Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

The Geometry of the Phase Retrieval Problem

Conference on Modern Challenges in Imaging
Tufts University, August 2019

Charles L. Epstein

Flatiron Institute of the Simons Foundation
and the University of Pennsylvania

August 8, 2019

Acknowledgment

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

I want to thank the organizers for inviting me to participate in this great event.

What I will discuss today is somewhat unconventional, ongoing work, at the intersection of pure mathematics, numerical analysis and mathematical physics.

This work was done jointly with Alex Barnett, Leslie Greengard, and Jeremy Magland and supported by the Centers for Computational Biology and Mathematics at the Flatiron Institute of the Simons Foundation, New York, NY.

Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

High Resolution Imaging

Today I will speak about inverse problems that arise in *Coherent Diffraction Imaging*. This is a technique that uses very high energy, monochromatic light, either electrons or x-rays, to form high resolution images of animate and inanimate materials. The goal is to achieve resolutions in the 1–10nm range.

What is Measured?

The photons used to illuminate the samples are in the 0.1-10KeV range; the light is typically assumed to be monochromatic, produced either by a laser or as synchrotron radiation. The object being imaged is many wavelengths across and the measurement is made very far from the object and light source, hence in the far field (the Fraunhofer regime). As is well known, if $\rho(\boldsymbol{x})$ describes the deviation of the object's refractive index from the vacuum, then the leading order term in the far field expansion of the scattered radiation is proportional to the square of the Fourier transform, $\widehat{\rho}(\boldsymbol{k})$, of ρ. At such small wavelengths, one can only measure the intensity of the scattered field, $|\widehat{\rho}(\boldsymbol{k})|^2$, and not its phase.

Note: this data is the Fourier transform of the autocorrelation function:

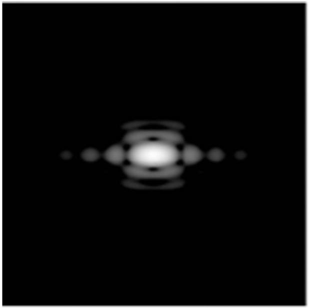
$$\rho \star \rho(\boldsymbol{x}) = \int_{\mathbb{R}^d} \rho(\boldsymbol{y})\rho(\boldsymbol{x}+\boldsymbol{y})d\boldsymbol{y}. \quad (1)$$

Some Physical Properties of ρ

For many x-ray energies and materials it is reasonable to assume that ρ is real valued, that is, there is no significant absorption. There are also x-ray bands where ρ can also be assumed to be real and non-negative. We often assume one, or both of these things.

Examples, I

Here are computed far field diffraction patterns produced by (a) a semi-circle, (b) an equilateral triangle.

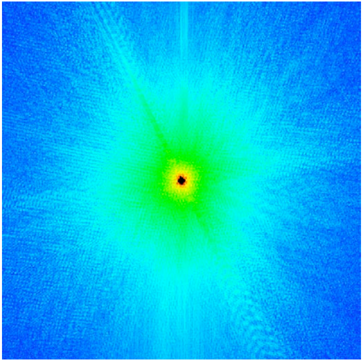


(a)

(b)

Examples, II

Here is an actual, but false colored, diffraction pattern produced by frozen-hydrated yeast spore at 520 eV. Courtesy E. Lima PhD thesis, Stony Brook U, from [8].



What is Needed?

We would like to reconstruct $\rho(\boldsymbol{x})$ from this measurement, but we need to, at least, estimate the unmeasured phase of $\widehat{\rho}(\boldsymbol{k})$. This is called the *phase retrieval problem*.

It also arises in x-ray crystallography, but is really quite a different problem for a periodic structure. In the context of a compact object, the possibility of solving this problem was first suggested by D. Sayre in 1952. He saw it as a consequence of Nyquist's sampling theorem! See [1].

Many attempts have been made to solve this problem, but with rather limited success. I'll first review some facts about the standard approaches to this problem, and then describe some novel methods for its solution.

Outline of this Lecture

The Geometry of the Phase Retrieval Problem

Charles L. Epstein

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

We will cover the following material:

- Auxiliary Data
- Standard Algorithms
- External Holography

Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

What is Measured

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

Without going into the details, if the measurement is made in the far field, then a classical calculation using Kirchoff's formula, shows that measurement can be interpreted as samples of $|\widehat{\rho}(\boldsymbol{k})|^2$, where $\rho(\boldsymbol{x})$ is the (frequency dependent) deviation of the refractive index from the vacuum caused by the object.
To have any hope of determining samples of $\widehat{\rho}(\boldsymbol{k})$ from these measurements, some additional information *must* be available.
The simplest, and most readily available information is an estimate on the **support** of the object $\rho(\boldsymbol{x})$. It is also sometimes assumed that $\rho(\boldsymbol{x}) \geq 0$, and has compact support.

This, or some other similar information, is referred to as **auxiliary information**, which is essential for this problem to be solvable, even in principle.

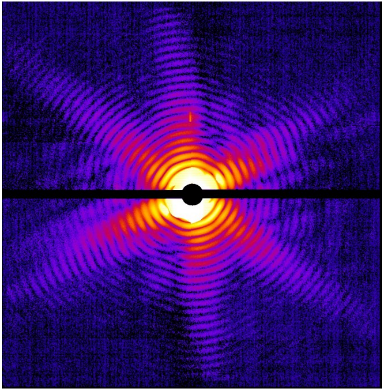
Measurement Issues

In fact, uniformly spaced samples collected on a planar detector do not lie on a uniformly spaced grid in $\boldsymbol{k}$-space. This has to do with the Ewald Sphere construction, which we do not have time to review. The maximum angle of scatter that impinges on the detector determines the highest spatial frequencies observed, and thereby the maximum resolution attainable in the reconstruction (even if we knew the phases).

