MARKOV CHAINS AND RANDOM OPERATOR SPLITTING: APPLICATION TO TOMOGRAPHY IN X-RAY IMAGING

Russell Luke

Universität Göttingen

CRM, Montréal,
May 27-30, 2025

*RTG*2088

CRC 1456
MATHEMATICS OF EXPERIMENT

Deutsche Forschungsgemeinschaft

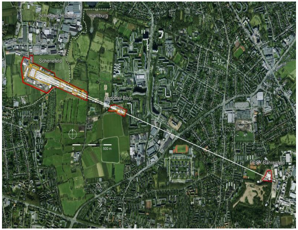
Collaborators

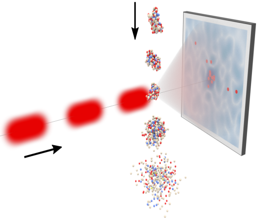
- Helmut Grubmüller (XFEL)
- Neal Hermer (stochastic fixed point problems)
- Titus Pinta (stochastic quasi-Newton methods)
- Steffen Schultze (XFEL)
- Anja Sturm (stochastic fixed point problems)

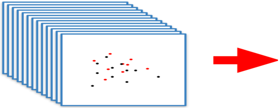
Outline I

Actual Application: X-FEL

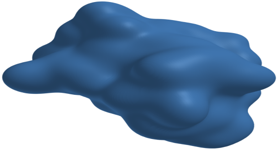
Single-shot Femtosecond X-ray Imaging







$10^6 \ldots 10^8$ *images*



(joint work with Helmut Grubmüller)

X-FEL

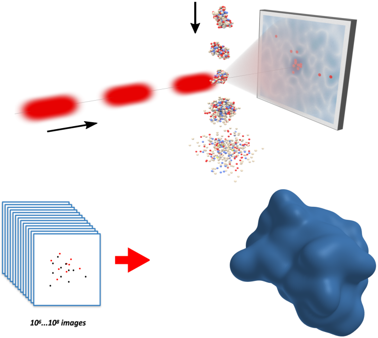
Single-shot Femtosecond X-ray Imaging with Density Functional Theory

Given a collection of measurements $\xi := \{\Omega_1, \ldots, \Omega_M\}$,

$$\text{Find } x \in \mathbb{R}^n \quad s.t.\ \rho(x) \text{ satisfying } \left(1 + e^{\beta(H_\rho - \mu)}\right)^{-1}(x) = \rho(x)$$

is the most likely cause of the data ξ.

- $\rho(x)$: electron density parameterized by $x \in \mathbb{R}^n$
- H_ρ: single-particle Hamiltonian + effective potential
- β, μ: physical constants



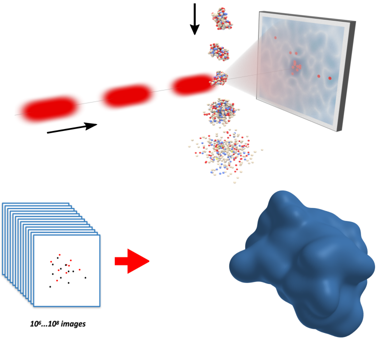
X-FEL

Single-shot Femtosecond X-ray Imaging

Given a collection of measurements $\xi := \{\Omega_1, \ldots, \Omega_M\}$,

$$(\mathcal{P}_\xi) \qquad \underset{x \in \Omega_0}{\text{minimize}} \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j\right)\right).$$

- $\phi : \mathbb{R}^n \to \mathbb{R}^m$: a smooth model of a probability distribution for the electrons parameterized by $x \in \mathbb{R}^n$
- $\Omega_0 \subset \mathbb{R}^n$: an abstract constraint set (self-consistent field or simpler)



X-FEL

Single-shot Femtosecond X-ray Imaging

Given a collection of measurements $\xi := \{\Omega_1, \ldots, \Omega_M\}$,

$$(\mathcal{P}_\xi) \qquad \underset{x \in \Omega_0}{\text{minimize}} \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j\right)\right).$$

$\Omega_j^{(k)}(z) \in \mathbb{N}$ is a photon count at $z \in \mathbb{R}^3$ and

$$\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right) := \int_{SO(3)} \prod_{z \in \operatorname{Supp}\left(\Omega_j^{(k)}\right)} \frac{\phi^s(z;x)^{\Omega_j^{(k)}(z)} \exp\left(-\phi^s(z;x)\right)}{\Omega_j^{(k)}(z)!} d\mu(s) \quad \text{for}$$

$$\phi^s(z;x) := \left| \left[\mathcal{F}\left(\underbrace{\sum_{i=1}^{N} \sigma * \exp\left(-\tfrac{1}{2}\|\cdot - x_i\|^2\right) \circ R(s)}_{\rho^s(\cdot;x)} \right) \right](z) \right|^2, \quad x \in \mathbb{R}^{3N}$$

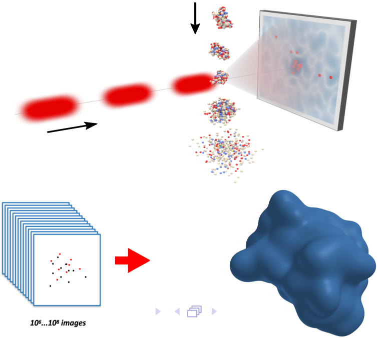
X-FEL

Single-shot Femtosecond X-ray Imaging

For $k = 1, 2, \ldots$ **do**

1. Sample a collection of measurements $\xi_k := \{\Omega_1^{(k)}, \ldots, \Omega_M^{(k)}\}$.
2. Compute

$$X_k \in \epsilon\text{−argmin}_{x \in \Omega_0} \left\{ \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)} \right) \right) \right\}.$$



