MATHEMATICAL ANALYSIS OF SOME NEURAL NETWORKS FOR SOLVING LINEAR AND QUADRATIC PROGRAMMING*

ZHANG XIANGSUN (章祥荪)

(*Institute of Applied Mathematics, the Chinese Academy of Sciences, Beijing* 100080, *China*)

Abstract

Artificial neural network techniques have been introduced into the area of optimization in the recent decade. Some neural network models have been suggested to solve linear and quadratic programming problems. The Kennedy and Chua model[5] is one of these networks. In this paper results about the convergence of the model are obtained. Another related problem is how to choose a parameter value $\bar{s}$ so that the equilibrium point of the network immediately and properly approximates the original solution. Such an estimation for the parameter is given in a closed form when the network is used to solve linear programming.

Key words. Neural network, linear programming, quadratic programming, penalty function method

1. Introduction

Artificial neural network techniques have been introduced into the area of optimization in the past ten years as a promising new methodology. It is the parallel and collective computational property of the neural network that makes its application to mathematical programming significant. Hopfield and Tank's works[1–3] greatly enhanced the ANN application in optimization. Among others, application in Linear Programming (LP) and Quadratic Programming (QP) has been given extra attention by many authors[3,5–13]. Nonlinear network with integrators as the basic components to solve nonlinear programming was developed by Chua and Lin in [4].

Compared with the ANN optimization model construction efforts, the mathematical analysis of the suggested ANN models is in the state of unsatisfactoriness. At least, it is necessary to compare the neural network applications to mathematical programming with the existing mature theory in the field of nonlinear programming algorithms in order to incorporate and improve the existing ANN analysis techniques and results. In [8], Maa

Received June 18, 1994.

* This work is supported by the National Natural Science Foundation of China (No. 69574034) and the Laboratory of Management, Decision and Information System, CAS.

and Shanblatt initiated such a research where traditional energy function research of ANN was then connected and compared with exterior penalty function method in mathematical programming. Secondary relationship between the equilibrium points (solutions of the equilibrium equations with different parameters) and the solution of given LP or QP was studied.

It should be noted that the penalty function method is not designed for solving LP or QP, it is preferably used to solve nonlinear programming problems with nonlinear constraints. LP and QP problems have much more efficient algorithms. It then suggests to us that when using penalty function method to solve LP or QP problems the assumptions or conditions for guaranteeing the convergence could be simplified or relieved, and the convergence mechanics would be clearer. It is the purpose of this paper to reorganize the results in [8] and present new convergence properties of the ANN models for solving LP or QP.

In Section 2 the related mathematical concepts and theorems for both in mathematical programming and ANN are given in a succinct way. Section 3 discusses the convergence property of the networks suggested by several authors[5-8] when they are applied to linear programming. Section 4 then deals with the quadratic programming. Some detailed proofs are put in the last section, an appendix.

2. Preliminaries

A general nonlinear programming problem (NLP) takes the following form:

$$\begin{array}{lll} \min & f(x), & \\ \text{subject to} & g_i(x) \le 0, & i \in I = \{1, 2, \cdots, m\}, \end{array} \tag{1}$$

where $x \in R^n$ and $f,\ g_i,\ i \in I,\ h_k,\ k \in K$ are real functions defined on R^n. Let the feasible set of the above problem be X. The Lagrange function is

$$L(x, \lambda) = f(x) + \lambda^T g(x), \tag{2}$$

where $g(x) = \big(g_1(x), \cdots, g_m(x)\big)^T$, and $\lambda \in R^m$ is the so-called Lagrange multiplier vector. For $\overline{x} \in X$, let $I(\overline{x}) = \big\{i:\ i \in I \text{ and } g_i(\overline{x}) = 0\big\}$. For $i \in I(\overline{x})$, $g_i(x)$ is said to be active at $\overline{x}$. We make two assumptions:

Assumption 1. $f, g_i,\ i \in I$ are differentiable.

Assumption 2 (Regularity Assumption). At the solution x^* of the problem, $\big\{\nabla g_i(x^*),\ i \in I(x^*)\big\}$ is an independent set.

