Beating the Clock

—

An Offline Clock Synchronization Method Inspired by Maximum Likelihood Techniques

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Li Luo

aus Guangzhou

Düsseldorf, September 2014

aus dem Mathematischen Institut
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Florian Jarre
Heinrich-Heine-Universität Düsseldorf

Korreferent: Prof. Dr. Björn Scheuermann
Humboldt-Universität zu Berlin

Tag der mündlichen Prüfung: 21. Oktober 2014

Mein größter Dank gilt Florian Jarre und Björn Scheuermann für Ihre langjährige Betreuung, meinen (ehemaligen) Kollegen am Institut, insbesondere Thomas Davi und Georg Jansing, für die aufschlussreichen Gespräche, sowie Mark Schellhase und Christian Kienle für das mühsame Korrekturlesen.

Nicht zu vergessen, bedanke ich mich bei meinen Eltern für Ihre liebevolle Förderung und bedingungslose Unterstützung.

Abstract

The analysis of network transmission log files is an important tool in the design process of network protocols and devices. As the logs record when which device sent or received a transmission. The very character of network communication makes log files of multiple devices hard to compare, as they suffer from random delays, incomplete transmissions, and inaccurate or non-synchronized clocks. Online clock synchronization protocols are not always applicable, require additional implementation effort, and can not eliminate certain random delays. Thus, offline clock synchronization, a post processing method which corrects the timestamps in log files, is very appealing.

In a local broadcast network a transmission is received by multiple devices. Exploiting this fact, this thesis models the timestamps generation as sampling of an unknown distribution which is parameterized by the inverse of clock functions. This allows us to apply maximum likelihood estimation techniques to create an offline clock synchronization method.

It is shown that the presented estimates are optimizers of linear programs with well-known sparsity structure. Thus, they can be computed efficiently with a tailored interior point method even for a large amount of samples. Further, under some weak regularity assumptions, the thesis establishes the strong consistency of the presented estimator. That is, the estimates converges almost surely to the inverse clock functions. The simulation displays the predicted asymptotic as expected. Even more, it indicates that consistency might remain valid under more general assumptions.

Zusammenfassung

Die Analyse von Übertragungslogdateien eines Netzwerkes ist ein wichtiges Werkzeug in der Entwicklung von neuen Netzwerkprotokollen und vernetzten Geräten. Logdateien speichern unter anderem zu welchen Zeitpunkten Übertragungen statt finden. Aber Logdateien von unterschiedlichen Geräten sind schlecht vergleichbar, da die gespeicherten Zeitstempeln durch zufällige Verzögerungen, unvollständigen Übertragungen, und ungenaue Uhren verfälscht werden. Online Synchronisierungsprotokollen sind nicht immer anwendbar, bereiten zusätzlichen Implementierungsaufwand, und können gewisse zufällige Verzögerungen nicht korrigieren. Aus diesen Gründen kann ein Offline Synchronisierungsverfahren, welches die Zeitstempeln im Nachhinein korrigiert, viel ansprechender sein.

In vielen Netzwerkarten erfolgt eine Übertragungen an alle benachbarten Geräten. Die vorliegende Arbeit nutzt diese Eigenschaft aus, um die Erzeugung der Zeitstempeln als das Ziehen einer Zufallstichprobe einer unbekannten Verteilung zu interpretieren. Wenn die Uhren als Funktionen modelliert werden, dann können die Inverse dieser Funktionen auf natürlicher Weise als Parameter dieser unbekannten Verteilung interpretiert werden. Dies erlaubt uns die inversen Uhren mit Hilfe von Schätzverfahren zu schätzen. Die vorliegende Arbeit verwendet hierfür einen Maximum-Likelihood Ansatz, um ein effizientes offline Synchronisierungsverfahren herzuleiten.

Es zeigt sich, dass die Schätzungen die Lösungen von dünnbesetzten linearen Programmen sind, die sich mit Hilfe eines speziell angepassten Innere-Punkte-Verfahrens unter Ausnutzung dieser Nullstruktur effizient berechnen

lassen, sogar für große Stichproben. Unter schwachen Regularitätsannahmen, zeigt die vorliegende Arbeit, dass das Schätzverfahren stark konsistent ist; das heißt, die Schätzungen konvergieren fast sicher gegen die inversen Uhren. Die Simulationen zeigen genau das vorhergesagte asymptotische Verhalten. Tatsächlich, gibt sie Anlass zu vermuten, dass das Schätzverfahren unter allgemeineren Annahmen ebenfalls konsistent ist.

Contents

Contents

Symbols

Set of numbers and indexed families:

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	positive integers, integers, and reals
$\mathbb{Z}_+, \mathbb{R}_+$	non-negative integers and reals
S^n, S^I	n-tuples and I-indexed families with values in some set S

Intervals with boundaries $\boldsymbol{a}, \boldsymbol{b} \in (\mathbb{R} \cup \{\pm\infty\})^n$:

$(\boldsymbol{a}, \boldsymbol{b}), [\boldsymbol{a}, \boldsymbol{b}]$	open, closed
$[\boldsymbol{a}, \boldsymbol{b}), (\boldsymbol{a}, \boldsymbol{b}]$	left-closed right-open, left-open right-closed

Sets of functions from M into N for some metric spaces M, N:

$\mathcal{C}(M, N)$	continuous functions
$\mathcal{L}ip(M, N)$	Lipschitz continuous functions

Spaces of function from a suitable subset $S \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$:

$\mathcal{C}^k(S, \mathbb{R}^m)$	functions with continuous k-the derivative
$\mathcal{C}^{k,\alpha}(S, \mathbb{R}^m)$	functions with α Hölder continuous k-th derivative

Spaces of order p integrable functions:

$\mathcal{L}^p(\mu, \mathbb{R}^m)$	from $\mathbb{R}^n$ into $\mathbb{R}^m$ with respect to a Borel measure μ
$\mathcal{L}^p(S, \mathbb{R}^m)$	from $S \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$ with respect to the Lebesgue measure
$\lVert\cdot\rVert_p, \lVert\cdot\rVert_{p,S}$	$\mathcal{L}^p$-norm, and restricted onto some subset S of the domain

Constants and named functions:

$\mathrm{e} = \exp(1)$	Euler's number
$\mathrm{id}_S, \mathbb{1}_S$	identity on S and indicator of S for some set S
$\boldsymbol{\lambda}, \boldsymbol{\lambda}_S$	the Lebesgue measure on $\mathbb{R}^n$ and on a measurable subset $S \subseteq \mathbb{R}^n$
P, E	probability and expected value

Direct product and separated functions:

$f \times g$	direct product of some functions f and g
$\mathcal{S}ep_m$	set of direct products of functions from $\mathbb{R}$ into $\mathbb{R}$ with m factors, the set of separated functions in $\mathbb{R}^m$

Extrem values of a function $f : D \to \mathbb{R}$ on a subset $S \subseteq D$:

$\inf_S f, \sup_S f$	infimum and supremum in $\mathbb{R} \cup \{\pm\infty\}$
$\min_S f, \max_S f$	minimum and maximum in $\mathbb{R}$
$\arg\min_S f, \arg\max_S f$	set of minimizer and maximizer, possibly empty

Chapter 1:

$\boldsymbol{Z} = (Z_1, \dots, Z_m)$	(censored) timestamp model
$\breve{\boldsymbol{u}} = \breve{u}_1 \times \cdots \times \breve{u}_m$	solution of the offline synchronization problem
$F_1, \dots, F_m$	(conditional) distribution functions of $Z_1, \dots, Z_m$
$Q_1, \dots, Q_m$	(conditional) quantile functions of $Z_1, \dots, Z_m$
$\boldsymbol{Q}_0 = (Q_1, \dots, Q_m)$	vector valued (conditional) quantile function of $\boldsymbol{Z}$
Q_0	average of $Q_1, \dots, Q_m$
P_0	distribution of $\boldsymbol{Z}$
P_n	sample distribution of P_0 with n samples
$\mathcal{E}$	$\mathcal{S}ep_m \cap \mathcal{L}^1(P_0, \mathbb{R}^m)$ itself or a proper subspace of it
T	delay model for $\boldsymbol{Z}$

Chapter 2:

$Q_{n,1}, \dots, Q_{n,m}$	sample (conditional) quantile functions of size n
$\boldsymbol{Q}_n = (Q_{n,1}, \dots, Q_{n,m})$	vector-valued sample (conditional) quantile function
$\mathcal{F}_n$	finite dimensional subset of $\mathcal{E}$

Chapter 3:

$\Phi'(\mathbf{0})[\boldsymbol{h}]$	directional derivative of Φ at $\mathbf{0}$ in direction $\boldsymbol{h}$
$\partial_x f(a, \dots)$	partial derivative of function f with respect to the parameter x at $(a, \dots)$

Chapter 1

Introduction

During the design process of network devices and communication protocols it is necessary to repeatedly simulate different use cases. Each run produces large event logs. The analysis of these logs is crucial to discover and to understand unexpected behavior and flawed designs. Intrinsic to network communication these logs suffer from random delays, drop outs, and inaccurate clocks, which complicate the analysis. Online clock synchronization protocols may gravely interfere with an experiment, depend on a stable network, and are unable to handle certain processing delays.

Exploiting the local broadcast character of certain networks, two efficient offline clock synchronization methods using an affine linear clock model based on maximum likelihood estimation and least squares estimation are introduced by Scheuermann et al. [22] and Jarre et al. [11] respectively. This thesis extends the maximum likelihood approach to non-linear clocks. The estimate is the solution of a sparse linear program with well-known structure, which can be solved efficiently by a customized Mehrotra predictor-corrector interior point method [17]; and the estimator is strongly consistent under weak assumptions.

1.1 A Motivation

An experiment within a computer network (or more general a system of networked devices) typically results in a huge set of event logs, a log entry

for each transmission event and a log for each involved device. Each log entry records a timestamp along with other parameters of interest. The analysis of these logs is crucial for the interpretation of the experimental results. Depending on the reliability of the medium and on the network connectivity, these logs are inherently incomplete: Not every device receives each transmission; multiple logs are necessary to reconstruct the complete experiment. Timestamps, based on the local clock on each device, suffer from random delays and non synchronized clocks though, which renders the logs almost incomparable between two devices.

Employing more accurate clocks is not always feasible, due to design constraints like manufacturing cost or power demand. *Global Positioning System* (GPS) disciplined oscillators for example provide highly accurate clocks comparing to simple crystal oscillators [29]. However, they also require more energy and cost much more. Not to mention, GPS reception is poor in buildings, in tunnels, under water, and so on. Thus, employing GPS receivers in huge numbers may conflict with other design decisions.

Keeping inaccurate clocks synchronized by constantly correcting them is much easier and more inexpensive in a network, a strategy deployed in mobile phones and networked computers for a long time. This is achieved by using Mill's well-known *Network Time Protocol* (NTP) [18], which is implemented in most modern computer operating systems. As NTP causes network traffic, employing NTP on the same transmission medium as the one of the proper experiment might cause unwanted side-effects and render the results useless. Also, NTP performs rather poorly in an unreliable network, where the transmission delay between two devices is not symmetric or varies frequently.

Ignoring the individual drawbacks of the methods mentioned above, both can not eliminate random delays between registering an event and time stamping, due to a device's internal working. Instead of altering the setup to improve the time measurement, the offline clock synchronization method presented in this thesis estimates the correct timestamps from event logs after an experiment. Such post processing has zero side-effects during an experiment and is ready to use without additional setup.

In local broadcast networks transmissions are usually received and recorded by multiple neighboring systems. Using such receptions as *anchors*, two efficient offline clock synchronization methods with strong theoretical properties

were introduced in [22] and [11]. Both methods model the clocks as affine linear functions. As clocks built into computers are nearly affine linear under constant conditions for a short time period, this approach shows high accuracy for logs that span up to 20 minutes. However, for longer time periods, changing environmental conditions leads to non-linear clock deviations. Therefore, generalizing the aforementioned approaches to a general clock model is a natural evolution.

1.2 Stochastic Formulation

A clock is a device which counts how often a specific time period is elapsed. If we take a date as a real number (for example as seconds passes since some reference date), then we can model a clock as a step function from $\mathbb{R}$ into $\mathbb{R}$. However, modeling clocks as continuous functions is much more appealing and is sufficient for most practical purposes. In particular, it is reasonable to assume that the clock functions are increasing homeomorphisms. For convenience, we say a clock is *correct* if it is the identity function on $\mathbb{R}$. However, in reality we can only compare clocks to other clocks. Thus, it would be more precise to say a clock is correct with respect to a reference clock.

The generation of timestamps can be modeled as follows. An event k happened, say at date $t_k \in \mathbb{R}$ (with respect to the correct clock), which is observed by device j after a delay $d_{j,k} \geq 0$, and is logged with the timestamp

$$z_{j,k} = C_j(t_k + d_{j,k})$$

using the clock C_j of device j. Notice that we are interest into the unknown data C_j and t_k. However, only the skewed timestamp $z_{j,k}$ is recorded. The goal of offline clock synchronization is to reconstruct C_j, and in process also t_k, for given multiple skewed timestamps.

As the occurrence of an event and the delays are non-deterministic, the timestamps and the process of their generation shall be modeled by stochastic means. In particular, a timestamp *is* a random number in a natural way. The timestamp generation is just sampling. A tuple of timestamps of the same

Homeomorphism: A continuous and bijective function with continuous inverse.

event is merely a vector of correlated random numbers. Even if the network devices are identical, the timestamps are not identically distributed due to clock deviations. Thus, *synchronization* is transforming the timestamp vector to an identically distributed random vector. We introduce the following formal definition:

1.1 Definition: *For $m \in \mathbb{N}$ let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ be a random vector in $\mathbb{R}^m$. The* (offline) synchronization problem *with the* timestamp model $\boldsymbol{Z}$ *is to find increasing homeomorphisms $u_1, u_2, \dots, u_m : \mathbb{R} \to \mathbb{R}$ such that*

1. *$u_1(Z_1), u_2(Z_2), \dots, u_m(Z_m)$ are identically distributed, and*
2. *the inverse functions $u_1^{-1}, u_2^{-1}, \dots, u_m^{-1}$ are the identity function in average:*

$$\frac{1}{m} \sum_{j=1}^{m} u_j^{-1} = \mathrm{id}_{\mathbb{R}} .$$

Then, the direct product $\boldsymbol{u} = u_1 \times u_2 \times \cdots \times u_m : \mathbb{R}^m \to \mathbb{R}^m$, *defined by*

$$\boldsymbol{u}(z_1, z_2, \dots, z_m) = (u_1(z_1), u_2(z_2), \dots, u_m(z_m)),$$

is called a solution *of $\boldsymbol{Z}$.*

Definition 1.1 reflects a network setup with m devices. The random vector $\boldsymbol{Z}$ models a vector of timestamps of one event, where Z_j models a timestamp transformed by the local clock at device j. A recorded timestamp vector is merely a realization of $\boldsymbol{Z}$, and the timestamp generation process is merely sampling $\boldsymbol{Z}$. Using a solution we can synchronize the skewed timestamps, make them comparable. In fact, we may interpret $u_1, u_2, \dots, u_m$ as the inverse functions of the clock functions. As we assume that a clock maps the entire $\mathbb{R}$ homeomorphic onto $\mathbb{R}$, and that the timestamps are supported on the entire $\mathbb{R}$, it makes sense also to assume that $u_1, u_2, \dots, u_m$ map onto the entire $\mathbb{R}$.

Given increasing homeomorphisms $u_1, u_2, \dots, u_m : \mathbb{R} \to \mathbb{R}$ satisfying Item 1 of Definition 1.1 and an increasing homeomorphism $\psi : \mathbb{R} \to \mathbb{R}$, the compositions $\psi \circ u_1, \dots, \psi \circ u_m$ also satisfy Item 1. Thus, to identify the correct global scale, Item 2 of Definition 1.1 is crucial. It can be replaced by a more general one of the form

$$\forall x \in \mathbb{R} : \phi(u_1^{-1}(x), \dots, u_m^{-1}(x)) = x,$$

where $\phi : \mathbb{R}^m \to \mathbb{R}$ is continuous and satisfies the following increase conditions: 1. $\phi(\boldsymbol{z}) \to +\infty$ if $\min_{1\leq j\leq m} z_j \to \infty$, 2. $\phi(\boldsymbol{z}) \to -\infty$ if $\max_{1\leq j\leq m} z_j \to -\infty$, and 3. for each $\boldsymbol{z}, \tilde{\boldsymbol{z}} \in \mathbb{R}^m$ with $\boldsymbol{z} \leq \tilde{\boldsymbol{z}}$ and $\boldsymbol{z} \neq \tilde{\boldsymbol{z}}$ it follows $\phi(\boldsymbol{z}) < \phi(\tilde{\boldsymbol{z}})$. For example, if the first device has a GPS receiver and its clock is assumed to be correct, then we can synchronize to that reference by using $\boldsymbol{z} \mapsto \phi(\boldsymbol{z}) := z_1$. Neither the algorithms nor the theoretical results of this thesis rely on the exact form of ϕ. Hence, ϕ is fixed to the arithmetic mean for the sake of simplicity.

Notice that Definition 1.1 can be extended to model incomplete log-sets by allowing $Z_1, Z_2, \dots, Z_m$ taking ∞ as value to signal a non-existent timestamp. For notational convenience, we will concentrate on complete log-sets first, and then extend all results to incomplete log-sets.

