Globally Convergent
Limited Memory Bundle Method
for Bound Constrained Large-Scale
Nonsmooth Optimization

Napsu Karmitsa
Department of Mathematics,
University of Turku, Finland,
email: napsu@karmitsa.fi
**Nonsmooth Optimization**

We consider the nonsmooth (nondifferentiable) bound constrained optimization problem of the form

\[
\begin{aligned}
& \text{minimize} & & f(x) \\
& \text{subject to} & & x^l \leq x \leq x^u,
\end{aligned}
\]

where the objective function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is supposed to be locally Lipschitz continuous and the number of variables \( n \) is supposed to be large.
NonSmooth Optimization (Cont’d)

Definition. Let a function $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz continuous at $x \in \mathbb{R}^n$. Then the subdifferential of $f$ at $x$ is the set $\partial f(x)$ of vectors $\xi \in \mathbb{R}^n$ such that

$$\partial f(x) = \text{conv}\{ \xi \in \mathbb{R}^n | \nabla f(x_i) \to \xi, \ x_i \to x \}$$

and $f$ is differentiable at $x_i$.

Each vector $\xi \in \partial f(x)$ is called a subgradient of $f$ at $x$.

Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$. If $f$ attains its local minimum at $x$, then

$$0 \in \partial f(x).$$

A point $x \in \mathbb{R}^n$ satisfying $0 \in \partial f(x)$ is called a critical or a stationary point for $f$. 

Napsu Karmitsa

EngOpt2008, Rio de Janeiro
Nonsmooth Optimization (Cont’d)

Theorem. (KKT-type optimality condition for bound constrained problems.) Let \( f : \mathbb{R}^n \to \mathbb{R} \) be locally Lipschitz continuous function at \( x \in \mathbb{R}^n \) and let us denote by \( g^l(x) = x^l - x \) and \( g^u(x) = x - x^u \). If \( x \) is a local minimum of problem (\ref{eq:bound_constrained_problem}), then there exist Lagrange multipliers \( \mu^l_i, \mu^u_i \geq 0 \) such that \( \mu^l_i g^l_i(x) = 0 \) and \( \mu^u_i g^u_i(x) = 0 \) for all \( i \in \{1, \ldots, n\} \) and

\[
0 \in \partial f(x) + \sum_{i=1}^{n} \mu^l_i \partial g^l_i(x) + \sum_{i=1}^{n} \mu^u_i \partial g^u_i(x).
\]

A feasible point \( x \) satisfying the KKT optimality condition above is said to be a KKT point associated to problem (\ref{eq:bound_constrained_problem}).

Note that in the previous theorem we have \( \partial g^l_i(x) = \{\nabla g^l_i(x)\} \) and \( \partial g^u_i(x) = \{\nabla g^u_i(x)\} \) due to differentiability of constraints. However, we usually do not know the whole subdifferential \( \partial f(x) \).
Limited Memory Bundle Method (LMBM)

- hybrid of the variable metric bundle method [Vlček and Lukšan 2001] and the limited memory variable metric method [Nocedal 1980, Byrd et.al. 1994],
- does not have to solve any quadratic direction finding problems appearing in the standard bundle methods,
- the size of bundle does not increase with the dimension of the problem,
- only a few vectors are kept to represent the variable metric approximation of the Hessian matrix,
- does not require knowledge of the structure of the problem or its Hessian,
- globally convergent method for locally Lipschitz continuous objective functions [Haarala et.al. 2007],
- efficient solver for both convex and nonconvex large-scale nonsmooth unconstrained optimization problems.
LMBM: Search Direction

The search direction $d_k$ is calculated by the formula

$$d_k = -D_k \xi_k,$$

where $D_k$ is a limited memory variable metric update that, in smooth case, represents the approximation of the inverse of the Hessian matrix, and $\xi_k \in \partial f(x_k)$, if the previous step was a serious step, and $\tilde{\xi}_k$ is an aggregate subgradient, otherwise.

Note that due to limited memory approach this search direction can be calculated using only $O(n)$ operations.
LMBM: Aggregation Procedure

Determine multipliers $\lambda_i^k$ satisfying $\lambda_i^k \geq 0$ for all $i \in \{1, 2, 3\}$, and $\sum_{i=1}^3 \lambda_i^k = 1$ that minimize the function

$$\varphi(\lambda_1, \lambda_2, \lambda_3) = (\lambda_1 \xi_m + \lambda_2 \xi_{k+1} + \lambda_3 \tilde{\xi}_k)^T D_k (\lambda_1 \xi_m + \lambda_2 \xi_{k+1} + \lambda_3 \tilde{\xi}_k) + 2(\lambda_2 \beta_{k+1} + \lambda_3 \tilde{\beta}_k),$$

where $\xi_m \in \partial f(x_k)$ is the current subgradient, $\xi_{k+1} \in \partial f(y_{k+1})$ is the auxiliary subgradient, and $\tilde{\xi}_k$ is a current aggregate subgradient from the previous iteration ($\tilde{\xi}_1 = \xi_1$). In addition, $\beta_{k+1}$ and $\tilde{\beta}_k$ are the subgradient locality measure and the aggregate subgradient locality measure ($\tilde{\beta}_1 = 0$), respectively.