The following theorem is a well known result in nonlinear programming:

Theorem 1. Under Assumptions 1 and 2, a necessary condition for x^* to be a local minimum solution of problem (1) is that there exists $\lambda^* \in R^m$ such that x^*, λ^* satisfy:

$$\nabla_x L(x, \lambda) = \nabla f(x) + \nabla g(x)\lambda = 0, \tag{3}$$

$$\nabla_\lambda L(x, \lambda) = g(x) \le 0, \tag{4}$$

$$\lambda \ge 0, \tag{5}$$

$$\nabla_\lambda L(x, \lambda)^T \lambda = \lambda^T g(x) = 0. \tag{6}$$

The set of conditions (3)–(6) is given by H. kuhn and A. Tucker, and hence called the K-T conditions. λ^* is called the optimal Lagrange multiplier. When f and g_i are convex functions, these conditions are also sufficient and any local minimal solution is a global minimal solution. When f, g_i and h_k are all linear functions, then (NLP) is reduced to a linear programming problem:

$$\begin{array}{ll} \min & f(x) = a^T x, \\ \text{s.t.} & g(x) = Dx - b \leq 0, \end{array} \tag{7}$$

where D is an $m \times n$ matrix $(m \geq n)$, $a, x \in R^n$, $b \in R^m$. In this case Theorem 1 can be rewritten as:

Corollary 1. If g_i, $i \in I$, of problem (1) are linear functions, Assumption 2 can be dropped and Theorem 1 is still valid.

A proof of this corollary is given in Appendix. In this paper we also consider the following QP:

$$\begin{array}{ll} \min & f(x) = x^T Q x/2 + q^T x, \\ \text{s.t.} & g(x) = Dx - b \leq 0, \end{array} \tag{8}$$

where Q is an $n \times n$ symmetric matrix and $q \in R^n$.

The penalty and barrier function methods in nonlinear programming theory are algorithms for approximating a constrained NLP by a sequence of unconstrained NLP's. In the case of penalty methods the approximation is accomplished by adding to the objective function a term that prescribes a high penalty for violation of the constraints. According to the theory, problem (7) can be approximated by a sequence of unconstrained minimization problems:

$$\min \quad F_1(x, s_k) = a^T x + s_k \sum_{i=1}^{m} (g_i^+(x))^2, \tag{9}$$

where $g_i^+(x) = \max\{0, g_i(x)\} = \max\{0, d^i x - b_i\}$. And problem (8) can be solved by treating the following sequence of unconstrained minimization problems:

$$\min \quad F_2(x, s_k) = x^T Q x/2 + q^T x + s_k \sum_{i=1}^{m} (g_i^+(x))^2. \tag{10}$$

Remark 1. Noticing that $\max\{0, d^i x - b_i\}$ is a convex function of x, then $F_1(x, s_k)$, $F_2(x, s_k)$ (if Q is positive semi-definite) are convex functions of x. So the solution of problem (9) is simply given by equation

$$\nabla_x F_1(x, s_k) = a + s_k \sum_{i=1}^{m} g_i^+(x) d^{iT} = 0, \tag{11}$$

and the solution of problem (10) is given by

$$\nabla_x F_2(x, s_k) = Qx + q + s_k \sum_{i=1}^{m} g_i^+(x) d^{iT} = 0. \tag{12}$$

Remark 2. Any limit point of the sequence of solutions $\{x^k\}$, produced by (11) or (12), is a solution of the respective original problem if the following Assumptions 3 and 4 are satisfied (a part of Assumption 4 is automatically met here by Remark 1).

Assumption 3. $f(x) \to \infty$ as $\|x\| \to \infty$, or, X_ε is compact, where

$$X_\varepsilon = \{x : \ g_i(x) \leq \varepsilon, \ i = 1, \cdots, m\}$$

for a given $\varepsilon > 0$.

Assumption 4. $F_i(x,s)$, $i=1,2$ is a convex function of x for a fixed s.

