DC programming

Frédéric Boileau

Contents

1 Introduction

1.1 TODO

1. Analysis of DCA, rate, stability, optimality, local vs global
2. specific cases (Generalized Fermat-Weber)
3. Using relative interior notions define other optimality criteria, "add structure for finite dimensional case
4. Analysis of the sets DC and DCf, starting with list of elementary options under which they are closed and after more subtle limiting arguments, e.g. smooth functions are DC
5. Explorations: Hilbert spaces, proximal methods, accelerated descent methods etc.

In this paper we investigate a type of non-convex program which allows us to leverage the tools from convex analysis in an elegant manner and which has many direct concrete applications. We are concerned with the minimization of the difference of convex functions and we use the following notation throughout:

$$\lambda := \min_{x \in \mathbb{E}} \{g(x) - h(x)\} \tag{1}$$

We usually assume that both g and h are convex, proper and lower continuous unless otherwise specified so.

2 Preliminaries

We define a couple of key concepts. First for exposition we use the following function notation throughout:

$$f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$$

We also use Rockafellar's notation (Rockafellar and Wets 2009) for sequences of real variables : x^ν where ν is always understood to be a natural number used as an index and not an exponent.

Definition 1. A function's *domain* is the subset of the initial set of the function where it doesn't attain infinity. Rockafellar and other's sometimes refer to this set as the *effective* domain to emphasize the difference with the naive set theory definition. However we will always refer to the domain in the first sense and will dispense from using the adjective "effective". Formally:

$$\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$$

Definition 2. A convex function is said to be *proper* if it has non empty domain and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. The set of all such functions is denoted Γ.

In optimisation smoothness assumptions often break-down. We thus need a different machinery than the one obtained from classical calculus. We introduce some of those notions now.

Definition 3. We define present the definition of the lower limit of a function f at some $\bar{x} \in \mathbb{R}^n$ as it is written by Rockafellar and Wets (Rockafellar and Wets 2009)

$$\begin{aligned} \liminf_{x \to \bar{x}} f(x) :&= \lim_{\delta \searrow 0} \left[\inf_{x \in B(\bar{x},\delta)} f(x) \right] \\ &= \sup_{\delta > 0} \left[\inf_{x \in B(\bar{x},\delta)} f(x) \right] \quad = \sup_{V \in \mathcal{N}(\bar{x})} \left[\inf_{x \in V} f(x) \right] \end{aligned}$$

Proposition 1.
(Rockafellar and Wets 2009)
The above definition of lower limit of a function is equivalent to

$$\liminf_{x \to \bar{x}} f(x) = \min\{\alpha \in \overline{\mathbb{R}} \mid \exists x^\nu \to \bar{x} \textit{ with } f(x^\nu) \to \alpha\}$$

We use min on the rhs as the definition presumes that the value is actually attained.

Definition 4. The concept of lower-limit allows us to talk about semicontinuity. A function f is said to be *lower semicontinuous* (lsc) at $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x}) \qquad \text{or equivalently} \qquad \liminf_{x \to \bar{x}} f(x) = f(\bar{x})$$

Where the equivalence comes from the fact that the reverse inequality always holds by definition of $\liminf$

Definition 5. Recalling the definition of proper convex functions; when $f \in \Gamma$ and is lsc we say that $f \in \Gamma_0$ which is set the of all such functions.

Those sets being of major importance we restate, more formally:

$$\Gamma := \{f : \mathbb{R}^n \longrightarrow \mathbb{R} \mid f \text{ proper and convex}\} \qquad \Gamma_0 := \{f : \mathbb{R}^n \longrightarrow \mathbb{R} \mid f \text{ proper , lsc and convex}\}$$

Definition 6. Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be convex and $x \in \mathbb{R}^n$. Then $g \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + \langle g, x - \bar{x} \rangle \quad \forall x \in \mathbb{R}^n$$

Moreover the set of all such subgradients of f at $\bar{x}$ is callled the subdifferential of f at $\bar{x}$ and we denote it as $\partial f(\bar{x})$.

Definition 7. The Fenchel conjugate of a function f, in the convex analysis setting often just referred to as the conjugate of a function, is an extended real valued functione defined as:

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

The mapping $f \mapsto f^*$ is called the Legendre-Fenchel transform. It is sometimes just refered to as the Legendre transform in physics which doesn't do justice to Fenchel's important work in the non-differentiable and variational case.

