ランダムプロジェクションとスパースネス

鈴木　大慈
東京大学情報理工学系研究科
数理情報学専攻

2010/7/16→7/26

- Compressed Sensing (CS) – encoding
  - ランダムプロジェクション
  - Johnson–Lindenstrauss Lemma

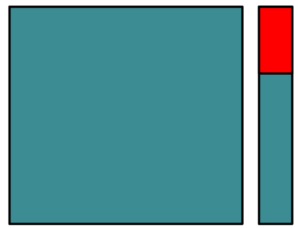
[Candes, Romberg, and Tao: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, 2006]
[Donoho: Compressed sensing. IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)]

- Lasso & Dantzig Selector – decoding
  - スパース表現のリカバリー

Compressed Sensing (CS)

[Logan-Shepp phantom]

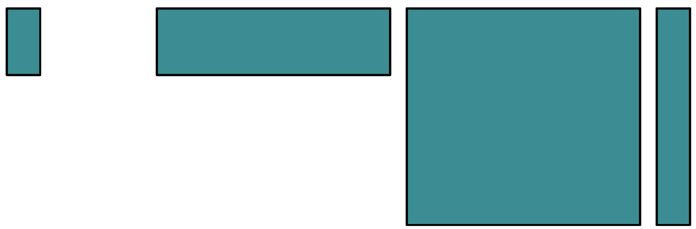
$$= \Psi x_0$$



d-sparse

ランダム

Random Sampling $$y = \Phi \Psi x_0 =: A x_0$$



Gaussian, Bernoulli, Hadamard matrix,…

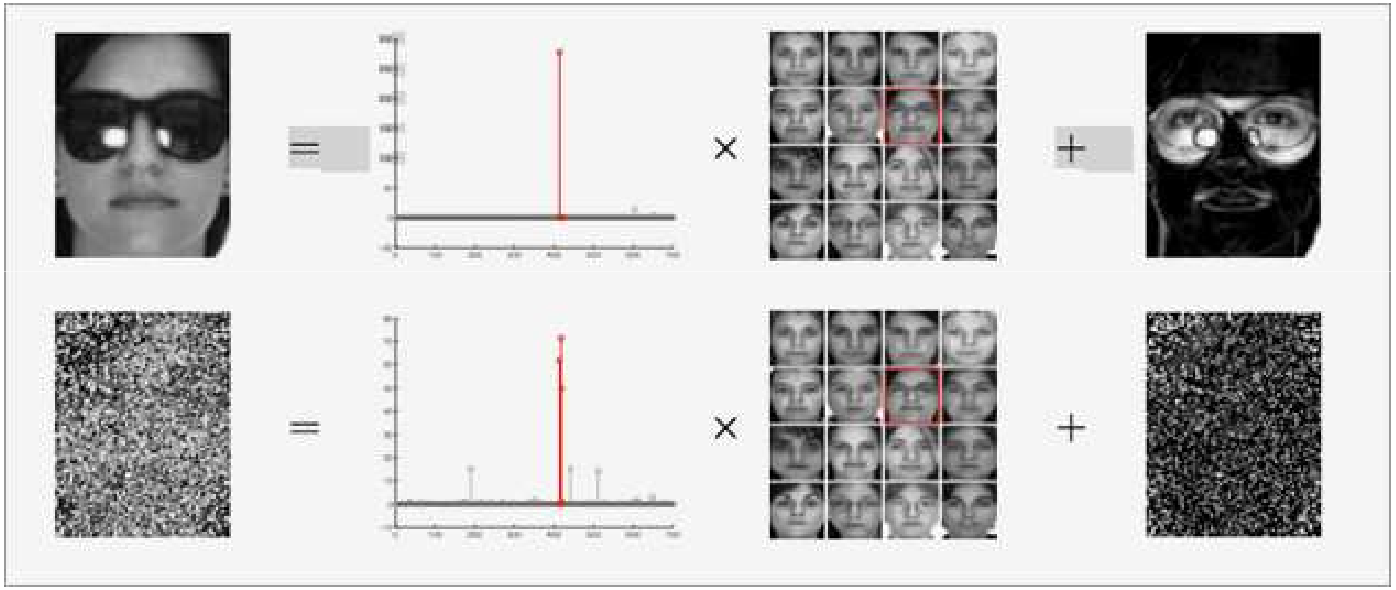
Decoding (L1-minimization)

$$\min \; ||x||_{\ell_1} \quad \text{s.t. } Ax = y$$

When we use (1.1) for the recovery problem illustrated in Figure 1 (with the popular Logan-Shepp phantom as a test image), the results are surprising. The reconstruction is *exact*; that is, $f^\sharp = f$! This numerical result is also not special to this phantom. In fact, we performed a series of experiments of this type and obtained perfect reconstruction on many similar test phantoms.

