Hybrid Coupled Local Minimizers

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

***Abstract*—This paper proposes an improved global optimization technique, named Hybrid Coupled Local Minimizers (HCLM), which is inspired on the method of coupled local minimizers (CLM). The HCLM method uses a set of search points, called "*particles*", initially spread over the search space, that are occasionally *impulsively coupled*, instead of permanently coupled. This approach leads to a much better optimization efficiency, because it combines the fast convergence (due to a Newton-based method that is used for each particle) with the capability of global optimization (resulting from the hybrid interaction which enables a parallel strategy). Such parallel and distributed computing process embedded with a trust region strategy relies on the hybrid interconnections of particles rather than individual particles, which is able to fully exploit the parallel nature of the computation and produce results which are more globally optimal. It is worth mentioning that the stability of each particle and the synchronization of all particles are also derived and proved by means of the *contraction theory*. The HCLM method is illustrated on several test functions with many local minima and applied to a problem of static nonlinear regression with multilayer perceptrons (MLPs) from given noisy measurement data.**

***Index Terms*—Global optimization, hybrid coupled local minimizers, Lagrange programming network, Newton-based trust region method, pulse-coupling.**

I. Introduction

SEVERAL practical design problems arising in many fields of engineering, biology, physics, chemistry and economics, can be formulated as an unconstrained optimization problem. Many of them (e.g., in learning, classification, identification, control etc.) are related to the minimization of an objective function with several local optima.

In order to obtain a locally optimal solution, there exist many well-developed numerical methods, such as steepest descent, conjugate gradient, Newton's, quasi-Newton and Levenberg-Marquardt [1]–[3]. Most of these approaches have a satisfactory convergence rate, but they are local optimization methods and possess a poor capability for 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 [4]. The past decades have witnessed a surge of development of global optimization methods that avoid being trapped in bad local minima. Examples of these methods include the genetic algorithm [5], and particle swarm optimization [6]. Nevertheless, 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.

The global optimization method that we propose in this paper is targeted at multimodal and multidimensional, but smooth problems with several local optima. The working principle of HCLM method is inspired on the CLM method originally conceived by Suykens *et al.* [7], which considers minimizing the average energy cost of the ensemble, subject to synchronization constraints between the state vectors of the individual search points. The CLM 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 in their work. Recently, Teughels *et al.* [8] has successfully applied the method to finite element model updating using experimental modal data, and shown that the global minimum is expected to be found more easily because of the simultaneous searching and cooperative behavior among the multiple particles. Very recently, the method has been combined with simulated annealing and shown an excellent reduction in the sensitivity to initialization parameters compared to the uncoupled cases [9]. Afterwards, we proved global asymptotical stability of the array of CLM within the Lagrange programming network framework and the stability of synchronization between any two minimizers via a directed graph method [10].

Here, we do not impose interactions between particles through the whole search process, but instead we consider a network-like coupling over all the particles in a *hybrid* mode. During the search process, the function and derivative values of each particle are first calculated. Then a local search implementation based on the Newton-based trust region method

Manuscript received February 03, 2013; revised April 27, 2013; accepted June 04, 2013. Date of publication August 23, 2013; date of current version January 24, 2014. This work is 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 & 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; ERC AdG A-DATADRIVE-B; Other: Helmholtz: viCERP, ACCM. This work is also partially supported by National Natural Science Foundation of China (61174021, 61104155), the Fundamental Research Funds for the Central Universities (JUSRP51322B) and the 111 Project (B12018). This paper was recommended by Associate Editor M. Frasca.

X. Lou is with the Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), 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).

Digital Object Identifier 10.1109/TCSI.2013.2278352

1549-8328 © 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

is performed for each particle so that every particle reaches a local minimum (i.e., a multi-start local optimization process) with a high convergence speed. After that, based on the function costs of the particles, the set of optimizers are coupled in order to create interaction between them, which results in a cooperative search mechanism. The coupling strength is *gradually* increased until some particle is escaping from its local minimum. Subsequently, the interactions among particles is removed and the search process is repeated. The contribution of this paper is four-fold:

- First, we present a new global optimization method called hybrid coupled local minimizers (HCLM). This method is composed of several parallel search processes with adaptive weighting coefficients and a hybrid coupling strength. It is more suitable for a parallel implementation than some other schemes with permanent coupling. Such parallel and distributed computing process relies on the hybrid interconnections of particles rather than individual particles, which is able to fully exploit the parallel nature of the computation, thereby significantly reducing the time required for each run.
- Second, we embed the Newton-based trust region method [12] in the parallel independent search process for particles, to realize a faster convergence instead of the steepest descent method used in the original CLM method.
- Third, we formulate the whole search process with a set of state-dependent impulsive control systems (or rather, hybrid dynamical systems [11]). In other words, the interactions among particles are only performed while all particles are not synchronized and are converging to different local minima. The coupling term consists of the particle states, the coupling strength and adaptive weighting coefficients which are functions of the current costs of all individual particles. Comparing to the original CLM scheme, in the framework of HCLM, there is no need to transform the unconstrained optimization problem into a nonlinear optimization problem with linear equality constraints.
- Fourth, we prove the convergence of the proposed algorithm and the synchronization of all states of the particles using *contraction theory*. Under some mild conditions, the stability of synchronization between any two particles can be ensured. The results show that the states of the particles usually can be synchronized if the coupling strength and adaptive weighting coefficients are chosen properly.

This paper is organized as follows. In Section II, we review the basic principle of CLM that is relevant to this paper. In Section III, we propose the HCLM method and describe in detail the algorithm for HCLM. In Section IV, we analyze the local stability of the HCLM and come up with a sufficient condition for the synchronization of HCLM ensemble. In Section V, we perform experiments on two multimodal functions to demonstrate the performance of the HCLM algorithm and compare the optimization results of HCLM with two projection-based neural network methods and the standard particle swarm optimization method on several benchmark functions. In Section VI, we apply the HCLM method to the optimization of training neural networks on a problem of static nonlinear regression with MLPs. Finally, in Section VII, we draw conclusions about the main contributions.

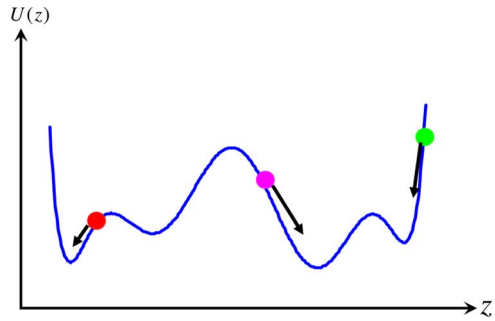


Fig. 1. Multi-start local optimization: No interaction.

Throughout this paper, we use the following notations. $\mathbb{R}^n$ denotes the n-dimensional real space; let $\mathbb{R}_+ = [0, \infty), \mathbb{Z}_+ = \{1, 2, \ldots\}$. $\mathbb{N}_q \triangleq \{1, 2, \cdots, q\}$. $\mathbf{z}^\top$ ($A^\top$, 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. $\bar{\lambda}(\cdot)$ denotes the maximal eigenvalues of a matrix, and $\bar{\sigma}(\cdot)$ denotes the maximal singular values of a matrix. $\mathbf{I}_n$ represents n-dimension identity matrix. $\|\mathbf{z}\|$ is the Euclidean vector 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 representing 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 time continuously differentiable functions. "$\otimes$" and "$\circ$" denote the Kronecker product and Hadamard product, respectively.

