتی که پایه ورودی $[v_1\ v_2]$ و پایه خروجی $[w_1\ v_2]$ است برابر با $B=W^{-1}V$ است زیرا:

$$\overbrace{[w_1\ \ w_2]}^{W}\,B=\overbrace{[v_1\ \ v_2]}^{V}$$

راه حل دیگر: فرض کنید بردار $\underline{u}$ در دو پایه $\{v_i\}_{i=1}^{n}$ ، $\{w_j\}_{j=1}^{n}$ به دو صورت زیر قابل نمایش است. در اینصورت داریم:

$$\underline{u}=\underline{V}\,\underline{c}=[v_1,\dots,v_n]\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}\Rightarrow Wd=Vc\Rightarrow d=W^{-1}Vc$$

$$\underline{u}=\underline{W}d=[w_1,\dots,w_n]\begin{bmatrix}d_1\\ \vdots\\ d_n\end{bmatrix}$$

توجه: بنابراین وقتی پایه استاندارد $V=I$ به یک پایه دیگر W تبدیل می‌شود، ماتریس تغییر پایه $B=W^{-1}$ (و نه $B=W$) است.

ساخت ماتریس تبدیل در حالت کلی (پایه ورودی $\{v_i\}_{i=1}^{n}$ ، پایه خروجی $\{w_j\}_{j=1}^{m}$ و تبدیل خطی T):

برای یافتن ستون i-ام ماتریس A_T ، می‌دانیم $T(v_i)$ برداری در فضای W است. در اینصورت می‌توان یک نمایش برای آن برحسب بردارهای پایه این فضا به دست آورد. فرض کنید این نمایش به صورت زیر باشد:

$$T(v_i)=a_{1i}w_1+a_{2i}w_2+\dots+a_{mi}w_m\ ,\ i=1,\dots,n$$

دراینصورت a_{ij} مؤلفه (i,j) -ام ماتریس A_T خواهد بود.

در این شرایط اگر ورودی خاص $(0,\dots,0,1,0,\dots,0)=v$ را در نظر بگیریم که برابر با e_k است، $Ve_k=v_k$ خواهد بود و $T(v_k)=a_k$ است که ستون k-ام ماتریس A_T است. بنابراین به ازای ورودی v_k ، نمایش خروجی در پایه $\{w_j\}_{j=1}^{m}$ به صورت a_k است. از طرفی می‌دانیم $Wa_k=T(v_k)$. بنابراین به ازای بردارهای پایه فضای ورودی، خروجی صحیح توسط ماتریس معرفی شده تولید می‌گردد. با استفاده از ویژگی خطی بودن می‌توان نشان داد این شرایط برای تمام بردارهای فضای V برقرار است.

مثال : فرض کنید پایه‌های فضای ورودی ۱ ، x ، x^2 , x^3 است و پایه‌های فضای خروجی ۱ ، x ، x^2 است. تبدیل T ، مشتق‌گیری است یعنی $T(v) = \frac{dv}{dx}$

در اینصورت ماتریس تبدیل متناظر با پایه‌های انتخابی برابر است با :

$$v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow u_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = a_1$$

$$v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow u_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = a_2$$

$$v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow u_3 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = a_3$$

$$v_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow u_4 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = a_4$$

$$\Rightarrow A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

$$v = c_1 + c_2 x + c_3 x^2 + c_4 x^3 \Rightarrow \underline{v} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$$

$$\underline{u} = A\underline{v} = \begin{bmatrix} c_2 \\ 2c_3 \\ 3c_4 \end{bmatrix} \Rightarrow u = c_2 + 2c_3 x + 3c_4 x^2$$

مثال : حال تبدیل $T^{+}(v)$ که انتگرال است را در نظر بگیرید. در این شرایط پایه‌های فضای ورودی ۱ ، x و x^2 است و پایه‌های فضای خروجی ۱، x ، x^2 و x^3 است. در اینصورت ماتریس تبدیل برای این پایه‌ها به صورت زیر است.

$$\underline{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow u_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = a_1$$

$$v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow u_2 = \begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 0 \end{bmatrix} = a_2$$

$$v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow u_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1/3 \end{bmatrix} = a_3$$

$$\Rightarrow A^{+} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

حال چندجمله‌ای $v = d_1 + d_2 x + d_3 x^2$ را در نظر بگیرید. با استفاده از ماتریس A^+ بردار خروجی متناظر با بردار ورودی $\underline{v} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$ برابر است با:

$$\underline{u} = A^+ v = \begin{bmatrix} 0 \\ d_1 \\ \frac{1}{2} d_2 \\ \frac{1}{3} d_3 \end{bmatrix} \Rightarrow u = d_1 x + \frac{d_2}{2} x^2 + \frac{d_3}{3} x^3$$

توجه شود که $AA^+ = I$ است یعنی اگر ابتدا انتگرال‌گیری و سپس مشتق‌گیری انجام شود فرآیند معکوس شده است اما $A^+A \neq I$ است یعنی اگر اول مشتق‌گیری انجام شود، دیگر فرآیند معکوس نمی‌شود.

علت معکوس‌ناپذیری $A^+A \leftarrow$ انتگرال مشتق پایه $v_1 = 1$ برابر صفر می‌گردد. این مشخصه در صفر شدن ستون اول ماتریس A^+A خود را نشان می‌دهد.

به عبارت دیگر ماتریس A یک متغیر آزاد داشته و در نتیجه بعد فضای پوچی آن یک است.

ضرب ماتریسی متناظر با ترکیب تبدیل‌ها

تعریف: تبدیل خطی S از فضای U به فضای V متناظر با ماتریس B و تبدیل خطی T از فضای V به فضای W متناظر با ماتریس A را در نظر بگیرید. در اینصورت ترکیب تبدیل‌های خطی T و S که خود یک تبدیل خطی است و با $T(S(u))$ نمایش می‌دهیم متناظر با ماتریس AB است (فرض کردیم که پایه‌های در نظر گرفته شده برای فضای V در خروجی تبدیل S و ورودی تبدیل T یکسان است.

مثال: می‌دانیم تبدیل خطی چرخش با زاویه θ که با T نمایش می‌دهیم متناظر با ماتریس A به صورت زیر است (

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

از طرفی براساس تعریف فوق می‌دانیم تبدیل T^2 برابر چرخش با زاویه 2θ است. ماتریس متناظر این چرخش را می‌توان از دو طریق به دست آورد:

۱) ضرب ماتریس های تبدیل $\Rightarrow$ ماتریس تبدیل $= AA = \begin{bmatrix} \cos^2\theta - \sin^2\theta & -2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \cos^2\theta - \sin^2\theta \end{bmatrix}$

۲) قرار دادن 2θ به جای θ در A $\Rightarrow$ ماتریس تبدیل $= \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}$

با برابر قرار دادن درایه‌های ماتریس تبدیل در دو حالت به دو تساوی مثلثی شناخته شده می‌رسیم.

$$\cos 2\theta = \cos^2\theta - \sin^2\theta$$
$$\sin 2\theta = 2\sin\theta\cos\theta$$

حال شرایطی را در نظر بگیرید که تبدیل T چرخش با زاویه θ، و تبدیل S چرخش با زاویه $-\theta$ باشد. در این صورت انتظار داریم ماتریس ترکیب دو تبدیل خطی برابر I باشد. از طرف دیگر با ضرب دو ماتریس رابطه زیر را داریم:

$$A^{(-\theta)}A^{(\theta)} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix}$$

از برابر عناصر ماتریس فوق و ماتریس I به تساوی مثلثی زیر می‌رسیم:

$$\cos^2\theta + \sin^2\theta = 1$$

انتخاب پایه مناسب

۱- بردارهای ویژه ماتریس تبدیل برای پایه‌های استاندارد ورودی و خروجی:

فرض کنید A ماتریس تبدیل متناظر با تبدیل خطی T است و A مربعی بوده و دارای n بردار ویژه مستقل خطی است. حال فرض کنید پایه ورودی و خروجی بردارهای ویژه A در نظر بگیریم. در اینصورت ماتریس تبدیل جدید که با $\hat{A}$ نمایش می‌دهیم برابر است با:

$$\hat{A} = [x_1 \dots x_n]^{-1} A [x_1 \dots x_n] = \Lambda$$

بنابراین $\hat{A}$ قطری بود. و برابر ماتریس مقدارهای ویژه است.

۲- انتخاب پایه‌های یکسان ورودی و خروجی:

اگر پایه‌های ورودی و خروجی یکسان باشند و در ستون های ماتریس B قرار داشته باشند، در اینصورت ماتریس تبدیل جدید برابر است با:

$$\hat{A} = \underset{\substack{\uparrow \\ \text{ماتریس تبدیل از استاندارد} \\ \text{به پایه انتخابی}}}{B^{-1}} \; \underset{\substack{\uparrow \\ \text{ماتریس تبدیل با} \\ \text{پایه‌های استاندارد}}}{A} \; \underset{\substack{\uparrow \\ \text{ماتریس تبدیل ورودی} \\ \text{از پایه انتخابی به استاندارد}}}{B}$$

بنابراین اگر برای ورودی و خروجی پایه یکسان انتخاب شود، ماتریس تبدیل حاصل مشابه با ماتریس تبدیل برای پایه های استاندارد یعنی A خواهد بود.

۳- انتخاب پایه‌ها براساس SVD :

می‌دانیم هر ماتریس دلخواه از جمله A که ماتریس تبدیل متناظر با پایه‌های استاندارد برای تبدیل خطی T است، دارای تجزیه SVD به صورت $A=U\Sigma V^T$ است. به عبارت دیگر می‌توانیم تساوی زیر را داشته باشیم:

$$\Sigma = U^{-1} A V$$

حال فرض کنید پایه ورودی را ستون‌های V، و پایه خروجی را ستون‌های U در نظر بگیریم. دراینصورت ماتریس تبدیل متناظر با این انتخاب برای پایه‌ها به صورت زیر است.

$$\hat{A} = \underset{\nearrow}{U^{-1}} \quad \underset{\uparrow}{A} \quad \underset{\nwarrow}{V}$$

تبدیل نمایش از پایه به استاندارد

تبدیل در پایه استاندارد

تبدیل نمایش از پایه استاندارد به پایه انتخابی

از طرفی براساس رابطه SVD می‌دانیم $\Sigma = \hat{A}$ است و درنتیجه قطری خواهد بود.

تعریف : ماتریس Σ را ایزومترک A گوییم اگر بتوان ماتریس‌های متعامد Q_1 و Q_2 را به گونه‌ای یافت که

$$\Sigma = Q_1^{-1} A Q_2$$

با تعریف فوق، اگر پایه‌های ورودی و خروجی را به ترتیب ستون‌های V و U انتخاب نماییم، ماتریس تبدیل حاصل، ایزومترک ماتریس تبدیل برای پایه‌های استاندارد یعنی A خواهد بود.

مثال: می‌دانیم برای تبدیل $T=\frac{d}{dx}$ از فضای چندجمله‌ای های درجه سوم به فضای چندجمله‌ای های درجه دوم اگر پایه های ورودی را ۱، x، x^2، x^3 و پایه های خروجی را به ترتیب ۱، x، x^2 در نظر بگیریم، ماتریس تبدیل به صورت زیر خواهد بود:

$$A=\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

حال فرض شود ترتیب پایه ها را در ورودی و خروجی معکوس نماییم. یعنی پایه های ورودی را x^3، x^2، x و ۱ و پایه های خروجی را x^2، x و ۱ را در نظر بگیریم. در اینصورت ماتریس تبدیل به صورت زیر خواهد شد:

$$\hat{A} = B_{out}^{-1} \; A \; B_{in}$$

تبدیل از پایه استاندارد به انتخابی

تبدیل در پایه استاندارد

تبدیل از پایه انتخابی به استاندارد

$$B_{in}=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

ستون اول نمایش $v_1=x^3$ در پایه استاندارد
" دوم نمایش $v_2=x^2$ در پایه استاندارد
" سوم نمایش $v_3=x$ در پایه استاندارد
" چهارم نمایش $v_4=1$ در پایه استاندارد.

$$B_{out}=\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

ستون اول نمایش $u_1=x^2$ در پایه استاندارد
" دوم نمایش $u_2=x$ در پایه استاندارد
" سوم نمایش $u_3=1$ در پایه استاندارد.

سایر پایه‌های مناسب

در بخش قبل دو انتخاب مناسب برای پایه‌های ورودی و خروجی تبدیل را دیدیم.

۱- برای تبدیل با ماتریس مربعی A که n بردار ویژه مستقل خطی دارد یک انتخاب مناسب برای پایه ورودی و خروجی ماتریس X است که ستون‌های آن، بردارهای ویژه هستند. در این حالت ماتریس جدید تبدیل برابر است با:

$$\hat{A} = X^{-1} A X = \Lambda$$

۲- برای تبدیل با ماتریس دلخواه A، یک انتخاب مناسب برای پایه ورودی ستون‌های V و برای پایه خروجی ستون‌های U است. در این حالت ماتریس جدید تبدیل برابر است با:

$$\hat{A} = U^{-1} A V = \Sigma$$

۳- فرم جردن: فرض کنید ماتریس مربعی A در اختیار شما قرار گرفته است و می‌خواهید ماتریس M را به گونه‌ای تعیین کنید که $M^{-1}AM$ تا حد امکان قطری باشد. می‌دانیم اگر در ماتریس مربعی $A: n\times n$، n بردار ویژه مستقل خطی داشته باشیم و ستون‌های M را برابر این بردارهای ویژه قرار دهیم، $M^{-1}AM$ قطری خواهد بود. حال شرایطی را در نظر بگیرید که A ناقص است. در اینصورت قطری‌شدنی نیست. در این شرایط فرم جردن نزدیکترین فرم برای A را ارائه می‌دهد.

فرض کنید A، s بردار ویژه مستقل خطی داشته باشد. در اینصورت می‌توان نشان داد این ماتریس مشابه یک ماتریس جردن با s بلوک مربعی روی قطر اصلی خواهد بود.

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_s \end{bmatrix}$$

که هر بلوک نیز متناظر با یک بردار ویژه و یک مقدار ویژه است و به صورت کلی زیر است:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{bmatrix}$$

مثال: ماتریس J به فرم جردن زیر را در نظر بگیرید.

$$J = \begin{bmatrix} \begin{bmatrix} 8 & 1 \\ 0 & 8 \end{bmatrix}^{J_1} & \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^{J_2} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \begin{bmatrix} 0 \end{bmatrix}^{J_3} \end{matrix} \end{bmatrix}$$

بر اساس فرم جردن بودن ماتریس فوق می‌توان بلافاصله نتایج زیر را گرفت:

۱- با توجه به وجود سه بلوک در ماتریس، سه بردار ویژه مستقل خطی وجود دارد.

۲- پنج مقدار ویژه ماتریس برابر ۸، ۸، ۰، ۰ و ۰ است.

۳- بردار ویژه‌های ماتریس برابر e_k برای $k = ۱, ۳, ۵$ است که این اندیس‌های شماره سطرهای شروع شده بلوک‌ها است.

تعمیم: اگر ماتریس مربعی $n \times n$، A، s بردار ویژه مستقل خطی داشته باشد، آنگاه ماتریس A مشابه با یک ماتریس J در فرم جردن خواهد بود. یعنی می‌توان ماتریس M را به گونه‌ای مشخص نمود که $J = M^{-1}AM$.

تعبیر فرم جردن: می دانیم که فرم جردن J در رابطه $J=M^{-1}AM$ صادق است.

بنابراین رابطه $AM = MJ$ را داریم. تعبیر ستونی این رابطه برای یک حالت خاص را در نظر بگیرید:

$$A\underset{\overset{\uparrow}{M}}{\begin{bmatrix} x_{۱} & x_{۲} & x_{۳} & x_{۴} & x_{۵} \end{bmatrix}} = \underset{\overset{\uparrow}{M}}{\begin{bmatrix} x_{۱} & x_{۲} & x_{۳} & x_{۴} & \lambda_{۵} \end{bmatrix}} \underset{\overset{\uparrow}{J}}{\begin{bmatrix} \overset{\text{بلوک ۱}}{\begin{bmatrix} \lambda & ۱ \\ ۰ & \lambda \end{bmatrix}} & & \\ & \overset{\text{بلوک ۲}}{\begin{bmatrix} ۰ & ۱ \\ ۰ & ۰ \end{bmatrix}} & \\ & & \overset{\text{بلوک ۳}}{\begin{bmatrix} ۰ \end{bmatrix}} \end{bmatrix}}$$

روابط زیر را داریم:

بلوک ۱: $Ax_{۱} = \lambda x_{۱}$, $Ax_{۲} = \lambda x_{۲} + x_{۱}$

بلوک ۲: $Ax_{۳} = ۰x_{۳}$, $Ax_{۴} = ۰x_{۴} + x_{۳}$

بلوک ۳: $Ax_{۵} = ۰x_{۵}$

همانطور که در روابط فوق مشاهده می کنید نتایج زیر برقرار است:

← همانطور که ماتریس به فرم جردن J، سه بردار ویژه مستقل دارد، ماتریس A نیز سه بردار ویژه مستقل $x_{۱}$، $x_{۳}$ و $x_{۵}$ را دارد.

← دو بردار خاص $x_{۲}$ و $x_{۴}$ نیز وجود دارند که بردارهای ویژه تعمیم یافته نامیده می شوند.

← هر بردار ویژه تعمیم یافته در یک رشته از بردارها قرار دارد که در رأس آن رشته بردار یک بردار ویژه قرار دارد. هر رشته نیز متناظر با یک بلوک در ماتریس جردن است. به عنوان مثال بلوک ۱ را در نظر بگیرید. در این بلوک بردار ویژه $x_{۱}$ و بردار ویژه تعمیم یافته $x_{۲}$ قرار دارد. این دو بردار ویژه در یک رشته بردار قرار دارند.

← بردارهای $x_{۳}$ و $x_{۴}$ در رشته بردار دوم قرار دارند

← بردار $x_{۵}$ در رشته بردار سوم است

← جستجو برای یافتن فرم جردن A معادل با یافتن رشته بردارهایی است که هرکدام شامل یک بردار ویژه و تعدادی بردار ویژه تعمیم یافته می شوند. بردارها در هر رشته بردار در یکی از روابط زیر صادق هستند:

$$Ax_i = \lambda_i x_i \qquad \text{یا} \qquad Ax_i = \lambda_i x_i + x_{i-1}$$

← ستون‌های ماتریس M از رشته بردارها ساخته می شود.

← اثبات قضیه بیان شده معادل با یافتن رشته بردارها است.

نحوه تعیین J , M : نحوه ساخت ماتریس‌های M و J در واقع از اثبات قضیه وجود ماتریس مشابه به فرم جردن به دست می آید. در این اثبات از استقراء استفاده شده است. گام های ساخت عبارتند از:

۱- ابتدا فرض کنید بعد ستونی ماتریس $A: n\times n$ برابر r است ($r<n$ است، در نتیجه می توان از فرض استقرا استفاده کرد) در نتیجه r بردار مستقل خطی x_i وجود دارد که در یکی از روابط زیر صادق است:

$$Ax_i = \lambda_i x_i \qquad Ax_i = \lambda_i x_i + x_{i-1}$$

۲- فرض کنید فضای پوچی و ستونی ماتریس A اشتراکی با ابعاد p داشته باشند.
می دانیم هر بردار x_i در فضای پوچی متناظر با مقدار ویژه $\lambda=0$ است. از طرف دیگر اگر این بردار در فضای ستونی نیز باشد رابطه $Ay_i = x_i$ برقرار خواهد بود. در نتیجه y_i یک بردار ویژه تعمیم یافته متناظر با مقدار ویژه صفر است (بردار ویژه متناظر با مقدار ویژه صفر در بخش ۱ استخراج شده است).

۳ـ علاوه بر p بردار استخراج شد در قسمت ۲ از فضای پوچی، در این فضا $n-r-p$ بردار پایه مستقل خطی مانند z_i نیز حضور دارند که $Az_i=0$ بود. و این بردارها در فضای ستون نیستند (این بردارها، بردار ویژه متناظر با مقدار ویژه صفر هستند).

مثال: فرم جردن مشابه با ماتریس A و ماتریس M متناظر را به دست آورید.

$$A=\begin{bmatrix} ۸ & ۰ & ۰ & ۸ & ۸ \\ ۰ & ۰ & ۰ & ۸ & ۸ \\ ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ & ۰ & ۸ \end{bmatrix}$$

گام ۱ ← رتبه A برابر با ۳ است.

ماتریس A، پنج مقدار ویژه ۰، ۰، ۰، ۸ و ۸ دارد.

در گام اول ما می بایست سه بردار مستقل خطی که در یکی از دو رابطه $\begin{cases} Ax_i=\lambda_i x_i \\ Ax_i=\lambda x_i+x_{i-1} \end{cases}$ صادق هستند را پیدا کنیم.

$$A-۸I \longrightarrow \text{پایه فضای پوچی} = \text{بردار ویژه متناظر با } \lambda=۸ = \begin{bmatrix} ۸ \\ ۰ \\ ۰ \\ ۰ \\ ۰ \end{bmatrix}$$

چون مقدار ویژه ۸، از مرتبه دوم بود اما پایه فضای پوچی آن یک بردار بود پس باید به دنبال یک بردار ویژه تعمیم یافته برای مقدار ویژه ۸ باشیم. یعنی بردار $x_۲$ را باید به گونه‌ای بیابیم که:

$$(A-۸I)x_۲=x_۱ \Rightarrow x_۲=\begin{bmatrix} ۰ \\ ۱ \\ ۰ \\ ۰ \\ ۱ \end{bmatrix}$$

حال به ازای $\lambda=۰$ داریم:

$$A \longrightarrow \text{پایه فضای پوچی} = \begin{matrix}\text{بردار ویژه متناظر با} \\ \lambda=۰\end{matrix} = \begin{bmatrix} ۰ \\ ۱ \\ ۰ \\ ۰ \\ ۰ \end{bmatrix} = x_۳$$

گام ۲: بعد فضای اشتراک پوچی و ستونی برابر با ۱ است. بردار x_3 را می توان به صورت ترکیب خطی ستون های A نوشت.

$$Ay = x_3 \Rightarrow y = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

بنابراین $\{y, x_3\}$ یک رشته بردار هستند.

گام ۳: بعد فضای پوچی ماتریس A برابر ۲ است. چون $p=1$ است بنابراین برداری مانند z در فضای پوچی وجود دارد که جزو فضای ستونی نیست. این بردار برابر است با:

$$z = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

جمع بندی:

$$M = \begin{bmatrix} x_1 & x_2 & x_3 & y & z \end{bmatrix}$$

$$J = \begin{bmatrix} \lambda & 1 & & & \\ 0 & \lambda & & & \\ & & 0 & 1 & \\ & & 0 & 0 & \\ & & & & 0 \end{bmatrix}$$

پایه براساس فرم جردن. اگر پایه ورودی و خروجی را برابر ستون های M در نظر بگیریم (M ماتریس حاصل از بردارهای ویژه و بردارهای ویژه تعمیم یافته است)، در اینصورت ماتریس تبدیل جدید به صورت زیر است:

$$\hat{A} = M^{-1} A M = J$$

بنابراین ماتریس تبدیل به فرم جردن است و حداقل فاصله از حالت قطری را خواهد داشت. در عین حال پایه های ورودی و خروجی نیز یکسان هستند.

نمایش تنک سیگنال‌ها در پایه فوق کامل

همانطور که می‌دانیم دستگاه معادلات $Ax=b$ در شرایطی که $n>m$ باشد لزوماً دارای متغیر آزاد بود. و بی‌نهایت جواب برای دستگاه وجود خواهد داشت (فرض ما براین است که حداقل یک جواب برای این دستگاه وجود دارد).

درنظر گرفتن ویژگی‌های دیگری برای بردار $\underline{x}$ می‌تواند منجر به یکتایی این بردار گردد. این شرایط را می‌توان به صورت مسأله بهینه‌سازی زیر فرمول‌بندی نمود:

$$\min_{\underline{x}} \; J(\underline{x}) \quad \text{subject to} \quad Ax=b$$

حداقل‌سازی $J(\underline{x})$ ویژگی دیگری است که برای $\underline{x}$ درنظر گرفتیم. یکی از این ویژگی‌ها می‌تواند حداقل طول بودن بردار $\underline{x}$ باشد. درانصورت مسأله زیر را خواهیم داشت:

$$\min_{\underline{x}} \; \|x\| \quad \text{subjelt to} \quad Ax=b$$

می‌دانیم مسأله فوق با استفاده از شبه معکوس قابل حل است و جواب آن به فرم بسته به دست می‌آید.

در این فصل به دنبال بردار $\underline{x}$ با یک ویژگی متفاوت از طول هستیم. برای تعریف این ویژگی ابتدا لازم است با تعریف نرم ها آشنا شویم که یک معیار از طول ارائه می‌دهند

تعریف: نرم l_p $(p \geqslant 0)$: نرم ها برای یک بردار دلخواه $\underline{x}$ با ابعاد n به صورت زیر تعریف می‌شود:

$$\|\underline{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

تعریف: نرم l_0: نرم l_0 حالت حدی از نرم l_p روی بردار $\underline{x}$ است که به صورت زیر تعریف می‌شود:

$$\|\underline{x}\|_0 = \lim_{p\to 0} \|x\|_p^p = \#\{i : x_i \neq 0\}$$

توجه: نرم l_2 همان طول بردار است و نرم l_0 برابر تعداد مؤلفه‌های غیر صفر بردار تعریف می‌شود.

مسئله نمایش تنک سیگنال: مسئله اصلی که در این فصل آن را مورد بررسی قرار می‌دهیم به صورت زیر فرمول‌بندی می‌شود:

$$P_0 : \min_{\underline{x}} \|\underline{x}\|_0 \quad \text{subject to} \quad Ax = b$$

تعبیر مسئله: ما به دنبال استفاده از حداقل ستون‌ها در A به منظور ساخت b هستیم.

تعریف: اسپارک ماتریس A را با $\text{spark}(A)$ نمایش می‌دهیم و برابر حداقل تعداد ستون‌هایی از A است که وابسته خطی هستند.

تضمین: اگر دستگاه معادلات $Ax = b$ یک جواب $\underline{x}^*$ داشته باشد به گونه‌ای که در رابطه $\|x^*\|_0 < \frac{1}{2} \text{spark}(A)$ صادق باشد، این پاسخ لزوماً تنک‌ترین پاسخ ممکن و پاسخ مسئله بهینه‌سازی P_0 است.

