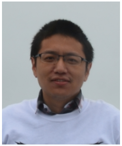
Sparse Bayesian Learning for Robust PCA

Jing Liu, Yacong Ding, Bhaskar D. Rao*

University of California, San Diego

ECE Department, DSP Lab

brao@ucsd.edu

May, 2019

Overview

- Background
- Model
- Algorithm & Analysis
- Experiments

Robust PCA

Candès et al.[1,2]:

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$$

$\boldsymbol{L} \in \mathbb{R}^{n_1 \times n_2}$: low-rank matrix;
$\boldsymbol{E} \in \mathbb{R}^{n_1 \times n_2}$: sparse matrix that captures outlier corruptions;
$\boldsymbol{N} \in \mathbb{R}^{n_1 \times n_2}$: inlier noise.

[1] E. J. Candès et al. "Robust Principal Component Analysis?". In: *J. ACM* 58.3 (June 2011), 11:1–11:37.

[2] Z. Zhou et al. "Stable Principal Component Pursuit". In: *2010 ISIT*. June 2010.

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t. \; \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t. \ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

Equivalent to ($n \triangleq \min(n_1, n_2)$):

$$\min_{\boldsymbol{U},\boldsymbol{V},\boldsymbol{s}\succeq 0,\boldsymbol{E}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{E}\|_0 \quad s.t. \ \|\boldsymbol{M} - \boldsymbol{U}\operatorname{diag}(\boldsymbol{s})\boldsymbol{V}^T - \boldsymbol{E}\|_F \leq \delta, \tag{2}$$

$$\boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \text{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

Equivalent to ($n \triangleq \min(n_1, n_2)$):

$$\min_{\boldsymbol{U},\boldsymbol{V},\boldsymbol{s}\succeq 0,\boldsymbol{E}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{U}\text{diag}(\boldsymbol{s})\boldsymbol{V}^T - \boldsymbol{E}\|_F \leq \delta, \\ \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.} \tag{2}$$

Further denote $\boldsymbol{m} = vec(\boldsymbol{M})$, $\boldsymbol{e} = vec(\boldsymbol{E})$, $\boldsymbol{U}_i$: ith column of $\boldsymbol{U}$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \\ \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.} \tag{3}$$

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t. \ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

(Candès et al., Chandrasekaran et al.):

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{E}\|_1 \quad s.t. \ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{4}$$

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

(Candès et al., Chandrasekaran et al.):

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{E}\|_1 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{4}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_{\cancel{0}1} + \lambda \|\boldsymbol{e}\|_{\cancel{0}1} \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{5}$$

$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

Robust PCA: $\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

(Candès et al., Chandrasekaran et al.):

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{E}\|_1 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{4}$$

Liu & Rao: Sparsity Regularized Principal Component Pursuit (SRPCP)[3]

$$\min_{\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{L}\|_* + \beta \|\boldsymbol{E}\|_0 + \lambda \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_1 \tag{6}$$

Exact recovery when no inlier noise, bounded error in the noisy case

[3] J. Liu and B. D. Rao. "Robust PCA via ℓ_0-ℓ_1 Regularization". In: *TSP* (Jan. 2019).

Our goal

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{3}$$

$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

Our goal

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \le \delta \tag{1}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \le \delta, \tag{3}$$
$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\cancel{\succeq 0},\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \le \delta, \tag{7}$$
$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1,\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d},\ \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}.$$

Our goal

$$\min_{\boldsymbol{L},\boldsymbol{E}} \text{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{3}$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\not\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{7}$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1,\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d},\ \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}.$$

Proposition

Set $d = n \triangleq \min(n_1, n_2)$ in (7). Then (1), (3) and (7) have the same global optimal solution(s) in terms of $\boldsymbol{L}$ and $\boldsymbol{E}$.

