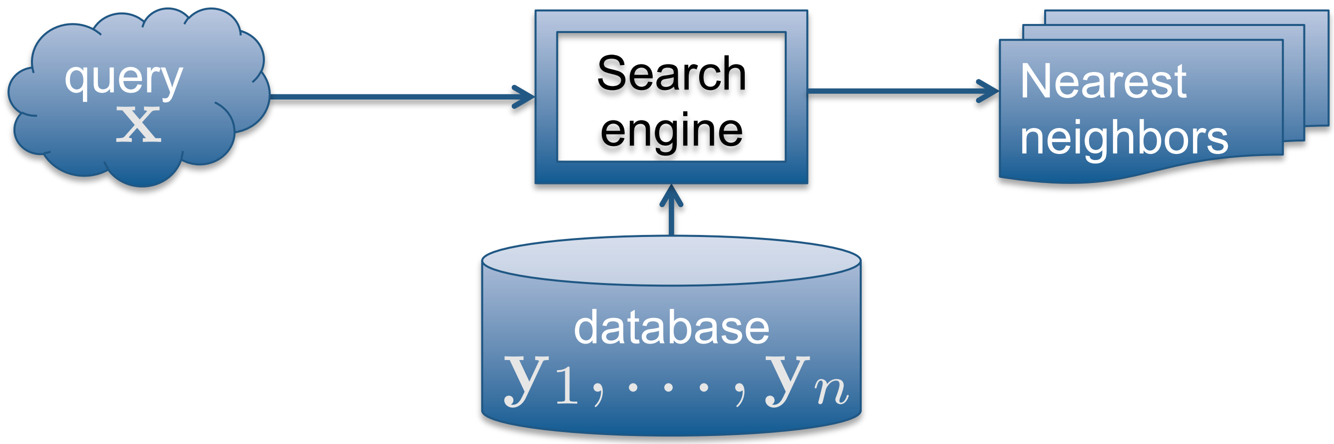
BEYOND "PROJECT AND SIGN" FOR COSINE ESTIMATION WITH BINARY CODES

Raghavendran Balu, Teddy Furon and Hervé Jégou
INRIA, Rennes

Problem statement: Nearest Neighbors search

– Finding the closest vector(s) from a database for a given query



– In this paper:

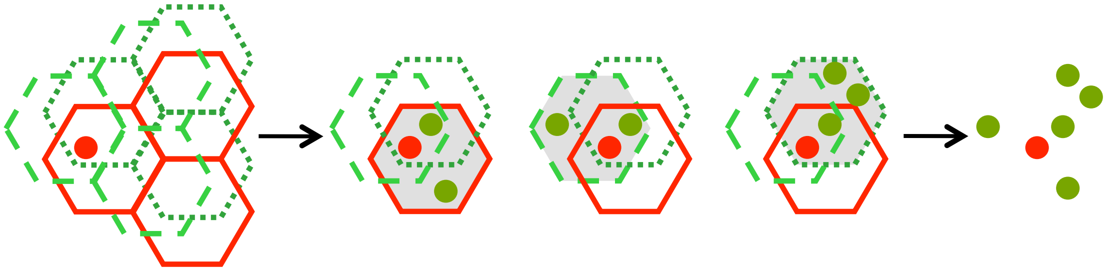
$$\mathbf{y}_i \in \mathbb{R}^D, \quad \|\mathbf{y}_i\| = 1$$

$$NN(\mathbf{x}) = \arg\min_{1 \leq i \leq n} \|\mathbf{x} - \mathbf{y}_i\| = \arg\max_{1 \leq i \leq n} \mathbf{x}^\top \mathbf{y}_i$$

