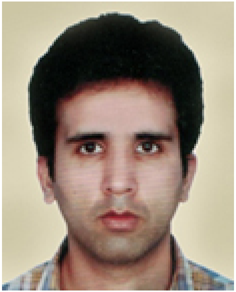
Structured recovery for imaging and image processing

Justin Romberg, Georgia Tech ECE
SPARS, Cambridge, UK
July 8, 2015

Collaborators



Ali Ahmed
(now at MIT)

Sohail Bahmani

Alireza Aghasi

Quadratic and bilinear equations

Simple (but only recently appreciated) observation:
Systems of bilinear equations, e. g.

$$u_1 v_1 + 5u_1 v_2 + 7u_2 v_3 = -12$$
$$u_3 v_1 - 9u_2 v_2 + 4u_3 v_2 = 2$$

can be recast as *linear system of equations on a matrix that has rank 1*:

$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 & \cdots & u_1v_N \\ u_2v_1 & u_2v_2 & u_2v_3 & \cdots & u_2v_N \\ u_3v_1 & u_3v_2 & u_3v_3 & \cdots & u_3v_N \\ \vdots & \vdots & & \ddots & \\ u_Kv_1 & u_Kv_2 & u_Kv_3 & \cdots & u_Kv_N \end{bmatrix}$$

Quadratic and bilinear equations

Simple (but only recently appreciated) observation:
Systems of bilinear equations, e. g.

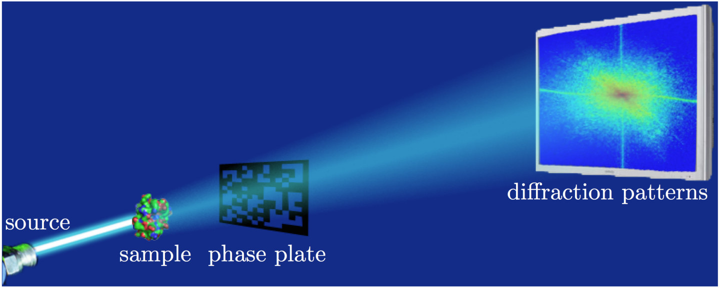
$$u_1v_1 + 5u_1v_2 + 7u_2v_3 = -12$$
$$u_3v_1 - 9u_2v_2 + 4u_3v_2 = 2$$

can be recast as *linear system of equations on a matrix that has rank 1*:

$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 & \cdots & u_1v_N \\ u_2v_1 & u_2v_2 & u_2v_3 & \cdots & u_2v_N \\ u_3v_1 & u_3v_2 & u_3v_3 & \cdots & u_3v_N \\ \vdots & \vdots & & \ddots & \\ u_Kv_1 & u_Kv_2 & u_Kv_3 & \cdots & u_Kv_N \end{bmatrix}$$

Compressive (low rank) recovery $\Rightarrow$
"Generic" quadratic/bilinear systems with cN equations and N unknowns can be solved using nuclear norm minimization

Phase retrieval



(image courtesy of M. Soltanolkotabi)

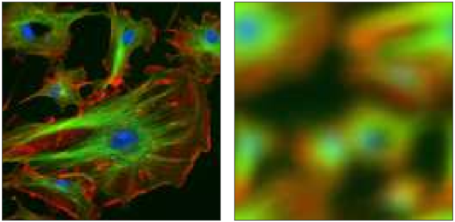
Observe the *magnitude* of the Fourier transform $|\hat{x}(\omega)|^2$

$\hat{x}(\omega)$ is complex, and its phase carries important information

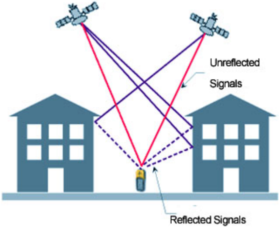
(Candes, Eldar, Li, Soltanolkotabi, Strohmer, and Voroninski)

Blind deconvolution

image deblurring



multipath in wireless comm



(image from EngineeringsALL)

We observe

$$y[\ell] = \sum_n s[n]\, h[\ell - n]$$

and want to "untangle" $\boldsymbol{s}$ and $\boldsymbol{h}$.

Blind deconvolution as low rank recovery

Each sample of $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h}$ is a bilinear combination of the unknowns,

$$y[\ell] = \sum_n s[n]h[\ell - n]$$

and is a *linear* combination of $\boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$:

$$\begin{array}{r} \\ \\ y_1[0] \\ y_1[1] \\ y_1[2] \\ \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ \bullet \\ y_1[9] \end{array} \begin{bmatrix} s[-2]h[0] & s[-2]h[1] & s[-2]h[2] \\ s[-1]h[0] & s[-1]h[1] & s[-1]h[2] \\ s[0]h[0] & s[0]h[1] & s[0]h[2] \\ s[1]h[0] & s[1]h[1] & s[1]h[2] \\ s[2]h[0] & s[2]h[1] & s[2]h[2] \\ s[3]h[0] & s[3]h[1] & s[3]h[2] \\ s[4]h[0] & s[4]h[1] & s[4]h[2] \\ s[5]h[0] & s[5]h[1] & s[5]h[2] \\ s[6]h[0] & s[6]h[1] & s[6]h[2] \\ s[7]h[0] & s[7]h[1] & s[7]h[2] \\ s[8]h[0] & s[8]h[1] & s[8]h[2] \\ s[9]h[0] & s[9]h[1] & s[9]h[2] \end{bmatrix}$$

Blind deconvolution as low rank recovery

Given $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h}$, it is impossible to untangle $\boldsymbol{s}$ and $\boldsymbol{h}$ unless we make some *structural assumptions*

Structure: $\boldsymbol{s}$ and $\boldsymbol{h}$ live in known *subspaces* of $\mathbb{R}^L$; we can write

$$\boldsymbol{s} = \boldsymbol{B}\boldsymbol{u}, \qquad \boldsymbol{h} = \boldsymbol{C}\boldsymbol{v}, \quad B : L \times K, \;\; C : L \times N$$

where $\boldsymbol{B}$ and $\boldsymbol{C}$ are matrices whose columns form bases for these spaces

Blind deconvolution as low rank recovery

Given $\boldsymbol{y} = \boldsymbol{s} * \boldsymbol{h}$, it is impossible to untangle $\boldsymbol{s}$ and $\boldsymbol{h}$ unless we make some *structural assumptions*

Structure: $\boldsymbol{s}$ and $\boldsymbol{h}$ live in known *subspaces* of $\mathbb{R}^L$; we can write

$$\boldsymbol{s} = \boldsymbol{B}\boldsymbol{u}, \qquad \boldsymbol{h} = \boldsymbol{C}\boldsymbol{v}, \quad B : L \times K, \;\; C : L \times N$$

where $\boldsymbol{B}$ and $\boldsymbol{C}$ are matrices whose columns form bases for these spaces

We can now write blind deconvolution as a *linear inverse problem with a rank contraint*:

$$\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0), \quad \text{where} \;\; \boldsymbol{X}_0 = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \;\; \text{has rank=1}$$

The action of $\mathcal{A}(\cdot)$ can be broken down into three linear steps:

$$\boldsymbol{X}_0 \;\rightarrow\; \boldsymbol{B}\boldsymbol{X}_0 \;\rightarrow\; \boldsymbol{B}\boldsymbol{X}_0\boldsymbol{C}^{\mathrm{T}} \;\rightarrow\; \text{take skew-diagonal sums}$$

Blind deconvolution theoretical results

We observe

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{s} * \boldsymbol{h}, \qquad \boldsymbol{h} = \boldsymbol{B}\boldsymbol{u}, \quad \boldsymbol{s} = \boldsymbol{C}\boldsymbol{v} \\ &= \mathcal{A}(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}), \qquad \boldsymbol{u} \in \mathbb{R}^K, \quad \boldsymbol{v} \in \mathbb{R}^N, \end{aligned}$$

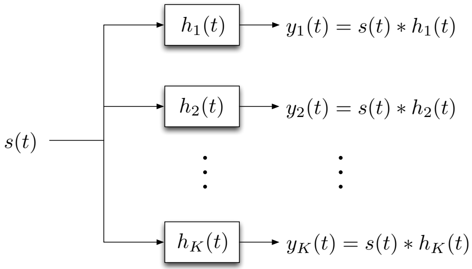
and then solve

$$\min_{\boldsymbol{X}} \ \|\boldsymbol{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}.$$

