INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCES

Pham Dinh Tao

National Institute for Applied Sciences-Rouen, France

Joint work with Le Thi Hoai An

23 octobre 2013
Plan

1. Outline DC programming & DCA
   - Convex Programming
   - Nonconvex Programming: Nonlinear Optimization – Global Optimization

2. Exact penalty techniques in DC Programming

3. Recent advances in DC programming
   - Improved Convergence of DCA for DC programming with subanalytic data
   - Extended DC Algorithms to general DC programs with DC constraints

4. Exact penalty and Error Bounds in DC programming
   - Exact penalty in concave programming
   - Exact penalty in DC programming via error bounds
   - Error bounds for DC inequality systems
   - Error bounds for concave inequality systems over polyhedral

INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCE
Outline DC programming & DCA
Exact penalty techniques in DC Programming
Recent advances in DC programming
Exact penalty and Error Bounds in DC programming

Convex Programming
Nonconvex Programming : Nonlinear Optimization – Global Optimization

Convex Programming $\rightarrow$ Nonconvex Programming

DC Programming & DCA
Convex Programming

\[ \Gamma_0(\mathbb{R}^n) := \{ f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}, \text{lower semicontinuous proper} \} \]

\((\text{dom} f \neq \emptyset) \text{ convex}\}, \text{a convex cone}\)

\[ \text{dom } \theta := \{ x \in \mathbb{R}^n : \theta(x) < +\infty \}. \]

Convex program : \( f \in \Gamma_0(\mathbb{R}^n) \)

\[ \alpha := \inf \{ f(x) : x \in \mathbb{R}^n \} \quad (P_c) \]
Constrained convex program: \( C \subseteq \mathbb{R}^n \) nonempty closed convex set \( \iff \chi_C \in \Gamma_0(\mathbb{R}^n) \)
\( \chi_C \) indicator function of \( C \)

\( \chi_C(x) := 0 \) if \( x \in C \), \( +\infty \) otherwise

Convex constrained program

\( \alpha := \inf \{ f(x) : x \in C \} \ (P_c) \)

can be rewritten as \( (f + \chi_C \in \Gamma_0(\mathbb{R}^n)) \)

\( \alpha := \inf \{ (f + \chi_C)(x) : x \in \mathbb{R}^n \} \)
In practice solvable $(P_c)$ takes the form: $f, f_i \in \Gamma_0(\mathbb{R}^n), i = 1, ..., m$

$$\alpha := \inf \{ f(x) : f_i(x) \leq 0, i = 1, ..., m \}$$

Convex programming is featured by:
- Local solution = global solution.
- Existence of optimality conditions verifiable $\rightarrow$ Iterative solution algorithms.
- No Lagrangian duality gap.

Extensively investigated from both a theoretical and an algorithmic point of views from the period 1960-1985 and after, with the development of Interior Point Methods and SDP.
(Hoang Tuy, R. Horst, H. Benson, H. Konno, P. Pardalos, Le Dung Muu, Le Thi Hoai An, Nguyen Van Thoai, Phan Thien Thach and Pham Dinh Tao)

Complexity (NP hard in general) : Finding Global solution the Holy Grail quest for mathematicians (optimizers)

Local solution $\neq$ global solution

Non existence of practically verifiable global optimality conditions : no iterative global solution algorithm
Two approaches (two schools) in Nonconvex Programming:

Local Approaches (Nonlinear Optimization, for smooth nonconvex programming) : Iterative algorithms based on local optimality conditions.

Global Approaches (Global Optimization) : consisting, generally, of

Bounding procedure : Computing a sequence of increasing lower bounds $\alpha_k$ for the optimal value $\alpha$ by solving relaxed convex programs and that of decreasing upper bound $\beta_k$. 
Branching procedure: Choosing a subdomain corresponding to the smallest lower bound to subdivide (usually bisect) and deleting those providing lower bounds greater than the current record (the current smallest upper bound).

Repeating until the gap $\beta_k - \alpha_k \leq \epsilon$ to carry out an $\epsilon$-global optimal solution.

Other methods: outer approximation using cuttings plane or sphere,...

Global algorithms are generally too expensive to handle large-scale nonconvex programs.
Local Approaches based on Convex Programming for Smooth/Nonsmooth Nonconvex Programming

http://lita.sciences.univ-metz.fr/~lethi/

- DC Programming and DCA: the backbone of Nonconvex Programming and Global Optimization.
- Key idea (Philosophy/Principle): Extension of Convex Analysis/Programming to Nonconvex Analysis/Programming, broad enough to cover most real-world nonconvex programs, but not too much to continue to use the powerful arsenal of convex analysis/programming.
- Unlike the sum, the difference of convex functions destroys convexity. The feature DC will be advantageously explored and exploited within DC Programming and DCA.
\( DC(\mathbb{R}^n) := \Gamma_0(\mathbb{R}^n) - \Gamma_0(\mathbb{R}^n) \), vector subspace spanned by \( \Gamma_0(\mathbb{R}^n) \)

DC program

\[
(P_{dc}) \alpha = \inf \{ f(x) := g(x) - h(x) : x \in \mathbb{R}^n \}
\]

DC Programming investigates: structure of the vector space \( DC(\mathbb{R}^n) \): it contains almost all realistic objective functions and is closed under all the operations usually considered in optimization.

Real-world nonconvex programs are DC programs.
DC duality : $g \in \Gamma_0(\mathbb{R}^n)$

$$g^*(y) := \sup \{\langle x, y \rangle - g(x) : x \in \mathbb{R}^n\}$$

$$g^* \in \Gamma_0(\mathbb{R}^n), g^{**} = g$$

$$\alpha = \inf \{h^*(y) - g^*(y) : y \in \mathbb{R}^n\} \quad (D_{dc})$$

Perfect symmetry : the dual of $(D_{dc})$ is $(P_{dc})$ itself.
Local optimality : 

For primal DC Program (for dual DC program by similarity) 

\[ \partial h(x^*) \cap \partial g(x^*) \neq \emptyset \] 

(such a point \( x^* \) is called critical point of \( g - h \) or generalized KKT point for \( (P_{dc}) \)), and 

\[ \emptyset \neq \partial h(x^*) \subset \partial g(x^*) . \]  \hspace{1cm} (1)

The condition (1) is also sufficient (for local optimality) in many important classes of DC programs. In particular it is sufficient for the next cases quite often encountered in practice :
In polyhedral DC programs with \( h \) being a polyhedral convex function. In this case, if \( h \) is differentiable at a critical point \( x^* \), then \( x^* \) is actually a local minimizer for \((P_{dc})\). Since a convex function is differentiable everywhere except for a set of measure zero, one can say that a critical point \( x^* \) is almost always a local minimizer for \((P_{dc})\).

In case the function \( f \) is locally convex at \( x^* \). In particular, if \( f \) is convex then \( x^* \) is a (global) optimal solution. Note that, if \( h \) is polyhedral convex, then \( f = g - h \) is locally convex everywhere \( h \) is differentiable.

The transportation of global solutions between \((P_{dc})\) and \((D_{dc})\) is expressed by:

\[
\begin{align*}
[ \bigcup_{y^* \in D} \partial g^*(y^*) ] & \subseteq P, \\
[ \bigcup_{x^* \in P} \partial h(x^*) ] & \subseteq D
\end{align*}
\]

(2)
equalities holds under technical assumptions.
Under technical assumptions, this transportation also holds for local solutions.

Solving a DC program amounts solving its dual and vice-versa.

The complexity of DC programs resides, of course, in the lack of verifiable conditions for global optimality.

**Subdifferential**

θ ∈ Γ₀(ℝⁿ) and x₀ ∈ dom θ, ∂θ(x₀) denotes the subdifferential of θ at x₀, i.e.,

\[
∂θ(x₀) := \{ y ∈ ℝⁿ : \theta(x) ≥ θ(x₀) + \langle x - x₀, y \rangle, \ ∀ x ∈ ℝⁿ \} \tag{3}
\]

The subdifferential ∂θ(x₀) is a closed convex set in ℝⁿ. It generalizes the derivative in the sense that θ is differentiable at x₀ if and only if ∂θ(x₀) is a singleton which is exactly {∇θ(x₀)}. 

Pham Dinh Tao

INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCES
Description of DCA

Two sequences \( \{x^k\} \) and \( \{y^k\} \) of trial solutions of the primal and dual programs respectively,

1) \( \{g(x^k) - h(x^k)\} \) and \( \{h^*(y^k) - g^*(y^k)\} \) are decreasing,

2) \( \{x^k\} \) (resp. \( \{y^k\} \)) converges to a primal feasible solution \( \tilde{x} \) (resp. a dual feasible solution \( \tilde{y} \)) satisfying local optimality conditions and

\[
\tilde{x} \in \partial g^*(\tilde{y}), \quad \tilde{y} \in \partial h(\tilde{x}).
\] (4)

The sequences \( \{x^k\} \) and \( \{y^k\} \) are determined in the way that \( x^{k+1} \) (resp. \( y^{k+1} \)) is a solution to the convex program \( (P_k) \) (resp. \( (D_{k+1}) \)) defined by (\( x^0 \in \text{dom} \partial h \) being a given initial point and \( y^0 \in \partial h(x^0) \) being chosen)

\[
(P_k) \quad \inf \{ (g - h)_k(x) := g(x) - [h(x^k) + \langle x - x^k, y^k \rangle] : x \in \mathbb{R}^n \},
\]

\[
(D_{k+1}) \quad \inf \{ (h^* - g^*)_k(y) := h^*(y) - [g^*(y^k) + \langle y - y^k, x^{k+1} \rangle] : y \in \mathbb{R}^n \}.
\] (5)
DCA deals with the convex DC components $g$ and $h$ but not with the DC function $f$ itself.

DCA is one of the rare algorithms for nonconvex nonsmooth programming.

DC function $f$ has infinitely many DC decompositions which have crucial implications for the qualities (speed of convergence, robustness, efficiency, globality of computed solutions,...) of DCA.