By definition we always have that

$$f(x) + f^*(y) \geq \langle x, y \rangle \quad \forall\ x, y \in \mathbb{R}^n$$

This relationship is known as the *Fenchel-Young Inequality*

Lemma 1.
Let $f \in \Gamma$, *then* $y \in \partial f(x) \iff f(x) + f^*(y) = \langle x, y \rangle$

Proof. Since we already have the Fenchel-Young inequality we only need the reverse:

$$\begin{aligned} y \in \partial f(x) &\overset{\text{def}}{\iff} f(z) \geq f(x) + \langle y, z - x \rangle \qquad \forall\, z \in \mathbb{R}^n \\ &\iff f(x) + \sup_z \{\langle y, z \rangle - f(z)\} \leq \langle x, y \rangle \\ &\iff f(x) + f^*(y) \leq \langle x, y \rangle \end{aligned}$$

□

The ϵ-*subdifferential* of a function. is a tool which has been mostly developped by J.B. Hiriart-Urruty in the second of his two volumes on convex analysis (Hiriart-Urruty and Lemaréchal 1993, p. 92)

Definition 8. Given $x \in \operatorname{dom} f$ the vector $s \in \mathbb{R}$ is called an $\epsilon - subgradient$ of f at x, written $s \in \partial_\epsilon f(x)$ when the following holds

$$f(y) \geq f(x) + \langle s, y - x \rangle - \epsilon \quad \forall\ y \in \mathbb{R}^n$$

The set of all ϵ-subgradients of a function at some point $x \in \mathbb{R}^n$ is called the ϵ-subdifferential of f at x. Moreover it is clear from the definition that $\partial f(x) = \bigcap_{\epsilon > 0} \partial_\epsilon f(x)$

We have an equivalent to the Fenchel-Young inequality for the $\epsilon-$ subdifferentials. We present it as a lemma here:

Lemma 2.
Let $f \in \Gamma$ and $x^o \in \operatorname{dom} f$ then we have that $y \in \partial_\epsilon f(x^o)$ if and only if $f(x^o) + f^*(y) \leq \langle x^o, y \rangle + \epsilon$

Proof.

$$\begin{aligned} y \in \partial_\epsilon f(x^o) &\iff f(x) \geq f(x^o) + \langle y, x - x^o \rangle - \epsilon \qquad x \in \mathbb{R}^n \\ &\iff \epsilon + \langle y, x^o \rangle \geq f(x^o) + \underbrace{\sup_x \{\langle x, y \rangle - f(x)\}}_{f^*(y)} \qquad \iff \epsilon + \langle y, x^o \rangle \geq f(x^o) + f^*(y) \end{aligned}$$

$\square$

3 Duality and Optimality

3.1 Toland-Singer duality

Let us recall the definition of a DC program and its associated notation.

$$\lambda := \min_{x \in \mathbb{E}} \{g(x) - h(x)\} \qquad g, h \in \Gamma_0$$

Let $f := g - h$ be the objective function of the above. We write $\mathcal{DC}$ for the set of all such objective functions and $\mathcal{DC}_f$ when they are finite. TODO: DCf closed under

With the DC program defined the first important result to establish is duality. It turns out there is a very natural dual to P that is due to Toland and Singer. Letting our primal problem as defined previously:

$$P \qquad \lambda \;:= \inf_{x \in \mathbb{E}} \{g(x) - h(x)\} \qquad \mathcal{P} := \operatorname{argmin}_{x \in \mathbb{E}} \{g(x) - h(x)\} \qquad g, h \in \Gamma_0$$

We define the dual as follows:

$$D \qquad \lambda^* := \inf_{y \in \mathbb{E}} \{h^*(x) - g^*(x)\} \qquad \mathcal{D} := \operatorname{argmin}_{y \in \mathbb{E}} \{h^*(y) - g^*(y)\} \qquad g, h \in \Gamma_0$$

Letting once again $f := g - h$ we write $f^\dagger$ for $h^* - g^*$ and refer to that function as the DC-dual of f. Note that in order to have a solution to the program (minimizer where the objective value is not ∞) we clearly need $x \in \operatorname{dom} g$, as even if we have the pathological case that both $x \notin \operatorname{dom} g$ and $x \notin \operatorname{dom} h$ then $f(x) = \infty - \infty = +\infty$ under the inf-addition convention. *Hence we assume $x \in \operatorname{dom} g$ unless otherwise specified.* We know present the first and probably most important result on DC programming due to Toland-Singer which establishes how a DC program is related to its DC-dual. The proof we present is from Hoheisel's notes on convex analysis. (Hoheisel 2016).

