# Maximum Entropy-type Methods, Projections, and (Non-)Convex Programming

Prepared for

**CARMA Colloquium**

March 12, 2009

... ... ...


Jonathan M. Borwein, FRSC

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

School of Math and Phys Sciences, Univ. of Newcastle, NSW

Faculty of Computer Science, Halifax, NS Canada

DALHOUSIE UNIVERSITY Inspiring Minds

URL: www.cs.dal.ca/∼jborwein



Revised: March 7, 2009

# BRAGG and BLAKE

*"I feel so strongly about the wrongness of reading a lecture that my language may seem immoderate. ··· The spoken word and the written word are quite different arts. ··· I feel that to collect an audience and then read one's material is like inviting a friend to go for a walk and asking him not to mind if you go alongside him in your car."*

**Sir Lawrence Bragg** (1890-1971)
Nobel Crystallographer (Adelaide)

Songs of Innocence and Experience (1825)
(We are both.)

## SANTAYANA

*"If my teachers had begun by telling me that mathematics was pure play with presuppositions, and wholly in the air, I might have become a good mathematician. But they were overworked drudges, and I was largely inattentive, and inclined lazily to attribute to incapacity in myself or to a literary temperament that dullness which perhaps was due simply to lack of initiation."*

(**George Santayana**)

Persons and Places, 1945, 238–9.

**TWO FINE REFERENCES:**

1. J.M. Borwein and Qiji Zhu, *Techniques of Variational Analysis*, CMS/Springer-Verlag, New York, 2005.

2. J.M. Borwein and A.S Lewis, *Convex Analysis and Nonlinear Optimization*, CMS/Springer-Verlag, 2nd expanded edition, New York, 2005.

# OUTLINE

I shall discuss in "tutorial mode" the formalization of inverse problems such as signal recovery and option pricing as (convex and non-convex) optimization problems over the infinite dimensional space of signals. I shall touch on* the following:

**1.** The impact of the choice of "entropy" (e.g., Boltzmann-Shannon, Burg entropy, Fisher information) on the *well-posedness* of the problem and the form of the solution.

**2.** Convex programming duality: what it is and what it buys you.

**3.** Algorithmic consequences.

**4.** Non-convex extensions: life is hard. But sometimes more works than should.

♠ Related papers at `http://docserver.cs.dal.ca/`

*More is an unrealistic task!

# THE GENERAL PROBLEM

• Many applied problems reduce to "**best**" solving (under-determined) systems of linear (or non-linear) equations $\boxed{Ax = b}$, where $b \in \mathbb{R}^n$, and the unknown $x$ lies in some appropriate function space.

Discretization reduces this to a finite-dimensional setting where $A$ is now a $m \times n$ matrix.

◇ In many cases, I believe it is better to address the problem in its function space home, discretizing only as necessary for computation.

• Thus, the problem often is *how do we estimate $x$ from a finite number of its 'moments'*? This is typically an under-determined linear inversion problem where the unknown is most naturally a function, not a vector in $\mathbb{R}^m$.

# EXAMPLE 1. AUTOCORRELATION

• Consider, extrapolating an *autocorrelation function* $R(t)$ given sample measurements.

◇ The Fourier transform $S(z)$ of the autocorrelation is the power spectrum of the data.

> Fourier moments of the power spectrum are the same as samples of the autocorrelation function, so by computing several values of $R(t)$ directly from the data, we are in essence computing moments of $S(z)$.

• We compute a finite number of moments of $S$, and estimate $S$ from them, and may *compute more moments* from the estimate $\widehat{S}$ by direct numerical integration.

• Thereby *extrapolating* $R$, without directly computing $R$ from potentially noisy data.

# THE ENTROPY APPROACH

• Following (B-Zhu) I sketch a maximum entropy approach to under-determined systems where the unknown, $x$, is a function, typically living in a *Hilbert space*, or more general space of functions.

> This technique picks a "best" representative from the infinite set of *feasible* functions (functions that possess the same $n$ moments as the sampled function) by minimizing an integral functional, $f$, of the unknown.

http://projects.cs.dal.ca/ddrive

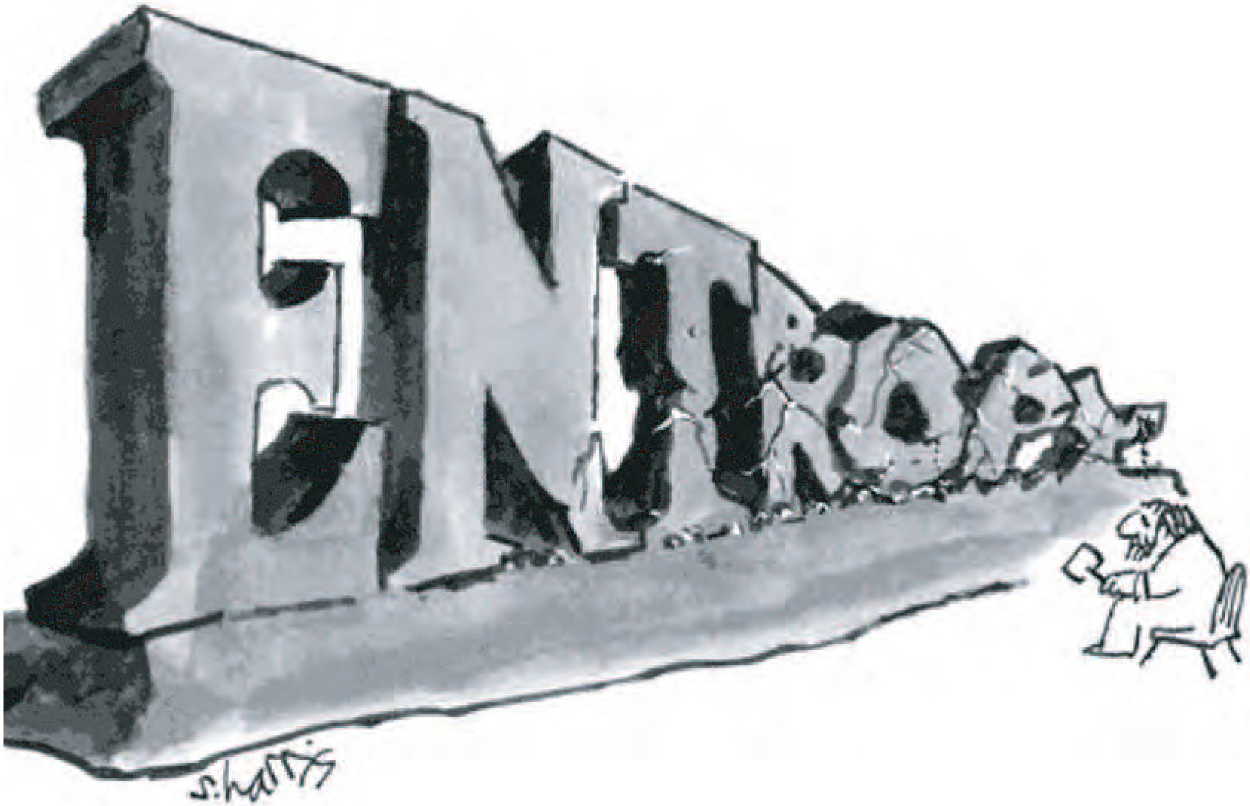
⋄ The approach finds applications in countless fields including:

> Acoustics, constrained spline fitting, image reconstruction, inverse scattering, optics, option pricing, multidimensional NMR, tomography, statistical moment fitting, and time series analysis, etc.
>
> (Many thousands of papers)

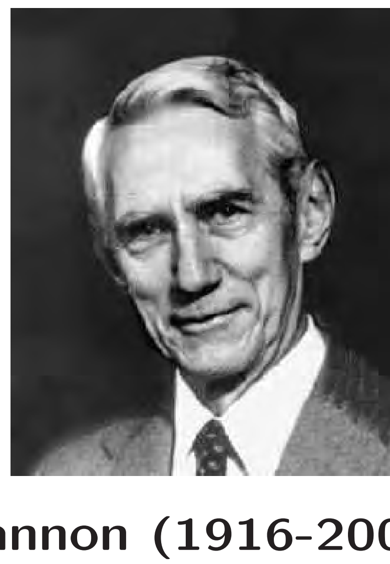
• However, the derivations and mathematics are fraught with subtle errors.

> I will discuss some of the difficulties inherent in infinite dimensional calculus, and provide a simple theoretical algorithm for correctly deriving maximum entropy-type solutions.

## WHAT is




**Boltzmann (1844-1906)**



**Shannon (1916-2001)**

# WHAT is ENTROPY?

*Despite the narrative force that the concept of entropy appears to evoke in everyday writing, in scientific writing entropy remains a thermodynamic quantity and a mathematical formula that numerically quantifies disorder. When the American scientist Claude Shannon found that the mathematical formula of Boltzmann defined a useful quantity in information theory, he hesitated to name this newly discovered quantity entropy because of its philosophical baggage.*

*The mathematician John von Neumann encouraged Shannon to go ahead with the name entropy, however, since "no one knows what entropy is, so in a debate you will always have the advantage."*

- **19C**: **Boltzmann**—thermodynamic *disorder*
- **20C**: **Shannon**—information *uncertainty*
- **21C**: **JMB**—potentials with *superlinear growth*

# CHARACTERIZATIONS of ENTROPY

- Information theoretic characterizations abound. A nice one is:

**Theorem** $H(\overrightarrow{p}) = -\sum_{k=1}^{N} p_k \log p_k$ is *the unique continuous function* (up to a positive scalar multiple) on finite probabilities such that

I. Uncertainty grows:

$$H\left(\overbrace{\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}}^{n}\right)$$

increases with $n$.

II. Subordinate choices are respected: for distributions $\overrightarrow{p_1}$ and $\overrightarrow{p_2}$ and $0 < p < 1$,

$$H\left(p\,\overrightarrow{p_1}, (1-p)\,\overrightarrow{p_2}\right) = p\,H(\overrightarrow{p_1}) + (1-p)\,H(\overrightarrow{p_2}).$$

## ENTROPIES FOR US

• Let $X$ be our *function space*, typically Hilbert space $L^2(\Omega)$, or the function space $L^1(\Omega)$ (or a Sobelov space).

◇ For $p \geq 1$,

$$L^p(\Omega) = \left\{ x \text{ measurable } : \int_\Omega |x(t)|^p dt < \infty \right\}.$$

It is well known that $L^2(\Omega)$ is a Hilbert space with *inner product*

$$\langle x, y \rangle = \int_\Omega x(t) y(t) dt,$$

(with variations in Sobelov space).

• A *bounded linear map* $A : X \to \mathbb{R}^n$ is determined by

$$(Ax)_i = \int x(t) a_i(t)\, dt$$

for $i = 1, \ldots, n$ and $a_i \in X^*$ the 'dual' of $X$ ($L^2$ in the Hilbert case, $L^\infty$ in the $L^1$ case).

• To pick a solution from the infinitude of possibilities, we may freely define "**best**".

⊗ The most common approach is to find the minimum norm solution*, by solving the *Gram system*

$$\boxed{AA^T\lambda = b}\ .$$

⊕ The solution is then $\hat{x} = A^T\lambda$. This recaptures all of *Fourier analysis*!

• This actually solved the following *variational problem*:

$$\inf\left\{\int_\Omega x(t)^2 dt : Ax = b \ \ x \in X\right\}.$$

*Even in the (realistic) infeasible case.

• We generalize the norm with a *strictly convex functional* $f$ as in

$$\min\{f(x) : Ax = b, \ \ x \in X\}, \qquad (P)$$

where $f$ is what we call, an *entropy functional*, $f : X \to (-\infty, +\infty]$. Here we suppose $f$ is a strictly convex integral functional* of the form

$$f(x) = \int_{\Omega} \phi(x(t))dt.$$

The functional $f$ can be used to include other constraints†.

For example, the constrained $L^2$ norm functional ('positive energy'),

$$f(x) = \begin{cases} \int_0^1 x(t)^2\,dt & \text{if } x \geq 0 \\ +\infty & \text{else} \end{cases}$$

is used in constrained *spline fitting*.

• Entropy constructions abound: *Bregman* and *Csizar distances* model statistical divergences.

*Essentially $\phi''(t) > 0$.

†Including nonnegativity, by appropriate use of $+\infty$.

• Two popular choices for $f$ are the *Boltzmann-Shannon* entropy (in image processing)

$$f(x) = \int x \log x,$$

and the *Burg entropy* (in time series analysis),

$$f(x) = -\int \log x.$$

◇ *Both implicitly impose a nonnegativity constraint* (positivity in Burg's non-superlinear case).

• There has been much information-theoretic debate about which entropy is best.

This is more theology than science!

• More recently, the use of *Fisher Information*

$$f(x, x') = \int_\Omega \frac{x'(t)^2}{2x(t)} \mu(dt)$$

has become more usual as it *penalizes* large derivatives; and can be argued for physically ('hot' over past five years).

# WHAT 'WORKS' BUT CAN GO WRONG?

• Consider solving $Ax = b$, where, $b \in \mathbb{R}^n$ and $x \in L^2[0,1]$. Assume further that $A$ is a continuous linear map, hence represented as above.

• As $L^2$ is infinite dimensional, so is $N(A)$: if $Ax = b$ is solvable, it is under-determined.

We pick our solution to *minimize*

$$f(x) = \int \phi(x(t))\,\mu(dt)$$

⊙ $\phi(x(t), x'(t))$ in Fisher-like cases [BN1, BN2, B-Vanderwerff (*Convex Functions*, CUP 2009)].