Our goal

$$\min_{\boldsymbol{L},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{L}) + \lambda \|\boldsymbol{E}\|_0 \quad s.t.\ \|\boldsymbol{M} - \boldsymbol{L} - \boldsymbol{E}\|_F \leq \delta \tag{1}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{3}$$

$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T),\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times n} \text{ and } \boldsymbol{V} \in \mathbb{R}^{n_2 \times n} \text{ orthonormal.}$$

$$\min_{\boldsymbol{A},\boldsymbol{s}\cancel{\succeq 0},\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta, \tag{7}$$

$$\boldsymbol{A}_i = \operatorname{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1,\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d},\ \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}.$$

Proposition

Set d ~~$= \min(n_1, n_2)$~~ $\in [\operatorname{rank}(\boldsymbol{L}_{opt}),\ \min(n_1, n_2)]$ in (7). Then (1), (3) and (7) have the same global optimal solution(s) in terms of $\boldsymbol{L}$ and $\boldsymbol{E}$.

Previous SBL approaches

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{E} + \boldsymbol{N}$$

- Ding et al. 2011: $\boldsymbol{L} = \boldsymbol{D}(\text{diag}(\boldsymbol{z})\text{diag}(\boldsymbol{s}))\boldsymbol{W}$, $\boldsymbol{E} = \boldsymbol{B} \circ \boldsymbol{X}$. Where $\boldsymbol{z}$, $\boldsymbol{B} \sim$ Bernoulli-Beta; $\boldsymbol{s}$, $\boldsymbol{X}$, $\boldsymbol{N}$ ∼Gaussian-Gamma; $\boldsymbol{D}, \boldsymbol{W} \sim$ Gaussian.
- Babacan et al. 2012: $\boldsymbol{L} = \boldsymbol{A}\boldsymbol{B}^T$, the columns of $\boldsymbol{A}$ and $\boldsymbol{B} \sim \mathcal{N}(0, \gamma_i^{-1}\boldsymbol{I})$, $\gamma_i \sim$ Gamma distribution. $\boldsymbol{E} \sim$ Gaussian-Jeffrey.
- Wipf 2012: columns of $\boldsymbol{L} \overset{i.i.d.}{\sim} \mathcal{N}(0, \Phi)$; $\boldsymbol{E}, \boldsymbol{N} \sim$ Gaussian.
- Jansson et al. 2015: $\text{vec}(\boldsymbol{L}) \sim \mathcal{N}(0, \Psi_{\boldsymbol{R}}^{-1} \otimes \Psi_{\boldsymbol{C}}^{-1}), \Psi_{\boldsymbol{R}}, \Psi_{\boldsymbol{C}}$ ∼Wishart. $\boldsymbol{E} + \boldsymbol{N}$ ∼Gaussian-Gamma.
- Wipf et al. 2016 (Pseudo-Bayes): $\text{vec}(\boldsymbol{L}) \sim \mathcal{N}(0, \Phi_{\boldsymbol{R}} \oplus \Phi_{\boldsymbol{C}})$; $\boldsymbol{E}, \boldsymbol{N} \sim$ Gaussian.

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A}, \boldsymbol{s} \not\geq 0, \boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t. \ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d}, \ \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}. \quad (7)$$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\cancel{\geq 0},\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \; s.t. \; \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d}, \; \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}. \quad (7)$$

Observation model:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \; \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \; \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; i = 1, ..., d.$$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\cancel{\succeq 0},\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \ s.t.\ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1,\ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1\times d},\ \boldsymbol{V} \in \mathbb{R}^{n_2\times d}. \quad (7)$$

Observation model:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t.\ \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T),\ \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1,\ i = 1, ..., d.$$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\cancel{\succeq 0},\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \; s.t. \; \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \le \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1\times d}, \; \boldsymbol{V} \in \mathbb{R}^{n_2\times d}. \quad (7)$$

Observation model:

$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \; \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \; \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; i = 1, ..., d.$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$
- $\boldsymbol{s} \sim \mathcal{N}(0, \boldsymbol{\Gamma}), \boldsymbol{\Gamma} \triangleq \text{diag}(\boldsymbol{\gamma})$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\not\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \; s.t. \; \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1\times d}, \; \boldsymbol{V} \in \mathbb{R}^{n_2\times d}. \tag{7}$$

Observation model:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \; \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \; \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; i = 1, ..., d.$$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$
- $\boldsymbol{s} \sim \mathcal{N}(0, \Gamma), \Gamma \triangleq \text{diag}(\gamma)$
- $\boldsymbol{e} \sim \mathcal{N}(0, \Lambda), \; \Lambda \triangleq \text{diag}(\alpha)$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\not\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \; s.t. \; \|\boldsymbol{m}-\boldsymbol{A}\boldsymbol{s}-\boldsymbol{e}\|_2 \le \delta,$$
$$\boldsymbol{A}_i = \mathrm{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1\times d}, \; \boldsymbol{V} \in \mathbb{R}^{n_2\times d}. \quad (7)$$