Now we turn to the ANN model that solves LP and QP problems. The Kennedy and Chua model can be described by the following dynamic system:

$$\dot{x} = C^{-1}\{-\nabla f(x) - s\nabla g(x)g^+(x)\}, \tag{13}$$

where C is an $n\times n$ diagonal matrix with positive elements (without loss of generality we take $C=I$ in the following discussion), $\nabla g(x) = D^T$ and $g^+(x) = \big(g_1^+(x),\cdots,g_m^+(x)\big)^T$. When using this model to solve LP or QP, the energy function of the network,

$$E(x) = f(x) + s\sum_{i=1}^{m}\big(g_i^+(X)\big)^2 \tag{14}$$

is a quadratic function and has been shown to be a Lyapunov function for the network represented by (13), that is,

$$\frac{dE}{dt} = \{\nabla f(x) + s\nabla g(x)g^+(x)\}^T\dot{x} = -\dot{x}^T C\dot{x} \le 0, \tag{15}$$

$$\frac{dE}{dt} = 0 \Longleftrightarrow \dot{x} = -\nabla f(x) - s\nabla g(x)g^+(x) = 0. \tag{16}$$

It is noted by Maa and Shanblatt that the energy function (14) is exactly the objective function of the unconstrained minimization problem in penalty method. Comparing (16) with (11) and (12) we can restate the discussion in their paper in terms of neural networks as follows:

Remark 3. If $f(x)$ is a convex function(in the case of QP, provided Q is positive semi-definite), the solution of (16) is a stable point that minimizes the energy function of the Kennedy-Chua model. The original solution can be obtained when the parameter s in (13) is infinite. But it is not practical to do that. Then how to estimate a proper s is an open and significant problem. To avoid this difficulty, Rodriguez-Vasquez et al.[6] proposed a network model which is formed by two mutually exclusive sub-networks. The shortcoming of their model is discussed by Maa and Shanblatt[9].

3. Linear Programming

In this paper we deal with LP problems of form (7) where rank$(D)=n$. (7) is in the dual form of an LP. The corresponding equations (3)–(6) are

$$a + \sum_{i=1}^{m}\lambda_i d^{iT} = 0, \tag{17}$$

$$d^i x - b_i \le 0, \qquad i=1,\cdots,m, \tag{18}$$

$$\lambda_i \ge 0, \qquad i=1,\cdots,m, \tag{19}$$

$$\lambda_i(d^i x - b_i) = 0, \qquad i=1,\cdots,m. \tag{20}$$

Equations (17) and (20) could be combined into

$$a + \sum_{i\in I(x)}\lambda_i d^{iT} = 0, \tag{21}$$

where

$$I(x) = \{i: \ d^i x - b_i = 0\}. \tag{22}$$

Now we discuss the convergence property when the Kennedy and Chua model is applied to problem (7). According to Remarks 1–3, we immediately have

Theorem 2. If the feasible set X of problem (7) is bounded, then for an increasing sequence $\{s_k\}$, the network produces a sequence of approximate solutions x^k in the meaning that any limit point of it is a solution x^* of (7).

Remark 4. If problem (7) has a unique solution x^*, then the sequence of equilibrium points produced by the network with increasing s_k will converge to the solution.

Remark 5. If the set X is not bounded, then it is possible that there is no limit point existing. Then, the energy function value $F_1(x^k, s_k)$ gives a lower bound of $f(x^*)$, where x^k is the equilibrium point at $s = s_k$, i.e., $f(x^*) \geq F_1(x^k, s_k) \geq f(x^k)$.

The assumption of a bounded X is obviously excessive, a somewhat reasonable assumption is that the solution set $\overline{X}$,

$$\overline{X} = \{x: \ a^T x \leq a^T y \ \text{ for any } \ y \in X\}, \tag{23}$$

is bounded.

Lemma 1. If problem (7)'s solution set $\overline{X}$ is bounded, then $F_1(x, s_k)$ in (9) satisfies Assumption 3, i.e., $F_1(x, s_k) \to \infty$ as $\|x\| \to \infty$.

See the proof of this lemma in Appendix.

Theorem 3. If the solution set $\overline{X}$ of problem (7) is bounded, then Theorem 2 and Remarks 4, 5 are still true.

This theorem covers the main result, Theorem 3, in [8], but in their paper Assumption 2 is needed. For most practical LP problems, there is a unique optimal solution, then the assumption of Theorem 2 is met and the successive equilibrium points of the network with increasing s_k converge to the solution. Hence a related problem is how to choose a parameter $\widetilde{s}$ so that the equilibrium point given by the network with this determined $\widetilde{s}$ approximates the original solution properly.

