

TWO TIME-SCALES IN GLOBAL OPTIMIZATION AND EQUILIBRIUM

---

A Dissertation

Presented to  
the faculty of the School of Engineering and Applied Science  
University of Virginia

---

in partial fulfillment  
of the requirements for the degree

Doctor of Philosophy

by

YUE SUN

August

2015

APPROVAL SHEET

The dissertation  
is submitted in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy



---

AUTHOR

The dissertation has been read and approved by the examining committee:

**ALFREDO GARCIA**

---

Advisor

**MICHAEL GALLMEYER**

---

**PETER BELING**

---

**GERARD LEARMONTH**

---

**ZONGLI LIN**

---

Accepted for the School of Engineering and Applied Science:



Dean, School of Engineering and Applied Science

August  
2015

## ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my academic advisor Prof. Alfredo Garcia. Thank you for offering me this great research opportunity on Global Optimization, Model Based Searching and Financial Engineering. Thank you for guiding me in my Ph.D. study which gave me lots of improvement. It is you who led me to the research world which is one of the most remarkable experiences of my life. I would like to thank my Ph.D. advisory committee members Prof. Gerard Learmonth, Prof. Peter Beling, Prof. Zongli Lin and Prof. Michael Gallmeyer. Thank you for your valuable advice and suggestions for my research work. I would like to thank Mingyi Hong, Yuting Wang and Jorge Barrera who are in our research group. Thank you for your ideas and inspiration for my projects. Lastly I would like to thank my parents Peishi Sun and Lili Liu. Thank you for your great support during my Ph.D. study.

## TABLE OF CONTENTS

|  | Page |
|--|------|
| LIST OF TABLES . . . . .                                     | v    |
| LIST OF FIGURES . . . . .                                    | vii  |
| ABSTRACT . . . . .   | viii |
| 1 Introduction . . . . .                                     | 1    |
| 2 Interactive Diffusions for Global Optimization . . . . .   | 6    |
| 2.1 Introduction and Literature Review . . . . .             | 6    |
| 2.2 Framework . . . . .                                      | 8    |
| 2.2.1 Problem Setting . . . . .                              | 8    |
| 2.2.2 Independent Diffusions . . . . .                       | 9    |
| 2.2.3 Interactive Diffusions . . . . .                       | 10   |
| 2.3 Local Escape Time Problem . . . . .                      | 10   |
| 2.3.1 Escape Time Theory . . . . .                           | 11   |
| 2.3.2 Escape Time Comparison . . . . .                       | 13   |
| 2.3.3 Discussion . . . . .                                   | 14   |
| 2.4 Global Asymptotic Behavior . . . . .                     | 15   |
| 2.4.1 Convergence Behavior . . . . .                         | 16   |
| 2.4.2 Unchanged Global Optimize Solution . . . . .           | 16   |
| 2.5 Improved Speed of Convergence . . . . .                  | 19   |
| 2.6 Illustration: Numerical Experiments . . . . .            | 24   |
| 2.6.1 Ackley Problem . . . . .                               | 24   |
| 2.6.2 Rastrigin Problem . . . . .                            | 30   |
| 2.6.3 Comparison with Parallel Simulated Annealing . . . . . | 33   |

|   | Page |
|---|------|
| 2.7 Chapter Summary . . . . .   | 36   |
| 3 Interactive Model-based Search with Reactive Resource Allocation . . . . .            | 39   |
| 3.1 Introduction and Literature Review . . . . .  | 39   |
| 3.2 Single Thread Model-based Search in Real-Time . . . . .                             | 42   |
| 3.2.1 Basic Single-Thread Computation . . . . .   | 44   |
| 3.2.2 Reacting to Incomplete Searches . . . . .   | 46   |
| 3.3 Interactive Model-based Search in Real Time . . . . .                               | 48   |
| 3.3.1 Reactive Resource Allocation: Avoiding Duplication of Incomplete Search . . . . . | 50   |
| 3.3.2 Analysis . . . . .  | 53   |
| 3.4 Numerical Experiments . . . . .   | 57   |
| 3.4.1 Ackley Problem . . . . .  | 57   |
| 3.4.2 Rastrigin Problem . . . . .   | 62   |
| 3.5 Chapter Summary . . . . .   | 64   |
| 4 High Frequency Traders in Inefficient Market . . . . .                                | 66   |
| 4.1 Introduction and Literature Review . . . . .  | 66   |
| 4.2 One Stage Kyle Model with High Frequency Trader . . . . .                           | 68   |
| 4.3 Dynamic Kyle Model with High Frequency Trader . . . . .                             | 74   |
| 4.3.1 Equilibrium Strategy and Information Update . . . . .                             | 76   |
| 4.3.2 Trading Process Convergence . . . . .   | 79   |
| 4.3.3 Profitability Analysis . . . . .  | 83   |
| 4.3.4 Market Impact of High Frequency Trader . . . . .                                  | 85   |
| 4.4 Illustrative Numerical Experiment . . . . .   | 89   |
| 4.5 Chapter Summary . . . . .   | 91   |
| 5 Conclusions . . . . .   | 94   |
| LIST OF REFERENCES . . . . .  | 96   |

## LIST OF TABLES

| Table   | Page |
|---|------|
| 2.1 Average iterations before entering $\rho$ -neighborhood of $x_*$ for 3-D Ackley function . . . . .                    | 28   |
| 2.2 Probability for reaching global optima within $10^4$ iterations for 3-D Ackley function . . . . .                     | 29   |
| 2.3 Average iterations before entering $\rho$ -neighborhood of $x_*$ for 5-D Ackley function . . . . .                    | 29   |
| 2.4 Probability for reaching global optima within $10^5$ iterations for 5-D Ackley function . . . . .                     | 30   |
| 2.5 Average iterations before entering $\rho$ -neighborhood of $x_*$ for 3-D Rastrigin function . . . . .                 | 32   |
| 2.6 Probability for reaching global optima within $10^5$ iterations for 3-D Rastrigin function . . . . .                  | 32   |
| 2.7 Probability for 2 trivial paralleled annealing threads reaching global optima within iterations cap . . . . .         | 34   |
| 2.8 Average iterations before 1 of 5 threads entering $\rho$ -neighborhood of $x_*$ for 3-D Ackley function . . . . .     | 34   |
| 2.9 Average iterations before 1 of 5 threads entering $\rho$ -neighborhood of $x_*$ for 5-D Ackley function . . . . .     | 35   |
| 2.10 Average iterations before 1 of 5 threads entering $\rho$ -neighborhood of $x_*$ for 3-D Rastrigin function . . . . . | 35   |
| 2.11 Probability for 1 of 5 threads reaching global optima within 5000 iterations for 3-D Ackley function . . . . .       | 35   |
| 2.12 Probability for 1 of 5 threads reaching global optima within $10^4$ iterations for 5-D Ackley function . . . . .     | 35   |
| 2.13 Probability for 1 of 5 threads reaching global optima within $10^4$ iterations for 3-D Rastrigin function . . . . .  | 36   |
| 3.1 Average number of iterations before reaching global optima . . . . .  | 59   |

| Table  | Page |
|--|------|
| 3.2 Average number of ticks before reaching global optima . . . . .      | 61   |
| 3.3 Probability of finding global optima within 5000 ticks . . . . .     | 61   |
| 3.4 Average number of iterations before reaching global optima . . . . . | 63   |
| 3.5 Average number of ticks before reaching global optima . . . . .      | 64   |
| 3.6 Probability of finding global optima within 2000 ticks . . . . .     | 64   |

## LIST OF FIGURES

| Figure  | Page |
|---|------|
| 2.1 Ackley function figures . . . . .                                     | 26   |
| 2.2 Current function value plot for 3-dimension Ackley function . . . . . | 27   |
| 2.3 Rastrigin function plots . . . . .                                    | 31   |
| 3.1 Independent model search method . . . . .                             | 58   |
| 3.2 Original and modified interactive model search method . . . . .       | 60   |
| 3.3 Independent model search method . . . . .                             | 62   |
| 3.4 Original and modified interactive model search method . . . . .       | 63   |
| 4.1 Value revealed path . . . . .   | 90   |
| 4.2 Variance expectation update path . . . . .                            | 90   |
| 4.3 Cumulative profit of informed trader . . . . .                        | 91   |
| 4.4 Cumulative profit (loss) of noise traders . . . . .                   | 92   |

## ABSTRACT

Sun, Yue Ph.D., University of Virginia, June 2015. Two Time-scales in Global Optimization and Equilibrium. Major Professor: Professor Alfredo Garcia.

In recent years, the performance improvement in computer architecture is shifting from making a single core faster to increasing the number of processors. Parallel computing becomes the dominant paradigm in computer architecture. In the global optimization and equilibrium community, parallel optimization algorithms have been developed to solve heavily computational intensive problems. One major associated problem is how to effectively utilize parallel computing power. In this dissertation, we consider two timescales parallelism in which tasks assigned to parallel threads are allowed to operate in two timescales. In chapter 2, we present an algorithmic design with interacting annealing processes in two timescales that guarantee a faster identification of a globally optimal solution. In chapter 3, we consider a parallel computing scheme for global optimization that combines a fast timescale multi-start local search with a slow timescale dynamic reallocation of computational resources. In chapter 4, we modified Kyle's informed trading model to include high frequency traders and show that these traders play a beneficial role in the market in which insider trading activity has also been detected.

## 1. INTRODUCTION

The efficient computation of optimal and equilibrium solutions is an important research domain at the intersection of several disciplines including operations research, computer science and economics. With the development of both hardware and software, computational power has been dramatically increased [1]. However, further increases of computing speed are prevented from the temperature of chips when the chips are clocked at higher speed and they become much less energy efficient [2]. In [3], the main techniques for increased clock frequency are described as hitting the "power wall" because of increasingly complex architectures. In light of these physical constraints, parallel computing is an alternative that has already been employed for many years in high-performance computing. The increasing number of processors is considered as a major source of future performance increases.

Compared to serial computing in which tasks are executed sequentially, parallel computing carries out many calculations simultaneously operating on the principle that large problems can often be divided into smaller ones which are then solved concurrently ("in parallel"). There are several different forms of parallel computing: bit-level [4], instruction level [5], data [6], and task parallelism [7]. The major form involved in this dissertation is task parallelism, which is a form of parallelization of computer code across multiple processors. This parallelism focused on distributing execution processes (threads) across different parallel computing nodes. Task parallelism is achieved when each processor executes a different thread (or process) on the

same or different data, in which different execution threads communicate with one another as they work.

Multiple thread computing has been used in optimization. The idea is to launch multiple optima or equilibrium identification threads simultaneously on the same or different searching domains. The optimization or equilibrium searching task is cut into small pieces operated on different threads. Algorithms derived from parallelization have allowed optimization or equilibrium problems requiring days or weeks of computation on a single-processor computer to be solved in a matter of hours on a multi-processor machine [8]. In order to effectively allocate threads and computational power, multiple threads with different communication layers such as MPI and PVM have been used to develop parallel optimization algorithms [9]. The searching threads of this kind of parallel optimization algorithms not only share information but also react to the information shared. Each thread receives attractive, repulsive or other kind of interactive force from accumulation of all other parallel threads. The most popular methods of this category are genetic algorithm [10], simulated annealing [11] and, most recently, particle swarm optimization [12]. Particle swarm optimization belongs to a population based optimization category which is particularly suited to continuous variable problems and has received increasing attention in the optimization and equilibrium community. This kind of algorithms work by having a population of particles following a few simple dynamic rules while the interaction between acts as a major optimization driver to keep particles formation and direction (see [13], [14] and [15]). In this dissertation, we implement an interactive technique to the particles with full scale searching dynamic instead of having interaction as the main driver. The communication and interaction benefit the efficient allocation computational power and avoid repetition of the searching effort. The parallelism which

related to this dissertation is similar to task parallelism, where different functions are assigned to population of threads.

While allowing different functions of different threads or the same thread, it is common that tasks are operated in different timescales. Inspired by many applications of science and engineering can be modeled by fast-slow processes [16, 17], multiple timescales parallel optimization algorithm is introduced [18]. The major applications of this category are in online learning and optimization (see [19] and [20]), the problems of which usually encounter objective function of uncertainty. Stochastic approximation is served as an objective function parameter update rule, the process includes parameter approximations which operated in a slow timescale and optimum identification, which is processed in a fast timescale. The idea of two timescales parallel optimization is to allocate a portion of threads in fast timescale to screen objective function and help parameters' approximation with the remaining threads in slow timescale to perform optima searching function. In this dissertation, we consider deterministic objective function problem and the parameters of objective function are assumed to be known. All computational threads in different timescales have the same category of functions of optimal and equilibrium identification while the threads in fast timescale concentrate more on exploration and the threads in slow timescale focus on exploitation. Threads in the same and different timescales communicate and interact in different layers which help efficiently allocate computational power and avoid duplication effort.

In the first part of the dissertation we present an algorithmic design with interacting annealing processes which guarantee a faster identification of a globally optimal solution. A first annealing process operates in a faster timescale and has a drift function that converges on a non-zero (but relatively small) noise level. A second annealing process (operating a slower timescale) is subject to a modified drift term in

which the steepest descent direction is perturbed with the density gradient associated to the first annealing process. In other words, a repulsive potential from fast screening thread is added to smooth the objective function of second annealing process. We show that this repulsive potential (based upon the first annealing process which quickly identifies locally optimal solutions) allows the second annealing process to bypass locally optimal solutions in a faster fashion.

In the second part, we revisit the interactive model-based approach to global optimization proposed in [21] in which parallel threads independently execute a model-based search method and periodically interact through a simple acceptance-rejection rule aimed at preventing duplication of search efforts. The local search functions operate in fast timescale with the periodic interactions process in slow timescale. We consider a real-time implementation of the interactive model-based approach which leads to the problem that when the acceptance-rejection rule is implemented, several threads may fail to identify a locally optimal solution. The acceptance-rejection rule is modified to achieve those alternates with enforcing diverse search (in order to prevent duplication) and reallocation of computational effort (in order to speed up the identification of local optima). We show this modification improves the rate of convergence in real-time which increases with the number of threads.

In the final part of this dissertation we consider a two timescales trading model in financial markets. A single high frequency trader who trading in fast timescale is added into Kyle's discrete dynamic trading model in [22] with the insider trader. The uninformed high frequency trader perturbs market structure and neutralizes the ability of observing knowledge of liquidity order low. During equilibrium trading process, market maker and the informed trader update their information about aggregated order low from the high frequency trader and noise traders by Bayes' rule. The belief of liquidity variance converges on real variance in limit. The high frequency

trader makes positive expected rent when market maker and informed trader's belief of liquidity is incorrect and this positive rent decreases with accuracy of liquidity expectation. We show that when the trading quantity of high frequency trader stay within a certain range, the high frequency trader protects noise traders by reducing the total expected loss and plays a beneficial role of the market by providing extra liquidity when market maker underestimated liquidity variance.

## 2. INTERACTIVE DIFFUSIONS FOR GLOBAL OPTIMIZATION

### 2.1 Introduction and Literature Review

Simulated annealing for global optimization (see [23], [24]) can be modeled in a continuous domain as a diffusion process in which the drift term is equal to the steepest descent direction and the drift function (i.e. the “cooling schedule”) approaches zero at a suitable rate so that weak convergence to a limiting distribution (concentrating almost all mass in the set of globally optimal solutions) is guaranteed (see, for example, [25–27]). The main drawback of this class of methods relates to speed of convergence as the cooling schedule can not be “too fast” in order to guarantee convergence to globally optimal solutions. In other words, the initial emphasis on “exploration” must slowly give way to emphasis on “exploitation” in order to guarantee globally optimal solutions are identified. Thus, many research efforts have been devoted to accelerating the convergence rate for certain classes of smooth objective functions (see, e.g., [28, 29] and, more recently, [30]). Given the technological challenges currently faced to speed-up processing, parallel implementation of simulated annealing appears as a sensible approach for speeding up convergence to globally optimal solutions (see, for instance, [31] for an application to molecular clustering).

In this chapter, we present a novel approach in which parallel annealing processes interact in a manner that enables a faster identification of a globally optimal solution. A first annealing process operates at a faster timescale and has a drift function that converges to a non-zero (but relatively small) noise level. A second annealing

process (operating at slower timescale) is subject to a modified drift term in which the steepest descent direction is perturbed with the density gradient associated to the first annealing process. In other words, a repulsive potential is added in order to ensure the second annealing process does not duplicate the first annealing process search effort. We show that this repulsive potential (based upon the first annealing process which quickly identifies locally optimal solutions) allows the second annealing process to *bypass* locally optimal solutions in a faster fashion.

Our work is related to the class of smoothing methods (see [32,33]) in which the original objective function evolves into a (smoothed) function possessing far fewer local minima. A potential drawback of these methods pertains to the possibility that the minimum of the modified objective function may shift away from the global minimum of the original objective function (see [33] for a smoothing method designed to avoid this difficulty). The approach proposed in this chapter can be seen as a time-varying smoothing of the original objective function which is related to the first diffusion density. The computational work by the first diffusion is used to construct a smoother objective function governing the second (slower) diffusion's search. Building up on [16], we show that it is easier for the second diffusion to escape the attraction basins of locally optimal solutions.

The structure for this chapter is as follows. In Section 2, we formalize the proposed approach in the context of a global optimization problem. In Section 3 we study the "escape time" of interactive annealing processes. It is shown that due to interaction, one of the annealing processes is able to escape the basins of attraction of local minima in less time (in expectation). The discussion on this speed-up effect is followed by the analysis of global asymptotic behavior in Section 4 in which convergence to global minima is proven. Building up on the results in Sections 3 and 4, we show that the speed of convergence is improved in Section 5. Finally in Section 6, we provide a

limited numerical testbed to illustrate the merits of the proposed approach (which are established in Sections 2 and 3 from a theoretical standpoint). Finally, in Section 7, we offer some conclusion and briefly comment upon future research.

## 2.2 Framework

### 2.2.1 Problem Setting

Consider the optimization problem  $\min\{H(x) : x \in \mathbb{R}^n\}$ , where  $H : \mathbb{R}^n \rightarrow \mathbb{R}$  is assumed to be continuously differentiable. As in [26], we make the following regularity assumption throughout the chapter:

$$\lim_{\|x\| \uparrow +\infty} H(x) \uparrow +\infty, \quad (2.1)$$

$$\lim_{\|x\| \uparrow +\infty} \|\nabla H(x)\| \uparrow +\infty, \quad (2.2)$$

$$\lim_{\|x\| \uparrow +\infty} \|\nabla H(x)\| - \Delta H(x) > -\infty. \quad (2.3)$$

Let  $y_l := \min\{H(x) : x \in B_l \subset \mathbb{R}^n\}$  for  $l = 1, 2, \dots, m$  denote the collection of local minima values, where  $B_l \subset \mathbb{R}^n$  is the basin of attraction for the  $l$ -th local minima. Let  $y_* := \min\{y_l : l = 1, \dots, m\}$ . Assume there exists a positive number  $M$  such that:

$$y_l - y_* > M \gg 0, \quad \forall y_l \neq y_*, l = 1, 2, \dots, m. \quad (2.4)$$

### 2.2.2 Independent Diffusions

We start by introducing the dynamic system for a diffusion in  $\mathbb{R}^n$  as follows:

$$\begin{cases} \dot{x}_p(t) = v_p(t) \\ \dot{v}_p(t) = \frac{1}{M} \left[ -\frac{\partial H(x)}{\partial x_p} - \gamma_0 v_p(t) + \sigma_0(t) \frac{dW_p(t)}{dt} \right], \end{cases}$$

according to the Fokker-Planck equation (see [34]), where  $p = 1, \dots, n$  and  $M$  is particle mass,  $\gamma_0$  is damping coefficient,  $\sigma_0(\cdot)$  is noise intensity and  $W_p(t)$  is  $p^{\text{th}}$  component of a standard  $n$ -dimensional Brownian motion,  $W(t)$ . Assume particle mass  $M$  is large enough relative to  $-\frac{\partial H(x)}{\partial x_p} - \gamma_0 v_p(t) + \sigma_0(t) \frac{dW_p(t)}{dt}$ , so that  $\dot{v}(t) \approx 0$  and the system can be reduced to:

$$\frac{dX(t)}{dt} = \frac{1}{\gamma_0} \left[ -\nabla H(x) + \sigma_0(t) \frac{dW(t)}{dt} \right].$$

Let  $U(x) = \frac{H(x)}{\gamma_0}$ ,  $\sigma(t) = \frac{\sigma_0(t)}{\gamma_0}$ . We obtain a diffusion model for simulated annealing as:

$$\frac{dX(t)}{dt} = -\nabla U(x) + \sigma(t) \frac{dW(t)}{dt}. \quad (2.5)$$

The partial differential equation of this diffusion process is given by

$$\frac{\partial}{\partial t} V(t, x) = \nabla \cdot (V(t, x) \nabla U(x)) + \frac{1}{2} \sigma^2(t) \Delta V(t, x),$$

where  $V(\cdot, \cdot)$  is the density of particles.

### 2.2.3 Interactive Diffusions

In our method, we shall make use of the notion of different time scales for interacting diffusions. In what follows, we shall consider a first diffusion (with subindex)  $i$  in a timescale  $t/\epsilon^2$  where  $\epsilon > 0$  with constant noise intensity  $\sigma_i$ , and a second diffusion  $j$  in slower timescale  $t$ , with annealing noise intensity  $\sigma_j(t) \sim C/\log(t)$ , interacting with diffusion  $i$ . The dynamic system for these interactive diffusions is

$$\begin{cases} dX_i(t) = -\frac{1}{\epsilon^2}\nabla U(X_i(t))dt + \frac{1}{\epsilon}\sigma_i dW(t) \\ dX_j(t) = -\nabla U(X_j(t))dt - k\nabla V_i(t, X_j(t))dt + \sigma_j(t)dW(t) \\ X_i(0) = \hat{x}_i, \quad X_j(0) = \hat{x}_j, \end{cases} \quad (2.6)$$

where  $V_i$  is density for diffusion  $i$ ,  $\hat{x}_i$  and  $\hat{x}_j$  are the initial conditions for diffusion  $i$  and  $j$ ,  $k > 0$  is a parameter that controls the strength of the “repulsive” potential, i.e.  $-\nabla V_i(t, X_j(t))$  affecting diffusion  $j$ . From Feynman-Kac formula [35], the associated Fokker-Planck type equation is:

$$\begin{cases} \frac{\partial}{\partial t}V_i(t, x) = \frac{1}{2\epsilon^2}\sigma_i^2\Delta V_i(t, x) + \frac{1}{\epsilon^2}\nabla \cdot (V_i(t, x)\nabla U(x)) \\ \frac{\partial}{\partial t}V_j(t, x) = \frac{1}{2}\sigma_j^2(t)\Delta V_j(t, x) + \nabla \cdot [V_j(t, x)(\nabla U(x) + k\nabla V_i(t, x))] \\ V_i(0, x) = \delta_{\hat{x}_i}(x), \quad V_j(0, x) = \delta_{\hat{x}_j}(x). \end{cases} \quad (2.7)$$

## 2.3 Local Escape Time Problem

In this section, we analyze the “escape times” for the interactive diffusions using the theory of large deviations introduced in [27]. We then proceed to compare the expected escape time properties for the interactive diffusions and the original simu-

lated annealing. Finally, we argue that such comparison is fair even when different time scales (used in the interactive diffusions approach) are taken into account.

