Algorithms for discrete nonlinear optimization in FICO Xpress

Pietro Belotti
FICO (Fair Isaac Corporation)
Birmingham, UK
Email: pietrobelotti@fico.com

Timo Berthold
FICO (Fair Isaac Corporation)
Berlin, Germany
Email: timoberthold@fico.com

Kelligton Neves
FICO (Fair Isaac Corporation)
Sao Paulo, Brazil
Email: kelligtonneves@fico.com

***Abstract*— The FICO Xpress-Optimizer is a commercial optimization solver for linear programming (LP), mixed integer linear programming (MIP), convex quadratic programming (QP), convex quadratically constrained quadratic programming (QCQP), second-order cone programming (SOCP) and their mixed-integer counterparts. Xpress also includes a general purpose non-linear solver, Xpress-NonLinear, which features a successive linear programming algorithm (SLP, first-order method), interior point methods and Artelys Knitro (second-order methods). This work explores algorithms for mixed-integer nonlinear programming problems (MINLPs), which are NP-hard in general, then it presents applications in signal processing and capitalizes advances in solving these problems with Xpress and its comprehensive suite of high-performance nonlinear solvers. Computational results show that signal processing nonlinear problems can be solved quickly and accurately, taking advantage of the algebraic modeling and procedural programming language, Xpress-Mosel, that allows to interact with the Xpress solver engines in a easy-to-learn way, and its unified modeling interface for all solvers, from linear to general nonlinear solvers.**

I. Introduction to MINLP

Optimization problems that feature, at the same time, nonlinear functions as constraints and integrality requirements for the variables are arguably among the most challenging problems in mathematical programming. These so-called *mixed-integer nonlinear programs (MINLPs)* are applied in various fields, e.g., in energy networks, power plant design, telecommunication networks, logistics, water distribution networks, chip design verification problems, traffic optimization, engineering, signal processing, manufacturing, and the chemical and biological sciences, see, e.g., [1], [2]. Real world problems regularly feature both nonlinear components and discrete decisions. Traditionally, these were often approximated by linear models and/or by using continuous variables, which was rather due to the fact that such models could be solved efficiently, whereas nonlinear, integer models cannot. Obviously, artificially linearized models are only a rough approximation of the reality and there is an high demand for having software available that can address nonlinear, discrete optimization problems directly.

Consequently, recent years have seen a strong interest in algorithms for *mixed-integer nonlinear programming* (MINLP). Advances in research are also reflected by the development and computational progress of several general-purpose solvers for MINLP or specific sub-classes, such as convex MINLP, *mixed-integer quadratically constrained quadratic programming* (MIQCQP), or *mixed-integer second order cone programming* (MISOCP) [3], [4], [5]. State-of-the-art solvers for MINLP and its subclasses comprise a variety of algorithmic techniques from several related fields such as *nonlinear programming*, *mixed-integer linear programming*, *global optimization*, and *constraint programming*, see, e.g., [6].

An MINLP is a mathematical optimization problem of the form

$$\begin{aligned} \min \quad & f(x) & & \\ \text{s.t.} \quad & g_j(x) \leq 0 & & \text{for } j = 1, \ldots, m, \\ & x_i \in \mathbb{Z} & & \text{for } i \in \mathcal{I}, \end{aligned} \tag{1}$$

where $\mathcal{I} \subseteq \{1, \ldots, n\}$ is the index set of the integer variables, $d \in \mathbb{R}^n$, $f, g_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, m$. Note that this definition is very general. First, maximization problems can be transformed to minimization problems by multiplying all objective function coefficients by -1. Similarly, "$\geq$" constraints can be multiplied by -1 to obtain "$\leq$" constraints. Equations can be replaced by two opposite inequalities. Among the most important subclasses of MINLP, is MIP, mixed-integer linear programming, for which both the *objective function* $f(x)$ and the constraint functions $g_j(x)$ are required to be linear.

Branch-and-bound [7] is the most widely used algorithm to solve mixed integer linear and nonlinear programs. State-of-the-art MIP solvers such as FICO Xpress [8], use *LP-based* branch-and-bound as a basic algorithm that is enhanced by various tricky subroutines to make the solvers efficient in practice.

The idea of branch-and-bound is simple, yet effective: an optimization problem is recursively split into smaller subproblems, thereby creating a search tree and implicitly enumerating all potential assignments of the integer variables.