Observation model:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \; \boldsymbol{A}_i = \mathrm{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \; \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; i = 1, ..., d.$$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$
- $\boldsymbol{s} \sim \mathcal{N}(0, \boldsymbol{\Gamma}), \boldsymbol{\Gamma} \triangleq \mathrm{diag}(\boldsymbol{\gamma})$
- $\boldsymbol{e} \sim \mathcal{N}(0, \boldsymbol{\Lambda}), \; \boldsymbol{\Lambda} \triangleq \mathrm{diag}(\boldsymbol{\alpha})$
- $\boldsymbol{n} \sim \mathcal{N}(0, \beta\boldsymbol{I})$

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A}, \boldsymbol{s} \cancel{\geq 0}, \boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \; s.t. \; \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d}, \; \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}. \tag{7}$$

Observation model:

$$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \; \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i \boldsymbol{V}_i^T), \; \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \; i = 1, ..., d.$$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$
- $\boldsymbol{s} \sim \mathcal{N}(0, \Gamma), \Gamma \triangleq \text{diag}(\gamma)$
- $\boldsymbol{e} \sim \mathcal{N}(0, \Lambda), \; \Lambda \triangleq \text{diag}(\alpha)$
- $\boldsymbol{n} \sim \mathcal{N}(0, \beta \boldsymbol{I})$

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$.

Sparse Bayesian Learning model

$$\min_{\boldsymbol{A},\boldsymbol{s}\not\succeq 0,\boldsymbol{e}} \|\boldsymbol{s}\|_0 + \lambda\|\boldsymbol{e}\|_0 \ s.t. \ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \le \delta,$$
$$\boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \ \forall i, \boldsymbol{U} \in \mathbb{R}^{n_1 \times d}, \ \boldsymbol{V} \in \mathbb{R}^{n_2 \times d}. \quad (7)$$

Observation model:

$\boldsymbol{m} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} + \boldsymbol{n}, s.t. \ \boldsymbol{A}_i = \text{vec}(\boldsymbol{U}_i\boldsymbol{V}_i^T), \ \|\boldsymbol{U}_i\|_2 = \|\boldsymbol{V}_i\|_2 = 1, \ i = 1, ..., d.$

- $\boldsymbol{A}$: a deterministic parameter that lies in the above constrained space $\mathscr{A}$
- $\boldsymbol{s} \sim \mathcal{N}(0, \Gamma), \Gamma \triangleq \text{diag}(\gamma)$
- $\boldsymbol{e} \sim \mathcal{N}(0, \Lambda), \ \Lambda \triangleq \text{diag}(\alpha)$
- $\boldsymbol{n} \sim \mathcal{N}(0, \beta\boldsymbol{I})$

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$.
Then $\boldsymbol{s}$ and $\boldsymbol{e}$ can be estimated via the posterior mean of $p(\boldsymbol{s}|\boldsymbol{m}; \hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ and $p(\boldsymbol{e}|\boldsymbol{m}; \hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$.

Inference

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$ via MAP-EM.

Inference

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$ via MAP-EM.

- **E-Step:**

$$\mathrm{Q}(\boldsymbol{A}, \gamma, \alpha | \boldsymbol{A}^{(k)}, \gamma^{(k)}, \boldsymbol{\alpha}^{(k)}) = \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\boldsymbol{\alpha}^{(k)},\beta}\{-\log p(\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{e} | \boldsymbol{A}, \gamma, \alpha, \beta)\}$$

$$= \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\boldsymbol{\alpha}^{(k)},\beta}\{-\log p(\boldsymbol{m}|\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A}, \beta) - \log p(\boldsymbol{e}|\alpha) - \log p(\boldsymbol{s}|\gamma)\}$$

Inference

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$ via MAP-EM.

