On function monotonicity in simplicial branch and bound

E.M.T. Hendrix[1,a),*], J.M.G. Salmerón[2,b),*] and L.G. Casado[2,*]

[1] *Computer Architecture, Universidad de Málaga.*
[2] *Departmento de Informática, Universidad de Almería (ceiA3).*

[a)] Corresponding author: eligius@uma.es
[b)] (leo,josemanuel)@ual.es

Abstract. Branch and bound methods in Global Optimization guarantee to find the set of global minimum points up to a certain accuracy. Partition sets typically take the shape of boxes, cones and simplices. In the tradition of interval arithmetic for generating bounds on box shaped partition sets, the concept of monotonicity test is used in order to reduce dimension. This paper shows how such a concept can be extended to simplicial partition sets and how to elaborate it for the Standard Quadratic Program.

INTRODUCTION

Branch and bound (B&B) algorithms in Global Optimization aim at providing a guarantee to find the set of global optimum points. [1, 2] provide an overview of the used partition sets and the mathematical structure of the optimization problems that can be used to derive bounds. We focus on the use of simplicial partition sets as outlined in [3] versus the use of boxes (intervals) which is the central point in interval arithmetic (IA) based bounding, see [4, 5]. We observe that in IA, one of the tests on having optimal solutions in a box is based on the so-called monotonicity test. Our research question is whether such a concept can easily be extended to simplicial partition sets. To investigate this question, we first embed the monotonicity idea in a general B&B framework. From this perspective, we formulate the monotonicity considerations for simplicial sets as bounds on directional derivatives and elaborate them for the case of the standard quadratic program.

MONOTONICITY IN A GENERAL FRAMEWORK

Algorithm 1 sketches a generic B&B scheme to find the minimum of continuous function $f : \mathbb{R}^n \to \mathbb{R}$ over set X for cases where lower bound calculation also involves the generation of a feasible point. It generates approximations of global minimum points that are less than δ in function value from the optimum. The method starts with a set C_1 enclosing the feasible set X of the optimization problem. For the ease, we assume minimization and the set X to be compact. At every iteration the B&B method has a list Λ of subsets (partition sets) C_k of C_1. As said, several geometric shapes can be used like cones, boxes and simplices. The method starts with C_1 as the first element and stops when the list is empty. For every set C_k in Λ, a lower bound f_k^L of the minimum objective function value on C_k is determined. For this, the mathematical structures discussed in [1] can be used.

At every stage, there also exists a global upper bound f^U of the minimum objective function value over the total feasible set defined by the objective value of the best feasible solution found thus far. The bounding (pruning) operation concerns the deletion of all sets C_k in the list with $f_k^L > f^U$, also called cut-off test. Besides this rule for deleting subsets from list Λ, a subset can be removed when it does not contain a feasible solution. In Algorithm 1, index r represents the number of subsets which have been generated. Note that r does not give the number of subsets on the list.

[*]This work has been funded by grants from the Spanish Ministry (TIN2015-66680), in part financed by the European Regional Development Fund (ERDF). J.M.G. Salmerón is a fellow of the Spanish FPU program.

Algorithm 1 General branch and bound algorithm for GO **B&B**(X, f, δ, ϵ)

Determine a set C_1 enclosing feasible set X, $X \subset C_1$,
Determine a lower bound f_1^L on C_1 and a feasible point $x_1 \in C_1 \cap X$
if (there exists no feasible point) STOP
else $f^U := f(x_1)$; Store C_1 in Λ; $r := 1$
while $(\Lambda \neq \emptyset)$
    Remove (selection rule) a subset C from Λ and split it
        into h new subsets $C_{r+1}, C_{r+2}, \ldots, C_{r+h}$
    Determine lower bounds $f_{r+1}^L, f_{r+2}^L, \ldots, f_{r+h}^L$
    for $(p := r + 1$ to $r + h)$ do
        if $(C_p \cap X$ contains no feasible point)
            $f_p^L := \infty$
        if $(f_p^L < f^U)$
            determine a feasible point x_p and $f_p := f(x_p)$
            if $(f_p < f^U)$
                $f^U := f_p$
                remove all C_k from Λ with $f_k^L > f^U$ cut-off test
            if $(f_p^L > f^U - \delta)$
                Save x_p as an approximation of the optimum
            else if $(Size(C_p) \geq \epsilon)$ store C_p in Λ
    $r := r + h$
endwhile

There are several reasons to remove subsets C_k from the list, or alternatively, not to put them on the list in the first place.

- C_k cannot contain a feasible solution.
- C_k cannot contain the optimal solution as $f_k^L > f^U$.
- C_k has been selected to be split.
- It has no use to split C_k any more. This may happen when the size $Size(C_k)$ of the partition set has become smaller than a predefined accuracy ϵ, where $Size(C) = \max_{v,w \in C} \|v - w\|$.