X_k is a random variable $\Longrightarrow$ convergence in distribution

In Contrast: Orbital Tomography

Angle-Resolved Photon Emission Spectroscopy

Given $\{\Omega_1, \ldots, \Omega_M\}$,

$$(\mathcal{P}) \qquad \underset{x \in \Omega_0}{\text{minimize}} \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j\right)\right).$$

$\Omega_j(z) >> 1$ large photon count at $z \in r\mathbb{S}$ and known orientation s

$$\mathbb{P}_{\phi(x)}\left(\Omega_j(z)\right) \quad := \quad exp\left(-\|b_j(z) - \phi(z;x)\|^2\right)$$

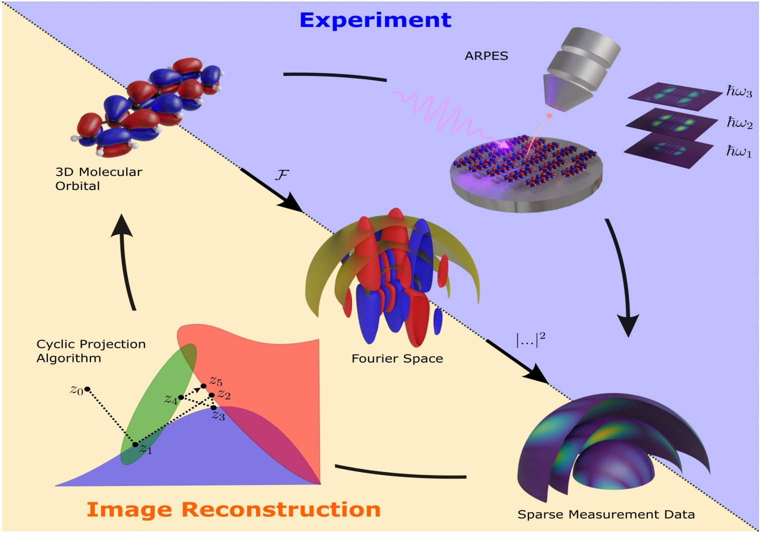
where $b_j(z) \in \mathbb{R}_+$ is the measured intensity

$$\phi(z;x) := \left|\left[\mathcal{F}\left(\rho(\cdot;x)\right)\right](z)\right|^2,$$

$\Longrightarrow$ Phase retrieval: $(\mathcal{P}) \qquad \underset{x \in \Omega_0}{\text{minimize}} \sum_{j=1}^{M} \|b_j(z) - \phi(z;x)\|^2.$

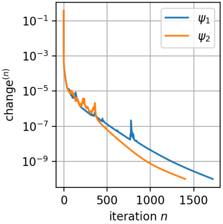
Example: Orbital Tomography

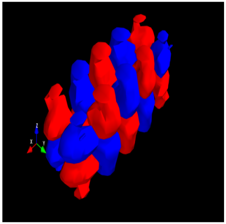
Orbital Imaging [Dinh-Matthijs-L.-Bennecke-Mathias, 2024]

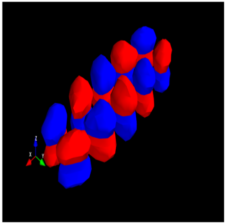


Example: Orbital Tomography [Dinh-Matthijs-L.-Bennecke-Mathias, 2024,2025]

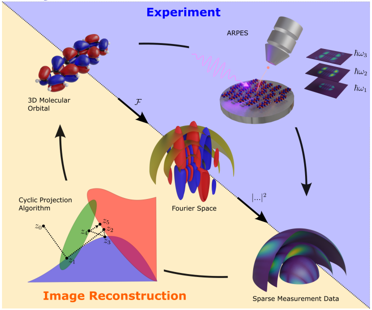
Cyclic Projections: $x^{k+1} = P_{\Omega_0} P_{\Omega_1} \cdots P_{\Omega_M} x^k$

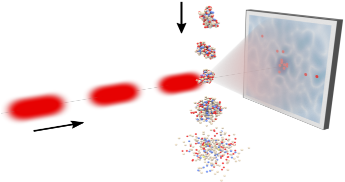


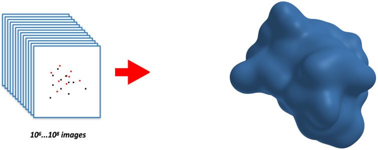




Key Differences to XFEL







- Acquisition time large $\implies$ M small ($\approx$ 7)
- Voxel parameterization $\rho \implies$ slightly over-parameterized
- Measurements inconsistent with Ω_0

- Acquisition time small $\implies$ M big ($\approx 10^7$)
- Highly underparameterized
- Extremely low-count images: $10 - 100$ photons per Ω_j

Summary

Main ideas:

- Stochastic iterations $\to$ Markov chains
- Regularity of mappings in application domain $\nearrow$ regularity of Markov operators in probability spaces

Main contributions so far:

- Calculus of nonmonotone mappings in measure spaces
- Convergence with rates for noncontractive/expansive Markov chains

Challenges

- characterization of invariant measures
- monitoring convergence
- concentration estimates with rates (complexity)
- scale-aware efficiency

Outlook

- new algorithms and methods for MCMC, randomized algorithms for large-scale optimization, Bayesian machine learning...

