Combinatorial Sparse PCA Beyond the Spiked Identity Model

Syamantak Kumar Purnamrita Sarkar

Kevin Tian Peiyuan Zhang

UT Austin and UW–Madison

Can we get fast sparse PCA guarantees under general covariance?

Sparse PCA: the model gap

Spiked identity model

$$\Sigma = 0.9(I - vv^\top) + vv^\top, \qquad \mathrm{nnz}(v) \le s.$$

- Isotropic background.
- Fast combinatorial algorithms are well understood.
- Examples: diagonal thresholding, covariance thresholding.

General covariance model

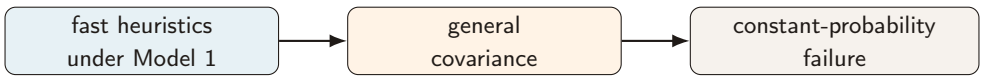
$$\lambda_2(\Sigma) \le (1-\gamma)\lambda_1(\Sigma), \qquad \mathrm{nnz}(v_1(\Sigma)) \le s.$$

- Minimal well-posedness assumptions: sparse top vector and relative eigengap γ.
- SDP-based methods handle this, but are expensive.
- Fast methods lacked global guarantees.

Takeaway. The general model only assumes a sparse top eigenvector and a relative eigengap γ, but this extra freedom breaks many model-specific heuristics.

What breaks beyond spiked identity?

Method	What it does	Counterexample intuition
Diagonal thresholding	choose coordinates with largest variances	the true support need not have the largest diagonal entries
Covariance thresholding	threshold entries of $\widehat{\Sigma}$ and take a top eigenvector	thresholding can expose a larger distractor block
Greedy correlation	grow support from row correlations	even a true seed can correlate more with fake coordinates



Takeaway. The paper gives explicit general-covariance instances that defeat each standard combinatorial rule.

Our algorithm: restart truncated power

RTPM

1. Form $\widehat{\Sigma} = n^{-1} \sum_i x_i x_i^\top$.
2. For every coordinate j, initialize $u_0^{(j)} = e_j$.
3. Iterate
$$u_t^{(j)} = \frac{\operatorname{top}_r(\widehat{\Sigma} u_{t-1}^{(j)})}{\| \operatorname{top}_r(\widehat{\Sigma} u_{t-1}^{(j)}) \|_2}.$$
4. Return the restart with largest $\langle u_T^{(j)}, \widehat{\Sigma} u_T^{(j)} \rangle$.

Why restarts help

Since v_1 is s-sparse, some coordinate e_j has

$$|\langle e_j, v_1 \rangle| \geq 1/\sqrt{s}.$$

Trying every j converts a local truncated-power guarantee into a global one.

Why oversample

Use $r \gg s$ during the power iterations so early truncation does not discard weak signal.

Main guarantee and proof idea

Informal theorem

Under the general model with constant gap, RTPM returns u with

$$\langle u, v_1(\Sigma)\rangle^2 \geq 0.9$$

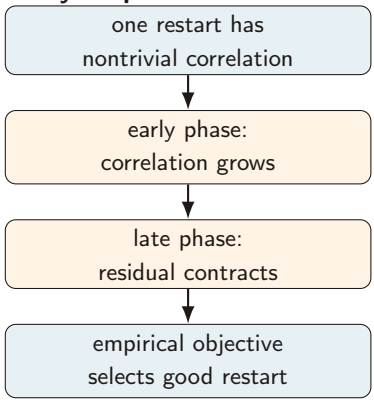
using

$$n = \Omega(s^3 \log(d/\delta)), \qquad r = \Omega(s^2),$$

$$T = \Omega(\log s).$$

Runtime: $O(nd^2T)$.

Analysis spine



Takeaway. The main technical step is a global convergence analysis of truncated power under general covariance.

Proof stage I: correlation growth

Informal stage theorem

Let

$$\Phi_t = \|V^\top u_t\|_2.$$

If the current iterate has weak but nontrivial signal,

$$\Phi_{t-1} \in \left[\Omega(1/\sqrt{s}),\, 1/\sqrt{2}\right],$$

then for a universal constant $c > 0$, a truncated power step gives

$$\Phi_t \geq (1 + c\gamma/k)\,\Phi_{t-1}.$$

Why this starts globally

- Because v_1 is s-sparse, one basis vector has correlation at least $1/\sqrt{s}$.
- RTPM tries all d basis vectors.
- Oversampling, $r \gg s$, prevents early truncation from deleting the signal.

