Stability of Coupled Local Minimizers Within the Lagrange Programming Network Framework

Xuyang Lou and Johan A. K. Suykens, *Senior Member, IEEE*

***Abstract*—Coupled local minimizers (CLMs) turn out to be a potential global optimization strategy to explore a search space, avoid overfitting and produce good generalization. In this paper, convergence properties of CLMs based on an augmented Lagrangian function in the context of equality constrained minimization, are studied. We first consider the augmented Lagrangian by taking the objective of minimizing the average cost of an ensemble of local minimizers subject to pairwise synchronization constraints. Then we study an array of CLMs within the Lagrange programming network framework and analyze the local stability of CLMs using a linearization strategy. We further show that, under some mild conditions, global asymptotical stability of the unique equilibrium point of the network can be guaranteed. Afterwards, some sufficient conditions are presented to ensure the stability of synchronization between any two minimizers via a directed graph method. The results show that the CLMs usually can be synchronized if the penalty factors in the array of CLMs are chosen large enough. It is worth pointing out that CLMs possess the capability of global exploration in the search space and the advantage of faster running time on convex problems in comparison with most of the neural network approaches, which is also illustrated through two test functions and their numerical simulations.**

***Index Terms*—Coupled local minimizers, Lagrange programming network, augmented Lagrangian, stability, synchronization.**

I. Introduction

OPTIMIZATION problems are abound in many fields of engineering, biology, physics, chemistry and economics. Many of them are related to the minimization of a cost function with several local optima. There are many well-developed methods for local optimization problems, such as steepest descent, Newton's, conjugate gradient, quasi-Newton, Levenberg-Marquardt and sequential quadratic programming [1]–[3].

For constrained optimization problems, the augmented Lagrangian method, which combines both the Lagrange and the penalty methods, has been serving as a fundamental solution methodology and effectively overcome the zigzagging problems or the possible infinity problems of penalty terms associated with the Lagrange method or the penalty method when used alone. Recently, research on convergence properties of augmented Lagrangian methods also has received a lot of attention. Nguyen and Strodiot [4] proved the local convergence of a modified exponential Lagrangian method. Luo *et al.* [5] presented new convergence properties of the primal-dual method based on four types of augmented Lagrangian functions in the context of constrained global optimization. Li [6] proposed an augmented Lagrange-Hopfield method based on the augmented Lagrange method and improved Hopfield-type neural network method in both the convergence and the solution quality in solving combinatorial optimization. Huang and Yang [7] showed that the first-order and second-order necessary optimality conditions of Rockafellar's augmented Lagrangian problems converge to those of the original problem. Hager [8] and Yamashita [9] investigated the global convergence of Rockafellar's augmented Lagrangian methods for nonconvex inequality-constrained problems.

Due to the parallel computational capacity and facilitation through electronic hardware implementation of recurrent neural networks, many Hopfield-type recurrent neural network models have been proposed to resolve various kinds of optimization problems since Tank and Hopfield's seminal work [10], such as linear variational inequalities [11]–[13], quadratic programming problem [14], [15], and nonlinear convex programming [16]–[20].

Among them, Zhang and Constantinides [17] proposed a recurrent neural network based on an augmented Lagrangian function for solving nonlinear optimization problems with equality constraints and examined its local asymptotical stability at Karush-Kuhn-Tucker (KKT) points that correspond to local optima under mild conditions. Xia [19] developed a Lagrange network for nonlinear convex programming problem with linear equality constraints and proved the global convergence of the Lagrange programming networks based on a basic Lyapunov function. In [15], Yang and Cao designed a neural-network-based solution to the quadratic programming problem with equality constraints. Afterwards, aiming at the improper definition of projection operator in [15], Cheng *et al.* [11] presented an improved delayed projection neural network for solving a class of linear variational inequalities, where the monotonicity assumption on the linear variational inequality

Manuscript received November 29, 2011; revised March 02, 2012; accepted April 02, 2012. Date of publication October 05, 2012; date of current version January 24, 2013. This work was supported by Research Council KUL: GOA/11/05 Ambiorics, GOA/10/09 MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM, several Ph.D./postdoc and fellow grants; Flemish Government: FWO: Ph.D./postdoc grants, projects: G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC), IWT: Ph.D. Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007–2011); EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940); Contract Research: AMINAL; Other: Helmholtz: viCERP, ACCM. This work was also supported in part by the National Natural Science Foundation of China (61174021, 61104155) and the 111 Project (B12018). This paper was approved by Associate Editor Xiang Li.

X. Y. Lou is with the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China and also with the School of IoT Engineering, Jiangnan University, Wuxi 214122, China (e-mail: Xuyang.Lou@gmail.com).

J. A. K. Suykens is with the Department of Electrical Engineering (ESAT-SCD/SISTA), Katholieke Universiteit Leuven, 3001 Leuven, Belgium (e-mail: Johan.Suykens@esat.kuleuven.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSI.2012.2215782

is removed. Recently, the authors in [21] introduced a simple neural network for a class of variational inequality problems and provided several sufficient conditions to ensure its asymptotical stability without estimating the Lipschitz constant which was required in [18]. To solve the nonlinear convex programming problem with linear and nonlinear constraints, Gao proposed a new neural network in [22], however, its stability required the initial point lying in a convex set. To overcome this drawback, Gao *et al.* presented a new neural network model for solving constrained variational inequality problems by converting the necessary and sufficient conditions for the solution into a system of nonlinear projection equations [12]. Although the achievement is rich and application of neural networks in the field of optimization problems has proliferated in recent years, most of these results focus on convergence of the methods without considering the capability of global optimization.

It is well known, however that there are many optimization problems requiring global exploration of the search space. For example, global minimum energy conformations of a molecule can have a dramatic effect on its activity [23]. The development of the global optimization method, which reaches global minima without being trapped at local minima, has been investigated extensively. Among popular methods for global exploration of the search space, meta-heuristics, in which heuristics are combined based on a very good search strategy, called diversification and intensification [24], [25], have received a great deal of attention. Most meta-heuristics are multipoint optimization methods, which use coupled multiple search points moving stochastically. Examples of these methods include the genetic algorithm [26], and particle swarm optimization [27]. Nevertheless, in these methods, interaction among all search points is mainly used as the driven force, and they require a large number of function evaluations since they are based on probabilistic searching without the use of any gradient information. Therefore, both of them share the disadvantage of losing diversity once all of the search points are attracted to one search point.

In 2001, Suykens *et al.* proposed a new method of coupled local minimizers (CLMs) [28], which is achieved by minimizing the average energy cost of the ensemble, subject to synchronization constraints between the state vectors of the individual local minimizers. This method can be formulated as a set of coupled Lagrange programming networks whose states exchange information via the coupling. It has been tested with the optimization of Lennard-Jones clusters and supervised training of neural networks. Recently, the method has been successfully applied to finite element model updating using experimental modal data [29], and shown that the global minimum is expected to be found more easily because of the simultaneous searching and cooperative behavior among multiple points. For more details about the underlying principles of this method, we refer the readers to [30]–[33]. Although some progress has been made in global optimization problems by using CLMs, the convergence of this method has not yet been established. We address the issue in this paper by providing a theoretical analysis for the method. To achieve this goal, a Lagrangian programming network is firstly constructed based on an augmented Lagrangian taking the objective of minimizing the average cost of an ensemble of local minimizers subject to pairwise synchronization constraints. Then, local stability analysis of the network is carried out through linearization techniques and eigenvalue analysis. Subsequently, we show that the global convergence to a unique optimal solution can be achieved under a mild condition. Finally, two test functions and their numerical simulations are provided to illustrate the effectiveness of the proposed method.

Throughout this paper, we use the following notations. $\mathbb{R}^n$ denotes the n-dimensional real space. $\mathbb{N}_q \triangleq \{1, 2, \ldots, q\}$. $\mathbb{1}_n = [1, 1, \ldots, 1]^\top$. $\mathbf{z}^\top$ ($A^\top$, respectively) is the transpose of vector $\mathbf{z} \in \mathbb{R}^n$ (matrix $A \in \mathbb{R}^{n\times m}$, respectively). $P > 0$ represents $P \in \mathbb{R}^{n\times n}$ is a positive definite matrix. $\mathbf{I}_n$ represents the n-dimensional identity matrix. $\|\mathbf{z}\|$ is the Euclidean norm of a vector $\mathbf{z}$ and $\|A\|$ denotes the spectral norm of a matrix A. For two vectors $\mathbf{z} \in \mathbb{R}^n$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$, $\langle \mathbf{z}, \boldsymbol{\lambda} \rangle = \mathbf{z}^\top \boldsymbol{\lambda}$. $\mathrm{diag}(\,\cdot\,)$ is used for represent a block diagonal matrix. $\nabla_{\mathbf{z}} f(\mathbf{z})$ and $\nabla^2_{\mathbf{z}} f(\mathbf{z})$ denote the gradient and Hessian of function f at $\mathbf{z}$, respectively. $\mathcal{C}^k$ is the set of k times continuously differentiable functions. "$\otimes$" denotes the Kronecker product.

The paper is organized as follows. In Section II, we recall the mechanism of CLMs and construct the array of CLMs within the Lagrange programming network framework. In Section III, we present a basic sufficient condition for optimality of our problem. In Section VI, we come up with some sufficient conditions for local and global stability of CLMs. In Section V, we analyze the synchronization of CLM ensemble by means of a directed graph method. We give an extension of our main results to the nonlinear optimization problem with equality constraints in Section VI. In Section VII, two test functions are presented to demonstrate the performance of the results. Finally, in Section VIII, we draw conclusions about the main contributions.

