# On the convergence of Parallel Simulated Annealing

Christian Meise
Universität Bielefeld

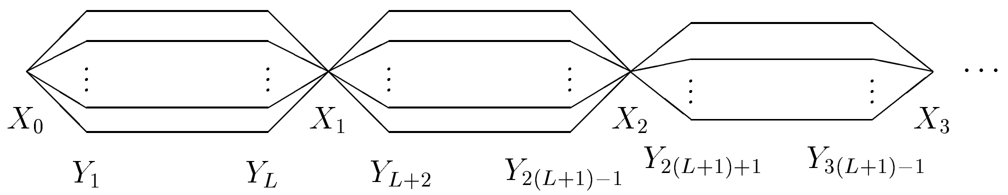
**Abstract.** Using the Matrix-Tree Theorem and coupling methods the convergence of the Parallel Chain (PC) algorithm to the set of global minima is established for various selection functions. It is illustrated that there may be convergence to a set of non-global minima when selecting one of the best states (Best-Wins strategy). In the latter case the convergence of the homogeneous PC algorithm is proved for sufficiently large length of the parallel Markov chains and number of parallel Markov chains. For any selection function we show that if the PC algorithm is convergent, then the cooling schedule must satisfy $\sum_{n=0}^{\infty} \exp(-\beta(n)\theta) = \infty$, $(\theta > 0)$.

**Keywords: Simulated Annealing, Weak ergodicity, Strong ergodicity, Coupling**

## 1 Introduction and Results

Let $X$ be a finite set and let $U : X \to \mathbb{R}_+$ be a non-constant function that should be minimized. We denote by $X_{\min}$ the set of global minima of $U$. Suppose that we have $p \geq 1$ processors. Choose any starting point $x_0 \in X$ and let each processor simulate a Markov chain of length $L \geq 1$ starting in $x_0$. After $L$ transitions the simulation is stopped and a state $x_1$ is selected from the $p$ states according to a selection strategy. Then each processor simulates a Markov chain of length $L$ starting in $x_1$. Again at the end of the simulation a state $x_2$ is selected from the $p$ states and the next simulation starts from $x_2$, etc.

**Figure 1.1** *Illustration of the algorithm's working method.*



Of course the number of parallel Markov chains that are simulated need not be fixed. We denote by $r(n) \geq 1$ the number of parallel Markov chains starting at time $n(L+1)$ and suppose that $(r(n))_n$ is monotone decreasing. Moreover we assume that $\lim_{n\to\infty} r(n) =: p \in \{1,\ldots,r(0)\}$ exists.

**Remark 1.2** *In our model the Markov chain transitions should depend on $U$: If only worst states, e.g. states with maximal costs, are selected, there will be no convergence to the set of global minima, when using Markov chain transitions independent of $U$. In comparison to the model of genetic algorithms in [C94], minimization of $U$ in the PC algorithm is achieved via the Simulated Annealing transition kernel and not by appropriate selection functions. Recall that the operations such as crossover and mutation used in the model of Cérf do not depend on $U$. Therefore optimization in the model of Cérf can only be achieved by using appropriate selection functions.*

In the sequel we specify the transition probabilities for each of the $r(n)$ Markov chains starting at time $n(L+1)$. Let $q$ be an irreducible exploration kernel on $X$, that also satisfies the following reversibility condition

$$q(x,y) > 0 \quad \Longleftrightarrow \quad q(y,x) > 0 \text{ for all } x,y \in X. \tag{1}$$

Moreover let $\beta(n,i)$ $(n \in \mathbb{N}_0,\ 0 \leq i < L)$ denote a sequence satisfying the following properties:

1. $\beta(n) \equiv \beta(n,0) =\leq \beta(n,i) \leq \beta(n,L-1) \leq \beta(n+1,0)$, for $n \in \mathbb{N}$ and $0 \leq i < L$
2. $\lim_{n\to\infty} \beta(n) = \infty$.
3. $\lim_{n\to\infty} \beta(n,L-1) - \beta(n) = 0$

The transition probabilities for the $r(n)$ Markov chains, being simulated at time $n(L+1)$, are given by the classical Simulated Annealing transition kernel, e.g. at time $n(L+1)+i$, $(0 \leq i \leq L-1)$ the transition probability for the $j$-th Markov chain, $j \in \{1,\ldots,r(n)\}$, is given by

$$P\left(Y^j_{n(L+1)+i+1} = z \mid Y^j_{n(L+1)+i} = y\right) = p_{\beta(n,i)}(y,z) \quad \text{for } y,z \in X, \tag{2}$$

where

$$p_\beta(x,y) = \begin{cases} \exp\left(-\beta\left(U(y)-U(x)\right)^+\right) q(x,y) & : \quad x \neq y \\ 1 - \sum_{z\neq x} p_\beta(x,z) & : \quad x = y. \end{cases} \tag{3}$$

The selection behavior is described by selection functions.

**Definition 1.3** *A function $F : \{1,\ldots,s\} \times X \times X^s \to [0,1]$ is called a selection function (of degree $s$), if*

$$\textstyle\sum_{m=1}^{s} F(m,x,y_1,\ldots,y_s) = 1 \textit{ for } (x,y_1,\ldots,y_s) \in X \times X^s.$$

**Remark 1.4** *The selection behavior is determined by selection functions in the following way: The probability for selecting the state* $y_j$ *from* $\{y_1, \ldots, y_s\}$ *is given by* $\sum_{l:y_l=y_j} F(l, x, y_1, \ldots, y_s)$.
*The second argument in the definition of selection functions is necessary for selection strategies that also depend on states visited before. For example this is the case for the First-Wins strategy, introduced below.*

Let $(F_m)_{m\geq 1}$ be a sequence of selection functions. At time $(n+1)(L+1)-1$ the selection function $F_{n+1}$ is used for choosing the next starting point in the following way:
We assume that the selection functions $(F_n)_n$ satisfy $H_0$:
There exists $N \in \mathbb{N}_1$ such that $F_n = F_m \equiv F$ for all $n, m \geq N$.

For an arbitrary selection function we have the following necessary condition.

**Theorem 1.5** *Necessary condition.*
*If* $\limsup_{n\to\infty} P_x(X_n \in X_{\min}) = 1$, *then* $\sum_{n=0}^{\infty} \exp(-\beta(n)\theta) = \infty$, *where*

$$0 \leq \theta \leq \max\{\, (U(y) - U(x)) \mid x \in X \setminus X_{\min},\ y \in N_1(x)\,\}.$$

*We have* $\theta = 0$ *if and only if for any* $x \in X \setminus X_{\min}$ *there exists a path* $x = \gamma_0, \gamma_1, \ldots, \gamma_l \in X_{\min}$ *satisfying* $U(\gamma_{i+1}) \leq U(\gamma_i)$ *for* $i \in \{0, \ldots, l-1\}$.

Theorem 1.5 follows from Corollary 3.2.

For $x = (x_1, \ldots, x_s) \in X^s$ we introduce

$$\begin{aligned}
[x] &:= \{\, x_i \mid i \in \{1, \ldots, s\}\} \\
x_* &:= \{\, x_i \mid U(x_i) = \min\{\, U(x_j) \mid j \in \{1, \ldots, s\},\ i \in \{1, \ldots, s\}\} \\
x^* &:= \{\, x_i \mid U(x_i) = \max\{\, U(x_j) \mid j \in \{1, \ldots, s\}\},\ i \in \{1, \ldots, s\}\}.
\end{aligned}$$

In this paper we investigate the PC algorithm for the following selection functions

1. First-Wins.

$$F_\beta(m, x, z_1, \ldots, z_s) = \begin{cases} 1 & : \quad z_1 = \ldots = z_{m-1} = x,\ z_m \neq x \\ 1 & : \quad z_1 = \ldots = z_s = x \\ 0 & : \quad z_m = x,\ \{x\} \neq [z] \end{cases}$$

   If there exists $i$ with $z_i \neq x$ the state $z_j$ with smallest such $j$ will be chosen, else the process will stay in $x$.

2. Chance-to-anyone.

$$F_\beta(m, x, z_1, \ldots, z_s) \geq c > 0 \text{ for all } m \in \{1, \ldots, s\},\ \beta \in \mathbb{R}_+.$$

3. Best-Wins.
Only the best states are chosen with equal probability, e.g.

$$F_\beta(m, x, z_1, \ldots, z_s) = \frac{1}{|z_*|} 1_{z_*}(z_m).$$

4. Worst-Wins.
Only the worst states are chosen with equal probability, e.g.

$$F_\beta(m, x, z_1, \ldots, z_s) = \frac{1}{|z^*|} 1_{z^*}(z_m).$$

Define

$$m_{\min} := \min \left\{ \sum_{(y \to z) \in g} (U(z) - U(y))^+ \mid g \in G_1\{x\} \right\},$$

where $G_1\{x\}$ denotes the set of $\{x\}$-graphs with respect to the transition kernel $p_{\beta(n)}$ and $x \in X_{\min}$ is arbitrary.

**Theorem 1.6** *Sufficient conditions.*
*a) Let $F$ be as in cases 1,2 or 4 and assume that the rectangle sequence satisfies $\beta(n) = \frac{1}{m_{\min}} \ln(n+1)$. Then for any initial distribution $\nu$ on $X$ the PC algorithm converges to the set of global minima, e.g. $\lim_{n\to\infty} \nu_n(X_{\min}) = 1$, where $\nu_n$ denotes the distribution of the algorithm at time $n$.*
*b) Let $F$ be as in case 3. Define $\delta := \min^+\{\ U(y) \mid y \in N_1(x),\ x \in X_{\min}\ \}$. Moreover let $\Gamma_{xy}^{\mathbb{N}_1}$ denote the paths of length $l \in \mathbb{N}_1$ leading from $x$ to $y$ with respect to the kernel $p_{\beta(n)}$. Define*

$$\theta := \max_{x \in X \setminus X_{\min}} \min \left\{ \sum_{e \in \gamma} (U(e^+) - U(e^-))^+ \mid z \in X_{\min},\ \gamma \in \Gamma_{xz}^{\mathbb{N}_1} \right\}.$$

*Then the homogeneous PC algorithm converges if $p\delta > \theta$ and $L$ is chosen such that for all $x \in X$ there exists $z \in X_{\min}$ and a path $\gamma \in \Gamma_{xy}^{\mathbb{N}_1}$ of length $L$ such that $\sum_{e\in\gamma}(U(e^+ - U(e^-)^+ \leq \theta$. If $L$ and $p$ are not large enough example (2.11) shows that the PC algorithm may converge to local but not global minima.*

**Remark 1.7** *A slightly different version of the PC algorithm has been used in [DLS]. In [DLS] the lengths of Markov chains tend to infinity while the number of parallel Markov chains decreases to one. Using the Boltzmann-distribution function and self-adapting cooling schedules, considerable speed-ups have been observed in several simulations.*

## 2 Convergence of the homogeneous algorithm

The sequence space is given by $\Omega = (X^{r(0)})^{L+1} \times (X^{r(1)})^{L+1} \times \ldots$. Let $Y_{n(L+1)+i}$ denote the projection, e.g. $Y^k_{n(L+1)+i}(\omega) = \omega_k^{n,i}$ for $0 \le i \le L$. By the definition of the algorithm we have $Y_{n(L+1)} \in \{(x, \ldots, x) \mid x \in X\}$ for all $n \ge 0$. Therefore by identifying the tuples $(x, \ldots, x)$ with $x$, we obtain a Markov chain $(X_n)_n$ on $X$ by the definition $X_n \equiv Y_{n(L+1)}$.

For $l \in \mathbb{N}_1$ and $x \in X$ introduce the set of $l$-step neighbors of $x$ under the kernel $p_{\beta(n)}$

$$N_l(x) := \left\{ \, y \in X \;\mid\; p^{(l)}_{\beta(n)}(x,y) > 0 \, \right\}.$$

For $x, y \in X$ denote by

$$\mathcal{N}_k(x,y) = \mathcal{N}_k^L(x,y) := \left\{ \, v \in N_L(x)^k \;\mid y \in [v] \, \right\}$$

the set of $k$-tuples whose components are $L$-step neighbors of $x$ containing $y$.

