Low-rank models

Optimization-Based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_spring16

Carlos Fernandez-Granda

5/2/2016

Matrix completion

- The matrix completion problem
- Nuclear norm
- Theoretical guarantees
- Algorithms
- Alternating minimization

Robust PCA

- Outliers
- Low rank + sparse model
- Theoretical guarantees
- Algorithms
- Background subtraction

Netflix Prize

Matrix completion

Bob	Molly	Mary	Larry	
1	?	5	4	The Dark Knight
?	1	4	5	Spiderman 3
4	5	2	?	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	?	5	Superman 2

Matrix completion as an inverse problem

$$\begin{bmatrix} 1 & ? & 5 \\ ? & 3 & 2 \end{bmatrix}$$

For a fixed sampling pattern, underdetermined system of equations

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} M_{11} \\ M_{21} \\ M_{12} \\ M_{22} \\ M_{13} \\ M_{23} \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 2 \end{bmatrix}$$

Isn't this completely ill posed?

Assumption: Matrix is low rank, depends on $\approx r\,(m+n)$ parameters

As long as data > parameters recovery is possible (in principle)

$$\begin{bmatrix} 1 & 1 & 1 & 1 & ? & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ ? & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Matrix cannot be sparse

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Singular vectors cannot be sparse

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}+\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\begin{bmatrix}1 & 2 & 3 & 4\end{bmatrix}=\begin{bmatrix}1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\1 & 1 & 1 & 1\\2 & 3 & 4 & 5\end{bmatrix}$$

Incoherence

The matrix must be incoherent: its singular vectors must be spread out

For $1/\sqrt{n} \leq \mu \leq 1$

$$\max_{1 \leq i \leq r, 1 \leq j \leq m} |U_{ij}| \leq \mu$$

$$\max_{1 \leq i \leq r, 1 \leq j \leq n} |V_{ij}| \leq \mu$$

for the left $U_1, \ldots, U_r$ and right $V_1, \ldots, V_r$ singular vectors

Measurements

We must see an entry in each row/column at least

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ ? & ? & ? & ? \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ ? \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

Assumption: Random sampling (usually does not hold in practice!)

Underdetermined inverse problems

	Measurements	*Class of signals*
Compressed sensing	Gaussian, random Fourier coeffs.	Sparse
Super-resolution	Low pass	Signals with min. separation
Matrix completion	Random sampling	Incoherent low-rank matrices

Matrix inner product

The trace of a $n \times n$ matrix is defined as

$$\text{Trace}\,(A) := \sum_{i=1}^{n} A_{ii}$$

The inner product between two $m \times n$ matrices is defined as

$$\langle A, B \rangle = \text{Trace}\left(A^T B\right) = \sum_{i=1}^{m} \sum_{i=1}^{n} A_{ij} B_{ij}$$

For any matrices A, B, C with appropriate dimensions

$$\text{Trace}\,(ABC) = \text{Trace}\,(BCA) = \text{Trace}\,(CAB)$$

Matrix norm

Let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ be the singular values of $M \in \mathbb{R}^{m \times n}$, $m \geq n$,

Operator norm

$$||M|| := \max_{||u||_2 \leq 1} ||M\,u||_2 = \sigma_1$$

Frobenius norm

$$||M||_{\mathrm{F}} := \sqrt{\sum_i M_{ij}^2} = \sqrt{\mathrm{Trace}\,(M^T M)} = \sqrt{\sum_{i=1}^n \sigma_i^2}$$

Nuclear norm

$$||M||_* := \sum_{i=1}^n \sigma_i$$

Characterization of nuclear norm

$$||A||_* = \sup_{||B|| \leq 1} \langle A, B \rangle$$

Consequence: Nuclear norm satisfies triangle inequality

$$||A + B||_* \leq ||A||_* + ||B||_*$$

Proof of characterization

For any $M \in \mathbb{R}^{m \times n}$, $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$, if $U^T U = I$, $V^T V = I$ then