II. Coupled Local Minimizers (CLMs)

Consider the following unconstrained problem of minimization of a twice continuously differentiable cost function

$$\min_{\mathbf{z}\in\mathbb{R}^n} U(\mathbf{z}). \tag{1}$$

By the steepest descent method, a simple continuous time local optimization algorithm can be carried out for this problem: $\dot{\mathbf{z}} = -\nu\nabla_{\mathbf{z}} U(\mathbf{z})$ with $\nu > 0$ the step size. Here, alternatively, using the CLM scheme in [28], we aim at minimizing the average energy cost $(1)/(q)\sum_{i=1}^{q} U[\mathbf{z}^{(i)}]$ of an ensemble consisting of q local minimizers, subject to pairwise state equality constraints:

$$\begin{aligned} \min_{\mathbf{z}^{(i)}\in\mathbb{R}^n} \quad & \frac{1}{q}\sum_{i=1}^{q} U\left[\mathbf{z}^{(i)}\right] \\ \text{subject to} \quad & \mathbf{z}^{(i)} - \mathbf{z}^{(i+1)} = 0, \quad i \in \mathbb{N}_{q-1}, \end{aligned} \tag{2}$$

with states $\mathbf{z}^{(i)} \in \mathbb{R}^n$. For the $q-1$ equality constraints in (2), we introduce $q-1$ Lagrange multipliers $\boldsymbol{\lambda}^{(i)} \in \mathbb{R}^n, i \in \mathbb{N}_{q-1}$ and give the augmented Lagrangian:

$$\begin{aligned} &\mathcal{L}\left(\mathbf{z}^{(i)}, \boldsymbol{\lambda}^{(i)}\right) \\ &\quad= \frac{\eta}{q}\sum_{i=1}^{q} U\left[\mathbf{z}^{(i)}\right] + \frac{1}{2}\sum_{i=1}^{q-1}\gamma_i \left\|\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right\|^2 \\ &\qquad+ \sum_{i=1}^{q-1}\left\langle \boldsymbol{\lambda}^i, \left[\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right]\right\rangle, \end{aligned} \tag{3}$$

where $\eta > 0$ represents the learning rate and $\gamma_i > 0$ ($i \in \mathbb{N}_{q-1}$) denote the penalty factors emphasizing the importance of each of the soft synchronization constraints. For further analysis convenience, let us rewrite (3) into a compact vector form:

$$\mathcal{L}(\mathbf{z},\boldsymbol{\lambda}) = \frac{\eta}{q}U^\top[\mathbf{z}]\mathbb{1}_q + \frac{1}{2}\mathbf{z}^\top\bar{G}_{\gamma_0}^\top\bar{G}_{\gamma_0}\mathbf{z} + \boldsymbol{\lambda}^\top\bar{G}_0^\top\mathbf{z}, \quad (4)$$

where

$$U(\mathbf{z}) = \begin{bmatrix} U\left[\mathbf{z}^{(1)}\right] \\ \vdots \\ U\left[\mathbf{z}^{(q)}\right]\end{bmatrix} \in \mathbb{R}^q, \quad \mathbf{z} = \begin{bmatrix}\mathbf{z}^{(1)} \\ \vdots \\ \mathbf{z}^{(q)}\end{bmatrix} \in \mathbb{R}^{nq},$$

$$\boldsymbol{\lambda} = \begin{bmatrix}\boldsymbol{\lambda}^{(1)} \\ \vdots \\ \boldsymbol{\lambda}^{(q-1)}\end{bmatrix} \in \mathbb{R}^{n(q-1)},$$

$$G_0 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{q\times(q-1)},$$

$$\bar{G}_{\gamma_0} = G_{\gamma_0}\otimes\mathbf{I}_n \in \mathbb{R}^{nq\times nq},$$
$$\bar{G}_0 = G_0\otimes\mathbf{I}_n \in \mathbb{R}^{nq\times n(q-1)},$$

G_{γ_0} is shown at the bottom of the page.

From this augmented Lagrangian (4), we can derive the following Lagrange programming network [17]:

$$\begin{cases} \dot{\mathbf{z}}^{(i)} = -\nabla_{\mathbf{z}^{(i)}}\mathcal{L}\left(\mathbf{z}^{(i)},\boldsymbol{\lambda}^{(i)}\right), & i\in\mathbb{N}_q, \\ \dot{\boldsymbol{\lambda}}^{(i)} = \nabla_{\boldsymbol{\lambda}^{(i)}}\mathcal{L}\left(\mathbf{z}^{(i)},\boldsymbol{\lambda}^{(i)}\right), & i\in\mathbb{N}_{q-1}, \end{cases} \quad (5)$$

By substituting (4) into (5), one can obtain the following array of CLMs in vector form:

$$\begin{cases} \dot{\mathbf{z}} = -\left(\frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma\mathbf{z} + \bar{G}_1\boldsymbol{\lambda}\right), \\ \dot{\boldsymbol{\lambda}} = \bar{G}_1^\top\mathbf{z}, \end{cases} \quad (6)$$

where

$$\nabla_{\mathbf{z}}U(\mathbf{z}) = \begin{bmatrix}\nabla_{\mathbf{z}^{(1)}}U\left[\mathbf{z}^{(1)}\right] \\ \vdots \\ \nabla_{\mathbf{z}^{(q)}}U\left[\mathbf{z}^{(q)}\right]\end{bmatrix} \in \mathbb{R}^{nq},$$
$$\bar{G}_\gamma = G_\gamma\otimes\mathbf{I}_n \in \mathbb{R}^{nq\times nq},$$
$$\bar{G}_1 = G_1\otimes\mathbf{I}_n \in \mathbb{R}^{nq\times n(q-1)},$$

G_γ is shown at the bottom of the page, and

$$G_1 = G_0 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & -1 \end{bmatrix} \in \mathbb{R}^{q\times(q-1)}.$$

Remark 1: It is worth mentioning that $\bar{G}_1^\top = G_1^\top\otimes\mathbf{I}_n$ has full column rank $n(q-1)$, and $\bar{G}_\gamma$ is symmetric and all of its eigenvalues are positive except that one of them equals zero. Hence, $\bar{G}_\gamma \geq 0$.

III. Sufficient Conditions for Optimality

Before giving our main results, we introduce the regularity condition and second order sufficient conditions for optimality, as a preparation for theoretical analysis.

Consider the following general nonlinear programming problem with equality constraints:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{subject to} \quad & h_i(\mathbf{x}) = 0\ i\in\mathbb{N}_m, \end{aligned} \quad (7)$$

where $f:\mathbb{R}^n\to\mathbb{R}$, $h_i:\mathbb{R}^n\to\mathbb{R}$, $i\in\mathbb{N}_m$, are twice continuously differentiable functions. Define a Lagrangian $\mathcal{L}_o : \mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ as

$$\mathcal{L}_o(\mathbf{x},\mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\top h(\mathbf{x}), \quad (8)$$

$$G_{\gamma_0} = \begin{bmatrix} \sqrt{\gamma_1} & -\sqrt{\gamma_1} & 0 & 0 & \cdots & 0 & 0 \\ -\sqrt{\gamma_1} & \sqrt{\gamma_1}+\sqrt{\gamma_2} & -\sqrt{\gamma_2} & 0 & \cdots & 0 & 0 \\ 0 & -\sqrt{\gamma_2} & \sqrt{\gamma_2}+\sqrt{\gamma_3} & -\sqrt{\gamma_3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -\sqrt{\gamma_{q-2}} & \sqrt{\gamma_{q-2}}+\sqrt{\gamma_{q-1}} & -\sqrt{\gamma_{q-1}} \\ 0 & 0 & \cdots & 0 & 0 & -\sqrt{\gamma_{q-1}} & \sqrt{\gamma_{q-1}} \end{bmatrix} \in \mathbb{R}^{q\times q}.$$

$$G_\gamma = \begin{bmatrix} \gamma_1 & -\gamma_1 & 0 & 0 & \cdots & 0 & 0 \\ -\gamma_1 & \gamma_1+\gamma_2 & -\gamma_2 & 0 & \cdots & 0 & 0 \\ 0 & -\gamma_2 & \gamma_2+\gamma_3 & -\gamma_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -\gamma_{q-2} & \gamma_{q-2}+\gamma_{q-1} & -\gamma_{q-1} \\ 0 & 0 & \cdots & 0 & 0 & -\gamma_{q-1} & \gamma_{q-1} \end{bmatrix} \in \mathbb{R}^{q\times q}$$

where $h(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_m(\mathbf{x})]^\top$ and $\mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_m]^\top$. Then we have the following classical optimization results.

Definition 1 [1]: Let $\mathbf{x}^*$ be a vector such that $h(\mathbf{x}^*) = 0$. We say that $\mathbf{x}^*$ is a regular point if the gradients $\nabla_{\mathbf{x}} h_1(\mathbf{x}^*), \ldots, \nabla_{\mathbf{x}} h_m(\mathbf{x}^*)$ are linearly independent.

Definition 2 [35]: A point $\mathbf{x}^*$ is said to be a strict minimum of the problem in (7) if $f(\mathbf{x}^*) < f(\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{N}(\mathbf{x}^*; \epsilon) \cap \mathcal{S}$, where $\mathcal{N}(\mathbf{x}^*; \epsilon)$ is a neighborhood of $\mathbf{x}^*$ with the radius $\epsilon > 0$ and $\mathcal{S}$ is the feasible region of the problem.

Definition 3 [19]: A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotonically increasing if, for each pair of points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\langle F(\mathbf{x}) - F(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0$$

where $\langle \cdot, \cdot \rangle$ denotes an inner product; F is said to be strictly monotonically increasing if the above strict inequality holds whenever $\mathbf{x} \neq \mathbf{y}$.

Proposition 1 [2]: Assume that f and h are twice continuously differentiable, and let $\mathbf{x}^* \in \mathbb{R}^n$ and $\mathbf{y}^* \in \mathbb{R}^m$ satisfy

$$\nabla_{\mathbf{x}} \mathcal{L}_o(\mathbf{x}^*, \mathbf{y}^*) = 0, \quad \nabla_{\mathbf{y}} \mathcal{L}_o(\mathbf{x}^*, \mathbf{y}^*) = 0, \tag{9}$$

$$\mathbf{y}^\top \nabla^2_{\mathbf{x}} \mathcal{L}_o(\mathbf{x}^*, \mathbf{y}^*) \mathbf{y} > 0 \tag{10}$$

for all $\mathbf{y} \neq 0$ with $\nabla h^\top(\mathbf{x}^*)\mathbf{y} = 0$. Then $\mathbf{x}^*$ is a strict local minimum of f subject to $h(\mathbf{x}) = 0$.

Lemma 1 [2]: Let P and Q be two symmetric matrices. Assume that Q is positive semidefinite and P is positive definite on the null space of Q, that is, $\mathbf{x}^\top P \mathbf{x} > 0$ for all $\mathbf{x} \neq 0$ with $\mathbf{x}^\top Q \mathbf{x} = 0$. Then there exists a scalar $\bar{c}$ such that

$$P + cQ : \text{positive definite}, \ \forall c > \bar{c}.$$

Now we give our sufficiency optimality condition for optimality of problem (2) as follows.

Lemma 2: Let $\mathbf{z}^*$ be a regular point and together with its associated Lagrange multiplier vector $\boldsymbol{\lambda}^*$ satisfies the sufficiency assumptions of Proposition 1, then $\mathbf{z}^*$ is a strict local minimum of $\langle U(\mathbf{z}) \rangle$ over $\bar{G}_1^\top \mathbf{z} = 0$, where $\langle U(\mathbf{z}) \rangle = (1)/(q) \sum_{i=1}^q U[\mathbf{z}^{(i)}]$.

