A dual approach for PhaseLift

Michael Friedlander
mpf@cs.ubc.ca

Ives Macêdo
ijamj@cs.ubc.ca

University of British Columbia
Department of Computer Science
Scientific Computing Laboratory

SINBAD Consortium Meeting Fall 2013
Whistler, December 3, 2013

Linear inverse problems

$$F : \mathbb{C}^N \longrightarrow \mathbb{C}^D \qquad \text{(linear forward/modeling operator)}$$

$$m \longmapsto d = Fm \qquad \text{(observed data)}$$

- Reverse Time Migration (RTM)

$$m_{RTM} := F^* d$$

- Least-Squares Migration (LSM)

$$m_{LSM} := \arg\min_{m \in \mathbb{C}^N} \frac{1}{2} \| Fm - d \|_2^2 \quad \left[= \left(F^* F \right)^{-1} F^* d \right]$$

where F^* is the adjoint operator

A case for quadratic measurements

- **Borcea, Papanicolaou, Tsogka (Inverse Problems 2005)**
  Coherent INTerferometric migration: $m_{INT} := \text{diag}(F^*(E \circ dd^*)F)$

- **Chai, Moscoso, Papanicolaou (Inverse Problems 2011)**
  Candès, Eldar, Strohmer, Voroninski (SIIMS 2013)
  Phaseless measurements: $\text{diag}(|d_1|^2, |d_2|^2, \ldots, |d_M|^2) \approx \text{diag}((Fm)\,(Fm)^*)$

- **Jugnon, Demanet (SEG 2013)**
  Interferometric inversion: find $m \in \mathbb{C}^N$ such that $d_i \overline{d_j} \approx (Fm)_i \overline{(Fm)_j}$

Matrix lifting

or, making it linear again!

find $\quad m \in \mathbb{C}^N$

such that

$(F(mm^*)F^*)_{i_k j_k} \approx (dd^*)_{i_k j_k}$

$k = 1, \ldots, K$

$$\xLeftrightarrow{M = mm^*}$$

find $\quad M \in \mathbb{C}^{N \times N}$

such that

$(FMF^*)_{i_k j_k} \approx (dd^*)_{i_k j_k}$

$k = 1, \ldots, K$

$M = M^*, M \succcurlyeq 0$

$\operatorname{rank} M = 1$

Matrix lifting

or, making it linear again!

$$\begin{aligned} &\text{find} \quad M \in \mathbb{C}^{N\times N} \\ &\text{such that} \\ &\quad (FMF^*)_{i_k j_k} \approx (dd^*)_{i_k j_k} \\ &\quad k = 1, \ldots, K \\ &\quad M = M^*, M \succcurlyeq 0 \\ &\quad \operatorname{rank} M = 1 \end{aligned}$$

$$\xRightarrow{\cancel{\operatorname{rank} M = 1}}$$

$$\begin{aligned} &\underset{M \in \mathcal{H}^{N\times N}}{\text{minimize}} \quad \operatorname{trace} M \\ &\text{subject to} \\ &\quad (FMF^*)_{i_k j_k} \approx (dd^*)_{i_k j_k} \\ &\quad k = 1, \ldots, K \\ &\quad M \succcurlyeq 0 \end{aligned}$$

Matrix lifting

or, making it linear again!

$$\begin{aligned}&\underset{M\in\mathcal{H}^{N\times N}}{\text{minimize}} \quad \operatorname{trace} M\\ &\text{subject to}\\ &\quad (FMF^*)_{i_k j_k} \approx (dd^*)_{i_k j_k}\\ &\quad k = 1,\dots,K\\ &\quad M \succcurlyeq 0\end{aligned} \quad \overset{\text{notation}}{\Longleftrightarrow} \quad \begin{aligned}&\underset{M\in\mathcal{H}^{N\times N}}{\text{minimize}} \quad \operatorname{trace} M\\ &\text{subject to}\\ &\quad \mathcal{F}M \approx b\\ &\quad M \succcurlyeq 0\end{aligned}$$

Matrix lifting

or, making it linear again!

