On Projection Based Operators in l_p space for Exact Similarity Search

Andreas Wichert

andreas.wichert@ist.utl.pt

Catarina Moreira

catarina.p.moreira@ist.utl.pt

Instituto Superior Técnico, INESC-ID

Av. Professor Cavaco Silva, 2744-016 Porto Salvo, Portugal

Abstract

We investigate exact indexing for high dimensional l_p norms based on the 1-Lipschitz property and projection operators. The orthogonal projection that satisfies the 1-Lipschitz property for the l_p norm is described. The adaptive projection defined by the first principal component is introduced.

Keywords: 1-Lipschitz property; curse of dimensionality; nearest neighbour; high dimensional indexing; l_p normCognition

1 Introduction

While there are relatively efficient approximate similarity search algorithms, it is widely supposed that the exact search suffers from dimensionality [11]. Thus, solving the problem in the most general case for an arbitrary dataset is impossible. We investigate exact indexing for a vector space V and a distance function d. Exact indexing is based on exact similarity search, and no data points are lost during range queries. For a range query vector $\mathbf{y}$ from a collection of s vectors,

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \cdots, \mathbf{x}_s$$

all vectors $\mathbf{x}_i$ that are ε-similar according to the distance function d are searched

$$d(\mathbf{x}_i, \mathbf{y}) < \varepsilon. \tag{1}$$

1

In approximate indexing, the data points that may be lost as some distances are distorted. Approximate indexing [8], [7] seems to be in some sense free from the curse of dimensionality, [11]. Distance-based exact indexing is based on the 1-Lipschitz property [11]. A mapping function $F()$ maps two vectors $\mathbf{x}$ and $\mathbf{y}$ into a lower dimensional space, where d is a metric in the original space and $d_{feature}$ is a metric in the feature space that satisfies the 1-Lipschitz property

$$d_{feature}(F(\mathbf{x}), F(\mathbf{y})) \leq d(\mathbf{x}, \mathbf{y}). \tag{2}$$

This equation is also known as the lower bounding postulate [5], [4]. Using the 1-Lipschitz property, a bound that is valid in both spaces can be determined. The distance from similar vectors to a query vector $\mathbf{y}$ is smaller or equal in the original space and, consequently, is smaller or equal in the lower dimensional space as well. During the computation, all the points below the bound are discarded. In the second step, the wrong candidates are filtered by comparisons in the original space. The application of the 1-Lipschitz property as used in metric trees and pivot tables does not resolve the curse of dimensionality, as shown in [10]. For high-dimensional spaces, the functions that obey the 1-Lipschitz property discard fewer points as the number of dimensions grows [10]. The number of points discarded drops as fast as the number of dimensions grows. As stated in [11], every 1-Lipschitz function concentrates sharply near its mean (or median) value, which results from the fact that a sphere with a constant radius increases exponentially with growing dimensions. A linear radial increase leads to an exponential increase of points inside the sphere [2], [11], which leads to a degradation of the method's performance. This situation leads to the "curse of dimensionality", which states that for an exact nearest neighbor, any algorithm for high dimension d and n objects must either use an n^d-dimension space or have a query time of $n \times d$ [2], [11].

However, [12], [13], and [14] show how the recursive application of the 1-Lipschitz property can be used to overcome the curse of dimensionality for certain cases of points equally distributed by subspace trees. A high-dimensional space is divided into low-dimensional sub-spaces [12], [13]. In the low-dimensional sub-spaces, 1-Lipschitz functions can be successfully applied. The main contributions of this paper are as follows:

- Introduction of a new adaptive projection. The optimal projection is not fixed by orthogonal projection but learned.

- Extension of the technique beyond the Euclidean norm (l_2). Many applications rely on the l_1

norm. It is shown that l_1 norm gives better results than l_2 norm.

- Simplification of the mathematical framework.

The paper is organized as follows:

- We review the projection operators.

- We introduce the adaptive projection and the l_p norm dependency.

- The adaptive projection and the l_p norm dependency are integrated into the subspace tree.

- We empirically compare the adaptive projection with the orthogonal mapping. We empirically compare the l_1, l_2, l_4 and l_∞ norms.