Proof: The Lagrangian $\mathcal{L}_o$ defined in (8) for problem (2) is given by

$$\mathcal{L}_o(\mathbf{z}, \boldsymbol{\lambda}) = \frac{\eta}{q} U^\top[\mathbf{z}]\mathbb{1}_q + \boldsymbol{\lambda}^\top \bar{G}_0^\top \mathbf{z}.$$

From (4), the gradient and Hessian of $\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda})$ with respect to $\mathbf{z}$ are

$$\nabla_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \frac{\eta}{q} \nabla_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \bar{G}_1 \boldsymbol{\lambda}, \tag{11}$$

$$\nabla^2_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \frac{\eta}{q} \nabla^2_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma, \tag{12}$$

where $\nabla^2_{\mathbf{z}} U(\mathbf{z}) = \operatorname{diag}\{\nabla^2_{\mathbf{z}^{(1)}} U[\mathbf{z}^{(1)}], \ldots, \nabla^2_{\mathbf{z}^{(q)}} U[\mathbf{z}^{(q)}]\}$.

In particular, if $\mathbf{z}^*$ and $\boldsymbol{\lambda}^*$ satisfy the conditions (9) and (10) of Proposition 1, we have

$$\begin{cases} \nabla_{\mathbf{z}} \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) = \dfrac{\eta}{q} \nabla_{\mathbf{z}} U(\mathbf{z}^*) + \bar{G}_\gamma \mathbf{z}^* + \bar{G}_1 \boldsymbol{\lambda}^* \\ \qquad = \nabla_{\mathbf{z}} \mathcal{L}_o(\mathbf{z}^*, \boldsymbol{\lambda}^*) = 0, \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) = \bar{G}_1^\top \mathbf{z}^* = \nabla_{\boldsymbol{\lambda}} \mathcal{L}_o(\mathbf{z}^*, \boldsymbol{\lambda}^*) = 0. \end{cases} \tag{13}$$

On the other hand, it follows from (12) that

$$\nabla^2_{\mathbf{z}} \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) = \frac{\eta}{q} \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \bar{G}_\gamma \geq \frac{\eta}{q} \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \hat{\gamma} \Gamma \bar{G}_\gamma,$$

where $\hat{\gamma} = \min_i(\gamma_i)$, $\Gamma = \operatorname{diag}(1/\gamma_1, \ldots, 1/\gamma_q) \otimes \mathbf{I}_n$.

By the condition (10), we have that

$$\mathbf{z}^\top \nabla^2_{\mathbf{z}} \mathcal{L}_o(\mathbf{z}^*, \boldsymbol{\lambda}^*) \mathbf{z} = \frac{\eta}{q} \mathbf{z}^\top \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) \mathbf{z} > 0, \tag{14}$$

for all $\mathbf{z} \neq 0$ with $\mathbf{z}^\top \Gamma \bar{G}_\gamma \mathbf{z} = 0$. By applying Lemma 1 with $P = (\eta)/(q) \nabla^2_{\mathbf{z}} U(\mathbf{z}^*)$ and $Q = \Gamma \bar{G}_\gamma$, it follows that there exists a scalar $\bar{\gamma} > 0$ such that

$$\begin{aligned} \nabla^2_{\mathbf{z}} \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) &= \frac{\eta}{q} \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \bar{G}_\gamma \\ &\geq \frac{\eta}{q} \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \hat{\gamma} \Gamma \bar{G}_\gamma > 0, \end{aligned} \tag{15}$$

for all $\hat{\gamma} > \bar{\gamma}$.

Using the sufficient optimality condition for unconstrained optimization [2], we conclude from (13) and (15), that for $\hat{\gamma} > \bar{\gamma}$, $\mathbf{z}^*$ is an unconstrained local minimum of $\mathcal{L}(\cdot, \boldsymbol{\lambda}^*)$. In particular, there exist $\kappa > 0$ and $\epsilon > 0$ such that

$$\begin{aligned} \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}^*) \geq \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) &+ \frac{\eta\kappa}{2} \|\mathbf{z} - \mathbf{z}^*\|^2, \\ &\forall \mathbf{z} \text{ with } \|\mathbf{z} - \mathbf{z}^*\| < \epsilon. \end{aligned} \tag{16}$$

Since for all $\mathbf{z}$ with $\bar{G}_1^\top \mathbf{z} = \bar{G}_0^\top \mathbf{z} = 0$, we have $\bar{G}_{\gamma_0}^\top \mathbf{z} = 0$ and $\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}^*) = \eta \langle U(\mathbf{z}) \rangle$, so it follows from (16) that

$$\langle U(\mathbf{z}) \rangle \geq \langle U(\mathbf{z}^*) \rangle + \frac{\kappa}{2} \|\mathbf{z} - \mathbf{z}^*\|^2, \ \forall \mathbf{z} \tag{17}$$

with $\bar{G}_1^\top \mathbf{z}^* = 0, \|\mathbf{z} - \mathbf{z}^*\| < \epsilon$. Therefore, $\mathbf{z}^*$ is a strict local minimum of $\langle U(\mathbf{z}) \rangle$ over $\bar{G}_1^\top \mathbf{z} = 0$. □

IV. Stability of CLMs

To make the array (6) of CLMs or the Lagrange programming network (6) of practical sense, the equilibrium point $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$ should furthermore be asymptotically stable, so that the network will always converge to $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$ from an arbitrary initial point within the attraction domain of $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$. In what follows, we shall analyze local stability and global stability of each CLM $\mathbf{z}^{(i)}$ $(i \in \mathbb{N}_q)$ in the network (6). For further analysis convenience, let us denote $\mathbf{u} = [\mathbf{z}^\top \quad \boldsymbol{\lambda}^\top]^\top \in \mathbb{R}^{n(2q-1)}$ and

$$\dot{\mathbf{u}} = F(\mathbf{u}) = \begin{bmatrix} \dfrac{\eta}{q} \nabla_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \bar{G}_1 \boldsymbol{\lambda} \\ -\bar{G}_1^\top \mathbf{z} \end{bmatrix}. \tag{18}$$

A. Local Stability

Let us first present local stability of the network (6).

Theorem 1: Let $\mathbf{u}^* = [\mathbf{z}^{*\top} \quad \boldsymbol{\lambda}^{*\top}]^\top$ be a stationary point of $\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda})$. Then the Lagrange programming network (6) is locally asymptotically stable at $\mathbf{u}^*$ for some penalty factors γ_i $(i \in \mathbb{N}_{q-1})$ satisfying $(\eta)/(q) \nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \bar{G}_\gamma > 0$, where $\mathbf{z}^*$ is a strict local minimum of the problem (2).

Proof: See *Appendix A*.

B. Global Stability

Next, we shall carry out the global stability analysis for the networks through the Lyapunov function method. Before doing this, we first provide a theorem for guaranteeing the uniqueness of the equilibrium point of the network (6).

Theorem 2: Suppose that the Hessian $(\eta)/(q)\nabla_{\mathbf{z}}^2 U(\mathbf{z}) + \bar{G}_\gamma$ is positive definite for all $\mathbf{z} \in \mathcal{S} \in \mathbb{R}^{nq}$, where $\mathcal{S}$ is the desired optimization domain, then the unique equilibrium point of the Lagrange programming network (6) is given by

$$\mathcal{E} = \left\{ \mathbf{u} = \begin{bmatrix} \mathbf{z}^{*\top} & \boldsymbol{\lambda}^{*\top} \end{bmatrix}^\top \mid \bar{G}_1 \boldsymbol{\lambda}^* = -\frac{\eta}{q}\nabla_{\mathbf{z}} U(\mathbf{z}^*) - \bar{G}_\gamma \mathbf{z}^* \right\}, \tag{19}$$

where $\mathbf{z}^*$ is a unique optimal solution to (2).

Proof: See *Appendix B*.

Theorem 3: Suppose that the Hessian $(\eta)/(q)\nabla_{\mathbf{z}}^2 U(\mathbf{z}) + \bar{G}_\gamma$ is positive definite for all $\mathbf{z} \in \mathcal{S} \in \mathbb{R}^{nq}$, where $\mathcal{S}$ is the desired optimization domain, then the Lagrange programming network (6) is stable in the Lyapunov sense and is globally convergent to an equilibrium point of (6), which corresponds to a unique optimal solution of (2).

Proof: See *Appendix C*.

V. Synchronization of CLM Ensemble

In order to derive a common local minimum, the CLM should converge to a common point so it is also necessary to analyze the synchronization of the CLM ensemble, that is, the stability of synchronization between $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(i+1)}$ $(i \in \mathbb{N}_{q-1})$. Since all local minimizers will be synchronized together ultimately, in the following, we shall generally prove the stability of synchronization between any two minimizers $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ for $i \neq j$ $(i, j \in \mathbb{N}_q)$ by using the directed graph method [36].

Definition 4 [36]: $\mathcal{W}_s$ is the class of irreducible symmetric real matrices with zero row sums and nonpositive off-diagonal elements.

Definition 5 [36]: Given an m by m matrix $\mathcal{V}$, a function $f(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^m$ is $\mathcal{V}$-uniformly decreasing if

$$(\mathbf{x} - \mathbf{y})^\top \mathcal{V}(f(\mathbf{x}) - f(\mathbf{y})) \leq -\vartheta \|\mathbf{x} - \mathbf{y}\|^2 \tag{20}$$

for some $\vartheta > 0$ and all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$.

Theorem 4: Let $P(t)$ be an n by n time-varying matrix and $\mathcal{V}$ be an n by n symmetric positive definite matrix such that $-(\eta)/(q)\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}) + P(t)\mathbf{z}^{(i)}$ $(i \in \mathbb{N}_q)$ is $\mathcal{V}$-uniformly decreasing. Then the network (6) synchronizes in the sense that $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| \to 0$ as $t \to \infty$ for all i, j, which means the asymptotical stability of synchronization between any two minimizers $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ for $i, j \in \mathbb{N}_q, i \neq j$ has been guaranteed, if there exists a q by q matrix Q in $\mathcal{W}_s$ such that

$$R \triangleq \left(\bar{G}_1 \bar{G}_1^\top\right)(Q \otimes \mathcal{V})(-\bar{G}_\gamma - \mathbf{I}_q \otimes P(t)) \leq 0. \tag{21}$$

Proof: See *Appendix D*.

Remark 2: Theorem 4 provides a condition of guaranteeing the stability of synchronization among CLMs, but it is difficult to find the matrix $P(t)$ satisfying the conditions and inefficient to verify the conditions at every instant in practice. However, if the penalty factors γ_i $(i \in \mathbb{N}_{q-1})$ are chosen large enough, the CLMs usually can be synchronized. More specifically, linearization of $-(\eta)/(q)\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}) + P(t)\mathbf{z}^{(i)}$ at $\mathbf{z}^{(i)}$ leads to $-(\eta)/(q)\nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}) + P(t)$. If $P(t) \leq (\eta)/(q)\nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})$ holds at some time, $-(\eta)/(q)\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}) + P(t)\mathbf{z}^{(i)}$ will be $\mathcal{V}$-uniformly decreasing at this time. Meanwhile, a sufficient condition for ensuring inequality (21) is $\mathbf{I}_q \otimes P(t) \geq -\bar{G}_\gamma = -G_\gamma \otimes \mathbf{I}_n$. Therefore, if we choose γ_i $(i \in \mathbb{N}_{q-1})$ appropriately such that

$$-G_\gamma \otimes \mathbf{I}_n \leq \frac{\eta}{q}\nabla^2_{\mathbf{z}^{(i)}} U\left(\mathbf{z}^{(i)}\right), \tag{22}$$

all coupled local minimizers will be synchronized.

