A GLOBALLY LINEARLY CONVERGENT METHOD FOR LARGE-SCALE POINTWISE QUADRATICALLY SUPPORTABLE CONVEX-CONCAVE SADDLE POINT PROBLEMS

Russell Luke
(Timo Aspelmeier, Charitha, Ron Shefi)

Universität Göttingen

LCCC Workshop, Large-Scale and Distributed Optimization, June 14-16, 2017, Lunds University

Outline

STimulated Emission Depletion

780 OPTICS LETTERS / Vol. 19, No. 11 / June 1, 1994

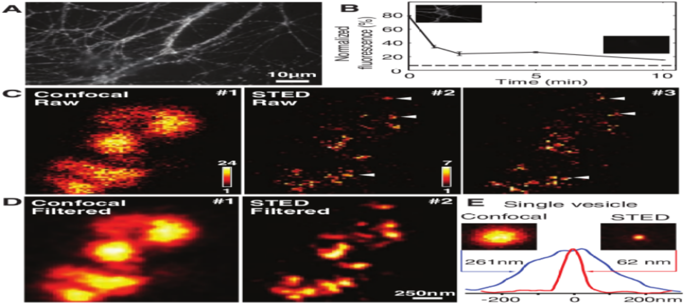
Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy

Stefan W. Hell and Jan Wichmann

Department of Medical Physics, University of Turku, Tykistökatu 6, 20521 Turku, Finland

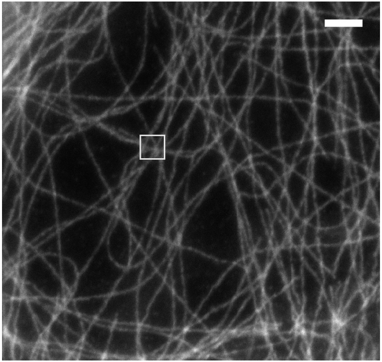
Received March 7, 1994

We propose a new type of scanning fluorescence microscope capable of resolving 35 nm in the far field. We overcome the diffraction resolution limit by employing stimulated emission to inhibit the fluorescence process in the outer regions of the excitation point-spread function. In contrast to near-field scanning optical microscopy, this method can produce three-dimensional images of translucent specimens.

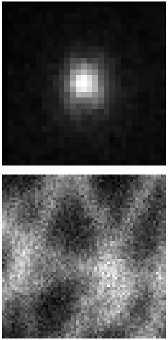


Science, 2008

STimulated Emission Depletion



$\approx 3nm$ per pixel



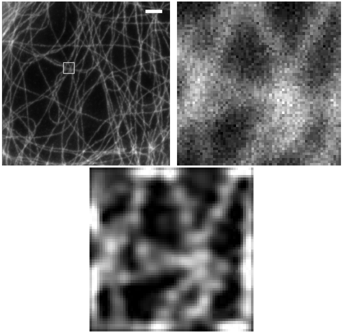
Statistical Image Denoising/Deconvolution

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g_\epsilon(Ax) \leq 0 \end{array}$$

where f is convex, piecewise linear-quadratic, $A : \mathbb{R}^n \to \mathbb{R}^n$, and

$$g_\epsilon : \mathbb{R}^n \to m = 2^{\mathbb{R}^n} := v \mapsto (g_1(v) - \epsilon_1, g_2(v) - \epsilon_2, \ldots, g_m(v) - \epsilon_m)^T$$

is convex and smooth



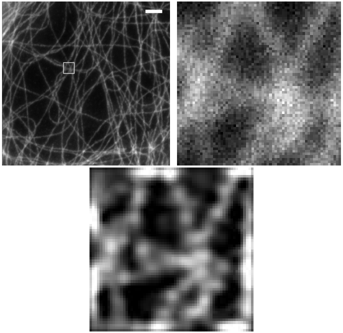
Statistical Image Denoising/Deconvolution

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g_\epsilon(Ax) \leq 0 \end{array}$$

where f is convex, piecewise linear-quadratic, $A : \mathbb{R}^n \to \mathbb{R}^n$, and

$$g_\epsilon : \mathbb{R}^n \to m = 2^{\mathbb{R}^n} := v \mapsto (g_1(v) - \epsilon_1, g_2(v) - \epsilon_2, \ldots, g_m(v) - \epsilon_m)^T$$

is convex and smooth



What is the scientific content of processed images?

