Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Lagrange-Type Neural Networks for Nonlinear Programming Problems with Inequality Constraints

Yuancan Huang

Abstract— **By redefining multiplier associated with inequality constraint as a positive definite function of the originally-defined multiplier, $u_i^2, i = 1, 2, \cdots, m$, say, the nonnegative constraints imposed on inequality constraints in Karush-Kuhn-Tucker necessary conditions are removed completely. In the construction of Lagrange-type neural networks, it is no longer necessary to convert inequality constraints into equality constraints by slack variables in order to reuse those results concerned only with equality constraints. Utilizing this technique, a new Lagrange-type neural network is devised, which handles inequality constraints directly without adding slack variables. Finally, the local stability of the proposed Lagrange neural networks is analyzed rigourously with Liapunov's first approximation principle, and its convergence is discussed with LaSalle's invariance principle.**

Index Terms–Nonlinear Programming, Inequality constraint, Lagrange-Type Neural Network, Stability, Convergence.

I. Introduction

Since Tank and Hopfield first proposed a recurrent neural network for solving linear programming problems[1], a new research branch for optimization–neural computation is initiated and has received a great deal of attention in the last two decades, see[1]-[13] and references therein. The eminent merit of neural computation is able to compute the optimal solution during the dynamical transient motion toward an equilibrium point which coincides with such an optimal solution. The approach is especially desirable for all those on-line applications where computing the optimum in real time is of fundamental importance, as in some signal processing and robotic problems.

Kennedy and Chua developed the dynamical canonical nonlinear programming neural network[2]. Because the network of Kennedy and Chua contains a penalty parameter, it generates only the approximate solutions and encounters a circuit implementation difficulty when the penalty parameter approaches infinite. To avoid using the penalty parameter, some methods are proposed in recent years. For instance, Rodriguez-Vázquez et al. proposed a switched-capacitor neural network. Bouzerdoum and Pattison proposed a neural network for solving quadratic problems[4], which can achieve exponential stability through an appropriate choice of self-feedback and lateral connection matrices. It, however, solves the quadratic optimal problems with bounded constraints only. Xia and Wang developed several primal-dual neural networks for solving linear and quadratic programming problems and a neural network for solving linear projection equations[7]-[9]. All of these neural networks are proved to be globally asymptotically stable to exact solutions. They also introduced a neural network for solving the nonlinear projection formulation[10], and analyze its global convergence and stability[11]. On the basis of Lagrange theory, Zhang and Constantinides proposed the Lagrange programming neural networks[18]. But in their construction inequality constraints need to be converted into the equality ones by using slack variables. We proposed a new convex programming neural network with the quadratic multiplier strategy to circumvent this shortcoming and to facilitate the circuit implementation of Lagrange neural network[13]. In this paper, we use the same technique to build the Lagrange neural networks for general nonlinear programming problems and prove its local stability with Liapunov's first approximation principle as well as discuss its convergence with LaSalle's invariance principle.

The rest of this paper is organized as follows: Section 2 presents a complete description on the construction and mechanism of the Lagrange-type neural networks. In Section 3, The stability of Lagrange-type neural networks is proved rigorously by Liapunov's first approximation Principle, and its convergence is discussed based on LaSalle's invariance principle. Section 4 contains the conclusions and the discussions.

II. Lagrange-Type Neural Networks for Nonlinear Programming Problems with Inequality Constraints

Consider the nonlinear programming problem with only inequality constraints:

$$\begin{array}{ll} minimize & f(x) \\ subject\ to & g(x) \leq 0 \end{array} \tag{1}$$

where assume that $f(x) : R^n \to R$ and $g(x) : R^n \to R^m$ are twice continuously differentiable scalar function and vector function, respectively.

Definition 0.1: Let x^* be a vector satisfying the constraint conditions, then $I(x^*)$ denotes a set of index i for which $g_i(x^*) = 0$, namely

$$I(x^*) = \{i \mid g_i(x^*) = 0, i = 1, 2, \cdots, m\}. \tag{2}$$

If the gradients $\nabla g_i(x^*), i \in I(x^*)$ are linearly independent, then x^* is called regular point.

