Identifiability with sparsity

Decompose a low-rank matrix with known coefficient sparsity.

$$\begin{cases} M = UV, \\ \operatorname{rank}(M) = \operatorname{rank}(U) = r, \\ \|V(:,j)\|_0 \leq k = r - s < r \, \forall j. \end{cases}$$

Identifiability with sparsity

Decompose a low-rank matrix with known coefficient sparsity.

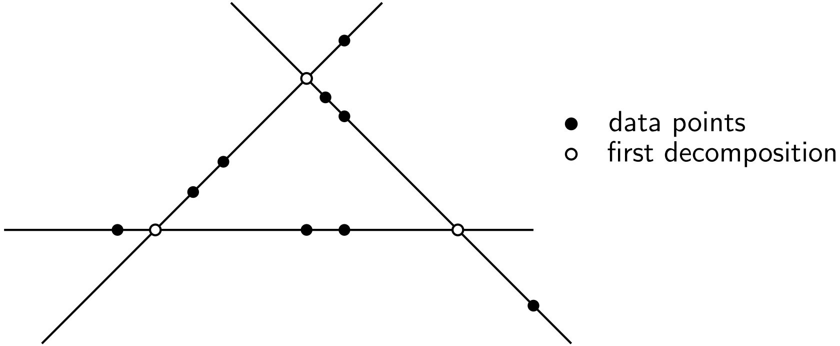
$$\begin{cases} M = UV, \\ \text{rank}(M) = \text{rank}(U) = r, \\ \|V(:,j)\|_0 \leq k = r - s < r \,\forall j. \end{cases}$$

Many existing theoretical results (see, e.g., [Gribonval 16]) and algorithms (Dictionary Learning). But:

- ✗ Not many results specific to the low-rank case
- ✗ Only two deterministic identifiability results [Elad 06, Georgiev 05]
- ✗ Not much in the NMF case except ℓ_1 regularization

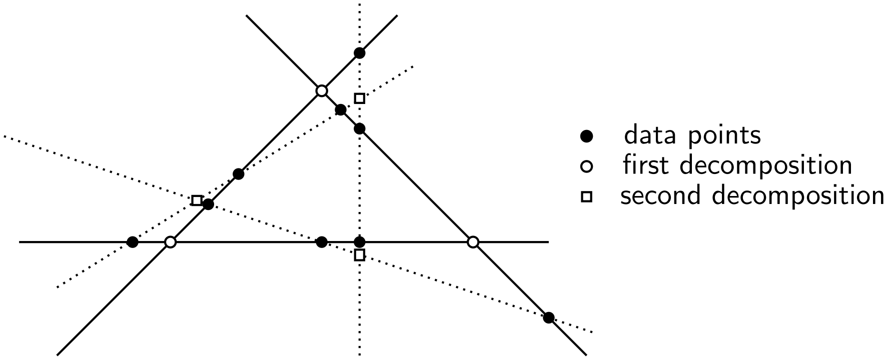
Identifiability with sparsity: example

Example: $p = 3$, $r = 3$, s=sparsity=1, $n = 9$.



Identifiability with sparsity: example

Example: $p = 3$, $r = 3$, s=sparsity=1, $n = 9$.



Identifiability results

Theorem

Let $M = UV$ where $\text{rank}(U) = \text{rank}(M) = r$ and each column of V has at least s zeros. The factorization (U, V) is essentially unique if on each hyperplane spanned by all but one column of U, there are $\left\lfloor \frac{r(r-2)}{s} \right\rfloor + 1$ data points with spark r.

[CG19] J.E. Cohen and N. Gillis, "Identifiability of Complete Dictionary Learning", SIAM J. on Mathematics of Data Science 1 (3), pp. 518-536, 2019.

Identifiability results

Theorem

Let $M = UV$ where $\text{rank}(U) = \text{rank}(M) = r$ and each column of V has at least s zeros. The factorization (U, V) is essentially unique if on each hyperplane spanned by all but one column of U, there are $\left\lfloor \frac{r(r-2)}{s} \right\rfloor + 1$ data points with spark r.

