

# Losowe projekcje

**Ewa Skubalska-Rafajłowicz**

Politechnika Wrocławska  
Wydział Elektroniki  
Katedra Informatyki Technicznej  
[ewa.rafajlowicz@pwr.edu.pl](mailto:ewa.rafajlowicz@pwr.edu.pl)

**Warszawa 2017**

# Spis treści:

- 1 Wstęp
- 2 Zanurzenia. Przybliżona izometria
- 3 Lemat Johnsona–Lindenstraussa
- 4 Dowody lematu JL
- 5 Projekcje ortogonalne
- 6 Redukcja rozmiaru sieci MLP
- 7 Rozpoznawanie twarzy zachowujące prywatność
- 8 Losowe projekcje wielowymiarowych rozkładów normalnych
- 9 Karty kontrolne - monitorowanie wielowymiarowego procesu
- 10 Phase I
- 11 Losowe projekcje z rzadkimi macierzami przekształcenia
- 12 Własność ograniczonej izometrii macierzy RIP

Losowe projekcje – losowe rzutowania – są jednym z narzędzi redukcji wymiaru.

Pozwalają rozwiązywać w efektywny sposób problemy oryginalnie sformułowane w wysokowymiarowych przestrzeniach.

Pokazane zostaną zastosowania **lematu**

**Johnsona-Lindenstraussa**, który gwarantuje istnienie funkcji transformującej zbiór punktów w przestrzeni euklidesowej do niższej wymiarowej przestrzeni przy zachowaniu przybliżonych odległości pomiędzy punktami.

Transformacja ta może być uzyskana drogą losowych projekcji, a twierdzenia o koncentracji gwarantują wysokie prawdopodobieństwo zachowania normy (bądź metryki) z niewielkim zniekształceniem jej wartości dla skończonej liczby wektorów (pomiedzy skończoną liczbą punktów).

**Omówione zostaną wybrane zastosowania losowych projekcji:**

- **w monitorowaniu wielowymiarowych danych,**
- **w projektowaniu sieci neuronowych oraz**
- **w przetwarzaniu obrazów.**

Niech  $x$  będzie pewnym wektorem w przestrzeni  $\mathcal{R}^d$ . Będziemy rozważać rzut, niekoniecznie ortogonalny, tego wektora na  $k$ -wymiarową losową podprzestrzeń  $\mathcal{R}^d$ .

Możemy to zrealizować w różny sposób. Wyznaczanie optymalnej z punktu widzenia pewnego kryterium podprzestrzeni jest jednym z ważniejszych problemów obliczeniowych [Absil].

W przestrzeniach wysokowymiarowych losowe metody generowania podprzestrzeni mogą być prawie równie skuteczne, a równocześnie realizowalne ze stosunkowo niewielkimi nakładami obliczeniowymi.

## Zanurzenia:

Dana jest pewna przestrzeń metryczna, którą chcemy odwzorować w innej przestrzeni metrycznej, na ogół jest to przestrzeń euklidesowa o niższym wymiarze. Wiadomo, że istnieją nawet skończone przestrzenie metryczne, które nie mogą być izometrycznie zanurzone w przestrzeni euklidesowej [Schoenberg], [Matousek90]. Z punktu widzenia zastosowań rozwiązania przybliżone mogą być wystarczająco satysfakcjonujące, czego przykładem jest wprowadzona przez Kruskala metoda wielowymiarowego skalowania, czy też metoda lokalnych liniowych zanurzeń LLE [Roweis].

# Zanurzenia

Zanurzanie jednej przestrzeni w drugiej prowadzi do:

– Może nastąpić zmiana wymiaru przestrzeni - najczęściej, choć nie koniecznie, jest to redukcja wymiaru przestrzeni.

—Może nastąpić zmiana metryki.

Z obliczeniowego punktu widzenia zadania sformułowane w różnych przestrzeniach izometrycznych mogą być mniej lub bardziej pracochłonne [Indyk], [Matousek], [Vempala].

Przybliżone zanurzenia mogą dopuszczać pewne zniekształcenia odległości w przestrzeni zanurzenia względem oryginalnych odległości zanurzanej przestrzeni, Zniekształcenia te nazywane są **dystorsją zanurzenia**.

# Zanurzenie izometryczne

Dane są przestrzenie metryczne  $(\mathcal{M}, D_{\mathcal{M}})$  oraz  $(Y, D_Y)$ .

## Odwzorowanie

$T : \mathcal{M} \rightarrow Y$  nazywamy zanurzeniem izometrycznym przestrzeni metrycznej  $(\mathcal{M}, D_{\mathcal{M}})$  w przestrzeni metrycznej  $(Y, D_Y)$ , jeśli

$$D_Y(T(x), T(y)) = D_{\mathcal{M}}(x, y), \quad x, y \in \mathcal{M}.$$

Jeśli ponadto  $T(\mathcal{M}) = Y$ , a nie tylko  $T(\mathcal{M}) \subseteq Y$ , to przestrzenie  $(\mathcal{M}, D_{\mathcal{M}})$  i  $(Y, D_Y)$  są przestrzeniami izometrycznymi.



# Zanurzenie z dystorsją

Niech  $F : \mathcal{M} \rightarrow Y$  będzie odwzorowaniem różnowartościowym z przestrzeni metrycznej  $(\mathcal{M}, D_{\mathcal{M}})$  do przestrzeni metrycznej  $(Y, D_Y)$ .

## Odwzorowanie

$F$  jest nazywane zanurzeniem z dystorsją  $1 + \epsilon_D$ , jeśli istnieje takie  $r > 0$ , że

$$rD_{\mathcal{M}}(x, y) \leq D_Y(F(x), F(y)) \leq r(1 + \epsilon)D_{\mathcal{M}}(x, y),$$
$$x, y \in \mathcal{M},$$

gdzie  $\epsilon_D$  jest kresem dolnym wszystkich wartości  $\epsilon$ ,

dla których zachodzą powyższe nierówności.

# Losowe projekcje - historia

Twórcą idei losowych projekcji jest Vitali Milman. Idee te stanowiły podstawę nowego, zrandomizowanego dowodu twierdzenia Dvoretzkiego [Milman 1971].

Twierdzenie to opisuje własności symetrycznych ciał, zwartych zbiorów (o niepustym wnętrzu) w  $d$ -wymiarowej unormowanej przestrzeni.

Zgodnie z tw. Dvoretzkiego przecięcie takiego zbioru z  $k$ -wymiarową liniową podprzestrzenią ma kształt bliski elipsoidzie, a zniekształcenie elipsoidy maleje asymptotycznie do zera wraz ze wzrostem wymiaru przestrzeni.

Milman w swoim dowodzie użył zarówno losowych projekcji jak i własności koncentracji miary na sferze, pokazując, że losowa podprzestrzeń z normą euklidesową jest w przybliżeniu izometryczna z oryginalną przestrzenią.