Outline I

Algorithms

Stochastic gradient descent [L.-Schultze-Grubmüller, 2024]

Initialization. Select $X_0 \sim \mu_0 \in \mathscr{P}(\mathcal{E})$, and $(\xi_k)_{k\in\mathbb{N}}$, an i.i.d. sequence, where X_0 and (ξ_k) are independently distributed, $\xi_k := \left(\Omega_1^{(k)}, \Omega_2^{(k)}, \ldots, \Omega_M^{(k)}\right)$. Define

$$T_{\xi_k}^{GD}(x) := \left(x - t\sum_{j=1}^{M} \nabla_x \left(-\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right)\right)\right)\right)^r$$

For $k = 0, 1, \ldots$ **do**

1. Sample ξ_k
2. Compute $X_{k+1} = T_{\xi_k}^{GD}(X_k)$.

The parameter t is the fixed step length, and r indicates doing r steps of this operation

Algorithms

Stochastic quasi-Newton [L.-Pinta, 2024]

Initialization. Select $X_0 \sim \mu_0 \in \mathscr{P}(\mathcal{E})$, and $(\xi_k)_{k\in\mathbb{N}}$, an i.i.d. sequence, where X_0 and (ξ_k) are independently distributed, $\xi_k := \left(\Omega_1^{(k)}, \Omega_2^{(k)}, \dots, \Omega_M^{(k)}\right)$. Define

$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)$$

For $k = 0, 1, \dots$ **do**

1. Sample ξ_k
2. Compute $X_{k+1} = T^{QN}_{\xi_k}(X_k)$.

The parameter r still indicates doing r steps of this operation, while $H_i(x)$ is the Quasi-Newton update at the ith inner iteration.

Algorithms

Remarks

- $\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right)$ is smooth, but nonconvex as a function of x
- ξ_k is a random variable
- $\implies X^{k+1}$ is a random variable

$\implies$

Algorithm 1 RFI: Random Function Iterations

Initialization: $X^0 \sim \mu$, $\xi_k \sim \xi$ iid
for $k = 0, 1, 2, \ldots$ **do**
$X^{k+1} = T_{\xi_k} X^k$
return $\{X^k\}_{k \in \mathbb{N}}$

(joint work with Anja Sturm and Neal Hermer [2019, 2023a, 2023b])

Outline I

Random Function Iteration

Algorithm 2 RFI: Random Function Iterations

Initialization: $X^0 \sim \mu$, $\xi_k \sim \xi$ iid
  for $k = 0, 1, 2, \ldots$ **do**
    $X^{k+1} = T_{\xi_k} X^k$
  return $\{X^k\}_{k \in \mathbb{N}}$

RFI as a Markov Chain

$X^{k+1} = T_{\xi_k} X^k$ $(k \in \mathbb{N}_0)$ is a time homogeneous Markov chain with transition kernel p:

$$(x \in \mathcal{E})(A \in \mathcal{B}(\mathcal{E})) \qquad p(x, A) := \mathbb{P}(T_\xi x \in A)$$

Markov Operators and the Stochastic Fixed Point Problem

Markov operator $\mathcal{P}$

- (dual) The Markov operator $\mathcal{P} : \mathscr{P}(\mathcal{E}) \to \mathscr{P}(\mathcal{E})$ acting on a measure $\mu \in \mathscr{P}(\mathcal{E})$ is defined via

$$(A \in \mathcal{B}(\mathcal{E})) \qquad \mu\mathcal{P}(A) := \int_{\mathcal{E}} p(x, A)\mu(dx).$$

Stochastic Feasibility

The fixed point approach:

$$\text{Find} \quad X \quad \text{such that} \quad \mathbb{P}(X = T_\xi X) = 1.$$

Generically, X does not exist $\Longrightarrow$ ill-posed.

$\Longrightarrow$

Stochastic Fixed Point Problem

Given a Markov operator $\mathcal{P}$, Find $\overline{\mu} \in \operatorname{inv} \mathcal{P} := \{\mu \mid \mathcal{P}\mu = \mu\}$ $(= \text{Fix } \mathcal{P})$

Markov Operators and the Stochastic Fixed Point Problem

$X^{k+1} = T_{\xi_k} X^k \sim \mu^{k+1}$

- The Markov operator $\mathcal{P} : \mathscr{P}(\mathcal{E}) \to \mathscr{P}(\mathcal{E})$ is defined by the transition kernel $p(x, \cdot) := \mathbb{P}(T_\xi x \in \cdot)$ where $T_\xi : \mathcal{E} \to \mathcal{E}$ is a randomly selected mapping with certain nice properties.
- Convergence of the sequence of measures $\mu^{k+1} = \mu^k \mathcal{P}$ depends on properties of T_ξ and
- is in the sense of distributions.

Metrizing convergence

Prokhorov-Lèvy&Wasserstein distance

Let $\mu, \nu \in \mathscr{P}(\mathcal{E})$.

(i) The Prokhorov-Lèvy distance, denoted by d_P, is defined by

$$\begin{aligned} d_P(\mu,\nu) \quad = \quad \inf \{\epsilon > 0 \mid \mu(A) \leq \nu(\mathbb{B}(A,\epsilon)) + \epsilon, \\ \nu(A) \leq \mu(\mathbb{B}(A,\epsilon)) + \epsilon \\ \forall A \in \mathcal{B}(\mathcal{E})\}\,. \end{aligned}$$

(ii) For $p \geq 1$ let

$$\mathscr{P}_p(\mathcal{E}) = \left\{ \mu \in \mathscr{P}(\mathcal{E}) \,\middle|\, \exists x \in \mathcal{E} \,:\, \int d^p(x,y)\mu(dy) < \infty \right\}.$$

The Wasserstein p-metric on $\mathscr{P}_p(\mathcal{E})$, denoted d_{W_p}, is defined by

$$d_{W_p}(\mu,\nu) := \left(\inf_{\gamma \in C(\mu,\nu)} \int_{\mathcal{E}\times\mathcal{E}} d^p(x,y)\gamma(dx,dy) \right)^{1/p} \quad (p \geq 1)$$

where $C(\mu,\nu)$ is the set of couplings of μ and ν (measures on the product space $\mathcal{E} \times \mathcal{E}$ whose marginals are μ and ν respectively).