Finally, the main result of this section shows that the offline synchronization problem admits a unique solution under mild regularity assumptions on the marginal distribution of the timestamp model:

1.2 Theorem:
Let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ be a random vector in $\mathbb{R}^m$. For each $j \in \{1, \dots, m\}$ assume that the marginal distribution function $F_j : \mathbb{R} \to [0, 1]$ of Z_j is a homeomorphism from $\mathbb{R}$ onto $(0, 1)$. Define $Q_0 : (0, 1) \to \mathbb{R}$ by

$$Q_0 := \frac{1}{m} \sum_{j=1}^{m} F_j^{-1}.$$

Then, the offline synchronization problem with the timestamp model $\boldsymbol{Z}$ has exactly one solution $\breve{\boldsymbol{u}} = \breve{u}_1 \times \cdots \times \breve{u}_m$, and for each $j \in \{1, \dots, m\}$ the function $\breve{u}_j$ is given by

$$\breve{u}_j = Q_0 \circ F_j.$$

Proof: Notice that Q_0 is continuous, strictly increasing, and surjective, as $Q_0(\tau) \to -\infty$ for $\tau \to 0$ and $Q_0(\tau) \to \infty$ for $\tau \to 1$. Thus, if we define $\breve{\boldsymbol{u}}$ as above, then $\breve{\boldsymbol{u}}$ is a solution of the synchronization problem by construction.

Let $u_1 \times u_2 \times \cdots \times u_m$ be any solution of the synchronization problem and let $F_0 : \mathbb{R} \to [0, 1]$ denote the common marginal distribution function of $u_1(Z_1), u_2(Z_2), \dots, u_m(Z_m)$. Then, for each $j \in \{1, 2, \dots, m\}$ and $x \in \mathbb{R}$ we have

$$F_0(x) = \mathbf{P}\big\{u_j(Z_j) \leq x\big\} = \mathbf{P}\big\{Z_j \leq u_j^{-1}(x)\big\} = F_j(u_j^{-1}(x)).$$

Thus, it follows

$$\mathrm{id}_{\mathbb{R}} = \frac{1}{m}\sum_{j=1}^{m} u_j^{-1} = \frac{1}{m}\sum_{j=1}^{m} F_j^{-1} \circ F_0 = Q_0 \circ F_0$$

and

$$u_j = Q_0 \circ F_j. \qquad \square$$

Although, an experiment lasts for a finite time period only, a small gaussian perturbation would make the timestamps unbounded in theory. Thus, it is a rather weak, and also technically convenient, assumption that the marginal distribution functions $F_1, F_2, \dots, F_m$ are strictly increasing and continuous on the entire real line. From a practical point of view, we may assume that the probability is mostly concentrated on some compact interval.

In the light of Theorem 1.2, the remaining thesis will impose some regularity conditions to ensure that the synchronization problem has a unique solution. Also, we require integrability for distribution estimation.

1.3 Assumption: *Assume that the timestamp model* $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$

- *admits an almost everywhere positive probability density function* $\rho_{\boldsymbol{Z}}$,
- *and is integrable.*

The regularity condition in Assumption 1.3 ensures that the marginal distribution functions $F_1, F_2, \dots, F_m$ of $\boldsymbol{Z}$ are homeomorphisms from $\mathbb{R}$ onto $(0, 1)$. In particular, the synchronization problem has exactly one solution $\breve{\boldsymbol{u}} = \breve{u}_1 \times \breve{u}_2 \times \cdots \times \breve{u}_m$, as stated in Theorem 1.2. Further, F_j, $\breve{u}_j$, and their inverse functions are locally absolutely continuous for each $j \in \{1, 2, \dots, m\}$.

Integrable random vector: The expectation of the norm is finite.

Finally, $u_j(Z_j)$ is also integrable for each $j \in \{1, 2, \dots, m\}$, as we have

$$\begin{aligned}\mathbf{E}\big[|\check{u}_j(Z_j)|\big] &= \int |\check{u}_j(z_j)| \,\mathrm{d}\left(\boldsymbol{\lambda}_{(0,1)} \circ F_j\right)(z_j) \\ &= \int_0^1 |\check{u}_j(F_j^{\text{-}1}(\tau))| \,\mathrm{d}\tau \\ &= \int_0^1 \left|\frac{1}{m}\sum_{j'=1}^{m} F_{j'}^{\text{-}1}(\tau)\right| \mathrm{d}\tau \\ &\leq \frac{1}{m}\sum_{j'=1}^{m} \int_0^1 |F_{j'}^{\text{-}1}(\tau)| \,\mathrm{d}\tau \leq \frac{1}{m}\sum_{j'=1}^{m} \mathbf{E}\big[|Z_{j'}|\big] < \infty.\end{aligned}$$

Here, we exploit the fact that the distribution of Z_j is given by $\boldsymbol{\lambda}_{(0,1)} \circ F_j$, the push-forward measure of the Lebesgue measure $\boldsymbol{\lambda}_{(0,1)}$ on $(0,1)$ with respect to $F_j^{\text{-}1}$. In an analogous fashion, we obtain

$$\mathbf{E}\big[\check{u}_j(Z_j)\big] = \frac{1}{m}\sum_{j'=1}^{m} \mathbf{E}\, Z_{j'}.$$

In applications, the marginal distribution functions are unknown. Thus, the results of this section are of purely theoretical interest. However, this section provides some key ideas to establish an approximation scheme for the offline synchronization problem in the following sections.

1.3 Distribution Estimation

If the distribution of the timestamp model is known, the solution of the offline synchronization problem is readily given by Theorem 1.2. Hence, the offline synchronization problem can be interpreted as a distribution estimation problem. Fix a probability measure P_0 on $\mathbb{R}^m$ and a *sample sequence* $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$ of P_0, that is an independent sequence of random variables in $\mathbb{R}^m$ distributed according to P_0. There are well-known non-parametric and parametric estimation techniques to reconstruct P_0 using $(\boldsymbol{X}_n)_{n\in\mathbb{N}}$. In fact, both techniques are used to solve the offline synchronization problem efficiently.

An example for non-parametric techniques is the so-called sample distribution. The *sample distribution* P_n of the first $n \in \mathbb{N}$ samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$

is defined by

$$P_n(A) = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{\delta}_{\boldsymbol{X}_k}(A).$$

Here, $\boldsymbol{\delta}_{\boldsymbol{x}}$ denotes the Dirac measure centered at $\boldsymbol{x} \in \mathbb{R}^m$. Noteworthy, $P_n(A)$ is itself a random number for each Borel set $A \subseteq \mathbb{R}^m$ and P_n is a so-called random measure. The well-known Glivenko-Cantelli theorem (compare to the literature like [27]) states that

$$\sup_{A \in \mathscr{A}} |P_n(A) - P_0(A)|$$

converges almost surely to 0, where $\mathscr{A} = \{(-\infty, \boldsymbol{a}] \mid \boldsymbol{a} \in \mathbb{R}^m\}$ denotes the set of closed right-bounded intervals on $\mathbb{R}^m$. That is, the distribution function of P_n converges uniformly to the distribution function of P_0 almost surely. If P_0 admits a continuous Lebesgue density ρ_0, there also exist many non-parametric methods to estimate ρ_0 instead of P_0 itself, like the so-called kernel density estimation [21].

Additionally assume that P_0 is a member of a known family $(P_\theta)_{\theta \in \Theta}$ of probability measures on $\mathbb{R}^m$ with a parameter in a non-empty set Θ. Then, exploiting the parametrization $\theta \mapsto P_\theta$ leads to parametric estimation techniques. A well-known example is the *maximum likelihood estimation*: Assume that P_θ admits a probability density ρ_θ for each $\theta \in \Theta$. The *maximum likelihood estimate* maximizes the expected value of the *log-likelihood* function

$$\theta \mapsto P_n[\log \rho_\theta] := \int \log \rho_\theta(\boldsymbol{x}) \, \mathrm{d}P_n(\boldsymbol{x}) = \frac{1}{n} \sum_{k=1}^{n} \log \rho_\theta(\boldsymbol{X}_k)$$

with respect to the sample distribution P_n. Under additional assumptions, every sequence of maximizers converges almost surely to the parameter of P_0 (compare to the literature, like [9]).

In the context of offline synchronization, estimating the distribution of the timestamp model directly using non-parametric methods leads to slow convergence and low accuracy. By regarding the solution of the synchronization problem as a parameter, we can also apply parametric methods. Combining both techniques, we can provide a consistent and efficient estimator.

1.4 Estimation for Offline Synchronization

For this section fix a timestamp model $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ which satisfies Assumption 1.3. For each $j \in \{1, 2, \dots, m\}$ let Q_j denote the *quantile function*, the inverse of the marginal distribution function, of Z_j and let

$$Q_0 = \frac{1}{m} \sum_{j=1}^{m} Q_j.$$

Let P_0 denote the distribution of $\boldsymbol{Z}$ and $\check{\boldsymbol{u}} = \check{u}_1 \times \check{u}_2 \times \cdots \times \check{u}_m$ the solution of $\boldsymbol{Z}$. For the sake of brevity, denote the set of $\mathbb{R}^m$-valued P_0-integrable functions by $\mathcal{L}^1(P_0, \mathbb{R}^m)$, the set of *separated functions* from $\mathbb{R}^m$ into $\mathbb{R}^m$ by,

$$\mathcal{S}ep_m = \{u_1 \times \cdots \times u_m \mid u_1, \dots, u_m : \mathbb{R} \to \mathbb{R}\},$$

and the set of separated P_0-integrable functions by

$$\mathcal{E} = \mathcal{S}ep_m \cap \mathcal{L}^1(P_0, \mathbb{R}^m).$$

Notice that $\mathcal{E}$ is a closed subspace of $\mathcal{L}^1(P_0, \mathbb{R}^m)$ and contains $\check{\boldsymbol{u}}$.

1.4.1 Error and Quantile Ansatz

The mean $\mathcal{L}^1$-error of any $u_1 \times u_2 \times \cdots \times u_m \in \mathcal{E}$ is given by

$$\frac{1}{m} \sum_{j=1}^{m} \mathsf{E}\left[|u_j(Z_j) - \check{u}_j(Z_j)|\right] = \frac{1}{m} \sum_{j=1}^{m} \int_0^1 |u_j(Q_j(\tau)) - Q_0(\tau)| \, \mathrm{d}\tau.$$

The right-hand side does not contain the unknown solution $\check{\boldsymbol{u}}$ and only requires the quantile functions $Q_1, \dots, Q_m$. Essentially, we obtain a *quantile ansatz* by minimizing the mean $\mathcal{L}^1$-error with sample quantile functions in place of quantile functions.

Simulation indicates that the quantile ansatz performs reasonably well on complete log-sets. Unfortunately, this ansatz exhibits substantially lower accuracy if the log-sets are incomplete. Noteworthy, the quantile ansatz estimates the marginal distributions of the timestamps $\boldsymbol{Z}$ only. It ignores any correlation between the components of $\boldsymbol{Z}$. The following maximum likelihood ansatz estimates the joint distribution and exploits the correlations in a canonic fashion.

1.4.2 Maximum Likelihood Ansatz

To apply maximum likelihood techniques, we need a parametric density for $\boldsymbol{Z}$. Here, the notation of the so-called synchronized timestamp model becomes useful, whose existence is imposed in the remaining part of this thesis:

1.4 Assumption: *Additionally to Assumption 1.3: There exists a random number T, called a* synchronized timestamp model *T of $\boldsymbol{Z}$, satisfying the following conditions:*

- *$(T, \boldsymbol{Z})$ admits a probability density;*
- *T is integrable and its density ρ_T is almost everywhere positive;*
- *the conditional density of $\breve{\boldsymbol{u}}(\boldsymbol{Z})$ given T is given by*

$$\rho_{\breve{\boldsymbol{u}}(\boldsymbol{Z})|T}(\boldsymbol{x}|t) = \prod_{j=1}^{m} \rho_\theta(x_j - t),$$

where ρ_θ denotes the density of the exponential distribution with mean $\theta^{-1} > 0$:

$$\rho_\theta(x) = \begin{cases} \theta \exp(-\theta x), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Notice that the random number T models the date at which the event occurs (with respect to a reference clock), and $\breve{u}_j(Z_j) - T$ models the delay at device j. Here, the delays are independent and identically exponentially distributed. Remember that we like to estimate the solution $\breve{\boldsymbol{u}}$ of the synchronization problem. However, it is not a typical regression setting, as we have only samples $\boldsymbol{Z}$ but no samples for T nor for $\breve{\boldsymbol{u}}(\boldsymbol{Z})$.

Under Assumption 1.4: Integration by substitution yields the conditional density of $\boldsymbol{Z}$ given T

$$\rho_{\boldsymbol{Z}|T}(\boldsymbol{z}|t; \breve{\boldsymbol{u}}) = \prod_{j=1}^{m} \rho_\theta(\breve{u}_j(z_j) - t) \cdot \breve{u}_j'(z_j),$$

where $\breve{\boldsymbol{u}}$ is considered as a parameter. Further, the density of $\boldsymbol{Z}$ is given by

$$\rho_{\boldsymbol{Z}}(\boldsymbol{z}; \breve{\boldsymbol{u}}) = \int_{\mathbb{R}} \prod_{j=1}^{m} \rho_\theta(\breve{u}_j(z_j) - t) \cdot \breve{u}_j'(z_j) \cdot \rho_T(t) \, \mathrm{d}t.$$

In general ρ_T is unknown and $\rho_{\boldsymbol{Z}}$ does not have a known analytic form. Also, we have no samples for T nor $\breve{\boldsymbol{u}}(\boldsymbol{Z})$. Thus, we can not estimate ρ_T nor $\rho_{\boldsymbol{Z}}$. In particular, we can not apply maximum likelihood techniques to $\rho_{\boldsymbol{Z}}$ directly. However, we can apply maximum likelihood techniques to $\rho_{\boldsymbol{Z}|T}$, which leads to *conditional maximum likelihood estimation*.

The *conditional log-likelihood* of $\boldsymbol{Z}$ given T with parameter $\boldsymbol{u} = u_1 \times u_2 \times \cdots \times u_m \in \mathcal{E}$, where $u_1, u_2, \dots, u_m$ are strictly increasing and locally absolutely continuous, is given for almost every $\boldsymbol{z} \in \mathbb{R}^m$ by

$$\log \rho_{\boldsymbol{Z}|T}(\boldsymbol{z}|t;\boldsymbol{u}) = \sum_{j=1}^{m} \log \rho_\theta(u_j(z_j) - t) + \sum_{j=1}^{m} \log u_j'(z_j).$$

The density of the exponential distribution, ρ_θ, is strictly decreasing on its support. Thus, for fixed $\boldsymbol{z}$ and $\boldsymbol{u}$, the function $t \mapsto \log \rho_{\boldsymbol{Z}|T}(\boldsymbol{z}|t;\boldsymbol{u})$ is strictly increasing on its proper domain and admits the unique maximum point

$$t^*(\boldsymbol{z};\boldsymbol{u}) = \min_{1 \le j \le m} u_j(z_j).$$

Further, it is well-known that the maximum likelihood estimate to θ is given by the reciprocal mean

$$\theta^*(\boldsymbol{z};\boldsymbol{u}) = \left(\frac{1}{m} \sum_{j=1}^{m} u_j(z_j) - \min_{1 \le j \le m} u_j(z_j) \right)^{-1}.$$

Thus, we conclude

$$\begin{aligned} &\sup_{\theta>0} \sup_{t \in \mathbb{R}} \log \rho_{\boldsymbol{Z}|T}(\boldsymbol{z}|t;\boldsymbol{u},\theta) \\ &\qquad \le -m \log \left(\frac{1}{m} \sum_{j=1}^{m} u_j(z_j) - \min_{1 \le j \le m} u_j(z_j) \right) \\ &\qquad\qquad - m + \underbrace{\sum_{j=1}^{m} \log u_j'(z_j)}_{(*)}. \end{aligned} \tag{1.1}$$

Notice that $u_1', u_2', \dots, u_m'$ are defined almost everywhere only. To remove the derivative term $(*)$ in the right-hand side of (1.1), the following constraint

Proper domain: The subset of the domain on which an extended real valued function is finite.

on $\boldsymbol{u}$ is introduced:

$$\frac{1}{m}\sum_{j=1}^{m} u_j \circ \breve{u}_j^{-1} = \mathrm{id}_{\mathbb{R}},$$

which implies that the transformed mean is estimated correctly. In that case, after some technical steps, we can establish

$$\mathsf{E}\left[\sum_{j=1}^{m} \log(u_j'(Z_j))\right] \leq \mathsf{E}\left[\sum_{j=1}^{m} \log(\breve{u}_j'(Z_j))\right];$$

that is, the derivative term $(*)$ in (1.1) is bounded in expectation by a constant. In fact, the consistency result in Chapter 3 shows that omitting $(*)$ does not alter the maximizer of the right-hand side of (1.1).

