# Zajęcie 0.2. Macierz Pseudoodwrotna. Najmniejsze kwadraty. Regresja

**Abstract**

Celem jest nabycie podstawowej znajomości użycia rozkładu SVD w celu rozwiązywania problemów liniowych o dużej skali.

## 1. Podstawowe pojęcia

Wiele układów fizycznych można przedstawić jako liniowy układ równań,

$$Ax = b$$

gdzie znana jest macierz ograniczeń $A$ i wektor $b$, a wektor $x$ jest nieznany. Jeśli $A$ jest kwadratową, odwracalną macierzą (tj. $A$ ma niezerowy wyznacznik), to istnieje unikalne rozwiązanie $x$ dla każdego $b$. Jednakże, gdy $A$ jest singularna lub prostokątna, może istnieć jedno, żadne lub nieskończenie wiele rozwiązań, w zależności od konkretnego $b$ oraz przestrzeni kolumn i wierszy $A$.

Najpierw rozważmy układ "niedookreślony", w którym $A \in \mathbb{R}^{n\times m}$ i $n << m$, tak że jest mniej równań niż niewiadomych. Ten typ systemu prawdopodobnie będzie miał kolumny obejmujące całe środowisko $\mathbb{R}^n$, ponieważ ma o wiele więcej kolumn, niż jest to wymagane w przypadku bazy liniowo niezależnej. Ogólnie rzecz biorąc, jeśli krótkie $A$ ma $n$ liniowo niezależnych kolumn (tj. jego przestrzeń kolumn obejmuje $\mathbb{R}^n$), to istnieje nieskończenie wiele rozwiązań $x$ dla każdego $b$. System nazywa się niedookreślonym, ponieważ w $b$ nie ma wystarczającej liczby wartości, aby można było go jednoznacznie określić określić $x$ o wyższym wymiarze.

Podobnie rozważmy układ "nadmiernie określony", gdzie $n >> m$ , w którym jest więcej równań niż niewiadomych. Macierz ta nie może mieć $n$ liniowo niezależnych kolumn, zatem jest pewne, że istnieją wektory $b$, które nie mają rozwiązania $x$. W rzeczywistości rozwiązanie $x$ będzie istniało tylko wtedy, gdy $b$ będzie znajdować się w przestrzeni kolumn $A$, tj. $b \in \mathrm{col}(A)$.

W przypadku nadmiernie określonym, gdy nie ma rozwiązania, często chcielibyśmy znaleźć rozwiązanie $x$, które minimalizuje błąd sumy kwadratów $\|Ax - b\|_2^2$ , tzw. rozwiązanie metodą **najmniejszych kwadratów**. Należy zauważyć, że rozwiązanie metodą najmniejszych kwadratów minimalizuje również $\|Ax-b\|_2$. W przypadku niedookreślonym, gdy istnieje nieskończenie wiele rozwiązań, możemy chcieć znaleźć rozwiązanie $x$ z minimalną normą $\|x\|_2$ tak, że $Ax = b$, co jest tzw. **rozwiązaniem o minimalnej normie**.

W przypadku tych ważnych problemów optymalizacyjnych używamy SVD. Po pierwsze, jeśli podstawimy dokładnie skrócony SVD $A = \tilde{U}\tilde{\Sigma}\tilde{V}^*$ w przypadku $A$ możemy "odwrócić" każdą z macierzy $\tilde{U}$,$\tilde{\Sigma}$, $\tilde{V}^*$ z kolei, co skutkuje lewą pseudoodwrotną macierz Moore'a – Penrose'a $A^+$ dla $A$:

$$A^+ := \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^*$$
$$A^+A = \tilde{V}\tilde{V}^*$$

Należy zauważyć, że $A^+A$ będzie równe tożsamości $I_{m\times m}$ tylko wtedy, gdy obcięte SVD przechwytuje wszystkie niezerowe wartości osobliwe; w przeciwnym razie $\tilde{V}\tilde{V}^* \neq Im \times m$ i będzie to jedynie przybliżenie tożsamości. Można to wykorzystać do znalezienia rozwiązań zarówno w zakresie normy minimalnej, jak i metody najmniejszych kwadratów

$$x = \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^*b$$

Założenie, że $\tilde{U}\tilde{U}^*$ jest równe tożsamości, jest jednym z najczęstszych przypadkowych nadużyć SVD. Jednak nadal prawdą jest, że $\tilde{U}\tilde{U}^* = I_{r\times r}$, gdzie $r$ jest rangiem $A$.

