

Entropy and Projection Methods

for

Convex and Nonconvex Inverse Problems

Prepared for

Univ of South Australia Distinguished Lecture

March 29, 2010

... ..

Jonathan M. Borwein, FRSC FAAAS FAA

Laureate Professor and Director



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

URL: www.carma.newcastle.edu.au/~jb616

Revised: **March 26, 2010**

I SHALL FOLLOW BRAGG

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)

AND 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**

In **Persons and Places**, 1945, 238–239.

THREE FINE REFERENCES:

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

BL1 J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, CMS/Springer, 2nd expanded edition, New York, 2005.

BV J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge Univ Press, 2010.

OUTLINE

I shall discuss in “tutorial mode” the formalization of **inverse problems** such as **signal recovery** and **option pricing**: **first** as (convex and non-convex) **optimization problems** and **second** as **feasibility problems**—each 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**: for design and implementation.
4. **Non-convex extensions and feasibility problems**: life is hard. But sometimes we observe that more works than we understand why it 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:

$$\text{Find } x \text{ such that } A(x) = b$$

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 numerical computation. And guided by our analysis.

- Thus, the problem often is *how do we estimate x from a finite number of its ‘moments’?* This is typically an **under-determined inverse problem** (linear or nonlinear) where the unknown is most naturally a function, not a vector in \mathbb{R}^m .

EXAMPLE 1. AUTOCORRELATION

- Consider, extrapolating an *autocorrelation function* from given sample measurements:

$$R(t) := E \left[(X_s - \mu)(X_{t+s} - \mu) \right] / \sigma$$

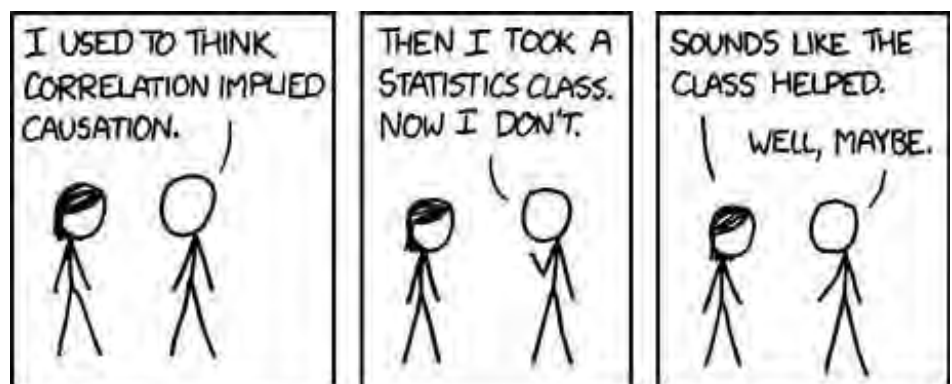
- ◇ (**Wiener-Khintchine**) Fourier moments of the power spectrum $S(\sigma)$ are samples of the autocorrelation function, so values of $R(t)$ computed directly from the data yields *moments* of $S(\sigma)$.

$$R(t) = \int_{\mathbb{R}} e^{2\pi i t \sigma} S(\sigma) d\sigma \quad S(\sigma) = \int_{\mathbb{R}} e^{-2\pi i t \sigma} R(t) dt$$

- Hence we may compute a finite number of moments of S , use them to make an estimate \hat{S} of S , and *estimate more moments* from \hat{S} by direct numerical integration. So we **dually extrapolate** R

...

- This avoids having to compute R directly from potentially noisy data.



PART ONE: THE ENTROPY APPROACH

- Following [BZ] 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, $f(x)$, of the unknown x .



www.carma.newcastle.edu.au

◇ The approach finds applications in countless fields including:

Acoustics, actuarial science, astronomy, biochemistry, constrained spline fitting, engineering, finance, image reconstruction, inverse scattering, optics, option pricing, multidimensional NMR, tomography, statistical moment fitting, and time series analysis, ...

(Many thousands of papers)

Medical Imaging

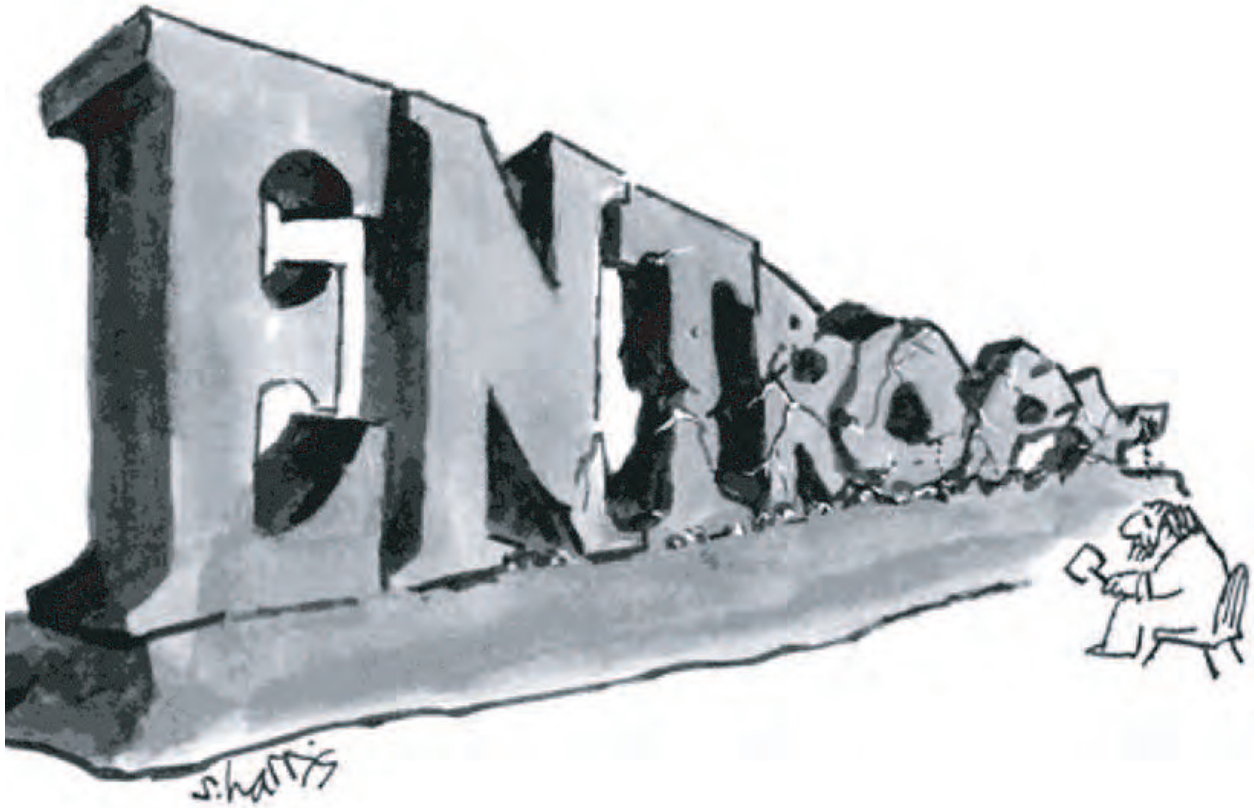
Modern HPC imaging techniques (such as PET using 'positrons' and SPECT using 'photons') provide non-invasive two- and three-dimensional real-time dynamic images for the brain, heart, kidney and other organs. They are revolutionizing research, surgery and disease management. A one-minute three-dimensional reconstruction requires enormous computing power to generate these images. [REF 6]



