EURO Journal on Computational Optimization 9 (2021) 100012

Contents lists available at ScienceDirect

EURO Journal on Computational Optimization

journal homepage: www.elsevier.com/locate/ejco

(Global) Optimization: Historical notes and recent developments

Marco Locatelli[a], Fabio Schoen[b,*]

[a] DIA - Università di Parma, Parco Area delle Scienze, 181/A, Parma - 43124, Italy
[b] DINFO - Università degli Studi di Firenze, via di S. Marta, 3, Firenze - 50139, Italy

ARTICLE INFO

Keywords:
Global optimization
Heuristics
Exact methods

ABSTRACT

Recent developments in (Global) Optimization are surveyed in this paper. We collected and commented quite a large number of recent references which, in our opinion, well represent the vivacity, deepness, and width of scope of current computational approaches and theoretical results about nonconvex optimization problems. Before the presentation of the recent developments, which are subdivided into two parts related to heuristic and exact approaches, respectively, we briefly sketch the origin of the discipline and observe what, from the initial attempts, survived, what was not considered at all as well as a few approaches which have been recently rediscovered, mostly in connection with machine learning.

1. introduction

It seems worthwhile, in a special issue like this one, to devote a few words to the foundation of (global) optimization as an independent research topic. Looking back to the early days might prove useful, as some old ideas which, in the beginning, did not lead to much development, might prove to be interesting for current research, when considered from a modern perspective. Moreover, it is interesting to notice how some of the sub-fields which have seen a very large set of contributions in recent years, were almost totally neglected in the beginning. We refer in particular to exact (global) optimization methods based on implicit enumeration and to the jungle of "nature inspired" population heuristics. For what concerns this kind of heuristics, we need of course to cite at least the book Holland (1975) where the basic ideas of genetic algorithms are nicely introduced. In what follows, we will first give a quick overview of the main approaches contained in the two classical books devoted to the subject in the 70's (Dixon and Szegö, 1975; 1978). Then, we will give more details on recent developments. We also refer to our book Locatelli and Schoen (2013) for a detailed discussion about (global) optimization topics, updated as of the date of publication of the book. After the historical introduction, the paper will consider two topics: recent, or re–discovered, heuristic approaches and recent developments in exact approaches. A final observation before the beginning of the paper: here and in the title we parenthesized the word (global). We will omit doing so in the paper, but we would like to observe that, although in the past the subject was considered somewhat exotic and off the main research streams, nowadays the richness of both theory as well as computational approaches gives to the subject a full recognition in the scientific community. We might then propose, with a slightly provocative style, to rename the whole subject simply as "optimization" – what else should we look for when optimizing, if not a global optimum?

The paper is structured as follows. After the brief introduction given in this section, in Section 2 we recall the contents of the two books edited in the70ıs by Lawrence Dixon and Giorgio Szegö, which gave a strong initial impulse to the whole discipline. In Section 3 we review some of the many heuristic computational approaches published in the literature in recent years (say, roughly after the publication of our book Locatelli and Schoen (2013), of which this paper might be considered as a continuation). Section 4 surveys recent literature dealing with structured optimization problems for which an exact procedure can be designed. In Section 5 we briefly discuss some computational aspects and suggest sites where exhaustive lists of GO test problems and solvers can be found. Some concluding remarks are finally presented.

2. On the origins of global optimization

Many early papers dealt with non convex optimization problems and outlined basic algorithms. We recall here, as particularly interesting examples, Dantzig (1960), McCormick (1972), Beale and Forrest (1976), Falk and Soland (1969), Soland (1971), McCormick (1976), Horst (1976). All of these papers had impact on the whole field, as well as many others we are not citing here. Despite the relevance of these as well as many other early papers, it can quite safely be assumed that the first "large scale" diffusion of the ideas of Global Optimization (denoted by GO in what follows) can be credited to the two "orange" books Dixon and Szegö (1975, 1978). These two books, although clearly not the first publications in GO, gave a fundamental impulse to the whole research field and since their publication, the term "Global Optimization" started to be a recognized and respected label, characterizing optimization the-

* Corresponding author.
E-mail addresses: marco.locatelli@unipr.it (M. Locatelli), fabio.schoen@unifi.it (F. Schoen).

https://doi.org/10.1016/j.ejco.2021.100012
Received 30 October 2020; Received in revised form 20 September 2021; Accepted 25 September 2021

ory and methods for nonconvex optimization problems. To insist on the fact that the field was just in its infancy at that time, it might be observed that only 9 of the 25 papers in the first volume and only 12 out of 24 in the second one were grouped in the "Global Optimization" chapters, all the others being local optimization papers, despite the title of the two books. It might be of interest to observe the topics which, at that time, were considered as the most promising ones. In this section we will briefly review some of the main ideas proposed at that time and comment on their success in the GO literature and practice.

We can group the first few papers on global optimization as follows:

Space covering. This area was related to methods aiming at implicitly exploring the whole feasible region. In particular, methods were described, mainly for 1–dimensional optimization, which, based on the knowledge of (an upper bound to) the Lipschitz constant of the objective function, could be built with a guaranteed maximum error in the approximation of $f^\star = \min_{x\in S} f(x)$. In the survey by Dixon (1975) the methods of Evtushenko (1971) and of Shubert (1972) (later known as Piyavskii-Shubert (Piyavskii, 1972)) were described. In the following years those approaches generated a stream of research in Lipschitz–optimization, with some interesting approaches but also severe limitations. In general, as nicely described in another milestone in the field, Horst and Tuy (1993), these can be seen as precursors and special cases of modern Branch & Bound algorithms. The idea of Branch & Bound for nonlinear optimization was not considered at all in the two original books.

Trajectory methods. At those times the idea of following the trajectory of a suitable set of differential equations seemed to be very promising. Some of the proposed approaches required locating saddle points in the boundary of the region of attraction of known local optima in order to be able to escape and explore new basins. A group of papers in the book, mostly from Joanna Gomulka and Giorgio Treccani, explored this idea. In the following years this approach received less attention, although some similarities can be traced with some approaches based on stochastic differential equations and with some recent approaches in the neural network literature. A recent survey on developments in this algorithmic family can be found in Alexandropoulos et al. (2020).

Bayesian Optimization. At those times, GO problems with as few as ten variables, with just box constraints, were considered as the frontier in computational GO approaches. Much of the effort then was devoted to very small scale problems; even at small scale, black-box problems were considered as very relevant. In fact, practical problems arise in which the objective function is not available in analytical form, but can only be evaluated at specific points, possibly through an expensive procedure. Among the most interesting and profound ideas for small dimensional, possibly 1–dimensional, GO problems, Bayesian Optimization (BO) was particularly relevant. The main ideas of BO can be traced back to papers published in the 60's by Harold Kushner (see, e.g., Kushner (1964)) and then generalized in Mockus (1975) and in many papers by the same author. The idea of BO is fascinating: assume the objective function is a realization of a stochastic process, typically a Gaussian one. Then, given a prior distribution on the possible sample paths of the process, after a few observations of the objective function, possibly affected by errors, have been performed, a posterior distribution can be computed, leading to a conditional stochastic model which is further updated as soon as the true objective function is observed at new sample points. This stochastic model can be analyzed in order to answer, through suitable numerical algorithms, queries like: find the point at which the expected value of the function is minimum, or find the point at which the expected improvement over the current best observation is maximum. These queries are themselves GO problems, but, differently from the original one, possess a known analytical expression, from which gradients can be analytically computed; the global optimization required can be carried out with standard GO methods, without any necessity of excessive precision. It is worth observing that, although born in a stochastic framework, these methods are indeed deterministic. A nice and relevant generalization and implementation of the BO ideas can be found in Jones et al. (1998), currently among the top cited papers in the whole GO literature. Among the reasons for the great success of this paper we can mention the fact that, after a period in which BO almost disappeared from the literature, as a consequence of the high increase in complexity per iteration in the multi–dimensional case, the implementation of EGO described in Jones et al. (1998) recently became a standard for hyperparameter optimization in machine learning (see, e.g., Frazier (2018) for a recent survey). The field has considerably expanded in the last years and it is the subject of very active research.

Random Search. Methods based on some form of random sampling were considered in the two books, as they are in general simple to implement and do not require first order information on the objective function. In Gaviano (1975) some general convergence results were proven; quite a few papers proposed random search algorithms, which, however, did not find great attention in the following years. The field was quite productive for some time, and new approaches have been published in recent years. Some of those we consider as the most interesting ones are reviewed in our book Locatelli and Schoen (2013).

Clustering. This, at the time, seemed to be one of the brightest and most innovative ideas. The efficiency of the most elementary GO method, Multistart (sample some random points and start a local search from each of them), is improved by carefully selecting from which sample points it seems worth starting an expensive local optimization. Ideally, a single local search should be started from each basin of attraction (and, possibly, not from all of them, but just from the most "promising" ones). This idea was first proposed in Becker and Lago (1970) and then has been expanded and made available to a larger audi-

ence in Törn (1978). The idea of the proposed approach is both simple and powerful. First draw a uniform sample of feasible points. Then "concentrate" the sample towards local optima either by temporarily discarding a fraction of highest value observations, or by performing a few descent steps. Finally, identify higher density regions (clusters), from each of which a single local search is started. The approach had immediately a great success and seemed to represent "the" GO method of choice, until a few years later, when it gradually got abandoned. There might be many reasons for the decline of clustering methods. We conjecture that this might be due to:

- the practical impossibility of dealing with large dimensional problems, say with more than 10 variables;
- the fact that the method could be used just to speed up Multistart, while more advanced algorithms appeared in the literature;
- the method became popular as it allows to save on local searches which, at the time, were very time consuming. With the advent of modern local optimization methods, although saving unnecessary local searches remains a good point, the efficiency of local search made those savings quite negligible in the overall method.

In recent years, some new methods have been proposed (see Bagattini et al., 2018; Bagattini et al., 2019; Schoen and Tigli, 2021) which revisit these ideas trying to soften as much as possible the above defects.

3. Recent heuristic GO methods

In this section we would like to present a few recent papers dealing with modern GO algorithms. The papers cited in this section come either from the vast field of evolutionary approaches or have been stimulated and, in some cases, re-discovered thanks to the exponential growth in the interest around machine learning. We point out that the review of recent literature will be partial and biased towards our knowledge and preferences. It is really an impossible task to track all the papers dealing with GO heuristics, and even more difficult to highlight those which are indeed relevant and truly innovative. But we tried to do our best in describing interesting new approaches in this wide field.

3.1. Population-based, evolutionary methods

An enormous quantity of papers dealing with variants of the basic evolutionary population schemes appeared in recent years. In the authors' opinion it is quite disturbing that most of the papers in this subfield justify themselves by some sort of an inspiration from nature, and, frequently, do not report any theoretical justification of the algorithmic choices, nor a fair and wide numerical comparison with state-of-the-art algorithms on well recognized benchmarks of test functions. A characteristic within this family of approaches is that they are mostly based on some variation of very basic schemes which include:

- a random generation of the initial set of solutions (initial population);
- a combination mechanism which takes parts of the components of the solutions represented in the current population and generates a new set of solutions;
- possibly a mutation, through which some solutions are randomly perturbed;
- possibly, in memetic algorithms, a local search applied to a selection of the elements in the current population;
- a substitution criterion, based on observed function values, by which from the original and the current population, a new set of solutions is built.

Some standard algorithms arose from this basic scheme. The most notable, in GO, are Differential Evolution (DE) and Particle Swarm (PS), for whose description we address the reader to the rich literature on the subject or to our chapter in Locatelli and Schoen (2013). Some recent surveys have been published on this subject (see, e.g., Das et al. (2016) or Del Ser et al. (2019)). In the recent literature some new proposals appeared like, e.g., Cui et al. (2016) or Wu et al. (2016), where multiple populations, with different evolution strategies, are evolved simultaneously and compete among themselves to improve the best population. It seems that memetic variants of DE stand out as a good compromise between simplicity and quality. For DE we can even cite some theoretical convergence results (Ghosh et al., 2012; Locatelli and Vasile, 2015) which can suggest guidelines for algorithm definition.

Among many variants of DE, those based on the exploitation of local optimization (memetic variants) are very interesting, as many GO problems allow for fast and reliable local search tools. In Cabassi and Locatelli (2016) an analysis of some variants of the basic memetic DE is introduced and numerically shown to be very effective on a wide range of GO test problems of varying dimension. In Schoen and Tigli (2021) those methods have been extended and mixed with a clustering approach in order to save unnecessary local searches; numerical results show that it is possible to significantly improve the efficiency of those population–based methods while keeping their good quality. On a different line of research, in Mansueto and Schoen (2021) a DE–based memetic approach is used to build an efficient GO method for optimal clustering in Euclidean spaces. In that paper a specialized local search is used, based on the well known K–means clustering algorithm, coupled with the exploration capabilities of DE.

A source for many evolutionary algorithms, as well as test problems, can be found in the various CEC (the IEEE Congress on Evolutionary Computation) competition websites. It appears that, in those competitions, a dominant role is played by population–based methods which in some way exploit the separability or partial separability of problems. The idea of, at least partially, decompose a GO into sub-problems is adopted by quite a large number of successful approaches in those competitions. Of course, for separable problems it is easy to forecast their good performance, but numerical results seem to be quite interesting also for non separable ones. The paper Ma et al. (2019) contains a survey on decomposition–based methods and presents the basic ideas of these approaches. There are many variants of the basic scheme, which typically differ in the decomposition strategy and the recombination one; many are based on a Gauss-Seidel–like strategy in which optimized subsets of variables are fixed and used while optimizing different subsets. We can cite here, as an example of an efficient GO approach, Hadi et al. (2019), where, starting from an adaptive DE scheme in which some hyperparameter is adjusted during the evolution, a hybrid, decomposition–based, memetic approach is proposed. In particular, at some stage of the computation, variables are randomly grouped and different local optimization algorithms are associated to each sub-group of variables. The winner of the 2019 CEC competition Sun et al. (2019) uses an innovative strategy to decompose non separable problems which leads to a highly efficient method. Their decomposition scheme is based on the identification of subsets of variables which partially overlap one another.