Ahmed, Recht, R, '12:
If $\boldsymbol{B}$ is "incoherent" in the Fourier domain, and $\boldsymbol{C}$ is randomly chosen, then we will recover $\boldsymbol{X}_0 = \boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$ exactly (with high probability) when

$$L \ \geq \ \mathrm{Const} \cdot (K+N) \cdot \log^3(KN)$$

Passive estimation of multiple channels

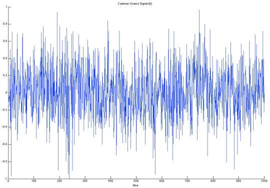


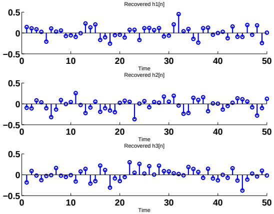
$$
\begin{array}{c|ccc|c|ccc|c|ccc|}
 & s[-2]h_1[0] & s[-2]h_1[1] & s[-2]h_1[2] & & s[-2]h_2[0] & s[-2]h_2[1] & s[-2]h_2[2] & & s[-2]h_3[0] & s[-2]h_3[1] & s[-2]h_3[2] \\
 & s[-1]h_1[0] & s[-1]h_1[1] & s[-1]h_1[2] & & s[-1]h_2[0] & s[-1]h_2[1] & s[-1]h_2[2] & & s[-1]h_3[0] & s[-1]h_3[1] & s[-1]h_3[2] \\
y_1[0] & s[0]h_1[0] & s[0]h_1[1] & s[0]h_1[2] & y_2[0] & s[0]h_2[0] & s[0]h_2[1] & s[0]h_2[2] & y_3[0] & s[0]h_3[0] & s[0]h_3[1] & s[0]h_3[2] \\
y_1[1] & s[1]h_1[0] & s[1]h_1[1] & s[1]h_1[2] & y_2[1] & s[1]h_2[0] & s[1]h_2[1] & s[1]h_2[2] & y_3[1] & s[1]h_3[0] & s[1]h_3[1] & s[1]h_3[2] \\
y_1[2] & s[2]h_1[0] & s[2]h_1[1] & s[2]h_1[2] & y_2[2] & s[2]h_2[0] & s[2]h_2[1] & s[2]h_2[2] & y_3[2] & s[2]h_3[0] & s[2]h_3[1] & s[2]h_3[2] \\
 & s[3]h_1[0] & s[3]h_1[1] & s[3]h_1[2] & \cdot & s[3]h_2[0] & s[3]h_2[1] & s[3]h_2[2] & \cdot & s[3]h_3[0] & s[3]h_3[1] & s[3]h_3[2] \\
\cdot & s[4]h_1[0] & s[4]h_1[1] & s[4]h_1[2] & \cdot & s[4]h_2[0] & s[4]h_2[1] & s[4]h_2[2] & \cdot & s[4]h_3[0] & s[4]h_3[1] & s[4]h_3[2] \\
\cdot & s[5]h_1[0] & s[5]h_1[1] & s[5]h_1[2] & \cdot & s[5]h_2[0] & s[5]h_2[1] & s[5]h_2[2] & \cdot & s[5]h_3[0] & s[5]h_3[1] & s[5]h_3[2] \\
\cdot & s[6]h_1[0] & s[6]h_1[1] & s[6]h_1[2] & \cdot & s[6]h_2[0] & s[6]h_2[1] & s[6]h_2[2] & \cdot & s[6]h_3[0] & s[6]h_3[1] & s[6]h_3[2] \\
\cdot & s[7]h_1[0] & s[7]h_1[1] & s[7]h_1[2] & \cdot & s[7]h_2[0] & s[7]h_2[1] & s[7]h_2[2] & \cdot & s[7]h_3[0] & s[7]h_3[1] & s[7]h_3[2] \\
\cdot & s[8]h_1[0] & s[8]h_1[1] & s[8]h_1[2] & & s[8]h_2[0] & s[8]h_2[1] & s[8]h_2[2] & & s[8]h_3[0] & s[8]h_3[1] & s[8]h_3[2] \\
y_1[9] & s[9]h_1[0] & s[9]h_1[1] & s[9]h_1[2] & y_2[9] & s[9]h_2[0] & s[9]h_2[1] & s[9]h_2[2] & y_3[9] & s[9]h_3[0] & s[9]h_3[1] & s[9]h_3[2]
\end{array}
$$

Recovery results

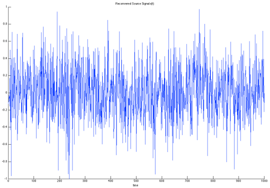
Source / output length: 1000
Number of channels: 100
Channel impulse response length: 50

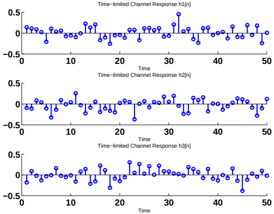
Original:



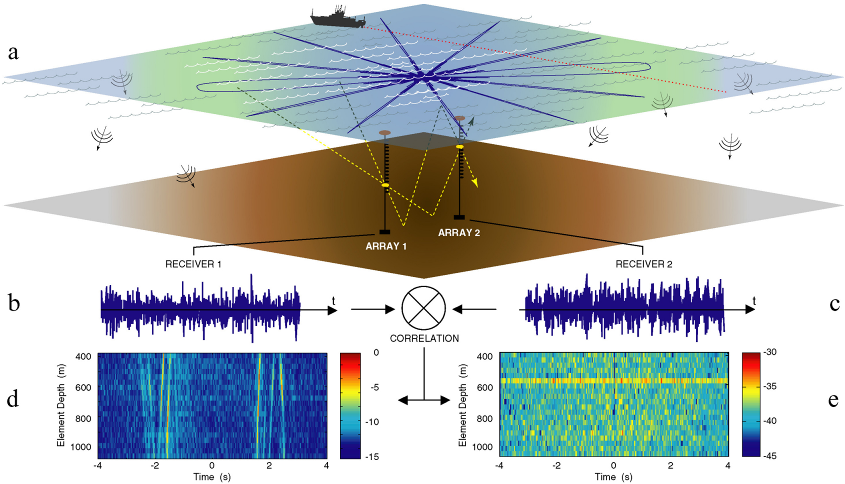


Recovered:

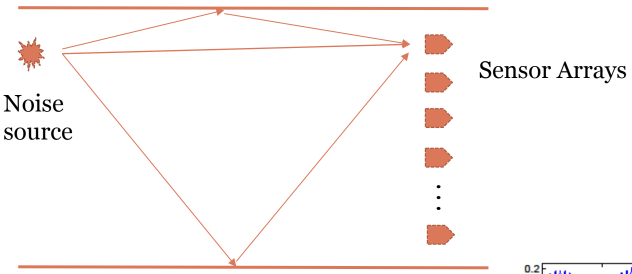




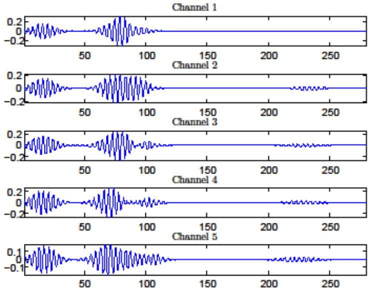
Passive imaging of the ocean



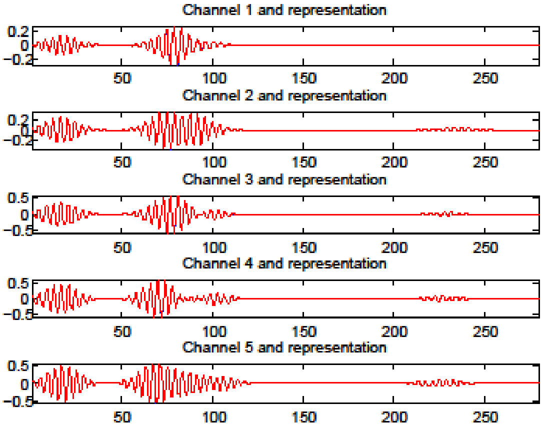
Realistic (simulated) ocean channels



- Noise signal is in the broad band 400~600Hz
- The distance between the noise source and sensor arrays is approximate 1km

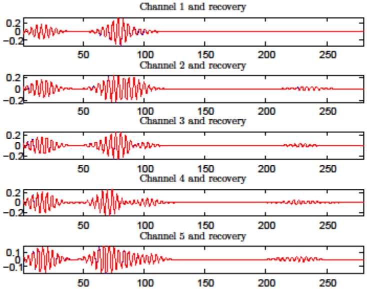


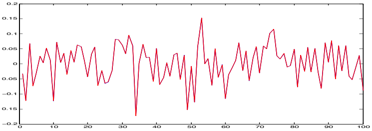
Realistic (simulated) ocean channels



Build a subspace model using bandwidth and approximate arrival times (about 20 dimensions per channel)