Remark 3: Due to the dynamic nature and the convenient conversion from an optimization problem to a dynamical system, at present, several neural networks have been developed to solve optimization problems [11], [21], [22], [12]–[14], [16]. Compared with the conventional numerical optimization method, the neural network approach has an advantage of a low computational complexity and faster running times thanks to the potential of electronic implementation. Though the fruitful achievements of neural network approaches in various kinds of optimization problems and related applications, most of them can be regarded as standard local optimization techniques and lack the capability of global optimization. In order to obtain a global optimal solution for nonconvex optimization problems, multistart searching processes are usually needed, that is, trying different starting points and running the processes independently from each other and selecting the best result from all trials. Different from neural network approaches, the CLM method is based on a cooperative search mechanism realised by minimizing the average cost of the population. Thanks to the coupling mechanism, the CLMs are able to exchange information which results in a better performance than multistart local optimization. Therefore, a global optimal solution of the original problem can be obtained. The advantages of CLM over several neural network approaches will be further illustrated in the next section.

Remark 4: In [38], Hou proposed a hierarchical recurrent neural network (LHONN) consisting of two hierarchically structured sub-networks. The two kinds of sub-networks can work simultaneously and the constraints of state equations are imbedded into the lower level sub-network. Though the states in subsystems decomposed from a large-scale system are coupled with each other, the neural network approach essentially differs from the proposed CLM scheme in three aspects. First, each state in the subsystems is updated based on one of its neighbors and the dynamic evolution of LHONN aims at the control variables instead of state variables, while the dynamic evolution of CLMs is with respect to the state variables and the evolution of each state is not only related to its own history but also its neighbors. Second, from the basic idea point of view, LHONN is related to hierarchical structures and more applicable to optimal control problems, while CLM comes from incorporating principles of master-slave dynamics and mainly focuses on a different and broader context of solving differentiable optimization problems. Third, since there are no interactions among control variables when regarding the control variable as decision variables in LHONN, LHONN is actually a local optimization method while CLM can be considered as a valuable alternative that combines the advantages of local gradient-based algorithms with global exploration.

VI. Extension to Problems With Equality Constraints

In this section, we shall extend our results to the nonlinear optimization problem with equality constraints:

$$\begin{aligned} \min \quad & U(\mathbf{z}),\ \mathbf{z} \in \mathbb{R}^n \\ \text{subject to} \quad & h(\mathbf{z}) = 0. \end{aligned} \tag{23}$$

where $U \in \mathcal{C}^2 : \mathbb{R}^n \to \mathbb{R}$, $h \in \mathcal{C}^1 : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$.

By applying the CLM scheme in Section II, the problem (23) can be formulated as follows by introducing q local minimizers:

$$\begin{aligned} \min_{\mathbf{z}^{(i)} \in \mathbb{R}^n, i \in \mathbb{N}_q} \quad & \frac{1}{q} \sum_{i=1}^{q} U\left[\mathbf{z}^{(i)}\right], \\ \text{subject to} \quad & \begin{cases} \mathbf{z}^{(i)} - \mathbf{z}^{(i+1)} = 0, & i \in \mathbb{N}_{q-1}, \\ h\left(\mathbf{z}^{(q)}\right) = 0. \end{cases} \end{aligned} \tag{24}$$

For these equality constraints, let us introduce q Lagrange multipliers $\boldsymbol{\lambda}_i \in \mathbb{R}^n, i \in \mathbb{N}_{q-1}$, $\boldsymbol{\lambda}_q \in \mathbb{R}^m$, and q penalty factors $\gamma_i > 0, i \in \mathbb{N}_q$. Then one defines the augmented Lagrangian:

$$\begin{aligned} \mathcal{L}&\left(\mathbf{z}^{(i)}, \boldsymbol{\lambda}^{(i)}\right) \\ &= \frac{\eta}{q} \sum_{i=1}^{q} U\left[\mathbf{z}^{(i)}\right] + \frac{1}{2} \sum_{i=1}^{q-1} \gamma_i \left\|\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right\|^2 \\ &\quad + \sum_{i=1}^{q-1} \left\langle \boldsymbol{\lambda}_i, \left[\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right] \right\rangle \\ &\quad + \frac{1}{2}\gamma_q \left\|h\left(\mathbf{z}^{(q)}\right)\right\|^2 + \left\langle \boldsymbol{\lambda}_q, h\left(\mathbf{z}^{(q)}\right) \right\rangle. \end{aligned}$$

Similar to the analysis in (3)–(6), we obtain the following Lagrange programming network

$$\begin{cases} \dot{\mathbf{z}} = -\left(\dfrac{\eta}{q}\nabla_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \mathcal{H}_1(\mathbf{z}) + \mathcal{H}_2(\mathbf{z})\boldsymbol{\lambda}\right), \\ \dot{\boldsymbol{\lambda}} = \mathcal{H}_3(\mathbf{z}), \end{cases} \tag{25}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}^{(1)} \\ \vdots \\ \mathbf{z}^{(q)} \end{bmatrix} \in \mathbb{R}^{nq}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}^{(1)} \\ \vdots \\ \boldsymbol{\lambda}^{(q)} \end{bmatrix} \in \mathbb{R}^{n(q-1)+m},$$

$$\mathcal{H}_1(\mathbf{z}) = \begin{bmatrix} 0_{n(q-1)\times 1} \\ \gamma_q \left(\nabla_{\mathbf{z}^{(q)}} h\left(\mathbf{z}^{(q)}\right)\right) h\left(\mathbf{z}^{(q)}\right) \end{bmatrix},$$

$$\mathcal{H}_2(\mathbf{z}) = \left[\bar{G}_1 \quad \left[0_{m\times n(q-1)} \ \left(\nabla_{\mathbf{z}^{(q)}} h\left(\mathbf{z}^{(q)}\right)\right)^\top\right]^\top\right],$$

$$\mathcal{H}_3(\mathbf{z}) = \begin{bmatrix} \bar{G}_1^\top \mathbf{z} \\ h(\mathbf{z}^{(q)}) \end{bmatrix},$$

and $\mathbf{z}, \bar{G}_\gamma, \bar{G}_1, \nabla_{\mathbf{z}} U(\mathbf{z})$ are the same with the definitions in (6).

For the stability of (25), we present the following theorem whose proof is similar to Theorem 3, thus we omit the detailed proof hereafter.

Theorem 5: Suppose that $\nabla_{\mathbf{z}}\mathcal{F}(\mathbf{z}, \boldsymbol{\lambda})$ is positive definite everywhere, where

$$\mathcal{F}(\mathbf{z}, \boldsymbol{\lambda}) = \frac{\eta}{q}\nabla_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \mathcal{H}_1(\mathbf{z}) + \mathcal{H}_2(\mathbf{z})\boldsymbol{\lambda},$$

then the Lagrange programming network (25) is stable in the Lyapunov sense and is globally convergent to an equilibrium point of (25), which corresponds to a unique optimal solution of (24) or (23).

VII. Illustrative Examples

The theoretical results about local and global stability of Lagrange programming networks generated by CLMs discussed in the previous sections are further illustrated now by the following numerical examples.

Example 1: Let us recall the double well cost function [28]

$$U(z) = z^4 - 16z^2 + 5z + 100 \tag{26}$$

with global minimum located at $z = -2.90$ and a local minimum $z = 2.76$.

Following the procedure in Section 2, one can derive a network (6) with q coupled local minimizers. Here,

$$\nabla_{\mathbf{z}} U(\mathbf{z}) = \begin{bmatrix} \nabla_{z^{(1)}} U\left[z^{(1)}\right] \\ \vdots \\ \nabla_{z^{(q)}} U\left[z^{(q)}\right] \end{bmatrix} \in \mathbb{R}^q,$$

$$\nabla^2_{z^{(i)}} U\left[z^{(i)}\right] = 12\left(z^{(i)}\right)^2 - 32, i \in \mathbb{N}_q.$$

It can be seen that $\nabla^2_{\mathbf{z}} U(\mathbf{z}) > 0$ does not always hold, which implies that $(\eta)/(q)\nabla^2_{\mathbf{z}} U(\mathbf{z}) + \bar{G}_\gamma > 0$ may not hold. Therefore, according to Theorem 3, it is insufficient to say that the Lagrange programming network (6) will globally converge to an equilibrium point. However, according to Theorem 1, the Lagrange programming network (6) is local asymptotically stable for penalty factors $\gamma_i = 100$ $(i \in \mathbb{N}_{q-1})$ satisfying $(\eta)/(q)\nabla^2_{\mathbf{z}} U(\mathbf{z}^*) + \bar{G}_\gamma > 0$ with $\eta = 40, q = 10$. Let us choose $\eta = 40$, $\gamma_i = 100$, $i \in \mathbb{N}_{q-1}$, and take $\mathcal{V} = 1, P(t) = 1, Q = \mathbf{I}_q - \mathbf{J}_q/q \in \mathcal{W}_s$ in Theorem 4 where $\mathbf{J}_q$ is a q by q matrix with all 1's. It is easy to verify that $R \leq 0$ by (21) (actually, the maximum eigenvalue of R is -2.0185×10^{-13}). Therefore, all CLMs $z^{(i)}$ $(i \in \mathbb{N}_q)$ will synchronize as $t \to \infty$. Fig. 1 shows the convergence behaviors of $q = 10$ local minimizers, with 8 local minimizers starting randomly from $[0 \quad 3]$ (blue solid line), 2 local minimizers starting randomly from $[-1 \quad 0]$ (red dashed line). We can find that all the 8 local minimizers near the local minimum have been pulled out the region and converge to the global minimum. This is because the $q = 10$ local optimizers have exchanged information through the synchronization constraints $z^{(i)} - z^{(i+1)} = 0$, $i \in \mathbb{N}_{q-1}$ during the optimization process, which helps finding the global minimum easily. Figs. 2, 3 and 4 show comparison results based on the same initial values using the neural network approaches proposed in [12], [16] and [11], respectively. It is shown that the global minimum is achieved from all initial values by the CLM scheme while not by the neural network approaches. Note that the convergence speed by the CLM method is much more slow compared to the neural

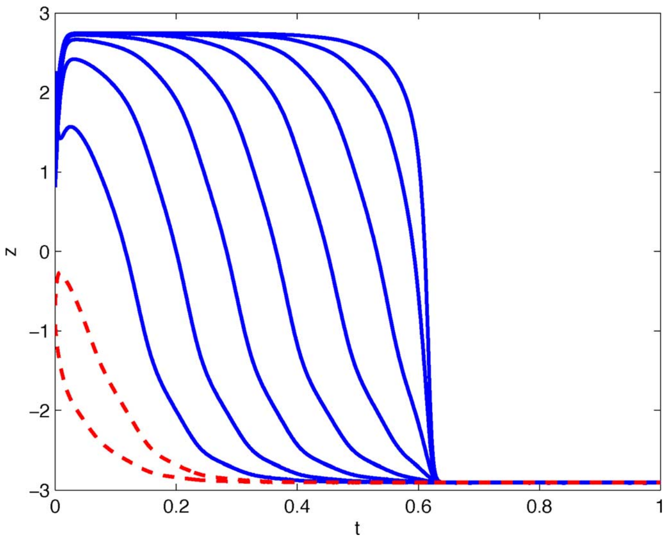


Fig. 1. Convergence behavior of the decision variable z using CLM for Example 1.