● However, the derivations and mathematics are fraught with subtle—and less subtle—errors.

I will next 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 example is:

Theorem. The function

$$H(\vec{p}) = - \sum_{k=1}^N p_k \log p_k$$

is *the unique continuous function* (up to a positive scalar multiple) on finite probabilities such that:

[I.] *Uncertainty grows:*

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

increases with n .

[II.] *Subordinate choices are respected:* for distributions \vec{p}_1 and \vec{p}_2 and $0 < p < 1$,

$$H(p\vec{p}_1, (1-p)\vec{p}_2) = p H(\vec{p}_1) + (1-p) H(\vec{p}_2).$$



ENTROPIES FOR US

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

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

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

for $i = 1, \dots, n$ and $a_i \in X^*$ the ‘dual’ of X (L^2 in the Hilbert case, L^∞ in the L^1 case).

- To pick a solution from the infinitude of possibilities, we may freely define “**best**” .

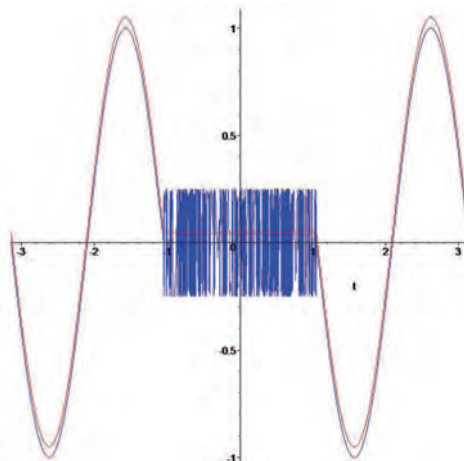
- ⊗ The most common approach is to find the **minimum norm solution*** by solving the *Gram system*:

$$\boxed{\text{Find } \lambda \text{ such that } AA^T \lambda = b} .$$

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

- This solved the following *variational problem*:

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



Continuous function with divergent Fourier series at 0.

*Even in the (realistic) infeasible case.

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

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

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

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

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* (based on $\phi(y) - \phi(x) - \phi'(x)(y - x)$) and *Csizar distances* (based on $x\phi(y/x)$) 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)

$$f(x) := \int x \log x d\mu,$$

(changes *dramatically* with μ) and the *Burg entropy* (in time series analysis),

$$f(x) := - \int \log x d\mu.$$

△ Includes the *log barrier* and *log det* functions from interior point theory.

◇ *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 !

- Use of Csizar distance based *Fisher Information*

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

(*jointly* convex) 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**, **BV10**].

- We introduce the *Lagrangian*

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

and the associated *dual problem*

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

- So we formally have a “dual pair” (BL1)

$$\min \{f(x) : Ax = b, x \in X\} = \min_{x \in X} \max_{\lambda \in \mathbb{R}^n} \{L(x, \lambda)\}, \quad (P)$$

and

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

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

- As $L(x, \hat{\lambda}) = \int_0^1 \phi(x(t))dt + \sum_{i=1}^n \hat{\lambda}_i (b_i - \langle x, a_i \rangle)$ this implies

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

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

A DANTZIG (1914-2005) ANECDOTE

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

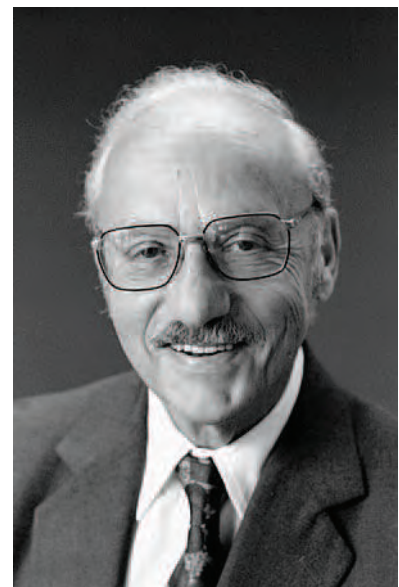
“The term Dual is not new. But surprisingly the term Primal, introduced around 1954, is. It came about this way. W. Orchard-Hays, who is responsible for the first commercial grade L.P. software, said to me at RAND one day around 1954: ‘We need a word that stands for the original problem of which this 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, 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 ϕ 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 way to get continuity/differentiability, is to

- work in $L^\infty(\Omega)$, or $C(\Omega)$ using essentially bounded, or continuous, functions. But, even with side qualifications, solutions to (P) may still not exist.

▽ Consider **Burg entropy** maximization in $L^1[T^3]$:

$$\max \int_{T^3} \log(x) dV \quad \text{s.t.} \quad \int_{T^3} x dV = 0$$

and

$$\begin{aligned} \int_{T^3} x \cos(a) dV &= \int_{T^3} x \cos(b) dV \\ &= \int_{T^3} x \cos(c) dV = \alpha. \end{aligned}$$

For feasible $1 > \alpha > \bar{\alpha}$, solutions only exist in $(L^\infty)^*$; a computable value [BL2].