# Lemat Johnsona–Lindenstraussa (zanurzenie skończonej przestrzeni z metryką euklidesową :

## Lemat Johnsona–Lindenstraussa 1984

Skończony zbiór punktów  $\mathcal{C} \subset R^d$  może zostać zanurzony w  $k$ -wymiarowej przestrzeni, gdzie  $k \sim O(\ln |\mathcal{C}|)$  i  $k$  nie zależy od  $d$ . Odległości między punktami zbioru  $\mathcal{C}$  zostaną przy tym zachowane z dokładnością do współczynnika  $(1 \pm \varepsilon)$ , co znaczy, że istnieje odwzorowanie  $F : \mathcal{R}^d \rightarrow \mathcal{R}^k$  które każdą parę punktów  $x, y \in \mathcal{C}$  przekształca z zachowaniem warunku

$$(1 - \varepsilon)\|x - y\|^2 \leq \|F(x) - F(y)\|^2 \leq (1 + \varepsilon)\|x - y\|^2,$$

gdzie  $\|\cdot\|$  oznacza normę euklidesową w odpowiedniej przestrzeni.

# Lemat Johnsona–Lindenstraussa

lub równoważnie

$$(1 - \varepsilon)\|x - y\| \leq \|F(x) - F(y)\| \leq (1 + \varepsilon)\|x - y\|$$

Czy mamy do tu czynienia z redukcją wymiaru (jaka jest relacja  $k$  i  $d$ )? Wszystko zależy od tego jak duże jest  $N$  w stosunku do wymiaru przestrzeni  $d$  oraz jak mała wartość dystorsji jest wymagana.

**Losowe rzutowania są podstawą wielu różnych wariantów konstruktywnych dowodów lematu JL:  
Frankl, Maehara 1987, 1990,  
Dasgupta, Gupta 1999,  
Indyk, Motwani 1998,  
Vempala 2004,  
które równocześnie wskazują na sposób konstrukcji zanurzenia.**

# Losowe gausowskie projekcje:

Projekcja:  $x \mapsto v = Sx, \mathcal{R}^d \rightarrow \mathcal{R}^k, k \ll d$  jest definiowana przez macierz  $S$ :

$$\begin{bmatrix} s_{11} & s_{12} & \dots & s_{1d} \\ s_{21} & s_{22} & \dots & s_{2d} \\ \vdots & \vdots & \dots & \vdots \\ s_{k1} & s_{k2} & \dots & s_{kd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix},$$

której elementy  $s_{ij}$  zostały wygenerowane niezależnie od siebie (iid) ze rozkładu  $\mathcal{N}(0, 1)$ .

(Rozpatrywane są także inne rozkłady o tzw. subgaussowskich ogonach [ Matousek 2008])

## Notacja

**S** - losowa macierz, **S** - jej realizacja

Po rzutowaniu wektora  $x$  w kierunku  $s_j$  określonym przez  $j$ -ty wiersz macierzy rzutowań **S** otrzymujemy nowy wektor w przestrzeni  $\mathcal{R}^k$  o współrzędnych

$$v_j = \sum_{i=1}^d s_{ji} x_i = \langle s_j, x \rangle = s_j^T x.$$

Każda z tych współrzędnych zawiera, w sensie statystycznym, informacje o długości całego wektora  $x \in \mathcal{R}^d$ .

Dla ustalonego wektora  $x$  jego projekcja  $V = \mathbf{S}x$  ma wielowymiarowy rozkład normalny  $\mathcal{N}_k(0, \|x\|^2 I)$ .  
Składowe tego rozkładu nie są wzajemnie skorelowane.

$$V_j^2 \sim \|x\|^2 \chi^2$$

$$\|V\|^2 \sim \|x\|^2 \chi_k^2$$

Wartość oczekiwana kwadratu długości pojedynczej współrzędnej ze względu na losową macierz projekcji **S**

$$E_S V_j^2 = \|x\|^2$$

dla każdego  $j = 1, \dots, k$ .

$$E_S \|V\|^2 = k \|x\|^2$$

Ze względu na liniowość transformacji zamiast wektora  $x$  można wstawić różnicę dwu wektorów i otrzymać analogiczne własności dla odległości a nie tylko dla norm.



# Vempala, Gupta i Dasgupta

Dla gaussowskiej macierzy projekcji (używając nierówności Chernoffa dla rozkładu  $\chi_k^2$ ) można pokazać, że

$$\mathcal{P}\left\{\left|\frac{\|V\|^2}{\|x\|^2} - 1\right| > \varepsilon\right\} \leq 2 \exp\left(-\frac{k}{4}\varepsilon^2 + \frac{k}{6}\varepsilon^3\right)$$

Dla

$$k \geq \frac{4 \ln N}{\varepsilon^2/2 - \varepsilon^3/3}$$

prawdopodobieństwo sukcesu, czyli zdarzenia, że  $\frac{1}{\sqrt{k}}S$  zachowuje odległości pomiędzy  $N$  ustalonymi punktami z dystorsją  $1 + \varepsilon$ , jest nie mniejsze niż  $1 - \frac{2}{N^2}$ .

Powtarzając w razie potrzeby losowanie macierzy projekcji  $S$  możemy wygenerować poszukiwane odwzorowanie w skończonym czasie (z prawdopodobieństwem 1).

# Ortogonalne losowe gausowskie projekcje:

Gausowskie ortogonalne projekcje:  $S_O = (SS^T)^{-1/2}S$   
uzyskane z użyciem

- ortogonalizacja Gram-Schmidt,
- metoda Householder,
- rotacje Givens,
- metoda reorthogonalizacji Lanczosa .

wymagają współczynnika  $\sqrt{\left(\frac{d}{k}\right)}$ .

Gupta Dasupta pokazali podobne oszacowania dla  $k$   
przy założeniu, że wektory projekcji są ortogonalne.

Indyk and Motwani wskazali, że warunki  
ortonormalności są w przybliżeniu spełnione przez  
wektory  $\frac{1}{\sqrt{d}}S$ , gdy  $d$  jest dostatecznie duże. W mojej  
ocenie dla  $d \sim 10000$  nie ma sensu ortogonalizować  
wektorów projekcji.

Wiadomo, że jeśli  $\theta$  jest kątem pomiędzy  $x \in R^d$  i losową  $k$ -wymiarową podprzestrzenią  $R^d$ , wtedy zmienna losowa  $\cos(\theta)^2$  ma rozkład beta  $\beta(k/2, (d-k)/2)$ .

Podobnie, zmienna losowa  $\sin(\theta)^2 = 1 - \cos(\theta)^2$  ma rozkład  $\beta(d-k)/2, k/2)$ .

$$\frac{\|S_O x\|^2}{\|x\|^2} \sim \beta(k/2, (d-k)/2)$$

o wartości oczekiwanej  $\frac{k}{d}$ , i stąd konieczność skalowania macierzy projekcji przez  $\sqrt{\frac{d}{k}}$ .

