A Generalization of Principal Component Analysis

Samuele Battaglino
Department of Computer Science
University of Illinois at Chicago

Erdem Koyuncu
Department of Electrical and Computer Engineering
University of Illinois at Chicago

Introduction

- PCA and variant methods are dimension reduction techniques that rely on orthogonal transformations, let $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^d$ be a set of d-dimensional input vectors with zero mean. In the conventional L^2-PCA, one "extracts" the first principal vector via:

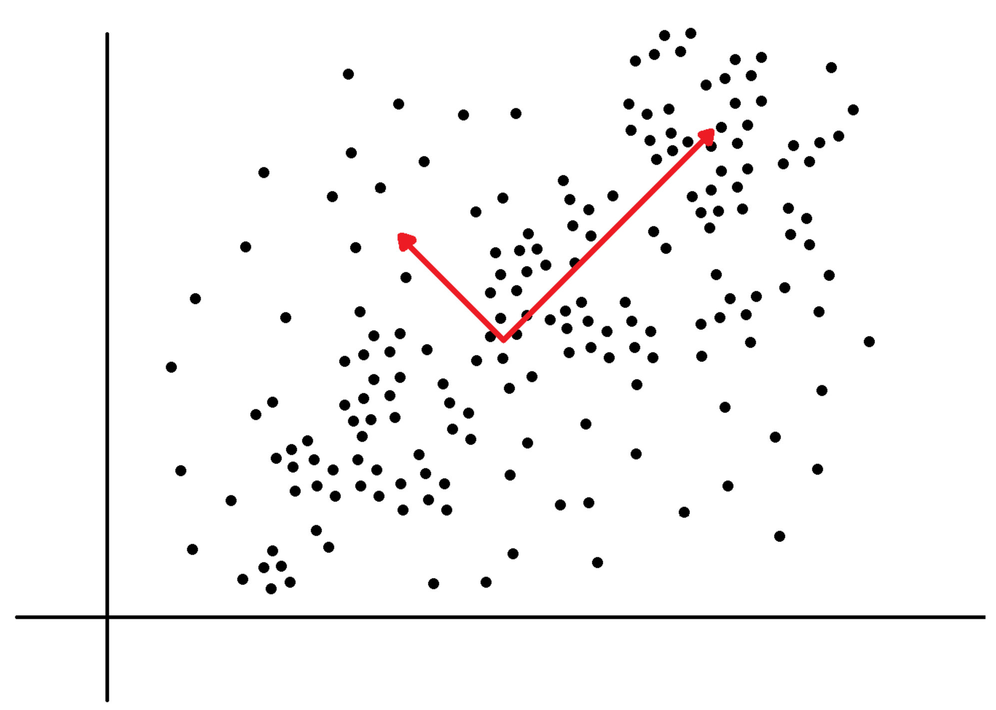
$$w_{L^2}^{(1)} \triangleq \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} (w^T x_i)^2$$

- In its original L^2-form, PCA is very susceptible to outliers or highly-noisy datasets. This is because a noisy sample or an outlier x_i may result in a large inner product $w^T x_i$, which will be further amplified as one considers the second powers of inner products in the objective function.
  Thus, by lowering the power it is possible to achieve a more robust L^p-PCA:

$$w_{L^p}^{(1)} \triangleq \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} |w^T x_i|^p$$

- **Subsequent principal vectors** of the PCA can be extracted by projecting X to a subspace orthogonal to the previous found ones through this equation:

$$w_{L^p}^{(j)} \triangleq \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} \left(w^T \left(I - \Sigma_{k=1}^{j-1} w_{L^p}^{(k)} \left(w_{L^p}^{(k)}\right)^T\right) x_i\right)^p$$



- Kernel Principal Component Analysis (KPCA) refers to performing PCA in a certain feature space. In the case of the L^2-form, we extract the first principal vector via:

$$w_{L^2,\phi}^{(1)} = \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} (w^T \phi(x_i))^2$$

Generalized PCA

In this work, we study the Generalized PCA (**GPCA**) problem:

$$w_f^{(1)} \triangleq \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} f(w^T x_i)$$

where f is an arbitrary convex function.

Theorem 1: Let $F(w)$ be a convex function. Let $||w'|| = 1$ and $w'' = \frac{\nabla F}{||\nabla F||}\,|_{w=w'}$. Then, $F(w') \geq F(w)$.
In particular, the update $w \leftarrow \frac{\nabla F}{||\nabla F||}$ provides a locally maximum solution to the problem $max_{w:||w||=1} F(w)$.