This work was supported by the Research Grants Committee of Beijing Institute of Technology

Y. Huang is with Department of Automatic Control, School of Information Science and Technology, Beijing Institute of Technology, Beijing, 100081, China yuancanhuang@bit.edu.cn

0-7803-9568-9/05/$20.00 ©2005 IEEE

If the Lagrangian function of problem is defined as

$$L(x,u) = f(x) + \sum_{i=1}^{m} u_i g_i(x), \tag{3}$$

there exists the following Karush-Kuhn-Tucker Theorem[14]-[17].

Theorem 0.2: Let x^* be a local minimum of problem (1) and assume that x^* is a regular point. Then there exists a unique vector u^* such that

$$\begin{aligned}
&\nabla_x L(x^*, u^*) = 0, && (4)\\
&u_i^* g_i(x^*) = 0, i = 1, 2, \cdots, m, && (5)\\
&g(x^*) \leq 0, &&\\
&u^* \geq 0. &&
\end{aligned}$$

From the above theorem, we observe that, if the multipliers are redefined as some positive function of original ones, $u_i^2, i = 1, 2, \cdots, m$, say, the nonnegative constraints imposed on the multipliers are removed completely. This implies that it is no longer necessary to convert inequality constraints into equality ones by slack variables in order to reuse those results concerned merely with equality constraints in the construction of Lagrange neural networks, and they are used directly[13]. The Lagrange-type neural network that is conceived with this method has two advantages over the original one: First, all the results on equality constraint can be transplanted in their original forms, and be proved similarly with minor modification; second, we may obtain the augmented Lagrangian functions with the same smoothness as that of objective function and constraints, therefore, they are more convenient to implement in circuits.

Define the augmented Lagrangian function as:

$$L_c(x,u) = f(x) + \sum_{i=1}^{m} u_i^2 g_i(x) + \frac{c}{2}\sum_{i=1}^{m}(u_i g_i(x))^2, \tag{6}$$

where c is a positive penalty parameter.

The aim is to construct a continuous-time dynamical system that will settle down to the KKT pair of nonlinear programming problem. Here is such a system:

$$\begin{aligned}
&\dot{x} = -\nabla_x L_c(x,u) && (7)\\
&\dot{u}_i = 2u_i g_i(x), i = 1, 2, \cdots, m. &&
\end{aligned}$$

From the system defined above, we may get u_i's analytical expression

$$u_i(t) = u_i(0)e^{\int_0^t 2g_i(x)dt}, i = 1, 2, \cdots, m, \tag{8}$$

where assume that $u_i(0)$ is a nonzero initial value of u_i.

This expression shows that, if x is outside the feasible region of problem (i.e. there is at least an i such that $g_i(x) > 0$) , then the corresponding multipliers will increase exponentially as time t increases. Thus the penalty terms containing those multipliers will increase continuously with time until the constraints are satisfied. This procedure will be repeated incessantly unless all the constraints are satisfied. Hence x will eventually be conducted into the feasible region whereas the multipliers associated with the inactive constraints approach to zero, and the remaining multipliers to constant.

III. Analysis of Stability and Convergence

Second-Order Sufficient Conditions 0.3: Let x^* be regular point for problem. If there exists vector u^* satisfying

$$\begin{aligned}
&\nabla_x L(x^*, u^*) = 0 && (9)\\
&u_i^* g_i(x^*) = 0, i = 1, 2, \cdots, m && (10)\\
&g(x^*) \leq 0 &&\\
&u^* \geq 0 &&
\end{aligned}$$

and for every $y \neq 0$ such that $\nabla g_i(x^*)^T y = 0$ for every $i \in I(x^*)$, it follows that

$$y^T[\nabla^2 f(x^*) + \sum_{i=1}^{m} u_i^* \nabla^2 g_i(x^*)]y > 0. \tag{11}$$

In addition, u^* satisfies the strict complementary assumption

$$u_i^* > 0, \forall i \in I(x^*), \tag{12}$$

then x^* is a strict local minimum of problem[14]-[17].

