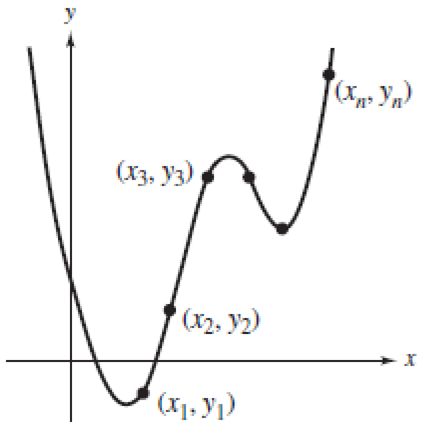
Matematyczne Metody Fizyki I

Wykład 14

Zastosowania układów równań liniowych



<u>Problem:</u> Niech zbiór danych będzie reprezentowany przez n punktów na płaszczyźnie:

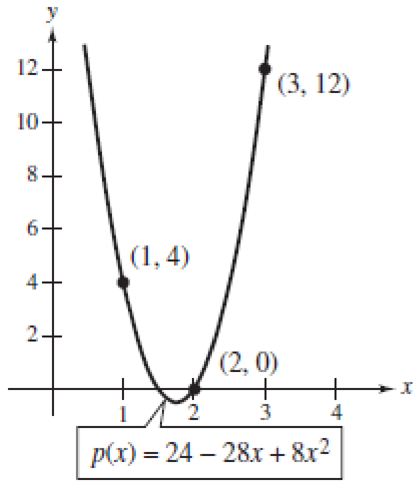
$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)$$

Chcemy znaleźć wielomian stopnia $n-1$: $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$

który przechodzi przez te punkty. Jeśli wszystkie współrzędne x są różne, wtedy istnieje dokładnie jeden wielomian stopnia $n-1$ lub niższego przechodzący przez te punkty.

Jego współczynniki wyznaczamy z układu równań:

$$\begin{cases} a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} = y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} = y_2 \\ \vdots \\ a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_{n-1} x_n^{n-1} = y_n \end{cases}$$



<u>Przykład:</u> Znajdź wielomian stopnia drugiego przechodzący przez punkty: $(1,4),\ (2,0),\ (3,12).$

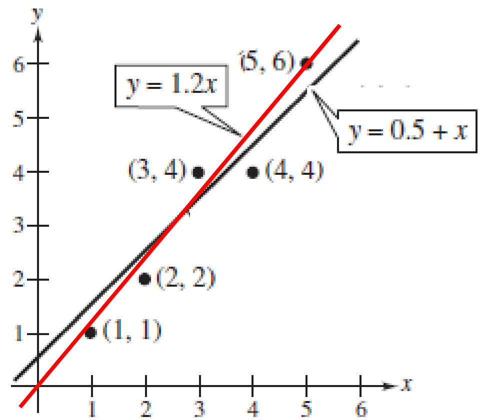
$$\begin{aligned} p(1) &= a_0 + a_1 + a_2 = 4 \\ p(2) &= a_0 + 2a_1 + 4a_2 = 0 \\ p(3) &= a_0 + 3a_1 + 9a_2 = 12 \end{aligned}$$

Rozwiązując układ równań znajdujemy współczynniki :

Dopasowanie linii prostej

Problem: Znajdź linię prostą, która jest najlepiej dopasowana do punktów:

$$(1,1),\ (2,2),\ (3,4),\ (4,4),\ (5,6)$$



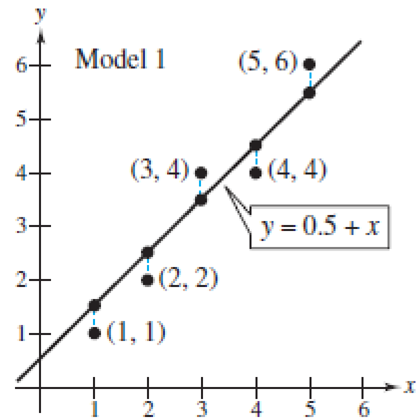
Jako miarę dopasowania linii prostej do punktów wyznaczamy sumę kwadratów odstępstw wartości funkcji od punktów doświadczalnych:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

Znajdujemy odpowiednie błędy dla powyższych dwóch prostych:

$$f(x) = 0.5 + x \;\Rightarrow\; \sum_{i=1}^{5} \left(y_i - f(x_i)\right)^2 = 1.25$$

$$f(x) = 1.2x \;\Rightarrow\; \sum_{i=1}^{5} \left(y_i - f(x_i)\right)^2 = 1.00$$



