Polar Incremental Matrix Completion

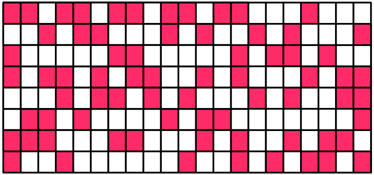
Ryan Kennedy*, C.J. Taylor*, and **Laura Balzano**†

*University of Pennsylvania and †University of Michigan

Signal Processing with Adaptive Sparse Structured Representations
SPARS July 2015

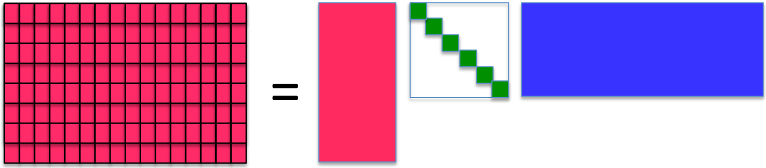
Identifying Subspaces from Partial Observations

In the problem of low-rank matrix completion, we observe a certain phenomenon in a high-dimensional ambient space, but the phenomenon lies on or near a low-dimensional subspace.
Each observed vector has missing elements.

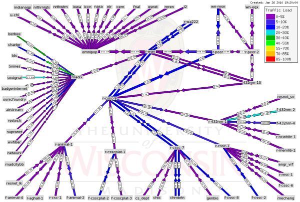


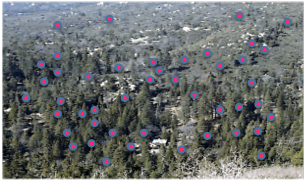
Identifying Subspaces from Partial Observations

In the problem of low-rank matrix completion, we observe a certain phenomenon in a high-dimensional ambient space, but the phenomenon lies on or near a low-dimensional subspace.
Each observed vector has missing elements.

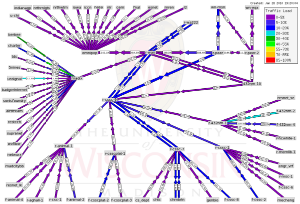


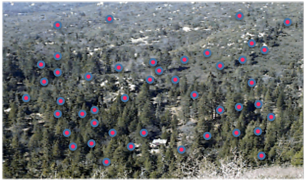
Common applications for the low rank assumption...

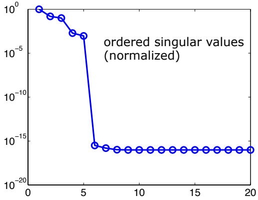


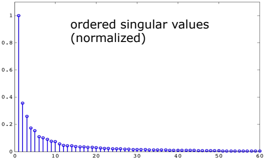


...often have data that are not isotropic in the subspace.

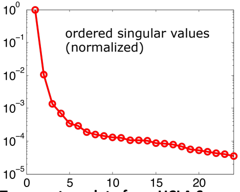








Byte Count data from UW network

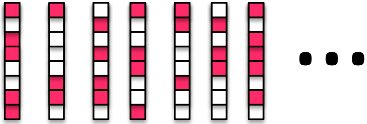


Temperature data from UCLA Sensornet

Identifying Subspaces from Partial Observations

- We seek the subspace $\mathcal{S} \subset \mathbb{R}^n$ of known dimension $d \ll n$.

- We observe only certain components $\Omega_t \subset \{1, 2, \dots, n\}$ of vectors $v_t \in \mathcal{S}$, $t = 1, 2, \dots$ — the subvector $[v_t]_{\Omega_t}$.



- The data are not drawn isotropically from $\mathcal{S}$.

Brief Interlude: Non-convex formulation for streaming data applications

Supposing we collect T vectors $v_1, \ldots, v_T$ into a matrix X. We could solve the convex problem:

$$\underset{M \in \mathbb{R}^{n \times T}}{\text{minimize}} \quad \|M\|_* + \lambda \|[M - X]_\Omega\|_F^2$$

Or we could solve this non-convex problem incrementally:

$$\underset{\text{span}(U) \in \mathcal{G}(d,n)}{\text{minimize}} \quad \|[UW^T - X]_\Omega\|_F^2 \;=\; \sum_{t=1}^{T} \|[Uw - v_t]_{\Omega_t}\|_2^2$$

With full data, use the Incremental SVD

Given matrix $X = U\Sigma V^T$, form the SVD of $\begin{bmatrix} X & v_t \end{bmatrix}$.

