Pattern Search for Mixed Variable Optimization Problems

Mark A. Abramson (Air Force Institute of Technology)
Charles Audet (École Polytechnique de Montréal)
John Dennis, Jr. (Rice University)

January 7, 2003

The views expressed here are those of the authors and do not reflect the official policy or position of the USAF, Department of Defense, or US government.
Outline

• Mixed Variable Problem Motivation and Formulation
• GPS for Linearly Constrained MVP Problems
• Filter GPS for General Constrained MVP Problems
• Results for Thermal Insulation System Design
• GPS Algorithms with Derivative Information
• NOMADm Software Demo
HeatShield: Thermal Insulation System

\[
\begin{align*}
\min \ & \text{power}(n, I, x, \overline{T}) \\
\text{subject to} \ & n \in \{1, 2, \ldots, n_{\text{max}}\}, \ I \in \mathcal{I}^{n+1} \\
& \overline{T}_{i-1} \leq \overline{T}_i \leq \overline{T}_{i+1}, \ i = 1, 2, \ldots, n \\
& \sum_{i=1}^{n+1} x_i = L, \quad x_i \geq 0, \ i = 1, 2, \ldots, n + 1
\end{align*}
\]
The General MVP Problem

\[
\min_{x \in X} f(x)
\]
subject to \(C(x) \leq 0,\)

where

- \(x = (x^c, x^d) \in \mathbb{R}^{n_c} \times \mathbb{Z}^{n_d}\)

- \(X = X^c \times X^d,\) where
  \(X^c(x^d) = \{x^c \in \mathbb{R}^{n_c} : \ell(x^d) \leq A(x^d)x^c \leq u(x^d)\}\)

- \(f\) and \(C = (c_1, c_2, \ldots, c_p)\) may be discontinuous, extended valued, costly
Literature: General Methods

- **MINLP Methods**: cannot handle categorical variables

- **Search Heuristics**: very limited convergence theory
  - Simulated Annealing
  - Tabu Search
  - Evolutionary Algorithms

- **Other methods**: SQP/direct search with 1 categorical variable
Literature: Pattern Search Methods


• Positive basis theory: Lewis & Torczon (based on Davis (1954))

• Bound/linearly constrained: Lewis & Torczon

• Bound/linearly constrained non-smooth: Audet & Dennis

• General NLPs using penalty functions: Lewis & Torczon

• General non-smooth NLPs (using filters): Audet & Dennis

• Bound constrained MVPs: Audet & Dennis
Generalized Pattern Searches

INITIALIZATION

For $k = 1, 2, \ldots$

- SEARCH a finite set of mesh points
- POLL neighboring mesh points
- UPDATE parameters:
  - Success: Accept new iterate
  - Failure: Refine mesh

End
Details of POLL Step $k$

Mesh: $M_k = \{p_k + \Delta_k D z : z \in \mathbb{Z}_+^{|D|}\}$,
Poll set: $P_k = \{p_k + \Delta_k d : d \in D_k \subseteq D\}$,
where
- $p_k$ is the current poll center
- $\Delta_k > 0$ is the mesh size parameter
- $D_k, D$ are positive spanning sets

Examples: $D = [I, -I]$  $D = [I, -e]$

Unless the current point is a stationary point, at least one of these directions should be a descent direction.

Derivative information may reduce the size of the poll set
Definition of Local Optimality

A point $x = (x^c, x^d) \in X$ is said to be a local minimizer of $f$ with respect to the set of neighbors $\mathcal{N}(x) \subset X$ if there exists an $\epsilon > 0$ such that $f(x) \leq f(v)$ for all $v$ in the set

$$X \cap \bigcup_{y \in \mathcal{N}(x)} (B(y^c, \epsilon) \times y^d).$$

\[X^c \times y^d\]

\[X^c \times x^d\]
HeatShield: The Set of Discrete Neighbors

- Replace any single insulator with a different type
- Remove any intercept with its left insulator, and increase the thickness of its right insulator to absorb the deficit
- Add an intercept with insulator to its right at any position:
  - The type of insulator is the same as the one to its right
  - The cooling temperature is set to the average of the two intercepts adjacent to it (rounded to the mesh)
  - The thickness of both the new insulator and the one to its right are both divided in half (rounded to the mesh)
Construction of the Poll Set

\[ P_k = \{u, v, w\} \]
Construction of the Poll Set

\[ P_k = \{u, v, w\} \]
\[ \mathcal{N}(x_k) = \{x_k, y_1^0, y_2^0\} \]

\[ y_1^0 \in \mathcal{N}(x_k) \] satisfies
\[ f(x_k) < f(y_1^0) < f(x_k) + \xi \]
Construction of the Poll Set

\[ P_k = \{u, v, w\} \]
\[ \mathcal{N}(x_k) = \{x_k, y^0_1, y^0_2\} \]
\[ x_k = \{y^1_1\} \cup \{a, b, c\} \]

\[ y^0_1 \in \mathcal{N}(x_k) \text{ satisfies} \]
\[ f(x_k) < f(y^0_1) < f(x_k) + \xi \]

Poll Set: \( P_k \cup \mathcal{N}(x_k) \cup x_k \)
Filter for Nonlinear Constraints

Poll center = best feasible point or least infeasible point.