تعریف: همبستگی متقابل: برای یک ماتریس دلخواه $\underline{A} : m \times n$، همبستگی متقابل برابر بزرگترین اندازه ضرب داخلی نرمالیزه شده بین ستون‌های مختلف $\underline{A}$ است. این مقدار را با $\mu(\underline{A})$ نمایش می‌دهیم و داریم:

$$\mu(\underline{A}) = \max_{\substack{1 \leq i,j \leq n \\ i \neq j}} \frac{|a_i^T a_j|}{\|a_i\| \, \|a_j\|}$$

توجه: برای هر ماتریس $A: m\times n$ می‌توان نشان داد رابطه زیر برقرار است:

$$\mathrm{spark}(A) \geq 1 + \frac{1}{\mu(\underline{A})}$$

قضیه: اگر دستگاه معادلات خطی $\underline{A}\underline{x} = \underline{b}$ دارای پاسخ $\underline{x}^*$ باشد که در رابطه $\|x^*\|_0 < \frac{1}{2}\left(1+\frac{1}{\mu(A)}\right)$ صادق باشد در اینصورت این پاسخ تنک‌ترین پاسخ ممکن و پاسخ مسأله بهینه‌سازی P_0 است.

الگوریتم‌های حریص برای تعیین $\underline{x}$ در مسأله P_0:

ایده اصلی: فرض کنید $\mathrm{spark}(A) > 2$ بود. و در مسأله P_0 به شکل $\underline{x}^*$ رسید. اینکه $\|x^*\|_0 = 1$ است (به عبارت دیگر $\underline{b}$ یک ضریب از یکی از ستون‌های ماتریس A است). طبق قضیه یکتایی براساس spark، این جواب تنک‌ترین جواب ممکن و پاسخ مسأله P_0 است.

الگوریتم‌های حریص به دنبال یافتن بردار $\underline{x}^*$ هستند، اگر این بردار طبق قضایای گفته شد، نرم صفر آن از حد آستانه کمتر باشد، پاسخ مسأله بهینه‌سازی P_0 خواهد بود. برای شرایطی که دنبال x^* با $\|x^*\|_0=1$ هستیم باید مقدار $\min_{z_j}\|a_j z_j - b\|^2$ را برای jهای مختلف محاسبه نماییم. برای یک j مشخص مقدار فوق از طریق نگاشت قابل محاسبه است.

$$\left.\begin{array}{l} \epsilon(j) = \min\limits_{z_j} \|\underline{a}_j z_j - b\|^2 \\ \text{براساس نگاشت می‌دانیم:} \quad z_j^* = \dfrac{a_j^T b}{a_j^T a_j} \end{array}\right\} \Rightarrow \epsilon(j) = \|b\|_2^2 - \frac{(a_j^T b)^2}{a_j^T a_j}$$

اگر در شرایطی که $\mathrm{spark}(A) > 2$ است به ازای j مشخص خطای فوق ($\epsilon(j)$) صفر گردد، طبق قضایای گفته شد، x^* حاصل پاسخ مسأله P_0 خواهد بود.

حال شرایطی که $\mathrm{spark}(A) > 2k_0$ است را در نظر بگیرید و فرض کنید توانسته‌ایم برداری مانند $\underline{x}^*$ را به دست آورید که $\|x^*\|_0 = k_0$ است. در اینصورت $\underline{x}^*$ پاسخ بهینه‌سازی P_0 خواهد بود.

الگوریتم‌های حریص روشی برای تعیین x^* هستند. در این فصل سه الگوریتم از خانواده الگوریتم‌های حریص را مورد بررسی قرار می‌دهیم.

الگوریتم تعقیب انطباقی (Matching Pursuit):

روال انجام این الگوریتم به صورت زیر است:

۱) مقداردهی اولیه:

← $k = 0$ قرار بده.

← $S^k = \{\}$ ، $\underline{r}^k = \underline{b} - \underline{A}\underline{x}^k = \underline{b}$ ، $\underline{x}^k = \underline{0}$

۲) تکرار اصلی (k را یک واحد افزایش دهید و گام‌های زیر را تکرار کنید):

← محاسبه خطا برای تمام ستون‌های A:

$$\epsilon(j) = \min_{z_j} \|\underline{a}_j z_j - \underline{r}^{k-1}\|^2 = \|\underline{r}^{k-1}\|^2 - \frac{(a_j^T r^{k-1})^2}{a_j^T a_j}$$

← مجموعه S^k را با افزودن اندیس j_0 به روز رسانی کنید که j_0 پاسخ مسأله زیر است: $\left(S^k = S^{k-1} \cup \{j_0\}\right)$

$$\epsilon(j_0) \leqslant \epsilon(j) \quad 1 \leqslant j \leqslant n$$

← $\underline{x}^k = \underline{x}^{k-1}$ قرار دهید و مؤلفه j_0 در $\underline{x}^k$ را به صورت زیر به روز رسانی کنید:

$$x^k(j_0) = x^k(j_0) + z_j^*$$

← مانده جدید را محاسبه نمایید:

$$\underline{r}^k = \underline{b} - \underline{A}\underline{x}^k = r^{k-1} - z_{j_0}^* a_{j_0}$$

← بررسی شرط توقف: اگر $\|r^k\|_2 < \epsilon_0$ است الگوریتم را خاتمه دهید در غیراینصورت یک تکرار دیگر را شروع نمایید.

الگوریتم تعقیب انطباقی ضعیف (Weak Matching Pursuit):

۱) مقداردهی اولیه:

← $k=0$ قرار دهید.

← $\underline{x}^k=0$ ، $\underline{r}=b-Ax=b$, $S^k=\{\}$

۲) تکرار اصلی (k را یک واحد افزایش دهید و گام های زیر را تکرار کنید)

← $\epsilon(j)=\|r^{k-1}\|^2-\frac{(a_j^T r^{k-1})^2}{a_j^T a_j}$ را تا زمانی محاسبه نمایید که $\frac{|a_j^T r^{k-1}|}{\|a_j\|}\geqslant t\|r^{k-1}\|$. اندیس حاصل را در j_0 قرار دهید.

← $S^k=S^{k-1}\cup\{j_0\}$

← $\underline{x}^k=x^{k-1}$, $x^k(j_0)=x^k(j_0)+z_j^*$

← $r^k=b-Ax^k=r^{k-1}-z_{j_0}^*a_{j_0}$

← اگر $\|r^k\|_2<\epsilon_0$ الگوریتم را خاتمه دهید در غیراینصورت یک تکرار دیگر را شروع نمایید.

تفاوت MP و WMP: همانطور که در الگوریتم MP دیدیم، برای انتخاب یک اندیس از ستون های ماتریس A به منظور اضافه کردن به مجموعه S، خطا $\epsilon(j)$ برای تمام ستون های ماتریس A محاسبه می گردد اما در روش WMP درصورتی که اولین ستونی که رابطه $\frac{|a_j^T r^{k-1}|}{\|a_j\|}\geqslant t\|r^{k-1}\|$ را برقرار می نماید، پیدا شود، اندیس همان ستون به مجموعه S اضافه می شود. از این نظر الگوریتم WMP سرعت بالاتری دارد اما از طرف دیگر این الگوریتم وابسته به انتخاب مناسب پارامتر t است.

مشکل الگوریتم‌های MP , WMP : در هر دو این الگوریتم‌ها در گامی که اندیس‌های مجموعه S به روز رسانی می‌شود امکان انتخاب مجدد یک ستون تکراری از ماتریس A وجود دارد. برای حل این مشکل الگوریتم تعقیب انطباقی متعامد پیشنهاد شده است.

الگوریتم تعقیب انطباقی متعامد (Orthogonal Matching Pursuit)

۱) مقداردهی اولیه :

← $k=0$ قرار بده.

← $\underline{x}^k = \underline{0}$ ، $\underline{r}^k = \underline{b} - \underline{A}\underline{x}^k = \underline{b}$, $S^k = \{\}$

۲) تکرار اصلی (k را یک واحد افزایش دهید و گام‌های زیر را تکرار کنید) :

← محاسبه خطا برای تمام ستون‌های A :

$$\epsilon(j) = \min_{z_j} \|\underline{a}_j z_j - r^{k-1}\|^2 = \|r^{k-1}\|^2 - \frac{(a_j^T r^{k-1})^2}{a_j^T a_j}$$

← مجموعه S^k را با افزودن اندیس j_0 به روز رسانی کنید که j_0 پاسخ مسئله زیر است.

$$S^k = S^{k-1} \cup \{j_0\}$$

$$\epsilon(j_0) \leqslant \epsilon(j) \quad 1 \leqslant j \leqslant n$$

← x_s^k برابر نگاشت b روی ستون‌های A_{S^k} قرار دهید.

$$x_S^k = \underset{\underline{x}}{\operatorname{argmin}} \|\underline{A}_{S^k} \underline{x} - \underline{b}\|^2 \longrightarrow$$

← مانده جدید را به صورت زیر حساب کنید:

$$\underline{r}^k = \underline{b} - \underline{A}_{S^k} \underline{x}_s^k$$

← بررسی شرط توقف : اگر $\|r^k\|_2 < \epsilon_0$ است الگوریتم را خاتمه دهید در غیر اینصورت یک تکرار دیگر را شروع کنید.

همانطور که مشاهده می‌شود در الگوریتم OMP امکان انتخاب یکی از ستون‌های موجود در A_{S^k} در مرحله بعد وجود ندارد زیرا r^k بر تمامی این ستون‌ها عمود است و با انتخاب آن ستون تغییری در مانده ایجاد نمی‌شود.

تضیح: الگوریتم‌های حریص (OMP , WMP ، MP)، برای ماتریس A و ماتریس نرمالیزه شده ستونی $\tilde{A}$ جوابهای یکسانی تولید می‌نمایند (ماتریس نرمالیزه $\tilde{A}$ از برابر با یک کردن طول ستون‌های A با استفاده از ماتریس قطری W حاصل می‌شود. یعنی $\tilde{A} = AW$ که عناصر قطر اصلی W، معکوس طول ستون‌های ماتریس $\tilde{A}$ است).

تضیح: برای دستگاه معادلات خطی $\underline{A}\underline{x} = \underline{b}$ ($n>m$ ، $A: m\times n$ و رتبه کامل) اگر جوابی مانند $\underline{x}^*$ وجود داشته باشد که $\|\underline{x}^*\|_0 < \frac{1}{۲}\left(1+\frac{1}{\mu(A)}\right)$ باشد، در اینصورت الگوریتم OMP با $\epsilon = ۰$ به آن جواب همگرا خواهد شد.

مثال: نمایش تنک $b = \begin{bmatrix} ۳/\sqrt{۲} \\ -۱/\sqrt{۲} \end{bmatrix}$ روی ماتریس $\underline{A} = \begin{bmatrix} +۱/\sqrt{۲} & ۱ & ۰ \\ -۱/\sqrt{۲} & ۰ & ۱ \end{bmatrix}$ را به دست آورید. مقدار spark(A) و $\mu(A)$ را تعیین نمایید.

$$\underline{A} = \begin{bmatrix} ۱/\sqrt{۲} & ۱ & ۰ \\ -۱/\sqrt{۲} & ۰ & ۱ \end{bmatrix} \xrightarrow[\text{ضریب} = -۱]{\text{ستون۱} = ۱/\sqrt{۲}} \begin{bmatrix} ۱/\sqrt{۲} & ۱ & ۰ \\ ۰ & ۱ & ۱ \end{bmatrix} \xrightarrow[\text{ستون۱} = ۱]{\text{ضریب} = ۱} \begin{bmatrix} ۱/\sqrt{۲} & ۰ & -۱ \\ ۰ & ۱ & \underset{\uparrow}{۱} \end{bmatrix}$$

ستون آزاد

← حداقل تعداد ستون‌هایی که وابسته خطی هستند برابر با ۳ است ← $\text{Spark}(A) = ۳$

← چون ستون‌های A نرمالیزه است، $\mu(A)$ حداکثر اندازه عنصرهای غیر قطر اصلی در ماتریس A^TA است که برابر با $\frac{۱}{\sqrt{۲}}$ است ← $\mu(A) = \frac{۱}{\sqrt{۲}}$

$k=۰$

$$\underline{x}^k=\underline{۰}\ ,\quad r^k=\underline{b}=\begin{bmatrix}۳/\sqrt{۲}\\ -۱/\sqrt{۲}\end{bmatrix},\quad s^k=\{\}$$

$k=۱:$

$$\epsilon(j)\longrightarrow\begin{cases}\epsilon(۱)=۱\\ \epsilon(۲)=۱/۲\\ \epsilon(۳)=۴٫۵\end{cases}\Rightarrow j_۰=۲$$

$$S^k=\{۲\}$$

$$x_S^k=\underset{\underline{x}}{\arg\min}\ \|A_{S^k}\underline{x}-\underline{b}\|^۲=\underset{\underline{x}}{\arg\min}\left\|\begin{bmatrix}۱\\ ۰\end{bmatrix}\underline{x}-\begin{bmatrix}۳/\sqrt{۲}\\ -۱/\sqrt{۲}\end{bmatrix}\right\|^۲$$

$$=۳/\sqrt{۲}$$

$$r^k=\begin{bmatrix}۳/\sqrt{۲}\\ -۱/\sqrt{۲}\end{bmatrix}-\begin{bmatrix}۱\\ ۰\end{bmatrix}۳/\sqrt{۲}=\begin{bmatrix}۰\\ -۱/\sqrt{۲}\end{bmatrix}$$

$k=۲:$

$$\epsilon(j)=\begin{cases}\epsilon(۱)=۰٫۲۵\\ \epsilon(۲)=۰٫۵\\ \epsilon(۳)=۰\end{cases}\Rightarrow j_۰=۳$$

$$S^k=\{۲,۳\}$$

$$x_s^k=\underset{\underline{x}}{\arg\min}\ \|A_{S^k}\underline{x}-\underline{b}\|^۲=\underset{\underline{x}}{\arg\min}\left\|\begin{bmatrix}۱&۰\\ ۰&۱\end{bmatrix}\underline{x}-\begin{bmatrix}۳/\sqrt{۲}\\ -۱/\sqrt{۲}\end{bmatrix}\right\|^۲$$

$$=\begin{bmatrix}۳/\sqrt{۲}\\ -\frac{۱}{\sqrt{۲}}\end{bmatrix}$$

$$r^k=\begin{bmatrix}\frac{۳}{\sqrt{۲}}\\ -۱/\sqrt{۲}\end{bmatrix}-\begin{bmatrix}۱&۰\\ ۰&۱\end{bmatrix}\begin{bmatrix}۳/\sqrt{۲}\\ -۱/\sqrt{۲}\end{bmatrix}=۰\rightarrow$$ خاتمه الگوریتم

بردار $\underline{b}$ در بیشتر کاربردها طی یک عملیات اندازه‌گیری در اختیار شما قرار می‌گیرد. این عملیات معمولاً در حالت ایده‌آل انجام نمی‌شود و در نتیجه بردار b با نویز همراه خواهد بود. بنابراین آنچه در اختیار شما قرار می‌گیرد $\underline{b} = A\underline{x}_0 + \underline{e}$ که e بردار نویز با طول محدود $\|e\|^2 = \epsilon^2$ است. در این شرایط مسأله P_0^ϵ مورد بررسی قرار می‌گیرد. به طور شهودی مسأله P_0^ϵ با هدف بازیابی بردار $\underline{x}_0$ در شرایط نویزی طراحی شده است و همان کاری را می‌کند که مسأله P_0 برای شرایط بدون نویز انجام می‌دهد.

$$P_0^\epsilon : \min_{\underline{x}} \|x\|_0 \quad \text{subject to} \quad \|b - Ax\| < \epsilon$$

قضیه (پایداری جواب P_0^ϵ): مسأله P_0^ϵ را در نظر بگیرید. فرض کنید یک بردار $\underline{x}_0$ قید تنکی $\|x_0\|_0 < (1+\frac{1}{\mu(A)})/2$ را برآورده کرده و نمایش Ax_0 در رابطه $\|b - Ax_0\| \leq \epsilon$ صادق است. در این صورت هر پاسخ x_0^ϵ به P_0^ϵ باید در رابطه زیر صادق باشد:

$$\|x_0^\epsilon - x_0\|^2 \leqslant \frac{4\epsilon^2}{1-\mu(A)(2\|x_0\|_0 - 1)}$$

تذکر: در الگوریتم‌های حریص معرفی شده می‌توان به سادگی با انتخاب $\epsilon = \epsilon_0$ مسأله P_0^ϵ را مورد بررسی قرار داد و تقریبی برای پاسخ آن ارائه نمود.

توجه: هنگامی که یک مسأله یکسان با استفاده از فرمول‌بندی P_0 و P_0^ϵ همزمان حل می‌شود مسأله P_0^ϵ همواره به جواب‌های تنک‌تری دست پیدا می‌کند زیرا مجموعه پاسخ‌های مجاز P_0 زیرمجموعه پاسخ‌های مجاز P_0^ϵ است.

کاربردهای نمایش تنک :

۱- آنالیز: فرض کنید بردار b که در اختیار شما قرار گرفته است ترکیبی تنک از ویژگی‌هایی است که این ویژگی‌ها در ستون‌های ماتریس A قرار گرفته‌اند. با استفاده از مسأله زیر می‌توان تقریبی از ویژگی‌های قرار گرفته در بردار b را به دست آورد:

$$P_0^\epsilon : \min_x \|x\|_0 \quad \text{subject to} \quad \|y - Ax\| < \epsilon$$

۲- فشرده‌سازی: با استفاده از مسأله P_0^ϵ می‌توان به یک نمایش فشرده برای بردار b بر اساس ستون‌های A دست یافت. فرض کنید که طول بردار x از بردار b بیشتر است اما به علت تنکی x می‌توان به جای ارسال خود آن، موقعیت و مقدار عناصر غیرصفر را ارسال نمود و به فشرده‌سازی دست یافت.

۳- نویززدایی: همانطور که دیدیم پاسخ x_0^ϵ تقریبی از پاسخ مسأله P_0 است. بنابراین با داشتن این پاسخ می‌توان تقریبی از بردار بدون نویز به صورت Ax_0^ϵ را ساخت.

یادگیری واژه‌نامه

تا اینجای بحث همواره فرض ما بر این بود که ماتریس A به عنوان اطلاعات پیشین در اختیار ما قرار گرفته است. حال فرض کنید یک مجموعه بردار $\{y_i\}_{i=1}^N$ در اختیار شما قرار گرفته است و می‌دانید این بردارها از مدل $y_i = Ax_i + \epsilon_i$ تولید شده‌اند که $\|x_i\|_0 \leq k_0$ است. در مسأله یادگیری واژه‌نامه ما به دنبال پاسخ این پرسش هستیم که آیا با داشتن $\{y_i\}_{i=1}^N$ می‌توان A را طراحی نمود. ماتریس A در واقع واژه‌نامه مورد نظر در این بخش است.

بنابراین مسأله مورد نظر در یادگیری واژه نامه به صورت زیر است :

$$\min_{A,\{x_i\}_{i=1}^N} \sum_{i=1}^N \|x_i\|_0 \quad \text{subject to} \quad \|y_i - Ax_i\| < \epsilon, \ 1\leq i\leq N$$

به طور معادل می‌توان نقش تابع هزینه و قید را جابه‌جا نیز کرد به گونه ای که مسأله زیر را داشته باشیم:

$$\min_{A,\{x_i\}_{i=1}^N} \sum_{i=1}^N \|y_i - Ax_i\|^2 \quad \text{subject to} \quad \|x_i\|_0 \leq k. \ 1\leq i\leq N \tag{1}$$

توجه : تضمین منحصر به فرد بودن واژه نامه برای یک دسته از بردارهای $\{y_i\}_{i=1}^N$ تحت شرایط مشخصی وجود دارد و قابل اثبات است اما ما در این درس به آن نمی‌پردازیم.

توجه : با توجه به عدم تغییر نمایش تنک با تغییر طول ستون های ماتریس واژه نامه، معمولاً در روش‌های مختلف ستون های این ماتریس به گونه‌ای نرمالیزه می‌شوند که طول آن‌ها برابر با واحد گردد.

روال کلی آموزش واژه نامه

مسأله (۱) را به عنوان مسأله مورد نظر در یادگیری واژه نامه در نظر بگیرید. این مسأله دو دسته متغیر دارد. دسته اول عناصر ماتریس A و دسته دوم $\{x_i\}_{i=1}^N$ است. یک روال کلی برای حل مسائل با بیش از یک دسته متغیر، حدآقل‌سازی ترتیبی است. یک گام بهینه‌سازی برای یک مسأله با t دسته متغیر، در t زیرگام انجام می‌شود.

شبه‌کد بهینه‌سازی نوبتی:

فرض کنید مسأله بهینه سازی زیر را داریم:

$$\min_{\underline{x}_1,\dots,\underline{x}_t} f(\underline{x}_1,\dots,x_t) \quad \text{sabject to} \quad C(\underline{x}_1,\dots,x_t)$$

دراینصورت شبه‌کد بهینه‌سازی نوبتی به صورت زیر است:

۱) مقداردهی اولیه:

← $k=0$

← $x_1^{(k)}$ ، $x_2^{(k)}$ ، … ، $x_t^{(k)}$

۲) تکرار اصلی (k را یک واحد افزایش دهید و گام‌های زیر را تکرار کنید)

← زیرگام ۱:

$$x_1^{(k)} = \arg\min_{\underline{x}_1} f(x_1, x_2^{(k-1)}, x_3^{(k-1)}, \dots, x_t^{(k-1)}) \text{ s.t. } C(x_1, x_2^{(k-1)}, \dots, x_t^{(k-1)})$$

⋮

← زیرگام i:

$$x_i^{(k)} = \arg\min_{\underline{x}_i} f(x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k-1)}, \dots, x_t^{(k-1)} \quad \text{s.t.} \quad C(x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_i, x_{i+1}^{(k-1)}, \dots, \underline{x}_t^{(k-1)})$$

⋮

← زیرگام t:

$$x_t^{(k)} = \arg\min_{\underline{x}_t} f(x_1^{(k)}, \dots, \underline{x}_{t-1}^{(k)}, \underline{x}_t) \quad \text{s.t.} \quad C(x_1^{(k)}, \dots, x_{t-1}^{(k)}, x_t)$$

بهینه‌سازی نوبتی برای یادگیری واژه نامه :

حال می‌خواهیم از روال معرفی شده برای یادگیری واژه نامه استفاده کنیم. این روال در هر گام، دو زیرگام خواهد داشت :

۱) مقداردهی اولیه :

← $k=0$

← عناصر ماتریس A را به صورت تصادفی از توزیع $U(0,1)$ مقداردهی اولیه و $A^{(0)}$ را بسازید.

← نمایش‌های $\{\underline{x}_i\}_{i=1}^{N}$ را به صورت صفر مقداردهی اولیه کنید و $\{x_i^{(0)}\}_{i=1}^{N}$ را بسازید.

۲) تکرار اصلی (k را یک واحد افزایش دهید و گام های زیر را تکرار کنید)

← بهینه‌سازی روی نمایش‌های تنک :

$$x_i^{(k)} = \underset{x_i}{\arg\min} \| y_i - A^{(k-1)} x_i \|^2 \quad s.t. \quad \|x_i\|_0 \leqslant k_0 \quad 1 \leqslant i \leqslant N$$

همانطور که مشخص است مسئله برای هر مقدار i می‌تواند به طور مستقل حل شود. برای تقریب پاسخ مسئله فوق می‌توان از هر یک از الگوریتم‌های حریص استفاده کرد اما پیشنهاد می‌شود OMP انتخاب گردد.

← بهینه سازی روی واژه نامه :

$$A^{(k)} = \underset{A}{\arg\min} \sum_{i=1}^{N} \| y_i - A x_i^{(k)} \|^2$$

با استفاده از تعبیر ستونی ضرب ماتریسی می‌توان مسئله فوق را به صورت ماتریسی زیر نیز بیان نمود:

$$A^{(k)} = \underset{A}{\arg\min} \| Y - A X^{(k)} \|_F^2$$

که در آن تعاریف زیر را داریم :

$$Y \triangleq [y_1, \dots, y_N]$$

$$X^{(k)} \triangleq [x_1^{(k)}, \dots, x_N^{(k)}]$$

توجه: شبه کد فوق برای بهینه‌سازی روی نمایش تنک، OMP را پیشنهاد داده است.
دو روش شناخته شده برای یادگیری واژه نامه روش جهت‌های بهینه (MOD) و K تجزیه مقادیر تکین (KSVD) هستند که منطبق بر شبه کد فوق بود. و روش‌های مختلفی برای زیرگام بهینه‌سازی روی واژه نامه ارائه می‌دهند.

توجه: در هر دو روش فوق ماتریس A پس از مقداردهی اولیه و پس از به روز رسانی به گونه‌ای نرمالیزه می‌شود که طول ستون‌های آن برابر با واحد گردد.

الگوریتم یادگیری واژه نامه MOD :

اگر پاسخ ماتریسی زیرگام به روز رسانی واژه نامه را درنظر بگیریم، داریم:

$$A^{(k)} = \underset{A}{\operatorname{argmin}} \| Y - A X^{(k)} \|_F^{2}$$

می‌توان نشان داد که مسأله فوق یک مسأله حداقل مربعات است و در نتیجه می‌توان با استفاده از شبه معکوس، پاسخ آن را به دست آورد. این پاسخ برابر است با:

$$A^{(k)} = Y X^{(k)T} \left(X^{(k)} X^{(k)T} \right)^{-1}$$

بنابراین در روش فوق ماتریس $A^{(k)}$ به صورت یکجا به دست می‌آید. یک ایده دیگر به روز رسانی ستون‌های A به طور مجزا است که در الگوریتم KSVD مورد استفاده قرار گرفته است.

