Hyperspectral Unmixing: Insights and Beyond

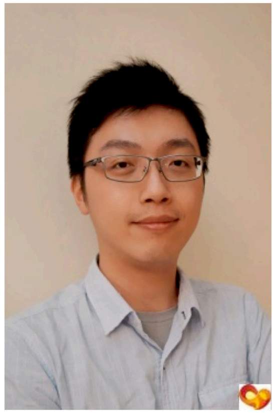
Wing-Kin (Ken) Ma

Department of Electronic Engineering, The Chinese University of Hong Kong

Plenary Talk, CAMSAP 2019, Guadeloupe, December 2019

Acknowledgment

José Bioucas-Dias
IST, Portugal

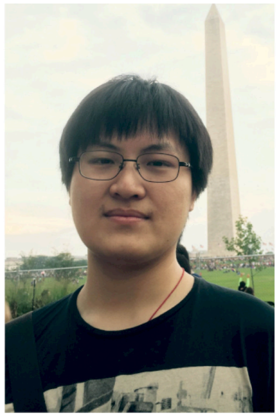


Tsung-Han Chan
MediaTek, Taiwan

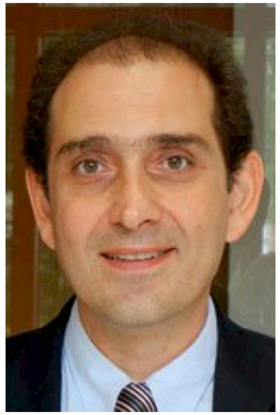
Chong-Yung Chi
NTHU, Taiwan

Xiao Fu
Oregon State Univ., US

Nicolas Gillis
Univ. Mons, Belgium



Chia-Hsiang Lin
NCU, Taiwan

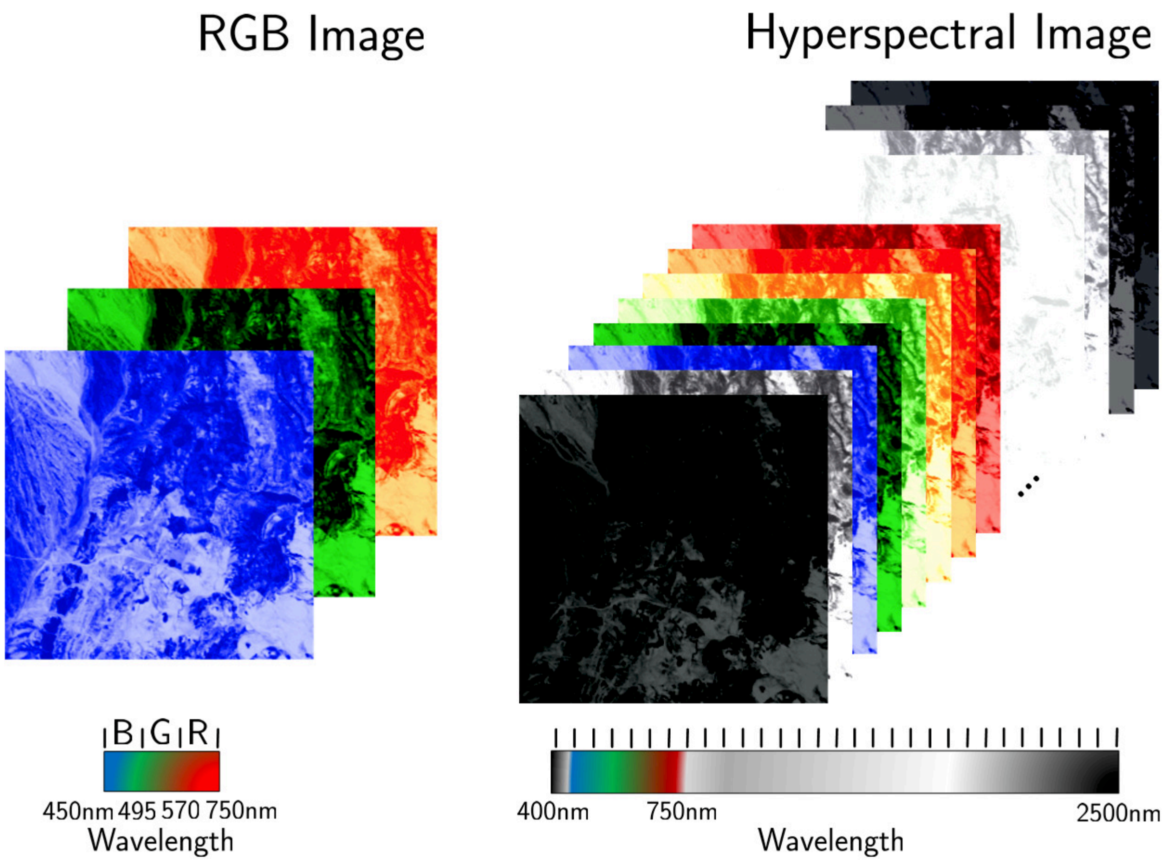


Nikos Sidiropoulos
Univ. Virgina, US

Background of Hyperspectral Unmixing

Hyperspectral Imaging

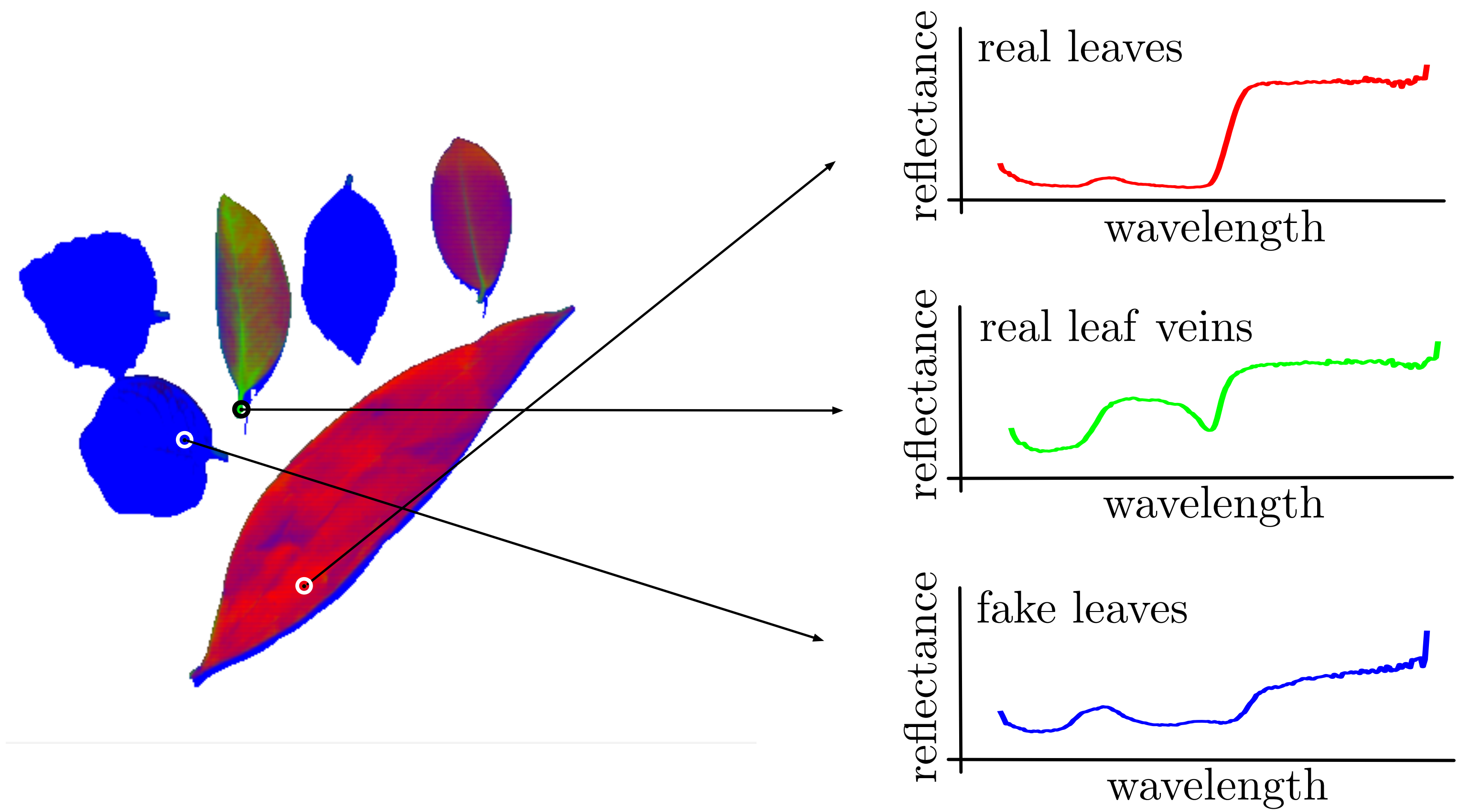
- cover visible to near-infrared wavelengths, with 10nm resolution and > 200 bands



A Hyperspectral Image Example: Real or Fake?

A hyperspectral image shown in RGB. Captured by SPCIM IQ HS camera.

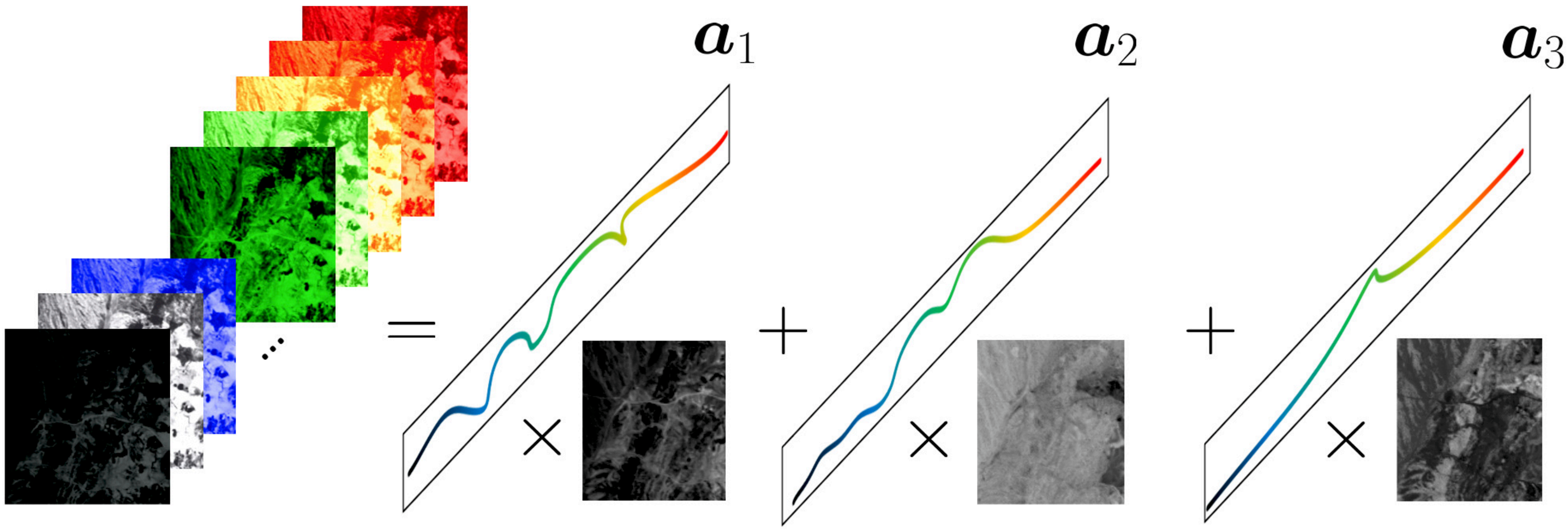
A Hyperspectral Image Example: Real or Fake?



A false colormap of its underlying materials. Taking out the background, the image is composed of real leaves, leaf veins, and fake leaves.

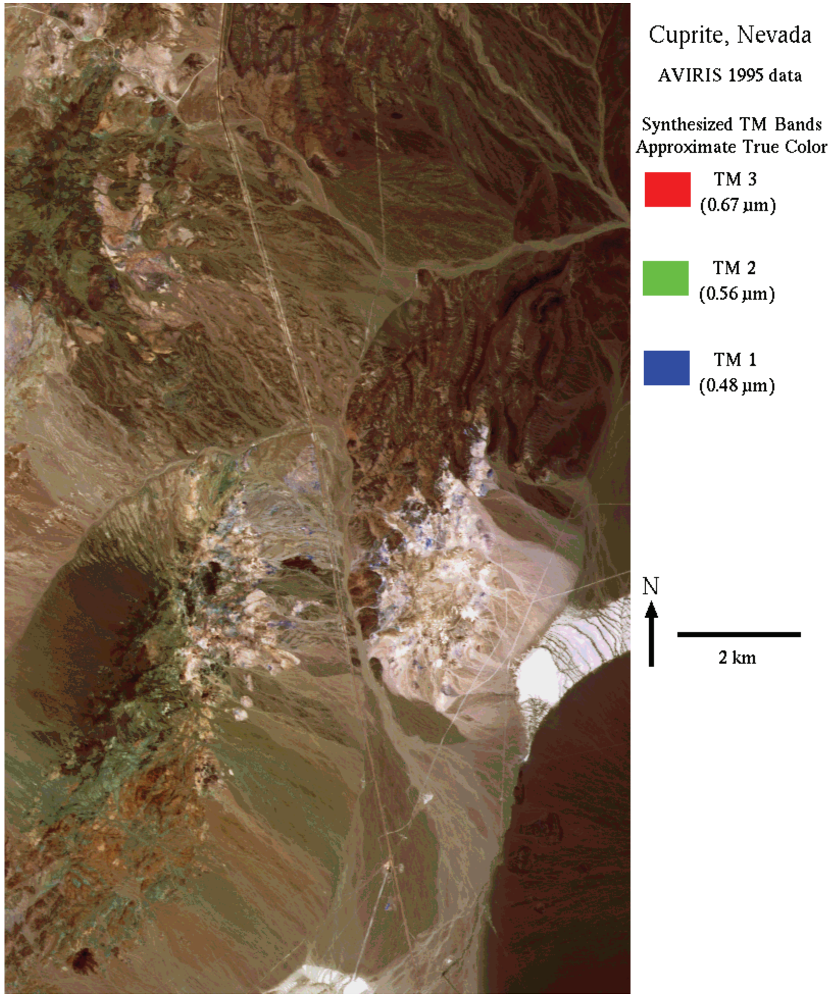
Why Hyperspectral?

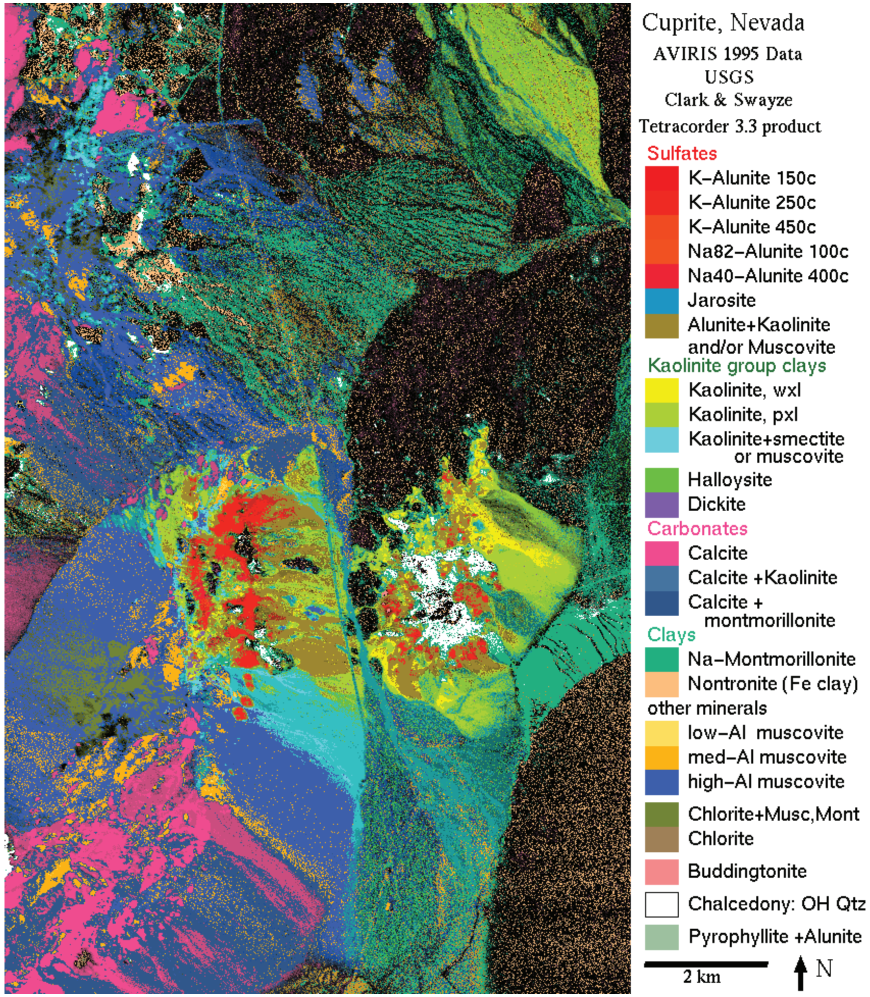
- allow us to "see" different materials, revealed by their spectral signatures



- **Hyperspectral Unmixing (HU):** to blindly identify
  - the underlying materials, in form of the spectral responses;
  - the abundance map of each material

Example: Mineral Identification in Remote Sensing

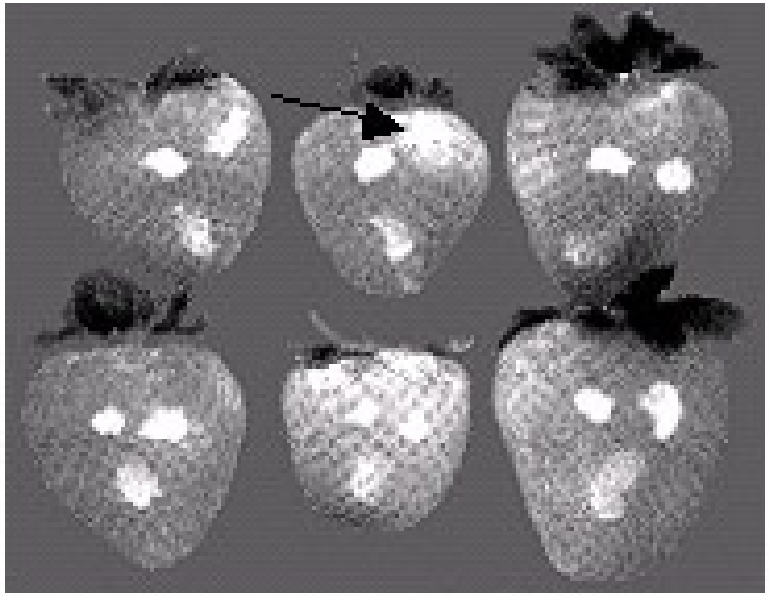




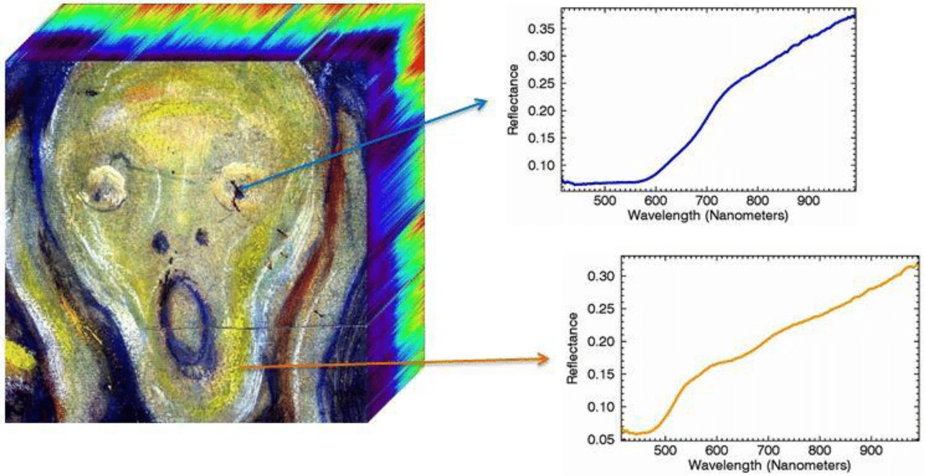
AVIRIS Cuprite image. Courtesy to USGS.