Obliczanie pseudoodwrotności $A^+$ jest wydajne obliczeniowo, po wysokich kosztach początkowych obliczenia SVD. Odwracanie macierzy unitarnych $\tilde{U}$ i $\tilde{V}^*$ polega na mnożeniu macierzy przez macierze transpozycji, które są operacjami $O(n^2)$. Odwracanie $\tilde{\Sigma}$ jest jeszcze bardziej efektywne, ponieważ jest to macierz diagonalna wymagająca $O(n)$ operacji. Natomiast odwrócenie gęstej macierzy kwadratowej wymagałoby operacji $O(n^3)$.

### *1.1. Liczba warunkowa*

Liczba warunkowa macierzy $A$ jest miarą wrażliwości mnożenia i odwracania macierzy na błędy na wejściu. Większa liczba warunkowa oznacza wyższą czułość i gorszą wydajność. Liczba warunkowa $\kappa(A)$ jest bezpośrednio powiązana z wartościami osobliwymi macierzy:

$$\kappa(A) := \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Jednym ze sposobów złagodzenia dużej liczby warunkowej jest bardziej agresywne obcięcie SVD, zasadniczo zwiększając efektywną minimalną wartość osobliwą $\sigma_{\min}$. Dzieje się to jednak kosztem zmniejszenia rozmiaru podprzestrzeni $\tilde{U}$ używanej do aproksymacji wyniku.

```
>> kappanew = 1.e-5; % Desired condition number
>> [U,S,V] = svd(A,'econ')
>> r = max(find(diag(S)>max(S(:))*kappanew));
>> invA = V(:,1:r)*inv(S(1:r,1:r))*U(:,1:r)'; % Approximate
```

*1.2. Liniowa regresja jednowymiarowa*

```
x = 3 # True slope
a = np.arange(-2,2,0.25)
a = a.reshape(-1, 1)
b = x*a + np.random.randn(*a.shape) # Add noise
plt.plot(a, x*a, Color='k', LineWidth=2, label='True line')
# True relationship
plt.plot(a, b, 'x', Color='r', MarkerSize = 10, label='Noisy
data') # Noisy measurements
# Compute least-squares approximation with the SVD
U, S, VT = np.linalg.svd(a,full_matrices=False)
xtilde = VT.T @ np.linalg.inv(np.diag(S)) @ U.T @ b # Leastsquare fit
plt.plot(a,xtilde * a,'--',Color='b',LineWidth=4, label='
Regression line')
# Alternative formulations of least squares
xtilde1 = VT.T @ np.linalg.inv(np.diag(S)) @ U.T @ b
xtilde2 = np.linalg.pinv(a) @ b
```

*1.3. Wiele-liniowa regresja*

```
# Load dataset
A = np.loadtxt(os.path.join('..','DATA','hald_ingredients.
csv'),delimiter=',')
b = np.loadtxt(os.path.join('..','DATA','hald_heat.csv'),
delimiter=',')
# Solve Ax=b using SVD
```

```
U, S, VT = np.linalg.svd(A,full_matrices=0)
x = VT.T @ np.linalg.inv(np.diag(S)) @ U.T @ b
plt.plot(b, Color='k', LineWidth=2, label='Heat Data')
plt.plot(A@x, '-o', Color='r', label='Regression')
x = np.linalg.pinv(A)*b # Alternative
```

## 2. Zadanie 2

Zadanie dotyczy obliczenia wieleliniowej regresji z użyciem macierzy psewdoodwrotnej dla zależności

$$y = a * x_1 + b * x_2,$$

gdzie $a$, $b$ są niewiadome, wartości $x_1, x_2, y_2$ określone wariantem zadania.

**Sprawozdania** w postaci:

1. Sprawozdanie (plik .pdf)

2. plik .ipynb

3. pdf-eksport pliku .pynb

zachować w zdalnym repozytorium (np Github) link na który umieściś w sprawozdaniu. Sprawozdanie należy wysłać na e-uczelnię w ustalonym terminem.

## References

## References

[pandasUG] Pandas User's Guide https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html

[DA2016] Data Analysis with Python and pandas using Jupyter Notebook https://dev.socrata.com/blog/2016/02/01/pandas-and-jupyter-notebook.html