Definicja: Prostą regresji metody najmniejszych kwadratów dla zbioru punktów $(x_i, y_i),\ i = 1,\dots,n$ nazywamy funkcję

$$f(x) = a_0 + a_1 x$$

która minimalizuje sumę kwadratów błędów

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2$$

Metoda najmniejszych kwadratów

Ogólnie możemy napisać:

$$\begin{aligned} y_1 &= f(x_1) + [y_1 - f(x_1)] = a_0 + a_1 x_1 + e_1 \\ y_2 &= f(x_2) + [y_2 - f(x_2)] = a_0 + a_1 x_2 + e_2 \\ &\vdots \qquad\qquad\qquad\qquad\qquad\quad \vdots \\ y_n &= f(x_n) + [y_n - f(x_n)] = a_0 + a_1 x_n + e_n \end{aligned}$$

lub w postaci macierzowej $\mathbf{Y} = \mathbf{AX} + \mathbf{E}$ gdzie

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ & \\ 1 & x_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, \quad \mathbf{E} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$

Twierdzenie: Dla modelu regresji liniowej $\mathbf{Y} = \mathbf{AX} + \mathbf{E}$ współczynniki linii regresji metody najmniejszych kwadratów dane są przez

$$\mathbf{X} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{T}}\mathbf{Y}$$

natomiast suma kwadratów błędów przez $\mathbf{E}^{\mathrm{T}}\mathbf{E}$.

Metoda najmniejszych kwadratów

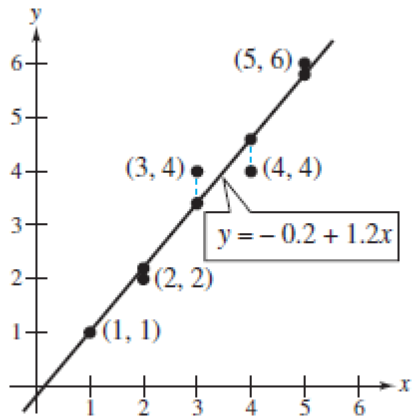
<u>Przykład:</u> Znajdź prostą regresji metody najmniejszych kwadratów dla punktów

$$(1,1),\ (2,2),\ (3,4),\ (4,4),\ (5,6)$$

$$\mathbf{Y} = \begin{pmatrix} 1 \\ 2 \\ 4 \\ 4 \\ 6 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \qquad \Rightarrow \quad \mathbf{X} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Y} = \begin{pmatrix} -0.2 \\ 1.2 \end{pmatrix}$$

A więc prosta regresji metody najmniejszych kwadratów ma postać:

$$y = -0.2 + 1.2x$$



Natomiast suma kwadratów błędów wynosi:

$$\mathbf{E}^T\mathbf{E} = \left(\mathbf{Y} - \mathbf{A}\mathbf{X}\right)^T\left(\mathbf{Y} - \mathbf{A}\mathbf{X}\right) =$$

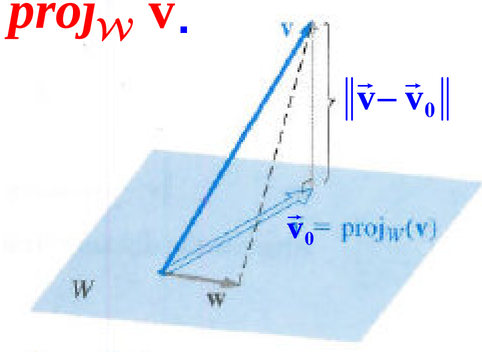
$$= \left(\mathbf{Y} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Y}\right)^T\left(\mathbf{Y} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Y}\right) = 0.8$$

I jest mniejsza od wyznaczonych wcześniej dla innych prostych.

Najlepsze przybliżenie

Definicja: Jeśli $\mathcal{W}$ jest podprzestrzenią przestrzeni liniowej $\mathcal{V}$ oraz $\vec{v} \in \mathcal{V}$ to najlepszym przybliżeniem wektora $\vec{v}$ w $\mathcal{W}$ jest wektor $\vec{v}_0$ taki, że dla każdego wektora $\vec{w} \in \mathcal{W}$ różnego od $\vec{v}_0$ zachodzi $\|\vec{v} - \vec{v}_0\| < \|\vec{v} - \vec{w}\|$

Twierdzenie: Jeśli $\mathcal{W}$ jest podprzestrzenią przestrzeni $\mathcal{V}$ z iloczynem skalarnym oraz $\vec{v} \in \mathcal{V}$ wówczas najlepszym przybliżeniem wektora $\vec{v}$ w $\mathcal{W}$ jest $proj_{\mathcal{W}}\,\vec{v}$.



