RECONSTRUCTION OF ARTICULATORY MEASUREMENTS WITH SMOOTHED LOW-RANK MATRIX COMPLETION

Weiran Wang[1] weiranwang@ttic.edu
Raman Arora[2] arora@cs.jhu.edu
Karen Livescu[1] klivescu@ttic.edu
[1]Toyota Technological Institute at Chicago
[2]Johns Hopkins University

1 Abstract

- Articulatory measurements have been used in a variety of speech science and technology applications, e.g., speech synthesis, articulatory inversion, and multi-view acoustic feature learning.
- Recording technologies (electromagnetic articulography, X-ray microbeam) typically have pellets attached to articulators. Limitations of the technologies lead to high rates of loss in this expensive and time-consuming data source.
- We propose a simple algorithm for reconstructing missing measurements based on **low-rank matrix factorization** combined with **temporal smoothness regularization**. The algorithm alternates between two steps, each having a **closed form** as the solution of a linear system.
- We demonstrate the algorithm on the Wisconsin X-ray microbeam database and achieve better root mean squared error and phonetic recognition performance than previous algorithms.

2 Smoothed Low-Rank Matrix Completion Objective

- **Notation:** We denote by $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{D\times N}$ the articulatory measurements over N successive frames, $\mathbf{M} \in \mathbb{R}^{D\times N}$ a binary matrix with $\mathbf{M}_{ij} = 1$ if $\mathbf{X}_{ij}$ is observed and 0 otherwise, $\odot$ the element-wise multiplication and $\otimes$ the Kronecker ("outer") product, $\mathbf{M}^i$ ($\mathbf{M}_j$) the i-th row (j-th column) of the matrix $\mathbf{M}$.
- Two important observations
  - Physical constraints imply **the data matrix is approximately low-rank**: $\mathbf{X} \approx \mathbf{U}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{D\times k}$, $\mathbf{V} \in \mathbb{R}^{N\times k}$, and $k < \max\{D, N\}$.
  - Articulatory trajectories are **smooth in time**, i.e., the difference between successive frames $\|\mathbf{x}_{i+1} - \mathbf{x}_i\|$ should be small. This suggests a smoothness penalty

$$\mathcal{S}(\mathbf{Y}) = \sum_{j=1}^{N-1} \|\mathbf{y}_{j+1} - \mathbf{y}_j\|^2 = \operatorname{tr}\left(\mathbf{Y}\mathbf{L}\mathbf{Y}^\top\right) \text{ with } \mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \text{ for } \mathbf{Y} \in \mathbb{R}^{D\times N}.$$

- Our objective function combines the two intuitions:

$$\min_{\mathbf{U},\mathbf{V}} \quad \left\|\mathbf{M} \odot (\mathbf{X} - \mathbf{U}\mathbf{V}^\top)\right\|_F^2 + \lambda(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) + \gamma \operatorname{tr}\left(\mathbf{U}\mathbf{V}^\top\mathbf{L}\mathbf{V}\mathbf{U}^\top\right), \qquad (1)$$

  where $\lambda, \gamma > 0$ are trade-off parameters selected by cross-validation.
- The $L2$ regularization functions like a Gaussian prior on $\mathbf{U}$ and $\mathbf{V}$, and helps avoid numerical instability.
- Special case of $\gamma = 0$: without smoothness penalty, our objective leads to the alternating least squares (ALS) algorithm which is widely used in the matrix completion and collaborative filtering literature.
- Reconstruction: once $(\mathbf{U}, \mathbf{V})$ are obtained by solving (1), missing entries of $\mathbf{X}$ are filled with the corresponding entries of $\mathbf{U}\mathbf{V}^\top$.

3 Solution via Alternating Optimization

- The objective function is **convex and quadratic** in $\mathbf{U}$ if $\mathbf{V}$ is fixed and vice versa.
- Our algorithm alternates between:
  - $\mathbf{U}$-step For fixed $\mathbf{V}$, compute the gradient of (1) wrt $\mathbf{U}$ and set it to zero to obtain a $k \times k$ linear system for each row i of $\mathbf{U}$:

$$\mathbf{U}^i\mathbf{V}^\top \operatorname{diag}\left(\mathbf{M}^i\right)\mathbf{V} + \lambda\mathbf{U}^i + \gamma\mathbf{U}^i(\mathbf{V}^\top\mathbf{L}\mathbf{V}) = \mathbf{X} \operatorname{diag}\left(\mathbf{M}^i\right)\mathbf{V},$$

    so that each row of $\mathbf{U}$ can be solved in closed form as

$$\mathbf{U}^i = \mathbf{X} \operatorname{diag}\left(\mathbf{M}^i\right)\mathbf{V}(\mathbf{V}^\top \operatorname{diag}\left(\mathbf{M}^i\right)\mathbf{V} + \lambda\mathbf{I} + \gamma\mathbf{V}^\top\mathbf{L}\mathbf{V})^{-1}.$$

  - $\mathbf{V}$-step For fixed $\mathbf{U}$, compute the gradient of (1) wrt $\mathbf{V}$ and set it to zero to obtain the following linear system

$$(\mathbf{M}^\top \odot (\mathbf{V}\mathbf{U}^\top - \mathbf{X}^\top))\mathbf{U} + \lambda\mathbf{V} + \gamma\mathbf{L}\mathbf{V}\mathbf{U}^\top\mathbf{U} = \mathbf{0}, \qquad (2)$$