الگوریتم KSVD:

تابع هزینه زیر گام به روز رسانی واژه نامه را در نظر بگیرید. در اینصورت اگر ماتریس A را به صورت ستونی و ماتریس X را به صورت سطری در نظر بگیریم، با استفاده از تعبیر چهارم معرفی شده برای ضرب ماتریسی داریم:

$$\left.\begin{array}{l} A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \\ X = \begin{bmatrix} z_1^T \\ \vdots \\ z_n^T \end{bmatrix} \end{array}\right\} \Rightarrow \| Y - AX \|_F^2 = \| Y - \sum_{j=1}^{n} a_j z_1^T$$

$$= \| \left(Y - \sum_{\substack{j=1 \\ j \neq j_0}}^{n} a_j z_j^T \right) - a_{j_0} z_{j_0}^T) \|_F^2$$

رابطه فوق تابع هدف بهینه‌سازی را به صورت صریح برحسب ستون j_0 از ماتریس A نشان می دهد. فرض کنید $E_{j_0} \triangleq Y - \sum_{\substack{j=1 \\ j \neq j_0}}^{n} a_j z_j^T$ در اینصورت تابع هدف بهینه‌سازی برابر است با:

$$\min_{a_{j_0}} \| E_{j_0} - a_{j_0} z_{j_0}^T \|_F^2$$

می دانیم ماتریس حاصلضرب $a_{j_0} z_{j_0}^T$ که دارای ابعاد m×N است، از رتبه اول است. از طرف دیگر می دانیم بهترین تقریب رتبه اول از منظر نرم فروبنیوس، از تجزیه به مقادیر تکین به دست می آید. بنابراین می توان به سادگی با استفاده از SVD، $a_{j_0}^{(k)}$ را به دست آورد.

در الگوریتم KSVD در این قسمت یک نوآوری اضافه شده است. می دانیم $z_{j_0}^T$ به روز رسانی شده است (در گام نمایش تنک). عناصر غیرصفر در $z_{j_0}^T$ اندیس نمونه‌هایی از مجموعه $\{y_i\}_{i=1}^{N}$ است که در گام نمایش تنک، از ستون j_0 استفاده کرده اند.

در الگوریتم KSVD این اندیس‌ها در ماتریس E_{j_0} نیز انتخاب می‌شوند. فرض کنیم اندیس‌های غیرصفر را در مجموعه $\mathcal{S}$ قرار داده‌ایم. در اینصورت مسأله زیر مورد نظر قرار می‌گیرد:

$$\min_{a_{j_0}} \| E_{j_0,\mathcal{S}} - a_{j_0} z^T_{j_0,\mathcal{S}} \|_F^2$$

که در رابطه فوق متغیرهایی که زیرنویس $\mathcal{S}$ دارند با انتخاب ستون‌هایی که اندیس آنها در مجموعه $\mathcal{S}$ است از متغیر اصلی به دست می‌آید ($E_{j_0,\mathcal{S}}$ ماتریسی با ابعاد $m\times|\mathcal{S}|$ است و $z^T_{j_0,\mathcal{S}}$ نیز برداری با ابعاد $1\times|\mathcal{S}|$ است).

حال مسأله فوق با استفاده از SVD حل شده و $a_{j_0}^{(k)}$ به دست می‌آید.

بنابراین در گام به‌روزرسانی ستون‌ها در مجموعه باید به تعداد ستون‌های ماتریس A، تجزیه مقادیر تکین را محاسبه نمود و با فرض اینکه این تعداد برابر K است، روش حاصل با نام KSVD شناخته می‌شود.

کاربردهای حداقل مربعات

دستگاه $\underline{A}\underline{x}=\underline{b}$ را در نظر بگیرید که $A: m\times n$ بوده و $m>n$ است. بردار $\underline{x}$ در این دستگاه می‌توان دو حالت داشته باشد:

۱- بردار $\underline{b}$ در فضای ستون‌های A قرار دارد. در این حالت دو وضعیت می‌تواند ایجاد شود:

حالت اول: ماتریس A ستون آزاد ندارد ← پاسخ $\underline{x}$ منحصر به فرد است.

حالت دوم: ماتریس A حداقل یک ستون آزاد دارد ← $\underline{x}$ بی‌نهایت پاسخ دارد.

۲- بردار $\underline{b}$ در فضای ستون‌های A قرار ندارد ← پاسخ $\underline{x}$ وجود ندارد.

در مسأله حداقل مربعات ما به دنبال بردار x هستیم به گونه‌ای که بتواند طول بردار $r=Ax-b$ را کمینه سازد (r را مانده می‌نامیم). بنابراین مسأله حداقل مربعات را می‌توان به صورت زیر فرمول‌بندی نمود:

$$\min_{x} \|Ax-b\|$$

- در مسأله فوق $\underline{x}$ متغیر و بردار $\underline{b}$ و ماتریس $\underline{A}$ داده هستند.
- مقدار $\|Ax-b\|$ که باید کمینه شود نیز تابع هدف نامیده می‌شود.
- چون بردار $\underline{r}$ یک تابع آفاین از بردار $\underline{x}$ است، مسأله حداقل مربعات فوق را خطی می‌نامیم.
- بردار $\hat{x}$ پاسخ تقریب حداقل مربعات (پاسخ حداقل مربعات) است اگر به ازای هر بردار $\underline{x}$ رابطه زیر برقرار باشد:

$$\|A\hat{x}-b\| \leqslant \|Ax-b\|$$

- مسأله بهینه‌سازی فوق را مسأله رگرسیون بردار $\underline{b}$ روی ستون های ماتریس A نیز می‌نامند.

- تعیین پاسخ حداقل مربعات:

در فصل‌های قبل با نحوه تعیین پاسخ حداقل مربعات با استفاده از نگاشت آشنا شدیم. در این فصل از مشتق‌گیری برای تعیین این پاسخ و توسعه آن به مسائل مختلف استفاده می‌کنیم.

توجه شود که فرض ما بر این است که ستون‌های ماتریس A مستقل خطی هستند. در غیر اینصورت می‌توان ماتریس A را با A_{ind} جایگزین نمود.

مسأله بهینه‌سازی مقابل را در نظر بگیرید که $f(\underline{x}) = \|A\underline{x} - b\|^2$:

$$\min_{\underline{x}} \; f(\underline{x})$$

از حساب دیفرانسیل می‌دانیم $\hat{x}$ حداقل کننده تابع $f(\underline{x})$ است اگر:

$$\frac{\partial f}{\partial x_i}(\hat{x}) = 0 \,, \quad i = 1, \dots, n$$

اگر بردار گرادیان را به صورت مقابل تعریف کنیم:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

در اینصورت شروط فوق معادل است با:

$$\nabla f(\hat{x}) = \underline{0}$$

بنابراین لازم است بردار $\nabla f(\underline{x})$ را به دست آوریم:

$$f(x) = \|Ax - b\|^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} \cdot x_i - b_i \right)^2$$

$$\nabla f(\underline{x})_k = \frac{\partial f}{\partial x_k}(x)$$

$$= \sum_{i=1}^{m} ۲\left(\sum_{j=1}^{n} A_{ij}x_j - b_i\right)$$

$$= \sum_{i=1}^{m} ۲(A^T)_{ki}\,(Ax-b)_i$$

$$= \left(۲A^T(Ax-b)\right)_k$$

بنابراین $\nabla f(\underline{x}) = ۲A^T(Ax-b)$. $\hat{x}$ باید در رابطه زیر صادق باشد:

$$\nabla f(\hat{\underline{x}}) = \underline{0} \Rightarrow ۲A^T(A\hat{x}-b) = \underline{0} \Rightarrow A^TA\hat{x} = A^Tb$$

- دستگاه فوق مشابه دستگاهی است که با استفاده از نگاشت برای تعیین جواب حداقل مربعات به دست آوردیم.
- معادلات دستگاه فوق با نام معادلات نرمال شناخته می‌شوند.
- ماتریس $G=A^TA$ ، ماتریس گرام متناظر با ماتریس A است و مؤلفه‌های آن ضرب داخلی ستون‌های A است (چون فرض کردیم ستون‌های A مستقل خطی هستند، G معکوس پذیر است)
- چون ماتریس گرام معکوس پذیر است، در نتیجه دستگاه معادلات نرمال جواب منحصر به فرد زیر را دارد:

$$\hat{x} = (A^TA)^{-1}A^Tb$$

- اصل تعامد: مانده حاصل از پاسخ حداقل مربعات به فضای ستونی A عمود است.

بردار n بعدی $\underline{z}$ را در نظر بگیرید. می‌دانیم فضای ستونی A تمام ترکیب‌های خطی ستون‌های A است که می‌تواند به صورت Az نمایش داده شود. باید نشان دهیم: $(Az)^T\hat{r}=0$ که $\hat{r}$ مانده متناظر با پاسخ حداقل مربعات $\hat{x}$ است:

$$(Az)^T\hat{r} = (Az)^T(A\hat{x}-b) = x^TA^T(A\hat{x}-b)$$

$$= z^T\underbrace{(A^TA\hat{x} - A^Tb)}_{=0} = 0$$

حداقل مربعات ماتریسی :

همانطور که قبلاً نیز دیدیم، در مسأله یادگیری واژه نامه به روش MOD با مسأله حداقل سازی به فرم مقابل مواجه شدیم.

$$\min_{X} \|AX-B\|_F^2$$

که $A: m\times n$ ، $X: n\times N$ ، $B: m\times N$ بودند.

مسأله فوق در حالت خاص $N=1$ به مسأله حداقل مربعات تبدیل می‌شود. با استفاده از تعبیر ستونی ضرب ماتریسی می‌توان فوق را به N مسأله حداقل مربعات معمولی تبدیل نمود. برای این کار روال زیر را داریم:

$$\|AX-B\|_F^2 = \|Ax_1-b_1\|^2+\|Ax_2-b_2\|^2+\cdots+\|Ax_N b_N\|^2$$

هرکدام از عبارت‌ها در سمت راست تساوی فوق تنها به یکی از ستون‌های ماتریس X وابسته است بنابراین هرستون را می‌توان به طور مستقل از رابطه زیر تعیین نمود:

$$\hat{x}_i = (A^TA)^{-1}A^Tb_i \quad 1\leqslant i\leqslant N$$

برای یافتن ماتریس پاسخ حداقل مربعات کافیست دوباره از تعبیر ستونی ضرب ماتریسی استفاده کنیم، یعنی:

$$\hat{X} = [\hat{x}_1 \cdots \hat{x}_N] = (A^TA)^{-1}A^T[b_1,\ldots,b_N]$$

$$\Rightarrow \hat{X} = (A^TA)^{-1}A^TB$$

مثال: نورپردازی: فرض کنید n لامپ داریم و می‌خواهیم m ناحیه را با آنها روشن کنیم. فرض کنید بردار $\underline{l}$ روشنایی را در کل نواحی ارائه دهد (l_i روشنایی ناحیه i است و بردار $\underline{l}$ دارای طول m است). بردار n تایی $\underline{p}$ را توان متناظر با لامپها در نظر بگیرید (p_j توان لامپ j-ام است). روشنایی تابعی خطی از توان لامپ‌ها به صورت $l = Ap$ است که A ماتریس با ابعاد $m \times n$ بوده و ستون j-ام آن، بردار روشنایی در نواحی مختلف را ارائه می‌دهد وقتی توان لامپ j-ام برابر $\underline{1}$ و توان سایر لامپ‌ها برابر با صفر است (فرض کنید $m > n$ بوده و A رتبه کامل ستونی دارد). هدف تعیین توان لامپ‌ها به منظور رسیدن به یک الگوی $\underline{l}^{des}$ در روشنایی نواحی است.

پاسخ: با توجه به شرایط ماتریس A، دستگاه $l^{des} = Ap$ لزوماً جواب ندارد اما می‌توان در حالت کلی با استفاده از حداقل مربعات یک جواب $\hat{p}$ برای دستگاه فوق به صورت زیر به دست آورد:

$$\hat{p} = (A^TA)^{-1}A^T l^{des}$$

توجه: در راه حل فوق تضمینی برای مثبت بودن عناصر $\hat{p}$ و کمتر بودن آنها از یک آستانه از پیش تعیین شده وجود ندارد.

برازش منحنی با حداقل مربعات

فرض کنید بردار $\underline{x}$ با ابعاد $n\times 1$، و عدد y را در اختیار دارید و می‌دانید تابع مجهولی مانند $f: \mathbb{R}^n \to \mathbb{R}$ به یکدیگر مرتبط هستند. در زمینه‌های مختلف برای $\underline{x}$ و y نام‌های مختلفی استفاده می‌شود که عبارتند از:

$\underline{x} \leftarrow$ بردار ویژگی، متغیرهای مستقل.

$y \leftarrow$ نتیجه، متغیر وابسته، متغیر پاسخ.

داده: در مسأله برازش منحنی ما تابع f را در اختیار نداریم. به جای f تعدادی داده به صورت زیر در اختیار داریم:

$\underline{x}^{(1)}, y^{(1)} \longrightarrow$ زوج داده اول

$\underline{x}^{(2)}, y^{(2)} \longrightarrow$ زوج داده دوم

$\vdots$

$\underline{x}^{(N)}, y^{(N)} \longrightarrow$ زوج داده N-ام

علاوه بر داده، مقادیر فوق با نام مشاهدات، نمونه‌ها و اندازه‌گیری‌ها نیز شناخته می‌شوند.

مدلسازی: به منظور تخمین تابع مجهول f، یک رابطه بین $\underline{x}$ و y با تابع $\hat{f}$ به صورت زیر شکل می‌دهیم:

$$y \simeq \hat{f}(\underline{x})$$

که $\hat{f}: \mathbb{R}^n \to \mathbb{R}$ است و مدل نامیده می‌شود.

خطی بودن مدل (برحسب پارامترها): ما بر روی مدل‌هایی متمرکز می‌شویم که به فرم کلی زیر هستند:

$$\hat{f}(x) = \theta_1 f_1(x) + \cdots + \theta_p f_p(x)$$

که $f_i: \mathbb{R}^n \to \mathbb{R}$ توابع پایه یا نگاشت‌های ویژگی نامیده می‌شوند و θ_iها پارامترهای مدل هستند که باید انتخاب گردند. مدل فوق تابعی خطی از پارامترهای θ_i است.

تابع‌های پایه معمولاً برحسب اطلاعاتی که از f داریم انتخاب می‌شوند. پس از انتخاب آنها سؤال اصلی متمرکز بر پارامترهای مدل است.

خطای پیش‌بینی: پارامترها باید به گونه‌ای انتخاب شوند که $\hat{f}$ تا حد امکان با داده‌ها سازگار باشد یعنی: $i=1,\dots,N$ ، $y^{(i)} \simeq \hat{f}(\underline{x}^{(i)})$.

برای بردار ویژگی شماره i-ام یعنی $x^{(i)}$ ، فرض کنیم خروجی مدل به صورت زیر باشد:

$$\hat{y}^{(i)} = \hat{f}(x^{(i)})$$

دراینصورت خطای پیش‌بینی باقیمانده در این بردار ویژگی برابر است با:

$$r^{(i)} = y^{(i)} - \hat{y}^{(i)}$$

نمایش برداری: حال کل داده‌ها را که پایگاه داده نیز می‌نامیم درنظر بگیرید و بردارهای N‌تایی زیر را بسازید:

$$y^d = (y^{(1)}, \dots, y^{(N)})$$

$$\hat{y}^d = (\hat{y}^{(1)}, \dots, \hat{y}^{(N)})$$

$$r^d = (r^{(1)}, \dots, r^{(N)})$$

توجه شود که بالانویس d از آن جهت اضافه شده است که بردارها در ارتباط با داده‌ها تشکیل شده‌اند.

می‌دانیم رابطه متقابل بین سه بردار فوق برقرار است:

$$\underline{r}^d = \underline{y}^d - \hat{\underline{y}}^d$$

یک روش برای محاسبه سازگاری مدل $\hat{f}$ با پایگاه داده استفاده از ریشه متوسط مربعات بردار $\underline{r}^d$ است که با $rms(\underline{r}^d)$ نمایش داده شده و به صورت زیر محاسبه می‌شود:

$$rms(r^d) = \frac{\|r^d\|}{N}$$

یک معیار نسبی نیز می‌تواند $\frac{rms(\underline{r}^d)}{rms(y^d)}$ است.

برازش مدل مبتنی بر حداقل مربعات: برازش مدل به معنی تعیین پارامترهای θ_1، ... و θ_p متناظر با مدل است. این پارامترها را می‌توان با حداقل کردن $\|r^d\|^2$ (یا به طور معادل حداکثر کردن سازگاری مدل و پایگاه داده) به دست آورد. اگر رابطه $\hat{y}^{(i)} = \hat{f}(x^{(i)})$ را برحسب پارامترهای مدل بنویسیم رابطه زیر را داریم:

$$\hat{y}^{(i)} = A_{i1}\,\theta_1 + \dots + A_{ip}\,\theta_p \quad i = 1, 2, \dots, N$$

اگر ماتریس $A: N \times p$ را به صورت زیر تعریف کنیم:

$$A_{ij} = \hat{f}_j(x^{(i)}) \quad \begin{matrix} i = 1, \dots, N \\ j = 1, \dots, p \end{matrix}$$

و بردار p تایی $\underline{\theta}$ را به صورت زیر تعریف کنیم:

$$\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$$

در اینصورت خروجی مدل برای بردارهای ویژگی پایگاه داده برابر است با:

$$\hat{y}^d = A\theta$$

حال پارامترهای مدل را به گونه‌ای تعیین می‌کنیم که $\|r^d\|^2$ حداقل گردد، یعنی

$$\|r^d\|^2 = \|y^d - \hat{y}^d\|^2 = \|A\theta - y^d\|^2$$

مسأله فوق یک مسأله حداقل مربعات است، و پاسخ θ نیز پاسخ حداقل مربعات خواهد بود یعنی (فرض می‌کنیم ستون‌های A مستقل خطی هستند):

$$\hat{\theta} = (A^T A)^{-1} A^T y^d = A^{\dagger} y^d$$

روش فوق را برازش حداقل مربعات بر روی پایگاه داده می‌نامیم.

- توجه: عبارت‌های موجود در $\|y^d - A\theta\|^2$ دارای تعاریف زیر هستند:

$\hat{y}^d = A\theta$ ← خروجی‌های پیش‌بینی شده توسط مدل با پارامترهای θ

y^d ← مقدارهای واقعی مشاهده شده در خروجی‌ها

$y^d - A\theta$ ← بردار Nتایی خطای پیش‌بینی

$\|y^d - A\theta\|^2$ ← مربع طول بردار خطای پیش‌بینی (مجموع مربعات مانده‌ها)

- توجه: عبارت $\|y^d - A\theta\|^2$ به ازای $\theta = \hat{\theta}$ کمینه می‌شود. مقدار $\|y^d - A\hat{\theta}\|^2$ حداقل مجموع مربعات خطا (برای پایه‌های مدل و پایگاه داده مفروض) است. مقدار $\frac{1}{N}\|y^d - A\hat{\theta}\|^2$ حداقل متوسط مربع خطا (MMSE) است، و ریشه دوم آن حداقل ریشه متوسط مربعات خطا (RMS) برازش نامیده می‌شود.

مثال: برازش حداقل مربعات با استفاده از عبارت ثابت:

در این شرایط ۱ = p است و $f_1(x) = 1$ و $\hat{f}(x) = \theta_1$ خواهد بود. از طرفی می‌دانیم ماتریس A دارای ابعاد $N \times 1$ بود. و تنها ستون آن به صورت مقابل است:

$$A = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{N\times 1} = \mathbf{1}$$

پارامترهای مدل یعنی بردار θ که در اینجا دارای طول ۱ است نیز برابر است با:

$$\hat{\theta} = (\hat{A} A)^{-1} A^T y^d = N^{-1} \mathbf{1} y^d = avg(y^d)$$

خطای برازش RMS برابر است با:

$$rms(y^d - avg(y^d)\underline{1}) = std(y^d)$$

بنابراین هنگام برازش کردن تعدادی داده با استفاده از یک عدد ثابت، بهترین مقدار آن عدد ثابت در معیار حداقل مربعات خطا برابر متوسط داده‌ها و خطای برازش RMS در این حالت برابر معیار داده‌ها است.

مثال: برازش حداقل مربعات با استفاده از خط مستقیم:

در این شرایط $p=2$ است و توابع پایه به صورت $f_1(x)=1$, $f_2(x)=x$ هستند. بنابراین ماتریس A برابر است با:

$$A=\begin{bmatrix}1 & x^{(1)}\\ 1 & x^{(2)}\\ \vdots & \vdots\\ 1 & x^{(N)}\end{bmatrix}_{N\times 2}=\begin{bmatrix}\underline{1} & \underline{x}^{d}\end{bmatrix}$$

$$\left.\begin{array}{l}\hat{\theta}=(A^{T}A)^{-1}A^{T}y^{d}\\ A^{T}A=\begin{bmatrix}N & \underline{1}^{T}x^{d}\\ \underline{1}^{T}x^{d} & (x^{d})^{T}x^{d}\end{bmatrix}\\ A^{T}y=\begin{bmatrix}\underline{1}^{T}y^{d}\\ (x^{d})^{T}y^{d}\end{bmatrix}\end{array}\right\}\Rightarrow\hat{\theta}=\frac{1}{N(x^{d})^{T}x^{d}-(\underline{1}^{T}x^{d})^{2}}\begin{bmatrix}(x^{d})^{T}x^{d} & -\underline{1}^{T}x^{d}\\ -\underline{1}^{T}x^{d} & N\end{bmatrix}\begin{bmatrix}\underline{1}^{T}y^{d}\\ (x^{d})^{T}y^{d}\end{bmatrix}$$

$$=\frac{1}{\text{rms}(x^{d})^{2}-\text{avg}(x^{d})^{2}}\begin{bmatrix}\text{rms}(x^{d})^{2} & -\text{avg}(x^{d})\\ -\text{avg}(x^{d}) & 1\end{bmatrix}\begin{bmatrix}\text{avg}(y^{d})\\ \frac{(x^{d})^{T}y^{d}}{N}\end{bmatrix}$$

از روابط فوق می‌توان نشان داد که:

$$\begin{cases}\hat{\theta}_{2}=\frac{\text{std}(y^{d})}{\text{std}(x^{d})}\rho\ ,\quad \rho=\frac{1}{N}\frac{(x^{d}-\text{avg}(x^{d})\underline{1})^{T}(y^{d}-\text{avg}(y^{d})\underline{1})}{\text{std}(x^{d})\,\text{std}(y^{d})}\\ \hat{\theta}_{1}=\text{avg}(y^{d})-\hat{\theta}_{2}\,\text{avg}(x^{d})\end{cases}$$

$$\Rightarrow\hat{f}(x)=\hat{y}=\text{avg}(y^{d})+\rho\frac{\text{std}(y^{d})}{\text{std}(x^{d})}(x-\text{avg}(x^{d}))$$

رابطه فوق را می‌توان به فرم متقارن زیر نیز بیان نمود:

$$\frac{\hat{y}-\text{avg}(y^{d})}{\text{std}(y^{d})}=\rho\frac{x-\text{avg}(x^{d})}{\text{std}(x^{d})}$$

تبدیل رگرسیون به برازش حداقل مربعات

مدل رگرسیون برای بردار nتایی $\underline{x}$ به صورت زیر مدل بندی می‌شود:

$$\hat{y} = \underline{x}^T \underline{\beta} + v$$

که در رابطه فوق β بردار وزن و v آفست نامیده می‌شود.

حال این مدل را با مدل در نظر گرفته شده برای برازش حداقل مربعات مقایسه کنید. در برازش حداقل مربعات داریم:

$$\left.\begin{array}{l} \hat{y} = \hat{f}(\underline{x}) \\ \hat{f}(x) = \theta_1 f_1(x) + \cdots + \theta_p f_p(x) \\ A_{ij} = \hat{f}_j(x^{(i)}) \ , \ \begin{cases} 1 \leqslant i \leqslant N \\ 1 \leqslant j \leqslant P \end{cases} \end{array}\right\} \Rightarrow \hat{y} = A\theta$$

برای تبدیل مسأله رگرسیون به برازش حداقل مربعات، توابع پایه را به صورت زیر در نظر بگیرید:

$$f_1(x) = 1$$

$$f_i(x) = x_{i-1} \quad i = 2, \ldots, n+1$$

بنابراین $P = n+1$ بردار پایه داریم. در این حالت مدل رگرسیون را می‌توان به صورت زیر بیان کرد:

$$\hat{y} = x^T \theta_{2:n+1} + \theta_1$$

در نتیجه $\beta = \theta_{2:n+1}$ و $v = \theta_1$. با استفاده از تعاریف فوق می‌توان ماتریس A با ابعاد $N \times (n+1)$ را به صورت زیر تشکیل داد:

$$A = \left[\underline{1} \ \ \underline{X}^T \right]$$

که $\underline{X}$ ماتریس ویژگی بوده و ستون‌های آن به ترتیب $x^{(1)}$ و ...، $x^{(N)}$ است. با استفاده از تعاریف فوق می‌توان $\hat{\theta}$ را با روش حداقل مربعات به دست آورد و از رابطه پارامترهای رگرسیون با بردار $\hat{\theta}$ استفاده کرد و $\hat{\beta}$، $\hat{v}$ را مشخص نمود.

تبدیل برازش حداقل مربعات به رگرسیون

می‌دانیم در برازش حداقل مربعات رابطه زیر را داریم:

$$\hat{y}^{d} = \begin{bmatrix} \hat{y}^{(1)} \\ \vdots \\ \hat{y}^{(N)} \end{bmatrix} = \begin{bmatrix} \hat{f}_1(x^{(1)}) & \hat{f}_2(x^{(1)}) & \cdots & \hat{f}_p(x^{(1)}) \\ \vdots & & & \\ \hat{f}_1(x^{(N)}) & \hat{f}_2(x^{(N)}) & \cdots & \hat{f}_p(x^{(N)}) \end{bmatrix} \theta$$

فرض کنیم تابع پایه $\hat{f}_1(x)=1$ باشد. دراینصورت برای سطر اول دستگاه فوق، رابطه زیر را داریم:

$$\hat{y}^{(1)} = \tilde{x}^{(1)T}\theta_{2:p} + \theta_1$$

که $\tilde{x}^{(1)} = \left(\hat{f}_2(x^{(1)}), \dots, \hat{f}_p(x^{(1)})\right)$. حال رابطه فوق را با مسأله رگرسیون مقایسه کنید:

$$\hat{y}^{(1)} = x^T\beta + v$$

بنابراین می‌توانیم مسأله حداقل مربعات را معادل رگرسیون بدانیم وقتی بردار ویژگی از رابطه $\tilde{x} = \left(f_2(x), \dots, f_p(x)\right)$ به دست می‌آید.

طبقه‌بندی حداقل مربعات

در این فصل شرایطی را در نظر می‌گیریم که خروجی مطلوب مدل $\underline{y}^d$، تنها می‌تواند مقادیر گسسته را اختیار کند. می‌خواهیم تقریب حداقل مربعات در این شرایط را مورد بررسی قرار دهیم. در ساده‌ترین شکل بردار $\underline{y}^d$ تنها می‌تواند مقادیر ۱- یا ۱+ را اختیار کند. به مقدار هر درایه در بردار $\underline{y}^d$ طبقه (برچسب) و مسأله حاصل را طبقه‌بندی می‌نامیم.

