Johnson-Lindenstrauss Transform

14-6-2011

- $\Phi \in M_{n\times N}$ con $n \leq N$ $\qquad \Phi : \mathbb{R}^N \to \mathbb{R}^n$
- (Ω, P) $\omega \to \Phi_\omega$
- Norma euclidea $\| x \|_2 = (\sum_{k=1}^n x_k^2)^{1/2}$
  Norma 1 $\| x \|_1 = \sum_{k=1}^n | x_k |$

Definition

Johnson-Lindenstrauss Transform se esiste $c > 0$ tale che per per ogni $\varepsilon \in (0,1)$ e per ogni $x \in \mathbb{R}^N$ si ha:

$$P\left(\{(1-\varepsilon) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1+\varepsilon) \| x \|_2^2\}\right) \geq 1 - 2e^{-c\varepsilon^2 n}$$

- $\Phi \in M_{n\times N}$ con $n \leq N$ $\quad\quad \Phi : \mathbb{R}^N \to \mathbb{R}^n$
- (Ω, P) $\omega \to \Phi_\omega$
- Norma euclidea $\| x \|_2 = (\sum_{k=1}^n x_k^2)^{1/2}$
  Norma 1 $\| x \|_1 = \sum_{k=1}^n | x_k |$

Definition

Johnson-Lindenstrauss Transform se esiste $c > 0$ tale che per per ogni $\varepsilon \in (0,1)$ e per ogni $x \in \mathbb{R}^N$ si ha:

$$P\left(\{(1-\varepsilon) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1+\varepsilon) \| x \|_2^2\}\right) \geq 1 - 2e^{-c\varepsilon^2 n}$$

- $\Phi \in M_{n \times N}$ con $n \leq N$ $\qquad \Phi : \mathbb{R}^N \to \mathbb{R}^n$
- (Ω, P) $\omega \to \Phi_\omega$
- Norma euclidea $\| x \|_2 = (\sum_{k=1}^{n} x_k^2)^{1/2}$
  Norma 1 $\| x \|_1 = \sum_{k=1}^{n} | x_k |$

Definition

Johnson-Lindenstrauss Transform se esiste $c > 0$ tale che per per ogni $\varepsilon \in (0,1)$ e per ogni $x \in \mathbb{R}^N$ si ha:

$$P\left(\{(1-\varepsilon) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1+\varepsilon) \| x \|_2^2\}\right) \geq 1 - 2e^{-c\varepsilon^2 n}$$

- $\Phi \in M_{n\times N}$ con $n \leq N$ $\quad \Phi : \mathbb{R}^N \to \mathbb{R}^n$
- (Ω, P) $\omega \to \Phi_\omega$
- Norma euclidea $\| x \|_2 = (\sum_{k=1}^n x_k^2)^{1/2}$
  Norma 1 $\| x \|_1 = \sum_{k=1}^n | x_k |$

Definition

Johnson-Lindenstrauss Transform se esiste $c > 0$ tale che per per ogni $\varepsilon \in (0,1)$ e per ogni $x \in \mathbb{R}^N$ si ha:

$$P\left(\left\{(1-\varepsilon) \| x \|_2^2 \leq \| \Phi x \|_2^2 \leq (1+\varepsilon) \| x \|_2^2\right\}\right) \geq 1 - 2e^{-c\varepsilon^2 n}$$

Esempi di JLT

Examples

- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. $N(0, \frac{1}{n})$
- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. uniformi $\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}$
- $$\Phi_{ij} = \begin{cases} +\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \\ 0 & \text{con probabilità } 2/3 \\ -\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \end{cases}$$
- Variabili Subgaussiane: $\mathbf{E}\exp(tX) \leq \exp(\sigma^2 t^2/2)$ per ogni t $(\sigma^2 = var(X))$

