**Motivation**

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

![](image1.png)

*Figure 1. Normalized explained variance vs principal-component (PC) sparseness.*

**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 $X \in \mathbb{R}^{D \times N}$

  $$L_2 \text{-SPCA: } q_{i}^{\text{opt}} = \arg \max_{q \in \mathbb{R}^{D}, \|q\|_2=1} \|Xq\|_2.$$  

- 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.

- Proposed optimal $L_1$-Sparse Solver

  - Initialize arbitrary unit vector $q_i^{(0)} \in \mathbb{R}^D$

  $$q_i^{(n+1)} = \arg \max_{q \in \mathbb{R}^{D}, \|q\|_2=1} \|Xq\|_2.$$  

  - Continue by

  $$q_i^{(n+1)} = \arg \max_{q \in \mathbb{R}^{D}, \|q\|_2=1} \|q\|_2.$$  

  - $b_i = \Delta(Xq^{(n)}), S \setminus I = \{0, 1, \ldots, S \} \setminus I$

  - $S$-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].

**Algorithmic Development**

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

  - $X^2 = \max_{b \in \{-1, 1\}^S} \max_{D \subseteq \{1, \ldots, D \}} \|Xb\|_2$.

- Case 1 ($N < D$)

  - Exhaustive $b \in \{-1, 1\}^S$ with complexity $O(2^S)$

  - $X^2 = \max_{D \subseteq \{1, \ldots, D \}} \|Xb\|_2$ with complexity $O(D^S)$.

- Case 2 ($N > D$)

  - Exhaustive $D \subseteq \{1, \ldots, D \}$, $S = \{1, \ldots, S \}$

  - $X^2 = \max_{D \subseteq \{1, \ldots, D \}} \|Xb\|_2$ with complexity $O(D^S)$.

- $\|X\|^2 = \max_{D \subseteq \{1, \ldots, D \}} \|Xb\|_2$.

**Numerical Studies**

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

- $D = 100$ snapshots, $N = 14$ antennas to form $X \in \mathbb{R}^{200 \times 14}$ in AWGN (N(0, 1)).

- 30 (out of 100) snapshots contain active signal $S \setminus I \subseteq\{1, \ldots, 14\}$.

- Corruption: $2$ antennas, $S$ entries, AWGN $N(0, 15)$.

- Performance metric: Normalized explained variance

  $$NEV = \frac{\|X^2 q_{\text{opt}}^2\|^2}{\|X^2 q\|^2},$$

where $q_{\text{opt}}^2$ is $S$-sparse PC over corrupted data and $q$ is standard $L_2$-PC over clean data.

![](image2.png)

*Figure 2. Normalized explained variance versus PC sparseness.*

**References**


