修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報システム学研究科 情報ネットワークシステム学 専攻 博士前期課程		
氏　　名	小林 哲也	学籍番号	1252014
論 文 題 目	The Capacity of Write-Once Memory with a Writing Cost Function （書込コスト関数を持つライト・ワンス・メモリの容量）		

要　　旨

書換に制限を有する記憶媒体というのは古今東西存在する．例えば，パンチカードは一度穴を開けたら塞ぐことができない．またフラッシュメモリにおいては，ブロック消去と呼ばれる特別な操作を行わなければ，記憶素子であるセル内部の電荷を減らすことができない．R.L. Rivest と A. Shamir は 1982 年に，そのような制限を有する記憶媒体のモデルとして Write-Once Memory (WOM) を導入した．そして彼らは，WOM を複数回にわたり書換える際，書込可能なメッセージの量をなるべく多くするための工夫とはどのようなものかを提示した．そのような工夫は，WOM 符号化と呼ばれている．

オリジナルの WOM では記憶素子の取りうる状態が二値であったが，その後 1984 年に A. Fiat と A. Shamir により，多値を扱える一般化 WOM が提案された．一般化 WOM では，素子の取りうる状態および状態間の可能な遷移を，無閉路有向グラフによって指定する．

WOM において興味深い問題の一つは，最良の符号化によりどれだけ多くのメッセージを書込めるかということである．この問題に対する解答として，F. Fu と A.J. Han Vinck は 1999 年に，任意の一般化 WOM における容量域と最大和率を決定した．

本研究では，一般化 WOM をさらに拡張した Write-Constrained Memory (WCM) を提案する．WCM の大きな特徴の一つは，それが状態遷移のコストを考慮することである．実用的には，状態遷移コストは時間やエネルギーといった物理量としての意味を持つものであり，何らかの理由でそれを制限しなければならないという状況が想定される．本研究ではそのような制限の一つとして，WCM に対する平均コスト制約というものを提案する．

本研究における主要な結果は，任意の WCM において，定数回にわたり書換える際，平均コスト制約を満足する容量域および最大和率を決定したことである．

Master's Thesis

The Capacity of Write-Once Memory with a Writing Cost Function

Department of Information Network Systems

Graduate School of Information Systems

The University of Electro-Communications

Tetsuya KOBAYASHI

Student Number: 1252014

Supervisory Committee:

Chair	:	Prof. Hiroyoshi MORITA
Member	:	Assoc. Prof. Hiroyuki KASAI
Member	:	Assoc. Prof. Tomohiro OGAWA

January 27, 2014

Contents

Chapter 1

Introduction

In all ages we find such recording media as have restrictions on rewriting. For example, a punched card represents information by the absence or presence of holes, and once we punch a hole we cannot undo it anymore. A vinyl record stores a sound by the depths of a spiral groove, and we can only deepen it. A flash memory stores messages by the electric charge in floating gates, and we cannot decrease it without block erasure.

In 1982, Rivest and Shamir [1] introduced a model of such rewrite-restricted media, called *Write-Once Memory (WOM)*, and formulated a notion of coding to rewrite on WOM. To be exact, their model of WOM should be called *the binary WOM*, and it consists of a sequence of storage elements called *wit*s. Each wit takes a state 0 or 1, and once we change a state of a wit from 0 to 1, we cannot change it anymore. And they presented a code that can store 2 bits of information twice in a 3-wit binary WOM (Table 1.1).

Fiat and Shamir [2] introduced *generalized WOM*'s. When a finite directed acyclic graph

decode	
000 or 111	0
100 or 011	1
010 or 101	2
001 or 110	3

encode			2nd: 0	1	2	3
1st	0	000	000	100	010	001
	1	100	111	100	101	110
	2	010	111	011	010	110
	3	001	111	011	101	001

Table 1.1: Store 2 bits of information twice in a 3-wit binary WOM.

$(\mathscr{V}, \mathscr{E})$ is given, $(\mathscr{V}, \mathscr{E})$-WOM is such that each storage element takes a state in $\mathscr{V}$, and that for two states i and j, a transition from i to j is possible iff there is a path in $(\mathscr{V}, \mathscr{E})$ from i to j.

An interesting question about WOM is how much information can be stored by rewriting on WOM multiple times. For example, using the code shown in Table 1.1, we can store 2 bits at the first writing and 2 bits at the second writing in a 3-wit binary WOM, and so the rate vector of this code is (2/3, 2/3), and the sum rate is 4/3. What a good code does there exist when the WOM length goes to infinity? What is the region that the rate vector of some code is in? How large can the sum rate of a code be? Fu and Han Vinck [3] answered these questions by determining the capacity region and the maximum sum rate.

We propose a further generalization of generalized WOM, called *Write-Constrained Memory (WCM)*. For a finite costed directed graph $(\mathscr{V}, \mathscr{E}, w)$, the $(\mathscr{V}, \mathscr{E}, w)$-WCM is such that each storage element, which we call *cell*, takes a state in $\mathscr{V}$, that for two states i and j, a state transition from i to j in one update is possible iff $(i, j) \in \mathscr{E}$, and that for each $(i, j) \in \mathscr{E}$, a state transition from i to j costs $w(i, j)$. The cost of WCM update is defined as the sum of the cost of each cell update. We restrict the average cost of WCM updates per cell per update.

But why does a state transition cost matter? In practice, a state transition cost is a physical quantity such as energy or time. For example, on punched cards, it is natural to think that the more holes must be punched, the more time the punching operation consumes. When you use an electric keypunch, the total electric energy to punch holes will be proportional to the number of holes to be punched. And it is a plausible story that for some reason you have to restrict the operation time or the operation energy.

In this thesis, we introduce WCM and determine the capacity region and the maximum sum rate of WCM with a certain cost constraint. This can be considered as an extension of Fu and Han Vinck's results for generalized WOM's in [3].

Chapter 2

Capacity of Write-Constrained Memory

In this chapter, we first give a mathematical definition of Write-Constrained Memory (WCM). Then we define and determine the capacity region and the maximum sum rate of WCM.

2.1 Preliminaries

2.1.1 Notation

For a nonempty set $\mathcal{S}$ of real numbers, we denote by $\min \mathcal{S}$ (resp. $\max \mathcal{S}$) the minimum (resp. maximum) element in $\mathcal{S}$.

2.1.2 Write-Constrained Memory

For $q \geq 2$, we fix $\mathscr{V} = \{0, 1, \ldots, q-1\}$, $\mathscr{E} \subset \mathscr{V} \times \mathscr{V}$, and $w : \mathscr{E} \to \mathbb{R}$, and then consider the $(\mathscr{V}, \mathscr{E}, w)$-WCM. The meaning of each parameter is as described in Chapter 1. Note that unlike in the case of generalized WOM's, the state transitions are associated with the edges of the graph $(\mathscr{V}, \mathscr{E})$, not with the paths.

In terms of the state transitions, we introduce an "arrow" notation for convenience. For i, $j \in \mathscr{V}$, let $i \to j$ mean $(i, j) \in \mathscr{E}$. For any $n \geq 1$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathscr{V}^n$, let $\boldsymbol{x} \to \boldsymbol{y}$ mean $(x_i, y_i) \in \mathscr{E}$ for every $1 \leq i \leq n$.

Formally, a sequence of pairs of functions $\{(\phi_t, \psi_t)\}_{t=1}^{T}$ is called an $[n, T, M_1, \ldots, M_T]$-code for the $(\mathscr{V}, \mathscr{E}, w)$-WCM when it satisfies the following conditions. The domain and the codomain of the functions ϕ_t and ψ_t are such that

$$
\begin{aligned}
&\phi_1 : I_1 \times \{\mathbf{0}\} \to \mathscr{V}^n, \\
&\phi_t : I_t \times \mathrm{Im}(\phi_{t-1}) \to \mathscr{V}^n \quad (2 \le t \le T), \\
&\psi_t : \mathscr{V}^n \to I_t \qquad\qquad\quad (1 \le t \le T),
\end{aligned}
$$

where we let $I_t = \{1, \ldots, M_t\}$ for $1 \le t \le T$, and for every $1 \le t \le T$ and for every $(a, \boldsymbol{x}) \in \mathrm{Dom}(\phi_t)$ it holds that

$$\psi_t(\phi_t(a, \boldsymbol{x})) = a, \quad \boldsymbol{x} \to \phi_t(a, \boldsymbol{x}).$$

The meaning is that ϕ_t is the t-th encoder and ψ_t is the t-th decoder, and $\mathbf{0} \in \mathscr{V}^n$ is the initial state of the WCM, where every cell takes state 0. Using this code, we can write T times on an n-cell WCM, with the number of messages being M_t at the t-th writing for $1 \le t \le T$.