In addition to the non-uniformity of the sampling, as a practical matter one needs to place a physical obstruction (called a beam stop) near the direct forward scattered direction, in order to prevent the detector from being destroyed by the intense forward scattered beam. This means that the measurements omit a neighborhood of $\boldsymbol{k} = 0$, where much of the energy in the Fourier data is located.

An Example

X-ray diffraction pattern of a single Mimivirus particle. Tomas Ekeberg, Uppsala University:



And we haven't even mentioned measurement noise, or the fact that the process of measurement destroys biological samples....

Simpler Model Problems

The phase retrieval problem encountered in coherent diffraction imaging is fraught with many practical difficulties. For the first part of the lecture we focus on a simpler model problem, which already proves very difficult to solve, is the *discrete, classical phase retrieval* problem.

At the end of the lecture we'll assume that we can sample the continuum Fourier transform of $\rho(\boldsymbol{x})$, with the goal of reconstructing a band-limited version of this function.

The Discrete Classical Phase Retrieval Problem

We imagine that the unknowns are samples of an object on a finite uniform grid:

$$\boldsymbol{x}_{\boldsymbol{j}} = \rho(j_1 \Delta x, j_2 \Delta x), \text{ where } \boldsymbol{j} \in J \subset \mathbb{Z}^2. \tag{2}$$

Here $J = [n_1 : N_1] \times [n_2 : N_2]$, is a rectangular grid. We call such collections of data, indexed by J, *images*. For the most part we restrict our discussion to real valued images, though most of what we say works just as well for complex valued images.

The Measurements

In our model problem, the measured data are the magnitudes of the DFT of these samples:

$$\{a_{\boldsymbol{j}} = |\hat{\boldsymbol{x}}_{\boldsymbol{j}}| : \boldsymbol{j} \in \widehat{J}\}, \tag{3}$$

where $\widehat{J}$ is a set of sample frequencies, with $|J| = |\widehat{J}|$. If $J = [0 : N-1]^d$, then

$$\hat{\boldsymbol{x}}_{\boldsymbol{j}} = \sum_{\boldsymbol{k}\in J} \boldsymbol{x}_{\boldsymbol{k}} e^{-\frac{2\pi i\, \boldsymbol{j}\cdot\boldsymbol{k}}{N}}. \tag{4}$$

There are important differences between these model measurements and the samples of a continuum Fourier transform that would actually be collected, which we don't have time to discuss.

The Magnitude Torus

The measured data $\{a_{\boldsymbol{j}}\}$ defines a torus in $\mathbb{R}^J$, which, in the Fourier representation, is a products of round circles in coordinate planes

$$\mathbb{A}_{\boldsymbol{a}} = \{\boldsymbol{x} \in \mathbb{R}^J : |\hat{\boldsymbol{x}}_{\boldsymbol{j}}| = a_{\boldsymbol{j}} \text{ for all } \boldsymbol{j} \in J\}. \tag{5}$$

We note that if we translate the image $\boldsymbol{x}^{(\boldsymbol{v})}_{\boldsymbol{j}} = \boldsymbol{x}_{\boldsymbol{j}-\boldsymbol{v}}$ or invert the image, $\check{\boldsymbol{x}}_{\boldsymbol{j}} = \boldsymbol{x}_{-\boldsymbol{j}}$, then the DFT magnitude data is unchanged. These are called "*trivial associates*". This makes the solution to the phase retrieval inherently non-unique.

The Support Condition, I

As noted above we need to have auxiliary information to be able to solve the phase retrieval problem. We imagine that the support of the unknown image $\boldsymbol{x}$, $S_{\boldsymbol{x}}$, is contained in a rectangular subset $R \subset J$ with the side lengths of R at most half those of J. In the literature it is often said that we "oversample," but really we just need to sample $\boldsymbol{k}$-space on a fine enough grid for the measured data to contain information about the support of $\boldsymbol{x}$.

Our auxiliary information will be an estimate for $S_{\boldsymbol{x}}$. This is a subset S with $S_{\boldsymbol{x}} \subset S \subset R \subset J$. We are therefore looking for an image, with the measured DFT magnitude data, with

$$\boldsymbol{x} \in B_S = \{\boldsymbol{x} : \boldsymbol{x}_{\boldsymbol{j}} = 0 \text{ for } \boldsymbol{j} \notin J\}. \tag{6}$$

This is a linear subspace. The standard formulation of the phase retrieval problem is therefore to find points in

$$\mathbb{A}_{\boldsymbol{a}} \cap B_S. \tag{7}$$

The Support Condition, II

For the purposes of analysis, this discrete model problem has a very important feature: it has exact solutions. If our measured data were instead finitely many samples of the continuum Fourier transform $\{\widehat{\rho}(\boldsymbol{k}_l) : l \in L\}$, (as in the real physical problem) then this would not uniquely specify (even up to translations and inversions) any function with compact support in a particular set. This means that one would first need to introduce a notion of complicated notion of approximate solution before any further analysis could be done....