برای حالت $y=-1$ یا $y=+1$ مسأله حاصل را مسأله طبقه‌بندی باینری بولین می‌نامیم. در این حالت فرض می‌کنیم $y=f(x)$ است که $f: \mathbb{R}^n \rightarrow \{-1,+1\}$ است. اگر تخمینی که از تابع f داریم را با $\hat{f}$ نشان دهیم، در این حالت $\hat{f}: \mathbb{R}^n \rightarrow \{+1,-1\}$ خواهد بود. $\hat{f}$ را طبقه‌بند می‌نامیم زیرا با دریافت یک ورودی $\underline{x}$ با بعد n، خروجی یکی از دو مقدار ۱+ یا ۱- را تولید می‌کند. به عبارت دیگر یک طبقه به بردار ورودی نسبت می‌دهد.

مثال‌ها: تشخیص ایمیل‌های هرز، تشخیص جعل، تشخیص بیماری

پاسخ حداقل مربعات برای مسأله طبقه‌بندی باینری: برای استفاده از تقریب حداقل مربعات فرض کنید شرط گسسته بودن y را نداریم. در این حالت با استفاده از حداقل مربعات پارامترها با استفاده از حداقل کردن عبارت زیر به دست می‌آیند:

$$\left(y^{(1)} - \tilde{f}(x^{(1)})\right)^{2}$$

که $\tilde{f}(x) = \sum_{i=1}^{p} \theta_i f_i(x)$، تقریب حداقل مربعات با یک داده بدون در نظر گرفتن فرض گسسته بودن y است. با تخمین زدن پارامترهای تابع $\tilde{f}$ به شیوه گفته شده در فصل قبل، مقدار $\hat{f}$ را می‌توان از رابطه زیر به دست آورد که با نام طبقه‌بند حداقل مربعات نیز شناخته می‌شود:

$$\hat{f}(\underline{x}) = \operatorname{sign}\left(\tilde{f}(\underline{x})\right)$$

طبقه‌بندی

همانطورکه در فصل قبل دیدیم، در این حالت تابع $\tilde{f}$ (تابعی که در آن گسسته بودن y در نظر گرفته نشده است) به صورت زیر است:

$$\tilde{f} = x^T\beta + v$$

می‌توانیم به روش گفته شده در فصل قبل بردار $\underline{\beta}$ و مقدار v را بدست آوریم. در این شرایط تابع $\hat{f}$ برابر خواهد بود با:

$$\hat{f}(x) = \text{sign}(x^T\beta + v)$$

توجه: با استفاده از رگرسیون می‌توان رابطه مستقیمی میان ویژگی‌های مختلف در بردار $\underline{x}$ و طبقه متناظر با آن به دست آورد. به عنوان مثال اگر β_2 منفی باشد، در نتیجه مقادیر بزرگ x_2 منجر به میل دادن $\hat{y}$ به سمت طبقه ۱- می‌گردد.

طبقه‌بندی چند برچسبی:

در این بخش حالتی را در نظر می‌گیریم که $y = f(x)$ بود. و $y \in \{1, 2, \dots, K\}$ است. این حالت را طبقه‌بندی چندگانه می‌نامیم. برای این حالت می‌خواهیم با استفاده از حداقل مربعات، طبقه‌بند را طراحی نماییم. به این منظور K حالت زیر را در نظر بگیرید و برای هر کدام یک طبقه‌بند باینری طراحی نمایید.

۱- برچسب جدید $\tilde{y}$ را به گونه‌ای تعریف کنید که

$$\tilde{y} = \begin{cases} +1 & \text{if } y = 1 \\ -1 & \text{if } y \neq 1 \end{cases}$$

برای این حالت، $\tilde{f}_1$ را طراحی نمایید.

⋮

M- برچسب جدید $\tilde{y}$ را به گونه‌ای تعریف کنید که:

$$\tilde{y} = \begin{cases} +1 & \text{if } y = M \\ -1 & \text{if } y = M \end{cases}$$

برای حالت فوق تابع $\tilde{f}_M$ را طراحی نمایید.

حال با داشتن توابع $\tilde{f}_۱$ ، $\tilde{f}_۲$ ، ... ، $\tilde{f}_M$ ، تابع طبقه‌بند چند کلاسه به صورت زیر قابل تعریف است:

$$\hat{f}(x) = \underset{k=1,\dots,M}{\arg\max}\ \tilde{f}_k(x)$$

مثال: مسأله طبقه‌بندی چند کلاسه زیر را در نظر بگیرید. برای این مسأله توابع $\tilde{f}_k(x)$ و $\hat{f}(x)$ را با در نظر گرفتن توابع پایه $f_۱(x) = x_۱x_۲$ ، $f_۲(x) = x_۱$ و $f_۳(x) = x_۲$ ، طراحی نمایید.

$\underline{x}^{(۱)} = (۲,۲) \quad y^{(۱)} = ۱$

$x^{(۲)} = (۳,۲) \quad y^{(۲)} = ۱$

$x^{(۳)} = (۰/۵,۰/۵) \quad y^{(۳)} = ۲$

$x^{(۴)} = (۰, ۰/۵) \quad y^{(۴)} = ۲$

$x^{(۵)} = (-۱, ۱) \quad y^{(۵)} = ۳$

$x^{(۶)} = (-۲, ۳) \quad y^{(۶)} = ۳$

طراحی $\tilde{f}_۱$: در این حالت اندیس‌های شماره ۱ و ۲ ، برچسب ۱+ و سایر اندیس‌ها ۱- در نظر گرفته می‌شوند.

$$A = \begin{bmatrix} ۴ & ۲ & ۲ \\ ۶ & ۳ & ۲ \\ ۰/۲۵ & ۰/۵ & ۰/۵ \\ ۰ & ۰ & ۰/۵ \\ -۱ & -۱ & ۱ \\ -۶ & -۲ & ۳ \end{bmatrix}$$

$$\Rightarrow \hat{\underline{\theta}}_۱ = A^{\dagger}\tilde{y}^d = A^{\dagger}\begin{bmatrix} +۱ \\ +۱ \\ -۱ \\ -۱ \\ -۱ \\ -۱ \end{bmatrix} = \begin{bmatrix} -۰/۰۲۷۵ \\ ۰/۴۹۴۷ \\ -۰/۱۳۹۳ \end{bmatrix}$$

به طور مشابه برای $\tilde{f}_۲$ و $\tilde{f}_۳$ داریم:

$$\hat{\underline{\theta}}_۲ = A^{\dagger}\tilde{y}^d = A^{\dagger}\begin{bmatrix} -۱ \\ -۱ \\ +۱ \\ +۱ \\ -۱ \\ -۱ \end{bmatrix} = \begin{bmatrix} -۰/۵۲۷۲ \\ ۱/۱۴۵۵ \\ -۰/۵۴۷۶ \end{bmatrix} \qquad \hat{\theta}_۳ = A^{\dagger}\tilde{y}^d = A^{\dagger}\begin{bmatrix} -۱ \\ -۱ \\ -۱ \\ -۱ \\ +۱ \\ +۱ \end{bmatrix} = \begin{bmatrix} ۰/۳۴۷۶ \\ -۱/۲۳۴۱ \\ ۰/۱۴۲۰ \end{bmatrix}$$

حال مسأله طبقه‌بندی، نگاشت داده جدید $x = \begin{bmatrix} ۳ \\ ۳ \end{bmatrix}$ را درنظر بگیرید. بر اساس روش طبقه‌بندی ارائه شده داریم:

$$\left.\begin{aligned} \tilde{f}_{۱}(x) &= [۹ \quad ۳ \quad ۳]\,\theta_{۱} = ۰٫۸۱۸۸ \\ \tilde{f}_{۲}(x) &= [۹ \quad ۳ \quad ۳]\,\theta_{۲} = -۲٫۹۵۰۷ \\ \tilde{f}_{۳}(x) &= [۹ \quad ۳ \quad ۳]\,\theta_{۳} = -۰٫۱۵۴۵ \end{aligned}\right\} \Rightarrow \hat{f}(x) = \arg\max_{i} \tilde{f}_{i}(x) = ۱$$

حداقل مربعات چندهدفه

در مسأله حداقل مربعات ما به دنبال بردار $\hat{x}$ بودیم که عبارت $\|Ax-b\|^2$ را حداقل می‌نماید. این عبارت را می‌توان به عنوان تابع هدف مسأله بهینه‌سازی در نظر گرفت و در نتیجه در شرایطی که بررسی کردیم مسأله تک هدفه بود. در مسأله حداقل مربعات چندهدفه به دنبال آن هستیم که تعدادی هدف همزمان کمینه گردند که این اهداف به صورت زیر هستند:

$$J_1 = \|A_1x-b_1\|^2$$
$$J_2 = \|A_2x-b_2\|^2$$
$$\vdots$$
$$J_k = \|A_kx-b_k\|^2$$

که در رابطه فوق A_i ماتریسی با ابعاد $m_i \times n$ و b_i برداری m_iتایی است.
اگر بخواهیم هرکدام از اهداف فوق را به طور جداگانه حداقل نماییم، می‌دانیم پاسخ $\hat{x}_i$ متناظر از رابطه مقابل به دست می‌آید:

$$\hat{x}_i = (A_i^TA_i)^{-1}A_i^Tb_i$$

اما در مسأله حداقل مربعات چندهدفه دنبال یک بردار $\hat{x}$ هستیم که بتواند ترکیب وزن‌دار از اهداف فوق به صورت زیر را حداقل نماید:

$$J = \lambda_1J_1 + \cdots + \lambda_kJ_k = \lambda_1\|A_1x-b_1\|^2 + \cdots + \lambda_k\|A_kx-b_k\|^2$$

که در رابطه فوق λ_iها ثابت‌های مثبت بوده و اهمیت هدف متناظر در J را نشان می‌دهند.
از آنجایی که مقیاس‌دهی در تابع هدف مسأله بهینه‌سازی نقطه جواب را عوض نمی‌کند، می‌توان تابع J را بر λ_1 تقسیم کرد و به تابع هدف زیر رسید:

$$J = \|A_1x-b_1\|^2 + \lambda_2\|A_2x-b_2\|^2 + \cdots + \lambda_k\|A_kx-b_k\|^2$$

که در تابع هزینه فوق عبارت اول معمولاً هدف اصلی در نظر گرفته می‌شود.

حال فرم کلی J به صورت زیر را در نظر بگیرید.

$$J = \lambda_1 \| A_1 x - b_1 \|^2 + \lambda_2 \| A_2 x - b_2 \|^2 + \cdots + \lambda_k \| A_k x - b_k \|^2$$

اگر بردار $r = \begin{bmatrix} \sqrt{\lambda_1}\,(A_1 x - b_1) \\ \vdots \\ \sqrt{\lambda_k}\,(A_k x - b_k) \end{bmatrix}$ را در نظر بگیریم به سادگی و براساس تعریف می‌توان نشان داد که:

$$J = \| r \|^2 = \left\| \begin{bmatrix} \sqrt{\lambda_1}\,(A_1 x - b_1) \\ \vdots \\ \sqrt{\lambda_k}\,(A_k x - b_k) \end{bmatrix} \right\|^2$$

از طرفی برای بردار r داریم (با استفاده از ضرب بلوکی)

$$r = \begin{bmatrix} \sqrt{\lambda_1}\, A_1 \\ \vdots \\ \sqrt{\lambda_k}\, A_k \end{bmatrix} x - \begin{bmatrix} \sqrt{\lambda_1}\, b_1 \\ \vdots \\ \sqrt{\lambda_k}\, b_k \end{bmatrix} = \tilde{A} x - \tilde{b}$$

که در رابطه فوق تعاریف زیر را داریم:

$$\tilde{A} \triangleq \begin{bmatrix} \sqrt{\lambda_1}\, A_1 \\ \vdots \\ \sqrt{\lambda_k}\, A_k \end{bmatrix}, \quad \tilde{A} : m \times n, \quad m = \sum_{i=1}^{k} m_i$$

$$\tilde{b} \triangleq \begin{bmatrix} \sqrt{\lambda_1}\, b_1 \\ \vdots \\ \sqrt{\lambda_k}\, b_k \end{bmatrix}, \quad \tilde{b} : m \times 1, \quad m = \sum_{i=1}^{k} m_i$$

بنابراین توانستیم مسأله حداقل مربعات چندهدفه را به مسأله حداقل مربعات معمولی تبدیل نماییم. مسأله حاصل به صورت زیر است:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \; \| \tilde{A} x - \tilde{b} \|^2$$

هزینه‌ای که برای این تبدیل پرداخت کردیم نیز افزایش ابعاد ماتریس $\tilde{A}$ و بردار $\tilde{b}$ است.
بنابراین پاسخ منحصر به فرد تقریب حداقل مربعات از رابطه زیر به دست می آید:

$$\hat{x} = (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b}$$

$$= (\lambda_1 A_1^T A_1 + \cdots + \lambda_k A_k^T A_k)^{-1}(\lambda_1 A_1^T b_1 + \cdots + \lambda_k A_k^T b_k)$$

توجه: پاسخ فوق به ازای $k=1$ به پاسخ شناخته شده برای تقریب حداقل مربعات مسأله یک هدفه تبدیل می گردد.

توجه: همانطور که در پاسخ حداقل مربعات دیدیم فرض ما براین است که ماتریس A دارای ستون های مستقل است. می توان نشان داد برای مستقل خطی بودن ستون های ماتریس $\tilde{A}$ کافی است فقط ستون های یکی از ماتریس های A_1، A_2، ...، A_k، مستقل خطی باشد.

توجه: ماتریس $\tilde{A}$ می تواند ستون های مستقل خطی داشته باشد حتی هنگامی که ستون های تمام ماتریس های A_1، A_2، ... و A_k وابسته خطی باشد. به عنوان یک حالت خاص فرض کنید برای تمام مقادیر i، $m_i < n$ باشد. در اینصورت ستون ها تمام ماتریس های A_i ها وابسته خطی است. اما ماتریس $\tilde{A}$ دارای ابعاد $m = \sum_{i=1}^{k} m_i$ است و چون $m > n$ است، ستون های $\tilde{A}$ می تواند مستقل خطی باشد.

منحنی مصالحه بهینه

حالت ساده شده مسأله حداقل مربعات دوهدفه زیر را در نظر بگیرید:

$$J = J_1 + \lambda J_2 = \|A_1x - b_1\|^2 + \lambda\|A_2x - b_2\|^2$$

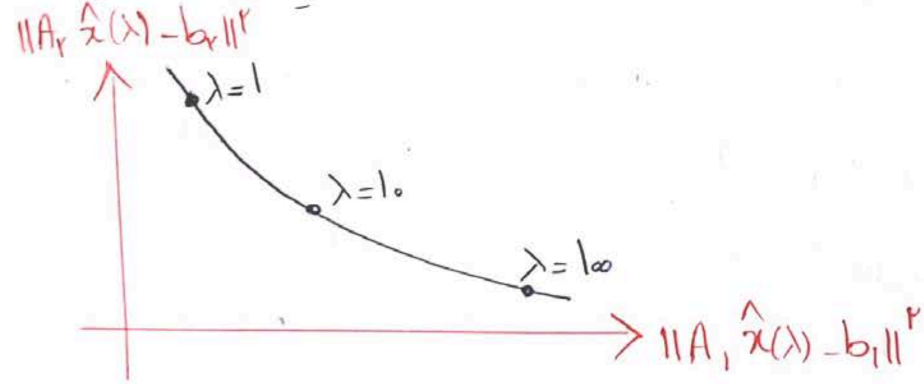
که $\lambda > 0$ اهمیت نسبی هر تابع هدف را مشخص می نماید.

فرض کنید $\hat{x}(\lambda)$ پاسخ حداقل مربعات چندهدفه $\hat{x}$ به صورت تابعی از λ باشد. همچنین فرض کنید ماتریس $\tilde{A}$ ستون‌های مستقل دارد. نقاط $\hat{x}(\lambda)$ با نام نقاط بهینه پارتو شناخته می‌شوند. یعنی بردار Z وجود ندارد به گونه‌ای که:

$$\|A_1 z - b_1\|^2 \leqslant \|A_1 \hat{x}(\lambda) - b_1\|^2 \text{ و } \|A_2 z - b_2\|^2 \leqslant \|A_2 \hat{x}(\lambda) - b_2\|^2$$

درحالیکه یکی از نامساوی‌ها اکید باشد. به عبارت دیگر بردار Z وجود ندارد که برای یکی از توابع هدف مقداری برابر $\hat{x}(\lambda)$ تولید کند و برای هدف دیگر مقدار آن کمتر از مقدار متناظر با $\hat{x}(\lambda)$ باشد.

اگر $\|A_1 \hat{x}(\lambda) - b_1\|^2$ را برحسب $\|A_2 \hat{x}(\ ) - b_2\|^2$ برای $\lambda \in (0,\infty)$ رسم نماییم. با منحنی به شکل زیر مواجه هستیم. این منحنی، منحنی مصالحه بهینه نامیده می‌شود. هیچ بردار z وجود ندارد که مقادیر J_1 و J_2 متناظر با آن پایین، سمت چپ منحنی مصالحه بهینه باشد.



کاربرد حداقل مربعات چند هدفه

کاربرد اصلی حداقل مربعات چند هدفه در اعمال نمودن اطلاعات پیشینی است که در ارتباط با بردار $\underline{x}$ وجود دارد. به عنوان نمونه ای از اطلاعات پیشین پرکاربرد می توان به موارد زیر اشاره کرد:

۱- $J_2 = \|x\|^2$ ← اطلاعات پیشین نشان دهنده طول کوچک بردار $\underline{x}$ است.

۲- $J_2 = \|\underline{x} - \underline{x}^{prior}\|^2$ ← اطلاعات پیشین نشان می دهد بردار $\underline{x}$ در حوالی بردار $\underline{x}^{prior}$ قرار دارد.

۳- $\|Dx\|^2$ که D ماتریس تفاضل مرتبه اول است ← اطلاعات پیشین نشان می دهد $\underline{x}$ برداری هموار است.

در موارد فوق عبارت J_2 با نام تنظیم کننده نیز شناخته می شود.

مثال: بردار x را به گونه ای بیابید که تابع $\|\underline{A}\underline{x}-\underline{y}\|^2+\lambda\|\underline{x}\|^2$ حداقل گردد ($\lambda>0$ است).

برای این مثال ماتریس $\tilde{A}$ برابر است با:

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\lambda}\, I \end{bmatrix} \qquad \tilde{b} = \begin{bmatrix} y \\ 0 \end{bmatrix} \Rightarrow \hat{x} = (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{b} = (A^TA+\lambda I)^{-1}A^Ty$$

توجه: مستقل از ماتریس A، ماتریس $\tilde{A}$ در رابطه فوق همواره دارای ستون های مستقل خطی است زیرا:

$$\tilde{A}\underline{x} = (A, \sqrt{\lambda}\, I)\underline{x} = (A\underline{x}, \sqrt{\lambda}\,\underline{x}) = 0 \Rightarrow \underline{x} = 0 \Rightarrow$$ ستون های $\tilde{A}$ مستقل خطی است.

توجه: تنظیم‌کننده فوق را تنظیم‌کننده تیخونوف (Tikhonov regularization) می‌نامند.

برازش منحنی تنظیم‌شده

در فصل قبل با برازش منحنی براساس حداقل مربعات آشنا شدیم. می‌توان برای افزودن اطلاعات بیشتر درارتباط با بردار $\underline{\theta}$ از تنظیم‌کننده استفاده نمود. فرض کنید مدل برازش به صورت زیر را داریم:

$$\hat{f}(x) = \theta_1 f_1(x) + \cdots + \theta_p f_p(x)$$

با نحوه به‌دست‌آوردن θ_i ها با استفاده از پایگاه داده قبلاً آشنا شدیم. می‌توان مقدار پارامتر θ_i را میزان وابستگی تخمین $\hat{f}(x)$ به پایه $f_i(x)$ دانست. بنابراین اگر مقدار θ_i خیلی بزرگ شود امکان دارد مدل نسبت به نویزهای موجود در پایه f_i به ازای بردار دلخواه $\underline{x}$ حساس گردد. بنابراین نیاز داریم که مقدار θ_i ها تا حدامکان کوچک باشد (توجه شود که اگر $f_i(x)=1$ است نیازی نیست اندازه θ_i کنترل شود زیرا نویز بردار $\underline{x}$ در تابع ثابت $f_i(x)=1$ اثری ندارد). بنابراین می‌توانیم مسأله زیر را درنظر بگیریم:

$$\hat{\theta} = \underset{\theta}{\arg\min} \; \|y - A\theta\|^2 + \lambda \|\theta_{2:p}\|^2 \quad , \lambda > 0$$

رگرسیون تنظیم شده

به طور مشابه در مسأله رگرسیون نیز می‌توان دامنه پارامترها را به صورت زیر کنترل نمود:

$$\{\hat{\beta}, \hat{v}\} = \underset{\beta, v}{\arg\min} \; \|y - X^T\beta - v\mathbf{1}\|^2 + \lambda\|\beta\|^2$$

که مسأله فوق نیز اثری مشابه برازش منحنی تنظیم شده دارد.

مثال: پایگاه داده زیر را در نظر بگیرید.

$x^{(۱)} = ۱ \quad y^{(۱)} = ۲,۳$

$x^{(۲)} = ۲ \quad y^{(۲)} = ۳,۴$

$x^{(۳)} = ۳ \quad y^{(۳)} = ۶,۸$

$x^{(۴)} = ۴ \quad y^{(۴)} = ۷$

$x^{(۵)} = ۵ \quad y^{(۵)} = ۱۱,۵$

فرض شد داده‌های فوق از مدل $y = ۲x + v$ تولید شده‌اند که v نویز اندازه‌گیری است. با استفاده از برازش منحنی مبتنی بر حداقل مربعات و تنظیم شده تیخونوف به ازای $\lambda = ۰$ ، $\lambda = ۰/۵$ و $\lambda = ۱۰$ ، یک چندجمله‌ای درجه ۴ بر داده‌ها برازش نمایید.

در این مسأله ماتریس A برابر است با:

$$A = \begin{bmatrix} ۱ & ۱ & ۱ & ۱ & ۱ \\ ۱ & ۲ & ۴ & ۸ & ۱۶ \\ ۱ & ۳ & ۹ & ۲۷ & ۸۱ \\ ۱ & ۴ & ۱۶ & ۶۴ & ۲۵۶ \\ ۱ & ۵ & ۲۵ & ۱۲۵ & ۶۲۵ \end{bmatrix}$$

ماتریس $\tilde{A}$ برابر است با:

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\lambda}\ I \end{bmatrix}$$

بردار y و $\tilde{b}$ برابر است با:

$$y = \begin{bmatrix} ۲,۳ \\ ۳,۴ \\ ۶/۸ \\ ۷ \\ ۱۱,۵ \end{bmatrix}$$

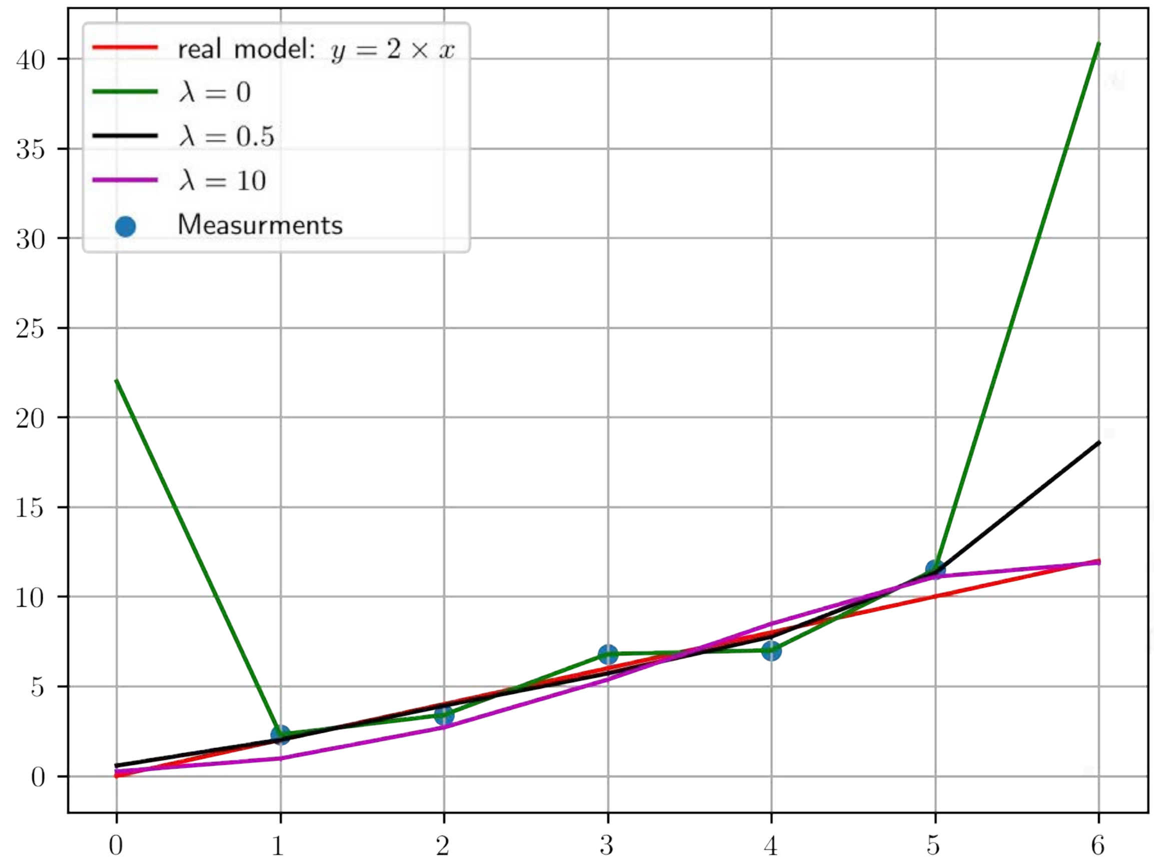
$$\tilde{b} = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

تعریف حداقل مربعات تنظیم شده با تیخونوف برابر است با:

$$\hat{\theta} = (A^T A + \lambda I)^{-1} A^T y$$
$$= (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T \tilde{b}$$

$$\lambda = ۰ \Rightarrow \hat{\theta} = \begin{bmatrix} ۲۲/۰۰ \\ -۳۹/۵۲ \\ ۲۵/۶۱ \\ -۶/۳۳ \\ ۰/۵۴ \end{bmatrix} \qquad \lambda = ۰/۵ \Rightarrow \hat{\theta} = \begin{bmatrix} ۰/۵۷ \\ ۰/۸۱ \\ ۰/۹۲ \\ -۰/۳۲ \\ ۰/۰۴ \end{bmatrix} \qquad \lambda = ۱۰ \Rightarrow \hat{\theta} = \begin{bmatrix} ۰/۲۴ \\ ۰/۳۱ \\ ۰/۳۴ \\ ۰/۱۰ \\ -۰/۰۲ \end{bmatrix}$$

نمودار حاصل از تخمین‌های مختلف در شکل زیر نشان داده شده است. همانطور که مشخص است تنظیم کردن می‌تواند باعث بهبود عملکرد در نقاط مختلف و کاهش فاصله تخمین از مدل واقعی داده‌ها در مقایسه با مدل حداقل مربعات بدون تنظیم شدن گردد



حداقل مربعات مقید

مسأله حداقل مربعات مقید خطی به صورت زیر بیان می‌شود :

$$\min_{x} \|A\underline{x} - b\|^2$$
$$\text{s.t.} \quad C\underline{x} = \underline{d}$$

که در رابطه فوق $\underline{x} \in \mathbb{R}^n$ ، $A \in \mathbb{R}^{m\times n}$ ، $b \in \mathbb{R}^m$ ، $C \in \mathbb{R}^{p\times n}$ ، $d \in \mathbb{R}^p$ است.