$$\underset{M \in \mathcal{H}^{N\times N}}{\text{minimize}} \quad \operatorname{trace} M \quad \text{subject to} \quad \mathcal{F}M \approx b \text{ and } M \succcurlyeq 0$$

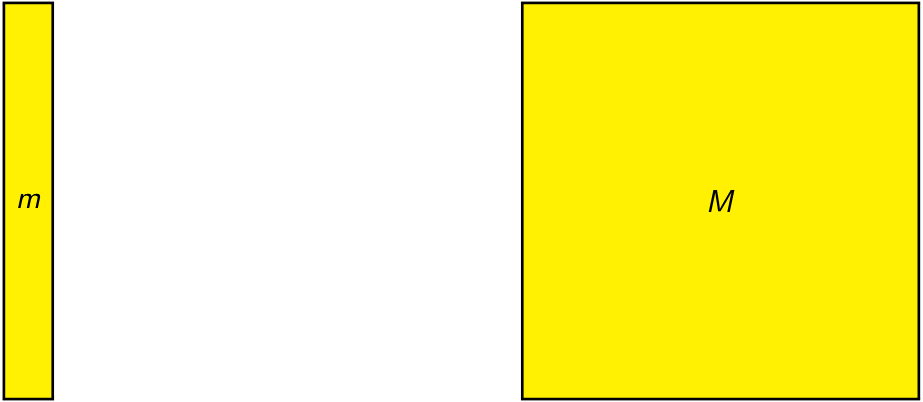
$$\begin{aligned} \mathcal{F} : \mathcal{H}^{N\times N} &\longrightarrow \mathbb{C}^K \\ M &\longmapsto (FMF^*)_{i_k j_k} \end{aligned} \qquad \text{linear operator!}$$

$$\mathcal{F}^* : \mathbb{C}^K \longrightarrow \mathcal{H}^{N\times N} \qquad \text{adjoint map}$$

Challenges

or, the cost of linearity. . .

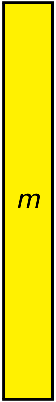
nonlinear program in $\mathbb{C}^N$ **vs.** linear-conic program in $\mathcal{H}^{N \times N}$

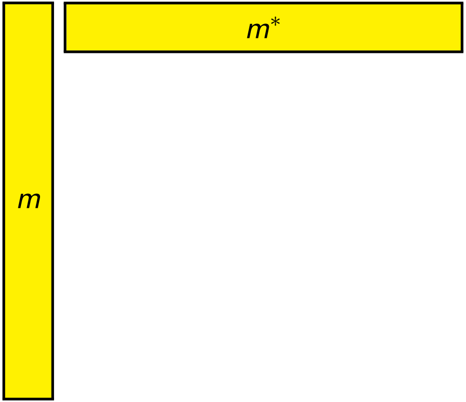


Challenges

or, the cost of linearity. . .

nonlinear program in $\mathbb{C}^N$ **vs.** linear-conic program in $\mathcal{H}^{N\times N}$





Challenges

or, the cost of linearity. . .

The convexified formulation in (5) is too costly to solve at the scale of even toy problems. Let N be the total number of degrees of freedom of your unknown model m; then the variable M of (5) is a $N \times N$ matrix, on which we want to impose positive semi-definiteness and approximate fit. As of 2013 and to our knowledge, there is no time-efficient and memory-efficient algorithm to solve this type of semi-definite program when N ranges from 10^4 to 10^6. **Jugnon, Demanet (SEG 2013)**

A measurement-centric perspective

or, what if K is (roughly) on the order of N?

Primal:

$$\underset{M\in\mathcal{H}^{N\times N}}{\text{minimize}} \quad \operatorname{trace} M \quad \text{subject to} \quad \|b-\mathcal{F}M\|\leqslant\sigma \text{ and } M\succcurlyeq 0$$

Lagrangian dual:

$$\underset{y\in\mathbb{C}^K}{\text{maximize}} \quad \Re\langle y,b\rangle-\sigma\|y\|_* \quad \text{subject to} \quad \lambda_{max}(\mathcal{F}^*y)\leqslant 1$$

Gauge dual: **[Freund (Math. Prog. 1987); Friedlander, M., Pong (arXiv 2013)]**

$$\underset{y\in\mathbb{C}^K}{\text{minimize}} \quad \lambda_{max}(\mathcal{F}^*y) \quad \text{subject to} \quad \Re\langle y,b\rangle-\sigma\|y\|_*\geqslant 1$$

Primal from dual

or, how does this solve my problem?