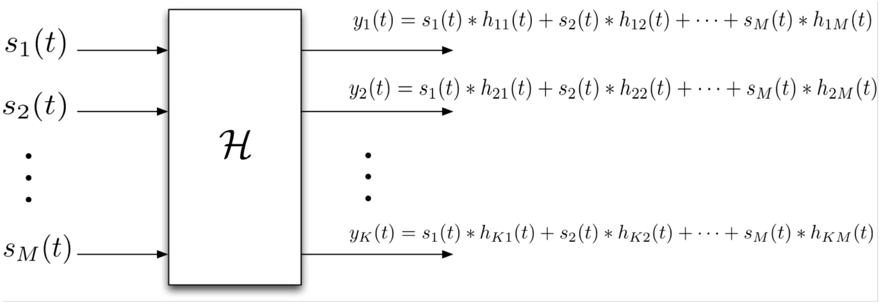
Simulated recovery





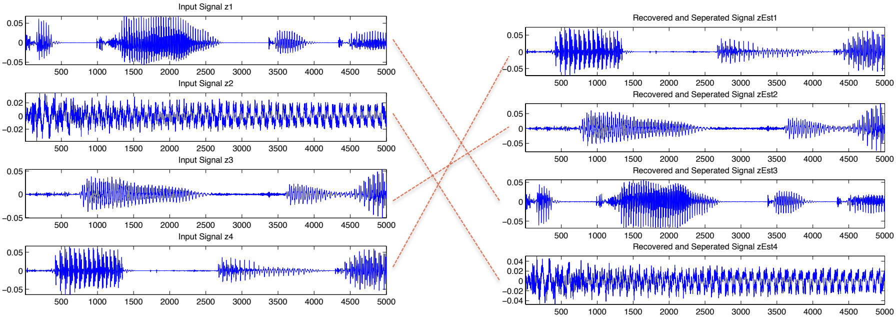
~ 100 channels total, ~ 2000 samples per channel,
Normalized error $\sim 10^{-4}$ (no noise), robust with noise

Multiple sources



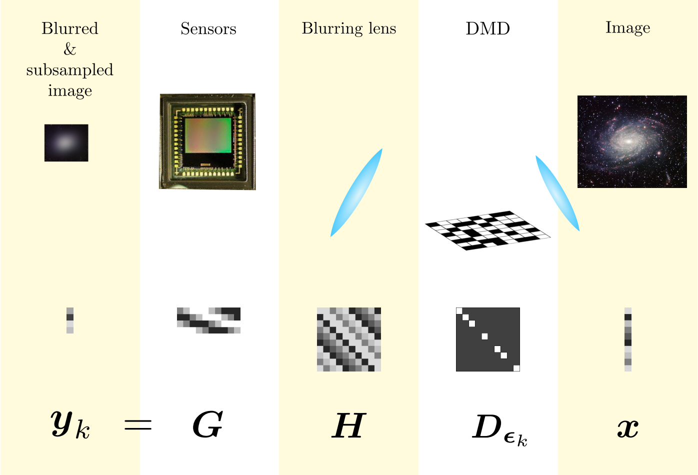
- Memoryless: structured matrix factorization (SMF) problem
  ICA, NNMF, dictionary learning, etc.
- Use matrix recovery to make convolutional channels "memoryless": recover rank M matrix, run SMF on column space

Low-rank recovery + ICA on broadband voice

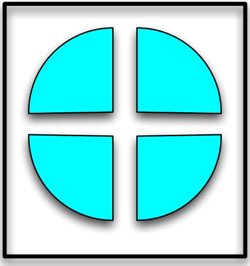


4 sources
30 channels (microphones)
2000 time samples
10 taps per channel

Imaging architecture



- Small number of sensors with gaps between them
- Blurring introduced to "fill in" these gaps
- Uncalibrated: blur kernel is unknown



Masked imaging linear algebra

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{GH} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix}$$

unknown

linear constraints

$$\boldsymbol{GH} = \sum_i a[i]\, \boldsymbol{B}_i$$

known

unknown image,

unconstrained

- Operator coefficients $\boldsymbol{a}$, image $\boldsymbol{x}$ unknown
- Observations: $\mathcal{A}(\boldsymbol{a}\boldsymbol{x}^{\mathrm{T}})$
- Alternative interpretation: *structured matrix factorization*

$$\boldsymbol{Y} = (\boldsymbol{GH})\, \mathrm{diag}(\boldsymbol{X}) \boldsymbol{\Phi}^{\mathrm{T}}$$

Masked imaging: theoretical results

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{GH} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix}$$

known

unknown
linear constraints

$$\boldsymbol{GH} = \sum_i a[i]\, \boldsymbol{B}_i$$

unknown image,
unconstrained

L pixels, N sensors, K codes

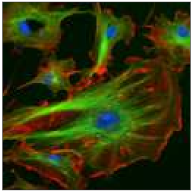
Theorem (Bahimani, R '14):
We can jointly recover the blur $\boldsymbol{H}$ and the image $\boldsymbol{X}$ for a number of codes:

$$K \gtrsim \mu^2 \frac{L}{N} \cdot \log^3(L) \log\log N$$

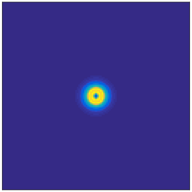
$\mu^2 \geq 1$ measures how spread out blur is in frequency

(Related work by Tang and Recht '14)

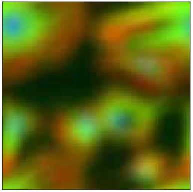
Masked imaging: numerical results



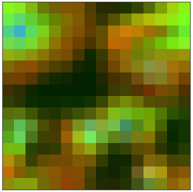
original



blur



blurred image

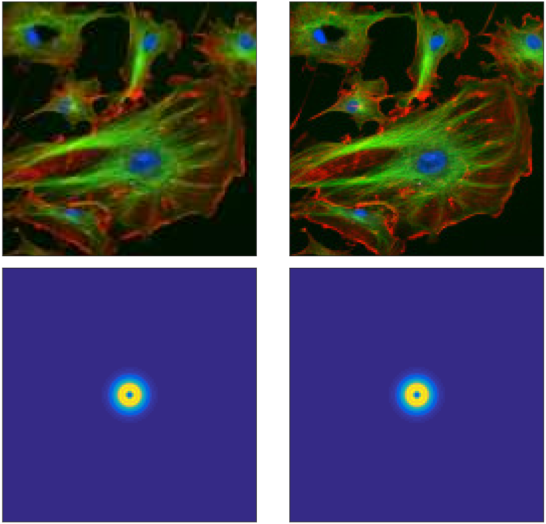


blurred, subsampled

- No structural model for the image
- Blur model: build basis from psfs over a range of focal lengths (EPFL PSF Generator, Born and Wolf model)

Masked imaging: numerical results

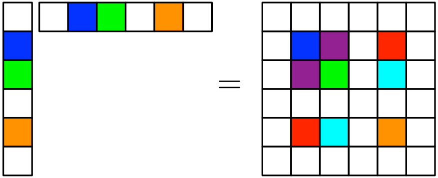
Recovery results: 16k pixels, 64 sensors, 200 codes



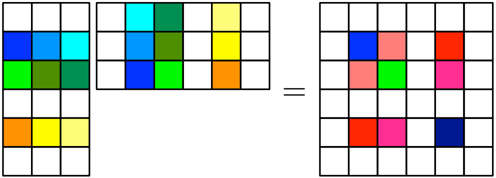
originals recovery

Simultaneous sparse and low rank

Problem 1: We want to recover $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$ when $\boldsymbol{x}$ is K sparse



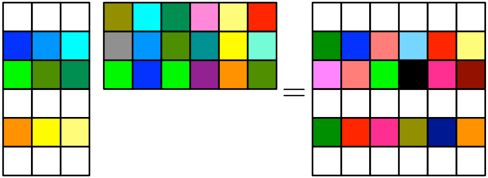
or more generally $\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}$, where $\boldsymbol{W} : N \times R$ is row sparse



We would like $\sim KR$ measurements instead of $\sim K^2$ or $\sim KN$

Simultaneous sparse and low rank

Problem 2: We want to recover $\boldsymbol{X}$ with only K active rows



Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Prior work

For $\boldsymbol{X}$ simultaneously sparse and low rank ...