Dowód: Niech $\vec{w} \in \mathcal{W}$ oraz $\vec{w} \neq proj_{\mathcal{W}}\,\vec{v}$

$$\vec{v} - \vec{w} = \underbrace{\left(\vec{v} - proj_{\mathcal{W}}\,\vec{v}\right)}_{\in \mathcal{W}^{\perp}} + \underbrace{\left(proj_{\mathcal{W}}\,\vec{v} - \vec{w}\right)}_{\in \mathcal{W}}$$

Korzystając z twierdzenia Pitagorasa mamy:

$$\|\vec{v} - \vec{w}\|^2 = \|\vec{v} - proj_{\mathcal{W}}\,\vec{v}\|^2 + \|proj_{\mathcal{W}}\,\vec{v} - \vec{w}\|^2 \quad \Rightarrow \quad \|\vec{v} - proj_{\mathcal{W}}\,\vec{v}\| < \|\vec{v} - \vec{w}\|$$

Przykład: Dane są wektory $\vec{u}_1 = (1,2,-1)^T$, $\vec{u}_2 = (5,-2,1)^T$, $\vec{v} = (3,2,5)^T$

Znajdź najlepsze przybliżenie wektora $\vec{v}$ w przestrzeni $\mathcal{W} = span(\vec{u}_1, \vec{u}_2)$

$$proj_{\mathcal{W}}\,\vec{v} = \left(\frac{\vec{u}_1 \cdot \vec{v}}{\vec{u}_1 \cdot \vec{u}_1}\right)\vec{u}_1 + \left(\frac{\vec{u}_2 \cdot \vec{v}}{\vec{u}_2 \cdot \vec{u}_2}\right)\vec{u}_2 = \frac{1}{5}\begin{pmatrix}15\\-2\\1\end{pmatrix}$$

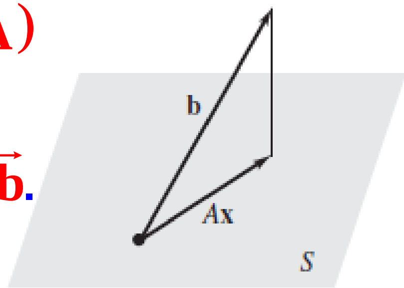
$$\|perp_{\mathcal{W}}\,\vec{v}\| = \|\vec{v} - proj_{\mathcal{W}}\,\vec{v}\| = \frac{12}{\sqrt{5}}$$

Metoda najmniejszych kwadratów

Rozważamy układ równań $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ gdzie $\mathbf{A}_{m\times n}$ i $\vec{\mathbf{b}} \in \mathcal{R}^m$

Ponieważ układ jest sprzeczny, więc można założyć, że $\vec{\mathbf{b}} \notin \mathcal{S} = \mathcal{R}(\mathbf{A})$

Szukamy wektora $\vec{\mathbf{x}}$ który minimalizuje wielkość $\left\|\mathbf{A}\vec{\mathbf{x}} - \vec{\mathbf{b}}\right\|$



Chcemy znaleźć taki wektor $\mathbf{A}\vec{\mathbf{x}} \in \mathcal{S}$ który jest najbliższy wektorowi $\vec{\mathbf{b}}$.

Wiemy, że takim wektorem jest $\mathbf{A}\vec{\mathbf{x}}_0 = proj_{\mathcal{S}}\,\vec{\mathbf{b}}$

Ponieważ wektor $\mathbf{A}\vec{\mathbf{x}}_0 - \vec{\mathbf{b}} = proj_{\mathcal{S}}\,\vec{\mathbf{b}} - \vec{\mathbf{b}}$ jest ortogonalny do $\mathcal{S}$, więc

$$\mathbf{A}\vec{\mathbf{x}}_0 - \vec{\mathbf{b}} \in \mathcal{R}(\mathbf{A})^{\perp} = \mathcal{N}(\mathbf{A}^{\mathrm{T}})$$

A to oznacza, że:

$$\mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\vec{\mathbf{x}}_0 - \vec{\mathbf{b}}\right) = 0 \quad \Rightarrow \quad \mathbf{A}^{\mathrm{T}}\mathbf{A}\vec{\mathbf{x}}_0 - \mathbf{A}^{\mathrm{T}}\vec{\mathbf{b}} = 0 \quad \Rightarrow \quad \vec{\mathbf{x}}_0 = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\vec{\mathbf{b}}$$

Metoda najmniejszych kwadratów sprowadza się do rozwiązania układu $n\times n$ równań liniowych $\mathbf{A}^{\mathrm{T}}\mathbf{A}\vec{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\vec{\mathbf{b}}$. Te równania nazywane są równaniami normalnymi problemu najmniejszych kwadratów $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$.

Układ równań normalnych nigdy nie jest sprzeczny, ale może być nieoznaczony.

Układ ten jest oznaczony gdy rząd macierzy $\mathbf{A}_{m\times n}$ wynosi n.