Finally, we would like to mention the family of CMA-ES (Covariance Matrix Adaptation Evolutionary Strategies) algorithms (see, among many papers on the subject, Hansen and Ostermeier (2001) for an early introduction to the approach, or Hansen (2006) for a survey on variants of the basic method). The idea behind CMA-ES is that of having a population of solutions to a GO problem which evolves through sampling from

a multinomial distribution whose mean and covariance matrix evolve during the iterations and in some sense adapts to the level sets of the objective function. The objective function is evaluated at each sample point and a new mean and covariance matrix are generated through an updating mechanism. Oversimplifying a description of this family of approaches, which is indeed quite vast, we might say that they try to adapt search directions and step sizes in order to favour changes in the current population members which are likely to contribute much to the improvement of the objective function. This is usually obtained by following directions associated to the principal components of the covariance matrix, which is built in such a way as to adapt to observed function values. In Diouane et al. (2015a,b) an interesting extension of the basic scheme is presented in which, imposing a criterion of sufficient decrease in the objective function, some convergence properties are obtained.

3.2. Basin-Hopping methods

Multistart is one of the simplest GO methods, where a local search is started from each point randomly generated within the feasible region of a GO problem. It is usually considered a low efficiency algorithm for GO due to the computational waste it produces in rediscovering the same local optima more than once and due to its absence of any learning mechanism (see Section 3.4 for further comments on these topics). However a relatively simple modification of Multistart, which goes under the name of Basin Hopping (BH), or Iterated Local Search, is quite an interesting approach for many hard GO problems. As in Multistart, also in BH local optimization is performed starting from a random initial point. However in BH, after a local optimum is found, new local searches are performed starting from a (suitably defined) neighborhood of the current one until possibly a better local optimum is found; in this case new local searches are performed starting from a neighborhood of the new local minimum, and the whole procedure is repeated until some stopping rule is satisfied. The version we sketch here is the monotonic version of the method: non monotonic ones have also been successfully implemented. BH has been used in Vinkó and Gelle (2017) as a tool to build a graph of neighboring local optima, much in the sense of what was described in Chapter 3 of our book Locatelli and Schoen (2013). Given a BH algorithm with specific parameters, a run on a specific test function generates a set of local optima which can be pairwise connected in a graph in which an oriented arc exists between two local minima if the second one has been reached through a BH step from the first one. The resulting graph associated to some classical test functions can be analyzed. It might be possible to exploit this information in order to build advanced BH methods, but this is still a subject of current research. A hybrid approach, mixing Differential Evolution with BH is presented in Di Carlo et al. (2020), where a simple criterion is presented to save unnecessary local searches.

Most recent literature on BH and its variations, including population–based variants as discussed, e.g., in Grosso et al. (2007), deals with application of these methods. Many of those applications are in computational chemistry, a field where BH was born and has found a prominent space in current research. As an example, in Zhao et al. (2017) a modified BH is applied to the optimization of atomic clusters. Among the many variations, in this paper it is suggested that some variables are kept fixed at certain iterations (those corresponding to the location of specific subsets of atoms), while some others are subject to perturbation in BH steps. Moreover, some acceptance criteria different from the improvement of the objective function are suggested, like the exploration of new geometrical configurations. In Ferreiro-Ferreiro et al. (2019) a simple modification of the BH scheme is introduced and tested on specific atomic clusters. The idea is to let BH explore deeper basins by reducing its “greediness”. To this aim, at each iteration, several neighboring local optima are generated and, instead of moving towards the first improving one, the best is chosen as the starting point of the next iteration. In Wales (2018) variants of BH are described in the context of energy landscapes for atomic clusters. Different acceptance rules are proposed that differ from the classical monotonic and Metropolis-like acceptance rules, generalizations to multiple objectives in the exploration phase are introduced, and further refinements based on molecular dynamics are proposed.

3.3. Methods for expensive black-box objective functions

This field has recently attracted enormous attention, driven by machine learning research, as algorithms within this framework are considered as especially suitable for hyperparameter calibration in training machine learning tools like, e.g., deep learning architectures. A survey on the subject of optimization–based hyperparameter choice in machine learning recently appeared in Yang and Shami (2020), where a list of available software is also provided. Also, in Tran et al. (2020) a discussion on the use of optimization tools in hyperparameter tuning is presented. In applications for hyperparameter setting, two levels are usually present. At the lower level, given a training set, learning proceeds by suitably choosing (in neural networks, as an example) connection weights in such a way that a loss function is minimized; at the upper level the loss function itself, the overall architecture and, sometimes, the optimization algorithm used for training, depend on a relatively small number of hyperparameters (e.g., the parameter associated to regularization, the number of layers, the “learning rate”,...). These hyperparameters are usually optimized considering the performance of the lower level trained system on a validation set, different from the one used for training. In general, thus, the objective function to be optimized at this level is neither known, nor cheap, and is modeled as a black box. GO methods for expensive cost functions are among the primary choices for this tuning. We will not even try to summarize the overwhelming quantity of reports recently published on this subject on arXiv. Although excellent research is sometimes found in that dynamically exploding repository, here we chose to refer exclusively to papers published in high quality refereed journals. A nice survey of state-of-the-art approaches in Bayesian Optimization can be found in Frazier (2018), while in Zhan and Xing (2020) several different objective functions based on variation of the originally proposed Expected Improvement are presented and discussed. In Mathesen et al. (2020) an interesting generalization of Bayesian Optimization models is proposed, in which the “acquisition function” takes into account the desire to improve the exploration capabilities, as the exploitation is delegated to a trust-region based local search which takes into account the available budget of function evaluations. It is interesting how, even in an expensive function setting, the idea of performing local optimization can be effectively exploited. In Ahmed et al. (2020) it is suggested to use an estimate of the Lipschitz constant in order to improve the search for the next observation point, an approach which, at least for moderately sized problems, might have some relevance. Instead, in Bemporad (2020) the proposed approach is not based on a stochastic model of the objective function, but, taking inspiration from those models, the author proposes a surrogate function composed of three elements: an interpolation of observed sample values, a distance from the points in the sample, distance of the objective function values. This acquisition function is optimized in order to decide where the next observation should be placed.

A recent stream of research has started exploring the possibility of applying the idea of Bayesian Optimization to large scale GO problems. The main tool used to greatly increase the dimension of problems which can be solved through this approach is the idea of *random embeddings*. The philosophy of these approaches (see Binois et al., 2020; Moriconi et al., 2020; Wang et al., 2016) relies on the assumption that for large dimensional problems there exists a true “low effective dimensionality”, which can be considered as a low–dimensional linear embedding such that the objective function does not change when moving along directions which are orthogonal to this embedding. A differ-

ent approach to large-scale adaptation of these methods is reported in Snoek et al. (2015), where a neural network is used to learn a set of basis functions to be used to build a Bayesian linear regressor which substitutes the computationally expensive classical one.

Among the most interesting approaches in the Bayesian Optimization literature, it seems worth citing extensions to the constrained case and, in particular, to the case in which also the constraints are expensive, black-box, functions. Hernández-Lobato et al. (2016) present a detailed survey of constrained Bayesian Optimization methods. A novel framework is also introduced for the case where at each iteration a new sample point needs to be generated at which both the objective function as well as the constraints have to be evaluated. However, in the proposed approach, it is assumed that functions (objective and constraints) might be at least partially separable into parts (so called "tasks") – as a simple example, the objective function might be evaluated on a CPU while the constraints are defined in a way that enables to exploit the parallelism of a GPU. In these cases, the merit (or "acquisition") function which guides the search for the next evaluation point, should be defined in such a way as to support a similar decomposition. In that paper, a novel acquisition function, called PESC (Predictive Entropy Search with Constraints) is proposed as particularly well suited in this framework. Constrained Bayesian Optimization is also considered in Feliot et al. (2017), mostly in the context of an extension of the basic approach to multi-objective optimization, an interesting topic which is, however, out from the scope of this paper.

On a different line of research, in Garrido-Merchán and Hernández-Lobato (2020) the problem is considered of adapting the approach to integer and categorical variables, both ordered as well as unordered. The interest on such a kind of extension has been greatly stimulated by the application of GO to hyperparameter calibration in machine learning, where, e.g., in a neural network, one might wish to calibrate the number of neurons, of layers, of filters on a convolution, which are integer–valued parameters. Otherwise hyperparametrs can be associated, e.g., to the activation function at a neuron, or to the loss function, or to the kernel function used in a Support Vector Machine; each of these can be seen as an example of categorical unordered parameter. The field has seen much expansion and interest in the scientific community and papers started to appear describing novel software implementations, like, e.g., Kandasamy et al. (2020); Martinez-Cantin (2014).

We conclude this subsection by citing Kim (2020), a recent theoretical paper which provides a rational explanation on how Bayesian Optimization algorithms proceed by alternating exploitation, or learning, phases with exploration, or optimization. Exploitation is connected with a greedy cost optimization strategy, while exploration is associated to variance regularization. The whole process, analyzed in the context of Dynamic Programming, is analyzed and it is shown how the temporal discount factor of a Markov Decision Process plays the role of trade–off parameter between exploration and exploitation.

3.4. Methods on the intersection between machine learning and GO

The title of this subsection refers to methods in which the search towards a global optimum, or, at least, a good local one, is guided by some form of "learning", much in the spirit of modern machine learning approaches. In this part we would like to include the re-discovery of clustering methods. These methods, as recalled in Section 2, were based on the idea of learning, from the sample, the shape of different basins of attractions of local optima, before starting expensive local searches. Clustering is one of the pillars of unsupervised learning and it might be of interest to look to those methods as early implementations of machine learning tools for GO. In Bagattini et al. (2018, 2019) and, Schoen and Tigli (2021) we proposed different strategies in order to overcome the limitations of original clustering methods. In particular, we showed that the idea of those algorithms can be successfully applied to much more refined algorithms than the standard Multistart, by showing how a memetic Differential Evolution variant, inspired by Cabassi and Locatelli (2016), can profitably save a very large number of useless local searches while maintaining the quality of the original. Moreover, we showed how to use the same ideas with methods based on local search methods which are, in a sense, more refined and "global" than standard, gradient-based, local optimization tools. Finally, some experiments on random projections enabled us to extend the approach to high dimensional problems.

It is worth noticing also that in the field of chemical physics many approaches are proposed in which GO methods (in particular, variations of basin hopping) are guided by knowledge on the problem domain obtained through machine learning. As an example, in Meldgaard et al. (2018), low dimensional features are extracted from the original variables of atomic clustering problems and regression is used to assign an energy contribution to each atom in order to let GO algorithms focus on the parts of the current solution which contribute most to the objective function.

We conclude this section citing a paper which does not belong to the literature on algorithms but is quite relevant to the subject. The paper Kawaguchi (2016) deals with a conjecture on the absence of local optima which are not global when using a quadratic loss in training a deep neural network. The results proven there are interesting, as the problem is neither convex nor concave, yet it seems that training with a local optimization algorithm can always lead to a global optimum solution, thus partially explaining the success of local methods in training neural networks.

4. Exact GO methods

While in the previous section we discussed heuristic approaches, in this section we deal with exact GO methods. We will first discuss "easy" GO problems, i.e., GO problems which can be reformulated as (tractable) convex problems or, at least, can be solved in polynomial time with respect to the size of the problem data and the inverse of the required precision (we refer, e.g., to Ben-Tal and Nemirovski (2001) for a detailed discussion about computational complexity in the context of continuous optimization problems and, in particular, in the context of convex optimization problems). Later on, we will discuss exact GO methods (basically, branch-and-cut approaches) which are usually applied to 'highly structured' GO problems. We will often refer to Quadratic Programming (QP) problems, which arise in many different contexts. Applications of QPs include the reformulation of some combinatorial optimization problems, like, e.g., max-clique and max-cut, portfolio optimization, packing problems, blending and pooling problems. More applications of QPs can be found in Furini et al. (2019), where a QP library is introduced.

4.1. "Easy" GO problems

Global optimization is, in general, a hard task. Even highly structured GO problems, like, e.g., Standard Quadratic Programming (StQP), where a quadratic form is minimized over the unit simplex, have been proved to be NP-hard. All the same, there are some classes of GO problems, in particular some QP problems, solvable in polynomial time. These include problems having the so called *hidden-convexity* property: though not convex, these problems can be reformulated as (tractable) convex problems. The best known of such problems is the *trust region* problem, where a quadratic function is minimized over the unit ball. It has been shown in Rendl and Wolkowicz (1997) that this problem admits a semidefinite reformulation. In fact, in Ben-Tal and den Hertog (2014) it has been shown that the trust region problem and, actually, a generalization where the unit ball constraint is replaced by a more generic quadratic constraint, can be reformulated as a simpler convex conic quadratic problem under the assumption that the Hessian matrices of the objective and constraint functions are simultaneously diagonalizable. After transforming both the objective and the

constraint into separable functions, the conic quadratic problem is obtained by first introducing new variables y_i and the related constraints $y_i = x_i^2$, and then relaxing these equalities into inequalities $x_i^2 \leq y_i$. The work Jiang et al. (2018) proposes a Second Order Cone Programming (SOCP) reformulation for this problem without requiring the assumption of simultaneous diagonalization. The paper Wang and Kilinç-Karzan (2020) proposes a convex reformulation in the original space of variables. Further problems with the property of hidden-convexity have been introduced in the last two decades. In Pong and Wolkowicz (2014) it is shown that, under suitable assumptions, a semidefinite relaxation for the problem of minimizing a quadratic function over a region defined by a two-sided quadratic constraint (i.e., a lower and upper bound are imposed over a quadratic function), is exact. For the same problem Jiang et al. (2018) presents a SOCP reformulation. The work Beck and Teboulle (2009) introduces a semidefinite reformulation for the problem of minimizing the ratio of two quadratic functions over a possibly degenerate ellipsoid. In Burer and Anstreicher (2013) and Sturm and Zhang (2003) it has been proven that problems with a quadratic objective function, a unit ball constraint and a single linear cut can be reformulated as semidefinite problems with an additional SOC-RLT (Second Order Cone - Reformulation Linearization Technique) constraint. The result has been extended in Burer and Anstreicher (2013) to the case of two parallel linear cuts and later on, in Burer and Yang (2015), generalized to the case of an arbitrary number of linear constraints, provided that such constraints do not intersect in the interior of the unit ball. For problems with a quadratic objective function and two quadratic constraints, in Ye and Zhang (2003) it is proved that if all the quadratic functions are homogeneous (i.e., there are no linear terms), then an exact SDP relaxation exists.