### 2.3.1 Escape Time Theory

For the dynamical system (2.5), define  $L$  as follows

$$L(\beta, x) := \sup_{\alpha} [\alpha' \beta - \alpha'(-\nabla U(x))], \quad \alpha, \beta \in \mathbb{R}^n.$$

The action functional  $S_x(T, \phi)$  for point  $x$  is defined as

$$S_x(T, \phi) := \int_0^T L(\dot{\phi}(s), \phi(s)) ds, \quad \phi(0) = x.$$

Finally, the action function between two points  $x$  and  $y$  is

$$S(x, y) = \inf_{\phi, T} \{S_x(T, \phi) : \phi(T) = y\}.$$

Let  $K_0$  be the set of local minima as

$$K_0 = \{x_l \in \mathbb{R}^n : | H(x_l) = y_l, l = 1, \dots, m\}.$$

Let  $G$  be a bounded open set containing  $K_0$ , with a piecewise differentiable boundary  $\partial G$  and  $\bar{G}$  in the domain of attraction of  $K_0$ , i.e.  $\bar{G} \subseteq \cup_{l=1}^m B_l$ , for  $x \in G$  define

$$S_G(x) := \inf_{y \in \partial G} S(x, y) = \inf_{\phi, T} \{S_x(T, \phi) : \phi(T) \in \partial G\}.$$

Define the action function for any open set  $B \subset G$  as follows:

$$S_G(B) := \inf_{x \in B} S_G(x).$$

The action function being the integral of difference between the potential gradient direction  $-\nabla U(\phi(s))$  and the path moving direction  $\dot{\phi}(s)$ , can be understood as the “energy” needed to resist the potential force through the path  $\phi$ .

**Assumption 3.1** *For  $\delta > 0$ , there is a  $\rho$ -neighborhood  $N_\rho(K_0)$  of  $K_0$  and  $\delta_\rho > 0$ ,  $T_\rho < +\infty$ , such that for each  $x, y \in N_\rho(K_0)$ , there is a path  $\phi(\cdot)$ :*

$$\phi(0) = x, \quad \phi(T) = y,$$

where  $T_y \leq T_\rho$  and  $S_x(T_\rho, \phi) \leq \delta$ . (see [27], A3.1)

**Lemma 3.1** *Let  $X^\sigma(t)$  be the solution of dynamic system (2.5) with noise intensity  $\sigma(t) = \sigma$  and escape time  $\tau^\sigma = \min\{t : X^\sigma(t) \notin G\}$ . Under Assumption 3.1, the expected escape time can be written as*

$$\lim_{\sigma} \sigma \log E_x \tau^\sigma = S_G(K_0),$$

which can be written as

$$E_x \tau^\sigma \sim \exp\left(\frac{S_G(K_0)}{\sigma}\right).$$

**Proof** See proof of [27], p.174, Theorem 1. ■

### 2.3.2 Escape Time Comparison

In this section, we will compare the local minima escape time of standard simulated annealing and that of the interactive diffusions.

**Theorem 3.1** *Let  $\bar{\tau}_j^\sigma = \min\{t : X_j^\sigma(t) \notin G\}$ , i.e. the escape time of the  $j$ -diffusion process defined in (2.6) when the noise intensity for the  $i$ -diffusion process is  $\sigma_i = \sigma$ . When  $\epsilon \downarrow 0$  (the timescale ratio for diffusion  $i$  in (2.6)), we have*

$$\lim_{\sigma} \sigma \log E_x \bar{\tau}_j^\sigma \leq \lim_{\sigma} \sigma \log E_x \tau^\sigma.$$

**Proof** Consider the behavior of fluid  $i$  of (2.7) in fast time scale  $t/\epsilon^2$  with constant noise intensity  $\sigma_i$ . let  $\hat{t} = t/\epsilon^2$ , the dynamic for fluid  $i$  is

$$\epsilon^2 \frac{\partial}{\partial t} V_i(t, x) = \frac{\partial}{\partial \hat{t}} V_i(t, x) = \frac{1}{2} \sigma_i^2 \Delta V_i(t, x) + \nabla \cdot (V_i(t, x) \nabla U(x)).$$

When  $\epsilon \downarrow 0$ , we have  $\hat{t} \uparrow +\infty$ , from standard results in the theory of diffusions (see [36], p.147), we have the fluid  $i$  stationary density  $\bar{V}_i(x) = \pi_\sigma(x)$ , where

$$\pi_\sigma(x) = C_0 \exp\left(-\frac{2U(x)}{\sigma^2}\right), \quad (2.8)$$

and

$$C_0 = \left( \int_{\mathbb{R}^n} \exp\left(-\frac{2U(x)}{\sigma^2}\right) dx \right)^{-1},$$

which is the Gibbs density.

For any  $K_0$  and  $\epsilon \downarrow 0$ , the fluid  $j$  can be seen as a diffusion governed by the modified potential function

$$\bar{U}(x) = U(x) + k\bar{V}_i(x) = U(x) + kV_0 \exp\left(-\frac{2U(x)}{\sigma^2}\right).$$

The corresponding  $L$ -function  $\bar{L}$  can be written as

$$\bar{L}(\beta, x) = \sup_{\alpha} \left[ \alpha' \beta + \alpha' \nabla U(x) \left( 1 - \frac{2k}{\sigma^2} \exp\left(-\frac{2U(x)}{\sigma^2}\right) \right) \right].$$

We have, for all  $x \in G$ ,  $\|\nabla \bar{U}(x)\| \leq \|\nabla U(x)\|$ ,

$$\bar{S}_G(K_0) = \inf_{x \in K} \inf_{y \in \partial G} \inf_{\phi, T} \left\{ \int_0^T \bar{L}(\dot{\phi}(s), \phi(s)) ds : \phi(0) = x, \phi(T) = y \right\} < S_G(K_0).$$

From Lemma 3.1, we have

$$\lim_{\sigma} \sigma \log E_x \bar{\tau}_j^{\sigma} = \bar{S}_G(K_0) \leq S_G(K_0) = \lim_{\sigma} \sigma \log E_x \tau^{\sigma}.$$

■

### 2.3.3 Discussion

According to the previous results, in expectation, diffusion  $j$  is able to escape the basin of attraction of locally optimal solutions in less time than a standard simulated annealing type diffusion. However, in this comparison the standard simulated annealing is assumed to diffuse in slow timescale  $t$ . In order to show that this comparison is fair, we need to consider the standard simulated annealing method in a faster timescale.

Consider the standard simulated annealing method in different time scale: a first diffusion  $\{X_a(t) : t > 0\}$  (with density  $V_a$ ) in slow timescale  $t$  and a second diffusion  $\{X_b(t) : t > 0\}$  (with density  $V_b$ ) in faster timescale  $t/\epsilon^2$  so that

$$\frac{\partial}{\partial t} V_a(t, x) = \frac{1}{2} \frac{c}{\log t} \Delta V_a(t, x) + \nabla \cdot (V_a(t, x) \nabla U(x))$$

$$\frac{\partial}{\partial t} V_b(t, x) = \frac{1}{2} \frac{c/\epsilon^2}{\log t/\epsilon^2} \Delta V_b(t, x) + \nabla \cdot (V_b(t, x) \frac{\nabla U(x)}{\epsilon^2}).$$

Let  $t_a$  and  $t_b$  the time required to reach a noise intensity  $\sigma_0 > 0$  It follows that

$$\sigma_0 = \frac{C/\epsilon^2}{\log t_b/\epsilon^2} = \frac{C}{\log t_a}.$$

From this we infer that

$$t_b = \epsilon^2 t_a^{1/\epsilon^2}.$$

For general situation, for  $t_a \gg 1$  and  $\epsilon \downarrow 0$ , we have  $t_b \gg t_a$ . That is, in a faster timescale it takes longer to arrive at given level of noise intensity. Using a faster timescale cannot accelerate convergence to global minima for the standard simulated annealing method.

## 2.4 Global Asymptotic Behavior

In this section, we will analyze the global asymptotic behavior of “slow” diffusion process, i.e.  $X_j(t)$ . As in [26], let  $S$  denote the set of all local minimas of  $U$  and  $S(\eta) = \{x | d(x, S) < \eta\}$ . Let

$$J(t, \eta) := \sup_{x, y \in S(\eta)} (I(t, x, y) - 2U(y)),$$

where

$$I(t, x, y) := \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t |\dot{\psi}(s) + \nabla U(\psi(s))|^2 ds.$$

Define

$$c_0 := \inf_{\eta} \limsup_{t \uparrow +\infty} J(t, \eta) \quad (2.9)$$

### 2.4.1 Convergence Behavior

**Lemma 4.1** *Under assumption (2.1), (2.2) and (2.3), and  $c > c_0$ ,*

$$V(t, x) \rightarrow \pi(x) \quad \text{as } t \uparrow +\infty,$$

where  $V(\cdot, \cdot)$  is the solution for equation:

$$\frac{\partial}{\partial t} V(t, x) = \frac{c}{2 \log t} \Delta V(t, x) + \nabla \cdot (V(t, x) \nabla U(x)), \quad V(0, x) = \delta_{x_0}(x)$$

and  $\pi(x) := \lim_{\sigma \downarrow 0} \pi_{\sigma}(x)$  concentrate on the global minima  $x_*$  of  $U(x)$ , with  $\pi_{\sigma}(x)$  defined as (2.8).

**Proof** See proof of [26], p.739, Theorem. ■

### 2.4.2 Unchanged Global Optimize Solution

For interactive diffusion (2.7), define the modified objective function

$$\tilde{U}(t, x) := U(x) + kV_i(t, x);$$

for slow diffusion process  $j$ , define constant  $\tilde{c}_0$  similar as (2.9)

$$\tilde{c}_0 := \inf_{\eta} \limsup_{t \uparrow +\infty} \tilde{J}(t, \eta),$$

where

$$\tilde{J}(t, \eta) := \sup_{x, y \in S(\eta)} (\tilde{I}(t, x, y) - 2\tilde{U}(y)),$$

and

$$\tilde{I}(t, x, y) := \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t \left| \dot{\psi}(s) + \nabla U(\psi(s)) \left( 1 - k \frac{V_i(x)}{\sigma^2} \right) \right|^2 ds.$$

**Theorem 4.1** *For fluid  $j$  with dynamic (2.7), and for fluid  $i$ , timescale ratio  $\epsilon \downarrow 0$ , let  $\sigma_j = c/\log t$  as  $c > \tilde{c}_0$ , under assumption (2.4), we have*

$$V_j(t, x) \rightarrow \tilde{\pi}(x),$$

where

$$\tilde{\pi}(x) = \lim_{\sigma \downarrow 0} \tilde{\pi}_\sigma(x) = \lim_{\sigma \downarrow 0} \left[ \tilde{C}_0 \exp\left(-\frac{2\tilde{U}_0(x)}{\sigma^2}\right) \right].$$

*This  $\tilde{\pi}(x)$  concentrated on the global minima  $\tilde{x}_*$  of modified objective function  $\tilde{U}(x)$ . Moreover  $\tilde{x}_*$  is in the attraction basin of the global minima  $x_*$  of original objective function  $U(x)$ .*

**Proof** From (2.8), with timescale ratio  $\epsilon \downarrow 0$ ,  $V_i(t, x) \rightarrow \bar{V}_i(x)$ , where

$$\bar{V}_i(x) = V_0 \exp\left(-\frac{2U(x)}{\sigma^2}\right).$$

Let  $\tilde{U}_0(x) = U(x) + k\bar{V}_i(x)$ , we have

$$\tilde{U}(t, x) \rightarrow \tilde{U}_0(x) \quad \text{as } t \uparrow +\infty.$$

Since  $U(x)$  is twice continuous differentiable, then  $\bar{V}_i(x)$  is at least twice differentiable. By assumption,  $U(x) \uparrow +\infty$  as  $\|x\| \uparrow +\infty$ , hence,  $\bar{V}_i(x) \downarrow 0$  as  $U(x) \uparrow +\infty$  and  $\lim_{\|x\| \uparrow +\infty} \bar{V}_i(x) = 0$ . We conclude that the stationary modified objective function  $\tilde{U}_0(x)$  satisfies assumptions (2.1), (2.2) and (2.3). From Lemma 4.1 for fluid  $j$  we have

$$V_j(t, x) \rightarrow \tilde{\pi}(x),$$

where

$$\tilde{\pi}(x) = \lim_{\sigma \downarrow 0} \tilde{\pi}_\sigma(x) = \lim_{\sigma \downarrow 0} \left[ \tilde{C}_0 \exp\left(-\frac{2\tilde{U}_0(x)}{\sigma^2}\right) \right];$$

and fluid  $j$  will concentrate on the global minima of modified objective function  $\tilde{U}_0(x)$ , which is different from the global minima of original objective function  $U(x)$ .

Now, we show the global minima of  $\tilde{U}_0(x)$  is in the attraction basin of the global minima of  $U(x)$  under assumption (2.4) with  $k < M$ . Let  $x_*$  be the global minima for  $U(x)$  with value  $y_*$ ,  $x_i$   $i = 1, 2, \dots, m$  be the local minima with value  $y_i$ . Assume  $\tilde{x}_*$  be a global minima for modified objective function  $\tilde{U}_0(x)$  with value  $\tilde{y}_*$ . Note that  $\bar{V}_j(t, x) \leq 1, \forall x \in \mathbb{R}^n$ . From  $k < M$ , we have

$$\tilde{U}_0(x_*) = y_* + k\bar{V}_i(x) \leq y_* + k \leq y_* + M.$$

From assumption (2.4), for all  $y_l \neq y_*$ ,  $l = 1, 2, \dots, m$ ,  $y_* + M < y_l$ , and  $\tilde{y}_*$  is the global minima value for  $\tilde{U}_0(x)$ , we get

$$\tilde{y}_* \leq \tilde{U}_0(x_*) \leq y_* + M < y_l \quad \forall y_l \neq y_*, l = 1, 2, \dots, m.$$

From

$$\tilde{y}_* = \tilde{U}_0(\tilde{x}_*) = U(\tilde{x}_*) + k\bar{V}(x) \geq U(\tilde{x}_*),$$

so that we have

$$U(\tilde{x}_*) \leq \tilde{y}_* \leq y_* + M < y_l \quad \forall y_l \neq y_*, l = 1, 2, \dots, m.$$

This means that  $U(\tilde{x}_*)$  is less than all the local minimas for original objective function. It follows that the global minima  $\tilde{x}_*$  for the modified objective function is in the attraction basin of the global minima of the original objective function  $U(x)$ .  $\blacksquare$

## 2.5 Improved Speed of Convergence

In Sections 3 and 4, we have shown that interactive annealing processes have shorter local escape time and converge to global minima attraction basin. To complement these results, in this section, we show that interactive annealing also exhibits improved speed of convergence.

The standard simulated annealing method requires an annealing schedule  $c_0/\log t$  in order to guarantee the fluid converge to global minima (see Lemma 4.1), where  $c_0$  is defined as follows:

$$c_0 := \inf_{\eta} \limsup_{t \uparrow +\infty} \sup_{x, y \in S(\eta)} (I(t, x, y) - 2U(y))$$

and

$$I(t, x, y) := \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t |\dot{\psi}(s) + \nabla U(\psi(s))|^2 ds.$$

In Theorem 4.1, it is shown that the interactive diffusion approach also requires an annealing schedule  $\tilde{c}_0/\log t$  in order to converge to global minima.

$$\tilde{c}_0 := \inf_{\eta} \limsup_{t \uparrow +\infty} \sup_{x, y \in S(\eta)} (\tilde{I}(t, x, y) - 2\tilde{U}(y)),$$

where

$$\tilde{I}(t, x, y) := \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t \left| \dot{\psi}(s) + \nabla U(\psi(s)) \left( 1 - k \frac{V_i(x)}{\sigma^2} \right) \right|^2 ds.$$

From (2.8), for fluid  $i$ , we have

$$V_i(y, t) = \pi_{\sigma_i}(y) + \delta(y, \epsilon), \quad \text{as } \lim_{\epsilon \downarrow 0} \delta(y, \epsilon) = 0.$$

For the Gibbs density

$$\pi_{\sigma_i}(x) := \frac{\exp\left(-\frac{2U(x)}{\sigma_i^2}\right)}{\int_{R^n} \exp\left(-\frac{2U(x)}{\sigma_i^2}\right) dx},$$

we have

$$\nabla V_i(y, t) = -\nabla U(y) \cdot \frac{2\pi_{\sigma_i}(y)}{\sigma^2} + \nabla \delta(y, \epsilon).$$

It follows that

$$\begin{aligned}
\tilde{I}(t, x, y) &= \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t \left| \dot{\psi}(\tau) + \nabla U(\psi(\tau)) \left( 1 - K \frac{2\pi_\sigma(\psi(\tau))}{\sigma^2} \right) + \nabla \delta(y, \epsilon) \right|^2 d\tau \\
&= \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \left\{ \frac{1}{2} \int_0^t \left| \dot{\psi}(\tau) + \nabla U(\psi(\tau)) \left( 1 - K \frac{2\pi_\sigma(\psi(\tau))}{\sigma^2} \right) \right|^2 d\tau + \frac{1}{2} \int_0^t |\nabla \delta(y, \epsilon)|^2 d\tau \right. \\
&\quad \left. + \int_0^t \left| \dot{\psi}(\tau) + \nabla U(\psi(\tau)) \left( 1 - K \frac{2\pi_\sigma(\psi(\tau))}{\sigma^2} \right) \right| \cdot |\nabla \delta(y, \epsilon)| d\tau \right\} \\
&= \inf_{\substack{\psi(0)=x \\ \psi(t)=y}} \frac{1}{2} \int_0^t \left| \dot{\psi}(\tau) + \nabla U(\psi(\tau)) \left( 1 - K \frac{2\pi_\sigma(\psi(\tau))}{\sigma^2} \right) \right|^2 d\tau + C(\epsilon) \\
&\leq I(t, x, y) + C(\epsilon).
\end{aligned}$$

where  $C(\epsilon) \downarrow 0$  as  $\epsilon \downarrow 0$ . If  $\epsilon \downarrow 0$ , then

$$\begin{aligned}
\tilde{c}_0 &= \inf_{\eta} \limsup_{t \uparrow +\infty} \sup_{x, y \in S(\eta)} (\tilde{I}_j(t, x, y) - 2(U(y) + K \cdot V_i(y, t))) \\
&< \inf_{\eta} \limsup_{t \uparrow +\infty} \sup_{x, y \in S(\eta)} (I(t, x, y) - 2U(y)) = c_0,
\end{aligned}$$

where the inequality follows from the fact that

$$2(U(y) + K \cdot V_i(y, t)) > 2U(y).$$

In other words, the cooling or annealing schedule for interactive diffusion can be decreased in constant level from  $c_0/\log t$  to  $\tilde{c}_0/\log t$ . We now restate a result ([26], p.740, Lemma 3) that establishes the relationship between the cooling schedule and the speed of convergence for an annealing process.

Let  $V^\sigma(x, t)$  be the solution of dynamic system

$$\frac{\partial}{\partial t} V(t, x) = \nabla \cdot (V(t, x) \nabla U(x)) + \frac{1}{2} \sigma^2 \Delta V(t, x).$$

**Lemma 5.1** *For noise intensity  $\sigma$*

$$\|V^\sigma(x, t) - \pi_\sigma(x)\| \leq \exp(t\lambda_2(\sigma)),$$

where  $\lambda_2(\sigma)$  is the second eigenvalue of  $L_\sigma = \frac{1}{2}\sigma^2\Delta - \nabla U \cdot \nabla$  and

$$\lambda_2(\sigma) \sim -\exp\left(-\frac{c_0}{\sigma}\right),$$

where  $c_0$  is the previous constant.

Let  $V^\sigma(x, t)$  denote the density associated with the standard simulated annealing method. We recall that  $V_j^\sigma(x, t)$  is the  $j$  fluid solution for the dynamic system associated to interactive annealing, i.e.:

$$\begin{cases} \frac{\partial}{\partial t} V_i(t, x) = \frac{1}{2\epsilon^2} \sigma_i^2 \Delta V_i(t, x) + \frac{1}{\epsilon^2} \nabla \cdot (V_i(t, x) \nabla U(x)) \\ \frac{\partial}{\partial t} V_j(t, x) = \frac{1}{2} \sigma^2 \Delta V_j(t, x) + \nabla \cdot [V_j(t, x) (\nabla U(x) + k \nabla V_i(t, x))] \end{cases} \quad (2.10)$$

**Proposition 5.1** *For cooling schedule  $\sigma \sim c/\log t$ , with  $c > c_0 > \tilde{c}_0$ , fluid  $j$  of interactive diffusion  $V_j^\sigma(x, t)$  has faster speed of convergence to global optima than standard annealing  $V^\sigma(x, t)$ .*

**Proof** Let  $S^*$  and  $\tilde{S}^*$  denote the set of global minima of  $U$  and  $\tilde{U}$ , respectively. Let  $B_1(\epsilon) = \{x \in \mathbb{R}^n \mid d(x, S^*) < \epsilon\}$  and  $B_2(\epsilon) = \{x \in \mathbb{R}^n \mid d(x, \tilde{S}^*) < \epsilon\}$ . For  $\delta > 0$  and  $\epsilon > 0$  with  $\int_{B_1(\epsilon)} dx = \int_{B_2(\epsilon)} dx > e\delta/2$ , there exists  $\sigma > 0$  such that

$$\min\left\{\int_{B_1(\epsilon)} \pi_\sigma(x) dx, \int_{B_2(\epsilon)} \tilde{\pi}_\sigma(x) dx\right\} > 1 - \delta/2.$$

From Lemma 5.1, we have

$$\|V^\sigma(x, t) - \pi_\sigma(x)\| \leq \exp[-t^{(1-\frac{c_0}{c})}], \quad \|V_j^\sigma(x, t) - \tilde{\pi}_\sigma(x)\| \leq \exp[-t^{(1-\frac{c_0}{c})}].$$

It follows that

$$\begin{aligned} \int_{B_1(\epsilon)} \pi_\sigma(x) dx - \int_{B_1(\epsilon)} V^\sigma(x, t) dx &\leq \int_{B_1(\epsilon)} \|V^\sigma(x, t) - \pi_\sigma(x)\| dx \\ &\leq \exp[-t^{(1-\frac{c_0}{c})}] \int_{B_1(\epsilon)} dx, \end{aligned}$$

hence

$$\int_{B_1(\epsilon)} V^\sigma(x, t) dx \geq \int_{B_1(\epsilon)} \left( \pi_\sigma(x) - \exp[-t^{(1-\frac{c_0}{c})}] \right) dx.$$

Similarly, we obtain

$$\int_{B_2(\epsilon)} V_j^\sigma(x, t) dx \geq \int_{B_2(\epsilon)} \left( \tilde{\pi}_\sigma(x) - \exp[-t^{(1-\frac{c_0}{c})}] \right) dx.$$

Define  $t_1(\delta)$  and  $t_2(\delta)$  as

$$t_1(\delta) := \left[ -\log \left( \frac{\delta/2}{\int_{B_1(\epsilon)} dx} \right) \right]^{\frac{c}{c-c_0}}, \quad t_2(\delta) := \left[ -\log \left( \frac{\delta/2}{\int_{B_2(\epsilon)} dx} \right) \right]^{\frac{c}{c-c_0}}.$$

It follows that, for  $t > t_1(\delta)$ ,

$$\int_{B_1(\epsilon)} V^\sigma(x, t) dx \geq 1 - \delta,$$

and, for  $t > t_2(\delta)$ ,

$$\int_{B_2(\epsilon)} V_j^\sigma(x, t) dx \geq 1 - \delta.$$

Finally, from  $c > c_0 > \tilde{c}_0 > 0$ , we get  $\frac{c}{c-c_0} > \frac{c}{c-\tilde{c}_0}$ . Together with

$$\int_{B_1(\epsilon)} dx = \int_{B_2(\epsilon)} dx > \epsilon\delta/2,$$

which implies

$$\left[ -\log \left( \frac{\delta/2}{\int_{B_1(\epsilon)} dx} \right) \right] = \left[ -\log \left( \frac{\delta/2}{\int_{B_2(\epsilon)} dx} \right) \right] > 1,$$

we have

$$t_1(\delta) = \left[ -\log \left( \frac{\delta/2}{\int_{B_1(\epsilon)} dx} \right) \right]^{\frac{c}{c-c_0}} > \left[ -\log \left( \frac{\delta/2}{\int_{B_2(\epsilon)} dx} \right) \right]^{\frac{c}{c-\tilde{c}_0}} = t_2(\delta).$$

■

## 2.6 Illustration: Numerical Experiments

In order to complement the theoretical results obtained, we now present an illustration of the application of interactive diffusions to global optimization. We use Ackley's Problem See [37] and extended Michalewicz function [38] to compare the performance of interactive annealing with the standard simulated annealing method. It can be verified that these two problems satisfy our standing assumptions (1) to (4).