Non-uniqueness, and other problems

- Unless $S = S_{\mathbf{x}}$, then this problem usually does not have a unique solution. Usually there are trivial associates that also have their support in S.
- There is a more insidious problem than this: unless $S = S_{\mathbf{x}}$, then the intersection between $\mathbb{A}_{\mathbf{a}}$ and B_S typically fails to be transversal. This is well known to make it difficult to find the intersection.
- There are other problems beyond this one: if $\mathbf{x}_0 \in \mathbb{A}_{\mathbf{a}} \cap B_S$, then there are usually many directions where $T_{\mathbf{x}_0}\mathbb{A}_{\mathbf{a}}$ and B_S make very small angles. The rates of convergence for all standard algorithms are determined by these angles. Smoother images lead to more small angles.

Non-uniqueness, and other problems

- Unless $S = S_{\mathbf{x}}$, then this problem usually does not have a unique solution. Usually there are trivial associates that also have their support in S.
- There is a more insidious problem than this: unless $S = S_{\mathbf{x}}$, then the intersection between $\mathbb{A}_{\mathbf{a}}$ and B_S typically fails to be transversal. This is well known to make it difficult to find the intersection.
- There are other problems beyond this one: if $\mathbf{x}_0 \in \mathbb{A}_{\mathbf{a}} \cap B_S$, then there are usually many directions where $T_{\mathbf{x}_0}\mathbb{A}_{\mathbf{a}}$ and B_S make very small angles. The rates of convergence for all standard algorithms are determined by these angles. Smoother images lead to more small angles.

Non-uniqueness, and other problems

- Unless $S = S_{\boldsymbol{x}}$, then this problem usually does not have a unique solution. Usually there are trivial associates that also have their support in S.
- There is a more insidious problem than this: unless $S = S_{\boldsymbol{x}}$, then the intersection between $\mathbb{A}_{\boldsymbol{a}}$ and B_S typically fails to be transversal. This is well known to make it difficult to find the intersection.
- There are other problems beyond this one: if $\boldsymbol{x}_0 \in \mathbb{A}_{\boldsymbol{a}} \cap B_S$, then there are usually many directions where $T_{\boldsymbol{x}_0}\mathbb{A}_{\boldsymbol{a}}$ and B_S make very small angles. The rates of convergence for all standard algorithms are determined by these angles. Smoother images lead to more small angles.

Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

Alternating Projection, I

The most classical approach to finding points in $\mathbb{A} \cap B_S$ is called *alternating projection*. In Functional Analysis it was introduced by von Neumann, and, in image reconstruction, by Gerchberg and Saxton, see [2]. We let $P_{B_S} : \mathbb{R}^J \to B_S$ be the orthogonal projection, and $P_{\mathbb{A}} : \mathbb{R}^J \to \mathbb{A}$ be the closest point map. This map is defined on the complement of a union of codimension 2 linear subspaces.

The idea for alternating projection is very simple. Choose a random collection of phases $\{e^{i\theta_{\boldsymbol{j}}} : \boldsymbol{j} \in \widehat{J}\}$ and use the inverse DFT to find the initial image $\boldsymbol{x}^{(0)} = \mathcal{F}^{-1}([\boldsymbol{a}_{\boldsymbol{j}} e^{i\theta_{\boldsymbol{j}}}])$, which is a random point on the magnitude torus.

Alternating Projection, II

Define the alternating projection sequence

$$x^{(n+1)} = P_{\mathbb{A}} \circ P_{B_s}(x^{(n)}). \tag{8}$$

It is clear that any point $x \in \mathbb{A} \cap B_S$ is a fixed point of this iteration, and it was hoped for many years that these iterates would converge to such a fixed point. In fact AP has many attracting fixed points that do not come from intersection points. These additional fixed points are local minima of the map from $\mathbb{A} \times B_S \to [0, \infty)$ defined by

$$d_{AB}(x, y) = \|x - y\|_2 \text{ (The Euclidean distance).} \tag{9}$$

From experiments with AP it is clear that there is a great abundance of such non-zero local minima.

Difference Maps, I

It has been known for a long time that algorithms based on the AP-map do not work very well. A different class of maps, which we call *difference maps*, were introduced to correct this problem. The first such algorithms were proposed by Fienup, using a map given by

$$D_{BA}^{\beta} = \boldsymbol{x} \mapsto \boldsymbol{x} + P_B[(1+\beta)P_A(\boldsymbol{x}) - \boldsymbol{x}] - \beta P_A(\boldsymbol{x}). \qquad (10)$$

Here $\beta \in (0, 1]$. This is call the “hybrid input-output (HIO) method;” there were later additions by Elser, Miao, etc. See [3], [4], [6], [7]. There are many variants, which all behave similarly.

Difference Maps, II

A representative case is given by the map:

$$D_{AB}(\boldsymbol{x}) = \boldsymbol{x} + P_A \circ R_B(\boldsymbol{x}) - P_B(\boldsymbol{x}), \tag{11}$$

which, in the notation of the previous slide, is D^1_{AB}. Here R_B is the "reflection around B" defined by

$$R_B(\boldsymbol{x}) = 2P_B(\boldsymbol{x}) - \boldsymbol{x}. \tag{12}$$

If B is a linear subspace, then R_B is just the usual orthogonal reflection with fixed point set equal to B.