Many papers in the literature provide conditions under which some convex relaxation of a class of nonconvex problems turns out to be exact. For QCQP problems (problems where objective and constraint functions are all quadratic), the work Kim and Kojima (2003) provides some sign conditions about the data under which SDP and SOCP relaxations are exact. The paper Sojoudi and Lavaei (2014) considers QCQP problems without linear terms, i.e., with objective function $x^\top A^0 x$ and constraints $x^\top A^k x \leq 0$, $k = 1, \ldots, m$ (but it also discusses a way to include also linear terms). For these problems a graph $\mathcal{G}$ is built with n nodes (one for each variable) and an edge (i, j) exists if and only if $A^k_{ij} \neq 0$ for some $k \in \{0, 1, \ldots, m\}$. Next, exactness of SDP and SOCP relaxations are related to some sign conditions and to the structure of this graph. For instance, exactness holds if for each edge (i, j) all entries A^k_{ij}, $k \in \{0, 1, \ldots, m\}$, have the same sign, and graph $\mathcal{G}$ is acyclic. The paper Jeyakumar and Li (2014) considers the problem of minimizing quadratic functions over a feasible region defined by a ball constraint and linear constraints. A so called dimension condition is introduced under which the SDP relaxation turns out to be exact. Following Ben-Tal and den Hertog (2014), in Locatelli (2015b, 2016a) a simpler convex conic quadratic relaxation is considered and in Locatelli (2016a) it is shown that such relaxation is equivalent to the SDP one. Then, a condition for exactness of the relaxation more general than the dimension condition is derived from the KKT conditions of the convex conic quadratic relaxation. The paper Ho-Nguyen and Kilinç-Karzan (2017) considers a SOCP relaxation in the original space of variables for problems with quadratic objective function, a unit ball constraint and constraints $Ax - b \in \mathcal{K}$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\mathcal{K} \subset \mathbb{R}^m$ is a closed convex cone. The paper introduces conditions under which the relaxation is tight. Moreover, given the epigraph of the problem, with the additional variable t and the additional constraint $h(x) \leq t$, where h is the objective function, conditions are provided for the derivation of its convex hull. In Burer and Ye (2020) diagonal QCQPs are considered, where all the Hessian matrices of the quadratic functions are diagonal. In this case the work provides some conditions related to the feasibility of suitably defined polyhedral sets, which guarantee the existence of rank-one solutions of the SDP relaxations and, thus, exactness of the relaxation. The work Wang and Kilinç-Karzan (2021) presents an exactness condition for a class of QCQPs which includes the diagonal ones, and it is shown that the condition implies the result proved in Burer and Ye (2020) when applied to diagonal QCQPs. In Jeyakumar and Li (2018) a class of minimax QCQPs is addressed (the objective is the maximum of a finite set of quadratic functions). A SOCP reformulation of the Lagrangian dual is presented. Exactness of this SOCP problem is proved when the epigraphical set (the epigraph of all the quadratic functions involved in the objective and in the constraints) is closed and convex. The Celis–Dennis–Tapia (CDT) problem, where a quadratic function is minimized over the intersection of two ellipsoids, has been investigated in some works. For instance, Ai and Zhang (2009) gives a necessary and sufficient condition for the exactness of the Lagrangian relaxation for this problem. Different papers investigate the possibility of narrowing (or closing) the duality gap by adding SOC-RLT constraints Burer and Anstreicher (2013), by solving two subproblems with SOC constraints Yuan et al. (2017), by adding lifted RLT cuts Yang and Burer (2016), by adding KSOC cuts Anstreicher (2017) (we refer to Section 4.2.4 for a discussion of all these cuts). It is also worthwhile to mention the result in Yang et al. (2018) stating that the addition to QCQPs of further reverse convex constraints, imposing that feasible points cannot lie in the interior of non-intersecting ellipsoids, does not lead to more difficult problems. More precisely, the existence of a tight SDP relaxation for the problem without such constraints implies the existence of a tight SDP relaxation also for the problem with these constraints.

For some problems exact convex reformulations are not known but still the problem can be solved in polynomial time. The polynomial methods usually enumerate all the (polynomially bounded) KKT points for these problems. An example of such enumerative methods is proposed in Bienstock and Michalka (2014b) for problems with a quadratic objective function, some ball constraints, some reverse ball constraints (i.e., constraints which impose that points cannot lie in the interior of a ball), and some linear constraints. The method runs in polynomial time provided that the number of ball and reverse ball constraints is fixed, and that the number of faces of the polyhedron defined by the set of linear inequalities having a nonempty intersection with the set defined by the ball constraints is polynomially bounded. The papers Consolini and Locatelli (2017) and Sakaue et al. (2016) provide a polynomial-time method to solve a generalization of the CDT problem, where one of the two quadratic constraints is allowed to be nonconvex. The proposed method identifies all the KKT points of the problem by solving a bivariate polynomial system with polynomials of degree at most $2n$ and with the two unknowns corresponding to the Lagrange multipliers of the two quadratic constraints. The algorithm has a polynomial complexity but with a large exponent (six) with respect to the number n of variables. Interestingly, a rather different polynomial-time approach for this problem has been presented in Bienstock (2016). In this approach a sequence of feasibility problems for systems of quadratic inequalities is solved by a polynomial-time algorithm based on Barvinok's construction Barvinok (1993). The algorithm is hard to implement but the approach can also be extended to any fixed number of quadratic constraints, provided that one of them is strictly convex.

We conclude this section by mentioning problems for which a polynomial-time solution algorithm is not available but for which polynomial-time approximation schemes are available. In Bomze and de Klerk (2002) a Polynomial Time Approximation Scheme (PTAS) is introduced for the StQP problem. The PTAS is based on the evaluation of the objective function over a uniform grid. In de Klerk et al. (2006), the result has been extended to the minimization of polynomials of fixed degree d over the unit simplex. In Depetrini and Locatelli (2011) a Fully Polynomial Time Approximation Scheme (FPTAS) has been proposed for Linear Fractional-Multiplicative Programming problems, where sums or products of a fixed number of ratios of affine functions are minimized over polytopes. The FPTAS is based on the solution of LP problems over a nonuniform grid. The approach has also been extended to a more general class of problems in Locatelli (2013) and Mittel and Schulz (2013).

4.2. Exact methods for "difficult" GO problems

Branch-and-Cut (B&C) methods are most widely employed for the exact solution of GO problems. In what follows we briefly sketch how they work. A collection $\mathcal{C}$ of subsets of the feasible region is maintained throughout the algorithm. The collection is initialized with a single set corresponding to the whole feasible region. Then, during the execution of the algorithm the problem is subdivided into subproblems by *branching* operations. Each branching operation replaces a subset in $\mathcal{C}$ by other subsets which cover it. For each subproblem a *relaxation*, often a convex relaxation, is defined whose solution gives a *lower bound* for the subproblem. The relaxation can be strengthened through the introduction of *cutting planes*, which are guaranteed not to remove feasible or, at least, optimal solutions. A global *upper bound* is possibly updated each time a feasible point is detected, e.g., as a result of the solution of the convex relaxation. *Fathoming* rules are applied, and, in particular, all subproblems in the collection $\mathcal{C}$ whose lower bound is not lower than the current upper bound (possibly decreased by a tolerance value ϵ) are removed from the collection $\mathcal{C}$. The algorithm stops as soon as the collection $\mathcal{C}$ is empty. In some cases, a further collection $\mathcal{Q}$ is maintained. Such collection contains all subsets over which the lower bound is not higher than the current upper bound. Then, when the algorithm stops (i.e., when the collection $\mathcal{C}$ is empty), the subsets in the collection $\mathcal{Q}$ contain all feasible solutions whose objective function value differ from the optimal value by at most the tolerance value ϵ. The theoretical issue of the finiteness of B&C methods has been widely investigated in the past and for this we refer to GO textbooks like Horst and Pardalos (1995), Horst et al. (2001) and Horst and Tuy (1993). In what follows we discuss the most recent developments related to the main operations of B&C approaches emphasized above. Note that we are not going to discuss operations like upper bounding and fathoming since these are rather standard operations for which significant developments cannot be expected. For what concerns fathoming we only observe that, besides the standard rule based on the comparison between lower and upper bounds, in some cases it is possible to introduce fathoming rules based on optimality conditions. In particular, one can remove subsets from $\mathcal{C}$ for which it is possible to guarantee that they do not contain points fulfilling necessary optimality conditions.

4.2.1. Branching

Branching operations can be subdivided into two broad classes, *spatial branching* and *KKT branching*. Spatial branching can be applied to generic GO problems. The feasible region of the GO problem is initially enclosed into a region with a simple geometrical form. Then, branching is performed by subdividing this region into smaller regions, often, but not necessarily, with the same geometrical form. The intersections of these smaller regions with the original feasible set give rise to the subsets entering the collection $\mathcal{C}$. The most common geometrical form is a box, since lower and upper bounds for the variables are often already part of the GO problem description or, alternatively, can be easily computed by solving auxiliary problems. But other geometrical forms have been adopted in the literature. For instance, simplices and polyhedral cones have been often employed, see again the textbooks (Horst and Pardalos, 1995; Horst et al., 2001; Horst and Tuy, 1993). Note that in all these cases (boxes, simplices, polyhedral cones) the branching operation subdivides a subset into smaller subsets, whose interiors do not overlap, but which can share some common face. Some papers (Cartis et al., 2015; de Angelis et al., 2004; Fowkes et al., 2013; Hager and Phan, 2009; Le Thi, 2000) employ ellipsoids, exploiting the fact that quadratic functions and also some cubic functions can be efficiently minimized over these sets. In this case the subsets generated by a branching operation may have overlapping interiors. We also mention Tóth et al. (2016), where a subdivision into *regular* simplices is proposed, which also leads to subsets with overlapping interiors.

For some specific problems further geometrical forms have been considered. For QCQP problems, Linderoth (2005) employs cartesian products of rectangles and right-angled triangles. For linear sum-of-ratios problems in Kuno (2005) cartesian products of trapezoids are used, while for the same problem Locatelli (2015a) employs cartesian products of rectangles and right-angled triangles as in Linderoth (2005). The use of these geometrical forms is strictly related to the development of tight under- and over-estimators for bilinear and bivariate fractional functions over the two-dimensional regions appearing in the cartesian products, i.e., rectangles, right-angled triangles and trapezoids.

The subdivision can be performed either in a problem-independent or in a problem-dependent way. In the former a subregion is subdivided into subregions of equal size. The most common problem-independent subdivision is bisection, where the subregion, say a box, is split into two subregions through a subdivision of the longest edge performed at its midpoint (for ellipsoids the longest edge is replaced by the longest axis). Instead, problem-dependent subdivisions take the solution of the convex relaxation into account. If the objective function is underestimated by a convex function, the subdivision is performed in such a way that the error at the optimal solution of the convex relaxation, i.e., the difference between the objective function and the convex underestimating function evaluated at such optimal solution, is reduced as much as possible in the newly generated subregions. Note that it is important to reduce the error at the optimal solution of the convex relaxation, since the value of the convex underestimating function at this point is equal to the lower bound computed over the subregion. Then, reducing the error usually means improving (increasing) the lower bound of the newly generated subregions with respect to the lower bound over the original subregion.

KKT branching has a more limited applicability with respect to spatial branching. It is employed for QP problems with linear constraints. In particular, for Box QP problems (problems with a quadratic objective function and box constraints) in Vandenbussche and Nemhauser (2005a,b) it is observed that these problems can be reformulated by replacing the original feasible region with the one defined by the KKT conditions. This requires the addition of the variables corresponding to the Lagrange multipliers of the box constraints, while the original quadratic objective function can be replaced by a linear one, involving also the additional variables. The only nonconvex constraints of the reformulation are those corresponding to the complementarity conditions. These are initially omitted, while KKT branching first selects a constraint according to some rule (usually the one with the largest violation of the complementarity condition at the solution of the convex relaxation), and then splits the current subregion into two new subregions by imposing that in one subregion the constraint is active, while in the other the corresponding Lagrange multiplier is set equal to 0. The work Burer and Vandenbussche (2008) extends the approach to feasible regions which are general polytopes. While the idea is the same, the simple approach of omitting the complementarity conditions in order to get a convex (linear) relaxation does not work in the general case since it leads to a trivial bound. Therefore, in Burer and Vandenbussche (2008) an SDP relaxation is introduced. A different SDP relaxation has been proposed in Burer and Vandenbussche (2009) for Box QP problems with further developments in Chen and Burer (2012), which reports very good computational results for problems with a dense Hessian matrix. KKT branching is also employed in Audet et al. (1999) for a special QP case, the case of disjoint bilinear programming. It is also worthwhile to point out here that the reformulation of QP problems based on KKT conditions has also been used to convert the solution of these problems into the solution of Mixed Integer Linear Programming (MILP) problems, where additional binary variables are used to establish whether a constraint is active or, alternatively, its Lagrange multiplier is equal to zero (see, e.g., Xia et al. (2020) where good results, especially for StQP problems, are reported).