Goals

Solve

$$0 \in F(x)$$

for $F : \mathbb{E} \rightrightarrows \mathbb{E}$ with $\mathbb{E}$ a Euclidean space.

- #1. Convergence (with a posteriori error bounds) of Picard iterations:

$$x^{k+1} \in Tx^k \quad \text{where} \quad \operatorname{Fix} T \approx \operatorname{zer} F$$

- #2. Algorithms:
  - (Non)convex Optimization: ADMM/Douglas-Rachford
  - Saddle-point Problems: Proximal Alternating Predictor-Corrector (PAPC)
- #3. Applications:
  - Image denoising/deconvolution
  - Phase retrieval

Building blocks

- Resolvent: $(\mathrm{Id} + \lambda F)^{-1}$
- Prox operator: for a function $f : X \to \overline{\mathbb{R}}$, define

$$\mathrm{prox}_{M,f}(x) := \mathrm{argmin}_y \left\{ f(y) + \frac{1}{2}\|y - x\|_M^2 \right\}$$

- Proximal reflector: $R_{M,f} := 2\,\mathrm{prox}_{M,f} - \mathrm{Id}$
- Projector: if $f = \iota_\Omega$ for $\Omega \subset X$ closed and nonempty, then $\mathrm{prox}_{M,f}(\overline{x}) = P_\Omega \overline{x}$ where

$$\begin{aligned} P_\Omega x &:= \{\overline{x} \in \Omega \mid \|x - \overline{x}\| = \mathrm{dist}(x, \Omega)\} \\ \mathrm{dist}(x, \Omega) &:= \inf_{y \in \Omega} \|x - y\|_M. \end{aligned}$$

- Reflector: if $f = \iota_\Omega$ for some closed, nonempty set $\Omega \subset X$, then $R_\Omega := 2P_\Omega - \mathrm{Id}$

Optimization

$$p_* = \min_x \left\{ f(x) + \sum_i^I g_i(A_i^T x) =: f(x) + g(\mathcal{A}^T x) : x \in \mathbb{R}^n \right\}. \qquad (\mathcal{P})$$

Reformulations:

Augmented Lagrangian

$$\min_{x\in\mathbb{R}^n} \min_{v\in\mathbb{R}^m} f(x) + \langle x,\ \mathcal{A}b\rangle - \langle b,\ v\rangle + g(v) + \tfrac{1}{2}\|\mathcal{A}^T x - v\|_M^2 \qquad (\mathcal{L})$$

Saddle-point

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} \left\{ K(x,y) := f(x) + \left\langle \mathcal{A}^T x,\ y \right\rangle - g^*(y) \right\}. \qquad (\mathcal{M})$$

Algorithms

ADMM

Initialization. Choose $\eta > 0$ and (x^0, v^0, b^0).
General Step ($k = 0, 1, \ldots$)

$$\begin{aligned} x^{k+1} &\in \operatorname{argmin}_x \left\{ f(x) + \langle b^k, Ax \rangle + \tfrac{\eta}{2}\|Ax - v^k\|^2 \right\}; && (1a)\\ v^{k+1} &\in \operatorname{argmin}_v \left\{ g(v) - \langle b^k, v \rangle + \tfrac{\eta}{2}\|Ax^{k+1} - v\|^2 \right\}; && (1b)\\ b^{k+1} &= b^k + \eta(Ax^{k+1} - v^{k+1}). && (1c) \end{aligned}$$

In the convex setting, the points in ADMM can be computed from the corresponding points in

Douglas-Rachford

$$y^{k+1} \in Ty^k \quad (k \in \mathbb{N})$$

for $T := \frac{1}{2}(R_{\eta B}R_{\eta D} + \mathrm{Id}) = \mathcal{J}_{\eta B}(2\mathcal{J}_{\eta D} - \mathrm{Id}) + (\mathrm{Id} - \mathcal{J}_{\eta D})$,
where $B := \partial\left(f^* \circ (-\mathcal{A}^T)\right)$ and $D := \partial g^*$