$$||UMV|| = ||M||$$

For any $M \in \mathbb{R}^{n \times n}$

$$\max_{1 \leq i \leq n} |M_{ii}| \leq ||M||$$

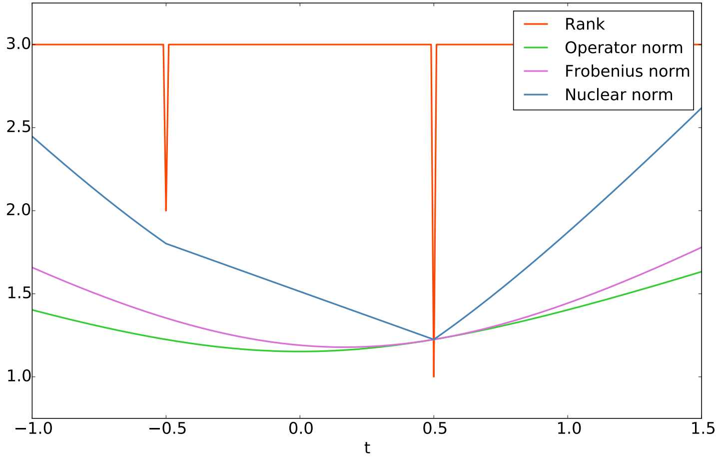
Experiment

Compare rank, operator norm, Frobenius norm and nuclear norm of

$$M(t) := \begin{bmatrix} 0.5+t & 1 & 1 \\ 0.5 & 0.5 & t \\ 0.5 & 1-t & 0.5 \end{bmatrix}$$

for different values of t

Matrix norms vs rank



Low-rank matrix estimation

First idea:

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \operatorname{rank}\left(\widetilde{X}\right) \quad \text{such that } \widetilde{X}_{\Omega} = y$$

Ω: indices of revealed entries
y: revealed entries

Computationally intractable because of missing entries

Tractable alternative:

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \left|\left|\widetilde{X}\right|\right|_{*} \quad \text{such that } \widetilde{X}_{\Omega} = y$$

Low-rank matrix estimation

If data are noisy

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \left|\left| \widetilde{X}_\Omega - y \right|\right|_2^2 + \lambda \left|\left| \widetilde{X} \right|\right|_*$$

where $\lambda > 0$ is a regularization parameter

Matrix completion via nuclear-norm minimization

Bob	Molly	Mary	Larry	
1	2 (1)	5	4	The Dark Knight
2 (2)	1	4	5	Spiderman 3
4	5	2	2 (1)	Love Actually
5	4	2	1	Bridget Jones's Diary
4	5	1	2	Pretty Woman
1	2	5 (5)	5	Superman 2

Exact recovery

Guarantees by Gross 2011, Candès and Recht 2008, Candès and Tao 2009

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \left| \left| \widetilde{X} \right| \right|_* \quad \text{such that } \widetilde{X}_\Omega = y$$

achieves exact recovery with high probability as long as the number of samples is proportional to $r(n+m)$ up to log terms

The proof is based on the construction of a dual certificate

Subgradient of nuclear norm

Let $M = U\Sigma V^T$. Any matrix of the form

$$G := UV^T + W$$

where

$$\begin{aligned} ||W|| &\leq 1 \\ U^T W &= 0 \\ W\,V &= 0 \end{aligned}$$

is a subgradient of the nuclear norm at M, so that

$$||M + H||_* \geq ||M||_* + \langle G, H \rangle \quad \text{for any } H$$

Proof

Follows from

$$||A||_* = \sup_{||B|| \leq 1} \langle A, B \rangle$$

Dual certificate

Let $M = U\Sigma V^T$. A dual certificate Q of the optimization problem

$$\min_{\widetilde{X}\in\mathbb{R}^{m\times n}} \left|\left|\widetilde{X}\right|\right|_* \quad \text{such that } \widetilde{X}_\Omega = y$$

is any matrix supported on Ω such that

$$Q = UV^T + W$$

$$\begin{aligned} &||W|| < 1 \\ &U^T W = 0 \\ &W\, V = 0 \end{aligned}$$

Dual certificate

$UV^T = Q - W$ where Q is supported on Ω and $||W|| < 1$

If U or V are not incoherent, UV^T might have large entries not in Ω

Proof of existence relies on concentration bounds

Proximal gradient method

Method to solve the optimization problem

$$\text{minimize} \quad f(x) + g(x),$$

where f is differentiable and prox_g is tractable

Proximal-gradient iteration:

$$x^{(0)} = \text{arbitrary initialization}$$
$$x^{(k+1)} = \text{prox}_{\alpha_k\, g}\left(x^{(k)} - \alpha_k \nabla f\left(x^{(k)}\right)\right)$$