- For $\mathcal{A}(\cdot)$ a random projection, it stably embeds this set of matrices (RIP)
  (Golbabaee '12, Lee et al '13)
- Convex relaxation for phase retrieval problem,
  $y_m = \langle \boldsymbol{X}, \boldsymbol{a}_m \boldsymbol{a}_m^{\mathrm{T}} \rangle, \quad m = 1, \ldots, M$, success for $M \sim K^2 \log N$
  (Li and Voron. '13)
  Extended to rank-R (Chen et al '13)
- Convex relaxation is generally not the best strategy (Oymak et al, '13)
  (M. Fazel's talk yesterday)
- Alternating minimization has similar guarantees (Netrapalli et al '13)
- Numerical results (Shechtman et al '11-'14, Schniter et al '15) suggest that we can do much better
- Identifiability (Li et al '15) conditions for blind deconvolution also suggest we can do much better
- Sparse power factorization (Lee et al '13) is efficient from observations that have RIP

Compressive phase retrieval with correlated measurements

Observe

$$y_\ell = |\langle \boldsymbol{x}, \boldsymbol{a}_\ell \rangle|^2 + \text{noise}, \quad \ell = 1, \ldots, L$$

$\boldsymbol{x} \in \mathbb{R}^N$ K-sparse, $\boldsymbol{a}_\ell$ *structured*

$$\boldsymbol{a}_\ell = \boldsymbol{\Psi}^{\mathrm{T}} \boldsymbol{w}_\ell,$$

$\boldsymbol{\Psi}^{\mathrm{T}}$ is $M \times N$, $\boldsymbol{w}_\ell$ are generic (random)

$\boldsymbol{a}_\ell$ all lie in a subspace $\Rightarrow$ we can decouple the recovery into two stages

Note

$$y_\ell = \langle \boldsymbol{x}\boldsymbol{x}^*, \boldsymbol{\Psi}^{\mathrm{T}} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^{\mathrm{T}} \boldsymbol{\Psi} \rangle = \langle \boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Psi}^{\mathrm{T}}, \boldsymbol{w}_\ell \boldsymbol{w}_\ell^{\mathrm{T}} \rangle$$

Two-stage decoding

Given

$$y_\ell = \langle \boldsymbol{x}_0 \boldsymbol{x}_0^*, \boldsymbol{\Psi}^{\mathrm{T}} \boldsymbol{w}_\ell \boldsymbol{w}_\ell^{\mathrm{T}} \boldsymbol{\Psi} \rangle = \langle \boldsymbol{\Psi} \boldsymbol{X}_0 \boldsymbol{\Psi}^{\mathrm{T}}, \boldsymbol{w}_\ell \boldsymbol{w}_\ell^{\mathrm{T}} \rangle, \quad \boldsymbol{y} = \mathcal{W}(\boldsymbol{\Psi} \boldsymbol{X}_0 \boldsymbol{\Psi}^{\mathrm{T}})$$

we solve

$$\hat{\boldsymbol{B}} = \arg\min_{\boldsymbol{B} \succeq \boldsymbol{0}} \ \operatorname{trace}(\boldsymbol{B}) \quad \text{subject to} \quad \|\mathcal{W}(\boldsymbol{B}) - \boldsymbol{y}\|_2 \leq \epsilon,$$

and then

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_1 \quad \text{subject to} \quad \|\boldsymbol{\Psi} \boldsymbol{X} \boldsymbol{\Psi}^{\mathrm{T}} - \hat{\boldsymbol{B}}\|_F \leq \frac{C\epsilon}{\sqrt{M}}$$

(We might use many different algorithms for these two steps, and get the same guarantees...)

Recovery guarantees

Observe

$$y_\ell = |\langle \boldsymbol{x}_0, \boldsymbol{a}_\ell \rangle|^2 + z_\ell, \quad \boldsymbol{a}_\ell = \boldsymbol{\Psi}^{\mathrm{T}} \boldsymbol{w}_\ell, \quad \ell = 1, \ldots, L$$

Suppose that

- The $\boldsymbol{w}_\ell \in \mathbb{R}^M$ are $\mathrm{Normal}(0, \mathbf{I})$
- The matrix $\boldsymbol{\Psi}$ embeds K-sparse vectors ($2K$-RIP)
- The noise $\boldsymbol{z}$ is bounded, $\|\boldsymbol{z}\|_2 \leq \epsilon$

Then if

$$L \geq C_1 K \log(N/K)$$

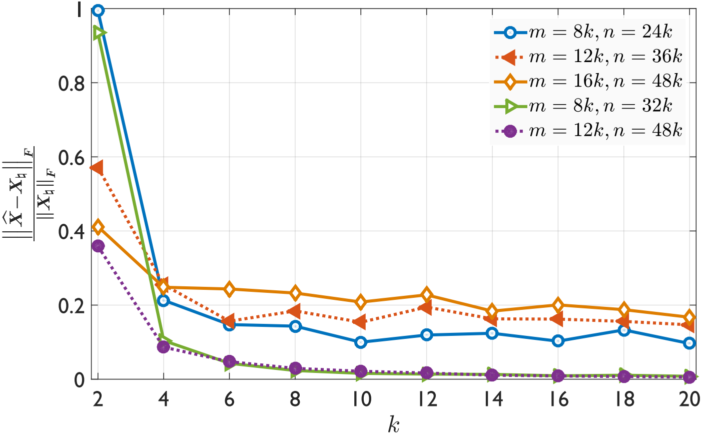
the two stage algorithm produces an estimate $\hat{\boldsymbol{X}}$ such that

$$\|\hat{\boldsymbol{X}} - \boldsymbol{x}_0 \boldsymbol{x}_0^{\mathrm{T}}\|_F \;\leq\; C_2 \frac{\epsilon}{\sqrt{L}}$$

with high probability, uniform over all rank-1 $k \times k$ sparse matrices.

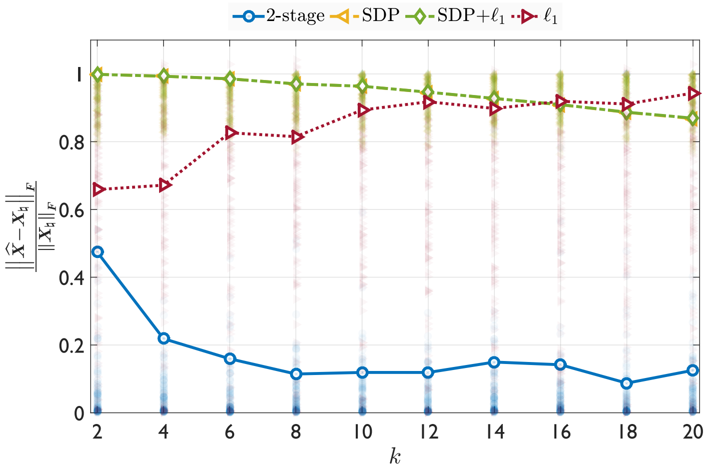
(Bahmani,R '15)

Numerical results



Recovery error vs. sparsity for different (M, L)

Numerical results



Recovery error vs. sparsity for two-stage, just ℓ_1,
just SDP low rank, SDP low rank, sparse

Covariance sketching

Data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_Q$ with covariance $\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] = \boldsymbol{R}$.

Compress/sketch data by correlating against one of L vectors $\boldsymbol{a}_\ell$, $\langle \boldsymbol{x}_t, \boldsymbol{a}_\ell \rangle$

Then if we used the same $\boldsymbol{a}_\ell$ for all $t \in \mathcal{T}_\ell$,

$$\frac{1}{|\mathcal{T}_\ell|} \sum_{t \in \mathcal{T}_\ell} |\langle \boldsymbol{x}_t, \boldsymbol{a}_\ell \rangle|^2 \approx \boldsymbol{a}_\ell^{\mathrm{T}} \boldsymbol{R} \boldsymbol{a}_\ell$$

$\Rightarrow$ estimating $\boldsymbol{R}$ from sketches is similar to phase retrieval (Chen et al '13)

Covariance sketching

Data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_Q$ with covariance $\mathrm{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}] = \boldsymbol{R}$.