2   Projection Operators

Ideally, the mapping function $F()$ should preserve the exact distances [5], [4]. An example of such a function for real vectors is a norm preserving linear operator Q. Such an operator can be represented by an orthogonal matrix with $Q^T = Q^{-1}$ performing a rotation or a reflection. An example of such an operator is the Karhunen-Loève transform, which rotates the coordinate system in such a way that the new covariance matrix will be diagonal, resulting in each dimension being uncorrelated. A mapping that reduces the dimensionality of a vector space can be represented by a projection operator in a Hilbert space, which extends the two or three dimensional Euclidean space to spaces with any finite or infinite number of dimensions. In such a space, the Euclidean norm is induced by the inner product

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}|\mathbf{x}\rangle}. \tag{3}$$

If W is a subspace of V, then the orthogonal complement of W is also a subspace of V. The orthogonal complement $W^\perp$ is the set of vectors

$$W^\perp = \{\mathbf{y} \in V | \langle \mathbf{y}|\mathbf{x}\rangle = 0 \ \mathbf{x} \in V\} \tag{4}$$

and

$$V = W \oplus W^\perp. \tag{5}$$

Each vector $\mathbf{x} \in V$ can be represented as $\mathbf{x} = \mathbf{x}_W + \mathbf{x}_{W^\perp}$ with $\mathbf{x}_W \in W$ and $\mathbf{x}_{W^\perp} \in W^\perp$. The mapping $P \cdot \mathbf{x} = \mathbf{x}_W$ is an orthogonal projection. Such a projection is always a linear transformation and can be represented by a projection matrix P. The matrix is self-adjoint with $P = P^2$. An orthogonal projection can never increase a norm

$$\|P \cdot \mathbf{x}\|^2 = \|\mathbf{x}_W\|^2 \leq \|\mathbf{x}_W\|^2 + \|\mathbf{x}_{W^\perp}\|^2 = \|\mathbf{x}_W + \mathbf{x}_{W^\perp}\|^2 = \|\mathbf{x}\|^2. \tag{6}$$

Using the triangle inequality

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{7}$$

setting

$$\|\mathbf{x}\| = \|\mathbf{y} + (\mathbf{x} - \mathbf{y})\| \leq \|\mathbf{y}\| + \|\mathbf{x} - \mathbf{y}\| \tag{8}$$

the tighten triangle inequality

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|. \tag{9}$$

follows. From the fact that the orthogonal projection can never increase the norm and the tightened triangle inequality, any orthogonal projection operator has the 1-Lipschitz property

$$\|P \cdot \mathbf{x}\| - \|P \cdot \mathbf{y}\| \leq \|P \cdot \mathbf{x} - P \cdot \mathbf{y}\| = \|P \cdot (\mathbf{x} - \mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|. \tag{10}$$

It follows that any projection satisfies the 1-Lipschitz property, which means that the lower bounding postulate [5],[4] and any orthogonal projection are satisfied. For example, the "Quadratic Distance Bounding" theorem is satisfied [5]. There is no the need for a more complicated proof based upon the unconstrained minimization problem using Lagrange multipliers [5].

2.1 Projection onto one-dimensional subspace

For $\|\mathbf{p}\| = 1$, $\mathbf{p} \cdot \mathbf{p}^\top$ is an orthogonal projection onto a one-dimensional space generated by $\mathbf{p}$. For example for the vector

$$\mathbf{p} = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}} \right) \tag{11}$$

the orthogonal projection from R^n onto one-dimensional space R is

$$P = \mathbf{p} \cdot \mathbf{p}^\top = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}. \tag{12}$$

For $f \leq n$ orthogonal subspaces

$$R^n = E_1 \oplus E_2 \oplus \ldots \oplus E_f \tag{13}$$

of the vector space R^n we can define a projection $P : R^n \mapsto R^f$ as a sum of f projections onto one dimensional space

$$P = \mathbf{p}_1 \cdot \mathbf{p}_1^\top + \mathbf{p}_2 \cdot \mathbf{p}_2^\top \ldots + \mathbf{p}_f \cdot \mathbf{p}_f^\top \tag{14}$$

with $\mathbf{p}_i \cdot \mathbf{p}_i^\top : E_i \mapsto R$ and $\|\mathbf{p}_i\| = 1$. The 1-Lipschitz property of the projection from the subspace E_i the one dimensional space R is

$$\left| \|\mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{x}\| - \|\mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{y}\| \right| \leq \left| \|\mathbf{x}\| - \|\mathbf{y}\| \right| \leq \|\mathbf{x} - \mathbf{y}\|. \tag{15}$$

The projection P, represented by Equation 14, should distort the distances between the vector space R^n and R^f as little as possible. As a consequence, the distortion for each subspace E_i should be minimized. Because of the 1-Lipschitz property for the one-dimensional space, according to the Equation 15, we need to minimize the distance in the one-dimensional space between the length of the vector and the length of its projected counterpart

$$\left| \|\mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{x}\| - \|\mathbf{x}\| \right|. \tag{16}$$