4.2.2. Convex underestimating functions

Convex underestimating functions or, analogously, concave overestimating functions, are essential for the definition of convex relaxations. The ability of detecting them and evaluating their tightness, i.e., how

close they are to the original function, strictly depends on the properties of the original function and on the region over which the underestimation takes place. For some functions and regions of relatively simple form it is possible to compute the tightest convex underestimator, also known as convex envelope. More formally, given a function f and a region X, the convex envelope of f over X is defined as follows for each $x \in X$:

$$conv_{f,X}(x) = \sup\{c(x) : f(y) \geq c(y),\ \forall y \in X, \quad c \text{ is convex}\},$$

i.e., $conv_{f,X}$ is the largest convex underestimator of f over X (note that in the above definition 'c is convex' can be replaced with 'c is affine'). Analogously, we can define concave envelopes, i.e., tightest concave overestimators. In what follows we will always refer to convex envelopes, since the concave envelope of f over X can easily be seen to be the opposite of the convex envelope of $-f$ over X. Besides some results for specific functions, the best known of which is probably McCormick convex envelope for bilinear terms over rectangles, some general results appeared in the literature. The first general results were mostly related to so called polyhedral convex envelopes. A function f admits a polyhedral convex envelope over a region X if its convex envelope is the maximum of a finite number of affine functions. In particular, if X is a polytope, a function f is said to admit a *vertex* polyhedral convex envelope if its convex envelope is equal to the convex envelope of the same function over the vertex set of X. In fact, according to Rikun (1997) polyhedral convex envelopes are always vertex polyhedral convex envelopes for continuously differentiable functions f. The condition of edge-concavity (concavity along each segment parallel to an edge of X) guarantees the existence of a vertex polyhedral convex envelope (Meyer and Floudas, 2005a; Tardella, 2003; 2008). For instance, multilinear functions always admit vertex polyhedral convex envelopes over boxes. In Meyer and Floudas (2005a) it is shown that these envelopes are strictly related to triangulations of the polytope X with a number of simplices which can be very large (e.g., of size $n!$ for n-dimensional boxes). Identification of the triangulation may be a rather difficult task, except for some special cases. In particular, when X is the unit box and f is submodular over the vertex set of X, the convex envelope is the Lovász extension of f (see Tawarmalani et al., 2013). Recently, different results about non-polyhedral convex envelopes have been presented. Most of them require that X is a box, while f is required to satisfy different assumptions. In Tawarmalani and Sahinidis (2001) f is assumed to be convex if we fix the value of one variable and concave if the remaining $n-1$ variables are fixed. In Jach et al. (2008) the Hessian of f is assumed to have at least one negative eigenvalue over the box and f is convex if the value of one variable is fixed. In Khajavirad and Sahinidis (2012, 2013) f is a product function $h(\mathbf{z})g(\mathbf{y})$, where h and g must fulfill some conditions (e.g., h must be nonnegative and convex and with some specific form, g must be nonnegative and component-wise concave). In Ballerstein and Michaels (2014) and Locatelli (2016b) f is required to be (strictly) convex if we fix the values of $n-1$ variables, while if we fix the value of the remaining variable, the minimum of the function is attained at a vertex of the corresponding $(n-1)$-dimensional box. In all the above works, the required properties for f imply that the convex envelope of f over the box is equivalent to the convex envelope of the same function over the border of the box. In Locatelli (2020) a case where this does not hold has been discussed, namely the convex envelope of bivariate cubic functions over rectangles.

The convex (and concave) envelopes of some functions of relatively simple form (usually, univariate or bivariate functions) can also be employed to underestimate and/or overestimate functions with a more complicated form. This is the case for factorable functions, i.e., functions which can be progressively decomposed into the sum and product of simple univariate and bivariate functions, for which convex and concave envelopes are available (see, e.g., Khajavirad et al., 2014; Scott et al., 2011; Tawarmalani and Sahinidis, 2004). For instance, given the factorable function $e^{x_1x_2}x_3$, we first introduce the new variables $x_4 = x_1x_2$, $x_5 = e^{x_4}$, and $x_6 = x_5x_3$, and then these equality constraints are replaced by inequalities where the left-hand side is imposed to be not lower than the convex envelope (over a suitable rectangle or interval) of the right-hand side, and not larger than its concave envelope. A simple example of factorable function is the sum of bilinear terms $\sum_{i,j=1,\ i\neq j}^{n} A_{ij}x_ix_j$. A possible convex underestimator (concave overestimator) for this function over the unit hypercube can be obtained by summing McCormick's convex envelope (McCormick's concave envelope) of each single bilinear term over the unit square. Unfortunately, the convex underestimator is *not* the convex envelope of the overall sum (in general, the convex envelope of a sum is not the sum of the convex envelopes of the single terms of the sum). Similarly, for the concave overestimator. However, there are some results which bound the difference between the concave overestimator and the convex underestimator. In particular, in Luedtke et al. (2012) it is shown that such difference cannot be larger than the difference between the concave and the convex envelope over the unit hypercube times a $O(n)$ constant, while in Boland et al. (2017) the order of magnitude of the constant has been refined to $O(\sqrt{n})$.

For a poorly structured function f the best known approach for underestimating f is the one proposed within the framework of the α-BB approach (see Adjiman et al. (1998a, 1996, 1998b)). In this case, f is only assumed to be twice-continuously differentiable and a nonnegative function q is introduced such that the Hessian of $f-q$ is semidefinite positive over a box X or, stated in another way, $f-q$ is a convex underestimator of f over X. In the original α-BB approach q was defined as

$$q(x) = \sum_{i=1}^{n} \alpha_i(x_i - \ell_i)(u_i - x_i),$$

where ℓ_i, u_i are, respectively, the lower and upper bound for variable x_i within the box X, while the values α_i are suitably chosen in order to guarantee convexity of $f-q$ over X. Later on, different functions q have been proposed in the literature, like, e.g., spline functions Meyer and Floudas (2005b), obtained by first subdividing the box X into smaller sub-boxes and then computing different (and sharper) α values over these sub-boxes, or exponential functions (Akrotirianakis and Floudas, 2004). In the recent work Kazazakis and Adjiman (2018) a further variant has been proposed, where the original function is replaced by a so called μ-subenergy function, with the property that its eigenvalues have a smaller magnitude with respect to those of the original function in the regions far away from global minimizers, thus allowing to choose tighter α values in those regions.

For poorly structured GO problems we should also mention methods based on interval arithmetic. These are rigorous methods, i.e., they guarantee the detection of a solution within a prescribed tolerance even in the presence of rounding errors. The overall number of contributions in this field in the last two decades is relatively limited, but there are some interesting works. We first mention the two surveys (Araya and Reyes, 2016; Neumaier, 2004), to which we refer for a more detailed discussion about the topic. In some works first and second order information are employed both to define linear and quadratic underestimating functions and to fathom (or shrink) boxes which are guaranteed not to include points satisfying first and second order necessary optimality conditions (see, e.g., Hansen et al., 2007; Markot and Schichl, 2014; Martinez et al., 2004). Note that the original α-BB approach exploits second order information to derive a convex quadratic underestimating function. In Borradaile and Van Hentenryck (2005), Kearfott (2011) and Neumaier and Shcherbina (2004) safe linear underestimators are discussed. These are linear relaxations which take into account numerical errors in the computation of the coefficients of the linear functions. The paper Berenguel et al. (2013) presents ways to take into account separability of the objective function. In Carrizosa et al. (2004) it is shown that simple translations of the variables allow to derive better inclusion functions through interval arithmetic. A discussion about different branching rules for the subdivision of a box into sub-boxes can be found in Csallner et al. (2000) and Markót et al. (2006). In

Jansson et al. (2007) interval arithmetic is applied to compute rigorous error bounds for the optimal value of semidefinite programs. In Domes and Neumaier (2016) it is shown that the information extracted from local optimization, namely the approximated local optimizer as well as the corresponding Lagrange multipliers, can be exploited to form an aggregated redundant constraint which turns out to be powerful in order to tight the bounds on the variables (see also the following Section 4.2.5). The authors also remark that such aggregate constraints are able to reduce the so called *cluster effect*, i.e., the presence of a large cluster of unfathomed small boxes in the regions around the global minimizer, which is a serious obstacle to the efficiency of B&C approaches. We also mention the successful application of an interval method to the circle packing problem (see Markót and Csendes, 2005).

We conclude this section by observing that up to now we have talked about *convex* underestimators. However, as discussed in Section 4.1, there are some nonconvex problems which can be solved efficiently. So, in some cases nonconvex underestimators can be employed. For instance, in Cartis et al. (2015) and Fowkes et al. (2013) the objective function is underestimated by quadratic and cubic functions, whose minimizers can be efficiently computed over spheres, which are the regions generated by the branching operation adopted in those works. If the branching operation generates polyhedral sets with a limited number of vertices, then concave underestimators can be employed. Indeed, the minimum value of a concave function over a polytope is attained at a vertex of the polytope. This is done, e.g., in the context of Lipschitz optimization (see, for instance, Hendrix and Tóth, 2010; Paulavičius and Žilinskas, 2014), where a concave lower bounding function based on the Lipschitz condition is minimized over a simple polyhedral region (in particular, a simplex).

4.2.3. Problem reformulations and convex relaxations

For some nonconvex problems a convex relaxation can be derived by first reformulating the problem, usually with the addition of new variables, and then removing or relaxing the nonconvex constraints of the reformulation. A typical example is the well known Shor relaxation for QCQPs, where the additional matrix variable X and the equality constraint $X = xx^\top$ are introduced. This way, each quadratic form $x^\top Qx$ can be replaced by the linear term $Q \bullet X = \sum_{i,j=1}^{n} Q_{ij} X_{ij}$. Nonconvexity only lies in the rank-one constraint $X = xx^\top$. A convex relaxation is obtained by replacing the equality constraint with the semidefinite constraint $X \succeq xx^T$, i.e., it is required that the matrix $X - xx^\top$ is positive semidefinite. The resulting SDP bound has a good quality but is costly. Conversely, SOCP relaxations, obtained by replacing quadratic functions with convex quadratic underestimating functions, have a poorer quality but are faster to compute. For this reason in Burer et al. (2014) it is proposed to construct mixed SOCP-SDP relaxations which allow for a balance between the quality of the bound and the time needed to compute it.

The work Burer (2009) introduces an exact reformulation for QP problems with quadratic objective function and linear constraints, possibly with some binary variables. The reformulation is based on the convex cone of $n \times n$ completely positive matrices

$$C_n^* = \{X \in \mathbb{R}^{n \times n} \ : \ X = \sum_{k \in K} x_k x_k^\top, \ |K| \text{ has finite cardinality}, \ x_k \in \mathbb{R}^n_+, \ \forall k \in K\},$$

and its dual cone, the dual cone of copositive matrices:

$$C_n = \{X \in \mathbb{R}^{n \times n} \ : \ x^\top X x \geq 0 \ \forall x \in \mathbb{R}^n_+\}.$$

In Burer (2009) it is proved that, under mild assumptions, any QP problem with linear constraints and, possibly, some binary variables, can be reformulated as a problem over the convex cone of completely positive matrices (we refer, e.g., to the survey paper Bomze et al. (2012) for a discussion about problems over this cone and over the dual cone of copositive matrices). This result has some predecessors and different successors. The predecessors are copositive reformulations for some subclasses of QP problems. In particular, the first of such results is a copositive reformulation of StQP problems presented in Bomze et al. (2000). In the successors of Burer (2009), completely positive and copositive reformulations have been extended to other classes of optimization problems. In fact, in Burer (2009) itself the result is extended also to cases where some quadratic equality constraints appear. In Burer and Dong (2012) it is shown that QCQPs can be reformulated as *generalized* copositive programs, i.e., problems defined over the generalized completely positive cone, where the requirement $x_k \in \mathbb{R}^n_+, \ \forall k \in K$, is replaced by $x_k \in \mathcal{K}, \ \forall k \in K$, where $\mathcal{K}$ is a convex cone. The paper Amaral et al. (2014) reformulates the problems of minimizing the ratio of two quadratic functions over a polyhedral region as problems over the cones of completely positive matrices and of copositive matrices. In Bai et al. (2016) it is shown that, under suitable assumptions, QCQP problems whose variables are constrained to belong to $\mathcal{K} \cap \mathcal{M}$, where $\mathcal{K}$ is a convex cone and $\mathcal{M}$ is a region defined by linear equality constraints, can be reformulated as problems over the cone of completely positive matrices. In Bomze et al. (2017) it is shown that, besides the one presented in Burer (2009), other equivalent completely positive reformulations for mixed-binary QPs are possible. The interest of these equivalent reformulations lies in the fact that, once the completely positive cone is relaxed into a tractable one, the new reformulations may lead to tighter bounds. In Bomze et al. (2018) it is shown that for QP problems with quadratic objective function, two quadratic constraints and some further linear constraints, under suitable assumptions a copositive reformulation exists. For some problems, a completely positive reformulation is not available but still they admit completely positive relaxations. In Bomze (2015) it is proved that for QCQPs bounds returned by completely positive relaxations dominate Lagrangian dual bounds. This result has been extended to polynomial programming problems in Kuang and Zuluaga (2018) after the introduction of completely positive tensors.

Unfortunately, though convex, the cones of copositive and completely positive matrices are not tractable, i.e., problems over these cones cannot be solved in polynomial time by interior point methods because the computation of their self-concordant barrier functions cannot be performed in polynomial time. All the same, reformulations over the completely positive or copositive cone allow to define polynomially solvable convex relaxations by replacing these cones with tractable cones. The cone of nonnegative matrices and the cone of semidefinite matrices are tractable and they outer approximate C_n^* and, being self-dual, they inner approximate C_n. The intersection of these two cones is the cone of doubly nonnegative matrices, which inner approximates C_n (in fact, equality holds for $n < 5$), while its dual cone, i.e., the cone made up by matrices which can be obtained by summing a nonnegative and a semidefinite matrix, outer approximates C_n^*. Different hierarchies of cones $\{\mathcal{K}_n^r\}$ have been proposed in the literature such that $\forall r \ : \ \mathcal{K}_n^r \supset \mathcal{K}_n^{r+1} \supset C_n^*$, i.e., the cones are finer and finer outer approximations of the completely positive cone and, conversely, their dual cones are finer and finer inner approximations of C_n (see de Klerk and Pasechnik, 2002; Parrillo, 2000; Peña et al., 2007). Although the bounds based on these hierarchies tend to become sharper as r increases, the main drawback is that the size of these problems tends to increase rapidly with r, since they involve $n^{r+1} \times n^{r+1}$ matrices or a comparable number of $n \times n$ matrices. Thus, from the computational point of view only the use of small values for r is feasible. In Bundfuss and Duer (2009), after observing that a matrix A is copositive if and only if $x^\top Ax \geq 0$ for all $x \in \Delta_n$, where Δ_n is the n-dimensional unit simplex (see also Tóth et al., 2021), polyhedral inner and outer approximations of the copositive cone are proposed, based on simplicial subdivisions of the unit simplex. It is shown that, under suitable assumptions, bounds computed by replacing the copositive cone with these approximations tend to converge to the optimal value of the copositive problem, provided that the length of the largest edge of the simplices in the partition converges to 0.

In order to strengthen the bound obtained by replacing C_n^* with a tractable cone, we may also proceed as follows. Let $X^\star$ be the optimal solution obtained by solving the relaxation over the tractable cone $\mathcal{K} \supset C_n^*$. Then, either $X^\star \in C_n^*$ or $X^\star \in \mathcal{K} \setminus C_n^*$. In the former case $X^\star$ is also an optimal solution of the completely positive problem. Otherwise, by definition of dual cone, there exists some $C \in C_n$ such that $C \bullet X^\star < 0$. Thus, adding the inequality $C \bullet X \geq 0$ to the previous relaxation, we are able to strengthen the previously computed bound. In other words, a separation problem is solved. Such an approach has been explored in Burer and Dong (2013) and also in Bomze et al. (2010) for what concerns the completely positive reformulation of the max-clique problem. Note that this topic would also fit into the following Section 4.2.4 about cutting planes.