Algorithms

Proximal Alternating Predictor-Corrector (PAPC) [Drori, Sabach&Teboulle, 2015]

Initialization: Let $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$, and choose the parameters τ and σ to satisfy

$$\tau \in \left(0, \frac{1}{L_f}\right), \quad 0 < \tau\sigma \leq \frac{1}{\|\mathcal{A}^T\mathcal{A}\|}.$$

Main Iteration: for $k = 1, 2, \ldots$ update x^k, y^k as follows:

$$\begin{aligned}
&p^k = x^{k-1} - \tau(\nabla f(x^{k-1}) + \mathcal{A}y^{k-1});\\
&\text{for } i = 1, \ldots, l,\\
&\qquad y_i^k = \text{prox}_{\sigma, g_i^*}\left(y_i^{k-1} + \sigma A_i^T p^k\right);\\
&x^k = x^{k-1} - \tau(\nabla f(x^{k-1}) + \mathcal{A}y^k).
\end{aligned}$$

Outline

Key abstract properties

Almost firm nonexpansiveness

$T : \mathbb{E} \rightrightarrows \mathbb{E}$ is pointwise almost firmly nonsexpansive at y when

$$\left\|x^+ - y^+\right\|^2 \leq \frac{\varepsilon}{2}\|x - y\|^2 + \langle x^+ - y^+, x - y\rangle$$

for all $x^+ \in Tx$, and all $y^+ \in Ty$ whenever $x \in U$.

Metric subregularity (Ioffe, Aze, Dontchev&Rockafellar)

$\Phi : \mathbb{E} \rightrightarrows \mathbb{Y}$ is metrically regular on $U \times V \subset \mathbb{E} \times \mathbb{Y}$ relative to $\Lambda \subset \mathbb{E}$ if $\exists$ a $\kappa > 0$ such that

$$\operatorname{dist}\left(x, \Phi^{-1}(y) \cap \Lambda\right) \leq \kappa \operatorname{dist}(y, \Phi(x)) \tag{2}$$

holds for all $x \in U \cap \Lambda$ and $y \in V$. When the set V consists of a single point, $V = \{\overline{y}\}$, then Φ is said to be metrically subregular for $\overline{y}$ on U relative to $\Lambda \subset \mathbb{E}$.

Abstract results

Linear convergence [L. Nguyen& Tam, 2017]

Let $g = \iota_\Omega$ for $\Omega \subset \mathbb{R}^n$ semi-algebraic and let $f : \mathbb{R}^n \to \mathbb{R}$ be linear-quadratic convex. Let $(x^k)_{k\in\mathbb{N}}$ be iterates of the Douglas–Rachford algorithm and let $\Lambda = \text{aff}\,(x^k)$. If $T_{DR} - \text{Id}$ is metrically subregular at all points $\overline{x} \in \text{Fix}\ T_{DR} \cap \Lambda \neq \emptyset$ relative to Λ then for all x^0 close enough to $\text{Fix}\ T_{DR} \cap \Lambda$, the sequence x^k converges linearly to a point in $\text{Fix}\ T \cap \Lambda$ with constant at most $c = \sqrt{1 + \varepsilon - 1/\kappa^2} < 1$ where κ is the constant of metric subregularity for $T_{\text{DR}} - \text{Id}$ on some neighborhood U containing the sequence and ε is the violation of almost firm nonexpansiveness on the neighborhood U.

Polyhedrality $\implies$ metric subregularity

If T is polyhedral and $\mathrm{Fix}\ T \cap \Lambda$ consists of isolated points, then $\mathrm{Id} - T$ is metrically subregular at $\bar{x}$ relative to Λ.

