On some variants of Johnson-Lindenstrauss lemma

Paweł Wolff

Case Western Reserve University

Perspectives in High Dimensions
Cleveland, August 2010

Linear Johnson-Lindenstrauss lemma

Theorem (Johnson-Lindenstrauss, 1984)

$x_1, \ldots, x_n \in \mathbb{R}^n$ — points in Euclidean space.
For any $\varepsilon > 0$ there exists a linear map $T\colon \mathbb{R}^n \to \mathbb{R}^k$ with

$$k \leq C\varepsilon^{-2} \log n$$

such that

$$(1-\varepsilon)\|x_i - x_j\| \leq \|Tx_i - Tx_j\| \leq (1+\varepsilon)\|x_i - x_j\|.$$

- For N points in $\mathbb{R}^n$, $N \geq n$,

$$k \leq C\varepsilon^{-2} \log N.$$

- $N = n$.

Linear Johnson-Lindenstrauss lemma

Theorem (Johnson-Lindenstrauss, 1984)

$x_1, \ldots, x_n \in \mathbb{R}^n$ — points in Euclidean space.
For any $\varepsilon > 0$ there exists a linear map $T\colon \mathbb{R}^n \to \mathbb{R}^k$ with

$$k \leq C\varepsilon^{-2} \log n$$

such that

$$(1-\varepsilon)\|x_i - x_j\| \leq \|Tx_i - Tx_j\| \leq (1+\varepsilon)\|x_i - x_j\|.$$

- For N points in $\mathbb{R}^n$, $N \geq n$,

$$k \leq C\varepsilon^{-2} \log N.$$

- $N = n$.

Linear Johnson-Lindenstrauss lemma

Theorem (Johnson-Lindenstrauss, 1984)

$x_1, \ldots, x_n \in \mathbb{R}^n$ — points in Euclidean space.
For any $\varepsilon > 0$ there exists a linear map $T\colon \mathbb{R}^n \to \mathbb{R}^k$ with

$$k \leq C\varepsilon^{-2} \log n$$

such that

$$(1-\varepsilon)\|x_i - x_j\| \leq \|Tx_i - Tx_j\| \leq (1+\varepsilon)\|x_i - x_j\|.$$

- For N points in $\mathbb{R}^n$, $N \geq n$,

$$k \leq C\varepsilon^{-2} \log N.$$

- $N = n$.

Typical proof (ideas)

We want $T : \mathbb{R}^n \to \mathbb{R}^k$ such that

$$\|T(x_i - x_j)\| \approx \|x_i - x_j\| \quad \text{for } \binom{n}{2} \text{ vectors } x_i - x_j.$$

Probabilistic proof:

- Take **random** $T : \mathbb{R}^n \to \mathbb{R}^k$
- Show that for any **fixed** vector $v \in \mathbb{R}^n$,

$$\|Tv\| \approx \|v\|$$

  with high probability $(> 1 - \frac{1}{n^2})$
- Take union bound over $\binom{n}{2}$ vectors $x_i - x_j$
- Conclusion: T works with probability $> \frac{1}{2}$.

Typical proof (ideas)

We want $T\colon \mathbb{R}^n \to \mathbb{R}^k$ such that

$$\|T(x_i - x_j)\| \approx \|x_i - x_j\| \quad \text{for } \binom{n}{2} \text{ vectors } x_i - x_j.$$

Probabilistic proof:

- Take **random** $T\colon \mathbb{R}^n \to \mathbb{R}^k$
- Show that for any **fixed** vector $v \in \mathbb{R}^n$,

  $$\|Tv\| \approx \|v\|$$

  with high probability ($> 1 - \frac{1}{n^2}$)
- Take union bound over $\binom{n}{2}$ vectors $x_i - x_j$
- Conclusion: T works with probability $> \frac{1}{2}$.