Reformulations and relaxations have also been proposed for polynomial programming problems. The unconstrained minimization of a polynomial f with degree d can easily be reformulated as:

$$\sup \quad \alpha$$
$$f(x) - \alpha \in \mathcal{Q}_{n,d},$$

where $\mathcal{Q}_{n,d}$ is the set of nonnegative polynomials of degree at most d. However, the set of nonnegative polynomials is not easily representable. Therefore, a tractable relaxation can be obtained by replacing it with the set of Sum-Of-Squares (SOS) polynomials of degree at most d, i.e., polynomials which can be written as a sum of a finite number of square of polynomials of degree at most $\frac{d}{2}$. The nice feature of SOS polynomials is that they are representable by a positive semidefinite condition imposed over matrices with dimension $O\left(n^{d/2}\right)$. In fact, Papp and Yildiz (2019) discusses an alternative and cheaper way to represent SOS polynomials. The bound based on the SOS relaxation (and its dual counterpart, the moment relaxation, see Lasserre, 2001) has been strengthened with the definition of hierarchies of bounds both for the unconstrained and the (polynomially) constrained case, which have been discussed in different papers such as, e.g., de Klerk et al. (2017a,b), Lasserre (2001, 2005, 2006), Laurent (2007), Nie (2013), Nie (2014), Nie et al. (2006) and Vui and So'n (2008).

4.2.4. Outer approximation and cutting planes

When the feasible region S is not convex, we need to outer approximate it with a convex region in order to derive a convex relaxation. The tightest convex outer approximation of S is called *convex hull* of S and is denoted by $chull(S)$. One possibility to derive a convex outer approximation is to replace each constraint $g(x) \leq 0$ defining S, g nonconvex, with a constraint $c(x) \leq 0$, where c is a convex underestimator of g over S, so that all the material discussed in Section 4.2.2 could be applied here. The tighter the convex underestimator, the tighter is the convex outer approximation. However, replacing each nonconvex function g with its convex envelope over a suitable region containing S leads to a convex outer approximation, but this is not necessarily the convex hull. A simple example is the following. Given the region $S = \{(x, y) \in [1, 2] \ : \ xy \geq 3\}$, according to the previous discussion the nonconvex function xy should be replaced by its concave envelope over $[1, 2]^2$, i.e., by $\min\{x + 2y - 2, 2x + y - 2\}$. But the resulting convex region $\{(x, y) \in [1, 2] \ : \ x + 2y - 2 \geq 3, \ 2x + y - 2 \geq 3\}$ is not the convex hull of S, since S is already a convex set (just note that $xy \geq 3$ can be rewritten as $x \geq 3/y$). More generally, in Anstreicher (2012) for QCQPs, with additional linear constraints defining a polytope P, it is shown that replacing the quadratic terms with their convex envelopes over P leads to a convex relaxation which is dominated by the one obtained through the convex hull of the set $\{(1 \ x)(1 \ x)^\top, \ x \in P\}$ (note, however, that both the convex envelope and the convex hull may be hard to compute).

Many results about convex hulls have been presented in the recent literature. The work Vandenbussche and Nemhauser (2005b), within the framework of a branch-and-bound approach for Box QP problems based on KKT branching (see Section 4.2.1), considers convex hulls of the regions defined by the KKT conditions related to each single variable x_i of the problem. Again for Box QP, in Burer and Letchford (2009) first the objective function is linearized with the addition of the variables X_{ij} and the related constraints $X_{ij} = x_i x_j$, and then the convex hull of the set

$$\left\{(x, X) \in [0, 1]^{n+n(n+1)/2} \ : \ X_{ij} = x_i x_j, \ 1 \leq i \leq j \leq n\right\},$$

is analyzed. Many facet-inducing inequalities for this set are derived from valid inequalities for the Boolean quadric polytope (see Padberg (1989)):

$$\left\{(x, X) \in \{0, 1\}^{n+n(n-1)/2} \ : \ X_{ij} = x_i x_j, \ 1 \leq i < j \leq n\right\}.$$

For QP problems or, more generally, for polynomial programming problems, a convex relaxation can be strengthened by the Reformulation-Linearization Technique (RLT), introduced in Sherali and Tuncbilek (1992). If the linear constraints $a_1^\top x \geq b_1$ and $a_2^\top x \geq b_2$ are present, then first the additional quadratic constraint $(a_1^\top x - b_1)(a_2^\top x - b_2) \geq 0$, implied by the two linear constraints, is added, and then this constraint is linearized by replacing the terms $x_i x_j$ with the additional variables X_{ij}. A typical example are the RLT constraints obtained by multiplying the box constraints $0 \leq x_i, x_j \leq 1$:

$$X_{ij} \geq 0, \ X_{ij} \geq x_i + x_j - 1, \ X_{ij} \leq x_i, \ X_{ij} \leq x_j,$$

which also correspond to the classical McCormick envelopes. In Anstreicher (2009) it is shown that for Box QPs and some QCQPs, including both the semidefinite condition on the variable matrix X and RLT constraints, leads to significatively better bounds than using the semidefinite condition or the RLT constraints alone. The paper Anstreicher and Burer (2010) derives convex hulls for quadratic forms over small-dimensional regions (triangles and boxes) based on semidefinite and nonnegative conditions over matrices and, possibly, additional RLT constraints. In Bienstock and Michalka (2014a) cutting planes are defined for the characterization of $chull(S)$ when

$$S = \{(x, q) \in \mathbb{R}^d \times R \ : \ q \geq Q(x), \ x \in \mathbb{R}^n \setminus int(P)\},$$

where Q is convex and differentiable and $int(P)$ denotes the interior of set P. A polynomial separation algorithm is proposed for the case when Q is quadratic and strictly convex, while P is a polyhedron or an ellipsoid. The work Burer and Kilinç-Karzan (2017) derives convex relaxations and, under suitable assumptions, even convex hulls for the intersection of the following sets:

- a SOC representable cone $\mathcal{K}$, i.e., given a matrix $B \in \mathbb{R}^{n \times (n-1)}$ and $b \in \mathbb{R}^n$, $\mathcal{K} = \{x \ : \ \|B^\top x\| \leq b^\top x\}$;
- a cone $\mathcal{Q}$ defined by a homogeneous quadratic function, i.e., $\mathcal{Q} = \{x \ : \ x^\top A x \leq 0\}$, for some matrix $A \in \mathbb{R}^{n \times n}$;
- an affine hyperplane.

Based on the observation that $A \succeq O$ and $B \succeq O$ implies $A \otimes B \succeq O$, where $\otimes$ denotes the Kronecker product, in Anstreicher (2017) so called Kronecker product constraints are introduced. Let $H(x) \succeq O$ and $G(x) \succeq O$ be semidefinite constraints where single components of both matrices are affine functions of x. Then, a Kronecker product constraint is obtained by replacing in $H(x) \otimes G(x) \succeq O$ each term $x_i x_j$ with X_{ij}. These constraints are generalization both of the classical RLT constraints obtained from two linear inequality constraints, and of the SOC-RLT constraints obtained from one linear inequality constraint and a SOC constraint. In Wang and Kilinç-Karzan (2021) for some QCQPs, conditions are given under which the convex hull of the epigraph of the QCQP is the projection of the epigraph of its Shor relaxation. In Del Pia and Khajavirad (2017, 2018) in the context of binary polynomial problems, i.e., problems with a polynomial objective function, some binary variables and some other variables constrained to belong to the interval [0,1], valid inequalities are derived to characterize the convex hull of multilinear sets:

$$S = \left\{(x, y) \ : \ y_I = \prod_{i \in I} x_i, \ I \in \mathcal{I}, \ x \in \{0, 1\}^n\right\},$$

where $\mathcal{I}$ is a collection of subsets of $\{1, \ldots, n\}$ with cardinality not lower than two.

The paper (Nguyen et al., 2018) presents the (non polyhedral) convex hulls for the intersection of cubes with regions $\{(x, y, z) : x^{b_1} y^{b_2} \geq z\}$ and with regions $\{(x, y, z) : xy^{b_2} \leq z\}$ for $b_1, b_2 \geq 1$. In Davarnia et al. (2017) a description is given for the convex hull of sets

$$S = \{(x, y, z) : x \in [0, 1]^n,\ Ex \geq f,\ y \in \mathbb{R}^m,\ y \geq 0,\ e_m^\top y \leq 1,$$
$$z_k = y^\top A_k x, k = 1, \ldots, K\},$$

where $E \in \mathbb{R}^{t \times n}$, $f \in \mathbb{R}^t$, e_m is the m-dimensional vector with all components equal to one, $A_k \in \mathbb{R}^{m \times n}$. Then, x belongs to a polyhedral subset of the unit box, y belongs to the m-dimensional unit simplex, while z is obtained by bilinear terms involving x and y. An application to network interdiction problems is presented. In Bonami et al. (2019) some cutting planes to strengthen linear relaxations of QPs with linear constraints are proposed. The cutting planes are based on a well known result by Motzkin and Straus (see Motzkin and Strauss (1965)) about max clique problems. For this reason, they are called Motzkin-Straus clique inequalities. The work Santana and Dey (2020) establishes that the convex hull of the intersection of the region defined by a quadratic equality constraint and a polytope is SOCP representable. In Bienstock et al. (2020) valid inequalities are introduced for sets $S \cap P$, where S is a closed set and P is a polyhedron. A cutting plane algorithm is proposed. Given an oracle returning the distance of some point from S, the algorithm generates cutting planes which are able to approximate $chull(S \cap P)$ in an arbitrarily precise way. An application to polynomial programming is presented. In Luedtke et al. (2020) non polyhedral convex hulls are given for subsets of $\mathbb{R}^5$, arising in pooling problems and defined by some linear constraints and a nonconvex bilinear constraint.

4.2.5. Bound tightening

Rather special cutting planes are those involving lower and upper bounds on single variables. Strengthening these bounds is also known as *bound tightening*. Given a lower and upper bound ℓ and u for variables x, a bound tightening procedure can be viewed as a function receiving the box $[\ell, u]$ in input and returning a box $[\ell', u'] \subseteq [\ell, u]$, such that $[\ell, u] \setminus [\ell', u']$ does not contain any feasible point (feasibility-based bound tightening) or, more powerfully, does not contain any optimal solution (optimality-based bound tightening). The importance of bound-tightening is due to the fact that it does not only reduce the feasible region of a convex relaxation, but it also improves the quality of (convex) underestimating functions for the objective and constraint functions. Indeed, in many cases the underestimating functions depend on the lower and upper limits for the variables and they tend to become tighter as these limits become tighter. The importance of bound-tightening for an efficient solution of nonconvex problems is proved by the fact that many GO solvers (such as `BARON` Sahinidis (2017), `BMIBNB` Lofberg (2004), `COUENNE` Belotti et al. (2006), `SCIP` Gamrath et al. (2020)) include procedures to perform it.

A simple way to perform bound tightening is by minimizing and maximizing each variable over a convex relaxation of the feasible region (feasibility-based bound tightening), or over the same relaxation with an additional constraint imposing that a convex underestimating function for the objective function is not larger than the current known upper bound (optimality-based bound tightening, since the additional constraint may remove feasible points but no optimal solution). Since, as previously commented, convex underestimating functions usually depend on variable bounds, once all bounds have been tightened, the new bounds improve the quality of the underestimating functions so that a further round of bound-tightening may allow to further reduce the bounds. This can be iteratively repeated until convergence. Such iterative procedure has been theoretically and computationally investigated in Caprara and Locatelli (2010) and Caprara et al. (2016). Note, however, that such iterative procedure, while providing tight bounds, is quite expensive, requiring the solution of many convex subproblems. In Gleixner et al. (2017) three techniques are introduced to keep the computational cost of these bound tightening techniques under control. The work Tawarmalani and Sahinidis (2004) presents a general theoretical framework for bound-tightening techniques and discusses ways to exploit dual solutions of convex subproblems in order to perform the tightening. Constraint propagation techniques are also widely employed to tighten bounds. For factorable functions, first a directed acyclic graph (DAG) is built, where the nodes correspond to the variables (both the original variables and the additional variables introduced to model factorable functions), while the arcs represent the dependencies between the variables. Then, bound tightening is performed by forward and backward propagation along this graph. This technique is discussed in different papers such as Belotti et al. (2009), Messine (2004), Schichl and Neumaier (2005), Vu et al. (2009) and Wechsung et al. (2015). In Puranik and Sahinidis (2017) it is observed that, in order to perform optimality-based bound tightening, besides the previously mentioned additional constraint involving a convex underestimating function of the objective, one could also add constraints imposing necessary optimality conditions or, more precisely, a convex relaxation of such conditions. This is also done in Zhang et al. (2020) where some topics are addressed such as how to bound the dual variables appearing in the optimality conditions, for which explicit bounds are not given.

5. Computational aspects, test problems and solvers

Most of the papers cited in this work report computational experiments on different sets of test problems and with different solvers. Discussing in detail computational experiences, test functions and solvers is beyond the scope of the current work. However, we make a few observations and, for the interested reader, we provide pointers to papers and web sites where these aspects are presented in more detail.

5.1. Computational aspects

Here we briefly discuss a couple of observations which are well known but should always be taken into account when evaluating and comparing different solution approaches.

The first observation is that we should never search for the 'best' approach to solve GO problems. This is obviously true for the whole class of GO problems, since such wide class encompasses problems with fairly different properties and characteristics, which lead to fairly different approaches for their solution. But it is also true for much narrower subclasses of GO problems. In particular, we mention here a subclass of problems which attracted a lot of attention in the recent literature, namely the class of nonconvex QP problems with linear constraints. Well known commercial software products like `CPLEX` and `GUROBI` have recently introduced solvers for the solution of problems within this class. These QP problems can be tackled in many different ways. As already mentioned at the end of Section 4.2.1, in Xia et al. (2020) the problem is reformulated as a Mixed Integer Linear Program (MILP) after reformulating it with the inclusion of the KKT conditions. Binary variables are included to model the nonconvex complementarity conditions. In Chen and Burer (2012) the problem is still reformulated with the inclusion of KKT conditions and KKT branching is applied, but semidefinite relaxations are considered. In Bonami et al. (2018) a spatial branch-and-cut approach is proposed with the addition of valid cuts for the Boolean Quadric Polytope. In Liuzzi et al. (2021) another spatial branch-and-bound approach with intensive bound-tightening has been applied to a class of QP problems arising from an application in game theory.