Esempi di JLT

Examples

- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. $N(0, \frac{1}{n})$
- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. uniformi $\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}$
- $\Phi_{ij} = \begin{cases} +\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \\ 0 & \text{con probabilità } 2/3 \\ -\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \end{cases}$
- Variabili Subgaussiane: $\mathbf{E}\exp(tX) \leq \exp(\sigma^2 t^2/2)$ per ogni t $(\sigma^2 = var(X))$

Esempi di JLT

Examples

- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. $N(0, \frac{1}{n})$
- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. uniformi $\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}$
- $$\Phi_{ij} = \begin{cases} +\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \\ 0 & \text{con probabilità } 2/3 \\ -\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \end{cases}$$
- Variabili Subgaussiane: $\mathbf{E}\exp(tX) \leq \exp(\sigma^2 t^2/2)$ per ogni t $(\sigma^2 = var(X))$

Esempi di JLT

Examples

- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. $N(0, \frac{1}{n})$
- $\Phi = \{\Phi_{ij}\}$ con Φ_{ij} v.a. i.i.d. uniformi $\{\frac{-1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}$
- $$\Phi_{ij} = \begin{cases} +\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \\ 0 & \text{con probabilità } 2/3 \\ -\sqrt{\frac{3}{n}} & \text{con probabilità } 1/6 \end{cases}$$
- Variabili Subgaussiane: $\mathbf{E}\exp(tX) \leq \exp(\sigma^2 t^2/2)$ per ogni t $(\sigma^2 = var(X))$

Lemma di Johnson-Lindenstrauss
(Riduzione dimensionale)

Theorem

Esiste una costante C tale che per ogni $\varepsilon \in (0, 1/2)$ e per ogni insieme di punti $\{x_1, x_2, \cdots, x_N\}$ in $\mathbb{R}^N$ per $n = C\varepsilon^{-2} \ln N$ esiste un operatore lineare $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ tale che per ogni $i \neq j$

$$(1-\varepsilon) \| x_i - x_j \|_2^2 \leq \| \Phi x_i - \Phi x_j \|_2^2 \leq (1+\varepsilon) \| x_i - x_j \|_2^2$$

- L'operatore è ottenuto da una *JLT*.
- Si può fare in modo che vi sia un'alta probabilità di ottenere una tale riduzione dimensionale.
- Scegliendo la terza famiglia di matrici si ha una notevole riduzione computazionale.

Lemma di Johnson-Lindenstrauss

(Riduzione dimensionale)

Theorem

Esiste una costante C tale che per ogni $\varepsilon \in (0, 1/2)$ e per ogni insieme di punti $\{x_1, x_2, \cdots, x_N\}$ in $\mathbb{R}^N$ per $n = C\varepsilon^{-2} \ln N$ esiste un operatore lineare $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ tale che per ogni $i \neq j$

$$(1-\varepsilon) \| x_i - x_j \|_2^2 \leq \| \Phi x_i - \Phi x_j \|_2^2 \leq (1+\varepsilon) \| x_i - x_j \|_2^2$$

- L'operatore è ottenuto da una *JLT*.
- Si può fare in modo che vi sia un'alta probabilità di ottenere una tale riduzione dimensionale.
- Scegliendo la terza famiglia di matrici si ha una notevole riduzione computazionale.

Lemma di Johnson-Lindenstrauss
(Riduzione dimensionale)

Theorem

Esiste una costante C tale che per ogni $\varepsilon \in (0, 1/2)$ e per ogni insieme di punti $\{x_1, x_2, \cdots, x_N\}$ in $\mathbb{R}^N$ per $n = C\varepsilon^{-2} \ln N$ esiste un operatore lineare $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ tale che per ogni $i \neq j$

$$(1-\varepsilon) \| x_i - x_j \|_2^2 \leq \| \Phi x_i - \Phi x_j \|_2^2 \leq (1+\varepsilon) \| x_i - x_j \|_2^2$$

- L'operatore è ottenuto da una *JLT*.
- Si può fare in modo che vi sia un'alta probabilità di ottenere una tale riduzione dimensionale.
- Scegliendo la terza famiglia di matrici si ha una notevole riduzione computazionale.