Werner Fenchel (1905-1988)

- Minerbo, e.g., posed tomographic reconstruction in $C(\Omega)$, with Shannon entropy. But, his moments are characteristic functions of strips across Ω , and the solution would be piecewise constant.

CONVEX ANALYSIS (AN ADVERT)

We will give a theorem that guarantees the form of solution found in the above faulty derivation

$$\hat{x} = (\phi')^{-1}(A^T \hat{\lambda})$$

is, in fact, correct. (Full derivation in [BL2, BZ].)

- We introduce the *Fenchel (Legendre) conjugate* (see [BL1]) of a function $\phi : \mathbb{R} \rightarrow (-\infty, +\infty]$:

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

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

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

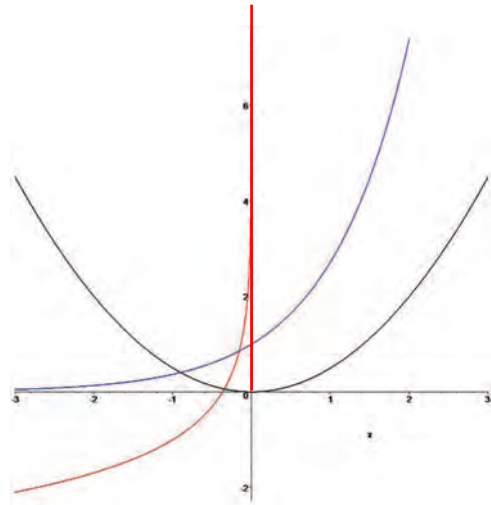
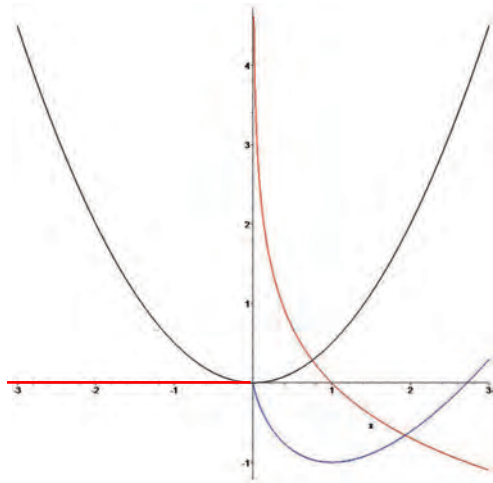
yield

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

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

- *Fisher* case is also explicit (via an integro-differential equation).

PRIMALS AND DUALS

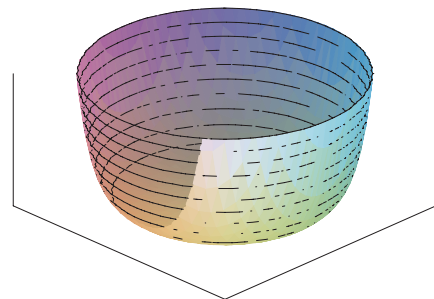
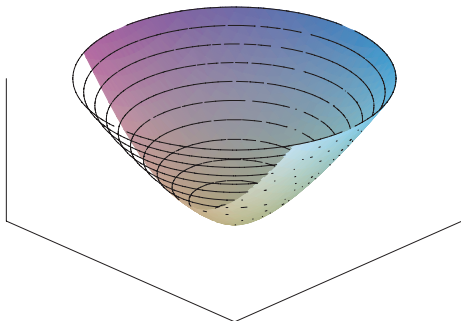


The three entropies below and their conjugates.

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

and

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



An NMR entropy and its conjugate.

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:

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

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 (see [BH] and Chris Hamilton's Dalhousie thesis package) produced:

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

- ◇ Compare the *Shannon entropy*:

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

FENCHEL DUALITY THEOREM (1951)

Theorem 1 (Utility Grade) Suppose $f: X \rightarrow R \cup \{+\infty\}$ and $g: Y \rightarrow R \cup \{+\infty\}$ are convex while $A: X \rightarrow Y$ is linear. Then

$$p := \inf_X f + g \circ A = \max_{Y^*} -g^*(-\cdot) - f^* \circ A^*,$$

if either $\text{int } A(\text{dom } f) \cap \text{dom } g \neq \emptyset$, (or if f, g are polyhedral).

- **indicator function** $\iota_C(x) := 0$ if $x \in C$ and $+\infty$ else.
- **support function** $\sigma_C(x^*) := (\iota_C)^*(x^*) = \sup_{x \in C} \langle x^*, x \rangle$.

Examples. include:

(i) $A := I$ is **equivalent** to **Hahn-Banach** theorem.

(ii) $g := \iota_{\{b\}}$ **yields** $p := \inf \{f(x) : Ax = b\}$.

Specializes to LP if $f := \iota_{R_n^+} + c$.

(iii) $f := \iota_C, g := \sigma_D$ **yields** **minimax theorem**:

$$\inf_C \sup_D \langle Ax, y \rangle = \sup_D \inf_C \langle Ax, y \rangle.$$

FENCHEL DUALITY (SANDWICH)