The outcome of the computational experiments reported in all these works is that none of the proposed methods strictly dominates the others. The method proposed in Xia et al. (2020), as well as problem specific ones proposed in Gondzio and Yildrim (2021), Liuzzi et al. (2019), are the best performing over the subclass of StQP problems. The approach proposed in Bonami et al. (2018) performs quite well over the subclass of Box QP problems and techniques proposed in that paper have

been successfully incorporated in CPLEX. However, for Box QP problems with dense Hessian matrices the best approach appears to be the one proposed in Chen and Burer (2012). Finally, the approach proposed in Liuzzi et al. (2021) is the best performing over the subclass of QP problems presented in that work (but ongoing experiments prove its effectiveness over more general QP problems with linear constraints).

The second observation we should always keep in mind is that whenever we deal with a class of special structured GO problems, incorporating as much as possible the special structure into a solution approach allows for significant improvements. That holds true both for exact and for heuristic methods. We just mention a couple of examples taken from our own personal experiences, but many other examples could be given.

Circle packing problems are special structured QCQPs with many nonconvex quadratic constraints, namely the nonoverlapping constraints. The best exact methods for this problem (see, e.g., Markót and Csendes (2005) for circle packing into a unit square) are those which incorporate tools based on special properties of the circle packing problem, such as symmetry-breaking tools. Special purpose methods strongly outperform general purpose methods for QCQPs when applied to circle packing problems.

Molecular conformation problems lead to challenging GO problems, with a number of local minimizers which grows exponentially with the number of atoms. Exact methods can be applied only at very small dimensions (i.e., with a small number of atoms). But many excellent heuristic approaches exist. Also in this case the best performing approaches strongly rely on special properties (in particular, geometrical properties) of the molecular conformation problems (see, e.g., Ferreiro-Ferreiro et al., 2019; Wales, 2018; Zhao et al., 2017).

5.2. Test problems

Test problems for GO have been proposed in different papers, like, e.g., Furini et al. (2019) for QP problems. We also recall the book Floudas et al. (1999). Currently, there are different web sites which provide large sets of GO test problems. In Neumaier (2021) many academic as well as real-life GO test problems are reported. Large collections of test functions have been reported in Gavana (2021); this quite recent web site reports results of some GO algorithms over the presented test functions and, based on these results, a classification of the difficulty of the test functions is proposed. Through different editions of the GECCO Workshop on Real-Parameter Black-Box Optimization Benchmarking (BBOB), a collection of test problems has been collected. Details can be found in Auger et al. (2019). Another conference, the IEEE Congress on Evolutionary Computation (CEC), organised different competitions on Large Scale Global Optimization, providing many test functions (see, e.g., Škvorc et al. (2019) and the web site IEEE TfLsgo (2021)).

5.3. Solvers and their comparison

In previous sections we mentioned some GO solvers such as BARON, BMIBNB, COUENNE, SCIP and for nonconvex QPs with linear constraints also CPLEX and GUROBI. But others are available (see, e.g., Neumaier (2021) for quite an extensive list). The work Biscani and Izzo (2020) describes pagmo and pygmo, C++/Python libraries for massively parallel global, possibly multi-objective, optimization. The web site Johnson (2021) makes available NLopt, another quite large set of software tools for nonlinear and global optimization. Even SciPy Jones et al. (2001–) includes a set of implemented general purpose GO algorithms. Most of the sites containing test problems also present detailed numerical comparisons among different solvers. Many papers compare a newly proposed approach with a limited set of existing GO approaches. But only few papers make a systematic comparison between a large set of different solvers. Here we mention Neumaier et al. (2005), where different GO exact methods are compared, and Rios and Sahinidis (2013), which presents many different derivative-free algorithms for bound constrained problems. We also mention Beiranvand et al. (2017), which discusses some guidelines to perform a fair comparison between different solvers. Finally, CEC competitions offer the opportunity of comparing with many different algorithms, while within the context of GECCO-BBOB workshops, the Comparing Continuous Optimizers (COCO) platform allows for automated benchmarking. Users can benchmark their own solvers over a wide set of test functions and compare their results with those of other solvers.

Conclusions

In this paper we presented our view on what we consider as relevant in the recent GO literature. It can be immediately seen, by simply browsing the rich list of references below, how the field is attracting more and more active research and novel computational paradigms. Large scale GO problems are no more out of the scope of solution algorithms, to the point that even professional solvers like GUROBI and CPLEX now include some GO solver, at least for nonconvex quadratic optimization. Although we tried our best to cover many recent advances, we do not claim to have given account of all relevant papers in the field. This survey reflects our personal points of view on the subject and we are perfectly aware that omissions are inevitable. Besides alternative approaches to the GO problems presented in this paper, we deliberately did not even mention other relevant fields, like multi-objective GO, stochastic GO, bi-level optimization, parallel, distributed, GPU based or quantum computing. These topics might become the subject of a different survey.

Acknowledgements

We would like to thank the co-editors of this issue, Miguel Anjos, Tibor Illés, and Tamás Terlaky, for their kind invitation to contribute to this important collection of papers. We also thank the anonymous reviewers for their careful reading and constructive suggestions.

References

Adjiman, C.S., Androulakis, I.P., Floudas, C.A., 1998. A global optimization method, αBB, for general twice-differentiable constrained NLPs - II. Implementation and computational results. Comput. Chem. Eng. 22, 1159–1179.
Adjiman, C.S., Androulakis, I.P., Maranas, C.D., Floudas, C.A., 1996. A global optimization method, αBB, for process design. Comput. Chem. Eng. Suppl. 20, S419–S424.
Adjiman, C.S., Dallwig, S., Floudas, C.A., Neumaier, A., 1998. A global optimization method, αBB, for general twice differentiable NLPs- I. Theoretical advances. Comput. Chem. Eng. 22, 1137–1158.
Ahmed, M.O., Vaswani, S., Schmidt, M., 2020. Combining Bayesian optimization and Lipschitz optimization. Mach. Learn. 109 (1), 79–102.
Ai, W., Zhang, S., 2009. Strong duality for the CDT subproblem: a necessary and sufficient condition. SIAM J. Optim. 19, 1735–1756.
Akrotirianakis, I.G., Floudas, C.A., 2004. A new class of improved convex underestimators for twice continuously differentiable constrained NLPs. J. Global Optim. 30, 367–390.
Alexandropoulos, S.-A.N., Pardalos, P.M., Vrahatis, M.N., 2020. Dynamic search trajectory methods for global optimization. Ann. Math. Artif. Intell. 88 (1), 3–37.
Amaral, P., Bomze, I., Júdice, J., 2014. Copositivity and constrained fractional quadratic problems. Math. Program. 146, 325–350.
Anstreicher, K., 2012. On convex relaxations for quadratically constrained quadratic programming. Math. Program. 136, 233–251.
Anstreicher, K., 2017. Kronecker product constraints with an application to the two-trust-region subproblem. SIAM J. Optim. 27, 368–378.
Anstreicher, K.M., 2009. Semidefinite programming versus the reformulation-linearization technique for nonconvex quadratically constrained quadratic programming. J. Global Optim. 43, 471–484.
Anstreicher, K.M., Burer, S., 2010. Computable representations for convex hulls of low-dimensional quadratic forms. Math. Program. 124, 33–43.
Araya, I., Reyes, V., 2016. Interval Branch-and-Bound algorithms for optimization and constraint satisfaction: a survey and prospects. J. Global Optim. 65, 837–866.
Audet, C., Hansen, P., Jaumard, B., Savard, G., 1999. A symmetrical linear maxmin approach to disjoint bilinear programming. Math. Program. 85, 573–592.
Auger, A., Brockhoff, D., Hansen, N., Tušar, T., GECCO workshop on real-parameter black-box optimization benchmarking (BBOB2019). http://numbbo.github.io/workshops/BBOB-2019/Accessed: 2021-04-02.
Bagattini, F., Schoen, F., Tigli, L., 2018. Clustering methods for the optimization of atomic cluster structure. J. Chem. Phys. 148 (14).
Bagattini, F., Schoen, F., Tigli, L., 2019. Clustering methods for large scale geometrical global optimization. Optim. Methods Softw. 34 (5), 1099–1122.
Bai, L., Mitchell, J.E., Pang, J.-S., 2016. On conic QPCCs, conic QCQPs and completely positive programs. Math. Program. 159, 109–136.