Applications

- remote sensing (studied extensively)
- food safety, art conservation, archaeology, medical imagery,...



(a) food safety

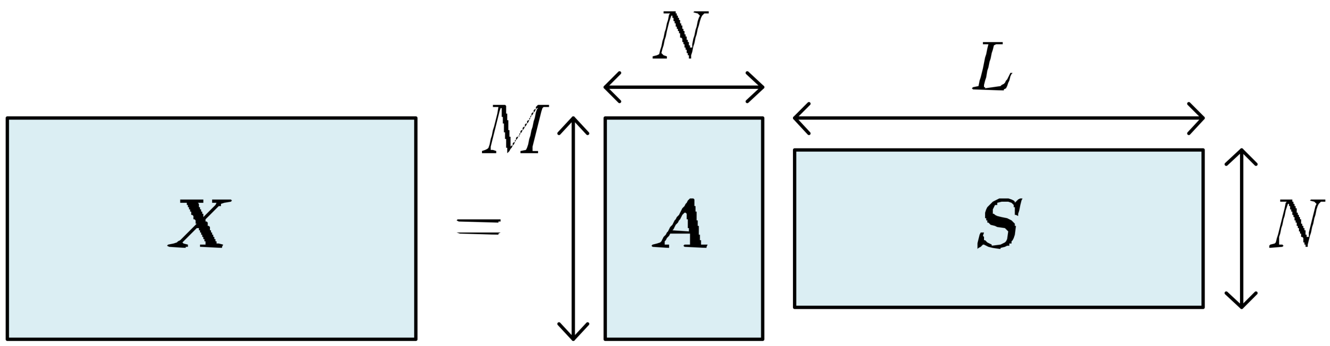


(b) pigment identification in art conservation

Source: (a) http://tao.umd.edu/html/fecal_contamination.html (b) H. Deborah, Pigment Mapping of Cultural Heritage Paintings Based on Hyperspectral Imaging, MSc Thesis, Gjøvik University College, Norway.

What We Will See in This Talk

- how remote sensing people solve a structured matrix factorization problem



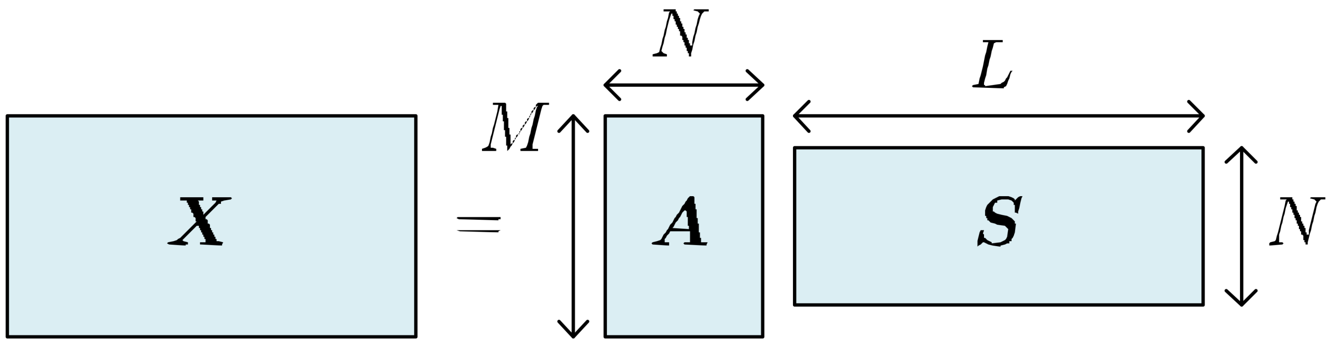
  where

  - $\boldsymbol{A}$ has full column rank;
  - every column of $\boldsymbol{S}$ lies in the unit simplex, i.e., $\boldsymbol{s}_i \geq \boldsymbol{0}, \boldsymbol{s}_i^T \boldsymbol{1} = 1$ for all i

- notes with HU:
  - a key topic in remote sensing, nearly 30 years of history
  - lead to a branch of provably good structured matrix factorization techniques
  - has connections to certain problems in machine learning, data science, etc

What Do You Mean by "Provably Good" Factorization?

- for example, consider non-negative matrix factorization (NMF)



where $\boldsymbol{A} \geq \mathbf{0}, \boldsymbol{S} \geq \mathbf{0}$

- we want the true $(\boldsymbol{A}, \boldsymbol{S})$ to be recoverable (subject to trivial effects)

- observe

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} = \underbrace{\boldsymbol{A}\boldsymbol{C}}_{=\tilde{\boldsymbol{A}}}\underbrace{\mathbf{C}^{-1}\boldsymbol{S}}_{=\tilde{\boldsymbol{S}}}, \qquad \text{for some invertible } \boldsymbol{C}$$

If $\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{S}} \geq \mathbf{0}$, $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{S}})$ is also an NMF solution; NMF may not recover $(\boldsymbol{A}, \boldsymbol{S})$

- NMF guarantees exact recovery under certain sufficient conditions [Donoho-Stodden2003], [Huang-Sidiropoulos-Swami2014]; they are arguably strong

Outline of this Talk

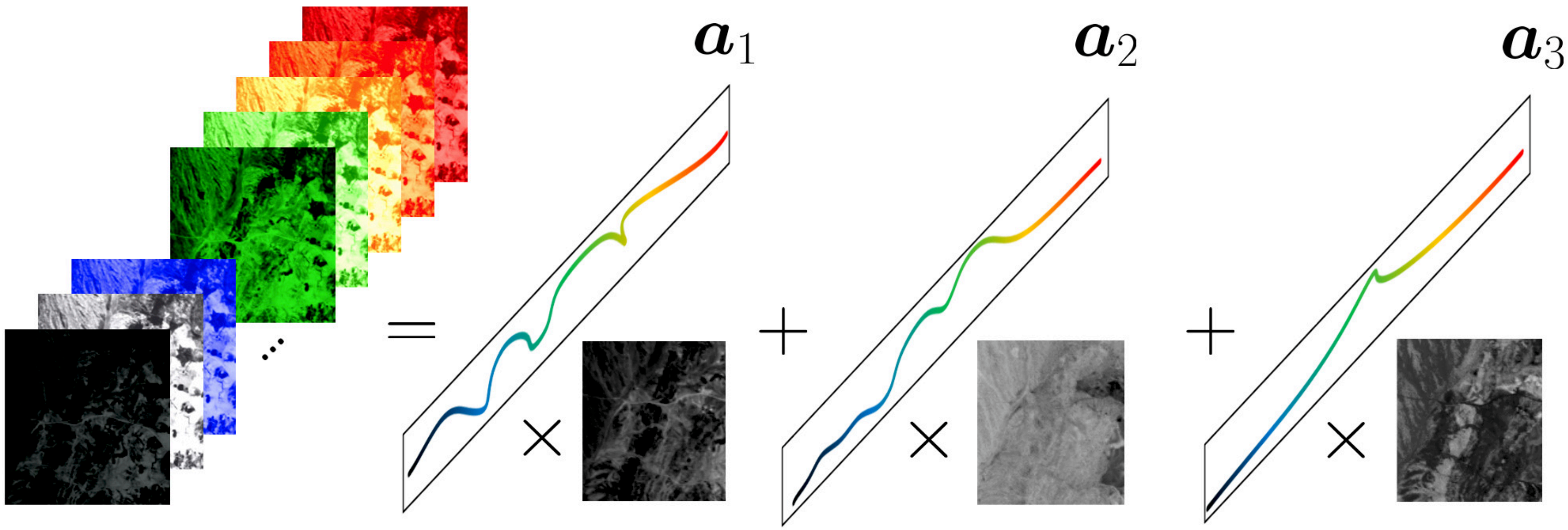
- HU model, the notion of convex geometry
- pure-pixel search
  - do you know separable NMF?
- applications beyond HU
- simplex volume minimization
  - goes beyond separable NMF—provably!
  - connections to regularized matrix factorization, probabilistic PCA
- we will not go through other approaches such as the Bayesian technique (via Monte Carlo), dictionary-aided sparse regression, simplex volume maximization, minimum volume enclosing ellipsoid, maximum volume inscribed ellipsoid, nonlinear unmixing, endmember variability,... Sorry!

HU: Linear Spectral Mixture Model

- **Postulate:** a pixel is a linear proportional combination of pure materials
- example:

  a hyperspectral pixel (as reflectance) $= X\%$ spectral signature of water$+$
  $Y\%$ spectral signature of soil $+ Z\%$ spectral signature of vegetation

where $X\% + Y\% + Z\% = 100\%$.

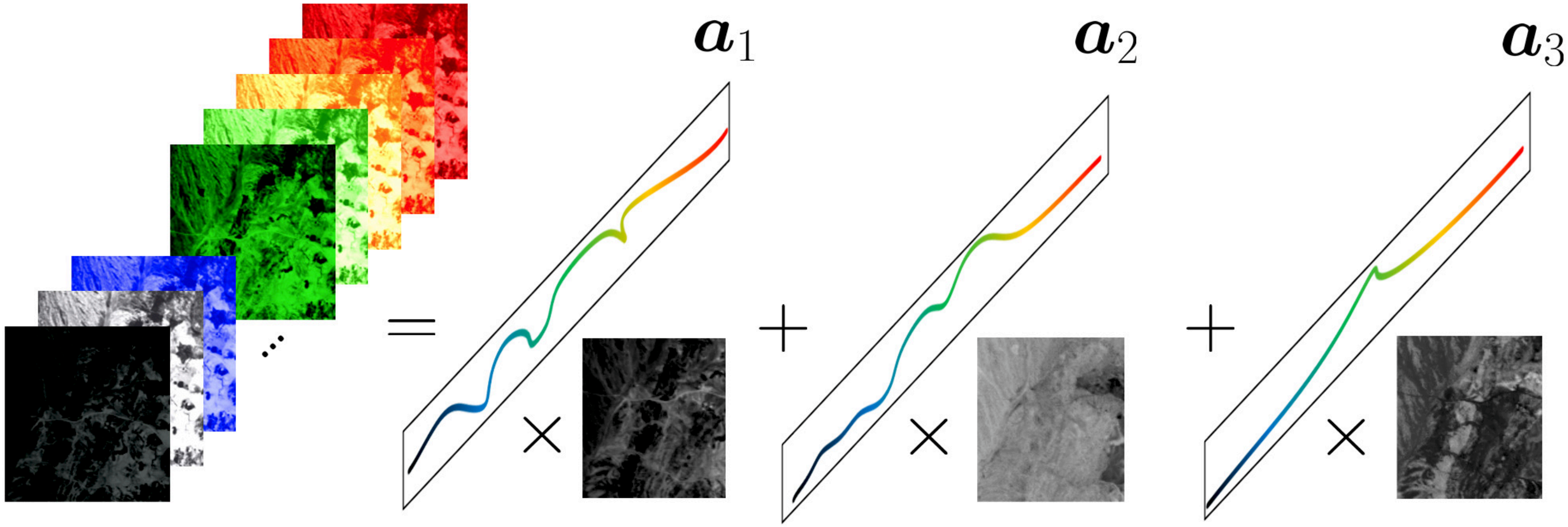


HU: Linear Spectral Mixture Model

- **Model:**

$$\boldsymbol{y}[n] = \sum_{i=1}^{N} \boldsymbol{a}_i s_i[n] + \boldsymbol{\nu}[n] = \boldsymbol{A}\boldsymbol{s}[n] + \boldsymbol{\nu}[n], \quad n = 1, \ldots, L,$$

  where
  - $\boldsymbol{y}[n] \in \mathbb{R}^M$ is the measured hyperspectral vector at pixel n;
  - $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$, $\boldsymbol{a}_i \in \mathbb{R}^M$ is an endmember signature vector;
  - $\boldsymbol{s}[n] \in \mathbb{R}^N$ is the abundance vector at pixel n, with $\boldsymbol{s}[n] \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{s}[n] = 1$;
  - $\boldsymbol{\nu}[n]$ is noise; N is the model order.

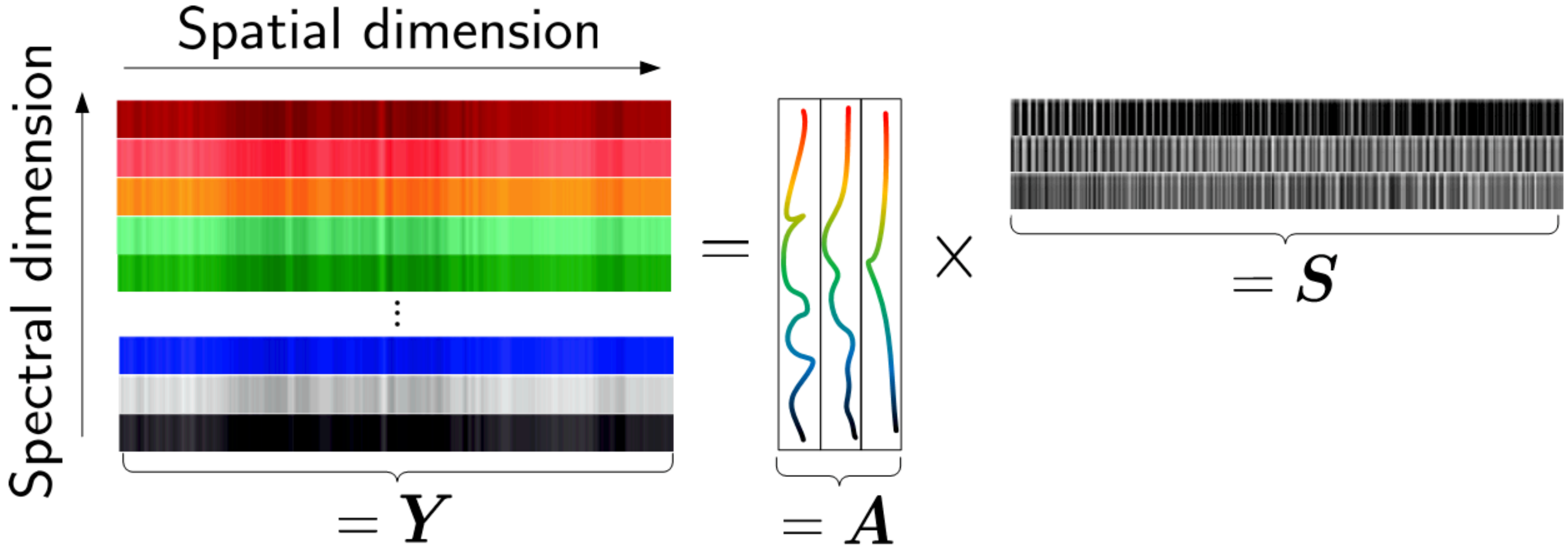


HU: Linear Spectral Mixture Model

- **Model:**

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{V}$$

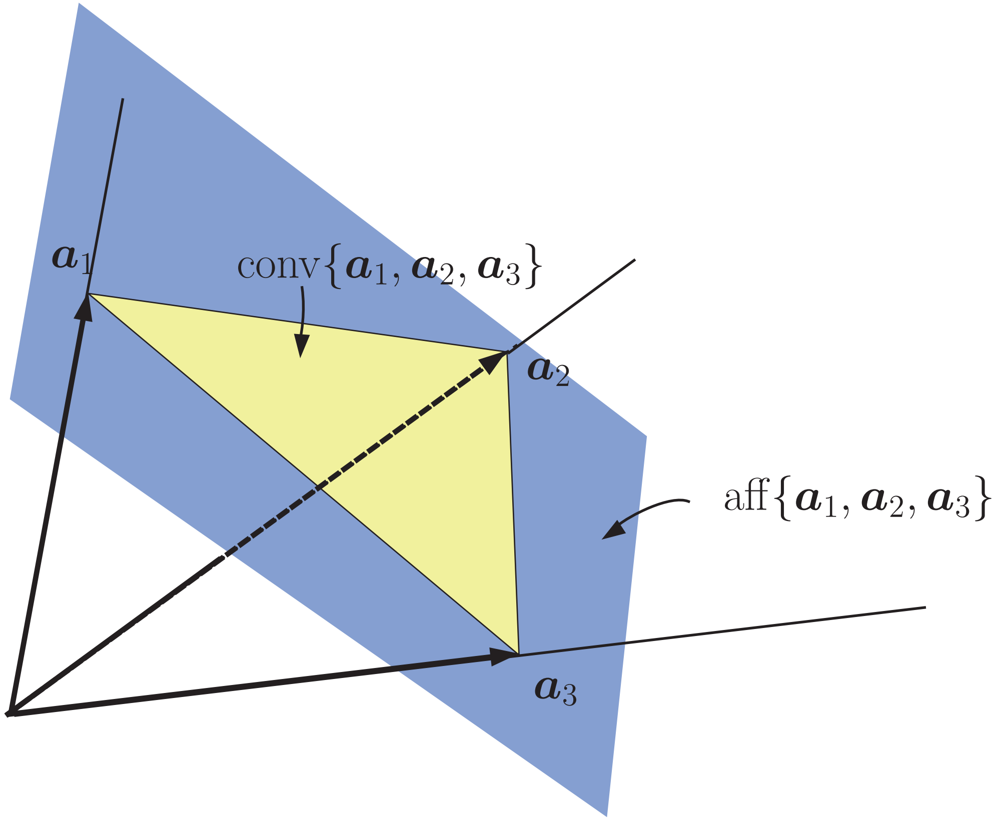
where $\boldsymbol{Y} = [\ \boldsymbol{y}[1], \ldots, \boldsymbol{y}[L]\ ]$; $\boldsymbol{S} = [\ \boldsymbol{s}[1], \ldots, \boldsymbol{s}[L]\ ]$; $\boldsymbol{V} = [\ \boldsymbol{v}[1], \ldots, \boldsymbol{v}[L]\ ]$;
recall $\boldsymbol{s}[n] \geq \boldsymbol{0}, \boldsymbol{1}^T\boldsymbol{s}[n] = 1$



- **HU:** recover $\boldsymbol{A}$ from $\boldsymbol{Y}$
  - once we have $\boldsymbol{A}$ we can get $\boldsymbol{S}$ by $\boldsymbol{S} = \boldsymbol{A}^\dagger\boldsymbol{Y}$

Convex Geometry

Convex Geometry Preliminaries



- convex hull: $\mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\} = \{\boldsymbol{y} = \sum_{i=1}^{N} \boldsymbol{a}_i \theta_i \mid \boldsymbol{\theta} \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{\theta} = 1\}$.
- simplex: $\mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ is a simplex if $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ are affinely independent.