During the course of the algorithm, *LP-relaxations* are solved for bounding. For (linear) MIPs, the LP relaxation is simply constructed by dropping the integrality conditions. For MINLP, an LP relaxation can be constructed from the bounds of the variables, gradient cuts for convex constraints and linear over- and underestimators of the nonconvex terms [9], [2].

II. MINLP problems in signal processing

Optimization is a standard tool used in signal processing applications. Configuration problems on networks, or more

Authorized licensed use limited to: Technische Universitaet Berlin. Downloaded on October 30,2020 at 08:49:01 UTC from IEEE Xplore. Restrictions apply.

generally graph structures, can be easily modeled as systems of linear equations and inequalities, with 0-1 integrality requirements on the variables. Resources which generate costs form the objective to be optimized. While relatively good solutions for such problems can be found by heuristic approaches or by simulation, finding an optimal solution, and proving its optimality, is the key advantage of MINLP.

Location problems for sensor networks are a typical example of an optimization problem that can be addressed by MINLP, see, e.g., [10]. In [11], [12], Hou, Shi, and Sherali introduce MINLP models to optimize the spectrum sharing for multi-hop networks with cognitive radios and software defined radios. Their model uses binary variables to indicate whether a certain sub-band is used by a link and continuous variables to determine the fraction a sub-band uses within a band and for the data rates. The model consists of capacity constraints, assignment constraints, linking constraints and interference constraints, the latter involving logarithmic expressions. The objective is to minimize the required network-wide radio spectrum resource for a set of user sessions.

Finally, there is a close connection between (non)linear programming and compressed sensing. Compressed sensing is searching for solutions of an underdetermined linear system, which is the typical setup for linear programming. In compressed sensing, the goal is to find a sparsest feasible solution, which corresponds to optimizing a linear or quadratic objective, depending on the chosen norm. Depending on the particular application, integer variables come into play. For recent work on using MIP and MINLP techniques in compressed sensing, see [13], [14].

III. Solving mixed integer quadratic & conic problems

Among the most practically important and approachable classes of MINLP problems solved by Xpress are two classes that share two main characteristics: the objective f and all constraint functions g_j are convex and either linear or quadratic. In the class of mixed-integer convex quadratically constrained quadratic programming (MIQCQP) problems, one seeks to minimize an objective function

$$f(x) = \frac{1}{2}x^\top Q_0 x + c_0^\top x$$

subject to constraints

$$g_j(x) = x^\top Q_j x + c_j^\top x + d_j \leq 0, j = 1, 2, \ldots, m,$$

where the matrices Q_j are positive semidefinite for all $j = 0, 1, \ldots, m$. Although the main problem is still nonconvex due to the discreteness of a subset of variables, convexity in both the objective and the constraints allows us to efficiently compute a valid lower bound by simply relaxing the integrality constraints and solving the convex relaxation.

For the special case where all g_j's are linear, Xpress implements the quadratic simplex method, which has the same structure as the simplex method for LP extended to deal with a quadratic objective function.

Similar to mixed integer linear programming, Xpress can solve MIQCQP problems to optimality by means of branch-and-bound (BB). This algorithm relies on the availability of a method that computes efficiently a lower bound z_{LB} on the optimal solution of (1) after relaxing the integrality constraints, and of a method that yields a feasible solution for the original problem, and as a result gives an upper bound z_{UB} on the optimal solution. The recursive partitioning performed by BB, outlined in section I, and the possibility to compare lower and upper bounds are the fundamental steps that limit dramatically the portion of the research space that needs to be explored to find a global optimum.

In the case of a MIP, a lower bound can be computed by solving a linear programming (LP) via the simplex method. This is a strong advantage in the solution of MIPs because BB algorithm generally create a large number of subproblem, but the corresponding LP problems do not change much; the simplex method allows for re-solving a previously solved LP very efficiently as the simplex method can be *warm-started*. If we used a BB based on the convex relaxation for solving convex MIQCQPs, this is no longer true. The reason for this is that all methods for solving QCQPs, especially the most sophisticated and efficient *interior point* methods [15], do not admit any warm-starting mechanism. As a consequence, the lower bound of every subproblem visited in a BB algorithm must be found by performing, without warm start, a full solve via an interior point method.