Suppose the dimensionality of the subspace E_i is m. We define the vector $\mathbf{a}$ as

$$\mathbf{a} = \mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{x}. \tag{17}$$

It follows that

$$a = \sqrt{m} \cdot \alpha = \|\mathbf{a}\| = \|\mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{x}\| \tag{18}$$

5

and with

$$a_1 = a_2 = ... = a_k = ... = a_m = \alpha$$

$$\mathbf{a} = (a_1, a_2, ..a_k, .., a_m). \tag{19}$$

With a being the length of the projected vector we preform the following operation

$$\min\{|a - \|\mathbf{x}\|\|\}. \tag{20}$$

From the tighten triangle inequality, it follows that

$$\min\{a - \|\mathbf{x}\|\} \leq \min\{\|\mathbf{a} - \mathbf{x}\|\} \tag{21}$$

according to the Euclidean distance function. To minimize the Euclidean metric $\|\mathbf{a} - \mathbf{x}\|$, how do we choose the value of α [14]? It follows that

$$\min_{\alpha} \left(\sqrt{(x_1 - \alpha)^2 + (x_2 - \alpha)^2 + ... + (x_m - \alpha)^2} \right) \tag{22}$$

$$0 = \frac{\partial d(\vec{x}, \vec{a})}{\partial \alpha} = \frac{m \cdot \alpha - (\sum_{i=1}^{m} x_i)}{\sqrt{m \cdot \alpha^2 + \sum_{i=1}^{m} x_i^2 - 2 \cdot \alpha \cdot (\sum_{i=1}^{m} x_i)}} \tag{23}$$

with the solution

$$\alpha = \frac{\sum_{i=1}^{m} x_i}{m} \tag{24}$$

which is the mean value of the vector $\mathbf{x}$. It follows

$$a = \sqrt{m} \cdot \alpha = \sqrt{m} \cdot \frac{\sum_{i=1}^{m} x_i}{m} = \|\mathbf{p}_i \cdot \mathbf{p}_i^\top \cdot \mathbf{x}\| \tag{25}$$

with the corresponding projection matrix P_i

$$P_i = \mathbf{p}_i \cdot \mathbf{p}_i^\top = \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \end{pmatrix}. \tag{26}$$

P_i is generated by the normalised vector $\mathbf{p}_i$

$$\mathbf{p}_i = \left(\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}, \cdots, \frac{1}{\sqrt{m}} \right). \tag{27}$$

which indicates the direction of the m-secting line, which is a continuous map from a one-dimensional space to an m-dimensional space given by

$$\begin{aligned} x_1 &= x_1 \\ x_2 &= x_1 \\ x_3 &= x_1 \quad \cdot \\ &\vdots \\ x_m &= x_1 \end{aligned} \tag{28}$$

For $m = 2$, this equation is the bisecting line with $x_1 = x_2$ or, represented as a curve,

$$\begin{aligned} x_1 &= x_1 \\ x_2 &= x_1 \end{aligned} \tag{29}$$

which, for uncorrelated data P_i, is the best projection onto one dimension, as indicated in next section. The projection can be computed efficiently without needing matrix operations as the mean value of the vector multiplied with the square root of its dimensionality.

$$\sqrt{m} \cdot \frac{\sum_{i=1}^m x_i}{m} = \left\| \begin{pmatrix} \frac{\sum_{i=1}^m x_i}{m} \\ \frac{\sum_{i=1}^m x_i}{m} \\ \vdots \\ \frac{\sum_{i=1}^m x_i}{m} \end{pmatrix} \right\| = \left\| \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \right\|. \tag{30}$$

A projection $P : R^n \mapsto R^f$ given the decomposition into f orthogonal spaces according to Equation 13 is composed of a sum of f projections onto a one-dimensional space. Each projection is a projection on an m-secting line with $P_i : R^m \mapsto R$. The method works with the space split in any way. For simplicity, we assume that the n-dimensional space is split into f equal-dimensional subspaces. In this case, the projections are efficiently computed as the mean value of each sub-vector. The corresponding mean values are multiplied with the constant $c = \sqrt{m} = \sqrt{\frac{n}{f}}$. The selection of the division can be determined by empirical experiments in which we relate m to n with the constraint that n is divisible by m.

2.2 The First Principal Component

The covariance matrix represents the tendency of two dimensions varying in the same direction as indicated by the data points. The Karhunen-Loève transform rotates the coordinate system in such a way that the new covariance matrix will be diagonal. Therefore, each dimension will be uncorrelated. The transformation is described by an orthonormal matrix, which is composed of the normalized eigenvectors of the covariance matrix. The squares of the eigenvalues represent the variances along the eigenvectors. The first principal component corresponds to the normalized eigenvector $\mathbf{z}$ with the highest variance.

