Nonconvex Optimization: from Global Optimality Conditions to Numerical Methods

Alexander Strekalovskiy[1]

[1] *Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Lermontov St., 134, Irkutsk, 664033, Russia*

strekal@icc.ru

Abstract. The paper addresses the nonconvex nonsmooth optimization problem with the cost function and equality and inequality constraints given by d.c. functions. The original problem is reduced to a problem without constraints with the help of the exact penalization theory. After that, the penalized problem is represented as a d.c. minimization problem without constraints, for which the new mathematical tools under the form of global optimality conditions (GOCs) are developed. The GOCs reduce the nonconvex problem in question to a family of convex (linearized with respect to the basic nonconvexities) problems. On the base of the proposed theory we develop numerical methods of local and global search for the problem in question.

INTRODUCTION

The paper addresses the following optimization problem

$$(\mathcal{P}): \qquad \left.\begin{array}{c} f_0(x) := g_0(x) - h_0(x) \downarrow \min\limits_x, \quad x \in S, \\ f_i(x) := g_i(x) - h_i(x) \le 0, \quad i \in I = \{1, \dots, m\}, \\ f_j(x) := g_j(x) - h_j(x) = 0, \quad j \in \mathcal{E} = \{m+1, \dots, l\}, \end{array}\right\}$$

where the functions $g_i(\cdot), h_i(\cdot)$ are convex and finite on $\boldsymbol{R}^n$, so that the functions $f_i(\cdot)$, $i \in \{0\} \cup I \cup \mathcal{E}$, are the d.c. functions [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Recall that any continuous functions can be approximated by a d.c. function at any desirable accuracy. Hence, any optimization problem of type $(\mathcal{P})$ with continuous functions $f_i(\cdot)$, $i \in \{0\} \cup I \cup \mathcal{E}$, can be approximated by correcponding problem $(\mathcal{P})$ with d.c. functions at any desirable accuracy. Besides, the set $S \subset \boldsymbol{R}^n$ is supposed to be closed and convex.

For instance, when all the functions $f_i(\cdot)$ are quadratic $q_i(x) := \langle x, Q_i x\rangle + \langle a_i, x\rangle + \alpha_i$, $i \in \{0\} \cup I \cup \mathcal{E}$, with indefinite matrices $Q_i \in \boldsymbol{R}^{n\times n}$, then we fall into the case, when well-known modern computational softwares (CPlex, Xpres, Gurobi, etc.) often fail to find ever a local solution to $(\mathcal{P})$ or a feasible vector. Therefore, there exists a need of new tools, robust and effective, capable to provide not only a local solution, but allowing to escape local pitfalls and, even, to rich a global solution. Besides, new methods should avoid the problem of so-called the "dimension curse", when the volume of computational efforts, for example, the solving time, increase exponentially together with the dimensions of the problem in question.

In what follows, we suppose that the optimal value $\mathcal{V}(\mathcal{P})$ of problem $(\mathcal{P})$ is supposed to be finite:

$$\mathcal{V}(\mathcal{P}) := \inf(f_0, \mathcal{F}) := \inf_x \{f_0(x) \mid x \in \mathcal{F})\} > -\infty,$$

where the feasible set

$$\mathcal{F} := \{x \in S \mid f_i(x) \le 0, \; i \in I, \; f_j(x) = 0, \; j \in \mathcal{E}\},$$

and the solutions set

$$Sol(\mathcal{P}) := \{z \in \mathcal{F} \mid f_0(z) = \mathcal{V}(\mathcal{P})\}$$

are supposed to be non-empty.

GLOBAL OPTIMALITY CONDITIONS

Let introduce the following penalty function

$$W(x) := \max\{0, f_1(x), \ldots, f_m(x)\} + \sum_{j\in\mathcal{E}} \left|f_j(x)\right|, \tag{1}$$

and, along with problem $(\mathcal{P})$, consider the penalized problem without equality and inequality constraints

$$(\mathcal{P}_\sigma): \qquad \theta_\sigma(x) \downarrow \min_x, \quad x \in S, \tag{2}$$

where $\sigma \geq 0$ is a penalty parameter, and $\theta_\sigma(x) = f_0(x) + \sigma W(x)$, $x \in S$.