In conclusion, we consider the following infinite dimensional optimization problem to characterize the solution of the synchronization problem:

1.5 Problem: *Under Assumption 1.4:*

$$\textit{Minimize} \quad \int f(\boldsymbol{u};\boldsymbol{z})\,\mathrm{d}P(\boldsymbol{z}) \quad \textit{subject to} \quad \operatorname*{ess\,sup}_{\tau\in(0,1)} g(\boldsymbol{u};\boldsymbol{Q}_0(\tau)) = 0,\ \boldsymbol{u}\in\mathcal{E},$$

where $f, g : \mathcal{E}\times\mathbb{R}^m \to \mathbb{R}$ *and* $\boldsymbol{Q}_0 : (0,1)\to\mathbb{R}^m$ *are defined by*

$$\begin{aligned} f(\boldsymbol{u};\boldsymbol{z}) &= \frac{1}{m}\sum_{j=1}^{m} u_j(z_j) - \min_{1\leq j\leq m} u_j(z_j), \\ g(\boldsymbol{u};\boldsymbol{z}) &= \frac{1}{m}\left|\sum_{j=1}^{m} u_j(z_j) - \sum_{j=1}^{m} z_j\right|, \\ \boldsymbol{Q}_0(\tau) &= \left(Q_1(\tau), Q_2(\tau), \ldots, Q_m(\tau)\right). \end{aligned}$$

Chapter 2 introduces an intuitive sample approximation to Problem 1.5. Any optimizer of such sample approximation can be interpreted as a *conditional maximum likelihood estimate* to the solution $\breve{\boldsymbol{u}}$ of the synchronization problem. Strictly speaking, the sample approximation of Problem 1.5 is not the conditional maximum likelihood estimator. We are in the more general setting of M-estimation (compare to the literature, like [26]). However, as the objective function is an upper bound of the conditional log-likelihood in expectation, we will tolerate that abuse of terminology and refer to the sample approximation of Problem 1.5 as conditional maximum likelihood estimator and its optimizer as conditional maximum likelihood estimate.

The parameter space $\mathcal{E}$ includes functions whose components are neither strictly increasing nor locally absolutely continuous. Thus, it is a-priori open whether Problem 1.5 characterizes the solution of the offline synchronization problem. Fortunately, Chapter 3 will confirm that Problem 1.5 does characterize $\check{\boldsymbol{u}}$, and that the estimator defined in Chapter 2 is strongly consistent. That is, the optimizers of the sample approximation of Problem 1.5 converge to $\check{\boldsymbol{u}}$ almost surely.

1.5 Incomplete Log-Sets

Problem 1.5 is developed for a complete synchronization problem. That is, each event is observed by each device and the log-sets are complete. The presented approach can be adapted easily to incomplete log-entries. For this purpose we modify Definition 1.1, Theorem 1.2, and Problem 1.5 accordingly:

1.6 Definition: *Let* $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ *be a random vector in* $(\mathbb{R} \cup \{\infty\})^m$ *with* $\mathbf{P}\{Z_j \in \mathbb{R}\} > 0$ *for all* $j \in \{1, 2, \dots, m\}$. *The* (offline) synchronization problem *with* censored timestamp model $\boldsymbol{Z}$ *is to find increasing homeomorphisms* $u_1, u_2, \dots, u_m : \mathbb{R} \to \mathbb{R}$ *such that*

1. $\mathbf{P}\{u_j(Z_j) \leq t \,|\, Z_j \in \mathbb{R}\}$ *does not depend on* j,

2. *and such that*

$$\frac{1}{m} \sum_{j=1}^{m} \check{u}_j^{-1} = \mathrm{id}_{\mathbb{R}} \,.$$

Then, the direct product $\boldsymbol{u} = u_1 \times u_2 \times \cdots \times u_m$ *is again called a solution.*

In the case of a censored timestamp model $\boldsymbol{Z}$, if the j-th device has not recorded the event, the timestamp Z_j is simply ∞. For illustration: Assume there are an uncensored model $\tilde{\boldsymbol{Z}} = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_m)$ and a random vector $\boldsymbol{R} = (R_1, R_2, \dots, R_m)$ in $\{0, 1\}^m$ such that

$$Z_j = \begin{cases} \tilde{Z}_j, & R_j = 0 \\ \infty, & R_j = 1. \end{cases}$$

In general $\tilde{\boldsymbol{Z}}$ and $\boldsymbol{R}$ are dependent. However, if $\tilde{\boldsymbol{Z}}$ and $\boldsymbol{R}$ are independent, then for each subset $J \subseteq \{1, 2, \dots, m\}$ and vector $\boldsymbol{z}_J \in \mathbb{R}^J$ we have

$$\mathbf{P}\big\{\boldsymbol{Z}_J \le \boldsymbol{z}_J \,|\, \boldsymbol{Z}_J \in \mathbb{R}^J\big\} = \mathbf{P}\Big\{\tilde{\boldsymbol{Z}}_J \le \boldsymbol{z}_J \,|\, \boldsymbol{R}_J = \boldsymbol{0}\Big\} = \mathbf{P}\Big\{\tilde{\boldsymbol{Z}}_J \le \boldsymbol{z}_J\Big\}.$$

That is, a solution of the censored model $\boldsymbol{Z}$ is also a solution of the uncensored model $\tilde{\boldsymbol{Z}}$.

To adapt Theorem 1.2 to a censored timestamp model, we simply replace F_j in that theorem with the conditional marginal distribution function of Z_j under $Z_j \in \mathbb{R}$. That is, for each $z \in \mathbb{R}$ and $j \in \{1, 2, \dots, m\}$ we have

$$F_j(z) = \mathbf{P}\big\{Z_j \le z \,|\, Z_j \in \mathbb{R}\big\} = \frac{\mathbf{P}\big\{Z_j \le z\big\}}{\mathbf{P}\big\{Z_j \in \mathbb{R}\big\}}.$$

Like in the uncensored case, assume that $F_1, F_2, \dots, F_m$ are homeomorphisms from $\mathbb{R}$ onto $(0, 1)$. Then, the censored synchronization problem has exactly one solution, which is again given by

$$\breve{u}_j = \left(\frac{1}{m}\sum_{j=1}^{m} F_j^{-1}\right) \circ F_j.$$

Adapting Problem 1.5 is slightly technical. Let a system $\mathscr{J}$ of subsets of $\{1, 2, \dots, m\}$ be given with

$$\bigcup \mathscr{J} = \{1, 2, \dots, m\} \quad \text{and} \quad \forall J \in \mathscr{J} : \mathbf{P}\big\{\boldsymbol{Z}_J \in \mathbb{R}^J\big\} > 0.$$

Consider the following problem:

1.7 Problem: *For each $J \in \mathscr{J}$ minimize*

$$\mathbf{E}\big[f_J(\boldsymbol{u}; \boldsymbol{Z}) \,|\, \boldsymbol{Z}_J \in \mathbb{R}^J\big]$$

subject to

$$\operatorname*{ess\,sup}_{\tau \in (0,1)} g_J(\boldsymbol{u}; \boldsymbol{Q}_0(\tau)) = 0,\ \boldsymbol{u} \in \mathcal{E},$$

where $f_J, g_J : \mathcal{E} \times \mathbb{R}^m \to \mathbb{R}$ are defined by

$$f_J(\boldsymbol{u}; \boldsymbol{z}) = \frac{1}{|J|}\sum_{j \in J} u_j(z_j) - \min_{j \in J} u_j(z_j),$$

$$g_J(\boldsymbol{u}; \boldsymbol{z}) = \left|\frac{1}{|J|}\sum_{j \in J} u_j(z_j) - \frac{1}{m}\sum_{j=1}^{m} z_j\right|.$$

Again, $\mathcal{E}$ is the space of separated $\mathcal{L}^1$ functions, and $\boldsymbol{Q}_0 = (Q_1, Q_2, \dots, Q_m)$ is the vector of the inverse functions of the conditional marginal distribution functions $F_1, F_2, \dots, F_m$. Notice that Problem 1.7 has the same structure as Problem 1.5. Thus, analyzing Problem 1.5 itself is sufficient. In particular, we can decompose the global synchronization problem into smaller subproblems by restricting the devices to the sets in $\mathscr{J}$.

Chapter 2

Maximum Likelihood Estimator

Chapter 1 sketches a conditional maximum likelihood ansatz to characterize the solution of the offline synchronization problem using a synchronized time-stamp model (see Problem 1.5 on page 12). In that chapter many technical questions crucial for the application are left unanswered. How do we evaluate integrals with respect to unknown measures? How do we minimize over a infinite dimensional function space? The answers to these questions will be provided in this chapter.

2.1 Sample Approximation

To provide an efficient scheme to approximate the solution of Problem 1.5, this section employs two well-known techniques: First, the integral and the quantile functions are replaced by their sample counterparts. Second, the infinite dimensional function space is approximated by well-chosen finite dimensional subset. The goal is to solve the following *sample approximation* of Problem 1.5:

2.1 Problem: *Let f, g, and $\mathcal{E}$ be defined as in Problem 1.5. Let P_n be a sample distribution of P_0 with sample size $n \in \mathbb{N}$, $Q_{1,n}, Q_{2,n}, \dots, Q_{m,n}$ be the sample quantile functions of its components, and $\boldsymbol{Q}_n = (Q_{1,n}, Q_{2,n}, \dots, Q_{m,n})$ be*

Finite dimensional subset: A subset of some linear space whose linear span is finite dimensional.

the vector-valued sample quantile function. For a finite dimensional subset $\mathcal{F}_n \subseteq \mathcal{E}$*, a finite subset* $I_n \subset (0,1)$*, and a positive number* $\eta_n > 0$*,*

$$\textit{minimize} \int f(\boldsymbol{u};\boldsymbol{z})\,\mathrm{d}P_n(\boldsymbol{z}) \;\; \textit{subject to} \;\; \max_{\tau\in I_n} g(\boldsymbol{u};\boldsymbol{Q}_n(\tau)) \leq \eta_n,\, \boldsymbol{u}\in\mathcal{F}_n.$$

Here, "sample" refers to the fact that $P_0, \boldsymbol{Q}_0$ are replaced by $P_n, \boldsymbol{Q}_n$. "Approximation" refers to the fact that $\mathcal{F}_n$ replaces $\mathcal{E}$, I_n replaces $(0,1)$, and η_n is positive instead of 0. With these relaxations, Problem 2.1 turns into a feasible finite dimensional problem.

Concerning I_n and η_n, notice that $\boldsymbol{Q}_n$ is a left continuous step function with jumps at $\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}$. Thus, we could settle with the constraint

$$\forall \tau \in \{\tfrac{1}{n}, \tfrac{2}{n}, \dots, \tfrac{n}{n}\} : g(\boldsymbol{u};\boldsymbol{Q}_n(\tau)) = 0. \tag{2.1}$$

This produces n constraints, unnecessarily many as the consistency analysis will show. Also, due to the random nature of the data, (2.1) might be inconsistent. That is, there might be no $\boldsymbol{u} \in \mathcal{F}_n$ which satisfies (2.1), even if $\mathcal{F}_n$ contains the solution $\check{\boldsymbol{u}}$. Instead, it is more convenient to impose the collocation conditions (2.1) at few predefined points given by I_n and bound the error:

$$\forall \tau \in I_n : g(\boldsymbol{u};\boldsymbol{Q}_n(\tau)) \leq \eta_n.$$

Replacing the infinite dimensional space $\mathcal{E}$ with a finite dimensional subset $\mathcal{F}_n$ is the standard technique to treat functions as unknowns. The upcoming consistency analysis implies that it is sufficient if $\mathcal{F}_n$ approximates the solution $\check{\boldsymbol{u}}$ well. That is, we do not need to approximate the complete space $\mathcal{E}$. Due to numerical and computational reasons it is important that $\mathcal{F}_n$ has a well-behaving basis. Desirable properties are for example positivity, uniform boundedness, or even compact support. The discussion of such implementation details is deferred to Section 2.4. The following general definition is sufficient for this section:

2.2 Definition: *Let* $\boldsymbol{d} = (d_1, d_2, \dots, d_m) \in \mathbb{N}^m$*. A tuple* $\boldsymbol{\phi} = (\boldsymbol{\varphi}_j)_{1\leq j\leq m}$ *is called a* $\boldsymbol{d}$*-dimensional* inverse clock model*, if* $\boldsymbol{\varphi}_j$ *is a Lipschitz continuous function from* $\mathbb{R}$ *to* $\mathbb{R}^{d_j}$ *with*

$$\mathrm{id} \in \mathrm{Span}\big\{\varphi_{j,1}, \varphi_{j,2}, \dots, \varphi_{j,d_j}\big\}$$

for each $j \in \{1, 2, \dots, m\}$*.*

A inverse clock model $\boldsymbol{\phi}$ defines the space

$$\mathcal{F}(\boldsymbol{\phi}) := \big\{\langle \boldsymbol{\vartheta}_1, \boldsymbol{\varphi}_1\rangle \times \cdots \times \langle \boldsymbol{\vartheta}_m, \boldsymbol{\varphi}_m\rangle \mid \boldsymbol{\vartheta}_j \in \mathbb{R}^{d_j},\, 1 \le j \le m\big\},$$

which is a $|\boldsymbol{d}| = d_1 + d_2 + \cdots + d_m$ dimensional subspace of $\mathcal{S}ep_m \cap \mathcal{L}^1(P_0; \mathbb{R}^m)$. Here, $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product of $\mathbb{R}^d$ and $\langle \boldsymbol{\vartheta}_j, \boldsymbol{\varphi}_j\rangle$ denotes the mapping from $\mathbb{R} \to \mathbb{R}$ defined by

$$\big\langle \boldsymbol{\vartheta}_j, \boldsymbol{\varphi}_j \big\rangle(z) = \big\langle \boldsymbol{\vartheta}_j, \boldsymbol{\varphi}_j(z) \big\rangle.$$

An illustrative example: If we assume that the solution $\check{\boldsymbol{u}}$ are affine linear functions, then for each $j \in \{1, 2, \dots, m\}$ we can use the basis function

$$\boldsymbol{\varphi}_j(z) = \begin{bmatrix} 1 \\ z \end{bmatrix}.$$

Using $\mathcal{F}(\boldsymbol{\phi})$ as $\mathcal{F}_n$, we can turn Problem 2.1 on page 17 into a sparse linear program.

2.2 Linear Program Formulation

Fix a realization $(\boldsymbol{z}_k)_{1\le k\le n} = (z_{j,k})_{1\le j\le m, 1\le k\le n}$ of a sample sequence of P_0 with sample size n. Fix a given inverse clock model $\boldsymbol{\phi} = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_m)$, a finite subset $I_n \subseteq (0, 1)$, and a positive number $\eta_n > 0$. Then, Problem 2.1 turns into the following equivalent finite dimensional linear program:

2.3 Problem: *Minimize*

$$\frac{1}{mn} \sum_{k=1}^{n} \sum_{j=1}^{m} \big(u_j(z_{j,k}) - t_k\big)$$

subject to

$$\forall (j,k) \in \{1, \dots, m\} \times \{1, \dots, n\} : t_k - u_j(z_{j,k}) \le 0, \tag{2.2}$$

$$\forall \tau \in I_n : \pm\frac{1}{m} \sum_{j=1}^{m} u_j\big(Q_{j,n}(\tau)\big) \le \pm\frac{1}{m} \sum_{j=1}^{m} Q_{j,n}(\tau) + \eta_n, \tag{2.3}$$

$$u_1 \times u_2 \times \cdots \times u_m \in \mathcal{F}(\boldsymbol{\phi}), \qquad (t_k)_{1\le k\le n} \in \mathbb{R}^n.$$

Although not explicitly stated, the coefficient vectors $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \dots, \boldsymbol{\vartheta}_m$ are the new unknowns in place of the unknown functions

$$u_j = \langle \boldsymbol{\vartheta}_j, \boldsymbol{\varphi}_j \rangle, \quad 1 \leq j \leq m.$$

The minimum term in f is eliminated by introducing the slack variables $t_1, t_2, \dots, t_n$ and the additional constraint (2.2). Trivially, Problem 2.3 is bounded from below by 0. Since $\mathcal{F}(\boldsymbol{\phi})$ contains the identity function on $\mathbb{R}^m$, Problem 2.3 has a strictly feasible point. Thus, the duality theory of linear programming states that Problem 2.3 has an optimizer (compare to the literature, like [12]). In conclusion, Problem 2.1 can be restated as a strictly feasible linear program featuring a sparse coefficient matrix, which is very appealing in applications.

2.2.1 Interior Point Method

The well-known interior point method is an efficient technique to solve linear programs, particularly sparse programs, like Problem 2.3. The problem can be restated in the standard form[a]

$$\text{maximize} \quad \boldsymbol{b}^{\mathsf{T}}\boldsymbol{y} \quad \text{subject to} \quad \boldsymbol{A}^{\mathsf{T}}\boldsymbol{y} \leq \boldsymbol{c}, \ \boldsymbol{y} \in \mathbb{R}^{\bar{m}},$$

with a suitable coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{\bar{m} \times \bar{n}}$ and vectors $\boldsymbol{b} \in \mathbb{R}^{\bar{m}}, \boldsymbol{c} \in \mathbb{R}^{\bar{n}}$.

The interior point method exploits the fact that the duality gap of a feasible linear program is zero at any solution. Usually, the Newton method is applied to solve a system of perturbed Karush–Kuhn–Tucker conditions, or to minimize some barrier function (see for example [19]). Only the key aspects concerning its application on the Problem 2.3 are highlighted here. The first one concerns the applicability, the second one concerns the efficiency.

Rank Assumption

To apply the interior point method, the coefficient matrix $\boldsymbol{A}$ needs to have rank $\bar{m}$. For particular inverse clock models and choice of I_n, the coefficient matrix $\boldsymbol{A}$ tends to have maximal rank. However, due to the random nature of the data, this is in general not guaranteed.

[a] In fact, that is just one of many popular standard forms. However, the standard forms founded in the literature are equivalent.

In applications, row elimination techniques are applied to $\boldsymbol{A}$ to reduce the problem to an equivalent one whose coefficient matrix satisfies the rank assumption. Fortunately, in case of Problem 2.3, the sparsity structure of $\boldsymbol{A}$ is preserved. Split $\boldsymbol{A}$ into two submatrices, where the first block $\boldsymbol{B}$ corresponds to the slack variables $\boldsymbol{t}$, and the second block $\boldsymbol{C}$ corresponds to the variables $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \dots, \boldsymbol{\vartheta}_m$:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{C} \end{bmatrix}.$$

Due to the structure of $\boldsymbol{B}$, the rows of $\boldsymbol{B}$ are linearly independent. Thus, we can restrict the row elimination on $\boldsymbol{C}$ only and preserve the sparse structure of $\boldsymbol{B}$. Thus, we will assume $\operatorname{rank} \boldsymbol{A} = \bar{m}$ without loss of generality for the remaining discussion.

