Gerchberg-Saxton and Misell-Gerchberg-Saxton phase retrieval

Anand Sivaramakrishnan

Wednesday, January 16 2019

Gerchberg-Saxton phase retrieval

The GS algorithm starts with nominally in-focus image data, and a known pupil geometry (even a throughput map for an apodized pupil).

We seed it with an initial guess at the phase over pupil. It then steps between pupil and image planes, enforcing the known pupil and the known image intensity constraints (as appropriate). A GS iteration goes from one estimate of the pupil phase to the next, even if is sent through a forward transform to image, then an inverse transform back to pupil.

If you want a wider view of GS you can skim the very readable Section 3.1 of Osherovich's 2012 Ph.D. thesis for an overview of GS and three similar methods that Fienup developed. Osherovich describes Fienup's GS modification (the Hybrid Input-Output algorithm), which is more robust than GS. Figure 1.3 (p.7) of the thesis is amusing and quite educational.

A monochromatic, better-than-Nyquist image sampling version is presented in the class exercises. For undersampled broadband images, see the article by **Jim Fienup**.

Gerchberg-Saxton loop

- Initial data: Pupil transmission array $A(x)$, real
- Initial data: Image intensity array $I(k) >= 0$, $a(k) \equiv I(k)^{\frac{1}{2}}$
- Initial data: A guess at pupil phase, $\phi_0(x)$

From these initial data and guesses one iteration of the GS loop brings you back to the pupil plane, with a next guess at the phase. The loop always converges, though not always to the correct phase. It can stagnate or converge to a different phase. If the pupil has point symmetry there are two possible "correct" phase solutions.

Monitor the changes in the phase guess until they are small. Try different initial guesses, see if you get the same final phase after convergence. If so that's good.

