Pattern Search Algorithms for Mixed Variable Programming

Charles Audet

John Dennis

Rice University, Houston 2000

www.caam.rice.edu/~charlesa/

Outline

- The optimization problem with categorical variables
- Example: A thermal insulation system
- Pattern search algorithms
- Theoretical convergence results
- Numerical results on the example
- Discussion

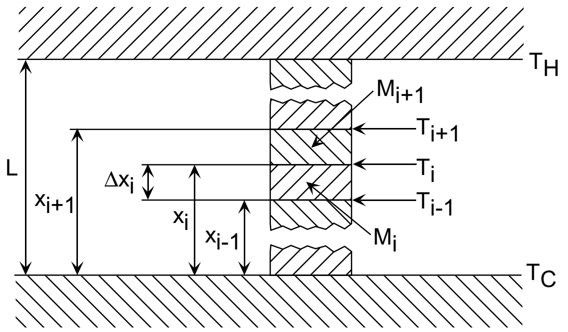
MVP optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \;=\; (x^c, x^d) \;\in\; \Omega \equiv [a,b] \;\subset\; \Re^{n^c} \times \mathcal{Z}^{n^d}, \end{array}$$

where f 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,
- continuous relaxation is unavailable.

Example: A thermal insulation system



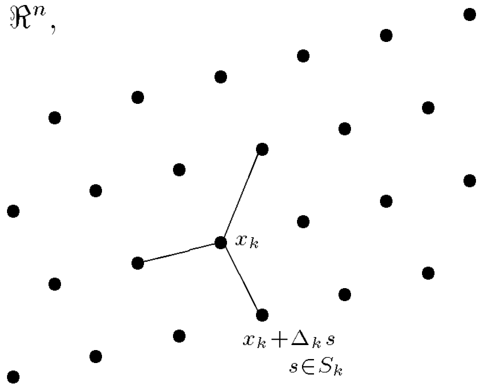
$$\begin{array}{ll} \displaystyle\min_{\mathbf{\Delta x}, \mathbf{T}, n, \mathbf{M}} & power(\mathbf{\Delta x}, \mathbf{T}, n, \mathbf{M}) \\ \text{subject to} & \mathbf{\Delta x} \geq \mathbf{0}, \quad T_C \leq \mathbf{T} \leq T_H, \\ & n \in \mathbb{N}, \qquad \mathbf{M} \in Materials. \end{array}$$

Meshes for unconstrained optimization

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

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 $\Delta_{k+1} = \Delta_k/2$.

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 $\|y' - x\| \leq \epsilon$.

Convergence results

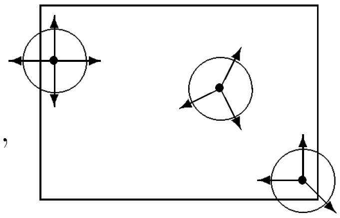
If all iterates are in a compact set then

- [Torczon 1997] The mesh gets infinitely fine (Δ_k goes to zero).
- For every limit point $\hat{x}$ of any subsequence $\{x_k\}_{k \in K}$ of unsuccessful iterates, and for the set $\hat{S}$ of search directions used infinitely many times in this subsequence
  - If f is Lipschitz near $\hat{x}$, then $f^\circ(\hat{x};s) \geq 0$ for every $s \in \hat{S}$

    PROOF: $$f^\circ(\hat{x};s) \ := \ \limsup_{y \to \hat{x}, \ t \downarrow 0} \frac{f(y+ts) - f(y)}{t} \ \geq \ \limsup_{k \in K} \frac{f(x_k + \Delta_k s) - f(x_k)}{\Delta_k} \ \geq \ 0.$$
  - If f is regular at $\hat{x}$, then $f'(\hat{x};s) \geq 0$ for every $s \in \hat{S}$.
  - If f is strictly differentiable at $\hat{x}$, then $\nabla f(\hat{x}) = 0$.

Bound constrained optimization

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



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

- $f^\circ(\hat{x}; s) \ge 0 \ \forall s \in \hat{S}$, 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 (e.g., linear constraints).
- 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 $\partial\Omega$.

Set of neighbors for POLL step

The pattern search algorithm for MVP consists again of two steps:

SEARCH and POLL.