Application: ADMM/Douglas-Rachford

Linear convergence of polyhedral DR/ADMM [Aspelmeier, Charitha, L., 2016]

Let $f : U \to \mathbb{R} \cup \{+\infty\}$ and $g : V \to \mathbb{R}$ be proper, lsc, convex, piecewise linear-quadratic functions and T the corresponding Douglas-Rachford fixed point mapping. Suppose that, for some affine subspace W, Fix $T \cap W$ is an isolated point $\{\overline{y}\}$. Then the Douglas-Rachford sequence $(y^k)_{k\in\mathbb{N}}$ converges linearly to $\overline{y}$ with rate bounded above by $\sqrt{1-\kappa^{-2}}$, where $\kappa > 0$ is a constant of metric subregularity of $\mathrm{Id} - T$ at $\overline{y}$ for the neighborhood $\mathcal{O}$. Moreover, the sequence $\left(b^k, v^k\right)_{k\in\mathbb{N}}$ generated by the ADMM Algorithm converges linearly to $\left(\overline{b}, \overline{v}\right)$ and the primal ADMM sequence $\left(x^k\right)_{k\in\mathbb{N}}$ converges to a solution to $\mathcal{P}$.

Remark

Compare to

Linear convergence with strong monotonicity

Let f and g be proper, lsc and convex. Suppose there exists a solution to $0 \in \left(\partial\left(f^* \circ (-\mathcal{A}^T)\right) + \partial g^*\right)(x)$ where $\mathcal{A}$ is an injective linear mappinig. Suppose further that, on some neighborhood of $\overline{y}$ g is strongly convex with constant μ and ∂g is β-inverse strongly monotone for some $\beta > 0$. Then any DR sequence initiated on this neighborhood converges linearly to a point in Fix T with rate at least $K = \left(1 - \frac{2\eta\beta\mu^2}{(\mu+\eta)^2}\right)^{\frac{1}{2}} < 1$.

[Lions&Mercier, 1979]

See also He&Yuan, (2012); Boley (2013); Hesse&L. (2013); Bauschke,BelloCruz,Nghia,Phan&Wang(2014); Bauschke&Noll(2014); Hesse, Neumann&L. (2014); Patrinos, Stella&Bemporad (2014); Giselsson (2015×2).

Strong monotonicity: nice when you have it...

- TV: $f(x) := \|\nabla x\|_1$
- modified Huber:

$$f_\alpha(t) = \begin{cases} \frac{(t+\epsilon)^2-\epsilon^2}{2\alpha} & \text{if } 0 \leq t \leq \alpha - \epsilon \\ \frac{(t-\epsilon)^2-\epsilon^2}{2\alpha} & \text{if } -\alpha + \epsilon \leq t \leq 0 \\ |t| + \left(\epsilon - \frac{\epsilon^2+\alpha^2}{2\alpha}\right) & \text{if } |t| > \alpha - \epsilon. \end{cases}$$

Beyond monotonicity

Pointwise quadratically supportable functions

(i) $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is pointwise quadratically supportable at y if it is subdifferentially regular there and $\exists$ a neighborhood V of y and a $\mu > 0$ such that

$$(\forall v \in \partial\varphi(y)) \quad \varphi(x) \geq \varphi(y) + \langle v,\ x - y\rangle + \frac{\mu}{2}\|x - y\|^2, \quad \forall x \in V.$$

(ii) $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is strongly coercive at y if it is subdifferentially regular on V and $\exists$ a neighborhood V of y and a constant $\mu > 0$ such that

$$(\forall v \in \partial\varphi(z)) \quad \varphi(x) \geq \varphi(z) + \langle v,\ x - z\rangle + \frac{\mu}{2}\|x - z\|^2, \quad \forall x, z \in V.$$

Strong convexity

Compare to:

(pointwise) strongly convex functions

(i) $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is pointwise strongly convex at y if there $\exists$ a convex neighborhood V of y and a constant $\mu > 0$ such that, $(\forall \tau \in (0,1))$

$$\varphi(\tau x + (1-\tau)y) \leq \tau\varphi(x) + (1-\tau)\varphi(y) - \frac{1}{2}\mu\tau(1-\tau)\|x-y\|^2, \quad \forall x \in V.$$

(ii) $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is strongly convex at y if $\exists$ a cvx neighborhood V of y and a constant $\mu > 0$ such that, $(\forall \tau \in (0,1))$

$$\varphi(\tau x + (1-\tau)z) \leq \tau\varphi(x) + (1-\tau)\varphi(z) - \frac{1}{2}\mu\tau(1-\tau)\|x-z\|^2, \ \forall x, z \in V.$$