2.1.3 Cost Constraint

Assume $\{(\phi_t, \psi_t)\}_{t=1}^{T}$ is an $[n, T, M_1, \ldots, M_T]$-code for the $(\mathscr{V}, \mathscr{E}, w)$-WCM. Fix $(a^{(1)}, \ldots, a^{(T)}) \in I_1 \times \cdots \times I_T$, where I_t's $(1 \le t \le T)$ are as in Subsect. 2.1.2, and consider writing $a^{(1)}, \ldots, a^{(T)}$ sequentially using this code. The state of the WCM after the t-th writing is given recursively by $\boldsymbol{x}^{(t)} = \phi_t(a^{(t)}, \boldsymbol{x}^{(t-1)})$, for $1 \le t \le T$, with $\boldsymbol{x}^{(0)} = \mathbf{0}$, and the cost per cell of the t-th writing is given by $w(\boldsymbol{x}^{(t-1)}, \boldsymbol{x}^{(t)})/n$, where we let $w(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} w(x_i, y_i)$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{V}^n$ such that $\boldsymbol{x} \to \boldsymbol{y}$. With the cost constraint we adopt in this paper, we restict the summation for $1 \le t \le T$ of the cost per cell of the t-th writing. We say that the code $\{(\phi_t, \psi_t)\}_{t=1}^{T}$ satisfies *c-average cost constraint* when it satisfies

$$\sum_{t=1}^{T} w(\boldsymbol{x}^{(t-1)}, \boldsymbol{x}^{(t)})/n \le Tc$$

for every $(a^{(1)}, \ldots, a^{(T)}) \in I_1 \times \cdots \times I_T$.

2.2 Capacity Region

For $T \geq 1$ and $c \in \mathbb{R}$, we define *the capacity region of the* $(\mathcal{V}, \mathcal{E}, w)$*-WCM with* T *writings and* c*-average cost constraint*, denoted by $\mathscr{A}_{T,c}(\mathcal{V}, \mathcal{E}, w)$, as

$$\begin{aligned}\mathscr{A}_{T,c}(\mathcal{V}, \mathcal{E}, w) \triangleq \mathrm{Clo}\{\,((\log M_1)/n, \ldots, (\log M_T)/n) : \quad & n \geq 1, \quad M_1, \ldots, M_T \geq 1, \\ & \text{there exists an } [n, T, M_1, \ldots, M_T]\text{-code for the } (\mathcal{V}, \mathcal{E}, w)\text{-WCM that satisfies} \\ & c\text{-average cost constraint}\},\end{aligned}$$

where $((\log M_1)/n, \ldots, (\log M_T)/n)$ is called the rate vector of the $[n, T, M_1, \ldots, M_T]$-code.

Informally, $\mathscr{A}_{T,c}(\mathcal{V}, \mathcal{E}, w)$ is the set of every T-vector that a sequence of the rate vectors of the suitable WCM codes converges to.

To describe $\mathscr{A}_{T,c}(\mathcal{V}, \mathcal{E}, w)$, we extend the "arrow" notation to the random variables on $\mathcal{V}$. For X, Y that are r.v.'s on $\mathcal{V}$, let $X \to Y$ mean that $\Pr\{X = i, Y = j\} > 0$ holds only if $(i, j) \in \mathcal{E}$. When we write such as $X_1 \to X_2 \to \cdots \to X_r$, we implicitly assume that $(X_i)_{i=1}^r$ forms a Markov chain.

Theorem 2.1. The following holds.

$$\begin{aligned}\mathscr{A}_{T,c}(\mathcal{V}, \mathcal{E}, w) = \{(R_1, \ldots, R_T) \in \mathbb{R}^T : \quad & \text{there exist r.v.'s } S_1, \ldots, S_T \text{ on } \mathcal{V} \text{ s.t.} \\ & 0 \to S_1 \to \cdots \to S_T, \quad \mathrm{E}[w(0, S_1) + \textstyle\sum_{t=2}^T w(S_{t-1}, S_t)] \leq Tc, \\ & 0 \leq R_1 \leq H(S_1), \quad 0 \leq R_t \leq H(S_t \mid S_{t-1}) \ \ (2 \leq \forall t \leq T)\},\end{aligned}$$

where $\mathrm{E}[\cdot]$ means the expectation and $H(\cdot)$ the entropy.

Proof. For the proof, we need only minor modifications to the proof of [3, Theorem 3.1]. In the following, we deliver a sketch of the proof, without a mathematical detail.

Direct Part: We only show the proof for $T = 2$. Extension to any T is easy.

The discussion utilizes the type theory. For a random variable X on a finite set $\mathcal{X}$, denote by

T_X^n the set of all n-vectors whose type is equal to X, that is, we define T_X^n as

$$T_X^n = \{(x_1, \ldots, x_n) \in \mathcal{X}^n \mid \#\{i \mid 1 \le i \le n,\ x_i = a\} = n \cdot \Pr\{X = a\}\ \ (\forall a \in \mathcal{X})\}.$$

For $n \in \mathbb{N}$, we say that the distribution of a r.v. X on $\mathcal{X}$ is an n-type when $n \cdot \Pr\{X = a\}$ is an integer value for every $a \in \mathcal{X}$. Similarly, for a r.v. X on $\mathcal{X}$ and a r.v. Y on $\mathcal{Y}$, the joint distribution of X and Y is called an n-type when $n \cdot \Pr\{X = a,\ Y = b\}$ is an integer value for every $a \in \mathcal{X}$ and $b \in \mathcal{Y}$.

Suppose r.v.'s S_1, S_2 on $\mathscr{V}$ are given such that $0 \to S_1 \to S_2$ and $\mathrm{E}[w(0, S_1) + w(S_1, S_2)] \le Tc$.

First we make an n-approximation $(\tilde{S}_1^{(n)}, \tilde{S}_2^{(n)})$ of (S_1, S_2) for $n \in \mathbb{N}$. It can be shown that there exists $\{(\tilde{S}_1^{(n)}, \tilde{S}_2^{(n)})\}_{n \in \mathbb{N}}$ such that for every $n \in \mathbb{N}$, $\tilde{S}_1^{(n)}$ and $\tilde{S}_2^{(n)}$ are r.v.'s on $\mathscr{V}$ whose joint distribution is an n-type, and which satisfy $0 \to \tilde{S}_1^{(n)} \to \tilde{S}_2^{(n)}$ and $\mathrm{E}[w(0, \tilde{S}_1^{(n)}) + w(\tilde{S}_1^{(n)}, \tilde{S}_2^{(n)})] \le \mathrm{E}[w(0, S_1) + w(S_1, S_2)]$, and that $(\tilde{S}_1^{(n)}, \tilde{S}_2^{(n)})$ converges in law to (S_1, S_2) as $n \longrightarrow \infty$. Now we explain how to make $(\tilde{S}_1^{(n)}, \tilde{S}_2^{(n)})$. We let $\mathcal{I} = \{(i_1, i_2) \in \mathscr{V}^2 \mid \Pr\{S_1 = i_1, S_2 = i_2\} > 0\}$, and take $(i_1^*, i_2^*) \in \mathcal{I}$ that satisfies $w(0, i_1^*) + w(i_1^*, i_2^*) \le w(0, i_1) + w(i_1, i_2)$ for every $(i_1, i_2) \in \mathcal{I}$. Define $p^{(n)} : \mathscr{V}^2 \to \mathbb{R}$ as

$$p^{(n)}(i_1, i_2) = \begin{cases} \lfloor n \cdot \Pr\{S_1 = i_1,\ S_2 = i_2\} \rfloor / n & \text{if } (i_1, i_2) \in \mathcal{I} \setminus \{(i_1^*, i_2^*)\}, \\ 1 - \displaystyle\sum_{(i_1', i_2') \in \mathcal{I} \setminus \{(i_1^*, i_2^*)\}} p^{(n)}(i_1', i_2') & \text{if } (i_1, i_2) = (i_1^*, i_2^*), \\ 0 & \text{otherwise.} \end{cases}$$

Then $p^{(n)}$ gives a probability distribution on $\mathscr{V}^2$. We give the joint probability distribution of $\tilde{S}_1^{(n)}$ and $\tilde{S}_2^{(n)}$ by $\Pr\{\tilde{S}_1^{(n)} = i_1,\ \tilde{S}_2^{(n)} = i_2\} = p^{(n)}(i_1, i_2)$ for every $(i_1, i_2) \in \mathscr{V}^2$.

Take $\delta > 0$ arbitrarily and set $M_1 := 2^{n[H(S_1) - \delta]}$, $M_2 := 2^{n[H(S_2|S_1) - \delta]}$. For sufficiently large n, if we can find a partition $\{F_m\}_{m=1}^{M_2}$ of $T^n_{\tilde{S}_2^{(n)}}$ that satisfies the following condition, then we can construct an $[n, 2, M_1, M_2]$-code for the $(\mathscr{V}, \mathscr{E}, w)$-WCM that satisfies c-average cost constraint.

The condition is that for any $\boldsymbol{x} \in T^n_{\tilde{S}_1^{(n)}}$ and for any $m \in I_2 = \{1, \ldots, M_2\}$ there exists $\boldsymbol{y} \in F_m$ such that $\boldsymbol{x} \to \boldsymbol{y}$. The code construction is as follows. For the first writing, let $\phi_1(\cdot, \boldsymbol{0})$ be any one-to-one mapping from $I_1 = \{1, \ldots, M_1\}$ to $T^n_{\tilde{S}_1^{(n)}}$, and for every $a \in M_1$, set $\psi_1(\phi_1(a, \boldsymbol{0})) := a$. For the second writing, set to $\phi_2(m, \boldsymbol{x})$ an element $\boldsymbol{y}$ in F_m that satisfies $\boldsymbol{x} \to \boldsymbol{y}$, and for every $m \in I_2$ and $\boldsymbol{y} \in F_m$, set $\psi_2(\boldsymbol{y}) := m$.