در مسأله فوق تابع $\|Ax - b\|^2$ تابع هدف مسأله و مجموعه قیدهای $Cx = d$ قیدهای مسأله نامیده می‌شوند. این قیدها را می‌توان به صورت منحصر زیر نیز نوشت.

$$c_i^T x = d_i , \quad i = 1, 2, \dots, p$$

که در رابطه فوق c_i^T سطر i-ام ماتریس C است.

بردار x مجاز: بردار $\underline{x}$ را مجاز گوییم اگر قیدها را برآورده نماید یعنی $Cx = d$ باشد.

بردار بهینه یا پاسخ: بردار $\hat{\underline{x}}$ را بردار بهینه یا پاسخ نامیم اگر $\hat{\underline{x}}$ مجاز بوده و به ازای تمام بردارهای مجاز دیگر $\underline{x}$ ، $\|A\hat{x} - b\|^2 \leq \|Ax - b\|^2$

تفسیر: مسأله حداقل مربعات مقید در واقع دو مسأله حل دستگاه معادلات خطی و حل مسأله حداقل مربعات را به صورت همزمان انجام می‌دهد.

پاسخ تقریبی: مسأله حداقل مربعات مقید را می‌توان حالت حدی یک مسأله حداقل مربعات چند هدفه در نظر گرفت که به صورت زیر است :

$$\min_{x} \|Ax - b\|^2 + \lambda \|Cx - d\|^2$$

و λ یک ثابت بسیار بزرگ است. پاسخ مسأله فوق می‌تواند تقریبی از پاسخ مسأله حداقل مربعات مقید باشد.

مثال: مسأله برازش تکه‌ای چندجمله‌ای تابع $\hat{f}(x)$ با استفاده از مجموع N نقطه (x_i, y_i) را در نظر بگیرید. تابع $\hat{f}(x)$ را به صورت زیر تعریف می‌کنیم:

$$\hat{f}(x) = \begin{cases} p(x) & x \leqslant a \\ q(x) & x > a \end{cases}$$

که در این مسأله فرض می‌کنیم $p(x)$ و $q(x)$ چندجمله‌ای‌هایی از درجه حداکثر ۳ هستند یعنی:

$$p(x) = \theta_{۱} + \theta_{۲} x + \theta_{۳} x^{۲} + \theta_{۴} x^{۳}$$

$$q(x) = \theta_{۵} + \theta_{۶} x + \theta_{۷} x^{۲} + \theta_{۸} x^{۳}$$

یک ویژگی مطلوب در برازش تکه‌ای - چندجمله‌ای، پیوستگی در نقطه تغییر برازش یعنی نقطه $x = a$ است. این پیوستگی می‌تواند پیوستگی خود تابع و مشتقات آن تا هر مرتبه‌ای باشد. برای این مثال فرض کنید پیوستگی تابع یعنی $p(a) = q(a)$ و پیوستگی مشتق یعنی $p'(a) = q'(a)$ مدنظر است. در اینصورت مجموع مربعات خطا برابر است با:

$$\sum_{i=1}^{M} (\theta_{۱} + \theta_{۲} x_i + \theta_{۳} x_i^{۲} + \theta_{۴} x_i^{۳} - y_i) + \sum_{i=M+1}^{N} (\theta_{۵} + \theta_{۶} x_i + \theta_{۷} x_i^{۲} + \theta_{۸} x_i^{۳} - y_i)^{۲}$$

از شرط $p(a) - q(a) = ۰$ داریم:

$$\theta_{۱} + \theta_{۲} a + \theta_{۳} a^{۲} + \theta_{۴} a^{۳} - \theta_{۵} - \theta_{۶} a - \theta_{۷} a^{۲} - \theta_{۸} a^{۳} = ۰$$

از شرط $p'(a) - q'(a) = ۰$ نیز داریم:

$$\theta_{۲} + ۲\theta_{۳} a + ۳\theta_{۴} a^{۲} - \theta_{۶} - ۲\theta_{۷} a - ۳\theta_{۸} a^{۲} = ۰$$

حال باید مجموع مربعات و قیدها را به فرم ماتریسی بیان نماییم. با توجه به مطالب فصل‌های گذشته رابطه ماتریسی زیر را داریم:

$$\min \quad \|A\theta - d\|^2 \qquad s.t. \quad C\theta = d$$

که :

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & 0 & 0 & 0 & 0 \\ 1 & x_2 & x_2^2 & x_2^3 & 0 & 0 & 0 & 0 \\ \vdots \\ 1 & x_M & x_M^2 & x_M^3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{M+1} & x_{M+1}^2 & x_{M+1}^3 \\ \vdots \\ 0 & 0 & 0 & 0 & 1 & x_N & x_N^2 & x_N^3 \end{bmatrix} \qquad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \\ y_{M+1} \\ \vdots \\ y_N \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & a & a^2 & a^3 & -1 & -a & -a^2 & -a^3 \\ 0 & 1 & 2a & 3a^2 & 0 & -1 & -2a & -3a^2 \end{bmatrix} \qquad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

توابع $\hat{f}$ به فرم تکه‌ای چندجمله‌ای با نام توابع اسپلاین (Splines) شناخته می‌شوند.

مسأله حداقل نرم : مسأله حداقل نرم یکی دیگر از مسائلی است که می‌توان آن را به صورت حداقل مربعات مقید فرمول‌بندی کرد. مسأله حداقل نرم به صورت زیر است :

$$\min_{x} \quad \|x\|^2$$
$$s.t. \quad Cx = d$$

همانطور که مشخص است مسأله فوق حالت خاص مسأله حداقل مربعات به ازای $A = I$ و $b = 0$ است. در این مسأله ما به دنبال جوابی از دستگاه $Cx = d$ هستیم که طول آن کمینه باشد.

پاسخ مسأله حداقل مربعات مقید :

در این بخش پاسخ مسأله حداقل مربعات متغیر را با استفاده از ضرایب لاگرانژ به دست می‌آوریم. ما در این درس به روش ضرایب لاگرانژ به عنوان یک ابزار نگاه می‌کنیم و مستقیماً به خود این ابزار نمی‌پردازیم.

برای استفاده از ضرایب لاگرانژ، قیدها را به صورت جداگانه زیر در نظر می‌گیریم:

$$\min \quad \|Ax_i - b\|^2$$
$$\text{s.t.} \quad c_i^T x = d_i \quad i = 1, \dots, p$$

تابع لاگرانژ را به صورت زیر تشکیل می‌دهیم:

$$L_1(x,z) = \|Ax - b\|^2 + z_1(c_1^T x - d_1) + \dots + z_p(c_p^T x - d_p)$$

که $z = (z_1, \dots, z_p)$ بردار ضرایب لاگرانژ است.

بر اساس روش ضرایب لاگرانژ، $\hat{x}$ پاسخ مسئله حداقل مربعات مقید است اگر یک دسته از ضرایب لاگرانژ $\hat{z}$ وجود داشته باشد به گونه‌ای که:

$$\frac{\partial L_1}{\partial x_i}(\hat{x}, \hat{z}) = 0 \quad i = 1, \dots, n$$

$$\frac{\partial L_1}{\partial z_i}(\hat{x}, \hat{z}) = 0 \quad i = 1, \dots, p$$

شرایط فوق را شرایط بهینگی برای مسئله حداقل مربعات مقید می‌نامیم. هر پاسخ مسئله حداقل مربعات مقید باید شرایط فوق را برآورده نماید.

حال نشان می‌دهیم شرایط بهینگی را می‌توان به صورت مجموعه‌ای از معادلات خطی نوشت.

$$\frac{\partial L_1}{\partial z_i}(\hat{x}, \hat{z}) = c_i^T \hat{x} - d_i = 0, \quad i = 1, \dots, p \qquad (1)$$

رابطه فوق بیان می‌کند که $C\hat{x} - d$ است (که از قبل می‌دانستیم). اما بر اساس رابطه اول داریم:

$$\frac{\partial L}{\partial x_i}(\hat{x},\hat{z}) = ۲\sum_{j=1}^{n}(A^TA)_{ij}\hat{x}_j - ۲(A^Tb)_i + \sum_{j=1}^{p}\hat{z}_j(c_j)_i = ۰$$

رابطه فوق را می‌توان به صورت ماتریسی زیر بازنویسی کرد:

$$۲(A^TA)\hat{x} - ۲A^Tb + C^T\hat{z} = ۰ \qquad (۲)$$

اگر رابطه (۱) و (۲) را در یک دستگاه قرار دهیم داریم:

$$\begin{bmatrix} ۲A^TA & C^T \\ C & ۰ \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} ۲A^Tb \\ d \end{bmatrix}$$

معادلات فوق با نام معادلات KKT برای مسأله حداقل مربعات مقید شناخته می‌شوند این معادلات توسعه معادلات نرمال . بنابراین -- حداقل مربعات مقید به حل دستگاه مربعی با ابعاد n+p تبدیل گردید.

معکوس پذیری ماتریس KKT : ماتریس ضرایب دستگاه معادلات خطی فوق که دارای ابعاد $(n\ p)X(n+p)$ است را ماتریس KKT می‌نامیم. این ماتریس معکوس پذیر است اگر و فقط اگر سطرهای ماتریس C مستقل خطی باشند و ستون‌های ماتریس $\begin{bmatrix} A \\ C \end{bmatrix}$ مستقل خطی باشند.

نکته : شرط استقلال خطی سطرهای C نتیجه می‌دهد که تعداد قیدها باید کوچکتر یا مساوی تعداد متغیرها باشد ($p \leqslant n$).

شرط دوم به A و C همزمان وابسته است و در شرایطی که A ستون‌های مستقل خطی ندارد نیز می‌تواند برآورده شود.

اثبات: اگر سطرهای ماتریس C و ستون های ماتریس $\begin{bmatrix}A\\C\end{bmatrix}$ مستقل خطی باشند، ماتریس KKT معکوس پذیر است.

فرض کنید ماتریس KKT معکوس پذیر نیست. در این صورت بردار غیر صفر $(\bar{x},\bar{z})$ وجود دارد به گونه‌ای که:

$$\begin{bmatrix} ۲A^TA & C^T \\ C & ۰ \end{bmatrix}\begin{bmatrix}\bar{x} \\ \bar{z}\end{bmatrix} = ۰$$

اگر رابطه بلوکی سطر اول را از سمت چپ در $\bar{x}^T$ ضرب کنیم داریم:

$$۲\|A\bar{x}\|^۲ + \bar{x}^T C^T \bar{z} = ۰$$

از طرفی می‌دانیم طبق رابطه بلوکی سطر دوم، $C\bar{x}=۰$ یا $\bar{x}^T C^T=۰$ است، یعنی داریم:

$$۲\|A\bar{x}\|^۲ = ۰ \Rightarrow A\bar{x} = ۰$$

اگر دو رابطه $C\bar{x}=۰$, $A\bar{x}=۰$ را به صورت بلوکی بنویسیم رابطه مقابل را داریم.

$$\begin{bmatrix}A\\C\end{bmatrix}\bar{x} = ۰$$

چون بنابر فرض ستون های ماتریس $\begin{bmatrix}A\\C\end{bmatrix}$ مستقل خطی هستند در نتیجه $\bar{x}=۰$ است.

اگر $\bar{x}=۰$ باشد، رابطه بلوکی سطر اول به صورت زیر تبدیل می‌شود:

$$C^T\bar{z} = ۰$$

چون سطرهای C یا ستون های C^T مستقل خطی هستند در نتیجه $\bar{z}=۰$.

بنابراین $(\bar{x},\bar{z})=۰$ بوده که در تناقض با فرض است.

اگر ماتریس KKT معکوس‌پذیر باشد، سطرهای ماتریس C، ستون‌های ماتریس $\begin{bmatrix} A \\ C \end{bmatrix}$ مستقل خطی است.

فرض کنید سطرهای C وابسته خطی باشد. در اینصورت بردار غیرصفر $\bar{z}$ وجود دارد به گونه‌ای که $C^T\bar{z}=0$ است. یعنی:

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \bar{z} \end{bmatrix}=0$$

رابطه بالا نشان می‌دهد ماتریس KKT معکوس‌پذیر نیست.

حال فرض کنید ستون‌های ماتریس $\begin{bmatrix} A \\ C \end{bmatrix}$ وابسته خطی باشند. در اینصورت بردار غیرصفر $\bar{x}$ وجود دارد که:

$$\begin{bmatrix} A \\ C \end{bmatrix}\bar{x}=0$$

می‌توان نشان داد بردار $\begin{bmatrix} \bar{x} \\ 0 \end{bmatrix}$ در رابطه زیر صادق است:

$$\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix}\begin{bmatrix} \bar{x} \\ 0 \end{bmatrix}=\underline{0}$$

که نشان می‌دهد ماتریس KKT معکوس‌پذیر نیست.

توجه: اگر سطرهای C و ستون‌های $\begin{bmatrix} A \\ C \end{bmatrix}$ مستقل خطی باشند، مسأله حداقل مربعات مقید جواب منحصر به فرد دارد و این جواب از رابطه زیر به دست می‌آید:

$$\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix}=\begin{bmatrix} 2A^TA & C^T \\ C & 0 \end{bmatrix}^{-1}\begin{bmatrix} 2A^Tb \\ d \end{bmatrix}$$

همانطور که مشاهده می‌کنید پاسخ ترکیب خطی b و d است.

بررسی پاسخ حداقل مربعات مقید:

در این بخش می‌خواهیم بدون استفاده از روش ضرایب لاگرانژ نشان دهیم پاسخ منحصربه‌فرد به دست آمده، حداقل کننده تابع هزینه و مجاز است.

فرض کنید $\hat{x}$ و $\hat{z}$ پاسخ‌های به دست آمده از رابطه صفحه قبل باشند در اینصورت:

$$\begin{cases} 2A^TA\hat{x} + C^T\hat{z} = 2A^Tb \\ C\hat{x} = d \end{cases}$$

فرض کنید $x \neq \hat{x}$ بردار دلخواه مجاز است (یعنی $Cx=d$). نشان می‌دهیم $\|Ax-b\|^2 > \|A\hat{x}-b\|^2$.

$$\begin{aligned}\|Ax-b\|^2 &= \|(Ax-A\hat{x})+(A\hat{x}-b)\|^2 \\ &= \|Ax-A\hat{x}\|^2 + \|A\hat{x}-b\|^2 + 2(Ax-A\hat{x})^T(A\hat{x}-b)\end{aligned}$$

حال عبارت سوم در رابطه فوق را توسعه می‌دهیم:

$$\begin{aligned}2(Ax-A\hat{x})^T(A\hat{x}-b) &= 2(x-\hat{x})^TA^T(A\hat{x}-b) \\ &= -(x-\hat{x})^TC^T\hat{z} \\ &= -(C(x-\hat{x}))^T\hat{z} \\ &= 0\end{aligned}$$

بنابراین:

$$\|Ax-b\|^2 = \|A(x-\hat{x})\|^2 + \|A\hat{x}-b\|^2$$

در نتیجه $\hat{x}$ حداقل کننده $\|Ax-b\|^2$ مشروط به $Cx=d$ است.

توجه: $\|A(x-\hat{x})\|^2 > 0$ است زیرا اگر اینطور نباشد $A(x-\hat{x})=0$ است. در نتیجه چون $C(x-\hat{x})=0$ است بنابراین رابطه زیر را داریم:

$$\begin{bmatrix} A \\ C \end{bmatrix}(x-\hat{x}) = 0$$

رابطه فوق با فرض معکوس پذیر بودن ماتریس KKT در تناقض است.

حداقل مربعات غیرخطی

تا این بخش از درس همواره با یک دستگاه معادلات خطی مواجه بودیم. در این فصل می‌خواهیم شرایطی را در نظر بگیریم که با دستگاه معادلات غیرخطی مواجه هستیم

تعریف: دستگاه معادلات غیرخطی: دستگاه معادلات با m معادله غیرخطی و n مجهول را در نظر بگیرید که مجهولات در بردار $\underline{x} \in \mathbb{R}^n$ قرار دارند، و معادلات به صورت زیر هستند:

$$f_i(\underline{x}) = 0 \quad i=1,\dots,m$$

که توابع f_i توابعی با خروجی عدد، به صورت $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ هستند.

توجه: در اینجا می‌توانستیم معادلاتی به صورت $f_i(\underline{x}) = b_i$ در نظر بگیریم اما به دلیل ماهیت غیرخطی f_i، ثابت b_i را نیز به سمت چپ تساوی می‌بریم و در واقع $f_i(\underline{x})$ خود نشان دهنده باقیمانده i-ام یعنی r_i است.

توجه: دستگاه معادلات را می‌توانیم به صورت برداری زیر نیز بازنویسی نماییم:

$$\underline{f}(\underline{x}) = \underline{0}$$

که $\underline{f}(\underline{x}) = \left(f_1(\underline{x}), \dots, f_m(\underline{x})\right)$ است. در اینصورت $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ است.

هدف مسائل حداقل مربعات غیرخطی: هدف پیدا کردن بردار $\underline{x}$ به گونه‌ای است که ماندهٔ متناظر با آن صفر گردد.

توجه: اگر $\underline{f}$ یک تابع آفاین باشد $(\underline{f} = A\underline{x} - b)$ در اینصورت با مسائل فصل‌های گذشته مواجهیم. در این فصل می‌خواهیم حالتی را در نظر بگیریم که f آفاین نبوده و غیرخطی است.

مسأله حداقل مربعات غیرخطی : با توجه به تعاریف گفته شد، مسأله حداقل مربعات غیرخطی به صورت زیر قابل تعریف است :

$$\min_{\underline{x}} \|f(x)\|^2$$

توجه: برای حل دستگاه $\underline{f}(\underline{x})=\underline{0}$ می‌بایست مقدار حداقل تابع فوق صفر گردد. در این شرایط بردار حاصل $\hat{\underline{x}}$ پاسخ دستگاه معادلات غیرخطی و پاسخ حداقل مربعات غیرخطی خواهد بود. اما در بیشتر مواقع پیدا کردن پاسخ دستگاه معادلات غیرخطی امکان پذیر نیست اما می‌توان با استفاده از مسأله حداقل مربعات غیرخطی تقریبی از پاسخ دستگاه معادلات در معیار حداقل مربعات را به دست آورد که می‌تواند در بسیاری از کاربردها مورد استفاده قرار گیرد.

شرایط لازم برای بهینگی : یک شرط لازم برای اینکه $\hat{\underline{x}}$ پاسخ مسأله حداقل مربعات غیر خطی باشد این است که :

$$\frac{\partial}{\partial x_i}\|f(\hat{x})\|^2 = 0 \qquad i=1,\dots,n$$

توجه: شرط فوق شرط لازم است. بنابراین می‌تواند برای نقاطی که حداقل کننده مسأله حداقل مربعات غیرخطی نیستند نیز برقرار باشد.

توجه: اگر شرایط بهینگی را به طور همزمان برای ابعاد مختلف بردار $\hat{\underline{x}}$ در نظر بگیریم، به رابطه $\nabla\|f(\hat{x})\|^2=\underline{0}$ می‌رسیم. این گرادیان برای بردار دلخواه x به صورت زیر قابل محاسبه است :

$$\nabla\|f(x)\|^2 = \nabla\left(\sum_{i=1}^{m} f_i(x)^2\right) = 2\sum_{i=1}^{m} f_i(x)\,\nabla f_i(x) = 2(Df(x))^T f(x)$$

که $Df(x)$ ماتریس ژاکوبین بوده و به صورت زیر تعریف می‌شود :

$$\left[D\underline{f}(\underline{z})\right]_{ij} = \frac{\partial f_i(z)}{\partial x_j}$$

نتیجه: اگر $\hat{x}$ حداقل کننده مسأله $\min_{x} \|f(x)\|^2$ باشد باید رابطه زیر برقرار باشد:

$$\left(Df(\hat{x})\right)^T f(\hat{x}) = 0$$

یعنی $f(\hat{x})$ در فضای پوچی ماتریس $\left(Df(\hat{x})\right)^T$ است یا $f(\hat{x})$ در فضای پوچی چپ ماتریس $Df(\hat{x})$ است.

حداقل مربعات غیرخطی در مقایسه با حداقل مربعات خطی:

حل کردن مسأله حداقل مربعات غیرخطی به طور کلی سخت‌تر از مسأله حداقل مربعات خطی است زیرا:

- برای مسأله حداقل مربعات غیرخطی امکان دارد صفر، یک یا بینهایت جواب وجود داشته باشد برخلاف مسأله حداقل مربعات خطی، تعیین اینکه کدام حالت در یک مسأله برقرار است، کار دشواری است.
- راه‌حل‌هایی برای تعیین دقیق پاسخ دستگاه معادلات غیرخطی وجود دارد اما این راه‌حل‌ها نیازمند توان محاسباتی بسیار بالا هستند و در کاربردهای عملیاتی مورد استفاده قرار نمی‌گیرند.
- روش‌هایی که برای حل دستگاه معادلات خطی دیدیم همواره پاسخ این دستگاه را تولید می‌کردند اما الگوریتم‌هایی که در این فصل برای حل دستگاه معادلات غیرخطی در معیار نرم دوم مانده بررسی خواهیم نمود به دنبال برداری مانند $\hat{x}$ هستند به گونه‌ای که شرایط بهینگی را داشته باشد. چنین نقطه‌ای حتی اگر پیدا شود جواب دستگاه، حداقل مربعات غیرخطی نخواهد بود.

- کاربردهای حداقل مربعات غیرخطی :

۱) پیدا کردن تعادل نش : یک بازی ریاضی را درنظر بگیرید که در آن n شرکت کننده. یک عدد x_i درنظر می‌گیرند هرشرکت کننده. براساس عدد انتخابی خود و دیگران یک جایزه دریافت می‌کند که مقدار آن برای فرد i-ام با تابع $R_i(x)$ به دست می‌آید و تابع بازده نامیده می‌شود. هرفرد تلاش می‌کند جایزه خود را بیشینه کند اما چون جایزه وی تنها تابع انتخاب خود او نیست مسأله پیچیده می‌گردد. نقطه تعادل نش جایی است که هیچکدام از اعضا نمی‌توانند با تغییر انتخاب خود، جایزه خود را افزایش دهند. این نقطه پایدار است زیرا در چنین جایی افراد تمایلی به تغییر انتخاب خود ندارند. در نقطه تعادل نش $\hat{x}$ رابطه زیر برقرار است :

$$\frac{\partial\ i}{\partial x_i}(\hat{x}) = 0 , \quad i = 1, \dots, n$$

بنابراین نقطه تعادل نش از یک دستگاه معادلات غیرخطی مربعی به دست می‌آید.

۲) پیدا کردن اجسام در فضای ۲ و ۳ بعدی : فرض کنید یک هدف در موقعیت x دارید و می‌خواهید آن را به دست آورید. ابزاری که دارید m رادار هستند که در موقعیت‌های a_i قرار دارند و می‌توانند با فعال شدن، فاصله خود از هدف را با یک خطای اندازه‌گیری به صورت $\rho_i = \|x - a_i\| + v_i$ تخمین نمایند (v_i خطای اندازه‌گیری رادار i-ام است). تخمین ما از موقعیت که با $\hat{x}$ نمایش می‌دهیم می‌تواند از طریق حداقل سازی مجموع مربعات مانده‌های زیر به دست آید.

$$\min_{x} \sum_{i=1}^{m} \left(\|x - a_i\| - \rho_i \right)^2$$

توجه شود که از مسأله فوق می‌توان برای تعیین موقعیت با استفاده از GPS نیز استفاده کرد که در آن a_i موقعیت معلوم ماهواره‌های GPS است که می‌تواند هدف را ببینند.

- **برازش منحنی غیرخطی** : مدلی به صورت $y \simeq \hat{f}(x;\theta)$ درنظر بگیرید که $\underline{x}$ یک بردار ویژگی و y خروجی عددی است. $\hat{f}$ تخمینی از مدل واقعی تولیدکننده داده‌ها است که دارای پارامترهای قرارگرفته در بردار $\underline{\theta}$ است (قبلاً با حالتی که $\hat{f}$ تبدیل امایی از پارامترها بود آشنا شدیم). فرض شد $\hat{f}$ تابعی غیرخطی از پارامترها است و شما می‌خواهید با استفاده از یک مجموعه‌داده شامل N نقطه، پارامترهای مدل را از طریق حداقل مربعات به دست آورید. در این شرایط باید تابع زیر را حداقل نمایید:

$$\min_{\theta} \sum_{i=1}^{N} \left(\hat{f}(x^{(i)};\underline{\theta}) - y^{(i)} \right)^2$$

همانطور که مشخص است مسأله فوق یک مسأله حداقل مربعات غیرخطی است.