Problem: Exhaustive search has complexity $O(nD)$

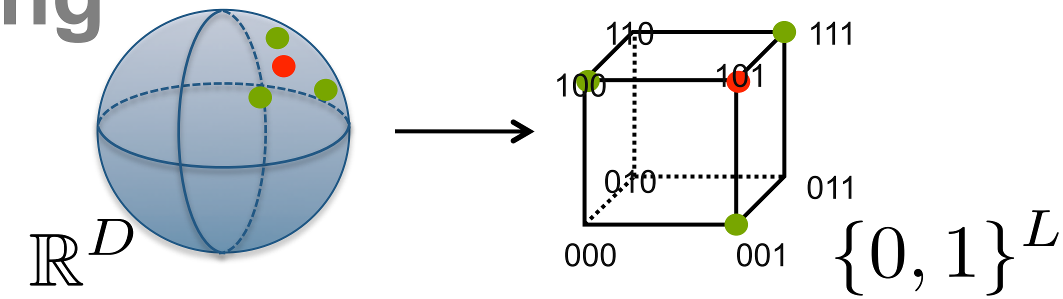
2 approaches to Nearest Neighbor Search

- Space partitioning
  - The search no longer exhaustive
  - Example: indexing technique involving several hash functions



- Approximate distance
  - Faster to compute but exhaustive
  - In this paper: we use an Hamming Embedding

Hamming embedding



- Design a mapping function $\mathbf{b} : \mathbb{R}^D \rightarrow \{0,1\}^L$

$$\mathbf{x} \rightarrow \mathbf{b}(\mathbf{x})$$

- Objective
  - neighborhood in Hamming space reflects true neighborhood

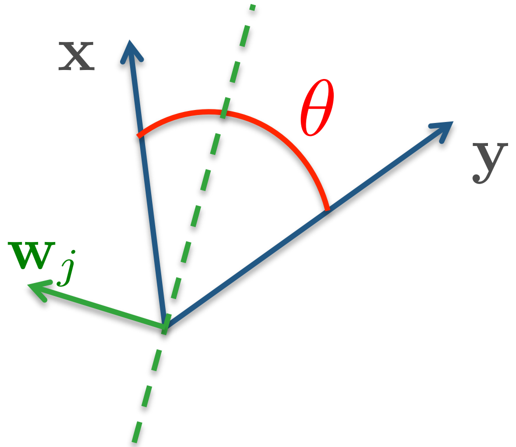
$$NN(\mathbf{x}) \approx \arg \min_{1 \leq i \leq n} d_H(\mathbf{b}(\mathbf{x}), \mathbf{b}(\mathbf{y}_i))$$

- Advantages
  - compact descriptor
  - fast distance computation

Locality Sensitive Hashing (LSH)

- Initialization: Randomly draw *L* directions $\{\mathbf{w}_j\}_{1 \le j \le L}$
- For a given vector $\mathbf{x}$, compute a bit for each direction, as
  1. Project $\quad p_j = \mathbf{x}^\top \mathbf{w}_j$
  2. And sign $\quad b_j(\mathbf{x}) = \mathrm{sign}(p_j)$

  $$\mathbf{b}(\mathbf{x}) = (b_1(\mathbf{x}), \ldots, b_L(\mathbf{x}))$$



- Properties
  - For two vectors $\mathbf{x}$ and $\mathbf{y}$

    $$\mathbb{P}(b_j(\mathbf{y}) \neq b_j(\mathbf{x})) = \theta / \pi$$

  - The Hamming distance is related *in expectation* to the angle as

    $$\theta = \pi \mathbb{E}(d_H(\mathbf{b}(\mathbf{y}), \mathbf{b}(\mathbf{x}))) / L$$

[Charikar 02]

Our approach

- Synthesis point of view
  - Reconstructed vector $\mathbf{c}(\mathbf{x}) = \frac{\mathbf{W}\mathbf{b}(\mathbf{x})}{\|\mathbf{W}\mathbf{b}(\mathbf{x})\|}$
  - If $\mathbf{c}(\mathbf{x})$ 'close' to $\mathbf{x}, \forall \mathbf{x}$ on the sphere, then

$$\cos(\mathbf{c}(\mathbf{y}), \mathbf{c}(\mathbf{x})) \approx \cos(\mathbf{y}, \mathbf{x})$$