✓ For $s = 1$, this requires $r^3 - 2r^2 + r$ data points and it is tight up to the constant r (counter examples for any $n = r^3 - 2r^2$).

[CG19] J.E. Cohen and N. Gillis, "Identifiability of Complete Dictionary Learning", SIAM J. on Mathematics of Data Science 1 (3), pp. 518-536, 2019.

Identifiability results

Theorem

Let $M = UV$ where $\text{rank}(U) = \text{rank}(M) = r$ and each column of V has at least s zeros. The factorization (U, V) is essentially unique if on each hyperplane spanned by all but one column of U, there are $\left\lfloor \frac{r(r-2)}{s} \right\rfloor + 1$ data points with spark r.

- ✓ For $s = 1$, this requires $r^3 - 2r^2 + r$ data points and it is tight up to the constant r (counter examples for any $n = r^3 - 2r^2$).
- ✓ For $s = r - 1$, this requires r data points and it is tight (one on each intersection of $r - 1$ hyperplanes).

[CG19] J.E. Cohen and N. Gillis, "Identifiability of Complete Dictionary Learning", SIAM J. on Mathematics of Data Science 1 (3), pp. 518-536, 2019.

Identifiability results

Theorem

Let $M = UV$ where $\text{rank}(U) = \text{rank}(M) = r$ and each column of V has at least s zeros. The factorization (U, V) is essentially unique if on each hyperplane spanned by all but one column of U, there are $\left\lfloor \frac{r(r-2)}{s} \right\rfloor + 1$ data points with spark r.

- ✓ For $s = 1$, this requires $r^3 - 2r^2 + r$ data points and it is tight up to the constant r (counter examples for any $n = r^3 - 2r^2$).
- ✓ For $s = r - 1$, this requires r data points and it is tight (one on each intersection of $r - 1$ hyperplanes).
- ✓ It is tight up to constant factors for any $s = \beta r$ for any fixed constant β.

[CG19] J.E. Cohen and N. Gillis, "Identifiability of Complete Dictionary Learning", SIAM J. on Mathematics of Data Science 1 (3), pp. 518-536, 2019.

Identifiability results

Theorem

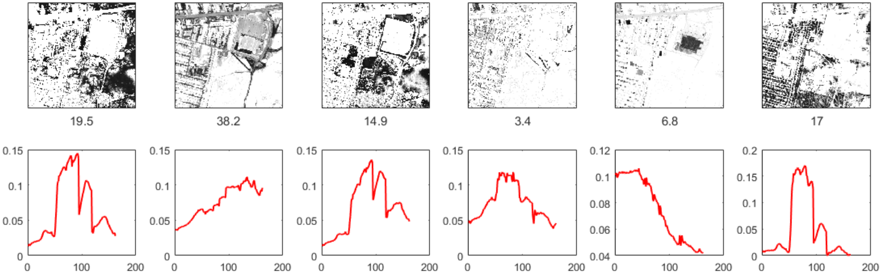
Let $M = UV$ where $\text{rank}(U) = \text{rank}(M) = r$ and each column of V has at least s zeros. The factorization (U, V) is essentially unique if on each hyperplane spanned by all but one column of U, there are $\left\lfloor \frac{r(r-2)}{s} \right\rfloor + 1$ data points with spark r.

- ✓ For $s = 1$, this requires $r^3 - 2r^2 + r$ data points and it is tight up to the constant r (counter examples for any $n = r^3 - 2r^2$).
- ✓ For $s = r - 1$, this requires r data points and it is tight (one on each intersection of $r - 1$ hyperplanes).
- ✓ It is tight up to constant factors for any $s = \beta r$ for any fixed constant β.
- ✓ Nonnegativity not taken into account in the analysis, it helps both in theory and in practice: further work.

[CG19] J.E. Cohen and N. Gillis, "Identifiability of Complete Dictionary Learning", SIAM J. on Mathematics of Data Science 1 (3), pp. 518-536, 2019.