<u>Definicja:</u> Jeśli macierz $\mathbf{A}_{m\times n}$ ma liniowo niezależne kolumny, to macierzą pseudo-odwrotną do macierzy A nazywamy macierz $\mathbf{A}^{+}_{n\times m} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$

Metoda najmniejszych kwadratów

Przykład: Korzystając z metody najmniejszych kwadratów dopasuj wielomian kwadratowy $y = a_0 + a_1 x + a_2 x^2$ do następujących danych:

rok	1980	1985	1990	1995	2000	2005
populacja	4.5	4.8	5.3	5.7	6.1	6.5

$$\mathrm{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \\ 1 & 20 & 400 \\ 1 & 25 & 625 \end{pmatrix}, \quad \vec{\mathrm{x}} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}, \quad \vec{\mathrm{b}} = \begin{pmatrix} 4.5 \\ 4.8 \\ 5.3 \\ 5.7 \\ 6.1 \\ 6.5 \end{pmatrix}$$

$$\mathrm{A^T A}\vec{\mathrm{x}} = \mathrm{A^T}\vec{\mathrm{b}} \Rightarrow \begin{pmatrix} 6 & 75 & 1375 \\ 75 & 1375 & 28125 \\ 1375 & 28125 & 611875 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 32.9 \\ 447 \\ 8435 \end{pmatrix} \Rightarrow \vec{\mathrm{x}} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \approx \begin{pmatrix} 4.5 \\ 0.08 \\ 0 \end{pmatrix}$$

A więc metodą najmniejszych kwadratów otrzymujemy zależność liniową $y = 4.5 + 0.08\,x$

Przewidywanie dla roku 2010: $y = 4.5 + 0.08 \cdot 30 \approx 6.9$

Wartości osobliwe macierzy

Z dowolnej macierzy rzeczywistej $\mathbf{A}_{m\times n}$ można utworzyć macierz symetryczną $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, a więc diagonalizowalną za pomocą ortogonalnej transformacji podobieństwa.
Wartości własne macierzy $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ są rzeczywiste i nieujemne:

$$0 \le \left\|\mathbf{A}\,\vec{\mathbf{v}}\right\|^2 = \left(\mathbf{A}\,\vec{\mathbf{v}}\right)\cdot\left(\mathbf{A}\,\vec{\mathbf{v}}\right) = \left(\mathbf{A}\,\vec{\mathbf{v}}\right)^T\left(\mathbf{A}\,\vec{\mathbf{v}}\right) = \vec{\mathbf{v}}^T\,\mathbf{A}^T\,\mathbf{A}\,\vec{\mathbf{v}} = \vec{\mathbf{v}}^T\,\lambda\,\vec{\mathbf{v}} = \lambda\left\|\vec{\mathbf{v}}\right\|^2 = \lambda$$

Definicja: Wartościami osobliwymi macierzy $\mathbf{A}_{m\times n}$ nazywamy pierwiastki kwadratowe wartości własnych macierzy $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, i oznaczamy przez $\sigma_1, \sigma_2, \ldots, \sigma_n$ gdzie $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_n$.

Twierdzenie: Niech $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$ będą wartościami własnymi symetrycznej macierzy $\mathbf{A}_{n\times n}$ stowarzyszonej z formą kwadratową $f\left(\vec{\mathbf{x}}\right) = \vec{\mathbf{x}}^T\,\mathbf{A}\,\vec{\mathbf{x}}$. Jeśli $\left\|\vec{\mathbf{x}}\right\| = 1$ to wówczas:

(a) $\lambda_1 \ge f\left(\vec{\mathbf{x}}\right) \ge \lambda_n$

(b) $\max f\left(\vec{\mathbf{x}}\right) = \lambda_1$ i występuje dla $\vec{\mathbf{x}}$ będącego wektorem własnym do w. własnej λ_1.

(c) $\min f\left(\vec{\mathbf{x}}\right) = \lambda_n$ i występuje dla $\vec{\mathbf{x}}$ będącego wektorem własnym do w. własnej λ_n.

D: (a) $f\left(\vec{\mathbf{x}}\right) = \vec{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{x}}^{\mathrm{T}}\,\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\mathbf{A}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\vec{\mathbf{x}} = \vec{\mathbf{y}}^{\mathrm{T}}\,\mathbf{D}\,\vec{\mathbf{y}} \;\Rightarrow\; \left\|\vec{\mathbf{y}}\right\| = \sqrt{\vec{\mathbf{y}}^{\mathrm{T}}\,\vec{\mathbf{y}}} = \sqrt{\vec{\mathbf{x}}^{\mathrm{T}}\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\vec{\mathbf{x}}} = \sqrt{\vec{\mathbf{x}}^{\mathrm{T}}\vec{\mathbf{x}}} = 1$

$$f\left(\vec{\mathbf{x}}\right) = \vec{\mathbf{x}}^{\mathrm{T}}\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{y}}^{\mathrm{T}}\,\mathbf{D}\,\vec{\mathbf{y}} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \le \lambda_1\left(y_1^2 + y_2^2 + \cdots + y_n^2\right) = \lambda_1$$