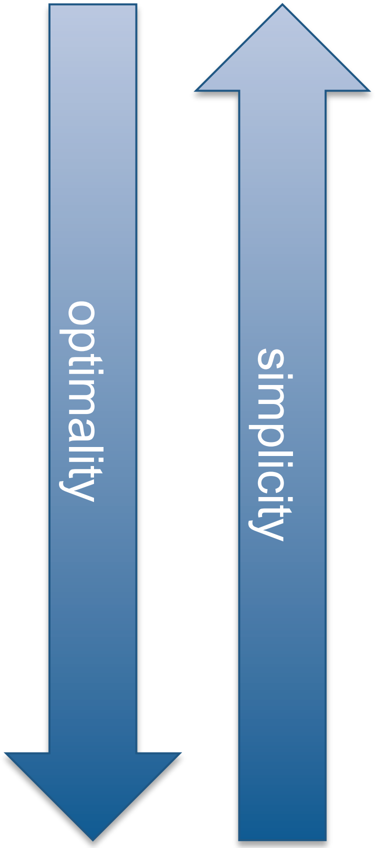
- Minimizing the quantization error $\|\mathbf{c}(\mathbf{x}) - \mathbf{x}\|$
  - If $L < D$ and $\mathbf{W}^\top \mathbf{W} \propto \mathbf{I}$, 'project and sign' is optimal
  - If $L > D$, it is a combinatorial problem
    - Not tractable for large D

Reconstruction point of view

- ‘Project and sign’ with a frame **W**
- ‘Project and sign’ with a tight frame **W**

- **Our algorithm qoLSH**
  - **quantization optimized LSH**

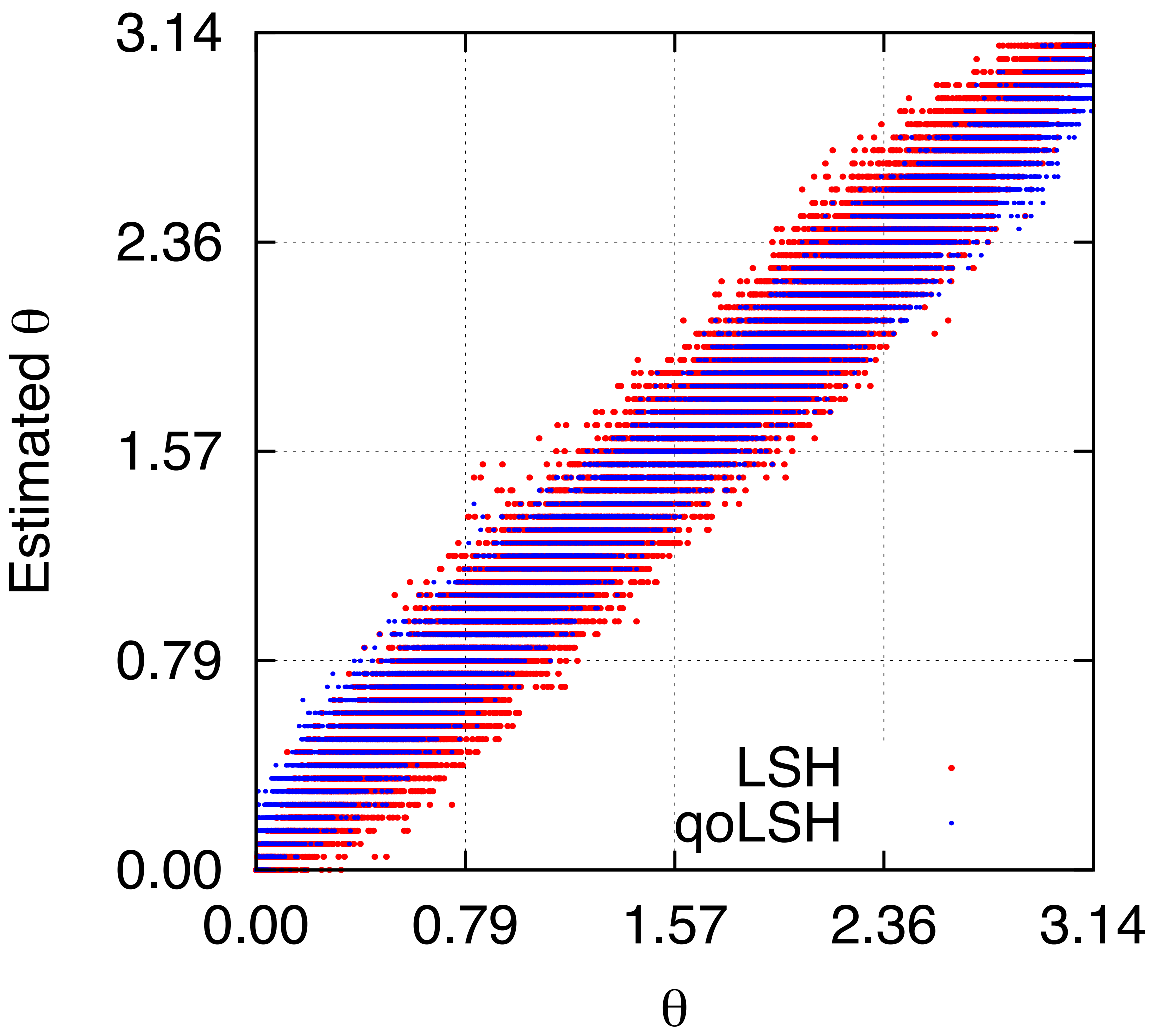
- ‘AntiSparse’ [Jégou 11]
  - Too slow for large *D*
- Optimal
  - Untractable for large *D*