II. Coupled Local Minimizers (CLM)

A. CLM Principle

The basic mechanism of CLM [7] is illustrated in Fig. 1 for a multimodal problem with three particles that are searching for the global minimum. Instead of performing the search processes independently, i.e., by multi-start local optimization (Fig. 1), the particles are coupled to each other in order to create an interaction (Fig. 2). This method can be formulated as a set of coupled Lagrange programming networks whose states exchange information via the coupling. As a result, the search space is in fact duplicated and the local optimizers are exchanging information through the interaction scheme. Such a coupling mechanism results in a better performance than multi-start local optimization.

B. CLM Algorithm

Consider the following unconstrained problem of minimization of a twice continuously differentiable 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:

$$\dot{\mathbf{z}} = -\nu \nabla_{\mathbf{z}} U(\mathbf{z}) \tag{2}$$

with $\nu > 0$ the step size. Alternatively, using the CLM scheme in [7], one aims at minimizing the average energy cost

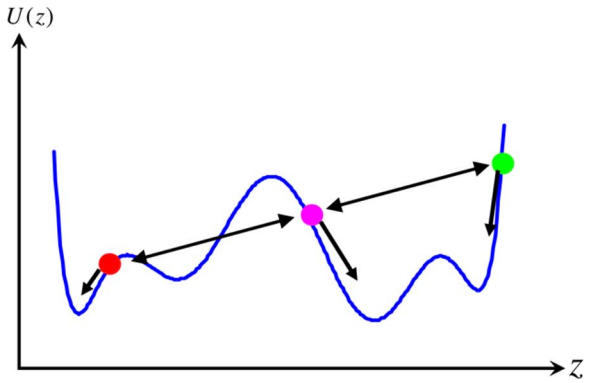


Fig. 2. Coupled local minimizers (CLM): Interaction and information exchange.

$(1)/(q)\sum_{i=1}^{q} U(\mathbf{z}^{(i)})$ of an ensemble consisting of q particles, subject to pairwise state equality constraints:

$$\min_{\mathbf{z}^{(i)}\in\mathbb{R}^n} \frac{1}{q}\sum_{i=1}^{q} U(\mathbf{z}^{(i)}) \quad \text{subject to } \mathbf{z}^{(i)} - \mathbf{z}^{(i+1)} = 0,\ i \in \mathbb{N}_q = \{1, 2, \ldots, q\}, \tag{3}$$

with states $\mathbf{z}^{(i)} \in \mathbb{R}^n$ and the boundary conditions $\mathbf{z}^{(0)} = \mathbf{z}^{(q)}, \mathbf{z}^{(q+1)} = \mathbf{z}^{(1)}$. For the q equality constraints, we introduce q Lagrange multipliers $\boldsymbol{\lambda}^{(i)} \in \mathbb{R}^n, i \in \mathbb{N}_q$ and give the augmented Lagrangian:

$$\mathcal{L}\left(\mathbf{z}^{(i)}, \boldsymbol{\lambda}^{(i)}\right) = \frac{\eta}{q}\sum_{i=1}^{q} U(\mathbf{z}^{(i)}) + \frac{1}{2}\sum_{i=1}^{q}\gamma_i \left\|\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right\|^2 + \sum_{i=1}^{q}\left\langle \boldsymbol{\lambda}^i, \left[\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right]\right\rangle, \tag{4}$$

where $\eta > 0$ represents 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.

From this augmented Lagrangian (4), one can derive the following Lagrange programming network [13]

$$\begin{cases} \dot{\mathbf{z}}^{(i)} = -\nabla_{\mathbf{z}^{(i)}}\mathcal{L}(\mathbf{z}^{(i)}, \boldsymbol{\lambda}^{(i)}), \\ \dot{\boldsymbol{\lambda}}^{(i)} = \nabla_{\boldsymbol{\lambda}^{(i)}}\mathcal{L}(\mathbf{z}^{(i)}, \boldsymbol{\lambda}^{(i)}), i \in \mathbb{N}_q. \end{cases} \tag{5}$$

This gives the following CLM array

$$\begin{cases} \dot{\mathbf{z}}^{(i)} = -\frac{\eta}{q}\nabla_{\mathbf{z}^{(i)}}U(\mathbf{z}^{(i)}) + \gamma_{i-1}\left[\mathbf{z}^{(i-1)} - \mathbf{z}^{(i)}\right] \\ \qquad -\gamma_i\left[\mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}\right] + \boldsymbol{\lambda}^{(i-1)} - \boldsymbol{\lambda}^{(i)}, \\ \dot{\boldsymbol{\lambda}}^{(i)} = \mathbf{z}^{(i)} - \mathbf{z}^{(i+1)}, i \in \mathbb{N}_q, \end{cases} \tag{6}$$

with learning rate η. In the simplest scheme that γ_i values are chosen to be constant. However, they have also been optimized for obtaining an optimal interaction in the ensemble, by rescheduling the values over time by solving linear programs [7].

III. Hybrid Coupled Local Minimizers (HCLM)

A. Model of HCLM

Instead of taking a single steepest descent gradient system (2), we consider now a type of CLM method where the particles are coupled in a hybrid mode according to the following equation

$$\begin{cases} \dot{\mathbf{z}}^{(i)} = -\frac{1}{q}\nabla_{\mathbf{z}^{(i)}}U\left(\mathbf{z}^{(i)}\right), & \mathbf{z} \notin \mathcal{M}, \\ \Delta\mathbf{z}^{(i)} = \rho(t)\sum\limits_{j\neq i} w_i(\mathbf{z})G_{ij}(\mathbf{z})\left(\mathbf{z}^{(j)} - \mathbf{z}^{(i)}\right), & \mathbf{z} \in \mathcal{M}, \end{cases} \tag{7}$$

where $\mathbf{z} = [(\mathbf{z}^{(1)})^\top \ldots (\mathbf{z}^{(q)})^\top]^\top \in \mathbb{R}^{nq}, \mathcal{M} = \{\mathbf{z} | \ \| \nabla_{\mathbf{z}}U(\mathbf{z})\| \leq \varepsilon\}, \varepsilon$ is a small enough constant, $\rho(t)$ is a coupling strength to be designed, $G_{ij}(\mathbf{z})$ is state-dependent and denotes the coupling weight between the ith and jth search point, and $w_i(\mathbf{z})$ is an adaptive weighting coefficient determined by the energy costs of particles. $\Delta\mathbf{z}^{(i)} := \Delta\mathbf{z}^{(i)}(t_k) = \mathbf{z}^{(i)}(t_k^+) - \mathbf{z}^{(i)}(t_k^-)$ is the control law in which $\mathbf{z}^{(i)}(t_k^+) = \lim_{t\to t_k^+}\mathbf{z}^{(i)}(t)$. Here, the time instants t_k are satisfying $0 < t_1 < t_2 < \cdots < t_k < \cdots$ for $k = 1, 2, \ldots$, and are representing the impulse instants for which $\mathbf{z} \in \mathcal{M}$. Without loss of generality, we assume that $\lim_{t\to t_k^-}\mathbf{z}^{(i)}(t) = \mathbf{z}^{(i)}(t_k)$, which means that the solution $\mathbf{z}^{(i)}(t)$ is continuous from the left.