[Candes, Romberg, and Tao: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.
IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, 2006]

応用例：顔識別



[Wright, Yang, Ganesh: Robust Face Recognition via Sparse Representation. IEEE TRANS. PAMI, vol. 31, no. 2, February 2009]

余談

一方、カリフォルニア大学バークレー校の工学部学部長を務めるShankar Sastry氏が指摘するように、Yang氏の新しい顔認識アプローチは、この分野における何年もの研究を用なしにしてしまうものでもある。
「研究者たちは本当に困惑している。あんまりではないか。顔のどの部分の特徴を選ぼうが構わない？ これは長年の研究と真っ向から対立するものだ」
と、Sastry氏は述べた。

[Wired Vision, WIRED NEWS]

必要なサンプル数

d: 真の非ゼロ成分の数
M: featureの次元
n:サンプル数（ランダムサンプリング）

サンプル数が

$$n \geq Cd\log(M/\delta)$$

を満たすなら，確率 $1-\delta$ で
真が正確に再現される $(\hat{x} = x_0)$.

$$\hat{x} = \arg\min_{x\in\mathbb{R}^M}\{\|x\|_{\ell_1} \mid Ax = y\ (= Ax_0)\}$$

A: ランダムサンプリング
n×M

本日のメイン

Johnson–Lindenstrauss Lemma

- **W. Johnson and J. Lindenstrauss**: Extensions of Lipschitz mappings into a Hilbert space. Contemporary Mathematics, 26:189--206, 1984.
- **S. Dasgupta and A. Gupta**: An Elementary Proof of a Theorem of Johnson and Lindenstrauss. Random Structures and Algorithms, 22(1):60--65, 2003.

Theorem 1 (Dasgupta, Gupta, '03). *Let $X \subset \mathbb{R}^M$ be a finite collection of k points, fix both $\delta \in (0,1)$ and $\alpha \in (0,1)$, and let Ψ be a random projection $\Psi : \mathbb{R}^M \to \mathbb{R}^n$ with*