A lemma is introduced first[14], [19]:

Lemma 0.4: Let P be a symmetric $n \times n$ matrix and Q a positive semidefinite symmetric $n \times n$ matrix. Assume that $x^T P x > 0$ for every $x \neq 0$ satisfying $x^T Q x = 0$, then there exists a scalar $c > 0$ such that

$$P + cQ > 0. \tag{13}$$

By straightforward calculation, the gradient and Hessian matrix of the augmented Lagrangian function with respect to x are respectively given as

$$\begin{aligned}
\nabla_x L_c(x,u) &= \nabla f(x) + \sum_{i=1}^{m} u_i^2 \nabla g_i(x)\\
&+ c\sum_{i=1}^{m} u_i^2 g_i(x) \nabla g_i(x)\\
&= \nabla_x L(x,u) + c\sum_{i=1}^{m} u_i^2 g_i(x) \nabla g_i(x)
\end{aligned}$$

and

$$\begin{aligned}
\nabla_{xx}^2 L_c(x,u) &= \nabla_{xx}^2 L(x,u) + c\sum_{i=1}^{m} u_i^2 g_i(x) \nabla^2 g_i(x)\\
&+ c\sum_{i=1}^{m} u_i^2 \nabla g_i(x) \nabla g_i(x)^T.
\end{aligned}$$

Let x^* and u^* be the same as those in second-order sufficient conditions. There is, for any c,

$$\nabla_x L_c(x^*, u^*) = 0 \tag{14}$$

and there exists a $\bar{c}$ by Lemma 0.4 such that

$$\begin{aligned}
\nabla_{xx}^2 L_c(x^*, u^*) &= \nabla_{xx}^2 L(x^*, u^*) && (15)\\
&+ c\sum_{i=1}^{m} u_i^2 \nabla g_i(x^*) \nabla g_i(x^*)^T > 0. &&
\end{aligned}$$

Equations (14) and (15) show that (x^*, u^*) is a strict local minimum of the augmented Lagrangian function $L_c(x, u)$.

Proposition 0.5: Let (x^*, u^*) is the KKT pair for problem. If the second-order sufficient conditions are satisfied, then system (7) is locally asymptotically exponentially stable.

Proof: We linearize firstly the system at its equilibrium point (x^*, u^*). By the principle of stability in the first approximation, the local characteristic of the system in the proximity of equilibrium point is determined completely by its linearized system.

For sake of convenience, we assume without loss of generality that the preceding s inequality constraints are active and the corresponding multipliers are denoted by u_s; the remaining multipliers are inactive, denoted by u_t. Taking the KKT conditions into account, the linearized system is given as follows:

$$\begin{bmatrix} \dot{x} \\ \dot{u}_s \\ \dot{u}_t \end{bmatrix} = H \begin{bmatrix} x - x^* \\ u_s - u_s^* \\ u_t - u_t^* \end{bmatrix}, \tag{16}$$

with

$$H = \begin{bmatrix} -\nabla^2_{xx} L_c(x^*, u^*) & -\nabla g_s(x^*)\Gamma(2u_s^*) & 0 \\ \Gamma(2u_s^*)\nabla g_s(x^*)^T & 0 & 0 \\ 0 & 0 & \Gamma(2g_t(x^*)) \end{bmatrix}. \tag{17}$$

where $\Gamma(\cdot)$ represents the diagonalized matrix of vector.

Now we shall show that the real part of every eigenvalue of H is negative.

For any complex vector v, denotes by v^H its complex conjugate transpose, and for any complex number α, denotes by $\Re(\alpha)$ its real part. Let β be an eigenvalue of H, and nonzero vector $P = (z^T, w^T, y^T)^T$ be a corresponding eigenvector. We have

$$\Re(P^H H P) = \Re(\beta)(|z|^2 + |w|^2 + |y|^2). \tag{18}$$

Expanding the left-hand side of the above equation, we obtain

$$\begin{aligned} \Re(P^H H P) = \Re\{&-z^H \nabla^2_{xx} L_c(x^*, u^*)z + y^H \Gamma(2g_t(x^*))y \\ &-z^H \nabla g_s(x^*)\Gamma(2u_s^*)w + w^H \Gamma(2u_s^*)\nabla g_s(x^*)^T z\}. \end{aligned} \tag{19}$$

Since there is $\Re(z^H \nabla g_s(x^*)\Gamma(2u_s^*)w) = \Re(w^H \Gamma(2u_s^*)\nabla g_s(x^*)^T z)$, it follows from Eq. (18) and Eq.(19) that

$$\begin{aligned} &\Re(\beta)(|z|^2 + |w|^2 + |y|^2) = \\ &\Re[-z^H \nabla^2_{xx} L_c(x^*, u^*)z + y^H \Gamma(2g_t(x^*))y] \leq 0. \end{aligned} \tag{20}$$