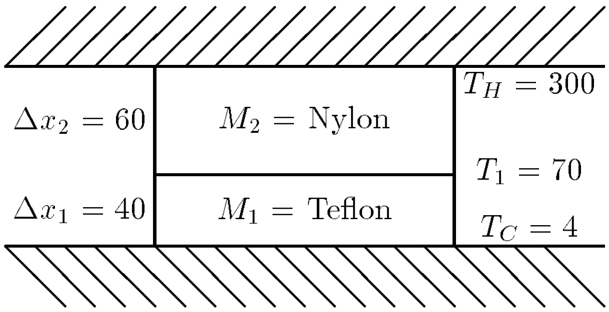
The SEARCH is still a finite search on the mesh, free of any rules.

The POLL is the failsafe step that evaluates function values at mesh neighbors for the continuous variables, and in a user-defined set of neighbors $\mathcal{N}(x_k)$.

This set of neighbors defines a notion of *local optimality.*

Illustration of a set of neighbors

Heat shield example:



Mesh neighbors: Vary either Δx_1 or Δx_2 or T_1 by $\pm\Delta_k$.

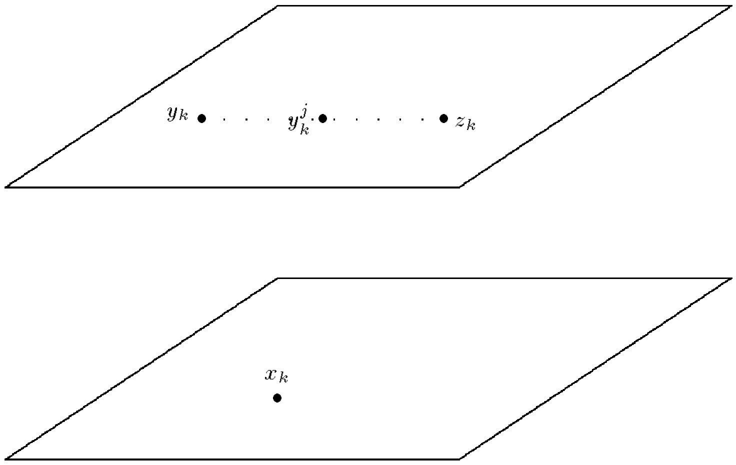
Set of neighbors $\mathcal{N}$:
- Change M_1 to Nylon or to Fiberglass.
- Change M_2 to Teflon or to Fiberglass.
- Add a shield and a material.
- Remove a shield and a material.

Pattern search algorithm for mixed variables

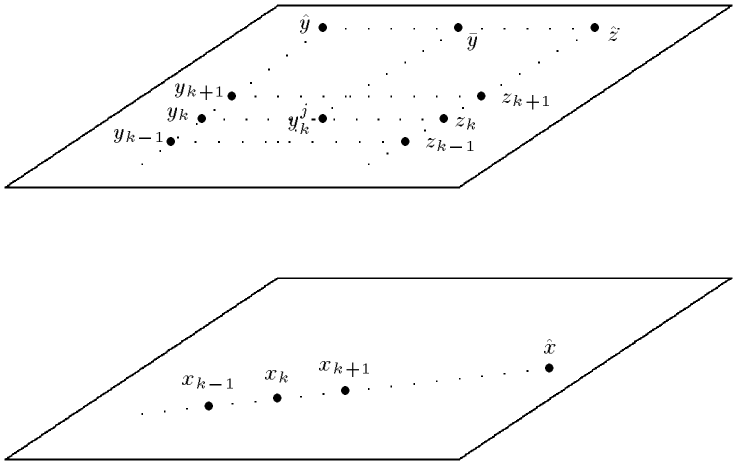
Given Δ_0, $x_0 \in M_0$ for $k = 0, 1, \cdots$, do

1. SEARCH on the mesh M_k;
2. POLL in the set of mesh neighbors X_k and on the set of neighbors $\mathcal{N}(x_k)$.
3. EXTENDED POLL STEP: If $f(y)$ is within ξ of $f(x_k)$ for some y in $\mathcal{N}(x_k)$, then fix the discrete variables at y^d and POLL until a local mesh minimizer is found or until a solution whose objective function value is less than $f(x_k)$ is found.