- **الگوریتم‌های حل مسأله** : در این بخش با دو الگوریتم حل مسأله حداقل مربعات غیرخطی آشنا می‌شویم. این دو الگوریتم هر دو الگوریتم‌های مبتنی بر تکرار هستند و در خروجی خود یک دنباله از نقاط $\underline{x}^{(1)}$، $\underline{x}^{(2)}$، ... را تولید می‌کنند و تعاریف زیر را در مورد آنها داریم:
  - $\underline{x}^{(1)}$ نقطه شروع الگوریتم است.
  - $x^{(k)}$ خروجی k-ام الگوریتم است.
  - تبدیل $x^{(k)}$ به $x^{(k+1)}$ طی فرآیندی با نام تکرار انجام می‌شود.
  - ارزیابی هر تکرار براساس مقدار $\|f(x^{(k)})\|^2$ صورت می‌پذیرد.
  - الگوریتم‌ها به گونه‌ای دنباله خروجی خود را تولید می‌کنند که رابطه $\|f(x^{(k)})\| \geqslant \|f(x^{(k+1)})\|$ معمولاً برقرار است.

- الگوریتم ها زمانی متوقف می‌شوند که $\| f(x^{(k)}) \|$ به اندازه کافی کوچک گردد یا فاصله میان دو نقطه $x^{(k)}$ و $x^{(k+1)}$ به اندازه کافی کوچک شود یا حداکثر تعداد تکرارها انجام شود.

- **الگوریتم گاوس-نیوتن**: ایده اصلی در این الگوریتم استفاده همزمان از دو ابزار حساب دیفرانسیل و جبر خطی است. به این منظور این الگوریتم از دو گام زیر بهره می‌برد.

۱- حساب دیفرانسیل ← دستگاه معادلات غیرخطی $\underline{f}(\underline{x})=0$ را حول نقطه $x=x^{(k)}$ بسط تیلور دهید:

$$\hat{f}(\underline{x}, x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x-x^{(k)})$$

که $Df(x^{(k)})$ ماتریس ژاکوبین با ابعاد $m \times n$ است. اگر $\|x-x^{(k)}\|$ کوچک باشد (برای $\underline{x}$ های به اندازه کافی نزدیک $x^{(k)}$)، تقریب فوق مناسب است.

۲- جبر خطی ← تقریب دستگاه معادلات غیرخطی یک دستگاه معادلات خطی است. با فرض استقلال ستون‌های ماتریس $Df(x^{(k)})$ پاسخ این دستگاه، که به عنوان $x^{(k+1)}$ نیز نامگذاری می‌شود برابر است با:

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

توجه: استقلال ستون‌های $Df(x^{(k)})$ لازم می‌دارد که $m \geqslant n$ باشد.

ـ شبه کد الگوریتم گاوس ـ نیوتن :

← ورودی ها :

- * تابع مشتق پذیر $f: \mathbb{R}^n \to \mathbb{R}^m$
- * نقطه حدس اولیه $\underline{x}^{(1)}$ (این نقطه می‌تواند تصادفی انتخاب شود).

← تکرار: برای $k=1,2,\dots,k^{max}$ گام های زیر را تکرار کن :

گام اول: تقریب افاین تابع f حول نقطه $x^{(k)}$ را تشکیل بده :

$$\hat{f}(\underline{x}\,;\,\underline{x}^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

گام دوم: $\underline{x}^{(k+1)}$ را نقطه کمینه $\|\hat{f}(\underline{x},\underline{x}^{(k)})\|^2$ قرار بده :

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

ـ نکاتی در مورد الگوریتم گاوس ـ نیوتن :

- * اگر ستون های $Df(x^{(k)})$ وابسته خطی باشند این الگوریتم با خطا مواجه می‌شود.
- * در این الگوریتم هنگامی $x^{(k+1)} = x^{(k)}$ می‌گردد که رابطه زیر برقرار باشد :

$$\left(Df(x^{(k)})^T Df(x^{(k)})\right)^{-1} Df(x^{(k)})^T f(x^{(k)}) = 0$$

چون فرض کردیم ستون های $Df(x^{(k)})$ مستقل خطی هستند رابطه فوق برقرار خواهد بود اگر و فقط اگر رابطه زیر برقرار باشد:

$$Df(x^{(k)})^T f(x^{(k)}) = 0$$

بنابراین الگوریتم گاوس ـ نیوتن زمانی متوقف می‌شود که شرایط بهینگی برقرار باشد.

* چون $x^{(k+1)}$ حداقل کننده $\|\hat{f}(x; x^{(k)})\|^2$ است و از طرف دیگر رابطه $\hat{f}(x^{(k)}; x^{(k)}) = f(x^{(k)})$ نیز برقرار است. در نتیجه

$$\|\hat{f}(x^{(k+1)}; x^{(k)})\|^2 \leqslant \hat{f}(x^{(k)}; x^{(k)})\|^2 = \|f(x^{(k)})\|^2$$

بنابراین اگرچه مانده تقریب ها در طول تکرارها نزولی است اما این نتیجه زیر را نمی دهد

$$\|f(x^{(k+1)})\|^2 \leqslant \|f(x^{(k)})\|^2$$

بنابراین مانده در طول تکرارها نزولی نیست.

- معایب الگوریتم گاوس-نیوتن:

* الگوریتم می تواند در مواردی واگرا باشد. همانطور که دیدیم تقریب افاین زمانی معتبر است که $x^{(k+1)}$ نزدیک $x^{(k)}$ باشد. بنابراین در شرایطی که $x^{(k+1)}$ فاصله زیادی از $x^{(k)}$ می گیرد امکان دارد مانده در $x^{(k+1)}$ بزرگتر گردد و موجب واگرایی الگوریتم گردد.

* فرض استقلال خطی ستون های $Df(x^{(k)})$ در بسیاری از کاربردها برقرار نمی شود و در نتیجه الگوریتم متوقف می شود.

- توجه: در شرایطی که $m = n$ است، گام دوم الگوریتم گاوس-نیوتن به صورت زیر ساده می شود:

$$x^{(k+1)} = x^{(k)} - (Df(x^{(k)})^T Df(x^{(k)}))^{-1} Df(x^{(k)})^T f(x^{(k)})$$

$$= x^{(k)} - Df(x^{(k)})^{-1} f(x^{(k)})$$

الگوریتم حاصل در این شرایط با عنوان الگوریتم نیوتن-رافسون نیز شناخته می شود.

- الگوریتم لونبرگ-مارکات :

همانطور که دیدیم یکی از معایب الگوریتم گاوس-نیوتن این بود که به علت فاصله زیاد میان $x^{(k)}$, $x^{(k+1)}$ تقریب می توانست مناسب نباشد و در نتیجه آن طول ماند در $f(x)$ کاهش نگردد. به منظور رفع این مشکل می توان $x^{(k+1)}$ را به عنوان حداقل کند تابع زیر در نظر گرفت :

$$x^{(k+1)} = \underset{x}{\operatorname{argmin}} \quad \| \hat{f}(x; x^{(k)}) \|^2 + \lambda^{(k)} \| x - x^{(k)} \|^2$$

که $\lambda^{(k)}$ پارامتری است که می تواند در طول تکرارها تغییر نماید. مقادیر بزرگ $\lambda^{(k)}$ منجر به نزدیکی $x^{(k+1)}$ به $x^{(k)}$ می گردد و هرچه $\lambda^{(k)}$ کوچکتر گردد پاسخ مسأله فوق به الگوریتم گاوس-نیوتن نزدیک می شود.

مسأله فوق یک مسأله حداقل مربعات چندهدفه است و قبلاً با نحوه حل آن آشنا شدیم. این مسأله به صورت زیر قابل بازنویسی است :

$$\left\| \begin{bmatrix} Df(x^{(k)}) \\ \sqrt{\lambda^{(k)}} \quad I \end{bmatrix} x - \begin{bmatrix} Df(x^{(k)})x^{(k)} - f(x^{(k)}) \\ \sqrt{\lambda^{(k)}} \quad x^{(k)} \end{bmatrix} \right\|^2$$

بنابراین در مسأله فوق ماتریس $\begin{bmatrix} Df(x^{(k)}) \\ \sqrt{\lambda^{(k)}} \quad I \end{bmatrix}$ حتماً ستون های مستقل دارد، به این ترتیب مشکل دوم الگوریتم گاوس-نیوتن نیز حل می شود. با استفاده از معادلات نرمال تساوی زیر را داریم :

$$\left(Df(x^{(k)})^T Df(x^{(k)}) + \lambda^{(k)} I \right) x^{(k+1)}$$

$$= Df(x^{(k)})^T \left(Df(x^{(k)}) x^{(k)} - f(x^{(k)}) \right) + \lambda^{(k)} x^{(k)}$$

$$= \left(Df(x^{(k)})^T Df(x^{(k)}) + \lambda^{(k)} I \right) x^{(k)} - Df(x^{(k)})^T f(x^{(k)})$$

بنابراین رابطه نهایی به صورت زیر خواهد بود :

$$x^{(k+1)} = x^{(k)} - \left(Df(x^{(k)})^T Df(x^{(k)}) + \lambda^{(k)} I \right)^{-1} Df(x^{(k)})^T f(x^{(k)})$$

که همواره ماتریس معکوس وجود دارد.

توجه: در الگوریتم لونبرگ-مارکات $x^{(k+1)} = x^{(k)}$ است اگر $2Df(x^{(k)})^T f(x^{(k)}) = 0$ گردد، و همانطور که می دانیم این شرایط همان شرایط بهینگی است.

چالش الگوریتم لونبرگ-مارکات: همانطور که دیدید الگوریتم لونبرگ-مارکات با افزودن تنظیم کند. بخصوص در هر گام مشکلات الگوریتم گاوس-نیوتن را مرتفع نمود، اما تنها چالش جدیدی که ایجاد شده است نحوه انتخاب $\lambda^{(k)}$ است. در مورد این پارامتر توجه به دو نکته زیر قابل توجه است (این پارامتر با نام پارامتر اطمینان نیز شناخته می‌شود) :

← اگر $\lambda^{(k)}$ خیلی کوچک باشد، هدف اصلی ما یعنی جلوگیری از فاصله افتادن میان $x^{(k)}$ و $x^{(k+1)}$ را برآورده نمی‌سازد. در نتیجه امکان دارد $\|f(x^{k+1})\|^2 > \|f(x^{(k)})\|^2$ گردد که مطلوب نیست.

← اگر $\lambda^{(k)}$ خیلی بزرگ باشد، اگرچه $x^{(k+1)}$ نزدیک $x^{(k)}$ خواهد بود و تقریب بطور معتبر است، اما کاهش در تابع هدف یعنی $\| f(x) \|^2$ کم خواهد بود و در نتیجه زمان زیادی لازم است تا الگوریتم همگرا گردد.

راه‌حل: روش‌های مختلفی به منظور تنظیم پارامتر $\lambda^{(k)}$ پیشنهاد شده است. یک راه‌حل ساده به این صورت است :

← براساس مقدار فعلی λ، $x^{(k+1)}$ را به دست آورید. اگر تابع هدف کاهش یافته است نقطه $x^{(k+1)}$ را به عنوان خروجی تکرار بپذیرید، در غیر اینصورت λ را کاهش دهید و دوباره $x^{(k+1)}$ را به دست آورید. این کار را تا زمانی که کاهش در تابع هدف مشاهده گردد ادامه دهید.

- شبه کد الگوریتم لونبرگ مارکات :

← ورودی‌ها :

* تابع مشتق پذیر $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

* نقطه حدس اولیه $x^{(1)}$

* پارامتر $\lambda^{(1)} > ۰$

← تکرار: برای $k = ۱, ۲, \dots, k^{max}$ گام‌های زیر را تکرار کن :

گام اول: تقریب افاین تابع f حول نقطه $x^{(k)}$ را تشکیل بده :

$$\hat{f}(x ; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

گام دوم : $x^{(k+1)}$ را نقطه کمینه مسئله بهینه‌سازی زیر قرار بده :

$$\underline{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \ \|\hat{f}(x ; x^{(k)})\|^{۲} + \lambda^{(k)} \|x - x^{(k)}\|^{۲}$$

گام سوم : تکرار انجام شده را چک کن :

اگر $\|f(x^{(k+1)})\|^{۲} < \|f(x^{(k)})\|^{۲}$ تکرار مورد پذیرش است و $\lambda^{(k+1)} = ۰/۸ \lambda^{(k)}$

در غیر اینصورت λ را افزایش بده و آپدیت را اعمال نکن یعنی :

$\lambda^{(k+1)} = ۲ \lambda^{(k)}$ ←

$x^{(k+1)} = x^{(k)}$ ←

- شرط توقف : الگوریتم لونبرگ مارکات قبل از رسیدن به تعداد تکرار بیشینه k^{max} متوقف می‌شود اگر یکی از دو حالت زیر رخ دهد :

۱) مانده کوچک : $\|f(x^{(k+1)})\|^{۲}$ به اندازه کافی کوچک شود.

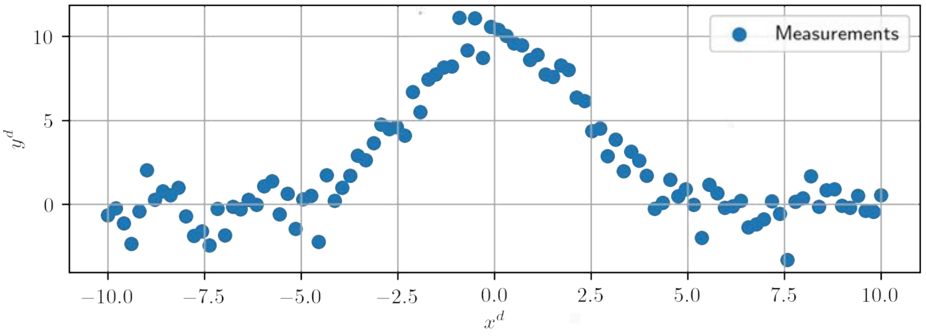
۲) مانده کوچک شرایط بهینگی : $\|۲Df(x^{(k+1)})^T f(x^{(k+1)})\|$ به اندازه کافی کوچک شود.

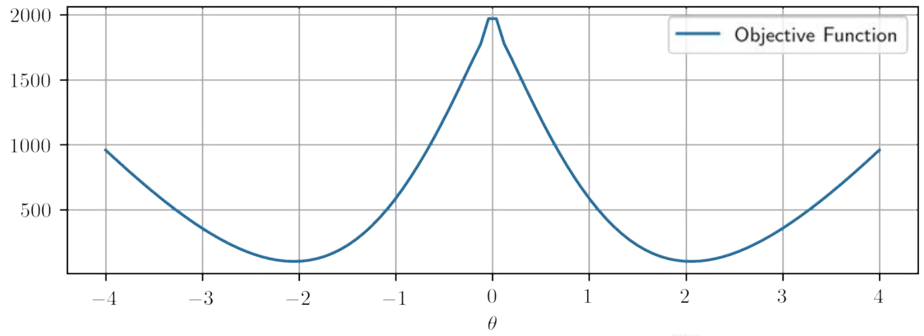
مثال: فرض کنید یک بانک داده در اختیار دارید که از تابع $y = a e^{-\frac{(x-\mu)^2}{2\theta^2}}$ تولید شده است و با یک نویز اندازه‌گیری در اختیار شما قرار گرفته است. همچنین فرض کنید دامنه (a) و متوسط (μ) تابع گاوسی فوق را می‌دانید و هدف تقریب θ است.

به این منظور مسأله برازش حداقل مربعات غیرخطی را داریم که در آن تابع f به صورت زیر است:

$$f(\theta) = \sum_{i=1}^{m} \left(a e^{-\frac{(x_i^{(d)}-\mu)^2}{2\theta^2}} - y_i^{(d)} \right)^2$$

به ازای $a=10$، $\mu=0$، $\theta=2$ و نویز اندازه‌گیری نمونه‌برداری شده توزیع نرمال استاندارد، تابع $f(\theta)$ به صورت زیر است ($m=100$ و نقاط $x_i^{(d)}$ به طور یکنواخت در بازه $[-10, 10)$ قرار گرفته‌اند):





در شرایط فوق اگر از الگوریتم گاوس-نیوتن برای تقریب زدن از داده‌ها استفاده نمایید، گام های زیر را خواهیم داشت.

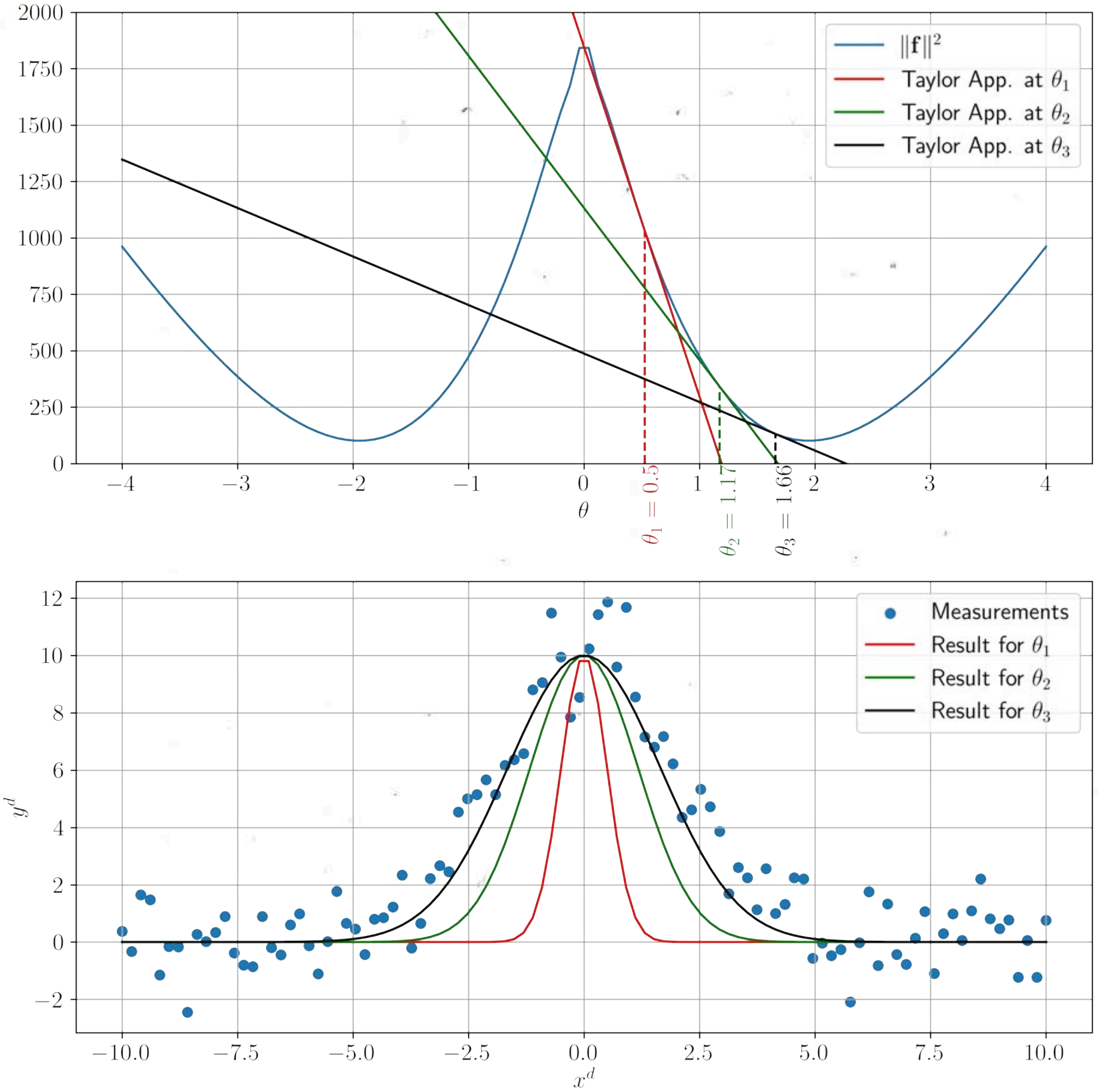
گام اول:

$$\hat{f}(\theta, \theta^{(k)}) = f(\theta^{(k)}) + f'(\theta^{(k)})(\theta - \theta^{(k)})$$

گام دوم:

$$\theta^{(k+1)} = \theta^{(k)} - \frac{f(\theta^{(k)})}{f'(\theta^{(k)})}$$

نتیجه اعمال دو گام از الگوریتم نیوتن بر داده‌های مورد نظر در شکل زیر نشان داده شده است. همانطورکه مشخص است، الگوریتم به سمت یافتن نقطه بهینه برای انحراف معیار پیش رفته است.



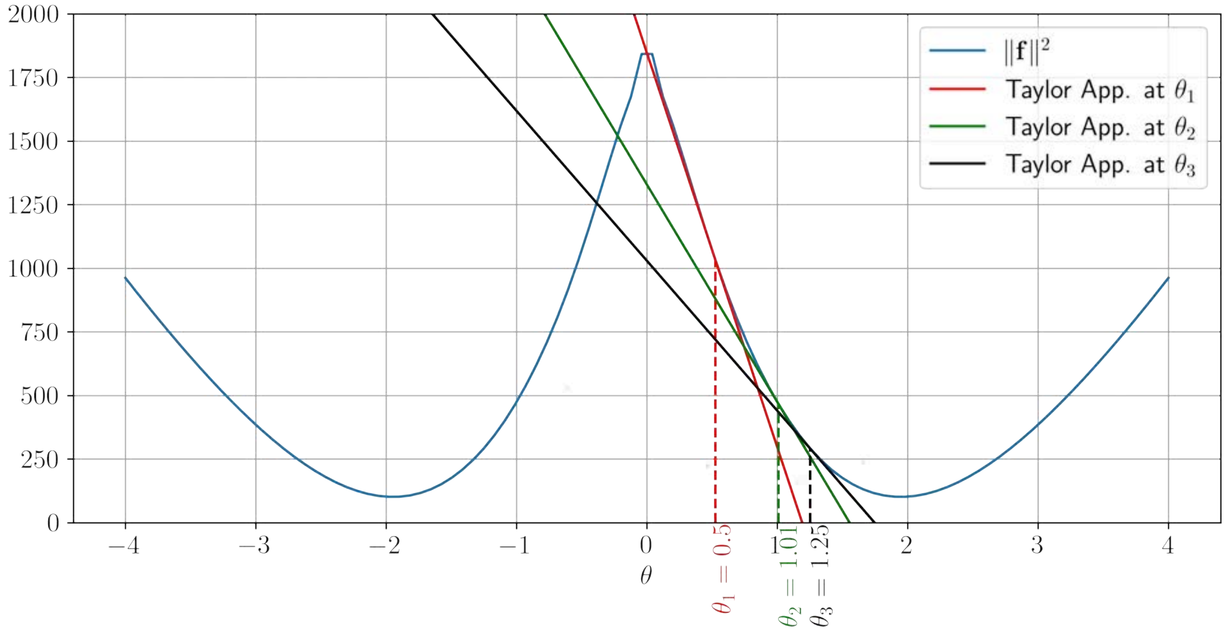
حال فرض کنید از الگوریتم لونبرگ مارکات استفاده نماییم. دراینصورت انتظار داریم که فاصله $x^{(2)}$ از $x^{(1)}$ بیشتر از الگوریتم گاوس نیوتن باشد. این انتظار ناشی از وجود عبارت $\lambda^{(k)}\|x-x^{(k)}\|^{2}$ در تابع هزینه است. برای این الگوریتم گام های هر تکرار به صورت زیر است:

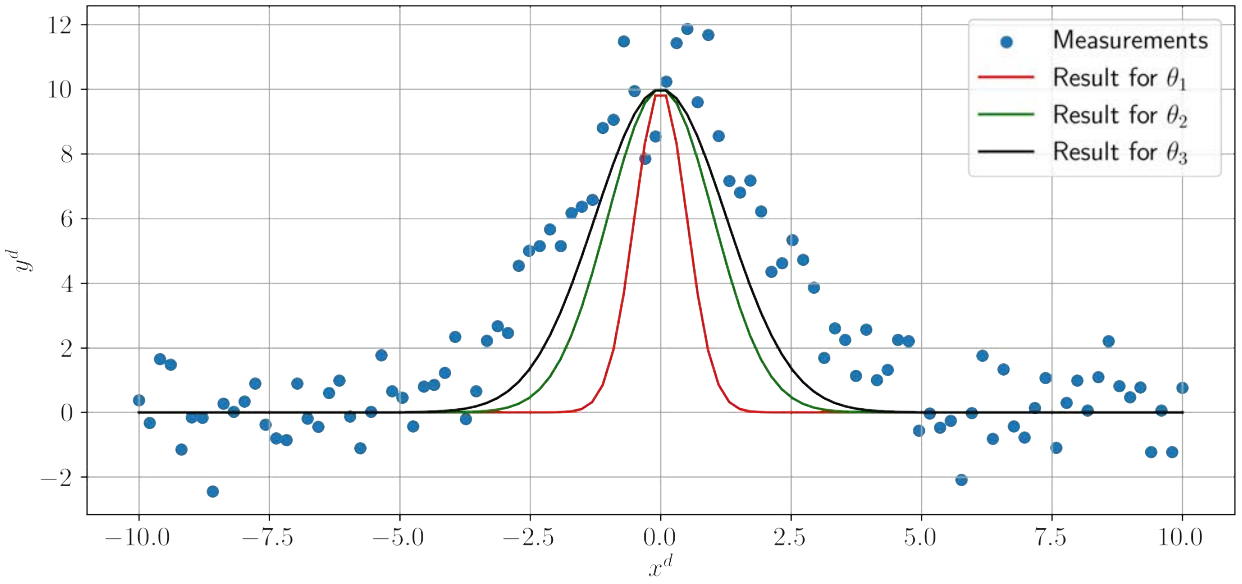
گام اول:

$$\hat{f}(\theta, \theta^{(k)}) = f(\theta^{(k)}) + f'(\theta^{(k)})(\theta - \theta^{(k)})$$

$$\theta^{(k+1)} = \theta^{(k)} - \frac{f'(\theta^{(k)})}{\lambda^{(k)} + (f'(\theta^{(k)}))^{2}} f(\theta^{(k)})$$

در رابطه بالا مشخص است که با افزایش $\lambda^{(k)}$، فاصله $\theta^{(k+1)}$ از $\theta^{(k)}$ کاهش می یابد. نتیجه دو تکرار الگوریتم لونبرگ-مارکات برای مسئله برازش منحنی گاوسی به صورت زیر است.





حداقل مربعات غیرخطی مقید

در این فصل توسعه مسأله حداقل مربعات غیرخطی به شرایطی که قیدهای غیرخطی نیز وجود دارند را در نظر می‌گیریم. مشابه مسأله حداقل مربعات غیرخطی، این مسأله نیز دارای چالش‌های جدی برای رسیدن به جواب دقیق است و در شرح استفاده از راهکارهای ابتکاری برای تقریب مورد نظر قرار می‌گیرد.