### 2.6.1 Ackley Problem

Ackley's Problem [37] is to find  $x \in \mathbb{R}^n$ , with  $x_i \in (-32.768, 32.768)$ , that minimizes the following function:

$$H(x) = -20 \cdot \exp(-0.2 \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \cdot \sum_{i=1}^n \cos(2\pi x_i)) + 20 + \exp(1).$$

When  $n = 3$ , the global minima for the Ackley function is  $x_* = (0, 0, 0)$  with function value  $y_* = 0$ . There are many local minima for this problem. Moreover, all local minimas have function value  $y_i > 2.1$ . Hence, the margin between global and local minima satisfies  $M \geq 2.5$ , and for any  $x$  with function value  $H(x) < 2.5$ ,  $x$  is in the attraction basin of the global minima  $x_*$ .

We set  $kC_0 = 1 < M$  for 3-dimension Ackley function, use constant temperature  $T = 5$  a ball of radius  $\rho = 0.118$  to ensure that  $H(x) \leq 0.5$  for all  $x \in B_\rho(x_*)$ .

In our implementation of the interactive diffusions approach, we have fluid  $i$  diffusing with constant temperature  $T = 5$  in a fast timescale. Convergence to the Gibbs density takes place in a relatively short time.

To gain some insight into the workings of the interactive diffusions approach we plot the objective function modified by the stationary distribution of fluid  $i$ , i.e.  $F(x) = H(x) + kC_0 \exp(-H(x)/T)$  where  $kC_0 = 1.5 < M$ . Note that for this modified objective function,  $x_* = (0, 0, 0)$  is still in the attraction basin of the new global minima. The original Ackley function, the modified Ackley function and fluid  $i$  stationary distribution are shown as Fig.2.1(Here we use 2-dimension Ackley function graph for demonstration).

We compare the performance of three methods: (1) standard simulated annealing (single thread) annealing; (2) independent diffusion (two independent, parallel threads, one running the standard simulated annealing and another with constant temperature diffusion); and (3) interactive diffusion (i.e. parallel annealing processes interacting with constant temperature diffusion in fast timescale). We set constant

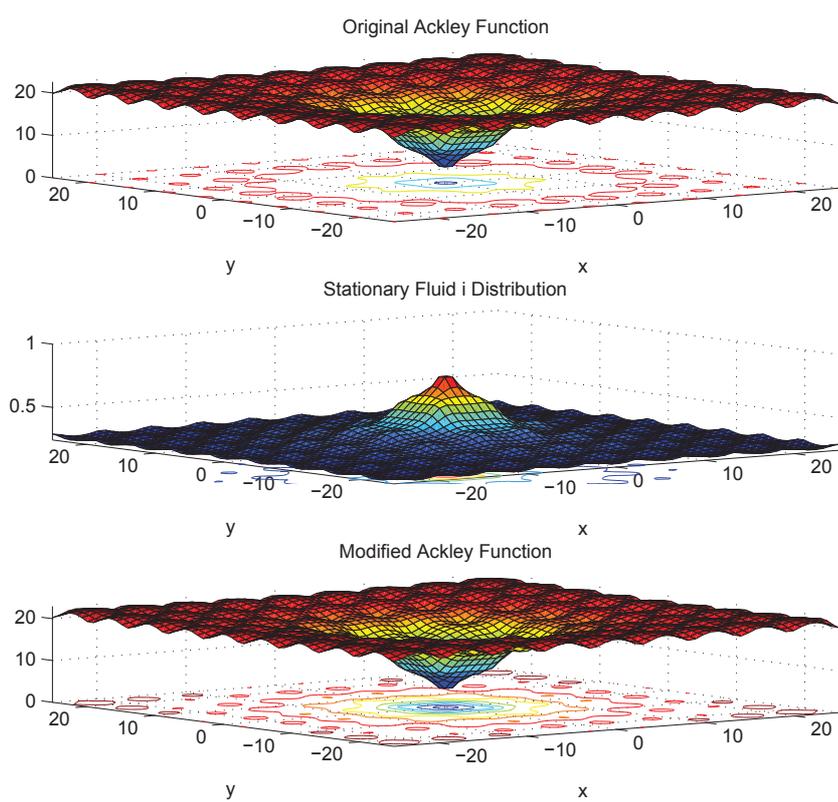


Fig. 2.1. Ackley function figures

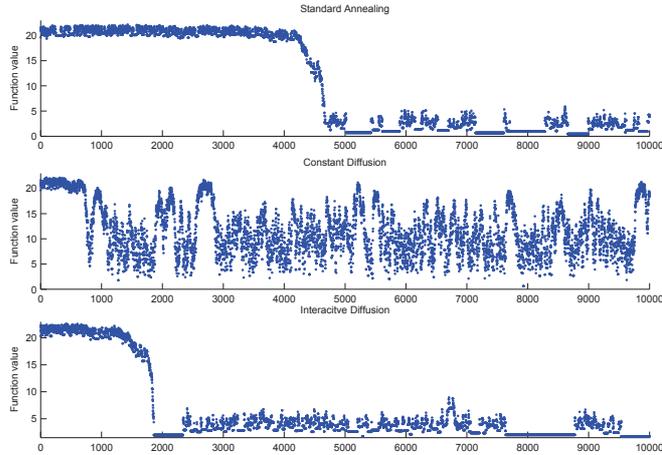


Fig. 2.2. Current function value plot for 3-dimension Ackley function

temperature  $T = 5$  for both the constant temperature diffusion in the second method (independent diffusion) and fluid  $i$  in the interactive diffusions approach.

For the standard simulated annealing method (in methods (1) and (2)) and for diffusion  $j$  (in method (3), the interactive diffusions approach) we use Boltzmann annealing with initial temperature  $c = 4$ , with logarithm temperature update without re-annealing. The convergence plot in a single run for three methods are given in current function value plot Fig. 2.2.

From Fig.2.2, we see that in a single run, the interactive diffusions method needs approximately 1800 iterations to reach the neighborhood of global minima, which is less than the approximately 4900 iterations required by the standard simulated annealing method. For constant temperature diffusion, it takes around 7900 iterations, and the current function value is (not surprisingly) not convergent toward global minima. The best performance of either the standard annealing method or the constant temperature annealing is  $\min\{4900, 7900\} = 4900$ .

To compare the speed of convergence, we report the average time (over 500 runs) to enter the  $\rho$ -neighborhood of the global minima for different temperature settings (see Table 2.1). For all three methods, we fixed initial point at  $(20, 20, 20)$ . The reported number of iterations for independent diffusion is the minimum number of iterations needed to enter the  $\rho$ -neighborhood of global minima by either the standard simulated annealing or the constant temperature annealing thread. The data supports the theoretical results regarding a speedier identification of global minima by the interactive diffusion method. This effect is more pronounced with lower initial temperatures. However, as we shall see below, the finite time performance of the interactive annealing method is superior for higher initial temperature.

Table 2.1  
Average iterations before entering  $\rho$ -neighborhood of  $x_*$  for 3-D Ackley function

| Initial Temperature $c$ | 7.5  | 9    | 10.5 | 12   | 13.5 | 15   |
|-------------------------|------|------|------|------|------|------|
| Standard Annealing      | 4696 | 5118 | 6150 | 6862 | 8336 | 9587 |
| Independent Diffusion   | 4555 | 4908 | 5843 | 6392 | 7603 | 8842 |
| Interactive Diffusion   | 4025 | 4281 | 4596 | 5188 | 6092 | 6699 |

To evaluate finite time performance (in this 3-dimension Ackley function) we run the three methods with a limit of  $10^4$  iterations. We report the probability (i.e. the empirical frequency over 500 runs) with which the three methods reached the  $\rho$ -neighborhood of global minima (see Table 2.2)

We now test the implications of increasing the dimensionality of the problem. When  $n = 5$ , the global minima for the Ackley function is  $x_* = (0, 0, 0, 0, 0)$  with function value  $y_* = 0$ , and all local minimas have function value  $y_i > 1.64$ . We set  $kC_0 = 1.2 < M$  for 5-dimension Ackley function, use constant temperature  $T = 5$  a ball of radius  $\rho = 0.176$  (again to ensure that that  $H(x) \leq 0.6$  for all  $x \in B_\rho(x_*)$ ).

Table 2.2

Probability for reaching global optima within  $10^4$  iterations for 3-D Ackley function

| Initial Temperature $c$ | 7.5   | 9     | 10.5  | 12    | 13.5  | 15    |
|-------------------------|-------|-------|-------|-------|-------|-------|
| Standard Annealing      | 93.4% | 91.6% | 84.0% | 77.6% | 68.4% | 63.0% |
| Independent Diffusion   | 93.8% | 93.0% | 85.8% | 80.6% | 73.6% | 67.0% |
| Interactive Diffusion   | 95.8% | 97.2% | 93.2% | 89.4% | 83.2% | 80.6% |

We report the average time needed (over 500 runs) to enter the  $\rho$ -neighborhood of the global minima for different temperature settings (see Table 2.3). For all processes of three methods, we fixed initial point at  $(5, 5, 5, 5, 5)$ .

Table 2.3

Average iterations before entering  $\rho$ -neighborhood of  $x_*$  for 5-D Ackley function

| Initial Temperature $c$ | 3    | 4.5   | 6     | 7.5   | 9     | 10.5  |
|-------------------------|------|-------|-------|-------|-------|-------|
| Standard Annealing      | 7279 | 18129 | 35733 | 52951 | 72908 | 79332 |
| Independent Diffusion   | 7270 | 18089 | 35484 | 52418 | 72019 | 78404 |
| Interactive Diffusion   | 3879 | 10401 | 21111 | 34508 | 52557 | 61078 |

Clearly, the gains in convergence speed by the interactive diffusion method are made even more apparent in higher dimensions. To evaluate finite-time performance (in this 5-dimension Ackley function), we report the probability (i.e. the empirical frequency over 500 runs), with which the three methods reached the  $\rho$ -neighborhood of global minima (see Table 2.4).

Table 2.4

Probability for reaching global optima within  $10^5$  iterations for 5-D Ackley function

| Initial Temperature $c$ | 3      | 4.5    | 6     | 7.5   | 9     | 10.5  |
|-------------------------|--------|--------|-------|-------|-------|-------|
| Standard Annealing      | 100.0% | 100.0% | 97.0% | 83.0% | 56.2% | 42.0% |
| Independent Diffusion   | 100.0% | 100.0% | 97.0% | 84.0% | 57.8% | 43.2% |
| Interactive Diffusion   | 100.0% | 100.0% | 99.8% | 97.6% | 83.6% | 71.6% |

### 2.6.2 Rastrigin Problem

The Rastrigin problem is the minimization of the function defined as follows:

$$F(x) := 10n + \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i)] \quad x_i \in [-5.12, 5.12], i = 1, 2, \dots, n.$$

When  $n = 3$ , the global minima for the Rastrigin function is  $x_* = (0, 0, 0)$  with function value  $y_* = 0$ . All the local minima have function value  $y_i > 0.95$ . The margin  $M$  (difference between function value of global minima and local minima) verifies  $M \geq 0.95$ . In our implementation of the interactive diffusions approach, we have fluid  $i$  diffusing with constant temperature  $T = 5$  in a fast timescale.

Again we plot the objective function modified by the stationary distribution of fluid  $i$ . The original Rastrigin function, and fluid  $i$ 's stationary distribution is shown as Fig.2.3(using 2 dimension Rastrigin function for demonstration.)

Here again, as in the previous section we compare the speed of convergence of (1) standard simulated annealing (single thread) annealing; (2) independent diffusion (two independent, parallel threads, one running the standard simulated annealing and another with constant temperature diffusion); and (3) interactive diffusion (i.e. parallel annealing processes interacting with constant temperature diffusion in fast timescale). We set the same temperature  $T = 5$  for both constant temperature

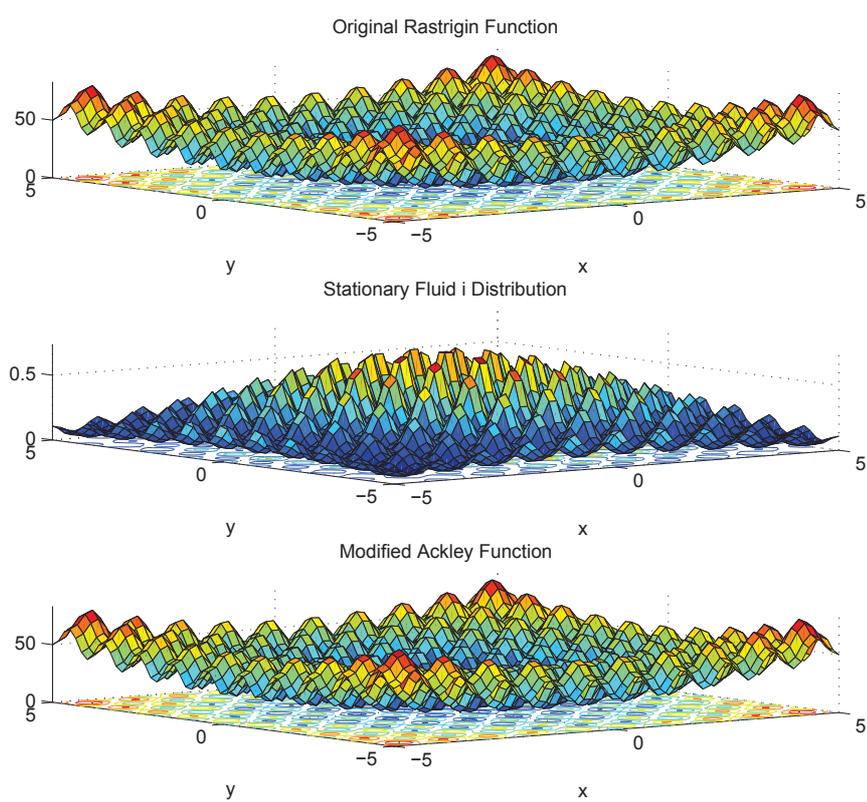


Fig. 2.3. Rastrigin function plots

diffusions in methods (2) and (3). We choose a ball of radius  $\rho = 0.05$  that  $H(x) \leq 0.5$  for all  $x \in B_\rho(x_*)$ . We report the average time (over 500 runs) to enter the  $\rho$ -neighborhood of the global minima for different temperature settings (see Table 2.5). For all processes of three methods, we fixed initial point at  $(5, 5, 5)$ .

Table 2.5  
Average iterations before entering  $\rho$ -neighborhood of  $x_*$  for 3-D Rastrigin function

| Initial Temperature $c$ | 7.5   | 9     | 10.5  | 12    | 13.5  | 15    |
|-------------------------|-------|-------|-------|-------|-------|-------|
| Standard Annealing      | 42170 | 11900 | 18458 | 39301 | 68866 | 77824 |
| Independent Diffusion   | 39785 | 11658 | 18260 | 38177 | 64837 | 73342 |
| Interactive Diffusion   | 23251 | 8885  | 14675 | 30855 | 54646 | 64104 |

To evaluate finite-time performance (in this 3-dimension Rastrigin function), we report the probability that process reached global minima  $\rho$ -neighborhood within  $10^5$  (see Table 2.6).

Table 2.6  
Probability for reaching global optima within  $10^5$  iterations for 3-D Rastrigin function

| Initial Temperature $c$ | 7.5   | 9      | 10.5   | 12     | 13.5  | 15    |
|-------------------------|-------|--------|--------|--------|-------|-------|
| Standard Annealing      | 62.2% | 99.6%  | 100.0% | 99.6%  | 70.6% | 33.4% |
| Independent Diffusion   | 66.0% | 99.8%  | 100.0% | 99.6%  | 73.8% | 40.6% |
| Interactive Diffusion   | 82.8% | 100.0% | 100.0% | 100.0% | 90.6% | 53.2% |

## Discussion

The limited computational evidence presented in this section confirms the theoretical prediction of the interactive diffusion approach's faster speed of convergence.

The results suggest that the relative gains in convergence speed are increasing with decreasing initial temperatures. It is often the case that practitioners using simulated annealing would attempt to reduce the value of the initial temperature as this parameter serves as a proxy for available computational budget. In this sense, the proposed method works best when there is a limited computational budget (in the form of a relatively low value for initial temperature). The finite-time performance evaluation is conducted by imposing a cap on the number of iterations. Here again, the interactive annealing method proposed exhibits a higher probability of finding the global optima.

### 2.6.3 Comparison with Parallel Simulated Annealing

We now compare the performance of independent parallel implementations of simulated annealing suggested in [39–41]. Let  $P_{c_i}$  be probability of single annealing threads reaching global optima within iteration cap for initial temperature  $c_i$ . Consider two annealing threads, let  $Q_{c_i}$  be the probability of one of paralleled annealing threads entering global optima. We have

$$Q_{c_i} = 1 - (1 - P_{c_i})^2,$$

the result shown in Table 2.7.

With only two threads, independent simulated annealing processes may outperform the interactive diffusions approach in certain cases. This can be explained as follows. In the interactive diffusions approach with two threads, one thread is "sacrificed" by having constant diffusion in order to speed up the interactive annealing thread. With only two threads, the opportunity cost of running a constant diffusion

Table 2.7  
Probability for 2 trivial paralleled annealing threads reaching global optima within iterations cap

|                                 |        |        |        |        |       |       |
|---------------------------------|--------|--------|--------|--------|-------|-------|
| Initial Temperature $c$         | 7.5    | 9      | 10.5   | 12     | 13.5  | 15    |
| 3-D Ackley Func., cap $10^4$    | 99.6%  | 99.3%  | 97.4%  | 95.0%  | 90.0% | 86.3% |
| Initial Temperature $c$         | 3      | 4.5    | 6      | 7.5    | 9     | 10.5  |
| 5-D Ackley Func., cap $10^5$    | 100.0% | 100.0% | 99.9%  | 97.1%  | 80.8% | 66.4% |
| Initial Temperature $c$         | 7.5    | 9      | 10.5   | 12     | 13.5  | 15    |
| 3-D Rastrigin Func., cap $10^5$ | 85.7%  | 100.0% | 100.0% | 100.0% | 91.4% | 55.6% |

thread to speed up a second diffusion is too high. However, with more reactive threads this cost is diluted. To make this point we now present the simulation results with five (5) threads (one (1) constant diffusion and four (4) interactive diffusion threads) and compare with the performance of five (5) independent simulated annealing processes. The performance is evaluated on the basis of (i) average iterations before first thread enters  $\rho$ -neighborhood of  $x_*$  in Table 2.8, Table 2.9 and Table 2.10 and (ii) the probability of 1 of 5 threads entering global optima within the total number of iterations in Table 2.11, Table 2.12 and Table 2.13.

Table 2.8  
Average iterations before 1 of 5 threads entering  $\rho$ -neighborhood of  $x_*$  for 3-D Ackley function

|                         |      |      |      |      |      |      |
|-------------------------|------|------|------|------|------|------|
| Initial Temperature $c$ | 7.5  | 9    | 10.5 | 12   | 13.5 | 15   |
| Paralleled Annealing    | 1916 | 2093 | 2207 | 2364 | 2806 | 3138 |
| Interactive Diffusion   | 1770 | 1887 | 1971 | 2270 | 2429 | 2559 |

Table 2.9

Average iterations before 1 of 5 threads entering  $\rho$ -neighborhood of  $x_*$  for 5-D Ackley function

| Initial Temperature $c$ | 3    | 4.5  | 6     | 7.5   | 9     | 10.5  |
|-------------------------|------|------|-------|-------|-------|-------|
| Paralleled Annealing    | 2399 | 5881 | 11049 | 17640 | 27869 | 41975 |
| Interactive Diffusion   | 1645 | 4072 | 7539  | 13222 | 20178 | 27819 |

Table 2.10

Average iterations before 1 of 5 threads entering  $\rho$ -neighborhood of  $x_*$  for 3-D Rastrigin function

| Initial Temperature $c$ | 7.5  | 9    | 10.5 | 12    | 13.5  | 15    |
|-------------------------|------|------|------|-------|-------|-------|
| Paralleled Annealing    | 3403 | 4450 | 8911 | 17767 | 32681 | 30926 |
| Interactive Diffusion   | 2251 | 3860 | 8124 | 15718 | 22697 | 20783 |

Table 2.11

Probability for 1 of 5 threads reaching global optima within 5000 iterations for 3-D Ackley function

| Initial Temperature $c$ | 7.5    | 9     | 10.5  | 12    | 13.5  | 15    |
|-------------------------|--------|-------|-------|-------|-------|-------|
| Paralleled Annealing    | 98.8%  | 98.6% | 96.8% | 94.6% | 88.8% | 84.8% |
| Interactive Diffusion   | 100.0% | 99.4% | 97.6% | 96.8% | 93.2% | 91.8% |

Table 2.12

Probability for 1 of 5 threads reaching global optima within  $10^4$  iterations for 5-D Ackley function

| Initial Temperature $c$ | 3      | 4.5   | 6     | 7.5   | 9     | 10.5  |
|-------------------------|--------|-------|-------|-------|-------|-------|
| Paralleled Annealing    | 100.0% | 83.6% | 51.8% | 33.6% | 19.0% | 12.2% |
| Interactive Diffusion   | 100.0% | 96.2% | 73.4% | 45.8% | 29.6% | 23.0% |

Table 2.13  
Probability for 1 of 5 threads reaching global optima within  $10^4$  iterations for 3-D Rastrigin function

| Initial Temperature $c$ | 7.5    | 9     | 10.5  | 12    | 13.5  | 15    |
|-------------------------|--------|-------|-------|-------|-------|-------|
| Paralleled Annealing    | 96.7%  | 98.0% | 61.3% | 21.3% | 31.0% | 57.0% |
| Interactive Diffusion   | 100.0% | 99.7% | 69.3% | 26.0% | 44.7% | 67.7% |

With five (5) threads, we observe the interactive diffusions approach exhibits better performance than independent simulated annealing processes. This is a further indication of the proposed method’s improved speed of convergence.