Frankl i Maehara (1990) podali następujące oszacowania na to jak  $X \sim \beta(p, q)$  odchyła się od swojej wartości średniej  $\mu = \frac{p}{p+q}$ .

Frankl Maehara 1990

Niech  $X$  będzie zmienną losową o rozkładzie  $\beta(p, q)$ . Dla  $\varepsilon \in (0, 1)$  oraz  $p\varepsilon^2 > 6$  zachodzi

$$\mathcal{P}\{X < (1 - \varepsilon)\mu\} \leq \frac{2}{\varepsilon} \sqrt{\frac{p+q}{2\pi pq}} \exp\left[-\frac{p\varepsilon^2}{2}\right]$$

i

$$\mathcal{P}\{X > (1 + \varepsilon)\mu\} \leq \frac{2}{\varepsilon - 2\varepsilon^2/3} \sqrt{\frac{p+q}{2\pi pq}} \exp\left[-p\left(\frac{\varepsilon^2}{2} - \frac{\varepsilon^3}{3}\right)\right].$$

## Losowe projekcje - redukcja rozmiaru sieci neuronowej

Niech

$$y(x) = v_0 + \sum_{i=1}^M v_i f\left(\sum_{j=1}^d w_{i(j)} x_j\right), \quad x \in D \subset \mathcal{R}^d, \quad (1)$$

definiuje sieć neuronową z jedną warstwą ukrytą i z sigmoidalnymi funkcjami aktywacji  $f : \mathcal{R} \rightarrow [-1, 1]$ , które są niemalejącymi, funkcjami Lipschitza takimi, że

$$f(t) \rightarrow 1, \text{ gdy } t \rightarrow \infty$$

oraz

$$f(t) \rightarrow -1, \text{ gdy } t \rightarrow -\infty$$

Ponadto założymy, że  $D$  jest zbiorem zwartym w  $\mathcal{R}^d$ .

Następnie zdefiniujemy nową sieć, która zawiera na wejściu dodatkową warstwę redukującą wymiar.

$$y_S(x) = v_0 + \sum_{i=1}^M v_i f\left(\frac{1}{k} \sum_{j=1}^k \hat{w}_{i(j)} \hat{x}_j\right), \quad x \in D,$$

gdzie

$$\hat{x}_j = \sum_{l=1}^d s_{jl} x_l, \quad j = 1, \dots, k$$

oraz

$$\hat{w}_{i(j)} = \sum_{l=1}^d s_{jl} w_{i(l)}, \quad j = 1, \dots, k \quad i = 1, \dots, M$$

są projekcjami wektorów  $x \in \mathcal{R}^d$  i  $w_i \in \mathcal{R}^d$  na podprzestrzeń  $\mathcal{R}^k$ .

Sieć ta zawiera mniej wag w warstwie wejściowej w stosunku do sieci początkowej. Pomijamy tu warstwę projekcji, która nie podlega uczeniu.

Dla każdej sieci sigmoidalnej, która przetwarza dane o wysokim wymiarze (ma bardzo wiele wejść) istnieje sieć o mniejszym wymiarze, która realizuje w przybliżeniu to samo odwzorowanie, co sieć początkowa.

### Twierdzenie 2009

Niech  $R = C_L \max_i |v_i| \|w_i\| \max_{x \in X} \|x\|$ , gdzie  $C_L > 0$  jest współczynnikiem Lipschitza funkcji aktywacji  $f$ . Niech  $x_{(i)} \in D$ ,  $i = 1, \dots, n$  będzie ustalonym zbiorem  $n$  punktów. Dla  $\varepsilon \in (0, 1)$ ,  $\delta > 0$  oraz  $k$  wybranego w taki sposób, że

$$k \geq \frac{\ln 2 + \ln M + \ln n - \ln \delta}{c_1 \varepsilon^2 - c_2 \varepsilon^3}, \quad (2)$$

gdzie  $c_1 = \frac{1}{4R^2}$  oraz  $c_2 = \frac{1}{6R^3}$  prawdopodobieństwo

$$\Pr\{|y(x_{(i)}) - y_S(x_{(i)})| > \varepsilon, \quad i = 1, \dots, n\} \leq \delta.$$

Dowód opiera się na własności zachowywania iloczynu skalarnego przez losowe projekcje. Dawno zaobserwowano [Kaski i inni], iż także iloczyn skalarny jest w przybliżeniu niezmienniczy ze względu na losowe projekcje, były to jednak obserwacje oparte głównie na wynikach eksperymentów obliczeniowych lub oszacowań momentów.

### Lemat ESR 2009

Dla dowolnego  $\epsilon \in (0, 1)$  oraz dla dowolnej ustalonej pary wektorów  $x$  i  $y \in \mathcal{R}^d$  oraz ich losowych normalnych projekcji zachodzi

$$\begin{aligned} &Pr \left\{ |\hat{I}_{norm}(x, y) - r| > \epsilon \right\} \\ &\leq 2 \exp \left( -k \left( \frac{\epsilon^2}{4} - \frac{\epsilon^3}{6} \right) \right), \end{aligned}$$

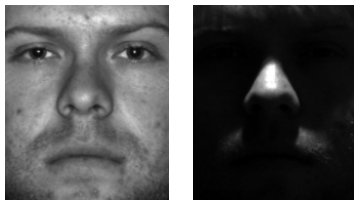
gdzie  $r = \frac{\langle x, y \rangle}{\|x\| \|y\|}$ .



## Rozpoznawanie twarzy zachowujące prywatność

The proposed method have been evaluated on the data taken from the Extended Yale Face Database B.

The sample images of the the same subject and different illumination conditions are presented in Fig. 1.



**Rysunek:** Sample images (cropped) of the Extended Yale B database (two different images of the same subject).

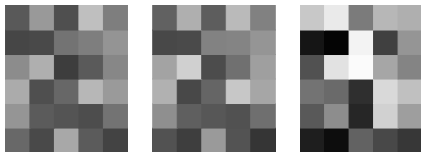
In our approach every image was into  $5 \times 6$  blocks of  $32 \times 32$  pixels. Each image block was projected into 10 dimensions.


All images were performed using the same transformations.



**Rysunek:** 10 projections of  $5 \times 6$  blocks (the left image from Fig. 1) .

Fig. 2 shows a subsequent 10 projections of  $5 \times 6$  blocks of the first image of the subject shown on the previous figure. Each projection is rearranged into the matrix form. Norms of transformed blocks of three subsequent similar images. Norm values are treated as pixel values and rearranged into the matrix form. According to the Johnson-Lindenstrauss we can claim that these values roughly approximate the mean pixel values of the corresponding image blocks.