The Fixed Point Set for Difference Maps

A easy calculation shows that $\boldsymbol{x}^*$ is a fixed point for D_{AB} if and only if

$$P_A \circ R_B(\boldsymbol{x}^*) = P_B(\boldsymbol{x}^*). \tag{13}$$

The fixed points do not necessarily belong to $\mathbb{A} \cap B_S$, but once a fixed point is found, then the point

$$\boldsymbol{x}^{**} = P_A \circ R_B(\boldsymbol{x}^*) = P_B(\boldsymbol{x}^*)$$

automatically does lie on the intersection. In our case the fixed points that "point to" $\boldsymbol{x}^{**}$ are a subset of

$$\mathcal{C}^{\boldsymbol{x}^{**}}_{\mathbb{A}_{\boldsymbol{a}} B_S} = N_{\boldsymbol{x}^{**}}\mathbb{A}_{\boldsymbol{a}} \cap B_S^{\perp}. \tag{14}$$

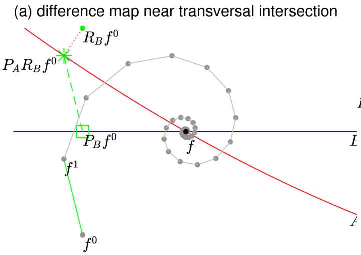
This is a subset of dimension about $|J|/4$, which we call the *center manifold*. It is important to note that $\boldsymbol{x}^{**}$ is essentially never an attracting fixed point itself. When these algorithms converge, they converge to points on the center manifold distant from the intersection that defines it.

Other Invariant Sets for Difference Maps

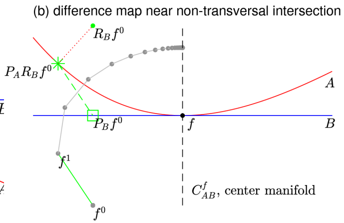
While it is true that the *fixed points* of difference maps are always related to points in $\mathbb{A} \cap B_s$, there are other invariant sets that are not. For example, if there exists points $\boldsymbol{x}_1 \in \mathbb{A}$ and $\boldsymbol{x}_2 \in B_s$, with $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, which are critical for $d_{\mathbb{A} B_S}$, the distance between $\mathbb{A}$ and B_S, then $\mathcal{C}_{\boldsymbol{x}_1 \boldsymbol{x}_2} = N_{\boldsymbol{x}_1} \mathbb{A} \cap B_S^{\perp}$ (as affine subspaces of $\mathbb{R}^J$) is non-empty. Many such critical points exist.

If $\mathcal{C}_{\boldsymbol{x}_1 \boldsymbol{x}_2} \cap \mathbb{A} \cap B_s = \emptyset$, then $\mathcal{C}_{\boldsymbol{x}_1 \boldsymbol{x}_2}$ is an invariant set without a fixed point. In examples, we have seen that this sort of set may be dynamically attracting. The map D_{AB} translates by a fixed vector along such an invariant set. One can study these maps when A and B are linear subspaces, to see that it is the angles between A and B that determine the rates of convergence. We will just look a some low dimensional examples.

Simple Examples of Difference Maps



(a) Difference Map with a transversal intersection.

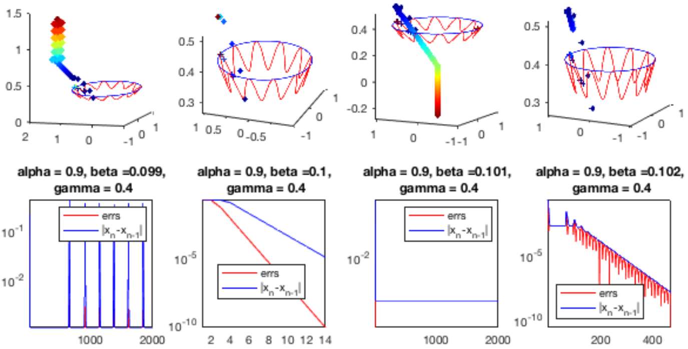


(b) Difference Map with a non-transversal intersection.

Figure: Cartoon of Difference Maps

Low Dimensional Examples for the HIO, I

We consider the effects of almost intersections, non-transversality, small angles and nearby intersections. Trajectories are ordered blue to red.



Near miss. Non-trans. int. 2 nearby ints. Trans. int.

Low Dimensional Examples, II

- The first images shows a near miss. There is a center manifold defined by the closest approach of A to B, which is attracting, and sends the trajectories spiraling out to infinity.
- The second example is a tangent intersection. Only a single coordinate needs to converge to zero. This is governed by a non-linear process that finds a point on the 2-dimensional center manifold near which this coordinate goes quickly to zero. This is not how phase retrieval works!
- The third example is the result of two very nearby intersection points. The center manifold defined by the local *maximum* of the d_{AB} between these points is attracting and, again, translates the iterates out to infinity.
- Finally a transversal intersection is attracting, but the angle between the two tangent spaces is very small, leading to slow, spiraling convergence.

Low Dimensional Examples, II

The Geometry of the Phase Retrieval Problem

Charles L. Epstein

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

- The first images shows a near miss. There is a center manifold defined by the closest approach of A to B, which is attracting, and sends the trajectories spiraling out to infinity.
- The second example is a tangent intersection. Only a single coordinate needs to converge to zero. This is governed by a non-linear process that finds a point on the 2-dimensional center manifold near which this coordinate goes quickly to zero. This is not how phase retrieval works!
- The third example is the result of two very nearby intersection points. The center manifold defined by the local *maximum* of the d_{AB} between these points is attracting and, again, translates the iterates out to infinity.
- Finally a transversal intersection is attracting, but the angle between the two tangent spaces is very small, leading to slow, spiraling convergence.