Convex Geometry Observation

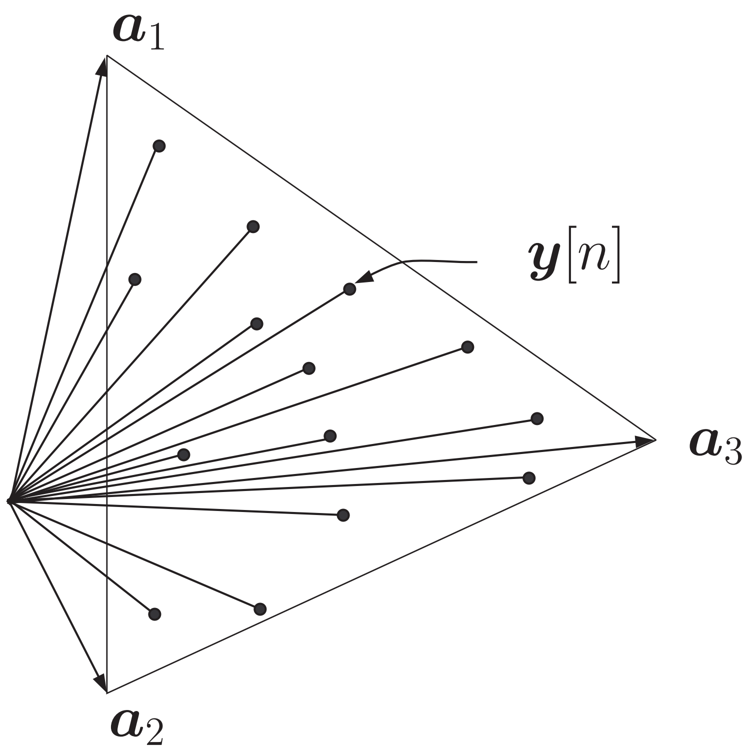
- consider the noiseless model

$$\boldsymbol{y}[n] = \sum_{i=1}^{N} s_i[n]\boldsymbol{a}_i.$$

Since $s_i[n] \geq 0$, $\sum_{i=1}^{N} s_i[n] = 1$,

$$\boldsymbol{y}[n] \in \text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}.$$

- assume linearly independent $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$



- **Observation:** each pixel $\boldsymbol{y}[n]$ lies in the simplex $\text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$.

Convex Geometry Observation

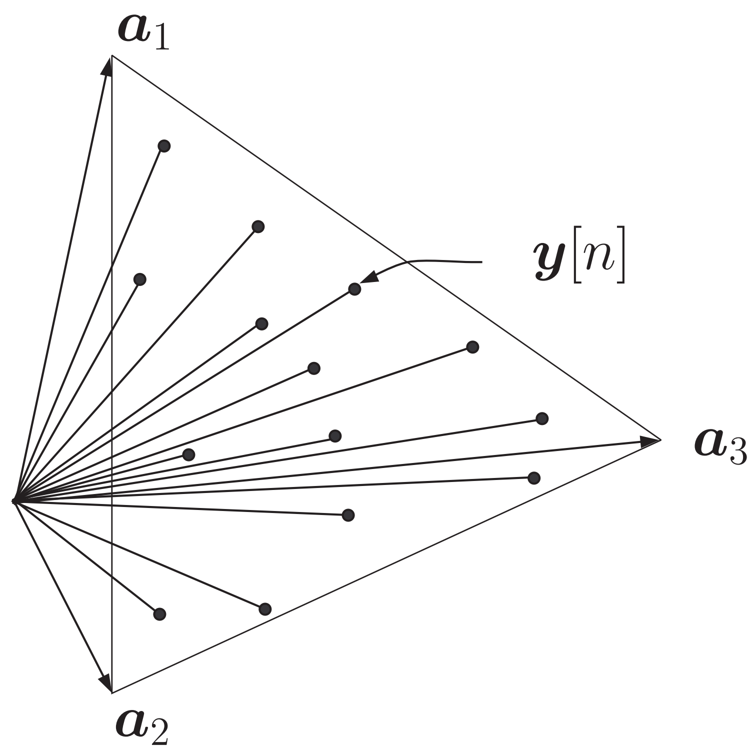
- consider the noiseless model

$$\boldsymbol{y}[n] = \sum_{i=1}^{N} s_i[n]\boldsymbol{a}_i.$$

Since $s_i[n] \geq 0$, $\sum_{i=1}^{N} s_i[n] = 1$,

$$\boldsymbol{y}[n] \in \mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}.$$

- assume linearly independent $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$



- **Observation:** each pixel $\boldsymbol{y}[n]$ lies in the simplex $\mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$.

- **Question (heart of this talk!):** can we identify the vertices of $\mathrm{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$ from $\boldsymbol{y}[1], \ldots, \boldsymbol{y}[L]$?

Craig's Seminal Work [Craig1994]

542 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 32, NO. 3, MAY 1994

Minimum-Volume Transforms for Remotely Sensed Data

Maurice D. Craig

***Abstract*—Scatter diagrams for multispectral remote sensing data tend to be triangular, in the two-band case, pyramidal for three bands, and so on. They radiate away from the so-called darkpoint, which represents the scanner's response to an unilluminated target. A minimum-volume transform may be described (provisionally) as a nonorthogonal linear transformation of the multivariate data to new axes passing through the dark point, with directions chosen such that they (for two bands), or the new coordinate planes (for three bands, etc.) embrace the data cloud as tightly as possible.**

The reason for the observed shapes of scatter diagrams is to be found in the theory of linear mixing at the subfootprint scale. Thus, suitably defined, minimum-volume transforms can often be used to unmix images into new spatial variables showing the proportions of the different cover types present, a type of enhancement that is not only intense, but physically meaningful. The present paper furnishes details for constructing computer programs to effect this operation. It will serve as a convenient technical source that may be referenced in subsequent, more profusely illustrated publications that address the intended application, the mapping of surface mineralogy.

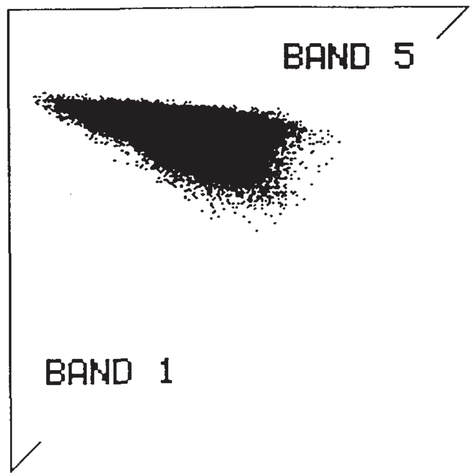


Fig. 1. Two-band triangular scatter plot for a 512 × 512 subscene of a Landsat Thematic Mapper image (actually WRS 111-075, Nullagine, W.A., acquired August 18, 1986).

I. Introduction

THIS paper describes processing algorithms for two closely related transformations, both applicable to radiance data from multispectral scanners. It supplies de-

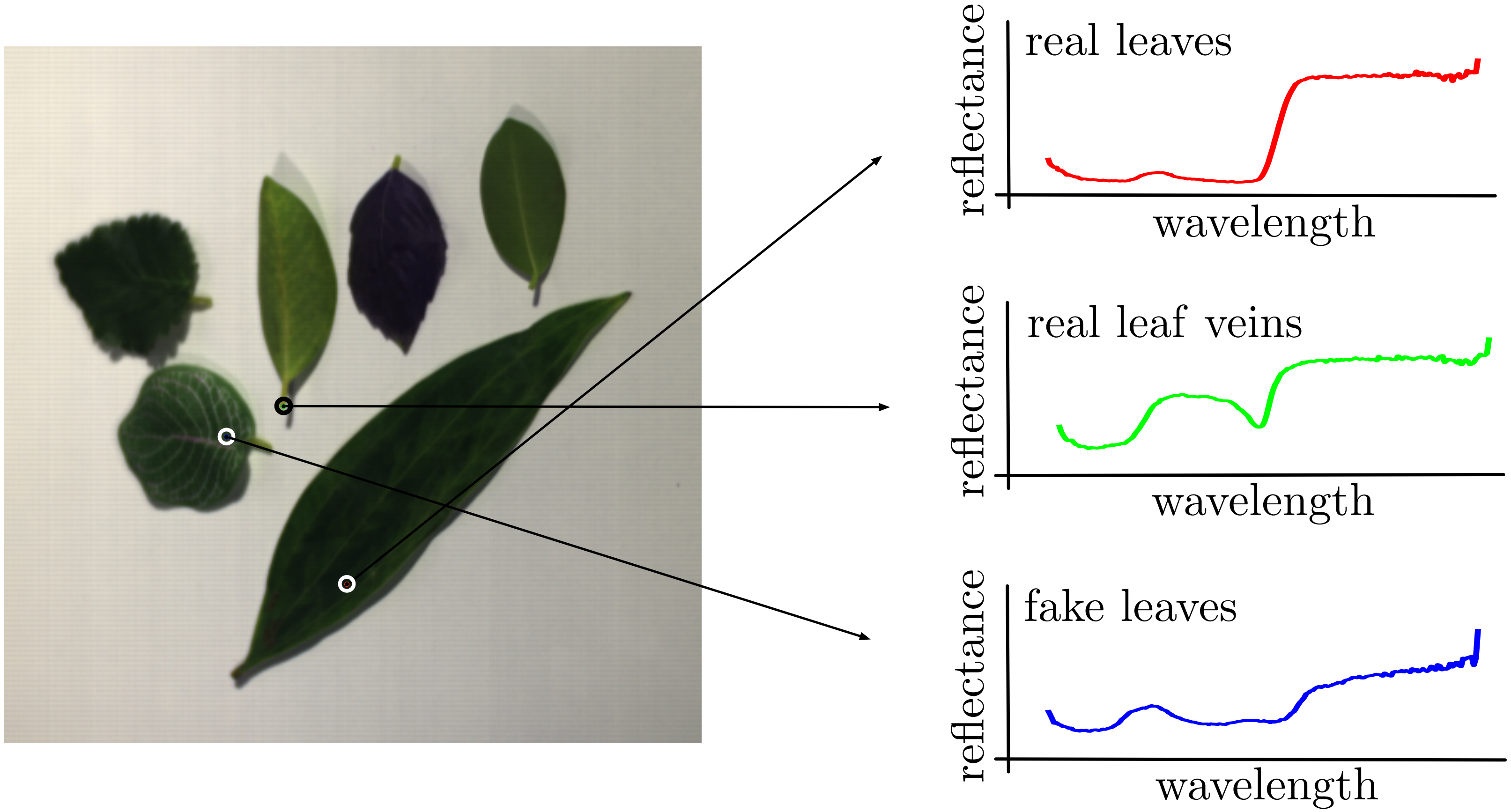
away from the so-called dark point, the scanner's response to a target of nil reflectance in all bands (see Fig. 1).

This appearance of bivariate scatter diagrams now sug-

Pure Pixel Search

Pure Pixels

- **Observation:** there are instances for which some pixels contain only one endmember; you may even read them out manually

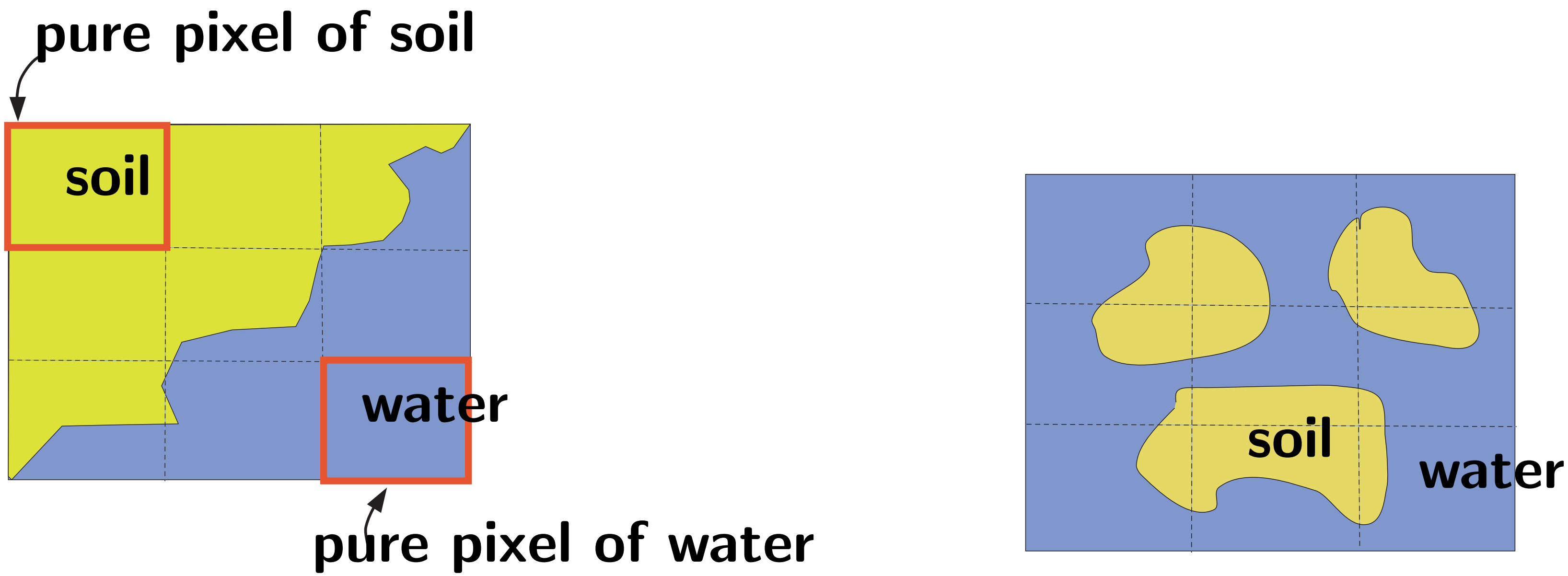


Pure Pixels

Definition: Endmember i is said to have a pure pixel if, without noise, there exists an index ℓ_i such that

$$\boldsymbol{y}[\ell_i] = \boldsymbol{a}_i$$

- or, endmember i has pure pixels if $\boldsymbol{s}[\ell_i] = \boldsymbol{e}_i$ for some ℓ_i; $\boldsymbol{e}_i$'s are unit vectors



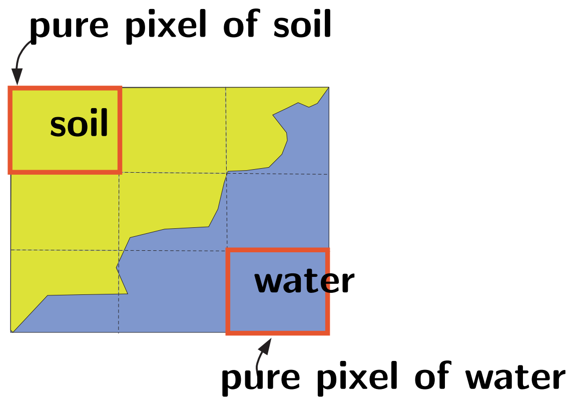
(a) pure pixel case (b) no-pure pixel case

Pure Pixels

Definition: Endmember i is said to have a pure pixel if, without noise, there exists an index ℓ_i such that

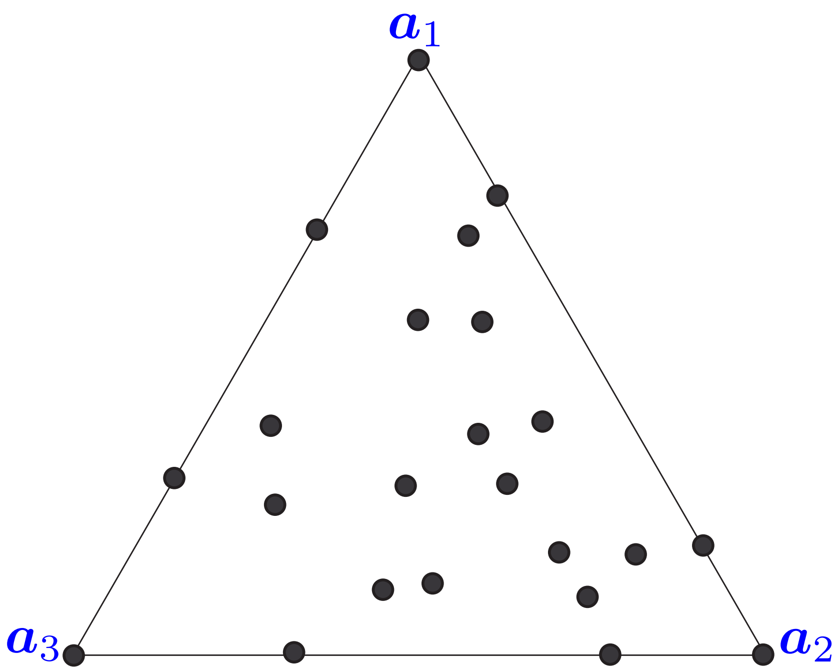
$$\boldsymbol{y}[\ell_i] = \boldsymbol{a}_i$$

- **Implication:**
  - Suppose that every endmember has a pure pixel, and there is no noise.
  - If we know $\ell_1, \ldots, \ell_N$, then $[\ \boldsymbol{y}[\ell_1], \ldots, \boldsymbol{y}[\ell_N]\ ] = \boldsymbol{A}$ — and the problem is solved!