Sparsity Structure

The main computational cost is caused by solving linear equations involving the coefficient matrix

$$\boldsymbol{M} = \boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^{\mathsf{T}},$$

where $\boldsymbol{D}$ is some diagonal matrix with positive diagonal entries depending on the current iterate. The matrix $\boldsymbol{M}$ is analogous to the coefficient matrix of the normal equations of a least squares problem; it is also numerically prone to rounding errors. Nonetheless, it is commonly used in interior point methods as it often admits a cheap and sparse Cholesky factorization after proper reordering. Further, the solutions are merely search directions, saying that inexact solutions are corrected in the successive iteration steps. Numerical results, like those presented by Fourer et al. [10], or by Lustig et al. [16], indicate that the inexact solutions obtained from a Cholesky factor of $\boldsymbol{M}$ yield efficient search steps in interior point methods.

For a general positive definite sparse matrix $\boldsymbol{M}$ of large size, iterative methods are usually preferred, like the preconditioned conjugate gradient method. Direct methods, like the Cholesky factorization, become favorable, if the structure of $\boldsymbol{M}$ is well-known and can be exploited. For the particular matrix $\boldsymbol{A}$ here, the matrix $\boldsymbol{M}$ is nearly diagonal, that is, computing the Cholesky factorization is tolerable, even for high dimensional $\boldsymbol{M}$. Noteworthy, we could also apply the preconditioned conjugate gradient method using

the Cholesky factor as preconditioner to increase the numerical accuracy. For example, this technique is implemented in the well-known optimization software package sedumi [25].

As $\boldsymbol{M}$ and $\boldsymbol{A}\boldsymbol{A}^\mathsf{T}$ share the same zero structure, it is sufficient to analyze the structure of the latter one. Let $\boldsymbol{A}^\mathsf{T}$ be split into two submatrices, where the first block $\boldsymbol{B}^\mathsf{T}$ corresponds to the slack variables $\boldsymbol{t}$, and the second block $\boldsymbol{C}^\mathsf{T}$ corresponds to the variables $\boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \dots, \boldsymbol{\vartheta}_m$:

$$\boldsymbol{A}^\mathsf{T} = \begin{bmatrix} \boldsymbol{B}^\mathsf{T} & \boldsymbol{C}^\mathsf{T} \end{bmatrix}.$$

Then, it follows

$$\boldsymbol{A}\boldsymbol{A}^\mathsf{T} = \begin{bmatrix} \boldsymbol{B}\boldsymbol{B}^\mathsf{T} & \boldsymbol{B}\boldsymbol{C}^\mathsf{T} \\ \boldsymbol{C}\boldsymbol{B}^\mathsf{T} & \boldsymbol{C}\boldsymbol{C}^\mathsf{T} \end{bmatrix}.$$

The columns of $\boldsymbol{B}$ are the columns of the $n \times n$ identity matrix $\boldsymbol{I}$, and each column of $\boldsymbol{I}$ appears m times in $\boldsymbol{B}$. Thus, it follows

$$\boldsymbol{B}\boldsymbol{B}^\mathsf{T} = m\boldsymbol{I}.$$

In particular, $\boldsymbol{A}\boldsymbol{A}^\mathsf{T}$ is nearly diagonal, if n is many magnitudes greater than $\bar{n} - n$. In general, the first diagonal block of $\boldsymbol{M}$ is a diagonal matrix only. It is not $m\boldsymbol{I}$ due to the scale factor $\boldsymbol{D}$.

2.3 Incomplete Log-Sets, continued

Adapting Problem 2.1 to a censored timestamp model is straightforward. Instead of Problem 1.5 on page 12, we need to approximate the solution of Problem 1.7 on page 14, which is basically a separated version of Problem 1.5. For the sake of completeness, let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ denote a censored timestamp model; that is, any Z_j takes values in $\mathbb{R} \cup \{\infty\}$. Also let $Q_{j,n}$ denote the conditional sample quantile function of Z_j under the condition $Z_j \in \mathbb{R}$. Let $\mathscr{J}_n$ be a system of subsets of $\{1, 2, \dots, m\}$ with $\bigcup \mathscr{J}_n = \{1, 2, \dots, m\}$ and

$$\min_{J \in \mathscr{J}_n} P_n\{(\,\cdot\,)_J \in \mathbb{R}^J\} > 0,$$

where $(\,\cdot\,)_J$ denotes the projection onto the J-subcomponent. That is, for each $J \in \mathscr{J}_n$ the random subvector $\boldsymbol{Z}_J$ has finite realizations. Then, we consider the following sample approximation to Problem 1.7:

2.4 Problem: *For each $J \in \mathscr{J}_n$ minimize*

$$P_n\left[f_J(\boldsymbol{u};\,\cdot\,) \mid (\,\cdot\,)_J \in \mathbb{R}^J\right]$$

subject to

$$\max_{\tau\in I_n} g_J(\boldsymbol{u};\boldsymbol{Q}_n(\tau)) \le \eta_n, \quad \boldsymbol{u} \in \mathcal{F}_n.$$

Here, $P_n[\,\cdot \mid (\,\cdot\,)_J \in \mathbb{R}^J]$ denotes the conditional sample expectation under the condition that the J-subcomponent is finite; f_J and g_J are defined as in Problem 1.7; I_n, η_n, $\mathcal{F}_n$, and $\boldsymbol{Q}_n$ are defined as in Problem 2.1.

As for an implementation, there are some open questions: 1. To exploit as many samples as possible, we could choose $\mathscr{J}_n$ as the set of realizations of following random set

$$J(\boldsymbol{Z}) = \big\{j \in \{1,2,\dots,m\} \mid Z_j \in \mathbb{R}\big\}.$$

In the extreme case, there are exponentially many realizations of $J(\boldsymbol{Z})$, which leads to exponentially many subproblems. Even for small m, this would be impractical. 2. To keep the computational cost low, the subsets in $\mathscr{J}_n$ shall be as small as possible. An extreme example would be

$$\mathscr{J}_n = \{\{j\} \mid j \in \{1,2,\dots,m\}\}.$$

In such a configuration, we do not exploit the correlation of the components of $\boldsymbol{Z}$. This leads to a poor rate of convergence. 3. For efficiency reason, we would prefer to solve the subproblem for each $J \in \mathscr{J}_n$ independently. However consider following example: In case of $J_1, J_2 \in \mathscr{J}_n$ with $1 \in J_1 \cap J_2$, the subproblem on J_1 and on J_2 both yield an estimator for the solution $\breve{u}_1$. Both estimators are consistent for $\breve{u}_1$, which will be shown in the upcoming Chapter 3. Deciding which one is "better" is not easy, in particular, if we have only few samples. Thus, we shall solve Problem 2.4 simultaneously to obtain a unique result:

2.5 Problem: *Minimize*

$$\sum_{J\in\mathscr{J}_n} P_n\left[f_J(\boldsymbol{u};\,\cdot\,) \mid (\,\cdot\,)_J \in \mathbb{R}^J\right]$$

subject to

$$\max_{J\in\mathscr{J}_n}\max_{\tau\in I_n} g_J(\boldsymbol{u};\boldsymbol{Q}_n(\tau)) \le \eta_n, \quad \boldsymbol{u} \in \mathcal{F}_n.$$

Fortunately, simulations indicate that the choice of $\mathscr{J}_n$ has little influence on the convergence.

2.4 A Spline Inverse Clock Model

In general, the solution $\check{\boldsymbol{u}}$ of the synchronization problem is not affine linear, or an element of a known finite dimensional space. In fact, we only have very weak assumptions on $\check{\boldsymbol{u}}$, namely it is locally absolutely continuous and integrable with respect to the distribution of the timestamp model. Approximating such a function is difficult, as it is not periodic and the domain is non-compact. If the solution $\check{\boldsymbol{u}}$ is square-integrable, series expansion techniques arise naturally. However, since the distribution of the timestamp model is unknown, explicitly constructing an orthonormal system and estimating the cut-off error would be difficult. To make standard approximation techniques viable, we assume that $\check{\boldsymbol{u}}$ is uniquely determined by its values on a known compact domain. To be precise, for each $j \in \{1, 2, \dots, m\}$ we assume

$$\check{u}_j = v_j + w_j,$$

where v_j is an affine function and w_j is a Lipschitz continuous function supported on some known compact interval, say $K_j = [\underline{z}_j, \bar{z}_j] \subseteq \mathbb{R}$. Notice that $\check{\boldsymbol{u}}$ is now globally Lipschitz continuous. The clock model can be generalized easily such that v_j is the sum of a polynomial and some function with periodic derivative, which allows non-globally Lipschitz continuous solutions. However, a higher polynomial degree imposes stronger integrability conditions on the timestamp model, and periodic terms produce a more dense coefficient matrix in the linear program formulation (Problem 2.3). For the sake of simplicity, we will assume v_j to be affine only.

Trivially, $\check{u}_j|_{K_j}$ uniquely determines v_j and w_j, as for each $z \in \mathbb{R}$ we have

$$\begin{aligned} v_j(z) &= \frac{\bar{z}_j - z}{\bar{z}_j - \underline{z}_j}\check{u}_j(\underline{z}_j) + \frac{z - \underline{z}_j}{\bar{z}_j - \underline{z}_j}\check{u}_j(\bar{z}_j), \\ w_j(z) &= \check{u}_j(z) - v_j(z). \end{aligned}$$

That is, we are dealing with approximations of Lipschitz continuous functions on a compact interval only, which is well studied and understood. Here, we apply spline approximation.

2.4.1 Splines and B-Splines

As usual, to approximate a Lipschitz continuous function u on some compact interval $[\underline{\xi}, \bar{\xi}]$, we divide $[\underline{\xi}, \bar{\xi}]$ into subintervals and approximate u by a polynomial on each interval. By decreasing the length of the subintervals we shall approximate u arbitrarily well.

2.6 Definition (Piecewise Polynomials and Splines): *Let $d \in \mathbb{N} \cup \{0\}$, $I \subseteq \mathbb{R}$ be an interval, and $\boldsymbol{\xi} = (\xi_0, \xi_1, \dots, \xi_\nu) \in \mathbb{R}^{1+\nu}$ be a strictly increasing sequence.*

- *The space of* polynomials *of degree $\leq d$ on I is denoted by*

$$\mathcal{P}_d(I) = \left\{ u : I \to \mathbb{R} \mid \exists \vartheta_0, \dots, \vartheta_d \in \mathbb{R}\ \forall z \in I : u(z) = \sum_{i=0}^{d} \vartheta_i z^i \right\}.$$

- *The space of (right continuous)* piecewise polynomials *of degree $\leq d$ with respect to the* knot sequence $\boldsymbol{\xi}$ *is denoted by*

$$\mathcal{P}_d(\boldsymbol{\xi}) = \left\{ u : [\xi_0, \xi_\nu) \to \mathbb{R} \mid \forall i \in \{1, \dots, \nu\} : u|_{[\xi_{i-1}, \xi_i)} \in \mathcal{P}_d[\xi_{i-1}, \xi_i) \right\}.$$

  Notice that each $f \in \mathcal{P}_d(\boldsymbol{\xi})$ can be extended uniquely to a function $\bar{f}$ on $[\xi_0, \xi_\nu]$ such that $\bar{f}$ is a polynomial of degree $\leq d$ on $[\xi_{\nu-1}, \xi_\nu]$. Thus, we identify f with $\bar{f}$.

- *The space of* splines *of degree $\leq d$ with respect to the knot sequence $\boldsymbol{\xi}$ is denoted by*

$$\mathcal{S}_d(\boldsymbol{\xi}) = \begin{cases} \mathcal{P}_0(\boldsymbol{\xi}), & d = 0, \\ \mathcal{P}_d(\boldsymbol{\xi}) \cap \mathcal{C}^{d-1}([\xi_0, \xi_\nu]; \mathbb{R}), & d \geq 1. \end{cases}$$

To solve Problem 2.3 with numerical methods, it is important that the splines admit a *well-behaved* basis, the so-called B-spline basis. In fact, the B-splines lead to an even more sparse coefficient matrix in Problem 2.3. Through out the literature, there are multiple equivalent definitions of the B-splines. Some are more convenient than the other in certain context. The following particular characterization demonstrates, according to Cox and de Boor [4, 5], that the B-splines can be evaluated as sum of non-negative numbers in a stable fashion.

2.7 Definition: *Let a knot sequence* $\boldsymbol{\xi} = (\xi_0, \dots, \xi_\nu)$ *be given, which we extend bi-infinitely to*

$$\cdots = \xi_{-1} = \xi_0 < \cdots < \xi_\nu = \xi_{\nu+1} = \cdots .$$

- *For* $i \in \mathbb{Z}$, *the* i*-th* B-spline $\varphi_{0,i} : \mathbb{R} \to \mathbb{R}$ *of order* 0 *is defined by*

$$\varphi_{0,i}(z) = \begin{cases} 1, & \xi_i \le z < \xi_{i+1}, \\ 0, & \textit{otherwise.} \end{cases}$$

- *For* $d \in \mathbb{N} \cup \{0\}$ *and* $i \in \mathbb{Z}$, *the* i*-th* B-spline $\varphi_{d+1,i} : \mathbb{R} \to \mathbb{R}$ *of degree* $d+1$ *is defined by*

$$\varphi_{d+1,i} = \gamma_{d,i}\varphi_{d,i} + (1 - \gamma_{d,i+1})\varphi_{d,i+1}$$

with

$$\gamma_{d,i}(z) = \begin{cases} \frac{z-\xi_i}{\xi_{i+d+1}-\xi_i}, & \xi_i \le z < \xi_{i+d+1}, \\ 0, & \textit{otherwise.} \end{cases}$$

Some properties of the B-splines follow directly from their recursive definition. For example, $\varphi_{d,i}$ is non-negative, zero outside of $[\xi_i, \xi_{i+d+1}]$, and positive on (ξ_i, ξ_{i+d+1}). Due to the construction of the knot-sequence, $\varphi_{d,i}$ is the zero function if and only if $i \notin \{-d, \dots, \nu - 1\}$. Also, $\varphi_{d,i}$ is a polynomial of degree d on the subinterval $[\xi_{\ell-1}, \xi_\ell)$ for $\ell \in \{1, 2, \dots, \nu\}$.

To establish that $\phi_{d,i}$ is globally $\mathcal{C}^{d-1}$, for $d \ge 1$, on $[\xi_0, \xi_\nu)$ is rather technical. The result is well-known. Thus, we refer to the standard literature, like the survey by de Boor [7]. Notice that due to the choice of the knot-sequence, $\varphi_{d,\nu-1}$ will be discontinuous at ξ_ν. By convention, we redefine $\varphi_{d,\nu-1}$ at ξ_ν for each $d \ge 0$:

$$\varphi_{d,\nu-1}(\xi_\nu) \leftarrow \lim_{\substack{z \to \xi_\nu \\ z < \xi_\nu}} \varphi_{d,\nu-1}(z) = 1,$$

which makes $\varphi_{d,\nu-1}$ a polynomial on $[\xi_{\nu-1}, \xi_\nu]$ and $\mathcal{C}^{d-1}$ on $[\xi_0, \xi_\nu]$.

In summary, $\varphi_{d,-d}, \dots, \varphi_{d,\nu-1}$ restricted onto $[\xi_0, \xi_\nu]$ are non-zero elements of $\mathcal{S}_d(\boldsymbol{\xi})$. Finally, Karlin showed in [13] that the non-zero B-splines are linearly independent. Counting the degrees of freedom yields that the linear space $\mathcal{S}_d(\boldsymbol{\xi})$ has the dimension $\nu + d$. Thus, restricted onto $[\xi_0, \xi_\nu]$, they form a basis for $\mathcal{S}_d(\boldsymbol{\xi})$, the *B-spline basis*.

2.4.2 Spline Approximation

The following result, according to , This section demonstrates that we can obtain arbitrarily high approximation order using splines. The result is well-known and a proof is given by de Boor [6], which we elaborate here for the sake of completeness.

It is well-known that a function $u \in \mathcal{C}([\underline{\xi}, \bar{\xi}], \mathbb{R})$ can be approximated by a step function, that is a spline $S_0(\boldsymbol{\xi}; u) \in \mathcal{S}_0(\boldsymbol{\xi})$, arbitrarily well by decreasing the mesh

$$\Delta\boldsymbol{\xi} = \max_{1 \le i \le \nu} \xi_i - \xi_{i-1},$$

where $\boldsymbol{\xi} = (\underline{\xi} = \xi_0, \xi_1, \dots, \xi_\nu = \bar{\xi})$ is some knot sequence in $[\underline{\xi}, \bar{\xi}]$. If u is Lipschitz continuous, we even obtain approximation order 1:

$$\|u - S_0(\boldsymbol{\xi}; u)\|_\infty \le \Delta\boldsymbol{\xi} \|u'\|_\infty.$$

If u satisfies stronger regularity conditions, for example $\mathcal{C}^{d,1}([\underline{\xi}, \bar{\xi}], \mathbb{R})$, similar results holds for $\mathcal{S}_d(\boldsymbol{\xi})$ with $d > 0$. The key is to construct a good approximation of $u^{(d)}$ explicitly, and then iteratively an approximation of $u^{(d-i)}$ by using the approximation of $u^{(d-i+1)}$ for $i = 1, 2, 3, \dots, d$. Here, the approximation is given by the so-called *quasi interpolant*.