By definition of the algorithm we conclude for the transition probabilities of the chain $(X_n)_n$:

$$\begin{aligned}
& P(X_{n+1} = y \mid X_n = x) \\
= & \sum_{v \in \mathcal{N}_{r(n)}(x,y)} P(X_{n+1} = y \mid Y_{nL-1} = v) P(Y_{nL-1} = v \mid X_n = x) \\
= & \sum_{v \in \mathcal{N}_{r(n)}(x,y)} \sum_{l : v_l = y} F_n(l, x, v_1, \ldots, v_{r(n)}) \prod_{i=1}^{r(n)} P_{n,L}(x, v_i),
\end{aligned}$$

where

$$P_{n,L}(x,y) := \sum_{(z_0,\ldots,z_L) \in \{x\} \times X^{L-1} \times \{y\}} \prod_{i=0}^{L-1} p_{\beta(n,i)}(z_i, z_{i+1}).$$

Define $Q_{n,L}(x,y) := P(X_{n+1} = y \mid X_n = x)$ for $x, y \in X$. Since $Q_{n,L}$ inherits the aperiodicity and irreducibility from $p_{\beta(n)}$, $Q_{n,L}$ has got exactly one stationary measure $\mu_n$.

For a finite subset $A \subset \mathbb{N}_1$ let $G_{n,A} = (X, E_{n,A})$ be the graph that is induced by $Q_{n,A}$, e.g.

$$(y \to z) \in E_{n,A} \quad :\Longleftrightarrow \quad \left( \max_{l \in A} Q_{n,l}(y,z) > 0 \right) \quad \Longleftrightarrow \quad \left( \max_{l \in A} p^{(l)}_{\beta(n)}(y,z) > 0 \right).$$

By condition (2), there exists $N \in \mathbb{N}_1$, such that $\beta(N) > 0$. This implies that the graph $G_{n,A}$ does not depend on $n$ for $n \ge N$. Therefore we write $G_A = G_{n,A}$ for such $n$. Often we have $A = \{L\}$.

Moreover denote by $G_A\{x\}$ the set of $\{x\}$-$A$-graphs, e.g. $G_A\{x\}$ is the set of $\{x\}$–graphs of $G_A$. For any arrow $(y \to z) \in g \in G_A\{x\}$ there exists a path $\gamma$

from $y$ to $z$ of length $l \in A$. We will $\gamma$ regard as $l+1$-tuple $\gamma = (y = \gamma_0, \dots, \gamma_l = z)$ as well as set of its edges

$$\gamma = \{\, (\gamma_i, \gamma_{i+1}) \;\mid\; 0 \leq i < l \,\}.$$

Denote by $\Gamma_{yz}^A$ the set of paths of lengths $l \in A$, that lead from $y$ to $z$.
Introduce the costs of $\gamma \in \Gamma_{yz}^A$

$$s(\gamma) := \sum_{e \in \gamma} \Big( U(e^+) - U(e^-) \Big)^+$$

and let

$$\Gamma_{yz,*}^A := \Big\{ \gamma \in \Gamma_{yz}^A \;\mid\; s(\gamma) = \min \Big\{\, s(\gamma') \;\mid\; \gamma' \in \Gamma_{yz}^A \,\Big\} \Big\} \tag{4}$$

be the set of $s$-minimizing paths.
Choose $\gamma_{yz} \in \Gamma_{yz,*}^A$ for $y, z \in X$ with $\Gamma_{yz}^A \neq \emptyset$. The costs of an $\{x\}$-$A$-graph is given by

$$h_A(g) := \sum_{e \in g} s(\gamma_e). \tag{5}$$

## 2.1 Preparations for the proof of $\lim_{n\to\infty} \mu_n(w_{*,L}) = 1$

In this section we will see, that $X_{\min} = m_{*,L}$, where $m_{*,L}$ is the set of states that minimize the function

$$x \mapsto \min\{\, h_L(g) \;\mid\; g \in G_L\{x\} \,\}.$$

Besides it will be shown that $m_{x,L} := \min\{\, h_L(g) \;\mid\; g \in G_L\{x\} \,\}$ does not depend on $L$.
In section (2.4) we will find out that for large $n$ the measures $\mu_n$ are concentrated on a subset $w_{*,L} \subset X$, where $w_{*,L}$ is determined by another cost definition $\widetilde{h}$.
The concentration of $\mu_n$ on $X_{\min}$ will be shown by establishing the inclusion $m_{*,L} \subset w_{*,L}$ for appropiate selection functions.
In the sequel we need the following definitions

$$m_{x,L} := \min \{\, h_L(g) \;\mid\; g \in G_L\{x\} \,\},$$

$$m_L := \min \{\, m_{x,L} \;\mid\; x \in X \,\}$$

and

$$m_{*,L} := \{\, x \in X \;\mid m_{x,L} = m_L \,\}.$$

Let $\gamma^{-1} := \{\, (e^+, e^-) \;\mid e \in \gamma \,\}$ denote the reverse path. Note that $\gamma^{-1}$ is a path by (1). The following Lemma gives a formula for the cost of the reverse path. This fact is used for deriving that $m_{*,L} = X_{\min}$.

**Lemma 2.1** *(Reverse path formula) Let $y \in X$ and $z \in N_L(y)$. For any path $\gamma \in \Gamma^L_{yz}$ from $y$ to $z$ we have*

$$s(\gamma) = s(\gamma^{-1}) + (U(z) - U(y)) \,. \tag{6}$$

**Proof:** Using $(-a)^+ = a^+ - a$ for $a \in \mathbb{R}$, we conclude

$$\begin{aligned} s(\gamma^{-1}) &= \sum_{e\in\gamma} \Big(U(e^-) - U(e^+)\Big)^+ \\ &= \sum_{e\in\gamma} \Big(U(e^+) - U(e^-)\Big)^+ - \Big(U(e^+) - U(e^-)\Big) \\ &= s(\gamma) - \sum_{e\in\gamma} \Big(U(e^+) - U(e^-)\Big) \\ &= s(\gamma) - (U(z) - U(y)) \quad \square \end{aligned}$$

The formula for reverse paths $\gamma \in \Gamma^L_{xy}$ yields a formula for reverse $\{x\}$-$L$-graphs.

**Lemma 2.2** *(Reverse graph formula) Let $x, y \in X$ and $g \in G_L\{x\}$. Moreover let $\eta \subset g$ be the uniquely determined path in $g$ from $y$ to $x$. Then for the $\{y\}$-$L$-graph $g' := g \setminus \eta \cup \eta^{-1}$ holds*

$$h_L(g') = h_L(g) + (U(y) - U(x)). \tag{7}$$

**Proof:** At first we show that

$$s(\gamma_{yz}^{-1}) = s(\gamma_{zy}) \text{ for } y, z \in X \text{ with } (y \to z) \in g \in G_L\{x\}. \tag{8}$$

**Proof** of (8): Since $\gamma_{yz}^{-1} \in \Gamma^L_{zy}$, it follows that $s(\gamma_{yz}^{-1}) \geq s(\gamma_{zy})$.
Applying (6) to $\gamma_{zy}$, yields

$$\begin{aligned} s(\gamma_{zy}) &= \underbrace{s(\gamma_{zy}^{-1})}_{\geq s(\gamma_{yz})} + (U(y) - U(z)) \geq s(\gamma_{yz}) + (U(y) - U(z)) \\ &= s(\gamma_{yz}^{-1}) \quad \square \text{ (8)} \end{aligned}$$

For $g$ and $g'$ as in the assertion, we conclude

$$\begin{aligned} h_L(g') &= h_L(g) - \sum_{e\in\eta} \Big(s(\gamma_e) - s(\gamma_e^{-1})\Big) \\ &= h_L(g) + \sum_{e\in\eta} \Big(U(e^-) - U(e^+)\Big) \\ &= h_L(g) + (U(y) - U(x)), \end{aligned} \tag{9}$$

where (6) has been used in (9). $\square$
Denote by

$$G_{L,*}\{x\} := \{\, g \in G_L\{x\} \mid h_L(g) = m_{x,L} \,\}$$

the set of $\{x\}$-$L$-graphs minimizing the costs $h_L$.
Next, we obtain a satisfying description of the set $m_{*,L}$.

**Proposition 2.3** *For all $L \geq 1$ we have $m_{*,L} = X_{\min}$.*

**Proof:** At first we show that $m_{*,L} \subset X_{\min}$.
Let $g \in G_{L,*}\{x\}$, $y \in X_{\min}$ and let $\eta \subseteq g$ be the path in $g$ leading from $y$ to $x$. By applying the reverse formula for graphs (7) to $g' := g \setminus \eta \cup \eta^{-1}$, we obtain for $x \notin X_{\min}$

$$m_{y,L} \leq h_L(g') = h_L(g) + \underbrace{(U(y) - U(x))}_{<0} < h_L(g) = m_{x,L}.$$

Hence $m_L \leq m_{y,L} < m_{x,L}$ and $x \notin m_{*,L}$.
For $x \in X_{\min}$, we obtain

$$m_{y,L} \leq h_L(g') = h_L(g) = m_{x,L}.$$

By symmetry it also holds $m_{y,L} \geq m_{x,L}$, hence $m_{y,L} = m_{x,L}$ if $x, y \in X_{\min}$. □

Finally we show that $m_{x,L}$ does not depend on $L$. This fact will be used for the establishement of $w_{*,L} = X_{\min}$ for Worst-Wins in (2.6).

**Lemma 2.4** *Let $x \in X$. Then for $L \geq 1$*

$$m_{x,1} = \min\{\, h_1(g) \mid g \in G_1\{x\}\,\} = \min\{\, h_L(g) \mid g \in G_L\{x\}\,\} = m_{x,L}.$$

**Proof:** Let $g \in G_L\{x\} \subseteq G_{\{1,L\}}\{x\}$. By definition of $h_{\{1,L\}}$, we have $h_{\{1,L\}}(g) \leq h_1(g)$.
We show that there exists $\tilde{g} \in G_1\{x\}$ such that $h_1(\tilde{g}) \leq h_{\{1,L\}}(g)$.
This is done by induction over $|\mathcal{K}(g)|$, where

$$\mathcal{K}(g) := \Big\{\, (y \to z) \in g \mid p_{\beta(n)}(y,z) = 0 \,\Big\} \text{ for } g \in G_{\{1,L\}}.$$

For $|\mathcal{K}(g)| = 0$ the graph $g$ is already element of $G_1\{x\}$, so that we can choose $\tilde{g} := g$. (Note that paths of length 1 have minimal costs, so that $h_1(g) \leq h_{\{1,L\}}(g)$).
Let $|\mathcal{K}(g)| = n + 1$. By induction hypothesis there exists a $\{x\}$-1-graph $\tilde{g}$ with $h_1(\tilde{g}) \leq h_{\{1,L\}}(g)$, if $g$ is a $\{x\}$-$\{1, L\}$-graph $g$ with $|\mathcal{K}(g)| \leq n$.
Let $(y \to z) \in \mathcal{K}(g)$.
Set $\eta := \gamma_{yz} \in \Gamma^L_{yz}$. For any $i \in \{0, \ldots, L\}$ there exists for $\eta_i$ exactly one $\eta'_i$, such that $(\eta_i \to \eta'_i) \in g$. Define the following $\{x\}$-$\{1, L\}$-graph

$$g' := g \setminus \left( \bigcup_{i=0}^{L-1} (\eta_i \to \eta'_i) \right) \cup \left( \bigcup_{i=0}^{L-1} (\eta_i \to \eta_{i+1}) \right).$$

Since $p_{\beta(n)}(\eta_i, \eta_{i+1}) > 0$, we can choose the path $\gamma_{\eta_i\eta_{i+1}} := (\eta_i, \eta_{i+1})$ of length 1 as minimizing path from $\eta_i$ to $\eta_{i+1}$.

