Direct Search Methods

Virginia Torczon

Department of Computer Science
College of William & Mary

Institute for Computer Applications in Science and Engineering

Center for Research on Parallel Computation

The Problem

$$\min_{x \in S \subseteq \mathbf{R}^n} \qquad f(x)$$

Assume:

$f : \mathbf{R}^n \to \mathbf{R}$

$f \in C^1$ (for analytical results)

$S = \mathbf{R}^n$ (for simplicity)

What are “direct search” methods?

Phrase was first coined in:

“Direct search” solution of numerical and statistical problems
Robert Hooke and T.A. Jeeves
1961

"We use the phrase "direct search" to describe sequential examination of trial solutions involving comparison of each trial solution with the "best" obtained up to that time together with a strategy for determining (as a function of earlier results) what the next trial solution will be."

"The phrase implies our preference, based on experience, for straightforward search strategies which employ no techniques of classical analysis except where there is a demonstrable advantage in doing so."

Hooke and Jeeves (1961)

What do they mean by "classical analysis"?

Method of Steepest Ascent

(Newton's Method)

What do they mean by "straightforward search"?

Do not construct a model of the function; instead, "query" or "sample" the function directly

(they specifically exclude polynomial approximations of the function)

Two approaches:

No Model of Function

classical:	grid search
modern:	pattern search methods
	("adaptive" grid search methods)

Polynomial Model of Function

classical:	steepest ascent/descent
	Newton's method
modern:	globalized quasi-Newton methods

Earlier History

On the Experimental Attainment of Optimum Conditions
G.E.P. Box and K.B. Wilson
1951

$\Rightarrow$ Experimental Design

$\Rightarrow$ Response Surface Methodology

Box's Scheme:

1. design a simple experiment (a "pattern")

2. sample process at points in pattern

3. use linear regression to build a first-order response surface (i.e., a linear model)

4. apply method of steepest ascent

5. repeat 1–4 until decide:
   "in the neighborhood of a solution"

6. design a more elaborate experiment (still a pattern, but with more points)

7. sample process at points in new pattern

8. use linear regression to build a second-order response surface (i.e., a quadratic model)

9. refine solution; do sensitivity analysis

A New Approach

Evolutionary Operation: A Method for Increasing Industrial Productivity
G.E.P. Box
1957

$\Rightarrow$ Direct Search Methods

Box's Evolutionary Operation:

1. start with a simple experimental design (a "pattern")

2. center the pattern at the current "best" point

3. sample process at points in pattern

4. choose "best" point in pattern (resample and average responses to account for error)

5. return to 2.

Subsequent Algorithmic Developments:

- *An automatic method for finding the greatest or least value of a function*
  H.H. Rosenbrock
  1960

- *"Direct search" solution of numerical and statistical problems*
  Robert Hooke and T.A. Jeeves
  1961

- *Sequential application of simplex designs in optimization and evolutionary operation*
  W. Spendley, G.R. Hext, and F.R. Himsworth
  1962

- *An efficient method for finding the minimum of a function of several variables without calculating derivatives*
  M.J.D. Powell
  1964

- *A simplex method for function evaluation*
  J.A. Nelder and R. Mead
  1965

•
•
•

Why?

- f is statistical in nature
- calculation of f is inaccurate
- derivatives are either not available or not reliable
- f is highly nonlinear
- f is nondifferentiable
- the first partial derivatives of f are discontinuous
- only rank (order) information is available

“Final” Judgment (Round One):

Direct Search Methods
W.H. Swann
in Numerical Methods for Unconstrained Optimization
W. Murray (ed.)
1972

“Although the methods described above have been developed heuristically and no proofs of convergence have been derived for them, in practice they have generally proven to be robust and reliable in that only rarely do they fail to locate at least a local minimum of a given function, although sometimes the rate of convergence can be slow.”