Then we derive that either $\Re(\beta) < 0$ or $z = 0, y = 0$. However, if $z = 0, y = 0$, the following equation

$$H \begin{bmatrix} z \\ w \\ y \end{bmatrix} = \beta \begin{bmatrix} z \\ w \\ y \end{bmatrix} \tag{21}$$

yields

$$-\nabla g_s \Gamma(2u_s^*)w = 0. \tag{22}$$

From the hypothesis, $\nabla g_s(x^*)$ has full row rank and $u_s^* \neq 0$, it follows that $w = 0$. This contradicts our earlier assumption that P is a nonzero vector. Consequently, we must have $\Re(\beta) < 0$. Thus (x^*, u^*) is the asymptotically exponentially stable point of system (7). ■

The results presented above are only concerned on the local stability. From the viewpoint of circuit implementation, we hope that the Lagrangian neural networks constructed are stable in the large or have larger attractive domain. In the sequel, we will discuss how to relax the stability conditions and to enlarge the attractive domain on the basis of LaSalle's invariance principle[20]-[22].

Consider the autonomous systems described by the equations

$$\dot{x} = f(x), \quad f : D \to R^n, \tag{23}$$

where D is an open and connected subset of R^n and f is a locally Lipschitz map from D into R^n.

Definition 0.6: A set M is said to be invariant set with respect to the dynamical systems $\dot{x} = f(x)$ if

$$x(0) \in M \Longrightarrow x(t) \in M, \forall t \in R^+.$$

In other words, M is the set of points such that a solution of $\dot{x} = f(x)$ belongs to M at some time,initialized at $t = 0$,then it belongs to M for all $t \geq 0$.

LaSalle's Invariance Principle 0.7: Let $V : D \to R$ be a continuously differentiable function and assume that

1) $M \subset D$ is a compact set, invariant with respect to the solution of system (23).
2) $\dot{V} \leq 0$ in M.
3) $E = \{x \in M | \dot{V} = 0\}$; that is, E is the set of all points of M such that $\dot{V} = 0$.
4) N is the largest invariant set in E.

Then every solution starting in M approaches N as $t \to \infty$.

If the augmented Lagrangian function is taken as the V-function in LaSalle's Invariance Principle, in order to make the system stable, the two following conditions should be satisfied : (i) the solution of system (7) lies in a bounded set M , and (ii) $\dot{L}_c \leq 0$ in M. Firstly, we remark on the first condition. Obviously, if the solution of system (7) is limited in some bounded set, this condition is true. In general case, the limitation is not very stringent, for we are only interested in the bounded solution of problem, and, in many cases, it is satisfied automatically, i.e., for some initial conditions $x(0), u(0)$, the solution trajectories of system (7) lie in some bounded set themselves without any limitation.

Now we begin to discuss the second condition. Differentiating the augmented Lagrangian function with respect to t, we have

$$\begin{aligned} \dot{L}_c &= \left(\frac{\partial L_c}{\partial x}\right)^T \dot{x} + \left(\frac{\partial L_c}{\partial u}\right)^T \dot{u} \\ &= -\parallel \frac{\partial L_c}{\partial x} \parallel^2 + 2\sum_{i=1}^{m} u_i^2 g_i^2(x)[2 + c g_i(x)]. \end{aligned} \tag{24}$$

In the above equation the first term is always negative and is positive factor for the stability of the system. The effect of each term in the second sum may be classified into three categories: (i) If $g_i(x) > 0$, the corresponding term in the sum is positive. It leads the trajectory leave from the equilibria, and thus it plays a negative role on the stability of the system. On the other hand, it may result in the increase of penalty and force x approach to the feasible set of problem. (ii) If $g_i(x) = 0$, the corresponding term is zero, no effect on the stability of the system. (iii) If $g_i(x) < 0$ and is a larger negative number, in this case, we may select an appropriate parameter $c > 0$ such that the corresponding term is negative, and then it plays a positive role on the stability of the system. Otherwise, if $g_i(x)$ is a very small negative number, the corresponding term will become positive. Fortunately, the effect of this term on the stability of the system can be negligible because its magnitude is very small. Note that if the solution of system (7) is confined to some bounded set, probably $\dot{x} \neq -\nabla_x L_c(x, u)$ in the boundary. This may lead to the first term in Eq.(24) nonnegative, and thus deteriorate the stability of the system.