$\|\mathbf{z}\| = 1$ with $Z = \mathbf{z} \cdot \mathbf{z}^\top$ is the best projection onto one-dimensional space because, in a Hillbert space, the first principal component passes through the mean and minimizes the sum of squares of the distances of the points from the line. It follows that

$$\|\mathbf{x}\| \geq \|P \cdot \mathbf{x}\| \geq \|Z \cdot \mathbf{x}\|. \tag{31}$$

For uncorrelated data, $Z = P$ represents the projection on the m-secting line. For correlated data contrary to the projection on the m-secting line, all the components of the vector $\mathbf{z}$ do not need to be equal, and the projection cannot be computed efficiently. For a vector $\mathbf{o}$ of length $\sqrt{m}$ in the direction of the m-secting line where P is the projection on the m-secting line,

$$\mathbf{o} = \underbrace{(1, 1, 1, \cdots, 1)}_{m} \tag{32}$$

it follows that

$$\sqrt{m} = \|\mathbf{o}\| = \|P \cdot \mathbf{o}\| \geq \|Z \cdot \mathbf{o}\| \geq 1. \tag{33}$$

The value

$$\sqrt{m} - \|Z \cdot \mathbf{o}\| \tag{34}$$

indicates the diversion from the m-secting line with value 0 corresponding to uncorrelated data and $\sqrt{m} - 1$ one dimension data with

$$\sqrt{m} - 1 \geq \sqrt{m} - \|Z \cdot \mathbf{o}\| \geq 0. \tag{35}$$

For a given decomposition into f orthogonal spaces according to the Equation 13 the data points are mapped into corresponding subspaces E_i. For each subspace E_i the covariance matrix C_i is computed. In the next step for each covariance matrix C_i the first principal component with the highest variance is determined. It is represented it by the normalised eigenvector $\mathbf{z}_i$. Each projection

$$Z_i = \mathbf{z}_i \cdot \mathbf{z}_i^\top \tag{36}$$

is a projection onto the first principal component with $Z_i : R^m \mapsto R$. An adaptive projection $A : R^n \mapsto R^f$, given the decomposition into f orthogonal spaces according to the Equation 13, is composed of a sum of f projections Z_i onto a one-dimensional space.

$$A = \mathbf{z}_1 \cdot \mathbf{z}_1^\top + \mathbf{z}_2 \cdot \mathbf{z}_2^\top \ldots + \mathbf{z}_f \cdot \mathbf{z}_f^\top. \tag{37}$$

The method works under any splitting of the space, such as the projection $P : R^n \mapsto R^f$.

2.3 l_p norm dependency

Some applications require distance functions that differ from the Euclidian distance function. In addition to the Euclidean distance function, the Manhattan distance and the Chebyshev distance function are commonly used. In the following, we generalize the Euclidean norm to the l_p norm that induces a corresponding metric. The l_p norm is defined as the following (for $p = 2$ it is the Euclidean norm):

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_m|^p)^{\frac{1}{p}} \tag{38}$$

l_p norms are equivalent and the following relation holds for $0 < q < p$

$$\|x\|_p \le \|x\|_q \le m^{\frac{1}{q}-\frac{1}{p}} \cdot \|x\|_p \tag{39}$$

and

$$m^{\frac{1}{p}-\frac{1}{q}} \cdot \|x\|_q \le \|x\|_p \le \|x\|_q. \tag{40}$$

The tighten triangle inequality is valid in any l_p norm due to the definition of norm. Because the l_p norms are equivalent the following equation is valid as well for any l_p norm

$$\|P \cdot \mathbf{x}\|_p - \|P \cdot \mathbf{y}\|_p \le \|P \cdot \mathbf{x} - P \cdot \mathbf{y}\|_p = \|P \cdot (\mathbf{x} - \mathbf{y})\|_p \le \|\mathbf{x} - \mathbf{y}\|_p \tag{41}$$

and

$$\|Z \cdot \mathbf{x}\|_p \le \|P \cdot \mathbf{x}\|_p \le \|\mathbf{x}\|_p. \tag{42}$$

The linear projection operator P has the 1-Lipschitz property in any l_p norm and

$$m^{\frac{1}{p}} \cdot \frac{\sum_{i=1}^{m} x_i}{m} = \left\| \begin{pmatrix} \frac{\sum_{i=1}^{m} x_i}{m} \\ \frac{\sum_{i=1}^{m} x_i}{m} \\ \vdots \\ \frac{\sum_{i=1}^{m} x_i}{m} \end{pmatrix} \right\|_p = \left\| \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \right\|_p. \tag{43}$$