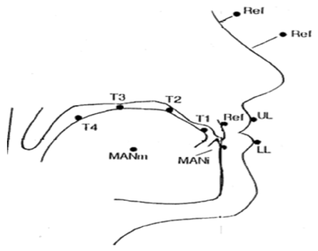
    where all rows of $\mathbf{V}$ are coupled due to the smoothness penalty. Nonetheless, $\mathbf{V}$ can be obtained efficiently by solving a sparse $NK \times NK$ linear system

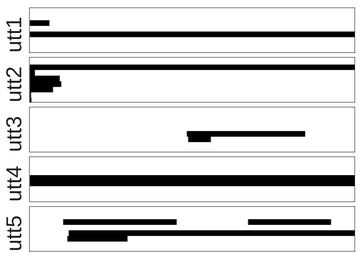
$$(\mathbf{K} + \lambda\mathbf{I} + \gamma\mathbf{L} \otimes (\mathbf{U}^\top\mathbf{U})) \cdot \operatorname{vec}(\mathbf{V}^\top) = \operatorname{vec}(\mathbf{U}^\top(\mathbf{M} \odot \mathbf{X})),$$

    where $\mathbf{K} = \begin{bmatrix} \mathbf{U}^\top \operatorname{diag}(\mathbf{M}_1)\,\mathbf{U} & & \\ & \cdots & \\ & & \mathbf{U}^\top \operatorname{diag}(\mathbf{M}_N)\,\mathbf{U} \end{bmatrix}$.
- Initialization: fill missing entries with zeros, compute the truncated SVD to obtain $\mathbf{U}$ and $\mathbf{V}$.
- Convergence: each $\mathbf{U}/\mathbf{V}$-step finds the **best solution** in $\mathbf{U}$ and $\mathbf{V}$ respectively given the other set of parameters are fixed, and decreases the overall objective.

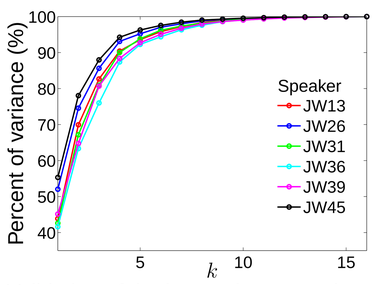
4 Data

- Wisconsin X-ray microbeam (XRMB) [1]: simultaneously recorded speech and articulatory measurements from 47 American English speakers.
- 53 utterances per speaker, 3.4% of the entries are missing, yet 23.6% of the frames contain at least one missing entry.





Missing data patterns



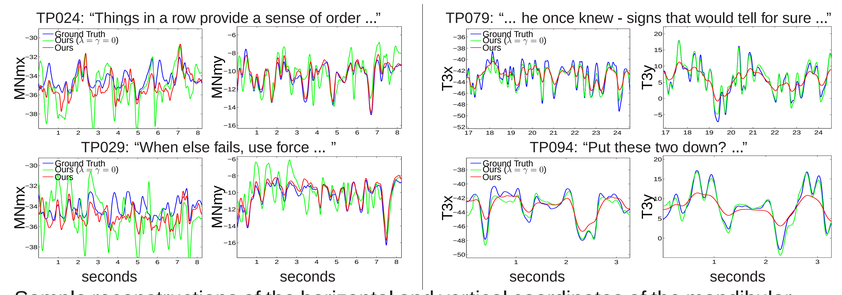
Validation of the low-rank assumption

5 Experimental Results

- We compare with two related approaches
  - The modified EM algorithm for PCA [2], corresponding to the **special unregularized case of our algorithm ($\lambda = \gamma = 0$)**.
  - Gaussian Mixture Model (GMM) [3]: model **fully observed frames** with GMM, fill in missing entries of each frame with conditional mean given observed entries.
- Reconstructing artificially blacked-out data by transferring missing patterns

Source speaker	Corrupted Frames (%)	Missing Entries (%)	Target speaker	Init.	GMM	Ours ($\lambda=0,\gamma=0$)	Ours ($\lambda=0$)	Ours ($\gamma=0$)	Ours
JW29	98.4	13.9	JW13	25.51	17.25	1.97	1.81	1.84	1.63
			JW26	23.13	13.21	2.10	1.99	1.33	1.32
			JW31	21.82	14.81	1.42	1.42	1.38	1.19
			JW45	24.95	13.67	1.88	1.38	1.45	1.20
JW30	20.6	3.4	JW13	21.65	2.59	6.51	1.69	6.38	1.69
			JW26	22.42	4.83	6.64	2.13	6.51	2.10
			JW31	19.72	7.14	5.87	1.85	5.76	1.83
			JW45	25.70	2.90	1.89	1.80	1.36	1.35

Reconstruction errors (RMSE) in millimeters obtained by different algorithms.



Sample reconstructions of the horizontal and vertical coordinates of the mandibular MNm (left) and mid-tongue T3 (right) pellets.

- Phonetic recognition with reconstructed data: reconstructed articulatory measurements are concatenated with 39D MFCC and used as inputs to a GMM-HMM phone recognizer.

Method	PER (%)
Baseline (MFCCs only)	31.1
GMM	22.0
Ours ($k=4,\lambda=0,\gamma=0$)	20.4
Ours ($k=6,\lambda=1,\gamma=1$)	**20.0**

6 Future Directions

Model articulatory measurements with nonlinear manifolds; use complementary information in acoustic data; use better dynamic models.

[1] J. R. Westbury, *X-Ray Microbeam Speech Production Database User's Handbook Version 1.0*, University of Wisconsin, Madison, 1994.
[2] S. T. Roweis, *Data Driven Production Models for Speech Processing*, Ph.D. Thesis, Cal. Inst. Of Tech., 1999.
[3] C. Qin and M. Á. Carreira-Perpiñán, *Estimating missing data sequences in X-ray microbeam recordings*, Interspeech 2010.