$$n \geq \frac{4+2\alpha}{\delta^2/2 - \delta^3/3} \log k.$$

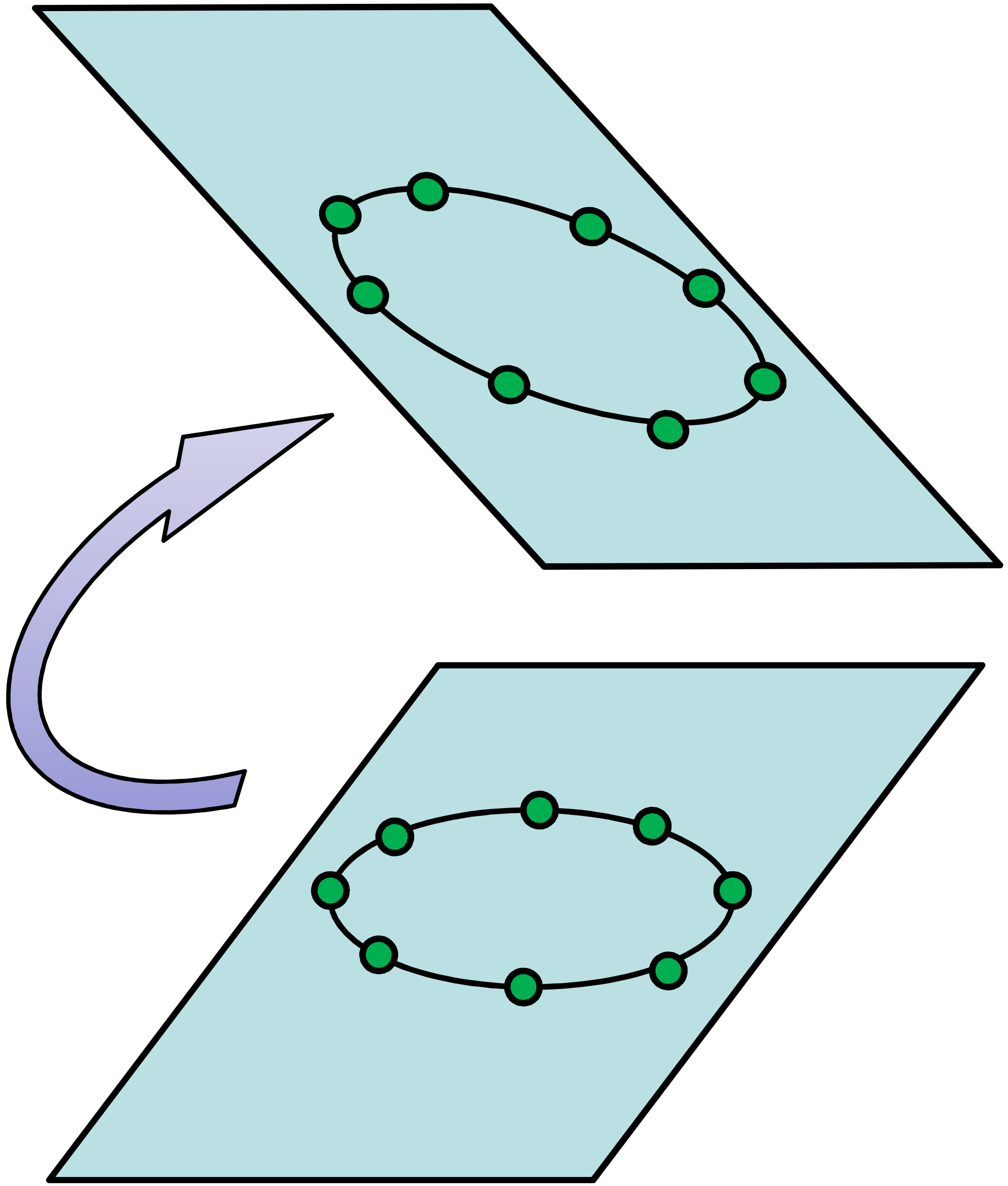
$$n \gtrsim \frac{1}{\delta^2} \log(k)$$

If $n \leq M$, then the following holds with probability exceeding $1 - k^{-\alpha}$:

$$(1-\delta) \leq \frac{\|\Psi x - \Psi y\|_2^2}{\|x - y\|_2^2} \leq (1+\delta)$$

for all $x, y \in X$

[http://www-users.math.umd.edu/~nstrawn/oralprelim.pdf]



J–L Lemma for CS

- **R. Baraniuk, M. Davenport, R. DeVore, M. Wakin**: A Simple Proof of the Restricted Isometry Property for Random Matrices. Constructive Approximation, 28(3), pp. 253-263, December 2008.

$$k \sim \delta^d \binom{M}{d}$$ のようにおく.

d次元空間に詰め込めるδボールの数 × d次元空間の数

Theorem 2 (Baraniuk, Davenport, DeVore, Wakin, '08). *Let Ψ be a random projection $\Psi : \mathbb{R}^M \to \mathbb{R}^n$ with*

$$n \geq \frac{c}{\delta^2} d \log(M/d).$$

$$n \gtrsim d \log(M) \quad !!$$

We have with probability $1 - e^{-c\delta n}$,

$$(1-\delta) \leq \frac{\|\Psi x - \Psi y\|_2^2}{\|x-y\|_2^2} \leq (1+\delta)$$

$$RI(\delta, d)$$

for any set T with $|T| = d$ and for all $x, y \in \mathbb{R}^T$.

Back to CS

$$RI(\delta, d) \ : \quad (1-\delta) \leq \frac{\|\Psi x\|_2^2}{\|x\|_2^2} \leq (1+\delta) \ \text{ for all } d\text{-sparse } x.$$

Ristricted Isometry

[Emmanuel Candès: The restricted isometry property and its implications for compressed sensing. Compte Rendus de l'Academie des Sciences, Paris, Series I, 346, pp. 589-592, 2008]

Theorem 3.