The projection P can be computed efficiently without needing a matrix operation as the mean value of the vector multiplied with the constant $c = m^{\frac{1}{p}}$. For the dimension m, for the l_1 norm, $c = m$, for the l_2 norm, $c = \sqrt{m}$, and for the l_∞ norm, $c = 1$. A lower l_p norm corresponds to a higher constant $m \ge c \ge 1$ and less information loss. We cannot gain any advantage of the 1-Lipschitz property using the different l_p norms. The behavior of the constant c is related to the equivalence of the norms relation. For example, the l_1 and l_2 relation is

$$\|x\|_2 \le \|x\|_1 \le \sqrt{m} \cdot \|x\|_2. \tag{44}$$

For $\|\mathbf{q}\|_p = 1$ with $Q = \mathbf{q}^\top \cdot \mathbf{q}$ is a mapping onto one dimensional space generated by $\mathbf{q}$. It is not a projection for $p > 2$ because the matrix is not self-adjoint with $Q = Q^2$. The mapping on the m-secting

line. the operator can be understood as

$$Q = \mathbf{q}^\top \cdot \mathbf{q} = \begin{pmatrix} \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \\ \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \end{pmatrix}. \tag{45}$$

Q is generated by the l_p normalized vector $\mathbf{q}$ indicating the direction of the m-secting line.

$$\mathbf{q}^\top = \left(\frac{1}{m^{\frac{1}{p}}}, \frac{1}{m^{\frac{1}{p}}}, \cdots, \frac{1}{m^{\frac{1}{p}}} \right). \tag{46}$$

The mapping can be computed efficiently without requiring matrix operations as the mean value of the vector multiplied with the constant $d = m^{\frac{p-1}{p}}$.

$$m^{\frac{p-1}{p}} \cdot \frac{\sum_{i=1}^{m} x_i}{m} = \left\| \begin{pmatrix} \frac{\sum_{i=1}^{m} x_i}{m^{\frac{2}{p}}} \\ \frac{\sum_{i=1}^{m} x_i}{m^{\frac{2}{p}}} \\ \vdots \\ \frac{\sum_{i=1}^{m} x_i}{m^{\frac{2}{p}}} \end{pmatrix} \right\|_p = \left\| \begin{pmatrix} \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \\ \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{m^{\frac{2}{p}}} & \frac{1}{m^{\frac{2}{p}}} & \cdots & \frac{1}{m^{\frac{2}{p}}} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \right\|_p. \tag{47}$$

However this mapping can increase a norm. For the norm l_p the induced matrix norm is

$$\|Q\|_p = \max_{\|x_-p|} \|Q \cdot \mathbf{x}\|_p \tag{48}$$

and for $\mathbf{x} = \mathbf{q}$

$$\|Q\|_p = m^{\frac{p-2}{p}}. \tag{49}$$

It follows that for $p > 2$

$$\|Q \cdot \mathbf{x}\|_p > \|\mathbf{x}\|_p \tag{50}$$

the norm is increased. Only for $p \leq 2$ the norm is not increased with l_2 the projection P and l_1 the simple mean value.

3   Subspace tree revisited

An adaptive projection, $A : R^n \mapsto R^f$, maps two vectors, $\mathbf{x}$ and $\mathbf{y}$, into a lower-dimensional space and satisfies the 1-Lipschitz property:

$$\|A \cdot \mathbf{x} - A \cdot \mathbf{y}\|_p \leq \|\mathbf{x} - \mathbf{y}\|_p. \tag{51}$$

Using the 1-Lipschitz property, a bound that is valid in both spaces can be determined. The distance of similar vectors to a query vector $\mathbf{y}$ is smaller or equal in the original space of dimensionality n, and, consequently, it is also smaller or equal in the lower-dimensional space of the dimensionality f. During the computation, all the points below the bound are discarded. In the second step, the wrong candidates are filtered by comparisons in the original space. The number of points discarded drops as fast as the relation between the dimensionalities $\frac{n}{f}$ grows. Depending on the correlation between the dimensionalities, the 1-Lipschitz property is only useful if the relation is sufficiently small with