It is well-known that, if $z \in Sol(\mathcal{P}_\sigma)$, and $W(z) = 0$, i.e. $z \in \mathcal{F}$, then $z \in Sol(\mathcal{P})$ [1, 2, 3, 12, 13, 14]. Notice that the inverse assertion, generally, does not hold. Hence, the crucial feature of the exact penalty (EP) theory is the existence of a threshold value $\sigma_* > 0$ of the penalty parameter $\sigma \geq 0$, for which $Sol(\mathcal{P}) = Sol(\mathcal{P}_\sigma)$ $\forall \sigma \geq \sigma_*$.

In other words, the existence of the threshold σ_* of the exact penalty parameter implies the existence of possibility to solve only a single unconstrained problem instead of solving a sequence of $(\mathcal{P}_{\sigma_k})$ with $\sigma_k \to \infty$ [2, 15, 16, 17].

Moreover, under various Constraint Qualification (CQ) conditions (MFCQ etc.), the error bound properties etc [18], one can prove the existence of the threshold penalty $\sigma_* \geq 0$ for a local solution as well as for a global one.

In what follows, we assume that some regularity conditions, ensuring the existence of such a threshold value σ_*, are fulfilled. Now let us show that the cost function $\theta_\sigma(\cdot)$ is a d.c. function.

Indeed, it can be readily seen that

$$\theta_\sigma(x) \overset{\triangle}{=} f_0(x) + \sigma \max\{0, f_i(x),\ i \in I\} + \sigma \sum_{i\in\mathcal{E}} |f_i(x)| = G_\sigma(x) - H_\sigma(x), \tag{3}$$

$$H_\sigma(x) := h_0(x) + \sigma\Big[\sum_{i\in I} h_i(x) + \sum_{j\in\mathcal{E}} (g_j(x) + h_j(x))\Big], \tag{4}$$

$$\begin{gathered} G_\sigma(x) := \theta_\sigma(x) + H_\sigma(x) = \\ = g_0(x) + \sigma \max\left\{\sum_{i\in I} h_i(x); \Big[g_p(x) + \sum_{j\in I}^{j\neq p} h_j(x)\Big],\ p \in I\right\} + \\ +2\sigma \sum_{j\in\mathcal{E}} \max\{g_j(x); h_j(x)\}. \end{gathered} \tag{5}$$

It is easy to see that $G_\sigma(\cdot)$ and $H_\sigma(\cdot)$ are both convex functions, so that $\theta_\sigma(\cdot)$ turns out to be a d.c. function, as claimed. Since for a feasible in $(\mathcal{P})$ point z one has $W(z) = 0$, and if $\zeta := f_0(z)$, we have

$$\theta_\sigma(z) = f_0(z) + \sigma W(z) = f_0(z) = \zeta. \tag{6}$$

Necessary optimality conditions can be represented as follows.

Theorem 1 *Let a feasible point $z \in \mathcal{F}$, $\zeta := f_0(z)$, be a solution to problem $(\mathcal{P})$, and $\sigma \geq \sigma_* > 0$, where $Sol(\mathcal{P}) = Sol(\mathcal{P}_\sigma)$ $\forall \sigma \geq \sigma_*$.*

Then, for every pair $(y, \beta) \in \boldsymbol{R}^n \times \boldsymbol{R}$ such that

$$H_\sigma(y) = \beta - \zeta, \tag{7}$$

the following inequality holds

$$G_\sigma(x) - \beta \geq \langle H'_\sigma(y), x - y\rangle \quad \forall x \in S, \tag{8}$$

for every subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the convex function $H_\sigma(\cdot)$ at the point y.

It is worth noting that Theorem 1 reduce the solution of the nonconvex problem $(\mathcal{P}_\sigma)$ to an investigation of the family of convex (linearized) problems

$$(\mathcal{P}_\sigma L(y)): \qquad \Phi_{\sigma y}(x) := G_\sigma(x) - \langle H'_\sigma(y), x\rangle \downarrow \min_x, \quad x \in S, \tag{9}$$

depending on a pair $(y,\beta) \in \boldsymbol{R}^{n+1}$, satisfying the equation (7).

Hence, the verification of the principal inequality (8) can be carried out by solving the linearized problems $(\mathcal{P}_\sigma L(y))$ and varying the parameters (y,β) satisfying (7). Besides, thanks to convexity of $H_\sigma(\cdot)$, it can be easily shown that, when the inequality (8) is violated by a triple (y,β,u) and a subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$, then it yields that $\theta_\sigma(z) > \theta_\sigma(u)$, $z \in \mathcal{F}$, $u \in S$, so that z is not a solution to $(\mathcal{P}_\sigma)$ as well as to $(\mathcal{P})$. Moreover, if $u \in \mathcal{F}$, then $f_0(u) < f_o(z)$, so that in $(\mathcal{P})$ the vector u is better in $(\mathcal{P})$, that $z \in \mathcal{F}$.