qoLSH algorithm

- Parameter: randomly draw a tight frame $\mathbf{W}$
- Initialization: input $\mathbf{x}$
  - 'project and sign': $\mathbf{b}_0(\mathbf{x}), \quad \mathbf{c}_0(\mathbf{x})$
- Iteration *k* + 1
  - For any *j*
    - Flip *j*-th bit: $\mathbf{c}^{(j)} = \mathbf{c}_k(\mathbf{x}) - 2b_{k,j}\mathbf{w}_j$
    - Measure cosine: $L_j = \mathbf{x}^\top \mathbf{c}^{(j)} / \|\mathbf{c}^{(j)}\|$
  - Keep best flip $j^\star = \arg\max_j L_j$
    - $\mathbf{b}_{k+1}(\mathbf{x}) = \mathbf{b}_k(\mathbf{x}), \quad b_{k+1,j^\star}(\mathbf{x}) = \overline{b_{k,j^\star}(\mathbf{x})}$

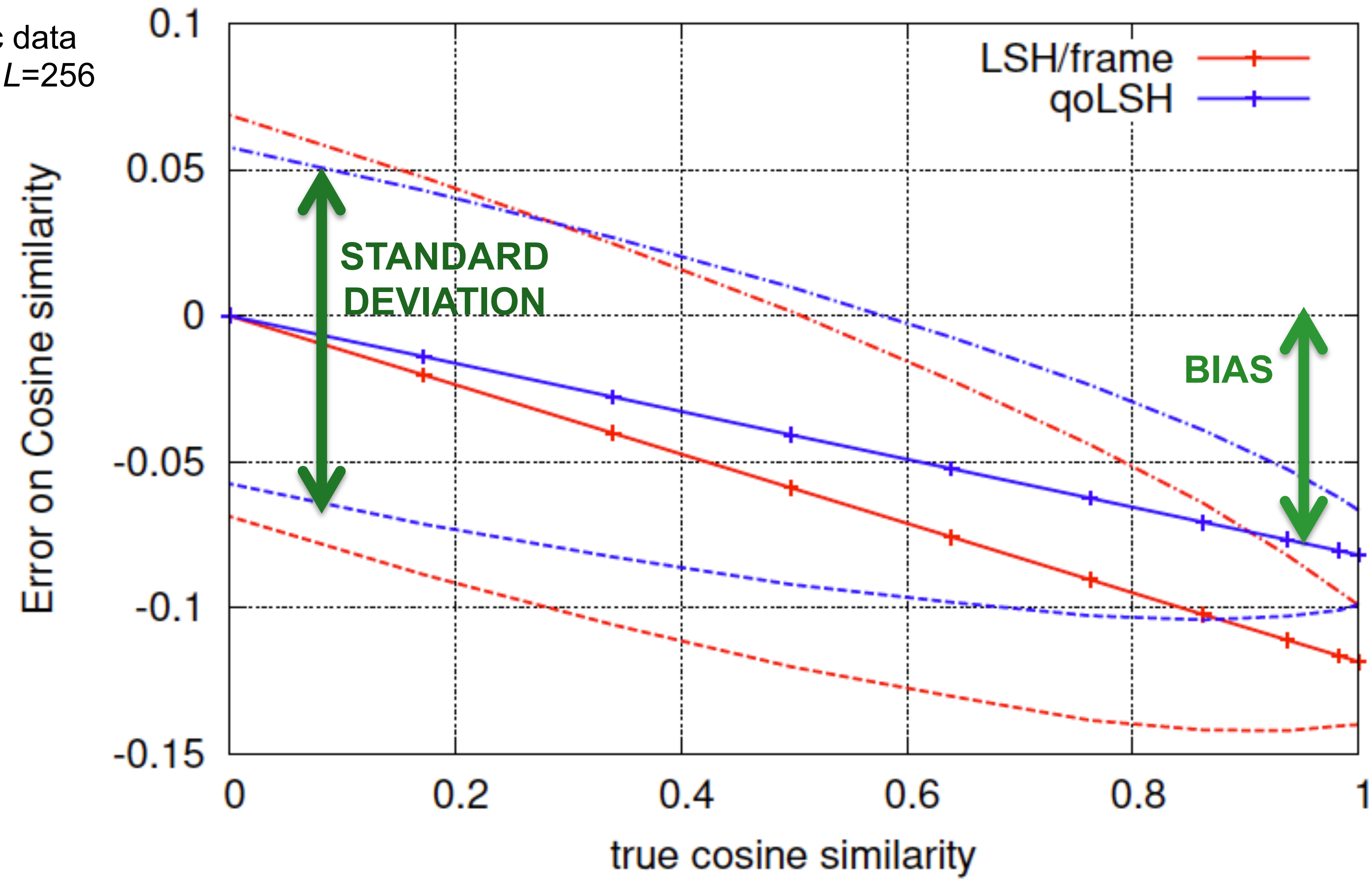