There are different criteria to describe the accuracy of an approximation to the solution, such as, (a) Given an $\varepsilon > 0$, to find an $\widetilde{s}$ such that there is an optimal solution x^*, for $s \geq \widetilde{s}$, and the equilibrium point $\overline{x}$ of the network satisfies $\|\overline{x} - x^*\| \leq \varepsilon$; (b) Given an $\varepsilon > 0$, to find an $\widetilde{s}$ such that for $\overline{s} \geq \widetilde{s}$, $|f(\overline{x}) - f(x^*)| \leq \varepsilon$; (c) To find an $\widetilde{s}$ such that for $s \geq \widetilde{s}$,

$$J(\overline{x}) = I(x^*), \tag{24}$$

where

$$J(\overline{x}) = \{i: \ g_i(\overline{x}) = d^i \overline{x} - b_i > 0\}. \tag{25}$$

Criteria (a) and (b) are easy to understand. But for (c), a little discussion is needed. Recall the Lagrange function (2) and the equilibrium equation (16). Let $\overline{x} \equiv \overline{x}(s)$ be a solution of (16) for a given s and compare

$$\nabla F_1(\overline{x}, s) = \nabla f(\overline{x}) + \nabla g(\overline{x})\{s g^+(\overline{x})\} = 0 \tag{26}$$

with the gradient of $L(x, \lambda)$ at $\overline{x}$:

$$\nabla_x L(\overline{x}, \lambda) = \nabla f(\overline{x}) + \nabla g(\overline{x})\lambda. \tag{27}$$

If we set

$$\overline{\lambda}_i \equiv \overline{\lambda}_i(s, x) = s g_i^+(x) \geq 0, \qquad i = 1, \cdots, m, \tag{28}$$

then $\overline{x}$ is an equilibrium point of $L(x, \overline{\lambda}(s,x))$ which is the energy function $F_1(x,s)$. When $\overline{\lambda}(s,x) \geq O$, $L(x, \overline{\lambda}(s,x))$ is a convex function of x, so $\overline{x}$ is a minimum point of $L(x, \overline{\lambda}(s,x))$ but is not necessarily a solution of the original problem. This is because $\overline{\lambda}$ defined in (28) does not necessarily satisfy equation (6), i.e., for some i, $\overline{\lambda}(s,\overline{x})_i g_i(\overline{x}) = s(g_i^+(\overline{x}))2 \neq 0$. In other words, $\overline{\lambda}$ is not the optimal multiplier. But one can conceive that if $\overline{\lambda}$ is approaching the optimal multiplier λ^*, the solution $\overline{x}$ will be approaching the solution x^*. Equation (24) then is a condition to force $\overline{\lambda}$ to approach λ^* in the meaning that if $i \notin I(x^*)$, then $\overline{\lambda}_i g_i(\overline{x}) = 0$. Geometrically, (24) says that $\overline{x}$ only violates the constraints that are active at x^*.

Without loss of generality we suppose in the following discussion that all elements of D, b, a are integers. Let $\overline{d} = \max\{|d_{ij}|\}$, $\overline{b} = \max\{|b_i|\}$ and $\overline{a} = \max\{|a_i|\}$.

Lemma 2. If the linear programming problem is neither degenerate nor dual-degenerate at its optimal solution x^*, there exists a sufficiently large $\widetilde{s}_1$ such that for any $s \geq \widetilde{s}_1$, the unique equilibrium point $\overline{x}$ of the network is given by the system

$$s(d^i x - b_i) = \lambda_i^*, \qquad i \in I^* = I(x^*), \tag{29}$$

where λ_i^*'s are the optimal multipliers.

Proof. First it can be proved that there exists an $\overline{s}_1$ such that for $s \geq \overline{s}_1$ the equilibrium equation

$$a + \sum_{i \in I} s g_i^+(x) d^{iT} = 0 \tag{30}$$

has a unique solution. Since problem (7) has a unique solution x^*,

$$x^* = D_{I^*}^{-1} b_{I^*}, \tag{31}$$

where D_{I^*} is the optimal basis, an $n \times n$ sub-matrix of D corresponding to the index set I^*, $|I^*| = n$, $|D_{I^*}| \neq 0$. And its corresponding optimal multiplier satisfies one of the K-T conditions, (20), and $\lambda_i^* = 0$, if $i \notin I^*$. Now set

$$\overline{s}_2 = \max\left\{\frac{d^i D_{I^*}^{-1} \lambda_{I^*}^*}{b_i - d^i D_{I^*}^{-1} b_{I^*}},\ i \notin I^*;\ \ 1\right\}. \tag{32}$$

Then for an $s > \widetilde{s}_1 = \max\{\overline{s}_1, \overline{s}_2\}$, we assert that the solution of system (30),

$$\overline{x} = D_{I^*}^{-1}(\lambda_{I^*}^*/s + b_{I^*}) \tag{33}$$

is the solution of (21). Furthermore, $\overline{x}$, the solution of system (29), is the equilibrium point.