We have

$$f_k(x) = f(x) + h(x) - [h(x^k) + \langle x - x^k, y^k \rangle] \geq f(x) \quad \forall x$$

$$\tag{6} (h^* - g^*)_k(y) = (h^* - g^*)(y) + g^*(y) - [g^*(y^k) + \langle y - y^k, x^{k+1} \rangle] \geq (h^* - g^*)(y) \quad \forall y$$

It is clear that the more tightly the convex function $(g - h)_k$ approximates $f$ the more efficient the resulting DCA will be. Consider now another DC decomposition of $f$ with $\varphi \in \Gamma_0(\mathbb{R}^n)$.
and compare the two convex approximation \((g - h)_k(x)\) and \([(g + \varphi) - (h + \varphi)]_k\). We have \((y^k \in \partial g(x^k), z^k \in \partial \varphi(x^k))\)

\[
[(g + \varphi) - (h + \varphi)]_k(x) = (g + \varphi)(x) - [(g + \varphi)(x^k) + \langle x - x^k, y^k + z^k \rangle]
\]

\[
= g(x) - [g(x^k) + \langle x - x^k, y^k \rangle] + \varphi(x)
- [\varphi(x^k) + \langle x - x^k, z^k \rangle]
\]

\[
= (g - h)_k(x) + \varphi(x) - [\varphi(x^k) + \langle x - x^k, z^k \rangle]
\]

\[
\geq (g - h)_k(x).
\]

Shift procedure for constructing new DC decompositions is not relevant. (see later the proximal DC decomposition).
The choice of *optimal* DC decompositions is still open. This depends strongly on the very specific structure of the problem being considered. In order to tackle the large-scale setting, one tries in practice to choose $g$ and $h$ such that sequences $\{x^k\}$ and $\{y^k\}$ can be easily calculated, *i.e.*, either they are in an explicit form or their computations are inexpensive.
DCA’s Convergence

\( \{x^k\} \subset C \) convex, \( \{y^k\} \subset D \) convex, \( \rho(g, C) \) (or \( \rho(g) \) if \( C = \mathbb{R}^n \)) the modulus of strong convexity of \( g \) on \( C \) given by:

\[
\rho(g, C) = \sup \{ \rho \geq 0 : g - (\rho/2)\|\cdot\|^2 \text{ be convex on } C \}.
\]

DCA’s convergence properties:

DCA is a descent method without linesearch (linesearch is actually dealt within DC decomposition) which enjoys the following properties:

i) The sequences \( \{g(x^k) - h(x^k)\} \) and \( \{h^*(y^k) - g^*(y^k)\} \) are decreasing and

- \( g(x^{k+1}) - h(x^{k+1}) = g(x^k) - h(x^k) \) iff \( y^k \in \partial g(x^k) \cap \partial h(x^k), y^k \in \partial g(x^{k+1}) \cap \partial h(x^{k+1}) \) and

\[
[\rho(g, C) + \rho(h, C)]\|x^{k+1} - x^k\| = 0.
\]

Moreover if \( g \) or \( h \) are strictly convex on \( C \) then \( x^k = x^{k+1} \).

In such a case DCA terminates at the \( k^{th} \) iteration (finite convergence of DCA).
• $h^*(y^{k+1}) - g^*(y^{k+1}) = h^*(y^k) - g^*(y^k)$ iff
  $x^{k+1} \in \partial g^*(y^k) \cap \partial h^*(y^k)$, $x^{k+1} \in \partial g^*(y^{k+1}) \cap \partial h^*(y^{k+1})$ and
  $[\rho(g^*, D) + \rho(h^*, D)]\|y^{k+1} - y^k\| = 0$. Moreover if $g^*$ or $h^*$ are strictly convex on $D$,
  then $y^{k+1} = y^k$.
  In such a case DCA terminates at the $k^{th}$ iteration (finite convergence of DCA).

ii) If $\rho(g, C) + \rho(h, C) > 0$ (resp. $\rho(g^*, D) + \rho(h^*, D) > 0$) then the series
  $\{\|x^{k+1} - x^k\|^2\}$ (resp. $\{\|y^{k+1} - y^k\|^2\}$) converges.

iii) If the optimal value $\alpha$ of problem $(P_{dc})$ is finite and the infinite
  sequences $\{x^k\}$ and $\{y^k\}$ are bounded then every limit point $\tilde{x}$ (resp.
  $\tilde{y}$) of the sequence $\{x^k\}$ (resp. $\{y^k\}$) is a critical point of $g - h$ (resp.
  $h^* - g^*$).

iv) DCA has a linear convergence for general DC programs.

v) DCA has a finite convergence for polyhedral DC programs.

For suitable DC decompositions DCA generates almost all standard algorithms in Convex and Nonconvex Programming.
An illustrative example

General nonconvex quadratic program

\[
\min \left\{ f(x) := \frac{1}{2} x^T Q x + q^T x : x \in K \right\},
\]

where \( Q \) is an \( n \times n \) symmetric matrix, \( q \in \mathbb{R}^n \), and \( K \subset \mathbb{R}^n \) is a nonempty polyhedral convex set.

\( f \) is a DC function with infinitely many DC decompositions. The following is particularly interesting because the resulting DC component are quadratic functions too:

\[
f(x) = \varphi(x) - \psi(x),
\]

\[
\varphi(x) := \frac{1}{2} x^T Q_1 x + q^T x, \quad \psi(x) := \frac{1}{2} x^T Q_2 x,
\]

where \( Q_1 \) and \( Q_2 \) are two \( n \times n \) positive semidefinite symmetric matrices such that

\[
Q = Q_1 - Q_2.
\]
This leads to the equivalent DC program
\[
\inf \{ g(x) - h(x) : x \in \mathbb{R}^n \},
\]
where
\[
g(x) := \varphi + \chi_K, \quad h := \psi
\]
The matrices $Q_1$ and $Q_2$ can have one of the following forms:
(i) Spectral DC decomposition: $Q_1 := PD^+ P^T$, $Q_2 := PD^- P^T$ where $P$ is an orthogonal matrix composed of eigenvectors of $Q$ related to its eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$ such that $Q = PDP^T$, with $D := \text{diag}(\lambda_1, \ldots, \lambda_n)$ being the diagonal matrix with diagonal entries $\lambda_i$. The matrices $D^+$ and $D^-$ are given by:
\[
D^+ := \text{diag}(\lambda_1^+, \ldots, \lambda_n^+), \quad D^- := \text{diag}(\lambda_1^-, \ldots, \lambda_n^-)
\]
(Here $\lambda_i^+ = \max\{0, \lambda_i\}$, $\lambda_i^- = \max\{0, -\lambda_i\}$).
(ii) Proximal DC decomposition: $Q_1 := Q + \rho l$, $Q_2 := \rho l$, with $\rho \geq 0$ such that $Q + \rho l$ is positive semidefinite, i.e., $\rho \geq -\lambda_1(Q)$.

(iii) Projection DC decomposition: $Q_1 := \rho l$, $Q_2 := \rho l - Q$, with $\rho \geq 0$ such that $\rho l - Q$ is positive semidefinite, i.e., $\rho \geq \lambda_n(Q)$.

It is clear that the spectral DC decomposition is expensive, especially in the large-scale setting, while there are inexpensive algorithms for computing the extreme eigenvalues $\lambda_1(Q)$ and $\lambda_n(Q)$. 
Let us describe now the DCA applied to the DC program (12) \((x^0\) being a given initial point) :

\[
y^k = Q_2 x^k; \quad x^{k+1} \in \arg \min \left\{ \frac{1}{2} x^T Q_1 x + q^T x - x^T Q_2 x^k : x \in K \right\}.
\] (13)

Note that the convex quadratic program

\[
\inf \left\{ \frac{1}{2} x^T Q_1 x + q^T x - x^T Q_2 x^k : x \in K \right\}
\]

is equivalent to \((P_k)\) above and can also be written as

\[
\min \left\{ \frac{1}{2} x^T Q x + q^T x + \frac{1}{2} (x - x^k)^T Q_2 (x - x^k) : x \in K \right\}.
\] (14)
We recognize here that, with the quadratic DC decomposition (10), the resulting DCA is an extended proximal-like method for nonconvex quadratic programming (applied to (9)). First result on proximal point algorithm for nonconvex programming. See below for extended result to non quadratic DC functions. The best choice of $\rho \geq -\lambda_1(Q)$ will be the smallest value $-\lambda_1(Q)$, accordingly with (8).

Consider the general DC program

$$\alpha = \inf \{ f_0(x) := g_0(x) - h_0(x) : x \in C, \ p(x) := g(x) - h(x) \leq 0 \} \quad (P_{dcg}) \quad (15)$$

where $C$ is a nonempty closed convex set in $\mathbb{R}^n$, $g, h, g_0, h_0 \in \Gamma_0(\mathbb{R}^n)$ and its feasible set $S = \{ x \in C : g(x) - h(x) \leq 0 \}$ assumed to be nonempty. The usual penalized DC program is the DC program defined by
\[ \alpha(\tau) = \inf \{ f_0(x) + \tau p^+(x) : x \in C \} \quad (P_{dcg}^\tau) \quad (16) \]

We will summarize main exact penalty results with/without error bounds of nonconvex constraint sets \( S \), which establish the equivalence of \( (P_{dcg}^\tau) \) and its penalized \( (P_{dcg}^\tau_{dcg}) \). These results are based on the concavity of objective/constraint functions and local error bounds [Le Thi, Pham Dinh and Huynh Van, 2012].

More precisely, by using the DC decomposition (ii) with \( \rho > 0 \), (14) becomes

\[
\min \left\{ \frac{1}{2} x^T Q x + q^T x + \frac{\rho}{2} \| x - x_k \|^2 : x \in K \right\}, \quad (17)
\]
and then the corresponding DCA is the extended standard proximal algorithm for nonconvex programming. Lastly, the DC decomposition (iii) leads to the following DCA:

\[ y^k = (\rho I - Q)x^k; \quad x^{k+1} \in \arg \min \left\{ \frac{1}{2} \left\| x - \frac{1}{\rho}(y^k - q) \right\|^2 : x \in K \right\}. \] (18)
In other words, the primal sequence \( \{x^k\} \) generated by this algorithm is given by

\[
x^{k+1} := P_K \left( x^k - \frac{1}{\rho} (Qx^k + q) \right),
\]

where \( P_K \) denotes the projection mapping on \( K \). One recognizes, in (19), the gradient projection algorithm of Goldstein-Levitin-Polyak in convex programming, i.e., when the objective function \( f \) is convex.

It is worth noting that the above DC decompositions can be extended to functions \( f \) twice continuously differentiable on an open convex set \( O \) containing \( K \) under the condition: There is \( \rho > 0 \) such that \( \frac{\rho}{2} \| \cdot \|^2 + f \) (resp. \( \frac{\rho}{2} \| \cdot \|^2 - f \)) is convex on \( K \). Clearly, these conditions are satisfied when \( K \) is bounded.