Here, we take

$$G_{ij}(\mathbf{z}) = 1 + \frac{U\left(\mathbf{z}^{(i)}\right) - U\left(\mathbf{z}^{(j)}\right)}{U_{\max}(\mathbf{z}) - U_{\min}(\mathbf{z})}$$

and

$$w_i(\mathbf{z}) = \begin{cases} w_{\min} + \mathcal{W}(\mathbf{z}), & U(\mathbf{z}^{(i)}) \leq \mathbb{U}(\mathbf{z}), \\ w_{\max}, & U(\mathbf{z}^{(i)}) > \mathbb{U}(\mathbf{z}), \end{cases}$$

with

$$\mathcal{W}(\mathbf{z}) = (w_{\max} - w_{\min})\frac{U(\mathbf{z}^{(i)}) - U_{\min}(\mathbf{z})}{\mathbb{U}(\mathbf{z}) - U_{\min}(\mathbf{z})},$$

$U_{\min}(\mathbf{z}) = \min_{i\in\mathbb{N}_q} U(\mathbf{z}^{(i)})$, $U_{\max}(\mathbf{z}) = \max_{i\in\mathbb{N}_q} U(\mathbf{z}^{(i)})$, $\mathbb{U}(\mathbf{z}) = (1)/(q)\sum_{i=1}^{q} U(\mathbf{z}^{(i)})$. Note that $G_{ij}(\mathbf{z}) \in [0, 2]$ and $w_i(\mathbf{z}) \in [w_{\min}, w_{\max}]$ for all $i, j \in \mathbb{N}_q$. Here, $w_i(\mathbf{z})$ is regarded as a time-varying inertia weight inspired by the particle swarm optimization (PSO) [14]. In many literatures on PSO algorithms, the linearly varying inertia weight evolves according to the iteration numbers, due to the discrete iteration characteristic of particle swarm optimization. Differently, the inertia weight here depends on the energy costs of particles. However, the essential principle of the inertia weight evolution is the same. It is designed as a tradeoff between the global and local search. When the energy cost of each particle tends to the average energy cost of the ensemble, the inertia weight will be increased, being beneficial for precise local search for the current search area; when all particles are relatively decentralized, the inertia weight will be reduced, facilitating the global exploration. Linear change in weight makes the inertia lie between $w_{\min}$ and $w_{\max}$ which can be chosen by the empirical study. However, through empirical

studies, Shi and Eberhart have observed that the optimal solution can be improved by varying the value of $w_{\max} = 0.9$ at the beginning of the search to $w_{\min} = 0.4$ at the end of the search for most problems [15]. This method of self-adaptive changing inertia weight is effective to improve the global and local exploration capabilities. Among our later simulations in Sections V and VI, we simply take $w_{\min} = 0.6$ and $w_{\max} = 0.9$. It is worth pointing out that there is no need to calculate $G_{ij}(\mathbf{z})$ and $w_i(\mathbf{z})$ all the time, since we only use them when $\mathbf{z} \in \mathcal{M}$. This is also addressed in the subsequent algorithm.

One can also rewrite (7) into the compact vector form

$$\dot{\mathbf{z}} = -\frac{\varrho}{q}\nabla_{\mathbf{z}} U(\mathbf{z}) + \bar{\rho}(t)\bar{G}(\mathbf{z})\mathbf{z}, \tag{8}$$

where

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}^{(1)} \\ \vdots \\ \mathbf{z}^{(q)} \end{bmatrix} \in \mathbb{R}^{nq},$$

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

$$\bar{G}(\mathbf{z}) = [W(\mathbf{z}) \circ G(\mathbf{z})] \otimes \mathbf{I}_n \in \mathbb{R}^{nq \times nq},$$

$$W(\mathbf{z}) = \begin{bmatrix} w_1(\mathbf{z}) & w_1(\mathbf{z}) & \cdots & w_1(\mathbf{z}) \\ w_2(\mathbf{z}) & w_2(\mathbf{z}) & \cdots & w_2(\mathbf{z}) \\ \vdots & \vdots & \vdots & \vdots \\ w_q(\mathbf{z}) & w_q(\mathbf{z}) & \cdots & w_q(\mathbf{z}) \end{bmatrix}_{q \times q},$$

$$G(\mathbf{z}) = \begin{bmatrix} \mathcal{G}_1(\mathbf{z}) & G_{12}(\mathbf{z}) & \cdots & G_{1q}(\mathbf{z}) \\ G_{21}(\mathbf{z}) & \mathcal{G}_2(\mathbf{z}) & \cdots & G_{2q}(\mathbf{z}) \\ \vdots & \vdots & \vdots & \vdots \\ G_{q1}(\mathbf{z}) & G_{q2}(\mathbf{z}) & \cdots & \mathcal{G}_q(\mathbf{z}) \end{bmatrix}_{q \times q},$$

$$\mathcal{G}_r(\mathbf{z}) = -\sum_{j \neq r} G_{rj}(\mathbf{z}), r = 1, 2, \ldots, q,$$

$$\varrho = \begin{cases} 1, & \mathbf{z} \notin \mathcal{M}, \\ 0, & \mathbf{z} \in \mathcal{M}, \end{cases}$$

$$\bar{\rho}(t) = \begin{cases} 0, & \mathbf{z} \notin \mathcal{M}, \\ \rho(t), & \mathbf{z} \in \mathcal{M}. \end{cases}$$

In principle there are two key differences between CLM and HCLM:

- First, the model of coupled Lagrange programming networks for the CLM is derived from the augmented Lagrangian method, while the model for HCLM is formulated as a set of impulsive control systems [16]–[19], or hybrid dynamical systems [11]. The HCLM consists of two parts: interactions are imposed when the ensemble of particles are converging to be stable and pure gradient information is performed otherwise.
- Second, the coupling actions in CLM are imposed through the whole search process, while in HCLM such actions only happen when the function gradients of the search points are approximately stationary. The gradient information does not help the particles escape from the current optimum if it's local. Hence, an impulsive coupling strategy is developed for bringing all states close to each other asymptotically. It is worth pointing out that during the parallel independent search process for particles, the Newton-based trust region method is meant for realizing a faster convergence instead of the steepest descent method used in the original CLM method.

Note that in HCLM the coupling term consists of all states, the coupling strength and an adaptive weighting coefficient which is a function of the current costs of all individual search points. From the definition of the weighting coefficient $w_i(\mathbf{z})(i \in \mathbb{N}_q)$, the larger the current cost of the particle i is, the stronger the coupling strength within its coupling term. Under this HCLM framework, each particle has information about the status of the costs of all others. The hybrid coupling term $\bar{\rho}(t)\bar{G}(\mathbf{z})\mathbf{z}$ defines how this information is used to bring the states close to each other and steer the optimization toward the global optimum. Moreover, in order to maximize the convergence, the Newton-based (trust region) method is independently applied most of the time.