2.8 Definition: *For a knot sequence* $\boldsymbol{\xi} = (\underline{\xi} = \xi_0, \xi_1, \dots, \xi_\nu = \bar{\xi})$*, which we extend bi-infinitely,* $d \in \mathbb{N} \cup \{0\}$*, and* $u : [\underline{\xi}, \bar{\xi}] \to \mathbb{R}$*, the B-spline* quasi interpolant $S_d(\boldsymbol{\xi}; u) \in \mathcal{S}_d(\boldsymbol{\xi})$ *of* u *of degree* d *is defined by*

$$S_d(\boldsymbol{\xi}; u) = \sum_{i=-d}^{\nu-1} u(\bar{\xi}_{d,i}) \varphi_{d,i},$$

where $\varphi_{d,i}$ *is the* i*-th B-spline of degree* d *and*

$$\bar{\xi}_{d,i} = \frac{\xi_i + \xi_{i+d+1}}{2}.$$

Up to constants, all results remain for a more general choice of $\bar{\xi}_{d,i}$ if the following condition is satisfied: For each $d \in \mathbb{N} \cup \{0\}$ there exists some constan t $C_d \ge 0$ depending on d only such that for each $z \in [\underline{\xi}, \bar{\xi}]$ and each $i \in \mathbb{Z}$ with $\varphi_{d,i}(z) > 0$ it follows

$$|z - \bar{\xi}_{d,i}| \le C_d \Delta\boldsymbol{\xi}.$$

As we do not want to explicitly construct a quasi interpolant in applications, the choice given in Definition 2.8 suffices. Here, we have $C_d = \frac{d+1}{2}$.

Exploiting some properties of the B-splines, we can show that the quasi interpolation error is of order 1.

2.9 Lemma: *For $d \geq 0$ and $u \in \mathcal{L}ip([\underline{\xi}, \bar{\xi}], \mathbb{R})$ it follows*

$$\|u - S_d(\boldsymbol{\xi}; u)\|_\infty \leq \frac{d+1}{2} \Delta\boldsymbol{\xi} \|u'\|_\infty.$$

Proof: Here, we exploit the fact that the sum of the B-splines is 1 on $[\xi_0, \xi_\nu]$. This is trivially true for $d = 0$. For the induction step $d \to d+1$: For each $i \in \{1, 2, \dots, \nu\}$ on $[\xi_{i-1}, \xi_i)$ we have

$$\begin{aligned}
\sum_{j \in \mathbb{Z}} \varphi_{d+1,j} &= \sum_{j=i-d-2}^{i-1} \varphi_{d+1,j} \\
&= \sum_{j=i-d-2}^{i-1} \big(\gamma_{d,j}\varphi_{d,j} + (1 - \gamma_{d,j+1})\varphi_{d,j+1}\big) \\
&= \gamma_{d,i-d-2} \underbrace{\varphi_{d,i-d-2}}_{=0} + \sum_{j=i-d-1}^{i-1} \gamma_{d,j}\varphi_{d,j} \\
&\quad + \sum_{j=i-d-2}^{i-2} (1 - \gamma_{d,j+1})\varphi_{d,j+1} + (1 - \gamma_{d,i}) \underbrace{\varphi_{d,i}}_{=0}.
\end{aligned}$$

That is

$$\sum_{j \in \mathbb{Z}} \varphi_{d+1,j} = \sum_{j=i-d-1}^{i-1} \big(\gamma_{d,j}\varphi_{d,j} + (1 - \gamma_{d,j})\varphi_{d,j}\big) = \sum_{j=i-d-1}^{i-1} \varphi_{d,j} = 1.$$

Now, by continuity the result also holds at ξ_ν.

Using that fact we can prove the statement. For each $z \in [\underline{\xi}, \bar{\xi}]$ let

$$I(z) = \big\{i \in \{0, 1, \dots, \nu - d - 1\} \mid \varphi_{d,i}(z) > 0\big\},$$

which is an interval in $\mathbb{Z}$ containing $d + 1$ indices at most. Then, for each

$z \in [\underline{\xi}, \bar{\xi}]$ it follows

$$\begin{aligned}|(u - S_d(\boldsymbol{\xi}; u))(z)| &= \left| \sum_{i \in I(z)} (u(z) - u(\bar{\xi}_{d,i}))\varphi_{d,i}(z) \right| \\ &\leq \max_{i \in I(z)} |z - \bar{\xi}_{d,i}| \|u'\|_\infty \\ &\leq \frac{d+1}{2} \Delta\boldsymbol{\xi} \|u'\|_\infty.\end{aligned}$$

Thus, it follows

$$\|u - S_d(\boldsymbol{\xi}; u)\|_\infty \leq \frac{d+1}{2} \Delta\boldsymbol{\xi} \|u'\|_\infty. \qquad \square$$

Iteratively applying the previous result, we can obtain higher approximation order using higher order derivatives:

2.10 Theorem:
Let $d \geq 0$ and $u \in \mathcal{C}^{d,1}([\underline{\xi}, \bar{\xi}], \mathbb{R})$. Then, there exists a spline $s_d \in \mathcal{S}_d(\boldsymbol{\xi})$ with

$$\|u - s_d\|_\infty \leq \frac{(d+1)!}{2^{d+1}} \Delta\boldsymbol{\xi}^{d+1} \|u^{(d+1)}\|_\infty.$$

Proof: By induction over d: For $d = 0$ the statement is trivially true. For $d - 1 \to d \geq 1$: Let $v \in \mathcal{S}_{d-1}(\boldsymbol{\xi})$ with

$$\|u' - v\|_\infty \leq \frac{d!}{2^d} \Delta\boldsymbol{\xi}^d \|u^{(d+1)}\|_\infty,$$

and let V denote the antiderivative of v:

$$V(z) = \int_{\xi_0}^{z} v(\tilde{z}) \, \mathrm{d}\tilde{z}.$$

Then, for

$$s_d = V + S_d(\boldsymbol{\xi}; u - V)$$

it follows

$$\begin{aligned}\|u - s_d\|_\infty &= \|u - V - S_d(\boldsymbol{\xi}; u - V)\|_\infty \\ &\leq \frac{d+1}{2} \Delta\boldsymbol{\xi} \|u' - v\|_\infty \\ &\leq \frac{(d+1)!}{2^{d+1}} \Delta\boldsymbol{\xi}^{d+1} \|u^{(d+1)}\|_\infty. \qquad \square\end{aligned}$$

Notice that the factor $(d+1)!$ of the error bound is rather pessimistic. For example, better error bounds are known for spline interpolation with particular boundary conditions. Also, usually high order quasi interpolants are constructed more efficiently by other means. However, we only want to display the approximation ability of the splines, and do not want to explicitly construct such a quasi interpolant.

Using similar techniques, we can also prove that for each $u \in \mathcal{C}^{d,1}([\underline{\xi}, \bar{\xi}], \mathbb{R})$ there exists a spline $s_{d+1} \in \mathcal{S}_{d+1}(\boldsymbol{\xi})$ with

$$\|u - s_{d+1}\|_\infty \leq C_d \Delta\boldsymbol{\xi}^{d+1} \|u^{(d+1)}\|_\infty$$

for some constant C_d depending on d only. This result is more suitable for our application, particularly for $d = 0$, since the approximation shall be continuous at least.

2.4.3 Knot Sequence

There are no optimal knot sequences in general. Approximation applications usually rely on heuristics to adaptively choose a suitable knot sequence. In the context of Problem 2.3, we have a knot sequence $\boldsymbol{\xi}_j = (\xi_{j,i})_{i=0,\dots,\nu_n}$ for each device $j \in \{1, \dots, m\}$. Intuitively, each subinterval $[\xi_{j,i-1}, \xi_{j,i}]$ shall have the same probability; that is, each subinterval has nearly the same number of samples. This consideration leads to the choice

$$\xi_{j,i} = Q_{j,n}(\underline{\tau} + \tfrac{i}{\nu_n}(\bar{\tau} - \underline{\tau})), \qquad i = 0, 1, \dots, \nu_n.$$

Here, $Q_{j,n}$ denotes the (conditional) sample quantile function of the j-th component with n samples, and $\underline{\tau}, \bar{\tau} \in (0,1)$ are some predefined values with $\underline{\tau} < \bar{\tau}$. The number of subintervals ν_n depends on the number of samples n. It makes less sense to have many breaks and thus also higher computational cost if only few samples are available. In case of incomplete log-sets, we may even need to vary the number of subintervals ν_n for each j, as the numbers of finite samples may differ between the components. A good heuristic can be deduced from the consistency results in the upcoming Chapter 3. It also implies that if the quantile functions $Q_1, Q_2, \dots, Q_m$ are locally Lipschitz continuous, then the mesh $\Delta\boldsymbol{\xi}_j$ converges to 0 for a proper choice of ν_n and $n \to \infty$.

Chapter 3

Consistency

Chapter 1 motivates Problem 1.5 to characterize the solution of an offline synchronization problem; an unproven claim yet. Chapter 2 introduces Problem 2.1 to approximate the solution of Problem 1.5; also an unproven claim. This chapter shows that these two claims are indeed valid under weak assumptions. That is the optimizers of Problem 2.1 approach the solution of the offline synchronization problem for increasing numbers of samples. In particular, the conditional maximum likelihood estimator is consistent.

3.1 Claim 1

This section assumes the standard setting as in Problem 1.5 on page 12 and establishes that the solution $\breve{\boldsymbol{u}}$ of the synchronization problem is also the unique solution of Problem 1.5. To simplify the complex notation, we change the variable as follows:

3.1 Remark: *Let $\breve{P}_0 = P_0 \circ \breve{\boldsymbol{u}}^{-1}$, that is the distribution of $\breve{\boldsymbol{u}}(\boldsymbol{Z})$. Then, Problem 1.5 is equivalent to the following problem:*

$$\textit{Maximize} \quad \int \min_{1\le j\le m} \tilde{u}_j(x_j)\,\mathrm{d}\breve{P}_0(\boldsymbol{x}) \quad \textit{subject to} \quad \frac{1}{m}\sum_{j=1}^{m} \tilde{u}_j = \mathrm{id} \;\; \boldsymbol{\lambda}\text{-}a.e.$$

over

$$\mathcal{S}ep_m \cap \mathcal{L}^1(\breve{P}_0, \mathbb{R}^m).$$

That is, we may assume $\breve{\boldsymbol{u}} = \mathrm{id}_{\mathbb{R}^m}$ without loss of generality.

Proof: Notice that for each measurable $\phi : \mathbb{R}^m \to \mathbb{R}$, we have $\phi \in \mathcal{L}^1(P_0, \mathbb{R})$ if and only if $\phi \circ \breve{\boldsymbol{u}}^{-1} \in \mathcal{L}^1(\breve{P}_0, \mathbb{R})$, and in that case

$$\mathbf{E}[\phi(\boldsymbol{Z})] = \int \phi(\boldsymbol{z}) \, \mathrm{d}P_0(\boldsymbol{z}) = \int \phi(\breve{\boldsymbol{u}}^{-1}(\boldsymbol{x})) \, \mathrm{d}\breve{P}_0(\boldsymbol{x}).$$

Also, a function $\boldsymbol{u} \in \mathcal{E}$ is feasible for Problem 1.5 if and only if almost everywhere

$$Q_0 = \frac{1}{m} \sum_{j=1}^{m} u_j \circ Q_j = \frac{1}{m} \sum_{j=1}^{m} u_j \circ \breve{u}_j^{-1} \circ Q_0.$$

Since Q_0 and its inverse function are locally absolutely continuous, it is equivalent to almost everywhere

$$\mathrm{id} = \frac{1}{m} \sum_{j=1}^{m} u_j \circ \breve{u}_j^{-1}.$$

Remember that $\breve{u}_1(Z_1), \breve{u}_2(Z_2), \ldots, \breve{u}_m(Z_m)$ are identically distributed, say according to the common marginal distribution $P_{\breve{u}_1(Z_1)}$. Thus, for each $\boldsymbol{u} \in \mathcal{E}$ we have

$$\begin{aligned}
\mathbf{E}\left[\frac{1}{m} \sum_{j=1}^{m} u_j(Z_j)\right] &= \frac{1}{m} \sum_{j=1}^{m} \int u_j(\breve{u}_j^{-1}(x_j)) \, \mathrm{d}\breve{P}_0(\boldsymbol{x}) \\
&= \frac{1}{m} \sum_{j=1}^{m} \int u_j(\breve{u}_j^{-1}(x)) \, \mathrm{d}P_{\breve{u}_1(Z_1)}(x) \\
&= \int_{\mathbb{R}} \underbrace{\frac{1}{m} \sum_{j=1}^{m} u_j(\breve{u}_j^{-1}(x))}_{x} \, \mathrm{d}P_{\breve{u}_1(Z_1)}(x) = \mathbf{E}[\breve{u}_1(Z_1)].
\end{aligned}$$

and

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{m} \sum_{j=1}^{m} u_j(Z_j) - \min_{1 \le j \le m} u_j(Z_j)\right] \\
&\qquad = \mathbf{E}\left[\frac{1}{m} \sum_{j=1}^{m} \breve{u}_j(Z_j)\right] - \int \min_{1 \le j \le m} u_j(\breve{u}_j^{-1}(\xi_j)) \, \mathrm{d}\breve{P}_0(x).
\end{aligned}$$

Now, naming $u_j \circ \breve{u}_j^{-1}$ to $\tilde{u}_j$ yields the statement. □

In the light of Remark 3.1, the following equivalent result implies that $\breve{\boldsymbol{u}}$ is the unique minimizer of Problem 1.5 (up to equality almost everywhere):

3.2 Theorem:
Let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)$ be a random vector in $\mathbb{R}^m$ with distribution P_0, let T be a random number, and let

$$\mathcal{E} = \mathcal{S}ep_m \cap \mathcal{L}^1(P_0, \mathbb{R}^m).$$

Define $\phi : \mathcal{E} \times \mathbb{R}^m \to \mathbb{R}$ by

$$\phi(h_1 \times \cdots \times h_m; z_1, \dots, z_m) = \min_{1 \le j \le m} z_j + h_j(z_j).$$

Assume that

- *$(T, \boldsymbol{Z})$ admits a probability density;*
- *T is integrable and its density ρ_T is almost everywhere positive;*
- *the conditional density of $\boldsymbol{Z}$ given T is given by*

$$\rho_{\boldsymbol{Z}|T}(\boldsymbol{z} \,|\, t) = \begin{cases} \prod_{j=1}^m \exp(-(z_j - t)), & z_1, z_2, \dots, z_m \ge t, \\ 0, & \text{else.} \end{cases}$$

Then, the zero function $\mathbf{0}$ is a maximizer of $\Phi : \mathcal{E} \to \mathbb{R}$, defined by

$$\Phi(\boldsymbol{h}) := \mathsf{E}[\phi(\boldsymbol{h}; \boldsymbol{Z})],$$

over the subset

$$\mathcal{S} := \left\{ h_1 \times \cdots \times h_m \in \mathcal{E} \mid \frac{1}{m} \sum_{j=1}^m h_j = 0 \ \ \boldsymbol{\lambda}\text{-a.e.} \right\}.$$

Further, for each $\boldsymbol{h} \in \mathcal{S}$ with $\Phi(\boldsymbol{h})$ it follows $\boldsymbol{h} = \mathbf{0}$ almost everywhere.

Remark: *The assumption of the previous theorem on $(T, \boldsymbol{Z})$ is exactly Assumption 1.4 on page 10 in case the solution $\breve{\boldsymbol{u}}$ is the identity function, and the parameter θ of the exponential distribution is 1. We can assume $\theta = 1$ without loss of generality, as we can rescale $\breve{\boldsymbol{u}}$ and $(T, \boldsymbol{Z})$ otherwise.*

Given the assumption above, P_0 admits the density

$$\rho_{\boldsymbol{Z}}(\boldsymbol{z}) = \int \rho_{\boldsymbol{Z}|T}(\boldsymbol{z}|t)\,\rho_T(t)\,\mathrm{d}t,$$

which is continuous, positive, and bounded by 1*. In particular, every P_0-integrable function is locally $\boldsymbol{\lambda}$-integrable and every P_0-null set is also a $\boldsymbol{\lambda}$-null set.*

The key to prove Theorem 3.2 is that Φ is concave and its *directional derivative*

$$\Phi'(\mathbf{0})[\boldsymbol{h}] := \lim_{\theta \to 0+} \frac{\Phi(\theta \boldsymbol{h}) - \Phi(\mathbf{0})}{\theta}$$

vanishes at $\mathbf{0}$ for any direction $\boldsymbol{h} \in \mathcal{S}$. Notice that $\mathbf{0}$ is a maximizer of Φ if and only if $\Phi'(\mathbf{0})[\boldsymbol{h}]$ is non-positive for every direction $\boldsymbol{h} \in \mathcal{S}$. Establishing the concavity of Φ is rather trivial, as the minimum function is concave and the expectation operator is monotone. However, calculating $\Phi'(\mathbf{0})[\boldsymbol{h}]$ is rather technical.

3.1.1 Preliminary

Starting with some elementary properties of the functional Φ:

3.4 Lemma: *With the same assumption as in Theorem 3.2 it follows:*

1. Φ is concave.

2. Φ admits the Lipschitz constant 1*.*

Proof: Notice that the minimum function

$$\mathbb{R}^m \ni (\xi_1, \xi_2, \dots, \xi_m) \mapsto \min\{\xi_1, \xi_2, \dots, \xi_m\}$$

is concave and Lipschitz continuous with respect to the maximum-norm $\|\cdot\|_\infty$ with Lipschitz constant 1. Thus, for each $\boldsymbol{z} \in \mathbb{R}^m$, the function $\phi(\,\cdot\,;\boldsymbol{z})$ is concave and admits the Lipschitz constant 1 with respect to $\|\cdot\|_\infty$.