Nonconvex Case

Almost α-firmly nonexpansive (aα-fne)

$T_\xi : \mathcal{E} \to \mathcal{E}$ is almost α-firmly nonexpansive (aα-fne) with constant $\alpha \in (0,1)$ and violation $\epsilon > 0$ on $D \subset \mathcal{E}$ whenever

$$d(T_\xi x, T_\xi y)^2 \leq (1+\epsilon) d(x,y)^2 - \frac{1-\alpha_\xi}{\alpha_\xi} d\left((\mathrm{Id} - T_\xi)\, x, (\mathrm{Id} - T_\xi)\, y\right)^2 \quad \forall x, y \in D \subset \mathcal{E}.$$

T_ξ is aα-fne in expectation with constant $\alpha \in (0,1)$:

$$\mathbb{E}_\xi \left[d^2(T_\xi(x), T_\xi(y)) \right] \leq (1+\epsilon) d^2(x,y) - \tfrac{1-\alpha}{\alpha} \mathbb{E}_\xi \left[d\left((\mathrm{Id} - T_\xi)\, x, (\mathrm{Id} - T_\xi)\, y\right)^2 \right], \quad \forall x, y \in D \subset \mathcal{E}.$$

Examples: projectors onto prox-regular sets, prox-mappings of smooth functions, gradient descent steps for smooth functions, proximal-gradient algorithms...

Main Result: Nonconvex Case

Define $\Psi : \mathscr{P}(D) \to \mathbb{R}_+ \cup \{+\infty\}$ by

$$\Psi(\mu) := \inf_{\pi \in \operatorname{inv} \mathcal{P}} \inf_{\gamma \in C_*(\mu,\pi)} \left(\int_{D\times D} \mathbb{E}_\xi \left[\underbrace{d\left((\mathrm{Id} - T_\xi)\, x, (\mathrm{Id} - T_\xi)\, y\right)^2}_{\psi(x,y,T_\xi x, T_\xi y)} \right] \gamma(dx, dy) \right)^{1/2}$$

Convergence rates [Hermer, L.& Sturm 2023b]

Suppose

(a) T_ξ is almost α-fne in expectation with constant $\alpha \in (0,1)$ on $D \subset \mathcal{E}$:

$$\mathbb{E}_\xi \left[d^2(T_\xi(x), T_\xi(y)) \right] \leq (1+\epsilon) d^2(x,y) - \frac{1-\alpha}{\alpha} \mathbb{E}_\xi \left[\psi_{T_\xi}(x,y) \right], \quad \forall x, y \in D;$$

(b) Ψ is gauge metrically subregular with respect to d_{W_2} for 0 on $\mathscr{P}_2(D)$ with gauge $\mathcal{G}$[1]:

$$d_{W_2}(\mu, \operatorname{inv} \mathcal{P}) \leq \mathcal{G}(\Psi(\mu)) \quad \forall \mu \in \mathscr{P}_2(D).$$

[1] [Klatte&Kummer, Ioffe, Ngai& Théra, Dontchev&Rockafellar]

Then,

for any $\mu_0 \in \mathscr{P}_2(D)$ the distributions (μ_k) of the iterates of Algorithm RFI satisfy the error bound

$$d_{W_2}\left(\mu_{k+1}, \operatorname{inv}\mathcal{P}\right) \le \theta_{\alpha,\epsilon}\left(d_{W_2}\left(\mu_k, \operatorname{inv}\mathcal{P}\right)\right) \quad \forall k \in \mathbb{N} \qquad (1)$$

where

$$\theta_{\alpha,\epsilon}(t) = \left((1+\epsilon)t^2 - \tfrac{1-\alpha}{\alpha}\left(\mathcal{G}^{-1}(t)\right)^2\right)^{1/2}.$$

If, in addition, $\theta_{\alpha,\epsilon}^{(k)}(t) \underset{k}{\to} 0$ for all t small enough, then the sequence converges in the Wasserstein metric with rate characterized by $\theta_{\alpha,\epsilon}$.

Main Result: Nonconvex Case

Linear case
Under the same assumptions as above if Ψ satisfies condition (b) with gauge $\mathcal{G}(t) = \kappa \cdot t$ and constant κ satisfying

$$\sqrt{\frac{1-\alpha}{\alpha(1+\epsilon)}} \leq \kappa \; < \sqrt{\frac{1-\alpha}{\alpha\epsilon}},$$

then the sequence of iterates (μ_k) converges R-linearly ($\mathcal{O}(1/\delta)$) to some $\pi^{\mu_0} \in \operatorname{inv}\mathcal{P} \cap \mathscr{P}_2(D)$:

$$d_{W_2}\left(\mu_{k+1}, \operatorname{inv}\mathcal{P}\right) \leq c\, d_{W_2}\left(\mu_k, \operatorname{inv}\mathcal{P}\right), \quad \text{where } c := \sqrt{1+\epsilon-\left(\tfrac{1-\alpha}{\kappa^2\alpha}\right)} < 1.$$

Proof sketch

Start with the inequality defining metric subregularity (b) and insert the inequality (a) to get the error bound in the Wasserstein metric. To get convergence of μ_k to a measure $\pi^{\mu_0} \in \operatorname{inv}\mathcal{P}$, show that it is a Cauchy sequence. $\square$

Convex Case

(α-firmly) nonexpansive

T_ξ is nonexpansive whenever

$$\mathrm{dist}(T_\xi x, T_\xi y)^2 \leq \mathrm{dist}(x, y)^2 \qquad \forall x, y \in \mathcal{E}.$$

T_ξ is α-firmly nonexpansive with constant $\alpha_\xi \leq \alpha \in (0, 1)$ whenever

$$\mathrm{dist}(T_\xi x, T_\xi y)^2 \leq \mathrm{dist}(x, y)^2 - \frac{1-\alpha_\xi}{\alpha_\xi} \mathrm{dist}\left((\mathrm{Id} - T_\xi)x, (\mathrm{Id} - T_\xi)y\right)^2 \quad \forall x, y \in \mathcal{E}.$$

Examples: gradient descents of convex functions, projectors onto convex sets, prox-mappings of convex functions, convex projection algorithms, convex Douglas-Rachford...

