# Geometric and algebraic structures in pattern recognition

Luke Oeding

Department of Mathematics, University of California, Berkeley

April 30, 2012

Rolf Bardeli

Multimedia Pattern Recognition

Fraunhofer

IAIS

mmprec.iais.fraunhofer.de/bardeli

Partial support from NSF Award No. 0853000: International Research Fellowship Program (IRFP).

# Something to solve by the end of the talk

What are the dimensions and degrees of the following two algebraic varieties? (take Zariski closure and $\blacksquare$ = anything)

$$
\color{red}{X} = \left\{ \begin{pmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{pmatrix} \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} + \begin{pmatrix} 0 & \blacksquare & 0 & 0 & 0 \\ \blacksquare & 0 & \blacksquare & 0 & 0 \\ 0 & \blacksquare & 0 & \blacksquare & 0 \\ 0 & 0 & \blacksquare & 0 & \blacksquare \\ 0 & 0 & 0 & \blacksquare & 0 \end{pmatrix} \right\}
$$

$$
\color{blue}{Y} = \left\{ \begin{pmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{pmatrix} \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} + \begin{pmatrix} 0 & \blacksquare & 0 & 0 & 0 \\ 0 & \blacksquare & \blacksquare & 0 & 0 \\ 0 & 0 & \blacksquare & \blacksquare & 0 \\ 0 & 0 & 0 & \blacksquare & \blacksquare \\ 0 & 0 & 0 & 0 & \blacksquare \end{pmatrix} \right\}
$$

(Thanks to Bernd Sturmfels and Zvi Rosen for these examples.)

# What is a musical score?

# What is a musical score?

- Typical 4 minute long song, 160 bpm, $32^{nd}$ note smallest increment = 20480 places to start playing a note.
- $T := \{1, \dots, N\}$ set of possible start times (ignore duration) (e.g. $N = 20480$)
- A simplified piano has 88 keys or pitches.
- $F := \{1, \dots, k\}$ set of possible notes (e.g. $k = 88$ notes)
- The simplest score $D$ is a collection of start times, and notes
- Naively: $D$ is a *sparse* matrix in $\{0, 1\}^{k \times N} \cong F \times T$
  – at a given time very few notes are simultaneously played.
- Better: Store $D$ as a collection of events: $D = (p_i, t_i)_i$.

# Musical Databases

- Let $\mathcal{M} = F \times T$ the set of musical notes.
- A database $\mathcal{D} = (D_1, \ldots, D_n)$ is a collection of scores $D_i \subset \mathcal{M}$. Usually well annotated.
- A query $Q \subset \mathcal{M}$ may be a short audio file, or, if the file has already been pre-processed, a short collection of notes (musical phrase).
- Find the set of documents $\{D_i\}$ such that $Q \subset D_i$ (approximate).
- To aid the search, try to find and exploit natural symmetries – group actions by scaling, translation, frieze symmetry, etc.

**Question**

*Given an $N \times k$ matrix $D$ with $0 < k << N$, develop efficient methods (or improve current ones) to find (useful) symmetry among the rows.*

[Bardeli R., *Similarity Search in Animal Sound Databases*, IEEE Transactions on Multimedia (2009).]

# Finding a bird by its song

Which birds are living in this German marshland?

# Finding a bird by its song

Which birds are living in this German marshland?

# Finding a bird by its song

Send groups of people (usually nature enthusiasts) out to the marsh with a pair of binoculars and a notepad to record what birds they spotted or heard.

The problem here is that there are fewer and fewer enthusiasts who have an understanding of bird song and those who are there are becoming older and thus harder of hearing.

# Finding a bird by its song

- New method: send one person with a microphone array that records geo-tagged audio files.
- For each file, annotate which bird songs are present.
- Using human volunteers, this is a huge task, and perhaps unreliable.
- Goal: Directly (computationally) compare field audio files to the annotated archive of audio files determine which birds are present.

[R Bardeli, D Wolff, F Kurth, M Koch, KH Tauchert, KH Frommolt, *Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring,* Pattern Recognition Letters (2010).]

# Tierstimmen archiv (animal sound archive)


- About 120,000 animal sound files. Well annotated, an extremely valuable resource.
- It can be used freely by anybody with a good (non-commercial) reason (e.g. science, nature conservation, art, teaching, ...)
- Even though the archive is big, for the more variable bird songs it hardly coves all the variability there is.
- There is a similar archive at Cornell (Chris Clark).

# Finding a bird by its song

- Analogous to searching the database of musical scores.
- This requires “feature extraction,” a topic for a different talk.
- Briefly, current methods involve taking the Fourier transform of the audio signal and studying the resulting data via visual properties.
- Informally, translate the audio file into a sequence of integer vectors representing the dominant features of the sound.
- For our purposes we will treat a query $Q$ as a long integer vector (subset of musical notes).