$$\inf_X f(x) - g(x) = \max_{Y^*} g_*(y^*) - f^*(y^*)$$

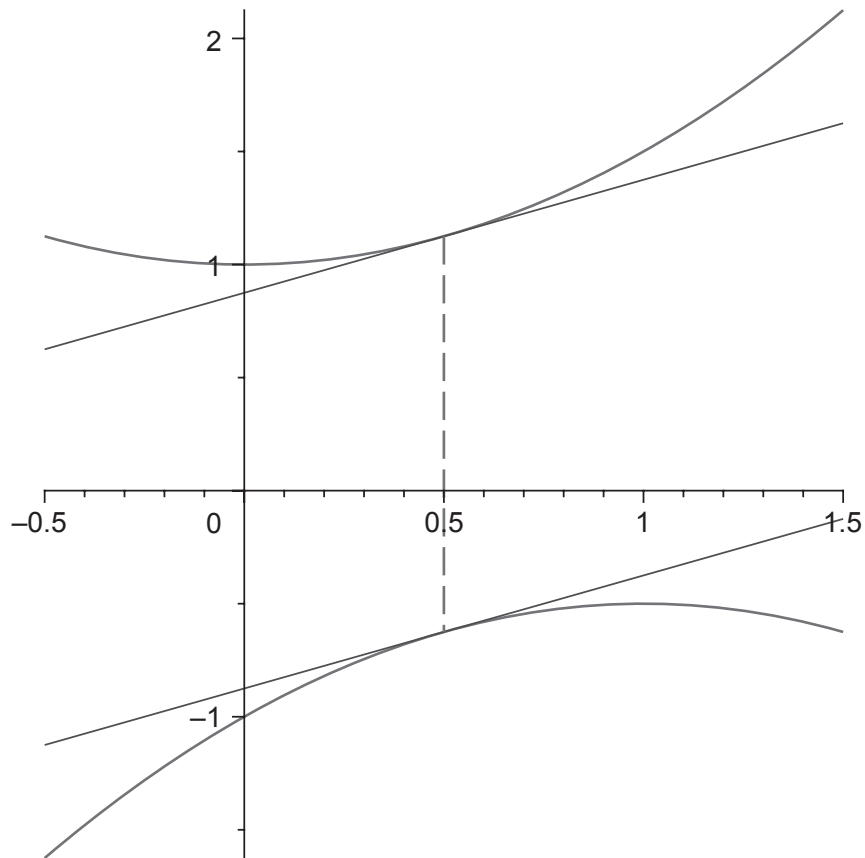


Figure 2.6 Fenchel duality (Theorem 2.3.4) illustrated for $x^2/2 + 1$ and $-(x - 1)^2/2 - 1/2$. The minimum gap occurs at $1/2$ with value $7/4$.

- Using concave conjugate: $g_* := -(-g)^*(-)$

COERCIVITY AND DUALITY

- We say ϕ possess *regular growth* if either $d = \infty$, or $d < \infty$ and $k > 0$, where

$$d := \lim_{u \rightarrow \infty} \phi(u)/u, \quad k := \lim_{v \uparrow d} (d - v)(\phi^*)'(v).$$

- Then $v \rightarrow v \log v$, $v \rightarrow v^2/2$ and the positive energy all have regular growth but $-\log$ does not.
- The *domain* of a convex function is $\text{dom}(\phi) = \{u : \phi(u) < +\infty\}$; ϕ is *proper* if $\text{dom}(\phi) \neq \emptyset$.
Let $\iota := \inf \text{dom}(\phi)$ and $\sigma := \sup \text{dom}(\phi)$.

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

$$\boxed{\begin{aligned} \exists \bar{x} \in L^1(\Omega), \text{ such that } A\bar{x} = b, \\ f(\bar{x}) \in \mathbb{R}, \quad \iota < \bar{x} < \sigma \text{ a.e.} \end{aligned}}$$

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

- The *Fenchel dual problem* for (P) is now:

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

*To ensure dual solutions. Standard *Slater* condition fails. Fenchel missed need for a (CQ) in his 1951 *Princeton Notes*.

Theorem 2 (BL2) Let Ω be a finite interval, μ Lebesgue measure, each a_k continuously differentiable (or just locally Lipschitz) and ϕ proper, strictly convex with regular growth.

Suppose (CQ) holds and also*

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

then the unique solution to (P) is given by

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

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

♠ We have a *functional reconstruction* for all $t \in \Omega$.

- 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 conjugacy formula

$$\boxed{(\int_{\Omega} \phi)^*(x^*) = \int_{\Omega} \phi^*(x^*)}.$$

*This is trivial if $d = \infty$.

- Thus, the form of the maximum entropy solution can be legitimated simply by validating the *easily* checked conditions of Theorem 2.

- ♠ 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

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

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

One can then apply a standard ‘industrial strength’ nonlinear equation solver, based say on Newton’s method, to this system, to find the optimal λ .

- 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 the problem (see [**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**] and Huang's 1993 thesis).

- The more nonlinear the optimization problem the more *dangerous* it is to treat it purely formally.

1949

- 1 H. Davenport, G. Pólya
On the product of two power series
- 6 Victor Lalan
Sur les surfaces à courbure moyenne isotherme
- 29 Alfred Schild
Discrete space-time and integral Lorentz transformations
- 48 H. W. Turnbull
Note upon the generalized Cayleyan operator
- 57 Hermann Weyl
Elementary algebraic treatment of the quantum mechanical symmetry problem
- 69 C. C. MacDuffee
Orthogonal matrices in four-space
- 73 W. Fenchel
On conjugate convex functions
- 78 K. Mahler
On the critical lattices of arbitrary point sets
- 88 R. H. Bruck, H. J. Ryser
The nonexistence of certain finite projective planes
- 94 Karl Menger
Generalized vector spaces. I. The structure of finite-dimensional spaces
- 105 Irving Kaplansky
Groups with representations of bounded degree
- 113 H. W. Ellis
Mean-continuous integrals
- 125 Ernst Snapper
Completely indecomposable modules
- 153 Marsten Morse, William Transue
Functionals of bounded Fréchet variation
- 166 G. de B. Robinson
On the disjoint product of irreducible representations of the symmetric group
- 176 M. H. Stone
Boundedness properties in function-lattices
- 187 Marshall Hall Jr.
Subgroups of finite index in free groups
- 191 E. C. Titchmarsh
On the uniqueness of the Green's function associated with a second-order differential equation
- 199 Garrett Birkhoff, Lindley Burton
Note on Newtonian force-fields
- 209 A. Einstein, L. Infeld
On the motion of particles in general relativity theory
- 242 S. Minakshisundaram, Å. Pleijel
Some properties of the eigenfunctions of the Laplace-Operator on Riemannian manifolds
- 257 J. L. Synge
On the motion of three vortices
- 271 Alexander Weinstein
On surface waves
- 279 Herbert Busemann
Angular measure and integral curvature
- 297 S. D. Chowla, John Todd
The density of reducible integers
- 300 Olga Taussky
On a theorem of Latimer and MacDuffee
- 303 J. S. Frame
Congruence relations between the traces of matrix powers
- 305 G. G. Lorentz
Direct theorems on methods of summability
- 320 S. Minakshisundaram
A generalization of Epstein zeta functions
- 328 N. S. Mendelsohn
Applications of combinatorial formulae to generalizations of Wilson's theorem
- 337 R. Rado
Axiomatic treatment of rank in infinite sets
- 344 Gordon Pall
Representation by quadratic forms
- 365 Robert Frucht
Graphs of degree three with a given abstract group
- 379 G. F. D. Duff
Factorization ladders and eigenfunctions

3. The theorem to be proved may now be formulated thus:

Let G be a convex point set in R^n and $f(x)$ a function defined in G convex and semi-continuous from below and such that $\lim_{x \rightarrow x^*} f(x) = \infty$ for each boundary point x^* of G which does not belong to G . Then there exists one and only one point set Γ in R^n and one and only one function $\phi(\xi)$ defined in Γ with exactly the same properties as G and $f(x)$ such that

$$(5) \quad \Sigma x \xi \leq f(x) + \phi(\xi),$$

where to every interior point x of G there corresponds at least one point ξ of Γ for which equality holds.