- **E-Step:**

$$Q(\boldsymbol{A}, \gamma, \alpha | \boldsymbol{A}^{(k)}, \gamma^{(k)}, \alpha^{(k)}) = \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\alpha^{(k)},\beta}\{-\log p(\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{e} | \boldsymbol{A}, \gamma, \alpha, \beta)\}$$

$$= \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\alpha^{(k)},\beta}\{-\log p(\boldsymbol{m}|\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A}, \beta) - \log p(\boldsymbol{e}|\alpha) - \log p(\boldsymbol{s}|\gamma)\}$$

- **M-Step:** $\min_{\gamma,\boldsymbol{\alpha},\boldsymbol{A}\in\mathscr{A}} Q(\boldsymbol{A}, \gamma, \alpha | \boldsymbol{A}^{(k)}, \gamma^{(k)}, \alpha^{(k)}) - \log p(\gamma)$

$$= \min_{\gamma,\boldsymbol{\alpha},\boldsymbol{A}\in\mathscr{A}} \frac{1}{2\beta}\langle \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2^2 \rangle + \frac{1}{2}\sum_i (\log \alpha_i + \frac{\langle \boldsymbol{e}_i^2 \rangle}{\alpha_i})$$

$$+ \frac{1}{2}\sum_i (\log \gamma_i + \frac{\langle \boldsymbol{s}_i^2 \rangle}{\gamma_i}) + \sum_i ((a+1)\log \gamma_i) + const$$

Inference

Goal: Infer $(\hat{\boldsymbol{A}}, \hat{\gamma}, \hat{\alpha})$ from the data $\boldsymbol{m}$ via MAP-EM.

- **E-Step:**

$$Q(\boldsymbol{A}, \gamma, \alpha | \boldsymbol{A}^{(k)}, \gamma^{(k)}, \boldsymbol{\alpha}^{(k)}) = \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\boldsymbol{\alpha}^{(k)},\beta}\{-\log p(\boldsymbol{m}, \boldsymbol{s}, \boldsymbol{e}|\boldsymbol{A}, \gamma, \alpha, \beta)\}$$

$$= \mathbb{E}_{\boldsymbol{s},\boldsymbol{e}|\boldsymbol{m};\boldsymbol{A}^{(k)},\gamma^{(k)},\boldsymbol{\alpha}^{(k)},\beta}\{-\log p(\boldsymbol{m}|\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A}, \beta) - \log p(\boldsymbol{e}|\alpha) - \log p(\boldsymbol{s}|\gamma)\}$$

- **M-Step:** $\min_{\gamma,\boldsymbol{\alpha},\boldsymbol{A}\in\mathscr{A}} Q(\boldsymbol{A}, \gamma, \alpha | \boldsymbol{A}^{(k)}, \gamma^{(k)}, \boldsymbol{\alpha}^{(k)}) - \log p(\gamma)$

$$= \min_{\gamma,\boldsymbol{\alpha},\boldsymbol{A}\in\mathscr{A}} \frac{1}{2\beta}\langle\|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2^2\rangle + \frac{1}{2}\sum_i(\log\boldsymbol{\alpha}_i + \frac{\langle\boldsymbol{e}_i^2\rangle}{\boldsymbol{\alpha}_i})$$

$$+ \frac{1}{2}\sum_i(\log\gamma_i + \frac{\langle\boldsymbol{s}_i^2\rangle}{\gamma_i}) + \sum_i((a+1)\log\gamma_i) + const$$

Employ Inverse-gamma prior on γ, i.e., $p(\gamma_i) = \text{IG}(a, b)$, with $b \to 0$.

Update rules

Update α: $\boldsymbol{\alpha}_i = \langle \mathbf{e}_i^2 \rangle = \boldsymbol{\mu}^2_{\mathbf{e}|\boldsymbol{m}}(i) + \boldsymbol{\Sigma}_{\mathbf{e}|\boldsymbol{m}}(i,i), \forall i.$

Update rules

Update α: $\boldsymbol{\alpha}_i = \langle \mathbf{e}_i^2 \rangle = \boldsymbol{\mu}_{\mathbf{e}|\mathbf{m}}^2(i) + \boldsymbol{\Sigma}_{\mathbf{e}|\mathbf{m}}(i,i), \forall i.$