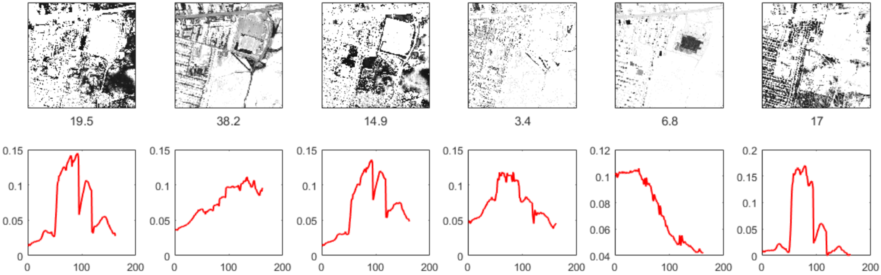
Sparsity in action

Spectral unmixing, $R = 6, s = 4$



Sparsity in action

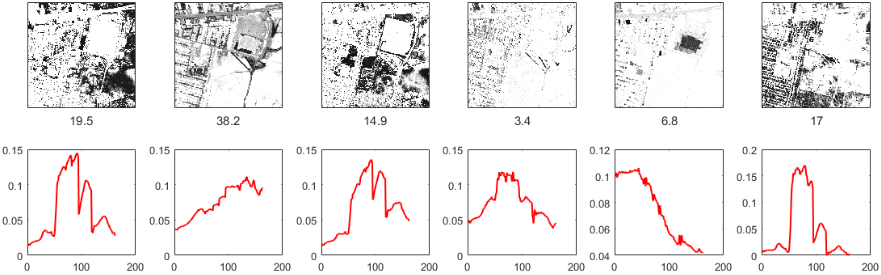
Spectral unmixing, $R = 6, s = 4$



✓ Sparsity is another way to obtain **identifiability** for matrix decompositions.

Sparsity in action

Spectral unmixing, $R = 6, s = 4$

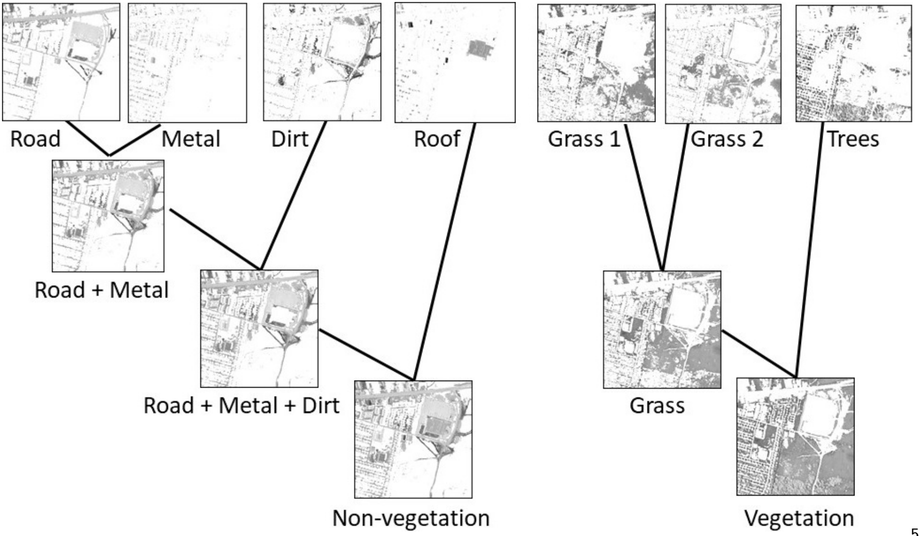


- ✓ Sparsity is another way to obtain **identifiability** for matrix decompositions.
- ✗ Hard combinatorial problems to solve. . .

What are we doing in Mons?