For each trial point $x$, $h(x)$ and $f(x)$ are plotted on the bi-loss map.

If $x$ is unfiltered, it is added to the filter; otherwise, the mesh is refined.

$h(x) = \|C(x)_+\|^2$
Construction of the Poll Set

$P_k = \{u, v, w\}$

$\mathcal{N}(p_k) = \{p_k, y_1^0, y_2^0\}$

$\mathcal{X}_k = \{y_1^1\} \cup \{a, b, c\}$

$y_1^0 \in \mathcal{N}(p_k)$ satisfies the extended poll criteria

Poll Set: $P_k \cup \mathcal{N}(p_k) \cup \mathcal{X}_k$
Local Filter for Extended Polling

Main Filter

Local Filter
Filter GPS Algorithm for MVP Problems

INITIALIZATION: Set $\Delta_0, \xi > 0$, and populate filter

For $k = 1, 2, \ldots$, do

- Update poll center $p_k$ and extended poll triggers $\xi_k^f \geq \xi, \xi_k^h \geq \xi$

- Compute incumbent values $f_k^F, f_k^I, h_k^I$

- Look for unfiltered mesh points (go to Update when found):
  - SEARCH: Any finite strategy
  - POLL: Evaluate points in $P_k \cup \mathcal{N}(p_k)$
  - EXTENDED POLL: Evaluate points in $\mathcal{X}_k(\xi_k^f, \xi_k^h)$

- Update:
  - If (found), set $\Delta_{k+1} \geq \Delta_k$ and update filter
  - If (not found), set $\Delta_{k+1} < \Delta_k$
Convergence Theory Assumptions

- All iterates lie in a compact set
- The linear constraint matrix $A$ is rational
- The mesh directions conform to the geometry of $X^c$
- All discrete neighbors lie on the current mesh
- The set-valued neighborhood function $\mathcal{N} : X \rightarrow 2^X$ satisfies a notion of continuity.
Limit Points of the Algorithm

∃ subsequence $K$ such that $\lim_{k \in K} \Delta_k = 0$, with limit points:

1. $\hat{p} = \lim_{k \in K} p_k$, where $p_k \in \{p_k^F, p_k^I\}$

2. $\hat{y} = \lim_{k \in K} y_k$, where $y_k \in \mathcal{N}(p_k)$ and $\hat{y} \in \mathcal{N}(\hat{p})$.

3. $\hat{z} = \lim_{k \in K} z_k$, where $z_k$ are extended poll endpoints.
**FMGPS Convergence Results**

Let \( D(\hat{p}) \) be the set of polling directions used infinitely often in \( K \).

- \( h \) continuous* at \( \hat{p} \) and \( \hat{y} \) \( \Rightarrow \) \( h(\hat{p}) \leq h(\hat{y}) \).
- \( f \) continuous* at \( \hat{p} \) and \( \hat{y} \) and \( p_k = p_k^F \) i. o. \( \Rightarrow \) \( f(\hat{p}) \leq f(\hat{y}) \).
- \( h \) Lipschitz* near \( \hat{p} \) \( \Rightarrow \) \( h^\circ(\hat{p}; (d, 0)) \geq 0 \) for all \( d \in D(\hat{p}) \).
- \( f \) Lipschitz* near \( \hat{p} \) and \( p_k = p_k^F \) i. o. \( \Rightarrow \) \( f^\circ(\hat{p}; (d, 0)) \geq 0 \) for all \( d \in D(\hat{p}) \).
- \( h \) strictly differentiable* at \( \hat{p} \) and \( \Rightarrow \) \( \hat{p} \) is a first-order stationary point* for \( h \).

Similar results hold for certain \( \hat{\hat{z}} \).

* with respect to the continuous variables
FMGIPS Convergence Results

$f$ strictly differentiable* at $\hat{p}$ and $p_k = p_k^F$ i. o. $\Rightarrow -\nabla^c f(\hat{p}) \in C^c_d$.

Similar results hold for certain $\hat{z}$.