If A satisfies $RI(\delta_{2d}, 2d)$ with $\delta_{2d} < \sqrt{2} - 1$, then CS recovers the truth exactly:

$$\hat{x} = x_0 \, ,$$

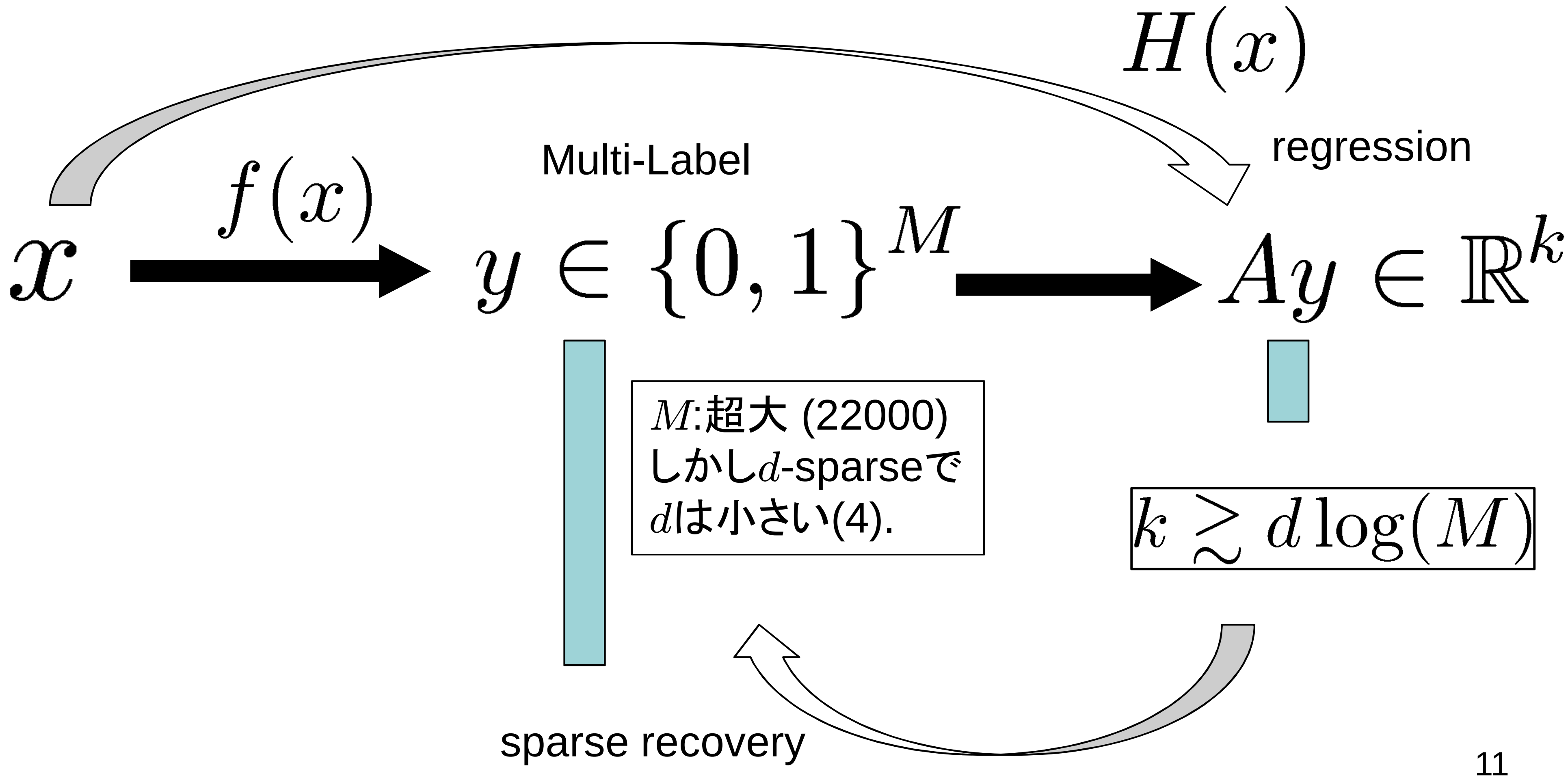
for all d-sparse x_0.

$$n \gtrsim d \log(M)$$

$$\hat{x} = \arg\min_{x \in \mathbb{R}^M} \{ \|x\|_{\ell_1} \mid Ax = y \ (= Ax_0) \}$$

応用例2

- D. Hsu, S. Kakade, J. Langford & T. Zhang: Multi-Label Prediction via Compressed Sensing. NIPS2009.