Estimate the weights: $w = \arg\min_a \| Ua - v_t \|_2^2$

Compute the residual: $r_t = v_t - Uw$.

Update the SVD:

$$\begin{bmatrix} X & v_t \end{bmatrix} = \begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and take SVD of the center matrix.

Incremental SVD with Missing Data

Suppose now we only observe entries of v on $\Omega_t \subset \{1, \dots n\}$. Let subscript Ω_t restrict to the corresponding rows.

Estimate the weights: $w = \arg\min_a \|[U_t a - v_t]_{\Omega_t}\|_2^2$.

Compute the residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

Update the SVD:

$$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and take SVD of the center matrix.

Incremental SVD with Missing Data: SAGE GROUSE

Suppose now we only observe entries of v on $\Omega_t \subset \{1, \ldots n\}$. Let subscript Ω_t restrict to the corresponding rows.

Estimate the weights: $w = \arg\min_a \|[U_t a - v_t]_{\Omega_t}\|_2^2$.

Compute the residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

Update the SVD:

$$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \mathcal{I}_d & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

and take the SVD of the center matrix. This is equivalent to the natural incremental gradient method on the Grassmannian (GROUSE) for a particular step size.

Incremental SVD with Missing Data Options

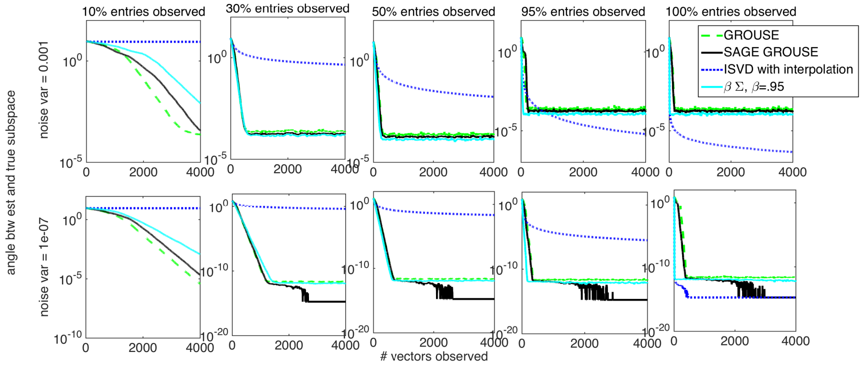
projection weights $w = \arg\min_a \|[U_t a - v_t]_{\Omega_t}\|_2^2$;
residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

$$\text{ISVD with interpolation: } \begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

$$\text{SAGE GROUSE: } \begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \mathcal{I}_d & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

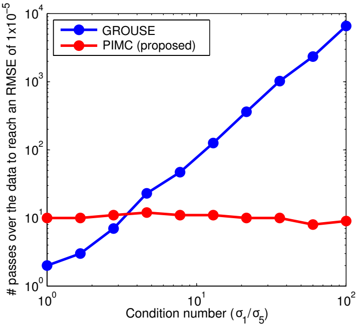
$$\text{Brand Algorithm } (\beta \leq 1) : \begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \beta\Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

Incremental SVD with Missing Data Performance



Data that are not isotropically distributed

500 $\times$ 500 matrices, rank 5, no added noise, no missing data.



We need a way to correct for a skew in singular values, but using the singular value estimate directly we get killed by missing data.

Polar Incremental Matrix Completion

To develop a different estimate for the singular values we use a thin version of the *polar decomposition*, $R = QS$, where Q is $n \times d$ with orthonormal columns and S is $d \times$d positive semi-definite.

We will track our data matrix with the decomposition $X = UR^T = USQ^T$; our estimate of the singular values is now flexibly represented with this PSD matrix S.

We also scale this matrix S at every step depending on the norm of the observed data only: let $s_t^2 = s_{t-1}^2 + \|v_{\Omega_t}\|_2^2$.