Ad Item 1: For $\boldsymbol{h}, \tilde{\boldsymbol{h}} \in \mathcal{E}$, $\boldsymbol{z} \in \mathbb{R}^m$, and $\theta \in [0,1]$ it follows

$$\begin{aligned}
\phi((1-\theta)\boldsymbol{h} + \theta\tilde{\boldsymbol{h}}; \boldsymbol{z}) &= \min_{1\le j\le m} z_j + (1-\theta)h_j(z_j) + \theta\tilde{h}_j(z_j) \\
&\ge (1-\theta)\left(\min_{1\le j\le m} z_j + h_j(z_j)\right) + \theta\left(\min_{1\le j\le m} z_j + \tilde{h}_j(z_j)\right) \\
&= (1-\theta)\phi(\boldsymbol{h};\boldsymbol{z}) + \theta\phi(\tilde{\boldsymbol{h}};\boldsymbol{z}).
\end{aligned}$$

Since the expectation operator is monotone and linear, it follows

$$\Phi((1-\theta)\boldsymbol{h}+\theta\tilde{\boldsymbol{h}}) \geq (1-\theta)\Phi(\boldsymbol{h})+\theta\Phi(\tilde{\boldsymbol{h}}).$$

That is, Φ is concave.

Ad Item 2: For $\boldsymbol{h}, \tilde{\boldsymbol{h}} \in \mathcal{E}$, and $\boldsymbol{z} \in \mathbb{R}^m$, we have

$$|\phi(\boldsymbol{h};\boldsymbol{z})-\phi(\tilde{\boldsymbol{h}};\boldsymbol{z})| \leq \max_{1\leq j\leq m}\left|z_j+h_j(z_j)-z_j-\tilde{h}_j(z_j)\right| \leq \sum_{j=1}^{m}\left|h_j(z_j)-\tilde{h}_j(z_j)\right|$$

and

$$|\Phi(\boldsymbol{h})-\Phi(\tilde{\boldsymbol{h}})| \leq \sum_{j=1}^{m}\mathsf{E}\left[|h_j-\tilde{h}_j|(Z_j)\right] = \|\boldsymbol{h}-\tilde{\boldsymbol{h}}\|_{\mathcal{L}^1(P_0,\mathbb{R}^m)}. \qquad \square$$

The next step is to establish the directional derivative of Φ. For that purpose, the following representation becomes useful:

3.5 Lemma: *Let $\boldsymbol{u}=u_1\times u_2\times\cdots\times u_m \in \mathcal{E}$. Assume $u_1, u_2, \ldots, u_m$ are increasing homeomorphisms. Define $v : \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ by*

$$v(t;x) = \sum_{j=1}^{m}(u_j^{-1}(x)-t)_+ = \sum_{j=1}^{m}\begin{cases} u_j^{-1}(x)-t, & u_j^{-1}(x)\geq t,\\ 0, & \textit{otherwise,}\end{cases}$$

and let $v^{-1}(t;\cdot)$ denote the inverse function of $v(t;\cdot)$ on $\mathbb{R}_+$ for every $t\in\mathbb{R}$. Then, it follows

$$\Phi(\boldsymbol{u}-\mathrm{id}_{\mathbb{R}^m}) = \iint_0^\infty v^{-1}(t;z)\,\exp(-z)\,\mathrm{d}z\,\rho_T(t)\,\mathrm{d}t.$$

Proof: Let $\tilde{\phi}=\phi(\boldsymbol{u}-\mathrm{id}_{\mathbb{R}^m};\cdot)$ and denote the standard exponential distribution by $\boldsymbol{\varepsilon}$; that is, for each measurable $A\subseteq\mathbb{R}$ it follows

$$\boldsymbol{\varepsilon}(A) = \int_{A\cap(0,\infty)}\exp(-x)\,\mathrm{d}x.$$

Like usual, the conditional distribution of $\tilde{\phi}(\boldsymbol{Z})$ given T is characterized by

its value on $(-\infty, x]$ for each $x \in \mathbb{R}$:

$$\begin{aligned}
\mathsf{P}\big\{\tilde{\phi}(\boldsymbol{Z}) \le z \mid T\big\} &= 1 - \mathsf{P}\big\{\forall 1 \le j \le m : u_j(Z_j) > x \mid T\big\} \\
&= 1 - \mathsf{P}\big\{\forall 1 \le j \le m : Z_j > u_j^{-1}(x) \mid T\big\} \\
&= 1 - \int_{u_1^{-1}(x)}^{\infty} \cdots \int_{u_m^{-1}(x)}^{\infty} \rho_{\boldsymbol{Z}|T}(\boldsymbol{z} \mid T)\, \mathrm{d}z_m \cdots\, \mathrm{d}z_1 \\
&= 1 - \prod_{j=1}^{m} \exp(-(u_j^{-1}(x) - T)_+) \\
&= 1 - \exp(-v(T; x)) \\
&= \boldsymbol{\varepsilon}(v(T; (-\infty, x])).
\end{aligned}$$

That is, we have

$$\mathsf{P}\big\{\tilde{\phi}(\boldsymbol{Z}) \in \cdot \mid T\big\} = \boldsymbol{\varepsilon} \circ v(T; \cdot)$$

and

$$\begin{aligned}
\mathsf{E}\big[\mathsf{E}\big[\tilde{\phi}(\boldsymbol{Z}) \mid T\big]\big] &= \iint x \, \mathrm{d}\, \boldsymbol{\varepsilon} \circ v(t; x)\, \rho_T(t)\, \mathrm{d}t \\
&= \iint_0^{\infty} v^{-1}(t; z)\, \exp(-z)\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t. \qquad \square
\end{aligned}$$

3.1.2 Directional Derivative of the Objective

The computation of the directional derivative of Φ is performed in two steps. First we assume that the direction $\boldsymbol{h}$ is smooth and compactly supported, then we extend the result to $\mathcal{L}^1(P_0, \mathbb{R}^m)$ functions by an approximation argument.

The Smooth Case

3.6 Assumption: *Let $\boldsymbol{h} = h_1 \times h_2 \times \cdots \times h_m : \mathbb{R}^m \to \mathbb{R}^m$ be smooth and compactly supported. For each $j \in \{1, 2, \ldots, m\}$ define $u_j : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by*

$$u_j(\theta; z) = z + \theta h_j(z)$$

and $\boldsymbol{u} : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ by

$$\boldsymbol{u}(\theta; \boldsymbol{z}) = \big(u_j(\theta; z_j)\big)_{1 \le j \le m}.$$

3.7 Remark:

1. *For each $\theta \in \mathbb{R}$ the random vector $\boldsymbol{u}(\theta; \boldsymbol{Z})$ is integrable, as $\boldsymbol{Z}$ is integrable and $\boldsymbol{h}$ is bounded.*

2. *For a given $\boldsymbol{h}$ there exists an open and bounded interval $\Theta \subseteq \mathbb{R}$ containing 0 such that for each θ in the closure $\bar{\Theta}$ of Θ and each $j \in \{1, 2, \dots, m\}$ the function $u_j(\theta; \cdot)$ is increasing and bilipschitz with constant $\frac{1}{2}$. Particularly, $u_j(\theta; \cdot)$ is an increasing homeomorphism.*

3.8 Lemma: *Under Assumption 3.6 on page 36: Let Θ be given as in Remark 3.7. For each $j \in \{1, 2, \dots, m\}$ and $\theta \in \Theta$ denote the inverse function of $u_j(\theta; \cdot)$ by $u_j^{-1}(\theta; \cdot)$. Further, define $v : \Theta \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by*

$$v(\theta, t; x) = \sum_{j=1}^{m} (u_j^{-1}(\theta; x) - t)_+.$$

Then, for each $(\theta, t) \in \Theta \times \mathbb{R}$ the function $v(\theta, t; \cdot) : \mathbb{R} \to \mathbb{R}_+$ is bilipschitz on its support

$$\operatorname{Supp} v(\theta, t; \cdot) = \Big\{ x \in \mathbb{R} \mid x \geq \min_{1 \leq j \leq m} u_j(\theta; t) \Big\}.$$

Further, $v^{-1} : \Theta \times \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is Lipschitz continuous and is smooth on the complement of

$$N = \bigcup_{j=1}^{m} \Big\{ \Big(\theta, t, v(\theta, t; u_j(\theta; t)) \Big) \mid (\theta, t) \in \Theta \times \mathbb{R} \Big\}.$$

Remark: *Since N is a union of graphs of continuous functions, N has measure zero in $\mathbb{R}^3$. Similarly, for each $\theta \in \Theta$ the set*

$$N_\theta := \{(t, z) \mid (\theta, t, z) \in N\}$$

has measure zero in $\mathbb{R}^2$. In particular, for each $\theta \in \Theta$ and almost every $(t, z) \in \mathbb{R} \times \mathbb{R}_+$ the partial derivative

$$\partial_\theta v^{-1}(\theta, t; z) := \frac{\mathrm{d}}{\mathrm{d}\theta} \left[v^{-1}(\theta, t; z) \right]$$

exists, and $\theta \mapsto \partial_\theta v^{-1}(\theta, \cdot\,; \cdot)$ is continuous with respect to convergence almost everywhere on Θ.

Proof: Under the given assumption, the function

$$\begin{bmatrix} \theta \\ t \\ x \end{bmatrix} \mapsto \begin{bmatrix} \theta \\ t \\ v(\theta, t; x) \end{bmatrix}$$

is a bilipschitz function from the support of v onto $\Theta \times \mathbb{R} \times \mathbb{R}_+$. Thus, v^{-1} is Lipschitz continuous.

As for smoothness: Fix $(\theta, t, z) \in (\Theta \times \mathbb{R} \times \mathbb{R}_+) \setminus N$. Due to the continuity of v and $u_1, u_2, \dots, u_m$, there exits a neighborhood U of (θ, t, z) which is disjunct from N. Particularly, v is smooth on

$$\tilde{U} = \left\{ (\tilde{\theta}, \tilde{t}, v^{-1}(\tilde{\theta}, \tilde{t}; \tilde{z})) \mid (\tilde{\theta}, \tilde{t}, \tilde{z}) \in U \right\}$$

and hence v^{-1} is smooth on U. □

Combining Lemmas 3.5 and 3.8, we can compute the directional derivative of Φ as follows:

3.10 Theorem:
Under Assumption 3.6 on page 36: Let Θ be given as in Remark 3.7. Then, the function

$$\theta \mapsto \Phi(\theta \boldsymbol{h})$$

is continuously differentiable on Θ with

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \left[\Phi(\theta \boldsymbol{h}) \right]_{\theta=0} = \sum_{j=1}^{m} \iint_t^\infty h_j(z)\, \mathrm{e}^{-m(z-t)}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t. \tag{3.1}$$

Remark: *Since Φ is Lipschitz continuous, $\theta \mapsto \Phi(\theta \boldsymbol{h})$ is also almost everywhere differentiable. However, on Θ, the derivative is continuous and admits the explicit integral form given by Equation* (3.1).

Proof: For each $\theta \in \Theta$, Lemma 3.5 states

$$\Phi(\theta \boldsymbol{h}) = \iint_0^\infty v^{-1}(\theta, t; z)\, \mathrm{e}^{-z}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t.$$

The idea is to differentiate v^{-1} under the integral sign with respect to θ (see Appendix A.2.2 on page 75).

Define N as in Lemma 3.8 and denote its complement by

$$N^{\mathsf{C}} = (\Theta \times \mathbb{R} \times \mathbb{R}_+) \setminus N.$$

That is, v^{-1} is smooth on N^{C}. For each $j \in \{1, 2, \dots, m\}$ and $(\theta, t, x) \in \Theta \times \mathbb{R} \times \mathbb{R}$ let

$$v_j(\theta, t; x) = (u_j^{-1}(\theta; x) - t)_+.$$

By the chain rule, for each $(\theta, t, z) \in S$ we have

$$\begin{aligned}\partial_\theta v^{-1}(\theta, t; z) &= -\partial_\theta v(\theta, t; v^{-1}(\theta, t; z)) \cdot \partial_z v^{-1}(\theta, t; z) \\ &= -\sum_{j=1}^{m} \partial_\theta v_j(\theta, t; v^{-1}(\theta, t; z)) \cdot \partial_z v^{-1}(\theta, t; z).\end{aligned}$$

Observe that for each (θ, t, x) in the interior of $\operatorname{Supp} v_j$ it follows

$$\partial_\theta v_j(\theta, t; x) = -h_j(t + v_j(\theta, t; x)) \cdot \partial_x v_j(\theta, t; x).$$

The formula remains valid for each (θ, t, x) in the complement of $\operatorname{Supp} v_j$, as $\partial_x v_j(\theta, t; x)$ is simply 0. Thus, we have

$$\begin{aligned}&\partial_\theta v^{-1}(\theta, t; z) \\ &\quad= \sum_{j=1}^{m} h_j(t + v_j(\theta, t; v^{-1}(\theta, t; z))) \cdot \partial_x v_j(\theta, t; v^{-1}(\theta, t; z)) \cdot \partial_z v^{-1}(\theta, t; z) \\ &\quad= \sum_{j=1}^{m} h_j(t + v_j(\theta, t; v^{-1}(\theta, t; z))) \cdot \frac{\mathrm{d}}{\mathrm{d}z} \left[v_j(\theta, t; v^{-1}(\theta, t; z)) \right].\end{aligned}$$

Integration by substitution with $w_j(\theta, t; z) = t + v_j(\theta, t; v^{-1}(\theta, t; z))$ yields

$$\begin{aligned}&\iint_0^\infty |h_j(w_j(\theta, t; z))| \cdot \partial_z w_j(\theta, t; z) \cdot \mathrm{e}^{-z} \,\mathrm{d}z \,\rho_T(t) \,\mathrm{d}t \\ &\qquad\qquad = \iint_t^\infty \underbrace{|h_j(z)| \,\mathrm{e}^{-w_j^{-1}(\theta, t; z)}}_{(*)} \,\mathrm{d}z \,\rho_T(t) \,\mathrm{d}t.\end{aligned}$$

The integrand $(*)$ on the right is continuous as a function of (θ, t, z). Further, as

$$w_j^{-1}(\theta, t; z) = v(\theta, t; u_j(\theta; z)) \geq v_j(\theta, t; u_j(\theta; z)) = (z - t)_+$$

holds for every $(\theta, t, z) \in \Theta \times \mathbb{R} \times \mathbb{R}$, the integrand $(*)$ is also dominated by $(t, z) \mapsto |h_j(z)|\mathrm{e}^{-(z-t)_+}$, which is integrable by assumption. Thus, $\theta \mapsto \partial_\theta v^{-1}(\theta, \cdot\,; \cdot\,)$ is a continuous function in $\mathcal{L}^1$. In particular, we may differentiate v^{-1} under the integral sign (Theorem A.7 on page 75) and obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\theta}[\Phi(\theta \boldsymbol{h})] &= \iint_0^\infty \partial_\theta v^{-1}(\theta, t; z)\, \mathrm{e}^{-z}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t \\ &= \sum_{j=1}^m \iint_t^\infty h_j(z)\, \mathrm{e}^{-v(\theta,t;u_j(\theta;z))}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t,\end{aligned}$$

which is continuous in θ on Θ. Finally, at $\theta = 0$, we have

$$v(0, t; u_j(0; z)) = m(z - t)$$

for $t \in \mathbb{R}$ and $z \in [t, \infty)$, which yields the statement. □

The General Case

3.12 Theorem:
Define $A : \mathcal{E} \to \mathbb{R}$ by

$$A\boldsymbol{h} = \sum_{j=1}^m \iint_t^\infty h_j(z)\, \mathrm{e}^{-m(z-t)}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t. \tag{3.2}$$

Then, A is continuous and is the (directional) derivative $\Phi'(\mathbf{0})$ of Φ at $\mathbf{0}$.