Lemma 3. Let a, h^i, $i \in I, \in R^n$, $\|h^i\| \geq 1$, $\{h^i,\ i \in I\}$ be an independent set and $a = \sum_{i \in I} \alpha_i h^i$. Then $|\alpha_i| \leq \|a\| \leq \sqrt{n}\,\overline{a}/2$, where $\overline{a} = \max\{|a_i|\}$.

The proof is simple, and hence is omitted. It is easy to see that the denominator, $b_i - d^i D_{I^*}^{-1} b_{I^*}$, of (32) is larger than zero. By integrality, the denominator is of value at least 1. The numerator can be estimated as follows: For $i \notin I^*$, let $d^i = \sum_{j \in I^*} \beta_j d^j = \beta^T D_{I^*}$. Using Lemma 3, we have

$$d^i D_{I^*}^{-1} \lambda_{I^*}^* = \beta^T D_{I^*} D_{I^*}^{-1} \lambda_{I^*}^* = \beta^T \lambda_{I^*}^* \leq n^2 \overline{a}\overline{d}/4. \tag{34}$$

Combining Lemma 2 and (34), we have the following theorem.

Theorem 4. With the same assumption of Lemma 2, if $s > \widetilde{s}_1$, the equilibrium point $\overline{x}$ of the network has property (24), i.e., $J(\overline{x}) = I(x^*)$. And $n^2\overline{a}\overline{d}/4$ is an estimate of $\widetilde{s}_1$, that is, the initial value of $\widetilde{s}_1$ could be taken as $n^2\overline{a}\overline{d}/4$.

The assumption of this theorem can be weakened. In fact, only the assumption of bounded optimal solution set $\overline{X}$ is needed to obtain a similar result. The extension induces more mathematical analysis than is described by Theorem A in Appendix, where a concrete estimation for $\widetilde{s}$ is no longer existing. Theorem 4 is a quantitative description of the criterion (c). If Criterion (a) is adopted, we have the following theorem.

Theorem 5. Under the same assumption of Theorem 4, for $s > \widetilde{s}_2 = n^2(m-1)!\,\overline{d}^{m-1}/\varepsilon$, the equilibrium point $\overline{x}$ has property

$$\|\overline{x} - x^*\| < \varepsilon. \tag{35}$$

Proof. If $s \geq \widetilde{s}_1$, according to (31) and (33),

$$\|\overline{x} - x^*\| = \|D_{I^*}^{-1}\lambda_{I^*}^*\|/s.$$

Note that each element of $D_{I^*}^{-1}$ is, by definition of the inverse, equal to an $(m-1)\times(m-1)$ determinant divided by a nonzero $m \times m$ determinant. By integrality, the denominator is of absolute value at least one. The determinant of the numerator is the sum of $(m-1)!$ elements of D. Therefore it has an absolute value not greater than $(m-1)!\,\overline{d}^{m-1}$. Using Lemma 3, we have

$$\|\overline{x} - x^*\| \leq n^2(m-1)!\,\overline{a}\overline{d}^{m-1}/s < \varepsilon;$$

then taking

$$\widetilde{s}_2 = n^2(m-1)!\,\overline{a}\overline{d}^{m-1}/\varepsilon \ (>> \overline{s}_2) \tag{36}$$

will satisfy the requirement of the theorem.

The estimation for $\widetilde{s}_2$ in (36) is not a tight one, and hence is not practical. In fact (36) is an upper bound for an estimation of the value of s such that the property (35) is valid. But when Criterion (b) is adopted, we would have a more practical estimation for $\widetilde{s}$.