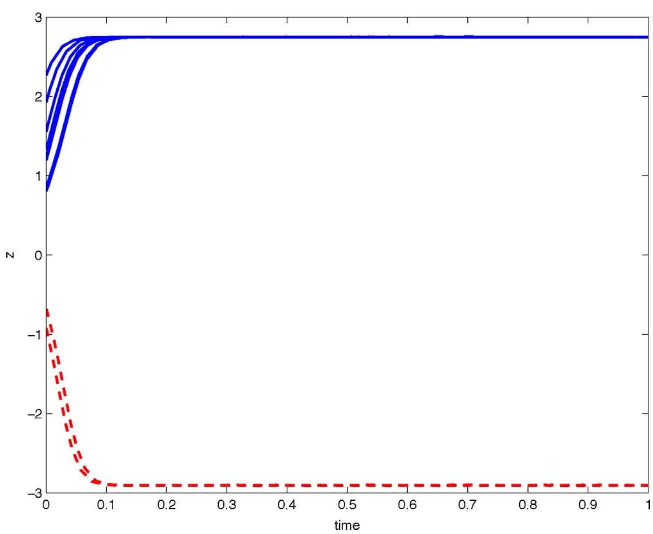


Fig. 2. Neural network approach in [12] (here, $\rho = 1$) for the same problem as in Fig. 1 with the same initial conditions.

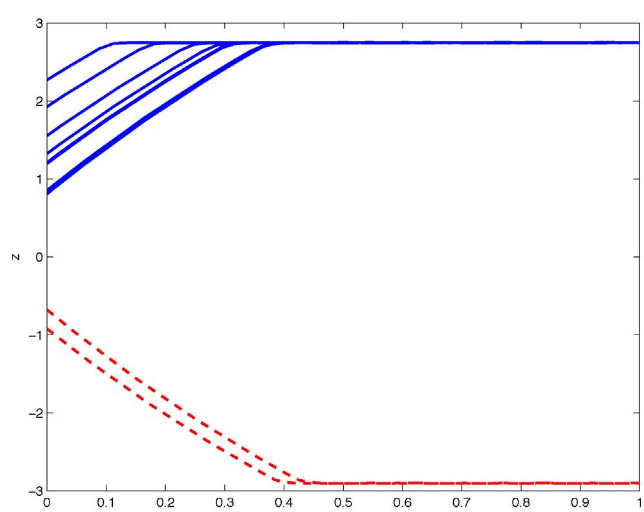


Fig. 3. Neural network approach in [16] for the same problem as in Fig. 1 with the same initial conditions.

network approaches. The reason is that for the nonconvex function (26), those local minimizers near the local optimum

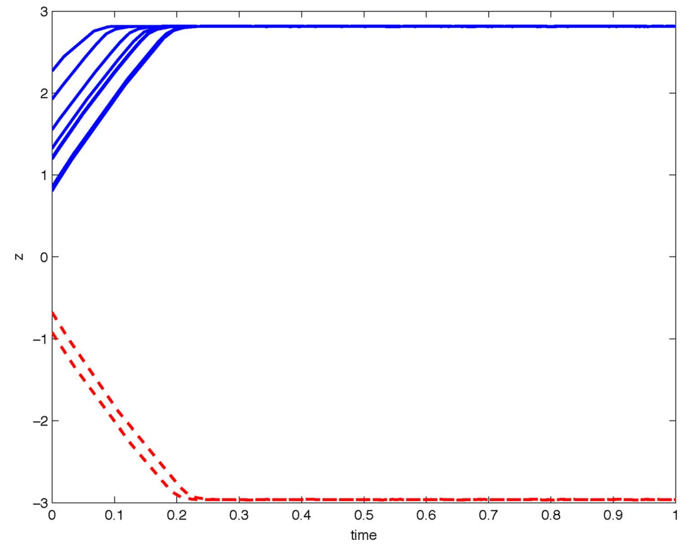


Fig. 4. Delayed neural network approach in [11] (here, $\alpha = 1, \tau = 1$) for the same problem as in Fig. 1 with the same initial conditions.

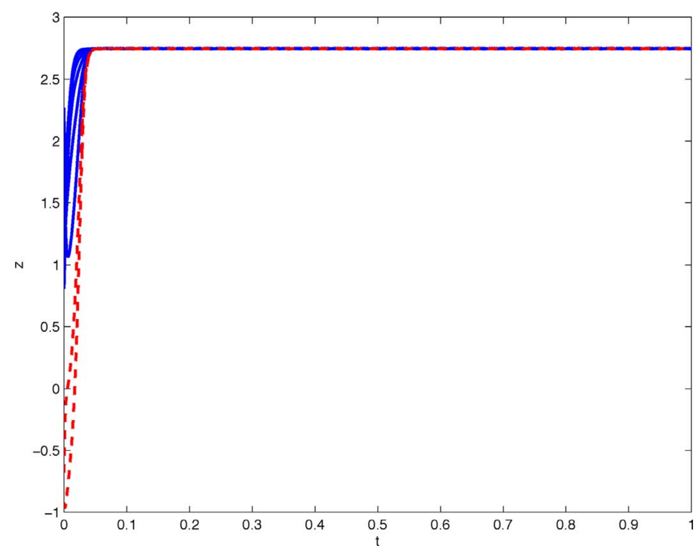


Fig. 5. Convergence behavior of the decision variable z using CLM with $\gamma_i = 200$ $(i \in \mathbb{N}_9)$ in Example 1.

basin move out of the basin one by one due to the ring-type interactions among local minimizers. In fact, in this example, the interactions among local minimizers do not help enhancing the convergence speed. It is only because the CLMs with coupling try to do more explorations and reach the global minimum. However, it is not implying that our method has slower convergence speed than those neural network approaches. To demonstrate this point, we do the above experiment with $\gamma_i = 200$ $(i \in \mathbb{N}_9)$ again and show the CLM results in Fig. 5. It is seen that due to the stronger interactions among CLMs, the local minimizers converge to an equilibrium point with a faster speed than the neural network approaches, but at the price of getting a local optimum. It is worth mentioning that the search points by the neural network approaches are not synchronized though they can also converge quickly. Here, the interactions existing in the CLM scheme help to do more explorations of the search space. As a comparison illustration, in the next example, the interactions in our scheme will help to increase the convergence speed due to the convexity of that function.

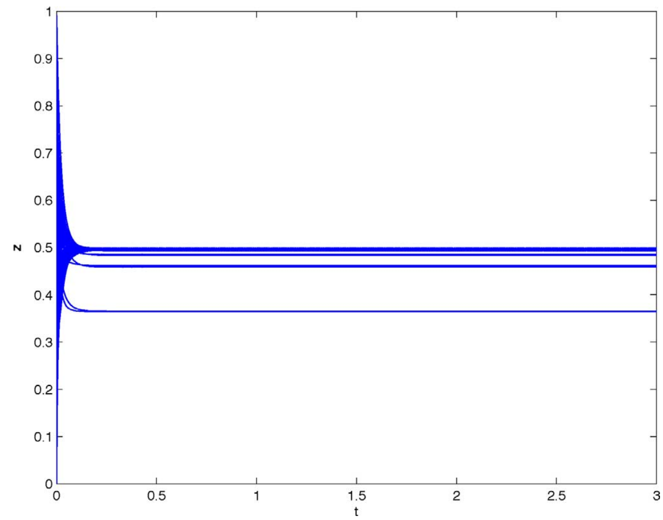


Fig. 6. Convergence behavior of the decision variable $\mathbf{z}$ using CLM for Example 2.

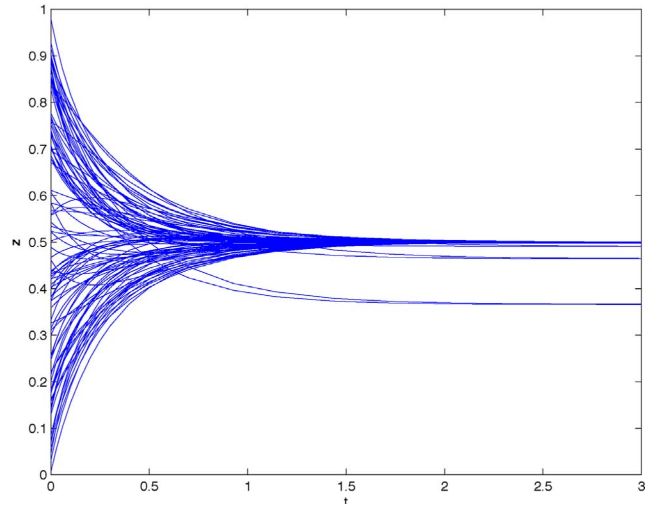


Fig. 7. Neural network approach in [11] for the problem in Example 2.

Example 2: Consider the following quadratic programming problem:

$$\min U(\mathbf{z}) = \frac{1}{2}\mathbf{z}^\top A\mathbf{z} + p^\top \mathbf{z}, \tag{27}$$

where $\mathbf{z} \in \mathbb{R}^{100}$, $p = [1 \quad 1 \quad \cdots \quad 1]^\top$,

$$A = \begin{bmatrix} 4 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 4 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 4 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 & 4 & -1 \\ 0 & 0 & \cdots & 0 & 0 & -1 & 4 \end{bmatrix}_{100\times 100}.$$

The quadratic programming problem has an optimal solution and the optimal value of the objective function is $U(\mathbf{z}^*) = -24.817$. While performing the CLM array (6) within the Lagrange programming network framework, since $\bar{G}_\gamma \geq 0$ and $\nabla^2_{\mathbf{z}} U(\mathbf{z}) = A > 0$, we have $(\eta)/(q)\nabla^2_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma > 0$. Therefore, by applying Theorem 3, the Lagrange programming network (6) is stable and globally convergent to an equilibrium point of (6), which corresponds to a unique optimal solution of (27). In simulation, we choose $q = 3$, $\eta = 40$ and $\gamma_i = 100$, $i \in \mathbb{N}_{q-1}$. Initial values are randomly chosen from $[0, 1]$. Similar to the choices in Example 1, let us take $\mathcal{V} = 1, P(t) = 1, Q = \mathbf{I}_q - \mathbf{J}_q/q \in \mathcal{W}_s$ in Theorem 4 where $\mathbf{J}_q$ is a q by q matrix with all 1's. One can also obtain that $R \leq 0$ according to (21). Therefore, all CLMs $z^{(i)}$ $(i \in \mathbb{N}_q)$ will synchronize as $t \to \infty$.