Takeaway. Stage I turns a coordinate-level warm start into constant correlation.

Proof stage II: residual contraction

Informal stage theorem

Once the iterate has entered the high-correlation regime,

$$\Phi_{t-1} \geq 1/\sqrt{2},$$

for a universal constant $c > 0$, the residual contracts:

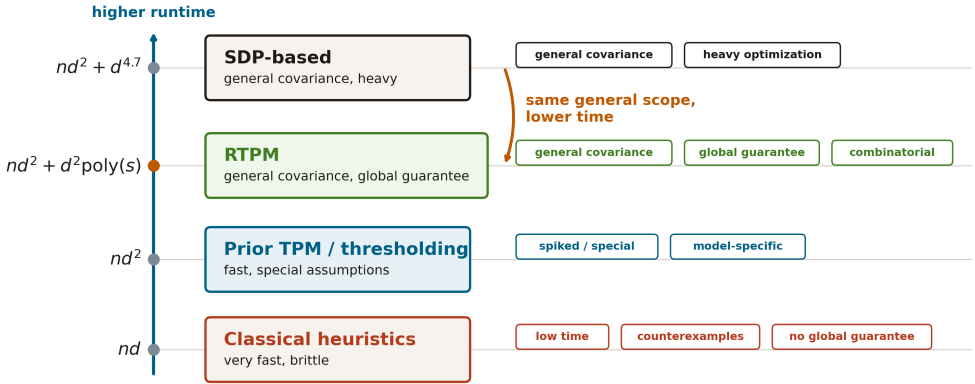
$$1 - \Phi_t^2 \leq (1 - c\gamma/k)\,(1 - \Phi_{t-1}^2).$$

From correlation to output

- The proof tracks population correlation, which RTPM cannot observe.
- The algorithm selects the restart with largest empirical quadratic form.
- Concentration converts this selection rule into a population guarantee.

Takeaway. Stage II upgrades constant correlation to the target error, then the empirical objective finds a good restart.

Time complexity landscape



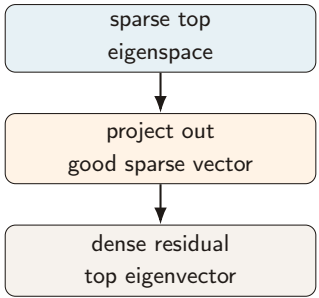
Takeaway. RTPM moves general-covariance sparse PCA down from SDP-style costs into combinatorial time.

One more lesson: deflation is subtle

Natural sparse k-PCA idea

1. recover one sparse component;
2. project it out;
3. recurse on the residual matrix.

Barrier. Even if the original top-2 eigenspace is 2-sparse, projecting out a highly accurate 3-sparse approximation to v_1 can make the residual leading eigenvector fully dense.



Takeaway. The paper solves sparse 1-PCA under general covariance and opens a careful path toward provable sparse k-PCA.

What to remember

1. **Model robustness matters:** standard fast sparse PCA heuristics can fail beyond spiked identity.
2. **RTPM is globally convergent:** restarts plus oversampling give a combinatorial method for general covariance.
3. **The cost is near the right computational regime:** $d^2 \cdot \mathrm{poly}(s, \log d)$ time, much lighter than SDP machinery.

Fast sparse PCA does not have to stop at the spiked identity model.

Thank you.

Backup: formal theorem shape

Under the sparse k-PCA model, with top-k eigenspace V, collective support size s, relative eigengap γ, and target error Δ, RTPM returns an r-sparse unit vector u with

$$\|V^\top u\|_2^2 \geq 1 - \Delta$$

when

$$n = \widetilde{\Omega}\left[\left(\frac{\sigma}{\lambda_k(\Sigma)}\right)^2 \left(\frac{s^3k^6}{\Delta^6\gamma^6} + \frac{s^2k^6}{\Delta^{10}\gamma^{10}}\right)\right],$$

$$r = \Omega\left(\frac{s^2k^2}{\Delta^2\gamma^2} + \frac{sk^2}{\Delta^4\gamma^4}\right), \qquad T = \Omega\left(\frac{k\log(s/(\Delta\gamma))}{\Delta\gamma}\right).$$