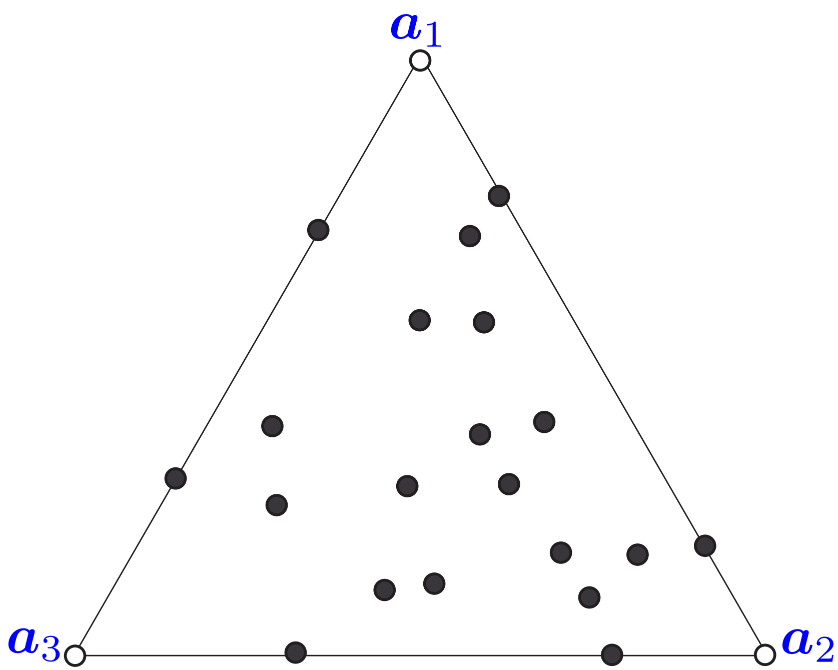


- **Problem:** find the pure pixel indices.

Convex Geometry With and Without Pure Pixels



(a) pure pixel case



(b) no pure pixel case

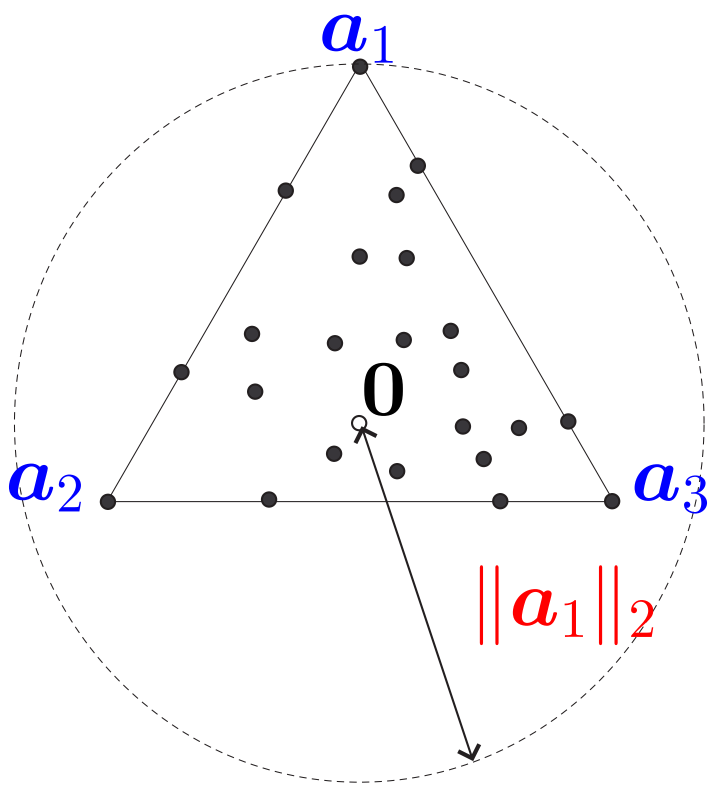
- the pure pixel case has points on the vertices
- the no pure pixel case does not.

Successive Projection Algorithm (SPA)

- there are numerous pure pixel search algorithms, e.g.,
  - pure pixel index (PPI) [Boardman-Kruse-Green1995], the first in HU
  - vertex component analysis (VCA) [Nascimento-Bioucas2003], the most popular
- we consider SPA, arguably the easiest one to understand the insights

A Simple Geometrical Question

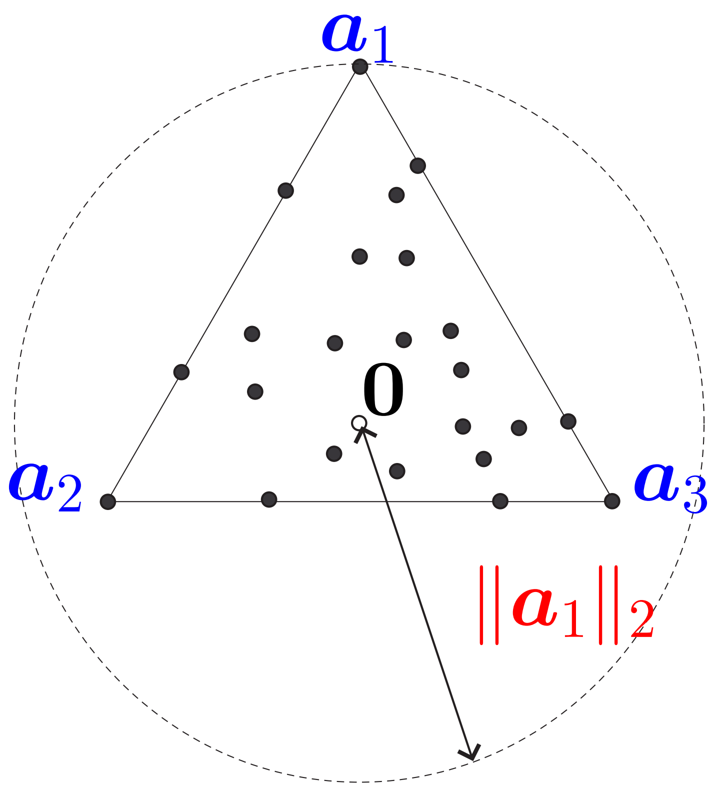
Question: the dark dots are the hyperspectral pixels $\boldsymbol{y}[n]$'s. Which $\boldsymbol{y}[n]$ gives the largest Euclidean norm $\|\boldsymbol{y}[n]\|_2$?



A Simple Geometrical Question

Question: the dark dots are the hyperspectral pixels $\boldsymbol{y}[n]$'s. Which $\boldsymbol{y}[n]$ gives the largest Euclidean norm $\|\boldsymbol{y}[n]\|_2$?

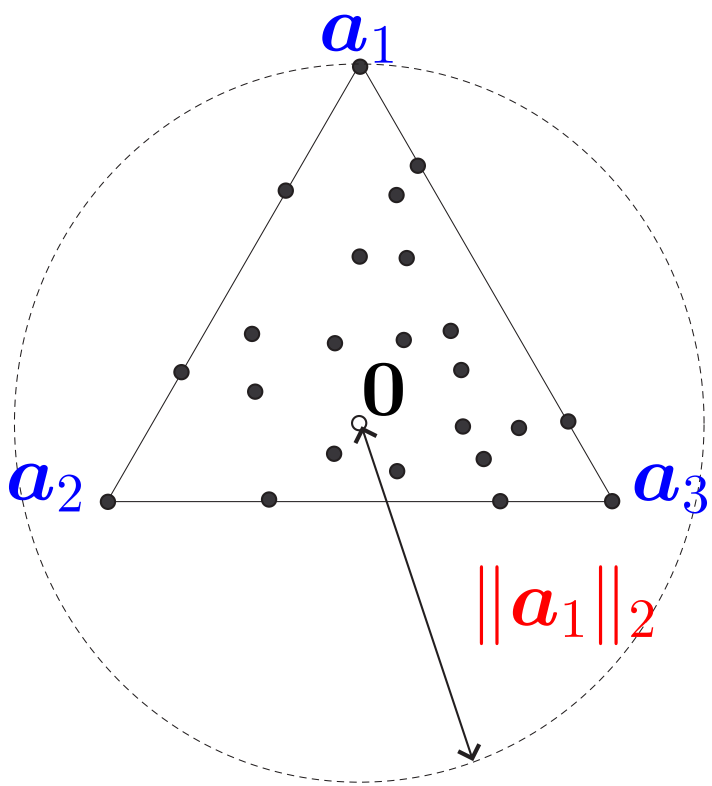
Answer: the pure pixel $\boldsymbol{y}[n] = \boldsymbol{a}_1$ of endmember 1



A Simple Geometrical Question

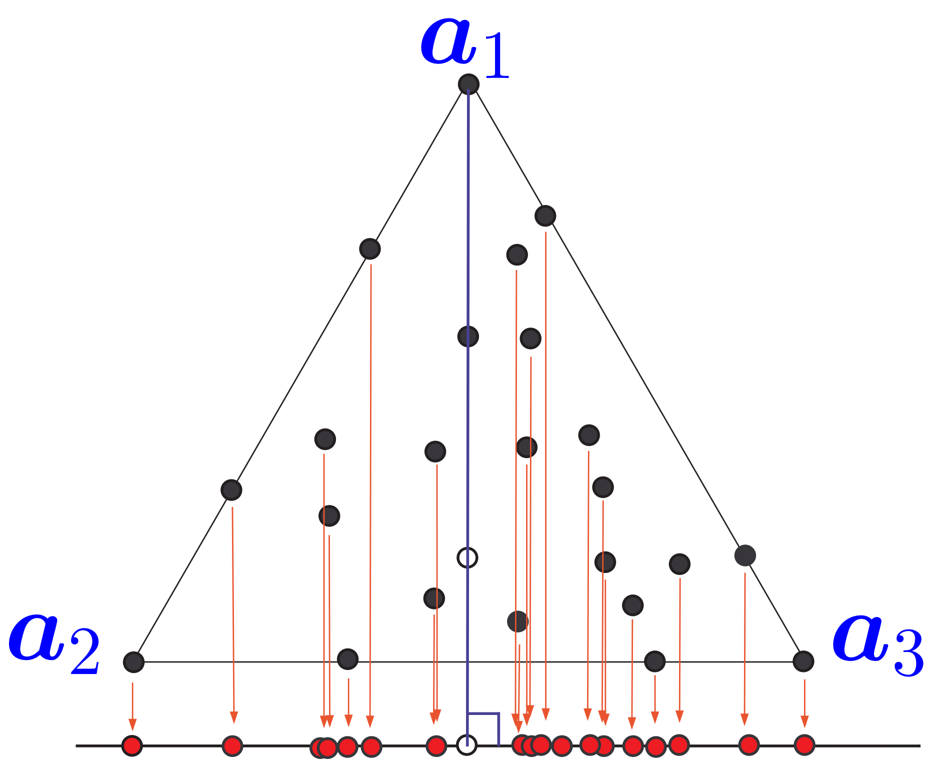
Question: the dark dots are the hyperspectral pixels $\boldsymbol{y}[n]$'s. Which $\boldsymbol{y}[n]$ gives the largest Euclidean norm $\|\boldsymbol{y}[n]\|_2$?

Implication: $\hat{\ell}_1 = \arg \max_{n=1,\ldots,L} \|\boldsymbol{y}[n]\|_2^2$ finds a pure pixel



Another Simple Geometrical Question

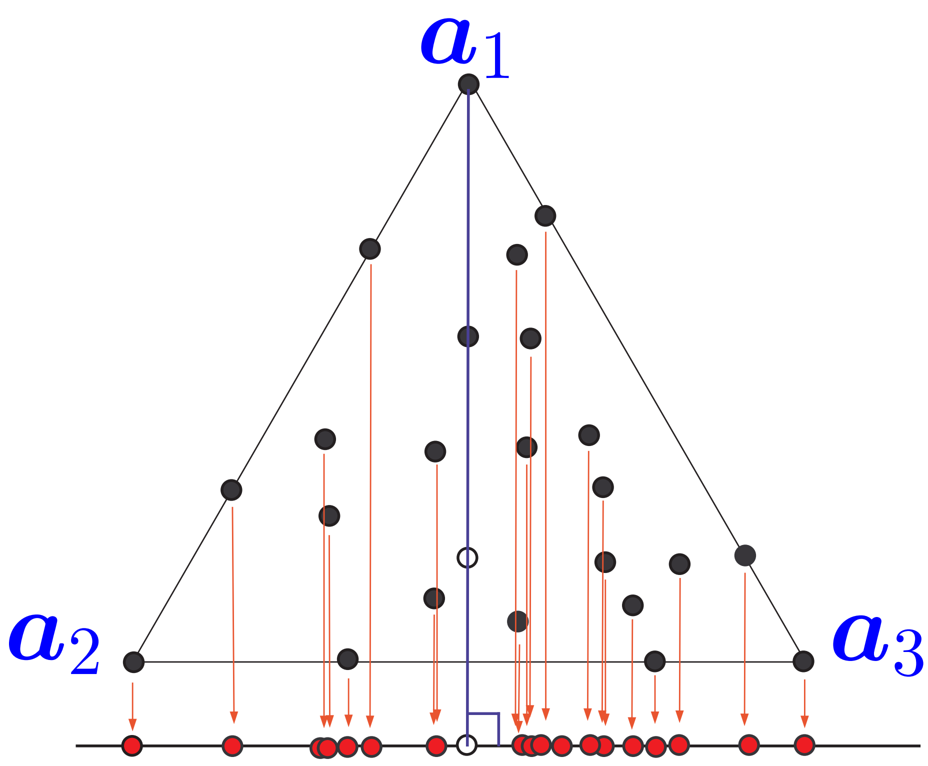
Question: suppose that $\boldsymbol{a}_1$ is known and we project $\boldsymbol{y}[n]$'s onto a line perpendicular to $\boldsymbol{a}_1$. Which $\boldsymbol{y}[n]$ has the largest Euclidean norm on that line?



Another Simple Geometrical Question

Question: suppose that $\boldsymbol{a}_1$ is known and we project $\boldsymbol{y}[n]$'s onto a line perpendicular to $\boldsymbol{a}_1$. Which $\boldsymbol{y}[n]$ has the largest Euclidean norm on that line?

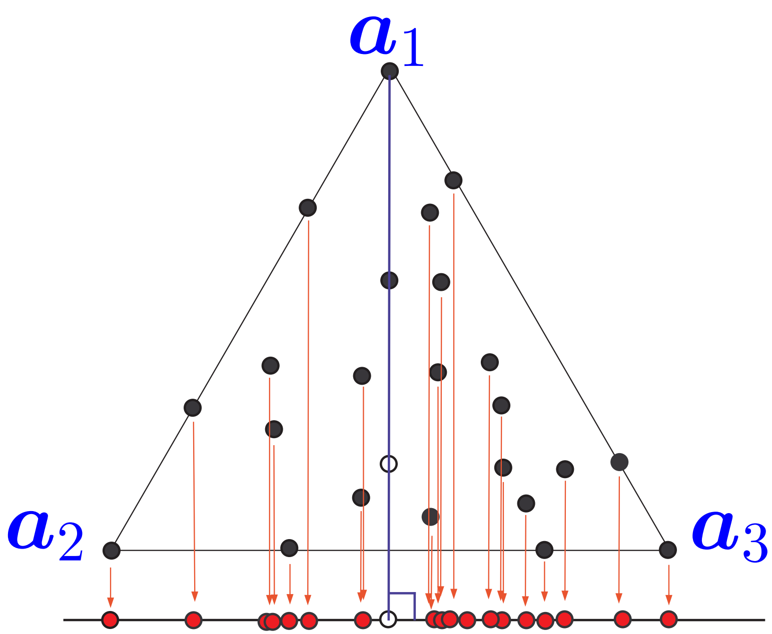
Answer: either the pure pixel $\boldsymbol{y}[n] = \boldsymbol{a}_2$ or $\boldsymbol{y}[n] = \boldsymbol{a}_3$



Another Simple Geometrical Question

Question: suppose that $\boldsymbol{a}_1$ is known and we project $\boldsymbol{y}[n]$'s onto a line perpendicular to $\boldsymbol{a}_1$. Which $\boldsymbol{y}[n]$ has the largest Euclidean norm on that line?

Implication: $\hat{\ell}_2 = \arg \max_{n=1,\ldots,L} \|\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} \boldsymbol{y}[n]\|_2^2$, where $\boldsymbol{P}_{\boldsymbol{a}_1}^{\perp} = \boldsymbol{I} - \boldsymbol{a}_1 \boldsymbol{a}_1^T / \|\boldsymbol{a}_1\|_2^2$ is the orthogonal complement projector of $\boldsymbol{a}_1$, finds a pure pixel



Successive Projection Algorithm (SPA)

Algorithm: SPA
input $\{\boldsymbol{y}[n]\}_{n=1}^{L}$, N.
$\hat{\ell}_1 = \arg \max_{n=1,\ldots,L} \|\boldsymbol{y}[n]\|_2^2$. % find the pixel with the largest Euclidean norm
$\hat{\boldsymbol{A}} = \boldsymbol{y}[\hat{\ell}_1]$.
for $k = 2, \ldots, N$
    % project pixels onto the orthogonal complement subspace of $\hat{\boldsymbol{A}}$, and find the projected pixel with the largest Euclidean norm
    $\hat{\ell}_k = \arg \max_{n=1,\ldots,L} \|\boldsymbol{P}_{\hat{\boldsymbol{A}}}^{\perp}\boldsymbol{y}[n]\|_2^2$, where $\boldsymbol{P}_{\hat{\boldsymbol{A}}}^{\perp} = \boldsymbol{I} - \hat{\boldsymbol{A}}(\hat{\boldsymbol{A}}^T\hat{\boldsymbol{A}})^{-1}\hat{\boldsymbol{A}}^T$
    $\hat{\boldsymbol{A}} := [\ \hat{\boldsymbol{A}},\ \boldsymbol{y}[\hat{\ell}_k]\ ]$.
end
output $\hat{\boldsymbol{A}}$.

- simple algorithm; computationally cheap

SPA is Theoretically Interesting

- guarantee exact recovery in the noiseless case [Chan-Ma-Ambikapathi-Chi2011]

Fact: If the pure pixel assumption holds and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ are linearly independent, SPA recovers $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ exactly in the noiseless case.

- proved to be robust to noise [Gillis-Vavasis2014]

Theorem: If the pure pixel assumption holds and the noise level $\epsilon = \max_{n=1,\ldots,L} \|\boldsymbol{\nu}[n]\|_2$ is sufficient small, SPA recovers $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ up to error $\mathcal{O}(\epsilon\kappa(\boldsymbol{A})^2)$ where $\kappa(\boldsymbol{A})$ is the condition number of $\boldsymbol{A}$.

- many extensions (with provable guarantees), has tight connections to simplex volume maximization [Chan-Ma-Ambikapathi-Chi2011], **self-dictionary sparse regression** [Fu-Ma-Chan-Bioucas2015]

A Self-Dictionary Example: Video Summarization in Computer Vision

The video summarization result shown in [Elhamifar-Sapiro-Vidal2012]. Courtesy to the above reference.