Fig. 1 from Gerchberg & Saxton 1972

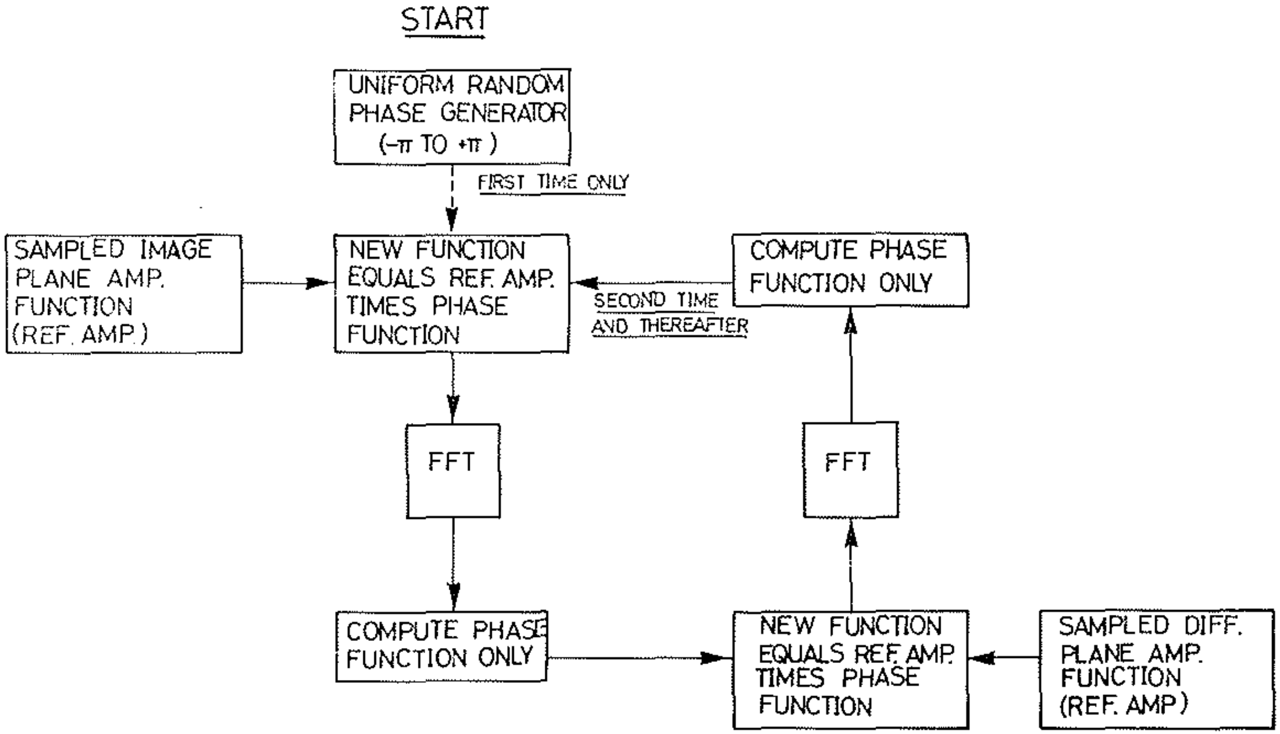

Fig. 1. Schematic drawing of phase determining algorithm.

GS loop flow in theory:

- Calculate image field: $a_0(k)\, e^{i\psi_0(k)} = \mathsf{FT}(A(x)\, e^{ik\phi_0(x)})$
- Enforce image intensity constraint: $a(k)\, e^{i\psi_0(k)}$
- Go back to pupil: $A_1(x)\, e^{i\phi_1(x)} = \mathsf{IFT}(a(k)\, e^{i\psi_0(k)})$
- Enforce pupil constraint: $A(x)\, e^{i\phi_1(x)}$
- Change in phase estimate: $E^2 = \dfrac{\int\int\ A(x)\,(\phi_1(x) - \phi_0(x))^2\ dx}{\int\int\ A(x)\ dx}$
- Are we done yet? Is $E^2 \leq E^2_{\mathsf{requirement}}$?
- NO: go through another loop, calculate $\phi_2(x)$, ...
- YES: Done. Or retry with several different initial phase guesses to be sure.

Real details to think about:

- Is your image data noisy? Low-pass filter out noise?
- Got negative intensity data? Zero it out?
- Differencing phase arrays: Zero-mean the differences. De-tilt?
- Missing/bad pixel data: Filters out easily with finely sampled data. More of a problem with Nyquist and coarser data (*e.g.*, NIRCam at 2um)
- How will incorrect pupil geometry show? Incorrect pixel scale? Incorrect wavelength?
- Finite pupil support leads to infinite image extent. Think about spatial filtering in the GS loop context.

Final comments on GS:

Even in the absence of noise the GS algorithm can converge to a radically wrong answer. To be more confident of the answer we get from it, we can initialize our pupil phase guess with various random phase arrays over the pupil and see if one phase answer comes out most often.

We can also try to start it close to the final answer if we know something about the situation (e.g. the Hubble case during science ops, but not during the earliest days of unknown aberration, before HST spherical aberration was characterized).

Or we can 'parametrize' the way we allow the iterations to select the next estimate of the phase given our 'current' estimate of the phase over pupil. Only vary the lowest order few Zernikes of the phase, by filtering the suggested 'next iteration of the pupil phase' to only change focus, or astigmatism, or a few low order Zernikes, in the next iteration of the pupil phase.

Misell-Gerchberg-Saxton

focus-diverse phase retrieval

Defocus senses curvature

Wednesday, January 23 2019

MGS focus-diverse PR loop (after J. E. Krist)

The MGS algorithm starts with two (or more) out-of-focus images, $I_\alpha(k)$ and $I_\beta(k)$, a known pupil geometry $A(x)$ as a **pupil field constraint**, and an **initial guess** at the phase $\phi_0(x)$ on the pupil.

Image field constraints are $a_\alpha(k) \equiv \sqrt{I_\alpha(k)}$ and $a_\beta(k) \equiv \sqrt{I_\beta(k)}$.

Conceptually we induce defocus by adding phase using hardware in the pupil plane, so we do not change the pixel scale of the images. If we move the detector to induce defocus the pixel scales of the two images will not be the same. Why is that? We can get around it but it's more detailed work.