A. *Outer approximation for MIQCQP*

To overcome this efficiency issue, Xpress exploits convexity to obtain an LP relaxation that, albeit generally loose, can lead to much better performance. This is based on the Duran and Grossmann [16] approach to approximate a convex constraint $g(x) \leq 0$ using its first-degree Taylor approximation: for any vector $\bar{x} \in \mathbb{R}^n$, the following inequality is valid for any x such that $g(x) \leq 0$:

$$g(\bar{x}) + \nabla g(\bar{x})^\top (x - \bar{x}) \leq 0,$$

where $\nabla g(\bar{x})$ is the gradient of g computed at $\bar{x}$. For any subset $\bar{X} = \{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^k\}$ of k vectors, the system of inequalities $g(\bar{x}^i) + \nabla g(\bar{x}^i)^\top (x - \bar{x}^i) \leq 0$ for $i = 1, 2, \ldots, k$ is a relaxation of the original constraint. By applying this technique, called *Outer Approximation* (OA), to all quadratic constraints, the Xpress solver obtains an LP relaxation of the original feasible set. The trivial trick of replacing the objective function with an auxiliary variable z and adding the constraint $f(x) - z \leq 0$ yields an LP relaxation once the same technique is applied to the latter constraint.

Although the resulting LP is larger, in terms of constraints, than the original formulation (several linear inequalities might be added for each nonlinear constraint), the branch-and-bound can now take advantage of warm starting mechanisms and, with an opportune policy of adding extra linear inequalities to tighten the LP relaxation, the problem can be solved much more efficiently. The Xpress solver employs this technique for

Authorized licensed use limited to: Technische Universitaet Berlin. Downloaded on October 30,2020 at 08:49:01 UTC from IEEE Xplore. Restrictions apply.

its MIQCQP problems as it leverages on its efficient LP-based branch-and-bound.

B. Second-order conic problems

A subclass of MIQCQP where the Outer Approximation implemented in Xpress is successful is one where the quadratic constraints determine a convex region even though for the quadratic term $x^\top Qx$ the matrix Q is not positive semidefinite. A so-called *second-order cone* (SOC), or Lorentz cone, is a constraint of the form

$$x_0 \geq \sqrt{x_1^2 + x_2^2 + \cdots + x_h^2}$$

or equivalently

$$\sum_{i=1}^{h} x_i^2 \leq x_0^2.$$

It represents a convex region given that $x_0 \geq 0$. A *rotated* SOC is a constraint of the form

$$\sum_{i=1}^{h} x_i^2 \leq 2yz$$

where $y, z \geq 0$. Regular and rotated SOCs admit a matrix $Q = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & \text{-1} \end{bmatrix}$ and $Q = \begin{bmatrix} \mathrm{I} & 0 & 0 \\ 0 & 0 & \text{-1} \\ 0 & \text{-1} & 0 \end{bmatrix}$, respectively, where I is the identity matrix of order h. This class of problems, although a very specific case of MIQCQP, finds applications in countless fields of engineering, finance, and economics. Furthermore, they are at the base of the *robust optimization* [17] paradigm.

The Xpress-Optimizer uses a branch-and-bound method based on Outer Approximation for MISOCP problems as it proved to be the most efficient for instances of all sizes. Among the techniques that contribute to the effectiveness of the solver are two that have been the subject of intense research in recent years: *perspective reformulations* [18] and *cone decomposition* [19].

IV. Solving mixed integer nonconvex MINLPs via Xpress-NonLinear

The most general case of MINLP (1) has nonlinear, possibly nonquadratic and nonconvex function f and g_i, for all $j = 1, 2, \ldots, m$. Adding integrality constraints to a subset of variables yields perhaps the hardest class of problems in Optimization, one that cannot be tackled with techniques such as Outer Approximation because of the nonconvexity of the problem.

The Xpress-Optimizer has a component for dealing with this class of problems, which is also known in the literature as the class of *global optimization* problems. This component, called SLP, does not guarantee to find a globally optimal solution for general nonconvex problems, i.e., if either f or one of the g_j is nonconvex, but it applies a branch-and-bound algorithm that utilizes a Nonlinear Programming (NLP) solver to find a local optimum.

Amongst the most important components of such a solver are heuristics to find a feasible point for the MINLP (see, e.g. [20], a *multistart* mechanism for repeating the search of a good local optimum in case the problem is nonconvex, and a set of *bound reduction* techniques.

The importance of good bound reduction algorithms (see e.g. [21]) is made apparent by the fact that solving a nonconvex problem requires exploring possibly all local optima in the search space. In order to exclude as much of the search space as possible, it is useful to reduce the range of the variables by exploiting the nonlinear information associated with constraints and, if a feasible solution is available, the objective function. Consider for example the constraint $x_1 x_2 \geq 6$ with $x_1 \in [1, 2]$, $x_2 \in [2, 4]$. It is clear that the constraint implies that the variable bounds can be further restricted so that $x_1 \geq \frac{6}{4} = 1.5$ and $x_2 \geq \frac{6}{3} = 2$.