A partition that satisfies the condition in the previous paragraph can be shown to exist for sufficiently large n, same as in the proof of [3, Theorem 3.1].

Converse Part: We show that for any $[n, T, M_1, \ldots, M_T]$-code for $(\mathscr{V}, \mathscr{E}, w)$-WCM that satisfies c-average cost constraint, there exist r.v.'s $S_1, \ldots, S_T$ on $\mathscr{V}$ such that $0 \to S_1 \to \cdots \to S_T$, $\mathrm{E}[w(0, S_1) + \sum_{t=2}^{T} w(S_{t-1}, S_t)] \le Tc$, $(\log M_1)/n \le H(S_1)$, and $(\log M_t)/n \le H(S_t \mid S_{t-1})$ for $2 \le t \le T$.

Let $\{(\phi_t, \psi_t)\}_{t=1}^{T}$ be such a code. Let $W_1, \ldots, W_T$ be independent r.v.'s uniformly distributed over $I_t = \{1, \ldots, M_t\}$ respectively. Denote $Y_0^n = \boldsymbol{0}$, $Y_t^n = (Y_{t,1}, \ldots, Y_{t,n}) = \phi_t(W_t, Y_{t-1}^n)$, $1 \le t \le T$. It follows that $Y_{t-1}^n \to Y_t^n$, and $\psi_t(Y_t^n) = W_t$, $Y_{t-1,i} \to Y_{t,i}$. Then we have $H(W_t) = H(Y_t^n|Y_{t-1}^n)$, $t = 1, \ldots, T$. Let L be an index r.v. which uniformly distributed over $\{1, \ldots, n\}$, independent of all other r.v.'s. Then we have $(\log M_t)/n \le H(Y_{t,L} \mid Y_{t-1,L})$ and $\mathrm{E}[\sum_{t=1}^{T} w(Y_{t-1,L}, Y_{t,L})] \le Tc$. Here, $(Y_{t,L})_{t=1}^{T}$ may not form a Markov chain, but we can take new r.v.'s $S_1, \ldots, S_T$ on $\mathscr{V}$ such that $(S_t)_{t=1}^{T}$ forms a Markov chain and for every $1 \le t \le T$, (S_{t-1}, S_t) and $(Y_{t-1,L}, Y_{t,L})$ have the same probability distribution. Therefore, $0 \to S_1 \to \cdots \to S_T$, $\mathrm{E}[w(0, S_1) + \sum_{t=2}^{T} w(S_{t-1}, S_t)] \le Tc$, $(\log M_1)/n \le H(S_1)$, and $(\log M_t)/n \le H(S_t \mid S_{t-1})$ for $2 \le t \le T$. □

2.3 Maximum Sum Rate

For $T \ge 1$ and $c \in \mathbb{R}$, we define *the maximum sum rate of the $(\mathscr{V}, \mathscr{E}, w)$-WCM with T writings and c-average cost constraint*, denoted by $C_{T,c}(\mathscr{V}, \mathscr{E}, w)$, as

$$C_{T,c}(\mathscr{V}, \mathscr{E}, w) \triangleq \max\left\{\sum_{t=1}^{T} R_t : (R_1, \ldots, R_T) \in \mathscr{A}_{T,c}(\mathscr{V}, \mathscr{E}, w)\right\}.$$

Preliminary to the determination of the maximum sum rate, here we introduce some notations. In what follows we fix $T \geq 1$. We define $\mathscr{P}_T \subset \mathscr{V}^T$ and $W_T : \mathscr{P}_T \to \mathbb{R}$ as

$$\mathscr{P}_T \triangleq \{(i_1, \ldots, i_T) \in \mathscr{V}^T \mid 0 \to i_1 \to \cdots \to i_T\} \text{ and}$$
$$W_T(i_1, \ldots, i_T) \triangleq w(0, i_1) + \sum_{t=2}^{T} w(i_{t-1}, i_t),$$

and we let $\mathscr{W}_T \triangleq \mathrm{Im} W_T$ and define the family $(\eta_u^{(T)})_{u \in \mathscr{W}_T}$ of positive integers as $\eta_u^{(T)} \triangleq \left| W_T^{-1}(\{u\}) \right|$ $(u \in \mathscr{W}_T)$. Note that $\mathscr{W}_T$ is a finite set. We denote $(\eta_u^{(T)})_{u \in \mathscr{W}_T}$ by $\boldsymbol{\eta}^{(T)}$, simply.

For $v \in \mathbb{R}$, we say that a family $(a_u)_{u \in \mathscr{W}}$ is *v-conformant* when it is a family of nonnegative real numbers indexed by a finite set of real numbers and there exist $u, u' \in \mathscr{W}$ such that $u < v < u'$, $a_u > 0$ and $a_{u'} > 0$.

Theorem 2.2. Suppose $|\mathscr{W}_T| \geq 2$ and fix $c \in \mathbb{R}$ arbitrarily such that $\min \mathscr{W}_T < Tc < \max \mathscr{W}_T$. Then the maximum sum rate $C_{T,c}(\mathscr{V}, \mathscr{E}, w)$ is given by

$$C_{T,c}(\mathscr{V}, \mathscr{E}, w) = \begin{cases} \zeta(\boldsymbol{\eta}^{(T)}) & \text{if } Tc \leq \dfrac{\sum_{u \in \mathscr{W}_T} u \eta_u^{(T)}}{\sum_{u \in \mathscr{W}_T} \eta_u^{(T)}}, \\ \log |\mathscr{P}_T| & \text{otherwise,} \end{cases} \tag{2.1}$$

where, for a Tc-conformant family $\boldsymbol{\eta} = (\eta_u)_{u \in \mathscr{W}_T}$ we denote by $\alpha(\boldsymbol{\eta})$ the unique positive root w.r.t. α of $g(\boldsymbol{\eta}, \alpha)$ that is defined as

$$g(\boldsymbol{\eta}, \alpha) \triangleq \sum_{u \in \mathscr{W}_T} \eta_u (u - Tc) \alpha^u, \tag{2.2}$$

and define $\zeta(\boldsymbol{\eta})$ as

$$\zeta(\boldsymbol{\eta}) \triangleq \log \left(\sum_{u \in \mathscr{W}_T} \eta_u \, \alpha(\boldsymbol{\eta})^u \right) - Tc \log \alpha(\boldsymbol{\eta}). \tag{2.3}$$

To prove Theorem 2.2, we first prepare a lemma and a corollary about zeroes of a certain type of functions, which we will use later.

Lemma 2.3. Let $\mathscr{W}$ be a finite set of real numbers and let $(a_u)_{u\in\mathscr{W}}$ be a family of real numbers indexed by $\mathscr{W}$ and suppose that there exists a real number v such that for every $u \in \mathscr{W}$, $a_u \geq 0$ if $u > v$ and $a_u \leq 0$ if $u < v$. Then for the function $f(x) = \sum_{u\in\mathscr{W}} a_u x^u$ on $\mathbb{R}$, we have either $f(x)$ is identically zero or $f(x)$ has at most one positive zero.

Proof. The proof is by mathematical induction on $|\mathscr{W}|$.

The statement clearly holds when $|\mathscr{W}| \leq 1$ because in this case either $f(x)$ is identically zero or $f(x)$ has no positive zero.

Fix $k \geq 1$ and assume the statement holds when $|\mathscr{W}| \leq k$. Then we consider the case when $|\mathscr{W}| = k + 1$. By induction hypothesis, the statement holds when $a_u = 0$ for some $u \in \mathscr{W}$, and so we assume otherwise. Set $l := \max \mathscr{W}$, $s := \min \mathscr{W}$. It is easy to see that the statement holds when $a_l < 0$ or $a_s > 0$, because in this case $f(x)$ has no positive zero. So we assume otherwise.

In the assumption made above, set $\tilde{f}(x) := x^{-s} f(x)$. Then $\tilde{f}(x)$ has the same positive zeroes as $f(x)$. Also $\tilde{f}(x)$ is continuous on $[0, +\infty)$ and differentiable on $(0, +\infty)$. We have $\tilde{f}(0) < 0$ and $\tilde{f}(x) \longrightarrow +\infty$ as $x \longrightarrow +\infty$, and so $\tilde{f}(x)$ has at least one positive zero due to the intermediate value theorem. To see that $\tilde{f}(x)$ has only one positive zero, assume by contradiction that $\tilde{f}(x)$ has at least two positive zeroes. Denote the two smallest positive zeroes of $\tilde{f}(x)$ by ξ_1 and ξ_2 $(\xi_1 < \xi_2)$, and consider the following three cases.

Case 1: When $\tilde{f}(c) < 0$ for every $c \in (\xi_1, \xi_2)$. Then, ξ_1 is a local maximum of $\tilde{f}$ and so we have $\tilde{f}'(\xi_1) = 0$, and Rolle's theorem assures the existence of $c_2 \in (\xi_1, \xi_2)$ such that $\tilde{f}'(c_2) = 0$.

Case 2: When $\tilde{f}(c) > 0$ for every $c \in (\xi_1, \xi_2)$, and for every sufficiently small $\varepsilon > 0$ it holds that $\tilde{f}(\xi_2 + \varepsilon) > 0$. Then, ξ_2 is a local minimum of $\tilde{f}$ and so we have $\tilde{f}'(\xi_2) = 0$, and by Rolle's theorem there exists $c_1 \in (\xi_1, \xi_2)$ such that $\tilde{f}'(c_1) = 0$.

