# CANONICAL DUAL SOLUTIONS TO NONCONVEX RADIAL BASIS NEURAL NETWORK OPTIMIZATION PROBLEM[1]

Vittorio Latorre[*] and David Yang Gao[+] [2]

[*]*University "Sapienza" of Rome, Rome, Italy*
[+]*University of Ballarat and Australian National University, Australia*

## Abstract

Radial Basis Functions Neural Networks (RBFNNs) are tools widely used in regression problems. One of their principal drawbacks is that the formulation corresponding to the training with the supervision of both the centers and the weights is a highly non-convex optimization problem, which leads to some fundamentally difficulties for traditional optimization theory and methods. This paper presents a generalized canonical duality theory for solving this challenging problem. We demonstrate that by sequential canonical dual transformations, the nonconvex optimization problem of the RBFNN can be reformulated as a canonical dual problem (without duality gap). Both global optimal solution and local extrema can be classified. Several applications to one of the most used Radial Basis Functions, the Gaussian function, are illustrated. Our results show that even for one-dimensional case, the global minimizer of the nonconvex problem may not be the best solution to the RBFNNs, and the canonical dual theory is a promising tool for solving general neural networks training problems.

## 1. Introduction

Radial Basis Function Neural Networks(RBFNN) are a tool introduced in the field of function interpolation [1] and then were adapted to the problem of regression [2]. During the last two decades RBFNN were applied in several fields. The problem of regression consists in trying to approximate a function $f : \mathbb{R}^n \to \mathbb{R}$ by means of an approximation function $g(\cdot)$ that uses a set of samples defined as:

$$\mathcal{T} = \{(x^p, y^p), x^p \in \mathbb{R}^n, y^p \in \mathbb{R}, p = 1, ..., P\}, \tag{1}$$

where $(x^p, y^p)$ are respectively arguments and values of the given function $f(x)$. In general the approximating function $g(\cdot)$ obtained by the RBFNNs with radial basis function $\phi(\cdot)$ has the following form:

$$g(x) = \sum_{i=1}^{N} w_i \phi(\|\mathbf{x} - \mathbf{c}_i\|), \tag{2}$$

where $N$ is the number of units used to approximate the function, or neurons of the network, $\mathbf{w}$ is the vector with components $w_i$ for $i = 1, \dots, N$ that is the vector of the weights associated with the connections between the units $\mathbf{x}$ and $\mathbf{c}_i \in \mathbb{R}^n$ for $i = 1, \dots, N$ are the centers of the RBFNNs.

Generally speaking, there are two main optimization strategies to train a RBFNN. The first consists in the optimization of only the weights of the neural network. In this case the centers are generally chosen by using clustering strategies [3]. This problem is a convex problem in the variable $\mathbf{w}$ and has the form:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{N} (w_i \phi(\mathbf{c}_i) - y_p)^2 + \frac{1}{2}\beta_w \|\mathbf{w}\|^2, \tag{3}$$

where $\beta_w$ is the regularization parameter for the weights.

The second strategy is to consider both weighter $\mathbf{w}$ and the centers $\mathbf{c}$ of the radial basis functions as variables. This strategy can be performed by solving the following unconstrained optimization problem:

$$E(\mathbf{w}, \mathbf{c}) = \frac{1}{2} \sum_{p=1}^{P} \sum_{i=1}^{N} (w_i \phi(\mathbf{c}_i) - y^p)^2 + \frac{1}{2}\beta_w \|\mathbf{w}\|^2 + \frac{1}{2}\beta \sum_{i=1}^{N} \sum_{j=1}^{n} c_{ji}^2. \tag{4}$$

This problem is non-convex, but from empirical experiments [4] it emerged that it generally yields neural networks with an higher precision than the ones trained with strategy (3). One of the most used strategies to solve this optimization problem is to apply decomposition algorithms [5]. However, due to the nonconvexity of the problem (4), there are some fundamental difficulties to find the global minimum of the problem and to characterize local minima. Indeed, the problem (4) is considered to be NP-hard even if the radial basis function $\phi(\mathbf{c})$ is a quadratic function and $n = 1$ [6, 7]. Another issue that characterizes this problem is the choice of the regularization parameters $\beta_w$ and $\beta$. In general a cross-validation strategy is applied in order to find these regularization parameters. Cross-validation consists in trying different values of the parameters in order to find the one that yields the neural network with the best prediction. Until now it was not possible to find a closed form for the optimal values of these parameters in the general case. If it is possible to find at least an upper bound for these parameters, the time needed to perform a cross validation would greatly decrease.

Canonical duality theory developed from nonconvex analysis

and global optimization [8, 9] is a potentially powerful methodology, which has been used successfully for solving a large class of challenging problems in biology, engineering, sciences [10, 14, 15], and recently in network communications [11, 13]. In this paper we study the canonical duality theory for solving the general Radial Basis Neural Networks optimization problem (4) and mainly analyze one-dimensional case in order to find properties and intuitions that can be useful for the multidimensional cases. The rest of this paper is arranged as follows. In Section 2, we first demonstrate how to rewrite the nonconvex primal problem as a dual problem by using sequential canonical dual transformation developed in [8, 12]. In Section 3 we prove the complementarity-dual principle showing that the obtained formulation is canonically dual to the original problem in the sense that there is no duality gap. In Section 4, we analyze the problem with the Gaussian function as radial basis in the neurons and show some examples. The last section presents some conclusions.

## 2. Primal problem for general Radial Basis Functions(RBF)

The general one dimensional non-convex function to be addressed in this paper can be proposed in the following form:

$$P(c) = W(c) + \frac{1}{2}\beta c^2 - fc, \tag{5}$$

where $\beta$ is the regularization coefficient and $f$ is a positive scalar close to zero. The term $-fc$ is not comprised in the original Radial Basis Neural Networks formulation but we consider it for the general mathematical case. The non-convex function $W(c)$ depends on the choice of the radial basis function $\phi(\cdot)$:

$$W(c) = \frac{1}{2}\left(w\phi(\|x - c\|^2) - y\right)^2, \tag{6}$$

where $x$, $y$ and $w$ belong to $\mathbb{R}$. In applications the parameter $w$ is also a variable, but the original problem (4) is convex in $w$ while non-convex in respect to the center of the radial basis function $c$. Therefore, the one-dimensional non-convex primal problem can be formulated as

$$(\mathcal{P}): \quad \min\left\{P(c) = \frac{1}{2}\left(w\phi(\|x - c\|^2) - y\right)^2 + \frac{1}{2}\beta c^2 - fc \quad | \forall c \in \mathbb{R}\right\}. \tag{7}$$

In order to apply the canonical duality theory to solve this problem, we need to choose the following geometrically nonlinear operator:

$$\xi = \Lambda(c) = w\phi(\|x - c\|^2): \quad \mathbb{R} \to \mathcal{E}_a. \tag{8}$$

