On DC and Local DC Functions

Liam Jemison
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/honorstheses

Part of the Mathematics Commons

Let us know how access to this document benefits you.

Recommended Citation
https://doi.org/10.15760/honors.931

This Thesis is brought to you for free and open access. It has been accepted for inclusion in University Honors Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
On DC and Local DC Functions

Liam Jemison

An undergraduate honors thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in University Honors and Mathematics

Thesis Advisor
Dr Mau Nam Nguyen
Portland State University
2020
1 Introduction

Convex analysis and convex optimization are well studied in the literature. Convex functions and convex sets are favorable for efficient optimization techniques and numerical algorithms that can be effectively implemented by programming languages such as Matlab and Python. However, optimization problems involving nonconvex functions also appear frequently and naturally in many applications. A straightforward example of such a problem is minimizing a quadratic form on a Euclidean sphere, given by Tao and An in [7]. Thus, it is desirable to study nonconvex functions and sets as well as to develop numerical algorithms for solving nonconvex optimization problems.

By considering the class of functions which can be represented as differences of convex (DC) functions, it is possible to develop optimization techniques for minimizing a broader class of functions, while taking advantage of available tools from convex analysis and optimization. We refer the reader to one of the very first papers on DC functions written by Hartman [3] presenting many nice algebraic and analytical properties of this class of functions. By using some elements of convex analysis, the DC Algorithm (DCA) was also introduced to minimize DC functions, that are not necessarily differentiable; see [7] and the references therein.

In this project, we study in some important known properties of DC functions. Our main goal is to provide detailed proofs for many existing results involving the class of DC functions in [3,5,8]. We lavish particular attention on the proof of the mixing lemma from [8] and the relationship between the class of DC functions and the class of local DC functions from [3,5]. We also study the DC Algorithm for optimizing DC functions and apply it to the support vector machine problem, contrasting its performance with other traditional convex optimization techniques.

We restrict our attention to real-valued functions defined on $\mathbb{R}^n$, and we denote the standard Euclidean norm in $\mathbb{R}^n$ by $\| \cdot \|$. However, the reader should be aware that many, although not all, of the results presented hold in more general settings as well. Throughout the thesis, $\langle \cdot, \cdot \rangle$ denotes the standard inner product; $B(x;r)$ and $B(x;r)$ denote the closed and open balls with center at $x$ and radius $r > 0$, respectively.

2 Preliminaries on DC Functions

We use many basic results from convex analysis, for which we generally cite [4], although there are undoubtedly many other places these well-known results can be found. We will define convex functions using Jensen’s inequality.

Definition 2.1. Suppose $A$ is a nonempty convex subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a
function such that for all \( x_1, x_2 \in \mathbb{R}^n \) and \( 0 < \lambda < 1 \) we have
\[
f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).
\]

Then we say \( f \) is \textit{convex}.

\textbf{Definition 2.2.} Let \( F : A \to \mathbb{R} \) be a function such that \( F = f_1 - f_2 \) where \( f_1, f_2 : A \to \mathbb{R} \) are convex functions and \( A \) is a nonempty convex subset of \( \mathbb{R}^n \). Then we will say \( F \) is a \textit{DC function} and we refer to the expression \( F = f_1 - f_2 \) as a \textit{DC decomposition} of \( F \).

Clearly any convex function is DC. In the next section we will give some examples to show the breadth of this class of functions. For simplicity, the remainder of our results will concern functions whose domain is the whole space \( \mathbb{R}^n \).

A given DC function will have many DC decompositions. For instance, we have the following result from [1, 5]:

\textbf{Proposition 2.3.} If \( F : \mathbb{R}^n \to \mathbb{R} \) is DC, then \( F \) can be decomposed by \( F = g_1 - g_2 \) where \( g_1, g_2 : \mathbb{R}^n \to \mathbb{R} \) are convex and nonnegative.

\textit{Proof.} Suppose \( f \) has a DC decomposition \( F = f_1 - f_2 \). Since \( f_1, f_2 : \mathbb{R}^n \to \mathbb{R} \) are convex, we can find affine functions \( a_1, a_2 : \mathbb{R}^n \to \mathbb{R} \) with \( f_1 \geq a_1 \) and \( f_2 \geq a_2 \) [4, Theorem 2.1.2]. Define \( a = a_1 - a_2 \). Then \( F = (f_1 - a_1) - (f_2 - a_2) + a \). Note that \( a \) is affine, and if we define \( a^+ = \max\{a, 0\} \) and \( a^- = \max\{-a, 0\} \), then \( a^+, a^- \) are convex and \( a = a^+ - a^- \). So then \( F = (f_1 - a_1 + a^+) - (f_2 - a_2 + a^-) \), where both \( (f_1 - a_1 + a^+) \) and \( (f_2 - a_2 + a^-) \) are nonnegative convex functions. \( \square \)

In the next proposition we give some equivalent definitions of DC functions that can be convenient.

\textbf{Proposition 2.4.} Suppose that \( F : \mathbb{R}^n \to \mathbb{R} \) is a continuous function. The following are equivalent:

(a) \( F \) is DC.

(b) There is a continuous function \( f \) such that \( -F + f \) and \( F + f \) are convex. If this condition is satisfied, then \( f \) must be convex.

(c) There is a continuous function \( f \) such that for any \( x_1, x_2 \in \mathbb{R}^n \) and \( 0 < \lambda < 1 \) we have
\[
|\lambda F(x_1) + (1 - \lambda)F(x_2) - F(\lambda x_1 + (1 - \lambda)x_2)| \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - f(\lambda x_1 + (1 - \lambda)x_2).
\]
Proof. (a) \(\implies\) (b): Suppose that \(F = f_1 - f_2\) is a DC decomposition of \(F\). Let \(f = f_1 + f_2\). Then \(-F + f = 2f_2\) and \(F + f = 2f_1\), both of which are clearly convex. Also we can see that \(f\) must be convex.

