Background Separation in Video Via Robust PCA

Data Analytics - Background Separation Module
Problem motivation

- Given a sequence of surveillance video frames, identify “actions” that stand out from the background.

- First step and the focus of this module: separate the background from the foreground.

- One possible approach, in line with many other modules in this course.
  - Statistical model for background, model for movements, occlusion, geometry arising from perspective view, etc.

- Instead, we shall take a more naive approach.

- But first a couple of videos illustrating the problem. (‘Subway’ and ‘restaurant’ videos)
The main idea

- Vectorise each frame into a column of numbers.
- Stack columns into a matrix.
- If camera does not move, if background is still, we expect to see
  \[ L = [v \ v \ v \ \cdots] \]
  
  \( L \) is a rank 1 matrix.
  Let us assume rank \( r \); captures slow background variations.

- With foreground movement, there can be occlusions of the background.

  \[ M = L + S \]

  \( S \) captures foreground variations across the frames. If movement is limited to a small region, \( S \) is sparse, i.e., very few nonzero entries, but don’t know where, and the nonzero entries can be arbitrary.

- Problem: Given \( M = L + S \), decompose into \( L \) and \( S \).
First try: Principal Component Analysis

- Minimise the following:

\[
\min ||M - L||_{op} \\
\text{subject to } \text{rank}(L) \leq r.
\]

- Here $||A||_{op}$ is the operator norm of $A$ and equals the largest singular value of $A$:

\[
||A||_{op} := \max_{x:x \neq 0} \frac{||Ax||_2}{||x||_2} = \sigma_1(A).
\]
PCA solution

- Solution: Obtain the singular value decomposition; pick the first $r$.

- Singular value decomposition:

$$M = U\Sigma V^T = [u_1 u_2 \cdots u_r] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix} = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

- $M$ is of size $m \times n$. $U$ is the matrix of left singular vectors. $V$ is the matrix of right singular vectors.

- First $r$ components: $\sum_{i=1}^{r} \sigma_i u_i v_i^T$.

- Works very well when $S$ is diffuse; a small perturbation of $L$. Also, it’s the maximum likelihood estimate when the entries of $S$ are random, i.i.d. Gaussian.

- If some pixels are grossly corrupted, won’t work well. This is the norm for us: foreground occludes parts of the background.
Relaxing the rank constraint

- If we do not know the rank ...
- Try rank 1, then rank 2, and so on, until all ‘significant’ components have been captured.
- There is a natural way to do this that also encourages sparsity in the number of components.
- The nuclear norm of a matrix:

\[
\|A\|_{nuc} = \min \{m,n\} \sum_{i=1}^{\min \{m,n\}} \sigma_i(A).
\]

- Relax the problem to

\[
\begin{align*}
\min & \quad \|M - L\|_{op} \\
\text{subject to} & \quad \|L\|_{nuc} \leq \tau
\end{align*}
\]

\[
\begin{align*}
\equiv & \quad \min \max_i \sigma_i(M - L) \\
\text{subject to} & \quad \|L\|_{nuc} \leq \tau
\end{align*}
\]
Encouraging sparsity in the entries

We would like to encourage the entries of the solution $S = M - L$ to be at the extreme points, in particular 0.

Thus

\[
\min \|M - L\|_1 \\
\text{subject to } \|L\|_{nuc} \leq \tau.
\]

Lagrangian relaxation of this problem is to minimise the following for a suitable weight parameter $\lambda$:

\[
\min [\|L\|_{nuc} + \lambda\|M - L\|_1]
\]

This encourages sparsity in the number of components (via nuclear norm of $L$) as well as sparsity in the number of nonzero entries of $S = M - L$ (via the 1-norm).
Discussion - Can we really recover $L$ and $S$?

- $M = e_1e_1^T$. It is both low-rank and sparse. Is this part of $L$ or $S$?

- For the 'recovery' problem to make sense, we need the low rank part to be 'diffuse' or 'incoherent'.

**Definition**

We say that a matrix $L$ is $\mu$-incoherent if the SVD $L = U\Sigma V^T$ satisfies the following:

$$
\|U^T e_i\|_2 \leq \frac{\mu \sqrt{r}}{\sqrt{m}} \quad i = 1, \ldots, m,
$$

$$
\|V^T e_j\|_2 \leq \frac{\mu \sqrt{r}}{\sqrt{n}} \quad j = 1, \ldots, n,
$$

where $L$ has dimensions $m \times n$ and has rank $r$.

- Sum of squares of all entries of $U$ is $r$. If this is spread out equally across rows, then each row has energy $r/m$ or norm $\sqrt{r/m}$. The above says there are no 'heavy-weight' rows.
Size of entries in $UV^T$

- To assess the sizes of entries in $UV^T$ ...

- if all singular values are the same, this would provide some measure the spread of entries of the low rank matrix.

\[
\|UV^T\|_\infty = \max_{i,j} |e_i^T UV^T e_j| \\
= \max_{i,j} |\langle U^T e_i, V^T e_j \rangle| \\
\leq \max_{i,j} \|U^T e_i\|_2 \cdot \|V^T e_j\|_2 \\
\leq \frac{\mu^2 r}{\sqrt{mn}} \quad \text{(by $\mu$-incoherence).}
\]
A surprising result (Candes, Li, Ma, Wright 2011)

- Impose $\mu$-incoherence on $L$ and additionally $\|UV^T\|_\infty \leq \frac{\mu\sqrt{r}}{\sqrt{mn}}$.

- Some mild randomness on the sparsity. Let $S_0$ be an arbitrary matrix. Identify (uniformly at random) a subset of $c$ entries. $S$ equals $S_0$ on these entries and is zero outside.

**Theorem**

Suppose $L$ is $\mu$-incoherent. Suppose further that $\|UV^T\|_\infty \leq \frac{\mu\sqrt{r}}{\sqrt{mn}}$. Let $S$ be any matrix whose support is uniformly distributed among sets of cardinality $c$.

There exist positive numerical constants $\rho_r$, $\rho_s$, and $\nu$ such that if $\text{rank}(L) \leq \rho_r m/(\mu \log n)^2$, if $c \leq \rho_s mn$, then with $\lambda = 1/\sqrt{n}$, the solution to

$$\min \ [\|L\|_{nuc} + \lambda \|M - L\|_1]$$

recovers $L$ and $S$ exactly with probability at least $1 - \nu/n^{10}$. 
Remarks

- A convex optimisation problem, ready-made tools available.

- Rank of $L$ can be quite large, as high as $n/(\log n)^2$, if $\mu$ is of the order of a constant.

- A fixed parameter $\lambda = 1/\sqrt{n}$ works. No tuning based on how many sparse entries, level of incoherence, etc., which one might anticipate is needed to balance the nuclear norm and sparsity objectives.

- The optimisation takes some computational effort (cubic). The main point is that exact recovery is possible under suitable assumptions. Perhaps one of you can take this up as a project.

- We will study an alternative method, a very natural one, and you will implement it.
An alternating projection approach

- $L$ is a low rank matrix, has rank $\leq r$.
  - Getting a low rank approximation of a matrix is relatively easy. Use SVD.

- $S$ is sparse.
  - Getting a sparse approximation of a matrix is also easy. Hard threshold at a suitable level and keep only the large values.

- So here’s a natural algorithm.
  (i) Start with the lowest rank approximation. $L^0 = 0$.
  (ii) Hard threshold $M$ to get $S^0$, a sparse matrix.
  (iii) Get a low rank approximation $L^1$ of $M - S^0$.
  (iv) Hard threshold $M - L^1$ to get $S^1$.
Repeat until convergence. [Picture on the board.]

- Some careful tweaking of thresholds needed; we will discuss them. (Netrapalli et al. 2014).
Notation

- $H_\tau(A)$ indicates hard-thresholding a matrix $A$ at level $\tau$.

- $P_r(A)$ indicates projection of a matrix $A$ into the space of matrices with rank $r$ or lower.

- $M$: Given matrix $L + S$ of size $m \times n$.
  - $\varepsilon$: convergence parameter.
  - $r$: rank of $L$.
  - $\beta$: a tuning parameter associated with the thresholding.

- $\hat{L}, \hat{S}$: estimated low rank and sparse components of given $M$. 
The Alternating Projection Algorithm: ALTPROJ

- Input: Matrix $M$, accuracy $\varepsilon$, rank $r$, tuning parameter $\beta$.
- Output: $\hat{L}$, $\hat{S}$.
- Initialise: $L^0 = 0$, $\tau_0 = \beta \sigma_1(M)$, $S^0 = H_{\tau_0}(M - L^0)$.

for ‘stage’ $k = 1$ to $r$ do:
  $T := 10 \log_2(n\beta \|M - S^0\|_{op}/\varepsilon)$
  for ‘iteration’ $t = 0$ to $T$ do:
    $\tau := \beta(\sigma_{k+1}(M - S^t) + 2^{-t}\sigma_k(M - S^t))$
    $L^{t+1} := P_k(M - S^t)$
    $S^{t+1} := H_\tau(M - L^{t+1})$
  end for
  if $\beta\sigma_{k+1}(L^{t+1}) < \varepsilon/(2n)$ then
    return: $L^T$, $S^T$
  else
    $S^0 := S^T$
  end if
end for
return: $L^T$, $S^T$
Remarks on the algorithm

- $r = 1$: Threshold changes in each iteration. Initial harsh thresholding, but threshold decreases to allow for a larger $S^t$.

- After the first stage, residuals are of size $\sigma_1$. Do not enter stage 2 (rank 2 approximation) until a good quality $L^T$ and $S^T$ at this rank. When entering stage 2, set a threshold for the next level of target residuals.

- $\beta$ enables tuning for spikiness.

- Complexity:
  In each iteration: $P_k$ takes $O(kmn)$ (PCA).
  Number of iterations in each stage: $O(\log(1/\varepsilon)) + O(\log(n\beta||M||_{op}))$.
  Number of stages: $r$.
  Total: $O(r^2 mn(\log(1/\varepsilon) + \log(n\beta||M||_{op})))$. 
ALTPROJ’s performance (Netrapalli et al. 2014)

**Theorem**

Suppose $L$ has rank at most $r$ and $L$ is $\mu$-incoherent. Suppose that each row and column of $S$ has at most $\alpha$ fraction of nonzero entries, where

$$\alpha \leq \frac{1}{512\mu^2r}.$$  

Fix $\varepsilon$ and take $\beta = 4\mu^2r/\sqrt{mn}$. Then the outputs $L^T, S^T$ of ALTPROJ satisfy

$$||L - \hat{L}||_{\infty} \leq \frac{\varepsilon}{\sqrt{mn}}$$

$$||S - \hat{S}||_{\infty} \leq \frac{\varepsilon}{\sqrt{mn}}$$

$$\text{Supp}(\hat{S}) \subseteq \text{Supp}(S).$$
A comparison of the two results

(Candes et al. 2011)
- Stricter constraint on $\|UV^T\|_\infty \leq \mu \sqrt{r}/\sqrt{mn}$.
- Randomness in the support set.
- But exact recovery w.h.p.

(Netrapalli et al. 2014)
- Do not impose the stricter constraint on $\|UV^T\|_\infty$.
- No randomness in the support set. But sparsity required on each row and each column.
- Approximate recovery only, but via an easier algorithm.
Main steps in the proof of ALTPROJ’s performance

- Focus on the symmetric case $m = n$.
- Let $L$ have eigenvalues $\sigma_1, \sigma_2, \ldots, \sigma_r$, indexed so that
  \[ |\sigma_1| \geq |\sigma_2| \geq \cdots \geq |\sigma_r|. \]
- $S^t$ and $L^t$ are the $t$th iterates in stage $k$ (suppressed). $E^t := S - S^t$, error in the sparse matrix.
- $M - S^t = L + S - S^t = L + E^t$.
- Let $M - S^t = L + E^t$ have eigenvalues $\lambda_1, \ldots, \lambda_n$, indexed so that
  \[ |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|. \]
  (Both $t$ and $k$ suppressed. Let’s get comfortable with this.)
Small low-rank projection error ensures small sparsity error

Lemma (LS)

If

$$\|L^{t+1} - L\|_\infty \leq \frac{2\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-t} |\sigma_k|),$$

then

$$\text{Supp}(E^{t+1}) \subseteq \text{Supp}(S)$$

$$\|E^{t+1}\|_\infty \leq \frac{7\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-t} |\sigma_k|).$$
Small sparsity error ensures small low-rank projection error

Lemma (SL)

If

\[ \text{Supp}(E^t) \subseteq \text{Supp}(S) \]
\[ \| E^t \|_\infty \leq \frac{8\mu^2 r}{n}(|\sigma_{k+1}| + 2^{-t}|\sigma_k|), \]

then

\[ \| L^{t+1} - L \|_\infty \leq \frac{2\mu^2 r}{n}(|\sigma_{k+1}| + 2^{-t}|\sigma_k|). \]

Note that the constraint on \( \| E^t \|_\infty \) on the previous page was tighter. We will need it when we do the induction and jump across stages.
Proof for symmetric matrices

- Start off induction at $k = 1$ and $t = -1$.

- To show: $\|L^0 - L\|_\infty = \|L\|_\infty \leq \frac{2\mu^2r}{n}(2|\sigma_2| + 2|\sigma_1|)$.

- Use $\mu$-incoherence:

  $|e_i^T L e_j| = |e_i U \Sigma V^T e_j| = |\langle U^T e_i, \Sigma V^T e_j \rangle| \\ 
  \leq \|U^T e_i\|_2 \cdot \|\Sigma V^T e_j\|_2 \\ 
  \leq |\sigma_1| \cdot \mu^2 r/n.$

- This enables induction, establishes $\|E^t\|_\infty$ and $\|L - L^t\|_\infty$ bounds for all $t$ in stage $k = 1$.

- Also, if we can ensure validity in the move from $(k, T)$ to $(k + 1, 0)$, the bounds hold for all $t$ and $k$ until termination.
We have established, for a particular stage $k$, for its last iteration $t = T$,

\[ \text{Supp}(E^T) \subseteq \text{Supp}(S) \]

\[ \|E^T\|_\infty \leq \frac{7\mu^2r}{n}(\sigma_{k+1} + 2^{-T}\sigma_k). \]

Claim: If $\beta\sigma_{k+1}(L^T) < \varepsilon/(2n)$, then the algorithm terminates, and

\[ \|L - L^T\|_\infty \leq \varepsilon/n, \quad \|S - S^T\|_\infty \leq \varepsilon/n. \]

Claim: If $\beta\sigma_{k+1}(L^T) \geq \varepsilon/(2n)$, then

\[ \{\|L - L^T\|_\infty, \|S - S^T\|_\infty\} \leq \frac{\{2, 8\}\mu^2r}{n}(\sigma_{k+2} + 2\sigma_{k+1}). \]

Note the $k + 2$ and $k + 1$, $t = -1$. This enables continuation of induction in the next stage.

The lemmas and the claims help us complete the proof for the symmetric case.
Lemma

Let $A + E = B$.

$A$ has eigenvalues $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

$B$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Then

$$|\lambda_i - \sigma_i| \leq ||E||_{op} \quad \text{for each } i.$$

We will take this as granted.
Lemma

Suppose $S$ is $\alpha$-sparse. Then $\|S\|_{op} \leq \alpha n \|S\|_{\infty}$.

Can all the nonzero entries of $S$ conspire to have a large operator norm (order larger than $\alpha n$)? No.

Proof: For the left and right singular unit vectors $u^T$ and $v$ associated with the top singular value, we have

$$\|S\|_{op} = u^T S v = \sum_{i,j} u_i S_{i,j} v_j \leq \frac{1}{2} \sum_{i,j} (u_i^2 + v_j^2) S_{i,j}.$$ 

Now it’s clear that each summation encounters at most $\alpha n$ nonzero entries.
The number of iterations is sufficiently large

- The number of iterations $T$ in a stage is sufficiently large to drive the error in $\|E^T\|_{op}$ comparable to $|\sigma_{k+1}|$.

- $T = \log_2(n/\beta \|M - S^0\|_{op}/\varepsilon)$.

\[
\|M - S^0\|_{op} \geq \|L\|_{op} - \|E^0\|_{op} \\
\geq |\sigma_k| - \alpha n \|E^0\|_{\infty} \\
\geq |\sigma_k| - \alpha n \cdot \frac{7\mu^2 r}{n} (|\sigma_{k+1}| + 2|\sigma_k|) \\
\geq |\sigma_k| - (a \text{ small fraction}) |\sigma_k| \\
\geq (3/4) |\sigma_k|.
\]

- Thus $T \geq \log \left(n \cdot \left(\frac{4\mu^2 r}{n}\right) \cdot ((3/4)|\sigma_k|) / \varepsilon \right) = \log(3\mu^2 r |\sigma_k| / \varepsilon)$.

- This implies $2^{-T} \leq \varepsilon/(3\mu^2 r |\sigma_k|)$.
  Increase multiplier inside log and we can make this even smaller.
Bound on $||E^T||_\infty$

Since we have enough iterations,

$$||E^T||_\infty \leq \frac{7\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-T}|\sigma_k|)$$

$$\leq \frac{7\mu^2 r}{n} \left(|\sigma_{k+1}| + \frac{\varepsilon}{3\mu^2 r|\sigma_1||\sigma_k|}\right)$$

$$\leq \frac{7\mu^2 r}{n} |\sigma_{k+1}| + \frac{7\varepsilon}{3n}.$$

Increasing the factor inside log, the 2nd term is \textit{(small fraction)} $\varepsilon/n$.

We also have, by Weyl,

$$|\sigma_{k+1}(M - S^T) - \sigma_{k+1}| \leq ||E^T||_{op} \leq \alpha n \times \text{above expression}$$

$$\leq 7\alpha \mu^2 r|\sigma_{k+1}| + \text{(small fraction)} \varepsilon.$$

The two cases $n\beta|\sigma_{k+1}(M - S^T)| = 4\mu^2 r|\sigma_{k+1}(M - S^T)| \geq \varepsilon/2$ discussion.
Recall the two claims

- **Claim:** If $\beta \sigma_{k+1}(L^T) < \varepsilon/(2n)$, then the algorithm terminates, and
  \[
  \|L - L^T\|_\infty \leq \varepsilon/n, \quad \|S - S^T\|_\infty \leq \varepsilon/n.
  \]

- **Claim:** If $\beta \sigma_{k+1}(L^T) \geq \varepsilon/(2n)$, then
  \[
  \{\|L - L^T\|_\infty, \|S - S^T\|_\infty\} \leq \left\{\frac{2, 8}{n}\mu^2 r \right\} (|\sigma_{k+2}| + 2|\sigma_{k+1}|).
  \]

- The discussion establishes how both are valid.
Towards Lemma LS, proximity of eigenvalues

Lemma
Recall $L$ has eigenvalues $\sigma_1, \ldots, \sigma_n$ in decreasing order.
$M - S^t = L + E^t$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ in decreasing order.
Suppose $E$ satisfies the $E$-conditions (used in the induction).
We are in stage $k$, iteration $t$.
Then
\[(7/8)(|\sigma_{k+1}| + 2^{-t}|\sigma_k|) \leq (|\lambda_{k+1}| + 2^{-t}|\lambda_k|) \leq (9/8)(|\sigma_{k+1}| + 2^{-t}|\sigma_k|).\]

Proof: The above is the same as absolute value of the difference is not greater than $1/8$ times $(|\sigma_{k+1}| + 2^{-t}|\sigma_k|)$.

We already saw
\[|\lambda_k - \sigma_k| \leq \|E^t\|_{op} \leq \alpha n \|E^t\|_{\infty} \leq 8\mu^r r \alpha (|\sigma_k| + 2^{-t}|\sigma_{k-1}|).\]

Discussion on how to use this.
Recall: Small low-rank projection error ensures small sparsity error

**Lemma (LS)**

If

$$\|L^{t+1} - L\|_\infty \leq \frac{2\mu^2 r}{n}(|\sigma_{k+1}| + 2^{-t}|\sigma_k|),$$

then

$$\text{Supp}(E^{t+1}) \subseteq \text{Supp}(S)$$

$$\|E^{t+1}\|_\infty \leq \frac{7\mu^2 r}{n}(|\sigma_{k+1}| + 2^{-t}|\sigma_k|).$$
Proof of Lemma LS - support

- Support. \( S^{t+1} = H_\tau(M - L^{t+1}) = H_\tau(S + L - L^{t+1}) \).

- Suppose \( S_{i,j} = 0 \). We must show \( E_{ij}^{t+1} = 0 \).

- \( E_{ij}^{t+1} = S_{ij} - S_{ij}^{t+1} = -S_{ij}^{t+1} = (L_{ij} - L_{ij}^{t+1})1_\{|L_{ij} - L_{ij}^{t+1}| > \tau\} \).

- But this can’t hold by assumption on \( L \) and proximity of \( \lambda \) and \( \sigma \).
Proof of Lemma LS - S error is bounded

- \( S^{t+1} = H_\tau(M - L^{t+1}) = H_\tau(S + L - L^{t+1}) \).

- Suppose \( |M_{ij} - L_{ij}^{t+1}| > \tau \).
  - Then \( S_{ij}^{t+1} = S_{ij} + L_{ij} - L_{ij}^{t+1} \), hard-thresholding does not affect entry.
  - \( E_{ij}^{t+1} = S_{ij} - S_{ij}^{t+1} = -(L_{ij} - L_{ij}^{t+1}) \) which is small.

- Suppose \( |M_{ij} - L_{ij}^{t+1}| \leq \tau \).
  - Then \( S_{ij}^{t+1} = 0 \) and \(|S_{ij} + L_{ij} - L_{ij}^{t+1}| < \tau\), hard-thresholding zeros entry.
  - \( E_{ij}^{t+1} = S_{ij} - S_{ij}^{t+1} = S_{ij} \).
  - So \(|S_{ij}| \leq \tau + |L_{ij} - L_{ij}^{t+1}|\). 
    - \( \tau \) is bounded by \( 4 \times (\cdots) \) and \( ||L - L^{t+1}||_\infty \) is bounded by \( 2 \times (\cdot) \). 
    - So the \( 7 \times (\cdots) \) bound holds.
Recall: Small sparsity error ensures small low-rank projection error

**Lemma (SL)**

If

\[ \text{Supp}(E^t) \subseteq \text{Supp}(S) \]

\[ ||E^t||_\infty \leq \frac{8\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-t} |\sigma_k|), \]

then

\[ ||L^{t+1} - L||_\infty \leq \frac{2\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-t} |\sigma_k|). \]
An attempt

- Recall that $L^{t+1} = P_k(M - S^t) = P_k(L + E^t)$.

- If

$$M - S^t = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U_1 \Lambda_1 U_1^T + U_2 \Lambda_2 U_2^T$$

then

$$L^{t+1} = \sum_{i=1}^{r} \lambda_i u_i u_i^T = U_1 \Lambda_1 U_1^T.$$ 

- Thus $L - L^{t+1} = M - S - (M - S^t - U_2 \Lambda_2 U_2^T) = U_2 \Lambda_2 U_2^T - E^t$.

- Bounding sup-norm of error $L - L^{t+1}$ via sup-norm of $E^t$ is not good enough.

- There is greater cancellation in $U_2 \Lambda_2 U_2^T - E^t$. We should leverage $\mu$-incoherence.
Another expression for the error

- For the first \( r \) eigenvectors (or fewer if some eigenvalues are zero)

\[
(L + E^t)u_i = \lambda_i u_i
\]

and by rearrangement

\[
u_i = \frac{1}{\lambda_i} \left( I - \frac{E^t}{\lambda_i} \right)^{-1} \quad \text{and} \quad Lu_i = \frac{1}{\lambda_i} \left( \sum_{p \geq 0} \frac{(E^t)^p}{\lambda_i^p} \right) Lu_i. \quad \text{Invertible?}
\]

- We can then write

\[
L^{t+1} = \sum_{i=1}^{r} \lambda_i u_i u_i^T
\]

\[
= \sum_{i=1}^{r} \lambda_i \left( \frac{1}{\lambda_i} \left( \sum_{p} \frac{(E^t)^p}{\lambda_i^p} \right) Lu_i \right) \left( \frac{1}{\lambda_i} \left( \sum_{q} \frac{(E^t)^q}{\lambda_i^q} \right) Lu_i \right)^T
\]

\[
= \sum_{p,q} (E^t)^p LU_1 \Lambda_1^{-(p+q+1)} U_1^T L((E^t)^q)^T
\]

\[
= LU_1 \Lambda_1^{-(p+q+1)} U_1^T L + \sum_{p,q:p+q > 0} (E^t)^p LU_1 \Lambda_1^{-(p+q+1)} U_1^T L((E^t)^q)^T.
\]
Another expression for the error (contd.)

So we can write the error as

\[ L - L^{t+1} = (L - LU_1 \Lambda_1^{-1} U_1^T L) + \sum_{p, q: p+q > 0} (E^t)^p L U_1 \Lambda_1^{-(p+q+1)} U_1^T L ((E^t)^q)^T. \]

Claim 3: First expression sup-norm bounded by

\[ \frac{\mu^2 r}{n} (|\sigma_{k+1}|) + \text{small frac} \cdot ||E^t||_\infty. \]

which then yields

\[ \leq \frac{\mu^2 r}{n} (|\sigma_{k+1}|) + \text{small frac} \frac{\mu^2 r}{n} (|\sigma_{k+1}| + 2 \cdot 2^{-t} |\sigma_k|) \]

\[ \leq \frac{\mu^2 r}{n} (|\sigma_{k+1}| + 2^{-t} |\sigma_k|) \times 2. \]

Claim 4: A similar (small frac \cdot ||E^t||_\infty) bound holds for the summation term.
Lemma
Under the $E$-bound,

$$\|E^t\|_{op} \leq \text{small frac} \ |\sigma_k| \quad \text{and} \quad |\sigma_k| \leq (1 + \text{small frac}) |\lambda_k|.$$ 

Proof: We already saw

$$\|E^t\|_{op} \leq \alpha n \|E^t\|_{\infty} \leq \text{small frac} \ |\sigma_k|$$

where the second inequality is because of $E$-bound assumption and the assumption on $\alpha \mu^2 r$.

Proximity of $\lambda_k$ and $\sigma_k$ is due to Weyl’s inequality, and the above bound on operator norm.
Proof steps for Claim 3. Claim 4 has a similar proof.

- Claim 3 says:

\[ \| L - LU_1 \Lambda_1^{-1} U_1^T L \|_\infty \leq \frac{\mu^2 r}{n} (|\sigma_{k+1}|) + \text{small frac} \cdot \| E^t \|_\infty. \]

- First: sup-norm bounded by operator-norm through a factor via \( \mu \)-incoherence:

\[ \| L - LU_1 \Lambda_1^{-1} U_1^T L \|_\infty \leq \frac{\mu^2 r}{n} \| L - LU_1 \Lambda_1^{-1} U_1^T L \|_{op} \]

- Second: substitute \( L = U_1 \Lambda_1 U_1^T + U_2 \Lambda_2 U_2^T - E^t \) and use \( U_1 \) and \( U_2 \) are made of orthogonal columns to get

\[ L - LU_1 \Lambda_1^{-1} U_1^T L = U_1 U_1^T E^t + (U_1 U_1^T E^t)^T - E^t U_1 \Lambda_1^{-1} U_1^T (E^t)^T + U_2 \Lambda_2 U_2^T - E^t. \]

- Operator norm \( \| L - LU_1 \Lambda_1^{-1} U_1^T L \|_{op} \) then bounded by

\[ 3\| E^t \|_{op} + \frac{\| E^t \|_{op}^2}{|\lambda_k|} + |\lambda_{k+1}| \leq |\sigma_{k+1}| + 6\| E^t \|_{op} \]

\[ \leq |\sigma_{k+1}| + 6\alpha n\| E^t \|_{\infty}. \]
References and data sets


Data sets:
http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html