Clearly, this is a nonlinear map from $\mathbb{R}$ to a subspace $\mathcal{E}_a \in \mathbb{R}$, which depends on the choice of the Radial Basis Function $\phi(\cdot)$. The *canonical function* associated with this geometrical operator is

$$V(\xi(c)) = \frac{1}{2}(\xi(c) - y)^2 = W(\Lambda(c)). \tag{9}$$

By the definition introduced in the canonical duality theory [9], $V : \mathcal{E}_a \to \mathbb{R}$ is said to be canonical function on $\mathcal{E}_a$ if for any given $\xi \in \mathcal{E}_a$, the duality relation

$$\sigma = V'(\xi) = \{\xi - y\} : \mathcal{E}_a \to \mathcal{S}_a \tag{10}$$

is invertible, where $\mathcal{S}_a$ is the range of the duality mapping $\sigma = \partial V(\xi)$, which depends on the choice of the Radial Basis Function $\phi(\cdot)$. The couple $(\xi, \sigma)$ forms a canonical duality pair on $\mathcal{E}_a \times \mathcal{S}_a$ with the Legendre conjugate $V^*(\sigma)$ defined by

$$V^*(\sigma) = \{\xi\sigma - V(\xi)|\sigma = V'(\xi)\} = \left(\frac{1}{2}\sigma^2 + y\sigma\right). \tag{11}$$

By considering that $W(c) = \Lambda(c)\sigma - V^*(\sigma)$, the primal function $P(c)$ can be reformulated as the so-called *total complementarity function* defined by

$$\begin{aligned}
\Xi(c, \sigma) &= \Lambda(w, c)\sigma - V^*(\sigma) + \frac{1}{2}\beta c^2 - fc \\
&= w\phi(\|x - c\|^2)\sigma - \left(\frac{1}{2}\sigma^2 + \sigma y\right) \\
&\quad + \frac{1}{2}\beta c^2 - fc.
\end{aligned} \tag{12}$$

The function $\phi(\cdot)$ can be a non convex function just like $W(c)$. For this reason we have to perform a sequential canonical dual transformation for the nonlinear operator $\Lambda(c)$. To this aim we choose a second nonlinear operator:

$$\epsilon = \Lambda_2(c) = \|x - c\|^2 \tag{13}$$

which is a map from $\mathbb{R}$ to $\mathcal{E}_b = \{\epsilon \in \mathbb{R}|\epsilon \geq 0\}$. In terms of $\epsilon$, the first level operator $\xi = \Lambda(c)$ can be written as

$$\xi = U(\epsilon) = w\phi(\epsilon). \tag{14}$$

We assume that $U(\epsilon)$ is a convex function on $\mathcal{E}_b$ such that the second-level duality relation

$$\tau = U'(\epsilon) = w\phi'(\epsilon) \tag{15}$$

is invertible, i.e.,

$$\epsilon = \left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1}, \tag{16}$$

where the term $\left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1}$ is the inverse of the function $\phi'(\epsilon)$. Thus, the Legendre conjugate of $U$ can be obtained uniquely by

$$U^*(\tau) = \tau\left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1} - w\phi\left(\left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1}\right). \tag{17}$$

We notice that $\xi = w\phi(\epsilon)$. By substituting the value of $\epsilon$ given by (16) we find a relation that connects the first level primal variable $\xi$ with the second level dual variable $\tau$:

$$\xi = w\phi\left(\left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1}\right). \tag{18}$$

By plugging this in (10) we obtain

$$\sigma = w\phi\left(\left(\phi'\left(\frac{\tau}{w}\right)\right)^{-1}\right) - y. \tag{19}$$

2

Generally speaking, it is possible, for certain functions $\phi$, to use the canonical dual transformation to find the relation between the first level dual variable $\sigma$ and the second level dual variable $\tau$ by means of the derivatives of $\phi(\cdot)$ and the first primal variable $\xi$. In general this relation is:

$$\tau = w\phi'\left(\phi^{-1}\left(\frac{\sigma + y}{w}\right)\right). \tag{20}$$

Therefore, replacing $U(\xi) = \Lambda(c)$ by its Legendre conjugate $U^*$, the total complementarity function becomes

$$\begin{aligned} \Xi(c, \sigma, \tau) &= \left(\|x_p - c_i\|^2 \tau - U^*(\tau)\right)\sigma \\ &\quad -V^*(\sigma) + \frac{1}{2}\beta c^2 - fc. \end{aligned} \tag{21}$$

It is also possible to rewrite the total complementary function (21) in the following form:

$$\begin{aligned} \Xi(c, \sigma, \tau) &= \frac{1}{2}c^2(2\tau\sigma + \beta) - c(2\tau\sigma x + f) \\ &\quad -U^*(\tau)\sigma - V^*(\sigma) + x^2\tau\sigma. \end{aligned} \tag{22}$$

By the criticality condition $\partial\Xi(c, \sigma, \tau)/\partial c = 0$ we obtain

$$c(\tau, \sigma) = \frac{2\tau x\sigma + f}{2\tau\sigma + \beta}. \tag{23}$$

Clearly, if $2\tau\sigma + \beta \neq 0$, the general solution of (23) is

$$c = \frac{2\tau x\sigma + f}{2\tau\sigma + \beta} \quad \forall(\sigma, \tau) \in \mathcal{S}_a = \{\sigma, \tau | \ 2\tau\sigma + \beta \neq 0\} \tag{24}$$

and the canonical dual function of $P(c)$ can be presented as

$$P^d(\sigma, \tau) = -\frac{1}{2}\frac{(2\tau x\sigma + f)^2}{2\tau\sigma + \beta} - U^*(\tau)\sigma - V^*(\sigma) + x^2\tau\sigma. \tag{25}$$

By considering dual relation given in (20), and by setting $s(\sigma) = \frac{\sigma + y}{w}$, we can write the total complementarity function in terms of only $c$ and $\sigma$

$$\begin{aligned} \Xi(c, \sigma) &= \tfrac{1}{2}c^2 G(\sigma) - cF(\sigma) - U^*(\sigma)\sigma - \\ &\quad V^*(\sigma) + x^2 w\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma, \end{aligned} \tag{26}$$

where

$$\begin{aligned} G(\sigma) &= 2w\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma + \beta, \\ F(\sigma) &= 2w\phi'\left(\phi^{-1}(s(\sigma))\right)x\sigma + f, \\ U^*(\sigma) &= w\phi'\left(\phi^{-1}(s(\sigma))\right)\phi^{-1}(s(\sigma)) - (\sigma + y). \end{aligned}$$

Therefore, in terms of $\sigma$ only, the canonical dual function can be written as

$$\begin{aligned} P^d(\sigma) &= -\frac{1}{2}\frac{F(\sigma)^2}{G(\sigma)} - U^*(\sigma)\sigma + V^*(\sigma) - \\ &\quad x^2 w\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma. \end{aligned} \tag{27}$$

## 3. Complementary-Dual Principle

**Theorem 3.1.** *If $\bar{\sigma}$ is a critical point of $(P^d)$ and the term:*

$$\begin{aligned} G'(\bar{\sigma}) &= \sigma\phi''\left(\phi^{-1}(s(\bar{\sigma}))\right)\left(\phi^{-1}(s(\bar{\sigma}))\right)' + \\ &\quad w\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right) \neq 0, \end{aligned} \tag{28}$$

*then the point*

$$\bar{c} = \frac{F(\bar{\sigma})}{G(\bar{\sigma})} \tag{29}$$

*is a critical point of $P(c)$ and $P(\bar{c}) = P^d(\bar{\sigma})$*

**Proof 3.1.** Suppose that $\bar{\sigma}$ is a critical point of $P^d$ then we have

$$\begin{aligned} P^d(\bar{\sigma})' &= \left[\bar{c}^2 - 2x\bar{c} + x^2 - \phi^{-1}(s(\bar{\sigma}))\right]G'(\bar{\sigma}) - \\ &\quad \sigma\left[\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right)\left(\phi^{-1}(s(\bar{\sigma}))\right)' - 1\right] = 0. \end{aligned} \tag{30}$$