Lemma di Johnson-Lindenstrauss
(Riduzione dimensionale)

Theorem

Esiste una costante C tale che per ogni $\varepsilon \in (0, 1/2)$ e per ogni insieme di punti $\{x_1, x_2, \cdots, x_N\}$ in $\mathbb{R}^N$ per $n = C\varepsilon^{-2} \ln N$ esiste un operatore lineare $\Phi : \mathbb{R}^N \to \mathbb{R}^n$ tale che per ogni $i \neq j$

$$(1-\varepsilon) \| x_i - x_j \|_2^2 \leq \| \Phi x_i - \Phi x_j \|_2^2 \leq (1+\varepsilon) \| x_i - x_j \|_2^2$$

- L'operatore è ottenuto da una *JLT*.
- Si può fare in modo che vi sia un'alta probabilità di ottenere una tale riduzione dimensionale.
- Scegliendo la terza famiglia di matrici si ha una notevole riduzione computazionale.

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

$$y \quad = \quad \Phi \quad x$$

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

$$y = \Phi x$$

Vettore dei dati (di grande dimensione)

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

$$y = \Phi x$$

Vettore dei dati (di grande dimensione)

Matrice di codifica (n-Osservazioni)

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

$$y = \Phi x$$

Vettore ottenuto (si spera di bassa dimensione)

Matrice di codifica (n-Osservazioni)

Vettore dei dati (di grande dimensione)

Una diversa interpretazione
($x \in \mathbb{R}^N$ e Φ matrice $n \times N$)

$$y = \Phi x$$

Vettore ottenuto (si spera di bassa dimensione)

Vettore dei dati (di grande dimensione)

Matrice di codifica (n-Osservazioni)

Un vettore è k-sparso se $\mid \{j : x_j \neq 0\} \mid \leq k$

Particolare decodificatore:

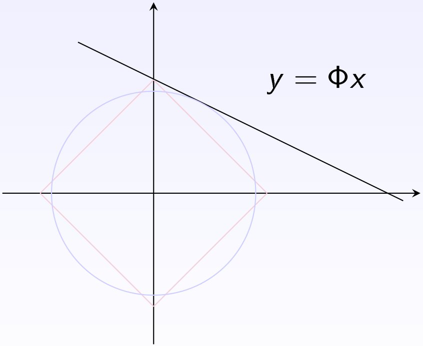
$$\Delta(y) := arg \min_{x:\ y=\Phi x} \| x \|_1$$

Particolare decodificatore:

$$\Delta(y) := arg \min_{x:\ y=\Phi x} \| x \|_1$$

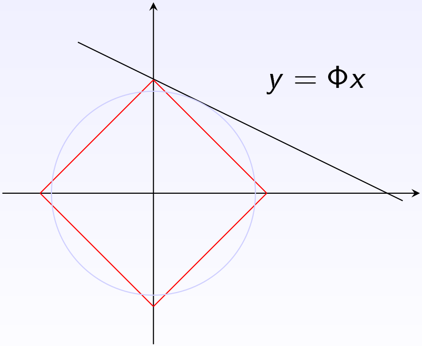
Particolare decodificatore:

$$\Delta(y) := arg \min_{x:\ y=\Phi x} \| x \|_1$$



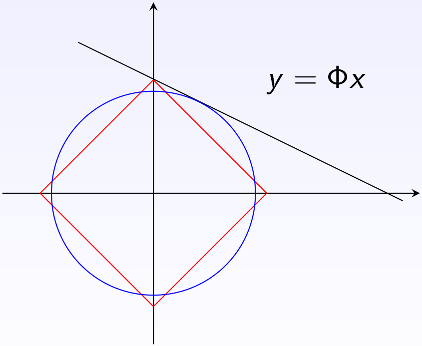
Particolare decodificatore:

$$\Delta(y) := arg \min_{x:\ y=\Phi x} \parallel x \parallel_1$$



Particolare decodificatore:

$$\Delta(y) := arg \min_{x:\ y=\Phi x} \| x \|_1$$



Definition

Una matrice Φ soddisfa alla *Restricted Isometry Property (RIP)* di ordine k se esiste un $\delta_k \in (0,1)$ tale che

$$(1-\delta_k) \parallel x \parallel_2^2 \leq \parallel \Phi x \parallel_2^2 \leq (1+\delta_k) \parallel x \parallel_2^2$$

per ogni vettore x k-sparso.

$$\sigma_k(x) = \inf\{\| x - y \|_1 : y \text{ k-sparso}\} = \sum_{j=k+1}^{N} | x_j^* |$$

Theorem

Se Φ soddisfa alla RIP di ordine $3k$ allora allora esiste una costante C tale che

$$\| x - \Delta(\Phi x) \|_2 \leq \frac{C\sigma_k(x)}{\sqrt{k}}$$

$$\sigma_k(x) = \inf\{\| x - y \|_1 : y \text{ k-sparso}\} = \sum_{j=k+1}^{N} | x_j^* |$$

Theorem

Se Φ soddisfa alla RIP di ordine $3k$ allora allora esiste una costante C tale che

$$\| x - \Delta(\Phi x) \|_2 \leq \frac{C\sigma_k(x)}{\sqrt{k}}$$

$JLT \Rightarrow RIP$

Theorem

Per ogni $\delta \in (0,1)$ se Φ è una JLT allora esiste una costante c (dipendenti solo da δ) tale che ogni $k \leq cn/\log(N/k)$ si ha che Φ soddisfa alla RIP di ordine k e $\delta_k = \delta$ (con "grande probabilità").

- $Ck \log(N/k) \leq n$
- x ha 300 componenti ed è 10-sparso. Bastano 30 osservazioni per ricostruirlo esattamente.

$JLT \Rightarrow RIP$

Theorem

Per ogni $\delta \in (0,1)$ se Φ è una JLT allora esiste una costante c (dipendenti solo da δ) tale che ogni $k \leq cn/\log(N/k)$ si ha che Φ soddisfa alla RIP di ordine k e $\delta_k = \delta$ (con "grande probabilità").

- $Ck\log(N/k) \leq n$
- x ha 300 componenti ed è 10-sparso. Bastano 30 osservazioni per ricostruirlo esattamente.

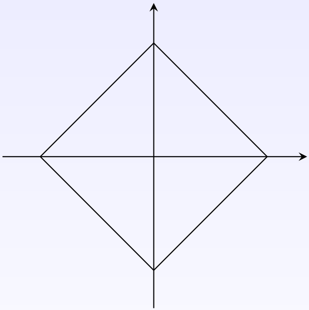
$JLT \Rightarrow RIP$

Theorem

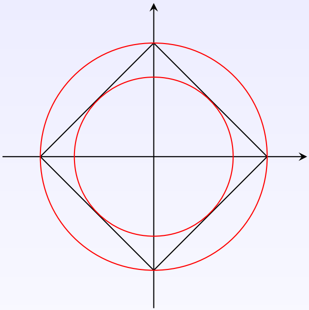
Per ogni $\delta \in (0,1)$ se Φ è una JLT allora esiste una costante c (dipendenti solo da δ) tale che ogni $k \leq cn/\log(N/k)$ si ha che Φ soddisfa alla RIP di ordine k e $\delta_k = \delta$ (con "grande probabilità").

- $Ck\log(N/k) \leq n$
- x ha 300 componenti ed è 10-sparso. Bastano 30 osservazioni per ricostruirlo esattamente.