Lasso
&
Dantzig Selector
(雑音あり)

Condition for Lasso Analysis

$$\min_{\beta \in \mathbb{R}^M} \frac{1}{n}\|X\beta - Y\|_2^2 + C\|\beta\|_{\ell_1}$$

$$A = \frac{X^\top X}{n}$$

[Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009]

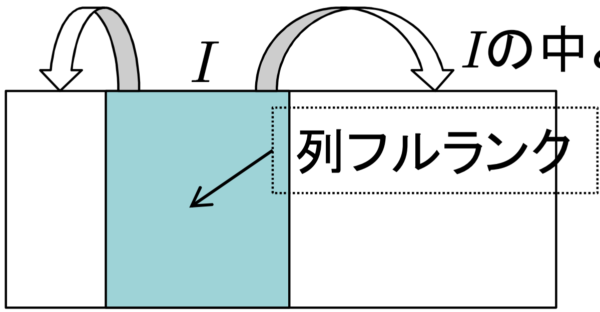
$RE(d, c_1)$: Restricted eigenvalue condition

$$\min_{\substack{I, v \in \mathbb{R}^M : |I| \le 2d \\ c_1 \|v_I\|_{\ell_1} \ge \|v_I\|_2^2}} \frac{v^\top A v}{\|v_I\|_2^2} > 0$$

十分条件

$\exists m \ge d \quad RI(\delta_m, m),\ RI(\delta_{d+m}, d+m)$

かつ $\frac{1-\delta_{d+m}}{1+\delta_m} > \frac{c_1^2 d}{m}$



Iの中と外で相関低い

Convergence Rate of Sparse Learning

– Candes & Tao: AS2007 (Dantzig selector)
– Bunea, Tsybakov & Wegkamp: AS2007 (Lasso)
– Meinshausen & Yu: AS2009 (Lasso)
– Bickel, Ritov & Tsybakov: AS2009 (Dantzig&Lasso)

$$\|\hat{\beta} - \beta^*\|_{\ell_2}^2 = O_p\left(\frac{d\log(M)}{n}\right)$$

– Raskutti, Wainwright & Yu: arXiv:0910.2042, 2009.

minimax rate $$\min_{\hat{\beta}} \max_{\beta^*} \mathrm{E}\|\hat{\beta} - \beta^*\|_{\ell_2}^2 \geq C\left(\frac{d\log(M/d)}{n}\right)$$

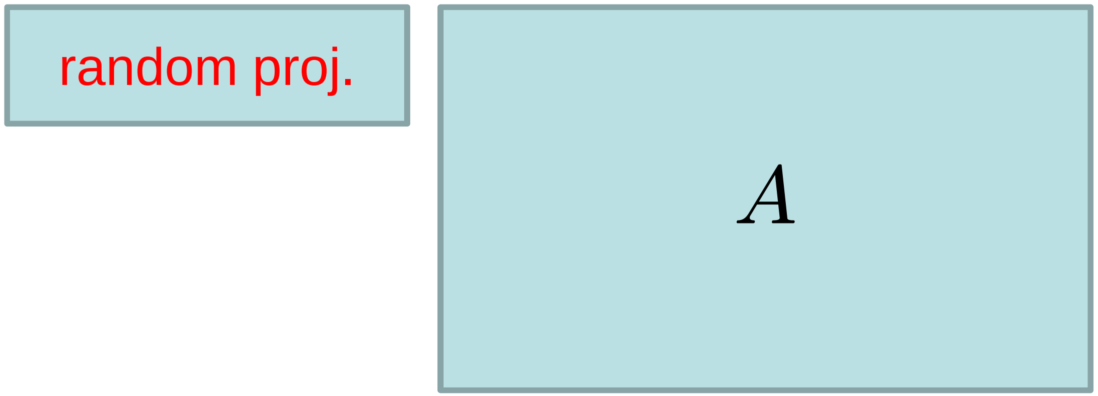
関連研究

- Gunnar Martinsson:

  “Randomization: Making Very Large-Scale Linear Algebraic Computations Possible”