Using $s(\gamma_{\eta_i\eta_{i+1}}) = (U(\eta_{i+1}) - U(\eta_i))^+$, we derive for the costs of $g'$

$$\begin{aligned}
h_{\{1,L\}}(g') &= h_{\{1,L\}}(g) - \sum_{i=0}^{L-1} s(\gamma_{\eta_i\eta_i'}) + \underbrace{\sum_{i=0}^{L-1} s(\gamma_{\eta_i\eta_{i+1}})}_{=s(\gamma_{yz})} \\
&= h_{\{1,L\}}(g) - s(\gamma_{yz}) - \sum_{i=1}^{L-1} \underbrace{s(\gamma_{\eta_i\eta_i'})}_{\geq 0} + s(\gamma_{yz}) \\
&\leq h_{\{1,L\}}(g).
\end{aligned}$$

Note that $\mathcal{K}(g')$ has at most $n$ elements. By induction hypothesis there exists $\tilde{g} \in G_1\{x\}$ with $h_1(\tilde{g}) \leq h_{\{1,L\}}(g')$. Hence $h_1(\tilde{g}) \leq h_{\{1,L\}}(g') \leq h_{\{1,L\}}(g) \leq h_L(g)$. This yields $m_{x,1} \leq m_{x,L}$.
We have to show that also $m_{x,1} \geq m_{x,L}$ holds.
Let $g \in G_{1,*}\{x\}$ and $(y \to z) \in g$.
If $L$ is an odd number any 1-path $\gamma_{yz}$ of $g$ can be replaced by a path $\gamma_{yz}^L$ of length $L$, that has the same costs. Therefore $h_1(g) = h_L(g)$ for $g \in G_{1,*}\{x\}$, $L$ odd.
Let $L$ be even.
Case a) $p_{\beta(n)}(x,x) > 0$.
If $p_{\beta(n)}(y,y) > 0$ resp. $p_{\beta(n)}(z,z) > 0$ we can choose the $L$-path

$$(y, \ldots, y, z) \in \Gamma_{yz,*}^L \text{ resp. } (y, z, \ldots, z) \in \Gamma_{yz,*}^L.$$

If $p_{\beta(n)}(y,y) = p_{\beta(n)}(z,z) = 0$, we have $U(y) = U(z)$, $z \neq x$ and there exists $(z \to z') \in g$. Moreover $U(z) \geq U(z')$ because $p_{\beta(n)}(z,z) = 0$. Hence we can choose the $L$-path $(y, z, y, z, \ldots, y, z, z') \in \Gamma_{yz',*}^L$.
Case b) $p_{\beta(n)}(x,x) = 0$.
For $g \in G_{\{1,L\}}(x)$ let $\mathcal{H}(g)$ be the set of arrows $(y \to z) \in g$ with $z \in N_1(y)$, such that there does not exist a path $\gamma_{yz}^L$ of length $L$ satisfying $s(\gamma_{yz}^L) \leq (U(z) - U(y))^+$.
By induction we assume that there exist $g' \in G_L\{x\}$ such that $h_L(g') \leq h_1(g)$, for $|\mathcal{H}(g)| \leq n$, $g \in G_{\{1,L\}}\{x\}$.
Since $U$ is non-constant there exists $m \in \mathbb{N}_1$ and a path $\eta = (\eta_{m+1}, \eta_m, \ldots, \eta_1, \eta_0 = x)$ in $g$, such that

$$U(\eta_{m+1}) \neq U(\eta_m) = U(\eta_{m-1}) = \ldots = U(\eta_0) = U(x).$$

Write $m = kL + i$ with $0 \leq i < L$.
If $U(\eta_{m+1}) > U(\eta_m)$ we have $p_{\beta(n)}(\eta_m, \eta_m) > 0$. Therefore the path $(\eta_m, \ldots, \eta_m, \eta_{m-1}, \ldots, \eta_{kL})$ is a path of length $L$, where we have $L - i$ loops in $\eta_m$. The other states are sent to $x$ via $\eta_m$.
More precise we take the arrows

$$V := \{(\eta_{m+1} \to \eta_m)$$

$$
\begin{aligned}
&(\eta_m \to \eta_{kL}), (\eta_{kL} \to \eta_{(k-1)L}), \ldots, (\eta_L \to \eta_0), \\
&(\eta_1 \to \eta_{L+1}), (\eta_{L+1} \to \eta_{2L+1}), \ldots, (\eta_{(k-1)L+1} \to \eta_{kL}), \\
&\quad \vdots \\
&(\eta_{L-1} \to \eta_{2L-1}), (\eta_{2L-1} \to \eta_{3L-1}), \ldots, (\eta_{kL-1} \to \eta_m), \\
&(\eta_{kL+1} \to \eta_m), (\eta_{kL+2} \to \eta_m), \ldots, (\eta_{m-1} \to \eta_m)\}
\end{aligned}
$$

If $U(\eta_{m+1}) < U(\eta_m)$ we use the arrows

$$
\begin{aligned}
V' \;:=\; &\big\{(\eta_{m+1} \to \eta_{kL}), (\eta_{kL} \to \eta_{(k-1)L}), \ldots, (\eta_L \to x), \\
&(\eta_1 \to \eta_{L+1}), (\eta_{L+1} \to \eta_{2L+1}), \ldots, (\eta_{kL+1} \to \eta_{m+1}), \\
&\quad \vdots \\
&(\eta_{L-1} \to \eta_{2L-1}), (\eta_{2L-1} \to \eta_{3L-1}), \ldots, (\eta_{(k-1)L-1} \to \eta_{kL-1}), \\
&(\eta_{kL+1} \to \eta_{m+1}), (\eta_{kL+2} \to \eta_{m+1}), \ldots, (\eta_m \to \eta_{m+1})\}
\end{aligned}
$$

By definition of $V$ and $V'$ we have $h_1(\eta) = h_L(V)$ resp. $h_1(\eta) = h_L(V')$.
For $y = \eta_j$ for some $j \in \{1, \ldots, m+1\}$ we define $\tilde{g} := g \backslash \eta \cup V$ resp. $\tilde{g} := g \backslash \eta \cup V'$.
Then $h_{\{1,L\}}(\tilde{g}) = h_{\{1,L\}}(g)$ in both cases.
Suppose that $y \neq \eta_i$ for all $i \in \{1, \ldots, m+1\}$. If $p_{\beta(n)}(y, y) > 0$ resp. $p_{\beta(n)}(z, z) > 0$, we can choose the $L$-path $(y, \ldots, y, z)$ resp. $(y, z, \ldots, z)$.
If $p_{\beta(n)}(y, y) = p_{\beta(n)}(z, z) = 0$, we have $U(y) = U(z)$.
There exist an arrow $(z \to z') \in g$ or $z = x$.
In the first case we get $U(z) \geq U(z')$ since $p_{\beta(n)}(z, z) = 0$. Therefore we can choose the path $(y, z, y, z, \ldots, y, z, z') \in \Gamma^L_{yz',*}$ with costs 0.
If $z = x$, we choose the path $(yxyx \ldots yx\eta_1)$ with costs 0.
Define $\tilde{g} := g \backslash (\eta \cup (y \to z)) \cup V \cup (y \to z')$ resp. $\tilde{g} := g \backslash (\eta \cup (y \to z)) \cup V \cup (y \to \eta_1)$.
We also obtain $h_{\{1,L\}}(\tilde{g}) = \tilde{h}_{\{1,L\}}(g)$. By induction hypothesis we get that there exists $g' \in G_L\{x\}$ satisfying $h_L(g') \leq h_{\{1,L\}}(\tilde{g}) = h_{\{1,L\}}(g) = h_1(g)$. $\square$

The above result justifies the simplifications: $m_x \equiv m_{x,L}$, $m_{\min} \equiv m_L$ and $m_* \equiv m_{*,L}$.

## 2.2 Asymptotics of $P_{n,L}$

In the following real sequences $(a_n)_n$, $(b_n)_n$ with $b_n \neq 0$ for almost all $n$ are said to be asymptotically equivalent, if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. In this case we also write $a_n \sim b_n$.
Introduce another costs for paths $\gamma \in \Gamma^L_{yz}$:

$$
s_n(\gamma) := \sum_{i=0}^{L-1} \beta(n, i) \left(U(\gamma_{i+1}) - U(\gamma_i)\right)^+.
$$

By monotonicity of $(\beta(n,i))_{0\le i<L}$ we have

$$\beta(n)s(\gamma) \le s_n(\gamma) \le \beta(n,L-1)s(\gamma) \le \beta(n+1)s(\gamma). \tag{10}$$

For $\gamma = (x=\gamma_0,\gamma_1,\dots,\gamma_L=y) \in \Gamma_{xy}^L$ we introduce

$$\begin{aligned} J_1 = J_1(\gamma) &:= \{\, i \mid 0 \le i < L,\ \gamma_i \neq \gamma_{i+1} \,\} \\ J_2 = J_2(\gamma) &:= \{\, i \mid 0 \le i < L,\ \gamma_i = \gamma_{i+1} \,\}. \end{aligned}$$

Moreover we define

$$c_n(\gamma) := \prod_{i\in J_1} q(e_i) \prod_{i\in J_2} p_{\beta(n,i)}(e_i),$$

and obtain for $P_{n,L}(x,y)$

$$P_{n,L}(x,y) = \sum_{\gamma\in\Gamma_{xy}^L} \prod_{i=0}^{L-1} p_{\beta(n,i)}(e_i) = \sum_{\gamma\in\Gamma_{xy}^L} \exp\left(-s_n(\gamma)\right) c_n(\gamma).$$

By (3) we have $\lim_{n\to\infty} s_n(\gamma) = \beta(n)s(\gamma)$, therefore

$$P_{n,L}(x,y) \sim \sum_{\gamma\in\Gamma_{xy}^L} \exp\left(-\beta(n)s(\gamma)\right) c_n(\gamma). \tag{11}$$

Let $N$ be such that $\beta(N) > 0$. Define

$$\begin{aligned} q_{\min} &:= \min{}^+\{\, q(x,y) \mid x,y\in X\} \\ p_{\min} &:= \min{}^+\{\, p_{\beta(N)}(x,x) \mid x \in X \,\}, \end{aligned}$$

and use $p_{\beta(n,i)}(x,x) \ge p_{\beta(N)}(x,x)$, in order to get the estimate

$$1 \ge c_n(\gamma) \ge (q_{\min}p_{\min})^L > 0 \quad \text{for } n \ge N,\ \gamma \in \Gamma_{xy}^L. \tag{12}$$

Since $c_n(\gamma) \sim c(\gamma) := \lim_{n\to\infty} c_n(\gamma) > 0$, we get from (11)

$$\begin{aligned} P_{n,L}(x,y) &\sim \sum_{\gamma\in\Gamma_{xy}^L} \exp\left(-\beta(n)s(\gamma)\right) c(\gamma) \\ &\sim \sum_{\gamma\in\Gamma_{xy,*}^L} \exp\left(-\beta(n)s(\gamma_{xy})\right) c(\gamma) \\ &= \exp\left(-\beta(n)s(\gamma_{xy})\right) \underbrace{\sum_{\gamma\in\Gamma_{xy,*}^L} c(\gamma)}_{=:B(x,y)} \\ &= B(x,y)\exp\left(-\beta(n)s(\gamma_{xy})\right). \end{aligned} \tag{13}$$

## 2.3 Asymptotics of $Q_{n,L}$

Using the asymptotics of $P_{n,L}$ and defining $\widetilde{B}(x,v) := \prod_{i=1}^{p} B(x,v_i)$ we get immediately

$$Q_{n,L}(x,y) \sim \sum_{v\in\mathcal{N}_p(x,y)} \widetilde{B}(x,v)\exp(-\beta(n)\sum_{i=1}^{p} s(\gamma_{xv_i}))\sum_{l:v_l=y} F_n(l,x,v) \tag{14}$$