## 2.7 Chapter Summary

Simulated annealing for global optimization is a well-known method for global optimization. Its main drawback pertains to the speed at which emphasis on “exploration” gives way to “exploitation”. This transition can not be “too fast” in order to guarantee convergence to globally optimal solutions. Evidently, with faster computing this drawback may become less critical. However, computing speed (which used to double every couple of years) has stopped increasing because as chips are clocked at higher speeds they become difficult to cool and much less energy-efficient. Parallel implementation of annealing-like search for global optima appears as a sensible approach for speeding up convergence to globally optimal solutions.

In this chapter, we have introduced a novel approach, in which parallel annealing processes interact in a manner which expedites the identification of a globally optimal solution. A first annealing process operates at a faster timescale and has a drift function that converges to a non-zero (but relatively small) noise level. A second

annealing process (operating at slower timescale) is subject to a modified drift term, in which the steepest descent direction is perturbed with the first process density gradient. This additional term ensures the second annealing process is “repelled” areas explored by the first process. As a result, the second annealing process is able to *bypass* locally optimal solutions that its own “cooling schedule” can be decreased. We have shown that when compared to independent diffusions, the proposed interactive diffusions approach can increase the speed of convergence at the expense of minimal additional computational overhead.

In a limited computational testbed, we provide numerical illustration of the (theoretical) speed-up effect. These numerical experiments suggest the relative gains in convergence speed are increasing with decreasing initial temperatures. Since the initial temperature often serves as a proxy for available computational budget, the proposed method works best when there is a limited computational budget in the form of a relatively low value for initial temperature. The finite-time performance evaluation also consistently indicates that the interactive annealing method exhibits a higher probability of finding the global optima in finite-time. Finally, the numerical illustrations suggest the relative performance of interactive annealing improves with higher dimensions.

There are a number of outstanding research questions that we continue to explore. First, in this chapter we have limited our analysis to the case of *two* interacting annealing processes. It would be desirable to characterize the extent to which faster convergence to globally optimal solutions is obtained as a function of *several* interacting annealing processes. Given the computational overhead involved, this type of analysis could help identify the “optimal” number of interacting annealing processes. Secondly, in this chapter, we have only considered one form of interaction amongst annealing processes, namely, one in which a second (slower) annealing process is re-

pelled by the density of the first (faster) process. In addition to a repulsive potential, the incorporation of an attractive potential in the differential equations governing the interaction across annealing process could conceivably yield a speedier exploration of neighborhoods of promising solutions. In a manner analogous to the results obtained in this chapter, this may yield faster identification of globally optimal solutions.

### 3. INTERACTIVE MODEL-BASED SEARCH WITH REACTIVE RESOURCE ALLOCATION

#### 3.1 Introduction and Literature Review

We consider a parallel computing scheme for global optimization that combines multi-start local search with the dynamic reallocation of computational resources (e.g. processing time). Our work builds upon the interactive model-based approach to global optimization proposed in [21] Wang and Garcia (2015) in which parallel threads independently execute a model-based search method (see [42]) and periodically interact through a simple acceptance-rejection rule aimed at preventing duplication of search efforts. In a model-based search method, the distribution of re-start points is adjusted at each iteration upon evaluating local search results which informs the selection of a new “model” (i.e. probability distribution) over the feasible region. This model is in turn used to randomly generate new re-start points. The degree to which the new probability distribution (or “model”) is concentrated around the best solutions identified so far reflects the relative emphasis on exploitation versus exploration. Diversity in multiple re-start points (i.e. exploration) is a desirable trait as it provides a form of insurance against operating with a poor model. However, too much diversity may slow down the identification of globally optimal solutions which could be accelerated by selecting models that lead to increased search effort in promising areas as determined by history (i.e. exploitation). This description encapsulates a wide variety of stochastic methods in the literature based upon a multi-start strategy featuring different resolutions to the exploration vs. exploitation tradeoff ([43])

and [44]). Invariably, in single-thread approaches to global optimization based on stochastic multi-start search, exploitation and exploration are substitutes. The main point of [21] is that in a parallel computing environment when duplication of search effort is prevented (or limited), exploitation and exploration are complements and not substitutes. This is shown to be the case as when the models governing each thread's multi-start local search strategy are subject to an acceptance-rejection rule. Assuming each thread successfully identifies a locally optimal solution every time the acceptance-rejection rule is implemented, it was shown in [21] that the rate of convergence to a globally optimal solution exponentially increases in the number of threads.

In practice however, the computational time required to identify a locally optimal solution varies greatly. Therefore, when the acceptance-rejection rule is implemented, several threads may fail to identify a locally optimal solution. Thus the main result in [21] relies on a highly stylized model of computational time. In this chapter we consider an implementation of the interactive model-based approach that accounts for real time; that is, it takes into account the possibility that several threads may fail to identify a locally optimal solution whenever the acceptance-rejection rule is implemented. We propose a modified acceptance-rejection rule that alternates between enforcing diverse search –in order to prevent duplication– and reallocation of computational effort –in order to speed up the identification of local optima– when one or more threads repeatedly fail to do so. We show that the rate of convergence in real-time increases with the number of threads. This result formalizes the idea that in parallel computing, exploitation and exploration are complements and not substitutes (as in most single-thread approaches to global optimization).

The promise of parallel computing for global optimization is more than ever a reality as computer manufacturers have continued to introduce more cores per chip and graphics processing units (GPUs) are increasingly popular. This trend implies significant multi-thread processing power is readily accessible to optimization practitioners which no longer need sophisticated or overly expensive infrastructure to run parallel algorithms for solving global optimization problems. This has motivated recent studies aiming to develop paralleled implementation of well-known global optimization algorithms (see for example, [45] and [46] for simulated annealing and [9] for particle swarm algorithm).

Parallel computing approaches to global optimization vary depending upon the level of coordination among threads (see [47] for a survey). Without any coordination among threads, a judicious choice of stopping rules is needed to fully accrue the benefits of parallelization (see [48]). Often some degree of coordination is desirable as real-time information by different threads can be used to improve performance at the expense of overhead. This is for example the case of a parallel implementation of simulated annealing (see [45] and [46]). Some degree of coordination also enables the real time re-allocation of computational resources among several instances of search algorithms (see for example, [49]) in order to improve performance. The optimal real-time allocation of computational resources can be modeled as a non-stationary multiple-armed bandit problem which –in and of itself– may be as complex as the underlying global optimization problem. For example, in [49] certain regularity assumptions are needed to obtain asymptotic bounds on algorithm’s performance.

In this chapter, we develop a real-time reallocation strategy that is based upon historical performance. There is no attempt at using sophisticated algorithmic variations in order optimally react to search outcomes. Instead, the main idea is to

leverage relatively simple ideas such as *i*) continuing searches that are promising because the end-points have lower objective values than all other solutions found so far and *ii*) avoiding duplication of failed searches and/or search effort across threads. The relatively small gains afforded by these simple ideas are then shown to be magnified by parallelization. Indeed, we show the rate of convergence for an interactive model-based search increases with the number of threads for a wide-class of local search techniques (i.e. model-based) when compared to independent parallel implementation. The structure of this chapter is as follows. In section 2 we review the single-thread model based search and provide a characterization of improved performance when the algorithm reacts to incomplete search outcomes. In section 3, we analyze an interactive multi-thread approach that in addition to reacting to incomplete searches (at each thread) incorporates a way to avoid duplication of failed searches and/or search effort across threads. In our main result we show the interactive scheme speeds up the search for global optimal solutions, i.e. the time needed to identify a global solutions decreasing with the number of threads. In section 4 we illustrate this effect by means of a computational testbed.

### 3.2 Single Thread Model-based Search in Real-Time

Consider a general optimization problem  $\min\{f(x) : x \in \Omega\}$  where  $f : \mathbb{R}^n \mapsto \mathbb{R}$  is continuous and  $\Omega \subset \mathbb{R}^n$  is such that  $X^* = \arg \min_{x \in \Omega} f(x)$  is well-defined. Assume further  $f$  has  $N$  local (non-global) minima, say  $X = \{x_1, x_2, \dots, x_N\}$ , that is,

$$f(x_i) \leq f(x) \quad \forall x \in N(x_i, \epsilon_i)$$

for some  $\epsilon_i > 0$  where  $N(x_i, \epsilon_i) = \{x \in \mathbb{R}^n \mid \|x - x_i\| \leq \epsilon_i\}$ .

The multi-start method that we shall describe later makes use of a local search algorithm. This algorithm takes an initial solution or “seed” as input, say  $x \in \Omega$  and produces an output in the form of a local minimum, say  $y \in X \cup X^*$ . The local search algorithm determines a map  $\ell : \Omega \rightarrow X \cup X^*$  such that  $\ell(x) = y$  if  $x \in B(y)$  where  $B(y) \subset \Omega$  is the “basin of attraction” of local minimum  $y \in X \cup X^*$ , i.e.:

$$B(y) := \{x \in \Omega \mid \ell(x) = y\}$$

The properties of the local search algorithm (e.g. the size of basins of attraction) are unknown. However, we assume the basins of attractions partition the solution space  $\Omega$ :

**Assumption 1:**  $B(x_i) \cap B(x_j) = \emptyset$ , for all  $x_i \neq x_j \in X \cup X^*$  and

$$\bigcup_{x_i \in X \cup X^*} B(x_i) = \Omega.$$

This assumption states that given any initial input on the solution space, the local search algorithm will produce one and only one local minimum as the output. Additionally, the local search algorithm is deterministic, i.e. the same output is always obtained when provided the same input.

In what follows we revisit the basic iteration scheme in Wang and Garcia (2015) so that each iteration is equivalent to  $T_0 > 0$  units of computational time. Let  $T(x)$  be the time required by local search  $\ell$  from  $x$  to identify  $x_i$  if  $x \in B(x_i)$ . We modify definition of operation  $\ell$  for real time as: if  $T(x) \leq T_0$ ,  $\ell(x)$  returns local optima  $x_i$  when  $x \in B(x_i)$ ; if  $T(x) > T_0$ ,  $\ell(x)$  returns ending location of local search from  $x$  at time  $T_0$ .

### 3.2.1 Basic Single-Thread Computation

Let  $\mathfrak{G}$  denote a class of probability density functions with support  $\Omega$  and  $g(J)$  the current “model”. As in [21] we assume that any  $g \in \mathfrak{G}$ ,  $g(x) \neq 0$  for  $x \in \Omega$  almost surely. Taking into account the computational time limit  $T_0$ , the basic iteration for each thread is as follows:

1. A sample  $x$  from the current “model”  $g(J)$  is drawn and a local search algorithm is launched.
2. At time  $T_0$ , the resulting state of information is

$$J' = \begin{cases} J \cup \ell(x) & \text{if } T(x) \leq T_0 \\ J & \text{otherwise} \end{cases}$$

3. A new model  $g(J') \in \mathfrak{G}$  is selected as follows:

$$g(x, J') = \arg \min_{g \in \mathfrak{G}} D_{KL}(h(x; J'), g) \quad (3.1)$$

where

$$h(x; J') = \frac{I(f(x), J')U(x)}{\int_{\Omega} I(f(x), J')U(x)dx}$$

where  $D_{KL}$  is the Kullback-Leibler divergence,  $U$  is the uniform probability density function on  $\Omega$  and the reference function  $I$  is defined as:

$$I(f(x), J') = \begin{cases} 1 & f(x) \leq \min_{x \in J'} f(x) + \epsilon \\ 0 & f(x) > \min_{x \in J'} f(x) + \epsilon \end{cases}$$

for  $\epsilon > 0$ .

The lower the value of  $\epsilon$  the more probability mass the reference density function posits around the best locally optimal solutions in the state of information  $J'$  and thus the selection of a new model emphasizes exploitation over exploration. To account for the possibility that the local search procedure may fail to identify a locally optimal solution within the allotted time  $T_0$  we define the set  $\mathfrak{C}(T_0) = \{x : T(x) \leq T_0\}$ . We obtain a Markov chain model for  $\{J^{nT_0} : n > 0\}$  with transition probabilities:

$$\Pr(J' | J) = \begin{cases} \int_{B(x) \cap \mathfrak{C}(T_0)} g(y; J) dy & J' = J \cup \{x\} \text{ and } x \in \{X \cup X^*\} \setminus J \\ \sum_{x \in J} \int_{B(x) \cap \mathfrak{C}(T_0)} g(y; J) dy \int_{\Omega \setminus \mathfrak{C}(T_0)} g(y; J) dy & J' = J \end{cases}$$

Define  $\mathcal{J}^*$  as the class of states with at least a globally optimal solution, i.e.  $J \in \mathcal{J}^*$  if and only if  $J \cap X^* \neq \emptyset$ . It follows that  $J \notin \mathcal{J}^*$ ,

$$\Pr(\mathcal{J}^* | J) = \int_{B(X^*) \cap \mathfrak{C}(T_0)} g(y; J) dy$$

**Lemma 1** *The transition probability matrix for the Markov chain  $\{J^{nT_0} : n > 0\}$  is upper triangular. The eigenvalues are  $\lambda_J = \Pr(J|J)$  and  $\Pr(\mathcal{J}^*|\mathcal{J}^*) = 1$ . Let  $\pi_J^{nT_0} = \Pr(J^{nT_0} = J|J^0)$  be the distribution at time  $nT_0$ , we have*

$$|\pi_{\mathcal{J}^*}^{nT_0} - 1| \leq C \lambda_{[2]}^n$$

where  $\lambda_{[2]} \in (0, 1)$  is the second-largest eigenvalue.

**Proof** The proof is essentially the same as in [21] Lemma 1 and Theorem 1 with modified transition probabilities accounting for the probability of incomplete search outcomes. ■

### 3.2.2 Reacting to Incomplete Searches

We consider a variation in which an incomplete search leading to a “high quality” end-point is allowed to continue. Specifically, if the objective function value associated with the end-point of an incomplete search is lower than the values associated with *all* discovered local optima, this end-point must be in the attraction basin of an undiscovered local optima. Hence, the search should be allowed to continue. The iteration of modified single-thread model based search is modified as follows:

1. A sample  $x$  from the current “model”  $g$  is drawn and a local search algorithm is launched.
2. At time  $T_0$ , the resulting state of information is

$$J' = \begin{cases} J \cup \{\ell(x)\} & T(x) \leq T_0 \\ J & T(x) > T_0 \end{cases}$$

Let  $y$  denote the end-point if search is incomplete.

3. A new model  $g' \in \mathfrak{G}$  is selected as follows:

$$g' = \begin{cases} g(J') & f(y) \geq \min_{x \in J'} f(x) \\ \mathbf{1}_y & f(y) < \min_{x \in J'} f(x) \end{cases} \quad (3.2)$$

$g(J')$  is computed as in (1) and  $\mathbf{1}_y$  is Dirac’s density on  $y$ .

Let  $y^{kT_0}$  denote the end point if search is incomplete at time  $kT_0$  where  $y^{kT_0} = \emptyset$  if the search is completed. The modified single-thread model based search is no longer

a *stationary* Markov-chain. The one-step transition probability matrices  $\{P^{nT_0} : n > 0\}$  is a stochastic process adapted to the filtration generated by the process  $\{y^{nT_0} : n > 0\}$ . Note that  $P^{nT_0}$  maintains upper triangular structure so that its eigenvalues (entries along diagonal) are of the form:

$$\lambda_J^{nT_0} = \Pr(J^{nT_0} = J | J^{(n-1)T_0} = J)$$

Let  $\bar{P}^{nT_0}$  denote the product of one-step transition probability matrices:

$$\bar{P}^{nT_0} = \prod_{k>0}^n P^{kT_0}$$

Note that  $\bar{P}^{nT_0}$  is also upper triangular so that its eigenvalues are the entries along the diagonal, i.e. they are of the form:

$$\bar{\lambda}_J^{nT_0} = \Pr(J^{nT_0} = J | J^0 = J) = \prod_{k>0}^n \lambda_J^{kT_0}$$

In our main result of this section we show that reacting to incomplete local searches provides an improved rate of convergence with probability 1. We will assume that the computational time required to identify any locally optimal solution is uniformly bounded.

**Assumption 2:**  $\bar{n} = \sup_{y \in \Omega} \lfloor \frac{T(y)}{T_0} \rfloor < \infty$ .

**Theorem 3.2.1** For all  $n \geq \bar{n}$  and  $J \in 2^X$ ,

$$(\lambda_J)^{n-\bar{n}} \geq \bar{\lambda}_J^{nT_0}$$

with probability one.

**Proof** We start by characterizing the process  $\{\lambda_J^{nT_0} : n > 0\}$  as a function of the history of the process  $\{y^{nT_0} : n > 0\}$ . Conditional upon  $J^{(n-1)T_0} = J$  and  $y^{nT_0}$ :

$$\lambda_J^{nT_0} = \begin{cases} \lambda_J & f(y^{nT_0}) \geq \min_{x \in J} f(x) \quad \text{or} \quad y^{nT_0} = \emptyset \\ 1 & f(y^{nT_0}) < \min_{x \in J} f(x) \quad \text{and} \quad T(y^{nT_0}) > T_0 \\ 0 & f(y^{nT_0}) < \min_{x \in J} f(x) \quad \text{and} \quad T(y^{nT_0}) \leq T_0 \end{cases}$$

To see why this is, recall that if the search is complete (i.e.  $y^{nT_0} = \emptyset$ ) or the end-point is not of “high quality” (i.e.  $f(y^{nT_0}) \geq \min_{x \in J} f(x)$ ) the single-thread model is not affected by the search outcome. The remaining cases are associated with an unsuccessful (respectively, successful) continuation of an incomplete search.

By Assumption 2, conditional upon  $y^{T_0}$  (the initial search outcome), we have

$$\bar{\lambda}_J^{nT_0} = \begin{cases} 0 & f(y^{T_0}) < \min_{x \in J} f(x) \\ \lambda_J \bar{\lambda}_J^{(n-1)T_0} & \text{otherwise} \end{cases}$$

for  $n \geq \bar{n}$ . Hence, by induction

$$\bar{\lambda}_J^{nT_0} \leq (\lambda_J)^{n-\bar{n}} \prod_{k>\bar{n}} \lambda_J^{kT_0} \leq (\lambda_J)^{n-\bar{n}}.$$

■

### 3.3 Interactive Model-based Search in Real Time

Having modified the single-thread model based search to react to incomplete searches we now consider a multiple-thread implementation, with  $M$  threads and denoting by  $\mathbf{J} = (J_1, \dots, J_M)$  the joint state of information. To avoid duplication of search efforts, all threads report their search outcome after  $T_0$  units of computational

time and the selection of a new model associated is subject to an acceptance-rejection test. The basic iteration of interactive model based search is as follows:

1. A sample  $y_i$  is drawn from the current model  $g_i(J_i)$  and a local search algorithm is launched for each thread  $i$ .
2. After  $T_0$  units of computational time, the current state  $J_i$  is updated as

$$J'_i = \begin{cases} J_i \cup \ell(y_i), & \text{if } T(y_i) \leq T_0 \\ J_i, & \text{otherwise} \end{cases}$$

3. A tentative new model  $g_i(J'_i) \in \mathfrak{G}$  is selected by each thread  $i$  as in (1). The acceptance-rejection test is implemented so that the new model  $g'_i \in \mathfrak{G}$  is determined as follows:

$$g'_i = \begin{cases} g(J'_i) & D_{KL}(g(J'_i), g(J'_\ell)) > \eta \quad \forall \ell > i \\ \tilde{g}(J'_i) & \text{otherwise} \end{cases}$$

where  $\tilde{g}(J'_i) \in \arg \min_{g \in \mathfrak{G}} D_{KL}(\tilde{h}(J'_i), g)$  and

$$\tilde{h}(J) = \frac{I(x, J)U(x)}{\int_{\Omega} I(x, J)U(x)dx}$$

$$I(x, J) = \begin{cases} 1, & x \notin B(J) \cap \mathfrak{C}(T_0) \\ 0, & \text{otherwise} \end{cases}$$

Here the intent of the reference density  $\tilde{h}$  is to choose a new model positing probability in areas not in the basin of attraction of the locally optimal solution in the state  $J'_i$ . In practice, an approximate solution to KL divergence minimization problem is needed. This is done by using moments of all empirical distributions. For

example, when the class of sampling densities is normal with fixed variance then the empirical mean is used to find the distribution that approximately minimizes the KL divergence.

In what follows we shall use an independent multi-thread implementation as a benchmark. The following lemma which is adapted from [21] characterizes the performance of this scheme.

**Lemma 2** *The interactive model based search can be modeled as a Markov chain  $\{\mathbf{J}^{nT_0} : n > 0\}$  with an upper triangular transition probability matrix so that the eigenvalues are of the form  $\lambda_{\mathbf{J}}^* = \Pr(\mathbf{J}|\mathbf{J})$ . The process with  $M$  independent threads is also a Markov chain with eigenvalues of the form  $\lambda_{\mathbf{J}} = \prod_{i \leq M} \Pr(J_i|J_i)$ . Moreover,*

$$(\lambda_{[2]})^M \geq \lambda_{\mathbf{J}} \geq \lambda_{\mathbf{J}}^*$$

where  $\lambda_{[2]}$  denotes the second largest eigenvalue of the single-thread model-based search.

**Proof** See [21] Theorem 3 and replace all  $B(J_i)$  terms to  $B(J_i) \cap \mathfrak{C}(T_0)$ . ■

### 3.3.1 Reactive Resource Allocation: Avoiding Duplication of Incomplete Search

In this section we propose a modification to the interactive model-based search scheme to reallocate resources in response to real-time information. We will use a relatively simple reactive strategy based upon the following ideas. Start points (or seeds) that have led to incomplete local searches should not continue to be used. We will show that the effect of this reallocation strategy is to speed up the identification of local optima when one or more threads repeatedly fail to do so. The reactive

reallocation strategy outlined above operates in a different (slower) time-scale  $\{nT_0 : n > 0\}$ .