Case 3: When $\tilde{f}(c) > 0$ for every $c \in (\xi_1, \xi_2)$, and for every sufficiently small $\varepsilon > 0$ it holds that $\tilde{f}(\xi_2 + \varepsilon) < 0$. Now that $\tilde{f}(x) \longrightarrow +\infty$ as $x \longrightarrow +\infty$, $\tilde{f}$ has a zero greater than ξ_2, the smallest of which we denote by ξ_3. Then, Rolle's theorem guarantees the existence of $c_1 \in (\xi_1, \xi_2)$ such that $\tilde{f}'(c_1) = 0$ and the existence of $c_2 \in (\xi_2, \xi_3)$ such that $\tilde{f}'(c_2) = 0$.

In either case, $\tilde{f}'(x)$ has at least two positive zeroes. But by mathematical induction, $\tilde{f}'(x) = \sum_{u\in\mathscr{W}\setminus\{s\}}(u-s)a_u x^{u-s-1}$ has at most one positive zero, which is a contradiction. □

By the proof of Lemma 2.3, we can immediately derive the following corollary.

Corollary 2.4. In the assumption of Lemma 2.3, assume also that there exist $u, u' \in \mathscr{W}$ such that $a_u > 0$ and $a_{u'} < 0$. Then the $f(x)$ in Lemma 2.3 has the unique positive zero.

The next lemma plays an important role in proving Theorem 2.2.

Lemma 2.5. Suppose $|\mathscr{W}_T| \geq 2$ and fix $c \in \mathbb{R}$ arbitrarily such that $\min \mathscr{W}_T < Tc < \max \mathscr{W}_T$. Consider the following optimization problem.

$$\begin{aligned} \text{Maximize} \quad & H(S_1) + \sum_{t=2}^{T} H(S_t \mid S_{t-1}) \\ \text{subject to} \quad & S_1, \ldots, S_T \text{ are r.v.'s on } \mathscr{V}, \\ & 0 \to S_1 \to \cdots \to S_T, \quad \mathrm{E}[w(0, S_1) + \textstyle\sum_{t=2}^{T} w(S_{t-1}, S_t)] = Tc. \end{aligned} \tag{2.4}$$

The optimal value of (2.4) is given by $\zeta(\boldsymbol{\eta}^{(T)})$.

Proof. The proof is by the Karush-Kuhn-Tucker (KKT) conditions [4]. (See also Appendix A.)

First we parametrize $(S_1, \ldots, S_T)$. We define $\mathscr{V}_1, \mathscr{E}_t, \mathscr{V}_t$ $(2 \leq t \leq T)$ recursively as follows.

$$\begin{aligned} \mathscr{V}_1 &= \{j \in \mathscr{V} \mid (0, j) \in \mathscr{E}\}, \\ \mathscr{E}_t &= \{(i, j) \in \mathscr{E} \mid i \in \mathscr{V}_{t-1}\} && (2 \leq t \leq T), \\ \mathscr{V}_t &= \{j \in \mathscr{V} \mid (i, j) \in \mathscr{E} \text{ for some } i \in \mathscr{V}_{t-1}\} && (2 \leq t \leq T). \end{aligned}$$

For each $i \in \mathscr{V}_1$ let $p_1(i) = \Pr\{S_1 = i\}$. For each $2 \leq t \leq T$ and $(i, j) \in \mathscr{E}_t$ let $b_t(i, j) = \Pr\{S_t = j \mid S_{t-1} = i\}$. For each $2 \leq t \leq T$ and $j \in \mathscr{V}_t$ let $p_t(j) = \Pr\{S_t = j\}$. Then (2.4) is equivalent to the following problem.

Problem (∗): Maximize

$$z = \sum_{i\in\mathscr{V}_1} -p_1(i)\log p_1(i) + \sum_{t=2}^{T}\sum_{i\in\mathscr{V}_{t-1}} \left\{ p_{t-1}(i) \sum_{j:(i,j)\in\mathscr{E}_t} -b_t(i,j)\log b_t(i,j) \right\}$$

subject to $p_1(i) \geq 0$ (for $i \in \mathscr{V}_1$), $b_t(i,j) \geq 0$ (for $2 \leq t \leq T$, $(i,j) \in \mathscr{E}_t$), and

$$K := -Tc + \sum_{i\in\mathscr{V}_1} p_1(i)w(0,i) + \sum_{t=2}^{T}\left[\sum_{i\in\mathscr{V}_{t-1}} p_{t-1}(i)\sum_{j:(i,j)\in\mathscr{E}_t} b_t(i,j)w(i,j)\right] = 0,$$

$$L := \sum_{i\in\mathscr{V}_1} p_1(i) - 1 = 0,$$

$$M_t(i) := \sum_{j:(i,j)\in\mathscr{E}_t} b_t(i,j) - 1 = 0 \qquad (2 \leq t \leq T,\ i \in \mathscr{V}_{t-1}),$$

$$N_t(j) := \sum_{i:(i,j)\in\mathscr{E}_t} p_{t-1}(i)b_t(i,j) - p_t(j) = 0 \quad (2 \leq t \leq T,\ j \in \mathscr{V}_t).$$

Second, we find all possible interior extrema of the above problem. We can confirm that at any interior point (a feasible point where all inequality constraints are inactive) the gradients of the equality constraints are linearly independent. To see it, suppose there exist real numbers k, l, $m_t(i)$ $(2 \leq t \leq T,\ i \in \mathscr{V}_{t-1})$ and $n_t(j)$ $(2 \leq t \leq T,\ j \in \mathscr{V}_t)$ that satisfy

$$k\nabla K + l\nabla L + \sum_{t=2}^{T}\sum_{i\in\mathscr{V}_{t-1}} m_t(i)\nabla M_t(i) + \sum_{t=2}^{T}\sum_{j\in\mathscr{V}_t} n_t(j)\nabla N_t(j) = \mathbf{0}. \tag{2.5}$$

Observing the system (2.5) componentwise, we get the following equations.

Equation for $\partial/\partial p_T(i)$: For $i \in \mathscr{V}_T$,

$$-n_T(i) = 0.$$

Equation for $\partial/\partial b_t(i,j)$: For $2 \leq t \leq T$ and $(i,j) \in \mathscr{E}_t$,

$$(n_t(j) + kw(i,j))p_{t-1}(i) + m_t(i) = 0.$$

Equation for $\partial/\partial p_t(i)$: For $2 \le t \le T-1$ and $i \in \mathscr{V}_t$,

$$-n_t(i) + \sum_{j:(i,\,j)\in\mathscr{E}_{t+1}} (n_{t+1}(j) + kw(i,\ j))b_{t+1}(i,\ j) = 0.$$

Equation for $\partial/\partial p_1(i)$: For $i \in \mathscr{V}_1$,

$$l + kw(0,\ i) + \sum_{j:(i,\,j)\in\mathscr{E}_2} (n_2(j) + kw(i,\ j))b_2(i,\ j) = 0.$$

From Equation for $\partial/\partial b_t(i,\ j)$, for $2 \le t \le T$ and $(i,\ j) \in \mathscr{E}_t$ we have $n_t(j) + kw(i,\ j) = -\dfrac{m_t(i)}{p_{t-1}(i)}$, so we have that when $i \in \mathscr{V}_{t-1}$ is fixed, then $n_t(j) + kw(i,\ j)$ takes the same value for every j such that $(i,\ j) \in \mathscr{E}_t$. With Equation for $\partial/\partial p_t(i)$ combined, this means that it holds that $n_t(i) = n_{t+1}(j) + kw(i,\ j)$ for $2 \le t \le T-1$ and $(i,\ j) \in \mathscr{E}_{t+1}$. Further, from Equation for $\partial/\partial p_1(i)$, we have for $(i,\ j) \in \mathscr{E}_2$ that $l + kw(0,\ i) = -(n_2(j) + kw(i,\ j))$. As a result, we have for every $(i_1, \ldots, i_T) \in \mathscr{P}_T$ that $l = -k\left(w(0,\ i_1) + \sum_{t=2}^{T} w(i_{t-1},\ i_t)\right) = -kW_T(i_1, \ldots, i_T)$, but this is only possible when $k = 0$, because $|\mathscr{W}_T| \ge 2$. And also we have $l = 0$. Now it is easy to see that every $m_t(i)$ and $n_t(j)$ must be equal to 0. Thus we have confirmed the linear independence.

Let κ, λ, $\mu_t(i)$ and $\nu_t(j)$ be the KKT multipliers of equality constraints K, L, $M_t(i)$ and $N_t(j)$, respectively. The KKT conditions require that at an interior extremum the following conditions be satisfied for some real numbers κ, λ, $\mu_t(i)$ and $\nu_t(j)$.