For $x,y\in X$ we introduce

$$\begin{aligned}
\mathcal{K}_L(x,y) &:= \left\{ v\in\mathcal{N}_p^L(x,y) \;\middle|\; \sum_{l:v_l=y} F(l,U(v))>0 \right\}\\
k_L(x,y) &:= \min\left\{ \sum_{i=1}^{p} s(\gamma_{xv_i}) \;\middle|\; v\in\mathcal{K}_L(x,y) \right\}\\
\mathcal{K}_{*,L}(x,y) &:= \left\{ v\in\mathcal{K}_L(x,y) \;\middle|\; k_L(x,y)=\sum_{i=1}^{p} s(\gamma_{xy}) \right\}.
\end{aligned}$$

By using $D_L(x,y) := \sum_{v\in\mathcal{K}_{*,L}(x,y)} \widetilde{B}(x,v)\sum_{l:v_l=y} F(l,x,v)$ and assuming $H_0$, we derive from (14)

$$\begin{aligned}
Q_{n,L}(x,y) &\sim \sum_{v\in\mathcal{N}_p(x,y)} \widetilde{B}(x,v)\exp\left(-\beta(n)\sum_{i=1}^{p} s(\gamma_{xv_i})\right)\sum_{l:v_l=y} F(l,x,v)\\
&= \sum_{v\in\mathcal{K}_L(x,y)} \widetilde{B}(x,v)\exp\left(-\beta(n)\sum_{i=1}^{p} s(\gamma_{xv_i})\right)\sum_{l:v_l=y} F(l,x,v)\\
&\sim \exp\left(-\beta(n)k_L(x,y)\right)\sum_{v\in\mathcal{K}_{*,L}(x,y)} \widetilde{B}(x,v)\sum_{l:v_l=y} F(l,x,v).\\
&\sim D_L(x,y)\exp\left(-\beta(n)k_L(x,y)\right).
\end{aligned} \tag{15}$$

## 2.4 Application of the Matrix-Tree Theorem to $Q_{n,L}$

In the sequel we show that the stationary measures $\mu_n$ corresponding to $Q_{n,L}$ are concentrated on a certain set $w_{*,L}$ for large $n$. The set $w_{*,L}$ is given by those $x\in X$, that minimize

$$x\mapsto \min\left\{ \sum_{(y\to z)\in g} k_L(y,z) \;\middle|\; g\in G_L\{x\} \right\}.$$

Using the the same notation as in [FW84] we define for $g\in G_L\{x\}$, $x\in X$

$$\begin{aligned}
\pi_n(g) &:= \prod_{(y\to z)\in g} Q_{n,L}(y,z)\\
Q_x^n &:= \sum_{g\in G_L\{x\}} \pi_n(g).
\end{aligned}$$

Applying the matrix-tree theorem yields for the stationary measure $\mu_n$ of the Markov chain $Q_{n,L}$

$$\mu_n(x) = \frac{Q_x^n}{\sum_{y \in X} Q_y^n} \text{ for } x \in X.$$

Define $K_L(g) := \prod_{(y \to z) \in g} D_L(y,z) > 0$ and use (15) in order to derive

$$\pi_n(g) \quad \sim \quad K_L(g) \exp\left( -\beta(n) \sum_{(y \to z) \in g} k_L(y,z) \right).$$

Therefore

$$Q_x^n \sim \sum_{g \in G_L\{x\}} K_L(g) \exp\left( -\beta(n) \sum_{(y \to z) \in g} k_L(y,z) \right). \tag{16}$$

Using the notations

$$\begin{aligned}
\widetilde{h}(g) = \widetilde{h}_L(g) \quad &:= \quad \sum_{(y \to z) \in g} k_L(y,z) && (17) \\
w_{x,L} \quad &:= \quad \min \left\{ \widetilde{h}(g) \mid g \in G_L\{x\} \right\} \\
w_{\min} \quad &:= \quad \min \{ w_{x,L} \mid x \in X \} \\
w_{*,L} \quad &:= \quad \{ x \mid w_{x,L} = w_{\min} \} \\
\widetilde{G}_{*,L}(x) \quad &:= \quad \left\{ g \in G_L\{x\} \mid \widetilde{h}(g) = w_{x,L} \right\} \\
\widetilde{K}_L(x) \quad &:= \quad \sum_{g \in \widetilde{G}_{*,L}\{x\}} K_L(g), && (18)
\end{aligned}$$

we conclude

$$\begin{aligned}
Q_x^n \quad &\sim \quad \sum_{g \in \widetilde{G}_{*,L}(x)} K_L(g) \exp\left(-\beta(n) w_{x,L}\right) \\
&= \quad \widetilde{K}_L(x) \exp(-\beta(n) w_{x,L}).
\end{aligned}$$

Since $\sum_{y \in X} Q_y^n \sim \sum_{y \in w_{*,L}} \widetilde{K}_L(y) \exp(-\beta(n) w_{\min})$, we derive

$$\mu_n(x) = \frac{Q_x^n}{\sum_{y \in X} Q_y^n} \sim \frac{\widetilde{K}_L(x)}{\sum_{y \in w_{*,L}} \widetilde{K}_L(y)} \exp\left(-\beta(n)(w_{x,L} - w_{\min})\right). \tag{19}$$

Finally we see that

$$\lim_{n \to \infty} \mu_n(x) = \begin{cases} 0 & \text{for} \quad x \notin w_{*,L} \\ \frac{\widetilde{K}_L(x)}{\sum_{y \in w_{*,L}} \widetilde{K}_L(y)} > 0 & \text{for} \quad x \in w_{*,L}. \end{cases}$$

## 2.5 Convergence for Chance-to-anyone

Denote by

$$\mathcal{F}_b \;:=\; \{\, F \mid \exists c > 0 \; F(l, x, y_1, \ldots, y_s) \geq c \;\forall\, (x, y_1, \ldots, y_s) \in X \times X^s \,\}$$

the set of selection functions that give a minimal chance to be selected to every individual.

**Proposition 2.5** *For any sequence $(F_n)_n$ of selection functions satisfying $\lim_{n\to\infty} F_n \in \mathcal{F}_b$ we have $w_{*,L} = X_{\min}$, $L \geq 1$.*

**Proof:** Let $g \in G_L\{x\}$. We have to show that $\widetilde{h}_L(g) = h_L(g)$.
Let $(y \to z) \in g$. By definition of $k_L(y, z)$ it holds $k_L(y, z) \geq s(\gamma_{yz})$. Since $\lim_n F_n \in \mathcal{F}_b$ we have $\mathcal{K}_L(y, z) = \mathcal{N}_p(y, z)$.
Therefore $k_L(y, z) = s(\gamma_{yz})$ and $\widetilde{h}_L(g) = h_L(g)$. Hence $w_{*,L} = m_* = X_{\min}$. □

## 2.6 Convergence for Worst-Wins

For $x \in X$, $y \in N_1(x)$ we say that $H_1(x, y)$ is satisfied, if $y \in N_L(x)$ and for $\gamma^L_{xy} \in \Gamma^L_{xy,*}$ the equality $s(\gamma^L_{xy}) = (U(y) - U(x))^+$ holds. Hypothesis $H_1(x, y)$ demands that the path $(x, y)$ of length 1 can be replaced by a path of length $L$ that is not more expensive.
We say that $H_1$ holds, if for all $x \in X$, $y \in N_1(x)$ $H_1(x, y)$ holds.
Hypothesis $H_1$ is already satisfied if one of the following assumptions is fulfilled:

1. $L$ odd.

2. $q(x, x) > 0$ for all $x \in \widehat{X}$, where

   $$\widehat{X} := \{\, x \in X \mid y \in N_1(x) \;\Rightarrow\; U(y) \leq U(x) \,\}$$

   denotes the set of local maxima.

For a finite subset $A \subset \mathbb{N}_1$, we define

$$\widetilde{h}_A(g) := \sum_{(y\to z)\in g} \min_{a\in A} k_a(y, z), \quad g \in G_A\{x\}.$$

If $\mathcal{K}_a(y, z) = \emptyset$ we let $k_a(y, z) = \infty$. Note that the definition of $\widetilde{h}_A(g)$ coincides with (17) for $A$ consisting of one element.
Under the assumption of $H_1$ we get immediately:

**Lemma 2.6** *Using the Worst-Wins strategy we have for any $x \in X$ and $g \in G_{\{1,L\}}\{x\}$ the equality $\widetilde{h}_L(g) = \widetilde{h}_{\{1,L\}}(g)$.*

For $L = 1$ we have the following result.

**Lemma 2.7** *Let $g \in G_1\{x\}$ for $x \in X$.*
*Then $\widetilde{h}_1(g) = h_1(g)$.*

**Proof:** Let $(y \to z) \in g \in G_1\{x\}$. Since $g$ is a 1-graph, we have $s(\gamma_{yz}) = (U(z) - U(y))^+$.
For $U(y) < U(z)$ we have $k_1(y, z) = (U(z) - U(y))^+$, beacuse the $p$-tuple $(z, y, \ldots, y)$ is contained in $\mathcal{K}_1(y, z)$.
For $U(y) \geq U(z)$ we have $k_1(y, z) = (U(z) - U(y))^+ = 0$, because the $p$-tuple $(z, \ldots, z)$ is contained in $\mathcal{K}_1(y, z)$.
Therefore

$$\widetilde{h}_1(g) = \sum_{(y \to z) \in g} k_1(y, z) = \sum_{(y \to z) \in g} (U(z) - U(y))^+ = h_1(g). \quad \square$$

Combining the above results, we get

**Proposition 2.8** *Let $F$ be as in case 1 and $L \geq 1$.*
*Then $w_{*,L} = X_{\min}$.*
*This implies $\lim_n \mu_n(X_{\min}) = 1$.*

**Proof:** We have the following inequality

$$\begin{aligned}
\min\left\{ \widetilde{h}_L(g) \mid g \in G_L\{x\} \right\} &= \min\left\{ \widetilde{h}_{\{1,L\}}(g) \mid g \in G_{\{1,L\}}\{x\} \right\} && (20) \\
&\leq \min\left\{ \widetilde{h}_1(g) \mid g \in G_1\{x\} \right\} \\
&= \min\left\{ h_1(g) \mid g \in G_1\{x\} \right\} && (21) \\
&= \min\left\{ h_L(g) \mid g \in G_L\{x\} \right\} && (22)
\end{aligned}$$

Here we have used Lemma 2.6 in (20), Lemma 2.7 in (21) and Lemma 2.4 in (22).
Using $\widetilde{h}_L(g) \geq h_L(g)$ we conclude

$$\min\left\{ \widetilde{h}_L(g) \mid g \in G_L\{x\} \right\} = \min\left\{ h_L(g) \mid g \in G_L\{x\} \right\}.$$

This implies $w_{x,L} = m_{x,L}$, $w_{\min} = m_{\min}$ and $w_{*,L} = m_*$.
Hence $w_{*,L} = m_* = X_{\min}$ by Lemma 2.3. $\square$

## 2.7 In general no convergence for Best-Wins

For $L = 1$ the convergence of the homogeneous algorithm is still true:

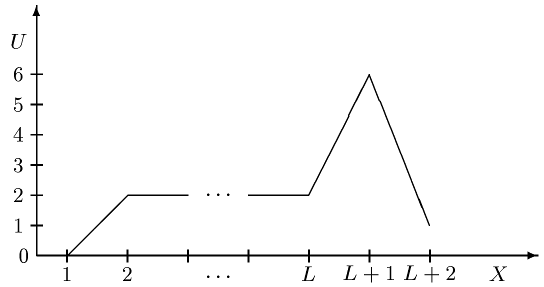
**Proposition 2.9** *Let $g \in G_1\{x\}$ for $x \in X$.*
*Then $\widetilde{h}_1(g) = ph_1(g)$ implying $w_{*,1} = m_{*,1} = X_{\min}$.*

**Proof:** Let $(y \to z) \in g \in G_1\{x\}$.
For $U(y) < U(z)$ we have $k_1(y,z) = p(U(z) - U(y))^+$ and for $U(y) \geq U(z)$, we get $k_1(y,z) = (U(z) - U(y))^+ = 0$.
Therefore

$$\widetilde{h}_1(g) = \sum_{(y \to z) \in g} k_1(y,z) = p \sum_{(y \to z) \in g} (U(z) - U(y))^+ = p h_1(g). \quad \square$$