(b) \(\implies\) (c): Pick any \(x_1, x_2 \in \mathbb{R}^n\), \(\lambda > 0\) and let \(x = \lambda x_1 + (1 - \lambda)x_2\). Then we have

\[
-F(x) + f(x) \leq \lambda(-F(x_1) + f(x_1)) + (1 - \lambda)(-F(x_2) + f(x_2))
\]

\[F(x) + f(x) \leq \lambda(F(x_1) + f(x_1)) + (1 - \lambda)(F(x_2) + f(x_2))\]

which can be rearranged to show

\[|\lambda F(x_1) + (1 - \lambda)F(x_2) - F(x)| \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - f(x).\]

(c) \(\implies\) (a): Choose arbitrary \(x_1, x_2 \in \mathbb{R}^n\), \(\lambda > 0\) and let \(x = \lambda x_1 + (1 - \lambda)x_2\). Then

\[|\lambda F(x_1) + (1 - \lambda)F(x_2) - F(x)| \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - f(x)\]

holds, which can be rearranged to Jensen’s inequality for \(-F + f\) and \(F + f\):

\[-F(x) + f(x) \leq \lambda(-F(x_1) + f(x_1)) + (1 - \lambda)(-F(x_2) + f(x_2))\]

\[F(x) + f(x) \leq \lambda(F(x_1) + f(x_1)) + (1 - \lambda)(F(x_2) + f(x_2)).\]

This means that \(-F + f\) and \(F + f\) are convex and \(F\) is DC since

\[F = \frac{1}{2}(F + f - (-F + f)).\]

This completes the proof. \(\square\)

If item (b) or (c) hold, we will use terminology that appears in [8] and say that \(f\) is a control function for \(F\).

In the next section we review the proof of the useful ‘mixing property’ of DC functions, for which we will need the next two propositions. The propositions and our proof of this property are adapted from [8].

**Proposition 2.5.** A continuous function \(f: \mathbb{R} \to \mathbb{R}\) is convex if, for any \(x \in \mathbb{R}\) and \(\delta > 0\) there exists \(0 < \lambda < 1\) and \(x_1, x_2 \in (x - \delta, x + \delta)\) such that \(x = \lambda x_1 + (1 - \lambda)x_2\) and \(f(x) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)\) and \(x_1 \neq x_2\)

**Proof.** Let \(g(x) = f(x) + \epsilon x^2\). To show that \(f\) is convex, it suffices to show that for any \(\epsilon > 0\), the function \(g\) is convex. Suppose by way of contradiction that \(g\) is not convex. Then there must exist some \(x_1, x_2 \in \mathbb{R}\) such that \(g(\lambda x_1 + (1 - \lambda)x_2) > \lambda g(x_1) + (1 - \lambda)g(x_2)\) for some \(0 < \lambda < 1\). Let \(l\) be the line segment connecting \((x_1, g(x_1))\) and \((x_2, g(x_2))\) and \(x_0 = \lambda x_1 + (1 - \lambda)x_2\).
Consider the continuous function $g - l$. This function must attain a maximum on $[x_1, x_2]$; let $\bar{x}$ be a point at which it does so. Define $h = l + (g(\bar{x}) - l(\bar{x}))$. Then for $x \in [x_1, x_2]$ we have the following:

\begin{align*}
g(x) &\leq h(x) \\
g(\bar{x}) = h(\bar{x}) \\
g(x_1) < h(x_1) \text{ and } g(x_2) < h(x_2).
\end{align*}

To see why (1) holds, pick any $x \in [x_1, x_2]$. Then

\[
g(x) - l(x) \leq g(\bar{x}) - l(\bar{x})
\]

that is,

\[
g(x) \leq l(x) + (g(\bar{x}) - l(\bar{x})) = h(x)
\]

Note that (2) follows from $h(\bar{x}) = l(\bar{x}) + g(\bar{x}) - l(\bar{x}) = g(\bar{x})$. The inequalities on (3) hold because $g(x_0) > l(x_0)$ so

\[
g(\bar{x}) - l(\bar{x}) \geq g(x_0) - l(x_0) > 0
\]

which means $h(x_2) = l(x_2) + (f(\bar{x}) - l(\bar{x})) > l(x_2) = g(x_2)$ and likewise with $x_1$. Figure 1 illustrates this situation. Next we will show that invoking the hypothesis yields a contradiction.

![Figure 1: The function $h$](image)

Note that (3) implies that $\bar{x} \in (x_1, x_2)$. By the hypothesis we can find $y_1, y_2 \in (x_1, x_2)$ and $0 < \lambda < 1$ such that $\bar{x} = \lambda y_1 + (1 - \lambda)y_2$ and $f(\bar{x}) \leq \lambda f(y_1) + (1 - \lambda)f(y_2)$ and $y_1 \neq y_2$. Then

\[
g(\bar{x}) = f(\bar{x}) + \epsilon \bar{x}^2
\]

\[
< \lambda f(y_1) + \epsilon \lambda y_1^2 + (1 - \lambda)f(y_2) + \epsilon(1 - \lambda)y_2^2
\]

\[
= \lambda g(y_1) + (1 - \lambda)g(y_2).
\]
Next we use the fact that $h$ is affine and (1) to show
\[
g(\bar{x}) < \lambda g(y_1) + (1 - \lambda)g(y_2) \\
\leq \lambda h(y_1) + (1 - \lambda)h(y_2) \\
= h(\lambda y_1 + (1 - \lambda)y_2) \\
= h(\bar{x}).
\]
So by (2) we have
\[
g(\bar{x}) < h(\bar{x}) = g(\bar{x}).
\]
This is a contradiction, so $g$ and therefore $f$ must be convex.