Note that, the DCA via (iii) can be viewed as a proximal regularization of gradient method for nonconvex programming.
Definition 1

(i) A subset $C$ of $\mathbb{R}^n$ is said to be semianalytic if each point of $\mathbb{R}^n$ admits a neighborhood $V$ such that $C \cap V$ is of the following form:

$$C \cap V = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \{ x \in V : f_{ij}(x) = 0, g_{ij}(x) > 0 \},$$

where $f_{ij}, g_{ij} : V \to \mathbb{R}$ $(1 \leq i \leq p, 1 \leq j \leq q)$ are real-analytic functions.

(ii) A subset $C$ of $\mathbb{R}^n$ is called subanalytic if each point of $\mathbb{R}^n$ admits a neighborhood $V$ such that

$$C \cap V = \{ x \in \mathbb{R}^n : \exists y \in \mathbb{R}^m, (x, y) \in B \},$$

where $B$ is a bounded seminalytic subset of $\mathbb{R}^n \times \mathbb{R}^m$ with $m \geq 1$.

(iii) A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be subanalytic if its graph $\text{gph } f$ is a subanalytic subset of $\mathbb{R}^n \times \mathbb{R}$. 

Pham Dinh Tao

National Institute for Applied Sciences-Rouen, France
The class of subanalytic sets (resp. functions) contains all analytic sets (resp. functions). The elementary properties of subanalytic sets and subanalytic functions (see, e.g., Bierstone and Milman; Lojasiewicz; Shiota):
- Subanalytic sets are closed under locally finite union and intersection (A collection of sets $C$ is locally finite if any compact set intersects only finitely many sets in $C$). The complement of a subanalytic set is subanalytic.
- The closure, the interior, the boundary of a subanalytic set are subanalytic.
- A closed $C \subseteq \mathbb{R}^n$ is subanalytic if and only if its indicator function $\chi_C$, defined by $\chi_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise, is subanalytic.
- Given a subanalytic set $C$, the distance function $d_C(x) := \inf_{z \in C} \|x - z\|$ is a subanalytic function.
- Let $f, g : X \to \mathbb{R}$ be continuous subanalytic functions, where $X \subseteq \mathbb{R}^n$ is a subanalytic set. Then the sum $f + g$ is subanalytic if $f$ maps bounded sets into bounded sets, or if both functions $f, g$ are bounded from below.
- Let $X \subseteq \mathbb{R}^n, T \subseteq \mathbb{R}^m$ be subanalytic sets, where $T$ is compact. If $f : X \times T \to \mathbb{R}$ is a continuous subanalytic function, then $g(x) := \min_{t \in T} f(x, t)$ is continuous subanalytic.
**Proposition** If \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{ +\infty \} \) is a lower semicontinuous subanalytic strongly convex function then its conjugate \( f^* \) is a \( C^{1,1} \) (the class of functions whose derivative is Lipschitz) subanalytic convex function.

\( S(\mathbb{R}^n) \) denotes the set of lower semicontinuous functions \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{ +\infty \} \).
For a function $f \in S(\mathbb{R}^n)$, the *Fréchet subdifferential* of $f$ at $x \in \text{dom } f$ is defined by

$$
\partial^F f(x) = \left\{ x^* \in \mathbb{R}^n : \liminf_{h \to 0} \frac{f(x + h) - f(x) - \langle x^*, h \rangle}{\|h\|} \geq 0 \right\}.
$$

For $x \notin \text{dom } f$, we set $\partial f(x) = \emptyset$. A point $x_0 \in \mathbb{R}^n$ is called a (Fréchet) *critical* point for the function $f$, if $0 \in \partial^F f(x_0)$. 
When $f$ is a convex function, then $\partial^F$ coincides with the subdifferential in the sense of Convex Analysis. Moreover, if $f$ is a DC function, i.e., $f := g - h$, where $g, h$ is convex functions, then

$$\partial^F f(x) \subseteq \partial g(x) - \partial h(x)$$

wherever $h$ is continuous at $x$. Moreover, if $h$ is differentiable at $x$, then one has the equality :

$$\partial^F f(x) = \partial g(x) - \nabla h(x).$$

Let us recall the nonsmooth version of the Lojasiewicz inequality established by Bolte-Daniliidis-Lewis, which is needed in the convergence analysis of DCA in the sequel.
(Theorem 3.1, Bolte et al.) Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) be a subanalytic function such that its domain \( \text{Dom} \, f \) is closed and \( f \big|_{\text{dom} \, f} \) is continuous and let \( x_0 \) is a Fréchet critical point of \( f \). Then there exist \( \theta \in [0, 1) \), \( L > 0 \) and a neighborhood \( V \) of \( x_0 \) such that the following inequality holds.

\[
|f(x) - f(x_0)|^\theta \leq L \|x^*\| \quad \text{for all } x \in V, \, x^* \in \partial^F f(x),
\]

where a convention \( 0^0 = 1 \) is used.

The number \( \theta \) in the theorem is called a Lojasiewicz exponent of the critical point \( x_0 \).
Theorem 3

Let us consider DC program \((P_{dc})\) with \(\alpha \in \mathbb{R}\). Suppose that the sequences \(\{x^k\}, \{y^k\}\) are defined by the DCA.

(i) Suppose that the DC function \(f := g - h\) is subanalytic such that \(\text{dom } f\) is closed; \(f|_{\text{dom } f}\) is continuous and that \(g\) or \(h\) is differentiable on \(\text{dom } h\) or \(\text{dom } g\), respectively with locally Lipschitz derivative. Assume that \(\rho(g) + \rho(h) > 0\), where \(\rho(g), \rho(h)\) are modulus of the strong convexity of \(g, h\), respectively. If either the sequence \(\{x^k\}\) or \(\{y^k\}\) is bounded then \(\{x^k\}\) and \(\{y^k\}\) are convergent to critical points of \((P_{dc})\) and \((D_{dc})\), respectively.

(ii) Similarly, for the dual DC program \((D_{dc})\), suppose that \(h^* - g^*\) is subanalytic such that \(\text{dom } (h^* - g^*)\) is closed; \((h^* - g^*)|_{\text{dom } (h^* - g^*)}\) is continuous and that \(g^*\) or \(h^*\) is differentiable on \(\text{dom } g^*\) or \(\text{dom } h^*\), respectively with locally Lipschitz derivative. If \(\rho(g^*) + \rho(h^*) > 0\) and either the sequence \(\{x^k\}\) or \(\{y^k\}\) is bounded then \(\{x^k\}\) and \(\{y^k\}\) are convergent to critical points of \((P_{dc})\) and \((D_{dc})\), respectively.
Corollary 4

Suppose that $g - h$ and $h^* - g^*$ are subanalytic functions with closed domain such that $(g - h) |_{\text{dom}(g-h)}$ and $(h^* - g^*) |_{\text{dom}(h^*-g^*)}$ are continuous. Assume that $\rho(g) + \rho(h) > 0$ as well as $\rho(g^*) + \rho(h^*) > 0$. If either the sequence $\{x^k\}$ or $\{y^k\}$ is bounded then these sequences converge to critical points of $(\mathcal{P})$ and $(\mathcal{D})$, respectively.
Theorem 5

Suppose that the assumptions of Theorem 3 (i) are satisfied. Let $x^\infty$ be the limit point of $\{x^k\}$ with a Lojasiewicz exponent $\theta \in [0, 1)$. Then there exists constants $\tau_1, \tau_2 > 0$ such that

$$
\|x^k - x^\infty\| \leq \sum_{j=k}^{\infty} \|x^j - x^{j+1}\| \leq \tau_1 \|x^k - x^{k-1}\| + \tau_2 \|x^k - x^{k-1}\|^{\frac{1-\theta}{\theta}}, \; k = 1, 2, \ldots \tag{20}
$$

As a result, one has

- If $\theta \in (1/2, 1)$ then $\|x^k - x^\infty\| \leq ck^{\frac{1-\theta}{2\theta}}$ for some $c > 0$.
- If $\theta \in (0, 1/2]$ then $\|x^k - x^\infty\| \leq cq^k$ for some $c > 0$; $q \in (0, 1)$.
- If $\theta = 0$ then $\{x^k\}$ is convergent in a finite number of steps.
In general, the Lojasiewicz exponent is unknown. But in several special cases, one can determine the Lojasiewicz exponent. Let us consider the case of trust-region subproblems, that is, the problems of minimizing a (nonconvex) quadratic function over an Euclidean ball in $\mathbb{R}^n$.

$$\min \left\{ \frac{1}{2} x^T Q x + \langle b, x \rangle : \ x \in \mathbb{R}^n , \ |x| \leq R \right\} ,$$

or equivalently,

$$\min \{ f(x) := \frac{1}{2} x^T Q x + \langle b, x \rangle + \chi_C(x) : \ x \in \mathbb{R}^n \} , \quad (21)$$

where, $Q$ is a $n \times n$ real symmetric matrix, $b \in \mathbb{R}^n$, $R$ is a positive scalar and $C := \{ x \in \mathbb{R}^n : |x| \leq R \}; \ \chi_C(x)$ stands for the indicator function of the set $C$.

The following theorem gives a Lojasiewicz exponent of some critical point of the objective function $f$ under a suitable assumption.
**Lemma 6**

Let $A$ be a $n \times n$ real symmetric matrix. Then there exists a constant $M > 0$ such that the following inequality holds

$$|x^T Ax|^{1/2} \leq M \|Ax\| \quad \text{for all } x \in \mathbb{R}^n.$$ 

**Theorem 7**

Suppose $x^0 \in C$ is a critical point of problem (21), i.e., there exists $\lambda_0 \geq 0$ such that

$$Qx^0 + b + \lambda_0 x^0 = 0; \quad \lambda_0(\|x^0\| - R) = 0. \quad (22)$$

In addition, assume that either $\|x^0\| < R$ or $\|x^0\| = R$ and

$$\langle v, Qv + \lambda_0 v \rangle > 0 \quad \text{for all } v \in \mathbb{R}^n \text{ with } \langle x^0, v \rangle = 0, \|v\| = 1. \quad (23)$$

The Lojasiewicz exponent of the critical point $x^0$ is $1/2$. 

---

**Improved Convergence of DCA for DC programming with subanalytic data**

**Extended DC Algorithms to general DC programs with DC constraints**
Introduction

Consider the following nonconvex optimization problem

\[(GP_{dc}) \quad \min f_0(x) \quad \text{subject to} \quad \begin{cases} f_i(x) \leq 0, & i = 1, \ldots, m; \\ x \in C, \end{cases} \]

where \( C \subseteq \mathbb{R}^n \) is a nonempty closed convex set; \( f, f_i : \mathbb{R}^n \to \mathbb{R} \) \((i = 0, 1, \ldots, s)\) are DC functions. Problem \((P_{dc})\) is a general DC program with DC constraints.