**Rysunek:** Norms images of transformed blocks of 3 subsequent images of 

As a standard method of classification the nearest neighbor (1-NN) classifier was employed.

The 1— PCA dimensionality reduction method was used for comparison purposes.

**Tabela:** Classification accuracy for 3 classes

Number of LE	<b>3X1</b>	<b>3X4</b>	<b>3X14</b>	<b>3X32</b>
Method				
1-PCA $k = 10$	-	0.694	0.667	0.7396
1-PCA $k = 30$				0.708
1-PCA $k = 90$				0.792
RDRP-NN <b>5X6, <math>k = 10</math></b>	0.704	0.756	0.847	0.792
RDRP-LR - <b>5X6, <math>k = 10</math></b>	-	0.733	0.84	0.854
RDRP-NN <b>10X12, <math>k = 10</math></b>	0.746	0.841	0.887	0.854

**Tabela:** Classification accuracy for RDRP-NN ( $M = 5 \times 6$ ) and different number of projections  $k$  (classification of 3 items based on  $3 \times 32$  prototype images)

$k =$	1	2	3	4	6	8
Method RDRP-NN $5 \times 6$	0.690	0.742	0.746	0.823	0.844	0.792

The results of a classification accuracy for projection dimension  $k = 1 - 6$  given in Table 2 are averaged over 10 experiments.

**Tabela:** Classification accuracy of RDRP-NN for 5 classes

Number of LE	$3 \times 1$	$3 \times 4$	$3 \times 14$	$3 \times 32$
RDRP-NN $5 \times 6, k = 10$	0.689	0.743	0.803	0.806

**Tabela:** Comparison of classification accuracy for 5 classes (based on  $5 \times 32$  learning examples)

RDRP-NN <b><math>5 \times 6</math> <math>k = 10</math></b>	RDRP-LR <b><math>5 \times 6</math> <math>k = 10</math></b>	<b>30</b> – PCA NN <b><math>k = 10</math></b>
0.806	0.925	0.694
<b>1D-PCA <math>k = 10</math></b>	<b>1D-PCA <math>k = 30</math></b>	<b>1D-PCA <math>k = 50</math></b>
0.606	0.656	0.694

Rozkład **SX** losowej projekcji rozkładu prawdopodobieństwa,

**X** – wielowymiarowy rozkład normalny  $\mathcal{N}_d(M, \Sigma)$

**SX** jest  $k$  wymiarowym wektorem losowym, którego współrzędne są zmiennymi losowymi o takim samym rozkładzie.

Bez straty ogólności możemy przyjąć, że macierz kowariancji  $\Sigma$  jest macierzą diagonalną postaci  $\sigma_1^2, \dots, \sigma_d^2$ . Jeżeli macierz  $\Sigma$  jest macierzą nieosobliwą, to wszystkie wartości  $\sigma_i > 0$ , a wszystkie zmienne losowe  $x_j$ , współrzędne wektora losowego  $X$ , są wzajemnie niezależne.

Craig 1936 badał rozkład prawdopodobieństwa zmiennej losowej  $xy$ , dla której zarówno  $x$  jak i  $y$  mają rozkład normalny.

Założmy, że  $x \sim \mathcal{N}(m_x, \sigma_x)$  oraz  $y \sim \mathcal{N}(m_y, \sigma_y)$ .

Funkcja generująca momenty takiej zmiennej losowej jest równa

$$\mathcal{M}_{xy}(t) = \frac{\exp \left[ \frac{(\sigma_x^2 m_y^2 + \sigma_y^2 m_x^2) t^2 + 2 m_x m_y t}{2(1 - \sigma_x^2 \sigma_y^2 t^2)} \right]}{(1 - \sigma_x^2 \sigma_y^2 t^2)^{1/2}}. \quad (3)$$

Macierz  $\Sigma_S = \frac{1}{k}S\Sigma S^T$  jest losową macierzą Wignera [Tropp], nie jest to jednak macierz o rozkładzie typu Wisharta. Jedynym wyjątkiem jest przypadek, gdy macierz kowariancji  $\Sigma$  jest macierzą diagonalną  $I$ .

Macierz  $\Sigma_S$  ma bardzo regularną strukturę.

Wszystkie jej diagonalne elementy  $\frac{1}{k}s_i^T \Sigma s_i$ , mają ten sam rozkład ze średnią równą  $tr[\Sigma]/k$ .

Wszystkie jej elementy znajdujące się poza diagonalą  $\frac{1}{k}s_i^T \Sigma s_j$ ,  $i \neq j$ , mają ten sam rozkład prawdopodobieństwa z wartością średnią równą zero.



Elementy na diagonalu macierzy mają rozkład będący sumą  $d$  niezależnych rozkładów  $\chi_1^2$  z wagami

$$\sim \frac{1}{k} \sum_{r=1}^d \sigma_r^2 \chi_1^2,$$

z funkcją generującą momenty

$$\mathcal{M}_D(t) = \prod_{r=1}^d (1 - 2\sigma_r^2 t/k)^{-1/2}.$$

Elementy poza diagonalą składają się z sumy ważonych iloczynów niezależnych standardowych rozkładów normalnych z funkcją generującą momenty postaci:

$$\mathcal{M}_P(t) = \prod_{r=1}^d \frac{1}{(1 - \sigma_r^4 t^2/k^2)^{1/2}}.$$

## Karta kontrolna $T^2$ , gdy wymiar problemu jest większy niż dane historyczne

Przyjmiemy tutaj, że obserwowany proces generuje wielowymiarowy szereg czasowy złożony z niezależnych realizacji wektorów losowych o tym samym rozkładzie (normalnym z nieznaną macierzą kowariancji).  
Macierz kowariancji rozkładu nie jest znana i nie może być dokładnie estymowana ze względu na bardzo duży wymiar problemu

We assume that  $s_{ij} \sim N(0, 1)$ ,  $i = 1, \dots, k$  and  $j = 1, \dots, d$  and are i.i.d.

**Remark:**

Once generated  $S$  is kept constant (the same for all data).

# Phase I (preparations):

In the first phase we project the collected data  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$  into  $k$ - dimensional ( $N < k \ll d$ ) space using random linear projections with projection matrix  $S$ .

Since,  $\mathbf{X}_i$  are i.i.d. random normal vectors  $\mathbf{X}_i \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$  we obtain:

$$\mathbf{V}_i = S \mathbf{X}_i, \quad \mathbf{V}_i \sim \mathcal{N}_k(\mathbf{0}, S \Sigma S^T), \quad i = 1, \dots, N.$$

For the sake of simplicity, we have assumed that  $E(\mathbf{X}_i) = E(\mathbf{V}_i) = \mathbf{0}$ .

If the sample size  $N \ll d$ , the covariance matrix  $\Sigma$  is not known and its sample based estimate  $C = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$  is a singular matrix (with probability one). Thus, it is not possible to use the Hotelling's  $T^2$  in  $d$  dimensions [?].