Proposition 2.6. A continuous function $F: \mathbb{R}^n \to \mathbb{R}$ is DC if, for all $x, v \in \mathbb{R}^n$ and $\delta > 0$ with $\|v\| = 1$, there exists $0 < \lambda < 1$ and distinct $x_1, x_2 \in B(x; \delta)$ such that $x = \lambda x_1 + (1 - \lambda)x_2$ and $x_2 - x_1 = \|x_2 - x_1\|v$ satisfying the inequality
\[
|\lambda F(x_1) + (1 - \lambda)F(x_2) - F(x)| \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - f(x).
\]
for a continuous function $f: \mathbb{R}^n \to \mathbb{R}$.

Proof. We will show that $F$ is DC by showing that $-F + f$ and $F + f$ are convex. To do this, we will show that for any choices of $a, b \in \mathbb{R}^n$ where $a$ is nonzero, the functions $G_1, G_2: \mathbb{R} \to \mathbb{R}$ given by $G_1(t) = (-F + f)(at + b)$ and $G_2(t) = (F + f)(at + b)$ are convex by Proposition 2.5.

Pick any $t \in \mathbb{R}$ and $\delta > 0$. We wish to find distinct $t_1, t_2 \in (t - \delta, t + \delta)$ such that $t = \lambda t_1 + (1 - \lambda)t_2$ for some $0 < \lambda < 1$ and
\[
G_1(t) \leq \lambda G_1(t_1) + (1 - \lambda)G_1(t_2) \\
G_2(t) \leq \lambda G_2(t_1) + (1 - \lambda)G_2(t_2).
\]
Define $x = at + b$ and $v = \frac{a}{\|a\|}$. By the hypothesis we can find $x_1, x_2 \in B(x; \delta\|a\|)$ such that $x_1 \neq x_2$ and
\[
\lambda x_1 + (1 - \lambda)x_2 = x \\
x_2 - x_1 = \|x_2 - x_1\|v \\
|\lambda F(x_1) + (1 - \lambda)F(x_2) - F(x)| \leq \lambda f(x_1) + (1 - \lambda)f(x_2) - f(x)
\]
for some $0 < \lambda < 1$ We can rearrange (6) to show
\[
-F(at + b) + f(at + b) \leq \lambda(-F(x_1) + f(x_1)) + (1 - \lambda)(-F(x_2) + f(x_2)) \\
F(at + b) + f(at + b) \leq \lambda(F(x_1) + f(x_1)) + (1 - \lambda)(F(x_2) + f(x_2)).
\]
(4) and (5) can be used to show that

\[ x_1 = a\left(t - (1 - \lambda)\frac{\|x_2 - x_1\|}{\|a\|}\right) + b \]

and

\[ x_2 = a\left(t + \lambda\frac{\|x_2 - x_1\|}{\|a\|}\right) + b. \]

Denote

\[ t_1 = \left(t - (1 - \lambda)\frac{\|x_2 - x_1\|}{\|a\|}\right) \]

\[ t_2 = \left(t + \lambda\frac{\|x_2 - x_1\|}{\|a\|}\right). \]

Then \( t = \lambda t_1 + (1 - \lambda)t_2 \) and (7) and (8) become

\[
(-F + f)(at + b) \leq \lambda(-F + f)(at_1 + b) + (1 - \lambda)(-F + f)(at_2 + b) \tag{9}
\]

\[
(F + f)(at + b) \leq \lambda(F + f)(at_1 + b) + (1 - \lambda)(F + f)(at_2 + b). \tag{10}
\]

Recall that \( x_1, x_2 \in B(x; \delta\|a\|) \). This means that

\[
\delta\|a\| > \|x - x_1\| \\
= \|at + b - a(t_1 + b)\| \\
= |t - t_1| \cdot \|a\|.
\]

Which implies \(|t - t_1| < \delta\) and by identical reasoning \(|t - t_2| < \delta\). So \( t_1, t_2 \in (t - \delta, t + \delta) \). Thus we have shown by Proposition 2.5 that \( G_1, G_2 \) are convex which means \( F \) is DC.

**Definition 2.7.** We will say that \( F: \mathbb{R}^n \to \mathbb{R} \) is a **local DC function** if for any \( x \in \mathbb{R}^n \) there exists a \( \delta > 0 \) such that \( F \) is DC on \( B(x; \delta) \).

It is fairly obvious that a DC function is a local DC function. Hartman also showed the converse under certain conditions [3]. The proof that follows originated with Hartman, but we refer to [5] for the detailed steps we show here. First we need to develop some tools to construct a control function for a local DC function. In the following lemma, we show that, given a convex function \( f \) defined near a point \( x_0 \), we can define a new function \( g \) that is convex on a larger domain and is equal to \( f \) near \( x_0 \).

**Lemma 2.8.** Suppose \( x_0 \) is a point in \( \mathbb{R}^n \). If \( f: U \to \mathbb{R} \) is a convex function, where \( U \) is an open convex set containing \( x_0 \), then we can construct a function \( g: \mathbb{R}^n \to \mathbb{R} \) with the property that \( g \) is convex and equal to \( f \) on \( V \), where \( x \in V \) and \( V \subset U \) is an open convex set.
Proof. Since $U$ is open we can find some $\delta > 0$ such that $f$ is bounded above by some $M \in \mathbb{R}$ on $B(x_0; \delta) \subset U$. Consider the function $h(x) = \alpha \|x - x_0\| + f(x_0) - 1$ for $x \in \mathbb{R}^n$. Certainly $h(x_0) = f(x_0) - 1 < f(x_0)$, which means there exists an open set $V \subset B(x_0; \delta) \subset U$ such that $h(x) \leq f(x)$ for $x \in V$. We can also choose $\alpha$ such that $f(x) < h(x)$ when $\|x - x_0\| = \delta$; for instance we could let $\alpha = \frac{2 - f(x_0) + M}{\delta}$.