The next result gives an answer to the natural question on the existence of such a triple (y,β,u).

Theorem 2 *Let a feasible in problem $(\mathcal{P})$ point z is not an ε-solution to $(\mathcal{P})$, i.e.*

$$\inf(f_0,\mathcal{F}) + \varepsilon = \mathcal{V}(\mathcal{P}) + \varepsilon < \zeta := f_0(z). \tag{10}$$

In addition, let a vector $v \in \boldsymbol{R}^n$ satisfy the following inequality

$$(\mathcal{H}): \qquad f_0(v) > \zeta - \varepsilon. \tag{11}$$

Then, for any penalty parameter $\sigma > 0$ there exists a triple (y,β,u), $(y,\beta) \in \boldsymbol{R}^{n+1}$, such that for every subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the function $H_\sigma(\cdot)$ at the point y the following conditions hold

$$\left.\begin{array}{ll} (a)\ \ H_\sigma(y) = \beta - \zeta + \varepsilon; & (b)\ \ G_\sigma(y) \le \beta, \\ (c)\ \ G_\sigma(u) - \beta < \langle H'_\sigma(y), u - y\rangle. & \end{array}\right\} \tag{12}$$

Notice, that Theorems 1 and 2 generalize the corresponding results from [19] respectively for the nonsmooth data and the presence of d.c. equalities. Besides, one has to point out that the penalty parameter $\sigma \ge 0$ plays different roles in Theorems 1 and 2. In Theorem 1 the penalty parameter $\sigma \ge 0$ should be greater than the threshold value: $\sigma \ge \sigma_* > 0$, which provides $Sol(\mathcal{P}) = Sol(\mathcal{P}_\sigma)$.

Meanwhile, in Theorem 2 the value σ can be arbitrary, but should remain positive, $\sigma > 0$. Regardless of that, one can find a triple (y,β,u) satisfying (12), i.e. violating the principal inequality (8) of Theorem 1, which allows to improve the current value $\theta_\sigma(z)$.

All this enables us to construct numerical methods of local and global searches in the original problem $(\mathcal{P})$.

Finally, let give the sufficient global optimality conditions.

Theorem 3 *Suppose that for a feasible in problem $(\mathcal{P})$ point z, $\zeta := f_0(z)$, the condition $(\mathcal{H})$–(11) is fulfilled. In addition, let some penalty parameter $\sigma > 0$ be given. Finally, assume that for every pair $(y,\beta) \in \boldsymbol{R}^{n+1}$, satisfying the relation*

$$(a) \quad H_\sigma(y) = \beta - \zeta + \varepsilon, \qquad (b) \quad G_\sigma(y) \le \beta, \tag{13}$$

there is a subgradient $H'_\sigma(y) \in \partial H_\sigma(y)$ of the function $H_\sigma(\cdot)$ at y, such that the following inequality holds

$$G_\sigma(x) - \beta \ge \langle H'_\sigma(y), x - y\rangle \quad \forall x \in S. \tag{14}$$

Then, the point $z \in \mathcal{F}$ turns out to be an ε-global solution to problem $(\mathcal{P}_\sigma)$ as well as to problem $(\mathcal{P})$.

We developed some scheme of Local Search with a procedure of penalty parameter update, solving, for instance, the following auxiliary problem:

$$(\mathcal{AP}_w L_k): \qquad \varphi_k(x) := G_W(x) - \langle y^k_W, x\rangle \downarrow \min_x, \quad x \in S, \tag{15}$$

where

$$G_W(x) := 2\sum_{j\in\mathcal{E}} \max\{g_j(x); h_j(x)\} + \max\{\sum_{p\in\mathcal{I}} h_p(x); [g_i(x) + \sum_{p\in\mathcal{I}}^{p\neq i} h_p(x)], i \in \mathcal{I}\}, \tag{16}$$

$$y^k_W := \sum_{i\in\mathcal{I}} h'_{ik} + \sum_{j\in\mathcal{E}} (g'_{jk} + h'_{jk}), \quad h'_{ik} \in \partial h_i(x^k), \ \ i \in \{0\} \cup \mathcal{I} \cup \mathcal{E}, \quad g'_{jk} \in \partial g_j(x^k), \ \ j \in \mathcal{E}. \tag{17}$$