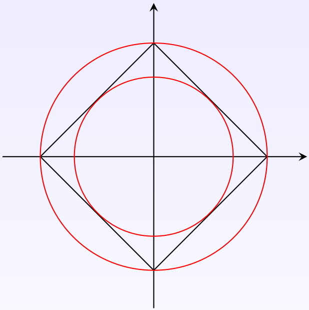
Teorema di Kashin



Teorema di Kashin

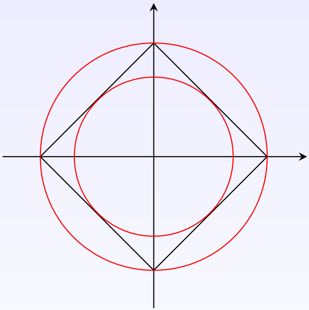


Teorema di Kashin



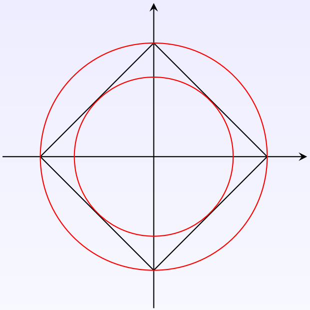
$$(x \in \mathbb{R}^2) \ \frac{1}{\sqrt{2}} B_2 \subseteq B_1 \subseteq B_2$$

Teorema di Kashin



$$(x \in \mathbb{R}^2)\ \frac{1}{\sqrt{2}} B_2 \subseteq B_1 \subseteq B_2 \qquad \frac{1}{\sqrt{n}} B_2 \subseteq B_1 \subseteq B_2\ (x \in \mathbb{R}^n)$$

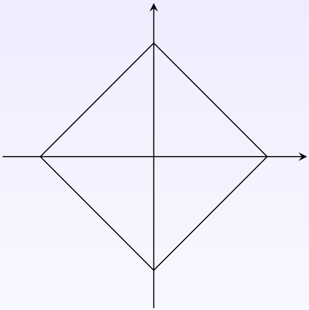
Teorema di Kashin



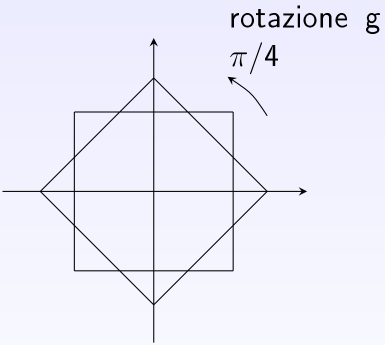
$(x \in \mathbb{R}^2)\ \frac{1}{\sqrt{2}} B_2 \subseteq B_1 \subseteq B_2$ $\qquad$ $\frac{1}{\sqrt{n}} B_2 \subseteq B_1 \subseteq B_2\ (x \in \mathbb{R}^n)$

$\| x \|_2 \leq \| x \|_1 \leq \sqrt{2} \| x \|_2$ $\qquad$ $\| x \|_2 \leq \| x \|_1 \leq \sqrt{n} \| x \|_2$