$$\frac{n}{f} \leq d \tag{52}$$

where d varies between $2 \leq d \leq 16$ in relation to the data set. However, high-dimensional indexing requires that the mapping $F : R^n \mapsto R^d$ with $n \gg d$ satisfies the 1-Lipschitz property. For such a function, only a tiny fraction of the points of a given set are below the bound. Thus, the majority of the points have to be filtered by comparisons in the original space. Therefore, no speed up, compared tp the use of a simple list matching, can be achieved, as proclaimed by the conjecture "the curse of dimensionality". If at least some points of a given set are below the bound, there is a way to build a recursive function that achieves a considerable speed up using a simple list matching. Motivated by the divide and conquer principle and the tree structure, one can build such a function recursively, indicating that the "the curse of dimensionality" conjecture is *wrong* for some data sets. It is well known that, for a dimensionality d ($2 \leq d \leq 16$), metric index trees operate efficiently. Thus, in the next step we define an efficient indexing structure that builds on the mapping $F : R^n \mapsto R^d$ that satisfies the 1-Lipschitz property, with F being a projection or an adaptive projection.

Suppose there exist a sequence of subspaces $U_0, U_1, U_2, \ldots, U_t$ with $R^n = U_0$ and $R^d = U_t$ in which

each subspace is a subspace of another space

$$U_0 \supset U_1 \supset U_2 \supset \ldots \supset U_t \qquad (53)$$

and with $dim(U_i)$ indicating the dimension of the subspace U_i

$$dim(U_0) > dim(U_1) > dim(U_2) \ldots > dim(U_t)$$

and the relation between neighbouring subspaces is sufficiently small with

$$\frac{dim(U_0)}{dim(U_1)} \leq d, \frac{dim(U_1)}{dim(U_2)} \leq d, \ldots \frac{dim(U_{t-1})}{dim(U_t)} \leq d. \qquad (54)$$

We define a family of projections for the sequence of subspaces (either adaptive or not) with the following

$$A_1 : U_0 \mapsto U_1; A_2 : U_1 \mapsto U_2; \ldots; A_t : U_{t-1} \mapsto U_t. \qquad (55)$$

The family of projections defines the sequence of subspaces. Given a bound ε and a query vector $\mathbf{y}$ for each subspace including the original space U_0, certain points are below the bound. For each subspace U_i, the number of points below the bound ε is indicated by the value σ_i. It follows that

$$\sigma_0 < \sigma_1 < \ldots < \sigma_t < s \qquad (56)$$

where s is the size of the data set. The resulting computing cost given a bound ε and a query vector $\mathbf{y}$ is

$$cost_s = \sum_{i=1}^{t} \sigma_i \cdot dim(U_{i-1}) + s \cdot dim(U_t). \qquad (57)$$

The cost of list matching is

$$cost_l = s \cdot dim(U_0) \qquad (58)$$

The saving $cost_s < cost_l$ is related to the bound ε. Empirical experiments suggest that $cost_s \ll cost_l$ for a bound with $\sigma_0 < d$.

The described projection based method cannot be applied to sparse representation, as present in the vector space model [1].

3.1 Tree isomorphy

The isomorphy to a tree results from the assumption that the value of σ_i is reciprocal to the preceding dimensionality [13]. Therefore, a bigger dimensionality $dim(U_{i+1})$ results in a smaller value σ_i, and vice versa. We can express this relation by

$$const \cdot \sigma_i = \frac{1}{dim(U_{i+1})}. \tag{59}$$

The value of *const* is dependent on the data set and its norm. The value of σ_i is reciprocal to the preceding dimensionality $dim(U_{i+1})$ (Equation 59), and the computing costs are expressed by

$$cost_s \approx 1/const \cdot \left(\frac{dim(U_0)}{dim(U_1)} + \frac{dim(U_1)}{dim(U_2)} + \ldots + \frac{dim(U_{n-1})}{dim(U_t)} \right) + dim(U_t) \cdot s. \tag{60}$$

Supposing $d = dim(U_t)$ and $n = dim(U_0)$

$$cost_s \approx 1/const \cdot d \cdot \log_d(n-d) + d \cdot s. \tag{61}$$

For a dimension d, the metric index trees operate efficiently with a nearly logarithmic search time. For the bound with $\sigma_0 < d$, the value $1/const \ll s$

$$cost_s \approx 1/const \cdot d \cdot (\log_d(n) - 1) + d \cdot \log_d(s). \tag{62}$$

It follows that the lower bound of the computational cost is

$$\Omega(\log(n) + \log(s)). \tag{63}$$

4 Examples of ε similarity

The ε range queries depends on the adequate value of ε. A method for the estimation of such a value is described in [12]. Let *DB* be a database of s multimedia objects $\mathbf{x}^{(i)}$ represented by vectors of dimensionality n in which the index i is an explicit key identifying each object

$$\{\mathbf{x}^{(i)} \in DB | i \in \{1..s\}\}. \tag{64}$$

For a query object $\mathbf{y}$, *all* objects $\mathbf{x}^{(i)}$ are searched that are ε-similar

$$\|\mathbf{x}^{(i)} - \mathbf{y}\|_p < \varepsilon. \tag{65}$$

For high dimensional vector space such a set can be determined by a list matching over the whole data set.