Update γ: $\gamma_i = \langle \mathbf{s}_i^2 \rangle / (2a+3) = \left(\boldsymbol{\mu}_{\mathbf{s}|\mathbf{m}}^2(i) + \boldsymbol{\Sigma}_{\mathbf{s}|\mathbf{m}}(i,i) \right) / (2a+3), \forall i.$

Update rules

Update α: $\alpha_i = \langle e_i^2 \rangle = \mu^2_{e|m}(i) + \Sigma_{e|m}(i,i), \forall i.$

Update γ: $\gamma_i = \langle s_i^2 \rangle / (2a+3) = \left(\mu^2_{s|m}(i) + \Sigma_{s|m}(i,i)\right)/(2a+3), \forall i.$

Update A_1: Given $A_2^{(k)}, A_3^{(k)}, \cdots, A_d^{(k)}$,

$$A_1^{(k+1)} = \arg \min_{\substack{A_1 = \mathrm{vec}(U_1 V_1^T) \\ \|U_1\|_2 = 1 \\ \|V_1\|_2 = 1}} \|h - A_1\|_2^2 \tag{8}$$

where $h = \frac{\langle s_1 \rangle m - \langle s_1 \rangle \langle e \rangle - \Sigma^T_{se|m}(1,:) - \sum_{i=2}^d [\langle s_1 \rangle \langle s_i \rangle + \Sigma_{s|m}(1,i)] A_i^{(k)}}{\langle s_1 \rangle^2 + \Sigma_{s|m}(1,1)}$.

$$(U_1^{(k+1)}, V_1^{(k+1)}) = \arg \min_{\substack{U_1, V_1 \\ \|U_1\|_2 = 1 \\ \|V_1\|_2 = 1}} \|\mathrm{Mat}(h) - U_1 V_1^T\|_F^2. \tag{9}$$

Update rules

Update α: $\alpha_i = \langle e_i^2 \rangle = \mu^2_{e|m}(i) + \Sigma_{e|m}(i,i), \forall i.$

Update γ: $\gamma_i = \langle s_i^2 \rangle / (2a+3) = \left(\mu^2_{s|m}(i) + \Sigma_{s|m}(i,i)\right)/(2a+3), \forall i.$

Update A_1: Given $A_2^{(k)}, A_3^{(k)}, \cdots, A_d^{(k)}$,

$$A_1^{(k+1)} = \arg \min_{\substack{A_1 = \text{vec}(U_1 V_1^T) \\ \|U_1\|_2 = 1 \\ \|V_1\|_2 = 1}} \|h - A_1\|_2^2 \tag{8}$$

where $h = \frac{\langle s_1 \rangle m - \langle s_1 \rangle \langle e \rangle - \Sigma^T_{se|m}(1,:) - \sum_{i=2}^{d}[\langle s_1 \rangle \langle s_i \rangle + \Sigma_{s|m}(1,i)] A_i^{(k)}}{\langle s_1 \rangle^2 + \Sigma_{s|m}(1,1)}$.

$$(U_1^{(k+1)}, V_1^{(k+1)}) = \arg \min_{\substack{U_1, V_1 \\ \|U_1\|_2 = 1 \\ \|V_1\|_2 = 1}} \|\text{Mat}(h) - U_1 V_1^T\|_F^2. \tag{9}$$

Solution: the 1st singular vector pair of Mat(h)

Complexity & guarantee

Update $\boldsymbol{\alpha}$: $\mathcal{O}(d^2 n_1 n_2)$
Update $\boldsymbol{\gamma}$: $\mathcal{O}(d^2 n_1 n_2)$
Update $\boldsymbol{A}$: $\mathcal{O}(d^2 n_1 n_2)$
Initialize d to the same order of the rank r for large-scale problems.

Complexity & guarantee

Update α: $\mathcal{O}(d^2 n_1 n_2)$
Update γ: $\mathcal{O}(d^2 n_1 n_2)$
Update $\boldsymbol{A}$: $\mathcal{O}(d^2 n_1 n_2)$
Initialize d to the same order of the rank r for large-scale problems.