Proximal operator of nuclear norm

The solution $\widehat{X}$ to

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \frac{1}{2} \left|\left| Y - \widetilde{X} \right|\right|_{\text{F}}^2 + \tau \left|\left| \widetilde{X} \right|\right|_*$$

is obtained by soft-thresholding the SVD of Y

$$\widehat{X} = \mathcal{D}_\tau (Y)$$

$$\mathcal{D}_\tau (M) := U \, \mathcal{S}_\tau (\Sigma) \, V^T \qquad \text{where } M = U \, \Sigma V^T$$

$$\mathcal{S}_\tau (\Sigma)_{ii} := \begin{cases} \Sigma_{ii} - \tau & \text{if } \Sigma_{ii} > \tau \\ 0 & \text{otherwise} \end{cases}$$

Proximal gradient method

Proximal gradient method for the problem

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \left|\left| \widetilde{X}_{\Omega} - y \right|\right|_2^2 + \lambda \left|\left| \widetilde{X} \right|\right|_*$$

$$X^{(0)} = \text{arbitrary initialization}$$

$$M^{(k)} = X^{(k)} - \alpha_k \left(X_{\Omega}^{(k)} - y \right)$$

$$X^{(k+1)} = \mathcal{D}_{\alpha_k \lambda} \left(M^{(k)} \right)$$

Low-rank matrix completion

Intractable problem

$$\min_{\widetilde{X} \in \mathbb{R}^{m \times n}} \operatorname{rank}\left(\widetilde{X}\right) \quad \text{such that } \widetilde{X}_\Omega \approx y$$

Nuclear norm: convex (☺) but computationally expensive (☹) due to SVD computations

Alternative

- Fix rank k beforehand
- Parametrize the matrix as AB where $A \in \mathbb{R}^{m \times k}$ and $B \in \mathbb{R}^{k \times n}$
- Solve

$$\min_{\widetilde{A} \in \mathbb{R}^{m \times k}, \widetilde{B} \in \mathbb{R}^{k \times n}} \left|\left| \left(\widetilde{A}\widetilde{B} \right)_{\Omega} - y \right|\right|_2$$

by alternating minimization

Alternating minimization

Sequence of least-squares problems (much faster than computing SVDs)

- To compute $A^{(k)}$ fix $B^{(k-1)}$ and solve

$$\min_{\widetilde{A} \in \mathbb{R}^{m \times k}} \left| \left| \left(\widetilde{A} B^{(k-1)} \right)_{\Omega} - y \right| \right|_2$$

- To compute $B^{(k)}$ fix $A^{(k)}$ and solve

$$\min_{\widetilde{B} \in \mathbb{R}^{k \times n}} \left| \left| \left(A^{(k)} \widetilde{B} \right)_{\Omega} - y \right| \right|_2$$

Theoretical guarantees: Jain, Netrapalli, Sanghavi 2013

Matrix completion

- The matrix completion problem
- Nuclear norm
- Theoretical guarantees
- Algorithms
- Alternating minimization

Robust PCA

- Outliers
- Low rank + sparse model
- Theoretical guarantees
- Algorithms
- Background subtraction

Collaborative filtering

$$A := \begin{pmatrix} 1 & 1 & 5 & 5 \\ 1 & 1 & 5 & 5 \\ 5 & 5 & 1 & 1 \\ 5 & 5 & 1 & 1 \\ 5 & 5 & 1 & 1 \\ 1 & 1 & 5 & 5 \end{pmatrix}$$

	Bob	Molly	Mary	Larry
The Dark Knight	1	1	5	5
Spiderman 3	1	1	5	5
Love Actually	5	5	1	1
Bridget Jones's Diary	5	5	1	1
Pretty Woman	5	5	1	1
Superman 2	1	1	5	5