The next aggregate values are given by

$$\tilde{\xi}_{k+1} = \lambda_1^k \xi_m + \lambda_2^k \xi_{k+1} + \lambda_3^k \tilde{\xi}_k$$

and

$$\tilde{\beta}_{k+1} = \lambda_2^k \beta_{k+1} + \lambda_3^k \tilde{\beta}_k.$$
Limited Memory Bundle Method for Bound Constrained Optimization (LMBM-B)

- constraint handling is based on the projected (sub)gradient method and dual subspace minimization [Byrd et. al. 1995],

- main differences from basic LMBM:
  - calculation of the search direction,
  - aggregation procedure,
  - line search procedure,

- main differences from basic limited memory BFGS method for bound constrained optimization:
  - capable of handling nonsmooth objectives,

- globally convergent method for locally Lipschitz continuous objective functions [Karmitsa, submitted],

- only generates feasible points.
Initialization.

Serious step initialization.

Desired accuracy? Yes STOP. No

Calculation of the generalized Cauchy point and determination of an active set. Direction finding using the limited memory BFGS update. Variables in the active set remains fixed.

Serious step

Line search and solution updating.

Calculation of the generalized Cauchy point and determination of an active set. Direction finding using the limited memory SR1 update. Variables in the active set remains fixed.

Null step

Projection and Aggregation.

Desired accuracy? Yes STOP. No
**LMBM-B: Generalized Cauchy Point**

We define the quadratic model function $q_k$ that approximates the objective function at the iteration point $x_k$ by

$$q_k(x) = f(x_k) + \tilde{\xi}_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k(x - x_k),$$

where $\tilde{\xi}_k$ is the aggregate subgradient of the objective function and $B_k = D_k^{-1}$.

In addition, we define the projection operator $\mathcal{P}_c[.]$ (component-wise) by

$$\mathcal{P}_c[x, x^l, x^u]_i = \begin{cases} x^l_i, & \text{if } x_i < x^l_i \\ x_i, & \text{if } x_i \in [x^l_i, x^u_i] \\ x^u_i, & \text{if } x_i > x^u_i. \end{cases}$$

LMBM-B: **Generalized Cauchy Point (Cont’d)**

The generalized Cauchy point at iteration $k$ is defined as the first local minimizer of the univariate piecewise quadratic function

$$
\hat{q}_k(t) = q_k(\mathcal{P}_c[x_k - t\tilde{\xi}_k, x^l, x^u]),
$$

along the projected gradient direction $\mathcal{P}_c[x_k - t\tilde{\xi}_k, x^l, x^u] - x_k$.

That is, if we denote by $t^{cp}_k$ the value of $t$ corresponding to the first local minimum of $\hat{q}_k(t)$, the generalized Cauchy point is given by

$$
x^{cp}_k = \mathcal{P}_c[x_k - t^{cp}_k\tilde{\xi}_k, x^l, x^u].
$$

The variables whose values at $x^{cp}_k$ are at lower or upper bound, comprise the active set

$$
\mathcal{I}_A = \{ i \mid x^{cp}_i = x^l_i \text{ or } x^{cp}_i = x^u_i \}.
$$
LMBM-B: Direction Finding

We solve \( \mathbf{d} \) from smooth quadratic problem

\[
\begin{align*}
\text{minimize} & \quad \tilde{\xi}_k^T \mathbf{d} + \frac{1}{2} \mathbf{d}^T \mathbf{B}_k \mathbf{d} \\
\text{such that} & \quad A_k^T \mathbf{d} = \mathbf{b}_k \quad \text{and} \\
& \quad \mathbf{x}^l \leq \mathbf{x}_k + \mathbf{d} \leq \mathbf{x}^u,
\end{align*}
\]

(DFP)

where \( A_k \) is the matrix of active constraints gradients at \( \mathbf{x}_k^{cp} \) and \( \mathbf{b}_k = A_k^T (\mathbf{x}_k^{cp} - \mathbf{x}_k) \).

The first order optimality conditions for (DFP) without bounds are

\[
\begin{align*}
\tilde{\xi}_k + \mathbf{B}_k \mathbf{d}^* + A_k \lambda^* &= 0 \\
A_k^T \mathbf{d}^* &= \mathbf{b}_k.
\end{align*}
\]
LMBM-B: Direction Finding (Cont’d)

From these, we can determine Lagrange multipliers $\mu^*$ by

\[(A_k^T D_k A_k) \mu^* = -A_k^T D_k \tilde{\xi}_k - b_k,\]

and the search direction $d^*$ can be calculated by

\[B_k d^* = -A_k \mu^* - \tilde{\xi}_k.\]

Remark:

- If there are no active bounds present, we obtain $d^* = -D_k \tilde{\xi}_k$, which is the formula used also in the original unconstrained version of LMBM.