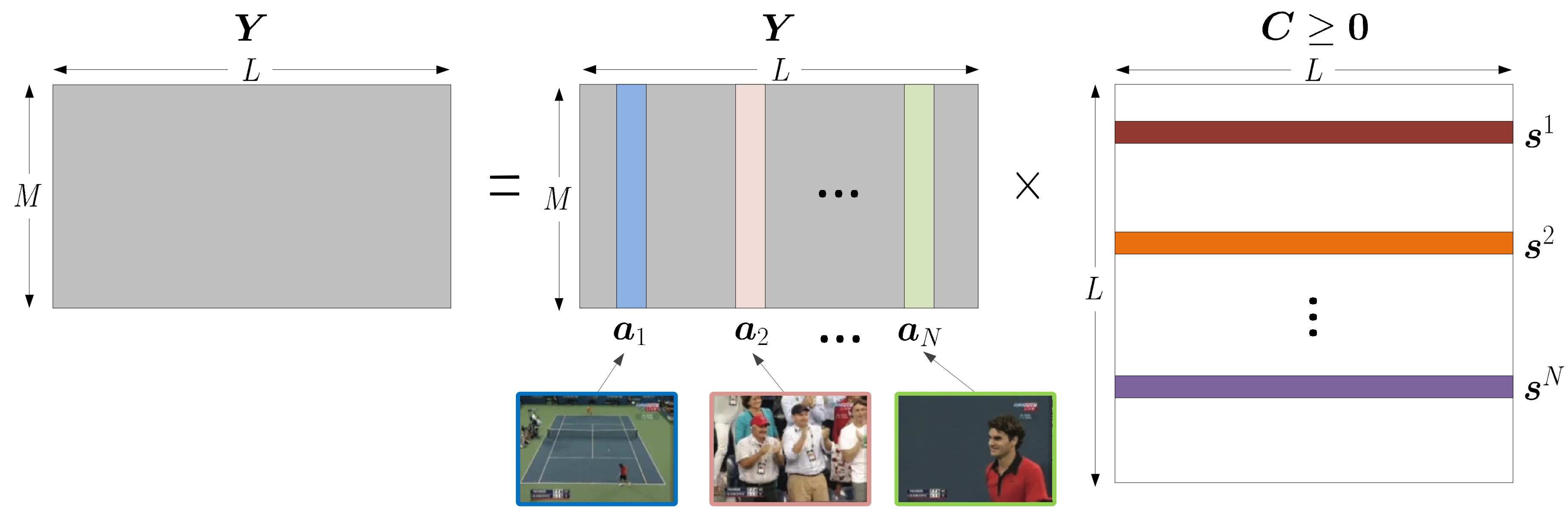
- **Problem:** find a (small) subset of video frames that best represents the whole set of video frames.

Self-Dictionary Sparse Regression

- **Problem:** use a smallest subset of measurements to represent all measurements

$$\min_{\boldsymbol{C}} \ \|\boldsymbol{C}\|_{\rm row-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \ \boldsymbol{C} \geq \boldsymbol{0}, \ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T,$$

where $\|\boldsymbol{C}\|_{\rm row-0}$ counts the number of nonzero rows of $\boldsymbol{C}$.

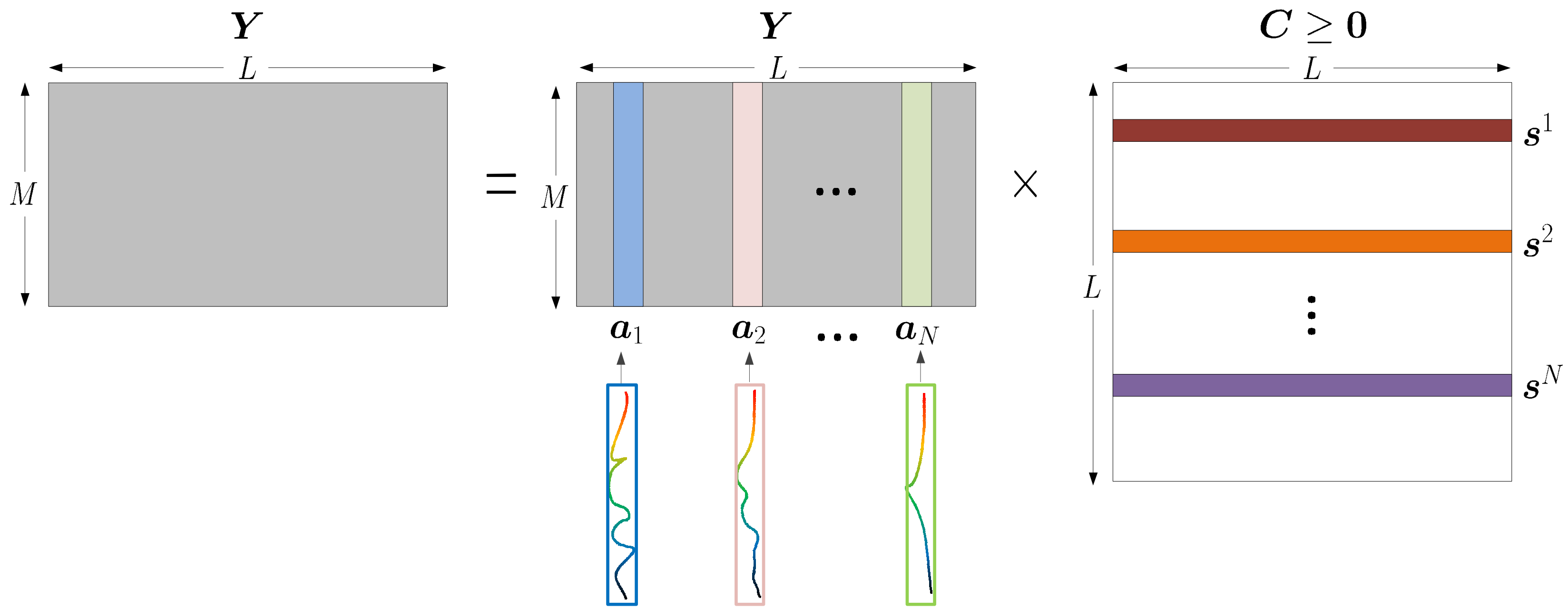


Self-Dictionary Sparse Regression

- **Problem:** use a smallest subset of measurements to represent all measurements

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \ \boldsymbol{C} \geq \boldsymbol{0}, \ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- turns out to be equivalent to pure pixel search— but without requiring knowledge of N [Esser-Moller-Osher-Sapiro-Xin2012]

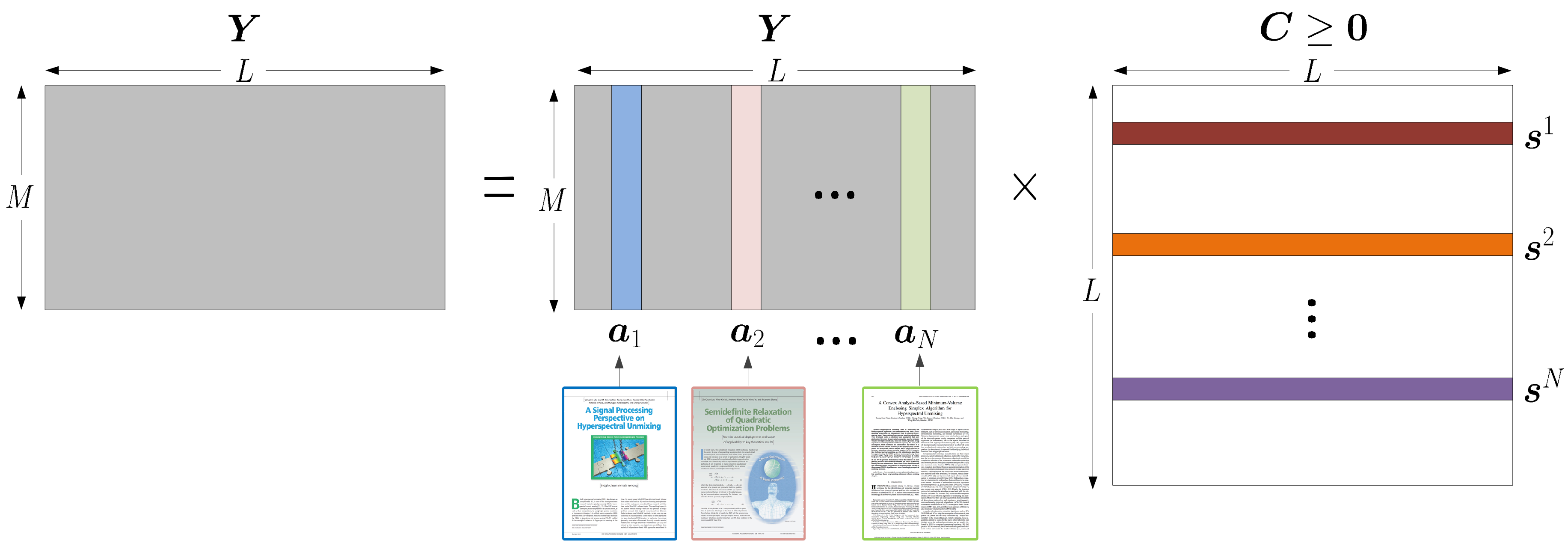


Self-Dictionary Sparse Regression

- **Problem:** use a smallest subset of measurements to represent all measurements

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C},\ \boldsymbol{C} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- same as separable NMF for topic modeling [Arora-Ge-Kannan-Moitra2012]; the separability assumption is the same as the pure pixel assumption



Self-Dictionary Sparse Regression

- **Problem:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C},\ \boldsymbol{C} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

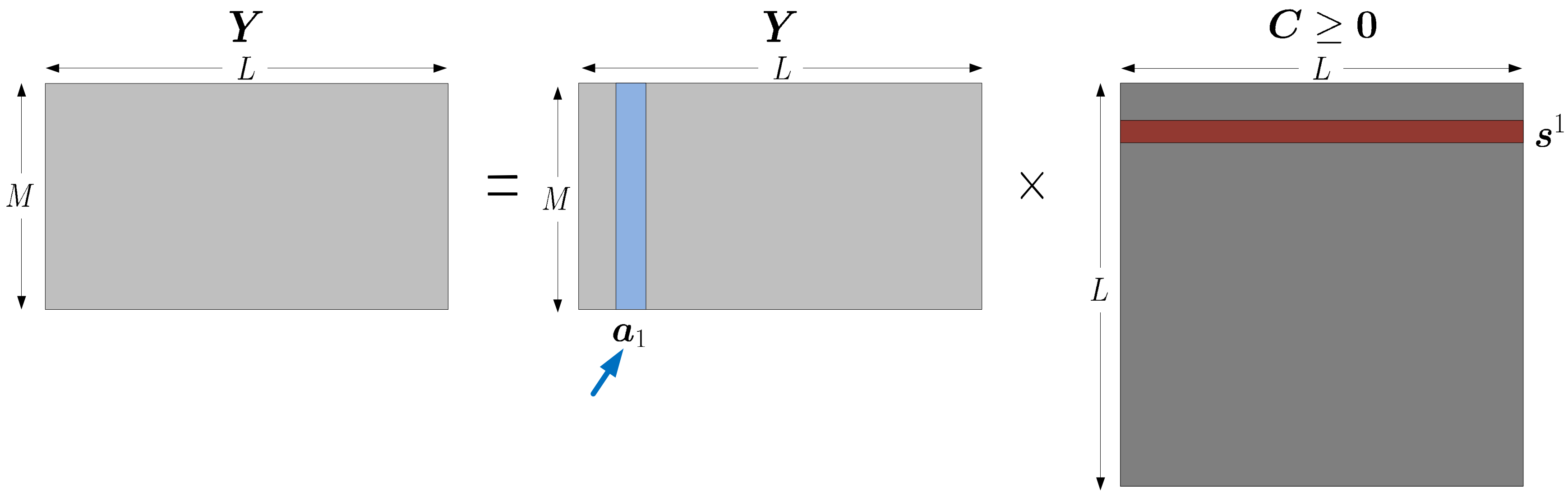
- how to (approximately) solve this problem?
  - convex relaxation: approximate $\|\cdot\|_{\text{row}-0}$ by a convex function, such as $\|\boldsymbol{C}\|_{2,1} = \sum_j \|\boldsymbol{c}^j\|_2$
  - references: [Esser-Moller-Osher-Sapiro-Xin2012], [Elhamifar-Sapiro-Vidal2012], [Recht-Re-Tropp-Bittorf2012], and more...

Self-Dictionary Sparse Regression

- **Problem:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C},\ \boldsymbol{C} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- how to (approximately) solve this problem?
  - greedy pursuit: greedily picks one atom at a time, and repeat



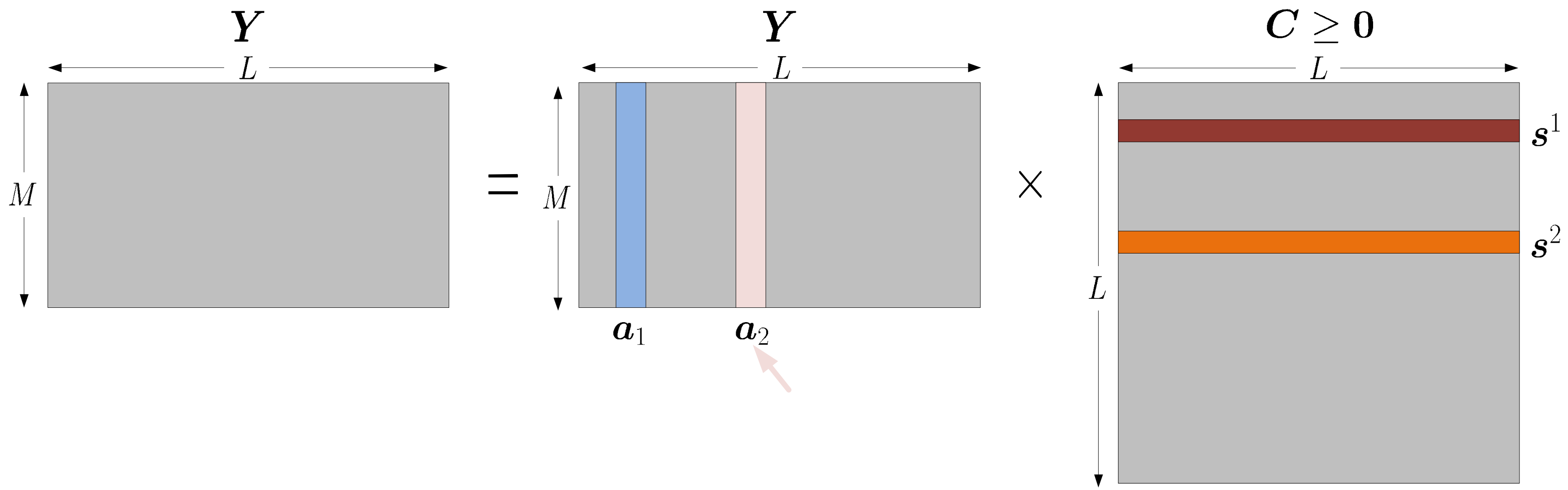
Self-Dictionary Sparse Regression

- **Problem:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$

$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \ \boldsymbol{C} \geq \boldsymbol{0}, \ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- how to (approximately) solve this problem?
  - greedy pursuit: greedily picks one atom at a time, and repeat

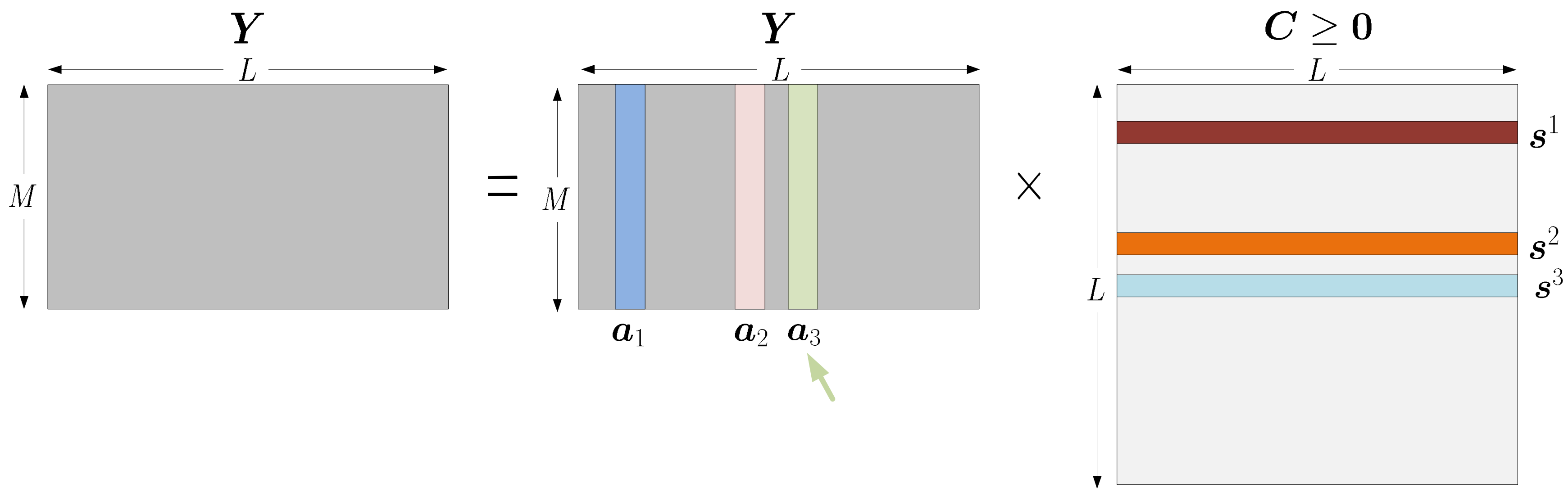


Self-Dictionary Sparse Regression

- **Problem:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C},\ \boldsymbol{C} \geq \boldsymbol{0},\ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- how to (approximately) solve this problem?
  - greedy pursuit: greedily picks one atom at a time, and repeat

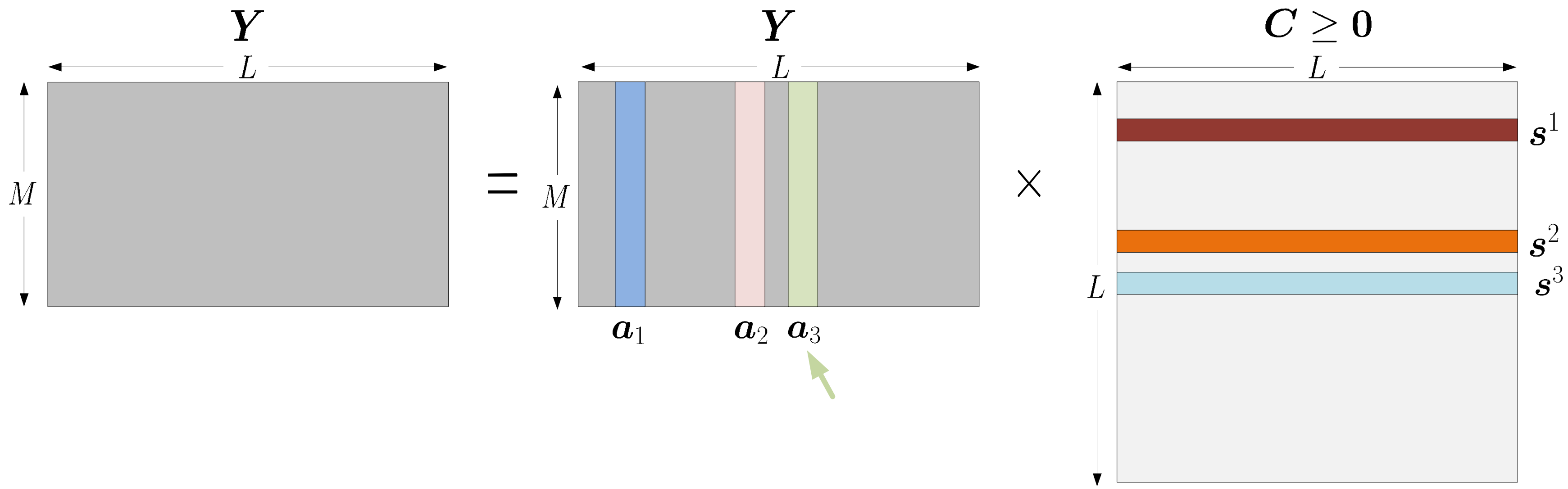


Self-Dictionary Sparse Regression

- **Problem:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{C}\|_{\text{row}-0}$$
$$\text{s.t. } \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C}, \ \boldsymbol{C} \geq \boldsymbol{0}, \ \boldsymbol{1}^T\boldsymbol{C} = \boldsymbol{1}^T.$$

- how to (approximately) solve this problem?
  - greedy pursuit: one instance of greedy pursuit, simultaneous orthogonal matching pursuit, is the same as SPA [Fu-Ma-Chan-Bioucas2015]



Applications Beyond HU Itself

Who Invented Convex Geometry?