Estimated angle vs True angle



Synthetic data
D = 8, *L*=64

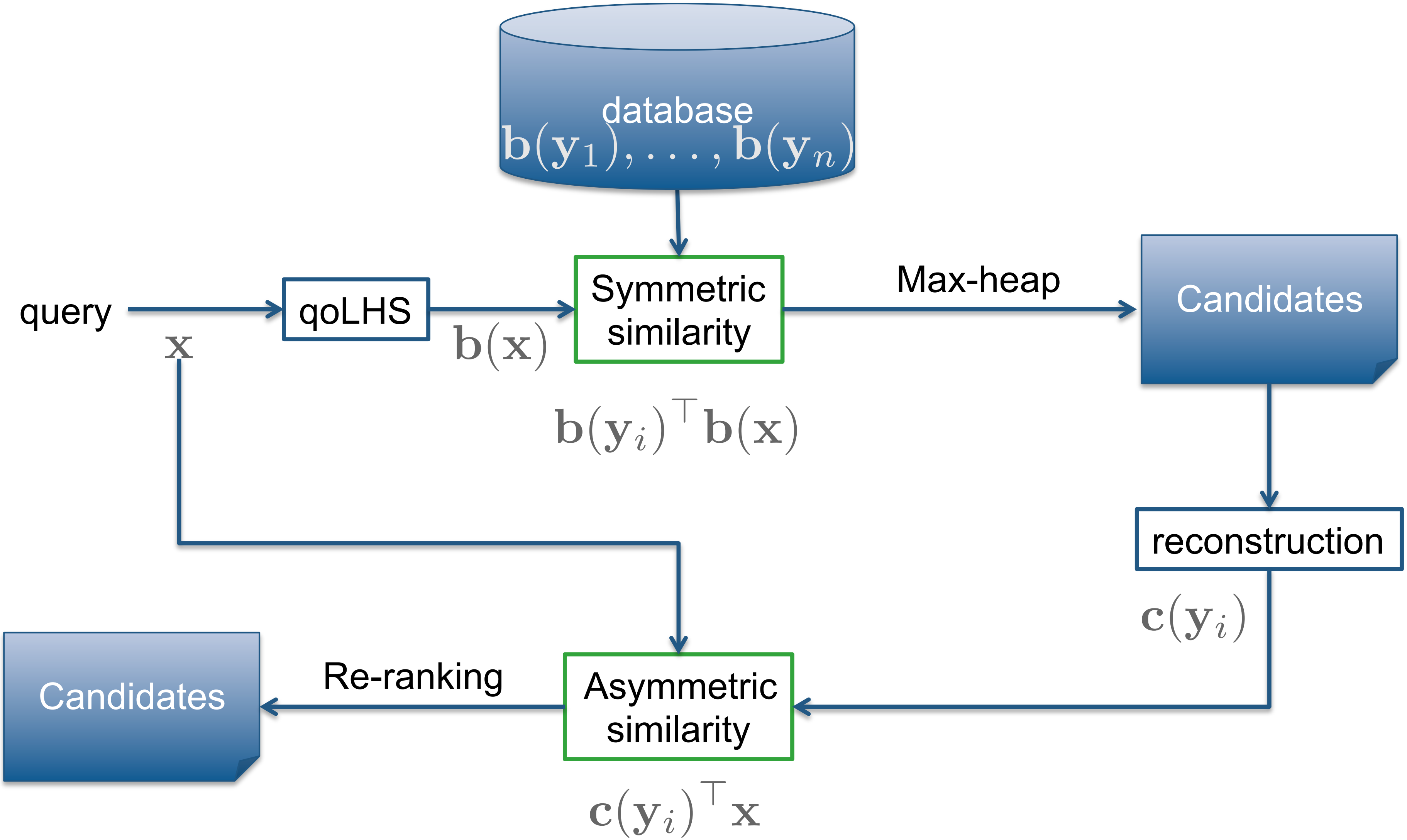
Angle estimation error analysis

Synthetic data
D = 128, L=256



qoLSH reduces estimation bias and variance compared to LSH

Application the Nearest Neighbor Search



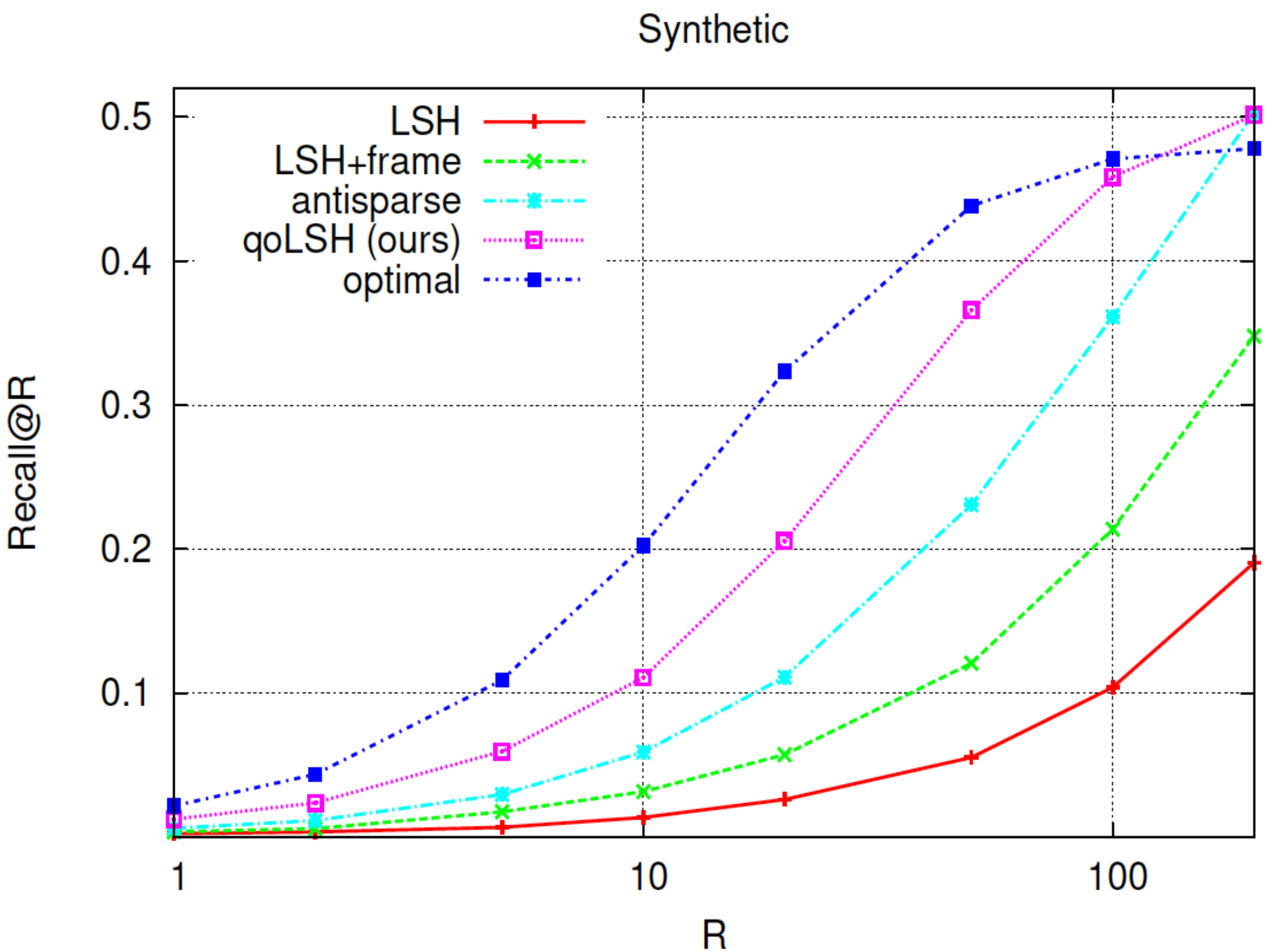