- If the vector $x_k + d^*$ violates the bounds in (??), we backtrack along the line joining the infeasible point $x_k + d^*$ and the generalized Cauchy point $x_k^{cp}$ to regain the feasible region.
LMBM-B: Aggregation Procedure

Let us define a very simple projection operator $\mathcal{P}_x[.]$ at point $x$ (component-wise) by

$$
\mathcal{P}_x[\xi]_i = \begin{cases} 
0, & \text{if } x_i^l - x_i \geq 0 \\
\xi_i, & \text{if } x_i \in (x_i^l, x_i^u) \\
0, & \text{if } x_i^u - x_i \leq 0.
\end{cases}
$$

The aggregation procedure is made by determining multipliers $\lambda_i^k$ satisfying $\lambda_i^k \geq 0$ for all $i \in \{1, 2, 3\}$, and $\sum_{i=1}^3 \lambda_i^k = 1$ that minimize the function

$$
\varphi(\lambda_1, \lambda_2, \lambda_3) = \mathcal{P}_{x_k}[\lambda_1 \xi_m + \lambda_2 \xi_{k+1} + \lambda_3 \tilde{\xi}_k]^T D_k \mathcal{P}_{x_k}[\lambda_1 \xi_m + \lambda_2 \xi_{k+1} + \lambda_3 \tilde{\xi}_k] \\
+ 2(\lambda_2 \tilde{\beta}_{k+1} + \lambda_3 \tilde{\beta}_k).
$$

The next aggregate values are given by

$$
\tilde{\xi}_{k+1} = \lambda_1^k \xi_m + \lambda_2^k \xi_{k+1} + \lambda_3^k \tilde{\xi}_k \\
\text{and} \quad \tilde{\beta}_{k+1} = \lambda_2^k \beta_{k+1} + \lambda_3^k \tilde{\beta}_k.
$$
LMBM-B: Convergence Analysis

Under some mild assumptions . . .

• If LMBM-B algorithm stops at iteration \( k \), then the point \( \mathbf{x}_k \) is a KKT-point for problem (??).

• LMBM-B algorithm does not stop at a non-KKT point.

• Every accumulation point \( \mathbf{\bar{x}} \) generated by LMBM-B algorithm is a KKT-point for problem (??).

• If the objective function is convex, the KKT-point is also a global minimum for problem (??).
Numerical Experiments

Table 1: Average results for 10 problems.

<table>
<thead>
<tr>
<th>Solver/n</th>
<th>1000</th>
<th></th>
<th></th>
<th>2000</th>
<th></th>
<th></th>
<th>4000</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ni/Nf</td>
<td>fail</td>
<td>time</td>
<td>Ni/Nf</td>
<td>fail</td>
<td>time</td>
<td>Ni/Nf</td>
<td>fail</td>
<td>time</td>
</tr>
<tr>
<td>LMBM-B</td>
<td>188/1167</td>
<td>20%</td>
<td>1.07</td>
<td>164/595</td>
<td>20%</td>
<td>1.11</td>
<td>176/1046</td>
<td>40%</td>
<td>5.57</td>
</tr>
<tr>
<td>LMBM-B-OLD</td>
<td>130/377</td>
<td>10%</td>
<td>0.21</td>
<td>68/147</td>
<td>30%</td>
<td>0.68</td>
<td>87/173</td>
<td>40%</td>
<td>1.90</td>
</tr>
<tr>
<td>PBNCGC</td>
<td>55/73</td>
<td>10%</td>
<td>67.48</td>
<td>33/39</td>
<td>20%</td>
<td>540.14</td>
<td>16/17</td>
<td>50%</td>
<td>1562.28</td>
</tr>
</tbody>
</table>

- Different variants of LMBM-B were superior when comparing the computational times: the computational times elapsed with LMBM-B and LMBM-B-OLD were usually hundreds of times shorter than those of PBNCGC.
- LMBM-B usually needed more function evaluations (and thus, also more computational time) than the older version LMBM-B-OLD.
- PBNCGC always used less function evaluations than the different variants of LMBM-B.
Final Remarks

- LMBM-B is an efficient method for solving both convex and non-convex large-scale nonsmooth bound constrained optimization problems.
- A drawback of the new variant is the increased number of function evaluations needed. ⇒ Different projection possibilities need to be studied.
- An advantage of LMBM-B is that it only generates feasible points.
- The basic LMBM (unconstrained version) is available online at

http://napsu.karmitsa.fi/lmbm/

Acknowledgments

- Prof. Marko M. Mäkelä (University of Turku, Finland).
- The work was financially supported by Magnus Ehrnrooth Foundation and the University of Turku.