Notice that

$$\left(\phi^{-1}(s(\bar{\sigma}))\right)' = \frac{1}{\phi'(\bar{\epsilon})} = \frac{1}{\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right)}, \tag{31}$$

The third term in (30) is zero. The term $G'(\bar{\sigma})$ is not zero from the hypothesis, so we obtain

$$(x - \bar{c})^2 - \phi^{-1}(s(\bar{\sigma})) = 0, \tag{32}$$

that is

$$\bar{\sigma} = w\phi\left(\|x - \bar{c}\|^2\right) - y. \tag{33}$$

The critical point condition for the primal problem $P'(c) = 0$ is

$$-2w(x - c)\phi'(\|x - c\|^2)(w\phi(\|x - c\|^2) - y) + \beta c - f = 0. \tag{34}$$

By considering that $\phi'(\|x - c\|^2) = \phi'\left(\phi^{-1}(s(\bar{\sigma}))\right)$ and $\sigma = w\phi\left((x - c)^2\right) - y$ we obtain

$$2w(x - c)\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma + \beta c - f = 0, \tag{35}$$

that is

$$c = \frac{2\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma + f}{2\phi'\left(\phi^{-1}(s(\sigma))\right)\sigma + \beta}. \tag{36}$$

By setting $\sigma = \bar{\sigma}$ in (36) we obtain (24) proving that $\bar{c}$ is a critical point of $P(c)$.

For the correspondence of the function values we start from the dual function

$$\begin{aligned} P^d(\bar{\sigma}) &= -\frac{1}{2}\frac{F^2(\bar{\sigma})}{G(\bar{\sigma})} - U^*(\bar{\sigma})\bar{\sigma} - V^*(\bar{\sigma}) + \\ &\quad x^2 w\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right)\bar{\sigma} \end{aligned} \tag{37}$$

add and subtract the term $\frac{1}{2}\frac{F^2(\bar{\sigma})}{G(\bar{\sigma})}$ and substitute the value of $\bar{c}$

$$\begin{aligned} &\tfrac{1}{2}\bar{c}^2 G(\bar{\sigma}) - \bar{c}F(\bar{\sigma}) - U^*(\bar{\sigma})\bar{\sigma} - V^*(\bar{\sigma}) + \\ &\quad x^2 w\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right)\bar{\sigma} \end{aligned} \tag{38}$$

by reordering the terms we obtain

$$\begin{aligned} &= \left(\|x - \bar{c}\|^2 w\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right) - U^*(\bar{\sigma})\right)\bar{\sigma} \\ &\quad -V^*(\bar{\sigma}) + \frac{1}{2}\beta\bar{c}^2 - f\bar{c}, \end{aligned} \tag{39}$$

3

Considering the (10), setting $\bar{\epsilon} = \|x - \bar{c}\|^2$ and $\phi'\left(\phi^{-1}(s(\bar{\sigma}))\right) = \phi'(\bar{\epsilon})$ we obtain:

$$[w\phi'(\bar{\epsilon})\bar{\epsilon} - w\phi'(\bar{\epsilon})\bar{\epsilon} + w\phi(\bar{\epsilon})][w\phi(\bar{\epsilon}) - y] -$$
$$\frac{1}{2}(w\phi(\bar{\epsilon}) - y)^2 + y(w\phi(\bar{\epsilon}) - y) + \frac{1}{2}\beta\bar{c}^2 - f\bar{c} =$$
$$w^2\phi(\bar{\epsilon})^2 - yw\phi(\bar{\epsilon}) - \frac{1}{2}(w\phi(\bar{\epsilon}) - y)^2$$
$$-yw\phi(\bar{\epsilon}) + y^2 + \frac{1}{2}\beta\bar{c}^2 - f\bar{c} \quad (40)$$

by collecting the terms we obtain:

$$(w\phi(\bar{\epsilon}) - y)^2 - \frac{1}{2}(w\phi(\bar{\epsilon}) - y)^2 + \frac{1}{2}\beta\bar{c}^2 - f\bar{c}, \quad (41)$$

that is

$$\frac{1}{2}\left(w\phi(\|x - \bar{c}\|^2) - y\right)^2 + \frac{1}{2}\beta\bar{c}^2 - f\bar{c} = P(\bar{c}). \quad (42)$$

that proves the theorem. $\qquad\square$

Theorem 3.1 shows that the problem $(\mathcal{P}^d)$ is canonically dual to the primal $(\mathcal{P})$ in the sense that the duality gap is zero.

## 4. Gaussian function

One of the most used RBF is the Gaussian function. In this section we will analyze the problem with $\phi(\|x - c\|^2) = \exp\left\{-\frac{\|x-c\|^2}{2\alpha^2}\right\}$, where $\alpha$ is a parameter that represents the standard deviation of the Gaussian function. In the RBFNN formulation normally there is no the linear term $fc$. The primal problem is:

$$\min P(c) = \frac{1}{2}\left(w\exp\left\{-\frac{\|x-c\|^2}{2\alpha^2}\right\} - y\right)^2 + \frac{1}{2}\beta c^2 \quad (43)$$

If we define the quantity $d(c) = \frac{\|x-c\|^2}{2\alpha^2}$, the nonlinear operator $\xi : \mathbb{R} \to \mathcal{E}_a$ from (8) becomes

$$\xi = w\exp\{-d(c)\}. \quad (44)$$

The expressions that define $\sigma$, $V$ and $V^*$ are the same as the general problem that is:

- $V(\xi(c)) = \frac{1}{2}(\xi - y)^2$;
- $\sigma = \xi - y$;
- $V^*(\sigma) = \left(\frac{1}{2}\sigma^2 + y\sigma\right)$.

The second order operator $\Lambda_2(c) : \mathbb{R} \to \mathcal{E}_b$ is

$$\epsilon = \Lambda_2(c) = \|x - c\|^2 = \epsilon \quad (45)$$

The second level canonical function becomes

$$U(\epsilon) = w\exp\left\{-\frac{\epsilon}{2\alpha^2}\right\}. \quad (46)$$

And the second order duality mapping $\tau$ is

$$\tau = w\phi'(\epsilon) = -\frac{w}{2\alpha^2}\exp\left\{-\frac{\epsilon}{2\alpha^2}\right\}. \quad (47)$$

So the Legendre conjugate $U^* : \mathcal{S}'_b \to \mathbb{R}$ is

$$U^*(\tau) = \tau\left(\phi^{-1}\left(\frac{\tau}{w}\right)\right)' - w\phi\left(\phi^{-1}\left(\frac{\tau}{w}\right)\right)'$$
$$= -2\alpha^2\tau\left(\ln\left(\frac{-2\alpha^2\tau}{w}\right) - 1\right). \quad (48)$$

The derivative of the exponential function is the exponential function itself. This simplifies the relation (18) between $\xi$ and $\tau$ making it linear, that is $\xi = -\frac{\tau}{2\alpha^2}$. The relation between $\sigma$ and $\tau$ is:

$$\tau = -\frac{(\sigma + y)}{2\alpha^2} \quad (49)$$

that is also linear. The total complementarity function becomes:

$$\Xi(c, \sigma) = \frac{1}{2}c^2 G(\sigma) - cF(\sigma) - U^*(\sigma)\sigma - V^*(\sigma) -$$
$$\frac{x^2(\sigma^2 + y\sigma)}{2\alpha^2} \quad (50)$$

where:

$$G(\sigma) = \beta - \frac{\sigma^2 + y\sigma}{\alpha^2}$$
$$F(\sigma) = -\frac{x\sigma^2 + xy\sigma}{\alpha^2}$$
$$U^*(\sigma) = (\sigma + y)(\ln(s(\sigma)) - 1)$$
$$s(\sigma) = \frac{\sigma + y}{w}$$