To summarize, the stability of system (7) is a consequence of the comprehensive action of versatile factors. If the solution of system (7) is constrained in some bounded set, and the time derivative of the augmented Lagrangian function is less than or equal to zero in the set, or this is true after some time, then all the trajectories of the system starting from the bounded set are convergent to the largest invariant set of system (7) by LaSalle's Invariance Principle.

Proposition 0.8: Assume that the conditions in LaSalle's invariance principle are satisfied. If $\dot{x}(t)$ is equal to the negative gradient direction of the augmented Lagrangian function for all t, i.e., $\dot{x} = -\nabla_x L_c(x, u)$, then the solutions of system (7) with initial conditions in M converge to the KKT pairs of problem.

Proof: We first show that every limit point x^* of $x(t)$ lies in the feasible set. To set up a contradiction, assume that x^* lies outside the feasible set. Then there exists at least some index i such that $g_i(x^*) > 0$. From the continuity assumption of $g_i(x)$, we can derive that the corresponding multiplier will tend to infinity. This contradicts to the boundedness of multipliers. Therefore, x^* must lie in the feasible set.

It is now shown that the solutions of system (7) approach to the KKT pairs of problem. Since $g(x^*) \leq 0$, we can choose an appropriate c such that the two terms in Eq.(24) are nonnegative. It follows that $\partial L_c(x^*, u^*)/\partial x = 0$ and $u_i^* g_i(x^*) = 0$. This concludes that (x^*, u^*) is the KKT pair of problem. ■

Assume that x and the multipliers u belong to a bounded set Ω. A projected dynamical system is constructed as follows:

$$\begin{aligned} \dot{x} &= P_\Omega(x - \alpha \nabla_x L_c(x, u)) - x, \ \ x \in R^n \\ \dot{u}_i &= P_\Omega(u_i + 2u_i g_i(x)) - u_i, i = 1, 2, \cdots, m. \end{aligned} \tag{25}$$

According to [23], we have

Proposition 0.9: There exists a unique continuous solution trajectory $(x(t), u(t))$ for the system (25) with the initial conditions $(x(t_0), u(t_0))$. Moreover, its solution $(x(t), u(t))$ will approach exponentially the bounded set Ω when the initial point $(x(t_0), u(t_0)) \ni \Omega$, and $(x(t), u(t)) \in \Omega$ when $(x(t_0), u(t_0)) \in \Omega$.

From the last two propositions, if we can reconstruct the augmented Lagrangian function using other type of penalty function and/or can design a regulating rule of the penalty parameter such that the augmented Lagrangian function is nonincreasing along the solution of the system (25), and the largest invariant set with the respect to the dynamical system (25) is just the KKT pair of problem (1), we conclude that the system is globally convergent to the KKT pair of problem (1).

IV. Conclusions and Discussions

Through redefining Lagrange multipliers as quadratic function of original ones, we create a novel method to deal with inequality constraints in Lagrangian neural networks. It is no longer necessary to convert inequalities into equalities with slack variables in order to reuse the results concerned only with equality constraints. Compared with the existing methods, the most remarkable feature of the new technique is that the great majority of results concerned with equality constraints may be transplanted in their original forms and be proved similarly with minor modification. Moreover, the derived augmented Lagrangian function has the same smoothness to objective function and constraints, they are, therefore, more convenient to be implemented in circuitry. Using Liapunov's first approximation principle, we prove that the dynamical system constructed is locally asymptotically exponentially stable. We also discuss its convergence by LaSalle's invariance principle. Our analysis shows that the stability of Lagrangian neural networks is a consequence of the common effect of stable and unstable factors. In order to obtain the optimal solution of problem, two factors are both indispensable. We must, however, design the system to guarantee that the stable factors are dominant ultimately to make the system convergent. For satisfying the boundedness assumption in LaSalle's invariance principle, a projected dynamical system is constructed. Furthermore, if we take the penalty parameter c as control variable and design a regulating rule such that the the augmented Lagrangian function is nonincreasing along the solution of system (25), then the the system is globally convergent to the KKT pair of problem (1). This statement illustrates that it is possible to use the control theory for conceiving an appropriate regulating rule of penalty parameter c such that the convergence of the dynamical system constructed is guaranteed. This will be our future research topic.