Compress/sketch data by correlating against one of L vectors $\boldsymbol{a}_\ell$, $\langle \boldsymbol{x}_t, \boldsymbol{a}_\ell \rangle$

Then if we used the same $\boldsymbol{a}_\ell$ for all $t \in \mathcal{T}_\ell$,

$$\frac{1}{|\mathcal{T}_\ell|} \sum_{t \in \mathcal{T}_\ell} |\langle \boldsymbol{x}_t, \boldsymbol{a}_\ell \rangle|^2 \approx \boldsymbol{a}_\ell^{\mathrm{T}} \boldsymbol{R} \boldsymbol{a}_\ell$$

$\Rightarrow$ estimating $\boldsymbol{R}$ from sketches is similar to phase retrieval (Chen et al '13)

If $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}}$, where $\boldsymbol{V}$ is $N \times R$ and K-row sparse, $L \sim RK \log(N/K)$ then we find $\hat{\boldsymbol{V}}$ such that

$$\|\boldsymbol{R} - \hat{\boldsymbol{V}}\hat{\boldsymbol{V}}^{\mathrm{T}}\|_F \leq \mathrm{Const} \frac{\epsilon}{\sqrt{L}}$$

We can do so while only manipulating matrices of size $\sim NR$

(Bahmani, R '15)

Simultaneously sparse and low rank recovery from nested measurements

Suppose $\boldsymbol{X}_0$ is $N \times N$, K-row sparse, rank R, $\boldsymbol{\Phi}$ is $M \times N$. Measure

$$\boldsymbol{y} = \mathcal{W}(\boldsymbol{\Psi}\boldsymbol{X}_0) + \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathrm{Normal}(0, \sigma^2\mathbf{I})$$

If $\mathcal{W}(\cdot)$ is $c_1 R$-RIP, $\boldsymbol{\Psi}$ is $c_2 K$-RIP, then two-stage recovery yields $\hat{\boldsymbol{X}}$ with

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \lesssim \sigma\sqrt{R\max(M,N)}.$$

If $\boldsymbol{\Phi}$ is a good CS matrix, take $M \sim K\log N$.

Simultaneously sparse and low rank recovery from nested measurements

Suppose $\boldsymbol{X}_0$ is $N \times N$, K-row sparse, rank R, $\boldsymbol{\Phi}$ is $M \times N$. Measure

$$\boldsymbol{y} = \mathcal{W}(\boldsymbol{\Psi}\boldsymbol{X}_0) + \boldsymbol{z}, \quad \boldsymbol{z} \sim \mathrm{Normal}(0, \sigma^2 \mathbf{I})$$

If $\mathcal{W}(\cdot)$ is $c_1 R$-RIP, $\boldsymbol{\Psi}$ is $c_2 K$-RIP, then two-stage recovery yields $\hat{\boldsymbol{X}}$ with

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \lesssim \sigma \sqrt{R \max(M, N)}.$$

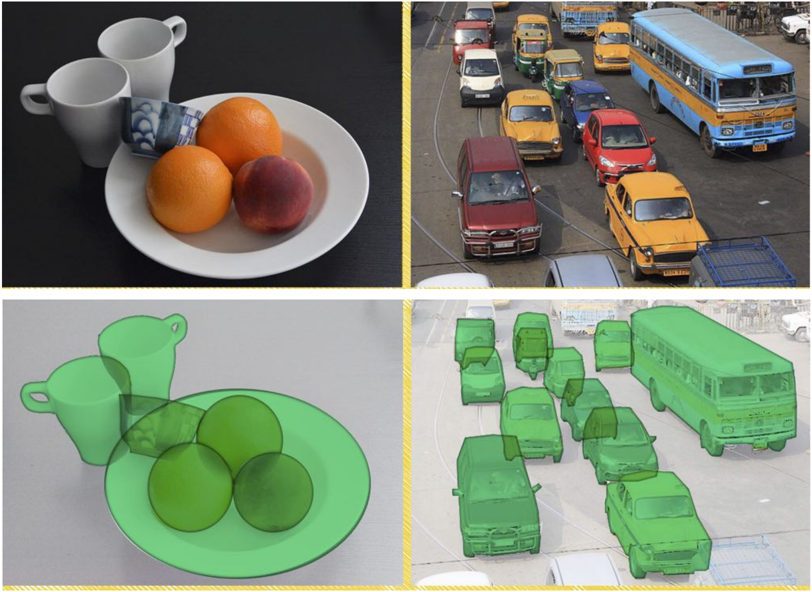
If $\boldsymbol{\Phi}$ is a good CS matrix, take $M \sim K \log N$.

Minimax lower bound:

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}_0\|_F \gtrsim \sigma \sqrt{K \log(N/K) + RK}$$

Convex shape composition

Image segmentation



Variational image segmentation

- Image domain $D \subset \mathbb{R}^d$, pixel values $u(x)$, $x \in D$
- **Binary segmentation**: partition D into Σ and $D \setminus \Sigma$

Variational image segmentation

- Image domain $D \subset \mathbb{R}^d$, pixel values $u(x)$, $x \in D$
- **Binary segmentation**: partition D into Σ and $D \setminus \Sigma$
- A standard variational model (Chan and Vese):

$$\Sigma^* = \arg\min_{\Sigma} \; \gamma(\Sigma) + \int_{\Sigma} \Pi_{in}(x) \, \mathsf{d}x + \int_{D \setminus \Sigma} \Pi_{ex}(x) \, \mathsf{d}x,$$

where $\Pi_{in}(.) \geq 0$ and $\Pi_{ex}(.) \geq 0$ are image-dependent inhomogeneity measures

$$\Pi_{in}(x) = (u(x) - \tilde{u}_{in})^2,$$

$$\Pi_{ex}(x) = (u(x) - \tilde{u}_{ex})^2$$

Variational image segmentation

- Image domain $D \subset \mathbb{R}^d$, pixel values $u(x)$, $x \in D$
- **Binary segmentation**: partition D into Σ and $D \setminus \Sigma$
- A standard variational model (Chan and Vese):

$$\Sigma^* = \arg\min_{\Sigma} \ \gamma(\Sigma) + \int_{\Sigma} \Pi_{in}(x)\, \mathsf{d}x + \int_{D \setminus \Sigma} \Pi_{ex}(x)\, \mathsf{d}x,$$

  where $\Pi_{in}(.) \geq 0$ and $\Pi_{ex}(.) \geq 0$ are image-dependent inhomogeneity measures

$$\Pi_{in}(x) = (u(x) - \tilde{u}_{in})^2,$$

$$\Pi_{ex}(x) = (u(x) - \tilde{u}_{ex})^2$$

- Other choices $\Pi_{in/ex}(x) = -\log \mathrm{P}\left(\upsilon(x) \middle|\, \theta_{in/ex}\right)$, where υ is a similarity feature of interest (intensity, texture, pattern)

Variational image segmentation

- Image domain $D \subset \mathbb{R}^d$, pixel values $u(x)$, $x \in D$
- **Binary segmentation**: partition D into Σ and $D \setminus \Sigma$
- A standard variational model (Chan and Vese):

$$\boxed{\Sigma^* = \arg\min_{\Sigma} \;\; \gamma(\Sigma) + \int_{\Sigma} \Pi_{in}(x) - \Pi_{ex}(x) \, \mathrm{d}x,}$$

where $\Pi_{in}(.) \geq 0$ and $\Pi_{ex}(.) \geq 0$ are image-dependent inhomogeneity measures

$$\Pi_{in}(x) = (u(x) - \tilde{u}_{in})^2,$$

$$\Pi_{ex}(x) = (u(x) - \tilde{u}_{ex})^2$$

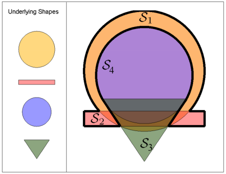
Dictionary of shapes

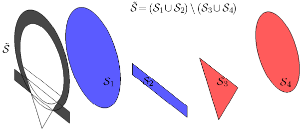
$$\Sigma^* = \arg\min_{\Sigma} \;\; \gamma(\Sigma) + \int_{\Sigma} \left(\Pi_{in}(x) - \Pi_{ex}(x)\right) \, \mathsf{d}x$$

Parameterize possible Σ using:

- Dictionary of shapes $\mathfrak{D} = \{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_{n_s}\}$
- Shape composition rule

$$\mathcal{R}_{\mathcal{I}_\oplus, \mathcal{I}_\ominus} \triangleq \Big(\bigcup_{j \in \mathcal{I}_\oplus} \mathcal{S}_j \Big) \setminus \Big(\bigcup_{j \in \mathcal{I}_\ominus} \mathcal{S}_j \Big),$$





Shape composition problem

$$\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \triangleq \Big(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j\Big)\Big\backslash\Big(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j\Big),$$

Shape composition problem:

$$\text{(SC)} \qquad \{{\mathcal{I}_\oplus}^*, {\mathcal{I}_\ominus}^*\} = \arg\min_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \big(\Pi_{in}(x) - \Pi_{ex}(x)\big)\ \mathsf{d}x$$

Cardinal shape composition problem:
(restrict the number of shapes used)

$$\text{(Cardinal-SC)} \qquad \min_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \big(\Pi_{in}(x) - \Pi_{ex}(x)\big)\ \mathsf{d}x \qquad \text{s.t.}: \quad |\mathcal{I}_\oplus| + |\mathcal{I}_\ominus| \le s$$