B. HCLM Algorithm

Here, we present an impulsively assigned strategy for the coupling strength $\rho(t)$ and the algorithm of the HCLM scheme. The main idea is to separate the search process into two *phases*. One is the local gradient-based search process based on the Newton-based trust region method. The other is a pulse-coupling process. After the first phase, the particles are located in different local minima and a pulse-coupling action will be imposed upon the particles. With this action, they are enforced to do some explorations by adjusting $\rho(t)$ and $\bar{G}(\mathbf{z})$ values which depend on the temporary minima of all particles. Hence, this method is called pulse-coupling searching. Such a searching scheme contains *two exquisite features*: (i) one is to explore unknown regions better; (ii) the other is to hold more knowledge of low-minimum regions of the cost function. The first feature will be realized by adjusting the $w_i(\mathbf{z})(i \in \mathbb{N}_q)$ values online. The second feature will be achieved by adaptively adjusting the coupling weight matrix $\bar{G}(\mathbf{z})$ depending on the current costs of the particles. In this way, the coupling term of each particle can be weighted with the temporary minima of the particles, such that those particles with low temporary energy cost have weak coupling strengths with the other particles, while others have relatively strong coupling strengths with their neighbors. Therefore, the knowledge of low-minimum regions of the cost function can be inherited.

In order to impose pulse-coupling actions on the particles and make sure that they have the trend of being synchronized, the coupling coefficient $\rho(t)$ should be assigned online to control the overall coupling term. It will be kept at zero when it is not time to impose a pulse-coupling action. The complete search algorithm can now be stated as follows:

Algorithm 1

1) *Initialization:* Set the number q of HCLM. Assign random and meaningful distributed initial starting points to $\mathbf{z}^{(i)}(0)(i \in \mathbb{N}_q)$. Initially, take the coupling coefficient $\rho(t) = 0$, so that the model (8) of HCLM exhibits uncoupled motions. Set $\Delta\rho = 0.1$ or other smaller positive constant.

2) *Generate* solutions by use of the Newton-based trust region method for minimizing $U(\mathbf{z}^{(i)})$ with respect to $\mathbf{z}^{(i)}(i \in \mathbb{N}_q)$, until all HCLM particles reach some local minima (that is, $\mathbf{z} \in \mathcal{M}$). More specifically, in each subiteration k, a quadratic approximation $m(\mathbf{p})$ of $U(\mathbf{z}^{(i)})$ at the current point $\mathbf{z}_k^{(i)}$ has to be minimized within a trust region Δ_k. The quadratic model $m(\mathbf{p})$ is usually defined as a quadratic function

$$\begin{aligned} \min_{\mathbf{p}} \quad & m(\mathbf{p}) = U\left(\mathbf{z}^{(i)}\right) + \left[\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})\right]^{\top} \mathbf{p} \\ & + \frac{1}{2}\mathbf{p}^{\top}\left[\nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})\right]\mathbf{p}, \\ \text{such that} \quad & \|\mathbf{p}\| \le \Delta, \end{aligned}$$

where $i \in \mathbb{N}_q$, $\mathbf{p}$ denotes a step-vector from $\mathbf{z}_k^{(i)}$, and $U(\mathbf{z}^{(i)})$, $\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})$ and $\nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})$ are the values of the function, the gradient and the Hessian of $U(\mathbf{z}^{(i)})$ at $\mathbf{z}_k^{(i)}$ respectively. If all of them enter the same valley (that is, both the variance of $[U(\mathbf{z}^{(1)}), U(\mathbf{z}^{(2)}), \cdots, U(\mathbf{z}^{(q)})]$ and $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\|$ for all $i, j \in N_q$ are small enough), *stop*; otherwise, go to next step.

3) *Assess* the costs $U(\mathbf{z}^{(i)})$ for all $i \in \mathbb{N}_q$, calculate $U_{\min}(\mathbf{z}), U_{\max}(\mathbf{z}), \mathbb{U}(\mathbf{z})$, and substitute them into $W(\mathbf{z})$ and $G(\mathbf{z})$. There may occur some cases that all HCLM particles are located in different local minima but have identical energy cost. For such cases, we endow $W(\mathbf{z})$ and $G(\mathbf{z})$ with random values that are randomly chosen in the interval [0, 1] to avoid meaningless denominators in $W(\mathbf{z})$ and $G(\mathbf{z})$. Then take $\rho(t) = 0.1$ to impose a pulse-coupling action on HCLM.

4) *Generate* a new solution by solving (8) in one iteration, and evaluate the function costs of this new solution. If none of the HCLM particles escape from their local minima, repeat to let $\rho(t) := \rho(t) + \Delta\rho$ and regenerate a new solution and reevaluate whether some HCLM particle escapes from its local minimum or not; *otherwise*, substitute the obtained control parameter $\rho(t)$ into (8) and activate the coupling term for another iteration, accept the obtained new solution and then invalidate the coupling term by letting $\rho(t) = 0$ again, go to next step.

5) *Stop* if all HCLM particles synchronize to the same minimum or the stopping criterion is met; *otherwise*, go to *Step 2*.

Remark 1: As a Newton-based trust region strategy is used when there is no pulse-coupling action, the search process in the HCLM algorithm has a high convergence rate. Moreover, the strategy of increasing $\rho(t)$ gradually helps those HCLM particles with high energy costs escape from local minima, do more explorations and move to the regions where other HCLM particles are located. In this way, all HCLM particles will be guided to the global minimum or a quasioptimal point. The HCLM algorithm is implemented in the optimization toolbox of MATLAB [19]. The Newton-based trust region method is applied by means of the command *fminunc*, for which the 'Trust Region' option is chosen.

Remark 2: Due to the dynamic nature and the convenient conversion of an optimization problem into a dynamical system, at present, several neural networks have been developed to solve optimization problems [20]–[23]. Though the fruitful achievements of neural network approaches in various kinds of optimization problems and related applications, most of them can be regarded as local optimization techniques and lack the capability of global optimization. Different from neural network approaches, the HCLM method in this paper is based on a cooperative search mechanism realised by minimizing the average cost of the population. Thanks to the coupling mechanism, the particles are able to exchange information which results in a better performance than multistart local optimization. Therefore, a global optimal solution of the original problem is more likely to be obtained. The comparisons of HCLM over several neural network approaches will be further illustrated in the next section.

IV. Stability and Synchronization

A. Local Stability

Let us first present local stability of the HCLM system (7).

Theorem 1: Let $\mathbf{z}_i^*$ be a stationary point of $U(\mathbf{z}^{(i)}), i \in \mathbb{N}_q$. Then for each particle $\mathbf{z}^{(i)}(i \in \mathbb{N}_q)$, the HCLM system (7) is locally asymptotically stable at $\mathbf{z}_i^*(i \in \mathbb{N}_q)$.

Proof: Since $\mathbf{z} \in \mathcal{M} = \{\mathbf{z} | \ \|\nabla_{\mathbf{z}} U(\mathbf{z})\| \le \varepsilon\}$ only happens in some instances and the set $\mathcal{M}$ equals to $\mathbf{z}^* = [(\mathbf{z}_1^*)^{\top} \cdots (\mathbf{z}_q^*)^{\top}]^{\top} \in \mathbb{R}^{nq}$ when $\varepsilon \to 0$, in order to ensure the local stability, we only need to consider the stability of the HCLM system (7) at $\mathbf{z} \notin \mathcal{M}$. Note that in this case, the HCLM system (7) performs the search process along the negative gradient direction. Therefore, each particle $\mathbf{z}^{(i)}$ starting with an arbitrary initial point will converge to its local minimizer $\mathbf{z}_i^*$. In other words, $\mathbf{z}$ always locally converges to $\mathbf{z}^*$. The proof is completed. □

B. Synchronization

To derive a common local minimum, all particles in the HCLM scheme should converge to a common point so it is also necessary to analyze the synchronization of the HCLM ensemble, that is, the stability of synchronization between $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}(i \ne j, i, j \in \mathbb{N}_q)$.

