Hybrid Derivative-Free Optimization for Mixed-Integer Functions

Juan J Torres[1], Emiliano Traversi[1], Roberto W Calvo[1] , Giacomo Nannicini[2]

[1] Université Sorbonne Paris Nord, Sorbonne Paris Cité, LIPN, CNRS, (UMR 7030), France

{torresfigueroa,traversi,wolfer}@lipn.univ-paris13.fr

[2] IBM T.J. Watson Research Center, NY, U.S.A.

nannicini@us.ibm.com

Mots-clés : *Derivative-free optimization, Mixed-integer programming, Discrete convex functions, Surrogate-based optimization, Difference of convex algorithmn*

1 Introduction

In this study we propose an algorithm for the solution of the following mixed-integer optimization problem :

$$\min_{x,y} f(x,y) \tag{1a}$$

$$x \in \Omega_c, \;\; y \in \Omega_z \tag{1b}$$

where $f : [\mathbb{R}^{n_1} \times \mathbb{Z}^{n_2}] \to \mathbb{R}$, $\Omega_c = \{x \in \mathbb{R}^{n_1} \mid x_{lb} \leq x \leq x_{ub}\}$ and $\Omega_z = \{y \in \mathbb{Z}^{n_2} \mid y_{lb} \leq y \leq y_{ub}\}$. We denote the mixed-integer box constraints as $\Omega_m = \{(x,y) \mid x \in \Omega_c, \;\; y \in \Omega_z\}$. In addition $f(x,y)$ is a mixed-integer black-box function which exhibits combinatorial properties at fixed values of x. Black-box are often expensive-to-evaluate and do not present analytical form, which means that no gradient nor second-order information can be used to optimize them. Black-box functions arise in several settings, such medical imaging, operations research, and, specially in computer simulation programs. Several methodologies have been developed to solve black-box instances including heuristics (i.e evolutionary algorithms, tabu search) and derivative-free optimization (DFO) methods. Within DFOs, surrogate-based approaches appear to be successful in the computation of local and global solution of black-box optimization programs.

Surrogate-based methods consist in the computation of a (surrogate) model which approximates the black-box function, via regression or interpolation. Different types of models can be used to approximate a black-box function, including low-order polynomials, radial basis functions (RBF) and kriging. To the extent of our knowledge, none of these approximations have been studied for mixed-integer functions with special combinatorial properties on their integer elements.

One example of such problems is the mixed-integer generalization of $M^\natural$ discrete functions [1]. $M^\natural$ are integrally convex functions that display interesting properties such as supermodularity, descent directions and minimizers. We aim to develop a methodology that takes advantage of such features and the existence of convergent zero-order algorithms for the optimization of given functions when only integer variables are considered.

2 Algorithm Overview

Our proposed algorithm is designed to solve problem (1) via hybrid surrogate approximation and the use of the difference of convex algorithm (DCA) [2]. The principle behind this dual methodology consists in the following reformulation :

$$\min_{x \in \Omega_c, y \in \Omega_z} f(x,y) = \min_{x \in \Omega_c} \psi(x)$$

Let $\psi : \mathbb{R}^{n_1} \to \mathbb{R}$ be defined as $\inf_{y \in \Omega_z} f(x, y)$. We emphasise that function $\psi(x)$ can be represented as the difference of two functions

$$\psi(x) = \frac{\lambda_k}{2}\|x\|^2 - \left[\frac{\lambda_k}{2}\|x\|^2 - \psi(x)\right]$$

where $\lambda_k > 0$. Let $\phi(x)$ be defined as $\phi(x) = \left[\frac{\lambda_k}{2}\|x\|^2 - \psi(x)\right]$. If λ_k is sufficiently large, $\phi(x)$ becomes convex, which allow us to use the standard DCA method for the optimization of $\psi(x)$. The DCA consists in the following repetition of operations :

1. Computation of a subgradient w of $\phi(x)$ at point x_k.
2. Computation of a new candidate solution $x_{k+1} = \text{argmin}_{x \in \Omega_c}\{\frac{\lambda_k \|x\|}{2} - \langle w, x - x_k \rangle\}$

These procedures are repeated until $\|x_{k+1} - x_k\| < \epsilon_{tol}$. The main difficulty for using DCA for the solution of problem (1) is the lack of a deterministic formula for the computation of the subgradients of $\phi(x)$. Nonetheless this problem can be overcome by considering accurate approximations of these subgradients (often called ϵ-subgradients) which can be estimated using fully-linear or fully-quadratic surrogates of $\psi(x)$. The DCA has been proved to be also convergent in the case where ϵ-subgradients are used [3, 4].
We modified the DCA and combined it with a first order model-based trust-region method, which consists in the repetition of the following list of operations :

1. Criticality test : Evaluation of convergence into stationary points.
2. Candidate computation : A new candidate is generated by solving a surrogate that considers the ϵ-subgradient information.
3. Candidate acceptance : If the new candidate yields a significant improvement in the objective it is accepted and the optimization domain (trust-region) is extended. Otherwise it is rejected.
4. Model maintenance : In case of a successful candidate we compute a new surrogate approximation at the point x_{k+1}. If the iteration was unsuccessful, we reduce the domain in which the optimization is performed aiming to reduce the error in the surrogate approximation.

We highlight that this algorithm is proved to be convergent to a type of mixed-integer stationary point as it preserves the basic properties of first-order surrogate approximation [5].

Références

[1] K. Murota, *Discrete Convex Analysis.* SIAM, 2003.

[2] P. D. Tao *et al.*, "The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems.," *Annals of Operations Research*, vol. 133, no. 1-4, pp. 23–46, 2005.

[3] X. T. Vo, *Learning with sparsity and uncertainty by difference of convex functions optimization.* PhD thesis, Université de Lorraine, 2015.

[4] H. A. Le Thi and T. P. Dinh, "DC programming and DCA : thirty years of developments," *Mathematical Programming*, vol. 169, no. 1, pp. 5–68, 2018.

[5] A. R. Conn, K. Scheinberg, and L. N. Vicente, "Global convergence of general derivative-free trust-region algorithms to first-and second-order critical points," *SIAM Journal on Optimization*, vol. 20, no. 1, pp. 387–415, 2009.