If  $k < N$  compute the new covariance matrix (after projection) estimate:

$$C_S = \frac{1}{N} \sum_{i=1}^N V_i V_i^T.$$

Estimate in  $k$ -dimensions

$C_S$  is  $k \times k$ , full rank (with probability 1 with respect to  $S$ ), even if  $\Sigma$  is a singular matrix.

Note, that  $C_S = SCS^T$ ,  $C = \frac{1}{N} \sum_{i=1}^N X_i X_i^T$ .

$N C_S$  is distributed as a Wishart distribution of dimensionality  $k$  with  $N$  degrees of freedom and associated parameter matrix  $S \Sigma S^T$  (not known):

$$N C_S \sim W_k(N, S \Sigma S^T)$$

and

$$N C \sim W_d(N, \Sigma)$$

has a pseudo-Wishart distribution ( $N < d$ ) (see [?], [?]).

## Phase II (detection):

In phase II, new independent observation  $X \sim \mathcal{N}_d(0, \Sigma)$  is projected using the same projection matrix  $S$ ,  $V = SX$ . Thus:

$$V \sim \mathcal{N}_k(0, S\Sigma S^T)$$

$$T_S^2 = (SX)^T C_S^{-1} SX \sim \frac{kN}{N - k + 1} F(k, N - k + 1)$$

$T_S^2 < h_\alpha$  with probability  $1 - \alpha$  (in-control state).



## Comparisons to $T^2$ statistic in $R^d$

$$T^2 = \mathbf{X}^T \mathbf{C}^{-1} \mathbf{X} \sim \frac{dN}{N-d+1} F(d, N-d+1),$$

$N \geq d.$

## Change in the mean :

If  $X \sim \mathcal{N}_d(m, \Sigma)$ ,  $V \sim \mathcal{N}_k(Sm, S\Sigma S^T)$  and

$$T_S^2 \sim \frac{kN}{N - k + 1} F(k, N - k + 1, (Sm)^T (S\Sigma S^T)^{-1} Sm)$$

i.e. noncentral F Snedecor distribution.

Power of the test ?

We do not know  $S\Sigma S^T$ , only  $C_S$  is given and the answer applies to  $Sm$  only.

## Diagnostic properties :

If  $X \sim \mathcal{N}_d(m, \Sigma)$ , then

$$E_x\{V^T C_S^{-1} V | S, X_1, \dots, X_N\} =$$

$$E_x\{X^T S^T C_S^{-1} S X | S, X_1, \dots, X_N\} =$$

Changes in  $\Sigma$

$$\text{Trace}[S^T C_S^{-1} S \Sigma] +$$

Changes in the mean

$$m^T (S^T C_S^{-1} S) m.$$

Note, that if we can replace  $C_S$  by  $S\Sigma S^T$ , then

$$\text{Trace}[S^T (S\Sigma S^T)^{-1} S \Sigma] = k.$$

$$\begin{aligned} \text{Var}_x\{V^T AV | S, X_1, \dots, X_N\} = \\ 2\text{Trace}[S^T AS\Sigma S^T AS\Sigma] + \\ 4m^T S^T AS\Sigma S^T ASm \end{aligned}$$

where  $A$  is symmetric, positive definite  $k \times k$  matrix ( $A = C_S^{-1}$ ).

## Properties of the covariance matrix:

We shall discuss properties of  $S\Sigma S^T$  and  $m^T S^T C_S^{-1} S m$ , since it has the main influence on the probability of detection a change of the mean vector.

- If  $k < d$ ,  $S^T (S\Sigma S^T)^{-1} S \neq \Sigma^{-1}$  since  $S^T A S$  is rank deficient.
- $E_S [S\Sigma S^T] = \text{trace}[\Sigma] I_k$  (eccentricity reduction)  
Note that  $\frac{1}{k} \text{trace}[C_S]$  estimates the trace of  $\Sigma$ .
- $E_S [m^T S^T S m] = k m^T m = k \|m\|^2$

## How large $k$ should be ?

- $k < N$
- $Pr \left\{ \frac{1}{k} \|Sm\|_k^2 < (1 - \epsilon) \|m\|_d^2 \right\} \leq \exp \left( -\frac{k}{4} \epsilon^2 + \frac{k}{6} \epsilon^3 \right)$   
(using chi-squared tail Chernoff bounds)

- **Different approach to projections of multivariate data is given in the papers of Runger 1996 [?], Bodnar, Schmid 2005 and others.**
- **Using of random projections can be considered as an implication of the Cramer–Wold theorem (J.A. Cuesta–Albertos, E. del Barrio, R. Fraiman, C. Matran: The random projection method in goodness of fit for functional data, Computational Statistics and Data Analysis 51 (2007)).**

# The main idea:

Here we propose to perform the procedure of random projection of data (historical data set and a new observation) **instantaneously**, just after a new observation appears.

This means that Phase I and Phase II of the control chart design (Mason and Young 2002) is repeated consecutively.

It is important to stress that this approach is different than the method proposed previously by the author (ESR 2011, 2013), where the projection matrix was generated only once as a starting point of the dimensionality reduction procedure.

The new approach, though more time-consuming, is also more efficient, because there is no direction in the observation space where changes cannot be detected. We propose instantaneous normal random projections.



# The problem statement(1):

We start by defining a historical dataset (HDS) consisting of  $N$  independent vectors of dimension  $d$  observed over time, where  $d$  is the dimension of these vectors.

We consider the case where the in-control observation vectors,  $X_i, i = 1, \dots, N$ , are iid random vectors with common mean vector  $\mu$  and covariance matrix  $\Sigma$ , having Gaussian distribution  $X_i \sim N_d(\mu_0, \Sigma)$ . Both,  $\mu_0$  and  $\Sigma$  are not known and should be estimated on the bases of the HDS.

## The problem statement(2):

If the distribution mean changes from  $\mu_0$  to  $\mu_1$  we should be able to detect this event with possibly the highest probability. We assume that the sample size  $N < d$ , so the sample covariance matrix, based on iid HDS,

$$S = \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T / (N - 1), \quad (4)$$

is singular, so  $S^{-1}$  does not exist and  $|S| = 0$ .

## The control chart :

We propose to use the Hotelling control chart based on individual observations and to apply it to multidimensional Gaussian data projected onto random  $k$  dimensional subspace of  $\mathcal{R}^d$ , where  $k \ll d$ .

Let  $R \in \mathcal{R}^{k \times d}$  be a matrix of projection with iid normal entries, i.e.,  $r_{ij} \sim \mathcal{N}(0, 1)$ . For a new observation matrix  $R$  is generated and applied to it as well as to HDS.

Conditioning on  $R$  the projected HDS are iid normal:

$$V_i = R X_i, \quad V_i | R \sim \mathcal{N}_k(R\mu_0, R\Sigma R^T), \quad i = 1, \dots, N. \quad (5)$$

## The control chart :

If  $k < N$ , the projected sample covariance matrix

$$S_R = \frac{1}{N-1} \sum_{i=1}^N (V_i - \bar{V})(V_i - \bar{V})^T, \quad (6)$$

where  $\bar{V} = \sum_{i=1}^N V_i / N$ , is non-singular with probability 1.

Let  $X$  denote a current observation in the monitoring sequence. For change-point detection of the mean vector we shall use the following statistic:

$$T_R^2 = (RX - \bar{V})^T S_R^{-1} (RX - \bar{V}) | R \sim \frac{k(N+1)(N-1)}{(N-k)N} F(k, N-k), \quad (7)$$

where  $F$  is the F-Snedecor distribution with  $k$  and  $N-k$  degrees of freedom. Notice that  $T_R^2$  depends on  $k$  and  $N$ ; it does not depend on  $d$ .

## The control chart :

When change in the mean occurred and

$$\mu_1 = \mu_0 + m, \quad m \neq 0.$$

$$V = RX | R \sim \mathcal{N}_k(R\mu_1, R\Sigma R^T).$$

$$T_R^2 \sim \frac{k(N+1)(N-1)}{(N-k)N} F(k, N-k, \lambda^2), \quad (8)$$

where  $F(k, N-k, \lambda^2)$  is a noncentral F–Snedecor distribution with non-centrality parameter  $\lambda^2 = (Rm)^T (R\Sigma R^T)^{-1} Rm$  and with the degrees of freedom  $k$  and  $N-k$ , respectively.

## The control chart :

The general idea is to test (based on individual observations):  $H_0 : \mu = \mu_0$  versus  $H_1 : \mu = \mu_1 \neq \mu_0$   
If  $T_R^2 < h_\alpha(k) = \frac{k(N+1)(N-1)}{(N-k)N} F_\alpha^{-1}(k, N-k)$ , where  $\alpha$  is a significance level, we decide that the process is in-control (no change in the mean is supposed) In other words,  $h_\alpha(k)$  is the upper control limit .

We propose here to use the following algorithm:  
For given HDS  $X_1, \dots, X_N$  and  $\alpha$  choose  $k$  and compute  $h_\alpha(k)$ . For a given new observation  $X$

- 1 Generate  $R$ . Compute  $RX$   $S_R$  and  $T_R^2$ .
- 2 If  $T_R^2 > h_\alpha(k)$  decide that the process is not in control. Otherwise, wait for the next observation.

## Comment:

Notice that for every new observation we should generate randomly a new projection matrix. This is very important, because there are no directions in  $\mathcal{R}^d$  where changes cannot be detected.

More precisely, for any orthogonal matrix  $Q \in O(d)$  it can be shown that  $RQ$  is also a normal random projection matrix.

Thus, although the test statistics  $T_R^2$  value depends on  $R$  and is not rotation invariant, its expected value with respect to  $R$  is invariant under the rotation group  $O(d)$ .

## Model of the changes in the mean vector :

Let

$$\delta = m^T \Sigma m, \quad m \in \mathcal{R}^d, \quad (9)$$

where  $m$  is a change in the mean vector.

Define a vector random variable  $Z \sim \mathcal{N}_d(0, I)$  and let  $Y = \Sigma^{1/2} Z$ . Recall that we have assumed that  $\Sigma$  is a nonsingular positive definite symmetric matrix.

$E\{Z^T Z\} = d$  since  $Z^T Z \sim \chi_d^2$ .

Similarly,  $Y^T \Sigma^{-1} Y \sim \chi_d^2$  and  $E\{Y^T \Sigma^{-1} Y\} = d$ .

Random vector  $M = \sqrt{\frac{\delta}{d}} Y$  is a random model of the changes in the mean vector with Mahalanobis distance which is close to  $\delta$ .



## Model of the changes in the mean vector :

It is easy to show that

$$(RM)^T(R\Sigma R^T)^{-1}MY = \frac{\delta}{d}Z^T R^T (RR^T)^{-1}RZ. \quad (10)$$

Although  $RR^T$  follows the Wishart distribution  $W_k(d, I)$  and  $U = RZ|Z = z \sim N_k(0, \|x\|^2 I)$ , the Gaussian random vector  $U = RZ|Z = z$  and the random matrix  $RR^T$  are dependent.

Let  $u$  be an eigenvector of  $R^T(RR^T)^{-1}R$ , and let  $\nu$  denote the eigenvalue corresponding to that eigenvector. Rank of the matrix  $R$  is  $k$  (with probability 1). Furthermore,

$$R^T(RR^T)^{-1}Ru = \nu u, \quad RR^T(RR^T)^{-1}Ru = \nu Ru, \quad Ru = \nu u \quad (11)$$

Thus, eigenvalues of matrix  $R^T(RR^T)^{-1}R$  are 1 or 0, with only  $k$  eigenvalues equal to 1 (with probability 1).

## Model of the changes in the mean vector :

Further,

$$R^T (RR^T)^{-1} R = Q^T A Q, \quad (12)$$

where  $Q \in O(d)$  is an appropriately chosen orthogonal matrix, and

$$A = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}. \quad (13)$$

It can be shown that the diagonal elements of  $R^T (RR^T)^{-1} R$  follow beta distribution  $\beta(k/2, (d - k)/2)$  and expected values of out-of-diagonal elements of this matrix are zero.

## Model of the changes in the mean vector :

Thus,

$$E_R[R^T (RR^T)^{-1} R] = E_Q\{Q^T A Q\} = \frac{k}{d} I$$

and

$$E_Z E_R \left\{ \frac{\delta}{d} Z^T R^T (RR^T)^{-1} R Z \mid Z = z \right\} \quad (14)$$
$$= \frac{\delta k}{d^2} E_Z(\|Z\|^2) = \frac{\delta k}{d}.$$

Furthermore, we can easily obtain that the random variable defined in (10) follows  $\frac{\delta}{d} \chi_k^2$  distribution.

## Comment:

Notice that for any  $Q \in O(d)$  vector  $AQZ$  is a  $k$ -dimensional orthogonal projection of normal random vector  $Z \sim \mathcal{N}_d(0, I)$  and using normal random projections we obtain a  $k$ -dimensional orthogonal vector basis according to a unique rotation-invariant probability measure (Haar measure) on compact Stiefel manifold (see for example (E. Meckes 2012)).

## Power of the test :

Power of the test  $T_R^2$  for individual observation  $X$  and given mean shift vector  $m$ , denoted as  $P_l(m)$  depends on dimension  $k$  of the normal random projection, noncentrality parameter

$\delta(k, R, m) = m^T R^T (R \Sigma R^T)^{-1} R m$  value obtained after projection and Type I error  $\alpha$  (recall that in our case  $ARL_0 = 1/\alpha$ ). Thus, conditionally on  $R$ , we obtain the following formula

$$P_l(m) = \int_{h(k)}^{\infty} f(k, N - k, \delta(k, R, m), x) dx, \quad (15)$$

where  $f(k, N - k, \delta(k, R, m), x)$  is the probability density of noncentral distribution

$F(k, N - k, \delta(k, R, m))$  and  $h(k)$  is the critical value of the test for a given  $\alpha$ . The new noncentrality parameter  $\delta(k, R, m)$  depends strongly on  $\Sigma$  which is unknown and in fact non-estimable due to small  $N$ .

## Power of the test :

Thus, assuming model (10) we shall concentrate on non-local power defined as:

$$P_{nl}(\delta, \mathbf{k}, N) = E_{M,R} \left\{ \int_{h(\mathbf{k})}^{\infty} f(\mathbf{k}, N - \mathbf{k}, \delta(\mathbf{k}, R, M), \mathbf{x}) d\mathbf{x} \right\} \quad (16)$$

$$= E_{Z,R} \left\{ \int_{h(\mathbf{k})}^{\infty} f(\mathbf{k}, N - \mathbf{k}, \frac{\delta}{d} \mathbf{Z}^T R^T (R R^T)^{-1} R \mathbf{Z}, \mathbf{x}) d\mathbf{x} \right\} \quad (17)$$

$$= \int_0^{\infty} f_{\chi^2}(\mathbf{k})(\mathbf{y}) \left[ \int_{h(\mathbf{k})}^{\infty} f(\mathbf{k}, N - \mathbf{k}, \frac{\delta}{d} \mathbf{y}, \mathbf{x}) d\mathbf{x} \right] d\mathbf{y}. \quad (18)$$

## Choosing dimension of random projection $k$ :

On the other hand the noncentrality parameter of the appropriate F- Snedecor distribution is larger in higher dimensions. Thus, the optimal choice of  $k$  is a compromise between both opposite tendencies. Under assumption of model (10) we have to solve the following optimization problem:

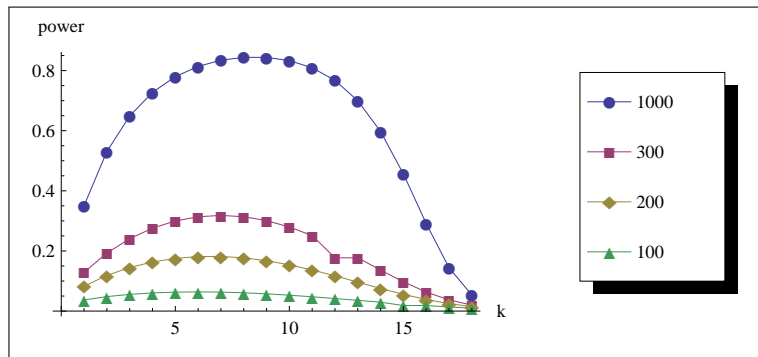
$$\text{maximise } P_{nl}(\delta, k, N) \text{ with respect to } k \leq N - 2. \quad (19)$$

## Choosing dimension of random projection $k$ :

For moderate  $N$  values the problem can be solved numerically. Figure 4 shows  $P_{nl}(\delta, k, N)$  for  $k = 1, \dots, 18$  and false alarm rate 0.005, i.e.  $\alpha = 0.005$ ,  $N = 20$ ,  $d = 100$  and  $\delta = 1000, 300, 200, 100$  respectively. It is important to indicate, that function (18) is very flat with respect to  $k$  and the proposed method of selecting  $k$  is robust to small and moderate changes of  $\alpha$  and  $\delta$ .



# Choosing dimension of random projection $k$ :



**Rysunek:** Non-local power values of the test obtained numerically according to the formula (18) for  $\alpha = 0.005$  ( $d = 100$ ,  $N = 20$ ,  $\delta = 1000, 300, 200, 100$  and  $k = 1, \dots, 18$ )

## Numerical experiments :

We perform some experiments for multivariate normal distributions with population covariance matrix with compound symmetry structure:  $\Sigma = (1 - \rho)I_d + \rho\mathbf{1}\mathbf{1}^T$ , where  $\mathbf{1}_d$  is the column vector with all entries one,  $\rho \in (0, 1)$  is a common correlation factor. The trace of  $\Sigma$  is  $d$ , independently of  $\rho$  value. Notice that the eigenvalues of the compound symmetry structure are  $\lambda_1 = 1 + (d - 1)\rho$  and  $\lambda_2 = \dots = \lambda_d = 1 - \rho$ . Without loss of generality we can assume that the true covariance matrix is diagonal since the proposed approach is rotation invariant in the mean.

## Numerical experiments :

In our study  $N = 20$ ,  $d = 100$  and  $\rho = 0.6, 0.4$  and  $0.1$  are selected. We fix the nominal significance level  $\alpha = 0.005$ . This significance level is equivalent to the average run length ARL in an in-control state, i.e., nominal  $ARL_0 = 1/\alpha = 200$ . Numerical experiments are performed for 10 different HDS. Each experiment consists of 10,000 replications. We select dimension of the projection according to (19) (see also Figure 4). In Table 5 the averages over 10 HDS of ARL to detection are shown for 6 different  $m$  and the corresponding  $\delta = m^T \Sigma^{-1} m$ .

# Numerical experiments :

**Tabela:** ARL's for  $d = 100$ ,  $N = 20$ ,  $k = 7$  and  $m_0 = (0, \dots, 0)$ ,  $m_1 = (10.99, 0, \dots, 0)$ ,  $m_2 = (5.495, 0, \dots, 0)$ ,  $m_3 = (0.0899, \dots, 0.0899)$ ,  $m_4 = (0, 0.6324, 0.6324, 0.6324, 0, \dots, 0)$ ,  $m_5 = (0, 0.6324, 0.6324, 0, \dots, 0)$ ,  $m_6 = (0, 0.6324, 0, \dots, 0)$

$m$	$\rho = 0.6$		$\rho = 0.4$		$\rho = 0.1$	
	$\delta$	<i>ARL</i>	$\delta$	<i>ARL</i>	$\delta$	<i>ARL</i>
$m_0$	0	208.3	0	217.4	0	217.4
$m_1$	200.	1.00	297.5	1.00	1108.	1.00
$m_2$	50.	1.30	74.4	1.11	277.	1.02
$m_3$	200.3	6.78	133.34	12.35	88.9	20.81
$m_4$	300.	3.97	200.	6.99	133.3	12.27
$m_5$	200	7.23	133.3	13.14	88.9	22.52
$m_6$	100.	21.87	66.7	37.11	44.4	56.95

# Conclusions:

As one can observe, ARL's to detection strongly depend on a direction of the change vector  $m$ , e.g., changes in the direction of the largest eigenvalue of  $\Sigma$  are the easiest to detect. In contrary, the most difficult to disclose is vector  $m_6$ , where the change is in the direction of the smallest eigenvalue.