For $L \geq 2$ the inclusion $w_{*,L} \subseteq X_{\min}$ is in general false:

**Example 2.10** *For the state space $X := \{1, \dots, L+2\}$ consider the subsequent energy landscape $(X, U)$.*




*The exploration kernel $q$ is given by nearest-neighbor random walk with uniform probability, e.g.*

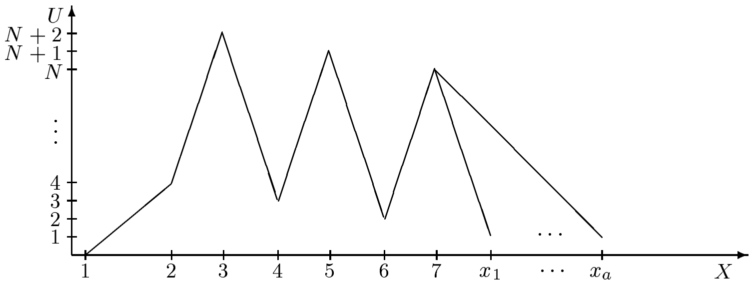
$$\begin{aligned} q(i,i+1) &= q(i,i-1) = \frac{1}{2} \quad \textit{for } i \in \{2, \dots, L+1\} \\ q(1,2) &= q(L+2, L+1) = 1. \end{aligned}$$

*If a moving particle is in the local minimum $L+2$, then after $L$ transitions it can only be in states that are worse except returning to $L+2$ again.*
*Therefore leaving the state $L+2$ is very expensive. We have $w_{L+2,L} = 2p+4$ and $w_{1,L} = 5p$. Thus for all $p \geq 2$*

$$\lim_{n \to \infty} \mu_n(X_{\min}) = \lim_{n \to \infty} \mu_n(1) = 0.$$

In the above example we have proposed an energy landscape $(X, U)$, for which $w_{*,L} \not\subset X_{\min}$ for any $p \geq 2$.
If we fix the above landscape and increase the simulation length $L$ of the parallel Markov chains, the inclusion $w_{*,L} \subseteq X_{\min}$ start to hold again.
Therefore one may conjecture that $w_{*,L} \subset X_{\min}$ can be established for large $L$. The next example shows that this does not suffice: In general one has to increase the number of parallel Markov chains $p$ as well. Compare the following example.

**Example 2.11** *Let $L \geq 7$ and $p \geq 3$. For $a \geq 1$ let the state space be given by $X := \{1, \dots, 7, x_1, \dots, x_a\}$. Consider the following energy landscape for $N \geq 7$*




*The exploration kernel $q$ is again given by nearest neighbor random walk with uniform probability.*
*A best $\{x_j\}$-L-graph, $j \in \{1,\ldots,a\}$ is the following*

$$\begin{aligned} g &= (X, \{(1 \to 2), (2 \to 4), (3 \to 2), (4 \to 6), (5 \to 4), (6 \to x_j), (7 \to 6), \\ &\quad (x_1 \to x_j), \ldots, (x_{j-1} \to x_j), (x_{j+1} \to x_j)(x_a \to x_j)\}) \end{aligned}$$

*with costs $\widetilde{h}(g) = 3(N-2) + 4p + (a-1)(N-1)$. A best $\{1\}$-L-graph is given by*

$$\begin{aligned} g' &= (X, \{(x_1 \to 1), (2 \to 1), (3 \to 2), (4 \to 6), (5 \to 4), (6 \to x_1), (7 \to 6) \\ &\quad (x_2 \to x_1), \ldots, (x_a \to x_1)\}) \end{aligned}$$

*with costs $\widetilde{h}(g') = (a+2)(N-1) + 2(N-2)$.*
*Hence the local minima $x_1, \ldots, x_a$ are contained in $w_{*,L}$ for $L \geq 7$ and all $3 \leq p < \frac{2N-1}{4}$.*

For $p$ and $L$ large we can establish $w_{*,L} \subset X_{\min}$.

**Proposition 2.12** *Set*

$$\delta := \min{}^{+} \{ (U(x) - U(y)) \mid x \in N(z),\ z \in X_{\min} \}$$

$$\theta := \max_{x \in X \setminus X_{\min}} \min \left\{ s(\gamma^l) \mid l \in \mathbb{N},\ z \in X_{\min},\ \gamma^l \in \Gamma^l_{xz} \right\}.$$

*Suppose that $p$ satisfies $p\delta > \theta$. Moreover let $L$ be chosen such that for all $x \in X$ holds: There exists $z \in X_{\min}$ and $\gamma^L \in \Gamma^L_{xz}$ such that $s(\gamma^L) \leq \theta$.*
*Then $w_{*,L} \subset X_{\min}$.*

**Proof:** Let $x \in X \setminus X_{\min}$ and $g \in G_{L,*}\{x\}$.
Without loss of generality we may assume that $g$ satisfies $(y \to z) \in g$ with $U(y) < U(z) \quad \Rightarrow z \in N_1(y)$.
By choice of $L$ there exist $z \in X_{\min}$ such that $s(\gamma_{zx}) \leq \theta$.
Let $\eta = (\eta_0, \ldots, \eta_{l(\eta)}) \subset g$ be the path from $z$ to $x$ and

$$V := \{ i \in \{0, \ldots, l(\eta) - 1\} \mid U(\eta_i) > U(\eta_{i+1}) \}$$

be the set of indices, that correspond to a decreasing arrow.

Let

$$j := \begin{cases} \min\{\, V \,\} & : \quad V \neq \emptyset \\ l(\eta) - 1 & : \quad V = \emptyset. \end{cases}$$

For $\eta_1 := \bigcup_{i=0}^{j}(\eta_i \to \eta_{i+1})$ define the $\{z\}$-$L$-graph

$$g' := g \setminus \eta_1 \cup \eta_1^{-1} \cup (x \to z).$$

The costs of $g'$ can be estimated by

$$\begin{aligned} \widetilde{h}_L(g') &= \widetilde{h}_L(g) - \widetilde{h}_L(\eta_1) + \widetilde{h}_L(\eta_1^{-1}) + s(\gamma_{xz}) \\ &\leq w_{x,L} - p\delta + \theta \\ &< w_{x,L} \quad \Box \end{aligned}$$

## 2.8 Convergence for First-Wins

**Proposition 2.13** *Using the First-Wins strategy we have $w_{*,L} = X_{\min}$.*

**Proof:** Let $(y \to z) \in g \in \Gamma_{*,L}$. The $p$-tuple $v = (z, v_2, \ldots, v_p)$ is contained in $\mathcal{K}_L(y,z)$, because $F(1,z,v) = 1$. Since $v_2, \ldots v_p$ can be chosen such that $s(\gamma_{yv_i}) = 0$ we have $k_L(y,z) = s(\gamma_{yz})$. Hence $\widetilde{h}_L(g) = h_L(g)$ and $w_{*,L} = m_* = X_{\min}$. $\Box$

We have proved that for any probability measure $\mu$ on $X$ the convergence

$$\lim_{n\to\infty} \lim_{m\to\infty} \mu Q_{n,L}^m(X_{\min}) = 1 \tag{23}$$

holds when using the Worst-Wins, First-Wins or Chance-to-anyone strategy. For the Best-Wins strategy (23) is only true for sufficiently large $p$ and $L$.
We will proceed similar to [C94] and give estimates for the transition probabilities $P_x(X_{n+k} = z \mid X_n = y)$. From these estimates we receive conditions on the sequence $\beta(n)$ for the probability of staying in a certain subset $A \subset X$.

# 3 Convergence of the inhomogeneous algorithm

## 3.1 Necessary condition for convergence

Let $k \geq 1$ such that each state has a connection to any other state in $kL$ steps. We need a lower estimate for the transition probability $P_x(X_{n+1} = z \mid X_n = y)$ of the chain $(X_n)_n$.

$$\begin{aligned} &P_x(X_{n+1} = z \mid X_n = y) \\ &= \sum_{v\in\mathcal{N}_{r(n)}(y,z)} \left[\prod_{i=1}^{r(n)} \sum_{\gamma\in\Gamma_{yv_i}^L} \exp\left(-s_n(\gamma)\right) c_n(\gamma)\right] \sum_{l:v_l=z} F_n(l,y,v) \end{aligned} \tag{24}$$

Since $(z,\ldots,z) \in \mathcal{N}_{r(n)}(y,z)$ for $P_{n,L}(y,z) > 0$, it follows using (10)

$$
\begin{aligned}
& P_x(X_{n+1} = z \mid X_n = y) \\
\geq\ & \left[\prod_{i=1}^{r(n)} \sum_{\gamma\in\Gamma^L_{yz}} \exp\left(-\beta(n+1)s(\gamma)\right) c_n(\gamma)\right] \underbrace{\sum_{l=1}^{r(n)} F_n(l,y,z,\ldots,z)}_{=1} \\
\geq\ & \prod_{i=1}^{r(n)} \left[ \sum_{\substack{\gamma\,\in\,\Gamma^L_{yz} \\ s(\gamma)\,=\,s(\gamma_{yz})}} \exp\left(-\beta(n+1)s(\gamma_{yz})\right) c_n(\gamma) \right] \\
=\ & \exp(-\beta(n+1)r(n)s(\gamma_{yz})) \prod_{i=1}^{r(n)} \sum_{\substack{\gamma\,\in\,\Gamma^L_{yz} \\ s(\gamma)\,=\,s(\gamma_{yz})}} c_n(\gamma)
\end{aligned}
$$

There exists a constant $c_1 > 0$ (compare (12) on page 11) such that

$$
P_x(X_{n+1} = z \mid X_n = y) \geq c_1 \exp\left(-\beta(n+1)r(n)s(\gamma_{yz})\right). \tag{25}
$$

For $y \in X$ and $z \in N_{(kL)}(y)$ we define

$$
\begin{aligned}
V(y,z)^k &:= \min\left\{ \sum_{i=0}^{k-1} s(\gamma_{x_i x_{i+1}}) \;\middle|\; x_0 = y,\ x_k = z,\ Q_{n,L}(x_i,x_{i+1}) > 0 \right\} \\
V_*(y,z)^k &:= \left\{ (x_1,\ldots,x_{k-1}) \;\middle|\; \sum_{i=0}^{k-1} s(\gamma_{x_i x_{i+1}}) = V(y,z)^k\ x_0 = y,\ x_k = z,\ Q_{n,L}(x_i,x_{i+1}) > 0 \right\}.
\end{aligned}
$$

From (25) we derive the following lower bound for the $k$-step transition probability of $(X_n)_n$

$$
\begin{aligned}
& P(X_{n+k} = z \mid X_n = y) \\
=\ & \sum_{x_1,\ldots,x_{k-1}} P(X_{n+1} = x_1 \mid X_n = y) \prod_{i=1}^{k-2} P(X_{n+i+1} = x_{i+1} \mid X_{n+i} = x_i) \\
& \quad \cdot P(X_{n+k} = z \mid X_{n+k-1} = x_{k-1}) \\
\geq\ & c_1^k \exp\left(-\beta(n+k)r(n)V(y,z)^k\right) |V_*(y,z)^k| \\
\geq\ & C_1 \exp\left(-\beta(n+k)r(n)V(y,z)^k\right)
\end{aligned}
$$

where $C_1 := c_1^k$ is independent of $y$ and $z$.
Do the same in order to get the upper bound

$$
\begin{aligned}
& P(X_{n+1} = z \mid X_n = y) \\
=\ & \sum_{v\in\mathcal{N}_{r(n)}(y,z)} \prod_{i=1}^{r(n)} \sum_{\gamma\in\Gamma^L_{yv_i}} \exp\left(- \underbrace{s_n(\gamma)}_{\geq\beta(n)s(\gamma)\geq\beta(n)s(\gamma_{yv_i})}\right) c_n(\gamma) \sum_{l:v_l=z} F_n(l,y,v)
\end{aligned}
$$

$$\leq \sum_{v\in\mathcal{N}_{r(n)}(y,z)} \exp\left(-\beta(n)\underbrace{\sum_{i=1}^{r(n)} s(\gamma_{yv_i})}_{\geq s(\gamma_{yz})}\right)\prod_{i=1}^{r(n)}\sum_{\gamma\in\Gamma^L_{yv_i}} c_n(\gamma)\underbrace{\sum_{l:v_l=z} F_n(l,y,v)}_{\leq 1}$$

$$\leq \exp\left(-\beta(n)s(\gamma_{yz})\right)\sum_{v\in\mathcal{N}_{r(n)}(y,z)}\prod_{i=1}^{r(n)}\sum_{\gamma\in\Gamma^L_{yv_i}} c_n(\gamma).$$

Hence there exists a constant $0 < c_2 \leq \sup\left\{\sum_{v\in\mathcal{N}_{r(n)}(y,z)}\prod_{i=1}^{r(n)}|\Gamma^L_{yv_i}| \mid y,z\in X\ n\geq 0\right\} < \infty$, such that

$$P(X_{n+1}=z \mid X_n=y) \leq c_2\exp(-\beta(n)s(\gamma_{yz})) \text{ for all } y,z\in X.$$

This yields the following upper bound for the $k$-step transition probability of $(X_n)_n$

$$P(X_{n+k}=z \mid X_n=y) \leq C_2\exp(-\beta(n)V(y,z)^k).$$

Summarizing the estimates we have

$$C_1\exp(-\beta(n+k)r(n)V(y,z)^k) \leq P_x(X_{n+k}=z \mid X_n=y) \leq C_2\exp(-\beta(n)V(y,z)^k). \tag{26}$$

The next Proposition leads to a necessary condition for convergence.