Random operator $T : \mathbb{R}^n \to \mathbb{R}^k$

Examples of random operators from $\mathbb{R}^n$ to $\mathbb{R}^k$:

- orthogonal projection onto a random k-dimensional subspace (original proof, 1984)
  (for proper scaling, $\times \sqrt{\frac{n}{k}}$)
- $T = \frac{1}{\sqrt{k}} (g_{ij})_{i \le k, j \le n}$, (g_{ij}) — i.i.d. $\mathcal{N}(0,1)$
  (Indyk-Motwani, 1998)
- $T = \frac{1}{\sqrt{k}} (\varepsilon_{ij})_{i \le k, j \le n}$, (ε_{ij}) — indep. ± 1
  Also for indep. 3-point r.v.'s: $-1, 0, 1 \quad \rightsquigarrow$ sparse matrix T
  (Achlioptas, 2003)

Objectives:

- simplicity of implementation (± 1 better than $\mathcal{N}(0,1)$)
- computing time of applying T to x_i
  (more complete picture for N points in $\mathbb{R}^n$, $N \gg n$)
- amount of randomness used to sample T

Random operator $T : \mathbb{R}^n \to \mathbb{R}^k$

Examples of random operators from $\mathbb{R}^n$ to $\mathbb{R}^k$:

- orthogonal projection onto a random k-dimensional subspace (original proof, 1984)
  (for proper scaling, $\times \sqrt{\frac{n}{k}}$)
- $T = \frac{1}{\sqrt{k}} (g_{ij})_{i \le k, j \le n}$, (g_{ij}) — i.i.d. $\mathcal{N}(0,1)$
  (Indyk-Motwani, 1998)
- $T = \frac{1}{\sqrt{k}} (\varepsilon_{ij})_{i \le k, j \le n}$, (ε_{ij}) — indep. ± 1
  Also for indep. 3-point r.v.'s: $-1, 0, 1 \quad \rightsquigarrow$ sparse matrix T
  (Achlioptas, 2003)

Objectives:

- simplicity of implementation (± 1 better than $\mathcal{N}(0,1)$)
- computing time of applying T to x_i
  (more complete picture for N points in $\mathbb{R}^n$, $N \gg n$)
- amount of randomness used to sample T

Random operator $T: \mathbb{R}^n \to \mathbb{R}^k$

Examples of random operators from $\mathbb{R}^n$ to $\mathbb{R}^k$:

- orthogonal projection onto a random k-dimensional subspace (original proof, 1984)
  (for proper scaling, $\times\sqrt{\frac{n}{k}}$)
- $T = \frac{1}{\sqrt{k}}\left(g_{ij}\right)_{i\le k, j\le n}$, (g_{ij}) — i.i.d. $\mathcal{N}(0,1)$
  (Indyk-Motwani, 1998)
- $T = \frac{1}{\sqrt{k}}\left(\varepsilon_{ij}\right)_{i\le k, j\le n}$, (ε_{ij}) — indep. ± 1
  Also for indep. 3-point r.v.'s: $-1, 0, 1 \quad \rightsquigarrow$ sparse matrix T
  (Achlioptas, 2003)

Objectives:

- simplicity of implementation (± 1 better than $\mathcal{N}(0,1)$)
- computing time of applying T to x_i
  (more complete picture for N points in $\mathbb{R}^n$, $N \gg n$)
- amount of randomness used to sample T

Random operator $T : \mathbb{R}^n \to \mathbb{R}^k$

Examples of random operators from $\mathbb{R}^n$ to $\mathbb{R}^k$:

- orthogonal projection onto a random k-dimensional subspace (original proof, 1984)
  (for proper scaling, $\times \sqrt{\frac{n}{k}}$)
- $T = \frac{1}{\sqrt{k}} (g_{ij})_{i \le k, j \le n}$, (g_{ij}) — i.i.d. $\mathcal{N}(0,1)$
  (Indyk-Motwani, 1998)
- $T = \frac{1}{\sqrt{k}} (\varepsilon_{ij})_{i \le k, j \le n}$, (ε_{ij}) — indep. ± 1
  Also for indep. 3-point r.v.'s: $-1, 0, 1 \quad \rightsquigarrow$ sparse matrix T
  (Achlioptas, 2003)

Objectives:

- simplicity of implementation (± 1 better than $\mathcal{N}(0,1)$)
- computing time of applying T to x_i
  (more complete picture for N points in $\mathbb{R}^n$, $N \gg n$)
- amount of randomness used to sample T

Random operator $T : \mathbb{R}^n \to \mathbb{R}^k$

Examples of random operators from $\mathbb{R}^n$ to $\mathbb{R}^k$:

- orthogonal projection onto a random k-dimensional subspace (original proof, 1984)
  (for proper scaling, $\times \sqrt{\frac{n}{k}}$)
- $T = \frac{1}{\sqrt{k}} (g_{ij})_{i \le k, j \le n}$, (g_{ij}) — i.i.d. $\mathcal{N}(0,1)$
  (Indyk-Motwani, 1998)
- $T = \frac{1}{\sqrt{k}} (\varepsilon_{ij})_{i \le k, j \le n}$, (ε_{ij}) — indep. ± 1
  Also for indep. 3-point r.v.'s: $-1, 0, 1 \quad \rightsquigarrow$ sparse matrix T
  (Achlioptas, 2003)

Objectives:

- simplicity of implementation (± 1 better than $\mathcal{N}(0,1)$)
- computing time of applying T to x_i
  (more complete picture for N points in $\mathbb{R}^n$, $N \gg n$)
- amount of randomness used to sample T

Fast Johnson-Lindenstrauss transform (Ailon-Chazelle, 2006)

- To speed up computation of Tx_i's — use really sprase random matrix:

$$T = (\xi_{ij}), \quad \text{where } \xi_{ij} = \begin{cases} g_{ij} \sim \mathcal{N}(0,1) \text{ w.p. } p \\ 0 \text{ w.p. } 1-p \end{cases}$$

  and p is small (actually $p \sim \frac{\log^2 n}{n} \rightsquigarrow k \log^2 n$ non-zero entries (on average)

- Problem arises — some of $x_i - x_j$ might be $(1, 0, \ldots, 0)$ which is likely to be in $\ker T$...
  and no hope for $\|T(x_i - x_j)\| \approx \|x_i - x_j\|$.
- On the other hand, T turns out to work well for "spread" vectors

Fast Johnson-Lindenstrauss transform (Ailon-Chazelle, 2006)

- To speed up computation of Tx_i's — use really sprase random matrix:

$$T = (\xi_{ij}), \quad \text{where } \xi_{ij} = \begin{cases} g_{ij} \sim \mathcal{N}(0,1) \text{ w.p. } p \\ 0 \text{ w.p. } 1-p \end{cases}$$

  and p is small (actually $p \sim \frac{\log^2 n}{n} \rightsquigarrow k \log^2 n$ non-zero entries (on average)
- Problem arises — some of $x_i - x_j$ might be $(1, 0, \ldots, 0)$ which is likely to be in $\ker T$...
  and no hope for $\|T(x_i - x_j)\| \approx \|x_i - x_j\|$.
- On the other hand, T turns out to work well for "spread" vectors

Fast Johnson-Lindenstrauss transform (Ailon-Chazelle, 2006)

- To speed up computation of Tx_i's — use really sprase random matrix:

$$T = (\xi_{ij}), \quad \text{where } \xi_{ij} = \begin{cases} g_{ij} \sim \mathcal{N}(0,1) \text{ w.p. } p \\ 0 \text{ w.p. } 1-p \end{cases}$$

  and p is small (actually $p \sim \frac{\log^2 n}{n} \rightsquigarrow k \log^2 n$ non-zero entries (on average)
- Problem arises — some of $x_i - x_j$ might be $(1, 0, \ldots, 0)$ which is likely to be in $\ker T$...
  and no hope for $\|T(x_i - x_j)\| \approx \|x_i - x_j\|$.
- On the other hand, T turns out to work well for "spread" vectors

Fast J.-L. transform (Ailon-Chazelle, 2006)

Theorem (Ailon-Chazelle, 2006; Matoušek, 2008)

Take $p = \frac{\log^2 n}{n}$ and

$$T = \frac{1}{\sqrt{knp}} MHD,$$

where

- $M = (b_{ij} X_{ij})_{i \le k, j \le n}$ is a sparse random matrix and
  - $\mathbb{P}(b_{ij} = 1) = p, \mathbb{P}(b_{ij} = 0) = 1 - p$
  - $X_{ij} \sim \mathcal{N}(0, 1)$ (Ailon-Chazelle)
  - $X_{ij} \sim \pm 1$ (Matoušek)
- H is an $n \times n$ Walsh-Hadamard matrix (assume $n = 2^l$)
- $D = \begin{pmatrix} \pm 1 & & \\ & \ddots & \\ & & \pm 1 \end{pmatrix}$

Then, for a fixed n points $x_1, \ldots, x_n \in \mathbb{R}^n$, $T : \mathbb{R}^n \to \mathbb{R}^k$ works with h.p.