Note that $f, h$ are convex. Define

$$g(x) = \begin{cases} \max\{h(x), f(x)\} & \text{if } x \in B(x_0; \delta), \\ h(x) & \text{if } x \notin B(x_0; \delta). \end{cases}$$

Observe that near the boundary of $B(x; \delta)$, we have that $g = h$ so $g$ is convex, as it is locally convex. Also $g$ has the desired property that $g|V = f$. \qed

Next we show that the local DC condition is sufficient to guarantee the DC condition on any compact and convex set in the domain.

**Lemma 2.9.** If $F: \mathbb{R}^n \to \mathbb{R}$ is local DC function, then for any compact convex set $K \subset \mathbb{R}^n$, it is the case that $F$ is DC on $K$.

**Proof.** By definition, we can find, for each $x \in K$, an open convex set $U_x$ containing $x$ such that $F + f_x$ and $-F + f_x$ are convex on $U_x$, where $f_x: U_x \to \mathbb{R}$. We can apply the previous lemma to find a convex function $g_x: \mathbb{R}^n \to \mathbb{R}$ and an open convex neighborhood $V_x$ of $x$ such that for $z \in V_x \subset U_x$ we have $g_x(z) = f_x(z)$. By the compactness of $K$ we can find a finite collection $\{x_1, \ldots, x_m\}$ where

$$K \subset \bigcup_{i=1}^m V_{x_i}.$$ 

Notice that for any $z \in K$, it must be the case that $z \in V_{x_j}$ for some $j \in \{1, \ldots, m\}$, which means $-F + \sum_{i=1}^m g_{x_i} = (-F + g_{x_j}) + \sum_{i \neq j} g_{x_i}$ and $F + \sum_{i=1}^m g_{x_i}$ are sums of functions convex on $V_{x_j}$. This means $-F + \sum_{i=1}^m g_{x_i}$ and $F + \sum_{i=1}^m g_{x_i}$ are convex on $K$ as they are locally convex on $K$ (see [5]), so $F$ is DC on $K$. \qed

Finally we show how to construct a sequence of functions on the domain that converge uniformly to our desired control function. This portion of the proof, based on [3, 5], requires the domain to be all of $\mathbb{R}^n$, although the result also holds when the domain of $f$ is an open or closed convex set of $\mathbb{R}^n$.

**Theorem 2.10.** If $F: \mathbb{R}^n \to \mathbb{R}$ is a local DC function, then it is also a DC function.
Proof. First define $C_n = \mathbb{B}(0; n)$ and $D_n = B(0; n + 1/2)$ for $n \in \mathbb{N}$ and observe that $C_1 \subset D_1 \subset C_2 \subset D_2 \subset \ldots$ and

$$\bigcup_{i=1}^{\infty} C_i = \mathbb{R}^n = \bigcup_{i=1}^{\infty} D_i.$$ 

It follows by the previous lemma that we can find a sequence of convex functions $\{g_n\}$ such that each $g_i : \mathbb{R}^n \to \mathbb{R}$ controls $F$ on $D_i$, i.e. $F + g_i$ and $-F + g_i$ are convex on $D_i$. Since $C_1$ is compact, so $g_1, g_2$ must be bounded on $C_1$, and thus we can find a constant $\alpha > 0$ such that $g_2(x) - \alpha < g_1(x)$ for $x \in C_1$. For $\beta \in \mathbb{R}$ let

$$d(x) = \begin{cases} 0 & \text{if } ||x|| \leq 1, \\ \beta(||x|| - 1) & \text{if } ||x|| > 1. \end{cases}$$

Note that if $\beta > 0$ then $d(x) = \max\{0, \beta(||x|| - 1)\}$, so $d$ is convex on $\mathbb{R}^n$. If $x \in C_1$, then $||x|| \leq 1$, so $d(x) = 0$ and $g_2(x) - \alpha + d(x) < g_1(x)$. Using the boundedness of $g_1, g_2$, we can pick $\beta$ such that $\beta > 0$ and for $x$ near the boundary of $D_1$ (i.e. $||x|| = 3/2$) we have $g_2(x) - \alpha + d(x) > g_1(x)$. Now define

$$f_1(x) = \begin{cases} \max\{g_1(x), g_2(x) - \alpha + d(x)\} & \text{if } ||x|| < 3/2, \\ g_2(x) - \alpha + d(x) & \text{if } ||x|| \geq 3/2. \end{cases}$$

We have shown that $f_1(x) = g_1(x)$ for $x \in C_1$ and $f_1(x) = g_2(x) - \alpha + d(x)$ for $x$ near the boundary of $D_1$; from this it follows that $f_1$ is convex on $\mathbb{R}^n$.

Recall that $F$ is controlled by $g_1$ on $D_1$ and by $g_2$ on $D_2$. Then $F$ is also controlled by $g_2 - \alpha + d$ on $D_2$ and it follows from the definition of $f_1$ that $F$ is controlled by $f_1$ on $D_1$. We can also show that $F$ is controlled by $f_1$ on a convex neighborhood of every point of $D_2 \setminus D_1$, and so $F$ is controlled by $f_1$ on the entirety of $D_2$.