The dual problem is

$$P^d(\sigma) = -\frac{1}{2}\frac{F(\sigma)^2}{G(\sigma)} - \ln(s(\sigma))\left(\sigma^2 + y\sigma\right) + \frac{1}{2}\sigma^2$$
$$-\frac{x^2(\sigma^2 + y\sigma)}{2\alpha^2} \quad (51)$$

The domains of the variables in the primal and dual problems are:

- $\mathcal{E}_b = \{\epsilon \in \mathbb{R} | \epsilon \geq 0\}$

- $\mathcal{S}_b = \{\tau \in \mathbb{R} | -\infty < \tau < 0\}$ if $w > 0$, $\mathcal{S}_b = \{\tau \in \mathbb{R} | -\infty < \tau < 0\}$ if $w < 0$

- $\mathcal{E}_a = \{\xi \in \mathbb{R} | 0 \leq \xi \leq w\}$

- $\mathcal{S}_a = \{\sigma \in \mathbb{R} | -y \leq \sigma \leq w - y\}$ if $w > 0$, $\mathcal{S}_a = \{\sigma \in \mathbb{R} | w - y \leq \sigma \leq -y\}$ if $w < 0$

**Remark 1.** *Parameters $\beta$, x, y, and w play important roles in solving the non-convex problem (P). In the original problem (7) one searches for the value of c that brings the term $w\exp\{-d(c)\}$ as closer as possible to y, that is $\sigma = w\exp\{-d(c)\} - y = 0$.*
*If $y < 0$ and $w > 0$ or $y > 0$ and $w < 0$ we will have that $|\sigma| > 0$. This means that in the case of the exponential function, it would be better to choose c as bigger as possible in order to make the exponential go to zero, but the result would never be satisfactory as the error committed by the approximation would go close to $-y$ as c goes to infinity. The value $-y$ is not a good value for the error as it is far from zero. On the other hand if y*

*and w have the same sign and $|y| > |w|$ the value of c will be x in order to have the exponential equal to 1 and to have the lowest value for $\sigma = w \exp\{-d(c)\} - y$.*

*In order to have a realistic problem, we will consider the case with y and w with the same sign, and with $|y| < |w|$. The cases with $y, w > 0$ and $y, w < 0$ are equivalent, so we will suppose that both y and w are positive without losing generality.*

**Theorem 4.1.** *Suppose that $\bar{\sigma} \in \mathcal{S}_a$ is a critical point of the dual problem (51) with the corresponding $\bar{c} = \frac{F(\bar{\sigma})}{G(\bar{\sigma})} \in \mathbb{R}$ and that $\bar{\sigma} \neq \frac{y}{2}$. Then $\bar{c}$ is a critical point of the primal problem and:*

$$P^d(\bar{\sigma}) = P(\bar{c}). \tag{52}$$

*moreover, there are the following relations between the critical points of the primal problem and the dual problem:*

1. *If $(2\bar{\sigma} + y) > 0$ and $G(\bar{\sigma}) \geq 0$ or $(2\bar{\sigma} + y) < 0$ and $G(\bar{\sigma}) \leq 0$ then if $\bar{\sigma}$ is a local minimum of the dual problem, the corresponding $\bar{c}$ is a local maximum of the primal problem; if $\bar{\sigma}$ is a local maximum of the dual problem the corresponding $\bar{c}$ is a local minimum of the primal problem;*

2. *If $(2\bar{\sigma} + y) > 0$ and $G(\bar{\sigma}) \leq 0$ or $(2\bar{\sigma} + y) < 0$ and $G(\bar{\sigma}) \geq 0$ then if $\bar{\sigma}$ is a local minimum of the dual problem the corresponding $\bar{c}$ is a local minimum of the primal problem; if $\bar{\sigma}$ is a local maximum of the dual problem the corresponding $\bar{c}$ is a local maximum of the primal problem.*

*Let $x_o = \sqrt{-2\alpha^2 \ln\left(\frac{y}{2w}\right)}$. If $\bar{\sigma} = -\frac{y}{2}$, then there is a corresponding critical point to $\bar{\sigma}$ in the primal problem if and only if the parameters x, y, $\beta$ and w satisfy one of the two following conditions:*

$$\begin{aligned} \beta x + \left(\beta + \frac{y^2}{4\alpha^2}\right) x_o &= 0 \\ \beta x - \left(\beta + \frac{y^2}{4\alpha^2}\right) x_o &= 0 \end{aligned} \tag{53}$$

*and the corresponding critical point $\bar{c}$ in the primal problem is always a local minimum. If neither of conditions (53) is satisfied, $\bar{\sigma} = -\frac{y}{2}$ is always a critical point of the dual problem, but it does not have any corresponding critical point in the primal problem.*

**Proof 4.1.** The first order derivative for the dual problem is:

$$P^d(\sigma)' = -\left[\left(x - \frac{F(\sigma)}{G(\sigma)}\right)^2 \frac{1}{2\alpha^2} + \ln(s(\sigma))\right][2\sigma + y] \tag{54}$$

so the term (28) is equal to $2\bar{\sigma} + y$. If $\bar{\sigma} \neq -\frac{y}{2}$, the critical point equivalency and condition (52) are consequences of Theorem 3.1.

To prove statements (*i*) and (*ii*) we use the second order derivatives of the problems $P(c)$ and $P^d(\sigma)$

$$\begin{aligned} P(c)'' &= \frac{(x-c)^2}{\alpha^4} \exp\{-d(c)\} (2w \exp\{-d(c)\} - y) \\ &+ \beta - \frac{1}{\alpha^2} w \exp\{-d(c)\} \left(w \exp\{-d(c)\} - y\right) \end{aligned} \tag{55}$$

$$\begin{aligned} P^d(\sigma)'' &= -\frac{1}{\alpha^2}\left(x - \frac{F(\sigma)}{\sigma}\right)^2 \left(1 + \frac{(2\sigma + y)^2}{\alpha^2 G(\sigma)}\right) \\ &\quad - \frac{2\sigma + y}{\sigma + y} - 2\ln(s(\sigma)). \end{aligned} \tag{56}$$

Since $\bar{\sigma}$ is a critical point of the dual, we have that $P^d(\sigma)' = 0$. Therefore when $\bar{\sigma} \neq -\frac{y}{2}$:

$$\left(x - \frac{F(\bar{\sigma})}{G(\bar{\sigma})}\right)^2 = -2\alpha^2 \ln(s(\bar{\sigma})) \tag{57}$$

By using condition (57) in (56) we obtain:

$$P^d(\bar{\sigma})'' = (2\bar{\sigma} + y)\left(\frac{2\ln(s(\bar{\sigma}))(2\bar{\sigma} + y)}{\alpha^2 G(\bar{\sigma})} - \frac{1}{\bar{\sigma} + y}\right). \tag{58}$$