- Craig is most widely recognized for introducing CG in hyperspectral remote sensing [Craig1990], [Craig1994]
  - worth mentioning: Boardman and Winter for pioneering pure pixel search [Boardman-Kruse-Green1995], [Winter1999]
- intriguingly, CG has been discovered or rediscovered several times in other fields
  - geology [Imbrie1964], also [Full-Ehrlich-Klovan1981]
  - chemometrics [Perczel et al. 1989]
  - nuclear magnetic resonance spectroscopy [Naanaa-Nuzillard2005]
  - signal processing theory and methods [Chan-Ma-Chi-Wang2008]
  - (in a way) machine learning [Arora-Ge-Kannan-Moitra2012]

Applications Beyond HU

- machine learning and data science: topic modeling, as mentioned; community detection; crowdsourcing

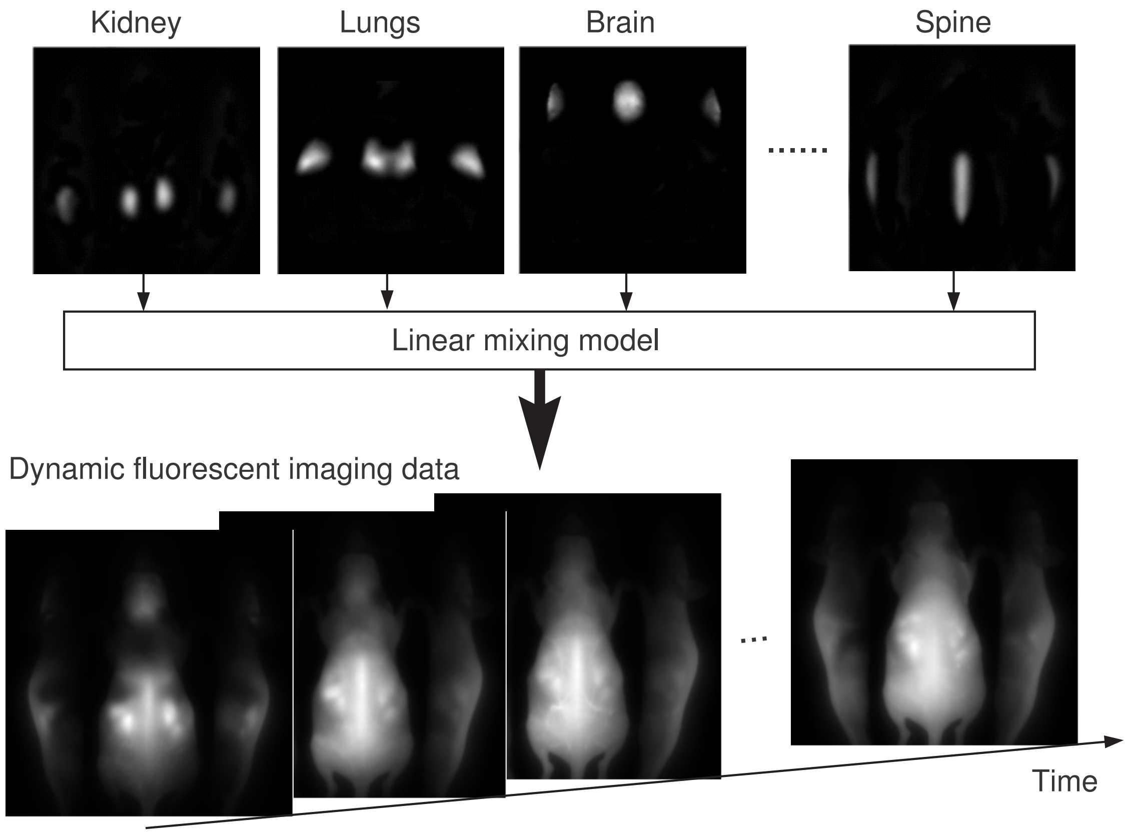
- biomedical imaging

- signal processing: classical blind source separation

- remote sensing: hyperspectral super-resolution

- many more...

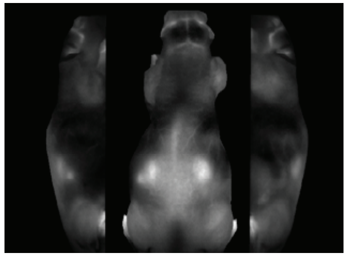
Dynamic Fluorescent Imaging (DFI) and My CG in 2008



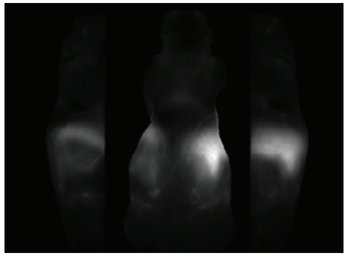
- DFI images are linear mixtures of the anatomical maps of different organs
- model: $\boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{s}[n]$, $\boldsymbol{s}[n] \geq \mathbf{0}$, $\mathbf{1}^T\boldsymbol{A} = \mathbf{1}^T$ (not $\mathbf{1}^T\boldsymbol{s}[n] = 1$ as in HU)

DFI and My CG in 2008

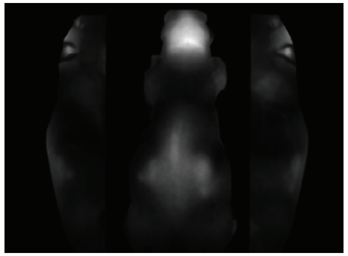
DFI and My CG in 2008



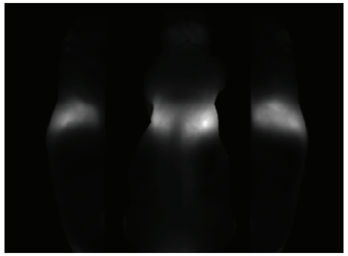
Kidney



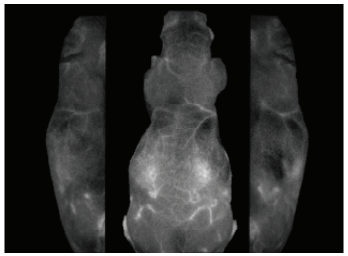
Liver



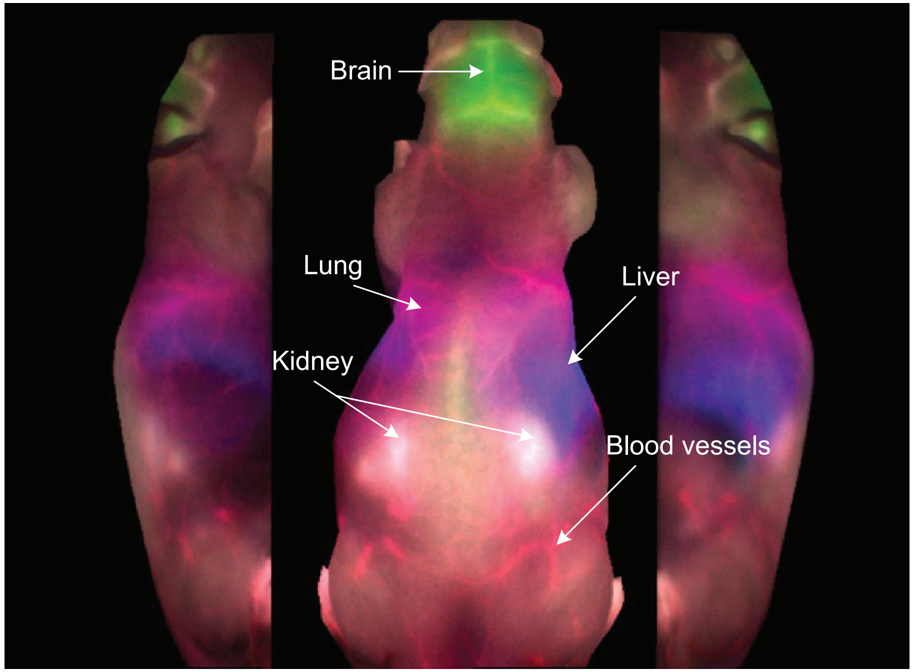
Brain



Lung



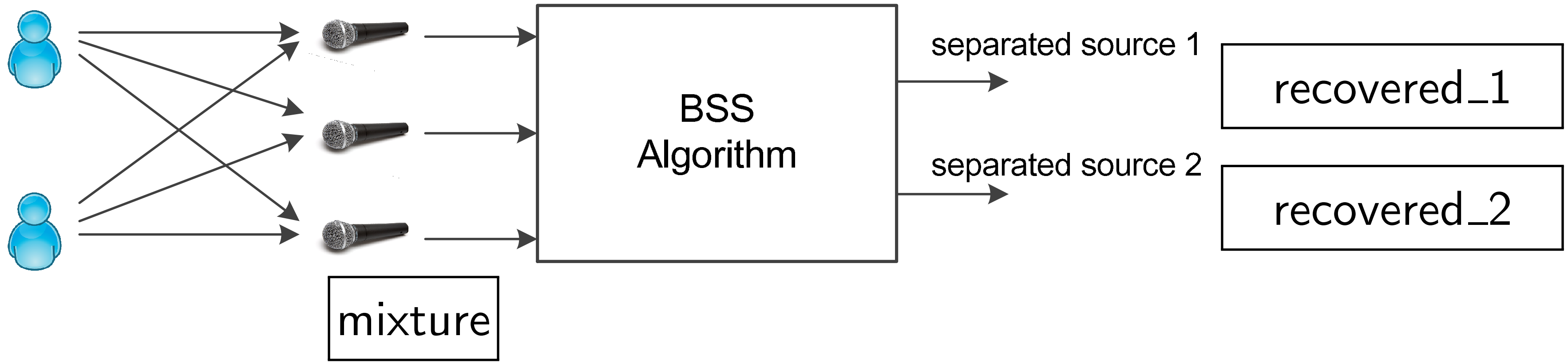
Blood vessels



Anatomical map

- anatomical maps recovered by a CG method [Chan-Ma-Chi-Wang2008]

Blind Source Separation (BSS)

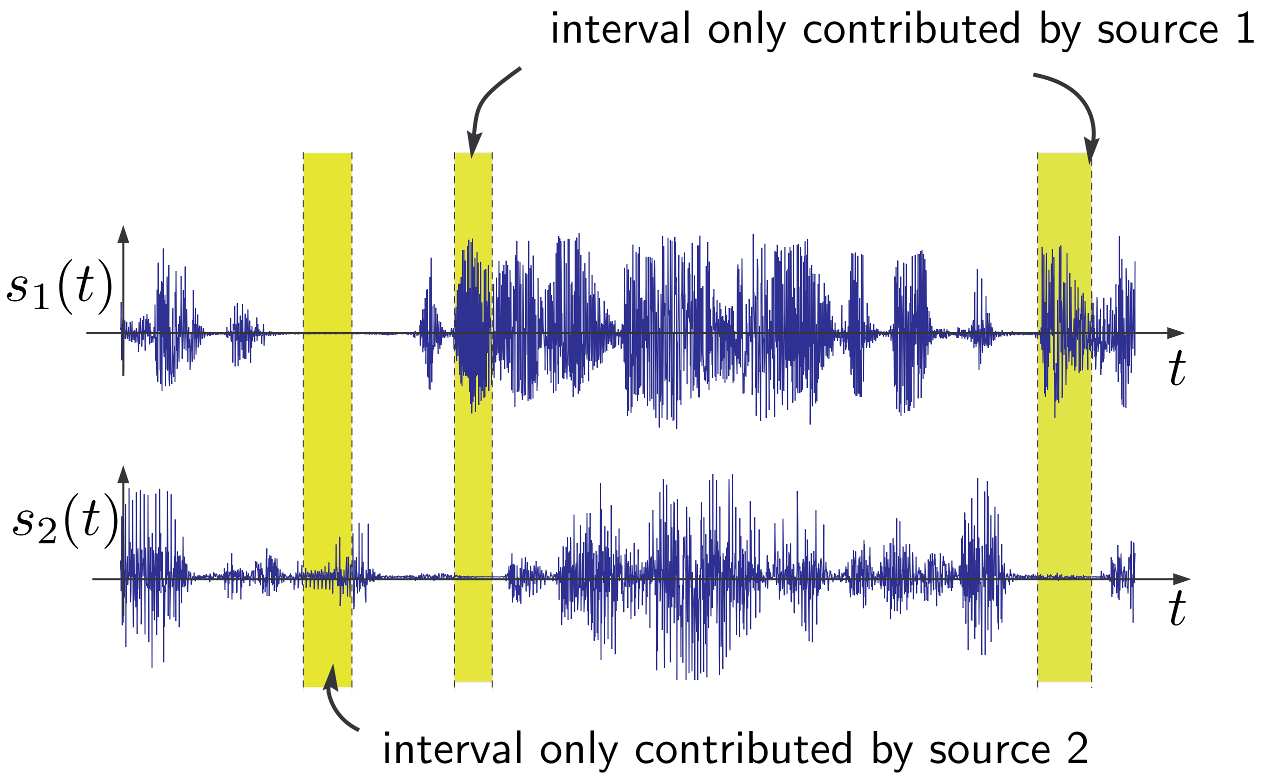


- a classic problem in signal processing, similar model and problem statement as HU

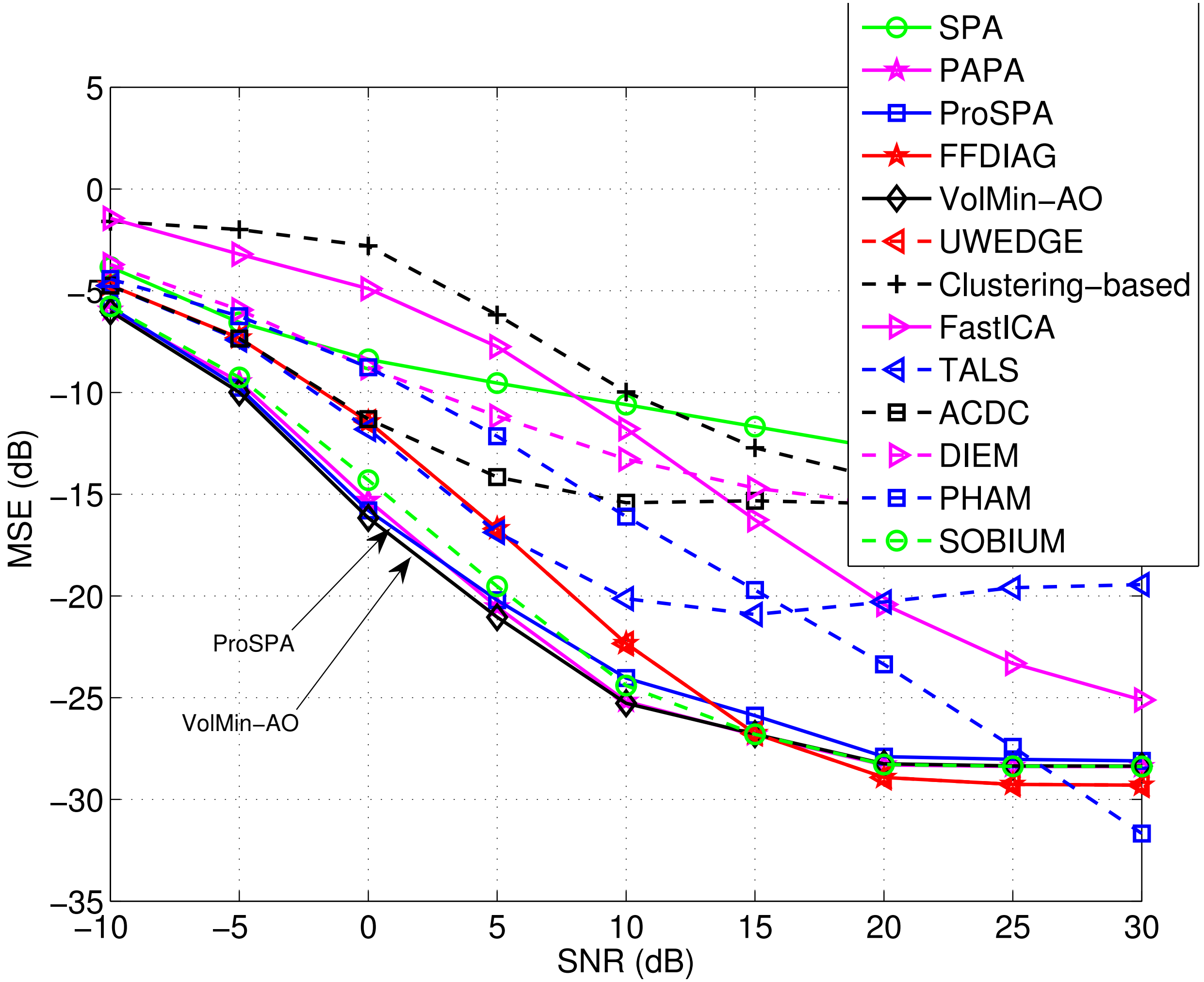
- can we apply convex geometry to BSS?