Fig. 6 shows the convergence behaviors of 100 decision variables $\mathbf{z}$ with $q = 3$ local minimizers each. Comparing to the results, which are shown in Figs. 7 and 8 and based on neural network approaches proposed in [11], [37], it is found that the minimum is reached much more quickly by the CLM scheme in comparison to the neural network approaches. Since the quadratic programming problem in this example is convex, all CLMs can individually converge to the same optimal solution. Thereby, the interactions here among the CLMs help to accelerate the converge speed.

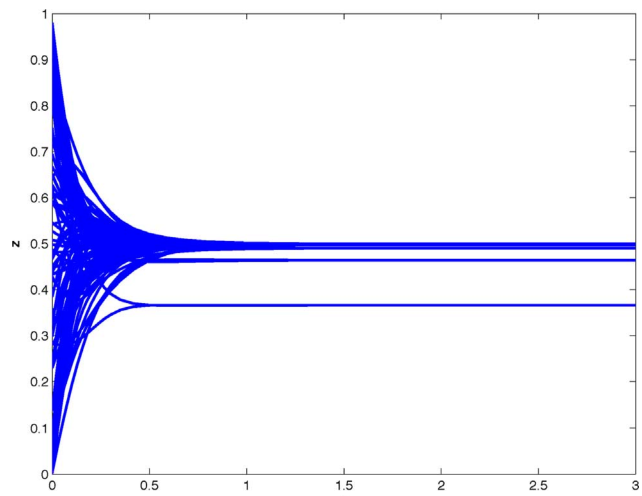


Fig. 8. Neural network approach in [37] for the problem in Example 2.

VIII. CONCLUSION

In this paper, we have analyzed the stability of CLMs and the synchronization of CLMs within the Lagrange programming network framework. We first dealt with the unconstrained optimization problem which can be transformed into a nonlinear optimization problem with linear equality constraints under the CLM scheme. We then presented conditions for stability of the Lagrange programming networks to the new optimization problem. Furthermore, we have also provided conditions for the synchronization of CLMs in order to ensure that all CLMs converge to a common optimization solution. Finally, simulations of two test functions have been performed to illustrate the effectiveness and advantages of the obtained results. We hope that these results will offer insight into CLMs within the Lagrange programming network framework, and consequently reveal deeper implications for the CLM scheme to unconstrained optimization problems. In future work, the current analytical work can be further extended to inequality constrained optimization problems.

APPENDIX A

Proof of Theorem 1: To begin with, let us linearize (6) at the equilibrium point $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$. The local characteristic of the equilibrium is determined by the linearized system. Taking

$$\begin{cases} \frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}^*) + \bar{G}_\gamma \mathbf{z}^* + \bar{G}_1\boldsymbol{\lambda}^* = 0 \\ -\bar{G}_1^\top \mathbf{z}^* = 0 \end{cases} \tag{28}$$

into account, the linearized system is given by

$$\begin{bmatrix} \dot{\mathbf{z}} \\ \dot{\boldsymbol{\lambda}} \end{bmatrix} = -\begin{bmatrix} \frac{\eta}{q}\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma & \bar{G}_1 \\ -\bar{G}_1^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{z} - \mathbf{z}^* \\ \boldsymbol{\lambda} - \boldsymbol{\lambda}^* \end{bmatrix}. \tag{29}$$

Define

$$\Theta = \begin{bmatrix} \frac{\eta}{q}\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma & \bar{G}_1 \\ -\bar{G}_1^\top & 0 \end{bmatrix}.$$

We shall prove that the real part of each eigenvalue of Θ is strictly positive provided that there exist some penalty factors $\gamma_i > 0$ $(i \in \mathbb{N}_q)$ satisfying $(\eta)/(q)\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma > 0$.

First, let us denote $\tilde{v}$ the complex conjugate of a complex vector v and $\mathrm{Re}(\beta)$ the real part of a complex number β. Let α be an eigenvalue of Θ, and let

$$\vartheta = \begin{bmatrix} v \\ w \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{30}$$

be a corresponding eigenvector, where v and w are complex vectors of dimension nq and $n(q-1)$, respectively. Then we have

$$\mathrm{Re}\{\vartheta^\top \Theta \vartheta\} = \mathrm{Re}\{\alpha \vartheta^\top \vartheta\} = \mathrm{Re}(\alpha)(\|v\|^2 + \|w\|^2). \tag{31}$$

On the other hand, using the definition of Θ

$$\mathrm{Re}\{\vartheta^\top \Theta \vartheta\} = \mathrm{Re}\left\{\tilde{v}^\top \left(\frac{\eta}{q}\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma\right) v + \tilde{v}^\top \bar{G}_1 w - \tilde{w}^\top \bar{G}_1^\top v\right\}. \tag{32}$$

Since for any real $nq \times n(q-1)$ matrix M,

$$\mathrm{Re}\{\tilde{v}^\top M w\} = \mathrm{Re}\{\tilde{w}^\top M^\top v\}.$$

it follows from (31) and (32) that

$$\mathrm{Re}\left\{\tilde{v}^\top \left(\frac{\eta}{q}\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma\right) v\right\} = \mathrm{Re}\{\vartheta^\top \Theta \vartheta\} = \mathrm{Re}(\alpha)(\|v\|^2 + \|w\|^2). \tag{33}$$

Moreover, for any positive definite matrix P, we have

$$\mathrm{Re}\{\tilde{v}^\top P v\} > 0,\ v \neq 0, \tag{34}$$

so it follows from (33) and the positive definiteness assumption on $(\eta)/(q)\nabla_{\mathbf{z}}^2 U(\mathbf{z}^*) + \bar{G}_\gamma$ that either $\mathrm{Re}(\alpha) > 0$ or else $v = 0$. But if $z = 0$, the equation $\Theta\vartheta = \alpha\vartheta$ yields $\bar{G}_1 w = 0$. From the definition of $\bar{G}_1$, it is easy to derive $w = 0$. This contradicts our earlier assumption (30). Consequently we must have $\mathrm{Re}(\alpha) > 0$. Thus the real part of each eigenvalue of Θ is strictly positive, which means that $-\Theta$ is strictly negative definite. Therefore, $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$ is locally asymptotically stable (see, [39]). □

APPENDIX B

Proof of Theorem 2: Since the Jacobian matrix of the mapping F given in (18)

$$\nabla_{\mathbf{u}} F(\mathbf{u}) = \begin{bmatrix} \frac{\eta}{q}\nabla_{\mathbf{z}}^2 U(\mathbf{z}) + \bar{G}_\gamma & \bar{G}_1 \\ -\bar{G}_1^\top & 0 \end{bmatrix} \tag{35}$$

is positive semidefinite, $F(\cdot)$ is monotonically increasing. Thus, the solution set $F(\mathbf{u}) = 0$ is convex [34], that is to say, for any two solutions $\mathbf{u}^{(a)}$ and $\mathbf{u}^{(b)}$ of $F(\mathbf{u}) = 0$, and any $\zeta \in [0,1]$, $\zeta\mathbf{u}^{(a)} + (1-\zeta)\mathbf{u}^{(b)}$ is also a solution of $F(\mathbf{u}) = 0$. Hence the set of the equilibrium points of the Lagrange programming network (6) is convex. Since these equilibrium points satisfy

$$\begin{cases} \frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \bar{G}_1\boldsymbol{\lambda} = 0, \\ -\bar{G}_1^\top \mathbf{z} = 0, \end{cases} \tag{36}$$

their set can be expressed as

$$\mathcal{E} = \left\{\mathbf{u} = \begin{bmatrix}\mathbf{z}^\top & \boldsymbol{\lambda}^\top\end{bmatrix}^\top \;\middle|\; \frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma \mathbf{z} + \bar{G}_1\boldsymbol{\lambda} = 0, \bar{G}_1^\top \mathbf{z} = 0\right\}. \tag{37}$$

To prove the uniqueness of the optimal solution, suppose that there exist $\mathbf{u}^{(1)} = (\mathbf{z}^{(1)}, \boldsymbol{\lambda}^{(1)}) \in \mathcal{E}$ and $\mathbf{u}^{(2)} = (\mathbf{z}^{(2)}, \boldsymbol{\lambda}^{(2)}) \in \mathcal{E}$. Substituting them into (36) and combining the equations yield

$$\begin{cases} \Xi\left(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}\right) + \bar{G}_1\left(\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}^{(2)}\right) = 0, \\ \bar{G}_1^\top\left(\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\right) = 0, \end{cases} \tag{38}$$

where $\Xi(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = (\eta)/(q)(\nabla_{\mathbf{z}}U(\mathbf{z}^{(1)}) - \nabla_{\mathbf{z}}U(\mathbf{z}^{(2)})) + \bar{G}_\gamma(\mathbf{z}^{(1)} - \mathbf{z}^{(2)})$. Multiplying by $\mathbf{z}^{(1)} - \mathbf{z}^{(2)}$ on the left side of (38), one can further derive

$$\left(\mathbf{z}^{(1)} - \mathbf{z}^{(2)}\right)^\top \Xi\left(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}\right) = 0. \tag{39}$$

Since $(\eta)/(q)\nabla_{\mathbf{z}}^2 U(\mathbf{z}) + \bar{G}_\gamma$ is positive definite, $(\eta)/(q)\nabla_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma\mathbf{z}$ is strictly monotonically increasing as $\mathbf{z}$ increases. Therefore, (39) implies that $\mathbf{z}^{(1)} = \mathbf{z}^{(2)}$. Then we can further get $\Xi(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = 0$. Substituting $\Xi(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) = 0$ into the first equation of (38), it follows that $\bar{G}_1(\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}^{(2)}) = 0$, so

$$\mathcal{E} = \left\{\mathbf{u} = [\mathbf{z}^{*\top} \quad \boldsymbol{\lambda}^{*\top}]^\top \;\middle|\; \bar{G}_1\boldsymbol{\lambda} = -\frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}^*) - \bar{G}_\gamma\mathbf{z}^*\right\}. \tag{40}$$

Furthermore, since $\bar{G}_1^\top$ has full rank (see, Remark 1), then the linear system $\bar{G}_1\boldsymbol{\lambda} = 0$ has only one zero solution. Therefore, $\bar{G}_1(\boldsymbol{\lambda}^{(1)} - \boldsymbol{\lambda}^{(2)}) = 0$ must imply that $\boldsymbol{\lambda}^{(1)} = \boldsymbol{\lambda}^{(2)}$. That is, $\mathbf{u}^{(1)} = \mathbf{u}^{(2)}$. So, the Lagrange programming network (6) has a unique equilibrium point. □

Appendix C

Proof of Theorem 3: Consider the following Lyapunov function:

$$V(\mathbf{u}) = \frac{1}{2}\|F(\mathbf{u})\|^2 + \frac{1}{2}\|\mathbf{u} - \mathbf{u}^*\|^2, \tag{41}$$

where $F(\mathbf{u})$ is defined in (18) and $\mathbf{u}^* = (\mathbf{z}^*, \boldsymbol{\lambda}^*)$ is a stationary point of the Lagrangian (4). Differentiating $V(\mathbf{u})$ with respect to time t gives

$$\begin{aligned}\frac{\mathrm{d}V(\mathbf{u})}{\mathrm{d}t} &= \nabla_{\mathbf{u}} V(\mathbf{u})\dot{\mathbf{u}} \\ &= -F^\top(\mathbf{u})\nabla_{\mathbf{u}}F(\mathbf{u})F(\mathbf{u}) \\ &\quad - (\mathbf{u} - \mathbf{u}^*)^\top F(\mathbf{u}),\end{aligned} \tag{42}$$

where $\nabla_{\mathbf{u}}F(\mathbf{u})$ has been given in (35).