مسأله حداقل مربعات غیرخطی مقید: این مسأله به صورت زیر قابل تعریف است:

$$\min_{x} \quad \|\underline{f}(\underline{x})\|^2$$
$$\text{subject. to.} \quad \underline{g}(\underline{x}) = \underline{0}$$

که $\underline{x} \in \mathbb{R}^n$ بردار متغیرها است. $\underline{f}(\underline{x})$ یک بردار m تایی و $\underline{g}(\underline{x})$ یک بردار p تایی است. مسأله فوق را می‌توان به فرم گسترده زیر نیز نوشت:

$$\min_{\underline{x}} \quad f_1(\underline{x})^2 + \cdots + f_m(\underline{x})^2$$
$$\text{subject to} \quad g_i(\underline{x}) = 0 \quad i = 1, 2, \ldots, p$$

که $f_i(\underline{x})$، i-امین مانده، و $g_i(x)=0$، i-امین قید تساوی است. اگر f و g افاین باشند، مسأله حداقل مربعات غیرخطی مقید به مسأله حداقل مربعات خطی مقید تبدیل می‌شود.

بردار مجاز: بردار $\underline{x}$ را مجاز می‌نامیم اگر تساوی $g(\underline{x})=0$ برقرار باشد.

پاسخ حداقل مربعات غیرخطی مقید: بردار $\hat{x}$ پاسخ حداقل مربعات غیرخطی مقید است اگر مجاز باشد و کمترین مقدار تابع هدف در میان تمام بردارهای مجاز را داشته باشد، یعنی به ازای تمام x های مجاز داشته باشیم:

$$\|f(x)\|^2 \geqslant \|f(\hat{x})\|^2$$

حداقل مربعات غیرخطی مقید با قید خطی : یک حالت مسأله حداقل مربعات غیرخطی مقید شرایطی است که قیود مسأله خطی (افاین) باشند. بنابراین مسأله زیر را داریم:

$$\min_{x} \ \|\underline{f}(\underline{x})\|^2$$

$$\text{Subject to} \quad Cx = d$$

که در رابطه فوق $d \in \mathbb{R}^p, C \in \mathbb{R}^{p \times n}$ است. مسأله فوق حداقل مربعات غیرخطی با قیود تساوی خطی نامگذاری می‌شود. برای حل این مسأله می‌توان از حالت توسعه یافته الگوریتم لونبرگ مارکوات استفاده کرد. به این منظور دو گام زیر را خواهیم داشت.

گام اول: تقریب افاین تابع f حول نقطه $x^{(k)}$:

$$\hat{f}(x; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

گام دوم: $x^{(k+1)}$ را نقطه کمینه مسأله بهینه‌سازی زیر قرار دهید:

$$\underline{x}^{(k+1)} = \underset{x}{\operatorname{argmin}} \ \|\hat{f}(x; x^{(k)})\|^2 + \lambda^{(k)} \|x - x^{(k)}\|^2$$

$$\text{subject to} \quad Cx = d$$

مسأله فوق یک مسأله حداقل مربعات چندهدفه مقید است که می‌توانیم پاسخ آن را به صورت زیر به دست آوریم:

$$x^{(k+1)} = \underset{x}{\operatorname{argmin}} \ \|\tilde{A}^{(k)} x - \tilde{b}^{(k)}\|^2 \quad \text{subject to} \quad Cx = d$$

$$\tilde{A}^{(k)} = \begin{bmatrix} Df(x^{(k)}) \\ \sqrt{\lambda^{(k)}} \ I \end{bmatrix}$$

$$\tilde{b}^{(k)} = \begin{bmatrix} Df(x^{(k)}) \, x^{(k)} - f(x^{(k)}) \\ \sqrt{\lambda^{(k)}} \ x^{(k)} \end{bmatrix}$$

مسأله‌ی نوین یک مسأله‌ی حداقل مربعات مقید است که در آن لزوماً ستون‌های ماتریس $\begin{bmatrix} \tilde{A}^{(k)} \\ C \end{bmatrix}$ مستقل خطی است (چون ستون $\tilde{A}^{(k)}$ مستقل خطی است). فرض کنید سطرهای C نیز مستقل خطی هستند. در این صورت ماتریس KKT معکوس‌پذیر بود. و می‌توان پاسخ حداقل مربعات $\hat{x}$ و ضرایب لاگرانژ متناظر $\hat{z}$ را به صورت زیر به دست آورد:

$$\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} 2\tilde{A}^{(k)}\tilde{A}^{(k)} & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} 2\tilde{A}^{(k)}\tilde{b}^{(k)} \\ d \end{bmatrix}$$

حال می‌توان به سادگی $x^{(k+1)} = \hat{x}$ قرار داد. با توجه به مقدار $\|f(x^{(k+1)})\|^2$ در مقایسه با $\|f(x^{(k)})\|^2$ نیز می‌توان مقدار $\lambda^{(k)}$ را تنظیم نمود.

- شبه کد الگوریتم لونبرگ-مارکات توسعه یافته برای مسأله‌ی حداقل مربعات غیرخطی با قیود تساوی خطی:

← **ورودی‌ها**: تابع مشتق‌پذیر $f: \mathbb{R}^n \to \mathbb{R}^m$، حدس اولیه $x^{(1)}$، پارامتر $\lambda^{(1)}$.

← **تکرار**: برای $k = 1, 2, \ldots, k^{max}$ گام‌های زیر را تکرار کن:

گام اول: تقریب افاین تابع f حول نقطه $x^{(k)}$ را تشکیل بده:

$$\hat{f}(x; x^{(k)}) = f(x^{(k)}) + Df(x^{(k)})(x - x^{(k)})$$

گام دوم: $x^{(k+1)}$ را نقطه کمینه مسأله‌ی بهینه‌سازی زیر قرار بده:

$$x^{(k+1)} = \underset{x}{\arg\min} \; \|\hat{f}(x; x^{(k)})\|^2 + \lambda^{(k)} \|x - x^{(k)}\|^2$$
$$\text{subject to} \quad Cx = d$$

گام سوم: تکرار انجام شده را چک کن:

اگر $\|f(x^{(k+1)})\|^2 < \|f(x^{(k)})\|^2$ است تکرار مورد پذیرش است و $\lambda^{(k+1)} = 0.8\,\lambda^{(k)}$

در غیر اینصورت λ را افزایش بده و آپدیت را اعمال نکن یعنی:

$$\lambda^{(k+1)} = 2\lambda^{(k)} \quad \leftarrow$$
$$x^{(k+1)} = x^{(k)} \quad \leftarrow$$

شرایط بهینگی در مسأله حداقل مربعات غیرخطی مقید :

مسأله حداقل مربعات غیرخطی مقید به صورت مقابل را درنظر بگیرید:

$$\min_{x} \|\underline{f}(\underline{x})\|^2$$

$$\text{subject to} \quad \underline{g}(\underline{x}) = \underline{0}$$

تابع لاگرانژین برای این مسأله به صورت زیر تعریف می‌شود:

$$L(x,z) = \|f(x)\|^2 + z_1 g_1(x) + \dots + z_p g_p(x) = \|f(x)\|^2 + g(x)^T z$$

که بردار pتایی z بردار ضرایب لاگرانژ است. براساس روش ضرایب لاگرانژ، برای هر $\hat{x}$ که پاسخ مسأله بهینه‌سازی فوق است می‌بایست یک دسته ضرایب لاگرانژ $\hat{z}$ وجود داشته باشد به گونه‌ای که:

$$\frac{\partial L}{\partial x_i}(\hat{x},\hat{z}) = 0 \quad i=1,\dots,n$$

$$\frac{\partial L}{\partial z_i}(\hat{x},\hat{z}) = 0 \quad i=1,\dots,p$$

مشروط براینکه سطرهای ماتریس $Dg(\hat{x})$ مستقل خطی باشند.

بردار $\hat{z}$ نیز ضرایب لاگرانژ بهینه است. براساس دو شرط فوق داریم:

$$\frac{\partial L}{\partial z_i}(\hat{x},\hat{z}) = 0 \Rightarrow g_i(\hat{x}) = 0 \quad i=1,\dots,p$$

بنابراین براساس شرط دوم رابطه زیر را داریم:

$$\underline{g}(\hat{\underline{x}}) = 0$$

و براساس این شرط باید $\hat{x}$ مجاز باشد. براساس شرط اول نیز داریم:

$$2Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 0$$

همانطور که به نظر می‌رسد رابطه فوق توسعه رابطه $Df(\hat{x})^T f(\hat{x}) = 0$ برای بهینگی در مسأله حداقل مربعات غیرخطی است.

بنابراین شرط لازم برای اینکه $\hat{x}$ پاسخ مسأله حداقل مربعات غیرخطی مقید باشد این است که بردار $\hat{z}$ وجود داشته باشد به گونه‌ای که روابط زیر برقرار باشد:

$$\begin{cases} \underline{g}(\hat{x}) = \underline{0} \\ 2Df(\hat{x})^T f(\hat{x}) + Dg(\hat{x})^T \hat{z} = 0 \end{cases}$$

- **الگوریتم جریمه**: در الگوریتم جریمه از تقریب مسأله مقید با مسأله غیرمقید استفاده می‌شود. مسأله حداقل مربعات غیرخطی مقید به صورت زیر را در نظر بگیرید:

$$\min_x \|f(x)\|^2 \quad \text{subject to} \quad \underline{g}(\underline{x}) = \underline{0}$$

می‌توان این مسأله را به صورت تقریبی برابر با مسأله زیر در نظر گرفت در حالیکه پارامتر μ به سمت بینهایت میل می‌کند:

$$\min_x \quad \|f(x)\|^2 + \mu \|g(x)\|^2$$

برای حل مسأله فوق از نمایش بلوکی دو تابع هدف فوق در یک عبارت به صورت زیر استفاده می‌نماییم:

$$\min_x \quad \left\| \begin{bmatrix} f(x) \\ \sqrt{\mu}\, g(x) \end{bmatrix} \right\|^2$$

مسأله فوق یک مسأله حداقل مربعات غیرخطی است که می‌توان با استفاده از الگوریتم لونبرگ-مارکوارت پاسخ آن را تقریب زد. نکته‌ای که وجود دارد با حل مسأله فوق لزوما پاسخ در رابطه $g(\underline{x}) = 0$ صادق نخواهد بود اما با انتخاب مقدار به اندازه کافی بزرگ برای پارامتر μ می‌توانیم به جایی برسیم که $g(x)$ به اندازه کافی کوچک است و در عین حال $f(x)$ نیز کوچک شده است.

در الگوریتم جریمه مسأله $\min_{x} \|f(x)\|^2+\mu\|g(x)\|^2$ برای یک دنباله از مقادیر افزایشی μ حل می‌شود. شبه کد این الگوریتم در ادامه آورده شده است:

- شبه کد الگوریتم جریمه برای حل مسأله حداقل مربعات غیرخطی مقید:

← ورودی‌ها:

* توابع مشتق‌پذیر $f: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^n \to \mathbb{R}^p$
* نقطه حدس اولیه $x^{(1)}$
* $\mu^{(1)}=1$

تکرار: برای $k=1,2,\ldots,k^{max}$، گام‌های زیر را تکرار کن:

گام اول: مسأله حداقل مربعات غیرخطی زیر را با استفاده از روش لونبرگ-مارکات و با شروع از نقطه $x^{(k)}$ حل کن و پاسخ آن را برابر $x^{(k+1)}$ قرار بده:

$$\min_{x} \|\underline{f}(\underline{x})\|^2+\mu^{(k)}\|\underline{g}(\underline{x})\|^2$$

گام دوم: $\mu^{(k)}$ را با استفاده از رابطه $\mu^{(k+1)}=2\mu^{(k)}$ به روز رسانی کن.

توجه: الگوریتم جریمه اگر $\|g(x^{(k)})\|$ به اندازه کافی کوچک شود متوقف می‌شود. در این شرایط قید تقریباً برآورده شده است.

مشکل الگوریتم جریمه: الگوریتم جریمه به سادگی قابل پیاده‌سازی است اما یک مشکل اصلی دارد. از آنجایی که مقدار $\mu^{(k)}$ با پیشرفت الگوریتم افزایش می‌یابد، به‌ازای مقادیر بزرگ μ تعداد تکرارهای لازم برای همگرایی الگوریتم لونبرگ-مارکات افزایش می‌یابد و در بعضی مواقع همگرا نشود. برای حل این مشکل می‌توان از روش لاگرانژین افزوده استفاده نمود.

توجه: در الگوریتم جریمه می‌دانیم که $x^{(k+1)}$ پاسخ مسأله بهینه‌سازی زیر است:

$$x^{(k+1)} = \underset{x}{\operatorname{argmin}} \left\| \begin{bmatrix} f(x) \\ \mu^{(k)} g(x) \end{bmatrix} \right\|^2$$

بنابراین شرایط بهینگی در آن برقرار خواهد بود. در اینصورت رابطه زیر را داریم:

$$2Df(x^{(k+1)})^T f(x^{(k+1)}) + 2\mu^{(k)} Dg(x^{(k+1)})^T g(x^{(k+1)}) = 0$$

اگر ($z^{(k+1)} = 2\mu^{(k)} g(x^{(k+1)})$ تعریف کنیم می‌توان نشان داد $x^{(k+1)}$, $z^{(k+1)}$ شرایط بهینگی را دارند اما می‌دانیم این شرایط، شرایط لازم است (شرط مجاز بودن $g(x^{(k)}) = 0$ تنها در حالت حدی $k \to \infty$ برقرار خواهد بود).

- الگوریتم لاگرانژین افزوده: الگوریتم لاگرانژین افزوده با هدف رفع مشکل الگوریتم جریمه برای مقادیر بزرگ μ ارائه گردید.

مسأله حداقل مربعات غیرخطی مقید را در نظر بگیرید.

$$\min_x \|f(x)\|^2 \quad \text{subject to} \quad g(x) = 0$$

لاگرانژین افزوده برای مسأله فوق به صورت زیر تعریف می‌شود:

$$\begin{aligned} L_\mu(x,z) &= L(x,z) + \mu\|g(x)\|^2 \\ &= \|f(x)\|^2 + g(x)^T z + \mu\|g(x)\|^2 \end{aligned}$$

در واقع لاگرانژین افزوده همان لاگرانژین است که عبارت $\mu\|g(x)\|^2$ به آن افزوده شده است. از یک نگاه دیگر نیز عبارت فوق همان تابع هدف الگوریتم جریمه است که عبارت $g(x)^T z$ به آن افزوده شده است.

لاگرانژین افزوده. را می‌توان لاگرانژین مساله زیر نیز در نظر گرفت :

$$\min_{x} \quad \|f(x)\|^{2} + \mu \|g(x)\|^{2}$$

$$\text{subject to} \quad g(x) = 0$$

توجه نمایید که مساله فوق معادل مساله حداقل مربعات غیرخطی مقید است .

در الگوریتم لاگرانژین افزوده، تابع لاگرانژین افزوده $L_{\mu}(x,z)$ به ازای یک دنباله از مقادیر z و μ روی x کمینه می‌شود. به این منظور ابتدا رابطه زیر را در نظر بگیرید :

$$\mu \|g(x) + \frac{z}{2\mu}\|^{2} = \mu \|g(x)\|^{2} + 2\mu g(x)^{T}\left(\frac{z}{2\mu}\right) + \mu \|\frac{z}{2\mu}\|^{2}$$

$$= g(x)^{T} z + \mu \|g(x)\|^{2} + \mu \|\frac{z}{2\mu}\|^{2}$$

بنابراین تساوی زیر را داریم :

$$g(x)^{T} z + \mu \|g(x)\|^{2} = \mu \|g(x) + \frac{z}{2\mu}\|^{2} - \mu \|\frac{z}{2\mu}\|^{2}$$

اگر از تساوی فوق در رابطه لاگرانژین افزوده استفاده نماییم به رابطه زیر می‌رسیم :

$$L_{\mu}(x,z) = \|f(x)\|^{2} + \mu \|g(x) + \frac{z}{2\mu}\|^{2} - \mu \|\frac{z}{2\mu}\|^{2}$$

مساله کمینه‌سازی تابع لاگرانژین بر روی متغیر x به صورت زیر است :

$$\min_{x} L_{\mu}(x,z) \equiv \min_{x} \|f(x)\|^{2} + \mu \|g(x) + z/2\mu\|^{2}$$

مساله فوق را می‌توان با استفاده از فرم بلوکی به صورت زیر بازنویسی کرد :

$$\min_{x} \left\| \begin{bmatrix} f(x) \\ \sqrt{\mu}\, g(x) + \frac{z}{2\sqrt{\mu}} \end{bmatrix} \right\|^{2}$$

یک تقریب از پاسخ مسأله فوق را می‌توان با استفاده از الگوریتم لونبرگ - مارکات به دست آورد. اگر پاسخ حاصل را $\tilde{x}$ بنامیم، براساس شرط بهینگی این نقطه باید در رابطه زیر صادق باشد:

$$0 = ۲Df(\tilde{x})^T f(\tilde{x}) + ۲\mu Dg(\tilde{x})^T \left(g(\tilde{x}) + \frac{z}{۲\mu} \right)$$

$$= ۲Df(\tilde{x})^T f(\tilde{x}) + Dg(\tilde{x})^T \left(۲\mu g(\tilde{x}) + z \right)$$

براساس رابطه فوق دو حالت می‌تواند پیش بیاید:

۱) اگر $\tilde{x}$ مجاز باشد ($g(\tilde{x}) = 0$) در اینصورت این بردار به همراه بردار دلخواه z شرایط بهینگی برای مسأله حداقل مربعات غیرخطی مقید را دارد.

۲) اگر $\tilde{x}$ مجاز نباشد ($g(\tilde{x}) \neq 0$) در اینصورت با انتخاب $\tilde{z} = ۲\mu g(\tilde{x}) + z$ شرایط بهینگی برای $\tilde{x}$ و $\tilde{z}$ برقرار خواهد شد.

الگوریتم لاگرانژین افزوده بین حداقل‌سازی تابع لاگرانژین با استفاده از الگوریتم لونبرگ مارکات روی متغیر x و به روز رسانی متغیر z به صورت $\tilde{z} = ۲\mu g(\tilde{x}) + z$ جابه‌جا می‌شود. پارامتر جریمه μ زمانی که نیاز است (زمانی که کاهش $\|g(x)\|$ به اندازه کافی نیست) افزایش می‌یابد.

توجه: الگوریتم لاگرانژین افزوده متوقف می‌شود اگر $g(x^{(k)})$ به اندازه کافی کوچک شود. توجه نمایید که دو بردار $x^{(k+1)}$ و $z^{(k+1)}$ شرایط بهینگی را برآورده می‌سازند.

توجه: الگوریتم لاگرانژین افزوده پیچیدگی خیلی بالاتری از الگوریتم جریمه ندارد اما در عمل بسیار بهتر عمل می‌کند. علت بهبود عملکرد نیز عدم نیاز به افزایش پیوسته مقدار $\mu^{(k)}$ است.

- شبه کد الگوریتم لاگرانژین افزوده برای حل مسأله حداقل مربعات مقید

← ورودی ها :

* توابع مشتق پذیر $f: \mathbb{R}^n \to \mathbb{R}^m$, $g: \mathbb{R}^n \longrightarrow \mathbb{R}^p$

* نقطه حدس اولیه $x^{(1)}$

* $\mu^{(1)}=1$, $z^{(1)}=0$

← تکرار: برای $k=1,2,\dots,k^{max}$ گام های زیر را تکرار کن :

گام اول: حل مسأله حداقل مربعات نامقید با استفاده از روش لونبرگ مارکات و باشروع از نقطه اولیه $x^{(k)}$ ، مسأله زیر را حل کن و پاسخ آن را در $x^{(k+1)}$ قرار دهید.

$$\min_{x} \quad \|f(x)\|^2 + \mu^{(k)} \|g(x) + \frac{z^{(k)}}{2\mu^{(k)}}\|^2$$

گام دوم: $z^{(k)}$ را به صورت زیر به روز رسانی کن :

$$z^{(k+1)} = z^{(k)} + 2\mu^{(k)} g(x^{(k+1)})$$

گام سوم: $\mu^{(k)}$ را به صورت زیر به روز رسانی کن :

$$\mu^{(k+1)} = \begin{cases} \mu^{(k)} & \|g(x^{(k+1)})\| < 0.25\|g(x^{(k)})\| \\ 2\mu^{(k)} & \|g(x^{(k+1)})\| \geqslant 0.25\|g(x^{(k)})\| \end{cases}$$

مثال: در این مسأله راه حل مبتنی بر لونبرگ-مارکارت برای مسأله حداقل مربعات غیرخطی مقید با قیود خطی را مورد بررسی قرار می دهیم. فرض کنید تعداد داده در قالب یک پایگاه داده در اختیار شما قرار گرفته است و می بایست این داده‌ها به صورت مجموع دو منحنی گاوسی برازش شوند. به عبارت دیگر رابطه برازش به صورت زیر است:

$$y^{(d)} = a_1 e^{-\frac{(x^{(d)}-\mu_1)^2}{2\sigma_1^2}} + a_2 e^{-\frac{(x^{(d)}-\mu_2)^2}{2\sigma_2^2}} = \bar{f}(\theta; x^{(d)})$$

علاوه بر مدل فوق موارد زیر نیز به عنوان اطلاعات پیشین در اختیار شما قرار گرفته است:

$$\sigma_1 = \sigma_2 = 2 \quad \leftarrow$$

$$\mu_1 - \mu_2 = 10 \quad \leftarrow$$

مسأله فوق را با استفاده از برازش حداقل مربعات غیرخطی با قیود خطی مدل کنید و نتیجه دو تکرار از الگوریتم را نمایش دهید.

می دانیم در مسأله با فرض $\theta = \begin{bmatrix} a_1 \\ \mu_1 \\ a_2 \\ \mu_2 \end{bmatrix}$ مسأله به صورت زیر است:

$$\min_{\underline{\theta}} \|\underline{f}(\underline{\theta})\|^2 \quad \text{s.t.} \quad \underbrace{\begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}}_{C} \theta = \underbrace{10}_{d}$$

که بردار $\underline{f}$ به صورت مقابل است:

$$\underline{f} = \begin{bmatrix} \tilde{f}(\theta; x_1^{(d)}, y_1^{(d)}) \\ \vdots \\ \tilde{f}(\theta; x_N^{(d)}, y_N^{(d)}) \end{bmatrix}, \quad \tilde{f}(\theta; x, y) = y - \bar{f}(\theta; x)$$

بنابراین گام‌های حل مسأله به صورت زیر خواهد بود :

گام اول :

$$\hat{\underline{f}}(\theta) = \underline{f}(\theta^{(k)}) + Df(\theta^{(k)})(\theta - \theta^{(k)})$$

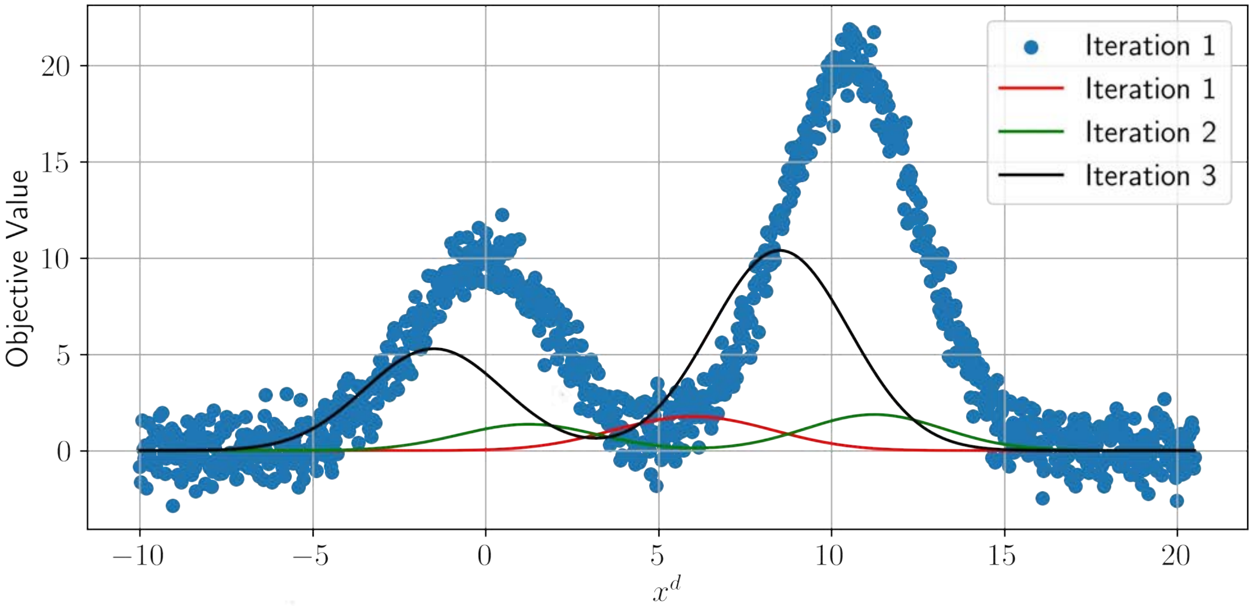
که $Df(\theta^{(k)})$ ماتریسی $N \times 4$ است .