Bound reduction, which is widely used in MIP as well [22], is heavily employed by Xpress SLP both as a standalone procedure and in *probing* procedures [23].

V. Computational results

To show the SLP capabilities, we consider a Wireless Sensor Network (WSN) deployment problem, as described in [10]. In this problem, we need to optimally determine the location of cluster-heads in order to minimize communication power, while taking into account a network reliability. So, in this problem is considered a two-layer network consisting of a lower layer of sensor nodes and a upper layer of cluster-heads. Sensor nodes are fixed and, for redundancy purposes, it's required that each sensor must connect to at least *p* cluster-heads. As a capacity constraint, due to wireless interference, computing power limits, etc., each cluster-head cannot connect to more than *q* sensors. So, considering that the allocation for each sensor is binary while the location of each cluster-head is continuous, the goal is to find: (*i*) the optimal cluster-head locations and (*ii*) the optimal network connectivity such that communication power consumption is minimized.

The cluster-head deployment problem is formulated as a MINLP problem. This problem belongs to the category of location-allocation (LA) problems, which is difficult because it is neither convex nor concave and possesses multiple local minima and is formulated as follow:

Notation:

- S_i — Known fixed location of sensor node i and $i \in I$ where I is a index set. The total number of sensor nodes is $|I|$. It's assumed that all sensor nodes and cluster-head are in a plane, so $S_i \in \mathbb{R}^2$.
- R_j — Decision variable controlling location of cluster-head j and $j \in J$ where J is a index set as well. The total number of cluster-heads to be allocated is $|J|$, and $|J| \geq p$. It's assumed that J is given and fixed and that all cluster-heads are in the same plane as the sensor nodes, so $R_j \in \mathbb{R}^2$.
- $c_{i,j}$ — Binary variable indicating that sensor node i is connected to cluster-head j when equals 1, and 0 otherwise.

Authorized licensed use limited to: Technische Universitaet Berlin. Downloaded on October 30,2020 at 08:49:01 UTC from IEEE Xplore. Restrictions apply.

$T_{i,j}$ Transmission power of sensor node i to cluster-head j.

As stated, the goal is to minimize the total end-point power:

$$\min \sum_{i \in I} \sum_{j \in J} T_{i,j} \tag{2}$$

Knowing that

$$T_{i,j} = \frac{N_o \gamma ||S_i - R_j||^d}{\alpha} \tag{3}$$

where α is a constant scaling factor for all i, j, d is the exponent characterizing the signal attenuation with respect to transmission distance (usually $2 \leq d \leq 3.5$), N_o indicates the signal noise ratio and γ is a threshold determined by the electronic characteristics of the receiver. For simplicity, we consider α, N_o and γ equals to 1 and d equals to 2.

That way, the optimization problem is formulated as follow:

$$\min \sum_{i \in I} \sum_{j \in J} c_{i,j} \frac{N_o \gamma ||S_i - R_j||^d}{\alpha} \tag{4}$$

s.t.

$$\sum_{j \in J} c_{i,j} \geq p, \qquad \forall i \in I \tag{5}$$

$$\sum_{i \in I} c_{i,j} \leq q, \qquad \forall j \in J \tag{6}$$

$$R_j \in \mathbb{R}^2 \tag{7}$$

$$c_{i,j} \in \{0, 1\} \tag{8}$$

The constraint 5 ensure that each sensor connects to at least p cluster-heads. The constraint 6 ensure that each head-cluster will not have more than q sensors connected. As said before, this problem is a MINPL problem which can be solved via SLP using Xpress which applies a branch-and-bound algorithm that utilizes a Nonlinear Programming (NLP) solver to find a local optimum.

We implemented the problem using the mathematical language Xpress-Mosel version 3.10. The computing plataform was a Dell Latitude E7440, Intel Core I7-4600U 2.1GHz CPU, 16GBytes RAM, Windows 10 Enterprise 64-Bit, Xpress 7.9. Instance cases was generated with increasing sizes and sensors positioned randomly. The results are given in the following table.