To summarize: we have constructed a convex function $f_1$ such that $f_1(x) = g_1(x)$ for $x \in C_1$ and $F$ is controlled by $f_1$ on $D_2$. Let us iterate this process, constructing, for each $k \in \mathbb{N}$ with $k > 1$ a corresponding function $f_k$ that is convex on $\mathbb{R}^n$, that controls $F$ on $D_{k+1}$ and is equal to $f_{k-1}$ on $C_k$, by defining

$$f_k(x) = \begin{cases} \max\{f_{k-1}(x), g_k(x) - \alpha_k + d_k(x)\} & \text{if } ||x|| < k + 1/2, \\ g_k(x) - \alpha_k + d_k(x) & \text{if } ||x|| \geq k + 1/2, \end{cases}$$

where $d_k, g_k$ and $\alpha_k$ are constructed in the same fashion as above.

Choose any $x \in \mathbb{R}^n$. Then there exists some $k_0$ such that $x \in C_k$ for all $k \geq k_0$. Then for any such $k$ we have that $f_k(x) = f_{k_0}(x)$, a constant. The sequence $\{f_k\}$ converges pointwise to a function $f$, and uniformly to $f$ on compact sets of $\mathbb{R}^n$, so $f$ is convex [6, Theorem 10.8]. Also $f$ controls $F$ on all $C_i$ for $i \in \mathbb{N}$, so $f$ controls $F$ on $\mathbb{R}^n$. Therefore, $F$ is DC. \qed
3 Closure Properties

The class of DC functions is very large. Here we will give many examples of DC functions.

**Proposition 3.1.** If \( F, G : \mathbb{R}^n \to \mathbb{R} \) is a DC function then the following are also DC functions

(a) \( F + G \)
(b) \( \alpha F \) for \( \alpha \in \mathbb{R} \)
(c) \( FG \).

**Proof.**

(a) Suppose \( F = f_1 - f_2 \) and \( G = g_1 - g_2 \) for convex functions \( f_1, f_2, g_1, g_2 \). Then \( f_1 + g_1 \) and \( g_2 + f_2 \) are convex so \( F + G = (f_1 + g_1) - (f_2 + g_2) \) is DC.

(b) Note that \( \alpha F = \alpha f_1 - \alpha f_2 = (-\alpha)f_2 - (-\alpha)f_1 \). If \( \alpha \geq 0 \) then \( \alpha f_1 \) and \( \alpha f_2 \) are convex and if \( \alpha < 0 \) then \( (-\alpha)f_2 \) and \( (-\alpha)f_1 \) are convex.

(c) First note that if \( H : \mathbb{R}^n \to \mathbb{R} \) is a DC function, then so is \( H^2 \). To see this, suppose \( H = h_1 - h_2 \) where \( h_1, h_2 \) are convex and nonnegative; such functions exist by Proposition 2.3. Then

\[
H^2 = h_1^2 + h_2^2 - 2h_1h_2 = 2(h_1^2 + h_2^2) - (h_1 + h_2)^2
\]

and clearly \( h_1^2 + h_2^2 \) and \( (h_1 + h_2)^2 \) are convex.

Now if \( F, G \) are DC, we have \( FG = \frac{1}{2}(F + G)^2 + \frac{1}{2}(F^2 + G^2) \) which is a sum of DC functions and thus DC.

**Proposition 3.2.** Let \( F_i : \mathbb{R}^n \to \mathbb{R} \) be DC functions for \( i = 1, \ldots, m \). Then

(a) \( \max \{ F_i : i = 1, \ldots, m \} \) is a DC function.
(b) \( \min \{ F_i : i = 1, \ldots, m \} \) is a DC function.

**Proof.**

(a) Since each \( F_i \) is a DC function, there exist convex functions \( g_i, h_i : \mathbb{R}^n \to \mathbb{R} \) such that

\[
F_i = g_i - h_i = (g_i + \sum_{j=1,j\neq i}^m h_j) - \sum_{i=1}^m h_i.
\]

It follows that

\[
\max \{ F_i : i = 1, \ldots, m \} = \max_{i=1,\ldots,m} (g_i + \sum_{j=1,j\neq i}^m h_j) - \sum_{i=1}^m h_i,
\]

which is a DC function since \( \max_{i=1,\ldots,m} (g_i + \sum_{j=1,j\neq i}^m h_j) \) and \( \sum_{i=1}^m h_i \) are both convex.
(b) We have the representation
\[
\min\{F_i : i = 1, \ldots, m\} = -\max\{-F_i : i = 1, \ldots, m\},
\]
which is a DC function by (a).

\[\square\]

**Corollary 3.3.** If \(F : \mathbb{R}^n \to \mathbb{R}\) is a DC function controlled by a continuous function \(f : \mathbb{R}^n \to \mathbb{R}\), then \(|F|\) is a DC function controlled by \(|F| + 2f\).

**Proof.** Since \(|F| = \max\{F, -F\}\), it follows immediately from Proposition 3.1(b) and Proposition 3.2(a) that \(|F|\) is DC. Since \(f\) controls \(F\), we have that \(F\) has the DC decomposition \(F = g - h\), where \(g = \frac{1}{2}(F + f)\) and \(h = \frac{1}{2}(-F + f)\). Then
\[
|F| = \max\{g - h, h - g\}
= \max\{2g, 2h\} - (h + g)
\]
and
\[
|F| + 2f = \max\{2g, 2h\} + h + g.
\]
From this it follows that \(|F| + 2f\) is a control function for \(|F|\) because
\[
|F| + |F| + 2f = 2(\max\{2g, 2h\})
\]
and
\[
-|F| + |F| + 2f = 2(h + g)
\]
which are both convex functions. \[\square\]