References

[1] D.W. Tank and J.J. Hopfield, Simple 'neural' optimization network: an A/D converter, signal decision circuit, and a linear programming circuit, *IEEE Transaction on Circuits and Systems*, vol.33, pp.533-541, 1986.

[2] M.P. Kennedy and L.O. Chuo, Neural networks for nonlinear programming, *IEEE Transaction on Circuits and Systems*, vol.35, pp.554-562, 1988.

[3] A. Rodriguez-Vázquez, R.Domínguez-Castro, A. Ruedam J.L. Huertas, and E. Sánchez-Sinencio, Nonlinear switched-capacitor "neural" networks for optimization problems, *IEEE Transaction on Circuits and Systems*, vol.37, pp.384-398, 1990.

[4] A. Bouzerdoum and T.R. Pattison, Neural network for quadratic optimization with bound constraints, *IEEE Transaction on Neural Networks*, vol.4, pp.293-304, 1993.
[5] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Singal Processing*, Chichester, U.K.: Wiley, 1993.
[6] E.K.P. Chong, S. Hui, and S.H. Żak, An analysis of a class of neural networks for solving linear programming problems, *IEEE Transaction on Automatic Control*, vol.44, pp.1095-2006, 1999.
[7] Y. Xia and J. Wang, Neuarl network for solving linear problems with bounded variables, *IEEE Transaction on Neural Networks*, vol.6, pp.515-519, 1995.
[8] Y. Xia, A new neuarl network for solving linear and quadratic programming problems, *IEEE Transaction on Neural Networks*, vol.7, pp.1544-1547, 1996.
[9] Y. Xia and J. Wang, A recurrent neuarl network for solving linear projection equations, *Neural Networks*, vol.13, pp.337-350, 2000.
[10] Y. Xia, H. Leung, and J. Wang, A projection neuarl network and its application to constrained optimization problems, *IEEE Transaction on Circuits and Systems*, vol.49, pp.447-458, 2002.
[11] Y. Xia, Further results on global convergence and stability of gobally projected dynamical systems, *Journal of Optimization Theory and Applications*, vol.122, pp.627-649, 2004.
[12] S. Zhang and A.G. Constantinides, Lagrange Programming Neural Networks, *IEEE Transaction on Circuits and Systems*, vol.35, pp.554-562, 1988.
[13] Y.C. Huang, A Novel Method to Handle Inequality Constraints for Convex Programming Neural Network, *Neural Processing Letters*, vol.16, pp.17-27, 20002.
[14] D. P. Bertsekas: *Constrainted Optimization and Lagrange Methods*. New York: Academic Press, 1982.
[15] M. Avriel: *Nonlinear Programming Analysis and Methods*. Englewood Cliffs, New Jersey: Prentice-Hall, 1976.
[16] R. T. Rockafellar: *Lagrange Multiplier and Optimality*. SIAM Review, 35(1993):183-238.
[17] D. P. Bertsekas: *Nonlinear Programming*, 2nd Edition. Belmont, Massachusetts: Athena Scientific, 1999.
[18] X.S. Zhang: *Neural Networks in Optimization*. Norwell: Kluwer Academic Publishers, 2000.
[19] J. Nocedal and S. J. Wright: *Numerical Optimization*. New York: Springer-Verlag, 1999.
[20] V. Lakshmikantham, V.M. Matrosov and S. Sivasundaram: *Vector Lyapunov Functions and Stability Analysis of Nonlinear Systems*. Dordrecht: Kluwer Academic Publisher, 1991.
[21] J. P. LaSalle: *The stability of dynamical systems* . New York: Springer, 1976.
[22] H.J. Marquez: *Nonlinear Control Systems-Analysis and Design*. New Jersey: John Eiley & Sons, 2003.
[23] Y. Xia and J. Wang, On the stability of globally projected dynamical systems, *Journal of Optimization Theory and Applications*, vol.106, pp.129-150, 2000.