RWTH Aachen

Department of Computer Science
Technical Report

Adjoint Subgradient Calculation for McCormick Relaxations

Markus Beckers, Viktor Mosenkis, Michael Maier and Uwe Naumann

ISSN 0935–3232 · Aachener Informatik-Berichte · AIB-2011-10

RWTH Aachen · Department of Computer Science · October 2011

Adjoint Subgradient Calculation for McCormick Relaxations

Markus Beckers and Viktor Mosenkis and Michael Maier and Uwe Naumann

Lehr- und Forschungsgebiet Informatik 12
Software and Tools for Computational Engineering
RWTH Aachen, Germany
Email: beckers@stce.rwth-aachen.de

Abstract. In [CMN$^+$11] the library modMC was presented which allows the propagation of McCormick relaxations and their corresponding subgradients based on the forward mode of Algorithmic Differentiation (AD). Subgradients are natural extensions of usual derivatives which allow the application of derivative based methods on possibly nondifferentiable convex and concave functions. These subgradients can be computed by AD, a method which allows the computation of derivatives with machine accuracy even for highly complex functions implemented by a computer program. In this article we present the advancement of modMC by reverse mode AD. Reverse mode AD is an adjoint method for the propagation of derivatives which is preferable when scalar functions are considered. We describe the theory behind the application of reverse mode in subgradient propagation as well as the improved library amodMC in detail. The calculated subgradients are used in an deterministic global optimization algorithm which is based on a branch-and-bound method. The improvements gained using Reverse instead of Forward mode AD are illustrated by examples.

1 Motivation & Context

Optimization problems in engineering often have nonconvex objective and constraints and require global optimization algorithms. Deterministic global optimization algorithms based on the branch-and-bound methods solve relaxations of the original program. These are constructed by convex/concave under-/overestimators of the functions involved. One of the alternative methods to construct these estimating functions are McCormick relaxations (see [McC76]). Without auxiliary variables this technique results in nonsmooth estimators, and thus to obtain derivative information, subgradients are needed. These can be calculated using techniques from AD [MCB09],[CMN$^+$11]. This is especially very useful if the functions are given by a long and complex computer program. AD allows the calculation of derivatives, and here additionally the relaxations, with machine accuracy. Hence numerical error based on finite difference approximations are avoided. Two important methods of AD are the forward (or tangent-linear) and the reverse (or adjoint) mode. The choice of the method depends on the dimensionality of the function to be relaxed/differentiated.

The combination of the afore mentioned methods into global optimization was first discussed in [MCB09] using forward mode AD. Enhancements of the implementation were presented in [CMN$^+$11] wherein runtimes were improved by using Fortran specifics and compiler support enabled AD by source transformation. However, [CMN$^+$11] was still lacking a reverse mode implementation, which is a disadvantage since sometimes the number of inputs (optimization variables) is much greater than the number of outputs (objective, constraints). Such an enhancement is given in this paper.

2 Theoretical Development

Let $F : Z \to I\!R$ be a function given on a convex set $Z \subseteq I\!R^n$. Then, a convex (concave) function F^{cv} (F^{cc}) for which $F^{cv}(\mathbf{z}) \leq F(\mathbf{z})$ ($F^{cc}(\mathbf{z}) \geq F(\mathbf{z})$) holds for all $\mathbf{z} \in Z$ is called a *convex (concave) relaxation* of F. As a special case we now observe McCormick relaxations of factorizable functions. In [MCB09] a procedure for propagating *subgradients* of relaxations is presented, which is based on the forward mode of *Algorithmic Differentiation* (AD). Our goal is to extend this procedure to a reverse mode.

Examine first the structure of propagating subgradients with AD methods. The propagation of the convex and concave relaxation of the function

$$F : Z \subseteq I\!R^n \to I\!R, \ \ (z_1, \ldots, z_n) \mapsto y$$

can be considered as the *composition* $g \circ f = \begin{pmatrix} F^{cv} \\ F^{cc} \end{pmatrix}$ of the two functions

$$f : Z \subseteq I\!R^n \to I\!R^{2n}, (z_1, \ldots, z_n) \mapsto (z_1^{\mathrm{cv}}, z_1^{\mathrm{cc}}, \ldots, z_n^{\mathrm{cv}}, z_n^{\mathrm{cc}}) = (z_1, z_1, \ldots, z_n, z_n)$$

and

$$g = (g^{cv}, g^{cc}) : Z^+ \subseteq I\!R^{2n} \to I\!R^2, \ \ (z_1^{\mathrm{cv}}, z_1^{\mathrm{cc}}, \ldots, z_n^{\mathrm{cv}}, z_n^{\mathrm{cc}}) \mapsto (y^{\mathrm{cv}}, y^{\mathrm{cc}}) \quad ,$$

where Z^+ denotes the set $\left\{(z_1, z_1, \ldots, z_n, z_n) \in I\!R^{2n} \mid \mathbf{z} = (z_1, \ldots, z_n) \in Z\right\}$.

This means $y^{cv} = F^{cv}(\mathbf{z}) = g^{cv}(f(\mathbf{z})) = g^{cv}(\mathbf{z}^+)$ and $y^{cc} = F^{cc}(\mathbf{z}) = g^{cc}(f(\mathbf{z})) = g^{cc}(\mathbf{z}^+)$. Here g really represents the simultaneous propagation process of the convex and concave relaxation, for which the duplication f of the variables is needed. (Note that both the convex and concave relaxation of the identity $f(\mathbf{z}) = \mathbf{z}$ are equal to f.)
The following Theorem explains our further proceedings. For a vector $\mathbf{z} = (z_1, \ldots, z_n) \in Z$, $\mathbf{z}^+$ denotes the corresponding vector $\mathbf{z}^+ = (z_1, z_1, \ldots, z_n, z_n) \in Z^+$.