Proof: For the first part, notice that $m(z-t)_+ \geq (z-t)_+$ for every $t, z \in \mathbb{R}$. Thus, A is linearly bounded:

$$|A\boldsymbol{h}| \leq \sum_{j=1}^m \iint_t^\infty |h_j(z)|\, \mathrm{e}^{-(z-t)}\, \mathrm{d}z\, \rho_T(t)\, \mathrm{d}t = \sum_{j=1}^m \mathsf{E}\big[|h_j(Z_j)|\big] = \|\boldsymbol{h}\|_{\mathcal{L}^1(P_0, \mathbb{R}^m)}.$$

As for directional differentiability, in case $\boldsymbol{h}$ is smooth and compactly supported, Theorem 3.10 implies the statement. For the general case: As P_0 is absolutely continuous with respect to the Lebesgue measure, the set of compactly supported smooth functions is dense in $\mathcal{L}^1(P_0, \mathbb{R}^m)$ (compare to the literature, like [8]). Let $\boldsymbol{h} \in \mathcal{E}$ and $\boldsymbol{h}_n \in \mathcal{S}ep_m \cap \mathcal{C}_c^\infty(\mathbb{R}^m; \mathbb{R}^m)$ with

$\boldsymbol{h}_n \to \boldsymbol{h}$ in $\mathcal{L}^1(P_0, \mathbb{R}^m)$. For every $\theta \in \mathbb{R} \setminus \{0\}$ it follows

$$\begin{aligned}
&\left|\frac{\Phi(\theta\boldsymbol{h}) - \Phi(\mathbf{0})}{\theta} - A\boldsymbol{h}\right| \\
&\qquad \le \frac{|\Phi(\theta\boldsymbol{h}) - \Phi(\theta\boldsymbol{h}_n)|}{|\theta|} + \left|\frac{\Phi(\theta\boldsymbol{h}_n) - \Phi(\mathbf{0})}{\theta} - A\boldsymbol{h}\right| \\
&\qquad \le \|\boldsymbol{h} - \boldsymbol{h}_n\|_{\mathcal{L}^1(P_0,\mathbb{R}^m)} + |A\boldsymbol{h}_n - A\boldsymbol{h}| + \left|\frac{\Phi(\theta\boldsymbol{h}_n) - \Phi(\mathbf{0})}{\theta} - A\boldsymbol{h}_n\right|.
\end{aligned}$$

In particular, for every $n \in \mathbb{N}$ we have

$$\limsup_{\theta\to 0}\left|\frac{\Phi(\theta\boldsymbol{h}) - \Phi(\mathbf{0})}{\theta} - A\boldsymbol{h}\right| \le 2\|\boldsymbol{h} - \boldsymbol{h}_n\|_{\mathcal{L}^1(P_0,\mathbb{R}^m)},$$

where the right-hand side converges to 0 for $n \to \infty$. □

3.1.3 The Proof of Claim 1

Turning to the actual proof of Theorem 3.2 on page 33: Theorem 3.12 already implies that $\mathbf{0}$ is *one* maximizer of Φ, as for every $\boldsymbol{h} \in \mathcal{S}$ we have

$$\begin{aligned}
\Phi(\boldsymbol{h}) - \Phi(\mathbf{0}) &\le \Phi'(\mathbf{0})[\boldsymbol{h}] \\
&= \iint_t^\infty \underbrace{\sum_{j=1}^m h_j(z)}_{0}\ \mathrm{e}^{-m(z-t)}\,\mathrm{d}z\,\rho_T(t)\,\mathrm{d}t = 0.
\end{aligned}$$

To ensure the uniqueness, we use the following result:

3.13 Lemma: *Let $\boldsymbol{h} \in \mathcal{E}$ with $\Phi(\boldsymbol{h}) = \Phi(\mathbf{0})$. Then, there exists a set $N \subseteq \mathbb{R}$ of Lebesgue measure zero such that*

1. *every $z \in \mathbb{R} \setminus N$ is a Lebesgue point*[a] *of $h_1, h_2, \dots, h_m$ and it follows*

$$\frac{1}{m} \sum_{j=1}^{m} h_j(z) = 0;$$

2. *for each $\boldsymbol{z} \in (\mathbb{R} \setminus N)^m$ the function $\theta \mapsto \phi(\theta \boldsymbol{h}; \boldsymbol{z})$ is affine on $[0, 1]$;*

3. *in particular, for each $\boldsymbol{z} \in (\mathbb{R} \setminus N)^m$ it follows*

$$J^*(\boldsymbol{z}) := \left(\operatorname*{arg\,min}_{1 \le j \le m} z_j \right) \cap \left(\operatorname*{arg\,min}_{1 \le j \le m} z_j + h_j(z_j) \right) \neq \emptyset.$$

Proof: Ad Item 1: Notice that h_j is locally Lebesgue integrable for each $j \in \{1, 2, \dots, m\}$. Thus, there exists a zero measure set $N_j \subseteq \mathbb{R}$ such that $\mathbb{R} \setminus N_j$ contains only Lebesgue points of h_j. Further, by the definition of $\mathcal{S}$, the mean of $h_1, h_2, \dots, h_m$ is the zero function on the complement of a zero measure set, say $N_0 \subseteq \mathbb{R}$. Now, the finite union

$$N := \bigcup_{j=0}^{m} N_j$$

has measure zero again and has the properties stated in Item 1.

Ad Item 2: Due to concavity, for every $\theta \in [0, 1]$ we have

$$\phi(\theta \boldsymbol{h}; \cdot) - ((1 - \theta)\phi(\boldsymbol{0}; \cdot) + \theta \phi(\boldsymbol{h}; \cdot)) \geq 0.$$

By assumption $\boldsymbol{0}$ and $\boldsymbol{h}$ are maximizers of Φ. Thus, we also have

$$\begin{aligned} 0 &\geq \Phi(\theta \boldsymbol{h}) - \Phi(\boldsymbol{0}) \\ &= \Phi(\theta \boldsymbol{h}) - ((1 - \theta)\Phi(\boldsymbol{0}) + \theta \Phi(\boldsymbol{h})) \\ &= \mathsf{E}[\phi(\theta \boldsymbol{h}; \boldsymbol{Z}) - ((1 - \theta)\phi(\boldsymbol{0}; \boldsymbol{Z}) + \theta \phi(\boldsymbol{h}; \boldsymbol{Z}))] \geq 0. \end{aligned}$$

[a] A point $\boldsymbol{x} \in \mathbb{R}^n$ is called a *Lebesgue point* (also known as *density point*) of a measurable function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ if

$$\lim_{\delta \to 0+} \frac{1}{\boldsymbol{\lambda}(B_\delta)} \int_{B_\delta} \| \boldsymbol{f}(\boldsymbol{x} + \boldsymbol{h}) - \boldsymbol{f}(\boldsymbol{x}) \| \, \mathrm{d}\boldsymbol{h} = 0,$$

where B_δ denotes the ball with radius $\delta > 0$ centered at $\boldsymbol{0}$. The definition does not depend on the norm on $\mathbb{R}^n$ and $\mathbb{R}^m$.

Trivially, if $\boldsymbol{f}$ is continuous at $\boldsymbol{x} \in \mathbb{R}^n$, then $\boldsymbol{x}$ is a Lebesgue point. Further, the *Lebesgue differentiation theorem* states, that if $\boldsymbol{f}$ is locally integrable, then almost every point of $\mathbb{R}^n$ is a Lebesgue point [8].

That is, we have almost surely

$$\phi(\theta \boldsymbol{h}; \boldsymbol{Z}) = (1-\theta)\phi(\mathbf{0}; \boldsymbol{Z}) + \theta\phi(\boldsymbol{h}; \boldsymbol{Z}).$$

Thus, it is sufficient to show that every $\boldsymbol{z} \in (\mathbb{R} \setminus N)^m$ is a Lebesgue point of $\phi(\theta \boldsymbol{h}; \cdot)$ for every $\theta \in [0,1]$. As the minimum function admits the Lipschitz constant 1 with respect to $\|\cdot\|_\infty$, we have for every $\tilde{\boldsymbol{z}} \in \mathbb{R}^m$

$$\begin{aligned}|\phi(\theta \boldsymbol{h}; \tilde{\boldsymbol{z}}) - \phi(\theta \boldsymbol{h}; \boldsymbol{z})| &\leq \sum_{j=1}^{m} \big|\tilde{z}_j + \theta h_j(\tilde{z}_j) - \big(z_j + \theta h_j(z_j)\big)\big| \\ &\leq \sum_{j=1}^{m} |\tilde{z}_j - z_j| + \theta \sum_{j=1}^{m} |h_j(\tilde{z}_j) - h_j(z_j)|.\end{aligned}$$

Now, the average of the right-hand side on the δ-neighborhood of $\boldsymbol{z}$ tends to 0 for $\delta \to 0$. Hence, $\boldsymbol{z}$ is a Lebesgue point of $\phi(\theta \boldsymbol{h}; \cdot)$. □

Concluding the proof of Theorem 3.2: Let $\boldsymbol{h} = h_1 \times h_2 \times \cdots \times h_m \in \mathcal{S}$ with $\Phi(\boldsymbol{h}) = \Phi(\mathbf{0})$. Let N and J^* be defined as in Lemma 3.13. Fix $\tilde{z} \in \mathbb{R} \setminus N$. For each $\varepsilon > 0$ and $j \in \{1, 2, \dots, m\}$ define

$$O_\varepsilon(\tilde{z}, j) = (\tilde{z}, \tilde{z} + \varepsilon)^{j-1} \times (\tilde{z} - \varepsilon, \tilde{z}) \times (\tilde{z}, \tilde{z} + \varepsilon)^{m-j}.$$

For each $\varepsilon > 0$, $j_1, j_2 \in \{1, 2, \dots, m\}$, and $\boldsymbol{z} \in (\mathbb{R} \setminus N)^m \cap O_\varepsilon(\tilde{z}, j_1)$ it follows

$$z_1, \dots, z_{j_1-1}, z_{j_1+1}, z_m > \tilde{z} > z_{j_1},$$

also

$$J^*(\boldsymbol{z}) = \{j_1\},$$

and finally

$$z_{j_1} + h_{j_1}(z_{j_1}) \leq \min_{1 \leq j \leq m} z_j + h_j(z_j) \leq z_{j_2} + h_{j_2}(z_{j_2}).$$

In particular, we have

$$\begin{aligned}0 &\leq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^m} \int_{O_\varepsilon(\tilde{z}, j_1)} (z_{j_2} + h_{j_2}(z_{j_2})) - (z_{j_1} + h_{j_1}(z_{j_1}))\, \mathrm{d}\boldsymbol{z} \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\tilde{z}}^{\tilde{z}+\varepsilon} z_{j_2} + h_{j_2}(z_{j_2})\, \mathrm{d}z_{j_2} - \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{\tilde{z}-\varepsilon}^{\tilde{z}} z_{j_1} + h_{j_1}(z_{j_1})\, \mathrm{d}z_{j_1} \\ &= h_{j_2}(\tilde{z}) - h_{j_1}(\tilde{z}),\end{aligned}$$

as $\tilde{z}$ is a Lebesgue point of h_{j_1} and h_{j_2}. We conclude

$$h_{j_1}(\tilde{z}) = h_{j_2}(\tilde{z}),$$

and as j_1 and j_2 are arbitrary, also

$$h_1(\tilde{z}) = h_2(\tilde{z}) = \cdots = h_m(\tilde{z}).$$

In particular, for every $j \in \{1, 2, \dots, m\}$ it follows

$$h_j(\tilde{z}) = \frac{1}{m} \sum_{j'=1}^{m} h_{j'}(\tilde{z}) = 0.$$

As $\tilde{z} \in \mathbb{R} \setminus N$ is arbitrary, it follows $h_j = 0$ almost everywhere for every $j \in \{1, 2, \dots, m\}$.

3.2 Claim 2

Does the solutions of the sample approximation, Problem 2.1 on page 17, converge to the solution of Problem 1.5 on page 12? Unfortunately, that question can not be answered in full generality. However, the answer is positive if the solution $\breve{\boldsymbol{u}}$ is determined by its value on a known compact set (comparing to Section 2.4 on page 24). The main idea is to apply the results about so-called extremum estimators described in Appendix B on page 77.

3.2.1 The Setting

This section follows the established notations and assumptions of Problem 2.1 on page 17. However, we need to strengthen some assumptions.

The Function Space

Unlike in Problem 2.1, let $\mathcal{E} \subseteq \mathcal{S}ep_m \cap \mathcal{L}^1(P_0, \mathbb{R}^m)$ be a Banach space with norm $\|\cdot\|_{\mathcal{E}}$ such that $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$ is continuously embedded into $\mathcal{L}^1(P_0, \mathbb{R}^m)$. Further, let $\mathcal{F} \subseteq \mathcal{E}$ be closed, convex, and locally compact as a topological subspace. For example, the following uniformly Lipschitz continuous subset

$$\{u \in \mathcal{L}ip([0,1], \mathbb{R}) \mid \|u'\|_\infty \leq 1\} \subseteq \mathcal{C}([0,1], \mathbb{R}),$$

is closed, convex, and locally compact by the Arzelà-Ascoli theorem (compare to the literature, like [1]).

Let $I_0 = [\underline{\tau}, \bar{\tau}] \subseteq (0,1)$ with $\underline{\tau} < \bar{\tau}$. Let $V \subseteq \mathbb{R}^m$ be bounded and open with

$$\forall n \in \mathbb{N} \cup \{0\} : \prod_{j=1}^{m} Q_{j,n}(I_0) \subseteq V.$$

Finally, let $\mathcal{F}_n \subseteq \mathcal{F}$ be a closed convex subset.

Extremum Estimator

Let f and g be given as in Problem 1.5. Let $(I_n)_{n\in\mathbb{N}}$ be a sequence of finite partitions of I_0. Let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\eta_n \to \eta_0 := 0$. For each $n \in \mathbb{N}$ let $\boldsymbol{u}_n^*$ denote a minimizer of

$$f_n(\boldsymbol{u}) := \int f(\boldsymbol{u}; \boldsymbol{z}) \, \mathrm{d}P_n(\boldsymbol{z})$$

over

$$\mathcal{S}_n := \left\{ \boldsymbol{u} \in \mathcal{F}_n \,|\, g_n(\boldsymbol{u}) := \max_{\tau \in I_n} g(\boldsymbol{u}; \boldsymbol{Q}_n(\tau)) \leq \eta_n \right\}.$$

3.2.2 Convergence

3.14 Theorem:
Under the setting given by Section 3.2.1 on page 44: Assume that

1. *(Identification)* $\check{\boldsymbol{u}} \in \mathcal{F}_0$ *and for each* $\boldsymbol{u} \in \mathcal{S}_0$ *it follows*

$$\forall \tau \in (0,1) : g(\boldsymbol{u}; \boldsymbol{Q}_0(\tau)) = 0;$$

2. *(Approximation)*

$$\operatorname{dist}(\mathcal{F}_n, \check{\boldsymbol{u}}) = o(\eta_n), \quad \operatorname{dist}(I_n, I_0) = o(\eta_n), \quad \sqrt{(\log\log n)/n} = o(\eta_n);$$

3. *there exists a neighborhood* $\mathcal{U} \subseteq \mathcal{F}$ *of* $\check{\boldsymbol{u}}$ *and some* $C_V > 0$ *with*

$$\forall \boldsymbol{u} \in \mathcal{U} : \|\boldsymbol{u}\|_{\infty,V} \leq C_V \|\boldsymbol{u}\|_{\mathcal{E}}$$

and

$$L := \sup_{\boldsymbol{u} \in \mathcal{U}} \|\boldsymbol{u}'\|_{\infty,V} < \infty;$$

4. and the quantile functions $Q_1, Q_2, \dots, Q_m$ are locally Lipschitz continuous.

Then, $\|\boldsymbol{u}_n^* - \check{\boldsymbol{u}}\|_{\mathcal{E}}$ converges to 0 almost surely.

Notice that o denotes the little-o *Landau symbol* and dist the Hausdorff distance in Item 2.

Proof: We need to check the assumptions of Theorem B.8 on page 86. Item 1–3 of Theorem B.8 are given by the assumptions.

Concerning Item 4: We show

$$\lim_{\delta \to 0} \int \omega_f(\delta, \mathcal{F} \times \{\boldsymbol{z}\}) \, \mathrm{d}P_0(\boldsymbol{z}) = 0$$

where

$$\omega_f(\delta, \mathcal{F} \times \{\boldsymbol{z}\}) = \sup\{|f(\tilde{\boldsymbol{u}}; \boldsymbol{z}) - f(\boldsymbol{u}; \boldsymbol{z})| \mid \boldsymbol{u}, \tilde{\boldsymbol{u}} \in \mathcal{F}, \|\tilde{\boldsymbol{u}} - \boldsymbol{u}\|_{\mathcal{E}} < \delta\}$$

denotes the *modulus of continuity* of f with respect to the first argument. Since $\mathcal{E}$ is continuously embedded in $\mathcal{L}^1(P_0, \mathbb{R}^m)$, there exists some $C_{\mathcal{L}^1} > 0$ with

$$\|\cdot\|_{\mathcal{L}^1(P_0, \mathbb{R}^m)} \leq C_{\mathcal{L}^1} \|\cdot\|_{\mathcal{E}}.$$

Thus, for each $\boldsymbol{u}, \tilde{\boldsymbol{u}} \in \mathcal{F}$ it follows

$$\begin{aligned} \int |f(\boldsymbol{u}; \boldsymbol{z}) - f(\tilde{\boldsymbol{u}}; \boldsymbol{z})| \, \mathrm{d}P_0(\boldsymbol{z}) &\leq 2 \int \max_{1 \leq j \leq m} \left|u_j(z_j) - \tilde{u}_j(z_j)\right| \mathrm{d}P_0(\boldsymbol{z}) \\ &\leq 2\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_{\mathcal{L}^1(P_0, \mathbb{R}^m)} \\ &\leq 2C_{\mathcal{L}^1} \|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_{\mathcal{E}}, \end{aligned}$$

which converges to 0 for $\|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_{\mathcal{E}} \to 0$.