SVD

$$A - \bar{A} = U\Sigma V^T = U \begin{bmatrix} 9.798 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} V^T$$

First left singular vector

$$U_1 = \begin{pmatrix} \overset{\text{D. Knight}}{-0.4082} & \overset{\text{Sp. 3}}{-0.4082} & \overset{\text{Love Act.}}{0.4082} & \overset{\text{B.J.'s Diary}}{0.4082} & \overset{\text{P. Woman}}{0.4082} & \overset{\text{Sup. 2}}{-0.4082} \end{pmatrix}$$

Interpretations:

- Score atom: Centered scores for each person are proportional to U_1
- Coefficients: They cluster movies into action (-) and romantic (+)

First right singular vector

$$V_1 = \begin{pmatrix} \text{Bob} & \text{Molly} & \text{Mary} & \text{Larry} \\ 0.5 & 0.5 & -0.5 & -0.5 \end{pmatrix}$$

Interpretations:

- Score atom: Centered scores for each movie are proportional to V_1
- Coefficients: They cluster people into action (-) and romantic (+)

Outliers

$$A := \begin{pmatrix} \color{red}{5} & 1 & 5 & 5 \\ 1 & 1 & 5 & 5 \\ 5 & 5 & 1 & 1 \\ 5 & 5 & 1 & 1 \\ 5 & 5 & 1 & 1 \\ 1 & 1 & 5 & \color{red}{1} \end{pmatrix}$$

	Bob	Molly	Mary	Larry
The Dark Knight	5	1	5	5
Spiderman 3	1	1	5	5
Love Actually	5	5	1	1
Bridget Jones's Diary	5	5	1	1
Pretty Woman	5	5	1	1
Superman 2	1	1	5	1

SVD

$$A - \bar{A} = U\Sigma V^T = U \begin{bmatrix} 8.543 & 0 & 0 & 0 \\ 0 & 4.000 & 0 & 0 \\ 0 & 0 & 2.649 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} V^T$$

Without outliers

$$A - \bar{A} = U\Sigma V^T = U \begin{bmatrix} 9.798 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} V^T$$

First left singular vector

	D. Knight	Sp. 3	Love Act.	B.J.'s Diary	P. Woman	Sup. 2
$U_1 =$	$(-0.2610$	-0.4647	0.4647	0.4647	0.4647	$-0.2610)$

Without outliers

	D. Knight	Sp. 3	Love Act.	B.J.'s Diary	P. Woman	Sup. 2
$U_1 =$	$(-0.4082$	-0.4082	0.4082	0.4082	0.4082	$-0.4082)$

First right singular vector

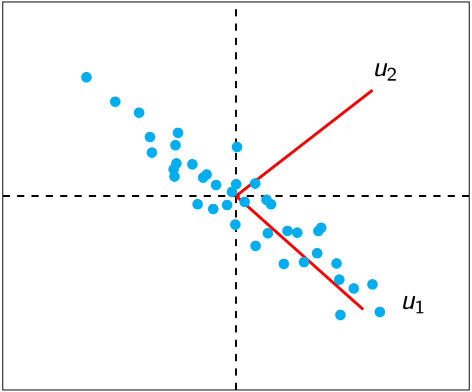
$$V_1 = \begin{matrix} \text{Bob} & \text{Molly} & \text{Mary} & \text{Larry} \\ (0.4352 & 0.5573 & -0.5573 & -0.4352) \end{matrix}$$

Without outliers

$$V_1 = \begin{matrix} \text{Bob} & \text{Molly} & \text{Mary} & \text{Larry} \\ (\,0.5 & 0.5 & -0.5 & -0.5) \end{matrix}$$