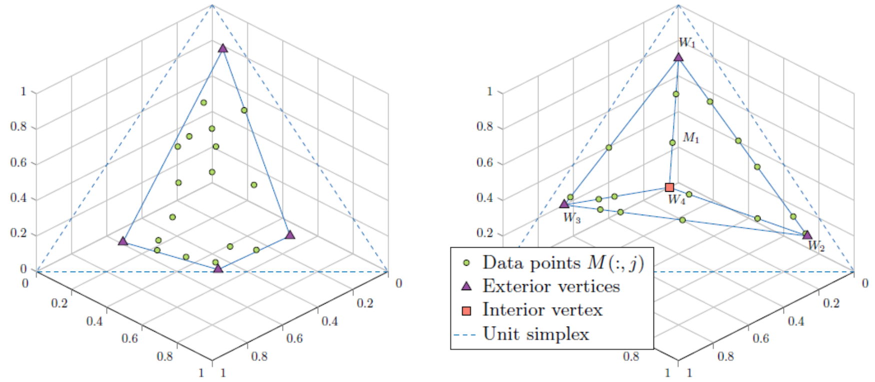
What are we doing in Mons?

Pierre DH is exploring deep NMF $M \approx UV_1 V_2 \dots V_\ell$.



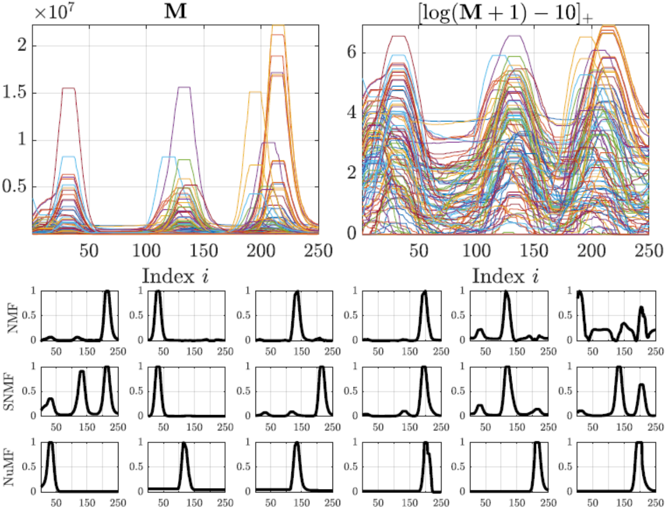
What are we doing in Mons?

Nicolas N is exploring sparse separable NMF



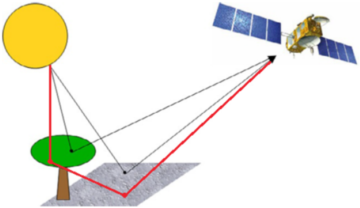
What are we doing in Mons?

Andersen is exploring unimodal NMF



What are we doing in Mons?

Christophe is exploring linear-quadratic NMF



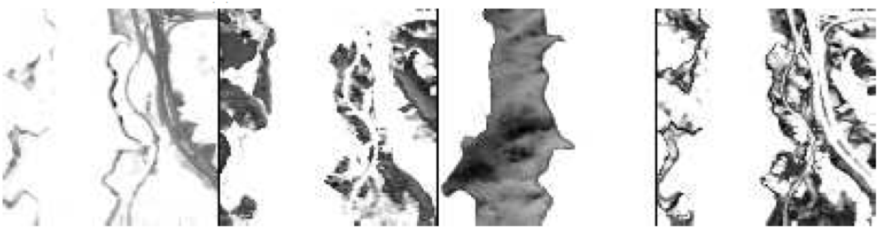
Linear-quadratic (LQ) model

$$M(:,j) \approx \underbrace{\sum_{k=1}^{r} U(:,k)V(k,j)}_{NMF} + \underbrace{\sum_{p=1}^{r}\sum_{l=p}^{r} \beta_{ipl}\Big(U(:,p) \odot U(:,l)\Big)}_{\text{double reflections}}.$$

What are we doing in Mons?

Valentin is exploring constrained β-divergence NMF

Jasper Ridge Data set



What are we doing in Mons?