Condition for $\partial/\partial p_1(i)$: For $i \in \mathscr{V}_1$,

$$-\kappa w(0,\ i) - \lambda - \log p_1(i) - \log e = \sum_{j:(i,\,j)\in\mathscr{E}_2} b_2(i,\ j)\{\kappa w(i,\ j) + \nu_2(j) + \log b_2(i,\ j)\}.$$

Condition for $\partial/\partial p_t(i)$: For $2 \le t \le T-1$ and $i \in \mathscr{V}_t$,

$$\nu_t(i) = \sum_{j:(i,\,j)\in\mathscr{E}_{t+1}} b_{t+1}(i,\ j)\{\kappa w(i,\ j) + \nu_{t+1}(j) + \log b_{t+1}(i,\ j)\}.$$

Condition for $\partial/\partial p_T(i)$: For $i \in \mathscr{V}_T$,

$$\nu_T(i) = 0.$$

Condition for $\partial/\partial b_t(i, j)$: For $2 \leq t \leq T$ and $(i, j) \in \mathscr{E}_t$,

$$-\mu_t(i) = p_{t-1}(i)\{(\kappa w(i, j) + \nu_t(j) + \log b_t(i, j)) + \log e\}.$$

Now in the equation of Condition for $\partial/\partial b_t(i, j)$, we have $p_{t-1}(i) \neq 0$ by assumption, and so the underwaved part of the equation takes the same value for every j. Thus also in each equation of Condition for $\partial/\partial p_1(i)$ and Condition for $\partial/\partial p_t(i)$, the underwaved part takes the same value regardless of j. Combining these facts with $\sum_{j:(i,j)\in\mathscr{E}_t} b_t(i, j) = 1$, we have the following modified version of conditions.

Condition′ for $\partial/\partial p_1(i)$: For $(i_1, i_2) \in \mathscr{E}_2$,

$$-\kappa w(0, i_1) - \lambda - \log p_1(i_1) - \log e = \kappa w(i_1, i_2) + \nu_2(i_2) + \log b_2(i_1, i_2).$$

Condition′ for $\partial/\partial p_t(i)$: For $2 \leq t \leq T-1$, $(i_t, i_{t+1}) \in \mathscr{E}_{t+1}$,

$$\nu_t(i_t) = \kappa w(i_t, i_{t+1}) + \nu_{t+1}(i_{t+1}) + \log b_{t+1}(i_t, i_{t+1}).$$

Condition′ for $\partial/\partial p_T(i)$: For $i_T \in \mathscr{V}_T$,

$$\nu_T(i_T) = 0.$$

So we have for every $\boldsymbol{i} = (i_1, \ldots, i_T) \in \mathscr{P}_T$ that

$$-\lambda - \log e - \kappa W_T(\boldsymbol{i}) = \log p_1(i_1) + \sum_{t=2}^{T} \log b_t(i_{t-1}, i_t). \tag{2.6}$$

When we denote $p(\boldsymbol{i}) = \Pr\{S_1 = i_1, \ldots, S_T = i_T\}$, then the RHS of (2.6) is equal to $\log p(\boldsymbol{i})$.

Thus if we let $a = \exp(-\lambda)/e$ and $\alpha = \exp(-\kappa)$, we have for every $\boldsymbol{i} \in \mathscr{P}_T$ that $p(\boldsymbol{i}) = a\alpha^{W_T(\boldsymbol{i})}$. It is derived from the equality constraints in Problem $(*)$ that $\sum_{\boldsymbol{i}\in\mathscr{P}_T} p(\boldsymbol{i}) = 1$, $\sum_{\boldsymbol{i}\in\mathscr{P}_T} W_T(\boldsymbol{i})p(\boldsymbol{i}) = Tc$. Thus using $\boldsymbol{\eta}^{(T)} = (\eta_u^{(T)})_u$ as defined before, we can write

$$a \sum_{u\in\mathscr{W}_T} \eta_u^{(T)}\alpha^u = 1, \quad a \sum_{u\in\mathscr{W}_T} \eta_u^{(T)} u\alpha^u = Tc. \tag{2.7}$$

From (2.7) we obtain that $g(\boldsymbol{\eta}^{(T)}, \alpha) = 0$, where g is as defined in (2.2), which, regarded as an equation in α, has the unique positive solution $\alpha(\boldsymbol{\eta}^{(T)})$ due to Corollary 2.4. Now a is determined uniquely from $\alpha = \alpha(\boldsymbol{\eta}^{(T)})$, this gives the sole candidate of interior extremum, the value at which is given by

$$\begin{aligned} z^* &= \sum_{\boldsymbol{i}\in\mathscr{P}_T} -p(\boldsymbol{i}) \log p(\boldsymbol{i}) \\ &= \sum_{u\in\mathscr{W}_T} \eta_u^{(T)} \{-a\alpha^u(\log a + u \log \alpha)\} = -\log a - Tc \log \alpha = \zeta(\boldsymbol{\eta}^{(T)}), \end{aligned} \tag{2.8}$$

where ζ is as defined in (2.3).

Finally to confirm that z^* in (2.8) gives the global maximum of Problem $(*)$, we assume by contradiction that z^* does not give the global maximum and hence there exists an optimal solution $\hat{P}$ that have parameters $\hat{p}_t(i)$ $(1 \le t \le T,\ i \in \mathscr{V}_t)$ and $\hat{b}_t(i, j)$ $(2 \le t \le T,\ (i, j) \in \mathscr{E}_t)$ such that one or more inequality constraints are active at $\hat{P}$ and that the value at $\hat{P}$, which we denote by $\hat{z}$, is greater than z^*. We make a modification to Problem $(*)$ so that every parameter variable is removed whose "hatted" value is equal to 0. More precisely, we let

$$\begin{aligned} \hat{\mathscr{V}}_t &= \{i \in \mathscr{V}_t \mid \hat{p}_t(i) > 0\} && (1 \le t \le T), \\ \hat{\mathscr{E}}_t &= \{(i, j) \in \mathscr{E}_t \mid i \in \hat{\mathscr{V}}_{t-1},\ \hat{b}_t(i, j) > 0\} && (2 \le t \le T) \end{aligned}$$

and construct Problem $(\hat{*})$ by replacing $\mathscr{V}_1$, $\mathscr{V}_{t-1}$, $\mathscr{E}_t$, $\mathscr{V}_T$, $\mathscr{V}_t$, K, L, $M_t(i)$ and $N_t(j)$ in Problem $(*)$ with $\hat{\mathscr{V}}_1$, $\hat{\mathscr{V}}_{t-1}$, $\hat{\mathscr{E}}_t$, $\hat{\mathscr{V}}_T$, $\hat{\mathscr{V}}_t$, $\hat{K}$, $\hat{L}$, $\hat{M}_t(i)$ and $\hat{N}_t(j)$, respectively. Note that a feasible solution of

Problem ($\circledast$) is also a feasible solution of Problem ($*$) when every undefined parameter is set to 0, and that they take the same value. It is obvious by assumption that $\hat{P}$ (with parameters properly removed) is a feasible solution of Problem ($\circledast$) where all inequality constraints are inactive. For technical reasons, we let $\hat{\mathscr{P}}_T = \{\boldsymbol{i} \in \mathscr{P}_T \mid p(\boldsymbol{i}) > 0\}$ and consider the following two cases.

Case (a): Consider when $|W_T(\hat{\mathscr{P}}_T)| \geq 2$. In this case, we can confirm the linear independence of the gradients of the equality constraints. An argument on the KKT conditions, which is very similar to what we have previously done, leads that when we define $\hat{\boldsymbol{\eta}} = (\hat{\eta}_u)_{u \in \mathscr{W}_T}$ as $\hat{\eta}_u = \left|W_T^{-1}(\{u\}) \cap \hat{\mathscr{P}}_T\right|$ for each $u \in \mathscr{W}_T$, then $\hat{\boldsymbol{\eta}}$ is Tc-conformant and the value of $\hat{z}$ is given by $\zeta(\hat{\boldsymbol{\eta}})$. As $\hat{\eta}_u \leq \eta_u^{(T)}$ clearly holds for every $u \in \mathscr{W}_T$, if we can show that *(♣): every partial derivative of ζ is nonnegative in* $\hat{\mathrm{H}} = \prod_{u \in \mathscr{W}_T} [\hat{\eta}_u, \eta_u^{(T)}]$, then it is derived that $\hat{z} = \zeta(\hat{\boldsymbol{\eta}}) \leq \zeta(\boldsymbol{\eta}^{(T)}) = z^* < \hat{z}$, which is a contradiction. Now we prove the claim (♣). Note that every $\boldsymbol{\eta} = (\eta_u)_u \in \hat{\mathrm{H}}$ is Tc-conformant. To calculate the partial derivatives of ζ, we first confirm that $\partial\alpha/\partial\eta_u$ exist for every $u \in \mathscr{W}_T$. As $g(\boldsymbol{\eta}, \alpha(\boldsymbol{\eta})) = 0$, for the purpose it is sufficient to show that

$$\frac{\partial g}{\partial \alpha} = \sum_{u \in \mathscr{W}_T} u(u - Tc)\eta_u \alpha^{u-1} = \frac{1}{\alpha} \sum_{u \in \mathscr{W}_T} (u - Tc)^2 \eta_u \alpha^u$$

is not equal to 0, which holds if $\boldsymbol{\eta} \in \hat{\mathrm{H}}$. Now, for each $u \in \mathscr{W}_T$ we have

$$\frac{\partial \zeta}{\partial \eta_u} = \frac{\left(\sum_{v \in \mathscr{W}_T} v \eta_v \alpha^{v-1}\right)(\partial\alpha/\partial\eta_u) + \alpha^u}{\sum_{v \in \mathscr{W}_T} \eta_v \alpha^v} - \frac{Tc}{\alpha}(\partial\alpha/\partial\eta_u) = \frac{\alpha^u}{\sum_{v \in \mathscr{V}} \eta_v \alpha^v},$$

which takes a positive value at every point in $\hat{\mathrm{H}}$, as desired.