In the same way $G, f(x)$ correspond to $\Gamma, \phi(\xi)$.

... a Modern Oak

- Theorem 2 works by relaxing the problem to $(L^1)^{**}$ — where solutions always exist — and using **Lebesgue decomposition**. Regular growth rules out **singular** part ($I_{\phi^{**}} = (I_{\phi})^{**} \upharpoonright_X$).
- More generally, for Ω an interval, consider $I_{\phi}(x) := \int_{\Omega} \phi(x) d\mu$ as a function on $L^1(\Omega)$. We say I_{ϕ} is **strongly rotund** (very well posed) if it is (i) **strictly convex** with (ii) weakly compact lower level sets (**Dunford-Pettis**) and (iii) **Kadec-Klee**:

$$I_{\phi}(x_n) \rightarrow I_{\phi}(x), x_n \rightharpoonup x \Rightarrow x_n \rightarrow_1 x.$$

Theorem 3 (BV) I_{ϕ} is strongly rotund as soon as ϕ^* is everywhere finite and differentiable on \mathbb{R} ; and conversely if μ is not purely atomic.

- Easy to check (holds for Shannon and energy but not Burg) and is the best **surrogate** for a reflexive norm on L^1 .

- **MomEnt+** (www.cecm.sfu.ca/interfaces/) provided 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.

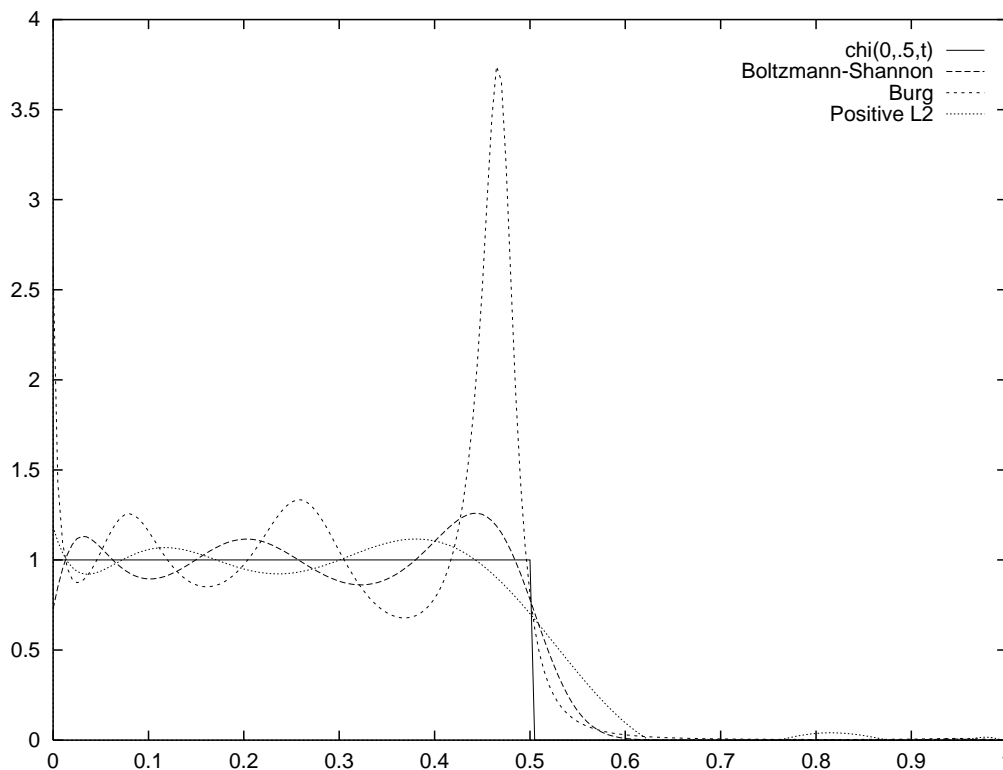
Several methods of solving the dual are possible, including *Newton and quasi-Newton methods (BFGS, DFP), conjugate gradients*, and the suddenly sexy *Barzilai-Borwein line-search free method*.

- For iterative methods as below, I recommend:

[**BaB**] H.H. Bauschke and J.M. Borwein, “**On projection algorithms for solving convex feasibility problems,**” *SIAM Review*, **38** (1996), 367–426 (over 265 ISI cites), 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: $\hat{x}(t) = (\phi^*)'(\sum_{i=1}^n \hat{\lambda}_i t^{i-1})$.

- Burg over-oscillates** since $(\phi^*)'(t) = 1/t$. But is still often the ‘best’ solution (with a closed form for Fourier moments)!

PART TWO: 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 $\{y: |\hat{y}| = c\}$, 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 2.

*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

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

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 2.

I am obscure, mainly 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.

GENERAL PHASE RECONSTRUCTION

The basic setup—more details follow.

- **Electromagnetic field:** $u : \mathbb{R}^2 \rightarrow \mathbb{C} \in L^2$
- **DATA:** Field intensities for $m = 1, 2, \dots, M$:

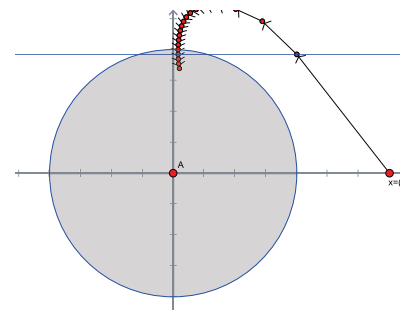
$$\psi_m : \mathbb{R}^2 \rightarrow \mathbb{R}_+ \in L^1 \cap L^2 \cap L^\infty$$

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

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

♠ INVERSE PROBLEM:

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



EXAMPLE 4. 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

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

(NCFP)

which is a **non-convex feasibility problem** as discussed below.