Fast J.-L. transform (Ailon-Chazelle, 2006)

$$H_{4\times 4} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad H_{8\times 8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Why does it work?

- $\frac{1}{\sqrt{n}}HD\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isometry and with h.p makes the vectors x_i spread:
  for any unit x, the vector $y = \frac{1}{\sqrt{n}}HDx$ has

$$\|y\|_\infty = O\left(\sqrt{(\log n)/n}\right) \text{ with h.p.}$$

- $z = \frac{1}{\sqrt{kp}}My, \qquad M = (b_{ij}X_{ij}),$

$$\mathbb{E}\|z\|^2 = \|x\|^2 = 1$$

- $\|z\|^2$ concentrates well enough around $\mathbb{E}\|z\|^2$ (this is true as long as y was spread)

Fast J.-L. transform (Ailon-Chazelle, 2006)

$$H_{4\times 4} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad H_{8\times 8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Why does it work?

- $\frac{1}{\sqrt{n}}HD\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isometry and with h.p makes the vectors x_i spread:
  for any unit x, the vector $y = \frac{1}{\sqrt{n}}HDx$ has

$$\|y\|_\infty = O\left(\sqrt{(\log n)/n}\right) \text{ with h.p.}$$

- $z = \frac{1}{\sqrt{kp}}My, \qquad M = (b_{ij}X_{ij}),$

$$\mathbb{E}\|z\|^2 = \|x\|^2 = 1$$

- $\|z\|^2$ concentrates well enough around $\mathbb{E}\|z\|^2$ (this is true as long as y was spread)

Fast J.-L. transform (Ailon-Chazelle, 2006)

$$H_{4\times 4} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad H_{8\times 8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Why does it work?

- $\frac{1}{\sqrt{n}}HD\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isometry and with h.p makes the vectors x_i spread:
  for any unit x, the vector $y = \frac{1}{\sqrt{n}}HDx$ has

$$\|y\|_\infty = O\left(\sqrt{(\log n)/n}\right) \text{ with h.p.}$$

- $z = \frac{1}{\sqrt{kp}}My, \quad M = (b_{ij}X_{ij}),$

$$\mathbb{E}\|z\|^2 = \|x\|^2 = 1$$

- $\|z\|^2$ concentrates well enough around $\mathbb{E}\|z\|^2$ (this is true as long as y was spread)

Fast J.-L. transform (Ailon-Chazelle, 2006)

$$H_{4\times 4} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad H_{8\times 8} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Why does it work?

- $\frac{1}{\sqrt{n}}HD\colon \mathbb{R}^n \to \mathbb{R}^n$ is an isometry and with h.p makes the vectors x_i spread:
  for any unit x, the vector $y = \frac{1}{\sqrt{n}}HDx$ has

$$\|y\|_\infty = O\left(\sqrt{(\log n)/n}\right) \text{ with h.p.}$$

- $z = \frac{1}{\sqrt{kp}}My, \quad M = (b_{ij}X_{ij}),$

$$\mathbb{E}\|z\|^2 = \|x\|^2 = 1$$

- $\|z\|^2$ concentrates well enough around $\mathbb{E}\|z\|^2$ (this is true as long as y was spread)

Fast J.-L. transform (Ailon-Chazelle, 2006)

- Computation time of $MHDx_i$:
  - $y_i = HDx_i$ — $O(n \log n)$ (DFT)
  - $z_i = My_i$ — $O(pkn) = O\left(k \log^2 n\right) = O\left(\varepsilon^{-2} \log^3 n\right)$
  - compare with "standard" approach — $O(kn) = O\left(\varepsilon^{-2} n \log n\right)$.
  - significantly beats previous approaches if # of points is $N \gg n$.
- Amount of randomness:
  - n random bits for $D = \begin{pmatrix} \pm 1 & & \\ & \ddots & \\ & & \pm 1 \end{pmatrix}$,
  - at least nk random bits for M...