Pure Pixels and Classical BSS

- Our work in [Fu-Ma-Huang-Sidiropoulos2015]:
  - hypothesis: existence of pure short-time frames; reasonable for speech
  - formulation: a tensor factorization problem with one factor having pure pixels



Pure Pixels and Classical BSS

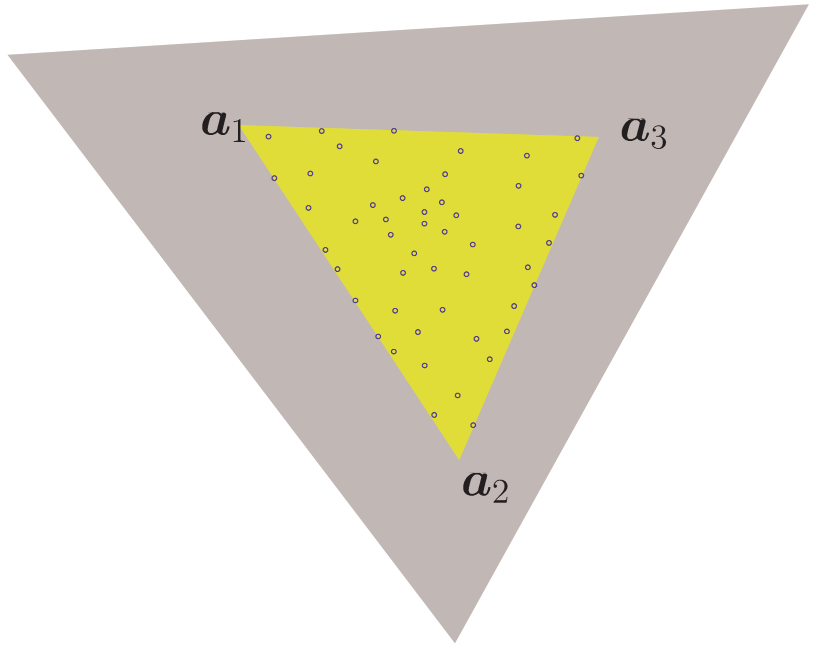


Performance comparison of various BSS algorithms. 'ProSPA' is a modified version of SPA, custom-designed for the blind speech separation application.

Simplex Volume Minimization

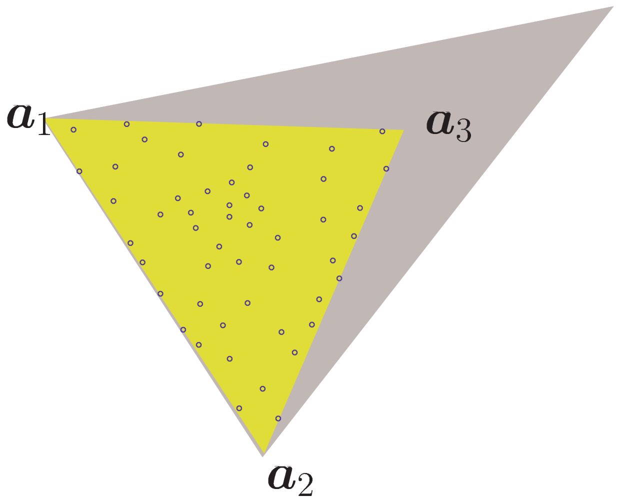
Simplex Volume Minimization: Intuition

Craig's belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.



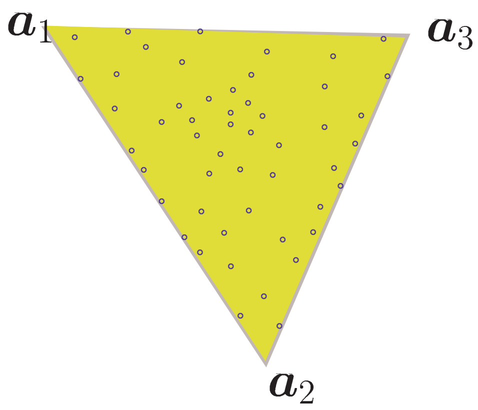
Simplex Volume Minimization: Intuition

Craig's belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.



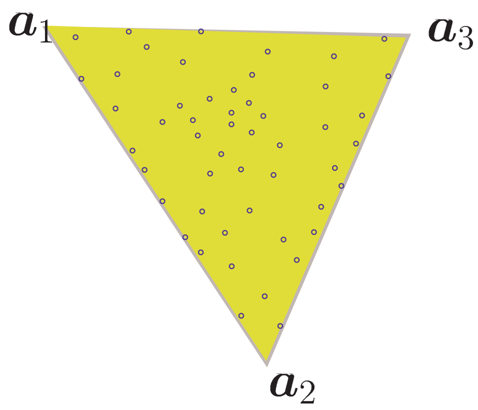
Simplex Volume Minimization: Intuition

Craig's belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.



Simplex Volume Minimization: Intuition

Craig's belief [Craig1994]: the true endmembers may be located by finding a data enclosing simplex whose volume is the smallest.

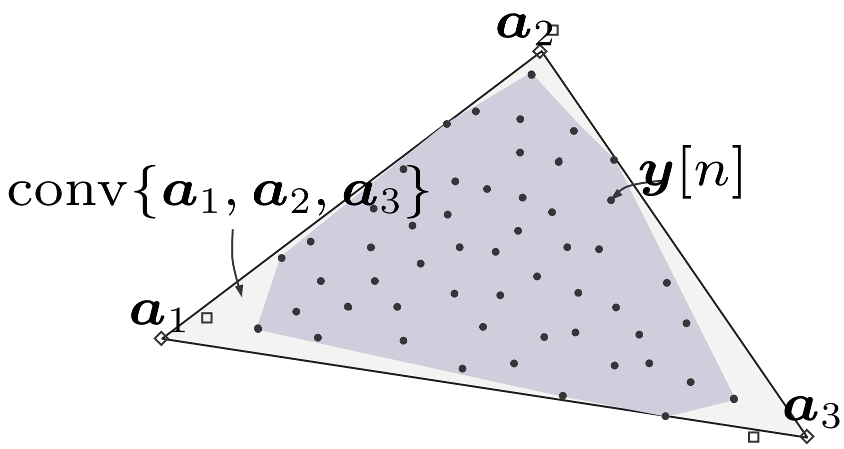


- **it seems volume min. (VolMin) can identify the true endmembers without pure pixels.**

Simplex Volume Minimization: Formulation

- **Formulation:**

$$\begin{aligned} \min_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N \in \mathbb{R}^M} \quad & \text{vol}(\boldsymbol{A}) \\ \text{s.t.} \quad & \boldsymbol{y}[n] \in \text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}, \\ & n = 1, \ldots, L. \end{aligned}$$



where

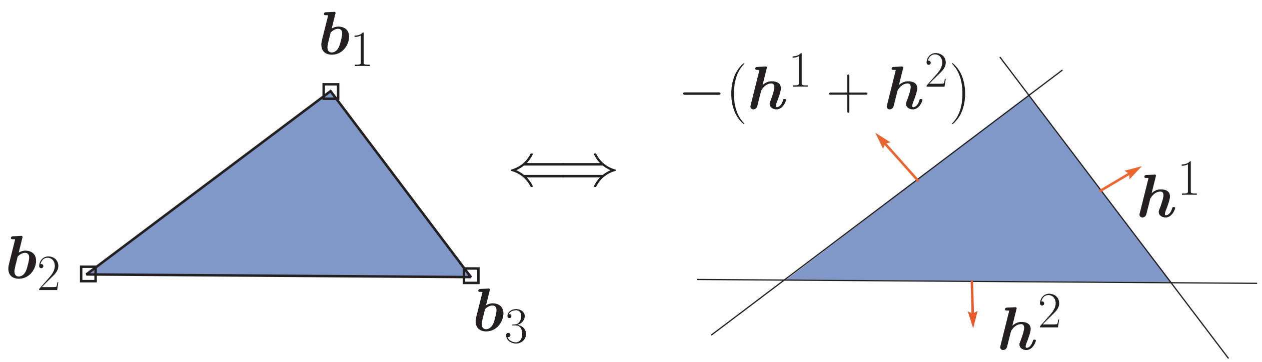
$$\text{vol}(\boldsymbol{A}) = \frac{1}{(N-1)!}\sqrt{\det(\bar{\boldsymbol{A}}^T \bar{\boldsymbol{A}})}, \quad \bar{\boldsymbol{A}} = [\ \boldsymbol{a}_1 - \boldsymbol{a}_N, \ldots, \boldsymbol{a}_{N-1} - \boldsymbol{a}_N\ ],$$

- non-convex, NP-hard [Packer2008]
- algorithms rely on non-convex optimization

Simplex Volume Minimization: Optimization

- dimensionally reduce $\{\boldsymbol{y}[n]\}$ to $\{\boldsymbol{x}[n]\}$ such that $\boldsymbol{x}[n] = \boldsymbol{B}\boldsymbol{s}[n]$, $\boldsymbol{B} \in \mathbb{R}^{N-1 \times N}$.
- by transformation of a simplex to a polyhedron, recast VolMin as

$$\begin{aligned} &\min_{\boldsymbol{B}} \ \mathrm{vol}(\boldsymbol{B}) \\ &\text{s.t. } \boldsymbol{x}[n] \in \mathrm{conv}\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}, \forall n \end{aligned} \iff \begin{aligned} &\max_{\boldsymbol{H},\boldsymbol{g}} \ |\det(\boldsymbol{H})| \\ &\text{s.t. } \boldsymbol{H}\boldsymbol{x}[n] - \boldsymbol{g} \geq \boldsymbol{0}, \\ &\qquad (\boldsymbol{H}\boldsymbol{x}[n] - \boldsymbol{g})^T \boldsymbol{1} \leq 1, \forall n \end{aligned}$$



where $\boldsymbol{H} = [\boldsymbol{b}_1 - \boldsymbol{b}_N, \ldots, \boldsymbol{b}_{N-1} - \boldsymbol{b}_N]^{-1}$, $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{b}_N$.

- algorithms: [Li-Bioucas2008], [Chan-Chi-Huang-Ma2009], [Bioucas2009]

Simplex Volume Minimization and Matrix Factorization

- recall the VolMin problem

$$\min_{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N} \mathrm{vol}(\boldsymbol{A}) \qquad \text{s.t. } \boldsymbol{y}[n] \in \mathrm{conv}\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N\}, \quad n = 1,\ldots,L$$

- by noting $\boldsymbol{y}[n] \in \mathrm{conv}\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_N\} \Longleftrightarrow \boldsymbol{y}[n] = \boldsymbol{A}\boldsymbol{s}_n$ for some $\boldsymbol{s}_n \geq \boldsymbol{0}, \boldsymbol{s}_n^T\boldsymbol{1} = 1$, VolMin can be equivalently written as

$$\min_{\boldsymbol{A},\boldsymbol{S}} \mathrm{vol}(\boldsymbol{A}) \qquad \text{s.t. } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{S},\ \boldsymbol{S} \geq \boldsymbol{0},\ \boldsymbol{S}^T\boldsymbol{1} = \boldsymbol{1}$$

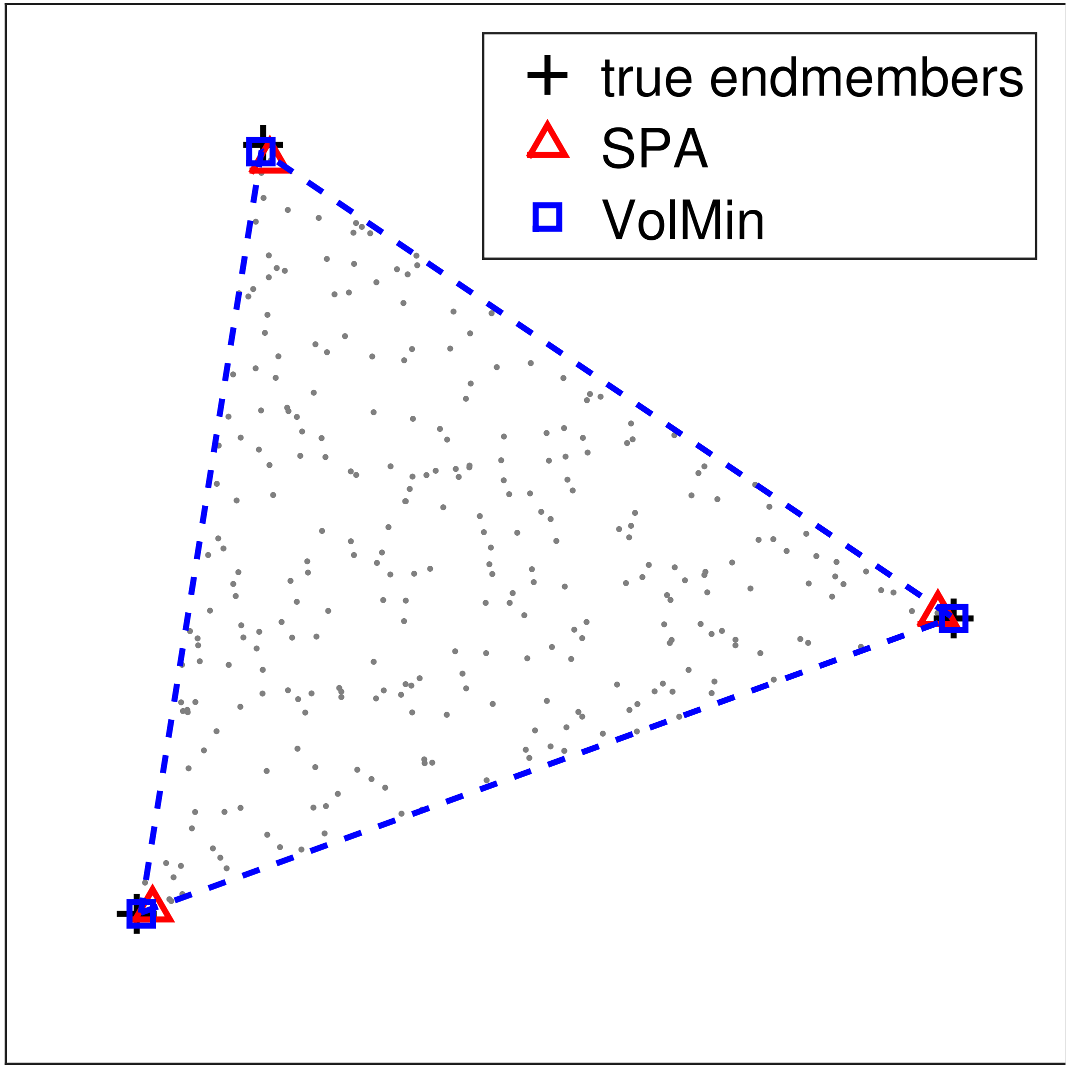
- or, a regularized form may be considered:

$$\min_{\boldsymbol{A},\boldsymbol{S}\geq\boldsymbol{0},\boldsymbol{S}^T\boldsymbol{1}=\boldsymbol{1}} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{S}\|_F^2 + \lambda \cdot \mathrm{vol}(\boldsymbol{A}); \qquad \lambda > 0 \text{ is given,}$$

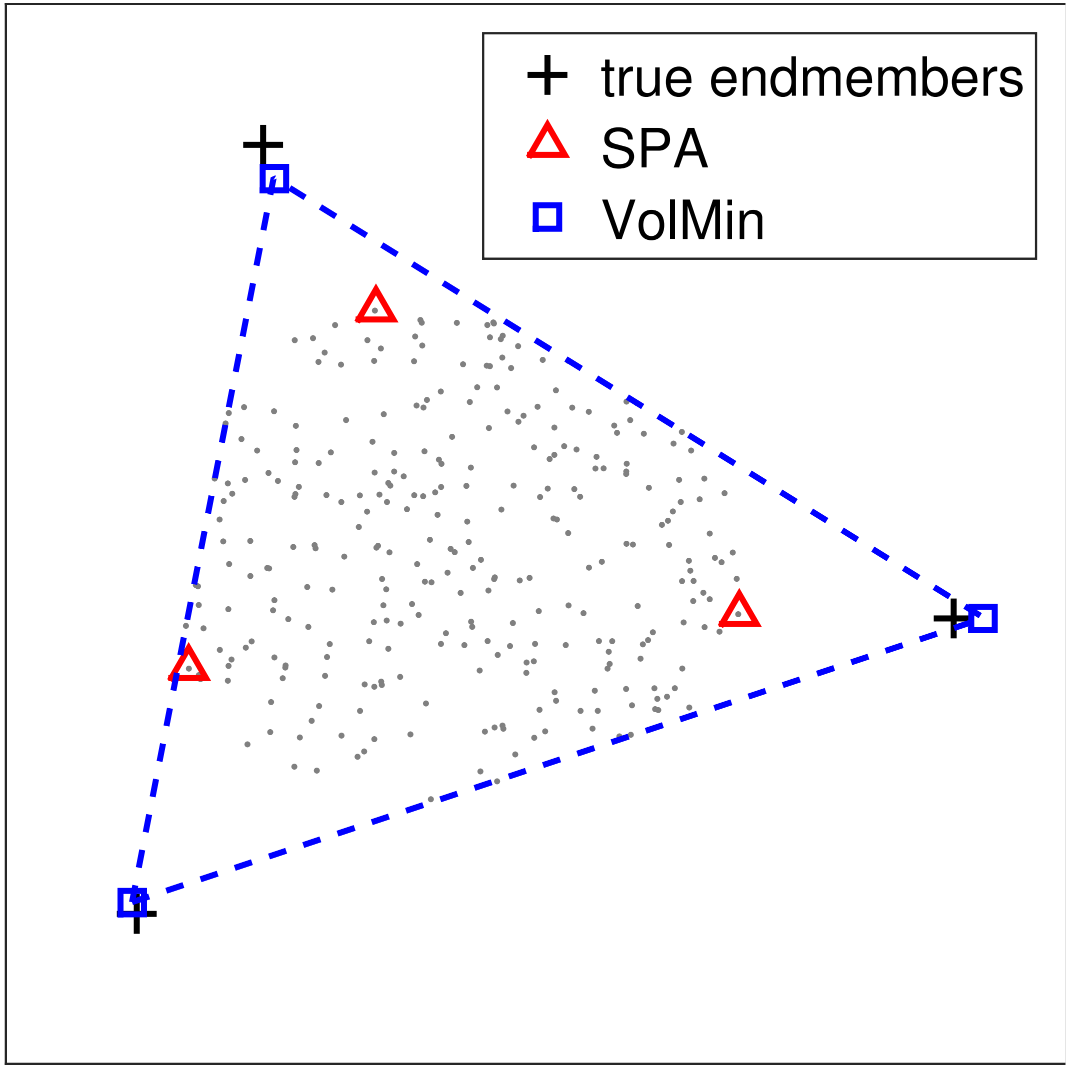
  which looks like a volume-regularized semi-NMF

  – algorithms: [Miao-Qi2007], [Fu-Huang-Yang-Ma-Sidiropoulos2016]