Experimental details

- Dataset
  - Synthetic (n = 1 million, D = 8)
  - SIFT (n = 1 million, D = 128)
    - http://corpus-texmex.irisa.fr

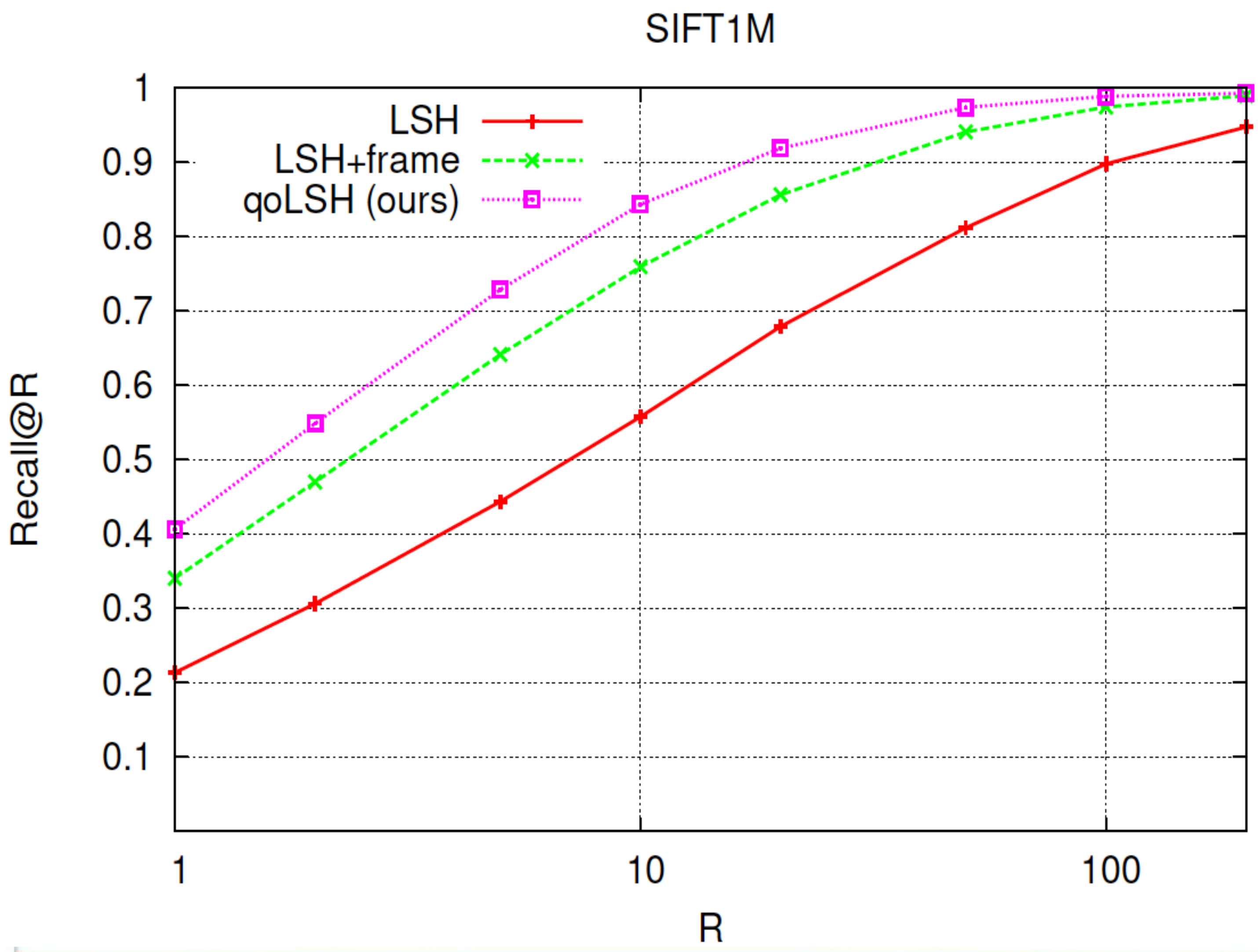
- Algorithms
  - LSH with or without tight frame
  - qoLSH
  - anti-sparse
  - quantization optimal (if tractable)