Branching concerns further refinement of the partition. This means that one of the subsets is selected to be split into new subsets. There exist several ways for doing so. The selection rule determines the subset to be split next, and influences the performance of the algorithm. One can select the subset with the lowest value for its lower bound (best first search) or for instance the subset with the largest size (relatively unexploited); breadth first search. The target is to obtain sharp bounds f^U soon, such that large parts of the search tree (of domain C_1) can be pruned.

Interval B&B algorithms ([4, 5]) follow the concept of enclosing the optimum by a box. The final result is a list of boxes of which the union certainly contains the global optimizers and not a list of global optimum points. The upper bound is not necessarily a result of evaluating the function value in a point, but can be based on the lowest (guaranteed) upper bound over all boxes. Moreover, in differentiable cases there is also a so-called monotonicity test. One focuses on partial derivatives $\frac{\partial f}{\partial x_j}(x) = \nabla_j f(x)$, which for a box C is included by $F_j'(C)$. If $0 \notin F_j'(C)$, then it cannot contain a stationary point and one should consider whether C_j contains a boundary point of the feasible area. If this is not the case, one can discard the box. The question is how to extend this analysis for simplicial subsets. What part of the boundary is of interest and can we really conclude that it cannot contain a global optimal solution?

MONOTONICITY IN SIMPLICIAL PARTITION SETS

Consider an $n \times m$ vertex matrix V with m affine independent columns and the corresponding convex hull simplex

$$\Delta = \{x = V\lambda, \sum \lambda_j = 1, \lambda_j \geq 0, j = 1, \ldots, m\} = \{x = V\lambda, \lambda \in S_m\} \tag{1}$$

with $S_m = \{\lambda, \sum \lambda_i = 1, \lambda_i \geq 0, i = 1, \ldots, m\}$ the m-dimensional unit simplex. Let us define the relative interior $\breve{S}_m$ of S_m as

$$\breve{S}_m = \{x \in \mathbb{R}^m \mid \sum_{j=1}^{m} x_j = 1;\ x_j > 0, \quad j = 1, \ldots, m\}. \tag{2}$$

The relative interior of $\breve{\Delta}$ is defined similarly and the relative boundary is $\partial S_m = S_m \setminus \breve{S}_m$. Besides the identity matrix $E_m = (e_1, \ldots, e_m)$ in m-dimensional space, we will also use a matrix $D_m = (d_1, \ldots, d_m) = E_m - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ of feasible directions over unit simplex facets, where $\mathbf{1}$ is the all-ones vector of appropriate dimension. Our focus is on a lower bound of the directional derivative $\underline{F}_j \leq d_j^T V^T \nabla f$ over Δ and facet $\Phi_j = \{x = V\lambda, \lambda \in S_m, \lambda_j = 0\}$ of Δ.

Proposition 1 *Let V be a vertex matrix, Δ from* (1)*, matrix D with columns d_j and lower bound of the directional derivative $\underline{F}_j \leq d_j^T V^T \nabla f$ over Δ. If $\underline{F}_j \geq 0$, $\exists x^* \in \operatorname{argmin}_\Delta f(x)$ at Φ_j.*

Proof. Assume the minimum is attained at $x^* = V\lambda^*$ with $\lambda_j^* > 0$. Now consider the point $x = V\lambda \in \Phi_j$ with $\lambda = \lambda^* - \lambda_j^* d_j$. According to the mean value theorem, there exists $0 \leq \xi \leq 1$

$$f(x) = f(x^*) + (x - x^*)^T \nabla f(V(\lambda^* - \xi d_j) = f(x^*) - \lambda_j^* d_j^T V^T \nabla f(V(\lambda^* - \xi d_j)). \tag{3}$$

Analyzing the right hand term, one can observe that $\lambda_j^* > 0$ and the directional derivative is nonnegative. This means that $f(x) \leq f(x^*)$ which shows that x is a minimum point on facet Φ_j. □

The importance of this theoretical result is that one can reduce the search for the minimum point to a certain lower dimensional facet. The monotonicity observations can be made sharper. If the relevant facet is not on the relative boundary ∂X of the search region and the directional derivative is strict positive, apparently the minimum is not in Δ. Formally:

Proposition 2 *Let simplex $\Delta \subset \breve{X}$ be defined by vertex matrix V and centroid $c = \frac{1}{m}V\mathbf{1}$. If $\exists j, \underline{F}_j > 0$, $\operatorname{argmin}_X f(x)$ cannot be found on Δ.*

Proof. According to Proposition 1, the local minimum point $x^* = V\lambda^*$ is attained at facet Φ_j where $\lambda_j^* = 0$. Now consider feasible direction $h = c - v_j = -Vd_j$ in point x^* defining one-dimensional function $\varphi(\mu) = f(x := x^* + \mu h)$. Notice that $\varphi'(0) = h^T \nabla f(x^*) = -d_i^T V^T \nabla f(x^*) < 0$. As $\Delta \subset \breve{X}$, $\exists x$ for $\mu > 0$, such that $f(x) < f(x^*)$ arbitrarily close to x^* This shows that x^* cannot be global minimum point on Δ. □

These conditions extend more or less the monotonicity test of IA for boxes. The question is of course how to elaborate the conditions for a specific case. Therefore we show that the conditions can even be a bit sharper for the Standard Quadratic program, where X is a convex set.