Extended poll



EXTENDED POLL



Convergence results

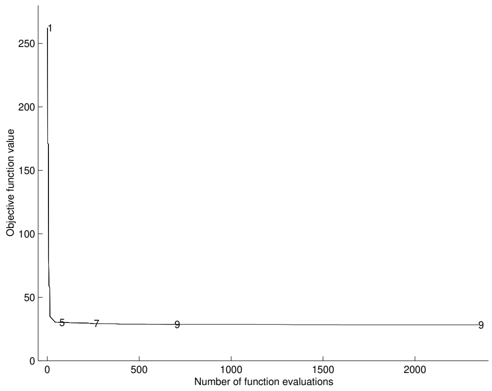
If f is strictly differentiable at $\hat{x}$

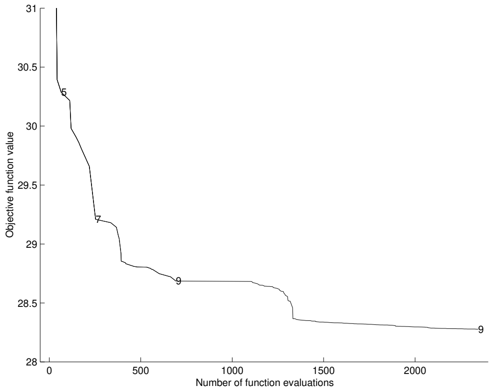
- $(x^c - \hat{x}^c)^T \nabla^c f(\hat{x}) \geq 0$ for any feasible $(x^c, \hat{x}^d)$.
- $f(\hat{x}) \leq f(\hat{y})$ for all $\hat{y} \in \Omega \cap \mathcal{N}(\hat{x})$.
- In the cases where $f(\hat{y}) < f(\hat{x}) + \xi$ (for a specified $\xi > 0$) then there exists a point $\hat{z} \in \Omega$ whose discrete components $\hat{z}^d$ are identical to $\hat{y}^d$ that satisfies
  - $f(\hat{x}) \leq f(\hat{z}) \leq f(\hat{y})$
  - $(z^c - \hat{z}^c)^T \nabla^c f(\hat{z}) \geq 0$ for any feasible $(z^c, \hat{z}^d)$
  - $f(\hat{x}) = f(\bar{y})$ whenever $f(\hat{x}) = f(\hat{y})$.

Numerical results: Thermal insulation system

Set of neighbors:
– change a material type
– add a shield
– remove a shield
(all are used).

Initial solution: one shield.
Local mesh optimizer at 65, 253, 692 and 2345 function evaluations with 5, 7, 9 and 9 shields respectively.
1861 more evaluations for local opt.

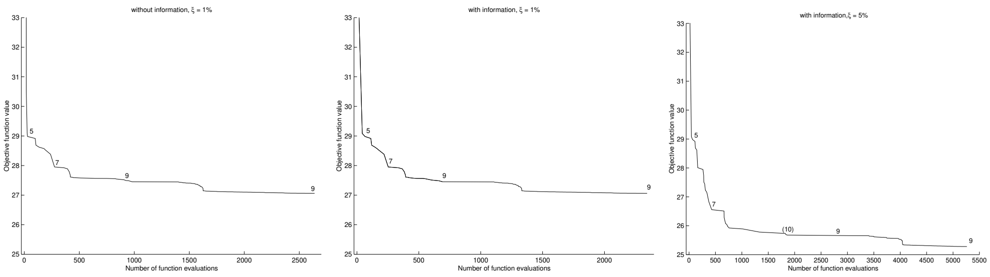




User-controlled "knobs"

- Search step
  Use heuristics, knowledge of the problem, surrogates of functions, interpolary models ...
- Set of neighbors $\mathcal{N}$
  Use information about the problem to define it.
- Extended polling trigger ξ
  Key control in finding a good local solution.

Numerical results: Thermal insulation system



Discussion

- The search step is the key to practical efficiency (it allows user's heuristics). Polling ensures robustness and provides rigor.
- Mixed variable pattern search is a descent method that guarantees first order optimality conditions with respect to the continuous variables, and is a local optimum with respect to the discrete variable.
- The user defines the set of neighbors and the constant that triggers the EXTENDED POLL step, i.e., he defines his/her notion of local optimality with respect to the discrete variables.
- We have a filter method that handles nonlinear constraints without using derivatives, Lagrange multipliers or penalty parameters. We will combine the filter method with MVP.