Shape composition problem

$$\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \triangleq \Big(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j \Big) \Big\backslash \Big(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j \Big),$$

Shape composition problem:

$$\text{(SC)} \qquad \{\mathcal{I}_\oplus{}^*, \mathcal{I}_\ominus{}^*\} = \arg\min_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \big(\Pi_{in}(x) - \Pi_{ex}(x)\big)\, \mathsf{d}x$$

Cardinal shape composition problem:
(restrict the number of shapes used)

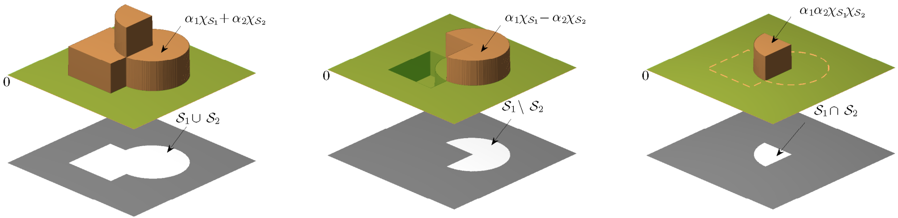
$$\text{(Cardinal-SC)} \qquad \min_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \big(\Pi_{in}(x) - \Pi_{ex}(x)\big)\, \mathsf{d}x \qquad \text{s.t. :} \quad |\mathcal{I}_\oplus| + |\mathcal{I}_\ominus| \leq s$$

Both of these are hard, combinatorial problems

Representing shapes

Basic idea: superimpose indicator functions, then cut between 0 and 1

$$\chi_{\mathcal{S}} = \text{indicator function for set } \mathcal{S}$$



$$\text{supp}^+(\alpha_1\chi_{\mathcal{S}_1} + \alpha_2\chi_{\mathcal{S}_2}) = \mathcal{S}_1 \cup \mathcal{S}_2 \qquad (\alpha_1 > 0,\ \alpha_2 > 0)$$

$$\text{supp}^+(\alpha_1\chi_{\mathcal{S}_1} - \alpha_2\chi_{\mathcal{S}_2}) = \mathcal{S}_1 \setminus \mathcal{S}_2 \qquad (\alpha_2 > \alpha_1 > 0)$$

$$\text{supp}^+(\alpha_1\chi_{\mathcal{S}_1}\alpha_2\chi_{\mathcal{S}_2}) = \mathcal{S}_1 \cap \mathcal{S}_2 \qquad (\alpha_1\alpha_2 > 0)$$

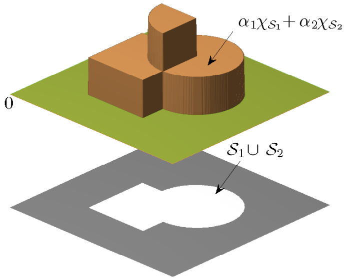
Shape composition → atomic decomposition

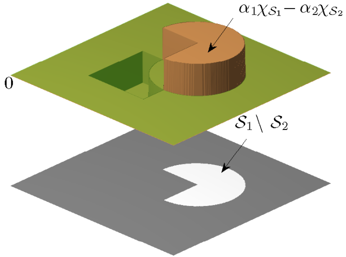
Given a collection of shapes $\{\mathcal{S}_j\}_{j=1}^{n_s}$, for any

$$\Sigma = \left(\bigcup_{j \in \mathcal{I}_\oplus} \mathcal{S}_j \right) \setminus \left(\bigcup_{j \in \mathcal{I}_\ominus} \mathcal{S}_j \right)$$

there exist scalars $\{\alpha_j\}$ such that

$$f_\alpha(x) = \sum_{j \in \mathcal{I}_\oplus \cup \mathcal{I}_\ominus} \alpha_j \cdot \chi_{\mathcal{S}_j}(x) \quad \text{satisfies} \quad \begin{cases} f_\alpha(x) \geq 1, & x \in \Sigma \\ f_\alpha(x) \leq 0, & x \notin \Sigma \end{cases}$$





Shape optimization

Then

$$\underset{\mathcal{I}_\oplus,\mathcal{I}_\ominus}{\text{minimize}} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \Pi_{in}(x) - \Pi_{ex}(x)\,\mathrm{d}x, \quad \mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} = \Big(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j\Big)\setminus\Big(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j\Big)$$

becomes

$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \int_D \big(\Pi_{in}(x)-\Pi_{ex}(x)\big)\cdot\mathrm{hev}(f_\alpha(x))\,\mathrm{d}x, \quad \text{s.t.} \quad f_\alpha(x) = \sum_{j=1}^{n_s} \alpha_j\,\chi_j(x),$$

where

$$\mathrm{hev}(u) = \begin{cases} 1, & u > 0 \\ 0, & u \le 0. \end{cases}$$

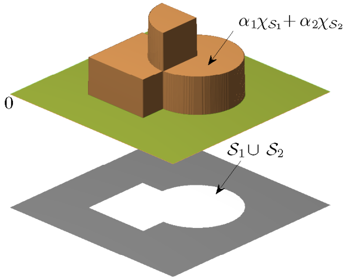
Shape optimization

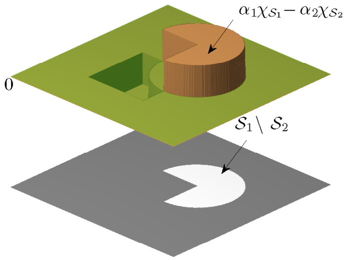
Then

$$\underset{\mathcal{I}_\oplus,\mathcal{I}_\ominus}{\text{minimize}} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} \Pi_{in}(x) - \Pi_{ex}(x)\,\mathrm{d}x, \quad \mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} = \Big(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j\Big)\setminus\Big(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j\Big)$$

becomes

$$\boxed{\underset{\boldsymbol{\alpha}}{\text{minimize}} \int_D g(x)\cdot \operatorname{hev}(f_\alpha(x))\,\mathrm{d}x, \quad \text{s.t.} \quad f_\alpha(x) = \sum_{j=1}^{n_s} \alpha_j\,\chi_j(x),}$$





Convexification

There is a natural convex proxy for the functional:

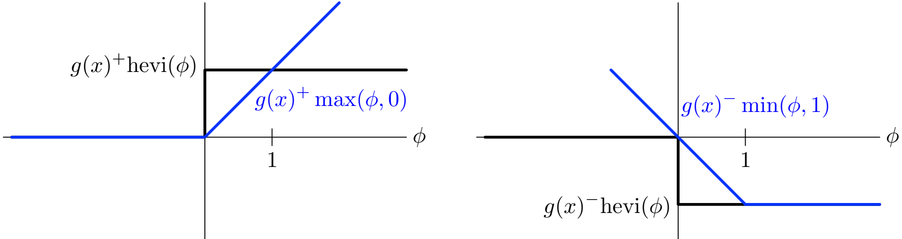
$$\int_D g(x) \cdot \mathrm{hev}(f_\alpha(x))\ \mathsf{d}x$$

$$\Downarrow$$

$$\int_D g(x)^+ \mathrm{hev}(f_\alpha(x)) \ + \ g(x)^- \mathrm{hev}(f_\alpha(x))\ \mathsf{d}x$$

$\Downarrow$ (relax)

$$\int_D g(x)^+ \max(f_\alpha(x), 0) \ + \ g(x)^- \min(f_\alpha(x), 1)\ \mathsf{d}x$$



Convexification

There is a natural convex proxy for the functional:

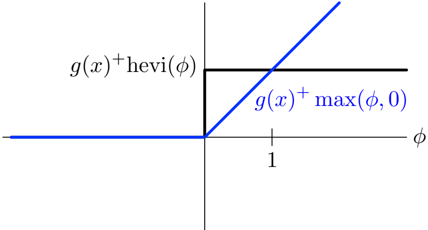
$$\int_D g(x) \cdot \mathrm{hev}(f_\alpha(x))\,\mathsf{d}x$$

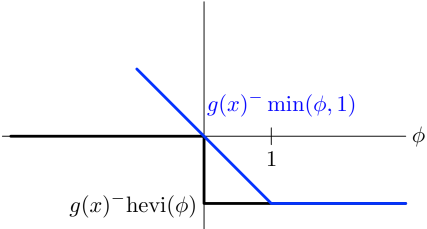
$$\Downarrow$$

$$\int_D g(x)^+ \mathrm{hev}(f_\alpha(x)) \;+\; g(x)^- \mathrm{hev}(f_\alpha(x))\,\mathsf{d}x$$