Low Dimensional Examples, II

- The first images shows a near miss. There is a center manifold defined by the closest approach of A to B, which is attracting, and sends the trajectories spiraling out to infinity.
- The second example is a tangent intersection. Only a single coordinate needs to converge to zero. This is governed by a non-linear process that finds a point on the 2-dimensional center manifold near which this coordinate goes quickly to zero. This is not how phase retrieval works!
- The third example is the result of two very nearby intersection points. The center manifold defined by the local *maximum* of the d_{AB} between these points is attracting and, again, translates the iterates out to infinity.
- Finally a transversal intersection is attracting, but the angle between the two tangent spaces is very small, leading to slow, spiraling convergence.

Low Dimensional Examples, II

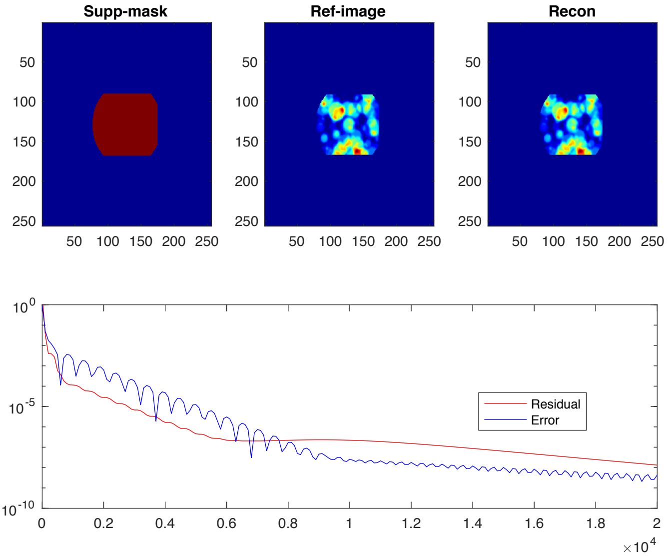
- The first images shows a near miss. There is a center manifold defined by the closest approach of A to B, which is attracting, and sends the trajectories spiraling out to infinity.
- The second example is a tangent intersection. Only a single coordinate needs to converge to zero. This is governed by a non-linear process that finds a point on the 2-dimensional center manifold near which this coordinate goes quickly to zero. This is not how phase retrieval works!
- The third example is the result of two very nearby intersection points. The center manifold defined by the local *maximum* of the d_{AB} between these points is attracting and, again, translates the iterates out to infinity.
- Finally a transversal intersection is attracting, but the angle between the two tangent spaces is very small, leading to slow, spiraling convergence.

The Geometry of the Phase Retrieval Problem
Charles L. Epstein
Introduction
Auxiliary Data
Standard Algorithms
External Holography
History of CDI
Bibliography

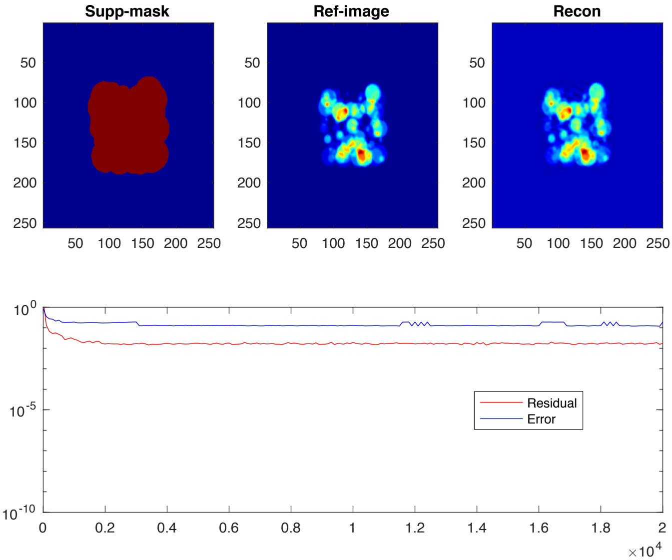
Using maps like $D_{\mathbb{A}B_S}$ to find points in $\mathbb{A} \cap B_S$ is the "industry" standard. Even with perfect data, these iterations rarely converge, but instead stagnate at points very distant from true limit points. The reconstructions from these stagnant orbits typically display a relative error in the 10^{-1} to 10^{-2} range, which, for many applications, is adequate.

This stagnation is caused by presence of many attracting basins, the non-transversailty of the intersections, $\mathbb{A} \cap B_S$, along with existence of many directions where the angles between $T_{\boldsymbol{x}}\mathbb{A}$ and B_S are very small. Algorithms that use positivity as auxiliary information tend to work somewhat better and are much less sensitive to the (softness) smoothness of the underlying object.

Phase Retrieval Example of the HIO, Hard Edge



Phase Retrieval Example of the HIO, Soft Edge



Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

External Holography, I

- We finish up today with a discussion of a different method, called *external holography* to reconstruct the phase, which entails a somewhat different protocol for collecting the data. It is easiest to describe these methods in the context of a continuum model.