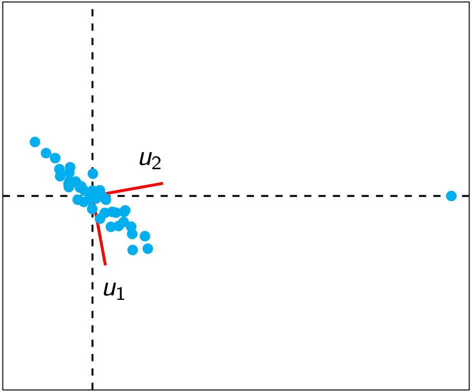
PCA without outliers

$$\frac{\sigma_1}{\sqrt{n}} = 1.042 \qquad \frac{\sigma_2}{\sqrt{n}} = 0.192$$



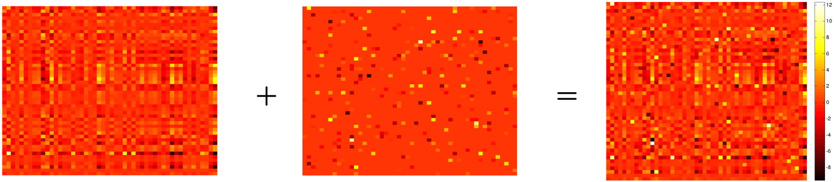
PCA with outliers

$$\frac{\sigma_1}{\sqrt{n}} = 1.774 \qquad \frac{\sigma_2}{\sqrt{n}} = 0.633$$



Low rank + sparse model

Sum of a low-rank component L and a sparse component S



L S M

Low-rank component cannot be sparse

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 47 & 0 \end{bmatrix}$$
$$=$$
$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 23 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 47 & 0 \end{bmatrix}$$

Incoherence

Low-rank component must be incoherent

For $L = U \Sigma V^T$

$$\max_{1 \leq i \leq r, 1 \leq j \leq m} |U_{ij}| \leq \mu$$

$$\max_{1 \leq i \leq r, 1 \leq j \leq n} |V_{ij}| \leq \mu$$

where $1/\sqrt{n} \leq \mu \leq 1$

Sparse component cannot be low rank

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}
$$

$$
=
$$

$$
\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

Assumption: Support distributed uniformly at random
(doesn't hold in practice!)

Nuclear norm + ℓ_1-norm

We want to promote a low-rank L and a sparse S

$$\min_{\widetilde{L},\widetilde{S} \in \mathbb{R}^{m \times n}} \left|\left|\widetilde{L}\right|\right|_* + \lambda \left|\left|\widetilde{S}\right|\right|_1 \quad \text{such that } \widetilde{L} + \widetilde{S} = Y$$

Here $||\cdot||_1$ is the ℓ_1 norm of the vectorized matrix

Choice of λ

$$\left\| \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \right\|_* = n \qquad \left\| \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \right\|_1 = n^2$$

$$\lambda > \frac{1}{n}$$

Choice of λ

$$\left\|\begin{bmatrix}0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\right\|_* = 1 \qquad \left\|\begin{bmatrix}0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\right\|_1 = 1$$

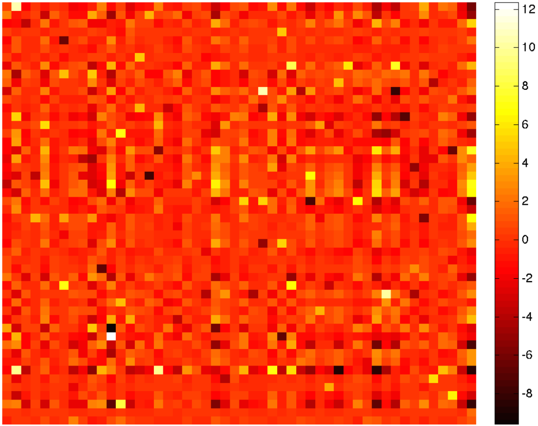
$$\lambda < 1$$

Choice of λ

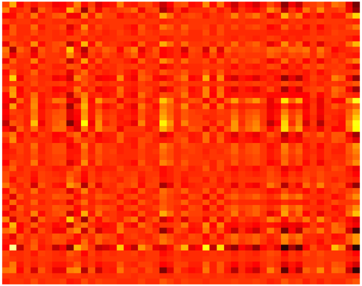
$$\left\| \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\|_* = \sqrt{n} \qquad \left\| \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right\|_1 = n$$