• We introduce the *Lagrangian*

$$L(x,\lambda) := \int_0^1 \phi(x(t))dt + \sum_{i=1}^n \lambda_i \left(b_i - \langle x, a_i \rangle\right),$$

and the associated *dual problem*

$$\max_{\lambda \in \mathbb{R}^n} \min_{x \in X} \{L(x,\lambda)\}. \qquad (D)$$

• So we formally have a "dual pair" (BL1)

$$\min\left\{f(x) : Ax = b,\ \ x \in X\right\}, \qquad (P)$$

and

$$\max_{\lambda \in \mathbb{R}^n} \min_{x \in X}\{L(x,\lambda)\}. \qquad (D)$$

• Moreover, for the solutions $\widehat{x}$ to $(P)$, $\hat{\lambda}$ to $(D)$, the derivative (w.r.t. $x$) of $L(x,\widehat{\lambda})$ should be zero, since $L(\widehat{x},\widehat{\lambda}) \le L(x,\widehat{\lambda}), \forall\, x$.

This implies

$$\begin{aligned}\widehat{x}(t) &= (\phi')^{-1}\left(\sum_{i=1}^{n} \widehat{\lambda}_i a_i(t)\right)\\ &= (\phi')^{-1}\left(A^T\widehat{\lambda}\right).\end{aligned}$$

- We can now reconstruct the primal solution (qualitatively and quantitatively) from a presumptively easier dual computation.

# A DANTZIG ANECDOTE

"George wrote in "Reminiscences about the origins of linear programming," 1 and 2, *Oper. Res. Letters*, April 1982 (p. 47):

> *"The term Dual is not new. But surprisingly the term Primal, introduced around 1954, is. It came about this way. W. Orchard-Hays, who is responsible for the first commercial grade L.P. software, said to me at RAND one day around 1954: 'We need a word that stands for the original problem of which this is the dual.'*
> *I, in turn, asked my father, Tobias Dantzig, mathematician and author, well known for his books popularizing the history of mathematics. He knew his Greek and Latin. Whenever I tried to bring up the subject of linear programming, Toby (as he was affectionately known) became bored and yawned.*

*But on this occasion he did give the matter some thought and several days later suggested Primal as the natural antonym since both primal and dual derive from the Latin. It was Toby's one and only contribution to linear programming: his sole contribution unless, of course, you want to count the training he gave me in classical mathematics or his part in my conception."*

A lovely story. I heard George recount this a few times and, when he came to the "conception" part, he always had a twinkle in his eyes. (Saul Gass, Oct 2006)

- In a Sept 2006 *SIAM book review*, I asserted George assisted his father—for reasons I believe but cannot reconstruct.

  I also called Lord Chesterfield, Chesterton (*gulp*!).

## PITFALLS ABOUND

There are 2 major problems to this approach.*

**1.** *The assumption that a solution $\hat{x}$ exists.* For example, consider the problem

$$\inf_{x \in L^1[0,1]} \left\{ \int_0^1 x(t)dt : \int_0^1 tx(t)\, dt = 1, x \geq 0 \right\}.$$

◇ The optimal value is not attained. Similarly, existence can fail for the Burg entropy with trig moments. Additional conditions on $\phi$ are needed to insure solutions exist.† (BL2)

**2.** *The assumption that the Lagrangian is differentiable.* In the above, $f$ is $+\infty$ for every $x$ negative on a set of positive measure.

◇ This implies the Lagrangian is $+\infty$ on a dense subset of $L^1$, the set of functions *not* nonnegative a.e.. The Lagrangian is *nowhere continuous*, much less differentiable.

*A third, the existence of $\hat{\lambda}$, is less difficult to surmount.

†The solution is actually the *absolutely continuous part of a measure* in $C(\Omega)^*$.

## FIXING THE PROBLEM

• One approach to circumvent the differentiability problem, is to pose the problem in $L^\infty(\Omega)$, or in $C(\Omega)$, the space of essentially bounded, or continuous, functions. However, in these spaces, even with additional side qualifications, we are not necessarily assured solutions to $(P)$ exist.

◇ In (BL2), an example is given of a one parameter problem on the torus in $\mathbb{R}^3$, using the first four Fourier coefficients, and Burg's entropy, where solutions fail to exist for certain feasible data values.

• Alternatively, Minerbo poses the problem of tomographic reconstruction in $C(\Omega)$ with the Boltzmann-Shannon entropy. Unfortunately, the functions $a_i$ are characteristic functions of strips across $\Omega$, and the solution is piecewise constant, not continuous.

# CONVEX ANALYSIS (AN ADVERT)

We prepare to state a theorem that guarantees that the form of solution found in the above faulty derivation $\boxed{\widehat{x} = (\phi')^{-1}(A^T\widehat{\lambda})}$ is, in fact, correct. A full derivation is given in (BL2) and (BZ05).

• We introduce the *Fenchel (Legendre) conjugate* (see BL1) of a function $\phi : I\!R \to (-\infty, +\infty]$:

$$\phi^*(u) = \sup_{v \in I\!R} \{uv - \phi(v)\}.$$

• Often this can be (pre-)computed explicitly, using Newtonian calculus. Thus,

$$\phi(v) = v \log v - v, -\log v \text{ and } v^2/2$$

yield

$$\phi^*(u) = \exp(u), -1 - \log(-u) \text{ and } u^2/2$$

respectively. The red is the *log barrier* of interior point fame!

• The Fisher case is similarly explicit.

## EXAMPLE 2. CONJUGATES & NMR

The *Hoch and Stern information measure*, or *neg-entropy*, is defined in complex $n-$space by

$$H(z) = \sum_{j=1}^{n} h(z_j/b),$$

where $h$ is convex and given (for scaling $b$) by:

$$\boxed{h(z) \triangleq |z| \log\left(|z| + \sqrt{1+|z|^2}\right) - \sqrt{1+|z|^2}}$$

for *quantum theoretic* (NMR) reasons.

- Recall the *Fenchel-Legendre conjugate*

$$f^*(y) := \sup_x \langle y, x\rangle - f(x).$$

- Our *symbolic convex analysis* package (stored at `www.cecm.sfu.ca/projects/CCA/`, also in Chris Hamilton's package at Dal) produced:

$$h^*(z) = \cosh(|z|)$$

◇ Compare the *Shannon entropy*:

$$(z \log z - z)^* = \exp(z).$$

## COERCIVITY AND DUALITY

- We say $\phi$ possess *regular growth* if either $d = \infty$, or $d < \infty$ and $k > 0$, where $d = \lim_{u \to \infty} \phi(u)/u$ and $k = \lim_{v \uparrow d}(d - v)(\phi^*)'(v)$.*

- The *domain* of a convex function is $\mathrm{dom}(\phi) = \{u : \phi(u) < +\infty\}$; $\phi$ is *proper* if $\mathrm{dom}(\phi) \neq \emptyset$. Let $\imath = \inf \mathrm{dom}(\phi)$ and $\sigma = \sup \mathrm{dom}(\phi)$.

- Our *constraint qualification*,† $(CQ)$, reads

$$\exists\, \overline{x} \in L^1(\Omega), \text{ such that } A\overline{x} = b,$$
$$f(\overline{x}) \in \mathbb{R}, \ \ \imath < \overline{x} < \sigma \ \ a.e.$$

◇ *In many cases, $(CQ)$ reduces to feasibility*, (e.g., spectral estimation) and trivially holds.