Case	$\|I\|$	$\|J\|$	Num. variables	Run time (secs)	Obj. Val.
1	25	4	108	0.224	230.28
2	50	8	416	0.325	1085.18
3	75	12	924	3.28	853.81
4	100	16	1632	17.043	576.34
5	150	24	3648	74.035	496.56
6	200	32	3264	126.46	446.718

It shows that, even to a extremely difficult MINLP, the Xpress solve can handle large-scale instances and find a local optimum solution in a acceptable time.

REFERENCES

[1] C. A. Floudas, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications (Topics in Chemical Engineering)*. Oxford University Press, USA, 1995.

[2] M. Tawarmalani and N. V. Sahinidis, *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications*, ser. Nonconvex Optimization and Its Applications. Kluwer Academic Publishers, 2002, vol. 65.

[3] P. Bonami, M. Kılınç, and J. Linderoth, "Algorithms and software for convex mixed integer nonlinear programs," in *Mixed Integer Nonlinear Programming*, ser. The IMA Volumes in Mathematics and its Applications, J. Lee and S. Leyffer, Eds. Springer New York, 2012, vol. 154, pp. 1–39.

[4] M. R. Bussieck and S. Vigerske, "MINLP solver software," in *Wiley Encyclopedia of Operations Research and Management Science*, J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith, Eds. John Wiley & Sons, Inc., 2010, online publication.

[5] C. D'Ambrosio and A. Lodi, "Mixed integer nonlinear programming tools: a practical overview," *4OR: A Quarterly Journal of Operations Research*, vol. 9, pp. 329–349, 2011.

[6] S. Vigerske, "Decomposition in multistage stochastic programming and a constraint integer programming approach to mixed-integer nonlinear programming," Ph.D. dissertation, Humboldt-Universität zu Berlin, 2012, submitted.

[7] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.

[8] "FICO Xpress Optimization Suite," http://www.fico.com/en/Products/DMTools/xpress-overview/Pages/Xpress-Optimizer.aspx.

[9] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems," *Mathematical Programming*, vol. 10, pp. 146–175, 1976.

[10] X. Ning and C. G. Cassandras, "Optimal cluster-head deployment in wireless sensor networks with redundant link requirements," in *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, p. 24.

[11] Y. T. Hou, Y. Shi, and H. D. Sherali, "Optimal spectrum sharing for multi-hop software defined radio networks," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*. IEEE, 2007, pp. 1–9.

[12] ——, "Spectrum sharing for multi-hop networking with cognitive radios," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, pp. 146–155, 2008.

[13] N. B. Karahanoğlu, H. Erdoğan, and S. I. Birbil, "A mixed integer linear programming formulation for the sparse recovery problem in compressed sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 5870–5874.

[14] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud-RAN," *Wireless Communications, IEEE Transactions on*, vol. 13, no. 5, pp. 2809–2823, 2014.

[15] S. J. Wright, *Primal-dual interior-point methods*. Siam, 1997.

[16] M. A. Duran and I. E. Grossmann, "An outer-approximation algorithm for a class of mixed-integer nonlinear programs," *Mathematical programming*, vol. 36, no. 3, pp. 307–339, 1986.

[17] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[18] O. Günlük and J. Linderoth, "Perspective reformulations of mixed integer nonlinear programs with indicator variables," *Mathematical programming*, vol. 124, no. 1-2, pp. 183–205, 2010.

[19] J. P. Vielma, I. Dunning, J. Huchette, and M. Lubin, "Extended formulations in mixed integer conic quadratic programming," *arXiv preprint arXiv:1505.07857*, 2015.

[20] T. Berthold, "Heuristic algorithms in global MINLP solvers," Ph.D. dissertation, Technische Universität Berlin, 2014.

[21] F. Messine, "Deterministic global optimization using interval constraint propagation techniques," *RAIRO-Operations Research*, vol. 38, no. 04, pp. 277–293, 2004.

[22] E. D. Andersen and K. D. Andersen, "Presolving in linear programming," *Mathematical Programming*, vol. 71, no. 2, pp. 221–245, 1995.

Authorized licensed use limited to: Technische Universitaet Berlin. Downloaded on October 30,2020 at 08:49:01 UTC from IEEE Xplore. Restrictions apply.

[23] H. S. Ryoo and N. V. Sahinidis, "A branch-and-reduce approach to global optimization," *Journal of Global Optimization*, vol. 8, no. 2, pp. 107–138, 1996.

Authorized licensed use limited to: Technische Universitaet Berlin. Downloaded on October 30,2020 at 08:49:01 UTC from IEEE Xplore. Restrictions apply.