To motivate the analysis, recall that for single-thread implementation of model-based search the eigenvalues of the associated Markov chain are of the form:

$$\Pr(J_i | J_i) = \sum_{x \in J_i} \int_{B(x) \cap \mathfrak{C}(T_0)} g(y; J_i) dy + \int_{\Omega \setminus \mathfrak{C}(T_0)} g(y; J_i) dy$$

The reallocation strategy outlined above reduces the magnitude of this eigenvalue by modifying  $g(\cdot; J_i)$  in order to reduce the probability to have new sample in  $\Omega \setminus \mathfrak{C}(T_0)$  where local search cannot finish in time  $T_0$ . In what follows, we shall describe how to achieve the goal by introducing a “repulsive” force to the start location of unfinished local search.

We will describe how to incorporate the need to avoid duplication in incomplete search effort into the overall interactive model-based search scheme. Suppose that upon executing the interactive model-based algorithm we keep track of the set (say  $K_s$ ) of starting points that have led to incomplete searches. In light of Assumption 1 we have:

$$\Omega \setminus \mathfrak{C}(T_0) = \cup_{x_i \in X \cup X^*} \tilde{B}(x_i)$$

where  $\tilde{B}(x_i) = B(x_i) \cap \Omega \setminus \mathfrak{C}(T_0)$ . Hence, for each  $y \in K_s$  there exists a unique  $x_i \in X \cup X^*$  such that  $y \in B(x_i) \cap \Omega \setminus \mathfrak{C}(T_0)$ . With a slight abuse of notation, in what follows, we shall refer to  $\tilde{B}(x_i)$  as  $\tilde{B}(y)$ . Consider the modified reference density  $H$  defined as follows:

$$H(x, J, K_s) = \frac{I(x, J, K_s)U(x)}{\int_{\Omega} I(x, J, K_s)U(x)dx}$$

where

$$I(x, J, K_s) = \begin{cases} 1, & x \notin \cup_{y \in K_s} \tilde{B}(y) \cup B(J), \\ 0, & \text{otherwise} \end{cases}$$

In order to minimize the likelihood of sampling a starting point that would lead to either no new locally optimal information (state  $J$ ) or another incomplete search outcome (the state  $K_s$ ) the new candidate model  $\tilde{g}$  is computed as follows:

$$\tilde{g}(J, K_s) \in \arg \min_{g \in \mathfrak{G}} [D_{KL}(H(x, J, K_s), g)]$$

Thus, sampling from this density gives a high likelihood to starting points or “seeds” that are more likely to result in new locally optimal solutions identified while avoiding incomplete searches.

We can now formally describe the modified interactive model-based search. Let  $t = nT_0$ , for  $M$  threads, with  $\mathbf{K}_s^t = (K_{s,1}^t, \dots, K_{s,M}^t)$  and  $\mathbf{J}^t = (J_1^t, \dots, J_M^t)$  as the current states, the interactive model-based search scheme with reactive allocation can be succinctly described as follows:

1. A sample  $y_i$  is drawn from the current model  $g_i^t$  and a local search algorithm is launched for each thread.
2. At time  $t + T_0$ , the result state is updated as

$$J_i^{t+T_0} = \begin{cases} J_i^t \cup \ell(y_i), & \text{if } T(y_i) \leq T_0 \\ J_i^t, & \text{otherwise} \end{cases}$$

The set of starting points leading to incomplete searches is updated as

$$K_{i,s}^{t+T_0} = K_{i,s}^t \cup \{y_i \mid T(y_i) > T_0\}$$

and record end-point  $y_i^{t+T_0}$

3. A new model  $g(J_i^{t+T_0}) \in \mathfrak{G}$  is selected as in (2). If  $y_i^{t+T_0} \geq \min_{x \in J_i^{t+T_0}} f(x)$ , the acceptance-rejection test is implemented and the resulting model  $g_i^{t+T_0}$  is computed as follows:

$$g_i^{t+T_0} = \begin{cases} g(J_i^{t+T_0}) & D_{KL}(g(J_i^{t+T_0}), g_\ell(J_\ell^{t+T_0})) > \eta \quad \forall \ell > i \\ \tilde{g}_i(J_i^{t+T_0}, K_{i,s}^{t+T_0}) & \text{Otherwise} \end{cases}$$

where  $\tilde{g}_i(J_i^{t+T_0}, K_{i,s}^{t+T_0}) \in \arg \min_{g \in \mathfrak{G}} D_{KL}(H(J_i, K_{i,s}^{t+T_0}), g)$ .

### 3.3.2 Analysis

Having enlarged the state of information to include points leading to incomplete searches, the interactive model based search can no longer be modeled as a *stationary* Markov-chain. In fact, the sequence of one-step transition probability matrix  $\{\mathbf{P}^{nT_0} : n > 0\}$  is a stochastic process adapted to the history of search outcomes, i.e. the processes  $\{\mathbf{K}_s^{nT_0} : n > 0\}$  and  $\{y_i^{nT_0} : n > 0\}$  for each thread  $i$ . Note that  $\mathbf{P}^{nT_0}$  maintains upper triangular structure so that its eigenvalues (entries along diagonal) are of the form:

$$\lambda_{\mathbf{J}}^{nT_0} = \Pr(\mathbf{J}^{nT_0} = \mathbf{J} | \mathbf{J}^{(n-1)T_0} = \mathbf{J})$$

Let  $\bar{\mathbf{P}}^{nT_0}$  denote the product of one-step transition probability matrices given the history of starting points resulting in incomplete local searches, i.e.:

$$\bar{\mathbf{P}}^{nT_0} = \prod_{k>0}^n \mathbf{P}^{kT_0}$$

Note that  $\bar{\mathbf{P}}^{nT_0}$  is also upper triangular so that its eigenvalues are the entries along the diagonal, i.e. they are of the form:

$$\bar{\lambda}_{\mathbf{J}}^{nT_0} = \Pr(\mathbf{J}^{nT_0} = \mathbf{J} | \mathbf{J}^0 = \mathbf{J}) = \prod_{k>0}^n \lambda_{\mathbf{J}}^{kT_0}$$

In the following result we show in finite time the probability of identifying a globally optimal solution is increased when the finite history of incomplete local searches is used to avoid an incomplete search.

**Theorem 3.3.1** *For any  $n > \bar{n}$ ,  $(\lambda_{\mathbf{J}}^*)^{n-\bar{n}} \geq \bar{\lambda}_{\mathbf{J}}^{nT_0}$  with probability one.*

**Proof** For any time  $nT_0$  and any thread  $i \leq M$ ,

$$\lambda_{J_i}^{nT_0} = \begin{cases} \Lambda_{J_i}(\hat{g}_i^{nT_0}, T_0) & f(y_i^{nT_0}) \geq \min_{x \in J_i} f(x) \quad \text{or} \quad y_i^{nT_0} = \emptyset \\ 1 & f(y_i^{nT_0}) < \min_{x \in J_i} f(x) \quad \text{and} \quad T(y_i^{nT_0}) > T_0 \\ 0 & f(y_i^{nT_0}) < \min_{x \in J} f(x) \quad \text{and} \quad T(y_i^{nT_0}) \leq T_0, \end{cases}$$

where

$$\hat{g}_i^{nT_0} = \begin{cases} g(J_i^{t+T_0}) & D_{KL}(g(J_i^{t+T_0}), g_\ell(J_\ell^{t+T_0})) > \eta \quad \forall \ell > i \\ \tilde{g}_i(J_i^{t+T_0}, K_{i,s}^{t+T_0}) & \text{Otherwise} \end{cases}$$

and  $g(J_i^{t+T_0}) \in \mathfrak{G}$  is selected as in (1) instead of (2). Define  $\hat{\lambda}_{\mathbf{J}}^{nT_0} = \prod_{i \leq M} \Lambda_{J_i}(\hat{g}_i^{nT_0}, T_0)$ , from theorem 1,

$$\prod_{k=1}^{n-\bar{n}} \hat{\lambda}_{\mathbf{J}}^{kT_0} \geq \bar{\lambda}_{\mathbf{J}}^{nT_0}, \quad \text{with probability one.}$$

For any  $n > \bar{n}$ , one sufficient condition of  $(\lambda_{\mathbf{J}}^*)^{n-\bar{n}} \geq \bar{\lambda}_{\mathbf{J}}^{nT_0}$  with probability one is:  $\lambda_{\mathbf{J}}^* \geq \hat{\lambda}_{\mathbf{J}}^{nT_0}$ ,  $\forall n > 0$  with probability one.

Consider first the case with two threads ( $M = 2$ ), recall that

$$\lambda_{\mathbf{J}}^* = \begin{cases} \Lambda_{J_1}(g_1, T_0)\Lambda_{J_2}(g_2, T_0) & D_{KL}(g_1, g_2) > \eta \\ \Lambda_{J_1}(\tilde{g}_1(J_1), T_0)\Lambda_{J_2}(g_2, T_0) & D_{KL}(g_1, g_2) \leq \eta \end{cases}$$

and

$$\hat{\lambda}_{\mathbf{J}}^{nT_0} = \begin{cases} \Lambda_{J_1}(g_1, T_0)\Lambda_{J_2}(g_2, T_0) & D_{KL}(g_1, g_2) > \eta \\ \Lambda_{J_1}(\tilde{g}_1(J_1, K_{s,1}^{nT_0}), T_0)\Lambda_{J_2}(g_2, T_0) & D_{KL}(g_1, g_2) \leq \eta \end{cases}$$

where

$$\Lambda_J(g, T_0) = \int_{B(J) \cap \mathfrak{C}(T_0)} g(y; J) dy + \int_{\Omega \setminus \mathfrak{C}(T_0)} g(y; J) dy$$

and

$$\tilde{g}_1(J_1, K_{s,1}) = \arg \min_{g \in \mathfrak{G}} D_{K,L}(H(x, J_1, K_{s,1}), g).$$

From the definition of  $\tilde{g}_1(J_1, K_{s,1})$  and  $\tilde{g}_1(J_1)$ , it follows that

$$\begin{aligned} \int_{\Omega} \ln \frac{\tilde{g}_1(x, J_1, K_{s,1}^{nT_0})}{H(x, J_1, K_{s,1}^{nT_0})} H(x, J_1, K_{s,1}^{nT_0}) dx &\geq \int_{\Omega} \ln \frac{\tilde{g}_1(x, J_1)}{H(x, J_1, K_{s,1}^{nT_0})} H(x, J_1, K_{s,1}^{nT_0}) dx \\ \int_{\Omega} \ln \frac{\tilde{g}_1(x, J_1, K_{s,1}^{nT_0})}{H(x, J_1)} H(x, J_1) dx &\leq \int_{\Omega} \ln \frac{\tilde{g}_1(x, J_1)}{H(x, J_1)} H(x, J_1) dx \end{aligned}$$

which implies

$$\begin{aligned} \int_{[B(J_1) \cap \mathfrak{C}(T_0)] \cup \tilde{B}(K_{s,1}^{nT_0})} \tilde{g}_1(J_1, K_{s,1}^{kT_0}, x) dx &\leq \int_{[B(J_1) \cap \mathfrak{C}(T_0)] \cup \tilde{B}(K_{s,1}^{nT_0})} \tilde{g}_1(J_1, x) dx \\ \int_{B(J_1) \cap \mathfrak{C}(T_0)} \tilde{g}_1(J_1, K_{s,1}^{kT_0}, x) dx &\geq \int_{B(J_1) \cap \mathfrak{C}(T_0)} \tilde{g}_1(J_1, x) dx \end{aligned}$$

When  $K_{s,1}^{nT_0} = \emptyset$ ,  $\tilde{g}_1(J_1, K_{s,1}^{nT_0}) = \tilde{g}_1(J_1)$  and  $\lambda_{\mathbf{J}}^* = \hat{\lambda}_{\mathbf{J}}^{nT_0}$ . When  $K_{s,1}^{nT_0} \neq \emptyset$  then:

$$[B(J_1) \cap \mathfrak{C}(T_0)] \subsetneq \{[B(J_1) \cap \mathfrak{C}(T_0)] \cup \tilde{B}(K_{s,1}^{nT_0})\} \subseteq \{[B(J_1) \cap \mathfrak{C}(T_0)] \cup [\Omega \setminus \mathfrak{C}(T_0)]\}$$

and

$$\int_{[(B(J_1) \cap \mathfrak{C}(T_0))] \cup [\Omega \setminus \mathfrak{C}(T_0)]} \tilde{g}_1(J_1, K_{s,1}^{nT_0}, x) dx \leq \int_{[(B(J_1) \cap \mathfrak{C}(T_0))] \cup [\Omega \setminus \mathfrak{C}(T_0)]} \tilde{g}_1(J_1, x) dx$$

thus

$$\Lambda_{J_1}(\tilde{g}_1(J_1, K_{s,1}^{nT_0}), T_0) \leq \Lambda_{J_1}(\tilde{g}_1(J_1), T_0)$$

Hence,  $\lambda_{\mathbf{J}}^* \geq \hat{\lambda}_{\mathbf{J}}^{nT_0}$ .

We now prove the case in which  $M > 2$  by induction. Assume the result holds for any  $M - 1$  dimensional state of information, say  $\mathbf{J}'$ . Let us construct an  $M$  dimensional state  $\mathbf{J}$  as follows:

$$\mathbf{J} = \{J_1\} \times \mathbf{J}'$$

where  $\mathbf{J}' = (J_2, \dots, J_M)$  is a  $M - 1$  dimension state variable. The induction hypothesis implies  $\lambda_{\mathbf{J}'}^* \geq \hat{\lambda}_{\mathbf{J}'}^{nT_0}$ . Recall that the interaction between threads is hierarchical: any given thread only interacts with *higher-indexed* threads. Thus, adding a new thread with say index 1 (as in the construction of  $\mathbf{J}$  above) has no impact on threads 2 to  $M$ . The eigenvalues  $\lambda_{\mathbf{J}'}^*$  and  $\hat{\lambda}_{\mathbf{J}'}^{nT_0}$  are independent of thread 1. Thus,

$$\lambda_{\mathbf{J}}^* = \begin{cases} \Lambda_{J_1}(g_1, T_0) \cdot \lambda_{\mathbf{J}'}^* & D_{KL}(g_1, g_j) > \eta \quad \forall j \in \{2, \dots, M - 1\} \\ \Lambda_{J_1}(\tilde{g}_1(J_1), T_0) \cdot \lambda_{\mathbf{J}'}^* & \text{otherwise} \end{cases}$$

and

$$\hat{\lambda}_{\mathbf{J}}^{nT_0} = \begin{cases} \Lambda_{J_1}(g_1, T_0) \cdot \lambda_{\mathbf{J}'}^{nT_0} & D_{KL}(g_1, g_j) > \eta \quad \forall j \in \{2, \dots, M-1\} \\ \Lambda_{J_1}(\tilde{g}_1(J_1, K_{s,1}^{nT_0}), T_0) \cdot \lambda_{\mathbf{J}'}^{nT_0} & \text{otherwise.} \end{cases}$$

Using the same argument as above we have:

$$\Lambda_{J_1}(\tilde{g}_1(J_1, K_{s,1}^{nT_0}), T_0) \leq \Lambda_{J_1}(\tilde{g}_1(J_1), T_0)$$

which together with  $\lambda_{\mathbf{J}'}^* \geq \hat{\lambda}_{\mathbf{J}'}^{nT_0}$  implies  $\lambda_{\mathbf{J}}^* \geq \hat{\lambda}_{\mathbf{J}}^{nT_0}$  with probability one. ■

### 3.4 Numerical Experiments

In this section we report the results from a series of numerical tests aimed at illustrating the improved performance enabled by an interactive approach to model-based search. The models used in the experiment correspond to class of multivariate normal distributions with the fixed covariance matrix  $\Sigma = 0.03I$ . The local search method set to BFGS Quasi-Newton method.

#### 3.4.1 Ackley Problem

Ackley's Problem [37] is to find  $x \in \mathbb{R}^n$ , with  $x_i \in (-32.768, 32.768)$ , that minimizes the following function:

$$H(x) = -20 \cdot \exp(-0.2 \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n x_i^2}) - \exp\left(\frac{1}{n} \cdot \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + \exp(1).$$

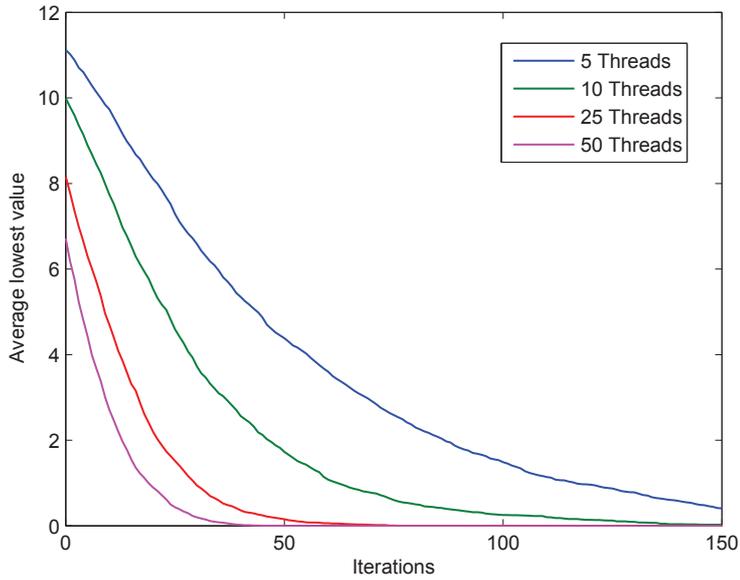


Fig. 3.1. Independent model search method

The performance of independent model search method is in figure 3.1.

To determine  $\tilde{g}$  of rejected thread, it is hard to calculate KL divergence precisely. We sample a finite set of trial random selected points (choosing 20 in this experiment). Calculate distance between the trailing points and the set of detected local optimas and take the one with the largest total distance as the mean of reference distribution. Let  $\epsilon = 0.1$ .

For the modified interactive model search method, we set the iteration with cap 350 function evaluation, the early finished threads wait other threads only within this cap. Same to interactive model based search, 20 trial points are sampled to determine new sample by choosing the one with largest total distance to set which contains starting points of incomplete search and detected local minimas. We report the average number of iterations of reaching global optima in table 3.1.

Table 3.1  
Average number of iterations before reaching global optima

| Number of Threads                 | 5     | 10    | 25    | 50    |
|-----------------------------------|-------|-------|-------|-------|
| Independent Model Search          | 68.75 | 44.73 | 26.34 | 19.11 |
| Interactive Model Search          | 66.77 | 43.95 | 25.83 | 18.91 |
| Modified Interactive Model Search | 69.28 | 42.81 | 26.90 | 19.37 |

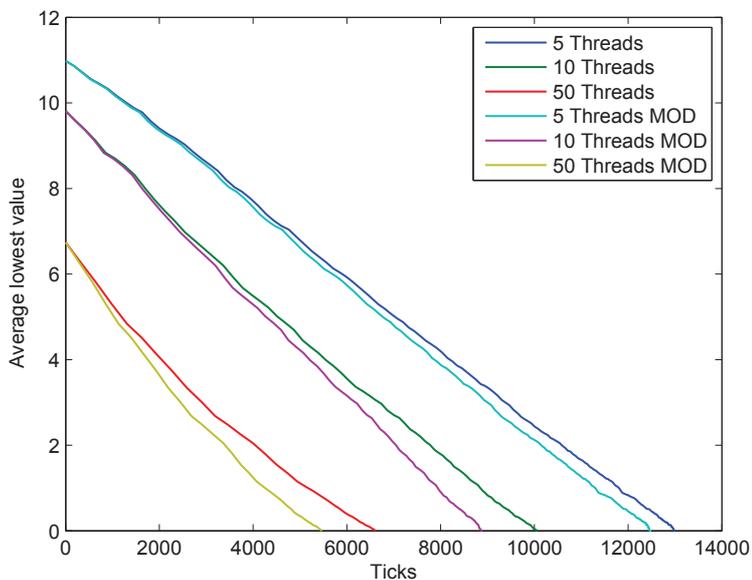


Fig. 3.2. Original and modified interactive model search method

Our modified interactive model search underperformed interactive model search in measure of iterations as a result of incomplete search iterations. To measure real local search time instead of numbers of iterations, the number of objective function evaluations is set as standardized time ticks. Performing 500 numerical experiments, we report the average number of maximum objective function evaluations cross threads before one of threads find the global optima in table 3.2. For finite time performance, we report the average best objective value detected over ticks eclipse in figure 3.2. We also report the probability of find global optima under ticks limitation 5000 in table 3.3.

The modified interactive model search need less ticks to reach to global optimas and achieve lower objective function value than independent and interactive model search. The acceleration becomes more significant as number of threads increases.

Table 3.2  
Average number of ticks before reaching global optima

| Number of Threads                 | 5     | 10    | 25   | 50   |
|-----------------------------------|-------|-------|------|------|
| Independent Model Search          | 13732 | 10698 | 8223 | 7194 |
| Interactive Model Search          | 13388 | 10491 | 8046 | 7123 |
| Modified Interactive Model Search | 12827 | 9236  | 7032 | 5820 |

Table 3.3  
Probability of finding global optima within 5000 ticks

| Number of Threads                 | 5     | 10    | 25    | 50    |
|-----------------------------------|-------|-------|-------|-------|
| Independent Model Search          | 7.6%  | 10.8% | 15.8% | 24.0% |
| Interactive Model Search          | 7.0%  | 12.6% | 21.4% | 26.2% |
| Modified Interactive Model Search | 10.2% | 18.8% | 28.8% | 43.6% |

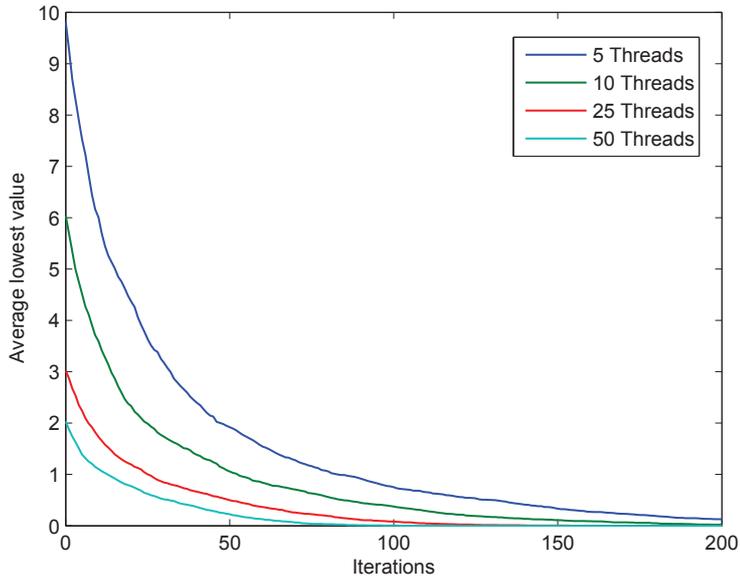


Fig. 3.3. Independent model search method

### 3.4.2 Rastrigin Problem

The Rastrigin problem is the minimization of the function defined as follows:

$$F(x) := 10n + \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i)] \quad x_i \in [-5.12, 5.12], i = 1, 2, \dots, n.$$

we have run numerical experiment on 3-D Rastrigin function with covariance matrix  $\Sigma = 0.03I$ .