Fast J.-L. transform (Ailon-Chazelle, 2006)

- Computation time of $MHDx_i$:
  - $y_i = HDx_i$ — $O(n \log n)$ (DFT)
  - $z_i = My_i$ — $O(pkn) = O\left(k \log^2 n\right) = O\left(\varepsilon^{-2} \log^3 n\right)$
  - compare with "standard" approach — $O(kn) = O\left(\varepsilon^{-2} n \log n\right)$.
  - significantly beats previous approaches if # of points is $N \gg n$.
- Amount of randomness:
  - n random bits for $D = \begin{pmatrix} \pm 1 & & \\ & \ddots & \\ & & \pm 1 \end{pmatrix}$,
  - at least nk random bits for M...

Circulant matrices (Hinrichs-Vybíral, 2010)

Take $T = \frac{1}{\sqrt{k}} CD \colon \mathbb{R}^n \to \mathbb{R}^k$, where

- $C = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ X_n & X_1 & \dots & X_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-k+1} & X_{n-k+2} & \dots & X_{n-k} \end{pmatrix}$
  is a $k \times n$ partial circulant matrix and X_i are $\mathcal{N}(0,1)$ (or ± 1),
- $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$.

Such random operator T works for J.-L. lemma if

- $k \sim \varepsilon^{-2} \log^3 n$ (Hinrichs-Vybíral, 2010)
- $k \sim \varepsilon^{-2} \log^2 n$ and $X_i \sim \mathcal{N}(0,1)$ (Vybíral, 2010)
- above also true for $X_i = \pm 1$.

Only $O(n)$ random bits!
But dimension reduction weaker...

Circulant matrices (Hinrichs-Vybíral, 2010)

Take $T = \frac{1}{\sqrt{k}} CD \colon \mathbb{R}^n \to \mathbb{R}^k$, where

- $C = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ X_n & X_1 & \dots & X_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-k+1} & X_{n-k+2} & \dots & X_{n-k} \end{pmatrix}$
  is a $k \times n$ partial circulant matrix and X_i are $\mathcal{N}(0,1)$ (or ± 1),
- $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$.

Such random operator T works for J.-L. lemma if

- $k \sim \varepsilon^{-2} \log^3 n$ (Hinrichs-Vybíral, 2010)
- $k \sim \varepsilon^{-2} \log^2 n$ and $X_i \sim \mathcal{N}(0,1)$ (Vybíral, 2010)
- above also true for $X_i = \pm 1$.

Only $O(n)$ random bits!
But dimension reduction weaker...

Circulant matrices (Hinrichs-Vybíral, 2010)

Take $T = \frac{1}{\sqrt{k}} CD \colon \mathbb{R}^n \to \mathbb{R}^k$, where

- $C = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ X_n & X_1 & \dots & X_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-k+1} & X_{n-k+2} & \dots & X_{n-k} \end{pmatrix}$
  is a $k \times n$ partial circulant matrix and X_i are $\mathcal{N}(0,1)$ (or ± 1),
- $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$.

Such random operator T works for J.-L. lemma if

- $k \sim \varepsilon^{-2} \log^3 n$ (Hinrichs-Vybíral, 2010)
- $k \sim \varepsilon^{-2} \log^2 n$ and $X_i \sim \mathcal{N}(0,1)$ (Vybíral, 2010)
- above also true for $X_i = \pm 1$.

Only $O(n)$ random bits!
But dimension reduction weaker...

Circulant matrices (Hinrichs-Vybíral, 2010)

Take $T = \frac{1}{\sqrt{k}} CD \colon \mathbb{R}^n \to \mathbb{R}^k$, where

- $C = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ X_n & X_1 & \dots & X_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-k+1} & X_{n-k+2} & \dots & X_{n-k} \end{pmatrix}$
  is a $k \times n$ partial circulant matrix and X_i are $\mathcal{N}(0,1)$ (or ± 1),
- $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$.

Such random operator T works for J.-L. lemma if

- $k \sim \varepsilon^{-2} \log^3 n$ (Hinrichs-Vybíral, 2010)
- $k \sim \varepsilon^{-2} \log^2 n$ and $X_i \sim \mathcal{N}(0,1)$ (Vybíral, 2010)
- above also true for $X_i = \pm 1$.