# Dictionary based methods: sparse represenation

*Dictionary based methods present a paradigm for feature extraction, compression, and source separation, and much more.*

Sample signal: $\mathbf{x} \in \mathbb{R}^k$, a column vector of length $k$.

Prescribed dictionary $D \in \mathbb{R}^{k \times N}$: column vectors $d_i$ (features) each of length $k$, and a large number $N$ of possible features ($k << N$)

$$D = \begin{pmatrix} | & & | \\ d_1 & \dots & d_N \\ | & & | \end{pmatrix}$$

Describe $\mathbf{x}$ as a linear combination of a small number of the columns of $D$, i.e. find (approximate) $\mathbf{s} \in \mathbb{R}^N$

$$\mathbf{x} = D\mathbf{s}.$$

Amounts to computing dot products of the rows of $D$ with $\mathbf{s}$.

Optimization methods require this computation to be done repeatedly.

(following [Sturm, Roads, et. al.], [Chandrasekaran, Parrilo, et. al.]

# Dictionary based methods: sparse representation

- Find "best" $\mathbf{s}$ subject to $\mathbf{x} = D\mathbf{s}$
- To compute $D\mathbf{s}$, must compute many dot products with $s$. Optimization methods require this computation to be done repeatedly.
- Idea: Find and exploit structure of $D$ to speed up the computation.
- For example, if $D$ has low rank, block structure, sparse structure, nice factorization, etc., this can speed up the computation considerably.

**Question**

*How can we automatically find the structure in D that will make this speed-up possible?*

# Dictionary based methods: dictionary learning

Alternately, let $r_1, \dots, r_k$ denote the rows vectors in $\mathbb{R}^N$, representing known bird songs.
We would like to learn a good dictionary structure as

$$D = \begin{pmatrix} — & r_1 & — \\ & \vdots & \\ — & r_k & — \end{pmatrix} \simeq \begin{pmatrix} | & & | \\ d_1 & \dots & d_N \\ | & & | \end{pmatrix}$$

We want to extract the dominant features $d_i$ and express $D$ as a sparse matrix (plus noise).
Learning a sparse structure for $D$ means that computations $\mathbf{x} = D\mathbf{s}$ may be done more quickly.

# Dictionary structures & sparsity

<u>Most restrictive:</u>

Assume dictionary atoms $d_i$ only differ by sparse additive error:

$$D = \begin{pmatrix} | & & | \\ d_1 & \ldots & d_N \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ d_1 & \ldots & d_1 \\ | & & | \end{pmatrix} + E,$$

where $d_i \in \mathbb{R}^k$, $E \in \mathbb{R}^{k \times N}$ is sparse in the sense that most of its entries are approximately zero.

Say the dictionary structure is "rank one plus sparse noise."

# Dictionary structures & sparsity

Relax assumptions:
Assume dictionary atoms $d_i$ only differ by sparse additive error and linear scaling

$$D = \begin{pmatrix} | & & | \\ d_1 & \dots & d_N \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ d_1\lambda_1 & \dots & d_1\lambda_N \\ | & & | \end{pmatrix} + E,$$
$$= (d_1)\lambda$$

where $d_i \in \mathbb{R}^k$, $\lambda$ is the column vector $\lambda = (\lambda_1, \dots, \lambda_N)$, and $E$ is sparse.
Dictionary structure is still "rank one plus sparse noise."

# Dictionary structures & sparsity

Relax assumptions more:

Assume dictionary $D = (d_1, \ldots, d_k)$ only depends on linear combinations of a subset of words $\{d_1, \ldots, d_r\}$ (after reordering) with $r < k << N$ and sparse additive error.

$$
\begin{aligned}
D &= \begin{pmatrix} | & & | \\ d_1 & \ldots & d_r \\ | & & | \end{pmatrix} \begin{pmatrix} f_{1,1} & f_{1,2} & \ldots & f_{1,N} \\ f_{2,1} & f_{2,2} & \ldots & f_{2,N} \\ \vdots & & & \\ f_{r,1} & f_{r,2} & \ldots & f_{r,N} \end{pmatrix} + E \\
&= \tilde{D}F + E.
\end{aligned}
$$

$E$ sparse, and $r << N$.

## Question

*Given a dictionary D, find efficient methods to determine the "best" structure of the low-rank-plus-sparse-noise format.*

# A motivating example: low rank plus diagonal

Let $N = k$, $D, R, E \in \mathbb{R}^{k\times k}$, $E$ diagonal, and $rank(R) = r$.
What are the algebraic relations imposed on the matrix $D$ by

$$D = R + E$$

If $r = 0$ – off diagonal entries vanish.
If $r = 1$ – off-diagonal $2 \times 2$ minors vanish.
If $r = 2$ – off-diagonal $3 \times 3$ minors vanish, but aren't all the relations.
[Drton, Sturmfels, Sullivant, *Algebraic factor analysis: tetrads, pentads, and beyond,* Probability Theory and Related Fields (2007).]