For our scenario, we first note that $w \leftarrow f(w^T x_i)$ is convex for any x_i, as the composition of a convex function and an affine function is always convex. It follows that $\sum_{i=1}^{N} f(w^T x_i)$ is convex and Theorem 1 is applicable. We thus use the update:

$$w \leftarrow \frac{\sum_{i=1}^{N} f'(w^T x_i) x_i}{||\sum_{i=1}^{N} f'(w^T x_i) x_i||}$$

The complete suggested algorithm is as follows:

1. Initialize $w \leftarrow \arg\max_i \frac{x_i}{||x_i||}$.
2. Iterate $w \leftarrow \frac{\sum_{i=1}^{N} f'(w^T x_i) x_i}{||\sum_{i=1}^{N} f'(w^T x_i) x_i||}$ until convergence.
3. Update $X \leftarrow (I_N - ww^T)X$ where I_N is a NxN identity matrix, this step is required in order to extract the next component.

Intuition for the choice for f:

Outliers typically induces a principal component with a large magnitude. Therefore, for the optimal performance, $f(\cdot)$ should behave as the L^1-norm $f(x) \approx x$ for abnormalities (large $|x|$), and it should behave as the L^2-norm $f(x) \approx x^2$ for normalities (small $|x|$).

Examples:

- $g_a(x) = \begin{cases} x^2, & x \leq a \\ |x|, & x > a \end{cases}$
- $\zeta_1(x) = (1 - \operatorname{sech}(|x|)sign(x)$
- $\zeta_2(x) = \tanh\text{\^{}}2(|x|)sign(x)$

Generalized Kernel PCA

We also study the Kernel version of the problem, the Generalized KPCA (**GKPCA**):

$$w_{f,\phi}^{(1)} \triangleq \underset{w:||w||=1}{\arg\max} \sum_{i=1}^{N} f(w^T \phi(x_i))$$

where $\phi(\cdot)$ is an arbitrary feature map.

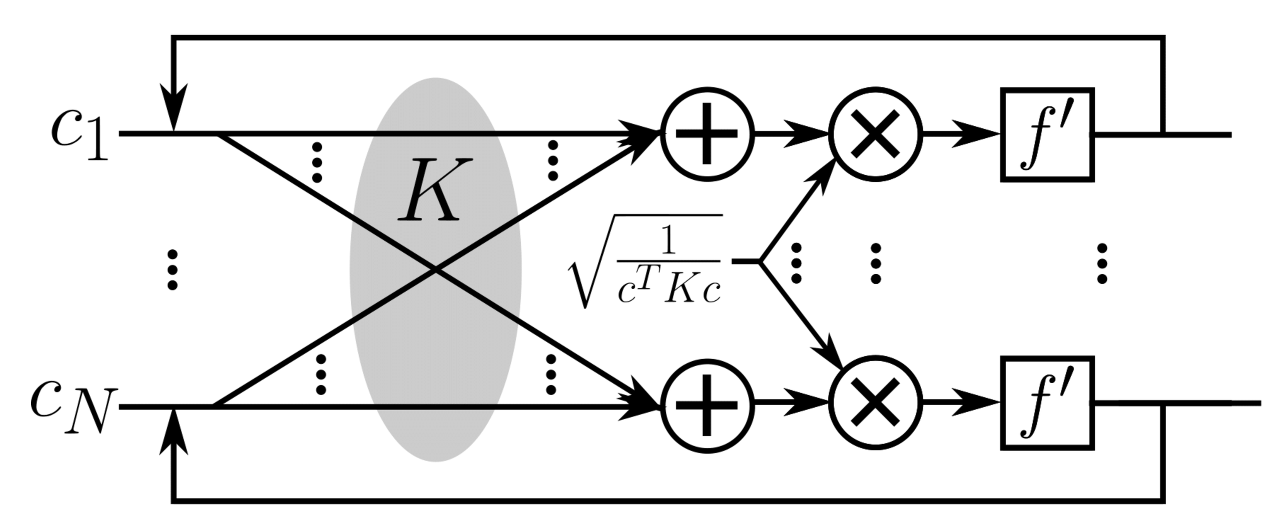
Let $c_i \triangleq f'(w^T \phi(x_i))$, $c = [c_1 \dots c_N]^T$, K define the Kernel matrix with entry $\phi(x_i)^T \phi(x_j)$ in the ith row, jth column, and Y be the all-1/n matrix. We propose the following algorithm to calculate the GKPCA:

$$c_i \leftarrow f'\left(\frac{\sum_{j=1}^{N} K_{ij} c_j}{\sqrt{\sum_{z,q=1}^{N} c_z c_q K_{zq}}}\right)$$