The above simulations were repeated for non-optimal  $k$  and in every case ARL's to detection were larger than those reported in Table 5. However, the results are not too sensitive to small inaccuracies in selecting optimal  $k$ . Furthermore,  $k$  selected according to (19), i.e., averaged with respect to (10), performed very well in each particular case listed in Table 5.

# Conclusions:

The method performs well, when a population covariance matrix has a compound symmetry structure. Due to the known distribution of the test statistic the control level of the chart is data independent, what is important in a small sample case. The proposed method of selecting the dimension of the normal projection provides accurate values when the number of samples is moderate. If the average change of the mean vector is assumed to be proportional to the dimension, the power of the test for a given projection dimension depends only on the number of samples and a significance level.

Realizacja projekcji wymaga  $O(kd)$  operacji arytmetycznych. Mimo możliwości zrównoleglenia, gdyż projekcje w kierunkach (wierszach macierzy projekcji) można wykonywać niezależnie od siebie, koszty projekcji mogą to być zbyt duże z praktycznego punktu widzenia. Macierze rzadkie, czyli takie, których zawierają wiele zerowych elementów, pozwalają zredukować te koszty, bez istotnej straty dokładności.

Achlioptas zaproponował użycie rozkładu dyskretnego, który prowadzi do do redukcji potrzebnych nakładów obliczeniowych:

$$s_{ij} = \sqrt{3} \begin{cases} 1 & \text{z prawdopodobieństwem } 1/6 \\ 0 & \text{z prawdopodobieństwem } 2/3 \\ -1 & \text{z prawdopodobieństwem } 1/6 \end{cases} \quad (20)$$

Mimo, że proponowany dyskretny rozkład ma tylko dwa pierwsze momenty identyczne jak rozkład  $\mathcal{N}(0, 1)$ , to wykazano, że z punktu widzenia zachowywania odległości otrzymujemy takie samo oszacowanie na wymiar  $k$  jak w przypadku normalnych projekcji.



## Losowe projekcje z rzadkimi macierzami przekształcenia

Można stosować dowolny rozkład z wartością średnią 0 oraz jednostkową wariancją, który należy do klasy rozkładów z subgaussowskimi ogonami [Ailon:06], [Matousek:08].

Zmienna losowa  $X$  ma subgaussowskie ogony, jeżeli zarówno  $X$  jak i  $-X$  zachowują się w taki sposób, że dla każdego  $\varepsilon > 0$ ,

$$\Pr \{|X| > \varepsilon\} \leq \exp(-a \varepsilon^2), \quad (21)$$

gdzie  $a > 0$  jest pewną stałą.

Macierz projekcji utworzona z elementów wygenerowanych niezależnie zgodnie z rozkładem postaci:

$$s_{ij} = \sqrt{c} \begin{cases} 1 & \text{z with probability } \frac{1}{2c} \\ 0 & \text{z with probability } 1 - \frac{1}{c} \\ -1 & \text{z with probability } \frac{1}{2c} \end{cases} \quad (22)$$

należy do klasy rzadkich z subgaussowskimi ogonami

Najprostszym reprezentantem tej klasy jest macierz projekcji, w której poszczególne elementy przyjmują wartości  $+1$  lub  $-1$  z prawdopodobieństwem  $1/2$ . Bardzo rzadkie macierze wymagają dodatkowych operacji "mieszania" lub spełnienia warunku

$$\max_j x_j / \|x\| \approx \frac{1}{\sqrt{d}}$$

(szybka transformacja Ailon Chazelle)

Własność ograniczonej izometrii macierzy RIP jest pojęciem wprowadzonym przez E. Candes, T. Tao i odnosi się do własności normy przekształceń definiowanych przez pewną macierz. Macierz  $\Phi \in \mathcal{R}^{p \times d}$  spełnia warunek RIP, jeżeli dla każdego naturalnego  $1 \leq k \leq d$  istnieje taka stała  $\delta_k \in (0, 1)$ , że dla każdej macierzy  $\Phi_k$ , która powstała z macierzy  $\Phi$  poprzez zastąpienie  $d - k$  kolumn macierzy wektorami zerowymi, dla każdego  $y \in Y \subseteq \mathcal{R}^d$  zachodzi

$$(1 - \delta_k) \|y\|^2 \leq \|\Phi_k y\|^2 \leq (1 + \delta_k) \|y\|^2.$$

Własność RIP jest ważna w zastosowaniach związanych z skompresowanym próbkowaniem CS.

N. Ailon and B. Chazelle, Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform, Proc 38th ACM Symp Theory of Computing, 2006, pp. 557–563.

C. C. Craig, On the Frequency Function of  $xy$ , Annals of Statistics 7(1), 1–15, 1936.

Cuesta-Albertos JA, del Barrio E, Fraiman R and Matran C (2007) The random projection method in goodness of fit for functional data. Computational Statistics and Data Analysis 51(10): 4814–4831

Frankl P and Maehara H (1990) Some geometric applications of the beta distribution. Annals of the Institute of Statistical Mathematics 42 (3):463–474

Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, STOC '98, pp. 604–613 (1998).

Johnson WB and Lindenstrauss J (1984) Extensions of Lipschitz mapping into Hilbert space. Contemporary Mathematics 26, 189–206

Matoušek, J.(2008) On variants of the Johnson–Lindenstrauss lemma, *Random Structures and Algorithms* 33(2): 142–156.

J. Matoušek, *Lectures on discrete geometry*, Springer, New York, 2002.

Meckes E (2012) Approximation of Projections of Random Vectors. *Journal of Theoretical Probability* 25(2):333–352

Milman V (1971) A new proof of the theorem of A. Dvoretzky on sections of convex bodies. *Functional Analysis and Its Applications* 5(4): 28–37 (English translation)

Skubalska-Rafajłowicz E (2013) Random projections and Hotelling's  $T^2$  statistics for change detection in high-dimensional data streams. *International Journal of Applied Mathematics and Computer Science* 23(2):447–461

Skubalska-Rafajłowicz, E., *Neural Networks with Sigmoidal Activation Functions – Dimension Reduction Using Normal Random Projection*, *Nonlinear Analysis* 71, e1255–e1263, 2009.

Vempala S (2004) *The Random Projection Method*. American Mathematical Society, Providence

D. P. Woodruff, “Sketching as a tool for numerical linear algebra”, *Foundations and Trends in Theoretical Computer Science*, vol. 10,

**D. P. Woodruff, “Sketching as a tool for numerical linear algebra”, Foundations and Trends in Theoretical Computer Science, vol. 10, pp. 1–157, 2014.**

**<http://libra.ibuk.pl/>**

**Ewa Skubalska-Rafajłowicz Losowe projekcje Metody algorytmy i wybrane zastosowania, Exit 2013**