In HCLM, the interactions among particles are only performed while all particles are not synchronized and are converging to different local minima, that is, the event $\mathbf{z} \in \mathcal{M}$ only happens at some special time instants t_k satisfying $0 < t_1 < t_2 < \cdots < t_k < \cdots$ for $k = 1, 2, \cdots$, and representing the impulse instants for which $\mathbf{z} \in \mathcal{M}$. Therefore, according to the definition of $\Delta\mathbf{z}^{(i)}$, the HCLM system (7) can be equivalently described as a resetting hybrid system:

$$\begin{cases} \dot{\mathbf{z}}^{(i)} = -\frac{1}{q}\nabla_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}), & t \ne t_k, \\ \mathbf{z}^{(i)}(t^+) = \mathbf{z}^{(i)}(t) + \rho(t) \sum\limits_{j \ne i} w_i(\mathbf{z}) & \\ \qquad \times G_{ij}(\mathbf{z})(\mathbf{z}^{(j)} - \mathbf{z}^{(i)}), & t = t_k, \end{cases} \tag{9}$$

where $t_k = \inf\{t | \|\nabla_{\mathbf{z}(t)} U(\mathbf{z}(t))\| \le \varepsilon, t > t_{k-1}\}$.

Before presenting our main results, some important results about contraction theory which will be used later are briefly reviewed, details of which can be found in [24], [25].

Definition 1 [25]: A resetting hybrid system is defined by the equations

$$\begin{cases} \dot{\mathbf{x}} = f(\mathbf{x}(t), t), & t \neq t_k, \\ \mathbf{x}(t^+) = h(\mathbf{x}(t), t), & t = t_k, \\ k(t^+) = I(\mathbf{x}(t), t, k(t)), \end{cases} \tag{10}$$

where $t \in \mathbb{R}_+, \mathbf{x} \in \mathbb{R}^n$ is the state vector, f and h are two $n \times 1$ vector functions, $k \in \mathbb{Z}_+$ is a piecewise constant signal, which is the resetting index. Assume that $f(\mathbf{x}(t), t)$ and $h(\mathbf{x}(t), t)$ are continuously differentiable.

A resetting hybrid system given by Definition 1 is associated with differential dynamics

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{x}(t) = J_c(\mathbf{x}(t), t)\delta\mathbf{x}(t), & t \neq t_k, \\ \delta\mathbf{x}(t^+) = J_d(\mathbf{x}(t), t)\delta\mathbf{x}(t), & t \neq t_k, \end{cases} \tag{11}$$

where $J_c(\mathbf{x}(t), t) = (\partial f(\mathbf{x}(t), t))/(\partial \mathbf{x})$ is the Jacobian of the vector field f, $J_d(\mathbf{x}(t), t) = (\partial h(\mathbf{x}(t), t))/(\partial \mathbf{x})$ is the Jacobian of the vector field h and t_j is the jth resetting time.

Definition 2 [25]: A resetting hybrid system of Definition 1 is said to be contracting if and only if the associated differential dynamics of (11) is uniformly exponential stable. Then all solutions associated with each resetting sequence $\{(t_k, h(t_k))\}$ will converge exponentially to a corresponding particular trajectory.

Assumption 1 [25]: For a resetting hybrid system, the set of reset R associated with a resetting sequence $\{(t_k, h(t_k))\}$ is infinitely countable and $\exists \epsilon > 0$ such that $\epsilon \leq t_k - t_{k-1} < +\infty, \forall k$.

Let $\bar{\sigma}(H) \leq \beta$ for some constant β where $H = \Theta_d J_d \Theta_d^{-1}$ is the generalized Jacobian associated with resetting dynamics $h(\mathbf{x}(t), t)$ in a metric $M_d = \Theta_d^{'}\Theta_d$ which is a symmetric positive-definite matrix and specifies the metric space, where $\Theta_d^{'}$ is the Hermitian matrix of matrix Θ_d. While $\bar{\lambda}(F_s) \leq \alpha$, where $F_s = (F^\top + F)/2$ is the symmetric part of the generalized Jacobian $F = (\dot{\Theta}_c + \Theta_c J_c)\Theta_c^{-1}$ associated with $f(\mathbf{x}(t), t)$ in the metric $M_c = \Theta_c^{'}\Theta_c$ which is also a symmetric positive-definite matrix and specifies the metric space, where $\Theta_c^{'}$ the Hermitian matrix of matrix Θ_c. Also, Δt_k denotes the period between two resets following the kth reset. Let $k^*(t)$ be the number of resets for a given system by time t such that $k^*(t) \to +\infty$ as $t \to +\infty$. We define the average dwell-time of a resetting system as a positive scalar function $\tau(t)$ such that

$$\tau(t) = \frac{\sum_{k=1}^{k^*(t)} \Delta t_k}{k^*(t)}. \tag{12}$$

Lemma 1 [25]: A resetting hybrid system is contracting if there exist constant metrics that are equal at the resetting times, which means that $\Theta_c(t_k) = \Theta_d(t_k), \forall k$, where t_k is the kth resetting time, and $\varpi < 0$ such that one of the following holds:

1) $\alpha + \ln \beta / \Delta t_k \leq \varpi, \forall k$;
2) there exists an average dwell-time $\tau(t) > 0$ such that $\alpha + \ln \beta / \tau(t) \leq \varpi, \forall t \geq t_0$.

Now we are ready to present out results.

Theorem 2: If one of the following holds:

(C1) $\bar{\alpha} + \ln \bar{\beta} / \Delta t_k < 0, \forall k$;

(C2) there exists an average dwell-time $\tau(t) > 0$ such that $\bar{\alpha} + \ln \bar{\beta} / \tau(t) < 0, \forall t \geq t_0$,

where $\bar{\alpha}$ is some constant satisfying $\max_{i \in \mathbb{N}_q}\{\bar{\lambda}(F_i)\} \leq \bar{\alpha}$, where $F_i = -[(\nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)}))^\top + \nabla^2_{\mathbf{z}^{(i)}} U(\mathbf{z}^{(i)})]/2$; $\bar{\beta}$ is some constant satisfying $\max_{i \in \mathbb{N}_q}\{\bar{\sigma}(H_i)\} \leq \bar{\beta}$, where H_i is the generalized Jacobian associated with resetting dynamics in (7). Then the HCLM system (7) 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.

Proof: According to Definition 2, if the resetting hybrid system (9) is contracting with respect to $\mathbf{z}$, then all trajectories of the particles will converge to a corresponding particular trajectory regardless of the initial conditions. Applying Lemma 1 and using the identity as a metric for both M_c and M_d yield that if one of the two conditions (C1) and (C2) is satisfied, the resetting hybrid system (9) is contracting with respect to $\mathbf{z}$. In other words, if any condition ((C1) or (C2)) in Theorem 1 is satisfied, all trajectories of the particles will converge to a corresponding particular trajectory

$$\begin{cases} \dot{\mathbf{z}}_\infty = -\frac{1}{q}\nabla_{\mathbf{z}_\infty} U(\mathbf{z}_\infty), & t \neq t_k, \\ \mathbf{z}_\infty(t^+) = \mathbf{z}_\infty(t), & t = t_k, \end{cases} \tag{13}$$

regardless of the initial conditions, which implies that the HCLM system (7) synchronizes in the sense that $\|\mathbf{z}^{(i)} - \mathbf{z}^{(j)}\| \to 0$ as $t \to \infty$ for all i, j. The proof is completed. □

Remark 3: Theorem 2 provides a condition of guaranteeing the stability of synchronization among the searching particles, but it is impossible to get the period Δt_k or the average dwell-time $\tau(t)$ in prior and inefficient to verify the conditions at every instant in practice. However, note that $G_{ij}(\mathbf{z})$ and $w_i(\mathbf{z})$ are limited, and thus if the coupling strength $\rho(t)$ is appropriately adjusted, the condition (C1) or (C2) will be satisfied. In fact, it is described in Algorithm 1 that at each pulse-coupling action on HCLM we choose $\rho(t)$ appropriately and try to draw the particles together.

V. Function Tests and Comparisons

We test the HCLM method introduced above on two simple multimodal functions and a suite of benchmark functions, and compare it with the projection-based neural network methods [20], [22].

A. Simple Multimodal Functions

Example 1: Consider the following one-dimensional nonconvex test function

$$\begin{aligned} \min \quad U_1(z) = &\frac{1}{50}z^8 + \frac{17}{500}z^7 - \frac{479}{500}z^6 - \frac{166}{125}z^5 \\ &+ \frac{1888}{125}z^4 + \frac{8261}{500}z^3 - \frac{43299}{500}z^2 - \frac{16479}{250}z. \end{aligned}$$

In Fig. 3(a), the test function with two initial search points is visualized. There are four local minima. One of them is the global minimum, located at $z^* = 1.925$. A HCLM run is carried out with $q = 2$ particles. Based on the HCLM algorithm above, we combine a trust region strategy and a hybrid coupling scheme to solve (8). Their initial values are $z^{(1)} = -4.0, z^{(2)} = 4.5$. It is each to check that Theorem 1

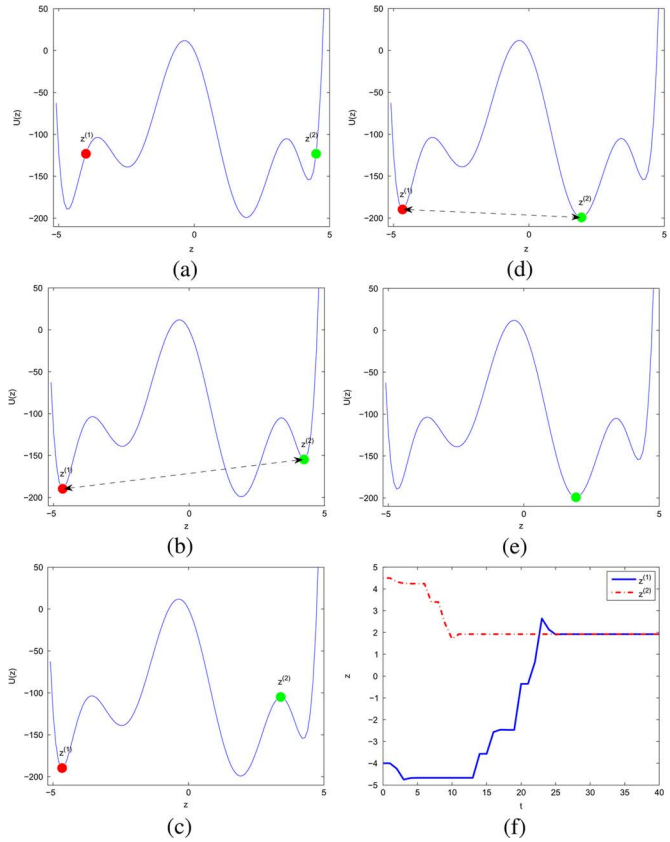


Fig. 3. Test function $U_1(z)$ with four local minima in Example 1 is illustrated. One of the minima is the global minimum at $z = 1.925$. Four snapshots of searching process for the function with two particles ($z^{(1)}$—red solid point, $z^{(2)}$—green solid point) are plotted. (a) Iteration$= 1$. (b) After 7 iterations, both particles converge to local minima and a pulse-coupling action is imposed with $\rho(t) = 52$. The double arrow dash line shows the coupling connection between $z^{(1)}$ and $z^{(2)}$. (c) At the 8th iteration, the second particle $z^{(2)}$ escapes from a local minimum and the pulse-coupling action is released. (d) After 14 iterations, $z^{(2)}$ converges to another valley. (e) Finally, both particles reach the global minimum after 28 iterations. (f) Convergence behavior of the two particles $z^{(1)}$ and $z^{(2)}$.

holds since any initial point will converge to a local optimum due to the negative gradient. Now, let us check the conditions of Theorem 2. By calculating the Hessian of the function, it is easy to verify that $\max_{i\in\mathbb{N}_q}\{\bar{\lambda}(F_i)\} \le \bar{\alpha} = 250$. The coupling weight matrix $G(\mathbf{z})$ can be obtained as $G(\mathbf{z}) = \begin{bmatrix} 0 & 0 \\ 2 & -2 \end{bmatrix}$, and thus $\max_{i\in\mathbb{N}_q}\{\bar{\sigma}(H_i)\} = 0$. Thereafter, as long as one takes a constant $\bar{\beta} < \exp(-\bar{\alpha}\Delta t_k)$ or $\bar{\beta} < \exp(-\bar{\alpha}\tau(t))$, the condition (C1) or (C2) in Theorem 2 can always be satisfied. It means that the two particles will finally be synchronized. Using the proposed search scheme (Algorithm 1), both minimizers end up in the global minimum (Fig. 3(e)) after 28 iterations. Fig. 3(b)–(e) shows four snapshots of the searching process at different iterations. Fig. 3(f) depicts the convergence behaviors of two particles $z^{(1)}$ and $z^{(2)}$. The nonsmoothness of the trajectories is due to several pulse-coupling actions.

Example 2: Consider the following optimization of cost function

$$\min U_2(\mathbf{z}) = \sum_{j=1}^{2} \left(z_j^4 - 16z_j^2 + 5z_j + 100\right). \qquad (14)$$

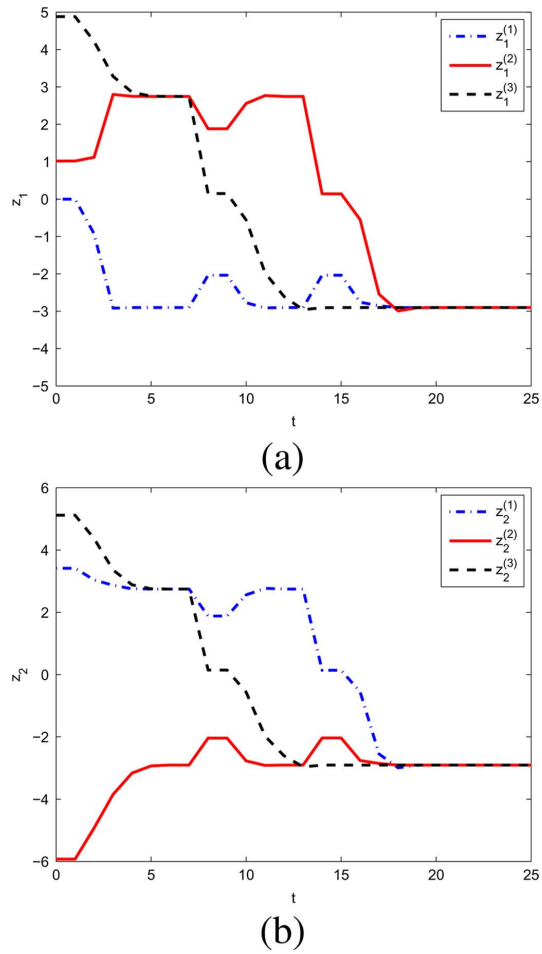


Fig. 4. (a) Convergence behavior of the decision variables z_1 with three particles. (b) Convergence behavior of the decision variables z_2 with three particles.