Theorem 6. Under the same assumption of Theorem 4, for $s > \widetilde{s}_3$ where

$$\widetilde{s}_3 = (n\overline{a})^2/\varepsilon, \tag{37}$$

the equilibrium point $\overline{x}$ satisfies

$$|a^T\overline{x} - a^Tx^*| < \varepsilon. \tag{38}$$

Proof. According to (31), (33),

$$a^T(\overline{x} - x^*) = a^TD_{I^*}^{-1}\lambda_{I^*}^*/s = \lambda_{I^*}^*D_{I^*}D_{I^*}^{-1}\lambda_{I^*}^*/s = \|\lambda^*\|^2/s \leq n^2\overline{a}^2/s < \varepsilon,$$

which implies (37).

4. Quadratic Programming

Now we consider using the Kennedy and Chua model for solving problem (8) with a positive semi-definite Q. The corresponding equations (3)–(6) for problem (8) are

$$Qx + q + \sum_{i=1}^{m}\lambda_i d^{iT} = 0, \tag{39}$$

$$d^ix - b_i \leq 0, \qquad i = 1,\cdots,m, \tag{40}$$

$$\lambda_i \geq 0, \qquad i = 1,\cdots,m, \tag{41}$$

$$\lambda_i(d^ix - b_i) = 0, \qquad i = 1,\cdots,m. \tag{42}$$

Equation (39) and (42) can be combined into

$$Qx + q + \sum_{i \in I(x)} \lambda_i d^{iT} = 0, \tag{43}$$

where $I(x) = \{i : \ d^i x - b_i = 0\}$. The equilibrium equation (16) now is

$$-Qx - q = s \sum_{i \in J(x)} (d^i x - b_i) d^{iT},$$

where $J(x) = \{i : \ d^i x - b_i > 0\}$, or

$$\Big(Q + s \sum_{i \in J(x)} d^i d^{iT}\Big) x = s \sum_{i \in J(x)} b_i d^{iT} - q. \tag{44}$$

According to Remarks 1–3, we have

Theorem 7. If the feasible set X of problem (8) is bounded, then for an increasing sequence $\{s_k\}$, the network produces a sequence of approximate solutions x^k in the meaning that any limit point of it is a solution x^* of (8).

Furthermore, conclusions similar to Remarks 4 and 5 in Section 3 are still valid. The assumption of a bounded X can also be removed by the following lemma that is similar to Lemma 1.

Lemma 4. If problem (8)'s solution set $\overline{X}$ is bounded, then $F_2(x, s_k)$ in (10) satisfies Assumption 3, i.e., $F_2(x, s_k) \to \infty$ as $\|x\| \to \infty$.

See the proof of the lemma in Appendix.

Theorem 8. If the solution set $\overline{X}$ of problem (8) is bounded, then Theorem 7 and similar conclusions of Remarks 4, 5 are still true.

This theorem covers another main result, Theorem 4, in [8] where the regularity assumption is needed. For quadratic problems, there is no closed form for a parameter $\widetilde{s}$ as discussed in Theorems 4, 5 and 6 such that the equilibrium point given by the network with that determined $\widetilde{s}$ properly approximates the original solution.

5. Appendix

1) Proof of Corollary 1

Farkas' Theorem[14, p.46]. Let B be an $m \times n$ matrix and $q \in R^n$. Then exactly one of the following two inequality systems has a solution:

$$Bx \le 0, \qquad q^T x > 0, \qquad x \in R^n, \tag{a1}$$

$$B^T y = q, \qquad y \ge 0, \qquad y \in R^m. \tag{a2}$$

Proof of Corollary 1. Consider problem (7), let $\overline{x}$ be a local minimum solution. We assert that the system of variable z:

$$\nabla f(\overline{x})^T z < 0, \qquad d^i z \le 0, \qquad i \in I(\overline{x}) = \{i : \ d^i x - b_i = 0\} \tag{a3}$$

has no solution. Otherwise suppose that $\overline{z}$ is a solution; then there exists a sufficiently small constant $\overline{\delta} > 0$, such that for every $\delta \le \overline{\delta}$, $\overline{x} + \delta \overline{z}$ satisfies

$$d^i(\overline{x} + \delta \overline{z}) \le b_i \ \text{ for every } \ i \in I, \text{ i.e., } \ \overline{x} + \delta \overline{z} \in X.$$

But one finds that $f(\overline{x}+\delta\overline{z}) = f(\overline{x}) + \delta\nabla f(\overline{x})^T\overline{z} + o(\delta) < f(\overline{x})$ if δ is small sufficiently. This contradicts the assumption that $\overline{x}$ is a local minimum solution.