(b) $\mathbf{A}\vec{\mathbf{q}}_1 = \lambda_1\,\vec{\mathbf{q}}_1 \;\Rightarrow\; f\left(\vec{\mathbf{q}}_1\right) = \vec{\mathbf{q}}_1^T\,\mathbf{A}\,\vec{\mathbf{q}}_1 = \vec{\mathbf{q}}_1^T\,\lambda_1\,\vec{\mathbf{q}}_1 = \lambda_1\left(\vec{\mathbf{q}}_1^T\,\vec{\mathbf{q}}_1\right) = \lambda_1$

Rozkład według wartości osobliwych

Przykład: Znajdź wartości osobliwe macierzy A.

$$\mathbf{A}=\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix} \Rightarrow \mathbf{A}^T\mathbf{A}=\begin{pmatrix}1&1&0\\1&0&1\end{pmatrix}\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}=\begin{pmatrix}2&1\\1&2\end{pmatrix} \Rightarrow \begin{cases}\sigma_1=\sqrt{\lambda_1}=\sqrt{3}\\ \sigma_2=\sqrt{\lambda_2}=1\end{cases}$$

Niech $\|\vec{\mathbf{x}}\|=1$ wówczas: $\|\mathbf{A}\vec{\mathbf{x}}\|^2=(\mathbf{A}\vec{\mathbf{x}})^T(\mathbf{A}\vec{\mathbf{x}})=\vec{\mathbf{x}}^T\mathbf{A}^T\mathbf{A}\vec{\mathbf{x}}=2x_1^2+2x_1x_2+2x_2^2$

Maksimum i minimum tej formy kwadratowej wynosi odpowiednio $\lambda_1=3$ oraz $\lambda_2=1$ i występuje dla odpowiadających im wektorów własnych:

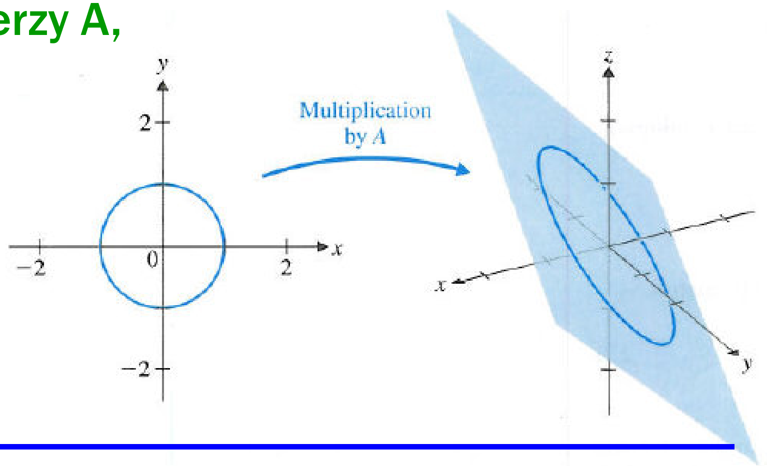
$$\vec{\mathbf{v}}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix} \text{ oraz } \vec{\mathbf{v}}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix}$$

Transformacja liniowa określona za pomocą macierzy A, przekształca $\mathcal{R}^2$ w płaszczyznę $x\text{-}y\text{-}z=0$ w $\mathcal{R}^3$.

Natomiast jednostkowy okrąg w $\mathcal{R}^2$ przechodzi w elipsę leżącą w płaszczyźnie $x\text{-}y\text{-}z=0$ w $\mathcal{R}^3$.

Długości półosi elipsy wynoszą odpowiednio:

$$\|\mathbf{A}\,\vec{\mathbf{v}}_1\|=\sigma_1=\sqrt{3} \text{ oraz } \|\mathbf{A}\,\vec{\mathbf{v}}_2\|=\sigma_2=1$$



Rozkład według wartości osobliwych

Twierdzenie: Niech macierz $\mathbf{A}_{m\times n}$ ma wartości osobliwe $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ oraz $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$. Wówczas istnieją ortogonalne macierze $\mathbf{U}_{m\times m}$ i $\mathbf{V}_{n\times n}$ oraz macierz $\Sigma_{m\times n}$ takie, że $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$

D: Konstruujemy macierz $\mathbf{V}$ z ortonormalnych wektorów własnych macierzy $\mathbf{A}^\mathrm{T}\mathbf{A}$:

$$\mathbf{V} = \left[\vec{\mathbf{v}}_1 \;\; \cdots \;\; \vec{\mathbf{v}}_n\right]$$