The performance of independent model search method is in figure 3.3 We take same 20 trial points to generate reference distribution mean of rejected threads. We also use  $\epsilon = 0.1$  in this experiment.

Table 3.4  
Average number of iterations before reaching global optima

| Number of Threads                 | 5       | 10     | 25     | 50     |
|-----------------------------------|---------|--------|--------|--------|
| Independent Model Search          | 121.404 | 86.224 | 52.676 | 33.828 |
| Interactive Model Search          | 122.842 | 80.316 | 40.492 | 20.636 |
| Modified Interactive Model Search | 127.842 | 78.628 | 42.046 | 22.266 |

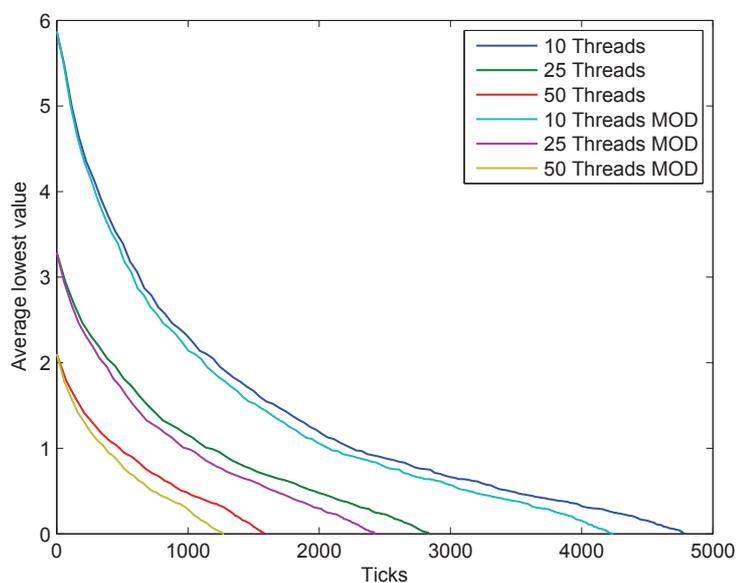


Fig. 3.4. Original and modified interactive model search method

For modified interactive model search method, we set the iteration cap as 60 function evaluation. Performing 500 numerical experiments, we report the average number of iterations of reaching global optimas by table 3.4.

Using the number of objective function evaluations as time measurement ticks, the performance is present in table 3.5, we report average number of maximum ticks over different threads before one of threads find global optima. We also report the

Table 3.5  
Average number of ticks before reaching global optima

| Number of Threads                 | 5      | 10     | 25     | 50     |
|-----------------------------------|--------|--------|--------|--------|
| Independent Model Search          | 5784.5 | 4664.6 | 3363.4 | 2406.6 |
| Interactive Model Search          | 6354.9 | 4866.6 | 2925.5 | 1671.4 |
| Modified Interactive Model Search | 6253.6 | 4304.4 | 2486.4 | 1335.2 |

Table 3.6  
Probability of finding global optima within 2000 ticks

| Number of Threads                 | 5     | 10    | 25    | 50    |
|-----------------------------------|-------|-------|-------|-------|
| Independent Model Search          | 11.0% | 15.2% | 28.8% | 44.6% |
| Interactive Model Search          | 10.6% | 16.2% | 37.6% | 62.0% |
| Modified Interactive Model Search | 9.4%  | 19.2% | 38.6% | 70.2% |

finite time performance in figure 3.4 and table 3.6 for best objective value found and probability of finding global optima within ticks limit.

### 3.5 Chapter Summary

In this chapter we consider a parallel computing scheme for global optimization that combines multi-start local search with the dynamic reallocation of computational resources (e.g. processing time). Our work builds upon the interactive model-based approach to global optimization proposed in [21] in which parallel threads independently execute a model-based search method (see [42]) and periodically interact through a simple acceptance-rejection rule aimed at preventing duplication of search efforts.

While sophisticated algorithmic variations can be designed in order to optimally react to search outcomes our focus is to leverage relatively simple ideas such as *i)* continuing searches that are promising because the end-points have lower objective values than all other solutions found so far and *ii)* avoiding duplication of failed searches and/or search effort across threads. The relatively small gains afforded by these simple ideas are shown to be magnified by parallelization: the rate of convergence for an interactive model-based search increases with the number of threads for a wide-class of local search techniques (i.e. model-based) when compared to independent parallel implementation.

## 4. HIGH FREQUENCY TRADERS IN INEFFICIENT MARKET

### 4.1 Introduction and Literature Review

High frequency trading is one of the most significant market structure developments in recent years. The computer hardware and software development shifts trading responsibilities for human agent to computer algorithms. Debate on the high frequency trading arbitraging is growing. In general, arbitraging is viewed positively as arbitrageurs actively search for mispricings, executing trades that eliminate mispricings thus increasing market efficiency. However, such positive consensus is yet to emerge on arbitraging enabled by computer algorithms at a very fast timescale. The increasing evidence shows that high frequency trading spreads widely in financial market including equity exchange, foreign exchange and future market (see [50], [51] and [52]). It has been argued that high-frequency trading to improve or at least does not hurt the overall quality of markets because high frequency trading increases liquidity and reduces discrepancies in prices across related markets [51,53]. Others argue that high frequency trading not only increases market volatility and the probability of mispricing but also makes abnormal profit from liquidity traders which discourage them from the market [54]. High frequency trading was names as one of the likely causes of the "flash crash" in 2010 but empirical evidence suggest this is not the case (see [55] Kirilenko et al. (2014)).

On low-latency communications and decision making, high frequency traders convey orders for electronic exchanges over intervals measured in micro- and milliseconds.

High frequency traders can front-run other traders by examining trading volume and quotes to detect the time when traders are using algorithms to split up large orders that will move the market [56]. This strategy, by which the high frequency trader can predict future flows into or out of a particular asset is called order anticipation in the order world, if high frequency traders are sufficiently fast, they might observe the first orders to arrive in an exchange and react by trading on other exchanges before the remaining orders from the original trader arrive there [57].

We study the effects of high frequency trading when there is also insider trading activity (trading activity involve possession of nonpublic information material) in the market. In earlier work, [22] Kyle (1985) had shown how insider trading induces systematic losses to uninformed liquidity (or noise) traders. The multiple informed traders model introduced in [58] extends Kyle's model to involve multiple imperfect information insider traders who competing and learning during the trading process. More recently, [59] and [60] introduce asymmetric information filtration for insider trader and market maker in which the liquidity order low is modeled as stochastic process, and the market maker assumed to know in which distribution liquidity low is drawn from while informed trader processes the information about underlying value of the asset but no accuracy information about liquidity. The informed trader update liquidity information during the trading process, and the profitability of the informed trader depends on this information updating.

In this chapter, we modify Kyle's discrete dynamic insider trading model by including a high frequency trader. The high frequency trader observes the combined order low from insider's and noise traders' demand ahead of the market marker. The presence of the high frequency trading makes liquidity information no longer available to the informed trader and market maker. The market maker and the informed trader have to learn the liquidity activity by observing the aggregated order low from noise

and high frequency traders. We show that high frequency traders make abnormal profit if the market maker and informed trader expectations about liquidity are incorrect. The profit associated with this incorrect expectation vanishes over time with accuracy of belief of liquidity improving. The expected profit of the informed trader decreases as a result of presence of a high frequency trader. High frequency trading quantity within a certain range makes the reduction of insider's profit exceeding the profit of the high frequency trader, thus the expected loss of noise traders is reduced. The high frequency trader plays beneficial role when market maker underestimates the liquidity variance. Under this situation, the high frequency trader makes market and provide extra liquidity. In section 2, we build a single stage trading model and provide conditions associated the beneficial role of the high frequency trader. In section 3, we introduce a discrete dynamic informed trading model with the high frequency trader. In our main results we show the convergence of belief about liquidity, the positive expected profit of the high frequency trader and the reduction of expected loss of noise traders. In section 4, we have an illustrative numerical experiment to demonstrate the theoretical findings.

## 4.2 One Stage Kyle Model with High Frequency Trader

In this section we introduce a modification of [22] Kyle's (1985) single auction model by including high frequency trader.

Assume there is an infinitely supplied and dividable asset available for trading for a single trading period. The fundamental value of is asset is given by

$$\tilde{v} \sim \mathcal{N}(0, \sigma_v^2)$$

At the end of trading period, any participant holding a share of the asset receives a liquidating dividend (liability) of  $\tilde{v}$  dollars.

There are 3 kinds of agents: one informed trader, one high frequency trader and noise traders. Let  $x$  denote order from the informed trader,  $u \sim \mathcal{N}(0, \sigma_u^2)$  denote the aggregate order from noise traders, and  $f$  denotes order from high frequency traders.

The trading takes place by following sequence:

1. The informed trader submit order  $x$  after observing fundamental value of asset  $\tilde{v}$ . Noise traders submit cumulated order  $u$ .
2. The HF trader observe order flow  $x + u$  and place  $f = \eta(x + u)$ , where  $|\eta| < 1$ .
3. The market maker receives total order flow  $y = x + u + f$  and clears the market at price  $p$ .

Define the profit of insider trader as  $\pi = (\tilde{v} - p)x$ .

**Assumption 1** *Both insider trader and market maker have same prior belief of aggregate liquidity orders from noise traders and high frequency trader as  $u + f \sim \mathcal{N}(0, \sigma_f^2)$ .*

Let  $X$  be the trading strategy of informed trader and  $P$  be the pricing rule of market maker. An equilibrium is defined as a pair  $(X, P)$  with two conditions:

1. Profit Maximization:

$$E\{\pi(X, P)\} \geq E\{\pi(X', P)\}$$

for any alternative trading strategy  $X'$ .

2. Market Efficiency:

$$p(X, P) = E[\tilde{v}|x + u + f]$$

**Lemma 3** *Under Assumption 1, there exists a unique equilibrium in which  $X$  and  $P$  are linear function. Define constants  $\beta = \frac{\sigma_f}{\sigma_v}$  and  $\lambda = \frac{\sigma_v}{2\sigma_f}$ . The equilibrium  $X$  and  $P$  are given by*

$$X(\tilde{v}) = \beta\tilde{v}, \quad P(x + u + f) = \lambda(x + u + f)$$

**Proof** See [22] Kyle (1985) Theorem 1. Replace  $u$  and  $\sigma_u$  by  $u + f$  and  $\sigma_f$ . ■

**Lemma 4** *The expected profit for informed trader is*

$$E\pi^i(\tilde{v}) = (1 - \eta)\frac{\sigma_f\sigma_v}{2}$$

*The high frequency trader receive no expected surplus if the expectations on  $\sigma_f^2$  are correct. If expectation on  $\sigma_f^2$  are incorrect, the expected profit of high frequency trader is as follows:*

- If  $\sigma_f^2 > \frac{1+\eta}{1-\eta}\sigma_u$ , with  $\eta > 0$ ,

$$E[\pi^f] = \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2] > 0$$

- If  $\sigma_f^2 < \frac{1+\eta}{1-\eta}\sigma_u$ , with  $\eta < 0$ ,

$$E[\pi^f] = \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2] > 0$$

**Proof** Based on equilibrium trading strategy, the expected profit for informed trader is

$$\begin{aligned} E[\pi^i(\tilde{v})] &= E[(\tilde{v} - p)x] = E[(\tilde{v} - \lambda[x + u + f])x] \\ &= E\left[\left(v - (1 + \eta)\frac{\sigma_v}{2\sigma_f}\left(u + \frac{\sigma_f}{\sigma_v}\tilde{v}\right)\right)\frac{\sigma_f}{\sigma_v}\tilde{v}\right] \\ &= (1 - \eta)\frac{\sigma_f\sigma_v}{2} \end{aligned}$$

The expected profit for high frequency trader is

$$\begin{aligned} E[\pi^f(\tilde{v})] &= E[(\tilde{v} - p)f] = E[(\tilde{v} - \lambda[x + u + f])f] \\ &= E\left[\left(v - (1 + \eta)\frac{\sigma_v}{2\sigma_f}\left(u + \frac{\sigma_f}{\sigma_v}\tilde{v}\right)\right)\eta\left(\frac{\sigma_f}{\sigma_v}\tilde{v} + u\right)\right] \\ &= \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2] \end{aligned}$$

If the expectations are fulfilled,

$$\sigma_f^2 = E[(u + f)^2] = E[(u + \eta[u + x])^2] = (1 + \eta)\sigma_u^2 + \eta^2\sigma_f^2$$

Hence,

$$\sigma_f^2 = \frac{(1 + \eta)^2}{(1 - \eta^2)}\sigma_u^2 = \frac{1 + \eta}{1 - \eta}\sigma_u^2$$

and  $\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2 = 0$ , thus

$$E[\pi^f(\tilde{v})] = \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2] = 0$$

If  $\sigma_f^2 > \frac{1+\eta}{1-\eta}\sigma_u^2$  and  $\eta > 0$ ,  $\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2 > 0$ ,

$$E[\pi^f] = \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1 - \eta) - (1 + \eta)\sigma_u^2] > 0$$

If  $\sigma_f^2 < \frac{1+\eta}{1-\eta}\sigma_u^2$  and  $\eta < 0$ ,  $\sigma_f^2(1-\eta) - (1+\eta)\sigma_u^2 < 0$ ,

$$E[\pi^f] = \eta \frac{\sigma_v}{2\sigma_f} [\sigma_f^2(1-\eta) - (1+\eta)\sigma_u^2] > 0$$

This two conditions follow the fact that the market maker over or under prices asset based on he or she overestimating or underestimating liquidity variance, the high frequency trader can make expected positive rent by observing this incorrect estimation and thus mispricing. ■

In the market without high frequency trader, we have  $\beta = \frac{\sigma_u}{\sigma_v}$  and  $\lambda = \frac{\sigma_v}{2\sigma_u}$ . The expected profit of informed trader is

$$E[\bar{\pi}^i] = E[(\tilde{v} - \lambda(x+u))x] = E[\beta\tilde{v}^2 - \lambda\beta^2\tilde{v}^2] = \frac{\sigma_u\sigma_v}{2}$$

In the zero-sum game environment, the total expected loss of noise trader in the market without high frequency trader is

$$E[\bar{L}] = E[\bar{\pi}^i] = \frac{\sigma_u\sigma_v}{2}$$

In the market with high frequency trader, the total expected loss of noise traders is

$$E[L] = E[\pi^i] + E[\pi^f] = (1-\eta)\frac{\sigma_f\sigma_v}{2} + \eta\frac{\sigma_v}{2\sigma_f}[\sigma_f^2(1-\eta) - (1+\eta)\sigma_u^2]$$

**Theorem 4.2.1** *The high frequency trader protect noise traders if expectation on  $\sigma_f^2$  are incorrect and one of following condition holds: (1)  $\eta > 0$  and  $\frac{1+\eta}{1-\eta}\sigma_u^2 < \sigma_f^2 < \sigma_{+\eta}^2$ ; (2)  $\eta < 0$  and  $\sigma_{-\eta}^2 < \sigma_f^2 < \frac{1+\eta}{1-\eta}\sigma_u^2$ , where*

$$\sigma_{+\eta} = \frac{\sigma_u(1 + \sqrt{1 + 4\eta(1-\eta^2)(1+\eta)})}{2(1-\eta^2)}, \quad \sigma_{-\eta} = \frac{\sigma_u(1 - \sqrt{1 + 4\eta(1-\eta^2)(1+\eta)})}{2(1-\eta^2)}.$$

Furthermore, when condition (2) holds, high frequency trader plays beneficial role by providing extra liquidity.

**Proof** The high frequency trader protect noise traders if

$$E[\bar{L}] - E[L] > 0,$$

the expected loss of noise traders is reduced due to presence of high frequency trader.

The expected loss reduction is

$$\begin{aligned} E[\bar{L}] - E[L] &= \frac{\sigma_u \sigma_v}{2} - (1 - \eta) \frac{\sigma_f \sigma_v}{2} - \eta \frac{\sigma_v}{2\sigma_f} [\sigma_f^2 (1 - \eta) - (1 + \eta) \sigma_u^2] \\ &= \frac{\sigma_u \sigma_v}{2} - \frac{\sigma_v}{2\sigma_f} [\sigma_f^2 (1 - \eta^2) - \eta(1 + \eta) \sigma_u^2] \\ &= \frac{\sigma_v}{2\sigma_f} [-(1 - \eta^2) \sigma_f^2 + \sigma_u \sigma_f + \eta(1 + \eta) \sigma_u^2] \end{aligned}$$

Condition  $E[\bar{L}] - E[L] > 0$  is equivalent to

$$(1 - \eta^2) \sigma_f^2 - \sigma_u \sigma_f - \eta(1 + \eta) \sigma_u^2 < 0$$

For fulfilled expectation  $\sigma_f$ ,

$$(1 - \eta^2) \sigma_f^2 - \sigma_u \sigma_f - \eta(1 + \eta) \sigma_u^2 = \frac{\sigma_u \sigma_v}{2} [1 - \sqrt{1 - \eta^2}] > 0$$

Consider incorrect expectation, the discriminant of equation  $(1 - \eta^2)x^2 - \sigma_u x - \eta(1 + \eta)\sigma_u^2 = 0$  is  $\sigma_u^2 + 4\eta(1 + \eta)(1 - \eta^2)\sigma_u^2$ . If  $\eta \in (0, 1)$ , we have  $\sigma_u^2 + 4\eta(1 + \eta)(1 - \eta^2)\sigma_u^2 > 0$ . If  $\eta \in (-1, 0)$ , we have  $(1 + \eta)\eta > -\frac{1}{4}$ , thus  $\sigma_u^2 + 4\eta(1 + \eta)(1 - \eta^2)\sigma_u^2 > 0$ . So the equation

$$(1 - \eta^2)x^2 - \sigma_u x - \eta(1 + \eta)\sigma_u^2 = 0$$

has two real solutions for all  $|\eta| < 1$ , which are

$$\sigma_{+\eta} = \frac{\sigma_u(1 + \sqrt{1 + 4\eta(1 - \eta^2)(1 + \eta)})}{2(1 - \eta^2)} \quad \text{and} \quad \sigma_{-\eta} = \frac{\sigma_u(1 - \sqrt{1 + 4\eta(1 - \eta^2)(1 + \eta)})}{2(1 - \eta^2)}$$

When  $\eta > 0$  and  $\frac{1+\eta}{1-\eta}\sigma_u^2 < \sigma_f^2 < \sigma_{+\eta}^2$ ; or  $\eta < 0$  and  $\sigma_{-\eta}^2 < \sigma_f^2 < \frac{1+\eta}{1-\eta}\sigma_u^2$ , we have

$$(1 - \eta^2)\sigma_f^2 - \sigma_u\sigma_f - \eta(1 + \eta)\sigma_u^2 > 0$$

and thus

$$E[\bar{L}] - E[L] > 0.$$

When  $\eta < 0$ , the high frequency trader takes order placed by slow agents, who makes the market and provides extra liquidity, and thus plays beneficial role. ■

### 4.3 Dynamic Kyle Model with High Frequency Trader

In this section, we consider the dynamic trading model with informed and high frequency trader. Assume both market maker and informed trader do not have full information about the liquidity orders and learn it during the trading process. We show that when the initial expectation about the liquidity order in a certain range, the high frequency trader plays beneficial role who reduce the total expected loss of noise trader.

Consider a single asset traded over time  $\{1, \dots, T\}$ . Similar to [59] H. Hong (2002), assume the value of the asset is gradually resolved over time as

$$v = \sum_{t=1}^T v_t$$

where  $v_t$  becomes public information after all trading activity end of time  $t$ . Let  $v_t \sim \mathcal{N}(0, \sigma_v^2)$  is a common knowledge. By the end of period  $T$ , the value of asset will be known by all participants.

There are 3 kinds of agents: one informed trader, one HF trader and noise traders. Informed trader observes  $v_t$  before  $v_t$  becomes public information. Let  $x_t$  denote order flow from the informed trader,  $u_t \sim \mathcal{N}(0, \sigma_u^2)$  denote the aggregate order flow from noise traders, and  $f_t$  denotes order flow from HF traders.

Assume the market maker and insider trader do not have full information about liquidity  $u_t + f_t$ , but have same prior guess  $u_t + f_t \sim \mathcal{N}(0, \sigma_f^2)$ , where  $\sigma_f^2$  is drawn from Inverse Gamma distribution as

$$\sigma_f^2 \sim \text{Inv} - \text{Gamma}(a_0, b_0)$$

To simplify analysis, assume  $a_0 > 2$ , the prior of  $\sigma_f^2$  has mean

$$\frac{b_0}{a_0 - 1}$$

and variance

$$\frac{b_0^2}{(a_0 - 1)^2(a_0 - 2)}$$

Let both informed trader and market maker update their belief  $(a_t, b_t)$  of  $\sigma_f^2$  based on Baye's rule.

In period  $t$ , trading takes place by following sequence:

1. The informed trader receives private information  $v_t$ . Informed trader and noise traders submit order  $x_t$  and  $u_t$  simultaneously.
2. The high frequency trader observe order flow  $x_t + u_t$  and place  $f_t = \eta(x_t + u_t)$

3. The market maker receives total order flow  $y_t = x_t + u_t + f_t$  and clear market at price  $p_t$ .
4. Information  $v_t$  and quantity  $p_t, y_t$  are revealed to public. Market maker and informed trader update belief  $(a_t, b_t)$  of liquidity order variance  $\sigma_f^2$ .

In reality, the market maker knows who are high frequency traders and can distinguish their orders. However, algorithmic trading is widely used by investment bank, pension funds, mutual funds and other institutional traders. Institutional traders can participate market at both high and low frequency. The market maker is hard to distinguish high frequency orders from slow orders, thus we assume market maker can only observe total flow  $y_t$ .

Let  $\Sigma_t$  be the belief of  $\sigma_f^2$  at time  $t$  as

$$\Sigma_t = \frac{b_t}{a_t - 1}$$

#### 4.3.1 Equilibrium Strategy and Information Update

An equilibrium requires two conditions for all time  $t \leq T$ :

1. Market Efficient: market maker sets price as

$$p_t = E[v | \Sigma_{t-1}, V_{t-1}, Y_t]$$

where  $V_{t-1} = (v_1, \dots, v_{t-1})$ ,  $Y_t = (y_1, \dots, y_t)$ .

2. Profit Maximization: the insider trader maximize the expected profit as

$$E[\pi_t] = E[x_t(v - p_t) | \Sigma_{t-1}, V_t, Y_{t-1}]$$

**Lemma 5** *At time  $t$ , if informed trader and market maker have same belief  $\Sigma_{t-1}$ .*

*Define  $\beta_t = \frac{1}{2\lambda_t}$  and  $\lambda_t = \frac{\beta_t^2 \sigma_v}{\beta_t^2 \sigma_v^2 + \Sigma_{t-1}}$ , the equilibrium trading strategy is*

$$p_t = \sum_{j=1}^{t-1} v_j + \lambda_t y_t, \quad x_t = \beta v_t$$

**Proof** At time  $t$ , given belief  $\Sigma_{t-1}$  the market maker can calculate the conditional expectation of  $v_t$  as

$$E[v_t | \Sigma_{t-1}, y_t] = \frac{\text{Cov}[v_t, \beta v_t + u_t + f_t | \Sigma_{t-1}]}{\text{Var}[\beta v_t + u_t + f_t | \Sigma_{t-1}]} y_t = \frac{\beta_t \sigma_v^2}{\beta_t^2 \sigma_v^2 + \Sigma_{t-1}} y_t$$

Market maker set the price rule as

$$p_t = E[v | \Sigma_{t-1}, y_t] = \sum_{j=1}^{t-1} v_j + E[v_t | \Sigma, y_t] = \sum_{j=1}^{t-1} v_j + \lambda_t y_t$$

where

$$\lambda_t = \frac{\beta_t \sigma_v^2}{\beta_t^2 \sigma_v^2 + \Sigma_{t-1}}$$

The profit maximization objective function of informed trader is

$$\begin{aligned} E[\pi_t] &= E[x_t(v - p_t)] = E[x_t(\sum_{j=1}^T v_j - \sum_{j=1}^{t-1} v_j - \lambda_t y_t)] \\ &= E[x_t(v_t + \sum_{j=t+1}^T v_j - \lambda_t[x_t + u_t + f_t])] \end{aligned}$$

From  $E[u_t + f_t] = 0$  and  $E[\sum_{j=t+1}^T v_j] = 0$ ,

$$E[\pi_t] = E[x_t(v_t - \lambda_t x_t)]$$

The FOC is

$$v_t - 2\lambda_t x_t = 0$$

thus

$$x_t = \beta_t v_t = \frac{1}{2\lambda_t} v_t$$

■

In trading period  $t$ , once  $v_t$  revealed, the market maker can deduce  $x_t$  by equilibrium strategy  $x_t = \beta_t v_t$ . Together with total order flow  $y_t$ , the total liquidity order  $u_t + f_t = y_t - \beta_t v_t$  is available to the market maker. The informed trader is able to access  $u_t + f_t$  at the end of trading period  $t$  as well.

The prior belief of  $\sigma_f^2$  at time  $t$  is

$$\sigma_f^2 \sim \text{Inv - Gamma}(a_{t-1}, b_{t-1})$$

The likelihood function for a single observation  $u_t + f_t$  is

$$f(u_t + f_t | \sigma_f^2) = \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{(u_t + f_t)^2}{2\sigma_f^2}\right)$$

The posterior distribution is

$$f(\sigma_f^2 | u_t + f_t) = \frac{f(\sigma_f^2) f(u_t + f_t | \sigma_f^2)}{\int f(\sigma_f^2) f(u_t + f_t | \sigma_f^2) d\sigma_f^2}$$

where  $f(\sigma_f^2)$  is the PDF of distribution  $\text{Inv - Gamma}(a_{t-1}, b_{t-1})$ .

From conjugate prior property of normal distribution with known mean, distribution associated with  $f(\sigma_f^2 | u_t + f_t)$  remains Inverse-Gamma with

$$a_t = a_{t-1} + \frac{1}{2}, \quad b_t = b_{t-1} + \frac{(u_t + f_t)^2}{2}$$

thus

$$a_t = a_0 + \frac{t}{2}, \quad b_t = b_0 + \sum_{j=1}^t \frac{(u_j + f_j)^2}{2}$$

After trading period  $t$ , both market maker and informed trader believe  $\sigma_f^2$  follows inverse gamma distribution with mean

$$\Sigma_t = \frac{b_t}{a_t - 1} = \frac{b_0 + \sum_{j=1}^t \frac{(u_j + f_j)^2}{2}}{a_0 + \frac{t}{2} - 1}$$

and variance

$$Var(\sigma_f^2) = \frac{\left(b_0 + \sum_{j=1}^t \frac{(u_j + f_j)^2}{2}\right)^2}{(a_0 + \frac{t}{2} - 1)(a_0 + \frac{t}{2} - 2)}$$

### 4.3.2 Trading Process Convergence

The high frequency trading quantity is

$$f_t = \eta(x_t + u_t).$$

The total liquidity order volume is

$$f_t + u_t = \eta\beta_t v_t + (1 + \eta)u_t$$

from  $u_t$  and  $v_t$  are independent,

$$f_t + u_t \sim \mathcal{N}(0, \eta^2 \beta_t^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2)$$

The variance of  $f_t + u_t$  is time variant which depends on  $\beta_t$ . Assume both market maker and insider trader have not prior information about the correlation between  $\beta_t$

and variance of  $f_t + u_t$ . The time variant fact is not available for market maker and informed trader who are fully confident that  $\sigma_f^2$  is constant over time.

At time  $t$ , given  $a_{t-1}$  and  $b_{t-1}$ ,

$$E[\Sigma_t | a_{t-1}, b_{t-1}] = E \left[ \frac{b_{t-1} + \frac{(u_t + f_t)^2}{2}}{a_{t-1} + \frac{1}{2} - 1} \middle| a_{t-1}, b_{t-1} \right] = \frac{2b_{t-1} + \eta^2 \beta_t^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2}{2a_{t-1} - 1}$$

thus

$$E[a_t] = a_{t-1} + \frac{1}{2} = a_0 + \frac{t}{2}$$

$$E[b_t] = E[b_{t-1}] + \frac{1}{2} [\eta^2 \beta_t^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2] = b_0 + \frac{1}{2} \left[ \sum_{i=1}^t \beta_i^2 \eta^2 \sigma_v^2 + t(1 + \eta)^2 \sigma_u^2 \right]$$

From equilibrium strategy, we have

$$\beta_t = \frac{1}{2\lambda_t} = \frac{\beta_t^2 \sigma_v^2 + \Sigma_{t-1}}{2\beta_t \sigma_v^2}$$

which implies

$$\beta_t^2 = \frac{\Sigma_{t-1}}{\sigma_v^2}$$

Thus,

$$E[b_t] = E[b_{t-1}] + \frac{1}{2} [\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2] = \left( 1 + \frac{\eta^2}{2a_0 + t - 3} \right) E(b_{t-1}) + \frac{(1 + \eta)^2 \sigma_u^2}{2}$$

We have  $\lim_{t \rightarrow \infty} a_t = \infty$  and  $\lim_{t \rightarrow \infty} E(b_t) = \infty$ . To determine whether  $E(\Sigma_t)$  is convergent, consider

$$\frac{|E(\Sigma_t)|}{|E(\Sigma_{t-1})|} = E \left( \frac{b_t}{b_{t-1}} \right) \frac{a_{t-1} - 1}{a_t - 1} = \frac{2a_0 + t - 3}{2a_0 + t - 2} \left[ \left( 1 + \frac{\eta^2}{2a_0 + t - 3} \right) + \frac{(1 + \eta)^2 \sigma_u^2}{2E(b_{t-1})} \right]$$

the ratio criteria is

$$r = \lim_{t \rightarrow \infty} \frac{|E(\Sigma_t)|}{|E(\Sigma_{t-1})|} = 1$$

The convergence is inclusive, it is possible that  $\lim_{t \rightarrow \infty} E(\Sigma_t) = \infty$ . However, if  $E(\Sigma_t) \rightarrow \infty$ , we have  $\lambda_t \rightarrow 0$ , trading quantity has no impact on price. We introduce an assumption to avoid this situation.

**Assumption 2** *Assume the market maker has a lower bound  $\underline{\lambda}$  for price parameter  $\lambda_t$ .  $\lambda_t \geq \underline{\lambda}$  for all  $t > 0$  is a public knowledge.*

**Theorem 4.3.1** *Under assumption 2, the expected liquidation variance belief  $E(\Sigma_t)$  converges. If  $\underline{\lambda} \leq \frac{\sqrt{1+\eta}\sigma_v}{2\sqrt{1-\eta}\sigma_u}$ , define  $\Sigma^*$  as the limit, we have*

$$\lim_{t \rightarrow \infty} E(\Sigma_t) = \Sigma^* = \frac{1+\eta}{1-\eta} \sigma_u^2$$

**Proof** Under Assumption 2, the informed trader can reach a upper bound of trading parameter as

$$\beta_t \leq \bar{\beta} = \frac{1}{2\underline{\lambda}}$$

and the expectation has

$$E(\Sigma_t) = \frac{2b_0 + \sum_{i=1}^t \beta_i^2 \eta^2 \sigma_v^2 + t(1+\eta)^2 \sigma_u^2}{2a_0 - 2 + t} \leq \frac{2b_0 + t\bar{\beta}^2 \eta^2 \sigma_v^2 + t(1+\eta)^2 \sigma_u^2}{2a_0 - 2 + t}$$

we have

$$\lim_{t \rightarrow \infty} E(\Sigma_t) \leq \bar{\beta}^2 \eta^2 \sigma_v^2 + (1+\eta)^2 \sigma_u^2 < \infty$$

Consider

$$\begin{aligned}
|E(\Sigma_t) - E(\Sigma_{t-1})| &= \left| \frac{2E(b_{t-1}) + \eta^2 \beta_t^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2}{2a_0 + t - 2} - \frac{2E(b_{t-1})}{2a_0 + t - 3} \right| \\
&\leq \left| \frac{2E(b_{t-1})}{(2a_0 + t - 2)(2a_0 + t - 3)} \right| + \left| \frac{\eta^2 \beta_t^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2}{2a_0 + t - 2} \right| \\
&\leq \left| \frac{2E(b_{t-1})}{(2a_0 + t - 2)(2a_0 + t - 3)} \right| + \left| \frac{\eta^2 \bar{\beta}^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2}{2a_0 + t - 2} \right|
\end{aligned}$$

we have

$$\lim_{t \rightarrow \infty} |E(\Sigma_t) - E(\Sigma_{t-1})| \leq \lim_{t \rightarrow \infty} \left| \frac{2E(b_{t-1})}{(2a_0 + t - 2)(2a_0 + t - 3)} \right| + \lim_{t \rightarrow \infty} \left| \frac{\eta^2 \bar{\beta}^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2}{2a_0 + t - 2} \right| = 0$$

then  $\lim_{t \rightarrow \infty} E(\Sigma_t) < \infty$  and  $\lim_{t \rightarrow \infty} |E(\Sigma_t) - E(\Sigma_{t-1})| = 0$ , thus  $\Sigma_t$  converges. Let  $\Sigma^* = \lim_{t \rightarrow \infty} \Sigma_t$ ,  $\beta^* = \lim_{t \rightarrow \infty} \beta_t$ . If  $\beta^* < \bar{\beta}$ , we have

$$\begin{cases} \Sigma^* = \beta^{*2} \eta^2 \sigma_v^2 + (1 + \eta)^2 \sigma_u^2 \\ \beta^{*2} = \frac{\Sigma^*}{\sigma_v^2} \end{cases}$$

and

$$\begin{cases} \beta^{*2} = \frac{(1 + \eta) \sigma_u^2}{(1 - \eta) \sigma_v^2} \\ \Sigma^* = \frac{1 + \eta}{1 - \eta} \sigma_u^2 \end{cases}$$

We can conclude that

$$\lim_{t \rightarrow \infty} \Sigma_t = \begin{cases} \frac{1 + \eta}{1 - \eta} \sigma_u^2, & \text{if } \lambda \leq \frac{\sqrt{1 + \eta} \sigma_v}{2\sqrt{1 - \eta} \sigma_u} \\ \frac{\sigma_v^2}{4\lambda^2}, & \text{otherwise} \end{cases}$$

■

### 4.3.3 Profitability Analysis

In this section, we assume  $\underline{\lambda}$  is small enough to satisfy  $\underline{\lambda} \leq \frac{\sqrt{1+\eta}\sigma_v}{2\sqrt{1-\eta}\sigma_u}$ .

Given equilibrium strategy  $\beta_t = \frac{1}{2\lambda_t}$  and  $\beta_t = \frac{\sqrt{\Sigma_{t-1}}}{\sigma_v^2}$  at time  $t$ , the expected profit for insider trader is

$$\begin{aligned}
E[\pi_t^i] &= E[(v - p_t)x_t] = E\left[\left(\sum_{i=1}^T v_i - \left(\sum_{i=1}^{t-1} v_i + \lambda_t y_t\right)\right)x_t\right] \\
&= E\left[\left(v_t + \sum_{i=t+1}^T v_i - \lambda_t(1 + \eta)(x_t + u_t)\right)x_t\right] \\
&= E[\beta_t v_t^2 - \lambda_t(1 + \eta)\beta_t^2 v_t^2] \\
&= \frac{\sigma_v^2}{2\lambda_t} - \frac{1 + \eta}{4\lambda_t}\sigma_v^2 = \frac{1 - \eta}{4\lambda_t}\sigma_v^2 = \frac{(1 - \eta)\sigma_v\sqrt{\Sigma_{t-1}}}{2}
\end{aligned}$$

**Theorem 4.3.2** *If  $\eta(\Sigma_0 - \Sigma^*) > 0$ , high frequency trader get positive rent for all trading period  $t > 0$ , the rent is vanished over time as market maker and informed trader learning liquidity information.*

**Proof** The profit for high frequency trader is

$$\begin{aligned}
E[\pi_t^f] &= E[(v - p_t)f_t] = E\left[\left(\sum_{i=1}^T v_i - \left(\sum_{i=1}^{t-1} v_i + \lambda_t y_t\right)\right)f_t\right] \\
&= E\left[\left(v_t + \sum_{i=t+1}^T v_i - \lambda_t(1 + \eta)(x_t + u_t)\right)\eta(x_t + u_t)\right] \\
&= \eta E[\beta_t v_t^2 - \lambda_t(1 + \eta)\beta_t^2 v_t^2 - \lambda_t(1 + \eta)u_t^2] \\
&= \eta \left[ \frac{\sigma_v^2}{2\lambda_t} - \frac{1 + \eta}{4\lambda_t}\sigma_v^2 - \lambda_t(1 + \eta)\sigma_u^2 \right] \\
&= \eta \left[ \frac{(1 - \eta)\sigma_v\sqrt{\Sigma_{t-1}}}{2} - \frac{\sigma_v}{2\sqrt{\Sigma_{t-1}}}(1 + \eta)\sigma_u^2 \right] \\
&= \frac{\eta}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta)\sigma_v\Sigma_{t-1} - (1 + \eta)\sigma_u^2\sigma_v \right]
\end{aligned}$$

When  $t \rightarrow \infty$ ,

$$\lim_{t \rightarrow \infty} E[\pi_t^f] = \frac{\eta}{2\sqrt{\Sigma^*}} \left[ (1 - \eta)\sigma_v \Sigma^* - (1 + \eta)\sigma_u^2 \sigma_v \right] = 0$$

If  $\Sigma_{t-1} > \Sigma^* = \frac{1+\eta}{1-\eta}\sigma_u^2$  and  $0 < \eta < 1$ ,

$$E[\pi_t^f] = \frac{\eta}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta)\sigma_v \Sigma_{t-1} - (1 + \eta)\sigma_u^2 \sigma_v \right] > 0$$

If  $\Sigma_{t-1} < \frac{1+\eta}{1-\eta}\sigma_u^2$  and  $-1 < \eta < 0$ ,

$$E[\pi_t^f] = \frac{\eta}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta)\sigma_v \Sigma_{t-1} - (1 + \eta)\sigma_u^2 \sigma_v \right] > 0$$

So for single period  $t$ , high frequency trader get positive rent if  $\eta(\Sigma_{t-1} - \Sigma^*) > 0$ .

For aggregate profit over time, we need to analyze the path of belief  $\Sigma_t$ . Consider any period  $t$ , assume  $\Sigma_{t-1} > \Sigma^*$ , we have

$$\begin{aligned} E[\Sigma_t] - E[\Sigma_{t-1}] &= \frac{2b_{t-1} + \eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} - \frac{b_{t-1}}{a_{t-1} - 1} \\ &= \frac{\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} - \frac{b_{t-1}}{2(a_t - 1)(a_{t-1} - 1)} \\ &= \frac{\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} - \frac{\Sigma_{t-1}}{2a_t - 2} \\ &= \frac{1}{2a_t - 2} \left[ (1 + \eta)^2 \sigma_u^2 - (1 - \eta^2) \Sigma_{t-1} \right] \\ &< \frac{1}{2a_t - 2} \left[ (1 + \eta)^2 \sigma_u^2 - (1 - \eta^2) \Sigma^* \right] = 0 \end{aligned}$$

and

$$\begin{aligned}
E[\Sigma_t] - E[\Sigma^*] &= \frac{2b_{t-1} + \eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} - \Sigma^* \\
&= \frac{\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} + \left( \frac{b_{t-1}}{a_{t-1} - 1} \right) \left( \frac{a_{t-1} - 1}{a_t - 1} \right) - \Sigma^* \\
&= \frac{\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} + \left( \frac{a_{t-1} - 1}{a_t - 1} \right) \Sigma_{t-1} - \Sigma^* \\
&> \frac{\eta^2 \Sigma_{t-1} + (1 + \eta)^2 \sigma_u^2}{2a_t - 2} - \frac{\Sigma^*}{2a_t - 2} = 0
\end{aligned}$$

Thus  $\Sigma_{t-1} > \Sigma^*$  implies

$$E[\Sigma_t] - E[\Sigma_{t-1}] < 0 \quad \text{and} \quad E[\Sigma_t] - \Sigma^* > 0$$

Given  $\Sigma_0 > \Sigma^*$ , we have  $\lim_{t \uparrow \infty} \Sigma_t \downarrow \Sigma^*$ , together with  $\eta > 0$ , we have

$$E(\pi_t^f) > 0 \quad \forall t > 0, \quad \text{and} \quad \lim_{t \rightarrow \infty} E(\pi_t^f) = 0$$

Similarly, we can have given  $\Sigma_0 < \Sigma^*$  and  $\eta < 0$ ,

$$E(\pi_t^f) > 0 \quad \forall t > 0, \quad \text{and} \quad \lim_{t \rightarrow \infty} E(\pi_t^f) = 0$$

■

#### 4.3.4 Market Impact of High Frequency Trader

In the market without high frequency trader, the equilibrium strategy is

$$\begin{cases} \beta_t = \frac{1}{2\lambda_t} \\ \lambda_t = \frac{\beta_t \sigma_v^2}{\beta_t^2 \sigma_v^2 + \sigma_u^2} \end{cases}$$

and  $\beta_t = \frac{\sigma_u}{\sigma_v}$ ,  $\lambda_t = \frac{\sigma_v}{2\sigma_u}$ . The expected profit of insider trader at time  $t$  is

$$E[\pi_t] = E[x_t(v - p_t)] = E[x_t(v_t - \lambda x_t)] = E[\beta_t v_t^2 - \lambda_t \beta_t^2 v_t^2] = \frac{\sigma_u \sigma_v}{2}.$$

From zero sum game property, the expected loss of noise traders is

$$E[L_t] = E[\pi_t^i] = \frac{\sigma_u \sigma_v}{2}$$

In the market with high frequency trader, the total expected profit of insider and high frequency trader is

$$\begin{aligned} E[\pi_t^i] + E[\pi_t^f] &= \frac{(1 - \eta)\sigma_v \sqrt{\Sigma_{t-1}}}{2} + \frac{\eta}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta)\sigma_v \Sigma_{t-1} - (1 + \eta)\sigma_u^2 \sigma_v \right] \\ &= \frac{\sigma_v}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta^2)\Sigma_{t-1} - \eta(1 + \eta)\sigma_u^2 \right] \end{aligned}$$

The expected loss of noise traders is

$$E[\tilde{L}_t] = E[\pi_t^i] + E[\pi_t^f] = \frac{\sigma_v}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta^2)\Sigma_{t-1} - \eta(1 + \eta)\sigma_u^2 \right]$$

**Theorem 4.3.3** *There are two sufficient conditions that high frequency trader protect noise traders*

1.  $\eta > 0$  and  $\Sigma_0 \in \left( \Sigma^*, \frac{\sigma_u^2 (1 + \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})^2}{4(1 - \eta^2)^2} \right)$ ;
2.  $\eta < 0$  and  $\Sigma_0 \in \left( \frac{\sigma_u^2 (1 - \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})^2}{4(1 - \eta^2)^2}, \Sigma^* \right)$ .

*Specially, under condition 2, high frequency trader plays beneficial role for the market.*

**Proof** The high frequency trader protects noise traders is equivalent to

$$\sum_{t \leq T} [E[L_t] - E[\tilde{L}_t]] > 0 \quad \forall T > 1$$

where

$$E[L_t] - E[\tilde{L}_t] = \frac{\sigma_u \sigma_v}{2} - \frac{\sigma_v}{2\sqrt{\Sigma_{t-1}}} \left[ (1 - \eta^2)\Sigma_{t-1} - \eta(1 + \eta)\sigma_u^2 \right]$$

Consider the limit situation, when  $t \rightarrow \infty$ ,  $\Sigma_{t-1} \rightarrow \Sigma^* = \frac{1+\eta}{1-\eta}\sigma_u^2$ , assume  $\eta \neq 0$ , we have  $E[\pi_\infty^f] = 0$  and

$$E[\tilde{L}_\infty] = \frac{(1 - \eta)\sigma_v \sqrt{\Sigma^*}}{2} = \frac{\sqrt{1 - \eta^2}\sigma_v \sigma_u}{2} < \frac{\sigma_v \sigma_u}{2} = E[L_\infty]$$

thus when  $t \rightarrow \infty$ , high frequency trader protects noise traders regardless  $\eta > 0$  or  $\eta < 0$ .

Now consider finite time situation,

$$E[L_t] - E[\tilde{L}_t] > 0$$

is equivalent to

$$(1 - \eta^2)\Sigma_{t-1} - \sigma_u \sqrt{\Sigma_{t-1}} - \eta(1 + \eta)\sigma_u^2 < 0$$

When  $\eta > 0$  and  $\Sigma_{t-1} > \Sigma^*$ , we have for all  $0 < \eta < 1$ ,

$$\sigma_u^2 + 4(1 - \eta^2)\eta(1 + \eta)\sigma_u^2 > 0$$

and

$$\frac{\sigma_u(1 - \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)} < \sqrt{\Sigma^*} < \frac{\sigma_u(1 + \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)}$$

If choose  $\Sigma_0 \in \left( \Sigma^*, \frac{\sigma_u^2 \left( 1 + \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)} \right)^2}{4(1 - \eta^2)^2} \right)$ , for all  $t > 0$ ,

$$\sqrt{\Sigma^*} < \sqrt{\Sigma_{t-1}} < \sqrt{\Sigma_0} < \frac{\sigma_u(1 + \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)}$$

thus

$$E[L_t] - E[\tilde{L}_t] > 0, \quad \forall t > 0$$

which implies

$$\sum_{t \geq T} E[L_t] - \sum_{t \leq T} E[\tilde{L}_t] > 0, \quad \forall T > 1$$

When  $\eta < 0$  and  $\Sigma_{t-1} < \Sigma^*$ , we have for all  $-1 < \eta < 0$ ,  $\eta(1 + \eta) \geq -\frac{1}{4}$  thus

$$\sigma_u^2 + 4(1 - \eta^2)\eta(1 + \eta)\sigma_u^2 > 0$$

and

$$\frac{\sigma_u(1 - \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)} < \sqrt{\Sigma^*} < \frac{\sigma_u(1 + \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)}$$

If choose  $\Sigma_0 \in \left( \frac{\sigma_u^2 \left( 1 - \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)} \right)^2}{4(1 - \eta^2)^2}, \Sigma^* \right)$ , for all  $t > 0$ ,

$$\frac{\sigma_u(1 - \sqrt{1 + 4(1 - \eta^2)\eta(1 + \eta)})}{2(1 - \eta^2)} < \sqrt{\Sigma_0} < \sqrt{\Sigma_{t-1}} < \sqrt{\Sigma^*}$$

thus

$$E[L_t] - E[\tilde{L}_t] > 0, \quad \forall t > 0$$

which implies

$$\sum_{t \geq T} E[L_t] - \sum_{t \leq T} E[\tilde{L}_t] > 0, \quad \forall T > 1$$

When  $\Sigma^t < \Sigma^*$  and  $\eta < 0$ , the high frequency trader takes order placed by slow agents, who makes the market and provides extra liquidity, and thus plays beneficial role. ■

#### 4.4 Illustrative Numerical Experiment

Consider the trading process with total  $T = 10^5$  trading periods. Asset value is reveal over time, while  $\tilde{v} = \sum_{i=1}^T v_i$  and  $\{v_i\}$  are i.i.d random variables drawn from  $\mathcal{N}(0, \sigma_v)$ , where we set  $\sigma_v = 0.025$ .

Set the high frequency trader parameter  $\eta = 0.1$  and noise trader variance  $\sigma_u = 0.1$ , the true limit liquidity variance is  $\sigma^* = \frac{1+\eta}{1-\eta} = 0.1222$ . The prior liquidity variance belief is  $a = 3$ ,  $b = 0.65$  and thus  $\Sigma_0 = 0.325$ .

This set of parameters setting satisfies the conditions to ensure high frequency trader makes positive expected rent and plays beneficial role during the trading process.

The profit of informed trader for each trading period is calculated as

$$\pi^i(t) = (\tilde{v} - p(t))x(t), \quad \pi^f(t) = (\tilde{v} - p(t))f(t)$$

The value revealed path is shown in figure 4.1. The price path is perturbed closely around this value path. Set Baye's update scale parameter to  $\epsilon = 0.01$ . We report the belief update path  $\Sigma_t$  in figure 4.2. The expectation of liquidity variance converges to the real limit variance  $\sigma^*$ .

In figure 4.3, we report the cumulative profit of informed trader in the market with and without high frequency trader. The presence of high frequency trader significantly reduce the cumulative profit of informed trader. The margin builds up mainly during

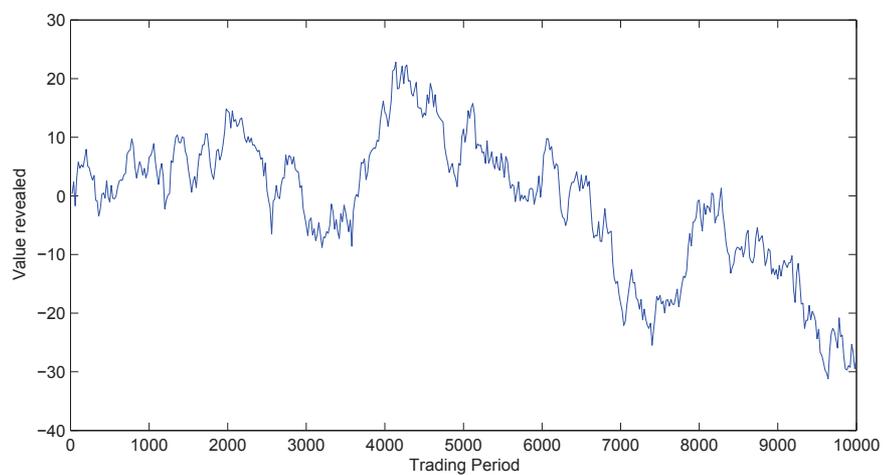


Fig. 4.1. Value revealed path

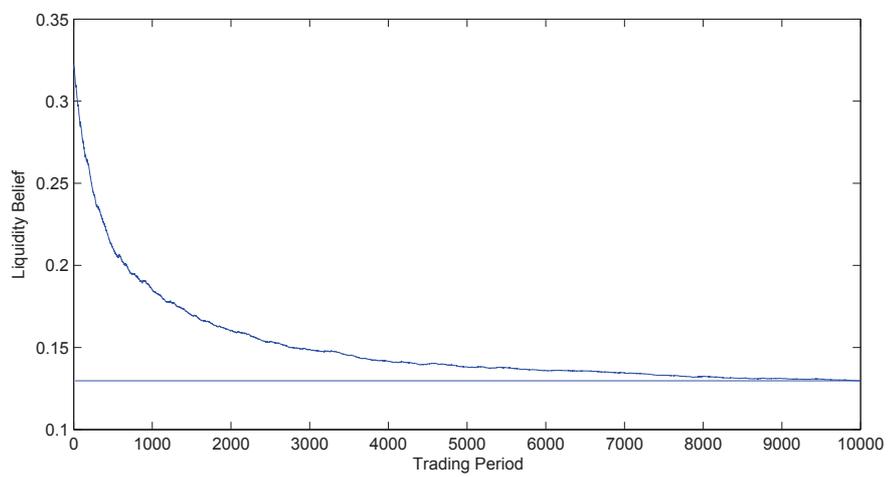


Fig. 4.2. Variance expectation update path

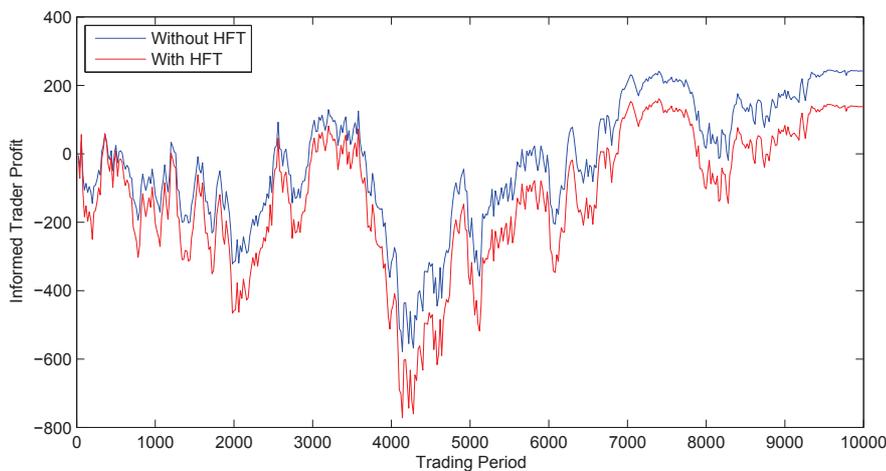


Fig. 4.3. Cumulative profit of informed trader

the first half of trading periods while the liquidity expectation is significant away from the true value. When expectation converges close to true value, the margin between profit with and without high frequency trading stops growing. In zero sum market, the loss of noise traders is equal to the total gain of insider traders and high frequency trader. The fast trader profit is less than reduction of insider trader's profit. All fast trader's profit is generated from insider traders, and the noise trader suffer less loss in market with fast traders. We report the cumulative loss of noise traders in figure 4.4.

#### 4.5 Chapter Summary

In this chapter, we focus on identification of the market impact of high frequency trading activities which provide additional liquidity but increase volatility and mispricing. Our work builds under the market environment involving significant insider trading activity in which high frequency trader clearly improves market quality as in-

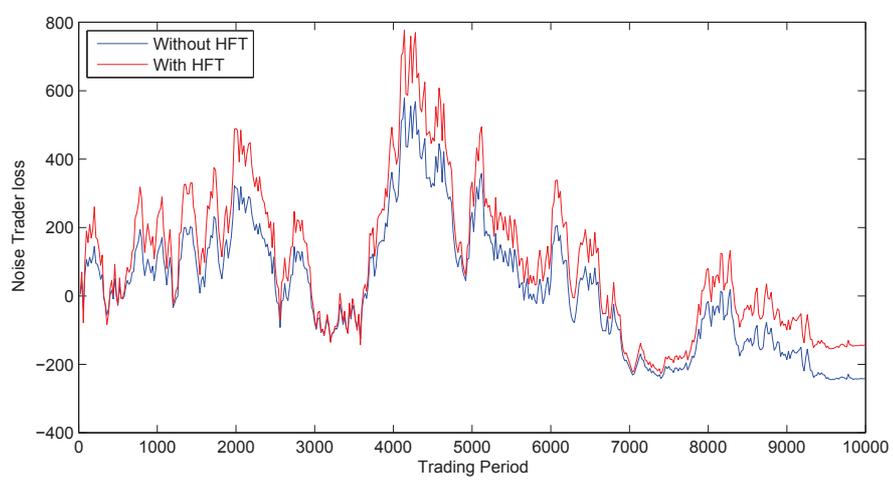


Fig. 4.4. Cumulative profit (loss) of noise traders

creasing the liquidity not only by directly trading but also by reducing the expected loss of liquidity trader to encourage liquidity trading activity.

While complicated, a dynamic model can more accurately describe competition between high frequency traders and informed traders, our goal is to utilize relative simple idea as: high frequency trading activity hides real liquidity variance from market maker and informed trader, that forces informed traders to choose conservative trading strategy without exploring all profit conveyed by private information. Instead of earning profit from noise traders, high frequency trader makes positive rent from the decreased profit portion of the informed trader.

## 5. CONCLUSIONS

In global optimization and equilibrium community, thanks to the development of computer architecture and algorithm such as task parallelism, multiple threads process has been widely applied. The time cost is reduced by trading with insensitive computing power. In decentralized parallel computing scheme, there is no constrain on the uniqueness of timescales of multiple tasks. Parallel threads can be assigned to operate tasks in two or more different timescales. In this dissertation, we introduced three applications facilitated by two timescales technique to demonstrate the effectiveness of this technique in stochastic perturbation control, computational power allocation and dynamic trading systems analysis.

In chapter 2, we introduced two timescales in global optimization scope, most specified, in simulated annealing. A first annealing process operates on a faster timescale and has a drift function that converges to a non-zero noise level. A second annealing process, which operating a slower timescale, is subject to a modified drift term in which the steepest descent direction is perturbed with the first process density gradient. The two time scale technique enables the second annealing process to be able to bypass locally optimal solutions. We have shown that when compared to independent diffusions, the proposed interactive diffusions approach can increase the speed of convergence at the expense of minimal additional computational overhead.

In chapter 3, we apply two timescales to a parallel computing scheme for global optimization that combines multi-start local search with the dynamic reallocation of computational resources. In fast time scale, parallel threads independently execute

a model-based search method, while in slow timescale, searching threads interact through a modified acceptance-rejection rule aimed at preventing duplication of search efforts. The modification of acceptance-rejection rule focus on two ideas related to local search that fails to complete: one is continuing searches that are promising because the end-points have lower objective values than all other solutions find so far and the other is avoiding duplication of failed searches and/or search effort across threads.

In chapter 4, we implement two timescales in a dynamic trading system. In fast timescale, the high frequency trader observes total trading quantity and places order, while the insider trader, the market maker and noise traders operate in a slow timescale. In the market environment involving significant insider trading activity, we show that high frequency trading activity hides real liquidity variance from market maker and the informed trader and forces informed trader to choose conservative trading strategy without exploring all profit generated by private information which implies the high frequency trader clearly improves market quality by increasing the liquidity by reducing the expected loss of liquidity traders.

## LIST OF REFERENCES

## LIST OF REFERENCES

- [1] G. E. Moore *et al.*, “Cramming more components onto integrated circuits,” 1965.
- [2] A. Sandberg, “The physics of information processing superobjects: daily life among the jupiter brains,” *Journal of Evolution and Technology*, vol. 5, no. 1, pp. 1–34, 1999.
- [3] S. Adve, V. S. Adve, G. Agha, M. I. Frank, M. Garzarán, J. Hart, W. Hwu, R. Johnson, L. Kale, R. Kumar, *et al.*, “Parallel computing research at illinois: The upcrc agenda,” *Urbana, IL: Univ. Illinois Urbana-Champaign*, 2008.
- [4] D. E. Culler, J. P. Singh, and A. Gupta, *Parallel computer architecture: a hardware/software approach*. Gulf Professional Publishing, 1999.
- [5] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2012.
- [6] W. D. Hillis and G. L. Steele Jr, “Data parallel algorithms,” *Communications of the ACM*, vol. 29, no. 12, pp. 1170–1183, 1986.
- [7] J. Quinn Michael, “Parallel programming in c with mpi and openmp mcgraw-hill inc. 2004,” tech. rep., ISBN 0-07-058201-7.
- [8] F. Anderson, J. Ziegler, M. Pandya, and R. Whalen, “Application of high-performance computing to numerical simulation of human movement,” *Journal of Biomechanical Engineering*, vol. 117, no. 1, pp. 155–157, 1995.
- [9] J. F. Schutte, J. A. Reinbolt, B. J. Fregly, R. T. Haftka, and A. D. George, “Parallel global optimization with the particle swarm algorithm,” *International Journal for Numerical Methods in Engineering*, vol. 61, no. 13, pp. 2296–2315, 2004.
- [10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: Nsga-ii,” *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 2, pp. 182–197, 2002.
- [11] W. L. Goffe, G. D. Ferrier, and J. Rogers, “Global optimization of statistical functions with simulated annealing,” *Journal of Econometrics*, vol. 60, no. 1, pp. 65–99, 1994.
- [12] R. C. Eberhart and Y. Shi, “Particle swarm optimization: developments, applications and resources,” in *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, vol. 1, pp. 81–86, IEEE, 2001.

- [13] N. Revol, Y. Denneulin, J.-F. Méhaut, and B. Planquelle, “A methodology of parallelization for continuous verified global optimization,” in *Parallel Processing and Applied Mathematics*, pp. 803–810, Springer, 2002.
- [14] A. Uğur and D. Aydin, “An interactive simulation and analysis software for solving tsp using ant colony optimization algorithms,” *Advances in Engineering software*, vol. 40, no. 5, pp. 341–349, 2009.
- [15] N. Bacanin, M. Tuba, and I. Brajevic, “Performance of object-oriented software system for improved artificial bee colony optimization,” *International Journal of Mathematics and Computers in Simulation*, vol. 5, no. 2, pp. 154–162, 2011.
- [16] R. Khasminskii and G. Yin, “Limit behavior of two-time-scale diffusions revisited,” *Journal of Differential Equations*, vol. 212, no. 1, pp. 85–113, 2005.
- [17] M. Jiaqi, Z. Weijiang, and H. Ming, “Asymptotic method of travelling wave solutions for a class of nonlinear reaction diffusion equation,” *Acta Mathematica Scientia*, vol. 27, no. 4, pp. 777–780, 2007.
- [18] S. Bhatnagar, M. C. Fu, S. I. Marcus, and S. Bhatnagar, “Two-timescale algorithms for simulation optimization of hidden markov models,” *IIE Transactions*, vol. 33, no. 3, pp. 245–258, 2001.
- [19] J. H. Chow, *Time-scale modeling of dynamic networks with applications to power systems*, vol. 46. Springer-Verlag, 1982.
- [20] L. Xiao, “Dual averaging method for regularized stochastic learning and online optimization,” in *Advances in Neural Information Processing Systems*, pp. 2116–2124, 2009.
- [21] Y. Wang and A. Garcia, “Interactive model-based search for global optimization,” *Journal of Global Optimization*, pp. 1–17, 2014.
- [22] A. S. Kyle, “Continuous auctions and insider trading,” *Econometrica: Journal of the Econometric Society*, pp. 1315–1335, 1985.
- [23] V. Černý, “Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm,” *Journal of optimization theory and applications*, vol. 45, no. 1, pp. 41–51, 1985.
- [24] S. Kirkpatrick, M. Vecchi, *et al.*, “Optimization by simulated annealing,” *science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [25] S. Geman and C.-R. Hwang, “Diffusions for global optimization,” *SIAM Journal on Control and Optimization*, vol. 24, no. 5, pp. 1031–1043, 1986.
- [26] T.-S. Chiang, C.-R. Hwang, and S. J. Sheu, “Diffusion for global optimization in  $\mathbb{R}^n$ ,” *SIAM Journal on Control and Optimization*, vol. 25, no. 3, pp. 737–753, 1987.
- [27] H. Kushner, “Asymptotic global behavior for stochastic approximation and diffusions with slowly decreasing noise effects: global minimization via monte carlo,” *SIAM Journal on Applied Mathematics*, vol. 47, no. 1, pp. 169–185, 1987.

- [28] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu, “Accelerating gaussian diffusions,” *The Annals of Applied Probability*, pp. 897–913, 1993.
- [29] G. Yin and K. Yin, “Global optimization using diffusion perturbations with large noise intensity,” *Acta Mathematicae Applicatae Sinica*, vol. 22, no. 4, pp. 529–542, 2006.
- [30] O. V. Poliannikov, E. Zhizhina, and H. Krim, “Global optimization by adapted diffusion,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 12, pp. 6119–6125, 2010.
- [31] T. Coleman, D. Shalloway, and Z. Wu, “A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing,” *Journal of Global Optimization*, vol. 4, no. 2, pp. 171–185, 1994.
- [32] Z. Wu, “The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation,” *SIAM Journal on Optimization*, vol. 6, no. 3, pp. 748–768, 1996.
- [33] M. S. Lau and C. Kwong, “A smoothing method of global optimization that preserves global minima,” *Journal of Global Optimization*, vol. 34, no. 3, pp. 369–398, 2006.
- [34] J. Elgin, “The fokker-planck equation: Methods of solution and applications,” *Journal of Modern Optics*, vol. 31, no. 11, pp. 1206–1207, 1984.
- [35] H. Pham, *Continuous-time stochastic control and optimization with financial applications*, vol. 61. Springer Science & Business Media, 2009.
- [36] C. Gardiner, “Handbook of stochastic methods for physics, chemistry and the natural sciences,” *Springer series in synergetics*, vol. 13, pp. 149–168, 1986.
- [37] D. H. Ackley, “A connectionist machine for genetic hillclimbing,” 1987.
- [38] Z. Michalewicz, *Genetic algorithms+ data structures= evolution programs*. Springer Science & Business Media, 1996.
- [39] C. P. Gomes, B. Selman, H. Kautz, *et al.*, “Boosting combinatorial search through randomization,” *AAAI/IAAI*, vol. 98, pp. 431–437, 1998.
- [40] M. Luby and W. Ertel, *Optimal parallelization of Las Vegas algorithms*. Springer, 1994.
- [41] O. V. Shylo, T. Middelkoop, and P. M. Pardalos, “Restart strategies in optimization: parallel and serial cases,” *Parallel Computing*, vol. 37, no. 1, pp. 60–68, 2011.
- [42] J. Hu, M. C. Fu, S. I. Marcus, *et al.*, “A model reference adaptive search algorithm for global optimization,” 2005.
- [43] F. Schoen, “Stochastic techniques for global optimization: A survey of recent advances,” *Journal of Global Optimization*, vol. 1, no. 3, pp. 207–228, 1991.

- [44] R. Martí, J. M. Moreno-Vega, and A. Duarte, “Advanced multi-start methods,” in *Handbook of metaheuristics*, pp. 265–281, Springer, 2010.
- [45] E. Onbaşıoğlu and L. Özdamar, “Parallel simulated annealing algorithms in global optimization,” *Journal of Global Optimization*, vol. 19, no. 1, pp. 27–50, 2001.
- [46] A. Ferreiro, J. García, J. López-Salas, and C. Vázquez, “An efficient implementation of parallel simulated annealing algorithm in gpus,” *Journal of Global Optimization*, vol. 57, no. 3, pp. 863–890, 2013.
- [47] M. D Apuzzo, M. Marino, A. Migdalas, P. Pardalos, and G. Toraldo, “Parallel computing in global optimization,” *STATISTICS TEXTBOOKS AND MONOGRAPHS*, vol. 184, p. 225, 2006.
- [48] C. G. E. Boender and A. R. Kan, “Bayesian stopping rules for multistart global optimization methods,” *Mathematical Programming*, vol. 37, no. 1, pp. 59–80, 1987.
- [49] A. György and L. Kocsis, “Efficient multi-start strategies for local search algorithms,” *Journal of Artificial Intelligence Research*, pp. 407–444, 2011.
- [50] A. Carrion, “Very fast money: High-frequency trading on the nasdaq,” *Journal of Financial Markets*, vol. 16, no. 4, pp. 680–711, 2013.
- [51] A. P. Chaboud, B. Chiquoine, E. Hjalmarrsson, and C. Vega, “Rise of the machines: Algorithmic trading in the foreign exchange market,” *The Journal of Finance*, vol. 69, no. 5, pp. 2045–2084, 2014.
- [52] N. P. Bollen and R. E. Whaley, “Futures market volatility: What has changed?,” *Journal of Futures Markets*, 2014.
- [53] J. Hasbrouck and G. Saar, “Low-latency trading,” *Journal of Financial Markets*, vol. 16, no. 4, pp. 646–679, 2013.
- [54] R. A. Jarrow and P. Protter, “A dysfunctional role of high frequency trading in electronic markets,” *International Journal of Theoretical and Applied Finance*, vol. 15, no. 03, 2012.
- [55] A. A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun, “The flash crash: The impact of high frequency trading on an electronic market,” *Available at SSRN 1686004*, 2014.
- [56] L. Harris, “What to do about high-frequency trading,” *Financial Analysts Journal*, vol. 69, no. 2, pp. 6–9, 2013.
- [57] M. Baldauf and J. Mollner, “High-frequency trade and market performance,” *Working paper*, 2014.
- [58] K. Back, C. H. Cao, and G. A. Willard, “Imperfect competition among informed traders,” *The journal of finance*, vol. 55, no. 5, pp. 2117–2155, 2000.
- [59] H. Hong and S. Rady, “Strategic trading and learning about liquidity,” *Journal of Financial Markets*, vol. 5, no. 4, pp. 419–450, 2002.
- [60] P. Collin-Dufresne and V. Fos, “Insider trading, stochastic liquidity and equilibrium prices,” tech. rep., National Bureau of Economic Research, 2012.