If $Q_{max} \in \mathbb{C}^{N \times r_{max}}$ matrix of λ_{max}-eigvecs of $\mathcal{F}^* y_{opt}$, then $M_{opt} = Q_{max} S_{opt} Q_{max}^*$,

$$S_{opt} \in \operatorname*{arg\,min}_{\substack{S \in \mathcal{H}^{r_{max} \times r_{max}} \\ S \succcurlyeq 0}} \left\| c - \mathcal{F}_{reduced} S \right\|,$$

where $\mathcal{F}_{reduced} := \mathcal{F}(Q_{max} \cdot Q_{max}^*)$ and $c := b - \sigma \frac{y_{opt}}{\|y_{opt}\|_*}$

Typically, $\quad r_{max} \lll N \ll N^2$!

Bundle methods in a nutshell

or, solid and customizable hammers for convex nails...

Gauge dual:

$$\min_{\Re\langle y,b\rangle - \sigma\|y\|_* \geqslant 1} \left\{ \lambda_{max}(\mathcal{F}^* y) \right\}$$

Proximal point method:

$$y^{k+1} := \operatorname*{arg\,min}_{\Re\langle y,b\rangle - \sigma\|y\|_* \geqslant 1} \left\{ \lambda_{max}(\mathcal{F}^* y) + \frac{1}{2\tau} \left\| y - y^k \right\|_H^2 \right\}$$

Proximal bundle method:

$$\hat{y}^{k+1} := \operatorname*{arg\,min}_{\Re\langle y,b\rangle - \sigma\|y\|_* \geqslant 1} \left\{ \max_j \left\{ \lambda_{max}\left(Q_j^* (\mathcal{F}^* y) Q_j \right) \right\} + \frac{1}{2\tau_k} \left\| y - y^k \right\|_{H_k}^2 \right\}$$

Bundle methods in a nutshell

or, solid and customizable hammers for convex nails...

The good:

- Provably convergent
- Bundle $\left\{ Q_j \in \mathbb{C}^{N \times r_j} \right\}_{j \in J}$ is very flexible w.r.t. both J and r_j
- Huge eigen-matrix-inequality substituted by a handful of tiny ones
- Admits Lanczos-based large-scale solvers, i.e. just need to code $(\mathcal{F}^* y)x$
- Known to compare well against other methods for nonsmooth convex programs

Bundle methods in a nutshell

or, solid and customizable hammers for convex nails...

The bad:

- Worst-case theoretical bounds far from theoretical optimal
- Uniformly "too small" J and r_j may impact speed

Bundle methods in a nutshell

or, solid and customizable hammers for convex nails...

The ugly:

- May require custom-tailored subproblem solvers for “production code”
- Require storage and “clever” management of the bundle

Case study: PhaseLift

X-ray crystallography

Candès, Eldar, Strohmer, Voroninski (SIIMS 2013)