To apply the DCA to \((GP_{dc})\), a penalization is used for DC constraints to equivalently transform this problem to a standard DC program without DC constraint. A penalty function largely used is \( l_\infty \)–penalty function of the form

\[ p^+(x) := \max\{0, f_i(x) : i = 1, \ldots, m\}, \quad x \in \mathbb{R}^n. \]
However, in the computational point of view, an inconvenience of this exact penalty method is that the penalty parameter is generally unknown. Moreover, there are practical optimization problems for which the exact penalization is not satisfied.

The purpose of this paper is to develop the DCA for solving general DC program \((GP_{dc})\) by using a penalty technique with updated parameter. The global convergence of the method proposed is investigated.
Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a locally Lipschitz function at a given $x \in \mathbb{R}^n$. The Clarke directional derivative and the Clarke subdifferential of $f$ at $x$ is given by the following formulas.

$$f^{\uparrow}(x, v) := \limsup_{(t, y) \to (0^+, x)} \frac{f(y + tv) - f(y)}{t}.$$ $$\partial^{\uparrow} f(x) := \{ x^* \in \mathbb{R}^n : \langle x^*, v \rangle \leq f^{\uparrow}(x, v) \forall v \in \mathbb{R}^n \}.$$ 

If $f$ is continuously differentiable at $x$ then $\partial^{\uparrow} f(x) = \nabla f(x)$ (the Fréchet derivative of $f$ at $x$). When $f$ is a convex function, then $\partial f(x)$ coincides with the subdifferential in the sense of Convex Analysis, i.e.,

$$\partial f(x) = \{ y \in \mathbb{R}^n : \langle y, d \rangle \leq f(x + d) - f(x), \ \forall d \in \mathbb{R}^n \}.$$
We list the following calculus rules for the Clarke subdifferential which are needed thereafter.

1. For given two locally Lipschitz functions $f, g$ at a given $x \in \mathbb{R}^n$, one has

\[
\partial(-f)(x) = -\partial f(x); \quad \partial(f + g)(x) \subseteq \partial f(x) + \partial g(x),
\]

and, the equality in the latter inclusion holds if $f$ is continuously differentiable at $x$.

2. For given functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ($i = 1, \ldots, m$), let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be the max-type function defined by

\[
f(x) := \max \{f_i(x) : i = 1, \ldots, m\}, \quad x \in \mathbb{R}^n.
\]
Then one has

$$ \partial f(x) \subseteq \left\{ \sum_{i=1}^{m} \lambda_i \partial f_i(x) : \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1; \lambda_i = 0 \text{ if } f_i(x) < f(x) \right\}, $$

and, the equality holds if $f_i : \mathbb{R}^n \to \mathbb{R}$ ($i = 1, \ldots, m$) are continuously differentiable at $x \in \mathbb{R}^n$.

Let $C$ be a nonempty closed subset of $\mathbb{R}^n$. Denote by $\chi_C(x)$ the indicator function of $C$, that is, $\chi_C(x) = 0$ if $x \in C$, otherwise $\chi_C(x) = +\infty$. For a closed subset $C$ of $\mathbb{R}^n$, the *normal cone* of $C$ at $x \in C$, denoted by $N(C, x)$, is defined by

$$ N(C, x) := \partial \chi_C(x) = \{ u \in \mathbb{R}^n : \langle u, y - x \rangle \leq 0 \ \forall y \in C \}. $$
DC Algorithm using $l_\infty$–penalty function with updated parameter

\[
(GP_{dc}) \quad \min \ f_0(x) \quad \text{subject to} \quad \begin{cases}
  f_i(x) \leq 0, & i = 1, \ldots, s; \\
  x \in C,
\end{cases}
\]

where $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set $f_i : \mathbb{R}^n \to \mathbb{R} \cap \{+\infty\}$, $i = 0, 1, \ldots, m$ are DC functions.

$D = \{x \in C : f_i(x) \leq 0, \ i = 1, \ldots, s\}$ : the feasible set of $(GP_{dc})$

\[
p(x) := \max\{f_1(x), \ldots, f_m(x)\}; \quad l(x) := \{i \in \{1, \ldots, m\} : f_i(x) = p(x)\}; \quad p^+(x) = \max\{p(x), 0\}.
\]

We consider the following penalty problems

\[
\begin{cases}
  \minimize \ \varphi_k(x) := f(x) + \beta_k p^+(x) \\
  \text{subject to} \quad x \in C,
\end{cases} \tag{26}
\]

where, $\beta_k$ is penalty parameters.
Note that $p^+$ is a DC function whenever $f_i$'s are DC function and that there are available DC decompositions for $p^+$. For example, if DC decompositions of $f_i$, ($i=1,...,m$) are given

$$f_i = g_i - h_i, \quad i = 1, ..., m,$$
then a well-known DC decomposition of $p^+$ is defined by

$$p^+(x) = \max_{i=1,...,m} \left\{ 0, g_i(x) + \sum_{j=1, j\neq i}^m h_j(x) \right\} - \sum_{j=1}^m h_j(x).$$

(27)

So the objective function $\varphi_k(x)$ of problem (26) is a DC function. Let DC decompositions of $f_0$ and $p^+$ be given by

$$f_0(x) = g_0(x) - h_0(x);$$

(28)

$$p^+(x) = p_1(x) - p_2(x),$$

(29)

where $g_0$, $h_0$, $p_1$, $p_2$ are convex functions defined on the whole space.
Then, we have the following DC decomposition for $\varphi_k$

$$\varphi_k(x) = g_k(x) - h_k(x), \quad x \in \mathbb{R}^n,$$

(30)

where

$$g_k(x) := g_0(x) + \beta_k p_1(x); \quad h_k(x) := h_0(x) + \beta_k p_2(x).$$

(31)

We make the following assumption that will be used in the sequel.

Assumption 1. $f_i'$s ($i = 0, ..., m$) are locally Lipschitz functions at every point of $C$.

Assumption 2. Either $g_k$ or $h_k$ is differentiable on $C$, and

$$\rho(g_0) + \rho(h_0) + \rho(p_1) + \rho(p_2) > 0.$$
DC Algorithm with updated parameter

**ALGORITHM 1.**

*Initialization:* Take an initial point $x^1 \in C$; $\delta > 0$; an initial penalty parameter $\beta_1 > 0$ and set $k := 1$.

1. Compute $y^k \in \partial h_k(x_k)$.
2. Compute $x^{k+1} \in \partial (g_k + \chi_C)^*(y^k)$, i.e., $x^{k+1}$ is a solution of the convex program

   $$\min \{g_k(x) - \langle x, y^k \rangle : x \in C \}.$$  (32)

3. **Stopping test.**
   Stop if $x^{k+1} = x^k$ and $p(x^k) \leq 0$.

4. **Penalty parameter update.**
   Compute $r_k := \min\{p(x^k), p(x^{k+1})\}$ and set

   $$\beta_{k+1} = \begin{cases} 
   \beta_k & \text{if either } \beta_k \geq \|x^{k+1} - x^k\|^{-1} \text{ or } r_k \leq 0, \\
   \beta_k + \delta & \text{if } \beta_k < \|x^{k+1} - x^k\|^{-1} \text{ and } r_k > 0,
   \end{cases}$$

5. Set $k := k + 1$ and go to Step 1.

---

**Pham Dinh Tao**
National Institute for Applied Sciences-Rouen, France

**Joint work with Le Thi Hoai An**

**INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCES**
Global convergence
Recall that a point \( x^* \in D \) (the feasible set of \((GP_{dc})\)) is a Karush-Kuhn-Tucker (KKT) point for the problem \((GP_{dc})\) if there exist nonnegative scalars \( \lambda_i, \ i = 1, \ldots, s \) such that

\[
\begin{align*}
0 & \in \partial f_0(x^*) + \sum_{i=1}^{m} \lambda_i \partial f_i(x^*) + N(C, x^*), \\
\lambda_i f_i(x^*) & = 0, \quad i = 1, \ldots, m.
\end{align*}
\]  

(33)

We say that the extended Mangasarian-Fromowitz constraint qualification (EMFCQ) is satisfied at \( x^* \in D \) with \( I(x^*) \neq \emptyset \) if

\((MFCQ)\) there is a vector \( d \in \text{cone}(C - \{x^*\}) \) (the cone hull of \( C - \{x^*\} \)) such that \( f_i^+(x^*, d) < 0 \) for all \( i \in I(x^*) \).

When \( f_i' \)'s are continuously differentiable, then \( f_i^+(x^*, d) = \langle \nabla f(x^*), d \rangle \). Therefore, (EMFCQ) becomes the well-known Mangasarian-Fromowitz constraint qualification.
It is well known that if the (extended) Mangasarian-Fromowitz constraint qualification is satisfied at a local minimizer $x^*$ of problem $(GP_{dc})$ then the KKT first order necessary condition (33) holds (see Mangasarian; Mangasarian and Fromovitz).

**Assumption 3.** The (extended) Mangasarian-Fromowitz constraint qualification (EMFCQ) is satisfied at any $x \in \mathbb{R}^n$ with $p(x) \geq 0$.

When $f_i, \ i = 1, ..., s$ are all convex functions, then it is obvious that this assumption is satisfied under the *Slater regular condition*, i.e., there exists $x \in C$ such that

$$f_i(x) < 0 \quad \text{for all } \ i = 1, ..., s.$$
Theorem 8

Suppose that $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $f_i, i = 1, ..., m$ are DC functions on $C$. Suppose further that Assumptions 1-3 are verified. Let $\delta > 0$, $\beta_1 > 0$ be given. Let $\{x_k\}$ be a sequence generated by Algorithm 1. Then Algorithm 1 either stops finitely at a KKT point $x^k$ for problem $(P)$ or generates an infinite sequence $\{x^k\}$ of iterates such that $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$ and every limit point $x^\infty$ of the sequence $\{x^k\}$ is a KKT point of problem $(P)$.