Case (b): Consider when $|W_T(\hat{\mathscr{P}}_T)| = 1$. In this case, $v = Tc$ is in $\mathscr{W}_T$ and it holds that $W_T(\hat{\mathscr{P}}_T) = \{v\}$. Now the equality constraint $\hat{K}$ in Problem ($\circledast$) is redundant and so we remove it. Then we can confirm that the gradients of the equality constraints (with $\hat{K}$ removed) are linearly independent. By an argument on the KKT conditions, it is derived that the value of $\hat{z}$ is given by $\log|\hat{\mathscr{P}}_T|$. Now we define $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_u)_u$ as follows.

$$\tilde{\eta}_u = \begin{cases} |\hat{\mathscr{P}}_T| & (\text{if } u = v), \\ (\max \mathscr{W}_T - v)/D & (\text{if } u = \min \mathscr{W}_T), \\ (v - \min \mathscr{W}_T)/D & (\text{if } u = \max \mathscr{W}_T), \\ 0 & (\text{otherwise}), \end{cases}$$

where we let $D = \max \mathscr{W}_T - \min \mathscr{W}_T$. Then we can see that the family $\tilde{\boldsymbol{\eta}}$ is Tc-conformant and $\alpha(\tilde{\boldsymbol{\eta}}) = 1$, and thus we have $\zeta(\tilde{\boldsymbol{\eta}}) = \log(\sum_u \tilde{\eta}_u)$. Now that $\tilde{\eta}_u \leq \eta_u^{(T)}$ holds for every $u \in \mathscr{W}_T$, we can derive, by an argument similar to that in Case (a), that $\zeta(\tilde{\boldsymbol{\eta}}) \leq \zeta(\boldsymbol{\eta}^{(T)})$, and hence $\hat{z} = \log|\hat{\mathscr{P}}_T| < \log(\sum_u \tilde{\eta}_u) \leq \zeta(\boldsymbol{\eta}^{(T)}) = z^* < \hat{z}$, which is a contradiction. □

Proof of Theorem 2.2. Due to Theorem 2.1, $C_{T,c}(\mathscr{V}, \mathscr{E}, w)$ is given by the optimal value of the following optimization problem.

$$\begin{aligned} \text{Maximize} \quad & H(S_1) + \sum_{t=2}^{T} H(S_t \,|\, S_{t-1}) \\ \text{subject to} \quad & S_1, \ldots, S_T \text{ are r.v.'s on } \mathscr{V}, \\ & 0 \to S_1 \to \cdots \to S_T, \quad \mathrm{E}[w(0, S_1) + \textstyle\sum_{t=2}^{T} w(S_{t-1}, S_t)] \leq Tc. \end{aligned} \tag{2.9}$$

We regard $\zeta(\boldsymbol{\eta}^{(T)})$ and $\alpha(\boldsymbol{\eta}^{(T)})$ as functions of c, and denote them by $\bar{\zeta}(c)$ and $\bar{\alpha}(c)$, respectively. Then we have $\bar{\alpha}(c^*) = 1$ for $c^* = \left(\sum_{u \in \mathscr{W}_T} u\eta_u\right) / \left(T \sum_{u \in \mathscr{W}_T} \eta_u\right)$. Furthermore, we regard $g(\boldsymbol{\eta}^{(T)}, \alpha)$ as a function of c and α, and denote it by $\bar{g}(c, \alpha)$. Then, using $\bar{g}(c, \bar{\alpha}(c)) = 0$, we have

$$\begin{aligned} \frac{d\bar{\alpha}}{dc}(c) &= -\frac{\frac{\partial \bar{g}}{\partial c}(c, \bar{\alpha}(c))}{\frac{\partial \bar{g}}{\partial \alpha}(c, \bar{\alpha}(c))} = \frac{T\bar{\alpha}(c) \sum_{u \in \mathscr{W}_T} \eta_u \bar{\alpha}(c)^u}{\sum_{u \in \mathscr{W}_T} (u - Tc)^2 \eta_u \bar{\alpha}(c)^u} > 0, \\ \frac{d\bar{\zeta}}{dc}(c) &= -T \log \bar{\alpha}(c), \end{aligned}$$

and hence $\bar{\zeta}(c)$ is increasing over $(\min \mathscr{W}_T, c^*)$ and decreasing over $(c^*, \max \mathscr{W}_T)$. Now that

$\bar{\zeta}(c^*) = \log|\mathscr{P}_T|$, it is derived that the optimal value for (2.9) is given by

$$z^* = \max_{\min \mathscr{W}_T < c' \leq c} \bar{\zeta}(c') = \begin{cases} \bar{\zeta}(c) & \text{if } c \leq c^*, \\ \bar{\zeta}(c^*) & \text{otherwise,} \end{cases}$$

from which we derive (2.1). □

It is not surprising that the $\log|\mathscr{P}_T|$ in (2.1), which is equal to $C_{T,c}(\mathscr{V}, \mathscr{E}, w)$ when c is sufficiently large, coincides with the value of the maximum sum rate of the corresponding WOM if it exists, which is determined in [3].

Chapter 3

Examples

In this chapter, we give fine examples of the capacity region and the maximum sum rate of WCM's.

As examples of the state transition rule $\mathscr{E}$ and the cost function w of $(\mathscr{V}, \mathscr{E}, w)$-WCM, we adopt $\mathscr{E} = \mathscr{E}^{(1)}$ or $\mathscr{E}^{(2)}$ and $w = w^{(1)}$ or $w^{(2)}$ that are defined as follows.

$$\mathscr{E}^{(1)} = \{(i,\ j) \in \mathscr{V} \times \mathscr{V} \mid i \leq j\},$$

$$\mathscr{E}^{(2)} = \{(i,\ j) \in \mathscr{V} \times \mathscr{V} \mid i = 0 \text{ or } i = j\},$$

$$w^{(1)}(i,\ j) = |j - i|,$$

$$w^{(2)}(i,\ j) = \begin{cases} 0 & \text{if } i = j, \\ 1 & \text{otherwise.} \end{cases}$$

These $\mathscr{E}^{(1)}$ and $\mathscr{E}^{(2)}$ are equivalent to the state transition rules of generalized WOM's that are adopted in examples of [3]. We believe cost functions $w^{(1)}$ and $w^{(2)}$ are practically natural ones. Note that if $q = 2$, then $\mathscr{E}^{(1)} = \mathscr{E}^{(2)}$ and $w^{(1)} = w^{(2)}$. For $q \geq 2$, we denote $\mathscr{V}^{(q)} = \{0,\ 1,\ \ldots,\ q-1\}$.

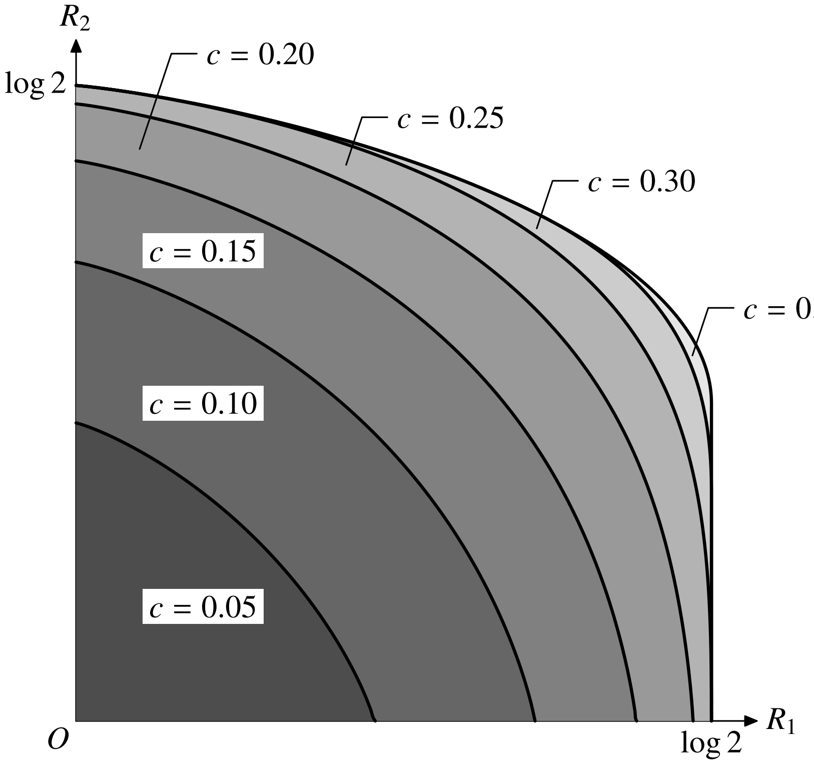

Figure 3.1: The capacity region of $(\mathscr{V}^{(2)}, \mathscr{E}^{(1)}, w^{(1)})$-WCM with 2 writings and c-average cost constraint.