- Is there **hidden convexity**?

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 exoplanet discoveries have become quotidian.
- There were 228 listed at www.exoplanets.org in March 09 and 432 a year later.



HUBBLE IS ALIVE AND KICKING, II

THE INDEPENDENT

Adrian Hamilton: People vote for competence not policy

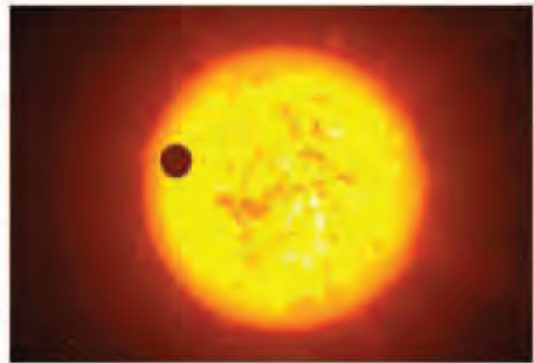
New planet Corot-9b has Earth-like temperatures

By Steve Connor, Science Editor

Thursday, 18 March 2010

The first planet with a "temperate" climate to orbit a distant star has been discovered by astronomers, who claim that the techniques used to study it will be critical in the search for Earth-like worlds beyond our own solar system.

Corot-9b, as the planet is called, is one of more than 400 "exoplanets" found to be orbiting other stars, but it is the first one with a near-normal temperature range that can be studied as it moves across (or "transits") the sun it orbits. "This is a normal, temperate exoplanet just like dozens we already know, but this is the first whose properties we can study in depth," said Claire Moutou, one of the team of astronomers at the European Southern Observatory who made the discovery. "It is bound to become a Rosetta stone in exoplanet research."



PA

An artist's impression of Corot-9b transit across its sun

Corot-9b passes in front of its host star, 1,500 light years away in the constellation Serpens, every 95 days, and the transit lasts about eight hours, which is when astronomers can make measurements of the planet's composition and temperature, estimated to range from minus 20C to 160C.

"Corot-9b is the first exoplanet that really does resemble planets in our solar system," said Hans Deeg, the lead author of the study published in the journal *Nature*. "It has the size of Jupiter and an orbit similar to that of Mercury."

- How reliable are these determinations (velocity, imaging, transiting, timing, microlensing)?

TWO RECONSTRUCTION APPROACHES

I. Error reduction of a nonsmooth objective

(‘entropy’) : for fixed $\beta_m > 0$

⊙ we attempt to solve

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

II. Non-convex (in)feasibility problem: Given

$\psi_m \neq 0$, define $\mathbb{Q}_0 \subset L^2$ convex, and

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

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

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

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 $\mathcal{F} \rightarrow X$ 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 $\varphi_z \notin \text{Range}(\mathcal{F})$ (which has a Fenchel conjugate characterization).

- 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

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

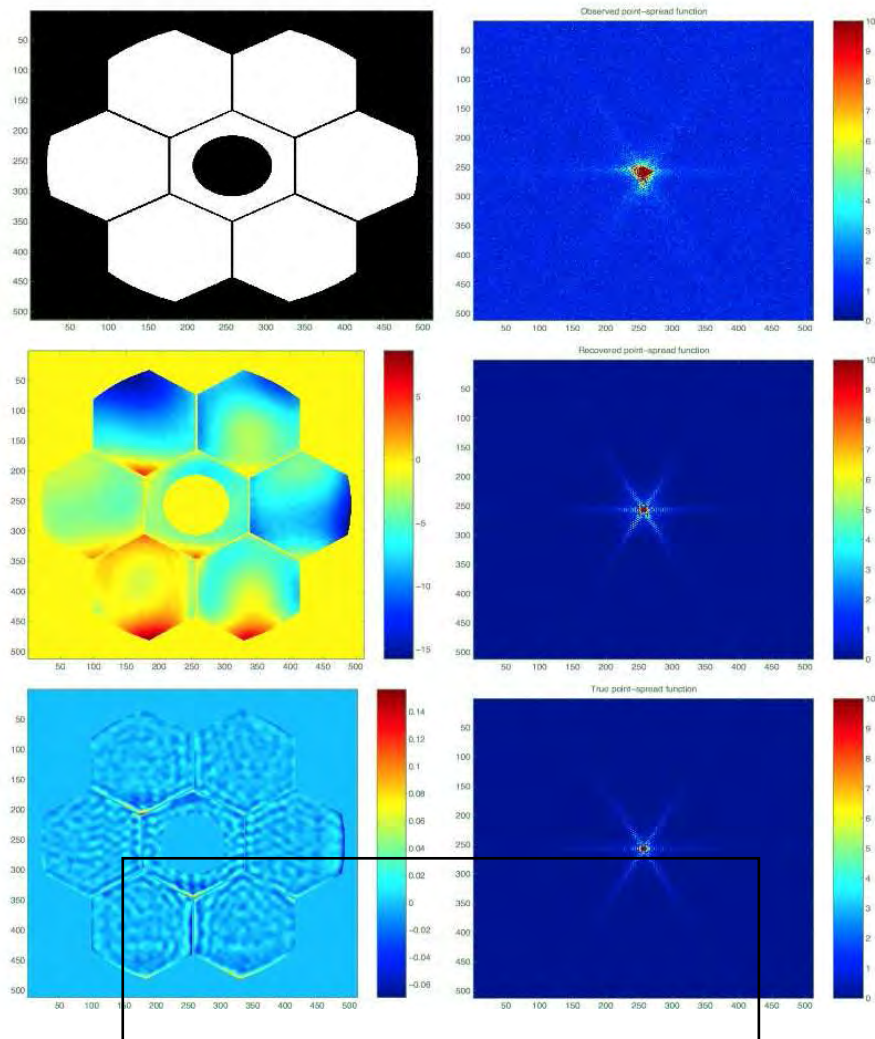
if and only if z is a point outside the scatterer.

- **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? (Ongoing work with Russell Luke [BLu] to appear in *Springer Handbook of Imaging*: also mentioned *Experimental Math in Action*, AKP, 2007.)

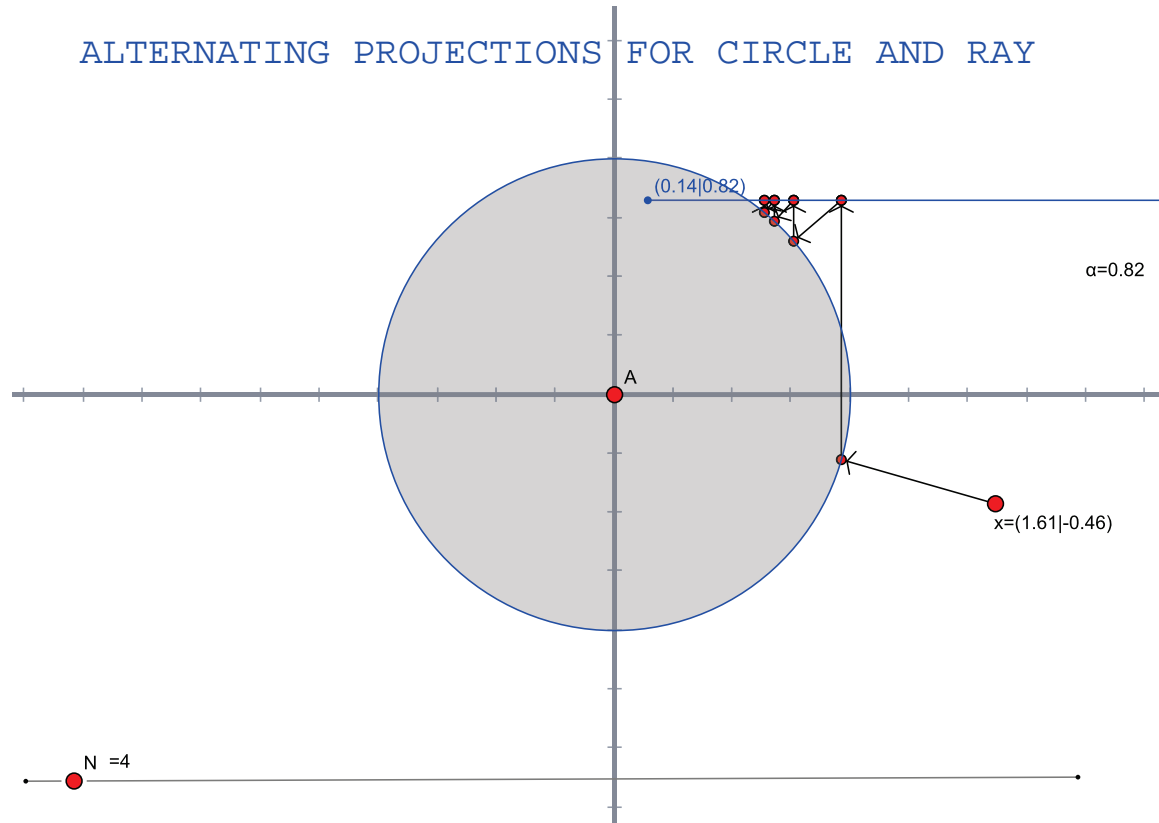
A SAMPLE RECONSTRUCTION (via I)

- The object and its spectrum



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

ALTERNATING PROJECTIONS



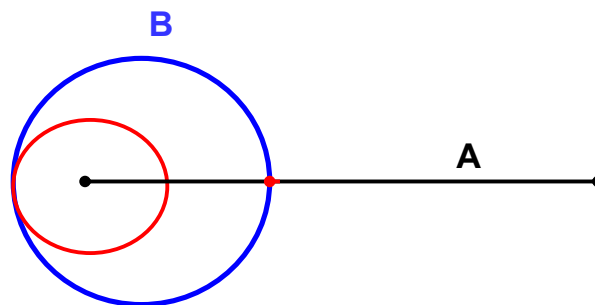
- The method is *fairly* well understood when all the sets are convex.
- If $A \cap B \neq \emptyset$ and A, B are closed convex then weak convergence (**only** 2002) is assured—von Neumann (1933) for subspaces, Bregman (1965).

II: NON-CONVEX PROJECTION CAN FAIL

If A is finite codimension and affine, B is the nonnegative cone in $\ell^2(N)$ and $A \cap B \neq \emptyset$,
is the method norm convergent?

⊙ 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'**.

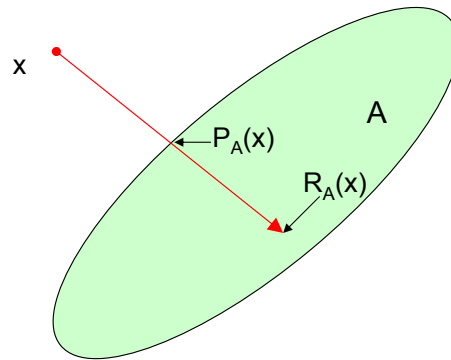


- Starting on line-segment outside **red circle**, we converge to unique feasible solution.
- Starting inside the red circle leads to a period-two locally 'least-distance' solution.

THE PROJECTION METHOD OF CHOICE

- For optical aberration correction this is the **alternating projection** method:

$$x \rightarrow P_A(P_B(x))$$



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

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

- Both parallelize neatly: $A := \text{diag}$, $B := \prod_i B_i$.
- Both are nonexpansive *in the convex case*.

NAMES CHANGE WHEN FIELDS DO...

- 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, see [E1] and [E2], at Cornell has had huge success (and press) using [divide-and-concur](#) on protein folding, sphere-packing, 3SAT, **Sudoku** (\mathbb{R}^{2916}), and more.

2008 Bauschke and Schaad likewise study **Eight queens problem** (\mathbb{R}^{256}) and image-retrieval (*Science News*, 08).

Given a partially completed grid, fill it so that each column, each row, and each of the nine 3×3 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. Brailey Sims and I [BS] have made some progress.