**Proposition 3.1** *Let $A\subset X$. Set*

$$\begin{aligned}\psi_A &:= \max_{y\in A}\min_{z\in X\setminus A} V(y,z)^k\\ \phi_A &:= \min_{y\in A}\min_{z\in X\setminus A} V(y,z)^k\end{aligned}$$

*a) If $\sum_{n=0}^{\infty}\exp(-\beta(n)p\psi_A)=\infty$, we have*

$$P(\exists N\ \forall n\geq N\ X_n\in A \mid X_0=x)=0.$$

*b) If $\sum_{n=0}^{\infty}\exp(-\beta(n)\phi_A)<\infty$, we have*

$$P(\exists N\ \forall n\geq N\ X_{N+nk}\in A \mid X_0=x)>0.$$

**Proof:** Summing up the inequality (26) over $X\setminus A$ yields

$$\begin{aligned}&C_1\sum_{z\in X\setminus A}\exp\left(-\beta(n+k)r(n)V(y,z)^k\right)\\ \leq\ & P_x(X_{n+k}\notin A \mid X_n=y)\\ \leq\ & C_2\sum_{z\in X\setminus A}\exp\left(-\beta(n)V(y,z)^k\right).\end{aligned}$$

Let $\theta(y) := \min\{ V(y,z)^k \mid z \in X \setminus A\}$. Then for $y \in A$

$$\begin{aligned}
P_x(X_{n+k} \notin A \mid X_n = y) &\geq C_1 \sum_{\substack{z \in X \setminus A \\ V(y,z)^k = \theta(y)}} \exp\left(-\beta(n+k)r(n)\theta(y)\right) \\
&= C_1 \underbrace{\#\{z \in X \setminus A \mid V(y,z)^k = \theta(y)\}}_{\geq 1} \exp\left(-\beta(n+k)r(n)\theta(y)\right) \\
&\geq C_1' \exp\left(-\beta(n+k)r(n)\theta(y)\right),
\end{aligned}$$

where

$$C_1' := C_1 \cdot \min\Big\{\#\{z \in X \setminus A \mid V(y,z)^k = \theta(y)\} \mid y \in A\Big\}.$$

Let $C_2' := C_2 \cdot |(X \setminus A)|$ and derive the following upper estimate

$$\begin{aligned}
P_x(X_{n+k} \notin A \mid X_n = y) &\leq C_2 \sum_{z \in X \setminus A} \exp\left(-\beta(n) \underbrace{V(y,z)^k}_{\geq \theta(y)}\right) \\
&= C_2' \exp\left(-\beta(n)\theta(y)\right).
\end{aligned}$$

Summarizing

$$C_1' \exp\left(-\beta(n+k)r(n)\theta(y)\right) \leq P_x(X_{n+k} \notin A \mid X_n = y) \leq C_2' \exp\left(-\beta(n)\theta(y)\right).$$

By definition of $\psi_A$ and $\phi_A$ we get

$$C_1' \exp(-\beta(n+k)r(n)\psi_A) \leq P_x(X_{n+k} \notin A \mid X_n = y) \leq C_2' \exp(-\beta(n)\phi_A)$$

and further

$$1 - C_2' \exp(-\beta(n)\phi_A) \leq P_x(X_{n+k} \in A \mid X_n = y) \leq 1 - C_1' \exp(-\beta(n+k)r(n)\psi_A). \tag{27}$$

For $M \geq 0$ we have

$$\begin{aligned}
& P_x(i \in \{0, \ldots, M\},\ X_{N+ik} \in A) \qquad (28) \\
= & \sum_{x_0,\ldots,x_M \in A} P_x(X_{N+Mk} = x_M, \ldots, X_N = x_0) \\
= & \sum_{x_0,\ldots,x_M \in A} P_x(X_N = x_0) \cdot \prod_{i=0}^{M-1} P_x(X_{N+(i+1)k} = x_{i+1} \mid x_{N+ik} = x_i).
\end{aligned}$$

Using (28), (27) and

$$\sum_{x_M \in A} P_x(X_{N+Mk} = x_M \mid X_{N+(M-1)k} = x_{M-1}) = P_x(X_{N+Mk} \in A \mid X_{N+(M-1)k} = x_{M-1})$$

we derive the inequality

$$\begin{aligned}
& (1-C_2'\beta(N+(M-1)k)\phi_A))\, P_x(i\in\{0,\ldots,M-1\},\ X_{N+ik}\in A) \\
\leq\ & P_x(i\in\{0,\ldots,M\},\ X_{N+ik}\in A) \qquad (29)\\
\leq\ & (1-C_1'\exp(-\beta(N+Mk)r(N+(M-1)k)\psi_A))) \\
& \cdot P_x(i\in\{0,\ldots,M-1\},\ X_{N+ik}\in A).
\end{aligned}$$

By induction we obtain from (29)

$$\begin{aligned}
& P_x(X_N\in A)\prod_{i=0}^{M-1}(1-C_2'\beta(N+ik)\phi_A)) \\
\leq\ & P_x(i\in\{0,\ldots,M\},\ X_{N+ik}\in A) \qquad (30)\\
\leq\ & P_x(X_N\in A)\prod_{i=0}^{M-1}(1-C_1'\exp(-\beta(N+(i+1)k)r(N+ik)\psi_A)))\,.
\end{aligned}$$

The divergence $\sum_{i=0}^{\infty}\exp(-\beta(i)p\psi_A)=\infty$ implies

$$\sum_{i=0}^{\infty}\exp(-\beta(N+ik)p\psi_A)=\infty \quad \text{for every } N\in\mathbb{N}. \qquad (31)$$

Since $r(n)=p$ for $n$ large, (31) implies

$$\sum_{i=0}^{\infty}\exp(-\beta(N+(i+1)k)r(N+ik)\psi_A)=\infty.$$

Therefore the product

$$\prod_{i=0}^{M-1}(1-C_1'\exp(-\beta(N+(i+1)k)r(N+ik)\psi_A)))$$

tends to zero for $M\to\infty$.
Hence (30) implies for any $N\in\mathbb{N}$

$$P_x(\forall n\geq 0\ X_{N+nk}\in A)=0.$$

Part a) of the Proposition follows from

$$P_x(\exists N\ \forall n\geq N\ X_n\in A)\leq\sum_{N=0}^{\infty}P_x(\forall n\geq N\ X_n\in A)=0.$$

For part b) note that $\phi_A<0$ because $\sum_{n=0}^{\infty}\exp(-\beta(n)\phi_A)<\infty$.
Hence there exist $N_0$, such that $P_x(X_{N_0}\in A)>0$ and $C_2'\exp(-\beta(n)\phi_A)<1$ for all $n\geq N_0$ holds.

The assumption $\sum_{n=N}^{\infty} C_2' \exp(-\beta(N_0+nk)\phi_A) < \infty$ implies

$$\prod_{n=0}^{\infty} (1 - C_2' \exp(-\beta(N_0+nk)\phi_A))) > 0.$$

Using the first inequality in (30) yields

$$\begin{aligned} 0 &< P_x(X_{N_0} \in A) \prod_{n=0}^{\infty} (1 - C_2' \exp(-\beta(N_0+nk)\phi_A)) \\ &\leq P_x(\forall n \geq 0,\ X_{N_0+nk} \in A) \\ &\leq P_x(\exists N\ \forall n \geq N\ X_{N+nk} \in A) \quad \square \end{aligned}$$

The next corollary formulates a necessary condition for the convergence of the PC algorithm using an arbitrary sequence of selection functions that satisfy the hypothesis $H_0$. If one takes a logarithmic sequence one has to choose $\beta(n) = c\ln(n+1)$ with $c \leq \max_{A\subset X\setminus X_{\min}} \min_{y\in A} \min_{z\in X\setminus A} V(y,z)^k$.

**Corollary 3.2** *Let* $\theta := \max\{\, \phi_A \mid A \subseteq X \setminus X_{\min} \,\}$*. If* $\limsup_{n\to\infty} P_x(X_n \in X_{\min}) = 1$*, then* $\sum_{n=0}^{\infty} \exp(-\beta(n)\theta) = \infty$*.*

**Proof:** Suppose that $\sum_n \exp(-\beta(n)\phi_A) < \infty$ for a subset $A \subseteq X \setminus X_{\min}$ with $\theta = \phi_A$. Apply part b) of Proposition 3.1 in order to get $P_x(\exists N\ \forall n \geq N\ X_n \in A) > 0$. Hence

$$\varepsilon := P_x(\exists N\ \forall n \geq N\ X_n \in X_{\min}^c) \geq P_x(\exists N\ \forall n \geq N\ X_n \in A) > 0.$$

On the other hand we have

$$\begin{aligned} 1-\varepsilon &= P_x(\forall N\ \exists n \geq N\ X_n \in X_{\min}) \\ &= \lim_{N\to\infty} P_x(\exists n \geq N\ X_n \in X_{\min}) \end{aligned}$$

Hence we obtain the contradiction

$$\begin{aligned} \limsup_N P_x(X_N \in X_{\min}) &\leq \limsup_N P_x\left(\bigcup_{n\geq N} \{X_n \in X_{\min}\}\right) \\ &= \limsup_N P_x(\exists n \geq N\ X_n \in X_{\min}) \\ &= 1-\varepsilon \quad \square \end{aligned}$$

## 3.2 Weak ergodicity of $(X_n)_n$ via coupling

Recall that an inhomogeneous Markov chain $(X_n)_n$ with transition kernels $Q_n$ is called weakly ergodic, if for all initial distributions $\mu$, $\nu$ on $X$ and all $m \geq 0$

$$\lim_{n\to\infty} \|(\mu-\nu)Q_m Q_{m+1} \dots Q_{m+n}\| = 0,$$

where $\|\cdot\|$ denotes the norm of the total variation.
An inhomogeneous Markov chain $(X_n)_n$ is called strongly ergodic, if there exist a probability measure $\nu$ on $X$ such that for all $m \geq 0$

$$\lim_{n\to\infty} \sup_{\mu} \|\mu Q_m Q_{m+1} \dots Q_{m+n} - \nu\| = 0.$$

Remind that the weak ergodicity of a Markov chain is equivalent to the existence of a successful coupling. Compare [T93], [GR78]. We will give sufficient conditions for the classical coupling for being successful. Let $\widetilde{\Omega} := \Omega \times \Omega$, where $\Omega = X^{\mathbf{N}}$. Let $\widetilde{P}_{\bar{x}_0}$ be the measure induced by the kernels

$$\begin{aligned} \widetilde{Q}_n((x_1,x_2),(y_1,y_2)) &:= Q_n(x_1,y_1)\cdot Q_n(x_2,y_2) \text{ for } y_1 \neq y_2 \\ \widetilde{Q}_n((x,x),(y,y)) &:= Q_n(x,y). \end{aligned}$$