Noticing $\sigma = w \exp\{-d(c)\} - y$, it is possible to rewrite $P(\bar{c})''$ in terms of $\bar{\sigma}$, i. e.:

$$P(c(\bar{\sigma}))'' = G(\bar{\sigma}) + \frac{2}{\alpha^2}(\bar{\sigma} + y)(2\bar{\sigma} + y)\left(x - \frac{F(\bar{\sigma})}{G(\bar{\sigma})}\right)^2. \tag{59}$$

by using again condition (57) we obtain:

$$P(c(\bar{\sigma}))'' = \frac{1}{\alpha^2}\left[\alpha^2 G(\bar{\sigma}) - 2(\bar{\sigma} + y)(2\bar{\sigma} + y)\ln(s(\bar{\sigma}))\right] \tag{60}$$

so it is possible to rewrite equation (58) in the following form:

$$P^d(\bar{\sigma})'' = -\frac{2\bar{\sigma} + y}{G(\bar{\sigma})(\bar{\sigma} + y)} P(c(\bar{\sigma}))''. \tag{61}$$

and to find the relations reported in Table 1. From these relations, we obtain:

- If $(2\sigma + y) > 0$ and $G(\sigma) \geq 0$ or $(2\sigma + y) < 0$ and $G(\sigma) \leq 0$ then the second order derivate of the primal problem and the second order derivate of the dual problem have opposite sign at their critical points;

- If $(2\sigma + y) > 0$ and $G(\sigma) \leq 0$ or $(2\sigma + y) < 0$ and $G(\sigma) \geq 0$ then the second order derivate of the primal problem and the second order derivate of the dual problem have the same sign at their critical points.

This proves statements 1 and 2.

| $(2\bar{\sigma} + y)$ | $G(\bar{\sigma})$ | $P(c(\bar{\sigma}))$ | $P^d(\bar{\sigma})$ |
|---|---|---|---|
| $> 0$ | $> 0$ | $\pm$ | $\mp$ |
| $> 0$ | $< 0$ | $\pm$ | $\pm$ |
| $< 0$ | $< 0$ | $\pm$ | $\mp$ |
| $< 0$ | $> 0$ | $\pm$ | $\pm$ |

Table 1: Relations between the second order derivatives of the primal problem and dual problem

The point $\bar{\sigma} = -\frac{y}{2}$ is a critical point of $P^d$ according to the second part of the (54). The point $\bar{c}$ corresponding to $\bar{\sigma} = -\frac{y}{2}$

5

is a critical point of the primal problem if and only if $P'(\bar{c}) = 0$. We can use the (10) to find the relation between $\bar{\sigma}$ and $\bar{c}$ that is:

$$\bar{\sigma} = \bar{\xi} - y \rightarrow \bar{\sigma} = w \exp\{-d(\bar{c})\} - y \qquad (62)$$

$$\bar{c} = x \pm \sqrt{-2\alpha^2 \left(\ln\left(s(\bar{\sigma})\right)\right)}. \qquad (63)$$

For $\bar{\sigma} = -\frac{y}{2}$ we obtain:

$$\bar{c} = x \pm x_o. \qquad (64)$$

Substituting these values in the first order derivative of the primal problem:

$$P'(\bar{c}) = \frac{1}{2}d(\bar{c})w \exp\{-d(\bar{c})\}\left(w \exp\{-d(\bar{c})\} - y\right) + \beta\bar{c} \qquad (65)$$

and considering that $w \exp\{-d(\bar{c})\} = \bar{\sigma} + y = \frac{y}{2}$ and $w \exp\{-d(\bar{c})\} - y = \bar{\sigma} = -\frac{y}{2}$ we obtain that the primal problem has a critical point at $\bar{c}$ corresponding to the critical $\bar{\sigma} = -\frac{y}{2}$ if and only if:

$$\beta x \pm \left(\beta + \frac{y^2}{4\alpha^2}\right) x_o = 0. \qquad (66)$$

This happens only for a particular configuration of the parameters $w$, $\beta$, $x$ and $y$ that makes one of the roots the first term of the derivative (54):

$$-\left[\left(x - \frac{F(\bar{\sigma})}{G(\bar{\sigma})}\right)^2 \frac{1}{2\alpha^2} + \left(\ln\left(s(\bar{\sigma})\right)\right)\right] = 0 \qquad (67)$$

be in $\bar{\sigma} = -\frac{y}{2}$.

To prove that at $\bar{\sigma} = -\frac{y}{2}$ the critical point of the dual problem corresponds to a minimum point of the primal problem we plug the value of $\bar{\sigma} = -\frac{y}{2}$ in the (59) and obtain

$$P''(\bar{\sigma}) = \beta + \frac{y^2}{4\alpha^2}, \qquad (68)$$

which is always a positive value. $\qquad\square$

**Remark 2.** *From now on we will refer to the critical point $\sigma_f = -\frac{y}{2}$ as pseudo dual critical point as it is a critical point of the dual problem that generally does not have a corresponding critical point for the primal problem.*

*4.1. Choice of the critical point*

In order to find the best solution among the critical points of problem (43) we introduce the following feasible spaces:

$$S_a^+ = \{\sigma \in S_a | G(\sigma) > 0\} \qquad (69)$$

$$S_a^- = \{\sigma \in S_a | G(\sigma) < 0\} \qquad (70)$$

The following theorem explains the relations between the critical points:

**Theorem 4.2.** *Suppose that the point $\bar{\sigma}_1 \in S_a^+$ and $\bar{\sigma}_2 \in S_a^-$ are critical points of the dual problem, that $\bar{\sigma}_i \neq -\frac{y}{2}$ for $i = 1, 2$ and that $\bar{c}_1$ and $\bar{c}_2$ are the corresponding critical points of the primal problem. Then if both $\bar{c}_1$ and $\bar{c}_2$ are local minima or local maxima of the primal problem, the following relation always holds:*

$$P(\bar{c}_1) = P^d(\bar{\sigma}_1) < P(\bar{c}_2) = P^d(\bar{\sigma}_2) \qquad (71)$$

**Proof 4.2.** This theorem is a consequence of the first theorem in triality theory [8]. $\qquad\square$

**Remark 3.** *The pseudo critical point $\sigma_f = -\frac{y}{2}$ is always in $S_a^+$.*

From the results in Theorem 4.2 it is always better to search for the dual critical point in $S_a^+$ that corresponds to a minimum in the primal problem. In order to characterize the solutions in $S_a^+$ and the domains in which search for the best solution, two theorems are proposed in the following:

**Theorem 4.3.** *Let $\sigma_f = -\frac{y}{2}$ be the pseudo critical point of the dual problem, $x_o = \sqrt{-2\alpha^2 \ln\left(\frac{y}{2w}\right)}$, x positive. Then:*

- *if $x \in (0, x_o)$ then $\sigma_f$ is always a local minimum of $P^d(\sigma)$;*

- *if $x > x_o$ then:*

  1. *if $\beta > 0$ and $\beta < \frac{y^2 x_o}{4\alpha^2(x-x_o)}$, $\sigma_f$ is a local minimum for the dual problem;*

  2. *if $\beta > 0$ and $\beta > \frac{y^2 x_o}{4\alpha^2(x-x_o)}$, $\sigma_f$ is a local maximum for the dual problem;*

  3. *if $\beta > 0$, $\beta = \frac{y^2 x_o}{4\alpha^2(x-x_o)}$, $\sigma_f$ is an inflection point in which the first order derivative is zero and that corresponds to a a local minimum of the primal problem.*