Theorem 3.

a)Let λ and its dual be defined as above, we have that $\lambda = \lambda^*$

b)Moreover we have that given an $\bar{x} \in \mathcal{P}$ every $y \in \partial h(\bar{x})$ is in $\mathcal{D}$.

Proof. a) Since $h \in \Gamma_0$ we directly have that h is equal to its biconjugate hence

$$\begin{aligned}\inf_x\{g(x) - h(x)\} &= \inf_x\{g(x) - h^{**}(x)\}\\ &= \inf_x\{g(x) - (h^*)^*(x)\}\\ &= \inf_x\left\{g(x) - \sup_y\{\langle x, y\rangle - h^*(y)\}\right\}\end{aligned}$$

Now let us note that

$$\inf_x\left\{g(x) - \{\langle x, y\rangle - h^*(y)\}\right\} < +\infty \iff y \in \operatorname{dom} h^* \tag{2}$$

$$\begin{aligned}\inf_x\{g(x) - h(x)\} &= \inf_x\left\{g(x) + \inf_y\{h^*(y) - \langle x, y\rangle\}\right\}\\ &= \inf_x\inf_y\left\{h^*(y) - \langle x, y\rangle + g(x)\right\}\\ &= \inf_y\left\{h^*(y) - \sup_x\{\langle x, y\rangle - g(x)\}\right\}\\ &= \inf_y\{h^*(y) - g^*(y)\}\end{aligned}$$

Hence the finiteness of λ boils down to

$$\operatorname{dom} g \subset \operatorname{dom} h \qquad \text{and} \qquad \operatorname{dom} h^* \subset \operatorname{dom} g^* \tag{3}$$

We thus assume 3 holds throughout the paper. Let us continue the computations, now anchored in the finite setting : b) By assumption $\bar{x} \in \mathcal{P}$ hence by definition:

$$g(\bar{x}) - h(\bar{x}) \leq g(x) - h(x)$$

Under the assumption that $x \in \operatorname{dom} g$ and the inclusions of 3 all terms are finite and we can re-arrange:

$$g(x) - g(\bar{x}) \geq h(x) - h(\bar{x})$$

Note that this would still be valid in the case where either just the LHS or both sides were infinite valued under the inf-addition convention; which is actually the main motivation behind said convention. Now using the assumption that some $\bar{y}$ is in the range of $\partial h(\bar{x})$ and the definition of the subdifferential

$$g(x) - g(\bar{x}) \geq \langle \bar{y}, x - \bar{x}\rangle \quad \Rightarrow \quad \bar{y} \in \partial g(\bar{x}) \quad (\text{and } y \in \partial h(\bar{x}) \text{ by assumption})$$

Recall that whenever $\bar{z} \in \partial f(z)$ for some $f \in \Gamma$, lemma 1 gives us the following equality $f(z) + f^\dagger(\bar{z}) = \langle z, \bar{z}\rangle$. This leads to the following observation: the point $\bar{y}$ being in the range of the subdifferentials of both g and h "binds" the value of the DC program and its DC-Dual.

$$g(\bar{x}) + g^*(y) = h(\bar{x}) + h^*(\bar{y}) = \langle \bar{x}, \bar{y}\rangle$$

Where the inner product is less relevant here than the fact that the value agree. Hence all in all we get :

$$g(\bar{x}) - h(\bar{x}) = h^*(\bar{y}) - g^*(\bar{y})$$

□

Note that the assumption on $g \in \Gamma_0$ is unnecessarily strong and the two previous proofs work by replacing it by any proper extended real valued function. However if we want a symmetric result, that is $\bar{y} \in \mathcal{D}$ and $\bar{x} \in \partial g^*(\bar{y})$ imply $\bar{x} \in \mathcal{P}$, then we do need g to be l.s.c. and the proof is essentially the same as the one above by symmetry.