3.1 Capacity Region

Example 3.1. The capacity region of $(\mathscr{V}^{(2)}, \mathscr{E}^{(1)}, w^{(1)})$-WCM with 2 writings and c-average cost constraint is plotted in Figure 3.1 for each $c = 0.05, 0.10, 0.15, 0.20, 0.25, 0.30$ and 0.5.

Note that in the figure, the area hidden by the capacity region with a smaller c is also a part of the capacity region with a larger c. Figure 3.1 is made using the fact that for $0 \le c \le 1/2$ it holds that

$$\mathscr{A}_{2,c}(\mathscr{V}^{(2)}, \mathscr{E}^{(1)}, w^{(1)}) = \Big\{(R_1, R_2) : 0 \le a \le \min\{2c,\ 1/2\}, \quad R_1 = h(a), \\ 0 \le R_2 \le (1-a) \cdot h\Big(\min\Big\{\tfrac{2c-a}{1-a},\ 1/2\Big\}\Big)\Big\},$$

where $h(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function.

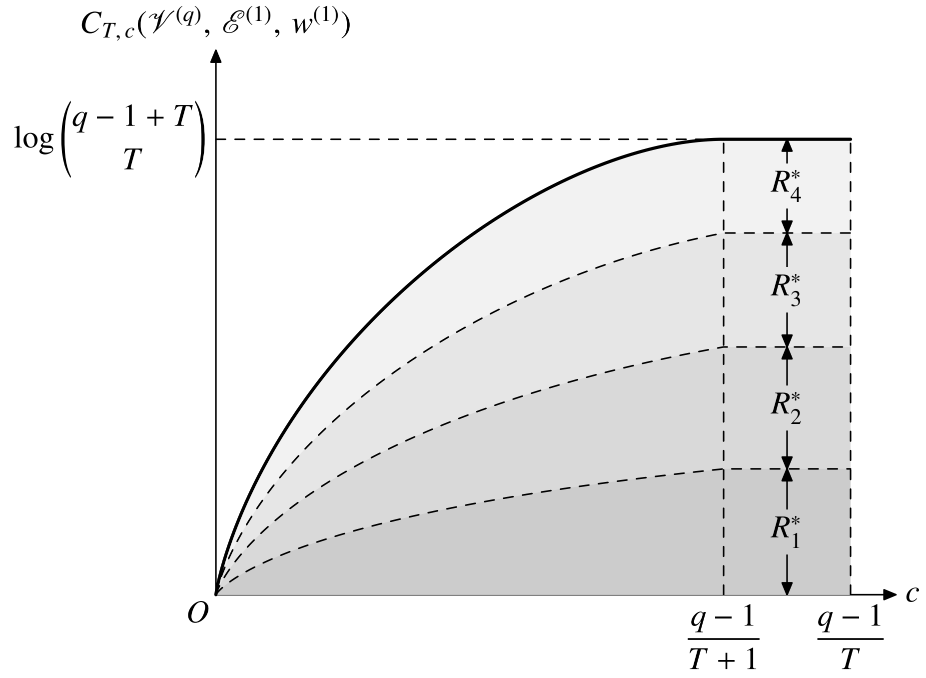

Figure 3.2: The maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(1)})$-WCM with T writings and c-average cost constraint (the plot is for $q = 8,\ T = 4$).

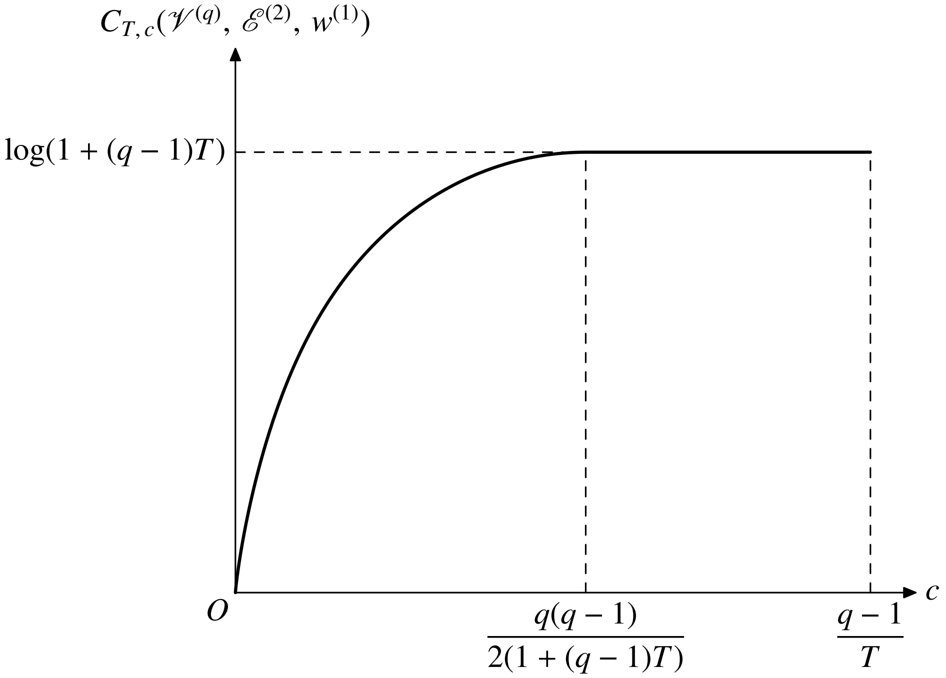

Figure 3.3: The maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(2)}, w^{(1)})$-WCM with T writings and c-average cost constraint (the plot is for $q = 8,\ T = 4$).

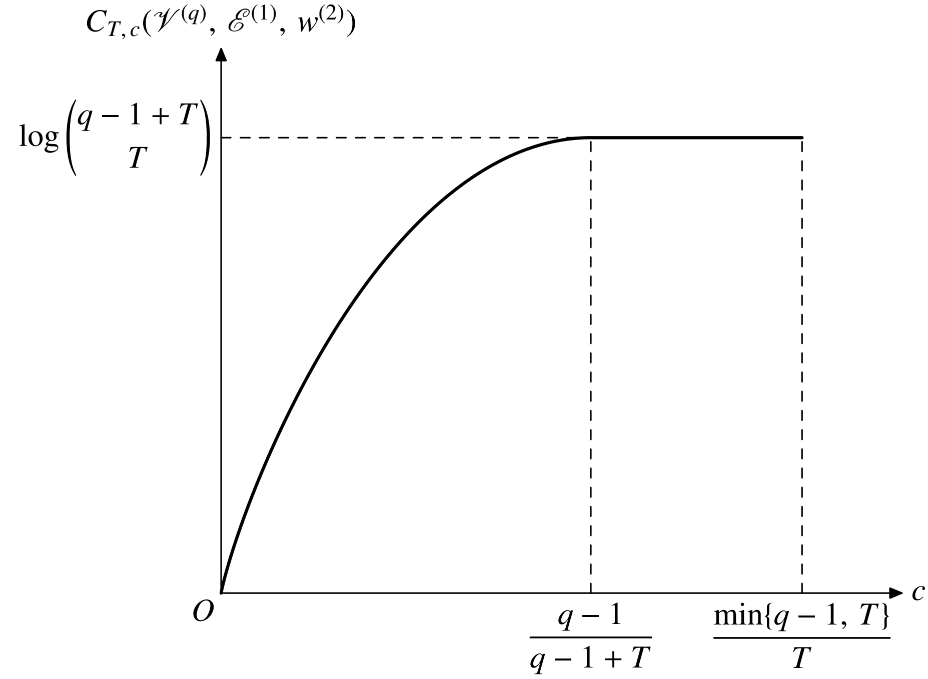

Figure 3.4: The maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(2)})$-WCM with T writings and c-average cost constraint (the plot is for $q = 8,\ T = 4$).

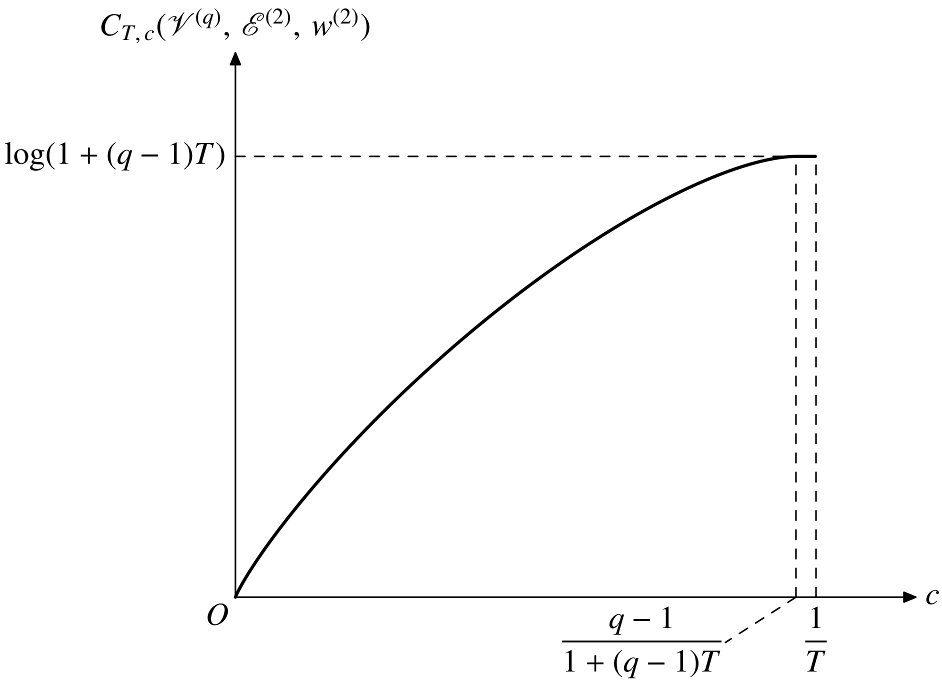

Figure 3.5: The maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(2)}, w^{(2)})$-WCM with T writings and c-average cost constraint (the plot is for $q = 8,\ T = 4$).