**Proof 4.3.** In order to understand that $\sigma_f = -\frac{y}{2}$ is a minimum or a maximum for the dual we have to plug its value in the second order derivative of $P^d(\sigma)$ that is equation (56) and analyze its sign. After the substitution we obtain

$$P^d(\sigma_f) = -\left[2\ln\left(-\frac{y}{2w}\right) + \frac{1}{\alpha^2}\left(\frac{x\beta}{\beta + \frac{y^2}{4\alpha^2}}\right)^2\right]. \qquad (72)$$

The first order derivate in $\beta$ of (72) is $-\frac{2x\beta^2}{\alpha^2\left(\beta + \frac{y^2}{4\alpha^2}\right)^2}$, that is the function is monotonic decreasing in $\beta$. The value of (72) in $\beta = 0$ is $-\ln\left(-\frac{y}{2w}\right)$ that is positive. If we make $\beta$ go to $+\infty$ we obtain:

$$\lim_{\beta \to +\infty} -\left[2\ln\left(-\frac{y}{2w}\right) + \frac{1}{\alpha^2}\left(\frac{x\beta}{\beta + \frac{y^2}{4\alpha^2}}\right)^2\right] = -2\ln\left(-\frac{y}{2w}\right) + \frac{x^2}{\alpha^2} \qquad (73)$$

that is the second order derivative of $P^d(\sigma)$ in $\sigma_f$ is non negative for any value of $\beta > 0$ if

$$x \in [-x_o, x_o] \qquad (74)$$

If $x$ does not satisfy this condition, from the (72) we have that the second order derivative of the dual problem is positive in $\sigma_f$ if $\beta$ satisfies:

$$\beta > \frac{-y^2 x_o}{4\alpha^2 (x + x_o)} \text{ and } \beta < \frac{y^2 x_o}{4\alpha^2 (x - x_o)}. \tag{75}$$

On the other hand if:

$$\beta < \frac{-y^2 x_o}{4\alpha^2 (x + x_o)} \text{ or } \beta > \frac{y^2 x_o}{4\alpha^2 (x - x_o)} \tag{76}$$

there will be a local maximum in $\sigma_f$. As $x$ is considered positive, the term $\frac{-y^2 x_o}{4\alpha^2(x+x_o)}$ is always negative, so $\beta$ will always be greater than it.

If the condition $\beta = \frac{y^2 x_o}{4\alpha^2(x-x_o)}$ is satisfied, the critical point $\sigma_f$ is an inflection point that also satisfies the first order condition and it has a corresponding minimum point in the primal problem for Theorem 4.1. □

**Remark 4.** *In the case of $x$ negative, the conditions are changed in the following way:*

- *if $x \in (-x_o, 0)$ then $\sigma_f$ is always a local minimum of $P^d(\sigma)$*

- *if $x < -x_o$ then:*

  1. *if $\beta > 0$ and $\beta < \frac{-y^2 x_o}{4\alpha^2(x+x_o)}$, $\sigma_f$ is a local minimum for the dual problem;*

  2. *if $\beta > 0$ and $\beta > \frac{-y^2 x_o}{4\alpha^2(x+x_o)}$, $\sigma_f$ is a local maximum for the dual problem;*

  3. *if $\beta > 0$, $\beta = \frac{-y^2 x_o}{4\alpha^2(x+x_o)}$, $\sigma_f$ is an inflection point in which the first order derivative is zero and that corresponds to a a local minimum of the primal problem.*

*The proof of these statement is similar to that of Theorem 4.3 and can be omitted.*

**Remark 5.** *Theorem 4.3 shows the effects of the parameter $\beta$ on the pseudo critical point $\sigma_f$. Similar effects can also be obtained in respect to $y$, $x$, $\alpha$, and $w$. The reason we choose $\beta$ is because it is an hyper-parameter that can be chosen by the practitioner before performing the optimization.*

For the next theorem, we introduce the two following subsets of $\mathcal{S}_a^+$:

$$\mathcal{S}_\sharp^+ = \left\{ \sigma \in \mathcal{S}_a^+ | \sigma > -\frac{y}{2} \right\} \tag{77}$$

$$\mathcal{S}_\flat^+ = \left\{ \sigma \in \mathcal{S}_a^+ | \sigma < -\frac{y}{2} \right\} \tag{78}$$

**Theorem 4.4.** *Let $\sigma_f = -\frac{y}{2}$ be the pseudo critical point in the dual problem and let the primal problem have a maximum of five critical points. Then*

- *if $\sigma_f$ is a local minimum for the dual function, there will be a local maximum in $\mathcal{S}_\sharp^+$ that corresponds to a minimum of the primal problem.*

- *if $\sigma_f$ is a local maximum then:*

  1. *there are no critical points in $\mathcal{S}_\sharp^+$;*

  2. *there is at least one critical point in $(\mathcal{S}_\flat^+$*

**Proof 4.4.** In the dual problem there must be a singularity point in $G(\sigma) = 0$ that goes to $-\infty$, so if $\sigma_f$ is a local minimum, there must be a local maximum in $\mathcal{S}_\sharp^+$.

If $\sigma_f$ is a local maximum, we prove condition (*i*) by negating the thesis and suppose that there is a least one critical point in $\mathcal{S}_\sharp^+$. As $P^d(\sigma)$ goes to $-\infty$ if $G(\sigma) \to 0$, there will be no one, but two critical points in $\mathcal{S}_\sharp^+$, a local minimum $\sigma_1$ and a local maximum $\sigma_2$ with the relation $P^d(\sigma_1) < P^d(\sigma_2)$. For Theorems 4.1 and 4.2, $\sigma_1$ corresponds to the second highest local maximum of the primal function $c_1$, and $\sigma_2$ corresponds to the lowest or second lowest local minimum of the primal function $c_2$, that is the relation $P(c_2) < P(c_1)$ is satisfied. By Theorem 3.1 we have:

$$P^d(\sigma_1) < P^d(\sigma_2) = P(c_2) < P(c_1) = P^d(\sigma_1) \tag{79}$$

that is a contradiction.

To prove condition (*ii*), it is sufficient to notice that if there are no critical points in $\mathcal{S}_\sharp^+$, for the triality theory there must be at least one critical point corresponding to the global minimum in $\mathcal{S}_a^+$ and this point will be in $\mathcal{S}_\flat^+$. □
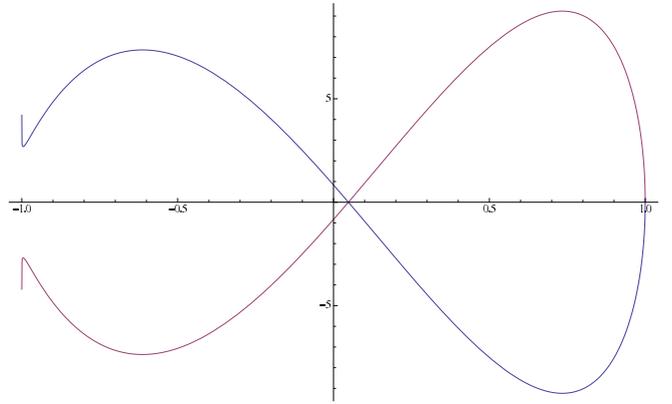


Figure 1: Dual algebraic curves with $y = 1$, $w = 2$, $\alpha = \frac{\sqrt{2}}{2}$ and $\beta = 0.1$ in respect to the internal input $x$