3.2 Optimality Conditions

TODO:Global optimality criteria with relative interiors (Pham Dinh Tao et al. 2005)

As mentionned previously global optimality is not within our reach in practical applications at this point and we usually have to resort to use local optimality criteria to derive algorithms eventhough they might in fact converge to a global solution. I now present different theorems and conditions for global optimality. To make sure both f and its DC-dual are finite valued we assume the following:

$$\operatorname{dom} g \subset \operatorname{dom} h \qquad \text{and} \qquad \operatorname{dom} h^* \subset \operatorname{dom} g^* \tag{4}$$

Theorem 4 (Characterization of global minima).
A point $\bar{x} \in \mathbb{R}^n$ is a solution to the primal problem, i.e. $\bar{x} \in \mathcal{P}$, if and only if $\partial_\epsilon h(\bar{x}) \subset \partial_\epsilon g(\bar{x}) \quad \forall \epsilon > 0$

Proof. Assume $x \in \operatorname{dom} g$ and $y \in \operatorname{dom} h^*$ for finiteness. Let $\bar{x} \in \mathcal{P}$. Since by DC-duality the primal and dual optimal values agree we have that $f(\bar{x}) \leq f^\dagger(y)$, hence :

$$g(\bar{x}) - h(\bar{x}) \ \leq h^*(y) - g^*(y) \tag{5}$$

$$\iff \qquad g(\bar{x}) + g^*(y) \leq h(\bar{x}) + h^*(y) \tag{6}$$

Now let $\bar{x} \in \mathcal{P}$, from Fenchel-Young inequality we have:

$$\langle \bar{x}, y \rangle \leq g(\bar{x}) + g^*(x) \leq h(\bar{x}) + h^*(x) \qquad \forall x \in \operatorname{dom} g \tag{7}$$

Then if $x \in \partial_\epsilon h(\bar{x})$ we have from "generalized Fenchel-Young" inequality (Lemma 2) that:

$$h(\bar{x}) + h^*(x) \leq \epsilon + \langle x, \bar{x} \rangle$$

Then clearly if we want (6) to potentially hold we need to have $g(\bar{x}) + g^*(x) \leq \epsilon + \langle x, \bar{x} \rangle$ as well. Hence $x \in \partial_\epsilon g(\bar{x})$ whenever $x \in \partial_\epsilon h(\bar{x})$ which shows the inclusion is necessary. For sufficiency we argue by contraposition. Assume (6) does not hold. Then by Fenchel-Young there exists some $\epsilon > 0$ such that

$$\langle x, \bar{x} \rangle \leq h(\bar{x}) + h^*(x) < \langle x, \bar{x} \rangle + \epsilon \leq g(\bar{x}) + g^*(x) \qquad \forall x \in \operatorname{dom} g$$

But then once again by lemma 2 we have that $x \in \partial_\epsilon h(x)$ and $x \notin \partial_\epsilon g(x)$ $\square$

An important remark is in order, notationnally the usage of ϵ has a heavy connotation as we expect to let it go to zero or to use it to get the strongest estimate possible. However here letting it to zero yields the much weaker condition $\partial h(x) \subset \partial g(x)$ The later actually yields a condition which while not sufficient is clearly *necessary* for global optimality.The strength of the condition imposed with the $\epsilon-$subdifferential for all $\epsilon > 0$ lies in the fact that we can also make ϵ as large as we'd like. This necessary and sufficient condition is in fact too strong and impractical as are gobal optimality conditions in general. Now I recall the property from the proof DC-duality that

$$\forall \bar{x} \in \mathcal{P} \quad \partial h(\bar{x}) \subset \mathcal{D} \subset \operatorname{dom} h^* \qquad \forall \bar{y} \in \mathcal{D} \quad \partial g^*(\bar{y}) \subset \mathcal{P} \subset \operatorname{dom} g \tag{8}$$

This property is commonly known as *transportation of global minimizers* as subdifferentials map minimizers between primal and dual spaces. This is a very good guide for our intuition as I will prove the next couple results about *local optimality* which are at the core of the DCA. First we introduce the following notation due to Tao and Souad (P. Tao and Souad 1988, p. 280):

$$\mathcal{P}_l := \{\bar{x} \in \mathbb{R}^n \mid \partial h(\bar{x}) \subset \partial g(\bar{x})\} \qquad \mathcal{D}_l := \{\bar{y} \in \mathbb{R}^n \mid \partial g^*(\bar{y}) \subset \partial h^*(\bar{y})\} \tag{9}$$