- As is always possible in a problem that involves a physical measurement, one can beat the devil by **changing the way the measurement is made!**

External Holography, I

The Geometry of the Phase Retrieval Problem

Charles L. Epstein

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

- We finish up today with a discussion of a different method, called *external holography* to reconstruct the phase, which entails a somewhat different protocol for collecting the data. It is easiest to describe these methods in the context of a continuum model.

- As is always possible in a problem that involves a physical measurement, one can beat the devil by **changing the way the measurement is made!**

External Holography, II

In fact there are 3 different approaches to external holography. The unifying idea is to replace the unknown object $\rho(\boldsymbol{x})$, with $\rho(\boldsymbol{x}) + \varphi(\boldsymbol{x} - \boldsymbol{c})$, where φ is some sort of approximation to the δ-function, at least as regards diffraction of X-rays.

In all cases it is assumed that we know the object φ and the diffraction pattern it alone would produce. The vector $\boldsymbol{c}$ indicates where this object is centered; which of the 3 methods we are using depends on this location. In all cases the measured data consists of samples of $|\widehat{\rho}(\boldsymbol{k}) + e^{-2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}\widehat{\varphi}(\boldsymbol{k})|^2$.

The Far Field Method
$|\widehat{\rho}(\boldsymbol{k}) + e^{-2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}\widehat{\varphi}(\boldsymbol{k})|^2$.

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

If the choice of $\boldsymbol{c}$ locates the δ-like function quite far from the support of ρ, then using the measured data, we directly compute the autocorrelation function. It takes the form

$$\begin{aligned}&\rho\star\rho+\varphi\star\varphi+\int[\rho(\boldsymbol{y})\varphi(\boldsymbol{x}+\boldsymbol{y}-\boldsymbol{c})+\rho(\boldsymbol{x}+\boldsymbol{y})\varphi(\boldsymbol{y}-\boldsymbol{c})]d\boldsymbol{y}\\ &\approx\rho\star\rho+\varphi\star\varphi+\rho(\boldsymbol{x}-\boldsymbol{c})+\rho(\boldsymbol{x}+\boldsymbol{c}).\end{aligned} \tag{15}$$

If $\|\boldsymbol{c}\|$ is large compared to the diameter of the support of ρ, then the supports of translates, $\rho(\boldsymbol{x}-\boldsymbol{c})$, $\rho(\boldsymbol{x}+\boldsymbol{c})$ are disjoint from each other and that of $\rho\star\rho+\varphi\star\varphi$. Hence these terms can simply be read off. In actual practice one needs to 'deconvolve' φ to get a sharp image. This method appears in a paper of Hohage et al.

The Mid-field Method $|\widehat{\rho}(\boldsymbol{k}) + e^{-2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}\widehat{\varphi}(\boldsymbol{k})|^2$.

For this method we assume that the support of $\varphi(\boldsymbol{x}-\boldsymbol{c})$ is at a moderate distance from that of $\rho(\boldsymbol{x})$. If we have a reasonable estimate for the support of ρ, and the object φ has a sharp edge, with its shape known to high precision, then the degeneracies of difference-map method can be avoided. Even if $\boldsymbol{c}$ is not known, a priori, then, using a fairly standard HIO algorithm, this data can be used quite successfully to reconstruct a high resolution image of ρ.

The Near-field Method, I
$\mathcal{H}(f)=\mathcal{F}^{-1}[-i\,\mathrm{sgn}(\xi)\widehat{f}(\xi)]$

If the support of $\varphi(\boldsymbol{x}-\boldsymbol{c})$ is just outside of, but very near to the support of ρ, then an entirely different algorithm which relies on the properties of the Hilbert transform, can be used. This method works in all dimensions; we first consider the $1d$-case.

Recall that the Hilbert transform is defined for a function $f\in L^2(\mathbb{R})$ by

$$\mathcal{H}(f)=\mathcal{F}^{-1}[-i\,\mathrm{sgn}(\xi)\widehat{f}(\xi)]. \tag{16}$$

The functions $f(x)\pm i\mathcal{H}(f)(x)$ are the boundary values of functions, $F_{\pm}(z)$, holomorphic in the upper (lower) half planes respectively.

The Near-field Method, II

- If $g(x)$ is a non-vanishing function, defined on $\mathbb{R}$ and $|g(x)| \to 1$ as $x \to \pm\infty$ sufficiently rapidly so that $f(x) = \log|g(x)|$ belongs to L^2, then the Hilbert transform, $\mathcal{H}(f)$ is well defined.
- If g is also the boundary value of a non-vanishing holomorphic function in the upper (or lower) half plane, so that $|g|$ tends to 1 at infinity, then

$$\mathcal{H}(f)(x) = \arg[\log g(x)]. \tag{17}$$

- It follows from the Paley-Wiener theorem, that since ρ has compact support, we can use an external object to directly determine the $\arg[\log \widehat{\rho}]$, which is equivalent to knowing the phase of $\widehat{\rho}$.

The Near-field Method, II

- If $g(x)$ is a non-vanishing function, defined on $\mathbb{R}$ and $|g(x)| \to 1$ as $x \to \pm\infty$ sufficiently rapidly so that $f(x) = \log|g(x)|$ belongs to L^2, then the Hilbert transform, $\mathcal{H}(f)$ is well defined.
- If g is also the boundary value of a non-vanishing holomorphic function in the upper (or lower) half plane, so that $|g|$ tends to 1 at infinity, then

$$\mathcal{H}(f)(x) = \arg[\log g(x)]. \tag{17}$$

- It follows from the Paley-Wiener theorem, that since ρ has compact support, we can use an external object to directly determine the $\arg[\log \widehat{\rho}]$, which is equivalent to knowing the phase of $\widehat{\rho}$.