4.1 Computational procedure

For the database DB that is projected into the subspace U_k,

$$\{U_k(\mathbf{x})^{(i)} \in U_k(DB) | i \in \{1..s\}\}. \tag{66}$$

The algorithm to determine all ε-similar objects is composed of two loops. The first loop iterates over the elements of the database DB, and the second iterates over their representation[1]. We can easily parallelize the algorithm over the first loop; different parts of the database can be processed by different processors, kernels, or computers.

Algorithm to determine NN

forall $\{\mathbf{x}^{(i)} \in DB | i \in \{1..s\}\}$

$\qquad \{$

$\qquad\quad for(k = t; k \neq 0, k--)$

$\qquad\qquad \{$

$\qquad\qquad\quad load(U_k(\mathbf{x})^{(i)});$

$\qquad\qquad\quad$ /* 1-Lipschitz property */

$\qquad\qquad\quad if\ (\|U_k(\mathbf{x})^{(i)} - U_k(\mathbf{y}))\|_p >= \varepsilon)$

$\qquad\qquad\qquad break ::;$

$\qquad\qquad\qquad if\ (k = 0)\ print\ \mathbf{x}^{(i)}\ is\ NN\ of\ \mathbf{y}$

$\qquad\qquad \}$

$\qquad \}$

---

[1]An implementation can be obtained upon request from the author

Each call of the 1-Lipschitz property costs $dim(U_k)$. The cost according to Equation 57 correspond to the number of 1-Lipschitz property calls, corresponding to the value σ_k.

4.2 960-dimensional vector space

We apply computational procedure on high-dimensional data set of 100000 vectors of dimensionality 960. The vectors represent the GIST global descriptor of an image and are composed by concatenated orientation image histograms [9]. The vector $\mathbf{x}$ of dimensionality 960 is split into 480 distinct subvectors of dimensionality 2. The data points are described by 480 covariance matrices C_i for each subspace. For all points, the covariance matrices are computed iteratively.

$$\mathbf{x} = \underbrace{x_1, x_2}_{C_1}, \underbrace{x_3, x_4}_{C_2}, \cdots \cdots, \underbrace{x_{479}, x_{959}}_{C_{480}}. \tag{67}$$

The resulting 480 projections, $\mathbf{z}_i \cdot \mathbf{z}_i^\top$, define the adaptive projection $A : R^{960} \mapsto R^{480}$. We apply the adaptive projection and the determination of the adaptive projection recursively. The resulting family of projections,

$$A_1 : U_0 \mapsto U_1; A_2 : U_1 \mapsto U_2; \ldots; A_7 : U_6 \mapsto U_7 \tag{68}$$

defines the dimensionalities of the subspaces.

$$dim(U_0) = 960 > dim(U_1) = 480 > dim(U_2) = 240 > dim(U_3) = 120$$

$$> dim(U_4) = 60 > dim(U_5) = 30 > dim(U_6) = 10 > dim(U_7) = 5.$$

In Table 1, we indicate the mean costs according to Equation 57 using the l_2 norm.

projection	ε for ≈ 52 NN	mean cost	ratio
orthogonal	6300	4584277	21.38
adaptive	6300	4393127	22.31

Table 1: Mean ratio of list matching to the mean computation costs according to Equation 57. The values were determined over a disjunct sample of $S \subseteq DB$ with size $|S| = 400$. The adaptive projection gives only a slight improvement. The diversion from the m-secting line according to Equation 34 is always $\ll 0.0001$.

4.3 12288-dimensional vector space

The 12288-dimensional vector space corresponds to an image database that consists of 9.876 3-band RGB (Red, Green, Blue) images of size 128×96. Each color is represented by 8 bits [12]. Each of the tree bands of size 128×96 is tiled with rectangular windows W of size 4×4. The data points are described by 32×24 covariance matrices C_i for each subspace, for each band. The resulting $768 = 32 \times 24$ projections $\mathbf{z}_i \cdot \mathbf{z}_i^\top$ define the adaptive projection $A : R^{12288} \mapsto R^{768}$ for each band (Red, Green, Blue). We apply the adaptive projection and the determination of the adaptive projection recursively. The resulting family of projections,

$$A_1 : U_0 \mapsto U_1; A_2 : U_1 \mapsto U_2; A_3 : U_2 \mapsto U_3 \tag{69}$$

defines the dimensionalities of the subspaces for each band

$$dim(U_0) = 12288 > dim(U_1) = 768 > dim(U_2) = 48 = 8 \times 6 > dim(U_3) = 12 = 4 \times 3.$$