Theoretical Developments

- *An efficient method for finding the minimum of a function of several variables without calculating derivatives*
  M.J.D. Powell, 1964
- *Lattice approximations to the minima of functions of several variables*
  Gerald Berman, 1969
- Optimisation: théorie et algorithmes
  J. Céa, 1971
- Computational Methods in Optimization: A Unified Approach
  E. Polak, 1971

- *Optimization of functions whose values are subject to small errors*
  Torkel Glad and Allen Goldstein, 1977

- *Positive basis and a class of direct search techniques*
  Yu Wen-ci, 1979

- *The convergence property of the simplex evolutionary techniques*
  Yu Wen-ci, 1979

- *Simplex direct search algorithms*
  A.S. Rykov, 1980

- *Simplex methods of direct search*
  A.S. Rykov, 1980

- *Simplex algorithms for unconstrained minimization*
  A.S. Rykov, 1983

- An Interactive Approach for Solving Multi-Objective Optimization Problems
  Daniel J. Woods, 1985

- *On the convergence of the multidirectional search algorithm*
  Virginia Torczon, 1989/1991

- *On the convergence of pattern search algorithms*
  Virginia Torczon, 1993/1994/1995
  (to appear in 1997)

- *Fortified-descent simplicial search method: a general approach*
  Paul Tseng, 1995

- *Pattern search algorithms for bound constrained minimization*
  Virginia Torczon and Robert Michael Lewis
  1996

- *Convergence properties of the Nelder-Mead simplex algorithm in low dimensions*
  Jeffrey C. Lagarias, James A. Reeds,
  Margaret H. Wright, and Paul E. Wright
  1996

$$\vdots$$

Hooke and Jeeves Revisited

"In discussing direct search with others, we have found that certain questions are repeatedly asked...

1. *Under what conditions does direct search converge to the* [sic] *solution?*

2. *Once an approximate solution is found, how good is it?*

3. *What is the effect of error, if any, in the function determination?"*

Hooke and Jeeves (1961)

“Direct Search” Methods:

Stochastic

- random search
- genetic algorithms

Deterministic

- Rosenbrock’s method
- Sequential Simplex Design of Spendley, Hext, and Himsworth
- Nelder–Mead simplex method

- grid search
- alternating variable method
- pattern search method of Hooke and Jeeves
- evolutionary operations
- multidirectional search algorithm
- DIRECT

Generalized Pattern Search

Given $x_0 \in \mathbf{R}^n$ and $\Delta_0 > 0$, for $k = 0, 1, \cdots$,

1. Compute $f(x_k)$.

2. Determine a step s_k using *Exploratory Moves*(Δ_k, P_k).

3. If $f(x_k) > f(x_k + s_k)$, then

   $$x_{k+1} = x_k + s_k.$$

   Otherwise

   $$x_{k+1} = x_k.$$

4. Update (Δ_k, P_k).

The Pattern

$$P_k = BC_k,$$

where

$$\begin{array}{ll} B \in \mathbf{R}^{n\times n} & \text{(nonsingular)} \\ C_k \in \mathbf{Z}^{n\times p},\ p > 2n,\ (p >> n) & \end{array}$$

with

$$C_k = [\underbrace{M_k}_{n}\ \underbrace{-M_k}_{n}\ \underbrace{L_k}_{p-2n}].$$

Trial Step

$s_k^i \in \Delta_k P_k$, where $\Delta_k \in \mathbf{R}$.

Hypotheses on Exploratory Moves

1. $s_k \in \Delta_k P_k$.

2. If
   $\min\{f(x_k + y), y \in \Delta_k B[M_k \ \ -M_k]\} < f(x_k)$,
   then
   $f(x_k + s_k) < f(x_k)$.

Theorem (Torczon ‘93,‘94,‘95)

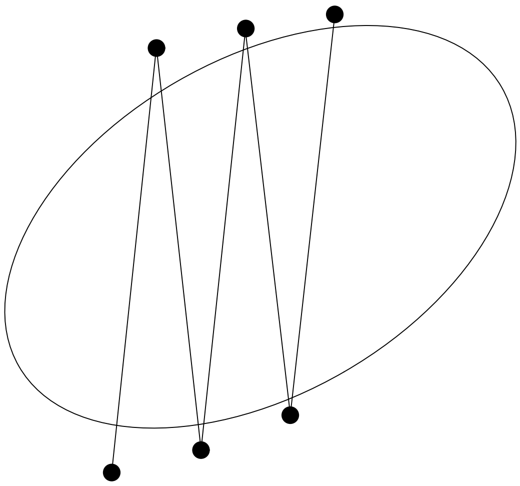
Assume that $L(x_0) = \{x : f(x) \leq f(x_0)\}$ is compact and that the function $f : \mathbf{R}^n \to \mathbf{R}$ is continuously differentiable on $L(x_0)$. Then for the sequence of iterates $\{x_k\}$ produced by the generalized pattern search method,

$$\liminf_{k \to +\infty} \|\nabla f(x_k)\| \quad = \quad 0.$$

Key ideas for proof:

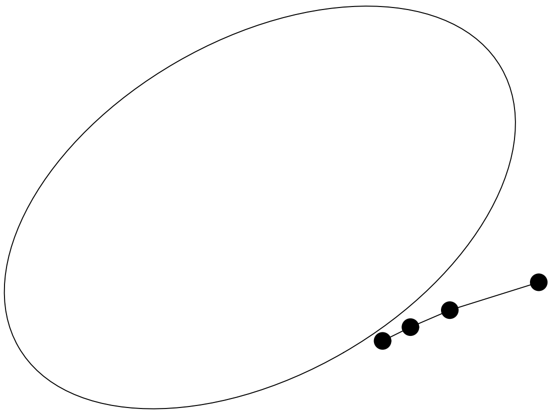
- descent methods
- gradient-related methods
- step length control

Possible pathologies:



decrease too small
relative to the
length of the step

$\Longrightarrow$ need for step length control



decrease too small
relative to the
norm of the gradient

Theorem (Torczon '93/'94/'95)

Any iterate x_N produced by a pattern search method can be expressed in the form:

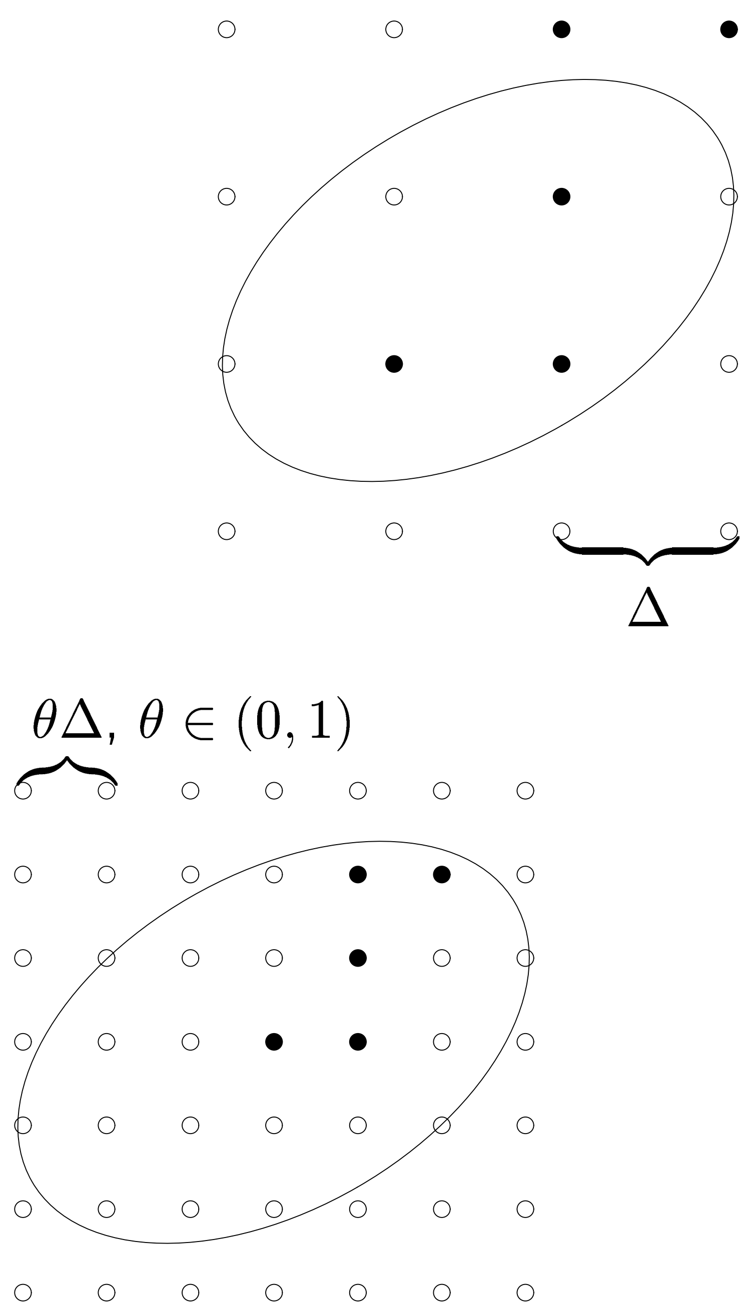
$$x_N \quad = \quad x_0 + (\beta^{r_{LB}} \alpha^{-r_{UB}}) \Delta_0 B \sum_{k=0}^{N-1} z_k$$