Now set $B = D_{I(\overline{x})}$, $q = -\nabla f(\overline{x})$, according to Farkas' Theorem, the system

$$D^T_{I(\overline{x})}\lambda u = -\nabla f(\overline{x}), \qquad \lambda \geq 0 \tag{a4}$$

has a solution. (a4) verifies equations (3) and (5). Equations (4) and (6) are obviously satisfied.

2) Proof of Lemma 1 and 4

We need only to prove Lemma 4, because Lemma 1 is a direct inference of Lemma 4.

Proof. Suppose the lemma is not true. Then there exists a sequence $\{x^k\}$ such that $\|x^k\| \to \infty$, but

$$x^{kT}Qx^k + q^Tx^k + s\sum_{i\in J}(d^ix^k - b_i)^2 \leq M, \tag{a5}$$

where M is a large positive number, $J = J(x^k) = \{i: \ d^ix^k - b_i > 0\}$. Without loss of generality J is assumed to be independent of k, otherwise, a subsequence of $\{x^k\}$ can be found to meet this requirement. Also we assume $\widetilde{x}^k = x^k/\|x^k\| \to \widetilde{x}$, $\|\widetilde{x}\| = 1$. (a5) can be written as

$$\widetilde{x^{kT}}Q\widetilde{x^k} + q^T\widetilde{x^k}/\|x^k\| + s\sum_{i\in J}\left(d^i\widetilde{x^k} - b_i/\|xk\|\right)^2 \leq M/\|x^k\|^2, \tag{a6}$$

which implies $d^i\widetilde{x} = 0$, $i \in J$ and $Q\widetilde{x} = 0$ by noticing that Q is a positive semi-definite matrix. For $i \notin J$, $d^ix^k \leq b_i$, then $d^i\widetilde{x} \leq 0$.

Now let $\overline{x} \in \overline{X}$ and $x^* = \overline{x} + \lambda\widetilde{x}$ for $\lambda > 0$; we have $d^ix^* = d^i\overline{x} + \lambda d^i\widetilde{x} \leq b_i$ for any $i \in I$, so $x^* \in X$ for any $\lambda > 0$. Furthermore, according to (43), one can find

$$\begin{aligned} x^{*T}Qx^* + q^Tx^* =& \overline{x}^TQ\overline{x} + q^T\overline{x} + \lambda q^T\widetilde{x} \\ =& \overline{x}^TQ\overline{x} + q^T\overline{x} + \lambda\sum_{i\in I(\overline{x})}\lambda_i d^i\widetilde{x} \\ \leq& \overline{x}^TQ\overline{x} + q^T\overline{x}. \end{aligned}$$

Hence x^* is an optimal solution for any $\lambda > 0$ which leads to a contradiction to the boundedness of the optimal solution set.

3) Extension of Theorem 4

Theorem A*. If the solution set $\overline{X}$ of problem (7) is bounded, and a sequence of equilibrium points x^k converges to a solution x^*, then there exists an $\widetilde{s}$ such that for $s_k > \widetilde{s}$, x^k satisfies $J(x^k) = I(x^*)$.

Proof. There are four cases: (i) The problem is both primal and dual non-degenerate, the result in this case has been proved in Theorem 4; (ii) The problem is not degenerate but is primal degenerate; (iii) The problem is degenerate but is not primal-degenerate; (iv) The problem is both primal and dual-degenerate.

Case (ii). If $\{x^k\} \to x^*$, x^* is a basic solution, then the proof is the same as that in Theorem 4; If x^* is not a basic solution, in this case we have a lemma that is similar to Lemma 2.

Lemma A1. In the case of (ii), for a sequence $\{x^k\} \to x^*$, x^* is not a basic solution, there exists a sufficiently large $\widetilde{k}$ such that for any $k \geq \widetilde{k}$, the equilibrium points x^k of the network with $s = s_k$ are given by the system

$$s_k(d^ix - b_i) = \lambda_i^*, \qquad i \in I^* = I(x^*), \tag{a7}$$

where λ_i^*'s are the optimal multipliers, $|I(x^*)| < n$.