- In this language, the *dual problem* for $(P)$ is

$$\sup \left\{ \langle b, \lambda \rangle - \int_\Omega \phi^*(A^T \lambda(t))dt \right\}. \qquad (D)$$

*-log does nor possess regular growth; $v \to v \log v$ does.

†The standard Slater's condition fails; this is what guarantees dual solutions exist.

**Theorem 1 (BL2)** *Let $\Omega$ be a finite interval, $\mu$ Lebesgue measure, each $a_k$ continuously differentiable (or just locally Lipschitz) and $\phi$ proper, strictly convex with regular growth.*

*Suppose $(CQ)$ holds and also*
(1)

$$\exists\, \tau \in \mathbb{R}^n \textit{ such that } \sum_{i=1}^{n} \tau_i a_i(t) < d \quad \forall t \in [a,b],$$

*then* the unique solution to $(P)$ is given by

(2) $$\widehat{x}(t) = (\phi^*)'(\sum_{i=1}^{n} \widehat{\lambda}_i a_i(t))$$

where $\widehat{\lambda}$ is any solution to dual problem $(D)$ *(and such $\widehat{\lambda}$ must exist).*

• This theorem generalizes to cover $\Omega \subset \mathbb{R}^n$, and more elaborately in Fisher-like cases. These results can be found in (BL2, BN1).

◇ 'Bogus' differentiation of a discontinuous function becomes the delicate $\boxed{(\int_\Omega \phi)^*(x^*) = \int_\Omega \phi^*(x^*)}$.

• Thus, the form of the maximum entropy solution can be legitimated simply by validating the *easily* checked conditions of Theorem 1.

♠ Also, any solution to $Ax = b$ of the form in (2) is automatically a solution to $(P)$.

So, solving $(P)$ is equivalent to finding $\lambda \in \mathbb{R}^n$ with

(3) $$\langle (\phi^*)'(A^T\lambda), a_i\rangle = b_i, \quad i = 1, \ldots, n,$$

a *finite dimensional* set of non-linear equations.

> One can then apply a standard 'industrial strength' nonlinear equation solver, like Newton's method, to this system, to find the optimal $\lambda$.

• Often, $\boxed{(\phi')^{-1} = (\phi^*)'}$ and so the 'dubious' solution agrees with the 'honest' solution.

Importantly, we may tailor $(\phi')^{-1}$ to our needs.

• **Note** that discretization is only needed to compute terms in (3). Indeed, *these integrals can sometimes be computed exactly* (e.g., in some tomography and option estimation problems). This is the gain of *not discretizing* early.

> By waiting to see the form of dual problem, one can customize one's integration scheme to the problem at hand.

• For European option pricing the constraints are based on 'hockey-sticks' of the form

$$a_i(x) := \max\{0, x - t_i\}$$

so the dual can be computed *exactly* and leads to a relatively small and explicit nonlinear equation to solve (BCM).

◇ Even when this is not the case one can often use the shape of the dual solution to fashion very*efficient heuristic reconstructions* that avoid any iterative steps (see BN2).

## $\mathcal{M}$om$\mathcal{E}$nt+

• $\mathcal{M}$om$\mathcal{E}$nt+ (`www.cecm.sfu.ca/interfaces/`) has code for entropic reconstructions as above. Moments (including wavelets), entropies and dimension are easily varied. It also allows for adding noise and relaxation of the constraints.

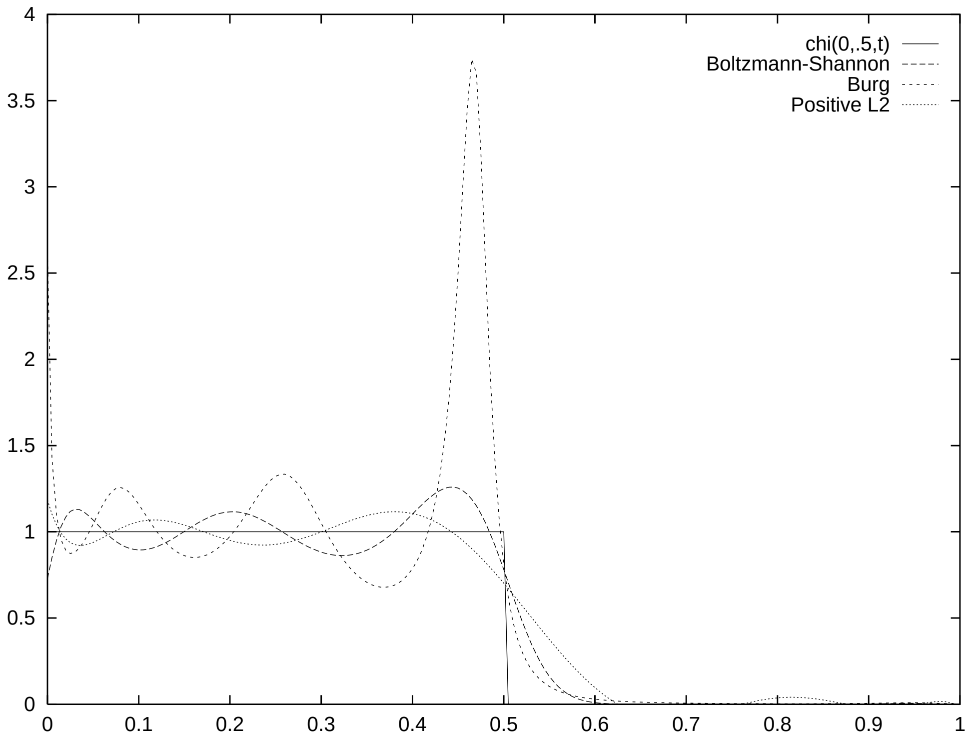
> Several methods of solving the dual are possible, including *Newton and quasi-Newton methods (BFGS, DFP), conjugate gradients*, and the suddenly sexy *Barzilai-Borwein line-search free method*.

• For iterative methods below, I recommend:

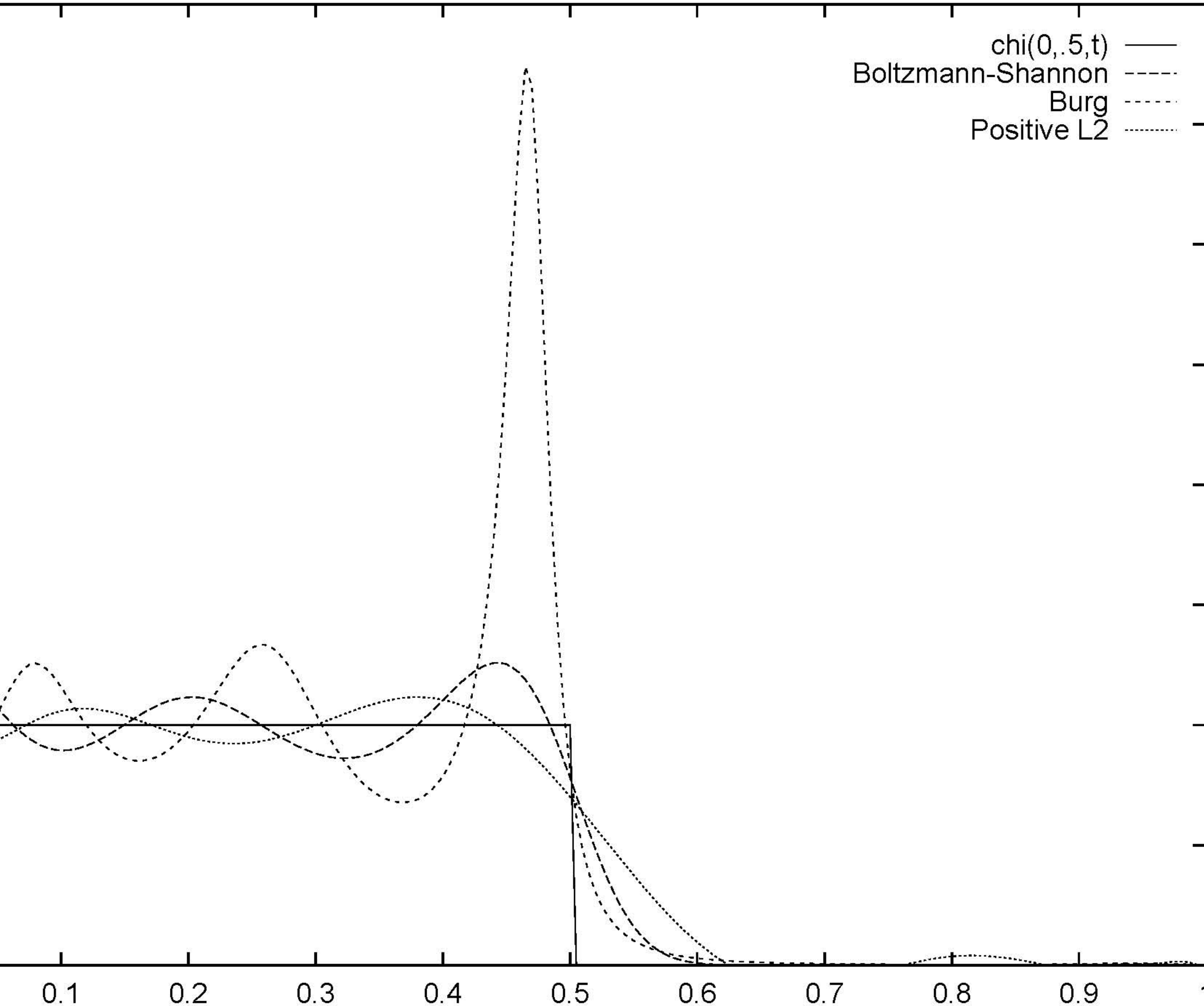
H.H. Bauschke and J.M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, **38** (1996), 367–426 (cited over 100 time by MathSciNet, 215 times in ISI, 350 in Google!), and a forthcoming CMS-Springer book written by *Bauschke and Combettes*.

# COMPARISON OF ENTROPIES

• The positive $L^2$, Boltzmann-Shannon and Burg entropy reconstruction of the **characteristic function** of $[0, 1/2]$ using *10 algebraic moments* ($b_i = \int_0^{1/2} t^{i-1} dt$) on $\Omega = [0, 1]$.




• **Solution**: $\widehat{x}(t) = (\phi^*)'(\sum_{i=1}^{n} \widehat{\lambda}_i t^{i-1})$.
Burg over-oscillates since $(\phi^*)'(t) = 1/t$.



chi(0,.5,t)
Boltzmann-Shannon
Burg
Positive L2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

## THE NON-CONVEX CASE

• In general non-convex optimization is a much less satisfactory field. We *can usually hope only to find critical points* ($f'(x) = 0$) or local minima. Thus, problem-specific heuristics dominate.

• **Crystallography**: We of course wish to estimate $x$ in $L^2(\mathbb{R}^n)$* Then the modulus $c = |\hat{x}|$ is known ($\hat{x}$ is the Fourier transform of $x$).†

Now $\{y \colon |\hat{y}| = c\}$, is not convex. So the issue is to find $x$ given $c$ and other convex information. An appropriate optimization problem extending the previous one is

$$\min \{ f(x) : Ax = b, \|Mx\| = c, \ \ x \in X \}, \quad (NP)$$

where $M$ models the modular constraint, and $f$ is as in Theorem 1.

*Here $n = 2$ for images, 3 for holographic imaging, etc.

†Observation of the modulus of the diffracted image in crystallography. Similarly, for optical aberration correction.

# EXAMPLE 3: CRYSTALLOGRAPHY

• My Parisian collaborator Combettes is expert on optimization perspectives of cognates to $(NP)$ and related feasibility problems.

◇ Most methods rely on a two-stage (easy convex, hard non-convex) decoupling schema— the following from Decarreau et al. (D). They suggest solving

$$\min \{f(x) : Ax = y, \|B_k y\| = b_k, \;\; x \in X\}, \tag{$NP^*$}$$

where $\|B_k y\| = b_k, \;\; k \in K$ encodes the hard modular constraints.

• They solve formal *first-order Kuhn-Tucker conditions* for a relaxed form of $(NP^*)$. The easy constraints are treated by Thm 1.

I am obscure, largely because the results were largely negative:

• They applied these ideas to **a prostaglandin molecule (25 atoms)**, with known structure, using quasi-Newton (which could fail to find a local min), truncated Newton (better) and trust-region (best) numerical schemes.

◇ They observe that the "*reconstructions were often mediocre*" and highly dependent on the amount of prior information – a small proportion of unknown phases to be satisfactory.

> "**Conclusion**: It is fair to say that the entropy approach has limited efficiency, in the sense that it requires a good deal of information, especially concerning the phases. *Other methods are wanted when this information is not available.*"

• Thus, I offer this part of my presentation largely to illustrate the difficulties.

# EXAMPLE 4. HUBBLE TELESCOPE

The basic setup—more details follow.

- **Electromagnetic field:** $u : \mathbb{R}^2 \to \mathbb{C} \in L^2$

- **DATA:** Field intensities for $m = 1, 2, \ldots, M$:

$$\psi_m : \mathbb{R}^2 \to \mathbb{R}_+ \in L^1 \cap L^2 \cap L^\infty$$

- **MODEL:** Functions $\mathcal{F}_m : L^2 \to L^2$, are *modified Fourier Transforms*, for which we can measure the modulus (intensity)

$$|\mathcal{F}_m(u)| = \psi_m \quad \forall m = 1, 2, \ldots, M.$$

$\oplus$ **INVERSE PROBLEM:** For the given transforms $\mathcal{F}_m$ and **measured** field intensities $\psi_m$ (for $m = 1, \ldots, M$), find a **robust estimate** of the underlying $u$.

## ... AND SOME HOPE FROM HUBBLE

• The (human-ground) lens with a micro-asymmetry was mounted upside-down. The perfect back-up (computer-ground) lens stayed on earth!

◇ NASA challenged ten teams to devise algorithmic fixes.