Depending on the parameters, the primal problem (43) can have at most five critical points. There are several cases:

**Case 1**: Three critical points for $P(c)$ and four critical points for $P^d(\sigma)$, two critical point in $\mathcal{S}_a^+$ and two critical points in $\mathcal{S}_a^-$, with $\sigma_f$ as local minimum. The values of the parameters are $y = 1$, $x = 1$, $w = 2$, $\alpha = \frac{\sqrt{2}}{2}$, $\beta = 0.1$ (see Figure 2). This case can be easily solved with the general canonical duality framework[8], as the local maximum in $\mathcal{S}_a^+$ corresponds to the global minimum of the problem, and the local minimum and maximum in $\mathcal{S}_a^-$ correspond to the local minimum and maximum in the primal problem.
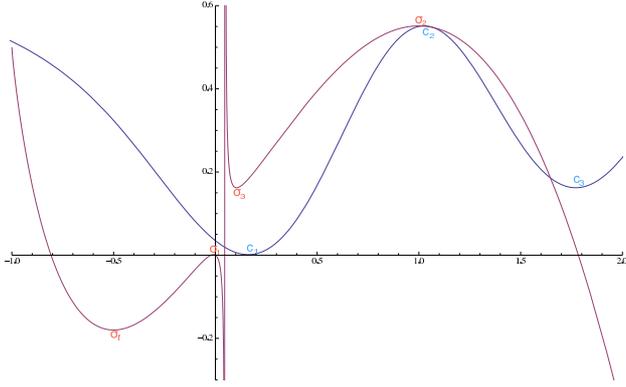
Figure 2: Primal(in blue) and dual(in red) functions for Case 1 with three critical points
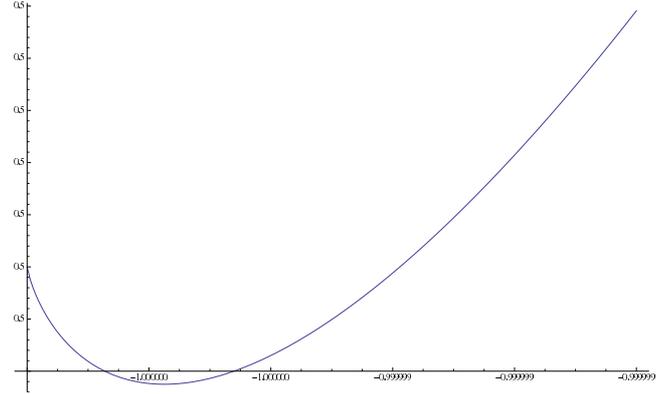


Figure 4: Critical point on the boundary of the dual function feasible set for Case 2.
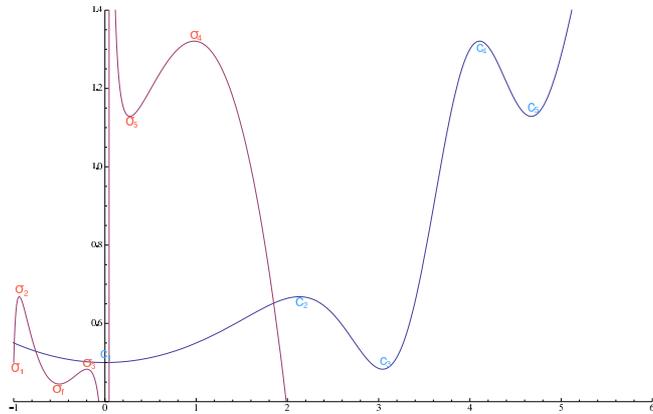


Figure 3: Primal(in blue) and dual(in red) functions for Case 2 with five critical points in the primal and six critical points in the dual.
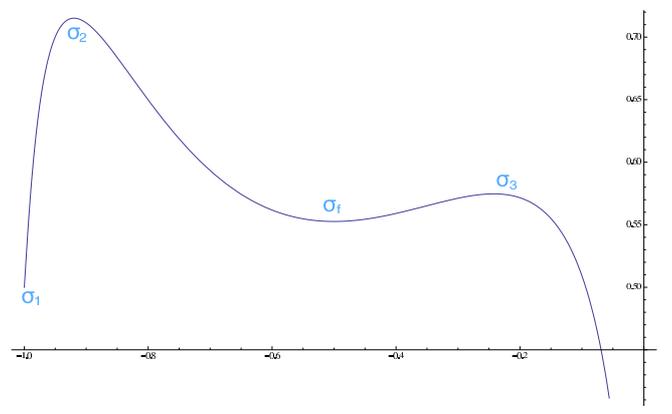


Figure 5: $\mathcal{S}_a^+$ of the dual problem in the case of $\beta = 0.12$. The minimum near the boundary $\sigma_1$ is a global minimum.

**Case 2**: Five critical points for $P(c)$, six critical points for $P^d(\sigma)$. The values of the parameters are $y = 1$, $x = 4$, $w = 2$, $\alpha = \frac{\sqrt{2}}{2}$ and $\beta = 0.1$ (see Figure 3). Notice that the only parameter that changed in respect to Case 1 is $x$. With these parameters the problem becomes multi-welled. The two critical points with the lowest value of the objective function belong to the same double well and their corresponding critical points are in $\mathcal{S}_a^+$. The critical point $\sigma = -0.999999$ of $P^d(\sigma)$ is corresponding to the second best minimizer $c = 0.00002$ of the primal problem and this $\sigma$ is situated near the boundary of $\mathcal{S}_b^+$ which is visible in Figure 4. It is also possible, for certain values of the parameters, that the local minimum on the boundary of $S_a$, corresponds to the global minimum of the problem (see Figure 5). In this case the choice of the value for $\sigma$ should be the critical point near the boundary. This critical point corresponds to a critical point in the primal with the value of $c$ near zero. This critical point is generated by the term $\frac{1}{2}\beta c^2$ that is the regularization term used to make the objective function coercive and more regular. On the other hand, this term doesn't have anything to do with the original aim of the problem. This point near zero in the primal function will always have the corresponding dual critical point near the boundary, because as $c$

gets close to zero, $\sigma = w \exp\{-d(c)\} - y$ gets close to $-y$. We also consider that $\sigma = w \exp\{-d(c)\} - y$ is the error that originally we want to minimize in problem (6) and that the critical point on the boundary will always have a $\sigma$ with an absolute value bigger than the other critical point closer to $\sigma = 0$. In other words the local minimum on the boundary has nothing to do with the original problem, has an high value of the error and should not be considered as a good solution. In order to find the optimal solution for the original problem, the local minimum in the primal problem corresponding to the critical point closer to zero in $S_a^+$ is preferable. By reducing the value of $\beta$ it is possible not only to make the critical point near $c = 0$ into a local minimum, but also to assure that $\sigma_f$ is a local minimum. In this way there is a critical point in $\mathcal{S}_\sharp^+$ and the domain of the solution is well defined. Basically if the critical point near the boundary of $S_a^+$ is the global minimum, a very big value of $\beta$ has been chosen.