Relations

- $$\begin{aligned} \{\text{str cvx fncts}\} &= \{\text{str coercive fncts}\} \\ &= \{\text{str mon fncts}\} \\ &\subset \{\text{cvx fncts}\} \end{aligned}$$

Relations

- $$\begin{aligned}\{\text{str cvx fncts}\} &= \{\text{str coercive fncts}\}\\ &= \{\text{str mon fncts}\}\\ &\subset \{\text{cvx fncts}\}\end{aligned}$$

- $\{$ptws str cvx fncts at $\overline{x}\} \subset \{$ptws quadr supportable fncts at $\overline{x}\}$

  $\{$ptws str mon fncts at $\overline{x}\} \subset \{$ptws quadr supportable fncts at $\overline{x}\}$

  f ptws quadratically supportable at $\overline{x} \not\Rightarrow f$ convex

Linear Convergence of PAPC

Recall

PAPC

Initialization: Let $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$, and choose the parameters τ and σ to satisfy

$$\tau \in \left(0, \frac{1}{L_f}\right), \quad 0 < \tau\sigma \leq \frac{1}{\|\mathcal{A}^T\mathcal{A}\|}.$$

Main Iteration: for $k = 1, 2, \ldots$ update x^k, y^k as follows:

$$\begin{aligned}
&p^k = x^{k-1} - \tau(\nabla f(x^{k-1}) + \mathcal{A}y^{k-1}); \\
&\text{for } i = 1, \ldots, l, \\
&\qquad y_i^k = \text{prox}_{\sigma, g_i^*}\left(y_i^{k-1} + \sigma A_i^T p^k\right); \\
&x^k = x^{k-1} - \tau(\nabla f(x^{k-1}) + \mathcal{A}y^k).
\end{aligned}$$

Saddle-point

$$\min_{x\in\mathbb{R}^n} \max_{y\in\mathbb{R}^m} \left\{K(x, y) := f(x) + \left\langle \mathcal{A}^T x, \ y\right\rangle - g^*(y)\right\}.$$

Convergence to unique solutions

Q-linear convergence of PAPC

For f convex, ptwise quadrat. supportable at all saddle-point solutions and differentiable with Lipschitz gradient, g convex and $\mathcal{A}$ full rank, the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ generated by the PAPC algorithm is Q-linearly convergent to every saddle-point solution with respect to a weighted Euclidean norm dependent on σ, τ and $\mathcal{A}$.

Uniqueness of saddle-points

For f convex, ptwise quadrat. supportable at all saddle-point solutions and differentiable with Lipschitz gradient, g convex and $\mathcal{A}$ full rank, the set of saddle points is a singleton.

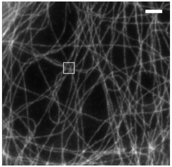
Outline

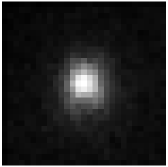
Statistical Image Denoising/Deconvolution

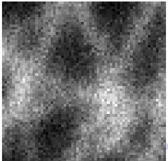
$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & g_\epsilon(Ax) \leq 0 \end{array} \quad \to \quad \underset{x \in \mathbb{R}^n}{\text{minimize}}\, f(x) + \rho \max\{g_\epsilon(Ax)\}.$$

exact regularization

Solve with
ADMM = Douglas-Rachford on the dual [Aspelmeier-Charitha-L. 2016]
Solve with Proximal Alternating Predictor-Corrector (primal-dual for saddle-point model) [L., Shefi 2017].







Structural assumptions

Reconstruct the estimator $\bar{x}$ of the observed signal b that is a solution to the convex optimization problem:

$$\inf_{x \in X} f(x) \quad \text{s.t} \quad \max_{s \in \mathcal{S}} \left| \sum_{\nu \in \mathcal{G}} \omega^s (Ax - b)_\nu \right| \leq q \tag{3}$$

The following blanket assumptions on the problem's data hold throughout:

Assumptions

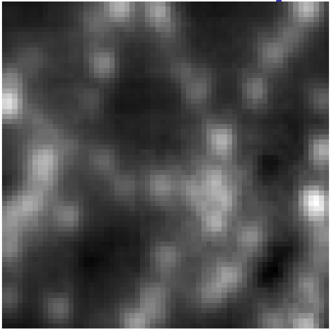
(i) The set of optimal solutions for problem $(\mathcal{P})$, denoted X^*, is nonempty.