Konstrukcja macierzy $\mathbf{U}$:

$$\left(\mathbf{A}\,\vec{\mathbf{v}}_i\right)\cdot\left(\mathbf{A}\,\vec{\mathbf{v}}_j\right) = \left(\mathbf{A}\,\vec{\mathbf{v}}_i\right)^T\left(\mathbf{A}\,\vec{\mathbf{v}}_j\right) = \vec{\mathbf{v}}_i^{\,T}\,\mathbf{A}^T\,\mathbf{A}\,\vec{\mathbf{v}}_j = \vec{\mathbf{v}}_i^{\,T}\,\lambda_j\,\vec{\mathbf{v}}_j = \lambda_j\left(\vec{\mathbf{v}}_i\cdot\vec{\mathbf{v}}_j\right) = 0$$

$$\sigma_i = \left\|\mathbf{A}\,\vec{\mathbf{v}}_i\right\| \;\Rightarrow\; \vec{\mathbf{u}}_i = \frac{1}{\sigma_i}\mathbf{A}\,\vec{\mathbf{v}}_i \quad \text{dla} \quad i = 1,\ldots,r$$

Zbiór wektorów $\vec{\mathbf{u}}_i$ uzupełniamy aby tworzył bazę ortonormalną w $\mathcal{R}^m$:

$$\mathbf{U} = \left[\vec{\mathbf{u}}_1 \;\; \cdots \;\; \vec{\mathbf{u}}_m\right]$$

$$\mathbf{AV} = \mathbf{A}\left[\vec{\mathbf{v}}_1 \;\; \cdots \;\; \vec{\mathbf{v}}_n\right] = \left[\mathbf{A}\,\vec{\mathbf{v}}_1 \;\; \cdots \;\; \mathbf{A}\,\vec{\mathbf{v}}_n\right] = \left[\mathbf{A}\,\vec{\mathbf{v}}_1 \;\; \cdots \;\; \mathbf{A}\,\vec{\mathbf{v}}_r \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}\right] =$$

$$= \left[\sigma_1\,\vec{\mathbf{u}}_1 \;\; \cdots \;\; \sigma_r\,\vec{\mathbf{u}}_r \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}\right] = \left[\vec{\mathbf{u}}_1 \;\; \cdots \;\; \vec{\mathbf{u}}_m\right]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{O} \\ 0 & \cdots & \sigma_r & \\ \hline & \mathbf{O} & & \mathbf{O} \end{array}\right] = \mathbf{U}\,\Sigma$$

Rozkład według wartości osobliwych

Przykład: Znajdź rozkład macierzy A według wartości osobliwych.

$$A=\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}\;\Rightarrow\; A^TA=\begin{pmatrix}2&1\\1&2\end{pmatrix}\;\Rightarrow\begin{cases}\sigma_1=\sqrt{\lambda_1}=\sqrt{3}\\ \sigma_2=\sqrt{\lambda_2}=1\end{cases}\;\Rightarrow\; \vec{v}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},\quad \vec{v}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix}$$

Macierze Σ oraz V mają postać:

$$\Sigma=\begin{pmatrix}\sqrt{3}&0\\0&1\\0&0\end{pmatrix}\qquad V=\begin{pmatrix}1/\sqrt{2}&-1/\sqrt{2}\\1/\sqrt{2}&1/\sqrt{2}\end{pmatrix}$$

Znajdujemy macierz U:

$$\vec{u}_1=\frac{1}{\sigma_1}A\vec{v}_1=\frac{1}{\sqrt{3}}\begin{pmatrix}1&1\\1&0\\0&1\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}=\frac{1}{\sqrt{6}}\begin{pmatrix}2\\1\\1\end{pmatrix}\qquad \vec{u}_2=\frac{1}{\sigma_2}A\vec{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}0\\-1\\1\end{pmatrix}$$

Aby wektory $\vec{u}_i$ tworzyły bazę ortonormalną w $\mathcal{R}^3$ znajdujemy trzeci wektor korzystając z metody Grama-Schmidta:

$$\vec{u}_3=\frac{1}{\sqrt{3}}\begin{pmatrix}-1\\1\\1\end{pmatrix}\;\Rightarrow\; U=\frac{1}{\sqrt{6}}\begin{pmatrix}2&0&-\sqrt{2}\\1&-\sqrt{3}&\sqrt{2}\\1&\sqrt{3}&\sqrt{2}\end{pmatrix}$$

Łatwo sprawdzić, że $A=U\Sigma V^T$

Rozkład według wartości osobliwych

Twierdzenie: Niech macierz $\mathbf{A}_{m\times n}$ ma wartości osobliwe $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ oraz $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$, oraz niech $\vec{\mathbf{u}}_1, \ldots, \vec{\mathbf{u}}_r$ i $\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_r$ będą lewo i prawostronnymi wektorami osobliwymi odpowiadającymi tym wartościom osobliwym, wówczas:

$$\mathbf{A} = \sigma_1 \vec{\mathbf{u}}_1 \vec{\mathbf{v}}_1^T + \cdots + \sigma_r \vec{\mathbf{u}}_r \vec{\mathbf{v}}_r^T$$

D:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \left[\vec{\mathbf{u}}_1 \;\cdots\; \vec{\mathbf{u}}_r \,\middle|\, \vec{\mathbf{u}}_{r+1} \;\cdots\; \vec{\mathbf{u}}_m\right] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{O} \\ 0 & \cdots & \sigma_r & \\ \hline & \mathbf{O} & & \mathbf{O} \end{array}\right] \left[\begin{array}{c} \vec{\mathbf{v}}_1^T \\ \vdots \\ \vec{\mathbf{v}}_r^T \\ \hline \vec{\mathbf{v}}_{r+1}^T \\ \vdots \\ \vec{\mathbf{v}}_n^T \end{array}\right] =$$

$$= \left[\vec{\mathbf{u}}_1 \;\cdots\; \vec{\mathbf{u}}_r\right] \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \vec{\mathbf{v}}_1^T \\ \vdots \\ \vec{\mathbf{v}}_r^T \end{bmatrix} + \left[\vec{\mathbf{u}}_{r+1} \;\cdots\; \vec{\mathbf{u}}_m\right] \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \vec{\mathbf{v}}_{r+1}^T \\ \vdots \\ \vec{\mathbf{v}}_n^T \end{bmatrix} =$$

$$= \left[\sigma_1 \vec{\mathbf{u}}_1 \;\cdots\; \sigma_r \vec{\mathbf{u}}_r\right] \begin{bmatrix} \vec{\mathbf{v}}_1^T \\ \vdots \\ \vec{\mathbf{v}}_r^T \end{bmatrix} = \sigma_1 \vec{\mathbf{u}}_1 \vec{\mathbf{v}}_1^T + \cdots + \sigma_r \vec{\mathbf{u}}_r \vec{\mathbf{v}}_r^T$$

Rozkład według wartości osobliwych

Twierdzenie: Niech macierz $\mathbf{A}_{m\times n}$ ma wartości osobliwe $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ oraz $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_n = 0$ oraz $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$. Wtedy mamy:

a) rząd macierzy A wynosi r,

b) $\left[\vec{\mathbf{u}}_1, \;\cdots,\; \vec{\mathbf{u}}_r\right]$ jest ortonormalną bazą w przestrzeni $\mathcal{R}$(A),

c) $\left[\vec{\mathbf{u}}_{r+1}, \;\cdots,\; \vec{\mathbf{u}}_m\right]$ jest ortonormalną bazą przestrzeni $\mathcal{N}$(A^T),

d) $\left[\vec{\mathbf{v}}_1, \;\cdots,\; \vec{\mathbf{v}}_r\right]$ jest ortonormalną bazą w przestrzeni $\mathcal{R}$(A^T),

e) $\left[\vec{\mathbf{v}}_{r+1}, \;\cdots,\; \vec{\mathbf{v}}_n\right]$ jest ortonormalną bazą w przestrzeni $\mathcal{N}$(A).

D: (a) rz(A) = rz(UΣV^T) = rz(Σ) = r

(b) zbiór wektorów $\vec{\mathbf{u}}_i$ (i=1, ..., r) ortonormalnych + dim($\mathcal{R}$(A)) = r czyli baza w $\mathcal{R}$(A)

(c) $\mathcal{R}$(A) oraz $\mathcal{N}$(A^T) są ortogonalnymi podprzestrzeniami $\mathcal{R}^m = \mathcal{R}(\mathbf{A}) \oplus \mathcal{N}(\mathbf{A}^T)$

czyli $\left[\vec{\mathbf{u}}_{r+1}, \;\cdots,\; \vec{\mathbf{u}}_m\right]$ jest ortonormalną bazą w $\mathcal{N}$(A^T)

(d) zbiór wektorów $\vec{\mathbf{v}}_i$ (i=1, ..., r) ortonormalnych + dim($\mathcal{R}$(A^T)) = r czyli baza w $\mathcal{R}$(A^T)

(e) $\mathcal{R}$(A^T) oraz $\mathcal{N}$(A) są ortogonalnymi podprzestrzeniami $\mathcal{R}^n = \mathcal{R}(\mathbf{A}^T) \oplus \mathcal{N}(\mathbf{A})$

czyli $\left[\vec{\mathbf{v}}_{r+1}, \;\cdots,\; \vec{\mathbf{v}}_n\right]$ jest ortonormalną bazą w $\mathcal{N}$(A^T)

Macierz pseudoodwrotna

Definicja: Niech macierz $\mathbf{A}_{m\times n}$ ma rozkład na wartości osobliwe $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$ gdzie $\Sigma = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ oraz $\mathbf{D}$ jest macierzą diagonalną $r\times r$ zawierającą niezerowe wartości osobliwe macierzy $\mathbf{A}$. Macierzą pseudoodwrotną do macierzy $\mathbf{A}$ nazywamy macierz:

$$\mathbf{A}^+_{n\times m} = \mathbf{V}\,\Sigma^+\,\mathbf{U}^T \quad \text{gdzie} \quad \Sigma^+_{n\times m} = \begin{bmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Uwaga: Powyższa definicja sprowadza się do poprzedniej w przypadku gdy kolumny macierzy $\mathbf{A}$ są liniowo niezależne.