Ballerstein, M., Michaels, D., 2014. Extended formulations for convex envelopes. J. Global Optim. 60, 217–238.
Barvinok, A., 1993. Feasibility testing for systems of real quadratic equations. Discr. Comput. Geom. 10, 1–13.
Beale, E.M.L., Forrest, J.J.H., 1976. Global optimization using special ordered sets. Math. Program. 10 (1), 52–69.
Beck, A., Teboulle, M., 2009. A convex optimization approach for minimizing the ratio of indefinite quadratic functions over an ellipsoid. Math. Program. 118, 13–35.
Becker, R.W., Lago, G.V., 1970. A global optimization algorithm. In: Proceedings of the 8th Allerton Conference on Circuits and Systems Theory, pp. 3–12.
Beiranvand, V., Hare, W., Lucet, Y., 2017. Best practices for comparing optimization algorithms. Optim. Eng. 18, 815–848.
Belotti, P., Bonami, P., Vigerske, S., Wächter, A., 2006. Couenne, an exact solver for nonconvex MINLPs.
Belotti, P., Lee, J., Liberti, L., Margot, F., Wächther, A., 2009. Branching and bounds tightening techniques for non-convex MINLP. Optim. Methods Softw. 24, 597–634.
Bemporad, A., 2020. Global optimization via inverse distance weighting and radial basis functions. Comput. Optim. Appl..
Ben-Tal, A., den Hertog, D., 2014. Hidden conic quadratic representation of some nonconvex quadratic optimization problems. Math. Program. 143, 1–29.
Ben-Tal, A., Nemirovski, A., 2001. Lectures on Modern Convex Optimization. Society for Industrial and Applied Mathematics.
Berenguel, J., Casado, L., Garcia, I., Hendrix, E., Messine, F., 2013. On interval branch-and-bound for additively separable functions with common variables. J. Global Optim. 56, 1101–1121.
Bienstock, D., 2016. A note on polynomial solvability of the CDT problem. SIAM J. Optim. 26, 488–498.
Bienstock, D., Chen, C., Muñoz, G., 2020. Outer-product-free sets for polynomial optimization and oracle-based cuts. Math. Program. 183, 105–148.
Bienstock, D., Michalka, A., 2014. Cutting-planes for optimization of convex functions over nonconvex sets. SIAM J. Optim. 24, 643–677.
Bienstock, D., Michalka, A., 2014. Polynomial solvability of variants of the trust-region subproblem. In: SODA 14 Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 380–390.
Binois, M., Ginsbourger, D., Roustant, O., 2020. On the choice of the low-dimensional domain for global optimization via random embeddings. J. Global Optim. 76 (1), 69–90.
Biscani, F., Izzo, D., 2020. A parallel global multiobjective framework for optimization: PAGMO. J. Open Source Softw. 5 (53), 2338.
Boland, N., Dey, S., Kalinowski, T., Molinaro, M., Rigterink, F., 2017. Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions. Math. Program. 162, 523–535.
Bomze, I., 2015. Copositive relaxation beats lagrangian dual bounds in quadratically and linearly constrained quadratic optimization problems. SIAM J. Optim. 25, 1249–1275.
Bomze, I., Cheng, J., Dickinson, P., Lisser, A., 2017. A fresh CP look at mixed-binary QPs: new formulations and relaxations. Math. Program. 166, 159–184.
Bomze, I., Frommlet, F., Locatelli, M., 2010. Copositive bounds for improving SDP bounds on the clique number. Math. Program. 124, 13–32.
Bomze, I., Jeyakumar, V., Li, G., 2018. Extended trust-region problems with one or two balls: exact copositive and Lagrangian relaxations. J. Global Optim. 71, 551–569.
Bomze, I., Schachinger, W., Uchida, G., 2012. Think co(mpletely)positive! matrix properties, examples and a clustered bibliography on copositive optimization. J. Global Optim. 52, 423–445.
Bomze, I.M., de Klerk, E., 2002. Solving standard quadratic optimization problems via linear, semidefinite and copositive programming. J. Global Optim. 24, 163–185.
Bomze, I.M., Duer, M., de Klerk, E., Roos, C., Quist, A., Terlaky, T., 2000. On copositive programming and standard quadratic optimization problems. J. Global Optim. 18, 301–320.
Bonami, P., Günlük, O., Linderoth, J., 2018. Globally solving nonconvex quadratic programming problems with box constraints via integer programming methods. Math. Program. Comput. 10, 333–382.
Bonami, P., Lodi, A., Schweiger, J., Tramontani, A., 2019. Solving quadratic programming by cutting planes. SIAM J. Optim. 29, 1076–1105.
Borradaile, G., Van Hentenryck, P., 2005. Safe and tight linear estimators for global optimization. Math. Program. 102, 495–517.
Bundfuss, S., Duer, M., 2009. An adaptive linear approximation algorithm for copositive programs. SIAM J. Optim. 20, 30–53.
Burer, S., 2009. On the copositive representation of binary and continuous nonconvex quadratic programs. Math. Program. 120, 479–495.
Burer, S., Anstreicher, K.M., 2013. Second-order-cone constraints for extended trust-region subproblems. SIAM J. Optim. 23 (1), 432–451.
Burer, S., Dong, H., 2012. Representing quadratically constrained quadratic programs as generalized copositive programs. Oper. Res. Lett. 40, 203–206.
Burer, S., Dong, H., 2013. Separation and relaxation for cones of quadratic forms. Math. Program. 137, 343–370.
Burer, S., Kilinç-Karzan, F., 2017. How to convexify the intersection of a second order cone and a nonconvex quadratic. Math. Program. 162, 393–429.
Burer, S., Kim, S., Kojima, M., 2014. Faster, but weaker, relaxations for quadratically constrained quadratic programs. Comput. Optim. Appl. 59, 27–45.
Burer, S., Letchford, A.N., 2009. On nonconvex quadratic programming with box constraints. SIAM J. Optim. 20, 1073–1089.
Burer, S., Vandenbussche, D., 2008. A finite branch-and-bound algorithm for nonconvex quadratic programming via semidefinite relaxations. Math. Program. 113, 259–282.
Burer, S., Vandenbussche, D., 2009. Globally solving box-constrained nonconvex quadratic programs with semidefinite-based finite branch-and-bound. Comput. Optim. Appl. 43, 181–195.
Burer, S., Yang, B., 2015. The trust region subproblem with non-intersecting linear constraints. Math. Program. 149, 253–264.
Burer, S., Ye, Y., 2020. Exact semidefinite formulations for a class of (random and non-random) nonconvex quadratic programs. Math. Program. 181, 1–17.
Cabassi, F., Locatelli, M., 2016. Computational investigation of simple memetic approaches for continuous global optimization. Comput. Oper. Res. 72, 50–70.
Caprara, A., Locatelli, M., 2010. Global optimization problems and domain reduction strategies. Math. Program. 125, 125–137.
Caprara, A., Locatelli, M., Monaci, M., 2016. Theoretical and computational results about optimality-based domain reductions. Comput. Optim. Appl. 64, 513–533.
Carrizosa, E., Hansen, P., Messine, F., 2004. Improving interval analysis bounds by translations. J. Global Optim. 29, 157–172.
Cartis, C., Fowkes, J., Gould, N., 2015. Branching and bounding improvements for global optimization algorithms with Lipschitz continuity properties. J. Global Optim. 61, 429–457.
Chen, J., Burer, S., 2012. Globally solving nonconvex quadratic programming problems via completely positive programming. Math. Program. Comput. 4, 33–52.
Consolini, L., Locatelli, M., 2017. On the complexity of quadratic programming with two quadratic constraints. Math. Program. 164, 91–128.
Csallner, A.E., Csendes, T., Markót, M.C., 2000. Multisection in interval branch-and-bound methods for global optimization I. Theoretical results. J. Global Optim. 16, 371–392.
Cui, L., Li, G., Lin, Q., Chen, J., Lu, N., 2016. Adaptive differential evolution algorithm with novel mutation strategies in multiple sub-populations. Comput. Oper. Res. 67, 155–173.
Dantzig, G.B., 1960. On the significance of solving linear programming problems with some integer variables. Econometrica 28 (1), 30–44.
Das, S., Mullick, S., Suganthan, P., 2016. Recent advances in differential evolution-an updated survey. Swarm Evol. Comput. 27, 1–30.
Davarnia, D., Richard, J.-P., Tawarmalani, M., 2017. Simultaneous convexification of bilinear functions over polytopes with application to network interdiction. SIAM J. Optim. 27, 1801–1833.
de Angelis, P.L., Bomze, I.M., Toraldo, G., 2004. Ellipsoidal approach to box-constrained quadratic problems. J. Global Optim. 28, 1–15.
de Klerk, E., Hess, R., Laurent, M., 2017. Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization. SIAM J. Optim. 27, 347–367.
de Klerk, E., Lasserre, J., Laurent, M., Sun, Z., 2017. Bound-constrained polynomial optimization using only elementary calculations. Math. Oper. Res. 42, 834–853.
de Klerk, E., Laurent, M., Parrillo, P.A., 2006. A PTAS for the minimization of polynomials of fixed degree over the simplex. Theor. Comput. Sci. 361, 210–225.
de Klerk, E., Pasechnik, D.V., 2002. Approximation of the stability number of a graph via copositive programming. SIAM J. Optim. 12, 875–892.
Del Pia, A., Khajavirad, A., 2017. A polyhedral study of binary polynomial programs. Math. Oper. Res. 42, 389–410.
Del Pia, A., Khajavirad, A., 2018. On decomposability of multilinear sets. Math. Program. 170, 387–415.
Del Ser, J., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P., Coello Coello, C., Herrera, F., 2019. Bio-inspired computation: where we stand and what's next. Swarm Evol. Comput 48, 220–250.
Depetrini, D., Locatelli, M., 2011. Approximation of linear fractional/multiplicative problems. Math. Program. 128, 437–443.
Di Carlo, M., Vasile, M., Minisci, E., 2020. Adaptive multi-population inflationary differential evolution. Soft comput 24 (5), 3861–3891.
Diouane, Y., Gratton, S., Vicente, L.N., 2015. Globally convergent evolution strategies. Math. Program. 152 (1–2), 467–490.
Diouane, Y., Gratton, S., Vicente, L.N., 2015. Globally convergent evolution strategies for constrained optimization. Comput. Optim. Appl. 62 (2), 323–346.
Dixon, L., 1975. Towards global optimisation. In: Dixon, L., Szegö, G. (Eds.), Towards Global Optimization. North-Holland, pp. 29–54.
Dixon, L., Szegö, G. (Eds.), Towards global optimisation. proceedings of a workshop at the University of Cagliari, Italy, October 1974, 1975. North-Holland.
Dixon, L., Szegö, G. (Eds.), 1978. Towards Global Optimisation 2. North-Holland.
Domes, F., Neumaier, A., 2016. Constraint aggregation for rigorous global optimization. Math. Program. 155, 375–401.
Evtushenko, Y.P., 1971. Numerical method for finding the global extremum of a function. Vyl. Mat. I Mat. Fiz. 11, 1390–1403. (in Russian)
Falk, J.E., Soland, R.M., 1969. An algorithm for separable nonconvex programming problems. Manage. Sci. 15 (9), 550–569.
Feliot, P., Bect, J., Vazquez, E., 2017. A Bayesian approach to constrained single- and multi-objective optimization. J. Global Optim. 67 (1), 97–133.
Ferreiro-Ferreiro, A.M., García-Rodríguez, J.A., Souto, L., Vázquez, C., 2019. Basin Hopping with synched multi L-BFGS local searches. Parallel implementation in multi-CPU and GPUs. Appl. Math. Comput. 356, 282–298.
Floudas, C.A., Pardalos, P.M., Adjiman, C.S., Esposito, W.R., Gumus, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A., Schweiger, C., 1999. Handbook of Test Problems in Local and Global Optimization. Nonconvex Optimization and its Applications, 33. Kluwer Academic Publishers, Dordrecht.
Fowkes, J., Gould, N., Farmer, C., 2013. A branch and bound algorithm for the global optimization of Hessian Lipschitz continuous functions. J. Global Optim. 56, 1791–1815.
Frazier, P.I., 2018. Bayesian optimization. In: Gel, E., Ntaimo, L., Shier, D., Greenberg, H.J. (Eds.), Recent Advances in Optimization and Modeling of Contemporary Problems. INFORMS, pp. 255–278.
Furini, F., Traversi, E., Belotti, P., Frangioni, A., Gleixner, A., Gould, N., Liberti, L., Lodi, A., Misener, R., Mittelmann, H., Sahinidis, N., Vigerske, S., Wiegele, A., 2019. QPLIB: A library of quadratic programming instances. Math. Program. Comput. 11, 237–265.

Gamrath, G., Anderson, D., Bestuzheva, K., Chen, W.-K., Eifler, L., Gasse, M., Gemander, P., Gleixner, A., Gottwald, L., Halbig, K., Hendel, G., Hojny, C., Koch, T., Le Bodic, P., Maher, S., Matter, F., Miltenberger, M., Mühmer, E., Müller, B., Pfetsch, M., Schlösser, F., Serrano, F., Shinano, Y., Tawfik, C., Vigerske, S., Wegscheider, F., Weninger, D., Witzig, J., 2020. The SCIP Optimization Suite 7.0. Technical Report. Optimization Online.

Garrido-Merchán, E.C., Hernández-Lobato, D., 2020. Dealing with categorical and integer–valued variables in Bayesian optimization with Gaussian processes. Neurocomputing 380, 20–35.

Gavana, A., Global Optimization Benchmarks. http://infinity77.net/go_2021/. Accessed: 2021-04-02.

Gaviano, M., 1975. Some general results on the convergence of random search algorithms in minimisation problems. In: Dixon, L., Szegö, G. (Eds.), Towards Global Optimization. North-Holland, pp. 149–157.

Ghosh, S., Das, S., Vasilakos, A.V., Suresh, K., 2012. On convergence of differential evolution over a class of continuous functions with unique global optimum. IEEE Trans. Syst. Man Cybern.Part B (Cybernetics) 42 (1), 107–124.

Gleixner, A., Berthold, T., Müller, B., Weltge, S., 2017. Three enhancements for optimization-based bound tightening. J. Global Optim. 67, 731–757.

Grosso, A., Locatelli, M., Schoen, F., 2007. A population-based approach for hard global optimization problems based on dissimilarity measures. Math. Program. 110 (2), 373–404.

Hadi, A.A., Mohamed, A.W., Jambi, K.M., 2019. LSHADE-SPA Memetic framework for solving large-scale optimization problems. Complex Intell. Syst. 5 (1), 25–40.

Hager, W., Phan, D., 2009. An ellipsoidal branch and bound algorithm for global optimization. SIAM J. Optim. 20, 740–758.

Hansen, N., 2006. The CMA evolution strategy: a comparing review. In: Lozano, J.A., Larrañaga, P., Inza, I., Bengoetxea, E. (Eds.), Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms. Springer, Berlin, Heidelberg, pp. 75–102.

Hansen, N., Ostermeier, A., 2001. Completely derandomized self-adaptation in evolution strategies. Evol. Comput. 9 (2), 159–195.

Hansen, P., Lagouanelle, J.-L., Messine, F., 2007. Comparison between Baumann and admissible simplex forms in interval analysis. J. Global Optim. 37, 215–228.

Hendrix, E.M.T., Tóth, B., 2010. Introduction to Nonlinear and Global Optimization. Springer, New York, NY.

Hernández-Lobato, J.M., Gelbart, M.A., Adams, R.P., Hoffman, M.W., Ghahramani, Z., 2016. A general framework for constrained bayesian optimization using information-based search. J. Mach. Learn. Res. 17 (160), 1–53.

Ho-Nguyen, N., Kilinç-Karzan, F., 2017. A Second-Order Cone based approach for solving the trust-region subproblem and its variants. SIAM J. Optim. 27, 1485–1512.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press.

Horst, R., 1976. An algorithm for nonconvex programming problems. Math. Program. 10 (1), 312–321.

Horst, R., Pardalos, P.M. (Eds.), 1995. Handbook of Global Optimization. Kluwer Academic Publishers, Dordrecht.

Horst, R., Pardalos, P.M., Thoai, N.V., 2001. Introduction to Global Optimization. Nonconvex Optimization and Its Applications, 48. Springer Verlag.

Horst, R., Tuy, H., 1993. Global Optimization: Deterministic Approaches, 2nd edition Springer Verlag.

Gondzio, J., Yildrim, E., 2021. Global solutions of nonconvex Standard Quadratic Programs via Mixed Integer Linear Programming reformulations. J. Glob. Optim 81, 293–321.

IEEE TfLsgo. IEEE Task Force on Large-Scale Global Optimization. https://www.tflsgo.orgAccessed: 2021-04-02.

Jach, M., Michaels, D., Weismantel, R., 2008. The convex envelope of (n–1)-convex functions. SIAM J. Optim. 19, 1451–1466.

Jansson, C., Chaykin, D., Keil, C., 2007. Rigorous error bounds for the optimal value in semidefinite programming. SIAM J. Numer. Anal. 46, 180–200.

Jeyakumar, V., Li, G., 2014. Trust-region problems with linear inequality constraints: exact SDP relaxation, global optimality and robust optimization. Math. Program. 147, 171–206.

Jeyakumar, V., Li, G., 2018. Exact second-order cone programming relaxations for some nonconvex minimax quadratic optimization problems. SIAM J. Optim. 28, 760–787.

Jiang, R., Li, D., Wu, B., 2018. SOCP reformulation for the generalized trust region subproblem via a canonical form of two symmetric matrices. Math. Program. 169, 531–563.

Johnson, S. G., The NLopt nonlinear-optimization package. http://github.com/stevengj/nlopt. Accessed: 2021-04-02.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. J. Global Optim. 13 (4), 455–492.

Jones, E., Oliphant, T., Peterson, P., et al., 2001–. SciPy: Open source scientific tools for Python. Accessed: 2021-04-02.

Kandasamy, K., Vysyaraju, K.R., Neiswanger, W., Paria, B., Collins, C.R., Schneider, J., Poczos, B., Xing, E.P., 2020. Tuning hyperparameters without grad students: scalable and Robust Bayesian optimisation with dragonfly. J. Mach. Learn. Res. 21, 1–27.

Kawaguchi, K., 2016. Deep Learning without Poor Local Minima. Technical Report arXiv:1605.07110v3. Massachusetts Institute of Technology.

Kazazakis, N., Adjiman, C., 2018. Arbitrarily tight α-BB underestimators of general non–linear functions over sub-optimal domains. J. Global Optim. 71, 815–844.

Kearfott, R.B., 2011. Interval computations, rigor and non-rigor in deterministic continuous global optimization. Optim. Methods Softw. 26, 259–279.

Khajavirad, A., Michalek, J., Sahinidis, N., 2014. Relaxations of factorable functions with convex-transformable intermediates. Math. Program. 144, 107–140.

Khajavirad, A., Sahinidis, N., 2012. Convex envelopes of products of convex and component-wise concave functions. J. Global Optim. 52, 391–409.

Khajavirad, A., Sahinidis, N.V., 2013. Convex envelopes generated from finitely many compact convex sets. Math. Program. 137, 371–408.

Kim, M.J., 2020. Variance regularization in sequential Bayesian optimization. Math. Oper. Res. 45 (3), 966–992.

Kim, S., Kojima, M., 2003. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. Comput. Optim. Appl. 26, 143–154.

Kuang, X., Zuluaga, L., 2018. Completely positive and completely positive semidefinite tensor relaxations for polynomial optimization. J. Global Optim. 70, 551–577.

Kuno, T., 2005. A revision of the trapezoidal branch-and-bound algorithm for linear sum-of-ratios problems. J. Global Optim. 33, 215–234.

Kushner, H.J., 1964. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. J. Basic Engi.g 97–106.

Lasserre, J.B., 2001. Global optimization with polynomials and the problem of moments. SIAM J. Optim. 11, 796–817.

Lasserre, J.B., 2005. Polynomial programming: LP-relaxations also converge. SIAM J. Optim. 15, 383–393.

Lasserre, J.B., 2006. A sum of squares approximation of nonnegative polynomials. SIAM J. Optim. 16, 751–765.

Laurent, M., 2007. Semidefinite representations for finite varieties. Math. Program. 109, 1–26.

Le Thi, H.A., 2000. An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. Math. Program. 87, 401–426.

Linderoth, J., 2005. A simplicial branch-and-bound algorithm for solving quadratically constrained quadratic programs. Math. Program. 103, 251–282.