Using the assumption $(\eta)/(q)\nabla^2_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma > 0$ and using the fact $F(\mathbf{u}^*) = 0$, it follows that

$$\frac{\mathrm{d}V(\mathbf{u})}{\mathrm{d}t} \le 0. \tag{43}$$

Therefore, the Lagrange programming network (6) is stable in the Lyapunov sense. From (41), we have

$$V(\mathbf{u}) \ge \frac{1}{2}\|\mathbf{u} - \mathbf{u}^*\|^2, \tag{44}$$

for any initial point $\mathbf{u}_0$, so there exist a convergent subsequence $\{\mathbf{u}(t_k)\}$ such that $\lim_{k\to\infty}\mathbf{u}(t_k) = \hat{\mathbf{u}}$, where $\hat{\mathbf{u}} = [\hat{\mathbf{z}}^\top \quad \hat{\boldsymbol{\lambda}}^\top]^\top$ satisfying

$$\frac{\mathrm{d}V(\hat{\mathbf{u}})}{\mathrm{d}t} = 0. \tag{45}$$

We now show that $\hat{\mathbf{u}}$ is an equilibrium point of (6). It can be seen that (45) holds if and only if

$$\begin{cases} F^\top(\hat{\mathbf{u}})\nabla_{\mathbf{u}}F(\hat{\mathbf{u}})F(\hat{\mathbf{u}}) = 0, \\ (\hat{\mathbf{u}} - \mathbf{u}^*)^\top F(\hat{\mathbf{u}}) = 0. \end{cases} \tag{46}$$

Because $\nabla^2_{\mathbf{z}}U(\hat{\mathbf{z}}) + \bar{G}_\gamma$ is positive definite, the first equation of (46) implies

$$\frac{\eta}{q}\nabla_{\mathbf{z}}U(\hat{\mathbf{z}}) + \bar{G}_\gamma\hat{\mathbf{z}} + \bar{G}_1\hat{\boldsymbol{\lambda}} = 0. \tag{47}$$

For the stationary point $\mathbf{u}^* = [\mathbf{z}^{*\top} \quad \boldsymbol{\lambda}^{*\top}]^\top$ of the Lagrangian (4), we know that

$$\frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}^*) + \bar{G}_\gamma\mathbf{z}^* + \bar{G}_1\boldsymbol{\lambda}^* = 0. \tag{48}$$

Combining (47) with (48) yields

$$\frac{\eta}{q}\nabla_{\mathbf{z}}U(\hat{\mathbf{z}}) + \bar{G}_\gamma\hat{\mathbf{z}} - \left(\frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}^*) + \bar{G}_\gamma\mathbf{z}^*\right) + \bar{G}_1(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*) = 0. \tag{49}$$

On one hand, by utilizing the fact $\bar{G}_1^\top\mathbf{z}^* = 0$, we have

$$\begin{aligned}(\hat{\mathbf{z}} - \mathbf{z}^*)^\top\bar{G}_1(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*) &= \left(\bar{G}_1^\top\hat{\mathbf{z}} - \bar{G}_1^\top\mathbf{z}^*\right)^\top(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*) \\ &= \hat{\mathbf{z}}^\top\bar{G}_1(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*).\end{aligned} \tag{50}$$

On the other hand, substituting (47) into the second equation of (46) yields

$$-(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*)^\top\bar{G}_1^\top\hat{\mathbf{z}} = 0 \tag{51}$$

or equivalently

$$\hat{\mathbf{z}}^\top\bar{G}_1(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*) = 0. \tag{52}$$

Therefore, substituting the above equation into (50), we derive

$$(\hat{\mathbf{z}} - \mathbf{z}^*)^\top\bar{G}_1(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}^*) = 0. \tag{53}$$

From (50) and (53), it follows that

$$(\hat{\mathbf{z}} - \mathbf{z}^*)^\top\left(\frac{\eta}{q}\nabla_{\mathbf{z}}U(\hat{\mathbf{z}}) + \bar{G}_\gamma\hat{\mathbf{z}} - \frac{\eta}{q}\nabla_{\mathbf{z}}U(\mathbf{z}^*) - \bar{G}_\gamma\mathbf{z}^*\right) = 0. \tag{54}$$

Because $(\eta)/(q)\nabla^2_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma$ is positive definite, $(\eta)/(q)\nabla_{\mathbf{z}}U(\mathbf{z}) + \bar{G}_\gamma\mathbf{z}$ is strictly monotonically increasing as $\mathbf{z}$ increases. Therefore, (54) implies that $\hat{\mathbf{z}} = \mathbf{z}^*$ and $\bar{G}_1^\top\hat{\mathbf{z}} = \bar{G}_1^\top\mathbf{z}^* = 0$. Moreover, it follows from (49) that $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^*$. Hence, $\bar{\mathbf{u}}$ is a stationary point of the Lagrangian (4) and thus is an equilibrium point of (6).

Now let us define another Lyapunov function

$$W(\mathbf{u}) = \frac{1}{2}\|F(\mathbf{u})\|^2 + \frac{1}{2}\|\mathbf{u} - \hat{\mathbf{u}}\|^2. \tag{55}$$

One can obtain that

$$\lim_{k\to\infty} W(\mathbf{u}(t_k)) = W(\hat{\mathbf{u}}) = 0. \tag{56}$$

Therefore, $\forall\epsilon > 0$ there exists $r > 0$ such that for all $t_k \ge t_r$, we have

$$W(\mathbf{u}(t_k)) < \epsilon. \tag{57}$$

Similarly, we can obtain $\mathrm{d}W/\mathrm{d}t \le 0$. It follows that for $t \ge t_r$

$$\frac{1}{2}\|\mathbf{u}(t) - \hat{\mathbf{u}}\|^2 \le W(\mathbf{u}(t)) \le \epsilon, \tag{58}$$

for $W(\mathbf{u}(t))$ decrease as $t \to +\infty$. So we have

$$\lim_{t\to\infty}\|\mathbf{u}(t) - \hat{\mathbf{u}}\| = 0 \tag{59}$$

and

$$\lim_{t\to\infty}\mathbf{u}(t) = \hat{\mathbf{u}}. \tag{60}$$

Hence, the Lagrange programming network (6) is globally convergent to an equilibrium point $\hat{\mathbf{u}} = [\hat{\mathbf{z}}^\top \quad \hat{\boldsymbol{\lambda}}^\top]^\top$, where $\hat{\mathbf{z}}$ is the optimal solution of (2). □

Appendix D

Proof of Theorem 4: Since the matrix $Q \in \mathcal{W}_s$ and $\mathcal{V}$ is symmetric positive definite, $Q \otimes \mathcal{V} \ge 0$. And from the definition of G_1 in (6), G_1 is a matrix with zero column sums. Hence,

$G_1 G_1^\top$ is a symmetric matrix with zero row sums and nonpositive off-diagonal elements, so is $\bar{G}_1 \bar{G}_1^\top$. Therefore, we can get $(G_1 G_1^\top)Q \in \mathcal{W}_s$ and $(\bar{G}_1\bar{G}_1^\top)(Q \otimes \mathcal{V}) = (Q \otimes \mathcal{V})(\bar{G}_1\bar{G}_1^\top) \geq 0$, and then construct the Lyapunov function

$$W(t) = \frac{1}{2}\mathbf{z}^T \left(\bar{G}_1\bar{G}_1^\top\right)(Q \otimes \mathcal{V})\mathbf{z} + \frac{1}{2}\boldsymbol{\lambda}^T \bar{G}_1^\top (Q \otimes \mathcal{V})\bar{G}_1 \boldsymbol{\lambda}.$$

Calculating the derivative of $W(t)$ along trajectories of (6) yields

$$\begin{aligned}
\dot{W}(t) &= \mathbf{z}^T \left(\bar{G}_1\bar{G}_1^\top\right)(Q \otimes \mathcal{V})\left(-\frac{\eta}{q}\nabla_{\mathbf{z}} U(\mathbf{z}) + \mathbf{I}_q \otimes P(t)\mathbf{z}\right) \\
&\quad + \mathbf{z}^T(t)(\bar{G}_1\bar{G}_1^\top)(Q \otimes \mathcal{V})(-\bar{G}_\gamma - \mathbf{I}_q \otimes P(t))\mathbf{z} \\
&\quad - \mathbf{z}^T(\bar{G}_1\bar{G}_1^\top)(Q \otimes \mathcal{V})\bar{G}_1\boldsymbol{\lambda} + \boldsymbol{\lambda}^T \bar{G}_1^\top (Q \otimes \mathcal{V})\bar{G}_1\bar{G}_1^\top \mathbf{z} \\
&= \mathbf{z}^T \left(\bar{G}_1\bar{G}_1^\top\right)(Q \otimes \mathcal{V}) \begin{pmatrix} -\frac{\eta}{q}\nabla_{\mathbf{z}^{(1)}} U\left(\mathbf{z}^{(1)}\right) + P(t)\mathbf{z}^{(1)} \\ \vdots \\ -\frac{\eta}{q}\nabla_{\mathbf{z}^{(q)}} U\left(\mathbf{z}^{(q)}\right) + P(t)\mathbf{z}^{(q)} \end{pmatrix} \\
&\quad + \mathbf{z}^T(t) R \mathbf{z}(t) \\
&\leq \sum_{i<j} -\Omega_{ij}\left(\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\right)^\top \mathcal{V}\left(-\frac{\eta}{q}\nabla_{\mathbf{z}^{(i)}} U\left(\mathbf{z}^{(i)}\right)\right. \\
&\quad \left. + P(t)\mathbf{z}^{(i)} + \frac{\eta}{q}\nabla_{\mathbf{z}^{(j)}} U\left(\mathbf{z}^{(j)}\right) - P(t)\mathbf{z}^{(j)}\right), \qquad (61)
\end{aligned}$$

where $\Omega = (G_1 G_1^\top)Q$. By means of the $\mathcal{V}$-uniformly decreasing condition and $R \leq 0$,

$$\dot{W}(t) \leq \sum_{i<j} -\Omega_{ij}\left(-\vartheta \left\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\right\|^2\right). \qquad (62)$$

Note that $\Omega = (G_1 G_1^\top)Q \in \mathcal{W}_s$, so $-\Omega_{ij} \geq 0$ for $i < j$. For each $-\Omega_{ij} \geq 0$ and $\delta > 0$, and sufficiently large t such that if $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| \geq \delta$, then

$$\dot{W}(t) \leq -\frac{\vartheta}{2}\left\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\right\|^2. \qquad (63)$$

This implies that for large enough t, $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| \leq \delta$. Therefore, $\lim_{t\to\infty} \|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| = 0$. Irreducibility of Ω implies that enough Ω_{ij} are nonzero to ensure $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| \to 0, \forall i, j$. The proof is completed. □

REFERENCES

[1] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press, 1982.