Theorem 1. *Let $g \circ f$ be defined as the above composition, $z \in Z$ and let*

$$s_{g^{cv}}\left(\mathbf{z}^+\right) = \left(\frac{\partial y^{cv}}{\partial z_1^{cv}}, \frac{\partial y^{cv}}{\partial z_1^{cc}}, \ldots, \frac{\partial y^{cv}}{\partial z_n^{cv}}, \frac{\partial y^{cv}}{\partial z_n^{cc}}\right)$$

denote a subgradient of the convex relaxation g^{cv} at z^+ and

$$s_{g^{cc}}\left(\mathbf{z}^+\right) = \left(\frac{\partial y^{cc}}{\partial z_1^{cv}}, \frac{\partial y^{cc}}{\partial z_1^{cc}}, \ldots, \frac{\partial y^{cc}}{\partial z_n^{cv}}, \frac{\partial y^{cc}}{\partial z_n^{cc}}\right)$$

denote a subgradient of the concave relaxation g^{cc} at z^+.
Then a subgradient of the convex relaxation F^{cv} of F at z is given by

$$s_{F^{cv}}(\mathbf{z}) := \left(\frac{\partial y^{cv}}{\partial z_1^{cv}} + \frac{\partial y^{cv}}{\partial z_1^{cc}}, \ldots, \frac{\partial y^{cv}}{\partial z_n^{cv}} + \frac{\partial y^{cv}}{\partial z_n^{cc}}\right) \quad . \tag{1}$$

Similarly the subgradient of the concave relaxation F^{cc} of F at z is given by

$$s_{F^{cc}}(\mathbf{z}) := \left(\frac{\partial y^{cc}}{\partial z_1^{cv}} + \frac{\partial y^{cc}}{\partial z_1^{cc}}, \ldots, \frac{\partial y^{cc}}{\partial z_n^{cv}} + \frac{\partial y^{cc}}{\partial z_n^{cc}}\right) \quad . \tag{2}$$

Proof. It is easy to see that Z^+ is a convex set and g^{cv} is a convex function for all convex sets $Z \subseteq I\!R^n$. This implies the existence of the above subgradients. According to Definition 1.1.4 on page 165 in [HUL93] we need to show that

$$\langle s_{F^{cv}}(z), d \rangle \leq (F^{cv})'(\mathbf{z}, \mathbf{d}) \quad \forall \mathbf{d} \in I\!R^n \quad , \tag{3}$$

where

$$(F^{cv})'(\mathbf{z}, \mathbf{d}) := \lim_{t \to +0} \frac{F^{cv}(\mathbf{z} + t \cdot \mathbf{d}) - F^{cv}(\mathbf{z})}{t} \quad .$$

We observe for arbitrary $\mathbf{d} \in I\!R^n$:

$$\begin{aligned}
\langle s_{F^{cv}}(\mathbf{z}), \mathbf{d} \rangle &= \left(\frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cv}}} + \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cc}}}, \ldots, \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cv}}} + \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cc}}} \right) \cdot \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} \\
&= d_1 \cdot \left(\frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cv}}} + \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cc}}} \right) + \cdots + d_n \cdot \left(\frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cv}}} + \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cc}}} \right) \\
&= d_1 \cdot \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cv}}} + d_1 \cdot \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cc}}} + \cdots + d_n \cdot \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cv}}} + d_n \cdot \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cc}}} \\
&= \left(\frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cv}}}, \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cc}}}, \ldots, \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cv}}}, \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cc}}} \right) \cdot \begin{pmatrix} d_1 \\ d_1 \\ \vdots \\ d_n \\ d_n \end{pmatrix} \\
&= \langle s_{g^{cv}}(\mathbf{z}^+), \mathbf{d}^+ \rangle \\
&\underset{\text{Def. of subgradient}}{\leq} \lim_{t \to +0} \frac{g^{cv}(\mathbf{z}^+ + t \cdot d^+) - g^{cv}(\mathbf{z}^+)}{t}) \\
&= \lim_{t \to +0} \frac{g^{cv}(f(\mathbf{z} + t \cdot \mathbf{d})) - g^{cv}(f(\mathbf{z}))}{t} \\
&= (F^{cv})'(\mathbf{z}, \mathbf{d}) \quad .
\end{aligned}$$

This shows (3) and completes the proof for the convex relaxation. The proof for the concave relaxation is analogue by considering the convex function $-F^{cc}$.

Remark 1. Theorem 1 yields the following view on the propagation of subgradients of convex and concave McCormick relaxations:
If we interprete the matrices

$$Dg := \begin{pmatrix} \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cv}}} & \frac{\partial y^{\mathrm{cv}}}{\partial z_1^{\mathrm{cc}}} & \cdots & \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cv}}} & \frac{\partial y^{\mathrm{cv}}}{\partial z_n^{\mathrm{cc}}} \\ \\ \frac{\partial y^{\mathrm{cc}}}{\partial z_1^{\mathrm{cv}}} & \frac{\partial y^{\mathrm{cc}}}{\partial z_1^{\mathrm{cc}}} & \cdots & \frac{\partial y^{\mathrm{cc}}}{\partial z_n^{\mathrm{cv}}} & \frac{\partial y^{\mathrm{cc}}}{\partial z_n^{\mathrm{cc}}} \end{pmatrix} \in I\!R^{2 \times 2n} \quad . \tag{4}$$