The Near-field Method, II

- If $g(x)$ is a non-vanishing function, defined on $\mathbb{R}$ and $|g(x)| \to 1$ as $x \to \pm\infty$ sufficiently rapidly so that $f(x) = \log|g(x)|$ belongs to L^2, then the Hilbert transform, $\mathcal{H}(f)$ is well defined.
- If g is also the boundary value of a non-vanishing holomorphic function in the upper (or lower) half plane, so that $|g|$ tends to 1 at infinity, then

$$\mathcal{H}(f)(x) = \arg[\log g(x)]. \tag{17}$$

- It follows from the Paley-Wiener theorem, that since ρ has compact support, we can use an external object to directly determine the $\arg[\log \widehat{\rho}]$, which is equivalent to knowing the phase of $\widehat{\rho}$.

The Near-field Method, $1d$-case

For simplicity lets assume that $\varphi(x) = M\delta(x)$, then the measured data is

$$|\widehat{\rho}(k) + Me^{-2\pi ick}| = M\left|e^{2\pi ick}\frac{\widehat{\rho}(k)}{M} + 1\right|. \tag{18}$$

If M is large enough, and c lies outside the support of ρ, then the function $\left|e^{-2\pi ick}\frac{\widehat{\rho}(k)}{M} + 1\right|$ satisfies the hypotheses for g on the previous slide. Hence

$$\mathcal{H}\left[\log\left|e^{2\pi ick}\frac{\widehat{\rho}(k)}{M} + 1\right|\right] = \arg\left[\log\left(e^{2\pi ick}\frac{\widehat{\rho}(k)}{M} + 1\right)\right], \tag{19}$$

from which we can easily compute $\widehat{\rho}(k)$.

The Near-field Method, $2d$-case

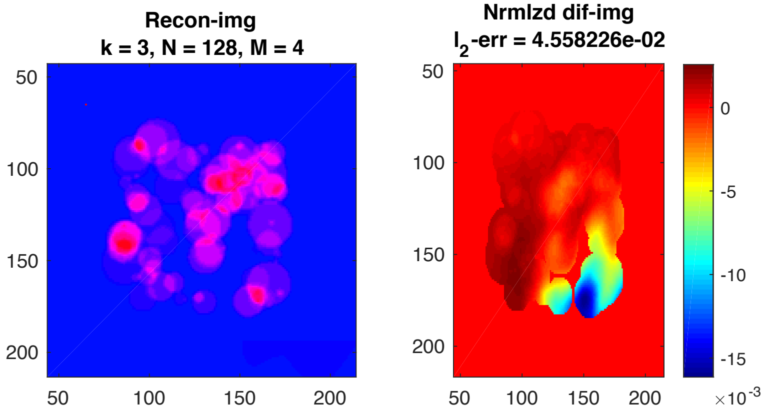
To extend this to the $2d$-case, we suppose that the unknown object takes the form $M\delta(x_1 - c_1, x_2 - c_2) + \rho(x_1, x_2)$. The measurement is then proportional to

$$|\widehat{\rho}(\boldsymbol{k}) + Me^{-2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}| = M\left|e^{2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}\frac{\widehat{\rho}(\boldsymbol{k})}{M} + 1\right|. \tag{20}$$

We can then apply the 1-variable Hilbert transform to $\log\left|e^{2\pi i \boldsymbol{c}\cdot\boldsymbol{k}}\frac{\widehat{\rho}(\boldsymbol{k})}{M} + 1\right|$ in one of the variables k_1 or k_2, with the other variable simply a parameter. Which variable is which depends on the location of $\boldsymbol{c}$ relative to the support of $\rho(\boldsymbol{x})$.

Of course we cannot make a δ-function, but it turns out not to be necessary. In the last slides we show numerical experiments.

Numerical Examples, I



Here $\varphi(\boldsymbol{x}) = \delta(\boldsymbol{x})$.

Theory with non-δ Object

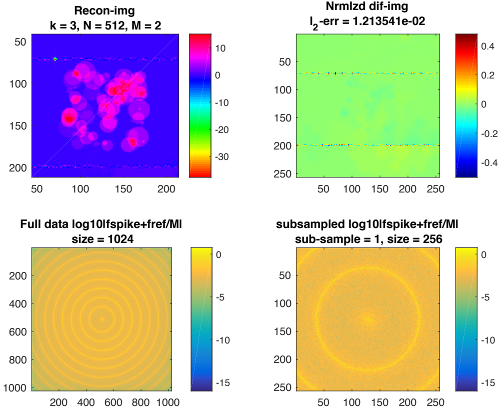
To employ a non-δ-like $\varphi(\boldsymbol{x})$, we need to be able to compute the quantity in the $|\cdot|$ on right hand side of:

$$|Me^{-2\pi i \boldsymbol{c}\cdot\boldsymbol{\xi}}\widehat{\varphi}(\boldsymbol{\xi})+\widehat{f}(\boldsymbol{\xi})|^2 = |M\widehat{\varphi}(\boldsymbol{\xi})|^2\cdot\left|1+M^{-1}e^{2\pi i \boldsymbol{c}\cdot\boldsymbol{\xi}}\frac{\widehat{f}(\boldsymbol{\xi})}{\widehat{\varphi}(\boldsymbol{\xi})}\right|^2. \tag{21}$$

For the theory outlined above to apply, the function in $|\cdot|$, on the right hand side, should be analytic in either the upper or lower half plane, and tend to 1 at ∞.