Theorem 5 (Necessary condition for local optimality)**.**
If $\bar{x}$ is a local minimizer of $f := g - h$ then $\bar{x} \in \mathcal{P}_l$

Proof. If $\bar{x}$ is a local minimizer of f then by definition there exists a neighborhood of $\bar{x}$, say $N(\bar{x})$ such that

$$g(x) - h(x) \geq g(\bar{x}) - h(\bar{x}) \quad \forall x \in N(\bar{x}) \tag{10}$$

Since by assumption g is proper we can take the intersection of the neighborhood with $\operatorname{dom} g$ without yielding an empty set. Taking this intersection forces g and h to be finite valued [a] so that we can rearrange the former inequatlity:

$$g(x) - g(\bar{x}) \geq h(x) - h(\bar{x}) \quad \forall x \in N(\bar{x}) \cap \operatorname{dom} g \tag{11}$$

Now for any $\bar{y} \in \partial h(\bar{x})$ we have that

$$h(x) - h(\bar{x}) \geq \langle x - \bar{x}, \bar{y} \rangle \qquad x \in \mathbb{R}^n \tag{12}$$

Combining (11) and (12) we get that given some $\bar{y} \in \partial h(\bar{x})$:

$$g(x) - g(\bar{x}) \geq \langle x - \bar{x}, \bar{y} \rangle \quad \forall\ \bar{x} \in N(\bar{x}) \cap \operatorname{dom} g \tag{13}$$

Note that the set $N(\bar{x}) \cap \operatorname{dom} g$ might not be open for an abritrary neighborhood however we can clearly make $N(\bar{x})$ small enough so that its intersection with $\operatorname{dom} g$ is open. Finally proving the inequality (13) on some neighborhood for a convex function actually implies it holds for all of $\mathbb{R}^n$ (Lemma) hence :

$$g(x) - g(\bar{x}) \geq \langle x - \bar{x}, \bar{y} \rangle \quad \forall\ \bar{x} \in \mathbb{R}^n$$
$$\therefore\ \bar{y} \in \partial g(\bar{x})$$

All in all we have just proven that local minimality implies the desired inclusion, i.e. that $\bar{x} \in \mathcal{P}_l$. □

[a] Recall that we always assume $\operatorname{dom} g \subset \operatorname{dom} h$

Definition 9 (Critical Point)**.** A point $\bar{x}$ is said to be a *critical point of* $g - h$ if

$$\partial g(\bar{x}) \cap \partial h(\bar{x}) \neq \emptyset$$

We now proceed to show a sufficient condition for local optimality. We first present the main result and then a more elegant corollary to test local optimality. Both results are drawn from Tao and An's paper (Pham Dinh Tao and An 1997). First we need a critical point and then explore the behaviour of the subdifferentials of g and h around it. In the following proof we restrict the neighborhood to the domain of the subdifferential of h as this mapping allows us to exploit the dual structure of DC programs.

Theorem 6 (Sufficient local optimality condition)**.**
Let $\bar{x}$ be a critical point of $f := g - h$, $\bar{y} \in \partial g(\bar{x}) \cap \partial h(\bar{x})$ and $\widetilde{N}(\bar{x}) := N(\bar{x}) \cap \operatorname{dom} \partial h$. Then, given the definitions above, we have that if

$$\forall x \in \widetilde{N}(\bar{x}) \quad \exists\, \hat{y} \in \partial h(x) \quad s.t. \quad f^{\dagger}(\hat{y}) \geq f^{\dagger}(\bar{y}) \tag{14}$$

Then $\bar{x}$ is a local minimizer on $\widetilde{N}(\bar{x})$

Proof. Once again taking the intersection of both subdifferentials allows us to bind the primal and dual problems.

$$\bar{y} \in \partial g(\bar{x}) \cap h(\bar{x}) \;\Rightarrow g(\bar{x}) + g^*(\bar{y}) = \langle \bar{x}, \bar{y} \rangle = h(\bar{x}) + h^*(\bar{y})$$