The principal component of an input x is then extracted via

$$w^T \phi(x) = \sum_{j=1}^{N} c_j K(x_j, x) / \sqrt{c^T K c}$$

A simple Recurrent Neural Network to find c and thus extract the principal components:



Algorithm 1: The Generalized KPCA Algorithm

Let K have entry $(\Phi(x_i))^T \Phi(x_j)$ in Row i, Column j.
$K \leftarrow K - YK - KY + YKY$.
for $i \leftarrow 1$ **to** *#(Principal Vectors to be Extracted)* **do**
  Set $c_j = 1, c_i = 0, i \neq j$, where $j = \arg\max_i K_{ii}$.
  Iterate $c \leftarrow f'(\frac{Kc}{\sqrt{c^T K c}})$ until convergence.
  Store c, K to analyze the ith principal component.
  $K \leftarrow K - Kcc^T K / c^T K c$
end

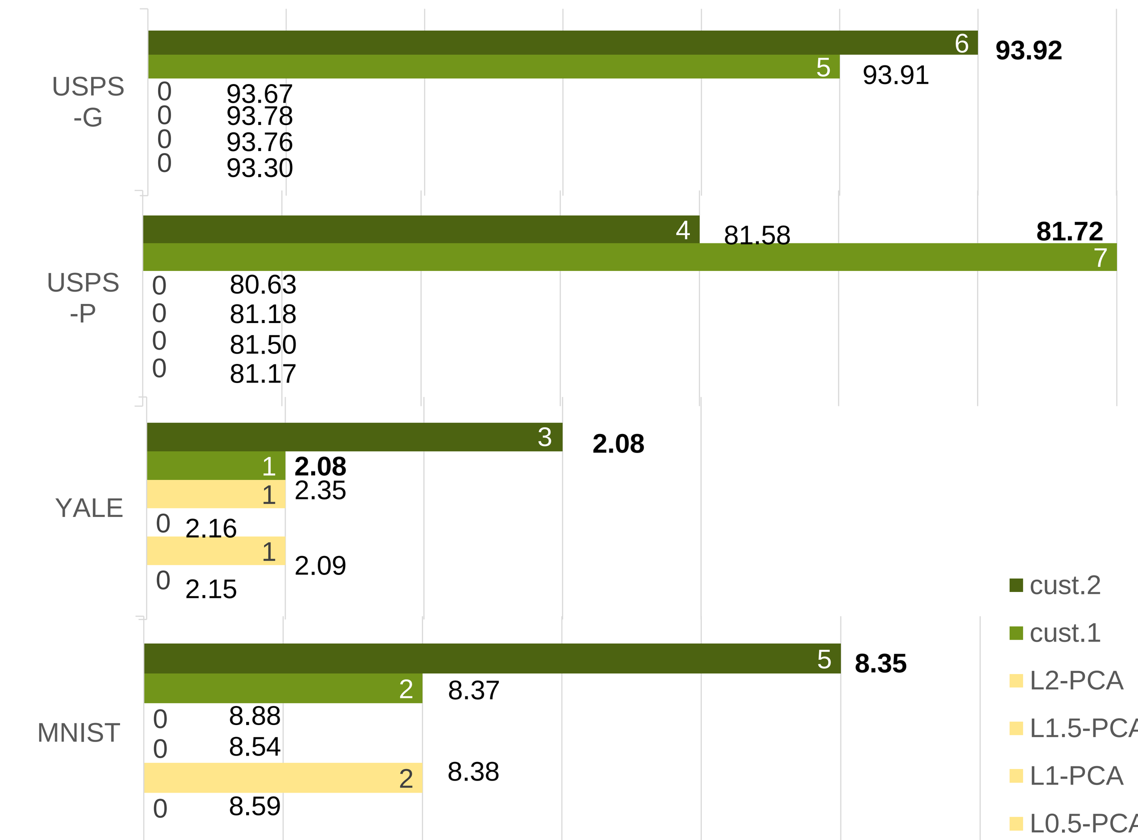
Results

Both methods have been tested extensively with 3 datasets (**USPS, MNIST and YALE faces**) combined with 4 different types of noise (**Gaussian, Salt & Pepper, Outlier Injection and Speckle**).

The graph will display the two best functions for each test suite in green against the methods available in the literature in yellow.

The length of each bar is defined by *how many times that method*, in a given dataset and noise, *outperformed the others* at a noise level. The *average performance* for the whole test suite is shown at the right of each bar (classification rate for USPS, reconstruction error for YALE and misclassification rate for MNIST).

GPCA Results:



GKPCA Results: