Analysis of Generalized Pattern Search

Charles Audet

John Dennis

Rice University, Houston 2000

www.caam.rice.edu/~charlesa/

Outline

- The optimization problem
- Pattern search algorithms and convergence results
- Generalized derivatives and stronger results
- Bound and linearly constrained optimization
- General constraints and filter methods
- GPS and filter convergence results
- Discussion

General optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \subset \Re^n, \end{array}$$

where $f : \Re^n \to \Re \cup \{\infty\}$ may be discontinuous or infinite valued, and:

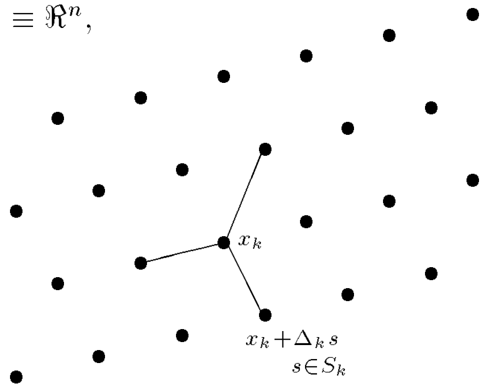
- $f(x)$ is usually given as a black box,
- $f(x)$ is expensive and has few correct digits,
- $f(x)$ may fail expensively and unexpectedly even for $x \in \Omega$,
- approximation of derivatives is problematic.

Meshes for unconstrained optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \equiv \Re^n, \end{array}$$

At iteration k, the algorithm searches on the current mesh M_k constructed by positive spanning set S_k.

The coarseness or fineness of the mesh is given by $\Delta_k > 0$.



A simple pattern search algorithm for unconstrained optimization

Let X_k be the mesh points adjacent to x_0 with respect to the positive spanning set spanning set S_k: $X_k = \{x_k + \Delta_k s : s \in S_k\}$.

Given Δ_0, $x_0 \in M_0$ for $k = 0, 1, \cdots$, do

1. Employ some finite strategy to try to choose $x_{k+1} \in M_k$ such that $f(x_{k+1}) < f(x_k)$ and then set $\Delta_{k+1} = \Delta_k$;
2. Else if x_k minimizes $f(x)$ for $x \in X_k$, then set $x_{k+1} = x_k$ and $\Delta_{k+1} = \Delta_k/2$.

Convergence results

Torczon shows that under the assumptions

1. f is continuously differentiable
2. all iterates are in a compact set

the following hold

1. the mesh gets infinitely fine (Δ_k goes to zero)
2. $\exists$ a limit point $\hat{x}$ for which $\nabla f(\hat{x}) = 0$.

Generalized derivatives (from Clarke)

Let f be *Lipschitz near*† *a given point* x and v be a vector in $\Re^n$.

The *generalized directional derivative* of f at x in the direction v is

$$f^\circ(x;v) \ := \ \limsup_{y \to x,\, t \downarrow 0} \frac{f(y+tv)-f(y)}{t}.$$

(†) i.e., there exists an $\epsilon > 0$ and a constant $L > 0$ such that $\|f(y) - f(y')\| \leq L\|y - y'\|$ whenever $\|y - x\| \leq \epsilon$ and $\|y' - x\| \leq \epsilon$.

Results on generalized derivatives

f is *regular* at x if for all v, the one-sided directional derivative exists and coincides with $f^\circ(x;v)$.

f *strictly differentiable* at x

$$\text{if} \lim_{y \to x,\, t \downarrow 0} \frac{f(y+tv)-f(y)}{t} = \nabla f(x)^T v \text{ for all } v.$$

If f is Lipschitz

$$\left.\begin{array}{c} f \text{ continuously} \\ \text{differentiable} \end{array}\right\} \Longrightarrow \left.\begin{array}{c} f \text{ strictly} \\ \text{differentiable} \end{array}\right\} \Longrightarrow f \text{ regular.}$$

Our new result: Stronger convergence

Since $\Delta_k \to 0 \Rightarrow \exists$ infinitely many unsuccessful iterations:

We let:

- $\hat{x}$ be the limit of a subsequence of unsuccessful iterates.
- $\hat{S}$ be the set of search directions used infinitely many times in this subsequence.

And then:

$$\left.\begin{array}{l} f \text{ is Lipschitz near } \hat{x} \\ \text{all iterates are in a compact set} \end{array}\right\} \quad \Longrightarrow \quad \begin{array}{l} f^\circ(\hat{x}; s) \geq 0 \\ \text{for every } s \in \hat{S}. \end{array}$$

Discussion of new results -I-

- f strictly differentiable at $\hat{x} \Rightarrow \nabla f(\hat{x}) = 0$ (Torczon's proof needs $\mathcal{C}^1$ - not suited to Clarke).
- Our proof adds that $\nabla f(x_k) \to 0$ for *any* convergent subsequence of unsuccessful iterates $\{x_k\}_{k \in K}$.

Key to the main lemma follows from the definition:

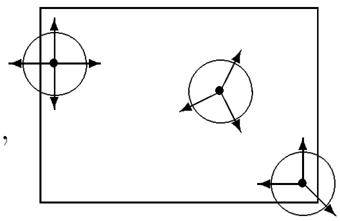
$$\begin{aligned} f^\circ(\hat{x}; s) &= \limsup_{y \to \hat{x},\ t \downarrow 0} \frac{f(y + ts) - f(y)}{t} \\ &\geq \limsup_{k \in K} \frac{f(x_k + \Delta_k s) - f(x_k)}{\Delta_k} \ \geq \ 0. \end{aligned}$$

Discussion of new result -II-

- Algorithm produces a limit point $\hat{x}$ even if f is discontinuous or extended valued.
- f is Lipschitz near $\hat{x}$ $\Rightarrow$ $f^\circ(\hat{x}; s) \geq 0 \;\; \forall s \in \hat{S}$.
- f is regular at $\hat{x}$ $\Rightarrow$ $f'(\hat{x}; s) \geq 0 \;\; \forall s \in \hat{S}$.
- f strictly differentiable at $\hat{x}$ $\Rightarrow$ $\nabla f(\hat{x}) = 0$.

Bound constrained optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \equiv \{x \mid a \leq x \leq b\}, \end{array}$$

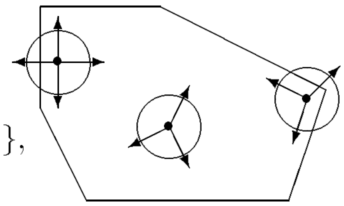


Reject infeasible points. Near a boundary point, use S_k that contains its tangent cone generators. Then:

- $f^\circ(\hat{x}; s) \geq 0 \; \forall s \in \hat{S}$, generators of $T_\Omega(\hat{x})$.
- f is strictly differentiable at $\hat{x}$ $\Rightarrow$ $\nabla f(\hat{x})^T w \geq 0 \; \forall w \in T_\Omega(\hat{x})$.
- $-\nabla f(\hat{x}) \in N_\Omega(\hat{x})$ and $\hat{x}$ is a KKT point.

Linearly constrained optimization

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \equiv \{x \mid \ell \le Ax \le u\}, \end{array}$$



Reject infeasible points. Near a boundary point, use S_k that contains its tangent cone generators. Then:

- $f^\circ(\hat{x}; s) \ge 0 \ \forall s \in \hat{S}$, generators of $T_\Omega(\hat{x})$.
- f is strictly differentiable at $\hat{x} \ \Rightarrow \ \nabla f(\hat{x})^T w \ge 0 \ \forall w \in T_\Omega(\hat{x})$.
- $-\nabla f(\hat{x}) \in N_\Omega(\hat{x})$ and $\hat{x}$ is a KKT point.

Positive spanning sets

- A richer set of S_k increases the number of directions in $\hat{S}$, and thus the directions for which $f^\circ(\hat{x}; s) \ge 0$.
- User's domain specific knowledge can help choose S_k.
- Theory limited to a finite number of different S_k, so the *barrier* approach (reject infeasible points) does not apply to general constraints because a finite number can not conform to $\partial\Omega$.

General nonlinear constrained optimization

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega \equiv \{x \mid C(x) \leq 0\}, \\ \text{where} \quad & C : \Re^n \to \Re^m \text{ with } C = (c_1, \ldots, c_m)^T \ . \end{aligned}$$

Define the nonnegative constraint violation function $h : \Re^n \to \Re$

$$h(x) = \sum_j \max(0, c_j(x))^2.$$

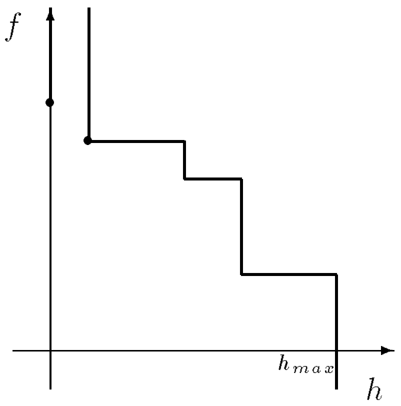
Note that $h(x) = 0$ iff $x \in \Omega$, and h inherits smoothness from C.

We look at the biobjective optimization problem where a priority is given to the minimization of h over the minimization of f.

Combining pattern search and filter methods

Each time a trial point x is considered, $h(x)$ and $f(x)$ are plotted the bi-loss map.

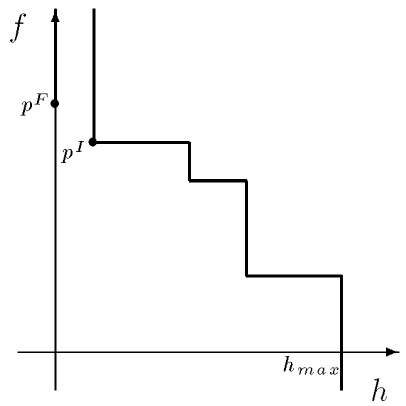
An iteration is declared *successful* if an unfiltered point is found in either the search or the poll step.



Polling centers

Before declaring an iteration *unsuccessful* polling must be completed around the poll center p_k.

p_k is chosen to be either the feasible incumbent solution p^F or the most feasible infeasible filter point p^I.



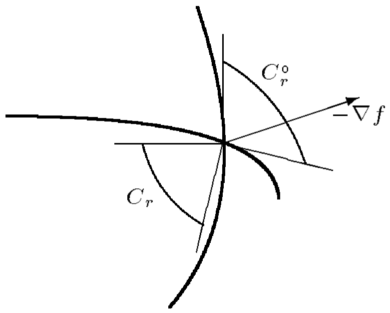
Results for constrained optimization

i) h Lipschitz near $\hat{x}$ $\Rightarrow h^{\circ}(\hat{x}; s) \geq 0$ for any $s \in \hat{S}$.

ii) h strictly differentiable at $\hat{x} \Rightarrow \nabla h(\hat{x}) = 0$.

iii) $h(p_k) = h(p_k + \Delta_k s)$ for a subsequence $\Rightarrow f^{\circ}(\hat{x}; s) \geq 0$.

iv) Let C_r be the cone generated by all the associated directions of $\hat{S}$ that satisfy the conditions above. If f is strictly differentiable, then $-\nabla f(\hat{x})$ belongs to the polar C_r° of C_r.



Discussion

- The search step is the key to practical efficiency.
  The Surrogate Management Framework seamlessly incorporates user provided heuristic searches.
- Polling ensures robustness and provides rigor.
- Using Clarke's tools simplifies, shortens, and strengthens pattern search convergence analysis.
- Filters handle nonlinear constraints without using derivatives, Lagrange multipliers or penalty parameters.