# Rank 2 plus sparse error

What are the dimensions and degrees of the following two algebraic varieties? (take Zariski closure and $\blacksquare$ = anything)

$$\color{red}{X}\color{black}{} = \left\{ \begin{pmatrix} \color{red}\blacksquare & \color{red}\blacksquare \\ \color{red}\blacksquare & \color{red}\blacksquare \\ \color{red}\blacksquare & \color{red}\blacksquare \\ \color{red}\blacksquare & \color{red}\blacksquare \\ \color{red}\blacksquare & \color{red}\blacksquare \end{pmatrix} \begin{pmatrix} \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare \\ \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare & \color{red}\blacksquare \end{pmatrix} + \begin{pmatrix} 0 & \color{red}\blacksquare & 0 & 0 & 0 \\ \color{red}\blacksquare & 0 & \color{red}\blacksquare & 0 & 0 \\ 0 & \color{red}\blacksquare & 0 & \color{red}\blacksquare & 0 \\ 0 & 0 & \color{red}\blacksquare & 0 & \color{red}\blacksquare \\ 0 & 0 & 0 & \color{red}\blacksquare & 0 \end{pmatrix} \right\}$$

Hypersurface of degree 3:

$$\det{}_{\blacksquare} \begin{pmatrix} \blacksquare & \square & \blacksquare & \square & \blacksquare \\ \square & \square & \square & \square & \square \\ \blacksquare & \square & \blacksquare & \square & \blacksquare \\ \square & \square & \square & \square & \square \\ \blacksquare & \square & \blacksquare & \square & \blacksquare \end{pmatrix}$$

(Thanks to Bernd Sturmfels and Zvi Rosen for these examples.)

# Rank 2 plus sparse error

What are the dimensions and degrees of the following two algebraic varieties? (take Zariski closure and $\blacksquare$ = anything)

$$\color{blue}{Y} = \left\{ \begin{pmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{pmatrix} \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{pmatrix} + \begin{pmatrix} 0 & \blacksquare & 0 & 0 & 0 \\ 0 & \blacksquare & \blacksquare & 0 & 0 \\ 0 & 0 & \blacksquare & \blacksquare & 0 \\ 0 & 0 & 0 & \blacksquare & \blacksquare \\ 0 & 0 & 0 & 0 & \blacksquare \end{pmatrix} \right\}$$

Hypersurface of degree 7 (perform elimination).

(Thanks to Bernd Sturmfels and Zvi Rosen for these examples.)

# Low rank plus sparse error

- Consider $P \in \{0,1\}^{k\times N}$ a zero-pattern, $|P| = \sum_{i,j} P_{i,j}$.
- Error matrix $E \in \mathbb{R}^{k\times N}$ has zero-pattern $P$ if (approximately) zero entries in $E$ occur at the zeros of $P$. Write $E \in \mathbb{R}\{P\}$.
- Define $X_r(P) \subset \mathbb{R}^{k\times N}$ as the (Zariski closure of) matrices that are rank $\leq r$ plus error with zero-pattern $P$.

$$X_r(P) := \overline{\left\{AB + E \in \mathbb{R}^{k\times N} \mid A \in \mathbb{R}^{k\times r},\ B \in \mathbb{R}^{r\times N},\ E \in \mathbb{R}\{P\}\right\}}$$

# Low rank plus sparse error

$$X_r(P) := \overline{\left\{AB + E \in \mathbb{R}^{k\times N} \mid A \in \mathbb{R}^{k\times r},\ B \in \mathbb{R}^{r\times N},\ E \in \mathbb{R}\{P\}\right\}}$$

## Questions

*Study the varieties $X_r(P)$ and $P$ varies keeping $|P|$ and $r$ fixed.*

1. *What are the dimensions of $X_r(P)$?*
2. *What are the degrees of $X_r(P)$?*
3. *Can we find "nice" defining equations for $X_r(P) \in \mathbb{R}^{k\times N}$?*
4. *What are sufficient conditions to separate $X_r(P)$ from $X_{r'}(P')$?*
5. *For which $k, N, r, P$, are these models identifiable?*
6. *Given that $D \in X_r(P)$, how can we effectively find $A, B$ such that $D = AB + E$?*

*Questions are most interesting when $r < k << N$ and $|P|$ small.*

Can ask the analogous questions for other matrix structures.

# Wrap-up

In multimedia pattern recognition, (and many other fields) data sets grow faster than we can process or make sense of them.

Most applications use some sort of optimization methods to “solve” the computational problem, with many successes, but there is much room for improvement.

Let’s study the geometric and algebraic structures that naturally arise in these applications, even for the smallest cases.

We hope that the new insights we find will shed light and lead to improvements and new developments for computational techniques.