This is a four-hump camel back function with local minima (2.747, 2.747), (2.747, −2.904), (−2.904, 2.747) and (−2.904, −2.904). The global minimum for this problem is known and located at $\mathbf{z}^* = (-2.904, -2.904)$ with $U_2(\mathbf{z}^*) = 43.34$. Three initial states $(q = 3)$ have been random uniformly generated in $[-6\ 6]^2$. Perform an HCLM run using the proposed search scheme (Algorithm 1). All three minimizers end up in the global minimum. Fig. 4 shows the convergence behavior of two decision variables z_1 and z_2. The nonsmoothness of the trajectories is due to two pulse-coupling actions at the 8th and 14th iteration, respectively.

B. Benchmark Functions

Four benchmark functions listed in Table I (shown at the end of the paper) are used for the experimental tests. The selected functions are very often encountered in optimization algorithm benchmarks. All of them are complex multimodal functions with a large number of local minima.

To establish a reference for comparison with the HCLM algorithm in this paper, we performed experiments with all four test functions and compared the results of the classical multi-start trust-region method (multi-TR) and the projection neural network (PNN) algorithms described in [20], [22]. The results of

TABLE I
STANDARD BENCHMARK FUNCTIONS

No.	Function	Equation (U^* gives the minima)	Parameters
U_3	Griewank	$U_3(\mathbf{z}) = \frac{1}{4000}\sum_{i=1}^{n} z_i^2 - \prod_{i=1}^{n}\cos\left(\frac{z_i}{\sqrt{i}}\right) + 1,$ $U_3^*(0,0,\cdots,0) = 0$	$n = 10, 30, 100$ $z_i \in [-600, 600]$
U_4	Rastrigin	$U_4(\mathbf{z}) = \sum_{i=1}^{n}(z_i^2 - 10\cos(2\pi z_i) + 10),$ $U_4^*(0,0,\cdots,0) = 0$	$n = 10, 30, 100$ $z_i \in [-5.12, 5.12]$
U_5	Ackley	$U_5(\mathbf{z}) = -ae^{-b\sqrt{\frac{\sum_{i=1}^{n} z_i^2}{n}}} - e^{\frac{\sum_{i=1}^{n}\cos(cz_i)}{n}} + a + e^1,$ $a = 20, b = 0.2, c = 2\pi, U_5^*(0,0,\cdots,0) = 0$	$n = 10, 30, 100$ $z_i \in [-32.768, 32.768]$
U_6	Styblinski	$U_6(\mathbf{z}) = \frac{1}{2n}a\sum_{j=1}^{n} z_j^2 + 8n - 4n\sum_{i=1}^{2}\prod_{j=1}^{n}\cos(\omega_i z_j),$ $a = 0.01, \omega_1 = 0.2, \omega_2 = 1, U_6^*(0,0,\cdots,0) = 0$	$n = 10, 30, 100$ $z_i \in [-10, 10]$

the multi-TR and PNN algorithms are chosen from the best performance of multiple runs.

For each function, we performed 100 optimization runs in 10, 30 and 100 dimensions, respectively. Each optimization run was composed of ten parallel processes (i.e., $q = 10$) for 10 and 30 dimensions, and thirty parallel processes (i.e., $q = 30$) for 100 dimensions. Generally speaking, the number of q to achieve a good performance, depends on the complex shape of the function or typically on the number of local minima per volume in the search space [8]. For the HCLM method, the parallel processes are performed with hybrid couplings, whereas for the multi-TR and PNN methods, each process consists of an independent optimization run. For the latter two methods, all initialization conditions of search points were the same with those in HCLM method, and at the end of all runs, only the best result was used as the output. Finally, the average results for 100 optimization runs with the same number of the overall cost-function evaluations were presented. Table II (shown at the end of the paper) presents the average (ave.) and variance (var.) summaries of the final results from the algorithms. The benefit of utilizing the HCLM method is clear from the table. It displays a good performance on all the test functions. The best results of the experiments for each function are shown in boldface.

For the case of "$\text{dimension} = 10$", all results in the framework of HCLM are best among these algorithms. For the cases of "$\text{dimension} = 30$" and "$\text{dimension} = 100$", the results under the HCLM method are comparable to those results under the other methods except for the variance value of U_5. The reason why the variance value from testing function U_5 is so large, is due to the special feature of the Ackley's function which has one narrow global optimum basin with many different minor shallow local optima around. For the HCLM method, each run with different initial values and different coupling actions will easily lead to different local optima. In contrast, because of lacking the exploration ability, the multi-TR and PNN methods may easily become trapped in local optima far from the global minimum. While according to the feature of the Ackley's function, the areas far from the global optimum basin are roughly flat, and thereby such local optima are not quite different from each other resulting into the smaller variance value of U_5 based on the multi-TR and PNN methods. Meanwhile, it seems that the multi-TR method achieves better performance than our HCLM method on Griewank function U_3 and Styblinski function U_6. Note that both functions contain cosine term causing linkages among variables and making this function difficult to optimize. For such functions, it has been interestingly found that they are more difficult to optimize in lower than higher dimensions [26]. From these experiments, we also observed that the global minimum for all benchmark functions were derived for many times under the HCLM method through the 100 optimization runs.

VI. APPLICATION

In this section we discuss the application of HCLM to the problem of training of multilayer perceptron networks analyzed in [7]. A multilayer perceptron network is described as

$$y = \mathbf{w}^\top \tanh(V\mathbf{u} + \boldsymbol{\vartheta})$$

with input vector $\mathbf{u} \in \mathbb{R}^m$, output $y \in \mathbb{R}$, interconnection weights $\mathbf{w} \in \mathbb{R}^{n_h}, V \in \mathbb{R}^{n_h \times m}$ and bias vector $\boldsymbol{\vartheta} \in \mathbb{R}^{n_h}$, where n_h denotes the number of hidden units. Denote the unknown parameter vector as $\boldsymbol{\theta} = [w;\ V(:);\ \vartheta]$ which stacks all the weights into one single vector. Generally, the backpropagation algorithm for training of the multilayer perceptron networks is performed by optimizing the cost function

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{k=1}^{N}[d_k - y_k(\boldsymbol{\theta})]^2, \tag{15}$$

where d_k are the desired target outputs (with given data set $\{u_k, d_k\}_{k=1}^N$ of N training data).