Lemma A1 says that in case (ii) there is an infinite number of equilibrium points for a fixed s_k. To prove the lemma, note that there exists a $\widetilde{k}$ such that for $i \notin I(x^*)$, $k > \widetilde{k}$,

$$b_i - d^i x^k = b_i - d^i x^* + d^i(x^* - x^k) < 0. \tag{a8}$$

Then the solutions of (a7) are the solutions of the equilibrium equation (20) by the same reasoning as in Theorem 4. (a7) and (a8) imply that $J(x^k) = I(x^*)$.

Case (iii). In this case problem (7) has a unique solution and $|I(x^*)| > n$. By Corollary 1 in Section 2, x^* still satisfies the K-T condition (20). But now there are infinitely many sets of optimal multipliers λ_i^*'s and it is not necessary that for any set of λ_i^*'s, (a7) has a solution. So a result similar to Lemma 2 now is not available. But because the solution of (20) does exist for a sufficiently large $\widetilde{k}$ such that for $i \notin I(x^*)$ and $k > \widetilde{k}$, (a8) is valid, so we have $J(x^k) \subseteq I(x^*)$.

Case (iv). When $\{x^k\} \to x^*$, x^* is not a basic solution, then condition "the problem is primal degenerate" works. This sub-case can be reduced to Case (ii). When x^* is a basic solution, then condition "the problem is degenerate" works and the discussion can be reduced to that of Case (iii).

References

[1] J.J. Hopfield. Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons. *Proc. Natl. Acad. Sci. USA*, 1984, 81: 3088–3092.

[2] J.J. Hopfield and D.D. Tank. "Neural" Computation of Decisions in Optimization Problems. *Biological Cybernetics*, 1985, 52: 141–152.

[3] D.W. Tank and J.J. Hopfield. Simple Neural Optimization Networks: An A/D Converter, Signal Decision Network, and a Linear Programming Circuit. *IEEE Trans. Circuits Syst.*, 1986, CAS-33: 533–541.

[4] L.O. Chua and G.N. Lin. Nonlinear Programming without Computation. *IEEE Trans. Circuits Syst.*, 1984, CAS-31: 182–188.

[5] M.P. Kennedy and L.O. Chua. Neural Networks for Nonlinear Programming. *IEEE Trans. Circuits Syst.*, 1988, 35: 554–562.

[6] A. Rodriguez-Vazquez et al. Nonlinear Switched-capacitor Neural Networks for Optimization Problems. *IEEE Trans. Circuits Syst.*, 1990, 37: 384–398.

[7] C.Y. Maa and M.A. Shanblatt. A Constrained Optimization Neural Net Technique for Power System Analysis. Proc. IEEE Int. Symp. Circuits Syst., 2946-2950, 1990.

[8] C.Y. Maa and M.A. Shanblatt. Linear and Quadratic Programming Neural Network Analysis. *IEEE Trans. Neural Net.*, 1992, 3: 580–594.

[9] C.Y. Maa and M.A. Shanblatt. A Two-phase Optimization Neural Network. *IEEE Trans. Neural Net.*, 1992, 3: 1003–1009.

[10] A. Cichocki and R. Unbehauen. Switched-capacitor Neural Networks for Differential Optimization. *International J. of Circuit Theory and Applications*, 1991, 19: 161–187.

[11] Zhang Shengwei and A.G. Constantinides. Lagrange Programming Neural Networks. *IEEE Trans. on Neural Networks*, 1992, 36: 441–452.

[12] A. Bouzerdoum and T.R. Pattison. Neural Network for Quadratic Optimization with Bound Constraints. *IEEE Trans. Neural Net.*, 1993, 4: 293–304.

[13] Xiang-sun ZHANG and Hui-can ZHU. A Neural Network Model for Quadratic Programming with Simple Upper and Lower Bounds and Its Application to Linear Programming. In: Algorithms and Computation, Lecture Notes in Computer Science, 834: 119–127, Springer-Verlag, Berlin, 1994.

[14] M.S. Bazaraa, C.M. Shetty. Nonlinear Programming: Theory and Algorithms. John Wiley, Sons, Inc., New York, 1979.