For an orthogonal projection, the sequence of subspaces $U_0 \supset U_1 \supset U_2 \supset U_3$ corresponds to the "image pyramid" [3], [6], which has a base that contains an image with a high-resolution and an apex that is the low-resolution approximation of the image. In Table 2, we indicate the mean costs according to Equation 57. The l_1 norm gives the best results.

projection	l_p	ε for ≈ 52 NN	cost	ratio
orthogonal	l_1	1240000	8571752	42.47
adaptive	l_2	8500	10343766	35.20
orthogonal	l_2	8500	10386043	35.05
orthogonal	l_4	825	12464281	29.32
orthogonal	l_∞	161	39639239	9.19

Table 2: Mean ratio of list matching to the mean computation costs according to Equation 57. The values were determined over a disjunct sample of $S \subseteq DB$ with size $|S| = 400$. The diversion from the m-secting line, according to Equation 34, is always $\ll 0.0001$.

5 Conclusion

An adaptive projection that satisfies the 1-Lipschitz property defined by the first principal component was introduced. We indicated the behavior of the projections for the l_p norms. The Manhattan distance

l_1 loses the least information, followed by Euclidean distance function. Most information is lost when using the Chebyshev distance function. Motivated by the tree structure, we indicated a family of projections that defines a mapping that satisfies the 1-Lipschitz property. It is composed of orthogonal or adaptive projections in the l_p space. Each projection is applied recursively in a low-dimensional space, where "the curse of dimensionality" conjecture does not apply.

ACKNOWLEDGEMENTS

References

[1] Baeza-Yates, R., Ribeiro-Neto, B.: Modeling, in: *Modern Information Retrieval* (R. Baeza-Yates, B. Ribeiro-Neto, Eds.), chapter 2, Addison-Wesley, 1999, 19–71.

[2] Böhm, C., Berchtold, S., Kei, D., A. K.: Searching in High-Dimensional Spaces—Index Structures for Improving the Performance of Multimedia Databases, *ACM Computing Surveys*, **33**(3), 2001, 322–373.

[3] Burt, P. J., Adelson, E. H.: The Laplacian Pyramidas a Compact Image Code, *IEEE Trans. Commin*, **COM-31**(4), 1983, 532–540.

[4] Faloutsos, C.: Modern Information Retrieval, in: *Modern Information Retrieval* (R. Baeza-Yates, B. Ribeiro-Neto, Eds.), chapter 12, Addison-Wesley, 1999, 345–365.

[5] Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and Effective Querying by Image Content, *Journal of Intelligent Information Systems*, **3**(3/4), 1994, 231–262.

[6] Gonzales, R. C., Woods, E. W.: *Digital Image Processing*, second edition, Prentice Hall, 2001.

[7] Indyk, P.: Nearest neighbours in high-dimensional spaces, in: *Handbook of Discrete and Computational Geometry* (J. Goodman, J. O'Rourke, Eds.), Boca Raton–London–New York–Wahington, 2004.

[8] Indyk, P., Motwani, R.: Approximate nearest neighbours: towards removing the curse of dimensionality, *Proc. 30th ACM Symp. Theory of Computing*, 1998.

[9] Jegou, H., Douze, M., Schmid, S.: Product quantization for nearest neighbor search, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**(1), 2011, 117–128.

[10] Pestov, V.: Lower bounds on performance of metric tree indexing schemes for exact similarity search in high dimensions, *Proceedings of the 4th International Conference on Similarity Search and Applications* (A. Ferro, Ed.), 2011.

[11] Pestov, V.: Indexability, concentration, and VC theory, *Journal of Discrete Algorithms*, **13**, 2012, 2–18.

[12] Wichert, A.: Content-based image retrieval by hierarchical linear subspace method, *Journal of Intelligent Information Systems*, **31**(1), 2008, 85–107.

[13] Wichert, A., Teixeira, P., Santos, P., Galhardas, H.: Subspace tree: High dimensional multimedia indexing with logarithmic temporal complexity, *Journal of Intelligent Information Systems*, **35**(3), 2010, 495–516.

[14] Wichert, A., Verissimo, A.: CBIR with a Subspacee-tree: Principal Component Analysis versus Averaging, *Multimedia Systems*, **18**(4), 2012, 283–293.