Hence we get the equality of the primal and dual objectives

$$g(\bar{x}) - h(\bar{x}) = h^*(\bar{y}) - g^*(\bar{y}) \tag{15}$$

Moreover we have that $\forall x \in \widetilde{N}(\bar{x})$ there is $\hat{y} \in \partial h(x)$ which majorizes the dual at $\bar{y}$, i.e.

$$h^*(\hat{y}) - g^*(\hat{y}) \geq h^*(\bar{y}) - g^*(\bar{y}) \qquad \text{In other words} \quad f^{\dagger}(\hat{y}) \geq f^{\dagger}(\bar{y}) \tag{16}$$

To complete the proof we need one last inequality which is direct from the Lemma 1 (as $\hat{y} \in \partial h(x)$) and the definition of the Fenchel dual of a function:

$$h(x) + h^*(\hat{y}) = \langle x, \hat{y} \rangle \leq g(x) + g^*(\hat{y}) \;\Rightarrow\; g(x) - h(x) \geq h^*(\hat{y}) - g^*(\hat{y}) \tag{17}$$

Combining (15), (16) and (17) we get the desired inequality:

$$g(x) - h(x) \geq g(\bar{x}) - h(\bar{x}) \qquad \forall x \in \widetilde{N}(\bar{x})$$

□

We can now leverage this rather technical result to get the following more elegant corollary which follows quite nicely. The idea however is quite in the same vein. For a critical point $\bar{x}$ we need a neighborhood around it that is contained in the subdifferential of g at $\bar{x}$ and such that every point of the neighborhood is a fixed point under the subdifferential of h.

Corollary 7 (sufficient local optimality condition)**.**
Let $\bar{x}$ be a critical point of ($\mathcal{DC} \ni$) $f := g - h$. If we can find a neighborhood $\widetilde{N}(\bar{x}) := N(\bar{x}) \cap \operatorname{dom} \partial h(x)$ such that

$$\forall x \in \widetilde{N}(\bar{x}) \qquad \partial g(\bar{x}) \cap \partial h(x) \neq \emptyset \tag{18}$$

then $\bar{x}$ is a local minimizer of f

Proof. Let $x \in \widetilde{N}(\bar{x})$ and $y \in \partial h(x) \ \cap \partial g(\bar{x})$

First we consider $\bar{y} \in \partial h(\bar{x}) \cap \partial g(\bar{x})$ and as before it binds dual and primal

$$g(\bar{x}) + g^*(\bar{y}) = \langle \bar{x}, \bar{y} \rangle = h(\bar{x}) + h^*(\bar{y}) \ \Rightarrow f(\bar{x}) = f^\dagger(\bar{y})$$

Now we have two inclusions which give us two inequalities, first $y \in \partial h(x)$ yields

$$h(x) + h^*(y) = \langle x, y \rangle \leq g(x) + g^*(y) \ \Rightarrow \ f(x) \geq f^\dagger(y)$$

And $y \in \partial g(\bar{x})$ hence

$$g(\bar{x}) + g^*(y) = \langle \bar{x}, y \rangle \leq h(\bar{x}) + h^*(y) \ \Rightarrow \ f^\dagger(y) \geq f(\bar{x})$$

Hence combining the three relationships above we get

$$f(x) \geq f(\bar{x}) \quad \forall x \in \widetilde{N}(\bar{x}) \qquad (:= N(\bar{x}) \cap \operatorname{dom} \partial h(x))$$

□

For both the theorem and its corollary the process of restricting the neighborhood to the domain of the subdifferential of h allows us to get satisfying inequalities in the dual space which can be linked elegantly to the primal. The corollary requires stronger assumptions as the neighborhood is not only restrained to $\operatorname{dom} \partial h(x)$ but also to the subdifferential of g at the critical point.