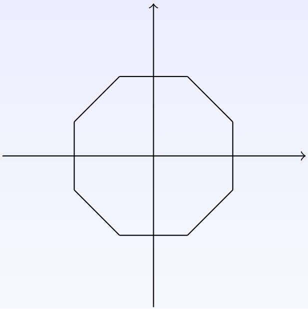
Teorema di Kashin



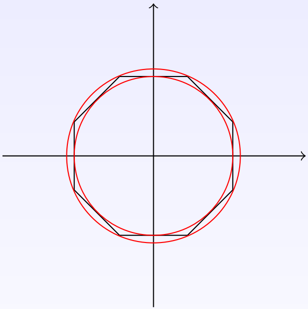
Teorema di Kashin



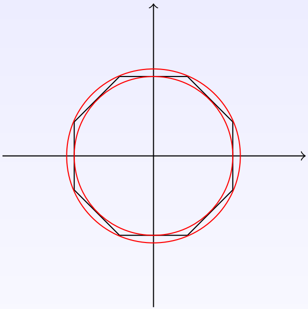
Teorema di Kashin



Teorema di Kashin

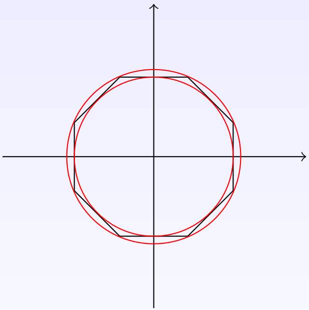


Teorema di Kashin



$$\frac{1}{\sqrt{2}}B_2 \subseteq B_1 \cap g(B_1) \subseteq \sqrt{2-\sqrt{2}}B_2$$

Teorema di Kashin



$$\frac{1}{\sqrt{2}}B_2 \subseteq B_1 \cap g(B_1) \subseteq \sqrt{2-\sqrt{2}}B_2$$

$$\frac{1}{\sqrt{2-\sqrt{2}}} \parallel x \parallel_2 \leq \max(\parallel x \parallel_1, \parallel g(x) \parallel_1) \leq \sqrt{2} \parallel x \parallel_2$$

Theorem

Esiste una costante C tale che per ogni n (sufficientemete grande) esiste una $g \in O_n$ $(g^t = g^{-1})$ per cui

$$\frac{1}{\sqrt{n}} B_2 \subseteq B_1 \cap g(B_1) \subseteq \frac{C}{\sqrt{n}} B_2$$

$$(c\sqrt{n} \parallel x \parallel_2 \leq \max(\parallel x \parallel_1, \parallel g(x) \parallel_1) \leq \sqrt{n} \parallel x \parallel_2)$$

Theorem

Esiste una costante C tale che per ogni n (sufficientemete grande) esiste una $g \in O_n$ $(g^t = g^{-1})$ per cui

$$\frac{1}{\sqrt{n}} B_2 \subseteq B_1 \cap g(B_1) \subseteq \frac{C}{\sqrt{n}} B_2$$

$$(c\sqrt{n} \parallel x \parallel_2 \leq \max(\parallel x \parallel_1, \parallel g(x) \parallel_1) \leq \sqrt{n} \parallel x \parallel_2)$$

Theorem

Esiste una costante C tale che per ogni n (sufficientemente grande) lo spazio $\mathbb{R}^{2n}$ *contiene due sottospazi ortogonali E e* $E^{\perp}$ *(dim*E *=dim*$E^{\perp} = n$*) tali che*

$$c\sqrt{n} \parallel x \parallel_2 \leq \parallel x \parallel_1 \leq \sqrt{n} \parallel x \parallel_2$$

per ogni $x \in E \cup E^{\perp}$ $\left(E = ker(I \mid g), E^{\perp} = ker(-g^t \mid I)\right)$

Corollary

Esiste una base ortonormale $\{\phi_n\}$ *di* L^2 *tale che la norma* L^2 *e* L^1 *sono equivalenti sui sottospazi* $\{\phi_n : n \text{ pari }\}$ *e* $\{\phi_n : n \text{ dispari }\}$

Theorem

Esiste una costante C tale che per ogni n (sufficientemente grande) lo spazio $\mathbb{R}^{2n}$ contiene due sottospazi ortogonali E e $E^{\perp}$ ($\dim E = \dim E^{\perp} = n$) tali che

$$c\sqrt{n} \parallel x \parallel_2 \leq \parallel x \parallel_1 \leq \sqrt{n} \parallel x \parallel_2$$

per ogni $x \in E \cup E^{\perp}$ $\left(E = ker(I \mid g), E^{\perp} = ker(-g^t \mid I)\right)$

Corollary

Esiste una base ortonormale $\{\phi_n\}$ *di* L^2 *tale che la norma* L^2 *e* L^1 *sono equivalenti sui sottospazi* $\{\phi_n : n \text{ pari }\}$ *e* $\{\phi_n : n \text{ dispari }\}$

- Grazie a metodi propri della teoria dei C-S si può dimostrare che altre mappe si possono usare al posto di g. Ad esempio tutte quelle indicate negli esempi di JLT.
- Si può ottenere il teorema di Kashin usando una JLT?