Only $O(n)$ random bits!
But dimension reduction weaker...

Circulant matrices (Hinrichs-Vybíral, 2010)

Take $T = \frac{1}{\sqrt{k}} CD \colon \mathbb{R}^n \to \mathbb{R}^k$, where

- $C = \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ X_n & X_1 & \dots & X_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-k+1} & X_{n-k+2} & \dots & X_{n-k} \end{pmatrix}$
  is a $k \times n$ partial circulant matrix and X_i are $\mathcal{N}(0,1)$ (or ± 1),
- $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$.

Such random operator T works for J.-L. lemma if

- $k \sim \varepsilon^{-2} \log^3 n$ (Hinrichs-Vybíral, 2010)
- $k \sim \varepsilon^{-2} \log^2 n$ and $X_i \sim \mathcal{N}(0,1)$ (Vybíral, 2010)
- above also true for $X_i = \pm 1$.

Only $O(n)$ random bits!
But dimension reduction weaker...

Back to FJLT — randomness reduction (W., 2010)

One can modify FJLT slightly and use (essentially) only $O(n)$ random bits:

- Recall: the sparse matrix M used in FJTL had **on average** $pn = \log^2 n$ non-zero entries in each row.
- What if we **fix** (or condition on) # of non-zero entries in each row to be **exactly** $m = \log^2 n$?
- Take $T = \frac{1}{\sqrt{mk}} MHD \colon \mathbb{R}^n \to \mathbb{R}^k$ with $k \sim \varepsilon^{-2} \log n$ where
  - H — $n \times n$ Walsh-Hadamard matrix
  - $D = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$
  - M — random $k \times n$ matrix with entries $1, 0, -1$:
    - rows of M are i.i.d. vectors in $\mathbb{R}^n$
    - each row of M contains exactly $m = \log^2 n$ non-zero entries which are indep. random signs (± 1)
    - in each row of M, the set of indices of non-zero entries is a random subset drawn from uniform distribution

Back to FJLT — randomness reduction (W., 2010)

One can modify FJLT slightly and use (essentially) only $O(n)$ random bits:

- Recall: the sparse matrix M used in FJTL had **on average** $pn = \log^2 n$ non-zero entries in each row.
- What if we **fix** (or condition on) # of non-zero entries in each row to be **exactly** $m = \log^2 n$?
- Take $T = \frac{1}{\sqrt{mk}} MHD \colon \mathbb{R}^n \to \mathbb{R}^k$ with $k \sim \varepsilon^{-2} \log n$ where
  - H — $n \times n$ Walsh-Hadamard matrix
  - $D = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$
  - M — random $k \times n$ matrix with entries $1, 0, -1$:
    - rows of M are i.i.d. vectors in $\mathbb{R}^n$
    - each row of M contains exactly $m = \log^2 n$ non-zero entries which are indep. random signs (± 1)
    - in each row of M, the set of indices of non-zero entries is a random subset drawn from uniform distribution

Back to FJLT — randomness reduction (W., 2010)

One can modify FJLT slightly and use (essentially) only $O(n)$ random bits:

- Recall: the sparse matrix M used in FJTL had **on average** $pn = \log^2 n$ non-zero entries in each row.
- What if we **fix** (or condition on) # of non-zero entries in each row to be **exactly** $m = \log^2 n$?
- Take $T = \frac{1}{\sqrt{mk}} MHD \colon \mathbb{R}^n \to \mathbb{R}^k$ with $k \sim \varepsilon^{-2} \log n$ where
  - H — $n \times n$ Walsh-Hadamard matrix
  - $D = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n)$
  - M — random $k \times n$ matrix with entries $1, 0, -1$:
    - rows of M are i.i.d. vectors in $\mathbb{R}^n$
    - each row of M contains exactly $m = \log^2 n$ non-zero entries which are indep. random signs (± 1)
    - in each row of M, the set of indices of non-zero entries is a random subset drawn from uniform distribution

Back to FJLT — randomness reduction (W., 2010)

Amount of randomness we use to sample T:

- D: n random bits
- M: for each row draw $m = \log^2 n$ (distinct) indices from the range $[1..n]$ — this can be done using $\sim m \times \log n$ random bits $\log^3 n$ random bits in each of k rows

Overall, T can be sampled using

$$O(n + km \log n) = O(n + \varepsilon^{-2} \log^4 n) \text{ random bits}$$

Back to FJLT — randomness reduction (W., 2010)

Amount of randomness we use to sample T:

- D: n random bits
- M: for each row draw $m = \log^2 n$ (distinct) indices from the range $[1..n]$ — this can be done using $\sim m \times \log n$ random bits $\log^3 n$ random bits in each of k rows

Overall, T can be sampled using

$$O(n + km \log n) = O(n + \varepsilon^{-2} \log^4 n) \text{ random bits}$$

Back to FJLT — randomness reduction (W., 2010)

Validity of the construction relies on a certain deviation inq. (cf. Matoušek, 2008 for a version with i.i.d. 3-point r.v.'s):

- $(\xi_1, \ldots, \xi_n) \in \{1, 0, -1\}^n$ — single row of M:

$$\mathbb{P}(\xi_1 = \pm a_1, \ldots, \xi_n = \pm a_n) = 2^{-m} / \binom{n}{m}$$

  for $(a_1, \ldots, a_n) \in \{0, 1\}^n$ with $\sum a_i = m$.
- Note: (ξ_i) are dependent r.v.'s. However, they are uncorrelated!
- (Recall the context:

$$y = \frac{1}{\sqrt{k}} HDx, x \in \mathbb{R}^n, \|x\| = 1 \quad \Longrightarrow \quad \|y\|_\infty = O(\frac{\sqrt{\log n}}{\sqrt{n}}) \text{ w.h.p.})$$

Proposition (W., 2010)

For a unit $y \in \mathbb{R}^n$ with $\|y\|_\infty \leq \alpha$, put $S = \sum_{i=1}^n y_i \xi_i$. Then $\mathbb{E} e^{uS} \leq e^{\frac{m}{n} u^2}$ for $|u| \leq \frac{\sqrt{2}}{\alpha}$. In consequence, for $0 \leq t \leq \frac{m}{n} \frac{2\sqrt{2}}{\alpha}$,

$$\mathbb{P}(|S| \geq t) \leq 2e^{-\frac{n}{m} t^2/4}.$$

Back to FJLT — randomness reduction (W., 2010)

Validity of the construction relies on a certain deviation inq. (cf. Matoušek, 2008 for a version with i.i.d. 3-point r.v.'s):

- $(\xi_1, \ldots, \xi_n) \in \{1, 0, -1\}^n$ — single row of M:

$$\mathbb{P}(\xi_1 = \pm a_1, \ldots, \xi_n = \pm a_n) = 2^{-m} / \binom{n}{m}$$

  for $(a_1, \ldots, a_n) \in \{0, 1\}^n$ with $\sum a_i = m$.
- Note: (ξ_i) are dependent r.v.'s. However, they are uncorrelated!
- (Recall the context:

$$y = \frac{1}{\sqrt{k}} HDx, x \in \mathbb{R}^n, \|x\| = 1 \quad \Longrightarrow \quad \|y\|_\infty = O(\frac{\sqrt{\log n}}{\sqrt{n}}) \text{ w.h.p.})$$

Proposition (W., 2010)

For a unit $y \in \mathbb{R}^n$ with $\|y\|_\infty \leq \alpha$, put $S = \sum_{i=1}^n y_i \xi_i$. Then $\mathbb{E} e^{uS} \leq e^{\frac{m}{n} u^2}$ for $|u| \leq \frac{\sqrt{2}}{\alpha}$. In consequence, for $0 \leq t \leq \frac{m}{n} \frac{2\sqrt{2}}{\alpha}$,

$$\mathbb{P}(|S| \geq t) \leq 2e^{-\frac{n}{m} t^2/4}.$$

Back to FJLT — randomness reduction (W., 2010)

Validity of the construction relies on a certain deviation inq. (cf. Matoušek, 2008 for a version with i.i.d. 3-point r.v.'s):

- $(\xi_1, \ldots, \xi_n) \in \{1, 0, -1\}^n$ — single row of M:
$$\mathbb{P}(\xi_1 = \pm a_1, \ldots, \xi_n = \pm a_n) = 2^{-m} / \binom{n}{m}$$
for $(a_1, \ldots, a_n) \in \{0, 1\}^n$ with $\sum a_i = m$.
- Note: (ξ_i) are dependent r.v.'s. However, they are uncorrelated!
- (Recall the context:
$y = \frac{1}{\sqrt{k}} HDx, x \in \mathbb{R}^n, \|x\| = 1 \implies \|y\|_\infty = O(\frac{\sqrt{\log n}}{\sqrt{n}})$ w.h.p.)

