Science Fiction in Nonsmooth Optimization

Robert Mifflin

http://www.math.wsu.edu/faculty/mifflin

(Work with C. Sagastizábal dedicated to the man who once said superlinear convergence in nonsmooth optimization is science fiction)

CAOA 2010, Les Houches

Supported by NSF Grant DMS 0707205 and AFOSR Grant FA9550-08-1-0370
1 Motivation with a max eigenvalue function 4

2 \( \mathcal{U} \)-theory & primal-dual tracks 5
   2.1 Primal track to \( \bar{x} \) ................................................. 6
   2.2 Dual track to \( 0 \in \partial f(\bar{x}) \) ........................................ 7

3 Approximating primal-dual tracks 8
   3.1 Bundle approximation ......................................................... 9

4 Newton-like corrector-predictor \( \mathcal{U} \) algorithm 10
   4.1 Ideal iteration and line search .............................................. 11
   4.2 Convergence properties ...................................................... 16
   4.3 Numerical results including new ones for quasi-Newton 17
1 Motivation

\[ \min_{x \in \mathbb{R}^n} f(x), \quad f \ \text{convex} \quad (\text{lower } C^2 \text{ in the future}), \]

know only one subgradient of \( f \) at each \( x \).

Fast algorithms need to identify some “curvature”;

only possible if Smooth Substructure exists.

Goal is to exploit natural structure,
including nonsmoothness,
without adding extraneous structure from barrier or smoothing functions.
2 \( \mathcal{V}\mathcal{U}\)-theory & primal-dual tracks

A pdg-structured example

\[
f(x_1, x_2, x_3) = \frac{1}{2} x_1^2 + \frac{1}{2} \sqrt{(x_1^2 - 2x_2)^2 + (x_3 - x_2)^2}
\]

minimizer \( \bar{x} = (0, 0, 0) \)

\[
\partial f((0, 0, 0))
\]

For \( \bar{g} \in \partial f(\bar{x}) \) \( \mathcal{V} = \text{lin}(\partial f(\bar{x}) - \bar{g}) \) and \( \mathcal{U} := \mathcal{V}^\perp \)
A view of $f$ on $\mathcal{V}$-space
2.1 Primal track to $\bar{x}$

$\mathcal{U}$-Lagrangian:

$$L_{\mathcal{U}}^\bar{g}(u) := \inf_{v \in \mathcal{V}} \{ f(\bar{x} + u \oplus v) - \langle \bar{g}, v \rangle \}$$

$$= f(\bar{x} + u \oplus v(u)) - \langle \bar{g}, v(u) \rangle$$

$f$ and $L_{\mathcal{U}}^0$ on the $\mathcal{U}$-space

$$\rightarrow L_{\mathcal{U}}^\bar{g}(0) = f(\bar{x}) \quad \rightarrow L_{\mathcal{U}}^\bar{g}(u) \in C^1(\mathcal{U})$$

$\rightarrow$ minimizer $v = v(u)$ generates trajectory smooth tangent to $\mathcal{U}$
if $\forall \bar{g} \in ri \partial f(\bar{x})$ the minimizer $v(u)$ is the same

$\exists$ primal track $\chi(u) := \bar{x} + u \oplus v(u)$

with $f(\chi(u)) = L^0_{\mathcal{U}}(u)$
2.2 Dual track to $0 \in \partial f(\bar{x})$

$$\gamma(u) = \text{argmin}\left\{|g|^2 : g \in \partial f(\chi(u))\right\}$$

$$\chi(u) = \bar{x} + u \oplus v(u)$$

$$\left(\chi(u), \gamma(u)\right) \to (\bar{x}, 0) \text{ as } u \to 0$$

**Good primal-dual track $\leftrightarrow \begin{array}{c} L_U^0 \in C^2 \quad + \quad 0 \in ri \partial f(\bar{x}) \end{array}$**

allows for Newton-like method to minimize $f$:

- **corrector** – predictor method
- **proximal point** – Newton method
- $\mathcal{V}$ – $\mathcal{U}$
3 Approximating primal-dual tracks

Fundamental theoretical result:
Proximal Points are on the primal track

If $\bar{g} = 0 \in \mathrm{ri}\partial f(\bar{x})$, then $\exists u(x) :

\[ p(x) := \arg\min \left\{ f(y) + \frac{1}{2}\mu|y - x|^2 \right\} = \chi(u(x)) \]

for all $x \approx \bar{x}$ with $\mu = \mu(x) : \mu(x)|x - \bar{x}| \to 0$ as $x \to \bar{x}$

$\Rightarrow$ use a bundle subroutine
to approximate the prox
and estimate the pair
$(\chi(u), \gamma(u))$
3.1 Bundle approximation