  NIPS 2009 turorial

→ 大きな行列をrandom projectionで小さな行列に落としてSVD

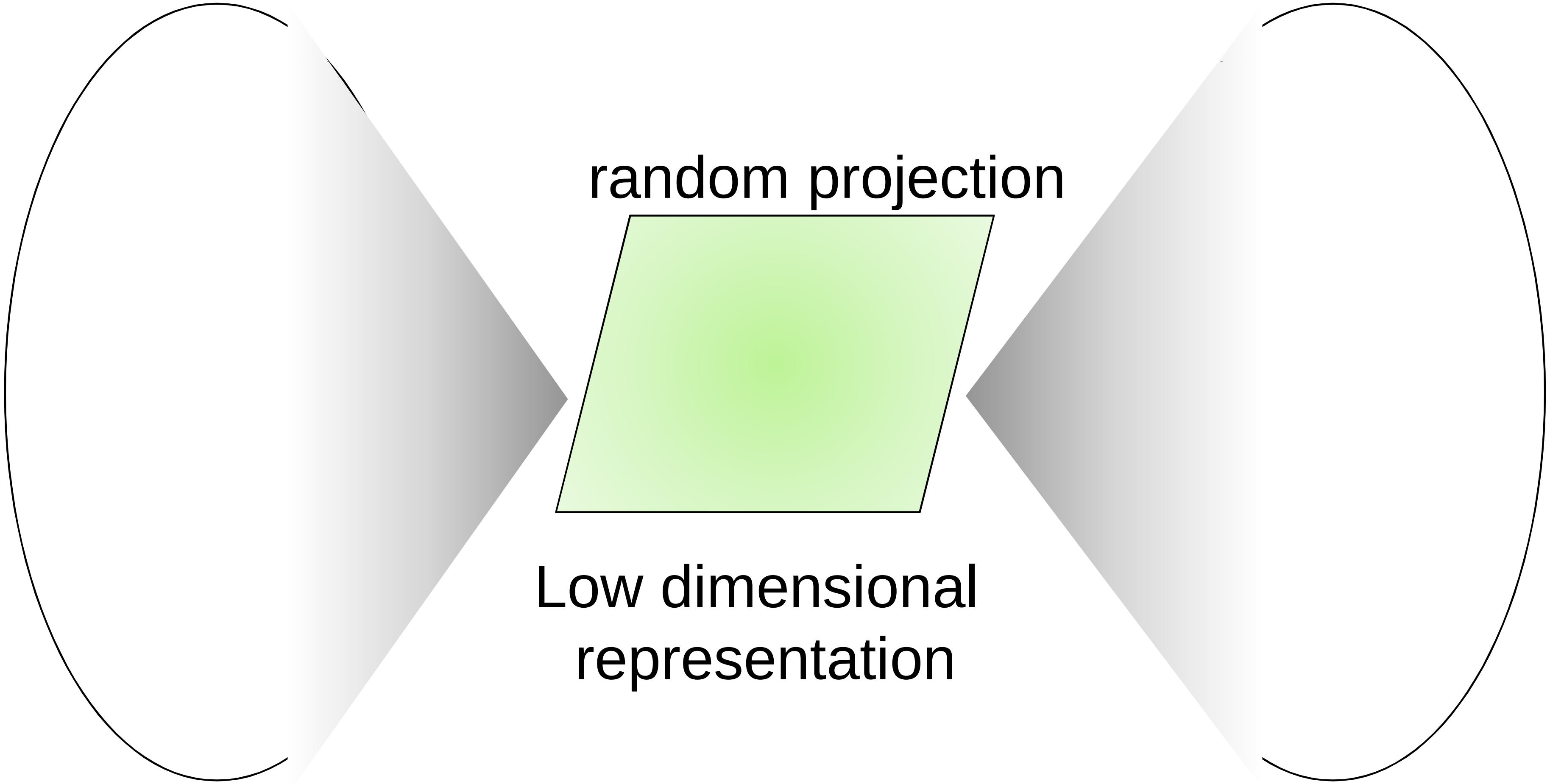


Multiple Input

- high dimension
- huge data

Multiple Output

- multi-label
- various tasks



今後の展開？

- 圧縮　⇔　Hashing
- 座標変換への不変性
- Non-sparse

- 深い理論と広い応用？