For $\bar{y}, \bar{z} \in X \times X$ define

$$\begin{aligned} \widetilde{V}(\bar{y},\bar{z}) &:= V(y_1,z_1) + V(y_2,z_2) \\ \widetilde{V}_*(\bar{y},\bar{z})^k &:= \left\{ (\bar{x}_1,\dots,\bar{x}_{k-1}) \mid \sum_{j=0}^{k-1} s(\gamma_{x_j^i x_{j+1}^i}) = V(y_i,z_i) \text{ for } i \in \{1,2\} \right\} \\ s(\gamma_{\bar{y}\bar{z}}) &:= s(\gamma_{y_1,z_1} \cup \gamma_{y_2 z_2}) = s(\gamma_{y_1 z_1}) + s(\gamma_{y_2 z_2}). \end{aligned}$$

The lower estimate (25) for $Q_n(y,z)$ yields for the $k$-step transition probability

$$\begin{aligned} &\widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = \bar{z} \mid \bar{X}_n = \bar{y}) \qquad (32) \\ \geq &\sum_{\bar{x}_1,\dots,\bar{x}_{k-1} \in X^2\setminus\Delta} \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+1} = \bar{x}_1 \mid \bar{X}_n = \bar{y}) \prod_{i=1}^{k-2} \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+i+1} = \bar{x}_{i+1} \mid \bar{X}_{n+i} = \bar{x}_i) \\ &\quad \cdot \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = \bar{z} \mid \bar{X}_{n+k-1} = \bar{x}_{k-1}) \\ \geq &\ \widetilde{C}_1 \exp\left(-\beta(n+k) r(n) \tilde{V}(\bar{y},\bar{z})^k\right) \end{aligned}$$

where $\widetilde{C}_1 := c_1^{2k}$.
Denote by $\Delta = \{\, (x,x) \mid x \in X \,\}$ the diagonal in $X \times X$.

**Proposition 3.3** *Let $A \subseteq X \times X$. Set*

$$\psi_A := \max_{\bar{y}\in A} \min_{\bar{z}\in A^c} \widetilde{V}(\bar{y},\bar{z}).$$

*Denote by $\widetilde{P}_{\bar{x}_0}$ the classical coupling starting in $\bar{x}_0 \in X^2$.*
*i) Let $F$ be any selection function.*
*For $\sum_{n=0}^{\infty} \exp(-\beta(n) p \psi_A) = \infty$ we have*

$$\widetilde{P}_{\bar{x}_0}(\exists N \ \forall n \geq N \ \bar{X}_n \in A) = 0.$$

*ii) Let $F \in \mathcal{F}_b$ or $F$ as in case 1.*
*For $\sum_{n=0}^{\infty} \exp(-\beta(n)\psi_A) = \infty$ we have*

$$\widetilde{P}_{\bar{x}_0}(\exists N \ \forall n \geq N \ \bar{X}_n \in A) = 0.$$

*For Worst-Wins the assertion holds for the special set $A = X \times X \setminus \Delta$.*

**Proof:** Conclude in the same way as in Proposition (3.1) by using the inequality

$$\widetilde{C}_1 \exp\left(-\beta(n+k)r(n)\widetilde{V}(\bar{y},\bar{z})^k\right) \leq \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = \bar{z} \mid \bar{X}_n = \bar{y})$$

in case i), resp.

$$\widetilde{C}_1 \exp\left(-\beta(n+k)\widetilde{V}(\bar{y},\bar{z})^k\right) \leq \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = \bar{z} \mid \bar{X}_n = \bar{y}) \tag{33}$$

in case ii). The lower estimate (33) is obtained in the following manner:
The inequality

$$P_x(X_{n+1} = z \mid X_n = y) \geq const \cdot \exp(-\beta(n+1)s(\gamma_{yz})) \text{ for } y \neq z \tag{34}$$

holds for $F \in \mathcal{F}_b$ or $F$ being the First-Wins strategy. This follows for $F \in \mathcal{F}_b$ from (24) by using that $F_n(l,y,v) \geq c > 0$ for all $v \in \mathcal{N}_{r(n)}(y,z)$. In the case of the First-Wins strategy use the invariance of $\mathcal{N}_{r(n)}(y,z)$ under permutations for $z \neq y$.
Using (34) in (32) yields (33).
In the case of the Worst-Wins strategy the estimate (33) can be obtained for $L = 1$ as above by using (24) and (32). But for $L \geq 2$ an estimate of the form (33) does not hold any more such that we we have to proceed differently.
Let $k \in \mathbb{N}$ be chosen such that for all $y, z \in X$ the costs of the path $\gamma_{yz}^k \in \Gamma_{yz}^k$ are minimal among all costs of paths of arbitrary lengths, e.g.

$$s(\gamma_{yz}^k) = \min\left\{ s(\gamma_{yz}^l) \ \mid \ \gamma_{yz}^l \in \Gamma_{yz}^L, \ l \in \mathbb{N} \right\}.$$

Let $y, z \in X$ and $\gamma = \gamma_{yz}^k$. Using $H_1$, for the $k$-step transition probability of $(X_n)_n$ holds

$$\begin{aligned}
P_x(X_{n+k} = z \mid X_n = y) &\geq \prod_{i=0}^{k-1} P_x(X_{n+i+1} = \gamma(i+1) \mid X_{n+i} = \gamma(i)) \\
&\geq const \cdot \exp\left(-\beta(n) \sum_{i=0}^{k} s(\gamma_{\gamma(i)\gamma(i+1)})\right) \\
&= const \cdot \exp(-\beta(n)V(y,z)^k).
\end{aligned} \tag{35}$$

The equality (35) is true by minimality of the costs of $\gamma$ and by the fact that there exists $\gamma_{\gamma(i)\gamma(i+1)}^L$ such that $(U(\gamma(i+1)) - U(\gamma(i)))^+ = s(\gamma_{\gamma(i)\gamma(i+1)}^L)$.

We maintain that

$$\widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = (z,z) \mid \bar{X} = \bar{y}) \geq \widetilde{C}_1 \exp(-\beta(n+k)\widetilde{V}(\bar{y},(z,z))^k). \tag{36}$$

Let $\gamma$ resp. $\eta$ be paths of lengths $k$ from $y_1$ to $z$ resp. from $y_2$ to $z$ with $y_1 \neq y_2$. If $\gamma(i) \neq \eta(i)$ for all $i \in \{0, \ldots, k-1\}$ the assertion follows from (32), since the $(k-1)$-tuple

$$(((\gamma(1),\eta(1)),\ldots,(\gamma(k-1),\eta(k-1))) \in (X^2)^{k-1}$$

occurs as summand in (32). Then (35) yields (36).
Otherwise let $i \in \{1,\ldots,k-1\}$ be minimal such that $\gamma(i) = \eta(i)$. Assume that the paths $\gamma$ and $\eta$ coincide after their first meeting, e.g. $\gamma(j) = \eta(j)$ for $j \geq i$. We conclude

$$\begin{aligned}
& \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+k} = (z,z) \mid \bar{X}_n = \bar{y}) \\
\geq\ & \prod_{j=0}^{i-1} \widetilde{P}_{\bar{x}_0}(\bar{X}_{n+j+1} = (\gamma(j+1),\eta(j+1)) \mid \bar{X}_{n+j} = (\gamma(j),\eta(j))) \\
& \cdot \prod_{j=i}^{k-1} P_{x_0^1}(X_{n+j+1} = \gamma(j+1) \mid X_{n+j} = \gamma(j)) \\
\geq\ & \prod_{j=0}^{k-1} P_{x_0^1}(X_{n+j+1} = \gamma(j+1) \mid X_{n+j} = \gamma(j)) \\
& \cdot \prod_{j=0}^{k-1} P_{x_0^2}(X_{n+j+1} = \eta(j+1) \mid X_{n+j} = \eta(j)) \\
\geq\ & \widetilde{C}_1 \exp(-\beta(n+k)\widetilde{V}(\bar{y},(z,z))^k),
\end{aligned}$$

where $\widetilde{C}_1$ is a strict positive constant and (37) follows from the minimality of $\gamma \in \Gamma^k_{y_1 z}$ and $\eta \in \Gamma^k_{y_2 z}$. Next, we use this inequality and follow the proof of Proposition 3.1 for $A = X \times X \setminus \Delta$. □

Proposition 3.3 yields a sufficient condition for the weak ergodicity of $(X_n)_n$.

**Proposition 3.4** *Let $\theta$ be given by*

$$\theta_\Delta := \psi_{X\times X\setminus\Delta} = \max_{y_1\neq y_2} \min_{(z,z)\in\Delta} \widetilde{V}((y_1,y_2),(z,z))^k.$$

*Moreover assume that*

$$\sum_{n=0}^{\infty} \exp(-\beta(n)p\theta_\Delta) = \infty. \tag{37}$$

*Then $(X_n)_n$ is weakly ergodic.*
*For $F \in \mathcal{F}_b$ or $F$ being the Worst-Wins or First-Wins strategy, the chain $(X_n)_n$ is weakly ergodic if*

$$\sum_{n=0}^{\infty} \exp(-\beta(n)\theta_\Delta) = \infty. \tag{38}$$

**Proof:** The divergence (37) resp. (38) implies by using Proposition 3.3

$$P_{\bar{x}_0}(\exists N \; \forall n \geq N \; \bar{X}_n \notin \Delta) = 0 \tag{39}$$

resp. equivalent to (39)

$$P_{\bar{x}_0}(\forall N \; \exists n \geq N \; \bar{X}_n \in \Delta) = 1.$$

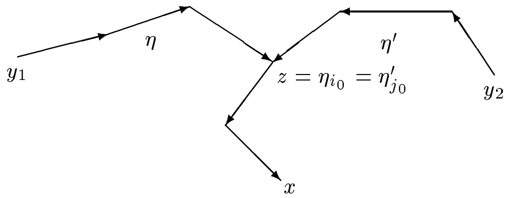
Therefore we have a successful coupling, e.g. $\widetilde{P}_{\bar{x}_0}(\tau_\Delta < \infty) = 1$, where $\tau_\Delta$ denotes the first entrance time into the diagonal. $\square$

**Lemma 3.5** *For the quantity $\theta_\Delta$ defined in Proposition 3.4 we have $\theta_\Delta \leq m_{\min}$.*

**Proof:** Let $x \in X$ and $g \in G_1\{x\}$ with $h_1(g) = m_{\min}$. For any $y_1 \neq y_2$ there exist paths $\eta$, $\eta'$ in $g$ from $y_1$ resp. $y_2$ to $x$, e.g.

$$\begin{aligned} \eta &= (y_1 = \eta_0, \eta_1, \ldots, \eta_n = x) \\ \eta' &= (y_2 = \eta'_0, \eta'_1, \ldots, \eta'_m = x) \end{aligned}$$

Let $i_0 \in \{0, \ldots, n\}$ be minimal satisfying the property $\eta_{i_0} = \eta'_j$ for some $j \in \{0, \ldots, m\}$. Note that there exist only one $j_0 \in \{0, \ldots, m\}$ with $\eta_{i_0} = \eta'_{j_0} =: z$.