* with respect to the continuous variables
HeatShield: Thermal Insulation System

\[
\begin{align*}
L & \quad \overline{T}_{n+1} = \overline{T}_H \\
& \quad \overline{T}_{i+1} \\
& \quad \overline{T}_i \\
& \quad \overline{T}_{i-1} \\
& \quad \overline{T}_0 = \overline{T}_C \\
\min & \quad power(n, I, x, \overline{T}) \\
\text{subject to} & \quad n \in \{1, 2, \ldots, n_{\max}\}, \quad I \in \mathcal{I}^{n+1} \\
& \quad \overline{T}_{i-1} \leq \overline{T}_i \leq \overline{T}_{i+1}, \ i = 1, 2, \ldots, n \\
& \quad \sum_{i=1}^{n+1} x_i = L, \quad x_i \geq 0, \ i = 1, 2, \ldots, n + 1
\end{align*}
\]
HeatShield: Previous Studies

- **Hilal & Boom**: 1-3 intercepts, single insulator type, constant cross-sectional areas

- **Hilal & Eyssa**: 1-3 intercepts, single insulator type, variable cross-sectional areas

- **Kokkolaras, Audet & Dennis**: Variable number of intercepts, multiple insulator types, constant cross-sectional areas

- **Current work**: Variable number of intercepts, multiple insulator types, variable cross-sectional areas, load-bearing nonlinear constraints
HeatShield: Objective and Nonlinear Constraints

- **Minimize power:**
  \[
  \sum_{i=1}^{n} P_i = \sum_{i=1}^{n} C_i \left( \frac{T_H}{T_i} - 1 \right) (q_i - q_{i-1})
  \]
  
  By Fourier’s law:
  \[
  q_i = \frac{A_i}{x_i} \int_{T_{i-1}}^{T_i} k(T; I_i) dT, \quad i = 1, 2, \ldots, n + 1
  \]

- **Stress:**
  \[
  \frac{F}{A_i} \leq \min \{ \sigma(T; I_i) : T_{i-1} \leq T \leq T_i \} \quad \text{(binding)}
  \]

- **Mass:**
  \[
  \sum_{i=1}^{n+1} \rho(I_i) A_i x_i \leq m_{\text{max}}
  \]

- **Contraction:**
  \[
  \sum_{i=1}^{n+1} \left( \frac{\int_{T_{i-1}}^{T_i} e(T; I_i) k(T; I_i) dT}{\int_{T_{i-1}}^{T_i} k(T; I_i) dT} \right) \left( \frac{x_i}{L} \right) \leq \frac{\delta}{100}
  \]
HeatShield: Implementation

- **Materials:**
  - Nylon 6063-T5 Aluminum
  - Teflon Fiberglass Epoxy (normal cloth)
  - 304 Stainless Steel Fiberglass Epoxy (plane cloth)
  - 1020 Low Carbon Steel

- **Material Data from Lookup Tables or Graphs:**
  - Thermal Conductivity
  - Unit Thermal Contraction
  - Maximum Allowable Stress (Tensile Yield Strength)

- **Interpolation/Integration:** Cubic splines, Simpson’s rule

- **Search/Poll:** No Search, Poll around $p_k^F$

- **Matlab Software:** NOMADm, available for download
# HeatShield: Computational Results

<table>
<thead>
<tr>
<th>Source</th>
<th>$\frac{PL}{A}$ $[\frac{W}{cm}]$</th>
<th>Insulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hilal &amp; Boom</td>
<td>68.6</td>
<td>$E_p E_p E_p$</td>
</tr>
<tr>
<td>Kokkolaras et al.</td>
<td>25.3</td>
<td>$NNNNNNNEEET$</td>
</tr>
<tr>
<td>Kokkolaras rerun</td>
<td>25.59</td>
<td>$NNNNNNTTEETT$</td>
</tr>
<tr>
<td>Hilal &amp; Eyssa</td>
<td>53.2</td>
<td>$E_p E_p E_p$</td>
</tr>
<tr>
<td>with stress constraint</td>
<td>24.55</td>
<td>$EEEEEEEEEEEE$</td>
</tr>
<tr>
<td>with all constraints</td>
<td>23.77</td>
<td>$EEEEEEEEEEEE$</td>
</tr>
</tbody>
</table>

Parameters:

- $\bar{T}_H = 300$ K, $\bar{T}_C = 4.2$ K, $L = 100$ cm, $n_{\text{max}} = 10$
- $F = 250$ kN, $m_{\text{max}} = 10$ kg, $\delta = 5\%$
- Termination: $\Delta_k \leq .15625$
HeatShield: FMGAPS Performance

![Graph showing FMGAPS Performance](image)

- **Number of function evaluations**: x 10^4
- **Power Required**: FMGAPS Performance

![Another graph showing power required vs. number of function evaluations](image)

- **Power Required**
- **Number of function evaluations**
HeatShield: Filter Performance

Filter after 150 evaluations

Filter after 200 evaluations

Filter after 500 evaluations

Zoom of filter on left

Zoom of filter on left

Zoom of filter on left
Gradient Pruning

Idea: Use the gradient to identify and prune ascent directions

Pruned Directions and Poll Set:

\[ D_k^p = \{ d \in D_k : d^T \nabla f(x_k) \leq 0 \} \]
\[ P_k^p = \{ x_k + \Delta_k d : d \in D_k^p \}, \]

Note: By convention, if \( \nabla f(x_k) \) is not used, \( D_k^p = D_k \).

