Generalized Power Method for Sparse Principal Component Analysis

M. Journée* Yu. Nesterov† P. Richtárik† R. Sepulchre*

†Université Catholique de Louvain *Université de Liège

- **Input:** Matrix $A = [a_1, \dots, a_n] \in \mathbf{R}^{p \times n}$, $p \le n$
- **Goal:** Find vector $z^* \in \mathbf{R}^n$ which simultaneously
  1. **maximizes variance** $z^T A^T A z$
  2. is **sparse**

If sparsity is not required, z^* is the dominant right singular vector of A. This is the single-unit ($m = 1$) case. Often $m > 1$ components (sparse dominant singular directions) are needed – block case.

Our approach:

1. Formulate sPCA as an optimization problem with sparsity-inducing penalty (ℓ_1 or ℓ_0) controlled by a single parameter γ
2. Reformulate to get problem of a suitable form
3. Solve reformulation using a gradient scheme
4. Do post-processing in the ℓ_1 case (will not detail it here)

Notation: $\|z\|_1 = \sum_i |z_i|$, $\|z\|_0 = \mathrm{Card}\{i \,:\, z_i \neq 0\}$.

Single-unit sPCA via ℓ_1-penalty

$$\phi_{\ell_1}(\gamma) \overset{\text{def}}{=} \max_{z^T z \le 1} \sqrt{z^T A^T A z} - \gamma \|z\|_1. \qquad (1)$$

1. To solve (1), first solve this reformulation

$$\phi^2_{\ell_1}(\gamma) = \max_{\substack{x \in \mathbf{R}^p \\ x^T x = 1}} \sum_{i=1}^{n} [|a_i^T x| - \gamma]_+^2. \qquad (2)$$

2. and then set

$$z_i = \mathrm{sign}(a_i^T x)[|a_i^T x| - \gamma]_+, \qquad z^* = z/\|z\|_2.$$

Single-unit sPCA via ℓ_0-penalty

$$\phi_{\ell_0}(\gamma) \overset{\text{def}}{=} \max_{z \in \mathcal{B}^n} z^T A^T A z - \gamma \|z\|_0, \qquad (3)$$

1. To solve (3), first solve this reformulation

$$\phi_{\ell_1}(\gamma) = \max_{\substack{x \in \mathbf{R}^p \\ x^T x = 1}} \sum_{i=1}^{n} [(a_i^T x)^2 - \gamma]_+. \qquad (4)$$

2. and then set

$$z_i = [\mathrm{sign}((a_i^T x)^2 - \gamma)]_+ a_i^T x, \qquad z^* = z/\|z\|_2.$$

Problems (2) and (4) (and their block generalizations) are of the form

$$f^* = \max_{x \in \mathcal{Q}} f(x). \qquad \text{(P)}$$

- $\mathbf{E}$ is a finite-dimensional vector space
- $f : \mathbf{E} \to \mathbf{R}$ is a convex function
- $\mathcal{Q} \subset \mathbf{E}$ is compact

In the single-unit case ($m = 1$), $\mathcal{Q}$ is the unit Euclidean sphere in $\mathbf{R}^p$, in the block case ($m > 1$), $\mathcal{Q}$ is the Stiefel manifold in $\mathbf{R}^{p \times m}$, i.e. the set of $p \times m$ matrices with orthonormal columns.

We will solve (P) using this **gradient algorithm (GA)**:

1. **Input:** Initial iterate $x_0 \in \mathbf{E}$
2. **For** $k \ge 0$ **repeat**
   - $x_{k+1} \in \mathrm{Arg\,max}\{f(x_k) + \langle f'(x_k), y - x_k \rangle \mid y \in \mathcal{Q}\}$
   - $k \leftarrow k + 1$

Theorem 1 (Convergence) Let f be convex with strong convexity parameter $\sigma_f \ge 0$ and $\mathrm{Conv}(\mathcal{Q})$ be strongly convex with parameter $\sigma_{\mathcal{Q}} \ge 0$. If $0 < \delta_f \le \inf_{x \in \mathcal{Q}} \|f'(x)\|_*$ and either $\sigma_f > 0$ or $\sigma_{\mathcal{Q}} > 0$, then

$$\sum_{k=0}^{N} \|x_{k+1} - x_k\|^2 \le \frac{2(f^* - f(x_0))}{\sigma_{\mathcal{Q}} \delta_f + \sigma_f}.$$

Our algorithm generalizes the power method for computing the largest eigenvalue of a symmetric positive definite matrix C:

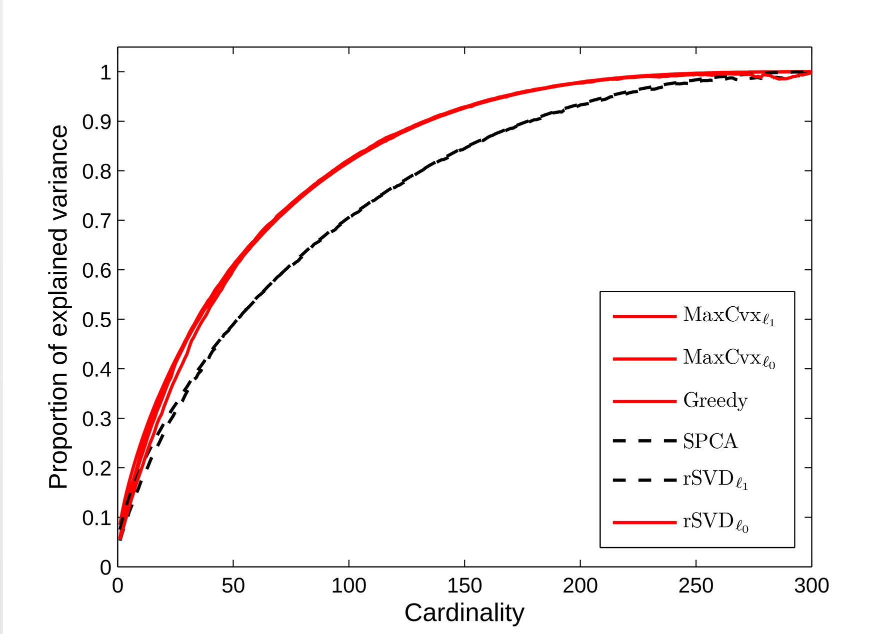
$$\max f(x) \equiv \tfrac{1}{2} x^T C x \quad \longrightarrow \quad x_{k+1} = \frac{C x_k}{\|C x_k\|_2}.$$

We compare the following Sparse PCA algorithms:

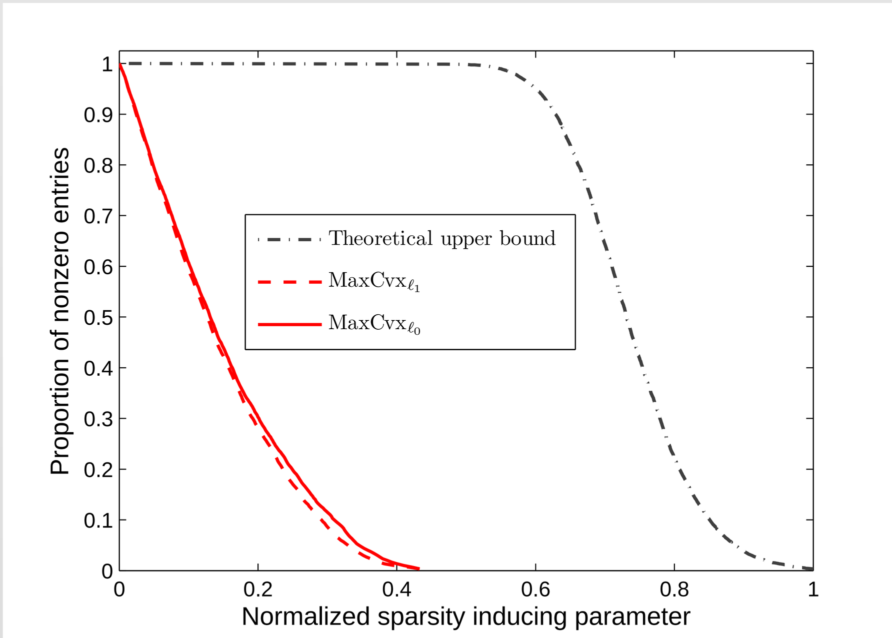
GPower$_{\ell_1}$	Single-unit sparse PCA via ℓ_1-penalty
GPower$_{\ell_0}$	Single-unit sparse PCA via ℓ_0-penalty
GPower$_{\ell_1,m}$	Block sparse PCA via ℓ_1-penalty
GPower$_{\ell_0,m}$	Block sparse PCA via ℓ_0-penalty
SPCA	SPCA algorithm [1]
Greedy	Greedy method [2]
rSVD$_{\ell_1}$	Method [3] with ℓ_1-penalty (“soft thresholding”)
rSVD$_{\ell_0}$	Method [3] with ℓ_0-penalty (“hard thresholding”)

Greedy slows down dramatically, compared to the other methods, if aimed at obtaining a component of higher cardinality.

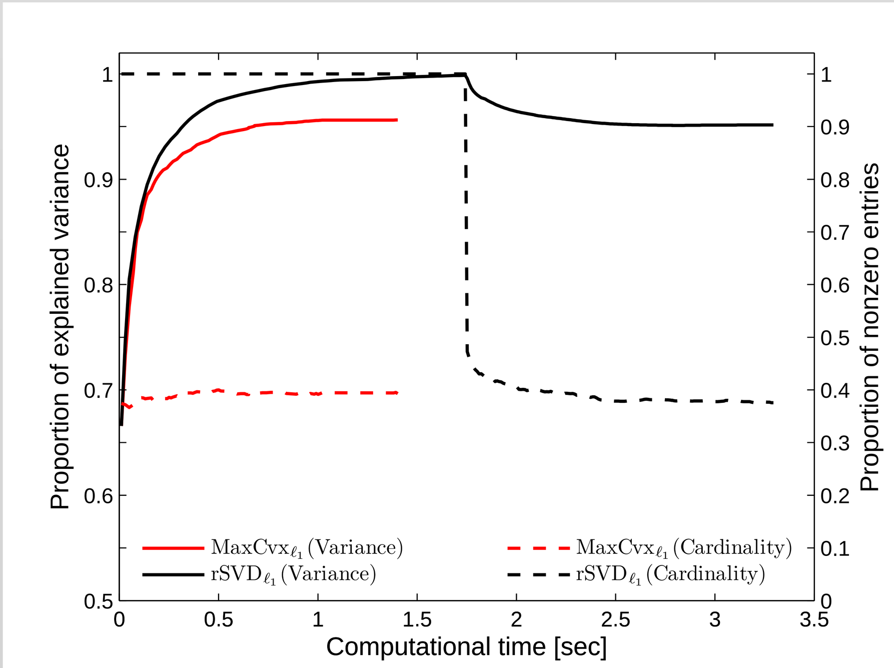
The entries of A are Gaussian with zero mean and unit variance. The first two plots are based on an average of 100 test problems of size $p = 100$ and $n = 300$.