• **Optical aberration correction**, using the *Misell algorithm*, a *method of alternating projections*, works much better than it should—given that it is being applied to find a member of a version of

$$\Psi := \bigcap_{k=1}^{M} \{x : Ax = b, \|M_k x\| = c_k, \;\; x \in X\}, \tag{NCFP}$$

which is a **non-convex feasibility problem** as on the next page.

Is there **hidden convexity**?

# HUBBLE IS ALIVE AND KICKING

## Hubble reveals most distant planets yet

**Last Updated: Wednesday, October 4, 2006 | 7:21 PM ET**
**CBC News**

Astronomers have discovered the farthest planets from Earth yet found, including one with a year as short as 10 hours — the fastest known.

Using the Hubble space telescope to peer deeply into the centre of the galaxy, the scientists found as many as 16 planetary candidates, they said at a news conference in Washington, D.C., on Wednesday.

The findings were published in the journal Nature.

Looking into a part of the Milky Way known as the galactic bulge, 26,000 light years from Earth, Kailash Sahu and his team of astronomers confirmed they had found two planets, with at least seven more candidates that they said should be planets.

The bodies are about 10 times farther away from Earth than any planet previously detected.

A light year is the distance light travels in one year, or about 9.46 trillion kilometres.

- From *Nature* Oct 2006. Hubble has since been reborn twice and exoplanets have become quotidian. There were 228 exoplanets listed at `www.exoplanets.org` in Sept 08 and March 09.

330 now?

# 5 Facts About Kepler (launch March 6)

-- Kepler is the world's first mission with the ability to find true Earth analogs -- planets that orbit stars like our sun in the "habitable zone." The habitable zone is the region around a star where the temperature is just right for water -- an essential ingredient for life as we know it -- to pool on a planet's surface.

-- By the end of Kepler's three-and-one-half-year mission, it will give us a good idea of how common or rare other Earths are in our Milky Way galaxy. This will be an important step in answering the age-old question: Are we alone?

-- Kepler detects planets by looking for periodic dips in the brightness of stars. Some planets pass in front of their stars as seen from our point of view on Earth; when they do, they cause their stars to dim slightly, an event Kepler can see.

-- Kepler has the largest camera ever launched into space, a 95-megapixel array of charge-coupled devices, or CCDs, as in everyday digital cameras.

-- Kepler's telescope is so powerful that, from its view up in space, it could see one person in a small town turning off a porch light at night.

NASA 05.03.2009

## TWO MAIN APPROACHES

**I. Non-convex (in)feasibility problem:** Given $\psi_m \neq 0$, define $\mathbb{Q}_0 \subset L^2$ convex, and

$$\mathbb{Q}_m := \left\{ u \in L^2 \mid |\mathcal{F}_m(u)| = \psi_m \ a.e. \right\} \quad \text{(nonconvex)}$$

we wish to find $u \in \bigcap_{m=0}^{M} \mathbb{Q}_m = \emptyset$.

⊙ via an *alternating projection method*: e.g., for two sets $A$ and $B$, **repeatedly compute**

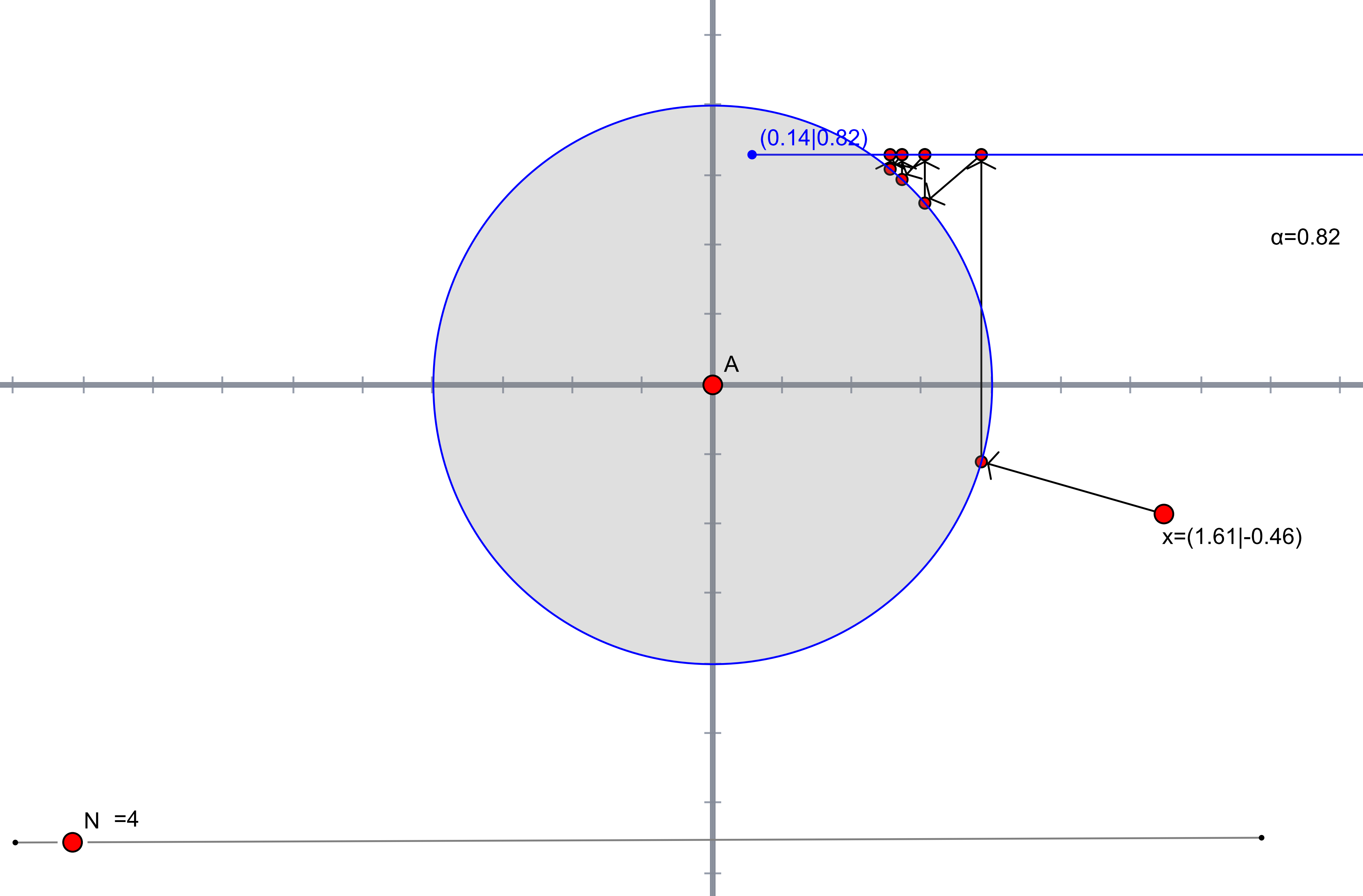
$$x \to P_B(x) =: y \to P_A(y) =: x.$$

**II. Error reduction of a nonsmooth objective** ('entropy') : for fixed $\beta_m > 0$

⊙ we attempt to solve

$$\text{minimize} \quad E(u) := \sum_{m=0}^{M} \frac{\beta_m}{2} \text{dist}^2(u, \mathbb{Q}_m)$$
$$\text{over} \quad u \in L^2.$$