Question: Is there a way to reduce the cardinality of \( P_k^p \) even further?
A Clever Choice of Spanning Set

Let $\mathbb{D} = \{-1, 0, 1\}^n$

Define $A_k = \{d \in \mathbb{D} : g^T d > 0\}$, where $g = \nabla f(x_k)$.

Let $\hat{d} \in \mathbb{D}$ satisfy the following:

- If $g_i = 0$, then $\hat{d}_i = 0$.
- If $g_i \neq 0$ then $\hat{d}_i \in \{0, \text{sign}(-g_i)\}$.
- If $|g_i| = \|g\|_{\infty}$, then $\hat{d}_i = \text{sign}(-g_i)$

Result: $\mathbb{D}_k = \{\hat{d}\} \cup A_k \subset \mathbb{D}$ positively spans $\mathbb{R}^n$, and $\mathbb{D}_k^p = \{\hat{d}\}$. 
An Illustration of Gradient Pruning

\[ \mathbb{D} = \{-1, 0, 1\}^n \]
An Illustration of Gradient Pruning

\[ \mathbb{D} = \{-1, 0, 1\}^n \]

\[ \nabla f(x_k) = (1, 4) \]

\[ A_k = \{(-1, 1), (1, 0), (1, 1), (0, 1)\} \]
An Illustration of Gradient Pruning

\[ \mathbb{D} = \{-1, 0, 1\}^n \]

\[ \nabla f(x_k) = (1, 4) \]

\[ \mathbb{A}_k = \{(-1, 1), (1, 0), (1, 1), (0, 1)\} \]

Then \( \hat{d} \in \{(0, -1), (-1, -1)\} \)
Examples of Directions that Prune to a Singleton

- $d^{(1)}$, defined by $d_i^{(1)} = \begin{cases} \text{sign}(-g_i) & \text{if } |g_i| = \|g\|_\infty, \\ 0 & \text{otherwise}, \end{cases}$

- $d^{(2)} \in \arg \max_{d \in \mathbb{D}\setminus\{0\}} \frac{-g^T d}{\|d\|_2}$

- $d^{(\infty)}$, defined by $d_i^{(\infty)} = \text{sign}(-g_i)$.

**Bottom line:** $d^{(1)}$ and $d^{(\infty)}$ are very rough approximations to the true gradient, but they still prune to a singleton and satisfy the convergence theory.
Pruning with Some Directional Derivatives

• Assume \( f'(x_k; v_i), \ i = 1, ..., \rho \leq n \) are known.

• Let \( V = \{v_1, v_2, \ldots, v_\rho\} \), and assume that \( \pm V \subseteq D \).

• Consider the ascent directions in \( D \), complete into a positive basis, and prune the ascent directions.

In general, \( |D^p_k| = n - \rho + 1 \).
Pruning with Some Derivatives: Example

Consider $f : \mathbb{R}^3 \to \mathbb{R}$, continuously differentiable near $x_k$, with $\frac{\partial f(x_k)}{\partial x_1} > 0$ and $\frac{\partial f(x_k)}{\partial x_2} < 0$. If we choose $D = \{-1, 0, 1\}^n$ and

$$D_k = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

then the pruning operation leaves only the last two columns:

$$D^p_k = \begin{bmatrix} 0 & -1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix},$$
NOMADm Demo Problem #1

minimize \( f(x, p) = (1 - p)(x_1^2 + x_2^2) + p(x_1^2 x_2 + x_1(1 - x_2)) \)
subject to
\[-2 \leq x_1 \leq 2 \]
\[-2 \leq x_2 \leq 2 \]
\( p \in \{0, 1\} \)

- Set of Neighbors: \( \mathcal{N}(x^k, p^k) = \{(x^k, 1 - p^k)\} \)
- Initial point: \( x^0 = (1, 0), p^0 = 0, f^0 = 1 \)
- Final solution: \( x^* = (-2, -2), p^* = 1, f^* = -14 \)
NOMADm Demo Problem #2

minimize $f(x) = x_1 + x_2 + x_3$
subject to $x_1^2 + x_2^2 + x_3^2 \leq 1$

• Initial point: $x^0 = (1.2, 3.8, 0.1), f^0 = 15.89$
• Final iterate: $\bar{x} = (-.8, -.4578, -.4), \bar{f} = -1.658$
• Actual solution: $x^* = (-.5774, -.5774, -.5774), f^* = -1.732$

NOMADm available at
http://www.caam.rice.edu/~abramson/NOMADm.html