**Lemma 3.4.** Let \(F_1, \ldots, F_m : \mathbb{R}^n \to \mathbb{R}\) be DC functions. Suppose that \(F : \mathbb{R}^n \to \mathbb{R}\) is a continuous function such that \(F(x) \in \{F_1(x), \ldots, F_m(x)\}\) for all \(x \in \mathbb{R}^n\). Then for any \(x, v \in \mathbb{R}^n\) with \(\|v\| = 1\) and \(\delta > 0\), there exist \(x_1, x_2 \in B(x; \delta), 0 < \lambda < 1\) and \(r, s \in \{1, \ldots, m\}\) with
\[
x_2 - x_1 = \|x_2 - x_1\|v, \quad x = \lambda x_1 + (1 - \lambda)x_2, \quad F_r(x_1) = F(x_1), \quad F_s(x_2) = F(x_2), \quad F_r(x) = F_s(x) = F(x).
\]

**Proof.** Let \(E = \{1, \ldots, m\}\) and let \(l\) be the line parallel to \(v\) running through \(x\), that is, \(l = \{x + \alpha v : \alpha \in \mathbb{R}\}\). Denote \(l^+ = \{x + \alpha v : \alpha > 0\}\) and \(l^- = \{x + \alpha v : \alpha < 0\}\).

Then for some \(r \in E\) it must be the case that \(F(x) = F_r(x)\) and \(F(x_1) = F_r(x_1)\) for some \(x_1 \in B(x; \delta) \cap l^+\). Indeed, suppose for a contradiction that this is not the case.
Thus there is some $x' \in B(x; \delta) \cap l^+$, although $F(x') = F_i(x')$ for some $i \in E$, we have that $F(x) \neq F_i(x)$, which means that $|F(x) - F_i(x)| > 0$ for all such indices $i$. Define
\[
d = \min \{|F(x) - F_i(x)| : i \in E, F(x) \neq F_i(x)\} > 0.
\]

Let $\{x_n\} \subset B(x; \delta) \cap l^+$ be a sequence that converges to $x$. Then $\{F(x_n)\}$ converges to $F(x)$ as $F$ is continuous. Suppose $i \in E$ with the property that $F(x') = F_i(x')$ for some $x' \in B(x; \delta) \cap l^+$. Then we have that $\{F_i(x_n)\}$ converges to $F_i(x)$, which means that we can find $\hat{n}_i \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ where $n \geq \hat{n}_i$ we have
\[
|F_i(x_n) - F_i(x)| < d/2.
\]

We can also find $\hat{n}_0 \in \mathbb{N}$ with the property that
\[
|F(x_n) - F(x)| < d/2.
\]

when $n \geq \hat{n}_0$. Define $\hat{n} = \max\{\hat{n}_0, \ldots, \hat{n}_m\}$ and let $k \in E$ such that $F_k(x_{\hat{n}}) = F(x_{\hat{n}})$. Then we have the following contradiction:
\[
d > |F_k(x_{\hat{n}}) - F_k(x)| + |F(x_{\hat{n}}) - F(x)|
\]
\[
= |F_k(x_{\hat{n}}) - F_k(x)| + |F_k(x) - F(x)|
\]
\[
\geq |F_k(x) - F(x)| \geq d.
\]

Thus there is some $x_1 \in l^+ \cap B(x; \delta)$ and $r \in E$ with $F(x_1) = F_r(x_1)$ and $F(x) = F_r(x)$. By the same reasoning we can find $x_2 \in B(x, \delta) \cap l^-$ and $s \in E$ with $F(x_2) = F_s(x_2)$ and $F(x) = F_s(x)$. Also there must be some $0 < \lambda < 1$ with $\lambda x_1 + (1-\lambda)x_2 = x$ and we can see that $x_2 - x_1 = \|x_2 - x_1\|v$.

\[\square\]

**Theorem 3.5.** Let $F_1, \ldots, F_m: \mathbb{R}^n \to \mathbb{R}$ be DC functions. If $F: \mathbb{R}^n \to \mathbb{R}$ is a continuous function such that $F(x) \in \{F_1(x), \ldots, F_m(x)\}$ for all $x \in \mathbb{R}^n$, then $F$ is a DC function.

**Proof.** We will proceed by Proposition 2.6 to show that $F$ is DC. Choose any $x, v \in \mathbb{R}^n$ with $\|v\| = 1$. By Lemma 3.4, there exist $x_1, x_2 \in B(x; \delta)$, $0 < \lambda < 1$ and $r, s \in \{1, \ldots, m\}$ with
\[
x_2 - x_1 = \|x_2 - x_1\|v, \quad x = \lambda x_1 + (1-\lambda)x_2,
\]
\[
F_r(x_1) = F(x_1), \quad F_s(x_2) = F(x_2), \quad F_r(x) = F_s(x) = F(x).
\]

For any $i, j \in \{1, \ldots, m\}$, let $f_i, f_j$ be the control functions of $F_i$ and $F_j$. Clearly $F_i - F_j + f_i + f_j$ and $F_j - F_i + f_i + f_j$ are convex and so $F_i - F_j$ is DC and controlled by $f_i + f_j$. Then from Corollary 3.3 we have that $|F_i - F_j|$ is DC and $f_i + f_j + \frac{1}{2}|F_i - F_j|$ is convex.