This condition can be considerably relaxed in practice. In the next example $\widehat{\varphi}(\boldsymbol{\xi}) = C\frac{1}{(R|\boldsymbol{\xi}|)^2}\left[\frac{\sin(R|\boldsymbol{\xi}|)}{R|\boldsymbol{\xi}|} - \cos(R|\boldsymbol{\xi}|)\right]$, which has infinitely many real zeros!

Numerical Examples, II

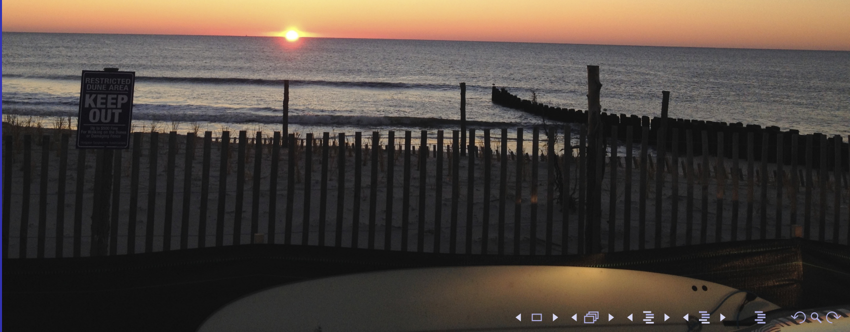


Here $\varphi(\boldsymbol{x}) = \sqrt{R^2 - \|\boldsymbol{x}\|^2}\chi_{[0,R]}(\|\boldsymbol{x}\|)$, which is the x-ray transform of a uniform spherical ball or radius R.

Thanks!

Thanks for your attention!

And thanks to Flatiron Institute of the Simons Foundation for supporting this research.



Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

The Geometry of the Phase Retrieval Problem

Charles L. Epstein

Introduction

Auxiliary Data

Standard Algorithms

External Holography

History of CDI

Bibliography

This slide courtesy Malcolm Howells, ESRF

http://www.esrf.eu/files/live/sites/www/files/events/conferences/Tutorials/slideslecture7.pdf

COHERENT X-RAY DIFFRACTION IMAGING: HISTORY

ESRF

Sayre (1952) - Fundamentals of sampling the wave amplitude and wave intensity

Gerchberg and Saxton (1972) - First phase-retrieval algorithm successful on test data

Sayre (1980) - Idea to do "crystallography" with non-periodic objects (i. e. attempt phase retrieval) and exploit the cross-section advantage of soft x-rays

Sayre, Yun, Chapman, Miao, Kirz (1980's and 1990's) - development of the experimental technique

Fienup (1978-) - Development of practical phase-retrieval algorithms including use of the combination of support constraint and oversampling of the amplitude pattern

Miao, Charalambous, Kirz and Sayre (1999) - first demonstration of 2-D CXDM using a Fienup-style algorithm at 0.73 keV x-ray energy, 75 nm resolution

Miao et al (2000) - imaging of a fixed biological sample in 2-D at 30 nm resolution

Miao et al (2001) - improved resolution in 2-D: 7 nm
achievement of 3-D with moderate resolution: 55 nm

Robinson et al (2001-3) - Application to microcrystals and defects - 3D reconstruction - hardest x-rays

ALS group 2002-5 - reconstruction without use of other micoscopes - 3D reconstructions with many (up to 280) views and 10 nm resolution

Outline

1 Introduction

2 Practical Measurements and Auxiliary Data

3 Standard Algorithms

4 External Holography

5 Short History of CDI

6 Bibliography

Some References

1. D. SAYRE, *Some implications of a theorem due to Shannon*, Acta Crystallogr., vol. 5, p. 843, 1952.
2. R. W. GERCHBERG AND W. O. SAXTON, *Practical algorithm for determination of phase from image and diffraction plane pictures*, Optik, vol. 35, pp. 237-246, 1972.
3. J.R.FIENUP, *Reconstruction of an object from the modulus of its Fourier transform*, Opt. Lett, vol. 3, pp. 27-29, 1978.
4. J.R.FIENUP, *Phase retrieval algorithms: a comparison*, 1982, Applied Optics. 21 (15) 2758-2769.
5. M. H. HAYES, *The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform*, IEEE Trans. Accoust., Speech Signal Process., vol. ASP-30, no. 2, pp. 140-154, Apr. 1982.
6. V. ELSER, *Solution of the crystallographic phase problem by iterated projections*, Acta Crystallogr. A, vol. 59, pp. 201-209, 2003.
7. JIANWEI MIAO, RICHARD L. SANDBERG, AND CHANGYONG SONG, *Coherent X-ray Diffraction Imaging*, IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS, DOI:10.1109/JSTQE.2011.2157306, 2011. **Contains a comprehensive bibliography on CXDI.**
8. MALCOLM HOWELLS, *Principles and Practice of Coherent X-ray Diffraction Imaging*, lecture notes, http://www.esrf.eu/files/live/sites/www/files/events/conferences/ Tutorials/slideslecture7.pdf, 2014.