With bundle \((y_i, f_i, g_i)\)
build \(\tilde{f}\), a \(\nu\)-model for \(f\) near \(x\), and find

\[ p := \text{argmin} \left\{ \tilde{f}(y) + \frac{1}{2} \mu |y - x|^2 \right\} \approx \chi(u(x)) \]

\[ s := \text{argmin} \left\{ |g|^2 : g \in \partial \tilde{f}(p) \right\} \approx \gamma(u(x)) \]

UNTIL “good enough”:
\[ \varepsilon \leq (\sigma/\mu)|s|^2 \]

By-product: local \(\nu U\)-decomposition, \(\Psi\)
4 Newton-like corrector-predictor $\mathcal{VU}$ algorithm

Given $x$ and a bundle:

- **Corrector step:** Solve ($\chi$ — and $\gamma$-qp)'s
  
  $\Rightarrow$ New $p, s, \mathcal{VU}$, and determine $H$ ($U$-Hessian $\approx \nabla^2 L_u$)

- **Predictor step:** Solve $H \Delta u = -u^\top s \Rightarrow x^+ = p + u \Delta u$
4.1 Ideal iteration

\[ \chi(u) \]

\[ x \]

\[ p^+ \]

\[ p \]

\[ U-\text{Newton step} \]

\[ \text{bundle steps} \]
4.1 Candidates for $x^+, p^+$

$\chi(u)$

$x$

$\mathbf{LS \ if \ } f(p_{cand}^+) \leq f(p)$
4.1 Line search if candidates fail descent

\[ f(x^+) < f(p) \]  

to find

\[ x_c(u) \]
4.1 Line search to good $x^+$

and repeat the process at most once
4.1 Second bundle run to good $p^+$

$\text{convergent}$ even if no $\chi(u)$
4.2 Convergence properties

1. If infinite number of inner bundle steps, this sequence converges to a minimizer of $f$

2. If the decreasing sequence $\{f(p)\}$ is infinite, then
   - either $f$ unbounded below,
   - or $s \to 0$ and any $\text{acc}(\{p\})$ minimizes $f$

3. If there is a good primal-dual track to $(\bar{x}, 0)$, and
   - $\frac{\sigma}{\mu^2} = O(|s^-|^2)$,
   - bounded $\{H^{-1}\}$,
   - $U \to U$,
   - Dennis-Moré-like condition for $\{H\}$,
   - $s$ approximates $\gamma$ superlinearly,

then $\{p\}$ converges superlinearly to $\bar{x}$
4.3 Preliminary Numerical results

Newton variant

for finite max-functions \( f(x) = \max_\ell f_\ell(x), f_\ell \in C^2 \)

\[
H = U^\top \sum_i \bar{\alpha}_i \nabla^2 f_\ell(y_i) U
\]

where \( \bar{\alpha} \) solves \( \gamma\)-qp: \[
\min_{\alpha \in \Delta^{||\cdot||^2}} \frac{1}{2} \left| \sum_i \alpha_i g_i \right|^2
\]

quasi-Newton variant

\[
H = U^\top H_{qN} U
\]

where \( H_{qN} = BFGS(p - p^-, s - s^-) \) is an \( n \times n \) matrix, with updating started at or after iteration 3 when a sufficiently large curvature is found for initial scaling of the identity.
## Summary of results

<table>
<thead>
<tr>
<th></th>
<th>2d-U1</th>
<th>3d-EX</th>
<th>3d-U2</th>
<th>3d-U1</th>
<th>3d-U0</th>
<th>MAXQUAD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f/g</td>
<td>Ac</td>
<td>f/g</td>
<td>Ac</td>
<td>f/g</td>
<td>Ac</td>
</tr>
<tr>
<td>N1CV2</td>
<td>38</td>
<td>7</td>
<td>103</td>
<td>7</td>
<td>55</td>
<td>7</td>
</tr>
<tr>
<td>N.VU</td>
<td>12</td>
<td>16</td>
<td>21</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>qN.VU</td>
<td>20</td>
<td>14</td>
<td>35</td>
<td>14</td>
<td>34</td>
<td>12</td>
</tr>
</tbody>
</table>


**N.VU**: good LS, special $\mu$-update, $\sigma = 0.5$, computed $\mu_1$

**qN.VU**: quasi-Newton variant of N.VU, with different $\mu$-update rules when $\mu$ too large, last $p$ replaced by best $p$ and a possible line search along $\mathcal{U}$-step direction

Current work: quasi-Newton, $\mu$-adjustment

Future work: lower $C^2$ functions
Nonsmooth Science Fiction
Smooth-BFGS 249  VU-BFGS 87

Lewis&Overton example, 2008
$sqrt(x'Ax)+x'Bx; A=\text{diag}(1,0,1,0,...), B=\text{diag}(1,...,1/n^2)$, $\text{dim} V=\text{dim} U=4$