where

- $\beta/\alpha \equiv \tau$, with $\alpha, \beta \in \mathbf{N}$ and relatively prime, and τ is as defined in the general algorithm for updating Δ_k,
- r_{LB} and r_{UB} depend on N, and
- $z_k \in \mathbf{Z}^n$, $k = 0, \ldots, N-1$.

$\Rightarrow$ iterates lie on a scaled translated integer lattice

Adaptive Grid Search Methods



Computational considerations:

In 1960's

− computers were expensive
− computer access was limited
+ mathematicians were (relatively) inexpensive

In 1990's

+ computers are relatively inexpensive
+ computer time and access (at least in the U.S.) is essentially unlimited
− technical people (computer scientists/ numerical analysts/programmers/statisticians) are relatively expensive

Computational Attributes
(of pattern search methods)

- quickly and easily programmed
- computationally compact
- sampling methods
(potential for large number of function evaluations)
- embarrassingly parallel

Strategy # 1

Dennis/Torczon
(circa 1991)

The “big hammer” approach:

use a large number of function evaluations per iteration—computed in parallel—for the basic search strategy

Goals:

- simple search strategy
- ease of use
- wide applicability

Strategy # 2

Dennis/Torczon and Dennis/Serafini/Torczon (circa 1995)

Return to modeling!

use a model of the function (probably *not* a linear or quadratic model)—which is provided by the user—to determine a search strategy

Goals:

- exploit features of specific problem
- require *very* few evaluations of the "true" function
- couple modeling concerns with optimization concerns to work in tandem

We are not alone:

- *A derivative-free algorithm for parallel and sequential optimization*
  A.G. Buckley and H. Ma, 1994

- *A direct search optimization method that models the objective and constraint functions by linear interpolation*
  M.J.D. Powell, 1994

- *A direct search method that models the objective function by quadratic interpolation*
  M.J.D. Powell, 1994

- *An algorithm using quadratic interpolation for unconstrained derivative-free optimization*
  A.R. Conn and Ph.L. Toint, 1995

- *A grid algorithm for bound constrained optimization of noisy functions*
  C. Elster and A. Neumaier, 1995

$$\vdots$$

Strategy # 3

Dennis/Serafini/Torczon/Trosset (circa 1996)

Some combination of both Strategy #1 and Strategy #2.

Goals:

- exploit computational parallelism
- determine a good initial candidate solution
- determine features of specific problem (i.e., build a good mid-range or global model)
- build a data base of information to support future queries

When to use:

- f is statistical in nature
- calculation of f is inaccurate
- derivatives are either not available or not reliable
- f is highly nonlinear
- f is nondifferentiable
- the first partial derivatives of f are discontinuous
- only rank (order) information is available

What to expect:

- potential for large number of function evaluations
  they are sampling methods

- effective for problems with relatively small number of variables
  curse of dimensionality

- good global behavior
  they are gradient-related methods

- slow local convergence
  they are not using a model of the function; no second-order information

- useful exploratory tool
  wide applicability, ease of use

Acknowledgments:

John Dennis

Michael Lewis

Michael Trosset

Air Force Office of Scientific Research
Army Research Office
Department of Energy
NASA Ames Research Center/IBM
NASA Langley Research Center/ICASE
National Science Foundation/CRPC
Sandia National Laboratory,
Lawrence Livermore
The State of Texas