Trade-off curves. Trade-off curve between explained variance and cardinality. The algorithms aggregate in two groups. The methods GPower$_{\ell_1}$, GPower$_{\ell_0}$, Greedy and rSVD$_{\ell_0}$ do better (black solid lines), and SPCA and rSVD$_{\ell_1}$ do worse (red dashed lines).



Controlling sparsity with γ. Dependence of cardinality on the value of the sparsity-inducing parameter γ. The horizontal axis shows a normalized interval of reasonable values of γ. The vertical axis shows percentage of nonzero coefficients of the resulting sparse loading vector z^*.



How does the trade-off evolve in time?. Evolution of the explained variance (solid lines and left axis) and cardinality (dashed lines and right axis) in time for the methods GPower$_{\ell_1}$ and rSVD$_{\ell_1}$ on a test problem of size $p = 250$ and $n = 2500$.

Speed (in seconds):

$p \times n$	250×2500	500×5000	750×7500	1000×10000
GPower$_{\ell_1}$	0.85	2.61	3.89	5.32
GPower$_{\ell_0}$	0.46	1.21	2.41	2.93
SPCA	2.77	14.0	41.0	81.6
rSVD$_{\ell_1}$	1.40	6.80	17.8	41.2
rSVD$_{\ell_0}$	1.33	6.20	15.4	36.3
$p \times n$	500×2000	500×4000	500×8000	500×16000
GPower$_{\ell_1}$	0.97	1.96	4.30	8.43
GPower$_{\ell_0}$	0.39	0.97	2.01	4.63
SPCA	7.37	11.4	22.4	44.6
rSVD$_{\ell_1}$	2.56	5.27	11.3	26.8
rSVD$_{\ell_0}$	2.30	4.70	10.3	23.8

Data sets (breast cancer cohorts):

Study	Samples (p)	Genes (n)	Reference
Vijver	295	13319	van de Vijver et al. [2002]
Wang	285	14913	Wang et al. [2005]
Naderi	135	8278	Naderi et al. [2006]
JRH-2	101	14223	Sotiriou et al. [2006]

Speed (in seconds):

	Vijver	Wang	Naderi	JRH-2
GPower$_{\ell_1}$	7.72	6.96	2.15	2.69
GPower$_{\ell_0}$	3.80	4.07	1.33	1.73
GPower$_{\ell_1,m}$	5.40	4.37	1.77	1.14
GPower$_{\ell_0,m}$	5.61	7.21	2.25	1.47
SPCA	77.7	82.1	26.7	11.2
rSVD$_{\ell_1}$	46.4	49.3	13.8	15.7
rSVD$_{\ell_0}$	46.8	48.4	13.7	16.5

PEI-values based on 536 cancer-related pathways:

	Vijver	Wang	Naderi	JRH-2
PCA	0.0728	0.0466	0.0149	0.0690
GPower$_{\ell_1}$	**0.1493**	0.1026	0.0728	0.1250
GPower$_{\ell_1}$	0.1250	0.1250	0.0672	0.1026
GPower$_{\ell_1,m}$	0.1418	0.1250	**0.1026**	**0.1381**
GPower$_{\ell_0,m}$	0.1362	**0.1287**	0.1007	0.1250
SPCA	0.1362	0.1007	0.0840	0.1007
rSVD$_{\ell_1}$	0.1213	0.1175	0.0914	0.0914
rSVD$_{\ell_0}$	0.1175	0.0970	0.0634	0.1063

Pathway Enrichment Index (PEI) measures the statistical significance of the overlap between two kinds of gene sets.

[1] H. Zou, T. Hastie, R. Tibshirani. ”Sparse Principal Component Analysis”. Journal of Computational and Graphical Statistics, 15(2):265–286, 2006.
[2] A. d'Aspremont, F. R. Bach, L. El Ghaoui. ”Optimal Solutions for Sparse Principal Component Analysis”. Journal of Machine Learning Research, 9:1269–1294, 2008.
[3] H. Shen, J. Z. Huang. Sparse principal component analysis via regularized low rank matrix approximation”. Journal of Multivariate Analysis, 99(6):1015–1034, 2008.