Moreover, it is clear that the problem (15) is related to the following problem

$$(\mathcal{AP}_w): \qquad W(x) \downarrow \min_x, \quad x \in S. \tag{18}$$

Indeed, it can be readily seen that the penalty function $W(\cdot)$ defined in (1) is nonconvex, more precisely, is d.c. function, since $W(x) = G_W(x) - H_W(x)$, where $G_W(\cdot)$ is defined in (16) and $H_W(\cdot)$ is as follows

$$H_W(x) = \sum_{i \in \mathcal{I}} h_i(x) + \sum_{j \in \mathcal{E}} [g_j(x) + h_j(x)].$$

In addition, the inclusion $y_W^k \in \partial H_W(x^k)$ is obvious, so that the problem $(\mathcal{AP}_w L_k)$ turns out to be the linearized (at the point x^k) problem of problem $(\mathcal{AP}_w)$.

To sum up, it is clear, that not all questions have been answered, but, apparently, Theorems 1, 2, 3 provide the new tools for development of new approaches, methods and algorithms for nonconvex optimization (see [20, 21, 22, 23]).

REFERENCES

[1] J. Nocedal and S. Wright, *Numerical Optimization* (Springer, New York, 2006).

[2] J.-F. Bonnans, J. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical Optimization. Theoretical and Practical Aspects* (Springer-Verlag, Berlin Heidelberg, 2006).

[3] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms* (Springer-Verlag, Berlin, 1993).

[4] H. Tuy, *Convex Analysis and Global Optimization* (Kluwer Academic Publishers, Dordrecht, 1998).

[5] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches* (Springer-Verlag, Berlin, 1996).

[6] R. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[7] J.-B. Hiriart-Urruty, "Generalized differentiability, duality and optimization for problems dealing with difference of convex functions," in *Convexity and Duality in Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 256, edited by J. Ponstein (Springer-Verlag, Berlin, 1985), pp. 37–70.

[8] J.-B. Hiriart-Urruty, *Optimisation et analyse convexe* (Presses Universitaires de France, 1998) [in French].

[9] H. A. Le Thi, "D.C. programming for solving a class of global optimization problems via reformulation by exact penalty," in *Global Optimization and Constraint Satisfaction*, Lecture Notes in Computer Science, Vol. 2861, edited by C. Bliek, C. Jermann, and A. Neumaier (Springer-Verlag, Berlin, 2003), pp. 87–101.

[10] H. A. Le Thi and T. Pham Dinh, Annals of Operations Research **133**, 23–46 (2005).

[11] H. A. Le Thi and T. Pham Dinh, Mathematical Programming **169**, 5–68 (2018).

[12] J. Burke, SIAM Journal on Control and Optimization **29**, 968–998 (1991).

[13] G. Di Pillo, S. Lucidi, and F. Rinaldi, Journal of Global Optimization **54**, 251–260 (2012).

[14] G. Di Pillo and L. Grippo, SIAM Journal on Control and Optimization **27**, 1333–1360 (1989).

[15] R. Byrd, J. Nocedal, and R. Waltz, Optimizationa Methods & Software **23**, 197–213 (2008).

[16] R. Byrd, G. Lopez-Calva, and J. Nocedal, Mathematical Programming, Ser. A **133**, 39–73 (2012).

[17] G. Di Pillo, S. Lucidi, and F. Rinaldi, Journal of Optimization Theory and Applications **164**, 862–882 (2015).

[18] A. Kruger, L. Minchenko, and J. Outrata, Positivity **18**, 171–189 (2014).

[19] A. Strekalovsky, Optimization Letters **DOI: 10.1007/s11590-017-1214-x** (2017).

[20] A. Strekalovsky, "On solving optimization problems with hidden nonconvex structures," in *Optimization in Science and Engineering*, edited by T. Rassias, C. Floudas, and S. Butenko (Springer, New York, 2014), pp. 465–502.

[21] A. Strekalovsky, Journal of Optimization Theory and Applications **173**, 770–792 (2017).

[22] A. Strekalovsky, Applied Mathematics and Computation **255**, 73–83 (2015).

[23] A. Strekalovsky and I. Minarchenko, Applied Mathematical Modelling **58**, 229–244 (2018).