3.2 Maximum Sum Rate

Example 3.2. In Figure 3.2, the maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(1)})$-WCM with T writings and c-average cost constraint is plotted for $\min \mathscr{W}_T \le Tc \le \max \mathscr{W}_T$.

In this case, we have $\mathscr{W}_T = \{0, 1, \dots, q-1\}$, and an enumerative argument gives $\eta_u^{(T)} = \binom{u-1+T}{T-1}$ for $u \in \mathscr{W}_T$. Thus we have $\sum_{u \in \mathscr{W}_T} \eta_u^{(T)} = \binom{q-1+T}{T}$, and if we take $c^* := \dfrac{\sum_{u \in \mathscr{W}_T} u\eta_u^{(T)}}{T \sum_{u \in \mathscr{W}_T} \eta_u^{(T)}} = \dfrac{q-1}{T+1}$, then the maximum sum rate attains $\log \binom{q-1+T}{T}$ when $c = c^*$ and remains the same value for $c \ge c^*$.

In Figure 3.2, we also shows $(R_1^*, \dots, R_T^*) \in \mathscr{A}_{T,c}(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(1)})$ that attains the maximum sum rate, that is, $\sum_{t=1}^{T} R_t^* = C_{T,c}(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(1)})$. It seems to hold that $R_1^* > \cdots > R_T^*$, which we have not yet succeeded in proving yet.

Example 3.3. In Figure 3.3, the maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(2)}, w^{(1)})$-WCM with T writings and c-average cost constraint is plotted for $\min \mathscr{W}_T \le Tc \le \max \mathscr{W}_T$.

In this case, we have $\mathscr{W}_T = \{0, 1, \dots, q-1\}$, and, by counting, $\eta_0^{(T)} = 1$ and $\eta_u^{(T)} = T$ for $u \in \mathscr{W}_T \setminus \{0\}$. Thus we have $\sum_{u \in \mathscr{W}_T} \eta_u^{(T)} = 1 + (q-1)T$ and $\dfrac{\sum_{u \in \mathscr{W}_T} u\eta_u^{(T)}}{T \sum_{u \in \mathscr{W}_T} \eta_u^{(T)}} = \dfrac{q(q-1)}{2(1+(q-1)T)}$.

Example 3.4. In Figure 3.4, the maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(1)}, w^{(2)})$-WCM with T writings and c-average cost constraint is plotted for $\min \mathscr{W}_T \le Tc \le \max \mathscr{W}_T$.

In this case, we have $\mathscr{W}_T = \{0, 1, \dots, \min\{q-1, T\}\}$, and an enumerative argument gives $\eta_u^{(T)} = \binom{q-1}{u}\binom{T}{u}$ for $u \in \mathscr{W}_T$. Thus we have $\sum_{u \in \mathscr{W}_T} \eta_u^{(T)} = \binom{q-1+T}{T}$ and $\dfrac{\sum_{u \in \mathscr{W}_T} u\eta_u^{(T)}}{T \sum_{u \in \mathscr{W}_T} \eta_u^{(T)}} = \dfrac{q-1}{q-1+T}$.

Example 3.5. In Figure 3.5, the maximum sum rate of $(\mathscr{V}^{(q)}, \mathscr{E}^{(2)}, w^{(2)})$-WCM with T writings and c-average cost constraint is plotted for $\min \mathscr{W}_T \le Tc \le \max \mathscr{W}_T$.

In this case, we have $\mathscr{W}_T = \{0, 1\}$, and, by counting, $\eta_0^{(T)} = 1$ and $\eta_1^{(T)} = (q-1)T$. Thus we have $\sum_{u \in \mathscr{W}_T} \eta_u^{(T)} = 1 + (q-1)T$ and $\dfrac{\sum_{u \in \mathscr{W}_T} u\eta_u^{(T)}}{T \sum_{u \in \mathscr{W}_T} \eta_u^{(T)}} = \dfrac{q-1}{1+(q-1)T}$.

Chapter 4

Conclusions

In this thesis, we introduced WCM and proved the capacity region and the maximum sum rate of WCM with such a cost constraint as restricts the average cost of rewrites.

It is another problem to construct a WCM code explicitly that satisfies some cost constraint, and it is completely out of the scope of this thesis.

As a future work, we plan to investigate the behavior of the maximum sum rate when q goes to infinity, and to make some proposition about an "optimal" value of q according to a given cost constraint.

Appendix A

The Karush-Kuhn-Tucker Conditions

In the proof of Lemma 2.5, we utilize the Karush-Kuhn-Tucker (KKT) conditions. In this chapter, we give an outline of the KKT conditions to the extent needed in the proof of Lemma 2.5.

Consider the following nonlinear optimization problem.

$$\begin{aligned} \text{Minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h_j(\boldsymbol{x}) = 0 \quad (j = 1, \ldots, m), \\ & g_i(\boldsymbol{x}) \geq 0 \quad (i = 1, \ldots, p). \end{aligned} \tag{A.1}$$

Here, $f : \mathbb{R}^n \to \mathbb{R}$ is called the objective function, $h_j : \mathbb{R}^n \to \mathbb{R}\ \ (j = 1, \ldots, m)$ is called the equality constraints, and $g_i : \mathbb{R}^n \to \mathbb{R}\ \ (i = 1, \ldots, p)$ is called the inequality constraints.

The set of points in $\mathbb{R}^n$ that satisfy the equality and the inequality constraints, which we denote by $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n \mid h_j(\boldsymbol{x}) = 0\ \ (j = 1, \ldots, m),\ g_i(\boldsymbol{x}) \geq 0\ \ (i = 1, \ldots, p)\}$, is called the feasible set of the problem (A.1), and every element in $\mathcal{F}$ is called a feasibile point (or feasible solution) of (A.1).

For a feasible point $\tilde{\boldsymbol{x}} \in \mathcal{F}$ and $1 \leq i \leq p$, if it holds that $g_i(\tilde{\boldsymbol{x}}) = 0$, then we say that the inequality constraint "$g_i(\boldsymbol{x}) \geq 0$" is active at $\tilde{\boldsymbol{x}}$. Conversely, if it holds that $g_i(\tilde{\boldsymbol{x}}) > 0$, then we say that the inequality constraint "$g_i(\boldsymbol{x}) \geq 0$" is inactive at $\tilde{\boldsymbol{x}}$. We denote by $I(\tilde{\boldsymbol{x}})$ the set of every index $1 \leq i \leq p$ such that the inequality constraint "$g_i(\boldsymbol{x}) \geq 0$" is active at $\tilde{\boldsymbol{x}}$, that is,

$$I(\tilde{\boldsymbol{x}}) \stackrel{\triangle}{=} \{i \mid 1 \leq i \leq p,\ g_i(\tilde{\boldsymbol{x}}) = 0\}.$$

We say that the Linear Independence Constraint Qualification (LICQ) is satisfied at $\tilde{\boldsymbol{x}} \in \mathcal{F}$ when gradients $\nabla h_j(\tilde{\boldsymbol{x}})\ \ (j = 1, \ldots, m),\ \nabla g_i(\tilde{\boldsymbol{x}})\ \ (i \in I(\tilde{\boldsymbol{x}}))$ exist and are linearly independent.

Suppose that $f,\ h_j\ \ (j = 1, \ldots, m)$ and $g_i\ \ (i = 1, \ldots, p)$ are all continuously differentiable at $\boldsymbol{x}^* \in \mathcal{F}$. Suppose also that $\boldsymbol{x}^*$ gives a local minimum of the problem (A.1) and that the LICQ is satisfied at $\boldsymbol{x}^*$. Then the KKT conditions say that there exist multipliers $\lambda_i\ \ (i = 1, \ldots, p)$ and $\mu_j\ \ (j = 1, \ldots, m)$ such that

$$\nabla f(\boldsymbol{x}^*) = \sum_{i=1}^{p} \lambda_i \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{m} \mu_j \nabla h_j(\boldsymbol{x}^*),$$

$$\lambda_i \geq 0, \quad \lambda_i g_i(\boldsymbol{x}^*) = 0\ \ (i = 1, \ldots, p).$$

Bibliography

[1] R.L. Rivest and A. Shamir, “How to reuse a write-once memory,” *Inform. and Contr.*, vol. 55, no. 1–3, pp. 1–19, December 1982.

[2] A. Fiat and A. Shamir, “Generalized ‘write-once’ memories,” *IEEE Trans. Inform. Theory*, vol. IT-30, no. 3, pp. 470–480, May 1984.

[3] F. Fu and A.J. Han Vinck, “On the capacity of generalized write-once memory with state transitions described by an arbitrary directed acyclic graph,” *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 308–313, 1999.

[4] M.S. Bazaraa and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, New York: Wiley, 1979.