Przykład: Znajdź macierz pseudoodwrotną do macierzy $\mathbf{A}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$$

$$\mathbf{A}^+ = \mathbf{V}\,\Sigma^+\,\mathbf{U}^T = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{pmatrix} =$$

$$= \begin{pmatrix} 1/3 & 2/3 & -1/3 \\ 1/3 & -1/3 & 2/3 \end{pmatrix}$$

Metoda najmniejszych kwadratów

<u>Uwaga:</u> **Gdy kolumny macierzy** $\mathbf{A}$ są liniowo zależne, wtedy $\mathbf{A}^T\mathbf{A}$ nie jest odwracalna i układ równań normalnych $\mathbf{A}^T\mathbf{A}\vec{\mathbf{x}} = \mathbf{A}^T\vec{\mathbf{b}}$ ma nieskończenie wiele rozwiązań.

Twierdzenie: Problem najmniejszych kwadratów $\mathbf{A}\vec{\mathbf{x}} = \vec{\mathbf{b}}$ ma zawsze jednoznaczne rozwiązanie w postaci wektora $\vec{\mathbf{x}}_0$ o najmniejszej długości danego przez: $\vec{\mathbf{x}}_0 = \mathbf{A}^+\vec{\mathbf{b}}$

D: Niech macierz $A_{m\times n}$ ma rząd r oraz rozkład na wartości osobliwe $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$

Oznaczenia: $\vec{\mathbf{y}} = [\vec{\mathbf{y}}_1 \,|\, \vec{\mathbf{y}}_2] = \mathbf{V}^T\vec{\mathbf{x}},\ \vec{\mathbf{y}}_1 \in R^r$ **oraz** $\vec{\mathbf{c}} = [\vec{\mathbf{c}}_1 \,|\, \vec{\mathbf{c}}_2] = \mathbf{U}^T\vec{\mathbf{b}},\ \vec{\mathbf{c}}_1 \in R^r$

Chcemy zminimalizować $\|\vec{\mathbf{b}} - \mathbf{A}\vec{\mathbf{x}}\|$

$$\|\vec{\mathbf{b}} - \mathbf{A}\vec{\mathbf{x}}\|^2 = \|\mathbf{U}^T(\vec{\mathbf{b}} - \mathbf{A}\vec{\mathbf{x}})\|^2 = \|\mathbf{U}^T(\vec{\mathbf{b}} - \mathbf{U}\Sigma\mathbf{V}^T\vec{\mathbf{x}})\|^2 = \|\mathbf{U}^T\vec{\mathbf{b}} - \mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T\vec{\mathbf{x}}\|^2 =$$

$$= \|\vec{\mathbf{c}} - \Sigma\vec{\mathbf{y}}\|^2 = \left\|\begin{bmatrix}\vec{\mathbf{c}}_1\\ \vec{\mathbf{c}}_2\end{bmatrix} - \begin{bmatrix}\mathbf{D} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\begin{bmatrix}\vec{\mathbf{y}}_1\\ \vec{\mathbf{y}}_2\end{bmatrix}\right\|^2 = \left\|\begin{bmatrix}\vec{\mathbf{c}}_1 - \mathbf{D}\vec{\mathbf{y}}_1\\ \vec{\mathbf{c}}_2\end{bmatrix}\right\|$$

Powyższe wyrażenie przyjmuje wartość minimalną dla $\vec{\mathbf{c}}_1 - \mathbf{D}\vec{\mathbf{y}}_1 = \mathbf{0} \quad\Rightarrow\quad \vec{\mathbf{y}}_1 = \mathbf{D}^{-1}\vec{\mathbf{c}}_1$

Rozwiązanie problemu najmniejszych kwadratów o minimalnej długości ma więc postać:

$$\vec{\mathbf{x}}_0 = \mathbf{V}\vec{\mathbf{y}}_0 = \mathbf{V}\begin{bmatrix}\mathbf{D}^{-1}\vec{\mathbf{c}}_1\\ \vec{\mathbf{0}}\end{bmatrix} = \mathbf{V}\begin{bmatrix}\mathbf{D}^{-1} & \mathbf{O}\\ \mathbf{O} & \mathbf{O}\end{bmatrix}\begin{bmatrix}\vec{\mathbf{c}}_1\\ \vec{\mathbf{c}}_2\end{bmatrix} = \mathbf{V}\Sigma^+\vec{\mathbf{c}} = \mathbf{V}\Sigma^+\mathbf{U}^T\vec{\mathbf{b}} = \mathbf{A}^+\vec{\mathbf{b}}$$