11
Define
\[ f = \sum_{i,j=1}^{m} (f_i + f_j + \frac{1}{2}|F_i - F_j|) \]
and
\[ h = \frac{1}{2}(f_r + f_s) + \sum_{i,j=1, i \neq r, j \neq s}^{m} (f_i + f_j + \frac{1}{2}|F_i - F_j|). \]

Being sums of convex functions, both \( f \) and \( h \) are convex. Then, for arbitrary \( A, B \in \mathbb{R} \) we have

\[ |\lambda F_r(x_1) + (1 - \lambda) F_s(x_2) - F(x)| = |\lambda F_r(x_1) + (1 - \lambda) F_s(x_2) - F(x)| \]
\[ = |\lambda F_r(x_1) + (1 - \lambda) F_s(x_2) + \frac{1}{2}A - \frac{1}{2}A - \frac{1}{2}B + \frac{1}{2}B - F(x)|. \]
\[ \leq \frac{1}{2} \left| \lambda F_r(x_1) + A - F_r(x) \right| + \frac{1}{2} \left| (1 - \lambda) F_s(x_2) + B - F_s(x) \right| \]
\[ + \frac{1}{2} \left| \lambda F_r(x_1) - B \right| + \frac{1}{2} \left| (1 - \lambda) F_s(x_2) - A \right| \]

We will choose \( A = (1 - \lambda) F_r(x_2) \) and \( B = \lambda F_s(x_1) \). Then by Proposition 2.4(c) we have

\[ \frac{1}{2} \left| \lambda F_r(x_1) + A - F_r(x) \right| = \frac{1}{2} \left| \lambda F_r(x_1) + (1 - \lambda) F_r(x_2) - F_r(x) \right| \]
\[ \leq \frac{1}{2} \left( \lambda f_r(x_1) + (1 - \lambda) f_r(x_2) - f_r(x) \right). \]

Similarly,

\[ \frac{1}{2} \left| (1 - \lambda) F_s(x_2) + B - F_s(x) \right| = \frac{1}{2} \left| \lambda F_s(x_1) + (1 - \lambda) F_s(x_2) - F_s(x) \right| \]
\[ \leq \frac{1}{2} \left( \lambda f_s(x_1) + (1 - \lambda) f_s(x_2) - f_s(x) \right). \]

Since \( |F_r(x) - F_s(x)| = 0 \), we have

\[ \frac{1}{2} \left| \lambda F_r(x_1) - B \right| + \frac{1}{2} \left| (1 - \lambda) F_r(x_2) - A \right| = \frac{1}{2} \lambda \left| F_r(x_1) - F_s(x_1) \right| + \frac{1}{2} (1 - \lambda) \left| F_r(x_2) - F_s(x_2) \right| \]
\[ + \frac{1}{2} F_r(x) - F_s(x). \]
Because $|F_i - F_j|$ is DC it follows that

$$\left| \lambda F(x_1) + (1 - \lambda)F(x_2) - F(x) \right| \leq \frac{1}{2} (\lambda f_r(x_1) + (1 - \lambda)f_r(x_2) - f_r(x))$$

$$+ \frac{1}{2} (\lambda f_s(x_1) + (1 - \lambda)f_s(x_2) - f_s(x)) + \frac{1}{2} \lambda |F_r(x_1) - F_s(x_1)|$$

$$+ \frac{1}{2} (1 - \lambda) |F_r(x_2) - F_s(x_2)| - \frac{1}{2} |F_r(x) - F_s(x)|$$

$$= \lambda \left( \frac{1}{2} (f_r + f_s + |F_r - F_s|)(x_1) \right) + (1 - \lambda) \left( \frac{1}{2} (f_r + f_s + |F_r - F_s|)(x_2) \right)$$

$$- \left( \frac{1}{2} (f_r + f_s + |F_r - F_s|)(x) \right)$$

$$= (\lambda f(x_1) + (1 - \lambda)f(x_2) - f(x)) - (\lambda h(x_1) + (1 - \lambda)h(x_2) - h(x))$$

$$\leq (\lambda f(x_1) + (1 - \lambda)f(x_2) - f(x)).$$

Note that the final step follows from the fact that

$$\lambda h(x_1) + (1 - \lambda)h(x_2) - h(x) \geq 0$$

which holds true by the convexity of $h$ and Jensen’s inequality. Since the function $f$ does not depend on the choice of $i, j$, this completes the proof that $F$ is DC. 

\[\square\]

**Proposition 3.6.** Suppose $F: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous and $|F|$ is DC. Then $F$ is DC.

**Proof.** Note that if $|F|$ is DC then so is $-|F|$. Define $A = \{ x \in \mathbb{R}^n : F(x) \geq 0 \}$. For any $x \in \mathbb{R}^n$ we have

$$F(x) = \begin{cases} |F(x)| & \text{if } x \in A, \\ -|F(x)| & \text{if } x \notin A. \end{cases}$$

Since $F$ is continuous, it is then DC by Proposition 3.5. \[\square\]

**Proposition 3.7.** Suppose $A$ is a symmetric $n \times n$ matrix. Then $F = x^T Ax$ is a DC function.

This proposition and its proof are based on [5].

**Proof.** Let $\lambda_1, ..., \lambda_n$ be the eigenvalues of $A$ and let $\gamma > \max_{i=1,\ldots,n} |\lambda_i|$. Then

$$F(x) = x^T Ax$$

$$= x^T \left( A + \frac{1}{2} \gamma I - \frac{1}{2} \gamma I \right) x$$

$$= \frac{1}{2} (x^T (A + \gamma I) x + x^T (A - \gamma I) x)$$

$$= \frac{1}{2} (x^T (A + \gamma I) x - x^T (\gamma I - A) x).$$
Suppose λ is an eigenvector for \(A + \gamma I\). Then for some \(v \in \mathbb{R}^n\), we have \((A + \gamma I)v = \lambda v\) which means \(Av = (\lambda - \gamma)v\) so \(\lambda = \lambda_i + \gamma > 0\) for some \(i = 1, \ldots, n\). Similarly we can show that an eigenvalue for \(\gamma I - A\) takes the form \(\gamma - \lambda_j > 0\). Since both matrices have positive eigenvalues, they are positive definite, so their quadratic forms are convex.

\[
4 \quad \text{Applications}
\]

\[
4.1 \quad \text{The DCA}
\]

First, we recall the following definitions.