Randomization→ weaker requirements for good behavior

Global Convergence in $\mathcal{E}$ [Hermer-L.-Sturm, 2023a]

- [2] Let $T_\xi : \mathcal{E} \to \mathcal{E}$ be nonexpansive and assume $\mathrm{inv}\,\mathcal{P} \neq \emptyset$. Let $\mu \in \mathscr{P}(\mathcal{E})$ and $\nu_k = \frac{1}{k}\sum_{j=1}^{k} \mu\mathcal{P}^j$, then this sequence converges in the Prokhorov-Lèvy metric to an invariant probability measure for $\mathcal{P}$, i.e. $\nu_k \to \pi^\mu \in \mathrm{inv}\,\mathcal{P}$.

- [3] If T_ξ is α-firmly nonexpansive and $\mathrm{inv}\,\mathcal{P} \neq \emptyset$, then $\mu^k \to \pi^\mu \in \mathrm{inv}\,\mathcal{P}$ in the Prokhorov-Lèvy metric.

Example

SGD for any smooth convex function and any fixed stepsize $t < 1/L$ initialized from any starting point converges in distribution with respect to the Prokhorov-Lèvy metric.

---

[2] This is the only convergence result we know of for nonexpansive mappings!

[3] The only result we know of where T_ξ don't have common fixed points.

Nonconvex Case: XFEL

Single-shot Femtosecond X-ray Imaging

Given a collection of measurements $\xi := \{\Omega_1, \ldots, \Omega_M\}$,

$$(\mathcal{P}_\xi) \qquad \underset{x\in\Omega_0}{\text{minimize}} \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j\right)\right).$$

$\Omega_j^{(k)}(z) \in \mathbb{N}$ is a photon count at $z \in \mathbb{R}^3$ and

$$\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right) := \int_{SO(3)} \prod_{z\in \mathrm{Supp}\left(\Omega_j^{(k)}\right)} \frac{\phi^s(z;x)^{\Omega_j^{(k)}(z)}\exp\left(-\phi^s(z;x)\right)}{\Omega_j^{(k)}(z)!} d\mu(s) \quad \text{for}$$

$$\phi^s(z;x) := \left|\left[\mathcal{F}\left(\underbrace{\sum_{i=1}^{N}\sigma * \exp\left(-\tfrac{1}{2}\|\cdot - x_i\|^2\right)\circ R(s)}_{\rho^s(\cdot;x)}\right)\right](z)\right|^2, \qquad x \in \mathbb{R}^{3N}$$

Nonconvex Case: XFEL

Recall:

Stochastic gradient descent/quasi-Newton (XFEL)

Initialization. Select $X_0 \sim \mu_0 \in \mathscr{P}(\mathbb{R}^n)$, and $(\xi_k)_{k\in\mathbb{N}}$, an i.i.d. sequence, where X_0 and (ξ_k) are independently distributed. Define

$$T^{GD}_{\xi_k}(x) := \left(x - t \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)^{r}$$

or

$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)$$

For $k = 0, 1, \ldots$ **do**

1. Sample ξ_k
2. Compute $X_{k+1} = T_{\xi_k}(X_k)$.

Stochastic gradient descent/quasi-Newton (XFEL)

Define

$$T_{\xi_k}^{GD}(x) := \left(P_{\Omega_0} \circ \left(x - t \sum_{j=1}^{M} \nabla_x \left(-\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right)\right)\right)\right)\right)^r$$

or

$$T_{\xi_k}^{QN}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j^{(k)}\right)\right)\right)\right)$$

The Quasi-Newton Update

$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} T^{\text{inner}}_{\xi_k}(x_i, H_i)$$

$$T^{\text{inner}}_{\xi_k}(x, H) = \begin{pmatrix} x - H \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \\ \underbrace{\left(I - \frac{sy^T}{y^T s} \right) H \left(I - \frac{ys^T}{s^T y} \right) + \frac{ss^T}{y^T s}}_{\mathcal{U}(H)} \end{pmatrix}$$

where

$$H_0 = I, \quad x^+ = x - H \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right),$$

$$s = x^+ - x, \quad y = \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x^+)} \left(\Omega_j^{(k)} \right) \right) \right) - \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right)$$

Gradient descent

Steepest descent mappings are aα-fne

Let $g : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable with locally Lipschitz, hypomonotone gradient, that is g satisfies

$$\exists L > 0 : \quad \|\nabla g(x) - \nabla g(y)\|^2 \quad \leq \quad L^2\|x - y\|^2 \quad \forall x, y \in D \subset \mathbb{R}^n,$$

and

$$\exists \tau \geq 0 : -\tau \|x - y\|^2 \quad \leq \quad \langle \nabla g(x) - \nabla g(y),\ x - y\rangle \quad \forall x, y \in D.$$

Then $T_{GD} := \mathrm{Id} - t\nabla g$ is aα-fne on D with violation at most

$$\epsilon_{GD} = \left\{2t\tau + \frac{t^2L^2}{\alpha}\right\} < 1, \quad \text{with constant} \quad \alpha$$

whenever the steps t satisfy

$$t \in \left(0, \frac{\alpha\sqrt{\tau^2 + L^2} - \alpha\tau}{L^2}\right).$$

Newton-like operators

Quasi Newton mappings are aα-fne

Let $g : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable with locally Lipschitz, hypomonotone gradient, that is g satisfies

$$\exists L > 0 : \quad \|\nabla g(x) - \nabla g(y)\|^2 \quad \leq \quad L^2\|x - y\|^2 \quad \forall x, y \in D \subset \mathbb{R}^n,$$

and

$$\exists \tau \geq 0 : -\tau \|x - y\|^2 \quad \leq \quad \langle \nabla g(x) - \nabla g(y),\ x - y\rangle \quad \forall x, y \in D.$$