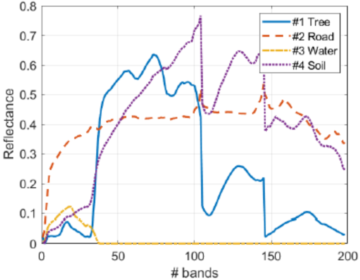
François is exploring ℓ_1 symNMF for document classification

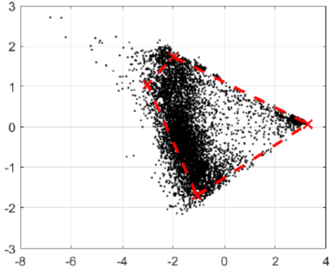
Data	SymNMF	ODsymNMF-ℓ_2	ODsymNMF-ℓ_1
classic	63.67	63.67	**66.33**
ohscal	**43.24**	43.16	38.08
hitech	49.07	49.24	**52.19**
reviews	49.37	49.55	**70.07**
sports	**51.46**	51.41	48.81
la1	**49.16**	48.81	40.61
la2	**48.94**	48.62	39.45
k1b	57.18	58.68	**66.45**
tr11	59.66	**59.90**	51.21
tr23	35.29	35.29	**36.76**
tr41	46.70	**47.15**	47.04
tr45	42.90	42.61	**43.04**

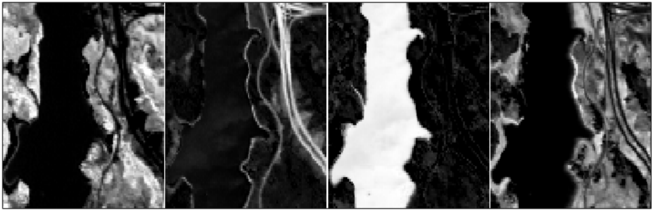
Table 4: Accuracy (in %) for each data set. The bold values are the best of each line.

What are we doing in Mons?

Maryam is exploring facet-based algorithms

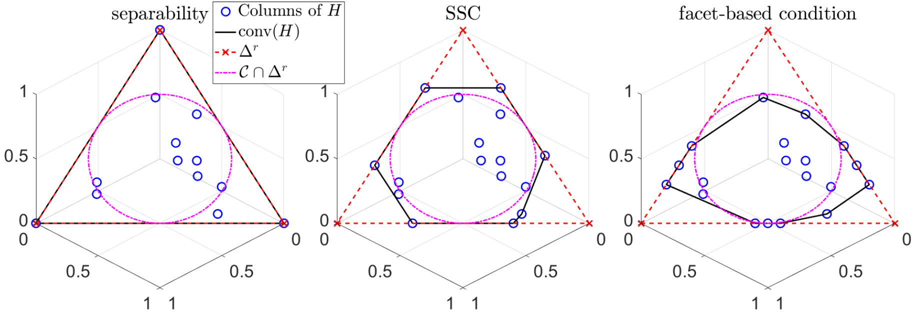






What are we doing in Mons?

Tim is exploring identifiability conditions



What are we doing in Mons?

Hien is developing a general class of highly efficient algorithms for non-convex non-smooth optimization

Algorithm 1 TITAN with cyclic update to solve Problem (1)

Input: Choose $x^{-1}, x^0 \in \mathcal{X}$ (x^{-1} can be chosen equal to x^0).
Output: x^k that approximately solves (1).

for $k = 0, 1, \ldots$ **do**
  Set $x^{k,0} = x^k$
  for $i = 1, ..., m$ **do**
    Choose a block i surrogate function u_i of f and an extrapolation $\mathcal{G}_i^k(x_i^k, x_i^{k-1})$.
    Update block i by

$$x_i^{k,i} \in \underset{x_i \in \mathcal{X}_i}{\operatorname{argmin}}\, u_i(x_i, x^{k,i-1}) - \langle \mathcal{G}_i^k(x_i^k, x_i^{k-1}), x_i \rangle + g_i(x_i), \tag{3}$$

    and set $x_j^{k,i} = x_j^{k,i-1}$ for all $j \neq i$.
  end for
  Set $x^{k+1} = x^{k,m}$.
end for

Thank you for your attention!

Code and papers available from
https://sites.google.com/site/nicolasgillis