$\Downarrow$ (relax)

$$\boxed{\int_D \max\left(g(x) f_\alpha(x),\; g(x)^-\right)\;\mathsf{d}x}$$





Convex shape composition

With the sparse regularization,

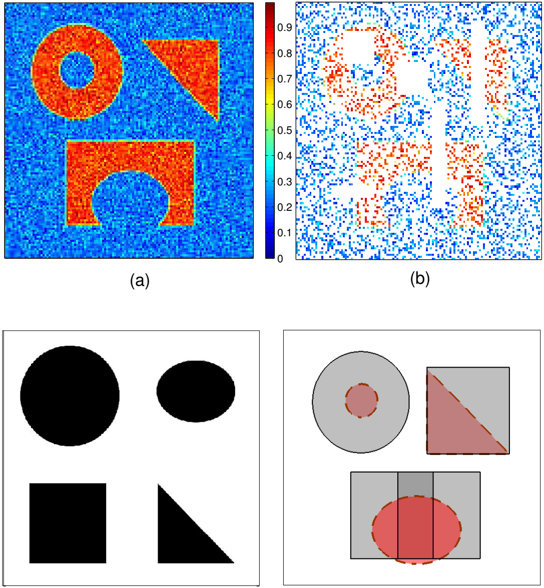
$$\underset{\mathcal{I}_\oplus,\mathcal{I}_\ominus}{\text{minimize}} \int_{\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus}} g(x)\,\mathrm{d}x \quad \text{subject to} \quad \begin{aligned} &\mathcal{R}_{\mathcal{I}_\oplus,\mathcal{I}_\ominus} = \Big(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j\Big)\setminus\Big(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j\Big) \\ &|\mathcal{I}_\oplus| + |\mathcal{I}_\ominus| \le S \end{aligned}$$

becomes

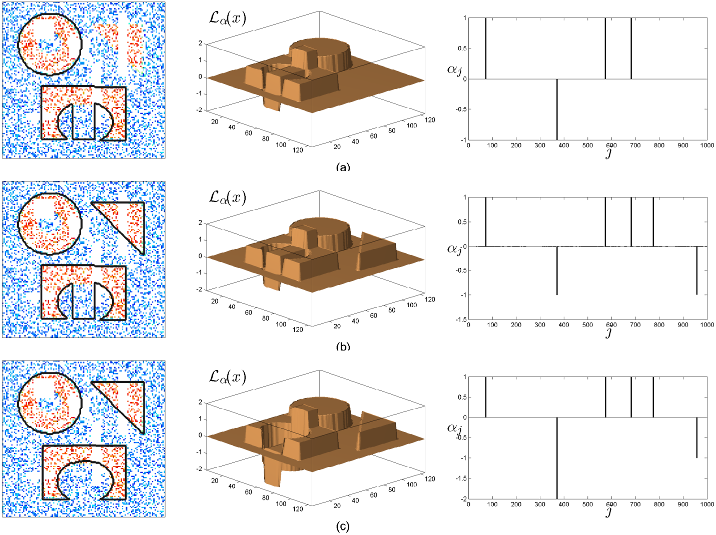
$$\underset{\boldsymbol{\alpha}}{\text{minimize}} \int_D \max\big(g(x) f_\alpha(x),\ g(x)^-\big)\ \mathrm{d}x \quad \text{subject to} \quad \begin{aligned} &f_\alpha(x) = \sum_{j=1}^{n_s} \alpha_j\, \chi_j(x) \\ &\|\boldsymbol{\alpha}\|_1 \le \tau \end{aligned}$$

τ is often an integer, related to the number of shapes used

Simulations

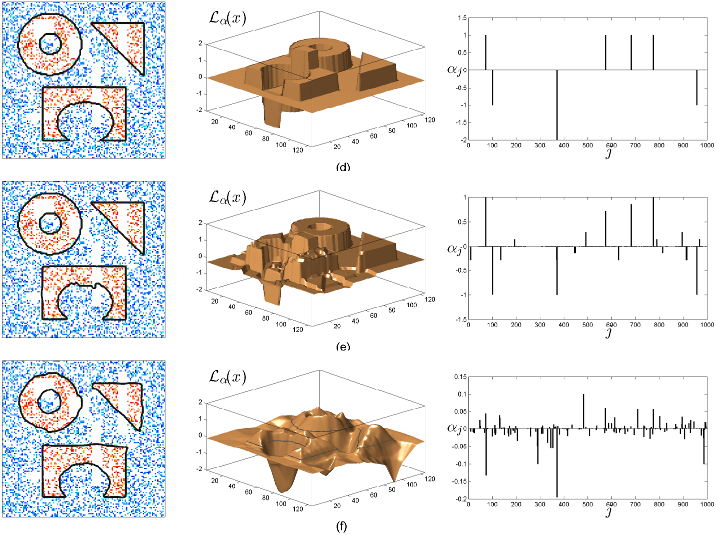


Simulations



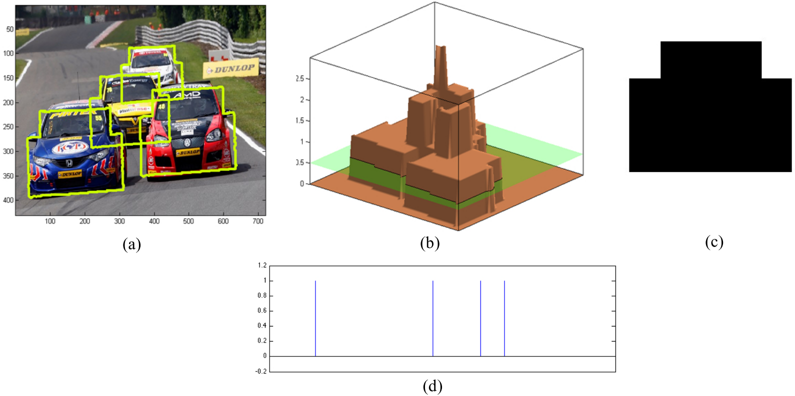
(a) $\tau = 4$; (b) $\tau = 6$; (c) $\tau = 7$;

Simulations



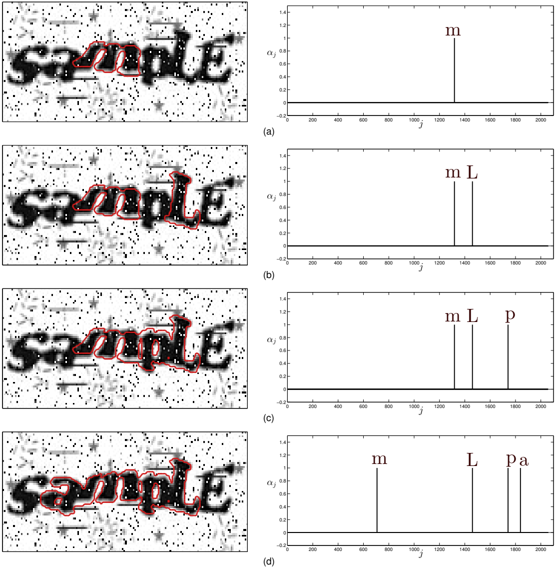
(d) $\tau = 8$; (e) $\tau = 10$; (f) previous method

Simulations



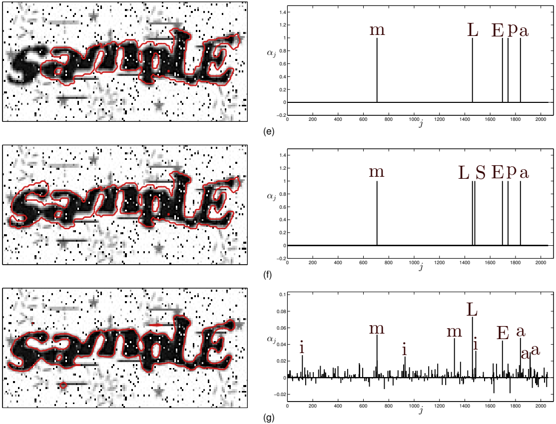
Identifying the objects inside an image

Simulations



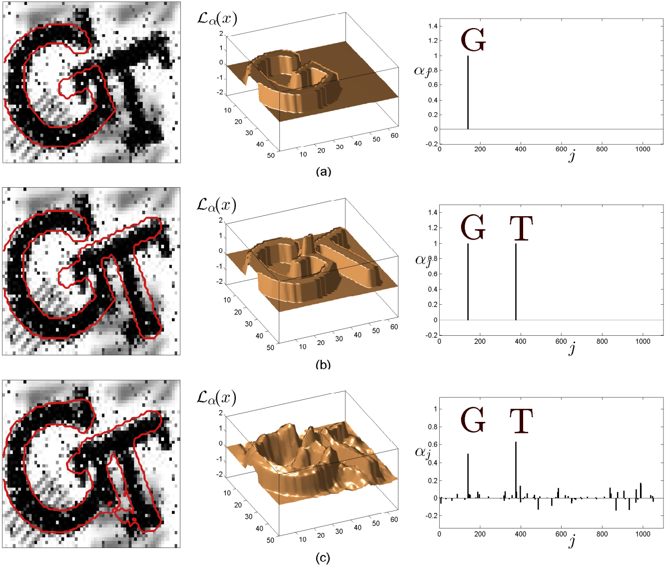
(a) $\tau = 1$; (b) $\tau = 2$; (c) $\tau = 3$; (d) $\tau = 4$

