Optimal Sparse $L_1$-Norm Principal-Component Analysis

Shubham Chamadia and Dimitris A. Pados
Signals, Communications, and Networking Research Group, University at Buffalo, The State University of New York
Email: {shubhamc, pados}@buffalo.edu

Motivation

- Consider genetic colon data $X \in \mathbb{R}^{2000 \times 62}$, 2000 gene expressions from 62 known healthy/cancerous subjects [1].
- **Aim:** Identify relevant genes.

![Figure 1. Normalized explained variance vs principal-component (PC) sparseness.](image)

**Introduction**

- In (“big”) data applications, not all dimensions (coordinates) equally important.
- **Goal:** Extract meaningful information from few undetermined coordinates [3].
- **Coordinate-based preference motivates sparsity:** Sparse principal-component analysis (SPCA).
- **Trade-off between statistical fidelity (close data representation) and interpretability** (few non-zero coordinates).
- **Applications as broad as:**
  - Financial stock market analysis
  - Microarray genomic-data classification
  - Improved clustering for biometric recognition
  - Ranking systems, movies, etc.

**Problem Setup**

- **Conventional $L_2$-SPCA:** Given data $X \in \mathbb{R}^{D \times N}$
  
  $$L_2\text{-SPCA: } q_i^{L_2} = \arg \max_{q_i \in \mathbb{R}^D, \|q_i\|_2=1} \|X^T q_i\|_2.$$  

- **Sub-optimal [2], [4] and optimal solvers [5].**

- **Con:** $L_2$-norm supported designs are highly sensitive to faulty/outlying data.

  - **Idea:** Switch to $L_1$-norm based computation.

- **$L_1$-SPCA:** Given data $X \in \mathbb{R}^{D \times N}$
  
  $$L_1\text{-SPCA: } q_i^{L_1} = \arg \max_{q_i \in \mathbb{R}^D, \|q_i\|_1=1} \|X^T q_i\|_1.$$  

**Prior Work**

- **Sub-optimal solver [6]:**

  - Initialize arbitrary unit vector $q_i^{(0)} \in \mathbb{R}^D$ and continue by
    
    $$b_i^{(t+1)} = \text{sgn}(X^T q_i^{(t)}),$$  

    $$q_i^{(t+1)} = \frac{(Xb_i^{(t+1)} - S)}{\|Xb_i^{(t+1)} - S\|_2}, \quad i = 0, 1, 2, \ldots.$$  

  - $\Delta(\cdot, S)$ returns the $S$ largest absolute values of input vector reduced by the $(S+1)$-largest absolute value.

- **Con:** Often converges to local-maxima and suffers heavy performance loss [6].

**Proposed Optimal $L_1$-Sparse Solver**

**Problem Translation:**

- From continuous search space of $q_i^{L_1} \in \mathbb{R}^D$ to discrete binary combinatorial pair $(I_{opt}, b_{opt})$.

  - **$I_{opt} \subseteq D = \{1, 2, \ldots, D\}$, $b_{opt} \in \{\pm 1\}^N$.**

  $$\max_{q_i \in \mathbb{R}^D, \|q_i\|_1=1} \|X^T q_i\|_1 = \max_{p_i \in \{\pm 1\}^D} \sum_i p_i |X_i b_i|_1.$$  

  $$\max_{p_i \in \{\pm 1\}^D} \sum_i p_i |X_i b_i|_1 = \max_{p_i \in \{\pm 1\}^D} \sum_i p_i |X_i b_i|_2 - \frac{\sum_i |X_i b_i|_2}{|D|}.$$  

- **Finally,** $q_i^{L_1} (I_{opt}) = X_{I_{opt}} b_{opt} \|X_{I_{opt}} b_{opt}\|_2$.

**Algorithmic Development**

**Q:** How to compute the optimal joint pair of index $I_{opt} \subseteq D$, $|I_{opt}| = S$ and binary vector $b_{opt} \in \{\pm 1\}^N$?

- **Case 1 ($N < D$):**

  $$\max_{b \in \{\pm 1\}^N} \max_{|I| = S} \|X_{I} b\|_2,$$

  - **Exhaustive $b \in \{\pm 1\}^N$ with complexity $O(2^N)$**

  - **solve**
    
    $$\max_{|I| = S} \|X_{I} b\|_2$$  

    - with complexity $O(DS)$

    - **O($2^N D$)**

- **Case 2 ($N > D$):**

  - **Exhaustive $I \subseteq D$, $|I| = S$ with complexity $O(D^S)$**

  - **solve**

    - $\max_{|I| = S} \|X_{I} b\|_2$ with complexity $O(N^S)$

    - **O($D^S N$)**

- **Numerical Studies**

**Experiment 1:** On-off signal detection

- **D = 100 snapshots, N = 14 antennas to form $X \in \mathbb{R}^{100 \times 14}$ in AWGN $\mathcal{N}(0, 1)$.

  - **30 (out of 100) snapshots contain active signal $\mathcal{N}(4, 2)$.

- **Corruption:** 2 antennas, 5 entries, $\mathcal{AWGN}(0, 15)$.

- **Performance metric:** Normalized explained variance

  $$\text{NEV} \triangleq \|X^T q_i^{\text{opt}}\|_2/\|X^T q_i\|_2$$

where $q_i^{\text{opt}}$ is $S$-sparse PC over corrupted data and $q_i$ is standard $L_2$-PC over clean data.

**Experiment 2:** Sparse $L_1$-image fusion

- **15 copies of 256 × 256 gray scale Lena image.**

  - **Noised by AWGN $\mathcal{N}(0, 100)$.**

  - **Arbitrarily chosen 8 (out of 16) square patches overwritten by salt and pepper corruption.**

  - **Collect vectorized patch from each image to form $X^p = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]\in \mathbb{R}^{P \times 256^2}$**

  - **Compute $q_i^{p}$ as $S$-sparse PC of $X_p$.**

  - **Normalized reliability-weight $w^{p}_i = r^p_i/\sum_i r^p_i$ where $r^p_i = \|\mathbf{x}_i - q_i^{p}q_i^{p\top}\|_2^2$**

  - **(d) RSPCA [6], (e) Tpower [2], (f) EMPCA [4], (g) proposed optimal $L_1$-sparse restored image.**

**References**