Incremental SVD with Missing Data Options

projection weights $w = \arg\min_a \|[U_t a - v_t]_{\Omega_t}\|_2^2$;
residual: $r_t = v_t - Uw$ on Ω_t ; zero otherwise.

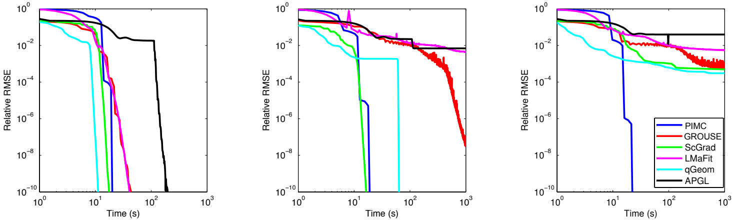
ISVD with interpolation: $$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

SAGE GROUSE: $$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \mathcal{I}_d & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

Brand Algorithm: $$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \beta\Sigma & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} V & 0 \\ 0 & 1 \end{bmatrix}^T$$

PIMC: $$\begin{bmatrix} U & \frac{r_t}{\|r_t\|} \end{bmatrix} \begin{bmatrix} \frac{s_t}{\|S\|_F} S & w \\ 0 & \|r_t\| \end{bmatrix} \begin{bmatrix} \frac{\|S\|_F}{s_t} Q & 0 \\ 0 & 1 \end{bmatrix}^T$$

Comparison of algorithms on ill-conditioned matrices.



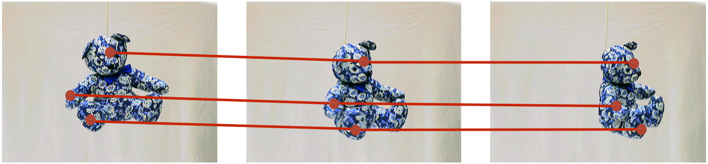
$\frac{\sigma_1}{\sigma_5} = 10$ $\frac{\sigma_1}{\sigma_5} = 100$ $\frac{\sigma_1}{\sigma_5} = 1000$

Matrices were 5000×5000, rank 5, no noise and 95% missing entries, with singular values that varied logarithmically from $\sigma_1 = 1 \times 10^3$ down to σ_5. For moderately ill-conditioned matrices, `ScGrad` and `qGeom` – both of which modify the metric on the Grassmannian order to perform well on ill-conditioned matrices – perform as well as `PIMC`, but with higher condition numbers even these algorithms have trouble.

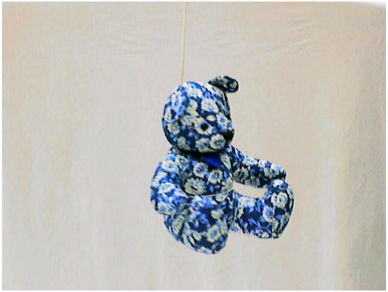
Structure from Motion

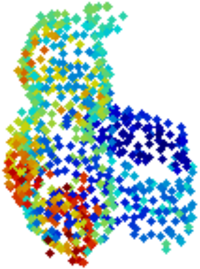
Observe an object from different camera angles, matching reference points on the object from image to image.

- Object is solid, so some reference points are occluded in each photo. **Missing data!**
- Matrix of 2d point locations has rank three, and the range subspace reveals 3-d location of reference points.

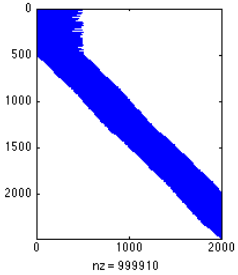


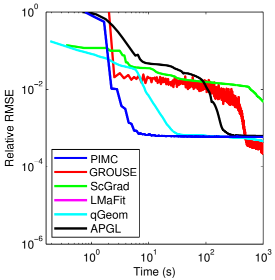
Structure from Motion





Structure from Motion: Synthetic Cylinder





Synthetic cylinder of radius 10 and height 5000 with 500 points tracked over 1000 frames. The cylinder rotated once every 500 frames, resulting in 80.13% missing data. This matrix has an exact rank-4 solution with a condition number $\sigma_1/\sigma_4 \approx 290$.

Thank you! Questions?