Projection methods
— the convex case

Heinz Bauschke

Mathematics and Statistics
Okanagan UC and U of Guelph
Canada

bauschke@cecm.sfu.ca
http://www.cecm.sfu.ca/~bauschke

"Phase Problem for Non-Periodic Objects"
Lawrence Berkeley National Laboratory
May 18, 2001, 2:10PM – 2:50PM
Building 50 Auditorium

Overview

1. Introduction
2. Fundamental results
3. Recent developments
4. Tools
5. Open problems

1. Introduction

Set up

X is a Hilbert space

with inner product $\langle \cdot, \cdot \rangle$ and norm $\|x\| = \sqrt{\langle x, x \rangle}$.

Example. X is some function space L^2.

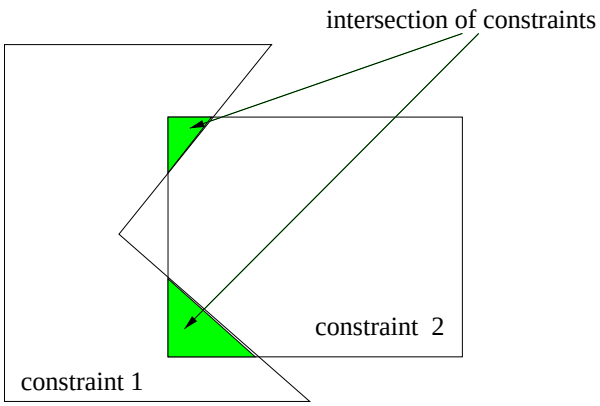
Given: closed sets ("constraints")

$$C_1, C_2, \ldots, C_N.$$

Wanted: a point ("solution")

$$x \in C_1 \cap C_2 \cap \cdots \cap C_N$$

of the Feasibility Problem.



The Feasibility Problem for $N = 2$.

Example. The Phase Retrieval Problem arises when $X = L^2$, $N = 2$, and the two constraints are as follows.

"space-limited image": for some set S,

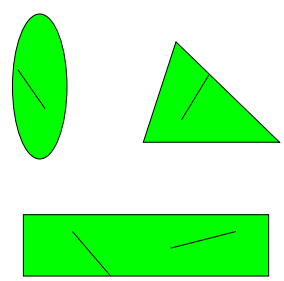
$$C_1 = \left\{ x \in L^2 : x \text{ is zero outside } S \right\};$$

"known Fourier magnitude": for some function $M(\omega)$,

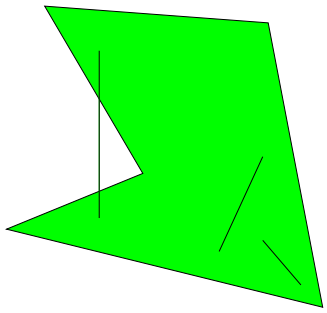
$$C_2 = \left\{ x \in L^2 : |\hat{x}(\omega)| = M(\omega), \quad \forall \omega \right\}.$$

C_1 is convex, but C_2 is not convex!

Definition. A set is called convex, if the line segment between any two points of the set lies entirely in the set.



Three convex sets.



A set that is not convex.

Phase Retrieval Problem revisited.
This problem is particularly difficult as it is a Feasibility Problem where one of the constraints — "known Fourier magnitude" — is not convex.

Here, I will focus on projection algorithms for solving convex feasibility problems. These methods can be viewed as "easier" variants of the Gerchberg-Saxton algorithm, and thus their convergence properties

- may suggest new approaches;
- adjust expectations for Gerchberg-Saxton.

2. Fundamental results

The Convex Feasibility Problem

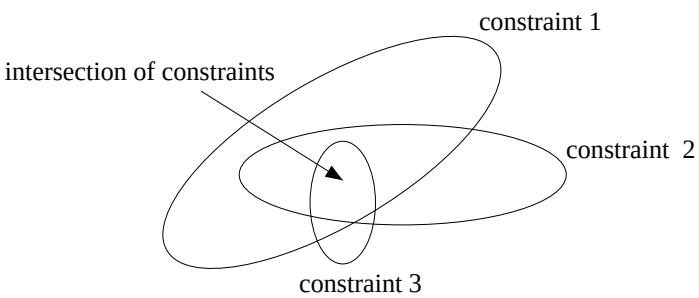
Given: closed convex sets ("constraints")

$$C_1, C_2, \ldots, C_N.$$

Wanted: a point ("solution")

$$x \in C_1 \cap C_2 \cap \cdots \cap C_N$$

of the Convex Feasibility Problem.

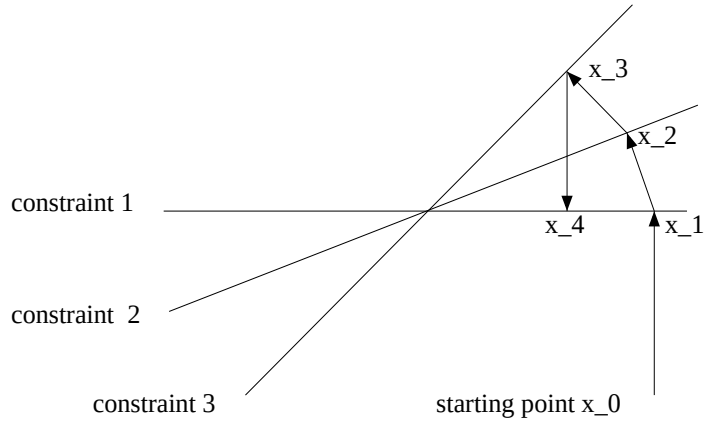


The Convex Feasibility Problem.

$\boxed{\text{The method of cyclic projections}}$

- P_i: (orthogonal) projection onto i^{th} set C_i.
- x_0: starting point of the algorithm.
- (x_n): sequence of cyclic projections given by

$$\begin{array}{lllllllll} x_0 & \xmapsto{P_1} & x_1 & \xmapsto{P_2} & x_2 & \xmapsto{P_3} & \cdots & \xmapsto{P_N} & x_N \\ & \xmapsto{P_1} & x_{N+1} & \xmapsto{P_2} & x_{N+2} & \xmapsto{P_3} & \cdots & \xmapsto{P_N} & x_{2N} \\ & \xmapsto{P_1} & x_{2N+1} & \xmapsto{P_2} & \cdots & . & & & \end{array}$$



The method of cyclic projections.

Some examples.

- each C_i is a hyperplane $\rightsquigarrow$ $Ax = b$;
- hyperplanes and orthant $\rightsquigarrow$ $Ax = b, x \geq 0$;
- functions that are band-limited and known on a segment $\rightsquigarrow$ Signal Recovery.

Some applications.

- Large-scale linear systems;
- Medical Imaging;
- Filter design;
- Enhancement of blocky JPEG images.

Phase Retrieval Problem revisited.
The Gerchberg-Saxton algorithm can be viewed as a method of cyclic projections for the two constraints.

"Fundamental Theorem" Bregman (1965). Suppose the problem is consistent, i.e.,

$$S := C_1 \cap C_2 \cap \cdots \cap C_N \neq \emptyset.$$

Then (x_n) converges weakly to a point in S.

- Hundal (2000): norm convergence can fail;
- von Neumann (1933) and Halperin (1962): norm convergence for subspaces;
- Gubin et al (1967): linear convergence in presence of "regularity";
- Youla, Stark (around 1980) popularized, applied and extended this to image restoration problems.

3. Recent developments

Unifying frameworks

Rather than studying cyclic projections, consider a much more flexibe general projection algorithm: starting point x_0; for $n \geq 0$, update via

$$x_{n+1} = x_n + \alpha_n \rho_n \sum_i w_{i,n}(P_{i,n}x_n - x_n).$$

- relaxation parameter $\alpha_n \in [0, 2]$.
- extrapolation parameter $\rho_n \geq 1$.
- weights $w_{i,n} \geq 0$, $\sum_i w_{i,n} = 1$.
- $P_{i,n}$ projection onto superset of C_i.

The general projection algorithm covers ...

⊕ the method of cyclic projections;
⊕ serial versions (under/over relaxations);
⊕ random projections;
⊕ parallel and block versions (e.g. Cimmino);
⊕ subgradient algorithms.

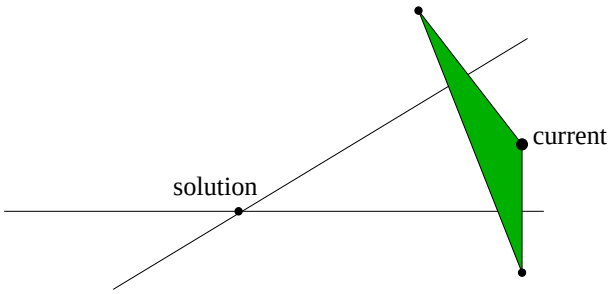
Many earlier results are covered by recent unifying work (B&Borwein, Censor, Kiwiel, ...).

Phase Retrieval Problem revisited.
If PRP were convex, then this framework would include the Fienup variant of Gerchberg-Saxton.

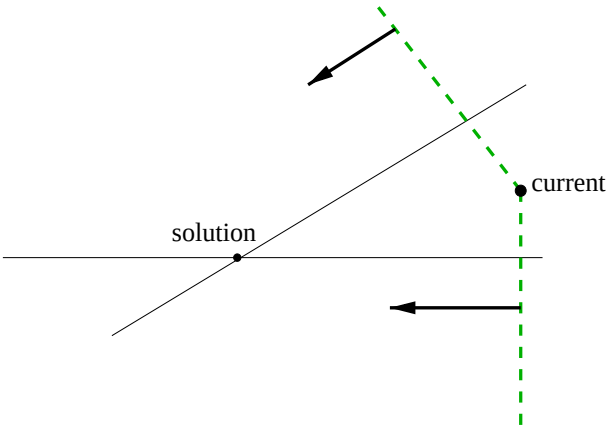
Extrapolation

Most projection methods had no extrapolation parameter ρ_n.

Given current x_n, the update x_{n+1} was chosen from "small" polytope (green triangle):



Pioneering work by Merzlyakov (1963), by Combettes (mid 1990s), and by Kiwiel and Lopuch (mid 1990s) introduced extrapolation. The next iterate x_{n+1} is chosen from a cone, which may already contain the solution:



Compared to more traditional methods,

projection algorithms with extrapolation

have the potential to converge

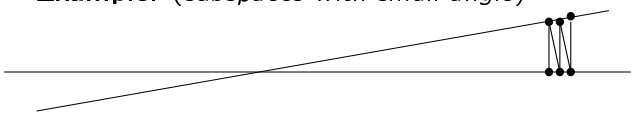
dramatically faster,

as demonstrated recently by Combettes. This feature is inherently parallel.

Regularity

The relationship among the constraints ("geometry of the problem") is responsible for how fast the method of cyclic projection converges.

Example. (*subspaces with small angle*)



Small "angle" yields slow convergence: "tunnels".

Fact. (B-Borwein-Lewis, 1996). Suppose each C_i is a subspace and $S = C_1 \cap C_2 \cdots \cap C_N$. Define the angle $\gamma(C_1, \ldots, C_N) \in [0, \pi/2]$ by

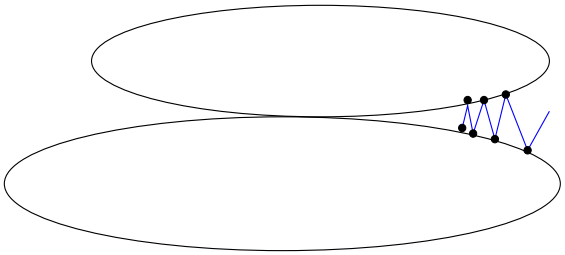
$$\cos\gamma = \|P_N \cdots P_1 - P_S\|.$$

Then

$$\|x_{kN} - P_S x_0\| \leq (\cos\gamma)^k, \quad \forall k.$$

$\oplus$ $\cos\gamma < 1$ in finite dimensions;
$\oplus$ $\cos\gamma < 1$ for hyperplanes;
$\ominus$ $\cos\gamma = 1$ and slow convergence is possible.

Example. (*only "touching" is bad*)



Mere touching yields slow convergence: "tunnels".

Fact. (Gubin-Polyak-Raik, 1967). If

$$C_N \cap \bigcap_{i=1}^{N-1} \operatorname{int}(C_i) \neq \emptyset,$$

then we have linear convergence: there exists $0 \leq \theta < 1$ and $s \in C_1 \cap C_2 \cap \cdots \cap C_N$ such that

$$\|x_{n+1} - s\| \leq \theta \|x_n - s\|, \quad \forall n.$$

Definition. (B-Borwein, 1992) The constraints $C_1, \ldots, C_N$ with intersection $S = C_1 \cap \cdots \cap C_N$ are regular, if for every bounded set B, there exists $\kappa > 0$ such that

$$d(x, S) \leq \kappa \max_i d(x, C_i), \quad \forall x \in B.$$

"If you are close to each constraint, then the intersection cannot be to far away."

Fact.
If the constraints are regular, then the method of cyclic projections — and more general variants — produces a sequence (x_n) that converges linearly to some solution.

This unifies and extends many earlier results on linearly convergent projection methods involving:
- subspaces;
- interiority assumptions;
- polyhedral sets.

4. Tools

Projections: existence and uniqueness

Fact. Suppose C is closed and convex, and $x \in X$. Then there exists a unique point in C, denoted $P_C(x)$ and called the projection of x onto C, that satisfies:

$$\|x - P_C(x)\| = d(x, C) = \min_{c \in C} \|x - c\|.$$

Phase Retrieval Problem revisited.
The "known Fourier magnitude" constraint is not convex. Projections are set-valued. The projection given in the G-S algorithm **well-defines one particular selection of the projection**. For this and new approaches to the PRP (perturbed Least Squares), see the recent preprint by Luke-Burke-Lyon (2001).

Convex projections are firmly nonexpansive

Fact. Suppose C is closed and convex. Then P_C is firmly nonexpansive: for all $x, y \in X$,

$$\|x-y\|^2 \geq \|P_C(x) - P_C(y)\|^2 + \|(I-P_C)(x) - (I-P_C)(y)\|^2.$$

Hence P_C is nonexpansive or "Lipschitz-1",

$$\|P_C(x) - P_C(y)\| \leq \|x-y\|, \quad \forall x, y \in X,$$

and thus continuous.

Phase Retrieval Problem revisited.
Continuity of the G-S projection is false.
May make G-S sensitive to numerical errors?

Fejér monotone sequences: basic properties

Definition. Suppose C is closed and convex, and (z_n) is a sequence. Then (z_n) is Fejér monotone with respect to C, if

$$\|z_{n+1} - c\| \leq \|z_n - c\|, \quad \forall n \geq 0, c \in C.$$

- (z_n) is bounded, and $d(z_n, C)$ decreases.
- (z_n) converges weakly to some point in C $\Leftrightarrow$ all weak cluster points of (z_n) lie in C.
- (z_n) converges in norm to some point in C $\Leftrightarrow$ $d(z_n, C) \to 0$.

Fejér monotonicity and projection methods

- all general projection methods are Fejér monotone with respect to $S = C_1 \cap \cdots \cap C_N$.
- design of extrapolation parameters is guided by Fejér monotonicity.

Phase Retrieval Problem revisited.
The Fienup measure, see also Levi&Stark,

$$\Delta(x) = d(x, C_1) + d(x, C_2)$$

is decreasing: $\Delta(x_{n+1}) \leq \Delta(x_n)$, $\forall n$.
This fails for $N \geq 3$, even in the convex case!

Remaining tools are ... convex ...

- Convex Relations Theory from Functional/Set-Valued Analysis (for regularity).
- Fixed Point Theory of nonexpansive maps — fixed point sets are convex.
- maximimal monotone operators (for inconsistent constraints): projection P_C onto a closed set C is always monotone:

$$\langle P_C(x) - P_C(y), x - y \rangle \geq 0, \quad \forall x, y;$$

P_C is "maximal monotone" $\Leftrightarrow$ C is convex.

5. Open problems

Random order

Suppose $N \geq 3$, and let

$$r : \mathbb{N} \to \{1, 2, \ldots, N\}$$

be a random map (an "N-die"), i.e., r is onto and takes up every value infinitely often.

When and in what sense does the sequence

$$x_n := P_{r(n)}(x_{n-1})$$

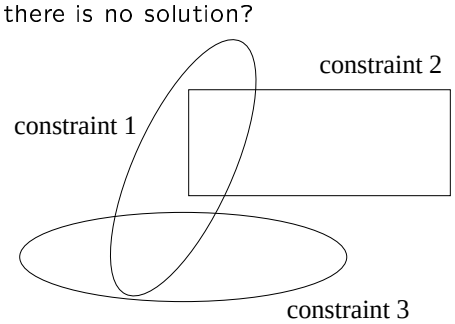
converge to a point in $C_1 \cap \ldots \cap C_N$?

$\oplus$ In Medical Imaging, random methods yield faster convergence than cyclic methods;
$\oplus$ known convergence results rely on interiority, regularity, or combinatorial arguments.

$\ominus$ weak convergence in general open;
$\ominus$ norm convergence open for subspaces.

Inconsistent constraints

What happens if we apply cyclic projections, but there is no solution?



The inconsistent case for $N = 2$

For clarity, write: A and B for the constraints, b_0 for the starting point in X, and

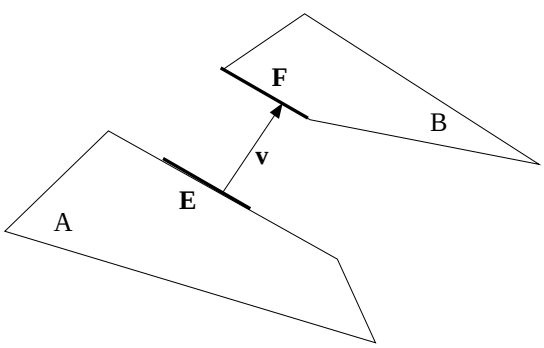
$$b_0 \overset{P_A}{\longmapsto} a_1 \overset{P_B}{\longmapsto} b_1 \overset{P_A}{\longmapsto} a_2 \overset{P_B}{\longmapsto} \cdots$$

for the sequence of alternating projections.

Fact. Define the displacement vector by

$$v := P_{\mathsf{cl}\,(B-A)}(0),$$

and let $E := A \cap (B - v)$ and $F := B \cap (A + v)$.
Then $E = \mathsf{Fix}\,(P_A P_B)$ and $F = \mathsf{Fix}\,(P_B P_A)$.
If $e \in E$, then $P_B(e) = P_F(e) = e + v$.



The inconsistent case $N = 2$.

Fact. We always have

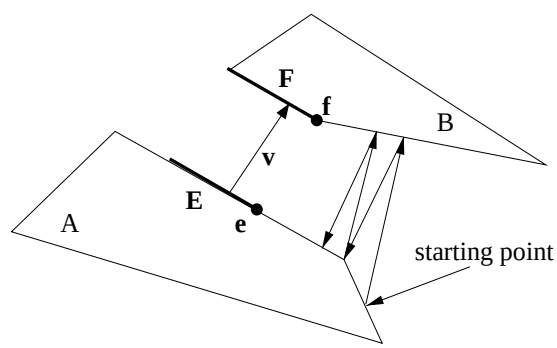
$$b_n - a_n, b_n - a_{n+1} \longrightarrow v.$$

Also, we have the following Dichotomy.
Either: E and F are both nonempty and

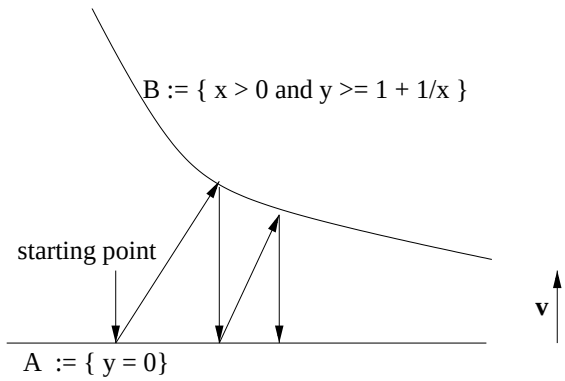
$$a_n \overset{\mathsf{w}}{\rightharpoonup} e, \quad b_n \overset{\mathsf{w}}{\rightharpoonup} f = e + v,$$

for some $e \in E$ and $f \in F$.
Or: E and F are both empty and

$$\|a_n\|, \|b_n\| \longrightarrow +\infty.$$



Dichotomy: the "Either" case.



Dichotomy: the "Or" case.

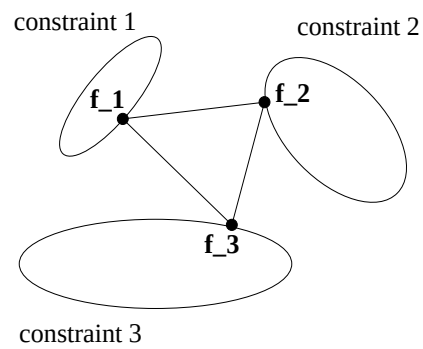
$\boxed{\text{The inconsistent case } N \geq 3}$

WLOG $N = 3$. $(f_1, f_2, f_3) \in C_1 \times C_2 \times C_3$ is a cycle, if

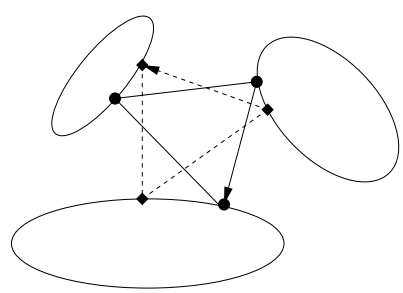
$$f_1 \xmapsto{P_2} f_2 \xmapsto{P_3} f_3 \xmapsto{P_1} f_1.$$

Set

$$F_1 := \mathsf{Fix}\,(P_1 P_3 P_2),$$
$$F_2 := \mathsf{Fix}\,(P_2 P_1 P_3),$$
$$F_3 := \mathsf{Fix}\,(P_3 P_2 P_1).$$



A cycle.



Cycles are order dependent.

Fact. (Dichotomy)

Either: each F_i is nonempty,

$$(x_{3n+1}, x_{3n+2}, x_{3n+3}) \xrightarrow{\mathrm{w}} (f_1, f_2, f_3)$$

for some cycle (f_1, f_2, f_3), and

$$x_{3n+i+1} - x_{3n+i} \to d_i,$$

for some difference vectors d_i independent of the cycle.

Or: each F_i is empty, and $\|x_n\| \to +\infty$.

Open problem: what are those vectors d_i?

Fact. (B, 2001). The composition

$$T := P_N P_{N-1} \cdots P_1$$

is asymptotically regular, i.e.,

$$T^k x - T^{k+1} x \to 0, \quad \forall x \in X.$$

Remark. In the language of Levi&Stark, "traps" or "tunnels" always exist.

Underrelaxations

What if, in the method of cyclic projections, we replace at every step the projection P_i by a more general underrelaxed projection

$$(1-\lambda)I + \lambda P_i,$$

for some fixed $\lambda \in (0, 1]$?

Remark. This is similar to the Fienup variant.

Fact. (Censor-Eggermont-Gordon; 1983)
If each C_i is an affine subspace, then:

- the sequence $\left(x_{Nn+1}, \dots, x_{Nn+N}\right)$ converges in norm to some $\left(x_1(\lambda), \dots, x_N(\lambda)\right)$.

- as $\lambda \to 0^+$, the net $\left(x_1(\lambda), \dots, x_N(\lambda)\right)$ converges to $(\xi, \dots, \xi) \in X^N$, where ξ is a least-squares solution of the optimization problem

$$\min_{x \in X} \sum_{i=1}^{N} d^2(x, C_i).$$

De Pierro's Conjecture.
This is essentially true for the non-affine case as well, although least-squares solutions may not exist.

Remark. Chris Hamilton, a recent graduate from OUC, has written nice Java code that gives further support to De Pierro's conjecture.

References

• Butnariu-Censor-Reich (editors), Conference proceedings of workshop on "inherently parallel algorithms" (Haifa 2000), to appear.

• B: The composition of projections onto closed convex sets in Hilbert space is asymptotically regular, *CECM Preprint* 01:163, 2001. To appear in *Proc AMS*.

• Contact info:

bauschke@cecm.sfu.ca