Since the paths

$$\begin{aligned} &(y_1 = \eta_0, \ldots, \eta_{i_0} = z) \\ &(y_2 = \eta'_0, \ldots, \eta_{j_0} = z) \end{aligned}$$

have disjoint set of edges, we conclude

$$\begin{aligned} \widetilde{V}((y_1, y_2), (z, z))^k &= V(y_1, z)^k + V(y_2, z)^k \\ &\leq \sum_{i=0}^{i_0 - 1} s(\gamma_{\eta_i \eta_{i+1}}) + \sum_{j=0}^{j_0 - 1} s(\gamma_{\eta'_j \eta'_{j+1}}) \\ &\leq \sum_{e \in g} s(\gamma_e) \\ &= m_{\min} \quad \square \end{aligned}$$

## 3.3 Strong ergodicity of $(X_n)_n$

Suppose that for $g \in G_L\{x\}$ and all $x \in X$ the convergence

$$\sum_{n=0}^{\infty} \Big| \prod_{(y\to z)\in g} Q_{n,L}(y,z) - \prod_{(y\to z)\in g} Q_{n+1,L}(y,z)\Big| < \infty \tag{40}$$

holds.
From (40) we conclude that

$$\begin{aligned}
& \sum_{n=0}^{\infty} \|\mu_n - \mu_{n+1}\| \\
= \; & \frac{1}{2}\sum_{n=0}^{\infty}\sum_{x\in X} |\mu_n(x) - \mu_{n+1}(x)| \\
= \; & \frac{1}{2}\sum_{x\in X}\sum_{n=0}^{\infty} \Big| \sum_{g\in G_L\{x\}} \prod_{(y\to z)\in g} Q_{n,L}(y,z) - \sum_{g\in G_L\{x\}} \prod_{(y\to z)\in g} Q_{n+1,L}(y,z)\Big| \\
\leq \; & \frac{1}{2}\sum_{x\in X}\sum_{n=0}^{\infty}\sum_{g\in G_L\{x\}} \Big| \prod_{(y\to z)\in g} Q_{n,L}(y,z) - \prod_{(y\to z)\in g} Q_{n+1,L}(y,z)\Big| \\
= \; & \frac{1}{2}\sum_{x\in X}\sum_{g\in G_L\{x\}} \underbrace{\sum_{n=0}^{\infty} \Big| \prod_{(y\to z)\in g} Q_{n,L}(y,z) - \prod_{(y\to z)\in g} Q_{n+1,L}(y,z)\Big|}_{<\infty} \\
< \; & \infty.
\end{aligned}$$

Proposition 3.4 and Lemma 3.5 imply the weak ergodicity of $(X_n)_n$ for temperature schedules like $\beta(n) = c\ln(n+1)$, where $c \leq \frac{1}{m_{\min}}$. We show in the sequel that (40) holds for $\beta(n) = \frac{1}{m_{\min}}\ln(n+1)$. Then the strong ergodicity of $(X_n)_n$ follows by the well-known result that an inhomogeneous weakly ergodic Markov chain with stationary measures $(\mu_n)_n$ is strongly ergodic if $\sum_{n=0}^{\infty}\|\mu_n - \mu_{n+1}\| < \infty$. Compare [IM76] p. 160.

**Definition 3.6** *A real sequence $(a_n)_n$ is called exponentially convergent with respect to $(\beta(n))_n$, if there exists $a \in \mathbb{R}$ such that there exists $N \in \mathbb{N}$, $h > 0$ and $K \geq 0$ such that*

$$|a_n - a| \leq K\exp(-\beta(n)h) \text{ for all } n \geq N. \tag{41}$$

*If (41) holds we also say that $(a_n)_n$ is exponentially convergent with rate $h$ with respect to $(\beta(n))_n$.*

Obviously the set of exponentially convergent sequences with respect to $(\beta(n))_n$ is a real algebra.

Hence there exists $h > 0$ and $K \geq 0$ such that for the sequences $(d_n)_n$ and $(d_{n+1})_n$ given by

$$d_n := \prod_{(y\to z)\in g} \sum_{v\in\mathcal{K}_{*,L}(y,z)} \prod_{i=1}^{p} c_n(\gamma_{yv_i}) \sum_{l:v_l=z} F_n(l,y,v)$$

holds

$$|d_n - d_{n+1}| \leq K\exp(-\beta(n)h) \quad \text{for all } n \geq N. \tag{42}$$

Choose $N$ such that (42) and $r(n) = p$ for $n \geq N$ holds. Furthermore we assume that $F_n = F$ for all $n \geq N$ according to the hypothesis $H_0$. The sequence $d_n$ is isotone for all $n \geq N$, because for any path $\gamma_{yv_i}$ the sequence $c_n(\gamma_{yv_i})$ is isotone. Define

$$\begin{aligned} a_n(y,z) &:= \sum_{v\in\mathcal{K}_{*,L}(y,z)} \exp\left(-\beta(n)\sum_{i=1}^{p} s(\gamma_{yv_i})\right) \prod_{i=1}^{p} c_n(\gamma_{yv_i}) \sum_{l:v_l=z} F(l,y,v) \\ b_n(y,z) &:= \sum_{v\in\mathcal{K}_L(y,z)\setminus\mathcal{K}_{*,L}(y,z)} \exp\left(-\beta(n)\sum_{i=1}^{p} s(\gamma_{yv_i})\right) \prod_{i=1}^{p} c_n(\gamma_{yv_i}) \sum_{l:v_l=z} F(l,y,v) \end{aligned}$$

and write the product $\prod_{(y\to z)\in g} Q_{n,L}(y,z)$ for $n \geq N$ in the following way

$$\prod_{(y\to z)\in g} Q_{n,L}(y,z) = \prod_{(y\to z)\in g} (a_n(y,z) + b_n(y,z)) = \tilde{a}_n(g) + \tilde{b}_n(g), \tag{43}$$

where

$$\begin{aligned} \tilde{a}_n(g) &:= \prod_{i=1}^{|g|} a_n(e_i) \\ \tilde{b}_n(g) &:= \prod_{(y\to z)\in g} Q_{n,L}(y,z) - \tilde{a}_n(g). \end{aligned}$$

This yields

$$\begin{aligned} &\sum_{n=N}^{\infty} |\prod_{(y\to z)\in g} Q_{n,L}(y,z) - \prod_{(y\to z)\in g} Q_{n+1,L}(y,z)| \\ = &\sum_{n=N}^{\infty} |\tilde{a}_n(g) + \tilde{b}_n(g) - (\tilde{a}_{n+1}(g) + \tilde{b}_{n+1}(g))| \\ \leq &\sum_{n=N}^{\infty} |\tilde{a}_n(g) - \tilde{a}_{n+1}(g)| + \sum_{n=N}^{\infty} |\tilde{b}_n(g) - \tilde{b}_{n+1}(g)|. \end{aligned}$$

Choose $\beta(n) = \frac{1}{m_{\min}} \ln(n+1)$ and note that $\tilde{h}_L(g) = \sum_{(y\to z)\in g} k_L(y,z) \geq m_{\min}$. Use

$$\tilde{a}_n(g) = \exp(-\beta(n)\tilde{h}_L(g))c_n.$$

and conclude

$$
\begin{aligned}
& \sum_{n=N}^{\infty} |\widetilde{a}_n(g) - \widetilde{a}_{n+1}(g)| \\
= & \sum_{n=N}^{\infty} |\exp\left(-\beta(n)\widetilde{h}(g)\right) c_n - \exp\left(-\beta(n+1)\widetilde{h}(g)\right) c_{n+1}| \\
\leq & \sum_{n=N}^{\infty} \exp(-\beta(n+1)\widetilde{h}(g))|c_n - c_{n+1}| \\
\leq & \sum_{n=N}^{\infty} K \exp(-\beta(n+1)(\widetilde{h}(g)+h)) \\
= & \sum_{n=N}^{\infty} \frac{1}{(n+1)^{\frac{\widetilde{h}(g)+h}{m_{\min}}}} < \infty.
\end{aligned}
$$

Any summand in $\prod_{(y\to z)\in g} Q_{n,L}(y,z)$, that is different from $\widetilde{a}_n(g)$ is of the form

$$\prod_{j\in J} a_n(e_j) \prod_{j\in I\setminus J} b_n(e_j),$$

where

$$
\begin{aligned}
I &:= \{1,\ldots,|g|\} \\
J &\subset I \text{ but } J \neq I \\
g &= \{\, e_1,\ldots,e_{|g|} \,\}.
\end{aligned}
$$

We maintain that for any of these summands the series

$$\sum_{n=0}^{\infty} |\prod_{j\in J} a_n(e_j) \prod_{j\in I\setminus J} b_n(e_j)|$$

is finite.
Set

$$\varepsilon(y,z) := \min\left\{ \sum_{i=1}^{p} s(\gamma_{yv_i}) \;\middle|\; v\in\mathcal{K}_L(y,z)\setminus\mathcal{K}_{*,L}(y,z) \right\} - k_L(y,z),$$

and

$$\varepsilon := \min\{\, \varepsilon(y,z) \mid (y\to z)\in g \,\} > 0.$$

For $j\in I\setminus J$ we have

$$
\begin{aligned}
b_n(e_j) &= \sum_{v\in\mathcal{K}_L(y,z)\setminus\mathcal{K}_{*,L}(y,z)} \exp\left(-\beta(n)\sum_{i=1}^{p} s(\gamma_{yv_i})\right) \prod_{i=1}^{p} c_n(\gamma_{yv_i}) \sum_{l:v_l=e^+} F_n(l,y,e) \\
&\leq \exp(-\beta(n)(k_L(e_j)+\varepsilon)) \sum_{v\in\mathcal{K}_L(y,z)\setminus\mathcal{K}_{*,L}(y,z)} \prod_{i=1}^{p} c_n(\gamma_{yv_i}) \sum_{l:v_l=e^+} F_n(l,y,e) \\
&\leq C\exp(-\beta(n)(k_L(e_j)+\varepsilon)),
\end{aligned}
$$

where $C$ is a positive constant.
This yields

$$\begin{aligned}
\sum_{n=0}^{\infty} \Big| \prod_{j\in J} a_n(e_j) \prod_{j\in I\setminus J} b_n(e_j) \Big| &\leq C' \sum_{n=0}^{\infty} \exp\left(-\beta(n)\left(\sum_{i\in I} k_(e_i) + \varepsilon\right)\right) \\
&= C' \sum_{n=0}^{\infty} \frac{1}{(n+1)^{\frac{\widetilde{h}(g)+\varepsilon}{m_{\min}}}} < \infty,
\end{aligned}$$

where $C'$ is a positive constant.
This establishes (40) for $\beta(n) = \frac{1}{m_{\min}} \ln(n+1)$.
In example 2.11 we have $m_{\min} = (a+2)(N+1)$ and $\theta_\Delta = 3(N+1)$. Hence Proposition 3.4 yields that $(X_n)_n$ is weakly ergodic for $\beta(n) = c\ln(n+1)$ with $c \leq \frac{1}{p\theta_\Delta}$. Therefore $(X_n)_n$ is strongly ergodic for $c \in [\frac{1}{(a+2)(N+1)}, \frac{1}{3p(N+1)}] \neq \emptyset$ for $a \geq 3p-2$.

# References


[D38] W. Doeblin, *Exposé de la théorie des chaines simples constantes de Markov a un nombre fini détats*, Rev. Math. Union Interbalkanique **2**, 77-105 (1938).

[IM76] Dean L. Isaacson, Richard W. Madsen, *Markov Chains Theory and Applications* Wiley (1976).

[GR75] D. Griffeath, *A Maximal Coupling for Markov Chains* Z. Wahrscheinlichkeitstheorie verw. Gebiete **31**, 95-106 (1975).

[P76] J.W. Pitman, *On Coupling of Markov Chains*,
Z. Wahrscheinlichkeitstheorie verw. Gebiete **35** 315-322 (1976).

[GR78] D. Griffeath, *Coupling Methods for Markov Processes* Studies in Probability and Ergodic Theory. Adv. Math. Supplementary Studies **2** (1978).

[FW84] M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer Verlag (1984).

[C94] R. Cérf, *Une Théorie Asymptotique Des Algorithmes Génétiques*, Thése de Doctorat, Université Montpellier II (1994).

[T93] H. Thorisson, *Coupling and Shift-Coupling Random Sequences* Contemporary Mathematics **149**, 85-95 (1993).

[DLS] R. Diekmann, R. Lüling, J. Simon, *Problem Independent Distributed Simulated Annealing and its Applications*, Lecture Notes in Economics and Math. Syst. 369, Springer, 1-16 (1993).