# ALTERNATING PROJECTIONS FOR CIRCLE AND RAY

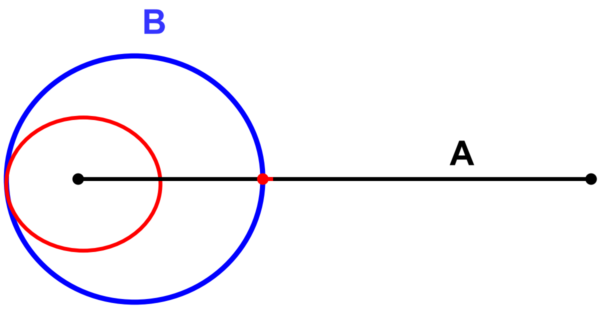


# I: NON-CONVEX PROJECTION CAN FAIL

• If $A \cap B \neq \emptyset$ and $A, B$ are closed convex then weak convergence (only 2002) is assured—von Neumann (1933) for subspaces, Bregman (1965).

⊙ Consider the **alternating projection method** to find the unique red point on the **line-segment A** (convex) and the blue **circle B** (non-convex).

• The **method is 'myopic'**.



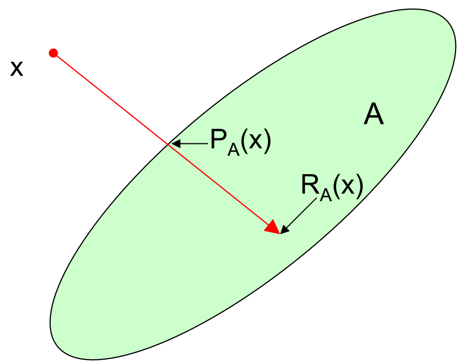

• Starting on line-segment outside the *red circle*, we converge to the unique feasible solution.

• Starting inside the red circle leads to a period-two locally 'least-distance' solution.

# I: PROJECTION METHOD OF CHOICE

• For optical abberation correction this is the **alternating projection** method:

$$x \to P_A\,(P_B(x))$$




• For crystallography it is better to use (HIO) **over-relax and average**: *reflect* to $R_A(x) := 2\,P_A(x) - x$ and use

$$x \to \frac{x + R_A\,(R_B(x))}{2}$$

• Both parallelize neatly: $A$ :=diag, $B := \prod_i C_i$.

• Both are nonexpansive *in the convex case*.

# APPROACH I: NAMES CHANGE . . .

• **The optics community** calls projection algorithms "*Iterative Transform Algorithms*".

> Hubble used *Misell's Algorithm*, which is just averaged projections. The best projection algorithm Luke* found was *cyclic projections* (with no relaxation).

• For the **crystallography problem** the best known method is called the *Hybrid Input-Output algorithm* in the optical setting. Bauschke-Combettes-Luke (JMAA, 2004) showed HIO, *Lions-Mercier* (1979), *Douglas-Rachford*, *Feinup*, and *divide-and-concur* coincide.

• When $u(t) \geq 0$ is imposed, Feinup's no longer coincides, and LM ('HPR') is still better.

*My former PDF, he was a *Hubble Graduate student.*

# ELSER, QUEENS and SUDOKU

**2006** Veit Elser at Cornell has had huge success (and press) using divide-and-concur on protein folding, sphere-packing, 3SAT, **Sudoku** ($\mathbb{R}^{2916}$), and more. Bauschke and Schaad likewise study **Eight queens problem** ($\mathbb{R}^{256}$) and image-retrieval (Science News, 08).

Veit Elser, Ph.D.

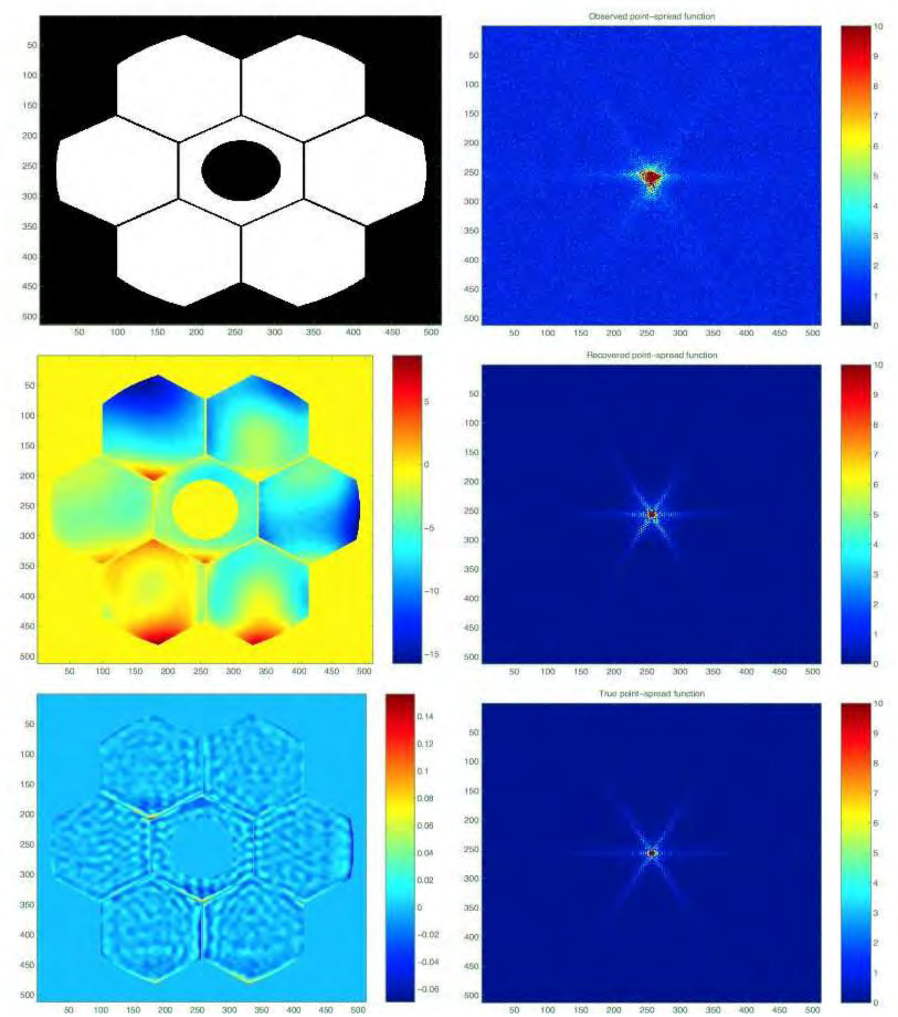
*Given a partially completed grid, fill it so that each column, each row, and each of the nine $3 \times 3$ regions contains the digits from $1$ to $9$ only once.*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 7 | 5 | | 9 | | | | 6 |
| | 2 | 3 | | 8 | | | 4 | |
| 8 | | | | | 3 | | | 1 |
| 5 | | | 7 | | 2 | | | |
| | 4 | | 8 | | 6 | | 2 | |
| | | | 9 | | 1 | | | 3 |
| 9 | | | 4 | | | | | 7 |
| | 6 | | | 7 | | 5 | 8 | |
| 7 | | | | 1 | | 3 | 9 | |

• This success (a.e.?) is not seen with alternating projections and cries out for explanation.