Then $T_{QN} := \mathrm{Id} - H\nabla g$ is aα-fne on D with violation at most

$$\epsilon_{QN} = \left\{2\|H\|\tau + \tfrac{\|H\|^2L^2}{\alpha}\right\} < 1, \quad \text{with constant} \quad \alpha$$

whenever

$$\|H\| \in \left(0, \frac{\alpha\sqrt{\tau^2 + L^2} - \alpha\tau}{L^2}\right).$$

Main Result: Nonconvex Case

Recall

$$\Psi(\mu) := \inf_{\pi \in \operatorname{inv} \mathcal{P}} \inf_{\gamma \in C_*(\mu,\pi)} \left(\int_{D \times D} \mathbb{E}_\xi \left[\underbrace{d\left((\mathrm{Id} - T_\xi)\, x, (\mathrm{Id} - T_\xi)\, y\right)^2}_{\psi_{T_\xi}(x,y)} \right] \gamma(dx, dy) \right)^{1/2}$$

Convergence rates [Hermer, L.& Sturm 2023b]

Suppose

(a) T_ξ is almost α-fne in expectation with constant $\alpha \in (0,1)$ on $D \subset \mathcal{E}$:

$$\mathbb{E}_\xi \left[d^2(T_\xi(x), T_\xi(y)) \right] \le (1+\epsilon) d^2(x,y) - \frac{1-\alpha}{\alpha} \mathbb{E}_\xi \left[\psi_{T_\xi}(x,y) \right], \quad \forall x, y \in D;$$

(b) Ψ is gauge metrically subregular with respect to d_{W_2} for 0 on $\mathscr{P}_2(D)$ with gauge $\mathcal{G}$:

$$d_{W_2}(\mu, \operatorname{inv} \mathcal{P}) \le \mathcal{G}(\Psi(\mu)) \quad \forall \mu \in \mathscr{P}_2(D).$$

Back to XFEL

Stochastic gradient descent/Quasi-Newton (XFEL)

The update function

$$T^{GD}_{\xi_k}(x) := \left(x - t \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)^r$$

or

$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)$$

is aα-fne with constant α_{GD} or α_{QN} and violation ϵ_{GD} or ϵ_{QN} for fixed step-lengths/Hessian approximations t or $\|H\| \in \left(0, \frac{\alpha_* \sqrt{\tau^2 + L^2} - \alpha_* \tau}{2L^2}\right)$ and inner iterations r.

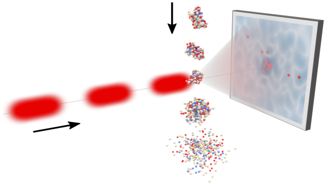
$\Longrightarrow$ we can verify condition (a) of the general convergence theorem. Metric subregularity, condition (b), is hard to verify, but was shown to be necessary for quantitative estimates [L.-Teboulle-Thao, 2020].

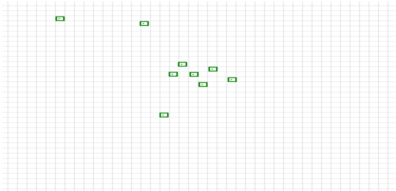
Outline I

Single-shot Femtosecond X-ray Imaging

Given a collection of measurements $\xi := \{\Omega_1, \ldots, \Omega_M\}$,

$$(\mathcal{P}_\xi) \qquad \underset{x \in \Omega_0 = \mathbb{R}^n}{\text{minimize}} \sum_{j=1}^{M} -\log\left(\mathbb{P}_{\phi(x)}\left(\Omega_j\right)\right).$$





Algorithms: Stochastic Gradient Descent

Recall:

SGD for XFEL
Initialization. Select $X_0 \sim \mu_0 \in \mathscr{P}(\mathbb{R}^n)$, and $(\xi_k)_{k\in\mathbb{N}}$, an i.i.d. sequence, where X_0 and (ξ_k) are independently distributed. Define

$$T_{\xi_k}(x) := \left(x - t \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)^{r}$$

For $k = 0, 1, \ldots$ **do**

1. Sample ξ_k
2. Compute $X_{k+1} = T_{\xi_k}(X_k)$.

Results

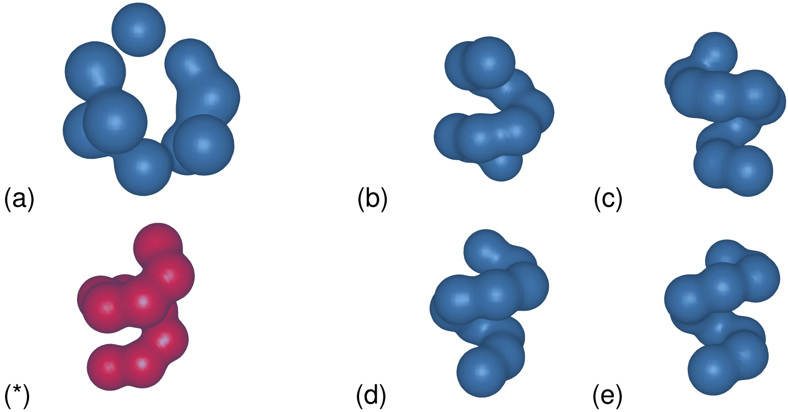


Figure: Random recovery of (*) starting from (a). Algorithm RFI for $r = 10$ and $t = 0.1$. Images are sampled with $|I_i| = M$ for (b) $M = 100$, (c) $M = 500$, (d) $M = 1000$, and (e) $M = 5000$. Shown are the computed average electron densities at iteration $k = 5000$.