Instead of using the penalty function $p^+$, we could also use the following one

$$\psi(x) := \sum_{i=1}^{m} f_i^+(x), \quad x \in \mathbb{R}^n.$$ 

It is easy to see that the convergence result also holds for this penalty function.
DC Algorithm by linearizing convex parts of DC constraints
By using the main idea of the DCA that linearizes the nonconvex part of
the DC structure, we can devise the extended DCA to DC constraints
which consists of the solution of a sequence of convex programs of the
form (Le Thi Hoai An, Pham Dinh Tao, The extended DCA to general
DC programs with DC constraints, Research Report, 2007, National
Insitute for Applied Sciences, Rouen) :

\[
\min g_0(x) - \langle y_0^k, x \rangle \quad \text{s.t.} \quad \begin{cases} 
  g_i(x) - h_i(x^k) - \langle y_i^k, x - x^k \rangle \leq 0, i = 1, \ldots, m; \\
  x \in C,
\end{cases}
\]

(34)
where, \( x^k \in \mathbb{R}^n \) is the current iterate, \( y_i^k \in \partial h_i(x^k) \) for \( i = 0, \ldots, m \).
However, this linearization can lead to infeasibility of convex subproblem (34). We propose a relaxation technique to deal with the feasibility of subproblems. Instead of (34), we consider the subproblem

\[
\min \ g_0(x) - \langle y^k_0, x \rangle + \beta_k t \quad \text{s.t.} \quad \begin{cases} 
  g_i(x) - h_i(x^k) - \langle y^k_i, x - x^k \rangle \leq t, \ i = 1, \ldots, m \\
  x \in C, \ t \geq 0;
\end{cases}
\]  

(35)

where \( \beta_k > 0 \) is a penalty parameter. Obviously, (35) is a convex problem that is always feasible.
Furthermore, the Slater constraint qualification is satisfied for the constraints of (35), thus the Karush-Kuhn-Tucker optimality condition holds for some solution \((x^{k+1}, t^{k+1})\) of (35): there exist some \(\lambda_i^{k+1} \in \mathbb{R}, i = 1, \ldots, m\), and \(\mu_i^{k+1} \in \mathbb{R}\) such that

\[
\begin{align*}
0 &\in \partial g_0(x^{k+1}) - y_0^k + \sum_{i=1}^m \lambda_i^{k+1} (\partial g_i(x^{k+1}) - y_i^k) + N(C, x^{k+1}), \\
\beta_k - \sum_{i=1}^m \lambda_i^{k+1} - \mu^{k+1} &\geq 0, \\
g_i(x^{k+1}) - h_i(x^k) - \langle y_i^k, x^{k+1} - x^k \rangle &\leq t^{k+1}, \lambda_i^{k+1} \geq 0 \quad i = 1, \ldots, m, \; x^{k+1} \in C, \\
\lambda_i^{k+1} (g_i(x^{k+1}) - h_i(x^k) - \langle y_i^k, x^{k+1} - x^k \rangle - t^{k+1}) &\geq 0, \; i = 1, \ldots, m, \\
t^{k+1} &\geq 0, \; \mu_i^{k+1} \geq 0, \; t^{k+1} \mu^{k+1} = 0.
\end{align*}
\] (36)
ALGORITHM 2.

Initialization: Take an initial point $x^1 \in C$; $\delta_1, \delta_2 > 0$; an initial penalty parameter $\beta_1 > 0$ and set $k := 1$.

1. Compute $y_i^k \in \partial h_i(x^k)$, $i = 0, ..., m$.

2. Compute $(x^{k+1}, t^{k+1})$ as the solution of (35), and the associated Lagrange multipliers $(\lambda^{k+1}, \mu^{k+1})$.

   Stop if $x^{k+1} = x^k$ and $t^{k+1} = 0$.

4. Penalty parameter update.
   Compute $r_k := \min\{\|x^{k+1} - x^k\|^{-1}, \|\lambda^{k+1}\|_1 + \delta_1\}$, where $\|\lambda^{k+1}\|_1 = \sum_{i=1}^m |\lambda_i^{k+1}|$, and set
   
   $$\beta_{k+1} = \begin{cases} 
   \beta_k & \text{if } \beta_k \geq r_k, \\
   \beta_k + \delta_2 & \text{if } \beta_k < r_k.
   \end{cases}$$

5. Set $k := k + 1$ and go to Step 1.
Note that the updated penalty parameter rule is inspired by Solodov in [Solodov M. V., On the sequential quadratically constrained quadratic programming methods, Mathematics of Oper. Research, 29, 64-79 (2004)]. That is to ensure that the unboundedness of \( \{\beta^k\} \) leads to the unboundedness of \( \{\|\lambda^k\|_1\} \) and \( \|x^{k+1} - x^k\| \to 0 \).

As in the preceding section, denote by

\[
\varphi_k(x) := f_0(x) + \beta_k p^+(x), \quad x \in \mathbb{R}^n.
\]

**Lemma 9**

The sequence \((x^k, t^k)\) generated by Algorithm 2 satisfies the following inequality

\[
\varphi_k(x^k) - \varphi_k(x^{k+1}) \geq \frac{\rho}{2} \|x^{k+1} - x^k\|^2, \quad \text{for all } k = 1, 2, \ldots
\]

where, \( \rho := \rho(g_0) + \rho(h_0) + \min\{\rho(g_i) : i = 1, \ldots m\} \).
Theorem 10

Suppose that $C \subseteq \mathbb{R}^n$ is a nonempty closed convex set and $f_i, \ i = 1, \ldots, m$ are DC functions on $C$ such that assumptions 1 and 3 are verified. Suppose further that for each $i = 0, \ldots, m$, either $g_i$ or $h_i$ is differentiable on $C$ and that

$$\rho := \rho(g_0) + \rho(h_0) + \min\{\rho(g_i) : i = 1, \ldots, m\} > 0.$$  

Let $\delta_1, \delta_2 > 0$, $\beta_1 > 0$ be given. Let $\{x_k\}$ be a sequence generated by Algorithm 2. Then Algorithm 2 either stops finitely at a KKT point $x^k$ for problem $(GP_{dc})$ or generates an infinite sequence $\{x^k\}$ of iterates such that $\lim_{k \to \infty} \|x^{k+1} - x^k\| = 0$ and every limit point $x^\infty$ of the sequence $\{x^k\}$ is a KKT point of problem $(GP_{dc})$. 

Pham Dinh Tao

National Institute for Applied Sciences-Rouen, France

Joint work with Le Thi Hoai An

INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCES
Note that, as shown in Theorem 8 and Theorem 10, the penalty parameter $\beta_k$ is constant when $k$ is sufficiently large. Observing from the proof of these convergence theorems, the sequence $\{\varphi(x^k)\}$ of values of the function $\varphi(x) = f_0(x) + \beta_k p^+(x)$ along with the sequence $\{x^k\}$ generated by Algorithms 1 and Algorithm 2 is decreasing. Note that these results remain valid if we replace, in (35), the variable $t$ by $t_i$ for $i = 1, \ldots, m$ and the function $\beta_k t$ by $\beta_k \sum_{i=1}^{m} t_i$. 
Consider the general DC program
\[
\alpha = \inf \{ f_0(x) := g_0(x) - h_0(x) : x \in C, \ p(x) := g(x) - h(x) \leq 0 \} \quad (P_{dcg}) \tag{38}
\]
where \( C \) is a nonempty closed convex set in \( \mathbb{R}^n \), \( g, h, g_0, h_0 \in \Gamma_0(\mathbb{R}^n) \) and its feasible set \( S = \{ x \in C : g(x) - h(x) \leq 0 \} \) assumed to be nonempty. The usual penalized DC program is the DC program defined by
\[
\alpha(\tau) = \inf \{ f_0(x) + \tau p^+(x) : x \in C \} \quad (P_{dcg}^\tau) \tag{39}
\]
We will summarize main exact penalty results with/without error bounds of nonconvex constraint sets \( S \), which establish the equivalence of \( (P_{dcg}) \) and its penalized \( (P_{dcg}^\tau) \). These results are based on the concavity of objective/constraint functions and local error bounds [Le Thi, Pham Dinh and Huynh Van, 2012].
We first deal with exact penalty for nonconvex programs having concave objective functions and bounded polyhedral convex constraint sets with additional concave constraint functions. These results are not derived from error bounds of feasible solution sets because the question concerning their error bounds is still open. We get back to it later. Let $K$ be a nonempty bounded polyhedral convex subset of $\mathbb{R}^n$ and let $f, g$ be finite functions on $K$. Recall that the function $f$ is said to be continuous relative to $K$, if its restriction to $K$ is a continuous function. The notions of relative upper semicontinuity and relative lower semicontinuity are defined similarly. In this section, we are concerned with the following nonconvex programs:

\[
\alpha := \min \{ f(x) : x \in K, \ g(x) = 0 \} \quad (P) \\
\min \{ f(x) : x \in K, \ g(x) \leq 0 \} \quad (P')
\]
In [Le Thi, Pham Dinh, Le Dung, 1999] it has been established an exact penalty result for the problem $(P)$ under the hypothesis that $g$ is nonnegative on $K$. We consider $(P)$ here without the nonnegativity assumption of $g$. Problem $(P)$ could be rewritten equivalently

\[ \alpha := \min \{ f(x) : x \in K, \ g(x) \leq 0, \ g(x) \geq 0 \} \quad (P) \tag{42} \]

Its penalized reads

\[ \alpha(\tau) := \min \{ f(x) + \tau g(x) : x \in K, \ g(x) \geq 0 \} \quad (P_\tau). \tag{43} \]

The following result constitutes a substantial extension of the key theorem in ([Le Thi, Pham Dinh, Le Dung, 1999]- Theorem 2) concerning the special case of $(P)\quad (42)$ when the concave function $g$ is assumed to be nonnegative on $K$. Let us denote by $V(K)$ the vertex set of $K$ and by $\mathcal{P}$ and $\mathcal{P}_\tau$ the optimal solution sets of $(P)\quad (42)$ and $(P_\tau)\quad (43)$, respectively. Here, the convention $\min_{\emptyset} g(x) = +\infty$ is used.
Theorem 11

Let $K$ be a nonempty bounded polyhedral convex set in $\mathbb{R}^n$ and let $f$, $g$ be finite concave functions continuous relative to $K$. Suppose that the feasible set of $(P)$ is nonempty. Then there exists $\tau_0 \geq 0$ such that for all $\tau > \tau_0$, the problems $(P)$—(42) and $(P_{\tau})$—(43) have the same optimal value and the same optimal solution set. Furthermore, we can take

$$\tau_0 = \frac{f(x_0) - \alpha(0)}{m}$$

with $m = \min\{g(x) : x \in V(K), g(x) > 0\}$ and any $x_0 \in K$, $g(x_0) = 0$. 
Obviously, \((P')\) is equivalent to the following one:

\[
\min\{f(x) : (x, t) \in K \times [0, \beta], \ g(x) + t = 0\} \ (P')
\]

where, \(\beta \geq \max\{-g(x) : x \in K\}\) and such the equivalence is in the
following usual sense: *If \(x\) is a solution of \((P')\) then \((x, -g(x))\) is a
solution of \((\overline{P}')\). Conversely, if \((x, t)\) is a solution of \((\overline{P}')\) then \(x\) is a
solution of \((P')\).*

In virtue of Theorem 11, there is \(\tau_0 \geq 0\) such that for all \(\tau > \tau_0\), the
problem \((\overline{P}')\) is equivalent to the following one:

\[
\min\{f(x) + \tau(t + g(x)) : (x, t) \in K \times [0, \beta], \ g(x) + t \geq 0\} \ (\overline{P}_{\tau}')
\]
Thus we obtain the following corollary.