We will use two out-of-focus images that correspond to adding parabolic phases $\alpha(x)$ and $\beta(x)$ to the (unknown) in-focus pupil phase $\phi(x)$.

It is advantageous if we can arrange that $\beta(x) = -\alpha(x)$: the amounts of defocus between the pair of images is the same, so they are either side of best focus. This symmetry mitigates non-uniform pupil illumination effects on wavefront estimate errors, apparently.

1. We add the defocus $\alpha(x)$ to our initial pupil phase $\phi_0(x)$.

2. We propagate this field, $A(x)e^{i\phi_0(x)+\alpha(x)}$, to the appropriate defocussed image plane with a Fourier transform. Our calculated image field is now $a_{0,\alpha}(k)e^{i\psi_{0,\alpha}}(k)$.

3. We apply one image field constraint: $a_\alpha(k)$ replaces the "amplitude spread function" of this propagated wave in the first plane in which we have image data. The image field is now $a_\alpha(k)e^{i\psi_{0,\alpha}}(k)$.

4. We inverse transform this modified image field back to pupil plane to get $A_{0,\alpha}(k)e^{i\phi_{0,\alpha}(k)}$. We modify this two ways: we enforce the pupil constraint by replacing $A_{0,\alpha}(k)$ by $A_(k)$. Then we subtract off the defocus phase $\alpha(x)$ from the pupil phase. This gives us an estimate of the pupil phase partway through the loop, $\phi_{0,\alpha}(x) - \alpha(x)$.

5. We then add in the second defocus, $\beta(x)$, to this phase estimate and transform the pupil field $A(x)e^{i\phi_{0,\alpha}(x)-\alpha(x)+\beta(x)}$ to the second defocussed image plane, $a_{0,\beta}(k)e^{i\psi_{0,\beta}(k)}$.

6. We apply our second image field constraint, $a_\beta(k)$. The modified image field becomes $a_\beta(k)e^{i\psi_{0,\beta}(k)}$.

7. We inverse Fourier transform the modified image field back to pupil, to get a pupil field $A_{0,\beta}(x)e^{i\phi_{0,\beta}(x)}$.

8. We subtract off the second image's defocus phase $\beta(x)$ from this pupil field's phase, yielding our new estimate of the pupil phase, $\phi_1(x) = \phi_{0,\beta}(x) - \beta(x)$.

9. We examine the difference between our earlier estimate $\phi_0(x)$ and the revised estimate $\phi_1(x)$, and decide whether we need to perform iteration or stop with this estimate.

Real details to think about:

- All the same ones as for GS, and...
- Images do not have the same centroid locations.
- Assumed defocus values are inexact. Residual focus shows in result..
- Choice of (multiple) defocus values. Dean & Bowers JOSA A 2003.
- Pupil illumination is non-uniform (or wrong). Use equal +/- defocus. See Krist & Burrows p4853 middle of RH column, and Figure 2.
- Strongly-illuminated PSF area is about $8\lambda/D$ diameter (Fienup gave me this rule-of-thumb). Impacts exposure times/saturation.

Dean & Bowers key points:

- The simple theory in this paper applies to "infinite apertures" - the simple but intuition-developing analytical Taylor expansion calculations neglect the aperture geomety itself, but highlight the way small ripple aberrations show up in defocussed images. A finite telescope aperture improves the ability to recover information from the theoretical worst case setup when a defocussed imgae has no information on the size of a certain spatial frequency phase ripple.

- Equation 8 describes the contrast in the image resulting from a ripple aberration in phase on the primary, at a given defocus. This is a shorthand for the information a given defocussed image has on a given spatial frequency (think "number of ripples across D") on the primary.

- Equation 10 gives you the conditions that hold when your defocussed image has the clearest information on the spatial frequency of a given sinusoidal phase aberration.

- Equation 12 tells us that for a given spatial frequency of aberation we cannot exceed a certain defocus - we will just not get any signal on that spatial frequency in our defocussed images.

- Dean & Bowers summary & discussion section is very helpful to practitioners.