- Grazie a metodi propri della teoria dei C-S si può dimostrare che altre mappe si possono usare al posto di g. Ad esempio tutte quelle indicate negli esempi di JLT.
- Si può ottenere il teorema di Kashin usando una JLT?

$$
\begin{tikzcd}
\boxed{\text{JLT (1-2-3)}} \arrow[r] \arrow[d, "(16)"'] & \boxed{\begin{array}{c}\text{J-L Lemma (4-5-12-13)}\\ \text{(riduzione dimensionale)}\end{array}} \\
\boxed{\begin{array}{c}\text{C-S (RIP)}\\ \text{(6-15)}\end{array}} \arrow[ur, "(8)"'] \arrow[r, "(14)"'] & \boxed{\text{Kashin splitting (7-9-10-11)}}
\end{tikzcd}
$$

1) Suresh Venkatasubramanian, Qiushi Wang
The Johnson-Lindenstrauss Transform: An Empirical Study
http://www.siam.org/proceedings/alenex/2011/ alx11_16_venkatasubramanians.pdf
2) Anirban Dasgupta, Ravi Kumar, Tamás Sarlós
A Sparse Johnson–Lindenstrauss Transform
http://arxiv.org/abs/1004.4240
3) Daniel M. Kaney Jelani Nelsonz
A Derandomized Sparse Johnson-Lindenstrauss Transform
eccc.hpi-web.de/report/2010/098/revision/1/download/
4) D. Achlioptas.
Database-friendly random projections.
(PODS), Santa Barbara, CA, May 2001.
5) S. Dasgupta and A. Gupta.
An elementary proof of the johnson-lindenstrauss lemma
TR-99-006, Univ. of Cal. Berkeley, Comput. Science Division, Mar. 1999.
6) D. Donoho.
Compressed sensing.
IEEE Trans. Inform. Theory, 52(4):12898211;1306, 2006.
7) A. Garnaev and E. Gluskin.
The widths of euclidean balls.
Dokl. An. SSSR, 277:10488211;1052, 1984.
8) F. Krahmer and R. Ward.
New and improved johnson-lindenstrauss embeddings via the restricted isometry property

Preprint, Sept. 2010.
9) B.S. Kashin,
The widths of certain finite-dimensional sets and classes of smooth functions.
Izv. Akad. Nauk SSSR (Ser. Mat.), 41 (2) (1977), 334 – 351. English transl.:
Math. USSR-Izv., 11 (2) (1977), 317 – 333.
10) G. Schechtman,
Special orthogonal splittings of L_1^{2k}.
Israel Journal of Mathematics, 139 (2004), 337 – 347.
11) S.J. Szarek,
On Kashin's almost Euclidean orthogonal decomposition of l_n^1
Bull. Acad. Polon. Sci. SŽer. Sci. Math. Astronom. Phys., 26 (8) (1978), 691 – 694
12) W. B. Johnson and J. Lindenstrauss
Extensions of Lipschitz maps into a Hilbert space
Contemp Math 26 (1984), 189 – 206.
13) Sanjoy Dasgupta, Anupam Gupta
An Elementary Proof of a Theorem of Johnson and Lindenstrauss
http://cseweb.ucsd.edu/∼dasgupta/papers/jl.pdf
14) http://www.math.drexel.edu/∼foucart/CSK.pdf
15) http://dsp.rice.edu/cs
16) Richard Baraniuk · Mark Davenport
A Simple Proof of the Restricted Isometry Property for Random Matrices
http://pages.cs.wisc.edu/∼brecht/cs838docs/Baraniuk_JL_RIP.pdf