**Corollary 12**

*Under the assumptions of Theorem 11, the problems \((P') - (??)\) and \((\overline{P'_T}) - (45)\) are equivalent in the sense given in Theorem 11.*

It is obvious that \((\overline{P'_T}) - (45)\) is a DC program. As shown by the corollary, it is worth noting that if \(f, g\) are finite concave functions on a nonempty bounded polyhedral convex set \(K\), then Problem \((P') - (??)\) is equivalent to \((\overline{P'_T}) - (45)\).

By applying Corollary 12, we list below the nonconvex programs (with DC objective functions and DC constraint functions) frequently encountered in practice that we can advantageously recast into more suitable DC programs equivalent (in the sense given Theorem 11):
(a) Polyhedral convex objective function and concave constraint function
Consider the problem with a nonempty feasible set:
\[
\min \{ f(x) : x \in K, \ g(x) \leq 0 \},
\]
here, \( K \) is a polyhedral convex set, \( f \in \Gamma_0(\mathbb{R}^n) \) is a polyhedral convex function, i.e., it is a sum of a pointwise supremum of a finite collection of affine functions and an indicator function of a nonempty polyhedral convex set \( D \)
\[
f(x) = \max\{ \langle x, y_i \rangle - \alpha_i : i = 1, \ldots, r \} + \chi_D(x) = \bar{f}(x) + \chi_D(x),
\]  
(46)
such that \( D \cap K \) is bounded and \( g \) is a finite concave function continuous relative to \( D \cap K \).
Obviously, this problem is equivalent to

$$\min \{ t : \bar{f}(x) \leq t, (x, t) \in [D \cap K] \times [a, b], g(x) \leq 0 \}, \quad (47)$$

where, $a \leq \min \{ f(x) : x \in [D \cap K] \}$ and $b \geq \max \{ f(x) : x \in [D \cap K] \}$.

According to Corollary 2, if the feasible set of there is $\tau_0$ such that for all $\tau > \tau_0$, Problem (47) is equivalent to following DC program (with $\beta \geq \max_{x \in [D \cap K]} (-g(x))$):

$$\min \{ t + \tau(s + g(x)) : \bar{f}(x) \leq t, (x, t, s) \in [D \cap K] \times [a, b] \times [0, \beta], \quad g(x) + s \geq 0 \} \quad (48)$$

that can be reduced to

$$\min \{ \bar{f}(x) + \tau(s + g(x)) : (x, s) \in [D \cap K] \times [0, \beta], \quad g(x) + s \geq 0 \} \quad (49)$$
(b) **Concave objective function and DC (whose first DC component is polyhedral convex function) constraint function**

Let $f$ be a finite concave function on a polyhedral convex set $K$. Let $g$ be a polyhedral convex function defined as in (46) with Dom $g := \{ x \in \mathbb{R}^n : g(x) < +\infty \} = D$ such that $D \cap K$ is bounded nonempty, and $h$ be a finite convex function on $D \cap K$. Let us now consider the following problem with a polyhedral DC constraint:

$$
\min \{ f(x) : x \in K, g(x) - h(x) \leq 0 \}. \tag{50}
$$

This problem is equivalently transformed into

$$
\min \{ f(x) : (x, t) \in [D \cap K] \times [a, b], \overline{g}(x) - t \leq 0, t - h(x) \leq 0 \}, \tag{51}
$$

where, $a \leq \min_{x \in [D \cap K]} g(x), b \geq \max_{x \in [D \cap K]} g(x)$. In virtue of Corollary 2, if the feasible set of (50) is nonempty then there is $\tau_0 \geq 0$ such that
for all $\tau > \tau_0$ Problem (51) is equivalent to the DC program
\begin{equation}
(\beta \geq \max_{x \in [D \cap K]} h(x) - a) \colon
\begin{align*}
\min \{ f(x) + \tau (s + t - h(x)) & : (x, t, s) \in [D \cap K] \times [a, b] \times [0, \beta], \\
\bar{g}(x) - t & \leq 0, \ s + t - h(x) \geq 0 \}.
\end{align*}
\end{equation}
(c) **DC objective function and DC constraint function (whose first DC component is polyhedral convex function)**

Let $K$ be a nonempty polyhedral convex set in $\mathbb{R}^n$. Consider the nonconvex program with a polyhedral DC objective function $f = g_0 - h_0$ and DC constraint function :

$$\min \{ g_0(x) - h_0(x) : x \in K, g_1(x) - h_1(x) \leq 0 \}$$  \hspace{1cm} (53)

where $g_0, g_1$ are polyhedral convex functions on $\mathbb{R}^n$, i.e.,

$$g_0(x) : = \max \{ \langle x, y_i \rangle - \alpha_i : i = 1, ..., q \} + \chi_D(x)$$ \hspace{1cm} (54)

$$= g_0^-(x) + \chi_D(x),$$

$$g_1(x) : = \max \{ \langle x, z_i \rangle - \alpha_i : i = 1, ..., r \} + \chi_R(x)$$

$$= g_1^-(x) + \chi_R(x),$$
with $D, R$ being nonempty polyhedral convex sets and $h_0, h_1$ are a finite convex function continuous relative to $D \cap K \cap R$. Assume that the polyhedral convex set $D \cap K \cap R$ is bounded nonempty. Problem (53) is equivalent to

$$\min \{ s - h_0(x) : (x, s, t) \in D \cap K \cap R \times [a_0, b_0] \times [a_1, b_1], \overline{g}_0(x) \leq s, \overline{g}_1(x) \leq t - h_1(x) \leq 0 \}$$

where $a_0, b_0, a_1, b_1$ are constant numbers verifying

$$a_0 \leq \min \{ \overline{g}_0(x) : x \in D \cap K \cap R \} \leq \max \{ \overline{g}_0(x) : x \in D \cap K \cap R \} \leq b_0$$

$$a_1 \leq \min \{ \overline{g}_1(x) : x \in D \cap K \cap R \} \leq \max \{ \overline{g}_1(x) : x \in D \cap K \cap R \} \leq b_1$$

Hence, it follows from Corollary 12 that if the feasible set of Problem (55) is nonempty then there is $\tau_0 \geq 0$ such that for all $\tau > \tau_0$ this problem is equivalent to
\[
\min \{ s - h_0(x) + \tau [u + t - h_1(x)] : (x, s, t, u) \in D \cap K \cap R \times [a_0, b_0] \times [a_1, \infty) \}
\]
\[
\overline{g}_0(x) \leq s, \overline{g}_1(x) \leq t, u + t - h_1(x) \geq 0
\]

with \( c \) being a constant number satisfying
\[
c \geq \max \{ h_1(x) : x \in D \cap K \cap R \} - a_1.
\]
Finally, Problem (53) is equivalent to the DC program
\[
\min \{ f(x, u, t) = \overline{g}_0(x) - h_0(x) + \tau [u + t - h_1(x)] : \}
\]
\[
(x, t, u) \in D \cap K \cap R \times [a_1, b_1] \times [0, c], \overline{g}_1(x) \leq t, u + t - h_1(x) \geq 0 \}
\]
To complement the results concerning exact penalty in concave programming of Subsection 1, we will use the error bounds results in Subsection 3 to establish below some exact penalty properties in DC programming. Recall that, throughout this section, two problems are said to be equivalent if they have the same optimal value and the same optimal solution set.

First, we give a general exact penalty result which is a refinement of that mentioned by Clarke in [?]. Recall that a real-valued function $f$ defined on a set $C$ in $\mathbb{R}^n$ is said to be Lipschitz on $C$, if there exists a nonnegative scalar $L$ such that

$$|f(x) - f(y)| \leq L \|x - y\|$$

for all $x, y \in C$. Also $f$ is said to be locally Lipschitz relative to $C$ at some $x \in C$ if for some $\epsilon > 0$, $f$ is Lipschitz on $B(x, \epsilon) \cap C$.

Let $\theta$ be a finite DC function on a closed convex set $C$, i.e.
\[ \theta(x) = \varphi(x) - \psi(x) \quad \forall x \in C, \tag{58} \]

where \( \varphi \) and \( \psi \) belong to \( \Gamma_0(\mathbb{R}^n) \) such that
\[ C \subset \text{Dom } \varphi := \{ x \in \mathbb{R}^n : \varphi(x) < +\infty \} \subset \text{Dom } \tag{59} \]

Note that if
(i) \( \text{Dom } \varphi \) and \( \text{Dom } \psi \) have the same dimension and
(ii) \( C \) is bounded and contained in the relative interior of \( \text{Dom } \varphi \),
then the DC function \( \theta \) is Lipschitz on \( C \).
Let $f$, $h$ be real-valued functions defined on $C$. Consider the minimization problem whose optimal solution set is denote by $\mathcal{P}$

$$\alpha = \inf \{ f(x) : x \in C, \ h(x) \leq 0 \} \quad (P), \quad (60)$$

that we can write as

$$\alpha = \inf \{ f(x) : x \in S \}, \quad (61)$$

where

$$S := \{ x \in C : h(x) \leq 0 \} \quad (62)$$

Let $g : C \to \mathbb{R}$ be a nonnegative function such that $S$ can be expressed by

$$S := \{ x \in C : g(x) \leq 0 \} \quad (63)$$

Such a function $g$ must verify

$$g(x) = 0 \text{ if and only if } x \in S \quad (64)$$
Exact penalty in mathematical programming usually deals with (Section 6)

\[ g(x) := \left[h^+(x)\right]^\alpha \]  \hspace{1cm} (65)

For \( \tau \geq 0 \), we define the problems \((P_\tau)\) by

\[ \alpha(\tau) = \inf \{ f(x) + \tau g(x) : x \in C \} \] \hspace{1cm} (P_\tau), \hspace{1cm} (66)

whose optimal solution set is denoted by \( P_\tau \).
Proposition 13

Let $f$ be a Lipschitz function on $C$ with constant $L$ and let $g$ be a nonnegative finite function on $C$ such that

$$S := \{ x \in C : h(x) \leq 0 \} = \{ x \in C : g(x) \leq 0 \}.$$  

If $S$ is nonempty and there exists some $\ell > 0$ such that

$$d(x, S) \leq \ell g(x)$$

for all $x \in C$,

then one has:

(i) $\alpha(\tau) = \alpha$ and $\mathcal{P} \subset \mathcal{P}_\tau$ for all $\tau \geq L\ell$

(ii) $\mathcal{P}_\tau = \mathcal{P}$ for all $\tau > L\ell$

The above proof is quite standard and follows the line in [?]. However, it is worth noting that our proof [?], unlike that of [?], needs neither the nonemptiness of the optimal solution set of (P) nor the closedness of its feasible set $S$. 