Liuzzi, G., Locatelli, M., Piccialli, V., 2019. A new branch-and-bound algorithm for standard quadratic programming problems. Optim. Methods Softw. 34, 79–97.

Liuzzi, G., Locatelli, M., Piccialli, V., Rass, S., 2021. Computing mixed strategies equilibria in presence of switching costs by the solution of nonconvex QP problems. Comput. Optim.. To appear

Locatelli, M., 2013. Approximation algorithm for a class of global optimization problems. J. Global Optim. 55, 13–25.

Locatelli, M., 2015. Alternative branching rules for some nonconvex problems. Optim. Methods Softw. 30, 365–378.

Locatelli, M., 2015. Some results for quadratic problems with one or two quadratic constraints. Oper. Res. Lett. 43, 126–131.

Locatelli, M., 2016. Exactness conditions for an SDP relaxation of the extended trust region problem. Optim. Lett. 10, 1141–1151.

Locatelli, M., 2016. Non polyhedral convex envelopes for 1-convex functions. J. Global Optim. 65, 637–655.

Locatelli, M., 2020. Convex envelope of bivariate cubic functions over rectangular regions. J. Global Optim. 76, 1–24.

Locatelli, M., Schoen, F., 2013. Global Optimization: Theory, Algorithms, and Applications. SIAM-MOS. SIAM.

Locatelli, M., Vasile, M., 2015. (Non) convergence results for the differential evolution method. Optim. Lett. 9 (3), 413–425.

Lofberg, J., 2004. YALMIP : a toolbox for modeling and optimization in MATLAB. In: 2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508), pp. 284–289.

Luedtke, J., D'Ambrosio, C., Linderoth, J., Schweiger, J., 2020. Strong convex nonlinear relaxations of the pooling problem. SIAM J. Optim. 30, 1582–1609.

Luedtke, J., Namazifar, M., Linderoth, J., 2012. Some results on the strength of relaxations of multilinear functions. Math. Program. 136, 325–351.

Ma, X., Li, X., Zhang, Q., Tang, K., Liang, Z., Xie, W., Zhu, Z., 2019. A survey on cooperative co-evolutionary algorithms. IEEE Trans. Evol. Comput. 23 (3), 421–441.

Mansueto, P., Schoen, F., 2021. Memetic differential evolution methods for clustering problems. Pattern Recognit 114, 107849.

Markot, M., Schichl, H., 2014. Bound constrained interval global optimization in the COCONUT environment. J. Global Optim. 60, 751–776.

Markót, M.C., Csendes, T., 2005. A new verified optimization technique for the "packing circles in a unit square" problem. SIAM J. Optim. 16, 193–219.

Markót, M.C., Fernandez, J., Casado, L.G., Csendes, T., 2006. New interval methods for constrained global optimization. Math. Program. 106, 287–318.

Martinez, J., Casado, L., Garcia, I., Sergeyev, Y., Toth, B., 2004. On an efficient use of gradient information for accelerating interval global optimization algorithms. Numer. Alg. 37, 61–69.

Martinez-Cantin, R., 2014. Bayesopt: a bayesian optimization library for nonlinear optimization, experimental design and bandits. J. Mach. Learn. Res. 15 (115), 3915–3919.

Mathesen, L., Pedrielli, G., Ng, S.H., Zabinsky, Z.B., 2020. Stochastic optimization with adaptive restart: a framework for integrated local and global learning. J. Global Optim..

McCormick, G., 1972. Attempts to calculate global solutions of problems that may have local minima. In: Lootsma, F. (Ed.), Numerical Methods for Non-Linear Optimization. Academic Press, pp. 209–221.

McCormick, G.P., 1976. Computability of global solutions to factorable nonconvex programs: Part I Convex underestimating problems. Math. Program. 10 (1), 147–175.

Meldgaard, S.A., Kolsbjerg, E.L., Hammer, B., 2018. Machine learning enhanced global optimization by clustering local environments to enable bundled atomic energies. J. Chem. Phys. 149 (13), 134104.

Messine, F., 2004. Deterministic global optimization using interval constraint propagation techniques. RAIRO-Oper. Res. 38 (4), 277–293.

Meyer, C.A., Floudas, C.A., 2005. Convex envelopes for edge-concave functions. Math. Programm. B 103, 207–224.

Meyer, C.A., Floudas, C.A., 2005. Convex underestimation of twice continuously differentiable functions by piecewise quadratic perturbation: spline αBB underestimators. J. Global Optim. 32, 221–258.
Mittel, S., Schulz, A., 2013. An FPTAS for optimizing a class of low-rank functions over a polytope. Math. Program. 141, 103–120.
Mockus, J., 1975. On Bayesian methods of optimisation. In: Dixon, L., Szegö, G. (Eds.), Towards Global Optimization. North-Holland, pp. 166–181.
Moriconi, R., Kumar, K.S.S., Deisenroth, M.P., 2020. High-dimensional Bayesian optimization with projections using quantile Gaussian processes. Optim. Lett. 14 (1), 51–64.
Motzkin, T.S., Strauss, E.G., 1965. Maxima for graphs and a new proof of a theorem of Turán. Can. J. Math. 17, 533–540.
Neumaier, A., Global optimiztion. https://www.mat.univie.ac.at/~neum/glopt.html Accessed: 2021-04-02.
Neumaier, A., 2004. Complete search in continuous global optimization and constraint satisfaction. In: Acta Numerica 2004. Cambridge University Press, pp. 271–369.
Neumaier, A., Shcherbina, O., 2004. Safe bounds in linear and mixed-integer programming. Math. Program. 99, 283–296.
Neumaier, A., Shcherbina, O., Huyer, W., Vinkó, T., 2005. A comparison of complete global optimization solvers. Math. Program. 103 (2), 335–356.
Nguyen, T., Richard, J.-P., Tawarmalani, M., 2018. Deriving convex hulls through lifting and projection. Math. Program. 169, 377–415.
Nie, J., 2013. Certifying convergence of Lasserres hierarchy via flat truncation. Math. Program. 142, 485–510.
Nie, J., 2014. Optimality conditions and finite convergence of Lasserres hierarchy. Math. Program. 146, 97–121.
Nie, J., Demmel, J., Sturmfels, B., 2006. Minimizing polynomials via sum of squares over the gradient ideal. Math. Program. 106, 587–606.
Padberg, M., 1989. The boolean quadric polytope: some characteristics, facets and relatives. Math. Program. 45, 139–172.
Papp, D., Yildiz, S., 2019. Sum-of-squares optimization without semidefinite programming. SIAM J. Optim. 29, 822–851.
Parrillo, P.A., 2000. Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization. California Institute of Technology, Pasadena.
Paulavičius, R., Žilinskas, J., 2014. Simplicial Global Optimization. Springer, New York, NY.
Peña, J., Vera, J., Zuluaga, L.F., 2007. Computing the stability number of a graph via linear and semidefinite programming. SIAM J. Optim. 18, 87–105.
Piyavskii, S.A., 1972. An algorithm for finding the absolute extremum of a function. USSR Comput. Math. Math.Phys. 12 (4), 57–67.
Pong, T., Wolkowicz, H., 2014. The generalized trust region subproblem. Comput. Optim. Appl. 58, 273–322.
Puranik, Y., Sahinidis, N., 2017. Bounds tightening based on optimality conditions for nonconvex box-constrained optimization. J. Global Optim. 67, 59–77.
Rendl, F., Wolkowicz, H., 1997. A semidefinite framework for trust region subproblems with applications to large scale minimization. Math. Program. 77, 273–299.
Rikun, A.D., 1997. A convex envelope formula for multilinear functions. J. Global Optim. 10, 425–437.
Rios, L., Sahinidis, N., 2013. Derivative-free optimization: a review of algorithms and comparison of software implementations. J. Global Optim. 56, 1247–1293.
Sahinidis, N., 2017. BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs, *User's Manual*.
Sakaue, S., Nakatsukasa, Y., Takeda, A., Iwata, S., 2016. Solving generalized CDT problems via two-parameter eigenvalues. SIAM J. Optim. 26, 1669–1694.
Santana, A., Dey, S., 2020. The convex hull of a quadratic constraint over a polytope. SIAM J. Optim. 30, 2983–2997.
Schichl, H., Neumaier, A., 2005. Interval analysis on directed acyclic graphs for global optimization. J. Global Optim. 33, 541–562.
Schoen, F., Tigli, L., 2021. Efficient large scale global optimization through clustering-based population methods. Comput. Oper. Res. 127 (105165).
Scott, J.K., Stuber, M.D., Barton, P.I., 2011. Generalized McCormick relaxations. J. Global Optim. 51, 569–606.
Sherali, H.D., Tuncbilek, C.H., 1992. A global optimization algorithm for polynomial programming problems using a reformulation-linearization technique. J. Global Optim. 2, 101–112.
Shubert, B.O., 1972. A sequential method seeking the global maximum of a function. SIAM J. Numer. Anal. 9.
Škvorc, U., Eftimov, T., Korošec, P., 2019. CEC real-parameter optimization competitions: Progress from 2013 to 2018. In: 2019 IEEE Congress on Evolutionary Computation (CEC), pp. 3126–3133.
Snoek, J., Ripped, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, Adams, R., 2015. Scalable Bayesian optimization using deep neural networks, 3, pp. 2161–2170.
Sojoudi, S., Lavaei, J., 2014. Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure. SIAM J. Optim. 24, 1746–1778.
Soland, R.M., 1971. An algorithm for separable nonconvex programming problems II: nonconvex constraints. Manage. Sci. 17 (11), 759–773.
Sturm, J.F., Zhang, S., 2003. On cones of nonnegative quadratic functions. Math. Oper. Res. 28 (2), 246–267.
Sun, Y., Li, X., Ernst, A., Omidvar, M.N., 2019. Decomposition for large-scale optimization problems with overlapping components. In: 2019 IEEE Congress on Evolutionary Computation (CEC), pp. 326–333.
Tardella, F., 2003. On the existence of polyhedral convex envelopes. In: Frontiers in Global Optimization. Kluwer Academic Publishers, pp. 563–574.
Tardella, F., 2008. Existence and sum decomposition of vertex polyhedral convex envelopes. Optim. Lett. 2, 363–375.
Tawarmalani, M., Richard, J.-P.P., Xiong, C., 2013. Explicit convex and concave envelopes through polyhedral subdivisions. Math. Program. 138, 531–577.
Tawarmalani, M., Sahinidis, N.V., 2001. Semidefinite relaxations of fractional programs via novel convexification techniques. J. Global Optim. 20, 137–158.
Tawarmalani, M., Sahinidis, N.V., 2004. Global optimization of mixed-integer nonlinear programs: a theoretical and computational study. Math. Program. 99 (3), 563–591.
Törn, A.A., 1978. A search–clustering approach to global optimization. In: Towards Global Optimization 2. North–Holland, pp. 49–62.
Tóth, B., Hendrix, E., Casado, L., Garcia, I., 2016. On refinement of the unit simplex using regular simplices. J. Global Optim. 64, 305–323.
Tóth, B., Hendrix, E.M.T., Casado, L.G., 2021. On monotonicity and search strategies in face-based copositivity detection algorithms. Central Eur. J. Oper. Res. available online.
Tran, N., Schneider, J.-G., Weber, I., Qin, A.K., 2020. Hyper-parameter optimization in classification: to-do or not-to-do. Pattern Recognit. 103, 107245.
Vandenbussche, D., Nemhauser, G.L., 2005. A branch-and-cut algorithm for nonconvex quadratic programs with box constraints. Math. Program. 102, 559–575.
Vandenbussche, D., Nemhauser, G.L., 2005. A polyhedral study of nonconvex quadratic programs with box constraints. Math. Program. 102, 531–557.
Vinkó, T., Gelle, K., 2017. Basin Hopping Networks of continuous global optimization problems. Central Eur. J. Oper. Res. 25 (4), 985–1006.
Vu, X.-H., Schichl, H., Sam-Haroud, D., 2009. Interval propagation and search on directed acyclic graphs for numerical constraint solving. J. Global Optim. 45, 499–531.
Vui, H.H., So'n, P.T., 2008. Global optimization of polynomials using the truncated tangency variety and sums of squares. SIAM J. Optim. 19, 941–951.
Wales, D.J., 2018. Exploring energy landscapes. Annu. Rev. Phys. Chem. 69 (1), 401–425.
Wang, A., Kilinç-Karzan, F., 2020. The generalized trust region subproblem: solution complexity and convex hull results. Math. Program. available online.
Wang, A., Kilinç-Karzan, F., 2021. On the tightness of SDP relaxations of QCQPs. Math. Program. to appear.
Wang, Z., Hutter, F., Zoghi, M., Matheson, D., Feitas, N.d., 2016. Bayesian optimization in a billion dimensions via random embeddings. J. Artif. Intell. Res. 55, 361–387.
Wechsung, A., Scott, J., Watson, H., Barton, P., 2015. Reverse propagation of McCormick relaxations. J. Global Optim. 63, 1–36.
Wu, G., Mallipeddi, R., Suganthan, P.N., Wang, R., Chen, H., 2016. Differential evolution with multi-population based ensemble of mutation strategies. Inf. Sci. 329, 329–345.
Xia, W., Vera, J., Zuluaga, L., 2020. Globally solving nonconvex quadratic programs via linear integer programming techniques. INFORMS J. Comput. 32, 40–56.
Yang, B., Anstreicher, K., Burer, S., 2018. Quadratic programs with hollows. Math. Program. 170, 541–553.
Yang, B., Burer, S., 2016. A two-variable approach to the two-trust region subproblem. SIAM J. Optim. 26, 661–680.
Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316.
Ye, Y., Zhang, S., 2003. New results on quadratic minimization. SIAM J. Optim. 14, 245–267.
Yuan, J., Wang, M., Ai, W., Shuai, T., 2017. New results on narrowing the duality gap of the extended Celis-Dennis-Tapia problem. SIAM J. Optim. 27, 890–909.
Zhan, D., Xing, H., 2020. Expected improvement for expensive optimization: a review. J. Global Optim. 78 (3), 507–544.
Zhang, Y., Sahinidis, N., Nohra, C., Rong, G., 2020. Optimality-based domain reduction for inequality-constrained NLP and MINLP problems. J. Global Optim. 77, 425–454.
Zhao, Y., Chen, X., Li, J., 2017. TGMin: a global-minimum structure search program based on a constrained basin-hopping algorithm. Nano Res. 10 (10), 3407–3420.