**Definition 4.1.** Given a function \(f : \mathbb{R}^n \to \mathbb{R}\), we define \(f^* : \mathbb{R}^n \to (-\infty, \infty]\) by \(f^*(x) = \sup\{\langle x, v \rangle - f(v) : v \in \mathbb{R}^n\}, x \in \mathbb{R}^n\), as the conjugate or Fenchel conjugate of \(f\).

**Definition 4.2.** Given a function \(f : \mathbb{R}^n \to (-\infty, \infty]\) with \(f(\bar{x}) < \infty\), we define \(\partial f : \mathbb{R}^n \to \mathbb{R}\) by

\[
\partial f(\bar{x}) = \{v : \langle v, \bar{x} - x \rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n\}
\]

as the subdifferential of \(f\) at \(\bar{x}\).

The following version of the DCA algorithm is based on [5]. We omit a discussion of convergence results and background for this algorithm, referring the interested reader to [5] and Tao.

Consider the problem of minimizing a function \(f\) with DC decomposition \(f = g - h\)

\[
\text{DCA}(N)
\]

1. Pick \(x_1 \in \mathbb{R}^n\)
2. for \(k = 1, \ldots, N\)
3. \quad Find \(y_k \in \partial h(x_k)\)
4. \quad Find \(x_{k+1} \in \partial g^*(y_k)\) or approximate \(x_{k+1}\) by solving the problem:
5. \quad minimize \(\phi_k(x) = g(x) - \langle x, y_k \rangle\)
6. return \(x_{N+1}\)

Notice that the minimization problem on 5 is convex, so even in situations where \(\partial g^*(y_k)\) is impossible or impractical to calculate, the DCA succeeds in reducing a nonconvex minimization problem to a series of convex minimization problems.

As we have shown, the class of DC functions is quite broad; in particular it contains many functions which are not convex. The DCA is a promising technique for optimizing non-convex functions, see for instance [2, 7]. However, in at least one instance, Tao et al have shown that the DCA can outperform traditional techniques for optimizing convex functions [7].

14
4.2 The Support Vector Machine Problem

Consider a set of points \( X = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^n \) with class labels \( Y = \{y_1, \ldots, y_m\} \) where each label is either 1 or \(-1\). The linear support vector machine approach to training a classifier on such a data set is to find for a hyperplane which maximizes the separation between the two classes of points; if such a hyperplane exists, the points are said to be linearly separable. We consider the least squares variant of the support vector machine problem (LSSVM), in which we seek to minimize the function

\[
J(w, b) = \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{m} (w^T x_i + b - y_i)^2
\]

where \( w \in \mathbb{R}^n \) and \( b \in \mathbb{R} \) and \( \lambda \) is a regularization parameter.

This technique can be used in supervised learning when we wish to use a representative set of labelled data to train a classifier. Assuming the classes are linearly separable, or close to linearly separable, a classifier that achieves good separation on a subset of the data will often generalize well.

Here we solve the LSSVM problem for the UIC Iris and Spambase data sets, comparing the performance of the subgradient method and the DCA. For convenience, let \( A = \{i \in \mathbb{N}: y_i = 1\} \) and \( B = \{i \in \mathbb{N}: y_i = 1\} \). In the DCA method, we represent \( J \) as

\[
J(w, b) = \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{m} (w^T x_i + b)^2 - 2y_i(w^T x_i + b)
\]

\[
= \lambda \|w\|^2 + \frac{1}{n} \sum_{i \in B} [(w^T x_i + b)^2 + 2(w^T x_i + b)] - \frac{1}{n} \sum_{i \in A} 2(w^T x_i + b)
\]

\[
= \lambda \|w\|^2 + \frac{1}{n} \sum_{i=1}^{m} [(w^T x_i + b)^2 + 2(w^T x_i + b)] - \frac{1}{n} \sum_{i \in A} 4(w^T x_i + b)
\]

\[
= g(w, b) - h(w, b).
\]

and our algorithm is

DCA\((N)\)

1. Pick \( w_1 \in \mathbb{R}^n \) and \( b_1 \in \mathbb{R} \)
2. for \( k = 1, \ldots, N \)
   3. Find \( y_k = \nabla_w h(w_k) \)
   4. Find \( z_k = \partial_b h(b_k) \)
   5. Approximate \( w_{k+1} \) by minimizing \( \phi_k(w) = g(w_k) - \langle w, y_k \rangle \)
   6. Approximate \( b_{k+1} \) by minimizing \( \rho_k(b) = g(w_k, b) - bz_k \)
3. return \( w_{N+1}, b_{N+1} \)
We implemented this algorithm in Python, using the numpy library to handle matrix math. We used a straightforward gradient descent approach with constant stepsize to solve the minimization problems on 5 and 6.

Both data sets were separated into disjoint training and validation sets, containing roughly 90% and 10% of the original data, respectively. In the case of the Iris data, we used a linearly separable subset of the data, and both algorithms yielded solutions that achieved 100% accuracy on the training and validation sets after only a few iterations. The Spanbase data is not linearly separable, but both algorithms yielded solutions that achieved comparable accuracy (roughly 88%).

In our implementation the DCA and the gradient method converged at almost exactly the same rate. This seems to be because the problem of minimizing $\phi$ is not much simpler than the original problem of minimizing $J$. This suggests that the main promise of the DCA in convex minimization is in scenarios where the sub-minimization problem is simpler than the original problem.

References