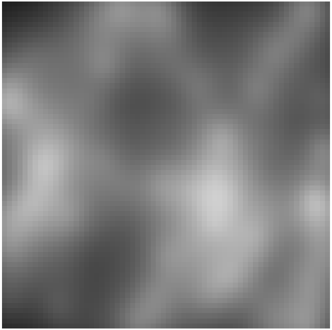
(ii) The function $f : \mathbb{R}^n \to \mathbb{R}$ is convex and continuously differentiable with Lipschitz continuous gradient ∇f (constant L_f) and pointwise quadratically supportable at points in X^*

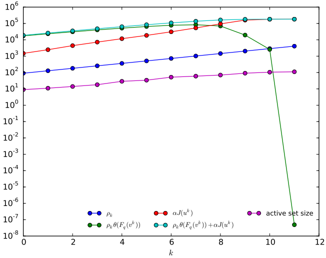
(iii) $g_i : \mathbb{R}^{m_i} \to (-\infty, +\infty]$, $(i = 1, \ldots, l)$ is proper, lsc, and convex.

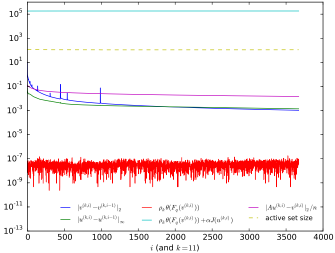
(iv) The linear mappings $A_i : \mathbb{R}^{m_i} \to \mathbb{R}^n$, $i = 1, \ldots, l$ are full rank, that is, $\sigma^2_{min}(A_i) = \lambda_{min}(A_i^T A_i) > 0$.

ADMM with exact penalization









(about 1 week cpu time)

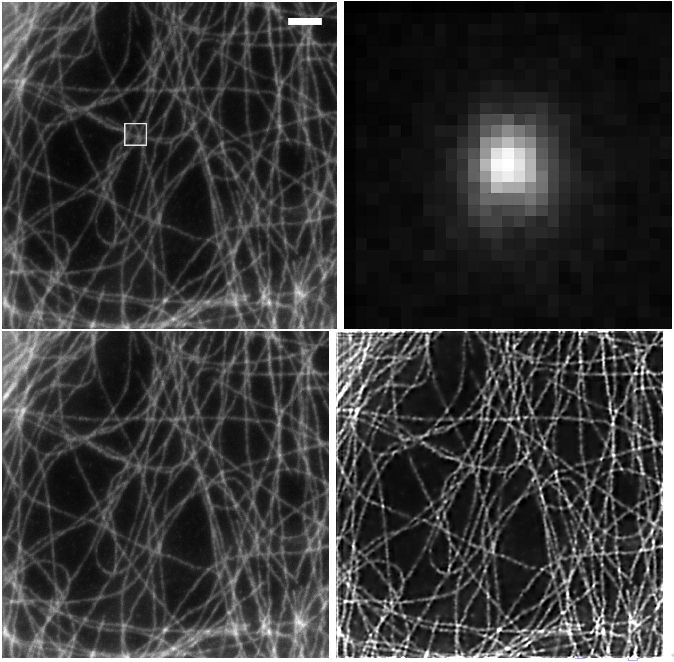
ADMM with exact penalization

What you can say about the reconstruction:

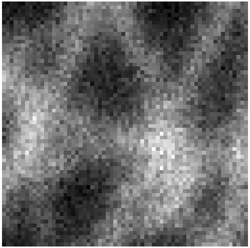
Under the assumption that the latter iterates are indeed in the region of local linear convergence and exact evaluation of prox mappings, the observed convergence rate is $c = 0.9997$, which yields an a posteriori upper estimate of the pixelwise error of about $8.9062e^{-4}$, or 3 digits of accuracy at each pixel for the computed solution to