Pham Dinh Tao
Throughout the paper $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$ unless otherwise specified. Let $C$ be a nonempty closed subset in $\mathbb{R}^n$ and let $h$ be a finite function on $C$. We consider the solution set of the following inequality system

$$S = \{ x \in C : h(x) \leq 0 \} \quad (67)$$
Recall that an error bound of $S$ is an inequality of the form

$$ d(x, S) \leq \tau h^+(x) \text{ for all } x \in C, \quad (68) $$

where, $d(x, S) = \inf_{z \in S} \| x - z \|$, and $\tau$ is a positive number. For $x_0 \in S$, if the inequality in (68) holds for all $x$ in a neighborhood of $x_0$, then we say that $S$ has an error bound around $x_0$. Instead of (68), an inequality of the form

$$ d(x, S) \leq \tau h^+(x)^\gamma \text{ for all } x \in C, \quad (69) $$

with $\tau > 0$ and $0 < \gamma < 1$ will be called an error bound for (67) with the exponent $\gamma$. Note that if an error bound of $S$ holds for a norm it holds for every norm because all norms on $\mathbb{R}^n$ are equivalent. Error bounds have important applications in many areas of mathematical programming, e.g., in sensitivity analysis, in convergence analysis of some algorithms, and in exact penalty, etc [Le Thi, Pham Dinh, Huynh Van, 2012].
In our work [?], we derive several new results on error bounds for the inequality system (67) when $C$ is a suitable closed convex set and $h$ is a concave function on $C$. These systems play an important role in global optimization.

In general, a system which has local error bounds at all points in its solution set does not necessarily attain the global error bound. However, under the compactness hypothesis, such an implication holds as will show the following proposition that will be needed thereafter.
Theorem 14

Let $C$ be a nonempty compact set in $\mathbb{R}^n$. Let $h$ be a finite function lower semicontinuous relative to $C$ and $S := \{x \in C : h(x) \leq 0\} \neq \emptyset$. Suppose that for each $z \in S$, there exist $\tau(z), \gamma(z) > 0$ and $\epsilon(z) > 0$ such that

$$d(x, S) \leq \tau(z)h^+(x)\gamma(z)$$ for all $x \in B(z, \epsilon(z)) \cap C,$ \hspace{1cm} (70)

where, as usual, $B(z, \epsilon(z))$ denotes the closed ball in $\mathbb{R}^n$ with center $z$ and radius $\epsilon(z)$. Then there exist $\tau, \gamma > 0$ such that

$$d(x, S) \leq \tau h^+(x)\gamma$$ for all $x \in C.$ \hspace{1cm} (71)

Remark 15

We see from the proof above that if (70) is satisfied for a common exponent $\alpha$ which does not depend on $z$, then (71) holds also with this exponential parameter.
As a consequence we can derive the following error bound

**Corollary 16**

Let $C$ be the Euclidean ball $\{x \in \mathbb{R}^n : \|x - c\| \leq r\}$ where $c \in \mathbb{R}^n$ and $r > 0$ and $S := \{x \in \mathbb{R}^n : \|x - c\| = r\}$. Then we have

$$d(x, S) \leq h(x) \text{ for all } x \in C,$$

where $h(x) := r - \|x - c\|$, and

$$d(x, S) \leq \frac{1}{r} k(x) \text{ for all } x \in C,$$

where $k(x) := r^2 - \|x - c\|^2$. 

**Pham Dinh Tao**

**National Institute for Applied Sciences-Rouen, France**

**Joint work with Le Thi Hoai An**
It is well known, see e.g. Proposition 13 in Section ?? , that error bounds of feasible solution sets of nonconvex programs provide, in an elegant and deep way, exact penalty for those problems with Lipschitz objective functions. This subsection is concerned with error bounds for DC inequality systems.

Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\} \) be a lower semicontinuous function on \( \mathbb{R}^n \). The lower directional derivative (or contingent derivative or Hadamard derivative) of \( f \) at \( x \in \text{Dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\} \) is given by

\[
f'(x; u) = \lim_{(t,v) \rightarrow (0+,u)} \inf \frac{f(x + tv) - f(x)}{t}.
\]
If $f$ is differentiable at $x$ with the derivative $\nabla h(x)$ then $f'(x; u) = \langle \nabla h(x), u \rangle$. When $f$ is convex or concave then the lower directional derivative coincides with the usual directional derivative.

We present first, in the context of finite dimension, a general sufficient condition for the error bound in terms of the lower directional derivative [Le Thi, Pham Dinh, Huynh Van, 2012].
Proposition 17

Let $C$ be a nonempty closed convex set in $\mathbb{R}^n$, let $h$ be a finite function lower semicontinuous relative to $C$ and $S := \{ x \in C : h(x) \leq 0 \} \neq \emptyset$.

(i) Let $x_0 \in S$. If there exists $\mu > 0$, $\epsilon > 0$ such that for each $y \in [B(x_0, \epsilon) \cap C] \setminus S$ there exists $z \in C$, $z \neq y$ such that $h'(y; z - y) < -\mu \| z - y \|$ then one has

$$d(x, S) \leq \frac{1}{\mu} h^+(x) \text{ for all } x \in B(x_0, \epsilon/2) \cap C.$$
Proposition 18

(ii) In addition, if $C$ is bounded, $h$ is continuous relative to $C$ and for all $x \in S$ with $h(x) = 0$, the set

$$\{ (\mu, \epsilon) > (0, 0) : \forall y \in [B(x, \epsilon) \cap C] \setminus S \quad \exists z_y \in C \text{ such that } h'(y; z_y - y) < -\mu \}$$

is nonempty, then there is $\tau > 0$ such that

$$d(x, S) \leq \tau h^+(x) \text{ for all } x \in C.$$
Remark 19

Proposition 17 remains valid if we replace the conditions (i) (resp. (ii)) by the following conditions (iii) (resp. (iv)):

(iii) Let $x_0 \in S$. If there exits $z \in C$ such that the function $x \rightarrow h'(x; z - x)$ is upper semicontinuous at $x_0$ and $h'(x_0; z - x_0) < 0$.

(iv) In addition, if $C$ is bounded, $h$ is continuous relative to $C$ and for all $x_0 \in S$ with $h(x_0) = 0$, there exits $z \in C$ such that the function $x \rightarrow h'(x; z - x)$ is upper semicontinuous at $x_0$ and $h'(x_0; z - x_0) < 0$. 
Next, we give a condition ensuring the error bound for DC inequality systems which is a slight extension the well known Slater one for convex systems (when $h = 0$). The systems of that type play an important role in DC programming. Let $C$ be a nonempty closed convex set in $\mathbb{R}^n$ and let $g, h$ be finite convex functions on $C$. Consider the following set:

$$S = \{ x \in C : g(x) - h(x) \leq 0 \}.$$
Theorem 20

Let $C \subset \mathbb{R}^n$ be a nonempty compact convex set, $g$ be a finite convex function continuous relative to $C$ and $h$ be a differentiable convex function on $C$. If for each $x_0 \in S$ with $g(x_0) - h(x_0) = 0$ the set

$$\{ z \in C : g(z) - \langle \nabla h(x_0), z - x_0 \rangle - h(x_0) < 0 \}$$

is nonempty, then there exists $\tau > 0$ such that

$$d(x, S) \leq \tau [g(x) - h(x)]^+ \text{ for all } x \in C.$$

In the case of systems consisting of quadratic inequalities, the conditions in the preceding proposition can be rewritten more simply as follows.
Corollary 21

Let $C$ be a nonempty compact convex subset of $\mathbb{R}^n$ and

$$f_i(x) := \frac{1}{2}\langle x, Q_i x \rangle + \langle q_i, x \rangle + r_i, \quad i = 1, ..., m$$

where $Q_i$ are symmetric $n \times n$ matrices, $q_i \in \mathbb{R}^n$, $r_i \in \mathbb{R}$. Let

$$f(x) := \max\{f_i(x) : i = 1, ..., m\}.$$ 

If for each $x \in S := \{x \in C : f(x) \leq 0\}$ with $f(x) = 0$, there holds

$$\{z \in C : f(z) + \frac{\lambda}{2}\|z - x\|^2 < 0\} \neq \emptyset,$$

where $\lambda := \max\{\rho(Q_i) : i = 1, ..., m\}$ and $\rho(Q_i)$ is the spectral radius of $Q_i$, then there exists $\tau > 0$ such that

$$d(x, S) \leq \tau f^+(x) \quad \text{for all} \quad x \in C.$$
Next, let $h_i : \mathbb{R}^n \to \mathbb{R}$, ($i = 1, ..., m$) be $m$ concave functions. Consider now the following system of the concave inequalities:

$$S = \{ x \in \mathbb{R}^n : h_i(x) \leq 0, i = 1, ..., m \}.$$

Setting ($\mathcal{G} := \{1, ..., m\}$)

$$h(x) := \max\{ h_1(x), ..., h_m(x) \}, \quad \mathcal{G}(x) := \{ i \in \{1, ..., m\} : h_i(x) = h(x) \}$$

**Proposition 22**

Let $h_i, (i = 1, ..., m), h$ and $S$ be defined as above. If $S$ and the set

$$\{ z \in \mathbb{R}^n : h_i(z) > 0 \text{ for } i = 1, ..., m \} \quad (74)$$

are nonempty, then for every $\rho > 0$, there exists $\tau > 0$ such that

$$d(x, S) \leq \tau h^+(x) \text{ for all } x \in \mathbb{R}^n, \|x\| \leq \rho.$$

Moreover, if $\mathbb{R}^n \setminus S$ is bounded then the inequality holds for all $x \in \mathbb{R}^n$. 