Simulations



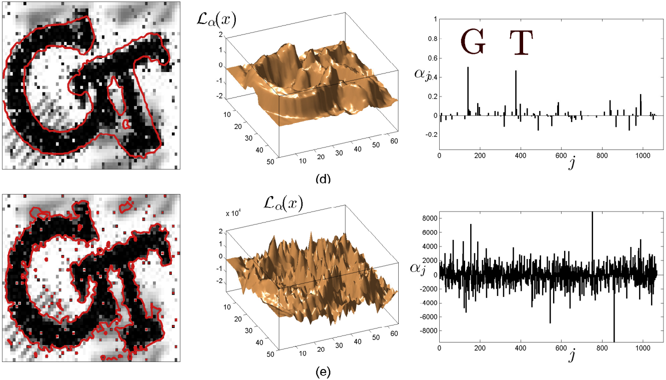
(e) $\tau = 5$; (f) $\tau = 6$; (g) previous technique

Simulations



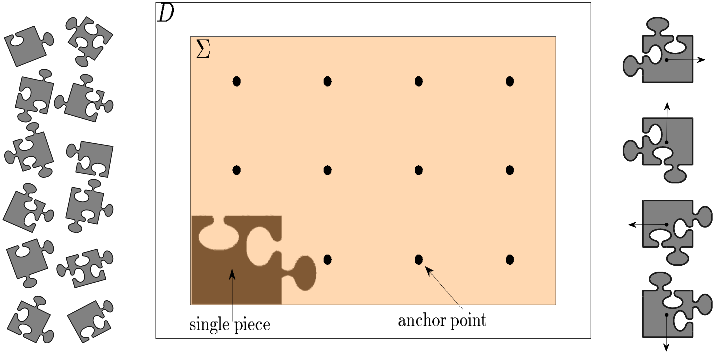
(a) $\tau = 1$; (b) $\tau = 2$; (c) $\tau = 4$; (d) $\tau = 5$; (e) $\tau = \infty$

Simulations

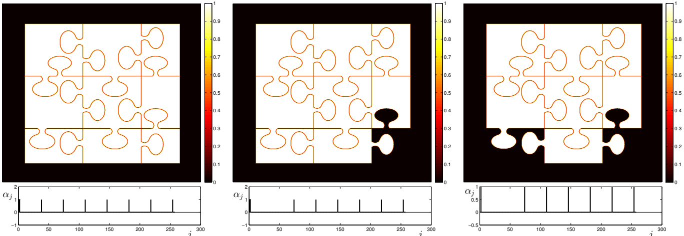


(a) $\tau = 1$; (b) $\tau = 2$; (c) $\tau = 4$; (d) $\tau = 5$; (e) $\tau = \infty$

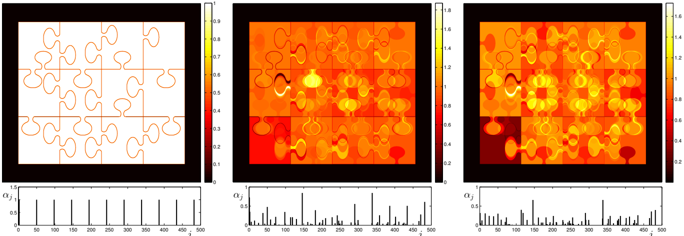
Simulations



Simulations

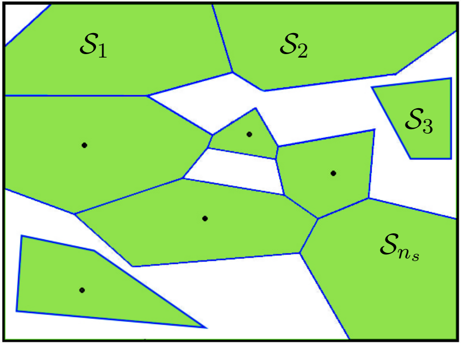


(a) (b) (c)



(a) (b) (c)

Theory: Disjoint dictionary elements



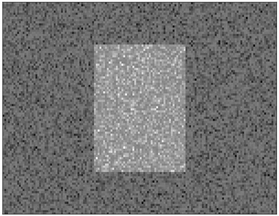
Disjoint dictionary elements $\Rightarrow$ Cardinal-SC and convex proxy ($\tau = s$) produce same result

Theory: Lucid objects

- For a given region $\Sigma \subset D$, the lucid object condition (LOC) holds if

$$\begin{cases} \Pi_{in}(x) < \Pi_{ex}(x) & x \in \Sigma \\ \Pi_{in}(x) > \Pi_{ex}(x) & x \in D \setminus \Sigma \end{cases}.$$

- Example: $\begin{cases} 0 < u(x) < \frac{1}{2} & x \in \Sigma \\ \frac{1}{2} < u(x) < 1 & x \in D \setminus \Sigma \end{cases}$ where $\Pi_{in}(x) = (u(x) - 1/4)^2$, $\Pi_{ex}(x) = (u(x) - 3/4)^2$

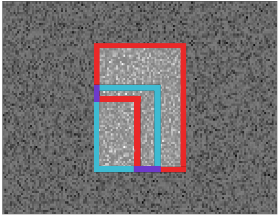


Theory: Lucid objects

If Σ is a "lucid object" and it can be composed of with s shapes $\mathcal{I}_{\oplus}^{\Sigma}, \mathcal{I}_{\ominus}^{\Sigma}$, then

- Cardinal-SC identifies $\mathcal{I}_{\oplus}^{\Sigma}$ and $\mathcal{I}_{\ominus}^{\Sigma}$ by selecting $s = |\mathcal{I}_{\oplus}^{\Sigma}| + |\mathcal{I}_{\ominus}^{\Sigma}|$
- Convex proxy produces $\boldsymbol{\alpha}^*$ such that

$$f_{\alpha^*}(x) \begin{cases} \geq 1, & x \in \Sigma \\ \leq 0, & x \notin \Sigma \end{cases}$$



Sketch of more general theory

- Image contains object $\Sigma = \left(\bigcup_{j\in\mathcal{I}_\oplus} \mathcal{S}_j\right) \setminus \left(\bigcup_{j\in\mathcal{I}_\ominus} \mathcal{S}_j\right)$
- We show that there is a bijection between non-redundant $\mathcal{I}_\oplus, \mathcal{I}_\ominus$ and coefficients $\boldsymbol{\alpha}$

$$\begin{tikzcd} \{\mathcal{I}_\oplus, \mathcal{I}_\ominus\} \arrow[r, bend left, "\mathcal{A}"] & \boldsymbol{\alpha}_\mathcal{R} \arrow[l, bend left, "\xi^+ \cup \xi^-"] \end{tikzcd}$$

- We develop sufficient conditions under which the output of the convex proxy is

$$\boldsymbol{\alpha}_{\mathcal{I}_\oplus \cup \mathcal{I}_\ominus} = \boldsymbol{\alpha}_\mathcal{R}, \qquad \boldsymbol{\alpha}^*_{(\mathcal{I}_\oplus \cup \mathcal{I}_\ominus)^c} = \mathbf{0}$$

  for appropriate τ
- Conditions mostly depend on how large energy $g(x)$ is in regions where non-included shapes overlap Σ

References

- A. Ahmed, B. Recht, and J. Romberg, "Blind Deconvolution using Convex Programming," *IEEE Transactions on Information Theory*, March 2014.
- S. Bahmani and J. Romberg, "Lifting for Blind Deconvolution in Random Mask Imaging: Identifiability and Convex Relaxation," submitted 2014, arXiv:1501.00046.
- S. Bahmani and J. Romberg, "Near-Optimal Estimation of Simultaneously Sparse and Low-Rank Matrices from Nested Linear Measurements," submitted 2015, arxiv.org:1506.08159.
- A. Aghasi and J. Romberg, "Sparse Shape Reconstruction," *SIAM J. Imaging Science*, December 2013.
- A. Aghasi and J. Romberg, "Convex Cardinal Shape Decomposition," submitted 2015, arXiv:1501.01347.