- Performance measurement
  - 1-Recall@R: probability that the true nearest neighbor belongs to a short list of R candidates

Recall on synthetic data (n = 1M, D = 8)



Recall on real SIFT data (n = 1M, D = 128)



Conclusion

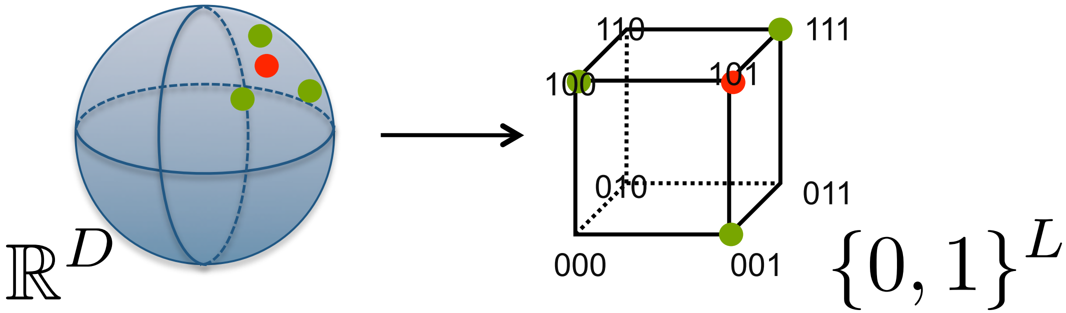
- Hamming embedding dedicated for cosine similarity estimation
- $L<D$
  - ‘Project and sign’ is optimal with orthogonal random projection
- $L>D$
  - Tight frame is a good choice
  - ‘Project and sign’ is suboptimal
  - Our reconstruction based approach
    - decreases quantization error
    - improves cosine similarity estimation
    - improves quality of approximate NN search
    - strikes a good trade-off between quality and complexity

Package Online!

http://people.rennes.inria.fr/Raghavendran.Balu/code/qolsh.zip

Thank You!

QUESTIONS?


$\mathbb{R}^D$
110
111
100
101
010
011
000
001
$\{0,1\}^L$

LSH suboptimality when *L* > *D*

- When *L*>*D*, $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_L)$ is not orthogonal
  - Entropy *H*(**B**) < *L* bits
- Example

$$\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \mathbf{w}_3] = \begin{bmatrix} 1 & 0 & \cos\frac{\pi}{3} \\ 0 & 1 & \sin\frac{\pi}{3} \end{bmatrix}$$

$$\mathbf{x} \propto \mathbf{w}_1 + \mathbf{w}_2 - \mathbf{w}_3$$

- LSH (sub optimal):

$$\mathbf{b}(\mathbf{x}) = [1, 1, 1]$$

- Optimal

$$\mathbf{c}^\star(\mathbf{x}) \propto \mathbf{W}.[1, 1, -1]^\top = \mathbf{x}$$