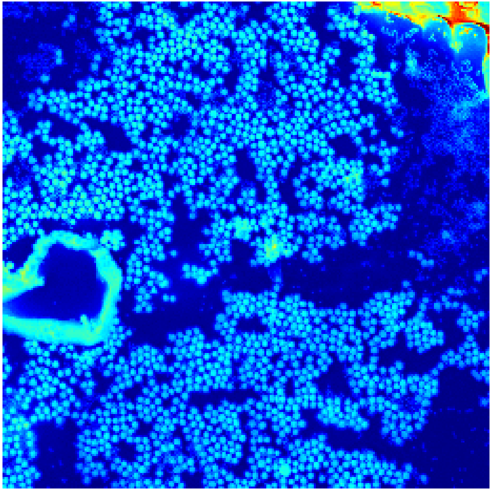
Phaseless measurements

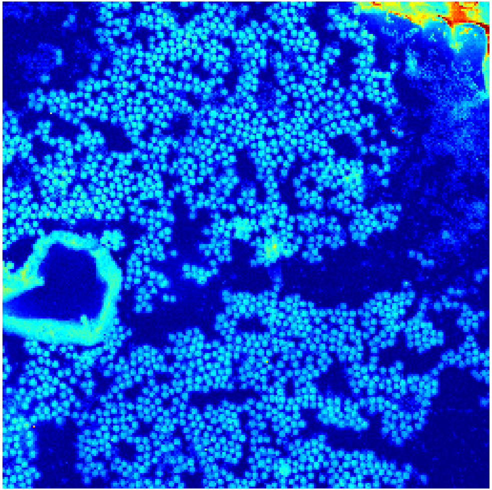
- $m \in \mathbb{R}^{n \times n}$ is a "latent" image
- we are provided with illumination patterns $f_p \in \mathbb{C}^{n \times n}$, $i = 1, \ldots, P$
- observables are magnitudes of Fourier coefficients $\left|\widehat{f_p \circ m}\right|^2$
- $N = n^2$ and $K = PN$

- Our (noisefree) experiments: $n = 256$, $N = 65536$, $P = 8$ (complex Gaussians) and $K = 524288$