Challenges

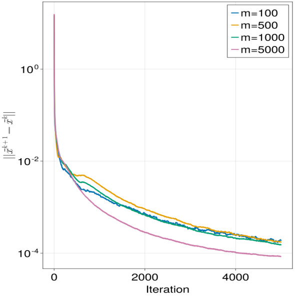
Monitoring

- Theorem: if metric subregularity holds, the sequence of measures converges in the Wasserstein metric with rate characterized by $\theta_{\alpha,\epsilon}$. If convergence is Q-linear, convergence can be estimated by monitoring $d_{W_p}(\mu_{k+1}, \mu_k)$.
-

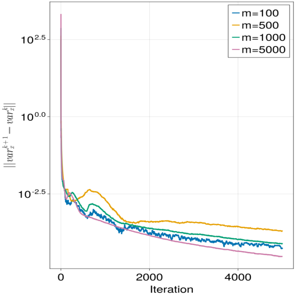
$$d_{W_p}(\mu,\nu) := \left(\inf_{\gamma \in C(\mu,\nu)} \int_{\mathcal{E}\times\mathcal{E}} d^p(x,y)\gamma(dx,dy) \right)^{1/p} \quad (p \geq 1)$$

where $C(\mu,\nu)$ is the set of couplings of μ and ν. How to calculate/estimate this?

SGD for XFEL



(a)



(b)

Figure: Convergence behavior of the mean (a) and variance (b) of the iterates. Algorithm RFI for $r = 10$ and $t = 0.1$. Images are sampled with $|I_i| = M$ for $M = 100, 500, 1000$, and 5000.

Compute times for the first 1000 outer iterations with 10 inner iterations:

- $M = 100$: 88s
- $M = 500$: 107s
- $M = 1000$: 139s
- $M = 5000$: 418s

Deterministic

$M = 10,000$, no sampling: 494s

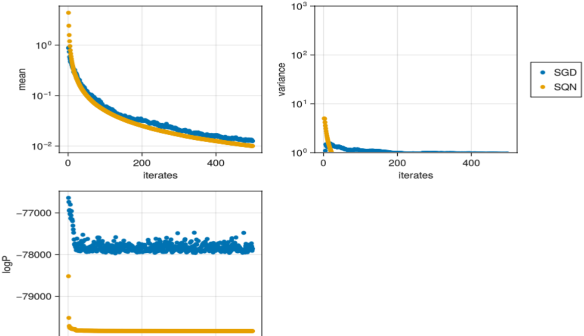
Stochastic Gradient Descent-Stochastic Quasi-Newton

Recall

$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)$$

For $k = 0, 1, \ldots$ **do**

1. Sample ξ_k
2. Compute $X_{k+1} = T_{\xi_k}(X_k)$.



Stochastic Gradient Descent-Stochastic Quasi-Newton

Recall

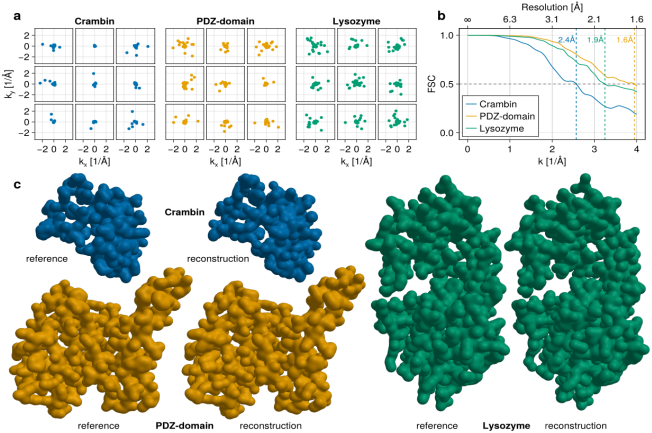
$$T^{QN}_{\xi_k}(x) := \prod_{i=1}^{r} \left(x - H_i \circ \sum_{j=1}^{M} \nabla_x \left(-\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right) \right)$$

For $k = 0, 1, \ldots$ **do**

1. Sample ξ_k
2. Compute
   $X_{k+1} = T_{\xi_k}(X_k) \in \text{argmin}_{x \in \Omega_0} \left\{ \sum_{j=1}^{M} -\log \left(\mathbb{P}_{\phi(x)} \left(\Omega_j^{(k)} \right) \right) \right\} + \epsilon \mathbb{B}.$

- Opens up a new analytic strategy: convergence of random sampling (well-established).
- Challenge: what are the analytic properties of the critical point mapping of the log-likelihood function?

SGD for XFEL: big molecules [Schultz,L. Grubmüller 2025]



- Crambin: $n = 1000$, $M = 10^7$
- PDZ: $n = 2400$, $M = 10^6$
- Lysozyme: $n = 3900$, $M - 10^6$

Summary

Main ideas:

- Stochastic iterations $\rightarrow$ Markov chains
- Regularity of mappings in application domain $\nearrow$ regularity of Markov operators in probability spaces

Main contributions so far:

- Calculus of nonmonotone mappings in measure spaces
- Convergence with rates for noncontractive/expansive Markov chains

Challenges

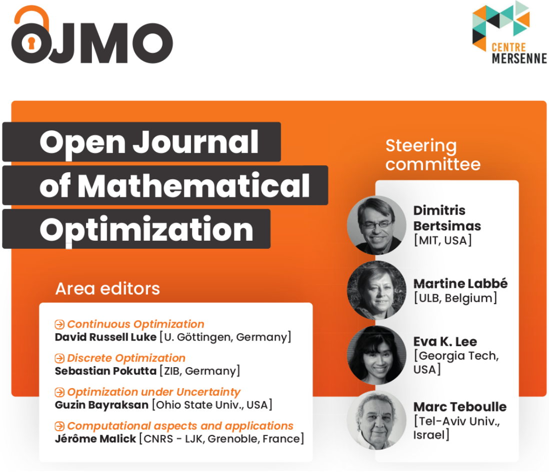
- characterization of invariant measures
- efficient computation/approximation of Wasserstein metrics
- variational analysis on probability measure spaces
- concentration estimates with rates (complexity)

Outlook

- new algorithms and methods for MCMC, randomized algorithms for large-scale optimization, Bayesian machine learning...