Proposition (W., 2010)

For a unit $y \in \mathbb{R}^n$ with $\|y\|_\infty \leq \alpha$, put $S = \sum_{i=1}^n y_i \xi_i$. Then $\mathbb{E} e^{uS} \leq e^{\frac{m}{n} u^2}$ for $|u| \leq \frac{\sqrt{2}}{\alpha}$. In consequence, for $0 \leq t \leq \frac{m}{n} \frac{2\sqrt{2}}{\alpha}$,

$$\mathbb{P}(|S| \geq t) \leq 2e^{-\frac{n}{m} t^2/4}.$$

Back to FJLT — randomness reduction (W., 2010)

Validity of the construction relies on a certain deviation inq. (cf. Matoušek, 2008 for a version with i.i.d. 3-point r.v.'s):

- $(\xi_1, \ldots, \xi_n) \in \{1, 0, -1\}^n$ — single row of M:
$$\mathbb{P}(\xi_1 = \pm a_1, \ldots, \xi_n = \pm a_n) = 2^{-m} / \binom{n}{m}$$
for $(a_1, \ldots, a_n) \in \{0, 1\}^n$ with $\sum a_i = m$.
- Note: (ξ_i) are dependent r.v.'s. However, they are uncorrelated!
- (Recall the context:
$y = \frac{1}{\sqrt{k}} HDx, x \in \mathbb{R}^n, \|x\| = 1 \implies \|y\|_\infty = O(\frac{\sqrt{\log n}}{\sqrt{n}})$ w.h.p.)

Proposition (W., 2010)

For a unit $y \in \mathbb{R}^n$ with $\|y\|_\infty \leq \alpha$, put $S = \sum_{i=1}^n y_i \xi_i$. Then $\mathbb{E} e^{uS} \leq e^{\frac{m}{n} u^2}$ for $|u| \leq \frac{\sqrt{2}}{\alpha}$. In consequence, for $0 \leq t \leq \frac{m}{n} \frac{2\sqrt{2}}{\alpha}$,

$$\mathbb{P}(|S| \geq t) \leq 2e^{-\frac{n}{m} t^2/4}.$$

Back to FJLT — randomness reduction (W., 2010)

Validity of the construction relies on a certain deviation inq. (cf. Matoušek, 2008 for a version with i.i.d. 3-point r.v.'s):

- $(\xi_1, \ldots, \xi_n) \in \{1, 0, -1\}^n$ — single row of M:
$$\mathbb{P}(\xi_1 = \pm a_1, \ldots, \xi_n = \pm a_n) = 2^{-m} / \binom{n}{m}$$
for $(a_1, \ldots, a_n) \in \{0, 1\}^n$ with $\sum a_i = m$.
- Note: (ξ_i) are dependent r.v.'s. However, they are uncorrelated!
- (Recall the context:
$y = \frac{1}{\sqrt{k}} HDx, x \in \mathbb{R}^n, \|x\| = 1 \implies \|y\|_\infty = O(\frac{\sqrt{\log n}}{\sqrt{n}})$ w.h.p.)

Proposition (W., 2010)

For a unit $y \in \mathbb{R}^n$ with $\|y\|_\infty \leq \alpha$, put $S = \sum_{i=1}^n y_i \xi_i$. Then $\mathbb{E} e^{uS} \leq e^{\frac{m}{n} u^2}$ for $|u| \leq \frac{\sqrt{2}}{\alpha}$. In consequence, for $0 \leq t \leq \frac{m}{n} \frac{2\sqrt{2}}{\alpha}$,

$$\mathbb{P}(|S| \geq t) \leq 2e^{-\frac{n}{m} t^2/4}.$$