[2] D. P. Bertsekas, *Nonlinear Progamming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[3] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. Chichester, U.K.: Wiley, 2000.

[4] V. H. Nguyen and J. J. Strodiot, "Convergence rate for a penalty function-method of exponential type," *J. Optimization Theory and Applications*, vol. 27, no. 4, pp. 495–508, 1979.

[5] H. Z. Luo, X. L. Sun, and D. Li, "On the convergence of augmented Lagrangian methods for constrained global optimization," *SIAM J. Optimization*, vol. 18, no. 4, pp. 1209–1230, 2008.

[6] S. Z. Li, "Improving convergence and solution quality of Hopfield-type neural networks with augmented Lagrange multipliers," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1507–1516, 1996.

[7] X. X. Huang and X. Q. Yang, "Further study on augmented Lagrangian duality theory," *J. Global Optimization*, vol. 31, no. 2, pp. 193–210, 2005.

[8] W. W. Hager, "Dual techniques for constrained optimization," *J. Optimization Theory and Applications*, vol. 55, no. 1, pp. 37–71, 1987.

[9] H. Yamashita, "A globally convergent constrained quasi-Newton method with an augmented Lagrangian type penalty function," *Math. Program.*, vol. 23, pp. 75–86, 1982.

[10] D. W. Tank and J. J. Hopfield, "Simple 'neural' optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit," *IEEE Trans. Circuits Syst.*, vol. CAS-33, no. 5, pp. 533–541, 1986.

[11] L. Cheng, Z. G. Hou, and M. Tan, "A delayed projection neural network for solving linear variational inequalities," *IEEE Trans. Neural Netw.*, vol. 20, no. 6, pp. 915–925, 2009.

[12] X. B. Gao and L. Z. Liao, "A new projection-based neural network for constrained variational inequalities," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 373–388, 2009.

[13] X. L. Hu and J. Wang, "A recurrent neural network for solving a class of general variational inequalities," *IEEE Trans. Syst. Man Cybern. B—Cybern.*, vol. 37, no. 3, pp. 528–539, 2007.

[14] S. Effati and M. Ranjbar, "A novel recurrent nonlinear neural network for solving quadratic programming problems," *Applied Mathematical Modelling*, vol. 35, no. 4, pp. 1688–1695, 2011.

[15] Y. Q. Yang and J. D. Cao, "Solving quadratic programming problems by delayed projection neural network," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1630–1634, 2006.

[16] Y. S. Xia, "An extended projection neural network for constrained optimization," *Neural Computation*, vol. 16, no. 4, pp. 863–883, 2004.

[17] S. W. Zhang and A. G. Constantinides, "Lagrange programming neural networks," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 39, no. 7, pp. 441–452, 1992.

[18] Y. S. Xia and J. Wang, "A general methodology for designing globally convergent optimization neural networks," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1331–1343, 1998.

[19] Y. S. Xia, "Global convergence analysis of Lagrangian networks," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 50, no. 6, pp. 818–822, 2003.

[20] Y. S. Xia and G. Feng, "On convergence conditions of an extended projection neural network," *Neural Comput.*, vol. 17, no. 3, pp. 515–525, 2005.

[21] X. B. Gao and L. Z. Liao, "A neural network for monotone variational inequalities with linear constraints," *Phys. Lett. A*, vol. 307, no. 2–3, pp. 118–128, 2003.

[22] X. B. Gao, "A novel neural network for nonlinear convex programming," *IEEE Trans. Neural Netw.*, vol. 15, no. 3, pp. 613–621, 2004.

[23] H. L. Wang and O. K. Ersoy, "A novel evolutionary global optimization algorithm and its application in bioinformatics," Dept. Electr. Comput. Eng., Purdue Univ., ECE Technical Reports, Aug. 2005.

[24] M. Mizukami, M. Hirano, and K. Shinjo, "Simultaneous alignment of multiple optical axes in a multistage optical system using Hamiltonian algorithm," *Opt. Eng.*, vol. 40, no. 3, pp. 448–454, 2001.

[25] T. Okamoto and E. Aiyoshi, "Global optimization using a synchronization of multiple search points autonomously driven by a chaotic dynamic model," *J. Global Optimization*, vol. 41, no. 2, pp. 219–244, 2008.

[26] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA: Addison-Wesley, 1989.

[27] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Networks*, 1995, pp. 1942–1948.

[28] J. A. K. Suykens, J. Vandewalle, and B. De Moor, "Intelligence and cooperative search by coupled local minimizers," *Int. J. Bifurcation and Chaos*, vol. 11, no. 8, pp. 2133–2144, 2001.

[29] A. Teughels, G. De Roeck, and J. A. K. Suykens, "Global optimization by coupled local minimizers and its application to FE model updating," *Comput. Struct.*, vol. 81, no. 24–25, pp. 2337–2351, 2003.

[30] S. Xavier-de-Souza, J. A. K. Suykens, J. Vandewalle, and D. Bollé, "Coupled simulated annealing," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 40, no. 2, pp. 320–335, 2010.

[31] J. A. K. Suykens and J. Vandewalle, "Coupled local minimizers: Alternative formulations and extensions," in *Proc. World Congr. Computational Intelligence—Int. Joint Conf. Neural Networks (WCCI-IJCNN 2002)*, Honolulu, HI, May 2002, pp. 2039–2043.

[32] J. A. K. Suykens, M. E. Yalçin, and J. Vandewalle, "Coupled chaotic simulated annealing processes," in *Proc. 2003 IEEE Int. Symp. Circuits and Systems (ISCAS)*, Bangkok, Thailand, May 2003, pp. III-582–III-585.

[33] S. Gunel, J. A. K. Suykens, and J. Vandewalle, "Global optimization with coupled local minimizers excited by Gaussian white noise," in *Proc. 2006 Int. Symp. Nonlinear Theory and Its Applications (NOLTA2006)*, Bologna, Italy, Sep. 2006, pp. 743–746.

[34] J. M. Ortega and W. G. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic, 1970.

[35] X. L. Hu and J. Wang, "Convergence of a recurrent neural network for nonconvex optimization based on an augmented lagrangian function," in *Proc. Advances in Neural Networks—ISNN 2007, Pt 3*, 2007, vol. 4493, pp. 194–203.

[36] C. W. Wu, "Synchronization in networks of nonlinear dynamical systems coupled via a directed graph," *Nonlinearity*, vol. 18, no. 3, pp. 1057–1064, 2005.

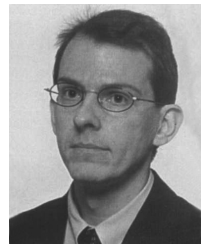
[37] Q. S. Liu, J. D. Cao, and Y. S. Xia, "A delayed neural network for solving linear projection equations and its analysis," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 834–843, 2005.

[38] Z. G. Hou, "A hierarchical optimization neural network for large-scale dynamic systems," *Automatica*, vol. 37, no. 12, pp. 1931–1940, 2001.

[39] P. A. Cook, *Nonlinear Dynamic Systems*. London, U.K.: Prentice-Hall International, 1986.

Xuyang Lou received the Ph.D. degree in control theory and control engineering from Jiangnan University, Wuxi, China, in 2009.

In 2010, he joined the School of Communication and Control Engineering, Jiangnan University. From October 2007 to October 2008, he was a visiting scholar in the CSIRO Division of Mathematical and Information Sciences, Adelaide, Australia. From July to November 2010, he was a Postdoctoral Fellow in the Department of Electrical Engineering (ESAT-SCD/SISTA) of the Katholieke Universiteit Leuven. He is currently an Associate Professor with Jiangnan University. His current research interests include computational intelligence, hybrid systems, and biological neural networks.



Johan A. K. Suykens (SM'05) was born in Willebroek, Belgium, on May 18, 1966. He received the M.Sc. degree in electromechanical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1989 and 1995, respectively.

He has been a Postdoctoral Researcher with the Research Foundation-Flanders (FWO-Vlaanderen) and is currently a Professor (Hoogleraar) with the Katholieke Universiteit Leuven. He is the author of the books *Artificial Neural Networks for Modeling and Control of Nonlinear Systems* (Kluwer, 1996) and *Least Squares Support Vector Machines* (World Scientific, 2002), a coauthor of the book *Cellular Neural Networks, Multiscroll Chaos and Synchronization* (World Scientific, 2005), and the Editor of the books *Nonlinear Modeling: Advanced Black-Box Techniques* (Kluwer, 1998) and *Advances in Learning Theory: Methods, Models and Applications* (IOS Press, 2003).

Dr. Suykens is a recipient of the Best Paper (Senior) Award from the IEEE Signal Processing Society in 1999, the Young Investigator Award from the International Neural Networks Society in 2000 for his significant contributions in neural networks, and several Best Paper Awards at international conferences. He was the Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS in 1997–1999 and 2004–2007, and since 1998, he has been an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS. He was the Director and an Organizer of the North Atlantic Treaty Organization Advanced Study Institute on Learning Theory and Practice (Leuven, 2002), a Program Committee Cochair of the 2004 International Joint Conference on Neural Networks and the 2005 International Symposium on Nonlinear Theory and Its Applications, an organizer of the International Symposium on Synchronization in Complex Networks 2007 and a co-organizer of the NIPS 2010 workshop on Tensors, Kernels and Machine Learning. He has been recently awarded an ERC Advanced Grant 2011.