Outline I

Open Journal of Mathematical Optimization



Outline I

References I

N. Hermer, D. R. Luke, and A. K. Sturm,
Nonexpansive Markov Operators and Random Function Iterations for Stochastic Fixed Point Problems,
J. Convex Analysis, (2023) 30:4, 1073–1114
https://arxiv.org/abs/2205.15897

N. Hermer, D. R. Luke, and A. K. Sturm,
Rates of convergence for chains of expansive Markov Operators
Trans. Math. Appl.(2023)7:1
https://doi.org/10.1093/imatrm/tnad001

N. Hermer, D. R. Luke, and A. K. Sturm,
Random Function Iterations for Consistent Stochastic Feasibility
Numerical Functional Analysis and Optimization 40(4):386–420 (2019)
DOI: 10.1080/01630563.2018.1535507.

References II

D. R. Luke, S. Schultze and H. Grubmüller,
Stochastic Algorithms for Large-Scale Composite Optimization: the Case of Single-Shot X-FEL Imaging (2025)
https://arxiv.org/abs/2401.13454

D. R. Luke
Convergence in distribution of randomized algorithms: the case of partially separable optimization
Mathematical Programming (2024).
DOI https://doi.org/10.1007/s10107-024-02124-w.

Outline I

ProxToolbox

https://num.math.uni-goettingen.de/proxtoolbox/
https://gitlab.gwdg.de/nam/ProxPython

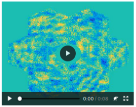
ProxToolbox

- ProxToolbox
  - Python version
    - Older versions
  - Matlab version
  - Additional binary data
  - Acknowledgements
- Getting started
  - Structure
    - I. proxtoolbox folder
    - II. InputData
    - III. Demos
      - Demos Accompanying Publications:

The ProxToolbox is a collection of modules for solving mathematical problems using fixed point iterations with proximal operators. It was used to generate many if not all of the numerical experiments conducted in the papers in the ProxMatlab Literature folder.

For a complete listing of papers with links, go to the [groups publications] (https://vaopt.math.uni-goettingen.de/en/publications.php).

This site is maintained by the Working Group in Variational Analysis of the Institute for Numerical and Applied Mathematics at the University of Göttingen.



Video: Representative iterate of a noisy JWST test wavefront recovered with the Douglas-Rachford algorithm

Python version

1.0.3:

- Sources
- Documentation

The documentation includes a tutorial.

Older versions

- **0.2.2**: Source, Documentation
- **0.1**: Source, Documentation

Matlab version

3.0

For help with the Matlab version, see the :code: `README` file in the ProxMatlab source.

Outline I

Examples

Deterministic consistent feasibility

Given $\{\Omega_1, \Omega_2, \ldots, \Omega_M\}$ closed and convex with $\Omega := \cap_j \Omega_j \neq \emptyset$.
Starting from $X_0 = \delta(x - x_0)$ for $x_0 \in \mathcal{E}$, generate

$$X_{k+1} = P_{\Omega_1} P_{\Omega_2} \cdots P_{\Omega_M} X_k$$

Here the RFI yields a sequence of point masses. The invariant measures are point masses supported on Ω

$$\operatorname{inv} \mathcal{P} \ni \delta(x - w^{x_0}) \quad \text{where } w^{x_0} \in \Omega.$$

Examples

Stochastic inconsistent feasibility

Given $\{\Omega_1, \Omega_2\}$ closed with $\Omega_1 \cap \Omega_2 = \emptyset$ and $X_0 := \delta(x - x_0)$. For $\xi_k \sim \xi$ with $\mathbb{P}(\xi = 1) = 1/2$ and $\mathbb{P}(\xi = 2) = 1/2$ generate

$$X_{k+1} = P_{\Omega_{\xi_k}} X_k$$

Here the RFI yields a sequence of averages of averaged point masses. The invariant measures, when these exist, are

$$\operatorname{inv}\mathcal{P} \ni \frac{1}{2}\delta\left(\cdot - w_1^{x_0}\right) + \frac{1}{2}\delta\left(\cdot - w_2^{x_0}\right) \quad \text{where } w_1^{x_0} \in P_{\Omega_1} w_2^{x_0}, \quad w_2^{x_0} \in P_{\Omega_2} w_1^{x_0}$$

$(w_1^{x_0}, w_2^{x_0})$ is a best approximation pair.

Low-Lying Fruit

Stochastic inconsistent feasibility:

Given $\{\Omega_1, \Omega_2, \dots, \Omega_M\}$ closed, convex and Ω_1 compact with $\cap_{j=1}^{M} \Omega_j = \emptyset$ and $X_0 := \delta(x - x_0)$. For $\xi_k \sim \xi$ with $\mathbb{P}(\xi = j) = 1/M$ ($j = 1, 2, \dots, M$) generate

$$X_{k+1} = P_{\Omega_{\xi_k}} X_k$$

- Prove that $\operatorname{inv}\mathcal{P}$ is nonempty (done, more or less).
- Prove that $\mu\mathcal{P}^k \to \pi \in \operatorname{inv}\mathcal{P}$ in an appropriate metric (done).

The following issues are open:

- Characterize $\operatorname{inv}\mathcal{P}$.
- Characterize the rate of asymptotic regularity: i.e. at what rate does $\operatorname{dist}_{W_2}\left(\mu\mathcal{P}^{k+1}, \mu\mathcal{P}^k\right) \to 0$?
- Characterize the rate of convergence: i.e. at what rate does $\operatorname{dist}_{W_2}\left(\operatorname{inv}\mathcal{P}, \mu\mathcal{P}^k\right) \to 0$?