CONSEQUENCE FOR THE STANDARD QUADRATIC PROGRAM

The standard quadratic program (StQP) is defined as $\min_{S_n} f(x) := x^T Ax$ for a symmetric matrix A. Although one can doubt whether a spatial B&B algorithm with simplicial partition sets is the most appropriate to find the set of minimum points, we discuss how the monotonicity propositions can be implemented for this class of problems. Notice that $\nabla f(x) = 2Ax$ and the feasible set being the unit simplex S_n is convex. Moreover, a lower bound F_j on the directional derivative is now provided by $F_j = 2\min_i d_j^T V^T A v_i$. The propositions and proofs are elaborated specifically for the StQP.

Proposition 3 *Let V be vertex matrix, Δ from* (1)*, matrix D with columns d_j and $Q = V^T AV$. If $d_j^T Q \geq 0$, $\exists x^* \in \operatorname{argmin}_\Delta f(x)$ at the facet Φ_j.*

Proof. Assume the minimum is attained at $x^* = V\lambda^*$ with $\lambda_i^* > 0$. Now consider the point $x = V\lambda$ with $\lambda = \lambda^* - \lambda_i^* d_i$ which is on the facet corresponding to $\lambda_i = 0$. One can write

$$f(x) = f(x^*) + (x - x^*)^T A(x + x^*) = f(x^*) + \lambda_i^* d_i^T Q(2\lambda^* - \lambda_i^* d_i). \tag{4}$$

Analyzing the right hand term, one can observe that $\lambda_i^* > 0$, $d_i^T Q \geq 0$ and component wise $2\lambda^* - \lambda_i^* d_i > 0$. This means that only in the specific case that $d_i^T Q = 0$ we have $f(x) = f(x^*)$, else $f(x) < f(x^*)$ which shows that x is a minimum point on facet Φ_j. □

The importance of this theoretical result is that, for problem StQP, one can reduce the search for minima to a lower dimensional facet, if the condition of Proposition 3 applies. Basically, one can remove vertex v_j from matrix V and proceed the search on a lower dimensional facet.

The monotonicity observation of Proposition 2 can be made sharper and implemented due to X being a convex set. It is not completely necessary to have $\Phi_j \subset \check{X}$, as long as we have $\Phi_j \not\subset \partial X$. Formally:

Proposition 4 *Let simplex $\Delta \subset S_n$ be defined by vertex matrix V and centroid c. If $\exists i, d_i^T Q > 0$ and $\forall \ell$ with $V_{\ell i} > 0$ and $V_{\ell j} > 0, \forall j \neq i$, the minimum point of the StQP cannot be found on Δ.*

Proof. According to Proposition 3, the minimum point $x^* = V\lambda^*$ is attained at the facet with $\lambda_i^* = 0$. Now consider direction $h = c - v_i = -Vd_i$ in point x^* defining one-dimensional function $\varphi(\mu) = f(x := x^* + \mu h)$.

Notice that $\sum_\ell h_\ell = 0$. Moreover, for elements ℓ with $V_{\ell i} = 0$, $x_\ell \geq 0$ and for elements ℓ with $V_{\ell i} > 0$, $x_\ell^* > 0$, such that $\exists \epsilon > 0$ with $\mu < \epsilon$ having also $x_\ell := x_\ell^* + \mu(c_\ell) - V_{\ell i} \geq 0$. So, for $x = x^* + \mu h$ and $\mu < \epsilon$, $\sum_\ell x_\ell = 1$ and $x_\ell \geq 0$ implies h is a feasible direction in S_n from point x^*.

Directional derivative $\varphi'(0) = h^T \nabla f(x^*) = -d_i^T V^T A x^* = -d_i^T Q \lambda^* < 0$. So, h is a feasible descent direction in point x^* on S_n implying that x^* cannot be the global minimum point in Δ. □

CONCLUSIONS

We asked ourselves the question how to extend concepts of monotonicity known in B&B applying Interval Arithmetic to B&B where simplicial partition sets are used. The analysis burns down to focusing on directional derivatives in the direction of centroid to vertices. In such cases, one can derive on which facets to focus the search for global minimum points. We investigated the implementation of the theoretical results to the Standard Quadratic Program.

REFERENCES

[1] R. Horst and H. Tuy, *Global Optimization (Deterministic Approaches)* (Springer, Berlin, 1990).

[2] E. M. T. Hendrix and B. G.-. Tóth, *Introduction to Nonlinear and Global Optimization* (Springer, New York, 2010).

[3] R. Paulavičius and J. Žilinskas, *Simplicial Global Optimization* (Springer, New York, 2014).

[4] R. E. Moore, *Interval analysis* (Prentice-Hall, New Jersey, 1966).

[5] E. Hansen, *Global Optimization Using Interval Analysis*, Pure and Applied Mathematics, Vol. 165 (Marcel Dekker, Inc, New York, 1992).