Here, let us apply the HCLM scheme by taking as state vectors $\mathbf{z}^{(i)} (i \in \mathbb{N}_q$ the unknown parameter vectors $\boldsymbol{\theta}^{(i)}$. Fig. 5 shows an example of estimating a sine function from given noisy data. Simulation results show that by HCLM one is able to obtain a smooth solution by optimizing the cost function (15) without regularization. As shown in [7] the initial distribution

TABLE II
RESULTS OF HCLM, MULTI-TR AND PNN METHODS ON FOUR BENCHMARK FUNCTIONS

			U_3	U_4	U_5	U_6
HCLM	dimension=10	ave.	**2.1123e-017**	**0.9353**	**7.4157e-008**	**4.6775e-023**
		std.	**1.2407e-032**	**1.0163**	**5.3574e-015**	**2.0328e-043**
	dimension=30	ave.	7.2328e-017	**13.2728**	**0.4544**	1.3856e-024
		std.	2.6952e-031	**3.4030**	0.5149	1.3679e-046
	dimension=100	ave.	**8.0115e-018**	**28.3466**	**0.3936**	**7.9053e-043**
		std.	**2.6609e-033**	**10.6089**	0.3046	**2.1096e-083**
multi-TR	dimension=10	ave.	0.0207	52.3047	18.4617	23.0637
		std.	0.0006	236.7678	0.2532	347.2781
	dimension=30	ave.	**8.0138e-030**	199.7170	19.0116	**1.8124e-035**
		std.	**4.1833e-057**	554.2156	0.0352	**1.4891e-068**
	dimension=100	ave.	1.0157e-017	714.2272	8.0361	8.6904e-043
		std.	5.0990e-033	1.5091e+003	**0.0275**	6.6103e-083
PNN [20]	dimension=10	ave.	180.2122	52.1356	18.4615	38.5365
		std.	1.8069e+003	232.5831	0.2547	25.4837
	dimension=30	ave.	667.9961	199.4782	19.0494	238.9246
		std.	7.0515e+003	575.7879	**0.0338**	143.8955
	dimension=100	ave.	2.4552e+003	272.3695	9.1089	800.1367
		std.	1.4470e+003	233.4527	0.0419	5.2762e-005
PNN [22]	dimension=10	ave.	180.2122	52.1356	18.4618	38.5531
		std.	1.8069e+003	232.5831	0.2549	25.5161
	dimension=30	ave.	667.9961	199.4782	19.0506	238.9246
		std.	7.0515e+003	575.7879	0.0340	143.8955
	dimension=100	ave.	2.4552e+003	272.3695	9.1089	800.1367
		std.	1.4470e+003	233.4527	0.0419	5.2762e-005

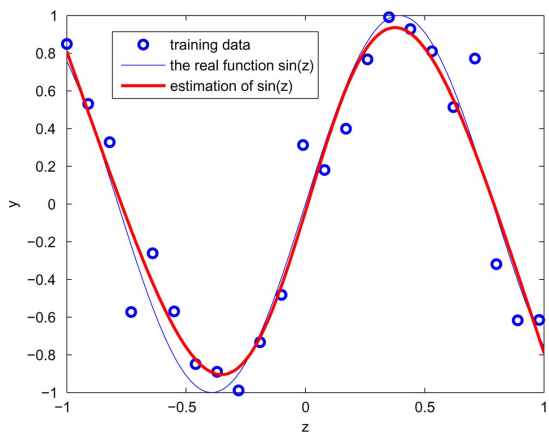


Fig. 5. Application of learning of multilayer perceptrons on training a sinusoidal function (blue line) given 23 training data points $\{u_k, d_k\}$ (blue circles) corrupted with Gaussian noise (zero mean, standard deviation 0.2) using a multilayer perceptron with one hidden layer consisting of 10 hidden units. The result by HCLM (red line) is shown which optimizes a sum squared error on the training data without regularization of the cost function provided the states of the initial particles are normally distributed with zero mean and standard deviation 0.1. Here, $q = 20$ particles are adopted to perform the searching process.

of the optimizers plays the role of a regularization term for (15).

VII. CONCLUSION

We have presented the design of a new global optimization method called hybrid coupled local minimizers (HCLM). This method is composed of several parallel search processes with adaptive weighting coefficients and a hybrid coupling strength. As the Newton-based trust region method is used, the search process in the HCLM method has a high convergence rate. In the framework of HCLM, there is no need to transform the unconstrained optimization problem into a nonlinear optimization problem with linear equality constraints as the original CLM scheme does. Meanwhile, the HCLM method and its implementation have been described extensively in the context. We have also proved the local stability of the coupled local minimizers and provided conditions for the synchronization of the HCLM system in order to ensure that all particles converge to a common optimization solution. Then, the HCLM method has been illustrated with several experiments including high-dimensional test functions possessing many local minima, and obviously superior to the PNN methods. Finally, the presented method has been applied to the training of multilayer perceptron networks.

REFERENCES

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

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

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

[4] H. L. Wang and O. K. Ersoy, "A Novel Evolutionary Global Optimization Algorithm and Its Application in Bioinformatics," ECE Technical Reports, Aug. 2005, Purdue University, Electrical and Computer Engineering.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. New York, NY, USA: Addison-Wesley, 1989.

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

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

[8] 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.

[9] 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.

[10] X. Lou and J. A. K. Suykens, "Stability of coupled local minimizers within the lagrange programming network framework," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 2, pp. 377–388, 2013.

[11] W. M. Haddad, V. Chellaboina, and S. G. Nersesov, *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control*. Princeton, NJ, USA: Princeton Univ. Press, 2006.

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 1999.

[13] 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.

[14] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 240–255, 2004.

[15] Y. Shi and R. C. Eberhart, "A modified particle swarm optimiser," in *Proc. IEEE Int. Conf. Evol. Comput.*, Piscataway, NJ, USA, 1998, pp. 69–73.

[16] S. G. Nersesov and W. M. Haddad, "Finite-time stabilization of nonlinear impulsive dynamical systems," *Nonlinear Analysis: Hybrid Systems*, vol. 2, no. 3, pp. 832–845, 2008.

[17] V. Lakshmikantham, D. D. Bainov, and P. S. Simeonov, *Theory of Impulsive Differential Equations*. , Singapore: World Scientific, 1989.

[18] J. A. K. Suykens, T. Yang, and L. O. Chua, "Impulsive synchronization of chaotic Lur'e systems by measurement feedback," *Int. J. Bifurcation Chaos*, vol. 8, no. 6, pp. 1371–1381, 1998.

[19] J. T. Sun and Y. P. Zhang, "Stability analysis of impulsive control systems," in *IEE Proc. Control Theory Appl.*, 2003, vol. 150, no. 4, pp. 331–334.

[20] 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.

[21] 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.

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

[23] S. Effati and M. Baymani, "A new nonlinear neural network for solving quadratic programming problems," *Appl. Math. Comput.*, vol. 165, no. 3, pp. 719–729, 2005.

[24] W. Lohmiller and J. J. E. Slotine, "On contraction analysis for nonlinear systems," *Automatica*, vol. 34, pp. 683–696, 1998.

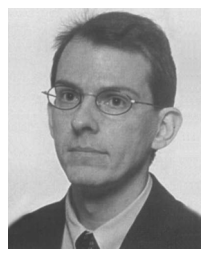
[25] K. E. Rifai and J. J. E. Slotine, "Compositional contraction analysis of resetting hybrid systems," *IEEE Trans. Autom. Contr.*, vol. 51, no. 9, pp. 1536–1541, 2006.

[26] D. Whitley, S. B. Rana, J. Dzubera, and K. E. Mathias, "Evaluating evolutionary algorithms," *Artif. Intell.*, vol. 85, no. 1–2, pp. 245–276, 1996.



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), 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 s an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS (1998–2009). 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, and chair of ROKS 2013. He has been awarded an ERC Advanced Grant 2011.