4 DCA

4.1 Intuition and Derivation of simplified DCA

We now have the technical results necessary to discuss the DC algorithm for general D.C. programs. Its aim is to generate sequences x^ν and y^ν which converge to a local minimum of the functions $f := g - h$ and $f^\dagger := h^* - g^*$ respectively where $g, h \in \Gamma_0$
We define two sub programs:

$$\widetilde{S}(x) \qquad \inf\{h^*(y) - g^*(y) \ : \ y \in \partial h(x)\} \tag{19}$$

$$\widetilde{T}(y) \qquad \inf\{g(x) - h(x) \ : \ x \in \partial g^*(y)\} \tag{20}$$

Let $\widetilde{\mathcal{S}}(x)$ and $\widetilde{\mathcal{T}}(y)$ denote the solution sets of $\widetilde{S}(x)$ and $\widetilde{T}(y)$ respectively. Then given a point $x^o \in \operatorname{dom} g$ the sequences are constructed as follows:

$$y^\nu \in \widetilde{\mathcal{S}}(x^\nu); \quad x^{\nu+1} \in \widetilde{\mathcal{T}}(y^\nu) \tag{21}$$

The main idea relies on a decomposition approach to the problem; we use DC duality to solve the initial program and its dual on subsets of the initial space using the subdifferential operators to move between primal and dual. Eventhough the the subprograms are simpler than the initial one they are still non-convex. Their main function is to give insights as to how to construct sequences with the desired behaviour.

Consider the program $T(y^\nu)$ where y^ν is the solution to some $\widetilde{S}(x^o)$. We have that $y^\nu \in \partial h(x^\nu)$ hence:

$$h(x^\nu) \leq h(x^{\nu+1}) + \langle y, x - x^{\nu+1} \rangle \tag{22}$$

This suggests a simplfication of the program $\widetilde{T}(y^\nu)$ by replacing $h(x^\nu)$ by its affine minorization through y^ν whereby we get the following program, which is convex:

$$\inf\{g(x) - [h(x^\nu) + \langle y^\nu, x - x^\nu \rangle]\} \tag{23}$$

Dually we can also replace g^* in the objective function of $\widetilde{S}(x^{\nu+1})$ by its affine minorant since we have that $x^{\nu+1} \in \partial g^*(y^\nu)$:

$$\inf\{h^*(y) - [g^*(y^\nu) + \langle x^{\nu+1}, y - y^\nu \rangle]\} \tag{24}$$

Note that we only needed the inclusion $x^{\nu+1} \in \partial g^*(y^\nu)$ and $y^\nu \in \partial h(x^\nu)$ to find a simplified versions of the two initial programs.

The programs (23) and (24) can actually be simplified further as the objective functions contain constants.

$$\operatorname{argmin}\{h^*(y) - [g^*(y^\nu) + \langle x^{\nu+1}, y - y^\nu\rangle]\} = \operatorname{argmin}\{h^*(y) - \langle x^{\nu+1}, y\rangle\} \tag{25}$$

Now note that

$$y^{\nu+1} \in \operatorname{argmin}\{h^*(y) - \langle x^{\nu+1}, y\rangle\} \iff y^{\nu+1} \in \partial h(x^{\nu+1}) \tag{26}$$

Similarly

$$\operatorname{argmin}\{g(x) - [h(x^\nu) + \langle y^\nu, x - x^\nu\rangle]\} = \operatorname{argmin}\{g(x) - \langle x, y^\nu\rangle\} \tag{27}$$

$$x^{\nu+1} \in \operatorname{argmin}\{g(x) - \langle x, y^\nu\rangle\} \iff x^{\nu+1} \in \partial g^* y^\nu) \tag{28}$$

References

Hiriart-Urruty, Jean-Baptiste and Claude Lemaréchal (1993). *Convex analysis and minimization algorithms II: Advanced theory and bundle methods, vol. 306 of Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, New York.

Hoheisel, Tim (2016). "Convex Analysis, University of Würzburg, Germany, Lecture Notes, Summer term 2016".

Rockafellar, R Tyrrell and Roger J-B Wets (2009). *Variational analysis.* Vol. 317. Springer Science & Business Media.

Tao, PD and EB Souad (1988). "Duality in DC optimization. Subgradient methods, Trends in mathematical optimization". In: *Internat. Ser. Numer. Math* 84, pp. 277–293.

Tao, Pham Dinh et al. (2005). "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems". In: *Annals of operations research* 133.1-4, pp. 23–46.

Tao, Pham Dinh and Le Thi Hoai An (1997). "Convex analysis approach to dc programming: Theory, algorithms and applications". In: *Acta Mathematica Vietnamica* 22.1, pp. 289–355.

A authors

Pham dih tao

El Bernoussi

JB Hiriart-Urruty

R. Ellaia

JF Tolland (1979 important paper)