**Case 3**: Three critical points for $P(c)$ and four critical points for $P^d(\sigma)$, all belonging to $\mathcal{S}_a^+$. The values of the parameters are $y = 1$, $x = 4$, $w = 2$, $\alpha = \frac{\sqrt{2}}{2}$ and $\beta = 0.22$ (see Figure 6). This case is similar to the previous one, and the solution of the dual problem should be the critical point that corresponds to a

8
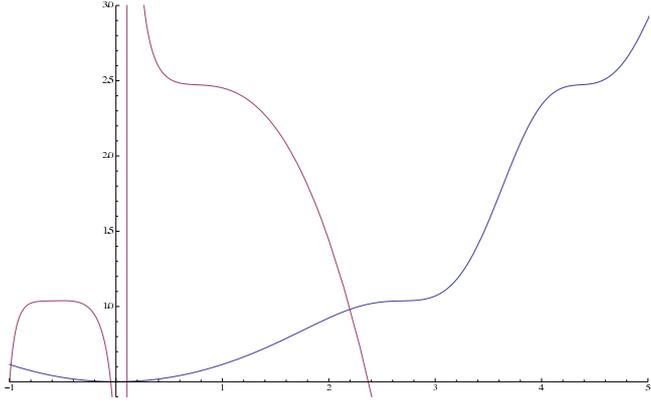
Figure 6: Primal(in blue) and dual(in red) functions for the Case 3 with three critical points in the primal and four critical points in $\mathcal{S}_a^+$.

minimum in the primal problem with the value of $\sigma$ closer to zero.
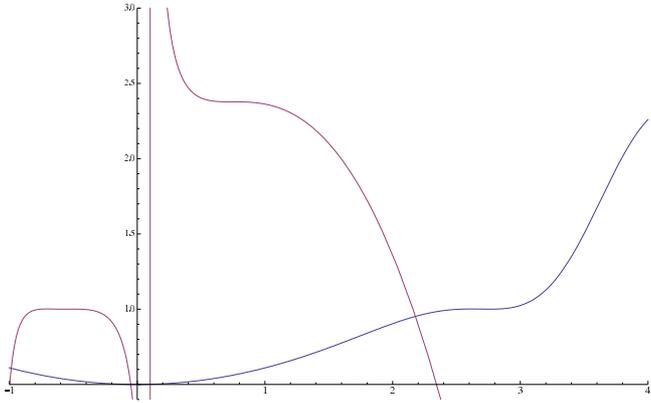


Figure 7: Primal(in blue) and dual(in red) functions for the Case 4 with three critical points in the primal and two critical points in $\mathcal{S}_a^+$ and two critical points in $\mathcal{S}_a^-$ and $\sigma_f$ as a local maximum.

**Case 4**: Three critical points in the primal and four critical points in the dual, but with two critical points in $\mathcal{S}_a^+$, two critical points in $\mathcal{S}_a^-$ and $\sigma_f$ as local maximum. The values of the parameters are $y = 1$, $x = 8$, $w = 2$, $\alpha = \frac{\sqrt{2}}{2}$ and $\beta = 0.25$ (see Figure 7). If the value of the hyper parameter $\beta$ is reduced it is possible to make $\sigma_f$ into a local minimum and return in one of the previous cases.

**Case 5**: One critical point in the primal problem and two critical points in the dual problem. This case occurs when the quadratic term with beta dominates the error function $W(x)$. If this case occurs, it means that the value of $\beta$ is too big and the problem is not related with the original anymore, so one should choose a smaller value of $\beta$ to have a problem related to the original.

Based on the study of these cases, we can obtain the general idea to find the best solution, i. e. the hyper parameter $\beta$ should be set to a value that satisfies condition (75) in order to have $\sigma_f$ as a local minimum, then search for the critical point in the

domain $\mathcal{S}_\sharp^+$.

## 5. Conclusions

In this paper we have presented an application of the canonical duality theory to function approximation using Radial Basis Functions. By using the sequential dual canonical transformation, the non convex problem with a general RBF function $\phi(\cdot)$ is reformulated in a canonical dual form. An associated strong duality theorem is also proposed.

Applications to one of the most used RBF, the exponential function, are illustrated. Due to the particular properties of the exponential function, we are able to find a linear relation between the dual variables, which leads to an explicit form of the canonical dual problem. We also found conditions on the hyper parameter $\beta$ in order to obtain a reliable domain where to search for the best solution. This research reveals an important phenomenon in complex systems, i.e. the global optimal solution may not be the best solution to the problem considered.

There are still several open topics on the application of the canonical duality theory to Radial Basis Error functions. For example there are other kinds of RBF that can be analyzed, like the multi quadratic and the multi quadratic inverse functions, a further development for future research is to expand the one dimensional case to the multidimensional case with also considering $w$ as a variable and not as a parameter. When this case is analyzed, we will be able to realize RBF neural networks based on canonical duality theory.

## References

[1] M. J. D. Powell, "Radial basis functions in 1990," *Adv. Numer. Anal.*, **2**, 105-210 (1992).

[2] S. Haykin,"Neural Networks, a Comprehensive Foundation," Prentice-Hall, (1999).

[3] L. Bruzzone, D. Prieto, "Supervised training techniques for radial basis function neural networks." *Electronic Letters*, **34**(11), 1115 1116 (1998).

[4] D. Wettschereck, T. Dietterich , "Improving the Performances of Radial Basis Functions Networks by Learning Center Locations," *Advances in Neural Information Processing Systems* (1992).

[5] C. Buzzi, L. Grippo, M. Sciandrone , "Convergent decomposition techniques for training RBF neural networks," *Neural Computation*, **13**, pp. 1891-1920 (2001).

[6] J. J. Moré, Z. J. Wu, "Global continuation for distance geometry problems," *SIAM Journal on Optimization*, **7**(3): 814-836 (1997).

[7] J. Saxe, "Embeddability of weighted graphs in k-space is strongly NP-hard," in *Proc. 17th Allerton Conference in Communications, Control, and Computing*, Monticello, IL, 1979: 480-489 (1979).

[8] D.Y. Gao , "Duality Principles in Nonconvex Systems: Theory, Methods, and Applications," *Nonconvex Optimization and Its Applications*, Kluwer Academic Publishers (2000).

[9] D.Y. Gao, "Canonical dual transformation method and generalized triality theory in nonsmooth global optimization," *J. Glob. Optim.* **17**(1/4), 127-160 (2000).

[10] D.Y. Gao , "Canonical duality theory: theory, method, and applications in global optimization," *Comput. Chem.* **33**, 1964-1972 (2009).

[11] D.Y. Gao, N. Ruan and P.M. Pardalos, "Canonical dual solutions to sum of fourth-order polynomials minimization problems with applications to sensor network localization," in *Sensors: Theory, Algorithms and Applications*, P.M. Pardalos, Y.Y. Ye, V. Boginski, and C. Commander (eds). Springer (2010).

[12] T.K. Gao, "Complete solutions to a class of 8th order polynomial optimization problems", to appear in *IMA J. Applied Mathematics*, published online at http://arxiv.org/abs/1205.6886 arXiv:1205.6886 (2012).

[13] N. Ruan and D.Y. Gao, "Global optimal solutions to a general sensor network localization problem", to appear in *Performance Evaluation* , published online at http://arxiv.org/submit/654731 (2013).

[14] Z.B. Wang, S.C. Fang, D.Y. Gao, W.X. Xing, "Canonical dual approach to solving the maximum cut problem," *J. Glob. Optim.*, **54**, 341-352 (2012).

[15] J. Zhang , D.Y. Gao, J. Yearwood, "A novel canonical dual computational approach for prion AGAAAAGA amyloid fibril molecular modeling," *Journal of Theoretical Biology*, **284**, 149-157 (2011).