Pham Dinh Tao  
National Institute for Applied Sciences-Rouen, France  
Joint work with Le Thi Hoai An
In the case of $m = 1$ concave inequality, if the assumption (74) doesn’t hold then $S = \mathbb{R}^n$ and it is trivial that $d(x, S) \leq \tau h^+(x)$ for all $x \in \mathbb{R}^n$. Hence one has

**Corollary 23**

Let $h$ be a finite concave function on $\mathbb{R}^n$ and let $S := \{ x \in \mathbb{R}^n : h(x) \leq 0 \}$ be nonempty. Then for every $\rho > 0$, there exists $\tau > 0$ such that

$$d(x, S) \leq \tau h^+(x) \quad \text{for all } x \in \mathbb{R}^n, \|x\| \leq \rho.$$  

Moreover if $\mathbb{R}^n \setminus S$ is bounded, then

$$d(x, S) \leq \tau h^+(x) \quad \text{for all } x \in \mathbb{R}^n.$$  

Pham Dinh Tao

INTRODUCTION TO DC PROGRAMMING & DCA AND RECENT ADVANCES
Remark 24

Let \( E := \{ x \in \mathbb{R}^n : \| x \| = r \} \) that can be written as \( E := \{ x \in C : h_1(x) = r - \| x \| \leq 0 \} \) where \( C := \{ x \in \mathbb{R}^n : \| x \| \leq r \} \). According to Corollary 24 there is \( \tau_1 > 0 \) such that

\[
d(x, E) \leq \tau_1 h_1(x) \quad \text{for all} \ x \in C. \tag{75}\]

Similarly using the concave function \( h_2(x) := r^2 - \| x \|^2 \geq 0 \) to define \( E := \{ x \in C : h_2(x) \leq 0 \} \) yields the error bound with \( \tau_2 > 0 \)

\[
d(x, E) \leq \tau_2 h_2(x) \quad \text{for all} \ x \in C. \tag{76}\]
Remark 25

Since

$$2rh_1(x) \geq h_2(x) := r^2 - \|x\|^2 = [r + \|x\|][r - \|x\|]$$

$$= [r + \|x\|]h_1(x) \geq rh_1(x) \quad \forall x \in C, \quad (77)$$

there hold

(i) If $\tau_1 > 0$ verifies (75) then $\tau_2 = \frac{\tau_1}{r}$ verifies (76)

(ii) Conversely if $\tau_2 > 0$ verifies (76) then $\tau_1 = 2r\tau_2$ verifies (75)
**Remark 26**

Consider now the feasible set
\[ F := \{ x \in B(c, r) : h_3(x) := r - \| x - c \| \geq 0 \} \] in Corollary 16. Since \( B(c, r) = c + C \), \( F = c + S \) and the distance is invariant with respect to translations, it follows from (75) that there \( \tau_3 > 0 \) such that
\[ d(x, F) \leq \tau_3 h_3(x) \quad \text{for all } x \in B(c, r) \]

Likewise, the concave function \( h_4(x) := r^2 - \| x - c \|^2 \) leads to the error bound with \( \tau_4 > 0 \)
\[ d(x, F) \leq \tau_4 h_4(x) \quad \text{for all } x \in B(c, r) \]

and the parameters \( \tau_3, \tau_4 \) satisfy the properties (i) and (ii) as \( \tau_1, \tau_2 \). Note at last that the error bound of \( F \) given by (72) and (73) are practically advantageous insofar as the parameters \( \tau_3, \tau_4 \) are known.
Let now \( K \) be a nonempty bounded polyhedral convex set and let \( h \) be a finite concave function on \( K \). Consider the following set:

\[
S = \{ x \in K : h(x) \leq 0 \}.
\]

The constraint "\( h(x) \leq 0 \)" is often called reverse convex constraint. Of course, the preceding results on exact penalty in Subsection 1 can be applicable for this system. However, under additional assumptions, we can establish some results on error bounds which are easy to verify in many applications. It is worth noting that the sets of that type figure in many practical problems (see, e.g., DC Programming and DCA: http://lita.sciences.univ-metz.fr/~lethi/DCA.html).
Theorem 27

Let $K$ be a nonempty bounded polyhedral convex set in $\mathbb{R}^n$ and let $h$ be a differentiable concave function on $K$. If $S$ is nonempty, then there exists $\tau > 0$ such that

$$d(x, S) \leq \tau h^+(x) \quad \text{for all } x \in K. \quad (80)$$

As before, let $K$ be a given nonempty bounded polyhedral convex set. We now consider the function $h$ defined by

$$h(x) = \sum_{j=1}^{m} \min \{ h_{ij}(x) : i \in \mathcal{S}_j \} \quad (81)$$

where, $\mathcal{S}_1, \ldots, \mathcal{S}_m$ are finite index sets and $h_{ij}$ are differentiable concave functions on $K$. The functions of this type are not differentiable.
By applying Theorem 27, we next show that the conclusion of Theorem 27 remains valid for the function $h$ given by (81).

**Theorem 28**

Let $K$ be a nonempty bounded polyhedral convex set and let $h$ be defined as above. Suppose that the set $S := \{x \in K : h(x) \leq 0\}$ is nonempty. Then there exists $\tau > 0$ such that

$$d(x, S) \leq \tau h^+(x) \text{ for all } x \in K.$$
Let us end this subsection with two other error bounds resulting from Theorem ???. The first one involves the convex quadratic constraints

\[ g_i(x) := \frac{1}{2} \langle Q_i x, x \rangle + \langle q_i, x \rangle + r_i, \quad i = 1, \ldots, m \]  

(82)

where \( q_i \in \mathbb{R}^n \), \( r_i \in \mathbb{R} \) and \( Q_i \) are positive semidefinite symmetric \( n \times n \) matrices. Let \( K \) be a nonempty bounded polyhedral convex set in \( \mathbb{R}^n \).

Consider the following inequality system:

\[ S = \{ x \in K : g_i(x) \leq 0, \ h_p(x) \leq 0, \ (i, p) \in \{1, \ldots, m\} \times \{1, \ldots, \ell\} \} , \]

where, \( h_p \) are functions of the same type as in Theorem 28, i.e.,

\[ h_p(x) := \sum_{j \in \mathcal{S}_p} \min\{ h_{ij}(x) : i \in \mathcal{S}_{j,p} \} , \]
where, $\mathcal{S}_p$, $\mathcal{S}_{j,p}$ are finite index sets and $h_{ij}$ are differentiable concave functions on $K$. Set

$$S_1 = \{ x \in K : g_i(x) \leq 0, \ i = 1, ..., m \}.$$  \hfill (83)

Theorem 28 yields the following result:

**Corollary 29**

Let $g_i$, $h_p$ be defined as above. In addition, suppose that $S$ is nonempty, $g_i, \ (i = 1, ..., m)$, are nonnegative on $K$ and $h_p, \ (p = 1, ..., \ell)$, are nonnegative on $S_1$. Then there exists $\tau > 0$ such that

$$d(x, S) \leq \tau \sum_{p=1}^{l} h_p(x) \text{ for all } x \in S_1.$$
Finally the second error bound derived from Theorem 28 is related to finite feasible sets

**Corollary 30**

Let $K$ be a nonempty bounded polyhedral convex set and let $h$ be as in Theorem 28. Assume that the set $S := \{x \in K : h(x) \leq 0\}$ has finitely many elements. Then for all closed set $C \subset K$ such that $S' = \{x \in C : h(x) \leq 0\}$ is nonempty, there exists $\tau > 0$ such that

$$d(x, S') \leq \tau h^+(x) \quad \text{for all} \quad x \in C.$$
Remark 31

It is easy to see that the assumption on the finiteness of the set $S$ is satisfied if $h$ is strictly concave and nonnegative on $K$. For example, consider the feasible set in $0-1$ programming:

$$S := \{ x \in C : x \in \{0, 1\}^n \}$$

where $C$ is a closed subset of $\mathbb{R}^n$. One can rewrite $S = \{ x \in C \cap [0, 1]^n : h(x) \leq 0 \}$ where $h(x) := \langle e, x \rangle - \|x\|_2^2$, and $e \in \mathbb{R}^n$ is the vector of ones. Obviously, the assumptions of Corollary 30 are satisfied. Note that the concave (but not strictly) penalty function

$$\bar{h}(x) = \sum_{i=1}^{n} \min\{x_i, 1 - x_i\}$$

also verifies the assumptions of Corollary 30.
Let us end the paper with new key results on exact penalty in mixed integer DC programming [Le Thi, Pham Dinh, 2012]

Consider the class of mixed integer DC programs

\[
\alpha := \inf \{ f(x, y) - h(x, y) : (x, y) \in K, x \in [l, u] \cap \mathbb{Z}^n \} \tag{84}
\]

where \(K\) is a bounded polyhedral convex set in \(\mathbb{R}^{n+p}\) :

\[K := \{ (x, y) \in \mathbb{R}^{n+p} : C(x, y) := Ax + By \leq b \} \] with \(C \in \mathbb{R}^{m \times (n+p)}, A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times p}, [l, u] := \prod_{i=1}^{n} [l_i, u_i] \subset \mathbb{R}^n\), with \(l_i, u_i \in \mathbb{Z}\),

\(l_i < u_i\) for \(i = 1, \ldots, n\).
This challenging class of nonconvex programs encompasses most combinatorial optimization problems, in particular usual mixed integer linear/quadratic programming problems. To our best knowledge, our exact penalty results concerning (84) are the first ones to date. Let us first summarize the penalty functions $p(x, y)$ used to penalize the integer variables $x$.

1) $p_1(x, y) := \sum_{j=1}^{n} \sin^2(\pi x_j)$ has been already used in some earlier works.

2) $p_2(x, y) := d_2^2(x, [l, u] \cap \mathbb{Z}^n) = \min\{\|x - z\|^2_2 : z \in [l, u] \cap \mathbb{Z}^n\} = \sum_{j=1}^{n} \min\{(x_j - z_j)^2 : z_j \in [l_j, u_j] \cap \mathbb{Z}\}$ is piecewise convex.

3) $p_3(x, y) := \sum_{j=1}^{n} |\sin \pi x_j|$ is piecewise concave

4) $p_4(x, y) := d_1(x, [l, u] \cap \mathbb{Z}^n) = \min\{\|x - z\|_1 : z \in [l, u] \cap \mathbb{Z}^n\}$
\[
= \sum_{j=1}^{n} \min \{ |x_j - z_j| : z_j \in [l_j, u_j] \cap \mathbb{Z} \} \text{ is piecewise concave.}
\]

5) \( p_5(x, y) = p_5(x) := \sum_{j=1}^{n} p_5^j(x_j), \) with

\[
p_5^j(x_j) := \max\{[x_j - (l_j + k)][(l_j + k + 1) - x_j] : k = 0, \ldots, (u_j - l_j) - 1\}
\]

is piecewise concave.

The five penalty functions are DC functions with explicit DC decompositions. By using Theorem 28, we have proved the exact penalty for the last three penalty functions [Le Thi, Pham Dinh, 2012]. Hence there is equivalence between (84) and its penalized by either of them.
Thank you for your attention!