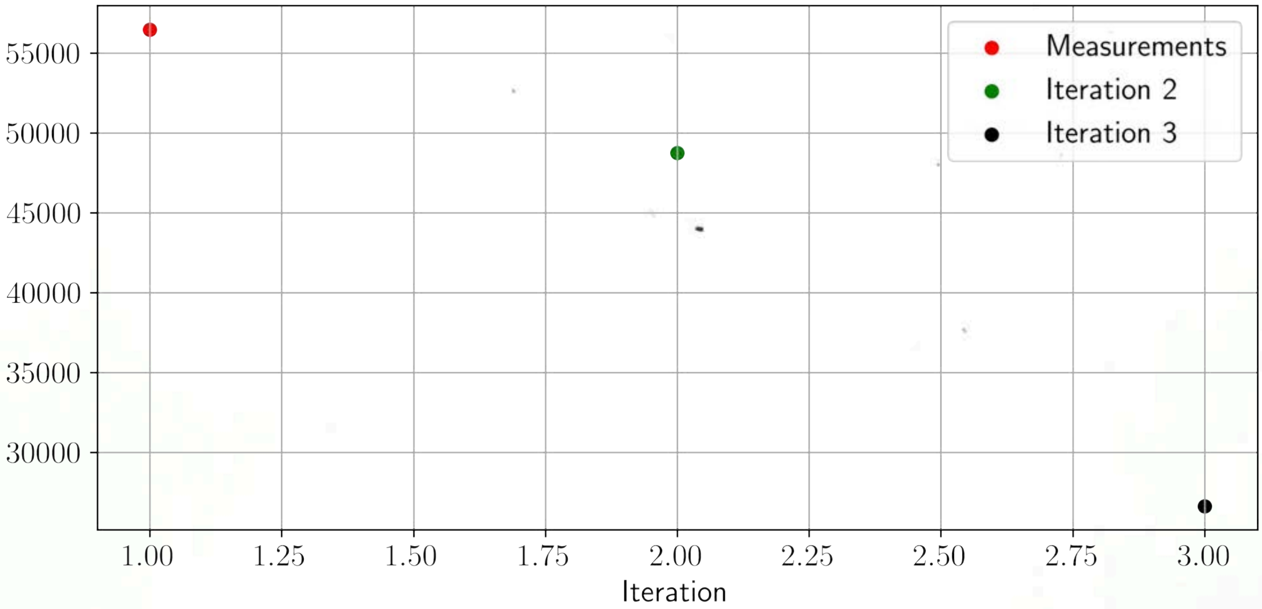
گام دوم : با استفاده از شرایط بهینگی مسأله زیر حل می‌شود :

$$\min_{\theta} \quad \hat{f}(\theta) + \lambda^{(k)} \|\theta - \theta^{(k)}\|^2$$

$$\text{s.t.} \quad C\theta = d$$

شکل زیر نتیجه اجرای دو تکرار از الگوریتم فوق را نشان می‌دهد .





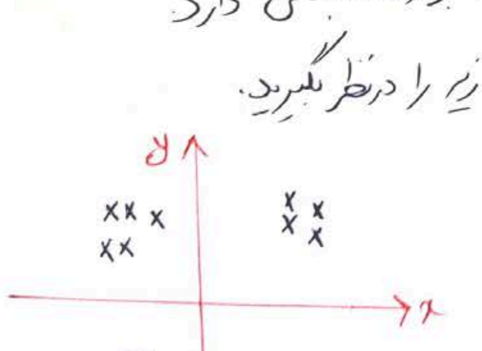
گروه‌بندی (خوشه‌بندی)

در این درس به طور عمده با بردارها و ماتریس‌ها سروکار داشتیم. در این فصل می‌خواهیم گروه‌بندی یک مجموعه از بردارها را مورد بررسی قرار دهیم. منظور از گروه‌بندی تقسیم بردارها به گروه‌هایی است که براساس معیارهای معرفی شد. در این فصل، بردارهای هم‌گروه دارای شباهت بالا به یکدیگر و بردارهای غیرهم‌گروه شباهت پایینی دارند. در این فصل یک الگوریتم شناخته شده گروه‌بندی با نام k-متوسط (k-means) را مورد بررسی قرار می‌دهیم. توجه داشته باشید که مشابه الگوریتم KSVD که در هر تکرار آن k بار تجزیه مقادیر تکین انجام می‌شد در الگوریتم k-means نیز در هر تکرار k بار متوسط حساب می‌شود.

- تعریف مسأله: فرض کنید مجموعه بردارهای $\{\underline{x}_i\}_{i=1}^{N}$ را در اختیار دارید. هدف گروه‌بندی تقسیم این بردارها به k گروه (خوشه) است به گونه‌ای که بردارهای هر گروه حتی‌الامکان به یکدیگر نزدیک باشند. معمولاً مقدار k بسیار کوچکتر از مقدار N است.

- توجه: اولین هدف در مسأله گروه‌بندی تعیین این موضوع است که آیا این کار امکان‌پذیر است یا خیر. امکان‌پذیری گروه‌بندی معمولاً به مقدار k و مجموعه بردارها بستگی دارد.

- مثال ساده: بردارهای دو بعدی نشان داده شده در شکل زیر را در نظر بگیرید.



همانطور که مشخص است این بردارها به صورت خوبی نشان دهنده ۳ گروه هستند.

- چالش‌ها: معمولاً مسأله گروه‌بندی به صورت ساده‌ای که در بالا به آن اشاره گردید نیست و موارد زیر می‌تواند چالش‌های مختلفی در مسیر آن ایجاد نماید.

← بعد بردارها بسیار بالاتر از ۲ است و در نتیجه امکان مصورسازی آنها وجود ندارد

← برخلاف مثال فوق، بردارها معمولاً به خوبی در گروه‌ها قرار نمی‌گیرند و بردارهایی وجود دارند که می‌توانند در دو گروه مختلف قرار بگیرند (مثلاً در مورد مثال فوق برداری را در نظر بگیرید که قرار داشته باشد. در اینصورت می‌تواند به هر دو کلاس بالا و پایین نیم‌صفحه تعلق گیرد)

- تعداد گروه‌های مناسب برای کل مجموعه داده معمولاً از قبل مشخص نیست و باید با امتحان مقادیر مختلف به کل مقدار مناسب رسید.

کاربردهای عملیاتی:

- تشخیص متن‌های مشابه ← فرض کنید بردارهای x_i نشان دهنده تعداد تکرار کلمات مختلف در متن i-ام باشد (مثلاً مؤلفه اول بردارها نشان‌دهنده تعداد تکرار کلمه "خانه" در متن است). در اینصورت با استفاده از گروه بندی می‌توان متن‌های مشابه را به دست آورد.

- گروه بندی بیمارها ← اگر بردار x_i نشان دهنده علائم بیمار i-ام باشد در اینصورت می‌توان با گروه بندی این بردارها بیماران با بیماری‌های تقریباً مشابه را مشخص نمود.

- گروه بندی مشتریان ← اگر بردار x_i نشان دهنده کالای خریداری شده توسط i-ام باشد (مثلاً مؤلفه اول بردارها نشان دهنده خرید یا عدم خرید "نان" باشد)، در اینصورت می‌توان با گروه بندی این بردارها، مشتریان هم گرایش را شناسایی نمود.

هدف گروه بندی: در مسأله گروه بندی هدف این است که اگر گروه‌ها را با برچسب‌های 1، 2، ...، k مشخص نمائیم، برداری با طول N مشخص گردد که مؤلفه i-ام آن برچسب گروه متناظر با بردار x_i را مشخص نماید. این بردار را معمولاً با $\underline{c}$ نمایش می‌دهیم.

علاوه براین به طور معادل می‌توان هدف در گروه بندی را تعیین مجموعه اندیس بردارهای متعلق به گروه j-ام به ازای j=1،2،...،k دانست. اگر این مجموعه را برای گروه j-ام با G_j نمایش دهیم در اینصورت:

$$G_j = \{i \mid c_i = j\}$$

- نماینده گروه: در مسائل گروه‌بندی معمولاً به بردارهای هر گروه یک نماینده مانند z_j نسبت داده می‌شود. این نماینده می‌تواند هرکدام از اعضای گروه باشد. اما نماینده معمولاً به گونه‌ای انتخاب می‌شود که به طور همزمان فاصله‌اش از تمام اعضای گروه حداقل گردد.

- تابع هدف گروه‌بندی: یک تابع هدف مناسب برای مسائل گروه‌بندی می‌تواند به صورت زیر باشد:

$$J^{clust} = \frac{1}{N}\left(\|x_1 - z_{c_1}\|^{۲} + \|x_۲ - z_{c_۲}\|^{۲} + \dots + \|x_N - z_{c_N}\|^{۲}\right)$$

که تابع فوق مربع فاصله هر بردار از نماینده گروه متناظر با آن بردار است. در تابع فوق نماینده‌های گروه‌ها و بردار $\underline{c}$ متغیر هستند. هرچه مقدار J^{clust} کوچکتر گردد، گروه‌بندی حاصل مناسب‌تر خواهد بود.

حالت حدی: اگر $J^{clust} = ۰$ گردد یعنی فاصله میان هر بردار و نماینده گروه متناظر با آن صفر است. این حالت در شرایطی رخ می‌دهد که بردارهای ورودی k مقدار منحصر به فرد داشته باشند و نماینده گروه‌ها نیز برابر همان k مقدار قرار گیرد.

- توجه: تعیین متغیرهای J^{clust} به گونه‌ای که کمترین مقدار تابع هدف به دست آید از نظر محاسباتی پیچیده است. بنابراین راه حل‌هایی که منجر به تقریب جواب بهینه (یا به دست آوردن جواب زیر بهینه)، معمولاً مورد بررسی قرار می‌گیرند.

- تقریب پاسخ بهینه گروه‌بندی: این کار با حداقل سازی ترتیبی روی متغیرهای تابع J^{clust} انجام می‌شود. همانطور که می‌دانیم مسائل گروه‌بندی به صورت زیر قابل فرمول‌بندی است:

$$\min_{\underline{c},\{z_i\}_{i=1}^{k}} J^{clust}$$

در مسائل یادگیری واژه‌نامه نیز با بهینه‌سازی ترتیبی آشنا شدیم. در ادامه دو گام مسائل فوق را مورد بررسی قرار می‌دهیم.

$\min_{\underline{c}} J^{clust}$

- گام اول: در این گام با فرض ثابت بودن نماینده‌های گروه‌ها (z_1 و ..., z_k)، بردار $\underline{c}$ را به گونه‌ای تعیین می‌کنیم که J^{clust} حداقل شود.

وابستگی J^{clust} به c_i (مؤلفه i-ام بردار گروه‌بندی)، به صورت زیر است:

$$\frac{1}{N}\|x_i - z_{c_i}\|^۲$$

و سایر عبارت‌های موجود در J^{clust} به c_i وابسته نیستند.

بنابراین برای حداقل کردن عبارت فوق کافیست از میان گروه‌ها، گروهی را انتخاب کنیم که نماینده آن فاصله کمتری با x_i دارد. در این حالت رابطه زیر را داریم:

$$\|x_i - z_{c_i}\| = \min_{j=۱,\ldots,k} \|x_i - z_j\|$$

با انجام گام فوق، مقدار J^{clust} برابر است با:

$$J^{clust} = \frac{1}{N}\left(\min_{j=۱,\ldots,k} \|x_1 - z_j\|^۲ + \cdots + \min_{j=۱,\ldots,k} \|x_N - z_j\|^۲\right)$$

عبارت فوق متوسط مربع فاصله هر بردار از نزدیکترین نماینده گروه است.

$\min_{\{z_i\}_{i=1}^{k}} J^{clust}$

- گام دوم: در این گام با فرض ثابت بودن گروه‌بندی، نماینده گروه‌ها به گونه‌ای به روز رسانی می‌شود که J^{clust} حداقل گردد. برای تعیین پاسخ این گام نمایش زیر را برای J^{clust} در نظر بگیرید:

$$J^{clust} = J_1 + J_۲ + \cdots + J_k \quad , \quad J_j = \frac{1}{N}\sum_{i \in G_j} \|x_i - z_j\|^۲$$

که J_j میزان تأثیر J^{clust} ناشی از بردارهای نسبت داده شده به کلاس j-ام است.

نمایش فوق باعث می‌شود بتوانیم مسأله این گام را به k زیر مسأله زیر تفکیک کنیم:

$$\min_{z_j} \frac{1}{N}\sum_{i \in G_j} \|x_i - z_j\|^۲$$

اما پاسخ این مسأله به سادگی به صورت زیر به دست می آید (با استفاده از مشتق گیری نسبت به z):

$$z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$$

که $|G_j|$ تعداد اعضای مجموعه G_j است.

- الگوریتم k-means : این الگوریتم در واقع دو گام فوق را به صورت تکراری انجام می دهد تا زمانی که همگرایی صورت پذیرد. شبه کد این الگوریتم به صورت زیر است:

- شبه کد الگوریتم k-means برای گروه بندی:

← ورودی ها: بردارهای $\{x_i\}_{i=1}^{N}$ ، نماینده گروه های اولیه $\{z_j\}_{j=1}^{k}$

← تکرار: تا زمان همگرایی گام های زیر را پشت سرهم تکرار کن:

گام اول: برای هر بردار $\underline{x}_i$ ، مقدار c_i را به صورت زیر به روز کن:

$$c_i = \underset{j=1,\dots,k}{\arg\min} \|\underline{x}_i - z_j\| , \quad i=1,\dots,N$$

گام دوم: برای هر گروه j، نماینده گروه را به صورت زیر به روز کن:

$$\underline{z}_j = \frac{1}{|G_j|} \sum_{i \in G_j} \underline{x}_i$$

- نکاتی در مورد الگوریتم k-means :

۱- این الگوریتم ضمانتی برای اینکه حتما تمامی گروه ها غیرتهی باشند ندارد. بنابراین امکان دارد در گام اول به یک گروه، هیچ برداری تعلق نگیرد. در اینصورت امکان به روز رسانی نماینده آن گروه در گام دوم وجود ندارد. در چنین شرایطی معمولاً گروه به طور کلی حذف می گردد و الگوریتم k-means به تعداد گروه های کمتری از k خواهد رسید.

۲- اگر بردار c حاصل در گام اول برای دو تکرار پشت سرهم ثابت گردد، در گام دوم نیز تغییری در مرکز گروه‌ها نخواهیم داشت و در ضمن شرایطی می‌گوییم الگوریتم همگرا شده است.

۳- در عمل معمولاً زمانی که کاهش مشاهده شده در J^{clust} از یک مقدار آستانه کمتر گردد الگوریتم متوقف می‌شود.

۴- یک انتخاب ساده برای نماینده‌ی گروه‌های اولیه، انتخاب k بردار از مجموعه $\{\underline{x}_i\}_{i=1}^{N}$ است.

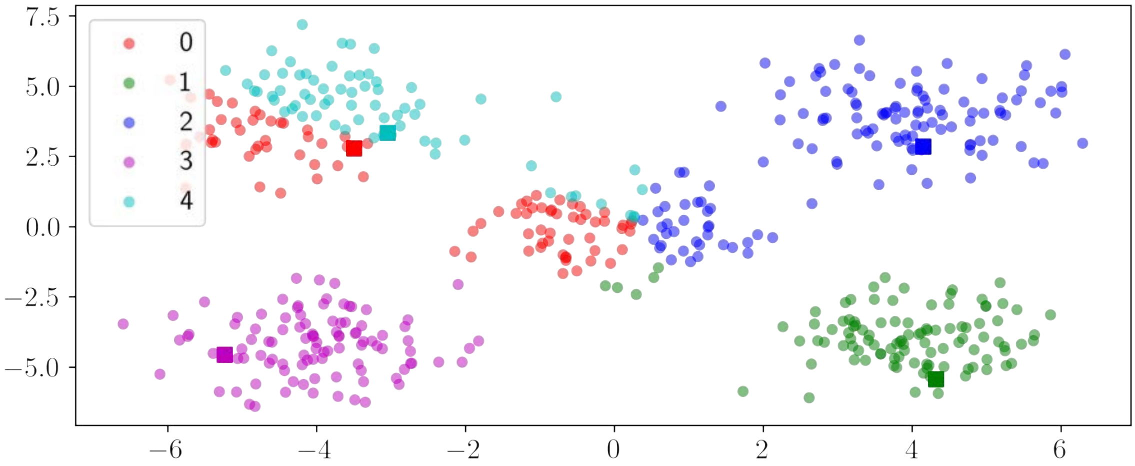
توجه: چون در الگوریتم k-means در هر تکرار J^{clust} را کاهش می‌دهد بنابراین در تعداد گام محدودی همگرا می‌شود. اما براساس مقداردهی اولیه، این نقطه همگرایی می‌تواند متفاوت باشد.

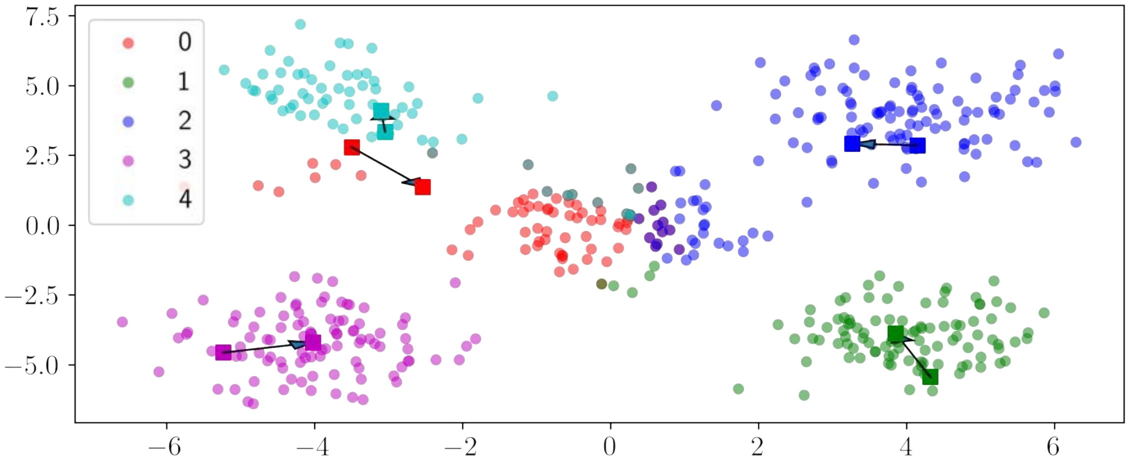
توجه: از آنجایی که الگوریتم k-means یک الگوریتم ابتکاری است (ضمانتی وجود ندارد که بتواند به حداقل J^{clust} همگرا شود)، معمولاً به ازای نقاط شروع مختلف اجرا می‌شود و در نهایت اجرایی که منجر به کمترین مقدار J^{clust} گردد، انتخاب می‌شود.

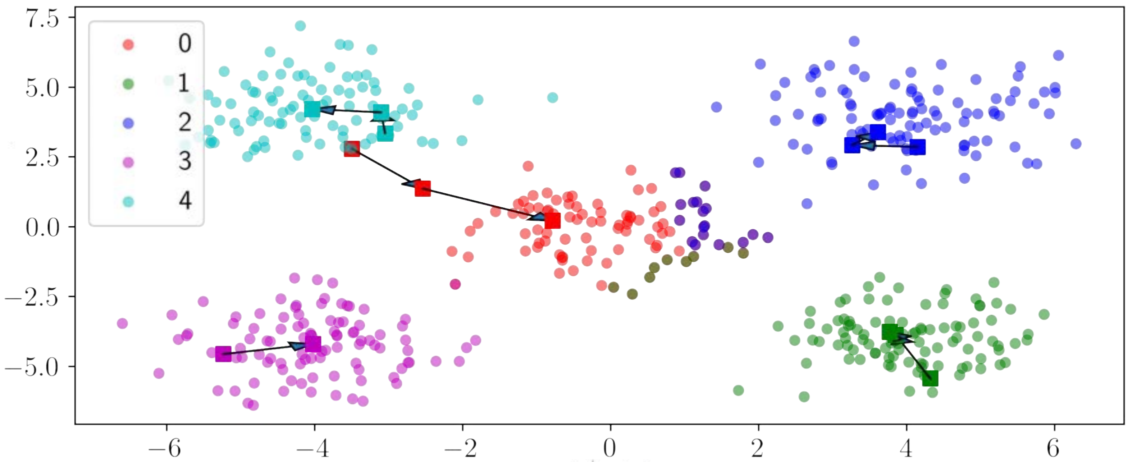
تفسیر نماینده‌ی گروه: نماینده‌های گروه پس از همگرایی الگوریتم k-means می‌توانند اطلاعات بسیار مفیدی از آن گروه ارائه نمایند. به عنوان مثال در کاربرد گروه‌بندی متن‌ها، مؤلفه‌ی اول در بردار z_1، متوسط تعداد تکرار کلمه "خانه" در آن گروه را نشان می‌دهد.

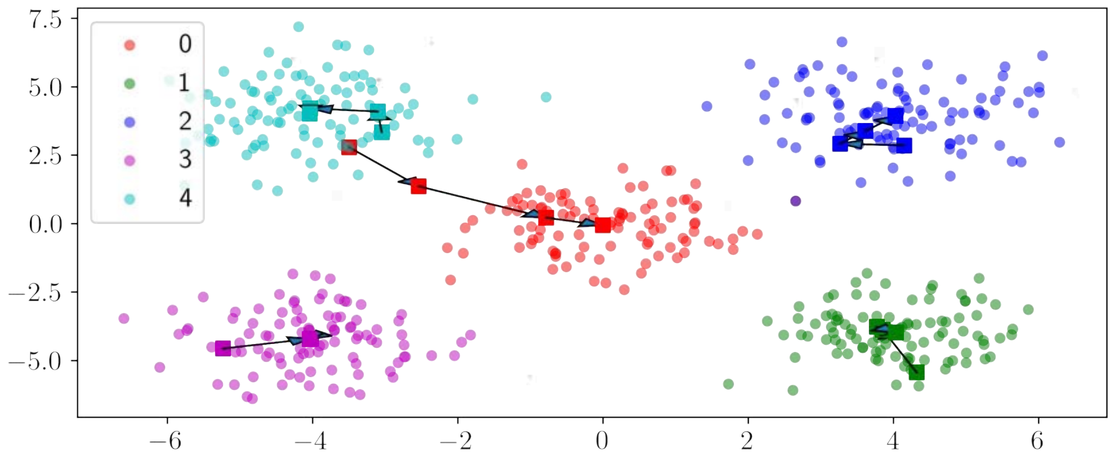
نحوه انتخاب k: با افزایش k، مقدار بهینه تابع J^{clust} کاهش می‌یابد و در حالت حدی $k = N$، این مقدار به کمترین مقدار مجاز یعنی ۰ می‌رسد. معمولاً برای انتخاب k الگوریتم k-means برای یک دسته مقدار پشت سرهم از k اجرا می‌شود. نمودار J^{clust} برحسب k به دست می‌آید. در این نمودار معمولاً می‌توان مقدار k_{best} را پیدا نمود به گونه‌ای که کاهش J^{clust} برای تغییر از $k_{best}-1$ به k_{best} چشمگیر و برای تغییر از k_{best} به $k_{best}+1$ اندک است.

مثال: پنج توزیع گاوسی دو متغیره با بردارهای میانگین $\begin{bmatrix}+4\\+4\end{bmatrix}$ ، $\begin{bmatrix}+4\\-4\end{bmatrix}$ ، $\begin{bmatrix}-4\\+4\end{bmatrix}$ ، $\begin{bmatrix}-4\\-4\end{bmatrix}$ و $\begin{bmatrix}0\\0\end{bmatrix}$ و ماتریس کوواریانس یکسان $\begin{bmatrix}1&0\\0&1\end{bmatrix}$ در نظر بگیرید. از هر کدام از توزیع های فوق ۱۰۰ داده تولید نمایید که در مجموع ۵۰۰ داده خواهد شد. با انتخاب تصادفی ۵ داده به عنوان نماینده گروه اولیه، شش تکرار از الگوریتم K-means را انجام دهید و نحوه تغییر موقعیت نماینده گروه‌ها را نشان دهید.

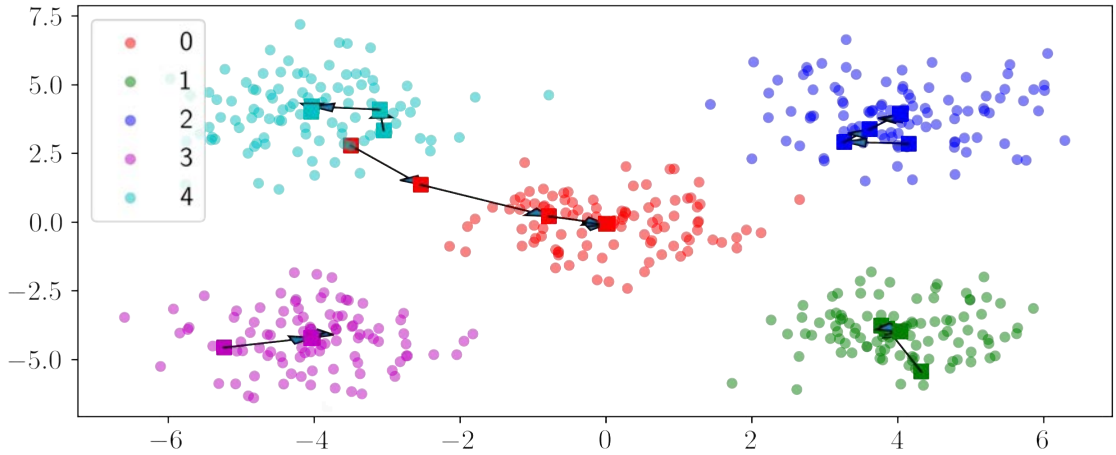




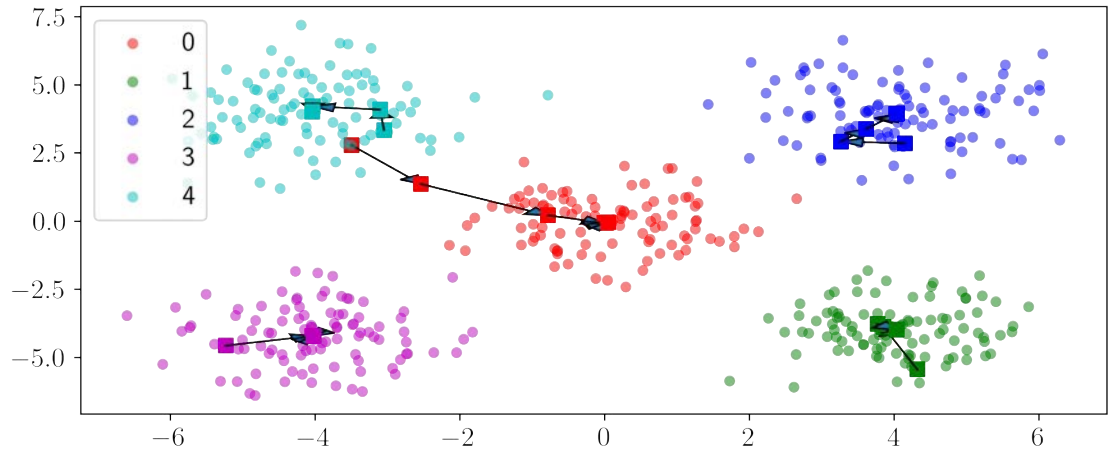




7.5
5.0
2.5
0.0
−2.5
−5.0
−6
−4
−2
0
2
4
6
● 0
● 1
● 2
● 3
● 4



7.5
5.0
2.5
0.0
−2.5
−5.0
−6
−4
−2
0
2
4
6
● 0
● 1
● 2
● 3
● 4



7.5
5.0
2.5
0.0
−2.5
−5.0
−6
−4
−2
0
2
4
6
● 0
● 1
● 2
● 3
● 4