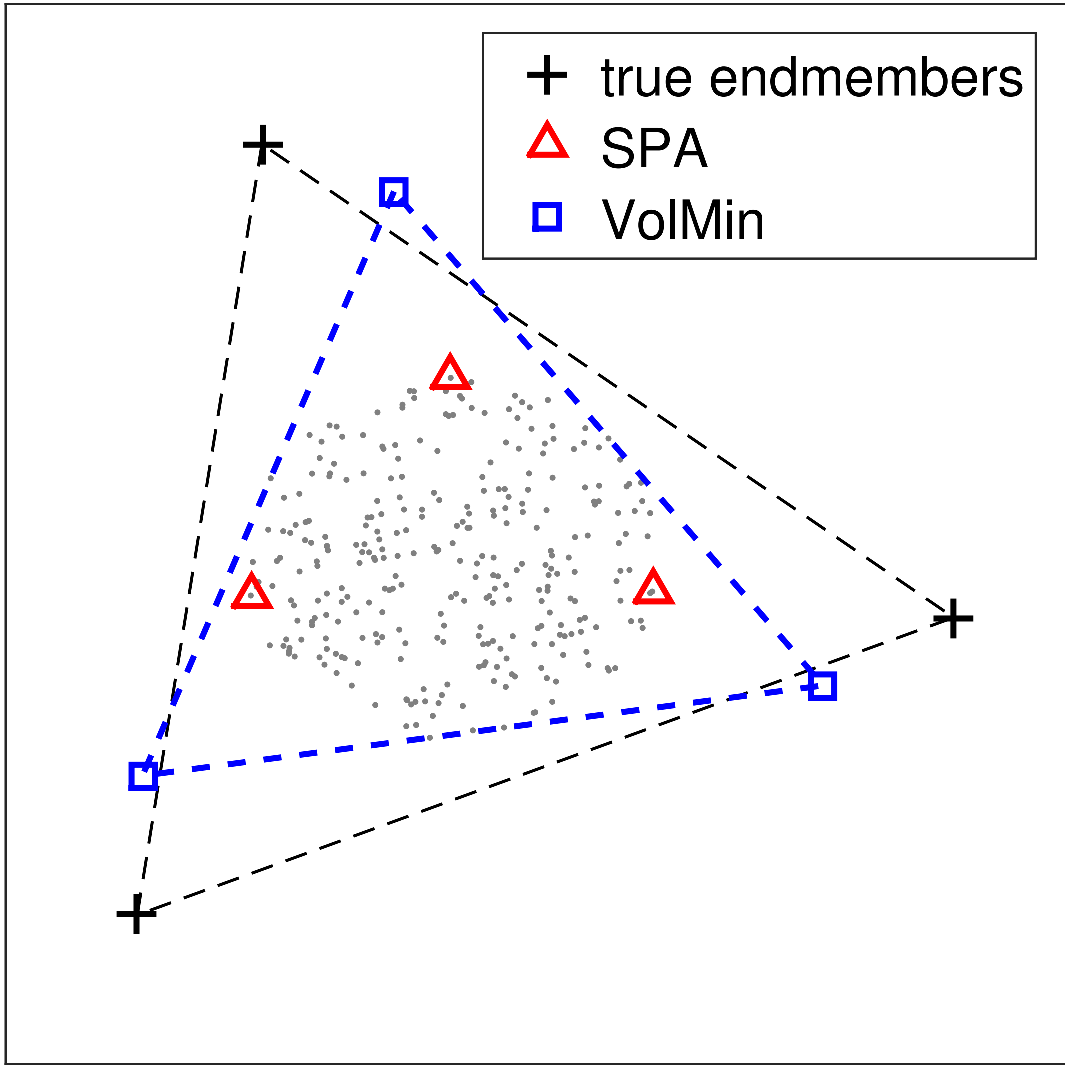
Numerical Demo.: Three Endmembers, Pure Pixel Case



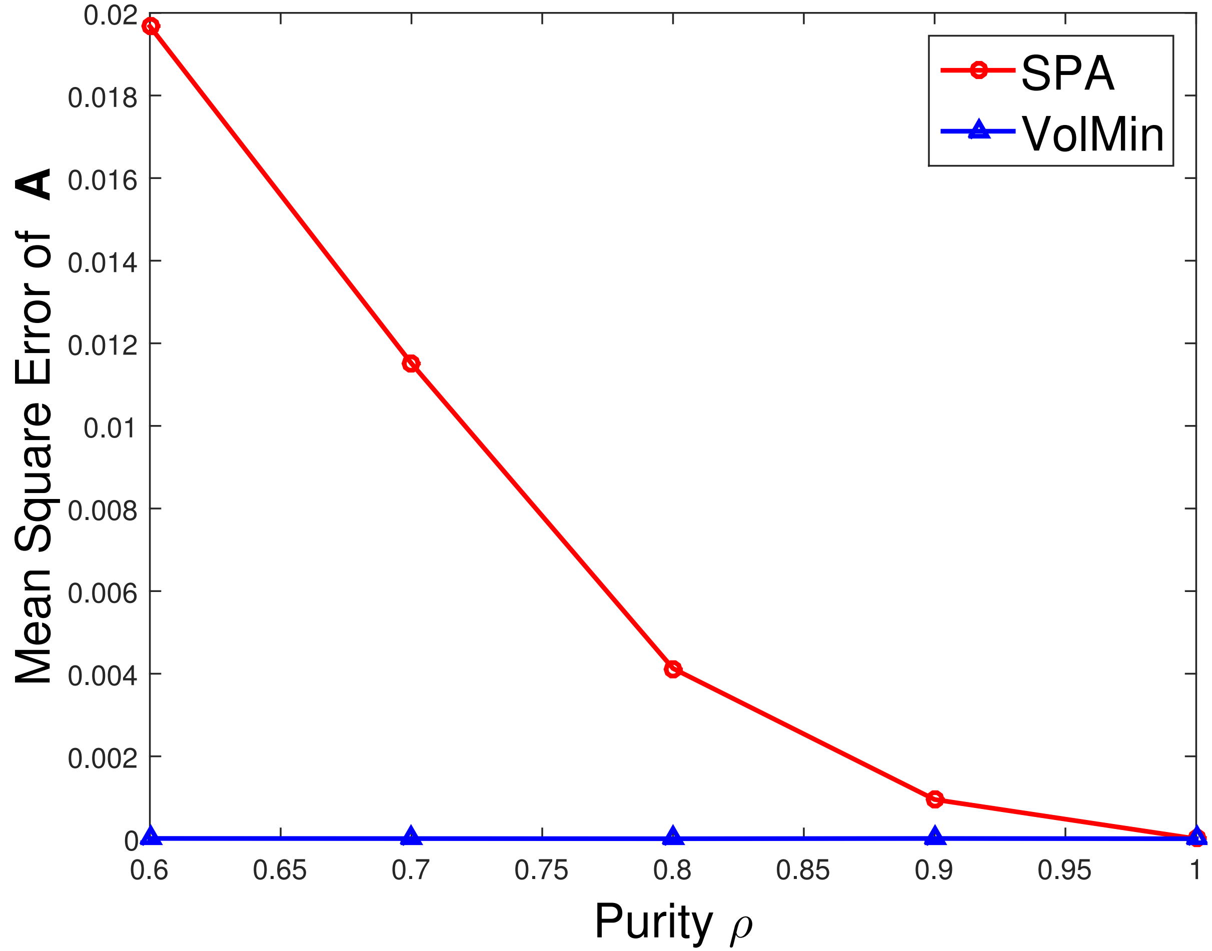
Numerical Demo.: Three Endmembers, No-Pure Pixel Case



Numerical Demo.: Three Endmembers, No-Pure Pixel Case



Simulation Results: Mean Squared Error Comparison



A Monte Carlo result. $N = 8$. "Purity ρ" describes the pixel purity: $\rho = 1$ corresponds to the pure pixel case, and $\rho = 1/\sqrt{N}$ the most heavily mixed case.

Unique Identifiability of VolMin

- numerical evidence suggests that VolMin works well without pure pixels

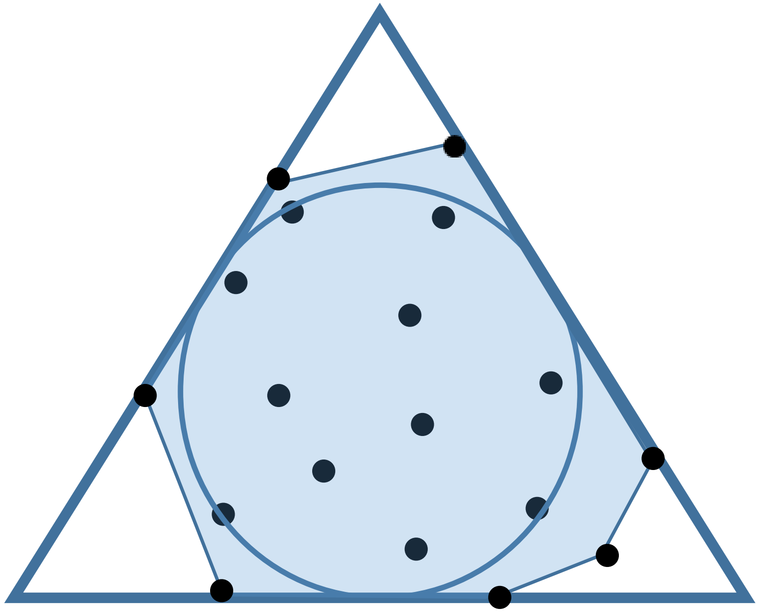
- **Question:** can VolMin uniquely recover the true endmembers—provably?

Unique Identifiability of VolMin

Theorem [Lin-Ma-Li-Chi-Ambikapathi2015]: Suppose no noise, $N \geq 3$ and that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ are linearly independent. Define

$$\gamma = \max\left\{r \leq 1 \mid (\mathrm{conv}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}) \cap \mathcal{B}(r) \subseteq \mathrm{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L\}\right\},$$

where $\mathcal{B}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_2 \leq r\}$. VolMin exactly recovers $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ if $\gamma > \frac{1}{\sqrt{N-1}}$.



Unique Identifiability of VolMin

Theorem [Lin-Ma-Li-Chi-Ambikapathi2015]: Suppose no noise, $N \geq 3$ and that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ are linearly independent. Define

$$\gamma = \max\left\{ r \leq 1 \mid (\text{conv}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}) \cap \mathcal{B}(r) \subseteq \text{conv}\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_L\} \right\}, \qquad (*)$$

where $\mathcal{B}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\|_2 \leq r\}$. VolMin exactly recovers $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N$ if $\gamma > \frac{1}{\sqrt{N-1}}$.

- much more relaxed than the pure pixel assumption (and separable NMF)
- (surprisingly) much more relaxed than known NMF recovery conditions
  - [Donoho-Stodden2003]: require the pure pixel assumption
  - [Huang-Sidiropoulos-Swami2014]: require *both* $\boldsymbol{A}$ and $\boldsymbol{S}$ to satisfy $(*)$, roughly speaking
- comparison: VolMin requires *only* $\boldsymbol{S}$ to satisfy $(*)$; more recent VolMin work even removes the condition of $\mathbf{1}^T\boldsymbol{s}[n] = 1$ [Fu-Huang-Sidiropoulos2018]

So, What's Next? Our Ongoing Research

- consider statistical inference akin to probabilistic PCA
- assume that $\boldsymbol{s}[n]$'s are i.i.d. distributed on the unit simplex
  - probabilistic PCA assumes i.i.d. Gaussian $\boldsymbol{s}[n]$'s
- in the noiseless case, the marginalized likelihood is

$$p_{\boldsymbol{A}}(\boldsymbol{y}[n]) = \int p_{\boldsymbol{A}}(\boldsymbol{y}[n]|\boldsymbol{s}[n])p(\boldsymbol{s}[n])d\boldsymbol{s}[n] = \frac{1}{\text{vol}(\boldsymbol{A})}\mathbb{1}_{\mathcal{A}}(\boldsymbol{y}[n])$$

where $\mathcal{A} = \text{conv}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\}$; $\mathbb{1}_{\mathcal{A}}(\boldsymbol{y}) = 1$ if $\boldsymbol{y} \in \mathcal{A}$, $\mathbb{1}_{\mathcal{A}}(\boldsymbol{y}) = 0$ if $\boldsymbol{y} \notin \mathcal{A}$

- maximum-likelihood estimator (alluded to in [Nascimento-Bioucas2012]):

$$\max_{\boldsymbol{A}} \sum_{n=1}^{L} \log(p_{\boldsymbol{A}}(\boldsymbol{y}[n])) = \max_{\boldsymbol{A}} -\log(\text{vol}(\boldsymbol{A})) \text{ s.t. } \boldsymbol{y}[n] \in \mathcal{A}\ \forall n = \text{VolMin}$$

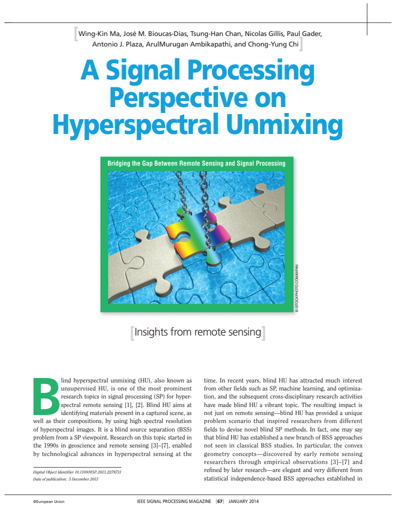
- ongoing work: the noisy case; challenge: intractable marginal likelihood

Conclusion and Discussion

- what are the great insights to learn from HU in remote sensing?
  - convex geometry, pure pixel search, volume minimization
- other than hyperspectral imaging, what is worthwhile to note?
  - its connections to important problems in machine learning and data science

Thank You! Main References of This Talk

W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, *et al.*, "A signal processing perspective on hyperpsectral unmixing," *IEEE SP Mag.*, Jan. 2014.



N. Gillis, "The why and how of nonnegative matrix factorization," in *Regularization, Optimization, Kernels, and Support Vector Machines*, 2014.

X. Fu, K. Huang, N. D. Sidiropoulos, and W.-K. Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE SP Mag.*, 2019.

References

Introduction:

[Donoho-Stodden2003] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Advances in Neural Information Processing Systems*, 2003, pp. 1141–1148.

[Huang-Sidiropoulos-Swami2014] K. Huang, N. D. Sidiropoulos and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 211–224, 2014.

[Craig1994] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

Pure Pixel Search:

[Boardman-Kruse-Green1995] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summ. JPL Airborne Earth Sci. Workshop*, vol. 1, Pasadena, CA, Dec. 1995, pp. 23–26.

[Nascimento-Bioucas2003] J. Nascimento and J. M. Bioucas-Dias, "Vertex component analysis: A fast algorithm to extract endmembers spectra from hyperspectral data," in *Pattern Recognition and Image Analysis: 1st Iberian Conf.*, 2003, pp. 626–635.

[Chan-Ma-Ambikapathi-Chi2011] T.-H. Chan, W.-K. Ma, A. Ambikapathi, and C.-Y. Chi, "A simplex volume maximization framework for hyperspectral endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4177–4193, 2011.

[Gillis-Vavasis2014] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, April 2014.

[Fu-Ma-Chan-Bioucas2015] X. Fu, W.-K. Ma, T.-H. Chan, and J. M. Bioucas-Dias, "Self-dictionary sparse regression for hyper spectral unmixing: Greedy pursuit and pure pixel search are related," *J. Select. Topics Signal Process.*, Sep. 2015.

[Elhamifar-Sapiro-Vidal2012] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE CVPR 2012*, 2012, pp. 1600–1607.

[Arora-Ge-Kannan-Moitra2012] S. Arora, R. Ge, R. Kannan, and A. Moitra, "Computing a nonnegative matrix factorization—Provably," in *Proc. 44th Symp. Theory of Computing*, ACM, 2012, pp. 145–162.

[Recht-Re-Tropp-Bittorf2012] B. Recht, C. Re, J. Tropp, and V. Bittorf, "Factoring nonnegative matrices with linear programs," in *Advances in Neural Information Processing Systems*, 2012, pp. 1214–1222.

[Esser-Moller-Osher-Sapiro-Xin2012] E. Esser, M. Moller, S. Osher, G. Sapiro, and J. Xin, "A convex model for nonnegative matrix factorization and dimensionality reduction on physical space," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3239–3252, Jul. 2012.

History of Convex Geometry and Applications:

[Craig1990] M. D. Craig, "Unsupervised unmixing of remotely sensed images," in *Proc. Australasian Remote Sens. Conf.*, Perth, Australia, 1990, pp. 324–330.

[Craig1994] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[Boardman-Kruse-Green1995] J. Boardman, F. Kruse, and R. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Proc. Summ. JPL Airborne Earth Sci. Workshop*, vol. 1, Pasadena, CA, Dec. 1995, pp. 23–26.

[Winter1999] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end- member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry*, Pasadena, CA, Oct. 1999, pp. 266–275.

[Imbrie1964] J. Imbrie, "Vector analysis of heavy-mineral data," *Geological Society of America Bulletin*, vol. 75, pp. 1131–1156, Nov. 1964.

[Full-Ehrlich-Klovan1981] W. E. Full, R. Ehrlich, and J. E. Klovan, "EXTENDED QMODEL—objective definition of external endmembers in the analysis of mixtures," *Mathematical Geology*, vol. 13, no. 4, pp. 331–344, 1981.

[Perczel et al. 1989] A. Perczel, M. Hollosi, G. Tusnady, and D. Fasman, "Convex constraint decomposition of circular dichroism curves of proteins,"' *Croatica Chim. Acta*, vol. 62, pp. 189–200, 1989.

[Naanaa-Nuzillard2005] "Blind source separation of positive and partially correlated data," *Signal Processing*, vol. 85, no. 9, pp. 1711–1722, 2005.

[Chan-Ma-Chi-Wang2008] T.-H. Chan, W.-K. Ma, C.-Y. Chi, & Y. Wang, "A convex analysis framework for blind separation of non-negative sources," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5120–5134, 2008.

[Fu-Ma-Huang-Sidiropoulos2015] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, "Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain," *IEEE Trans. Signal Process.*, 2015.

Simplex Volume Minimization:

[Craig1994] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.

[Packer2008] A. Packer, "NP-hardness of largest contained and smallest containing simplices for V- and H-polytopes," Discrete and Computational Geometry, vol. 28, no. 3, pp. 349–377, 2002.

[Li-Bioucas2008] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *Proc. IEEE IGARSS*, Aug. 2008, pp. 250–253.

[Chan-Chi-Huang-Ma2009] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.,* 2009.

[Bioucas2009] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. IEEE WHISPERS*, Aug. 2009.

[Miao-Qi2007] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, 2007.

[Fu-Huang-Yang-Ma-Sidiropoulos2016] X. Fu, K. Huang, B. Yang, W.-K. Ma, and N. D. Sidiropoulos, "Robust volume minimization-based matrix factorization for remote sensing and document clustering," *IEEE Trans. Signal Process.*, 2016.

[Lin-Ma-Li-Chi-Ambikapathi2015] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, Oct. 2015.

[Fu-Huang-Sidiropoulos2018] X. Fu, K. Huang, and N. D. Sidiropoulos, "On identifiability of nonnegative matrix factorization," *IEEE Signal Process. Lett.*, 2018.

[Nascimento-Bioucas2012] J. Nascimento and J. Bioucas, ""'Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, 2012.