Power Iteration and Its Block Generalization

From the spectral theorem to randomized subspace recovery

Three questions

1. When does repeated multiplication recover a top eigenvector?
2. Why does a random Gaussian start solve the unknown-alignment problem?
3. What should we recover when the leading eigenvectors are not individually identifiable?

One mechanism answers all three: powers amplify ratios of eigenvalues.

Spectral coordinates

Spectral theorem

For every real symmetric $A \in \mathbb{R}^{n \times n}$, there are an orthogonal matrix $U = [u_1 \ \cdots \ u_n]$ and a real diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ such that

$$A = U \Lambda U^{\mathsf{T}}, \qquad A u_i = \lambda_i u_i.$$

Orthogonality
$U^{\mathsf{T}} U = I$, so the u_i form an orthonormal basis.

Ordering for power iteration
$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$.

Convention: eigenvectors are columns of U, so the factorization is $U \Lambda U^{\mathsf{T}}$.

Every start is a mixture

For any $z \in \mathbb{R}^n$, orthonormality gives the unique expansion

$$z = \sum_{i=1}^{n} \alpha_i u_i, \qquad \alpha_i = u_i^{\mathsf{T}} z.$$

Applying A changes the coefficients, not the directions:

$$Az = \sum_{i=1}^{n} \lambda_i \alpha_i u_i, \qquad A^m z = \sum_{i=1}^{n} \lambda_i^m \alpha_i u_i.$$

In the eigenbasis, multiplying by A is coordinatewise scaling.

Power iteration is one line

Algorithm

Choose $v_0 = z/\|z\|_2$. For $t = 0, 1, \ldots, m-1$:

$$y_{t+1} = Av_t, \qquad v_{t+1} = \frac{y_{t+1}}{\|y_{t+1}\|_2}.$$

In exact arithmetic,

$$v_m = \frac{A^m z}{\|A^m z\|_2}.$$

Normalization prevents overflow and underflow; it does not change the direction.

Cost per step: one matrix-vector product plus one normalization.

What does “usually” require?

To converge to one top eigenvector, two facts must hold:

A unique dominant magnitude

$$|\lambda_1| > |\lambda_2|, \qquad \rho := \frac{|\lambda_2|}{|\lambda_1|} < 1.$$

A visible dominant component

$$\alpha_1 = u_1^{\mathsf{T}} z \neq 0.$$

If either fails, no amount of iteration can create the missing information.

If $\lambda_1 < 0$, the normalized iterates alternate sign; the one-dimensional subspace still converges.

Powers preserve directions

With $z = \sum_i \alpha_i u_i$,

$$A^m z = U \Lambda^m U^{\mathsf{T}} z$$
$$= \sum_{i=1}^{n} \lambda_i^m \alpha_i u_i.$$

Each eigenvector is an invariant direction. The coefficient ratio between u_i and u_1 becomes

$$\frac{\lambda_i^m \alpha_i}{\lambda_1^m \alpha_1} = \left(\frac{\lambda_i}{\lambda_1}\right)^m \frac{\alpha_i}{\alpha_1}.$$

Iteration turns a spectral ratio into an exponential filter.

Factor out the winner

Assume $\alpha_1 \neq 0$ and define

$$\widetilde{w}_m := \frac{A^m z}{\lambda_1^m \alpha_1}.$$

Then

$$\widetilde{w}_m = u_1 + \sum_{i>1} \left(\frac{\lambda_i}{\lambda_1}\right)^m \frac{\alpha_i}{\alpha_1} u_i.$$

By orthogonality, the error has the exact squared norm

$$\|\widetilde{w}_m - u_1\|_2^2 = \sum_{i>1} \left|\frac{\lambda_i}{\lambda_1}\right|^{2m} \left|\frac{\alpha_i}{\alpha_1}\right|^2.$$

The eigengap sets the clock

Write the relative gap as

$$\rho = \frac{|\lambda_2|}{|\lambda_1|} \leq 1 - \gamma, \qquad 0 < \gamma \leq 1.$$

Since $(1-\gamma)^m \leq e^{-\gamma m}$,

$$\|\tilde{w}_m - u_1\| \leq e^{-\gamma m} C,$$

where $C = \|\alpha_{2:n}\|_2 / |\alpha_1|$. To achieve $\|\tilde{w}_m - u_1\| \leq \delta$, it is enough that

$$\boxed{m \geq \frac{\log C + \log(1/\delta)}{\gamma}}$$

A small gap makes one eigenvector intrinsically hard to identify.

Normalization loses only scale

Suppose an unavailable rescaling satisfies

$$\|\widetilde{w} - u_1\|_2 \le \delta.$$

The reverse triangle inequality gives

$$\left| \|\widetilde{w}\|_2 - 1 \right| \le \delta.$$

Therefore the computable normalized vector obeys

$$\begin{aligned} \left\| \frac{\widetilde{w}}{\|\widetilde{w}\|_2} - u_1 \right\|_2 &\le \left\| \frac{\widetilde{w}}{\|\widetilde{w}\|_2} - \widetilde{w} \right\|_2 + \|\widetilde{w} - u_1\|_2 \\ &\le 2\delta. \end{aligned}$$

Two quick checks

Negative leader

$A = \mathrm{diag}(-4, 2)$, $z = (1, 1)^{\mathsf{T}}$.

$$v_m \to (-1)^m e_1$$

Invisible leader

$A = \mathrm{diag}(4, 2)$, $z = e_2$.

$$A^m z = 2^m e_2$$

Which problem is fixed by randomness, and which is not?

Randomness fixes missing alignment with probability one; it cannot create a spectral gap.

Failure modes are informative

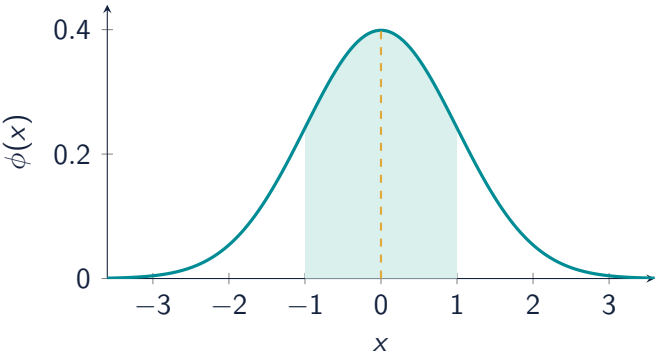
Symptom	Mathematical cause	Better target or remedy
No top component	$u_1^{\mathsf{T}} z = 0$	Random restart
Sign alternation	$\lambda_1 < 0$	Compare one-dimensional subspaces
No vector limit	$\lvert\lambda_1\rvert = \lvert\lambda_2\rvert$	Recover the dominant subspace
Slow progress	$\lvert\lambda_2/\lambda_1\rvert \approx 1$	Accelerate, shift, or change target

Convergence failures often say that the requested object is not identifiable.

A random start

Rotational invariance removes the alignment assumption

The standard normal

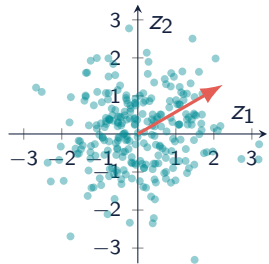


$$g \sim \mathcal{N}(0,1)$$

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

- symmetric about 0
- constant-scale coordinates
- continuous: $\Pr[g = 0] = 0$

An isotropic Gaussian cloud



For $z \sim \mathcal{N}(0, I_2)$:

- level sets are circles;
- every unit direction has the same projected law;
- rotating the axes does not change the cloud.

Rotations preserve the law

Let $z \sim \mathcal{N}(0, I_n)$ and let U be orthogonal. Then

$$\mathbb{E}[Uz] = 0, \qquad \mathrm{Cov}(Uz) = UIU^{\mathsf{T}} = I.$$

Because a multivariate Gaussian is determined by its mean and covariance,

$$\boxed{Uz \sim \mathcal{N}(0, I_n).}$$

In the unknown eigenbasis,

$$\alpha = U^{\mathsf{T}} z \sim \mathcal{N}(0, I_n),$$

so $\alpha_1, \ldots, \alpha_n$ are again independent standard Gaussians.

A Gaussian start is equally random in every orthonormal basis.

A random start is useful

For $z \sim \mathcal{N}(0, I_n)$, with constant probability simultaneously,

$$|\alpha_1| \geq 0.1, \qquad \|\alpha_{2:n}\|_2 \leq 2\sqrt{n}.$$

Hence the starting-condition factor satisfies

$$C(z, u_1) = \frac{\|\alpha_{2:n}\|_2}{|\alpha_1|} \leq 20\sqrt{n}.$$

Combining with the gap bound gives

$$m = O\left(\frac{\log n + \log(1/\delta)}{\gamma}\right).$$

A failed start is cheap: restart independently or use a small random block.

Alignment costs a logarithm

The deterministic bound is

$$m \geq \frac{\log C + \log(1/\delta)}{\gamma}.$$

Even if $|\alpha_1|$ is polynomially small, C is polynomially large and $\log C = O(\log n)$.

Power iteration is forgiving about the start, but unforgiving about the gap.

Eigenvalue only

Why value estimation is easier than vector recovery

Rayleigh gives the value

For $v = \sum_i \beta_i u_i \neq 0$, define the Rayleigh quotient

$$R_A(v) := \frac{v^T A v}{v^T v}, \qquad R_A(v) = \sum_{i=1}^{n} \lambda_i \beta_i^2 \quad \text{when } \|v\|_2 = 1.$$

Thus $R_A(v)$ is a weighted average of the eigenvalues. For positive semidefinite A, $0 \leq R_A(v) \leq \lambda_1$.

Concentrate mass near the top eigenspace to estimate the top value.

Top clusters are enough

Assume $A \succeq 0$ and fix a relative accuracy $\varepsilon \in (0, 1)$. Define

$$S_\varepsilon := \{j : \lambda_j \geq (1 - \varepsilon)\lambda_1\}.$$

The eigenvalues inside S_ε may be arbitrarily close or equal. All we need is that the normalized iterate places little squared mass outside S_ε:

$$\sum_{j \notin S_\varepsilon} \beta_j^2 \leq \varepsilon.$$

Then every direction carrying most of the mass already has nearly optimal value.

Vector recovery asks "which direction?" Value recovery asks only "how large?"

The tail loses energy

Let

$$w_m := \frac{A^m z}{\lambda_1^m} = \sum_j \left(\frac{\lambda_j}{\lambda_1}\right)^m \alpha_j u_j.$$

For $j \notin S_\varepsilon$, $\lambda_j/\lambda_1 < 1 - \varepsilon$, so

$$\sum_{j \notin S_\varepsilon} |u_j^{\mathsf{T}} w_m|^2 \le e^{-2\varepsilon m} \|z\|_2^2 .$$

Meanwhile the top component does not decay:

$$|u_1^{\mathsf{T}} w_m|^2 = |\alpha_1|^2.$$

Thus tail mass is at most an ε fraction once

$$m \gtrsim \frac{\log(\|z\|_2 / |\alpha_1|) + \frac{1}{2}\log(1/\varepsilon)}{\varepsilon}.$$

A cluster suffices

Let $v_m = w_m / \|w_m\|_2$ and write $v_m = \sum_j \beta_j u_j$. If the tail mass is at most ε, then

$$\begin{aligned} R_A(v_m) &= \sum_j \lambda_j \beta_j^2 \\ &\geq (1-\varepsilon)\lambda_1 \sum_{j \in S_\varepsilon} \beta_j^2 \\ &\geq (1-\varepsilon)^2 \lambda_1 \geq (1-2\varepsilon)\lambda_1. \end{aligned}$$

Always $R_A(v_m) \leq \lambda_1$, hence

$$(1-2\varepsilon)\lambda_1 \leq R_A(v_m) \leq \lambda_1.$$

Indefinite matrices need care

For a general symmetric A, power iteration follows the largest *magnitude* eigenvalue.

Estimate the spectral norm

Apply the method to $B = A^2 \succeq 0$:

$$\lambda_{\max}(B) = \|A\|_2^2 .$$

One B-product is $x \mapsto A(Ax)$.

Recover a signed eigenvalue

Use $v^{\mathsf{T}} A v$ only after v resolves a single eigendirection.
Equal-magnitude $+\lambda$ and $-\lambda$ modes can mix.

Squaring preserves magnitude but discards sign.

A whole subspace

Block power iteration and invariant-subspace recovery

Near-ties change the target

Suppose

$$\lambda_1 \approx \lambda_2 \approx \cdots \approx \lambda_k, \qquad \lambda_{k+1} < (1-\gamma)\lambda_k,$$

for a positive semidefinite A. Individual vectors $u_1, \ldots, u_k$ may rotate dramatically under tiny perturbations, but the invariant subspace

$$\mathcal{U}_1 = \mathrm{span}\{u_1, \ldots, u_k\}$$

can still be stable and identifiable.

Place the gap after the object you want: between λ_k and λ_{k+1}.

Block power iteration

Start with $V_0 \in \mathbb{R}^{n \times k}$ having orthonormal columns. For $t = 0, \ldots, m-1$:

$$Y_{t+1} = AV_t, \qquad V_{t+1} = \text{gs}(Y_{t+1}).$$

Here $\text{gs}(Y)$ is a result of applying Gram-Schmidt orthogonalization to Y. As a V_{t+1} is a matrix with orthonormal columns, but the same column span as Y_{t+1}, and therefore

$$\text{range}(V_m) = \text{range}(A^m V_0).$$

Gram-Schmidt orthogonalization

Given vectors $y_1, y_2, \ldots y_k$, produce a sequence of orthonormal vectors $v_1, \ldots v_k$, s.t. for all j, $\text{span}\{v_1, \ldots v_j\} = \text{span}\{y_1, \ldots y_j\}$.

Algorithm

For $t = 1, \ldots k$:

$$\tilde{v}_j := y_t - \sum_{i<t} \langle y_t, v_j \rangle$$
$$v_j := \tilde{v}_j / \|\tilde{v}_j\|$$

Subspace similarity

Let $U, V \in \mathbb{R}^{n \times k}$ have orthonormal columns spanning $\mathcal{U}$ and $\mathcal{V}$. Define

$$I(\mathcal{U}, \mathcal{V}) := \frac{1}{k} \left\| U^{\mathsf{T}} V \right\|_F^2 = \sum_{i,j \le k} \langle u_i, v_j \rangle^2,$$

where u_i are columns of U, and v_j are columns of V.

$$I = 1 \iff \mathcal{U} = \mathcal{V}. \qquad\qquad I = 0 \iff \mathcal{U} \perp \mathcal{V}.$$

Note that this is indeed well-defined for two subspaces of $\mathcal{U}$ and $\mathcal{V}$: this value do not depend on the choice of basis U and V.

Work in the eigenbasis

Assume $A \succeq 0$ and split

$$A = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \qquad V_0 = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix},$$

where $\Lambda_1 \in \mathbb{R}^{k \times k}$ contains $\lambda_1, \ldots, \lambda_k$. Assume C_1 is nonsingular. For a Gaussian start this holds with probability one. Then

$$A^m V_0 = \begin{bmatrix} \Lambda_1^m C_1 \\ \Lambda_2^m C_2 \end{bmatrix}.$$

C_1 measures how well the starting block sees the top-k subspace.

The iterate is a graph

Right multiplication by an invertible matrix does not change a column span. Multiply $A^m V_0$ by $C_1^{-1}\Lambda_1^{-m}$:

$$\operatorname{range}(A^m V_0) = \operatorname{range}\begin{bmatrix} I \\ E_m \end{bmatrix},$$

where

$$E_m := \Lambda_2^m C_2 C_1^{-1} \Lambda_1^{-m}.$$

The exact top subspace is $\operatorname{range}\left[\begin{smallmatrix} I \\ 0 \end{smallmatrix}\right]$. Thus E_m is the "tilt" of the computed subspace away from the target.

The block gap contracts

Let

$$\rho_k := \frac{\lambda_{k+1}}{\lambda_k} < 1.$$

Submultiplicativity gives

$$\begin{aligned} \|E_m\|_F &\le \|\Lambda_2^m\|_2 \left\|C_2 C_1^{-1}\right\|_F \left\|\Lambda_1^{-m}\right\|_2 \\ &\le \rho_k^m \left\|C_2 C_1^{-1}\right\|_F . \end{aligned}$$

If $\rho_k \le 1 - \gamma$, then $\|E_m\|_F \le e^{-\gamma m} \left\|C_2 C_1^{-1}\right\|_F =: \delta \ll 1$.

Scalar power iteration reappears with the gap moved from 1 to k.

Graph error controls similarity

Let $V := \begin{bmatrix} I \\ E_m \end{bmatrix}$, and $U_0 := \begin{bmatrix} I \\ 0 \end{bmatrix}$, is a matrix columns of which spans the top k eigenspace. We would like to show that $I(\text{range } V, \text{range } U_0) \approx 1$.

Orthormalization?

Matrix $W := V(V^T V)^{-1/2}$ is an orthonormal matrix with $\text{range}(W) = \text{range}(V)$.

Indeed $WW^T = V(V^T V)^{-1} V^T = VV^{-1}(V^T)^{-1} V^T = I$.
The $\text{range}(W) = \text{range}(V)$ is immediate.

Graph error controls similarity

$$I(\text{range } V, \text{range } U_0) = I(W, U_0) = \frac{1}{k}\|WU_0^T\|_F^2$$
$$= \frac{1}{k}\|V(V^TV)^{-1/2}U_0^T\|_F^2$$

Since $V = \begin{bmatrix} I \\ E_m \end{bmatrix}$, we have $V^TV = I + E_m^TE_m$, hence

$$\|V^TV - I\| = \|E_m^TE_m\| \leq \|E_m\|_F^2 \leq \delta^2,$$

and

$$(V^TV)^{-1/2} = I + \tilde{E},$$

where $\|\tilde{E}\| \leq O(\delta^2)$.

Graph error controls similarity

$$
\begin{aligned}
I(\text{range } V, \text{range } U_0) &= \frac{1}{k}\| V(V^T V)^{-1/2} U_0^T \|_F^2 \\
&= \frac{1}{k}\| V(I + \tilde{E}) U_0^T \|_F^2 \\
&= \frac{1}{k}\| VU_0^T + V\tilde{E}U_0^T \|_F^2
\end{aligned}
$$

Now $\| VU_0^T \|_F^2 = k$, and $V\tilde{E}U_0$ is small (since $\tilde{E}$ already is small). Indeed

$$
|\|A + B\|_F^2 - \|A\|_F^2| \leq \|B\|_F(\|A + B\|_F + \|A\|_F),
$$

so

$$
I(\text{range } V, \text{range } U_0) \geq 1 - 3\| V\tilde{E}U_0^T \|_F \geq 1 - O(k^3\delta^2).
$$