Case study: PhaseLift

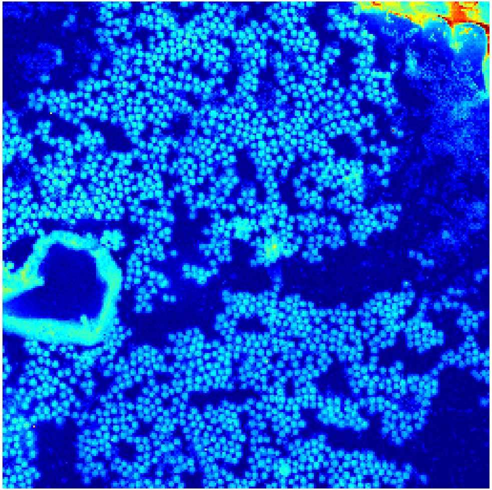
Ground-truth vs. our reconstruction

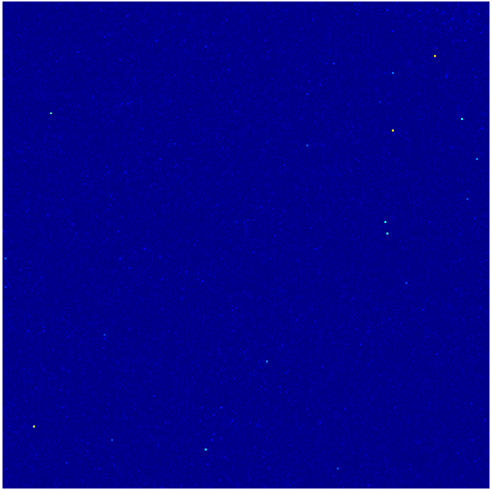




Case study: PhaseLift

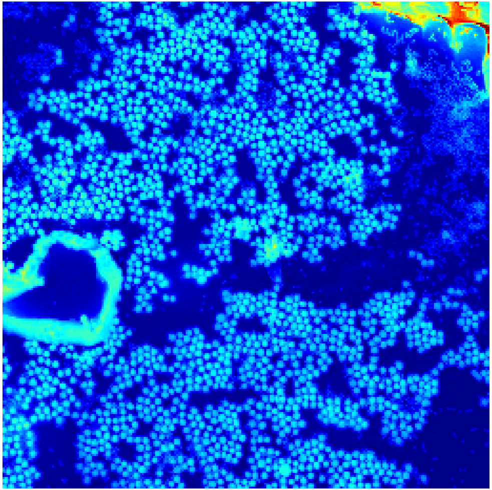
Our reconstruction vs. error

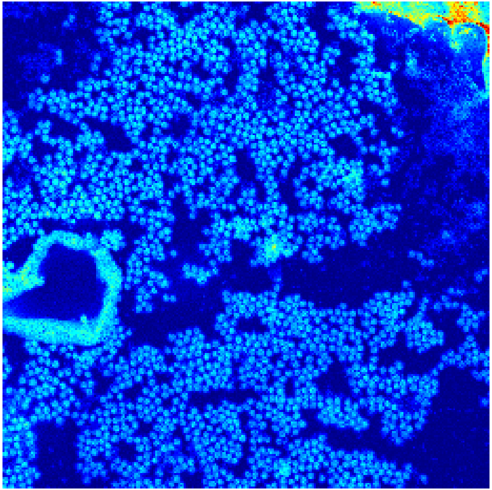




Case study: PhaseLift

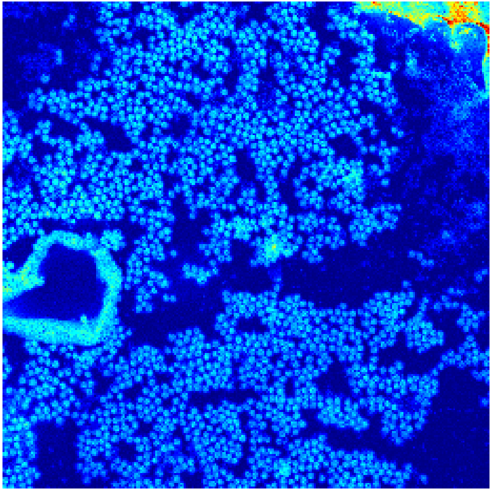
Ground-truth vs. reconstruction from TFOCS

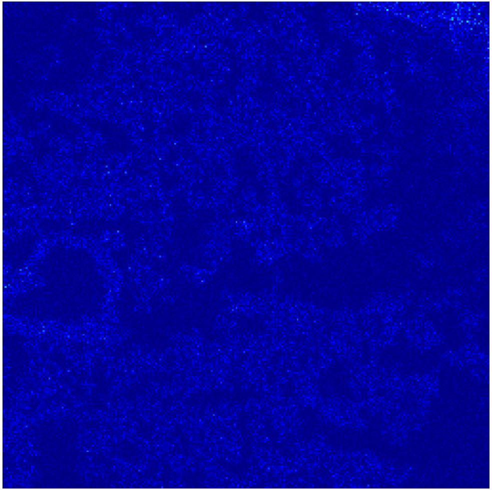




Case study: PhaseLift

Reconstruction from TFOCS vs. error





Case study: PhaseLift
TFOCS vs. our solver

	TFOCS	TFOCS	Ours
# $(\mathcal{F}^* y)x$	7371	9941	7366
max # eig-pairs	2	10	2
total # eig-pairs	454	700	44
time	706s	486s	478s
$\|\hat{m} - m\|_2 / \|m\|_2$	$17.0 \cdot 10^{-2}$	$17.8 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$
$\|\hat{m} - m\|_1 / \|m\|_1$	$15.4 \cdot 10^{-2}$	$16.2 \cdot 10^{-2}$	$6.5 \cdot 10^{-2}$
$\|b - \mathcal{F}\hat{M}\|_2 / \|b\|_2$	$17.6 \cdot 10^{-2}$	$19.7 \cdot 10^{-2}$	$9.9 \cdot 10^{-2}$
$\|b - \mathcal{F}\hat{M}\|_1 / \|b\|_1$	$19.7 \cdot 10^{-2}$	$20.9 \cdot 10^{-2}$	$9.1 \cdot 10^{-2}$

The road ahead

- Inexact bundle for adaptive tolerances in eigensolver (i.e., less operator evals)
- Exploit primal-from-dual result and nonconvex heuristics for better bundles
- Approximate 2nd-order info to accelerate bundle convergence *à la* **Helmberg, Overton, Rendl (Opt.-Online 2012)**
- Exploit different problems and their structures to specialize into better methods
- Experiment and compare more...

Thank you!

☺

A dual approach for PhaseLift

Michael Friedlander
mpf@cs.ubc.ca

Ives Macêdo
ijamj@cs.ubc.ca

University of British Columbia
Department of Computer Science
Scientific Computing Laboratory

SINBAD Consortium Meeting Fall 2013
Whistler, December 3, 2013