# A SAMPLE RECONSTRUCTION (via II)

- The object and its spectrum




**Top row**: data
**Middle**: reconstruction
**Bottom**: truth and error

# EXAMPLE 5. INVERSE SCATTERING

• **Central problem:** determine the location and shape of buried objects from measurements of the *scattered field* after illuminating a region with a known *incident field.*

• **Recent techniques:** determine if a point $z$ is inside or outside of the scatterer by determining *solvability* of the linear integral equation

$$\mathcal{F} g_z \stackrel{?}{=} \varphi_z$$

where $\mathcal{F} \to X$ is a compact linear operator constructed from the observed data, and $\varphi_z \in X$ is a known function parameterized by $z$.

• $\mathcal{F}$ has *dense range*, but if $z$ is on the exterior of the scatterer, then $\varphi_z \notin$Range($\mathcal{F}$).

• Since $\mathcal{F}$ is compact, any numerical implementation to solve the above integral equation will need some *regularization scheme*.

• If *Tikhonov regularization* is used—in a restricted physical setting—the solution to the regularized integral equation, $g_{z,\alpha}$, has the behaviour

$$||g_{z,\alpha}|| \to \infty \quad \text{as} \quad \alpha \to 0$$

*if and only if* $z$ is a point outside the scatterer.

• **An important open problem** is to determine the behavior of regularized solutions $g_{z,\alpha}$ under different regularization strategies.

In other words, when can these techniques fail? (On going joint work with Russell Luke for a 2009 IMA Summer School: also in *Experimental Math in Action*, AKP, 2007).

## FINIS: REFLECTIONS IN THE CIRCLE

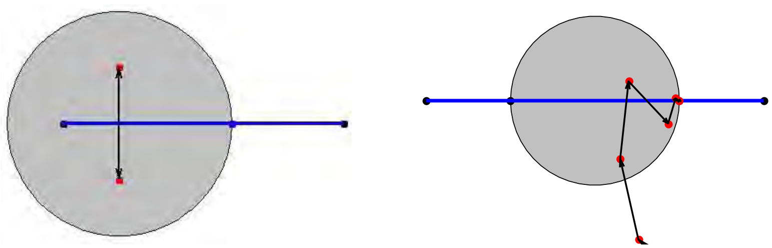
- **Dynamics** when $B$ is the unit circle and $A$ is the blue horizontal line at height $\alpha \geq 0$ are already fascinating. Steps are for

$$T := \frac{I + R_A \circ R_B}{2}:$$

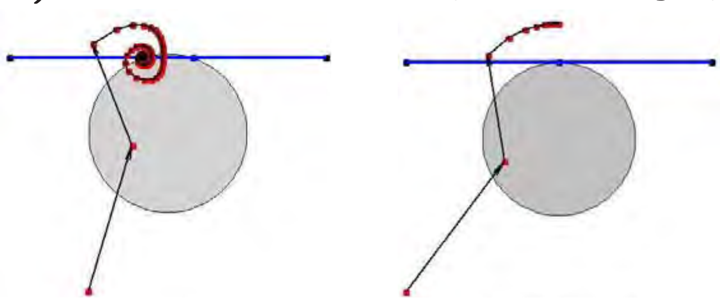
with $\theta_n$ the argument this becomes set

$$\boxed{x_{n+1} := \cos\theta_n, y_{n+1} := y_n + \alpha - \sin\theta_n.}$$

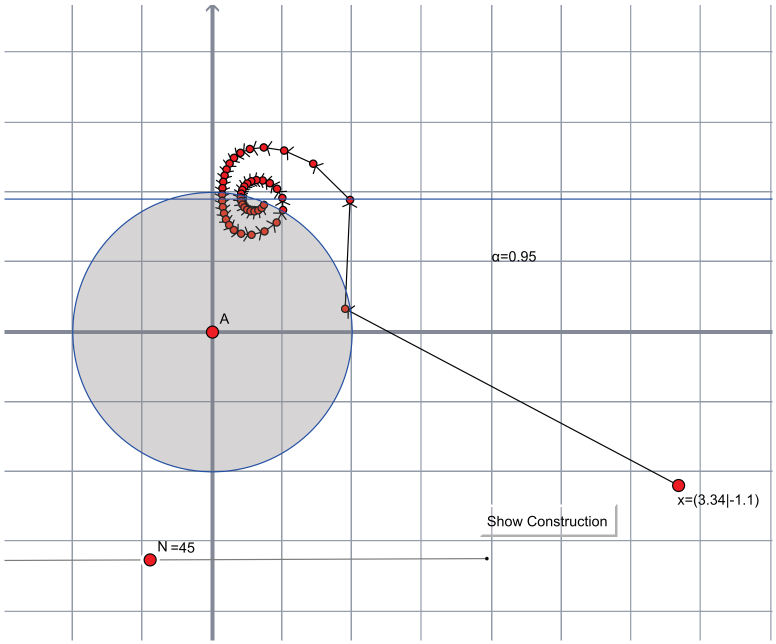
- $\alpha = 0$: converge **iff** start off $y$-axis ('chaos'):



- $\alpha > 1 \Rightarrow y \to \infty$, while $\alpha = 0.95$ $(0 < \alpha < 1)$ (**unproven**) and $\alpha = 1$ respectively produce:



# DYNAMIC GEOMETRY




- I finish with a *Cinderella* demo developed with Chris Maitland.

- Next week a proper introduction to the package will be given by Ulli Kortenkamp ...

# OTHER REFERENCES


**BCM** J.M. Borwein, R. Choksi and P. Maréchal, "Probability distributions of assets inferred from option prices via the Principle of Maximum Entropy," *SIAM J. Optimization*, **4** (2003), 464–478.

**BL1** J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS Books, Springer-Verlag, New York, 2005.

**BL2** J. M. Borwein and A. S. Lewis, "Duality relationships for entropy–like minimization problems," *SIAM J. Control and Optim.*, **29** (1991), 325–338.

**BLi** J.M. Borwein and M.A. Limber, "Under-determined moment problems: a case for convex analysis," *SIAMOpt*, Fall 1994.

**BN1** J.M. Borwein, A.S. Lewis, M.N. Limber and D. Noll, "Maximum entropy spectral analysis using first order information. Part 2," *Numerische Mathematik*, **69** (1995), 243–256.

**BN2** J. Borwein, M. Limber and D. Noll, "Fast heuristic methods for function reconstruction using derivative information," *App. Anal.*, **58** (1995), 241–261.

**D** A. Decarreau, D. Hilhorst, C. LeMaréchal and J. Navaza, "Dual methods in entropy maximization. Application to some problems in crystallography," *SIAM J. Optim.* **2** (1992), 173–197.