$$\lambda \approx \frac{1}{\sqrt{n}}$$

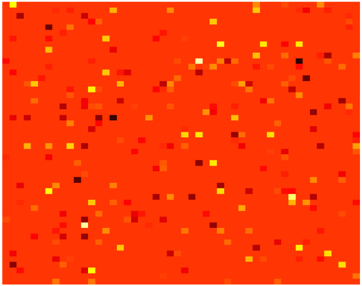
$L + S$



$\lambda = \frac{1}{\sqrt{n}}$

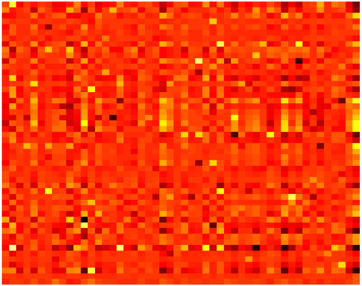


L



S

Large λ

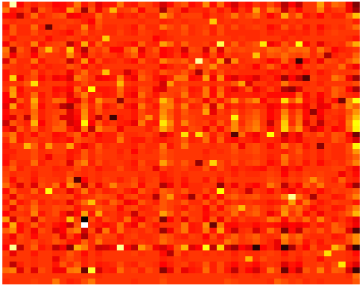


L

S

Small λ

L



S

Exact recovery

Guarantees by Candès, Li, Ma, Wright 2011

$$\min_{\widetilde{L},\widetilde{S}\in\mathbb{R}^{n\times n}} \left|\left|\widetilde{L}\right|\right|_* + \lambda \left|\left|\widetilde{S}\right|\right|_1 \quad \text{such that } \widetilde{L}+\widetilde{S}=Y$$

achieves an exact decomposition with high probability for

- $\operatorname{rank}(L)$ of order n if L is incoherent
- a sparsity level of S of order n^2 if its support is random

The proof is based on the construction of a dual certificate

Dual certificate

Let $L = U\Sigma V^T$ and Ω be the support of S

A dual certificate Q of the optimization problem

$$\min_{\widetilde{L},\widetilde{S}\in\mathbb{R}^{m\times n}} \left|\left|\widetilde{L}\right|\right|_* + \lambda \left|\left|\widetilde{S}\right|\right|_1 \quad \text{such that } \widetilde{L} + \widetilde{S} = Y$$

is any matrix such that

$$Q = UV^T + W = \lambda \operatorname{sign}(S) + F$$

$$||W|| < 1 \qquad U^T W = 0 \qquad W\,V = 0$$

$$F_\Omega = 0 \qquad ||F||_\infty < \lambda$$

Convex program with equality constraints

Canonical problem with linear equality constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = y \end{aligned}$$

Lagrangian

$$\mathcal{L}(x, z) := f(x) + \langle z, Ax - y \rangle$$

z is a Lagrange multiplier

Dual function

$$g(z) := \inf_x f(x) + \langle z, Ax - y \rangle$$

Convex program with equality constraints

If strong duality holds for the optimal x^* and z^*

$$\begin{aligned} f(x^*) &= g(z^*) \\ &= \inf_x \mathcal{L}(x, z^*) \\ &\leq f(x^*) \end{aligned}$$

If x^* is unique and we know z^*, we can compute x^* by solving

$$\text{minimize} \quad \mathcal{L}(x, z^*)$$

Dual-ascent method

Find z^* using gradient ascent

Iterations:

- Primal variable update

$$x^{(k)} = \arg\min_x \mathcal{L}\left(x, z^{(k)}\right)$$

- Compute gradient of dual function at $z^{(k)}$

$$\nabla g\left(z^{(k)}\right) = Ax^{(k)} - y$$

- Dual variable update

$$z^{(k+1)} = z^{(k)} + \alpha^{(k)} \nabla g\left(z^{(k)}\right)$$

Augmented Lagrangian

Aim: Make dual-ascent method more robust

Augmented Lagrangian

$$\mathcal{L}_{\rho}(x, z) := f(x) + \langle z, Ax - y \rangle + \frac{\rho}{2} ||Ax - y||_2^2$$

Lagrangian of modified problem

$$\begin{aligned} \text{minimize} \quad & f(x) + \frac{\rho}{2} ||Ax - y||_2^2 \\ \text{subject to} \quad & Ax = y \end{aligned}$$

Method of multipliers

Iterations:

- Primal variable update

$$x^{(k)} = \arg\min_x \mathcal{L}_\rho \left(x, z^{(k)} \right)$$

- Compute $z^{(k+1)}$ such that

$$\nabla_x \mathcal{L} \left(x^{(k)}, z^{(k+1)} \right) = 0$$

Dual update

We have

$$\nabla_x \mathcal{L}_\rho \left(x^{(k)}, z^{(k)} \right) = 0$$

$$\nabla_x \mathcal{L}_\rho \left(x^{(k)}, z^{(k)} \right) = \nabla_x f \left(x^{(k)} \right) + A^T \left(z^{(k)} + \rho \left(Ax - y \right) \right)$$

$$\nabla_x \mathcal{L} \left(x^{(k)}, z^{(k)} + \rho \left(Ax - y \right) \right) = \nabla_x f \left(x^{(k)} \right) + A^T \left(z^{(k)} + \rho \left(Ax - y \right) \right)$$

So we can use the dual-ascent update with $\alpha_k = \rho$

$$z^{(k+1)} = z^{(k)} + \rho \left(Ax - y \right)$$

Alternating direction method of multipliers (ADMM)

Apply same ideas to

$$\begin{aligned} &\text{minimize} && f_1(x_1) + f_2(x_2) \\ &\text{subject to} && Ax_1 + Bx_2 = y \end{aligned}$$

Alternating direction method of multipliers (ADMM)

Iterations:

- Primal variable updates

$$x_1^{(k)} = \arg\min_x \mathcal{L}_\rho \left(x, x_2^{(k-1)}, z^{(k)} \right)$$
$$x_2^{(k)} = \arg\min_x \mathcal{L}_\rho \left(x_1^{(k)}, x, z^{(k)} \right)$$

- Dual variable update

$$z^{(k+1)} = z^{(k)} + \rho \left(A x_1^{(k)} + B x_2^{(k)} - y \right)$$

ADMM for robust PCA

Robust PCA problem

$$\min_{L,S \in \mathbb{R}^{n \times n}} ||L||_* + \lambda ||S||_1 \quad \text{such that } L + S = Y$$

Augmented Lagrangian

$$||L||_* + \lambda ||S||_1 + \langle Z, L + S - Y \rangle + \frac{\rho}{2} ||L + S - M||_F^2$$

Primal updates

$$
\begin{aligned}
L^{(k)} &= \arg\min_{L} \mathcal{L}_{\rho}\left(L, S^{(k-1)}, Z^{(k)}\right) \\
&= \arg\min_{L} ||L||_{*} + \left\langle Z^{(k)}, L \right\rangle + \frac{\rho}{2} \left|\left| L + S^{(k-1)} - M \right|\right|_{\mathsf{F}}^{2} \\
&= \mathcal{D}_{1/\rho}\left(\frac{1}{\rho} Z^{(k)} + S^{(k-1)} - M\right)
\end{aligned}
$$

$$
\begin{aligned}
S^{(k)} &= \arg\min_{S} \mathcal{L}_{\rho}\left(S, L^{(k-1)}, Z^{(k)}\right) \\
&= \arg\min_{S} \lambda\, ||S||_{1} + \left\langle Z^{(k)}, S \right\rangle + \frac{\rho}{2} \left|\left| L^{(k-1)} + S - M \right|\right|_{\mathsf{F}}^{2} \\
&= \mathcal{S}_{\lambda/\rho}\left(\frac{1}{\rho} Z^{(k)} + L^{(k)} - M\right)
\end{aligned}
$$

ADMM for robust PCA

Iterations:

- Primal variable updates

$$L^{(k)} = \mathcal{D}_{1/\rho}\left(\frac{1}{\rho}Z^{(k)} + S^{(k-1)} - M\right)$$

$$S^{(k)} = \mathcal{S}_{\lambda/\rho}\left(\frac{1}{\rho}Z^{(k)} + L^{(k)} - M\right)$$

- Dual variable update

$$Z^{(k+1)} = Z^{(k)} + \rho\left(L^{(k)} + S^{(k)} - M\right)$$

Background subtraction

Frame 17

Low-rank component

Sparse component

Frame 42

Low-rank component

Sparse component

Frame 75

Low-rank component

Sparse component