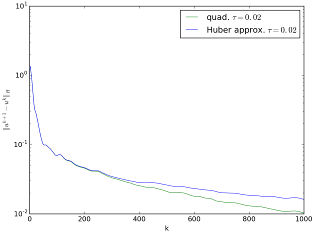
$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & F_\epsilon(Ax) \le 0 \end{array} .$$

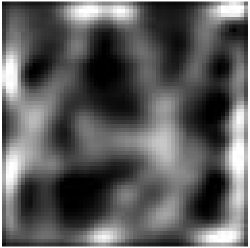
PAPC with exact constraints

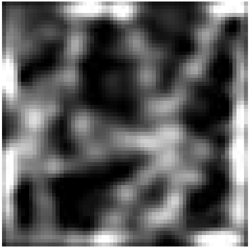


PAPC with exact constraints









(about 2 hours cpu time)

PAPC with exact constraints

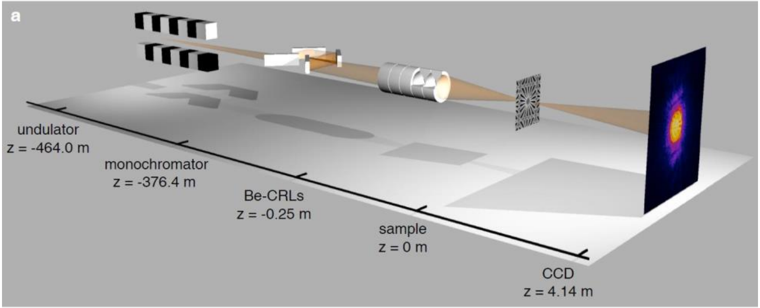
What you can say about the reconstruction:

With an estimated convergence rate of $c = 0.9993$ for the Huber objective this corresponds to an a posteriori upper estimate of the error at iteration $k = 800$ of $2.4 * 10^{-3}$. With an estimated convergence rate of $c = 0.9962$ for the quadratic objective function this corresponds to an a posteriori upper estimate of the error at iteration $k = 800$ of $1.5 * 10^{-3}$ – about two digits of accuracy at each pixel for the computed solution to

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & F_\epsilon(Ax) \le 0 \end{array} .$$

Blind Phase Retrieval

Ptychographic Imaging [Hegerl&Hoppe, (1970)]



[Institute for X-Ray Physics, Göttingen]

Blind Phase Retrieval

Mathematical Model:

Let $\mathcal{F} : \mathbb{C}^n \to \mathbb{C}^n$ denote the discrete Fourier transform. Given $b_j \in \mathbb{R}^n_+$ and the linear shift operator $S_j : \mathbb{C}^n \to \mathbb{C}^n$,
find $x, y \in \mathbb{C}^n$ satisfying

$$\left|\left(\mathcal{F}\left(S_j\left(x\right) \odot y\right)\right)_l\right| = b_{jl}, \;\; (j = 1, 2, \ldots, m)(l = 1, 2, \ldots n).$$

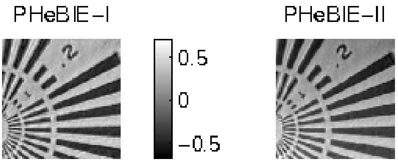
Typical problem sizes:

$n = 9.6 \times 10^5, \;\; m = 400$

$\Longrightarrow$

3.86×10^8 nonlinear equations
in 3.86×10^6 unknowns.

Algorithms
must be simple (no parameters)
and must say more than
the standard techniques can say.



[Hesse, L. Sabach, Tam (2015)]

ProxToolbox

http://num.math.uni-goettingen.de/proxtoolbox

(Python version coming soon!)

Outline

References I

T. Aspelmeier, Charitha, D. R. Luke,
Local Linear Convergence of the ADMM/Douglas–Rachford Algorithms without Strong Convexity and Application to Statistical Imaging
SIAM J. on Imaging Sciences (2016)

D. R. Luke, Nguyen H. T., M. K. Tam
Quantitative convergence analysis of iterated expansive, set-valued mappings
submitted (arXiv:1605.05725)

D.R. Luke and R. Shefi,
A Globally Linearly Convergent Method for Pointwise Quadratically Supportable Convex-Concave Saddle Point Problems
J. Math. Anal. and Appl. (2017).