Under MAP-EM framework (Chen et al.'10)

Theorem

$$p(\boldsymbol{A}^{(k+1)}, \gamma^{(k+1)}, \alpha^{(k+1)} | \boldsymbol{m}) \geq p(\boldsymbol{A}^{(k)}, \gamma^{(k)}, \alpha^{(k)} | \boldsymbol{m})$$

Underlying SBL objective function

Type-II MAP, i.e., maximize $p(\boldsymbol{A}, \boldsymbol{\gamma}, \boldsymbol{\alpha} | \boldsymbol{m}) \propto p(\boldsymbol{m} | \boldsymbol{A}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) p(\boldsymbol{\gamma})$
Apply $-2\log(\cdot)$ transformation:

$$\begin{aligned}
&\min_{\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{A} \in \mathscr{A}} -2\log[p(\boldsymbol{m} | \boldsymbol{A}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) p(\boldsymbol{\gamma})] \\
&= \min_{\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{A} \in \mathscr{A}} m^T \boldsymbol{\Sigma}_{\boldsymbol{m}}^{-1} m + \log |\boldsymbol{\Sigma}_{\boldsymbol{m}}| + 2(a+1) \log |\boldsymbol{\Gamma}| + C \\
&= \min_{\boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{A} \in \mathscr{A}} \Big\{ \min_{\boldsymbol{s}, \boldsymbol{e}} [\frac{1}{\beta} \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2^2 + \boldsymbol{s}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{s} + \boldsymbol{e}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{e}] \\
&\quad + \log |\boldsymbol{\Sigma}_{\boldsymbol{m}}| + 2(a+1) \log |\boldsymbol{\Gamma}| \Big\} + C \\
&= \min_{\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A} \in \mathscr{A}} \Big\{ \frac{1}{\beta} \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2^2 \\
&\quad + \underbrace{\min_{\boldsymbol{\gamma}, \boldsymbol{\alpha}} [\boldsymbol{s}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{s} + \boldsymbol{e}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{e} + \log |\boldsymbol{\Sigma}_{\boldsymbol{m}}| + 2(a+1) \log |\boldsymbol{\Gamma}|]}_{g_{SBL}(\boldsymbol{A}, \boldsymbol{s}, \boldsymbol{e})} \Big\} + C
\end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{m}} = \boldsymbol{A} \boldsymbol{\Gamma} \boldsymbol{A}^T + \boldsymbol{\Lambda} + \beta \boldsymbol{I}$

Underlying SBL objective function

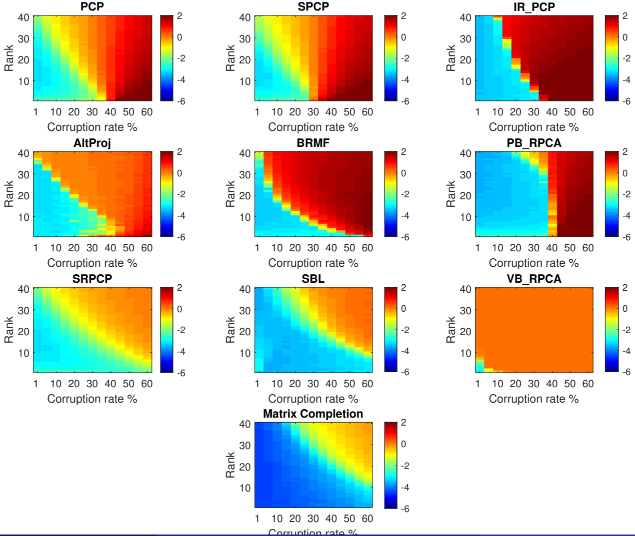
Type-II MAP, i.e., maximize $p(\boldsymbol{A}, \gamma, \boldsymbol{\alpha} | \boldsymbol{m}) \propto p(\boldsymbol{m} | \boldsymbol{A}, \gamma, \boldsymbol{\alpha}) p(\gamma)$

$$\min_{\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A} \in \mathscr{A}} \|\boldsymbol{s}\|_0 + \lambda \|\boldsymbol{e}\|_0 \ s.t. \ \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2 \leq \delta. \tag{7}$$

$$\min_{\boldsymbol{s}, \boldsymbol{e}, \boldsymbol{A} \in \mathscr{A}} \Bigg\{ \frac{1}{\beta} \|\boldsymbol{m} - \boldsymbol{A}\boldsymbol{s} - \boldsymbol{e}\|_2^2 + \underbrace{\min_{\gamma, \boldsymbol{\alpha}} [\boldsymbol{s}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{s} + \boldsymbol{e}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{e} + \log |\boldsymbol{A} \boldsymbol{\Gamma} \boldsymbol{A}^T + \boldsymbol{\Lambda} + \beta \boldsymbol{I}| + 2(a+1) \log |\boldsymbol{\Gamma}|] \Bigg\}}_{g_{SBL}(\boldsymbol{A}, \boldsymbol{s}, \boldsymbol{e})} + C$$