THE KEPLER SATELLITE

5 Facts About Kepler (launch March 6)

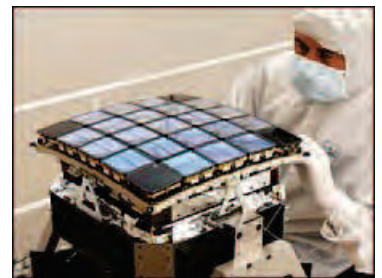
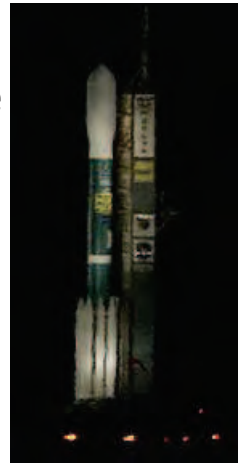
-- Kepler is the world's first mission with the ability to find true Earth analogs -- planets that orbit stars like our sun in the "habitable zone." The habitable zone is the region around a star where the temperature is just right for water -- an essential ingredient for life as we know it -- to pool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

FINIS: REFLECTIONS IN THE CIRCLE

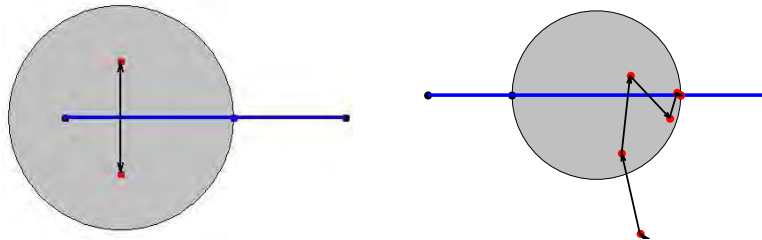
- **Dynamics** when B is the unit circle and A is the blue horizontal line at height $\alpha \geq 0$ are already fascinating. Steps are for

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

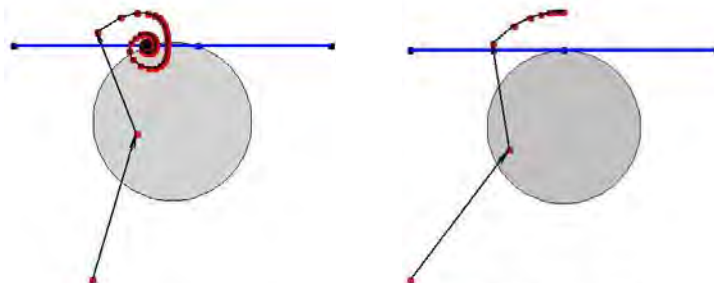
with θ_n the argument this becomes set

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

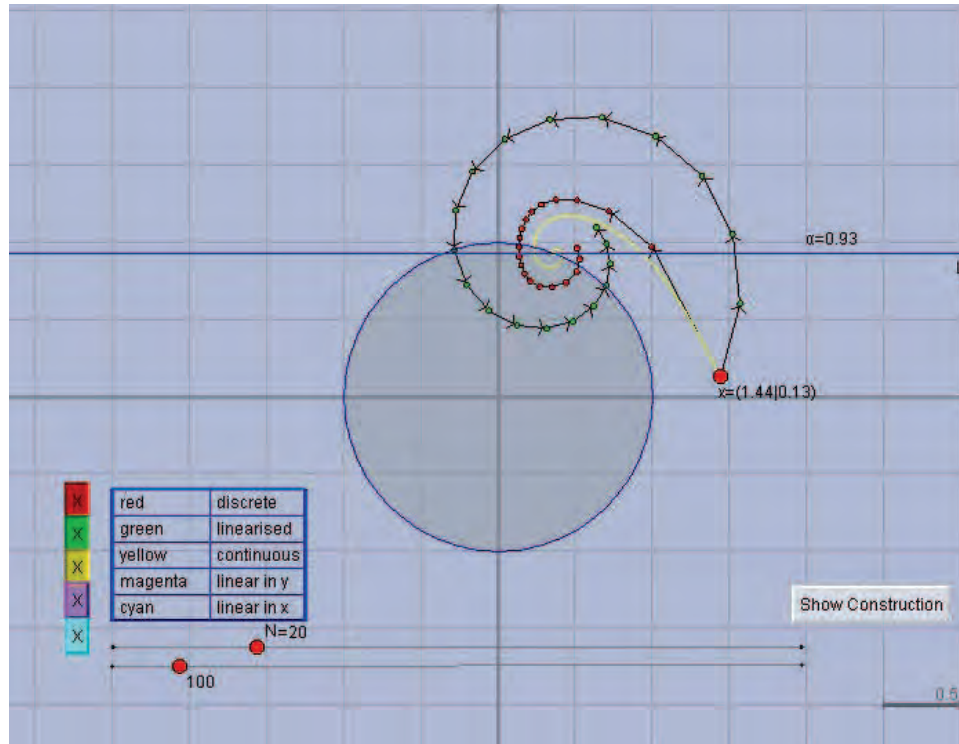
- $\alpha = 0$: converge **iff** start off y -axis ('chaos'):



- $\alpha > 1 \Rightarrow y \rightarrow \infty$, while $\alpha = 0.95$ ($0 < \alpha < 1$) (largely proven) and $\alpha = 1$ respectively produce:



DYNAMIC GEOMETRY



- I finish with a *Cinderella* demo built with Chris Maitland and based on recent work with Brailey Sims [BS]. The applets are at:

www.carma.newcastle.edu.au/~jb616/composite.html



www.carma.newcastle.edu.au/~jb616/expansion.html

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,” *SIAMOpt*, **4** (2003), 464–478.
- BH** J.M. Borwein and C. Hamilton, “Symbolic Convex Analysis: Algorithms and Examples,” *Math Programming*, **116** (2009), 17–35.
- BL2** J. M. Borwein and A. S. Lewis, “Duality relationships for entropy-like minimization problems,” *SIAM Control and Optim.*, **29** (1991), 325–338.
- BLi** J Borwein and M. 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,” *Numer Math*, **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.
- BLu** J.M. Borwein and R.L. Luke, “Duality and Convex Programming.” *Handbook of Mathematical Methods in Imaging*, O. Scherzer (ed.), Springer. In press, 2010.
- BS** J.M. Borwein and B. Sims, “The Douglas-Rachford algorithm in the absence of convexity.” In *Springer Optimization and Appl.*, 2010.
- E2** Gravel, S. and Elser, V., “Divide and concur: A general approach constraint satisfaction,” preprint, 2008, <http://arxiv.org/abs/0801.0222v1>.
- 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.
- E1** Elser, V., Rankenburg, I., and Thibault, P., “Searching with iterated maps,” *Proceedings of the National Academy of Sciences* **104** (2007), 418–423.