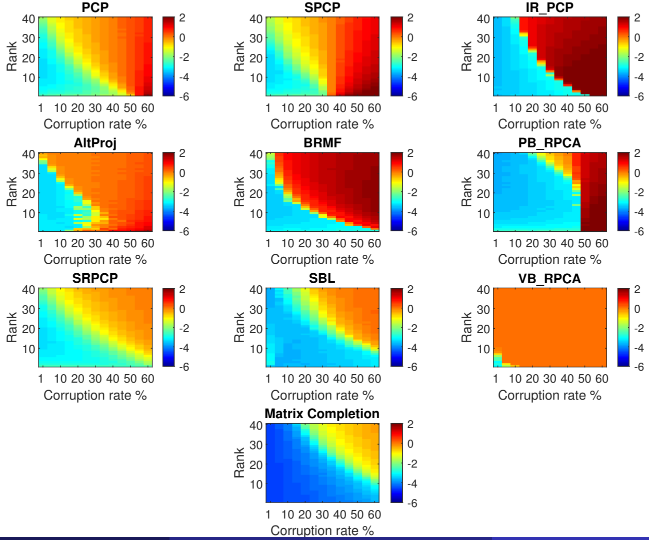
$\frac{\|\hat{L}-L_0\|_F^2}{\|L_0\|_F^2}$ of each method in log scale

$L_0 = AB^T$; $A^{n\times r}$, $B^{n\times r}$: standard Gaussian matrices; Corruptions drawn from U[0,100]



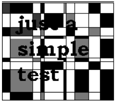
$\frac{\|\hat{L}-L_0\|_F^2}{\|L_0\|_F^2}$ of each method in log scale

$L_0 = AB^T$; $A^{n\times r}$, $B^{n\times r}$: standard Gaussian matrices; Corruptions drawn from U[-100,100]

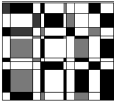


Recovered text mask (F-measure) and low-rank background

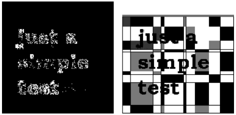
Input



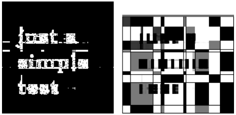
Ground Truth



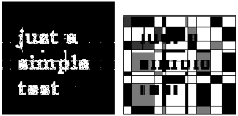
PCP, F=0.537, error=81.7



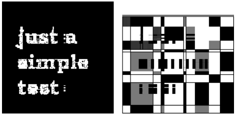
IR_PCP, F=0.799, error=70.6



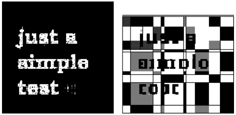
BRMF, F=0.867, error=73.1



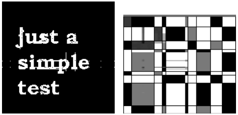
AltProj, F=0.890, error=71.9



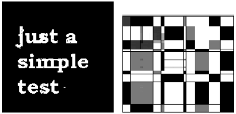
PB_RPCA, F=0.903, error=76.4



SRPCP, F=0.967, error=17.4



SBL, F=0.971, error=17.4



$$\text{F-measure}=2(\textit{precision}\times\textit{recall})/(\textit{precision}+\textit{recall})$$

Conclusion

1. Proposed a Robust Sparse Linear Regression objective, equivalent to the fundamental minimizing "rank+sparsity" objective of Robust PCA;
2. To solve this objective, a concise SBL method is proposed, which has minimum assumptions and effectively deals with the requirements of the problem, and allows exact inference;
3. The underlying cost function of the proposed SBL method is shown to lead to "sparse and low-rank decomposition";
4. Empirical studies demonstrate the superiority of the proposed method.