Concerning Item 5: Notice that g is Lipschitz continuous on $\mathcal{U} \times V$, as for each $\boldsymbol{u}, \tilde{\boldsymbol{u}} \in \mathcal{U}$ and $\boldsymbol{z}, \tilde{\boldsymbol{z}} \in V$ it follows

$$|g(\boldsymbol{u}; \boldsymbol{z}) - g(\tilde{\boldsymbol{u}}; \boldsymbol{z})| \leq \max_{1 \leq j \leq m} |u_j(z_j) - \tilde{u}_j(z_j)| \leq C_V \|\boldsymbol{u} - \tilde{\boldsymbol{u}}\|_{\mathcal{E}},$$

and

$$|g(\boldsymbol{u}; \boldsymbol{z}) - g(\boldsymbol{u}; \tilde{\boldsymbol{z}})| \leq \max_{1 \leq j \leq m} |u_j(z_j) - u_j(\tilde{z}_j)| \leq L \|\boldsymbol{z} - \tilde{\boldsymbol{z}}\|_{\infty}.$$

Thus, we only need to prove

$$\operatorname{dist}(\boldsymbol{Q}_n(I_n), \boldsymbol{Q}_0(I_0)) = \mathcal{o}(\eta_n).$$

Let $\bar{I} \subseteq (0,1)$ be a compact interval with $V \subseteq \prod_{j=1}^{m} Q_{j,0}(\bar{I})$. By Theorem C.3 on page 91, it follows almost surely

$$\|\boldsymbol{Q}_n - \boldsymbol{Q}_0\|_{\infty,\bar{I}^m} = \mathcal{O}\left(\sqrt{\frac{\log\log n}{n}}\right),$$

where $\mathcal{O}$ denotes the big-o Landau symbol. Let C be the Lipschitz constant of $\boldsymbol{Q}_0$ on $\bar{I}^m$. Notice that for each $n \in \mathbb{N}$ we have $I_n \subseteq I_0 \subseteq \bar{I}$ and thus

$$\operatorname{dist}(\boldsymbol{Q}_n(I_n), \boldsymbol{Q}_0(I_0)) \leq \|\boldsymbol{Q}_n - \boldsymbol{Q}_0\|_{\infty,\bar{I}^m} + C \operatorname{dist}(I_n, I_0) = \mathcal{o}(\eta_n). \qquad \square$$

3.3 Examples

To illustrate the consistency result, this section presents some examples for $\mathcal{E}$, $\mathcal{F}_0, \mathcal{F}_1, \ldots$ We will begin with an affine linear inverse clock model, and then expand the example to the B-spline based inverse clock model found in Section 2.4 on page 24.

For the sake of completeness: Let $\underline{\tau}, \bar{\tau} \in (0,1)$ with $\underline{\tau} < \bar{\tau}$. For $\underline{z}, \bar{z} \in \mathbb{R}$ with

$$\underline{z} < \min_{1 \leq j \leq m} Q_{j,0}(\underline{\tau}) < \max_{1 \leq j \leq m} Q_{j,0}(\bar{\tau}) < \bar{z}$$

define $V = (\underline{z}, \bar{z})^m$. For each $n \in \mathbb{N}$ let

$$\eta_n = \sqrt{(\log\log n)/\sqrt{n}}$$

and

$$I_n = \left\{ \underline{\tau} + (\bar{\tau} - \underline{\tau}) \tfrac{k}{p_n} \,|\, 0 \leq k \leq p_n \right\}$$

with

$$p_n = \operatorname{ceil}\Big(\sqrt{n/(\log\log n)} \Big).$$

Notice that the values for η_n and p_n are for illustration only and are not practical. The value of η_n should be multiplied by a small number in applications, as the asymptotic is much too pessimistic. Also, p_n should depend on the

dimension of the inverse clock models, which again depend on n and irregularity of the assumed clock deviations. Applicable examples are discussed later with simulation results.

Since $\boldsymbol{Q}_n$ converges locally to $\boldsymbol{Q}_0$ almost surely, we have almost surely

$$\max_{1\le j\le m} \operatorname{dist}\big(Q_{j,n}(I_0), Q_{j,0}(I_0)\big) \to 0.$$

That is, for almost each sample and n sufficiently large, we have

$$\prod_{j=1}^{m} Q_{j,n}(I_0) \subseteq V.$$

Further, by construction, it follows

$$\sqrt{(\log\log n)/n} = o(\eta_n),$$

$$\operatorname{dist}(I_n, I_0) = \frac{\bar\tau - \underline\tau}{p_n} = \mathcal{O}(\sqrt{(\log\log n)/n}),$$

and thus

$$\operatorname{dist}(I_n, I_0) = o(\eta_n).$$

The exact values of V, η_n, and I_n have little influence from a theoretical point of view. However, the values of $\underline\tau$ and $\bar\tau$ do have some implication, as shown later.

3.3.1 Affine Linear Inverse Clock Model

Let P be some probability measure on $\mathbb{R}$ with $\int |z| \,\mathrm{d}P(z) < \infty$, and let $K = [\underline{z}, \bar{z}]$. Denote the set of affine linear functions from $\mathbb{R}$ into $\mathbb{R}$ by $\mathcal{A}$, which is trivially a finite dimensional subspace of $\mathcal{L}^1(P, \mathbb{R})$. As $\underline{z} < \bar{z}$, the sup-norm $\|\cdot\|_{\infty,K}$ restricted onto K is also a norm on $\mathcal{A}$. Further, as $\mathcal{A}$ is finite dimensional, and therefore every pair of norms on $\mathcal{A}$ are equivalent, there exist $c, C > 0$ with

$$\forall u \in \mathcal{A} : c\|u\|_{\infty,K} \le \int |u| \,\mathrm{d}P \le C\|u\|_{\infty,K}.$$

In fact, an upper bound C can be given easily: For each $u \in \mathcal{A}$ and $z \in \mathbb{R}$ it follows

$$\begin{aligned}\int |u(z)| \, \mathrm{d}P(z) &\leq |u(\underline{z})| + \left| \frac{u(\bar{z}) - u(\underline{z})}{\bar{z} - \underline{z}} \right| \int |z - \underline{z}| \, \mathrm{d}P(z) \\ &\leq \|u\|_{\infty,K} \underbrace{\left(1 + 2|\bar{z} - \underline{z}|^{-1} \int |z - \underline{z}| \, \mathrm{d}P(z)\right)}_{=:C}.\end{aligned}$$

Also, notice that any bounded subset $\mathcal{U} \subseteq \mathcal{A}$ is uniformly Lipschitz continuous: For each $u \in \mathcal{U}$ it follows

$$\|u'\|_\infty = \left| \frac{u(\bar{z}) - u(\underline{z})}{\bar{z} - \underline{z}} \right| \leq 2|\bar{z} - \underline{z}|^{-1} \|u\|_{\infty,K}.$$

Since this approach does not generalize well to infinite dimensional spaces, we will equip $\mathcal{A}$ with the norm $\|\cdot\|_{\infty,K}$ instead of $\|\cdot\|_{\mathcal{L}^1(P,\mathbb{R})}$.

Adapting for Theorem 3.14

Define

$$\mathcal{E} = \mathcal{F} = \mathcal{F}_0 = \cdots = \{u_1 \times u_2 \times \cdots \times u_m \mid u_1, u_2, \ldots, u_m \in \mathcal{A}\}$$

with

$$\|\boldsymbol{u}\|_{\mathcal{E}} := \max_{1 \leq j \leq m} \|u_j\|_{\infty,K}.$$

If $\mathcal{F}_0$ contains the solution $\check{\boldsymbol{u}}$, then the extremum existmator $\boldsymbol{u}_n^*$ converges to $\check{\boldsymbol{u}}$ almost surely.

In applications, we can simply choose $\underline{\tau}$ and $\bar{\tau}$, as the values of $\underline{z}$ and $\bar{z}$ are not needed for estimation. In general, we can not provide a finite dimensional space which already contains $\check{\boldsymbol{u}}$, and approximation techniques are required. It that case, the specific values of $\underline{\tau}$ and $\bar{\tau}$ become more important.

3.3.2 A Lipschitz Continuous Inverse Clock Model

The affine linear inverse clock model can be generalized in a natural way by adding continuous functions whose support is in $K = [\underline{z}, \bar{z}]$, which coincides with the approach stated in Section 2.4 on page 24. Again, the approach can be extended to include periodic terms.

Let P be some probability measure on $\mathbb{R}$ with $\int |z| \, \mathrm{d}P(z) < \infty$, denote the set of affine linear functions from $\mathbb{R}$ into $\mathbb{R}$ by $\mathcal{A}$, and denote the set of continuous functions from $\mathbb{R}$ into $\mathbb{R}$ with support in K by $\mathcal{C}_0(K, \mathbb{R})$. Let $\mathcal{K}$ be a closed and convex subset of $\mathcal{C}_0(K, \mathbb{R})$ which is uniformly Lipschitz continuous; that is

$$\sup_{v \in \mathcal{K}} \|v'\|_\infty < \infty.$$

Now, consider the Banach space

$$\mathcal{A} + \mathcal{C}_0(K, \mathbb{R}) = \{v + w \mid v \in \mathcal{A}, w \in \mathcal{C}_0(K, \mathbb{R})\}$$

equipped with the sup-norm $\|\cdot\|_{\infty, K}$ on K, and the closed convex subset

$$\mathcal{A} + \mathcal{K} = \{v + w \mid v \in \mathcal{A}, w \in \mathcal{K}\}.$$

Notice that $\mathcal{A} + \mathcal{C}_0(K, \mathbb{R})$ is continuously embedded into $\mathcal{L}^1(P, \mathbb{R})$, and $\mathcal{A} + \mathcal{K}$ is locally compact, as $\mathcal{A}$ is finite dimensional and $\mathcal{K}$ is compact by Arzelà-Ascoli theorem[1]. In particular, each bounded subset of $\mathcal{A} + \mathcal{K}$ is compact and uniformly Lipschitz continuous. Further, $\mathcal{A} + \mathcal{C}_0(K, \mathbb{R})$ is isometrically isomorphic to $\mathcal{C}(K, \mathbb{R})$, again by restricting onto K, and $\mathcal{A} + \mathcal{K}$ is isometrically isomorphic to some locally compact subset of $\mathcal{C}(K, \mathbb{R})$.

Adapting for Theorem 3.14

Let $\mathcal{L}ip_0(K, \mathbb{R}; L)$ denote the set of Lipschitz continuous function from $\mathbb{R}$ into $\mathbb{R}$ with constant L and support in K. Define

$$\mathcal{E} = \{u_1 \times u_2 \times \cdots \times u_m \mid u_1, u_2, \ldots, u_m \in \mathcal{A} + \mathcal{C}_0(K, \mathbb{R})\},$$
$$\mathcal{F} = \big\{u_1 \times u_2 \times \cdots \times u_m \mid u_1, u_2, \ldots, u_m \in \mathcal{A} + \mathcal{L}ip_0(K, \mathbb{R}; L)\big\}.$$

For each $n \in \mathbb{N} \cup \{0\}$ and $j \in \{1, 2, \ldots, m\}$ let $\mathcal{K}_{j,n} \subseteq \mathcal{L}ip_0(Q_{j,n}(I_0); L_n)$ be closed and convex, and define

$$\mathcal{F}_n = \big\{u_1 \times u_2 \times \cdots \times u_m \mid \forall j \in \{1, 2, \ldots, m\} : u_j \in \mathcal{A} + \mathcal{K}_{j,n}\big\}.$$

Assume that almost surely

$$L_0 \leq \liminf_{n \to \infty} L_n \leq \limsup_{n \to \infty} L_n \leq L$$

and that almost surely

$$\forall j \in \{1, 2, \dots, m\}, u \in \mathcal{K}_{j,0} : \operatorname{dist}\big(\mathcal{K}_{j,n}, u\big) = \mathcal{O}(\sqrt{(\log\log n)/n}).$$

If $\mathcal{F}_0$ contains $\breve{\boldsymbol{u}}$, then the extremum estimator $\boldsymbol{u}_n^*$ converges to $\breve{\boldsymbol{u}}$ almost surely.

Concerning the Lipschitz constants, L and L_0 are deterministic, and L_n is an estimate for L_0. The value of L is irrelevant. However, L_n needs to be consistent in the sense stated above. Again, estimating L_0 has little practical value. In applications we simply guess L_0 and set L_n to a large value accordingly, or completely ignore L_n.

Concerning the compact subsets $\mathcal{A} + \mathcal{K}_{j,n}$: They are random, as they depend on $Q_{j,n}(I_0)$, a random compact interval, and on L_n. A simple and practical choice for $\mathcal{K}_{j,n}$ are the space of splines, as laid out in Section 2.4.

3.4 Incomplete Log-Sets, continued II

Concerning the case of incomplete log-sets (Problem 1.7 on page 14 with its sample approximation Problem 2.5 on page 23), the solutions of Problem 2.5 converge to the unique solution of Problem 1.7 almost surely under similar assumptions imposed in this chapter. The structure of Problems 1.7 and 2.5 is the same as one of Problems 1.5 and 2.1 on page 12 and on page 17.

Notice that we can simplify Problem 2.5 by exploiting Theorem 3.12 and the linearity of differentiation. Assume that the network does not split into two independent parts; that is

3.15 Assumption: *For each non-empty $J \subseteq \{1, 2, \dots, m\}$ there exist $j_1 \in J$ and $j_2 \in \{1, 2, \dots, m\} \setminus J$ with*

$$\mathbf{P}\big\{Z_{j_1}, Z_{j_2} \in \mathbb{R}\big\} > 0.$$

Further, if the sets in $\mathscr{J}$ of Problem 1.7 have the same cardinality, then Problem 1.7 and the following problem have the same optimizer:

3.16 Problem: *With the same notation as in Problem 1.7 and assumed that $m_0 = |J|$ for each $J \in \mathscr{J}$: Minimize*

$$\frac{1}{|\mathscr{J}|} \sum_{J \in \mathscr{J}} \mathbf{E}\big[f_J(\boldsymbol{u}; \boldsymbol{Z}) \,\big|\, \boldsymbol{Z}_J \in \mathbb{R}^J\big]$$

subject to

$$\operatorname*{ess\,sup}_{\tau\in(0,1)} \tilde{g}(\boldsymbol{u}; \boldsymbol{Q}_0(\tau)) = 0,\ \boldsymbol{u} \in \mathcal{E},$$

where $\tilde{g} : \mathcal{E} \times \mathbb{R}^m \to \mathbb{R}$ *is defined by*

$$\tilde{g}(\boldsymbol{u}; \boldsymbol{z}) = \left| \frac{1}{m_0 |\mathscr{J}|} \sum_{J\in\mathscr{J}} \sum_{j\in J} u_j(z_j) - \frac{1}{m} \sum_{j=1}^{m} z_j \right|.$$

Analogously to Problems 1.7 and 2.5, we can construct a consistent sample approximation of Problem 3.16 using the same techniques. However, such a sample approximation of Problem 3.16 will have substantially fewer constraints in its linear program formulation comparing to Problem 2.5.

Concerning Assumption 3.15, if the network does split into two independent parts, say $J_1 \cup J_2 = \{1, 2, \dots, m\}$ with $J_1 \cap J_2 = \emptyset$ and

$$\forall (j_1, j_2) \in J_1 \times J_2 : \mathbf{P}\big\{Z_{j_1}, Z_{j_2} \in \mathbb{R}\big\} = 0,$$

then Problem 3.16 has infinitely many solutions. For the sake of simplicity, this is illustrated in case of $\mathscr{J} = \{J_1, J_2\}$ with $|J_1| = |J_2|$. The general case is analogous with additional scaling. Let $\breve{u}_1 \times \breve{u}_2 \times \cdots \times \breve{u}_m$ be a solution of Problem 3.16, and let $\alpha > 0$. Then, the function $\tilde{\boldsymbol{u}} = \tilde{u}_1 \times \tilde{u}_2 \times \cdots \times \tilde{u}_m$ defined by

$$\tilde{u}_j = \begin{cases} \breve{u}_j + \alpha \breve{u}_j, & j \in J_1, \\ \breve{u}_j - \alpha \breve{u}_j, & j \in J_2, \end{cases}$$

is again feasible and has the same objective value as $\breve{\boldsymbol{u}}$. In fact, $\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_m$ fail to be monotonic for sufficiently large α. However, if Assumption 3.15 holds, then we can establish that Problem 3.16 has exactly one solution in a similar fashion as in Section 3.1.3 on page 41.

Chapter 4

Simulations

Chapter 3 ensures that the conditional maximum likelihood estimates defined in Chapter 2 converge to the solution of the offline synchronization problem. However, for applications, it is also important to know how fast the estimates converge, and when the asymptotic become effective. Notice that the estimation error is mainly comprised of two factors, caused by the sample average approximation and the inverse clock approximation. Thus, the estimation error is expected to decrease with order $\frac{1}{2}$ at least due to the central limit theorem, and then stagnates as the approximation error of the inverse clocks becomes dominant. The quality of the inverse clock approximation is obviously determined by the regularity of the actual clocks, the choice of the inverse clock model, and the distribution of the delays.

4.1 Methods

In the following simulations, clock functions and timestamps are generated randomly for various parameters. Then, we compute the conditional maximum likelihood estimate and analyze the estimation error to find a dependency with respect to the varying parameters. For the analysis, 100 log-sets of each setup are generated randomly. In the following figures the bars show the median estimation errors, and on the top of them the left and the right hooks indicate the 50% and 95% inter quantile range.

4.1.1 Generating Log-Sets

We will simulate a setup with m devices and n events, where the correct timestamps (without clock deviations and delays) are concentrated on the compact interval $[0, \bar{\xi}]$ for some $\bar{\xi} > 0$. Every time related value is stated in second. We consider $\bar{\xi} = 20 \cdot 60$ as an example for shorter time periods and $\bar{\xi} = 8 \cdot 60 \cdot 60$ for longer time periods. If not mentioned explicitly, we have $m = 100$ devices and $n = 10^5$ events.

Random Clock Functions

We simulate the clock functions as twice continuous differentiable functions with nearly constant first derivative and slightly oscillating second derivative. This assumption is rather common in the synchronization community (compare to [18]). Here, we choose the clock functions $C_1, C_2, \dots, C_m : \mathbb{R} \to \mathbb{R}$ randomly with

$$\frac{1}{m} \sum_{j=1}^{m} C_j = \mathrm{id}_{\mathbb{R}}$$

and such that each C_j is the sum of an affine linear function and a cubic spline with respect to some equidistant knot sequence $\boldsymbol{\xi} = (\xi_0, \dots, \xi_\nu)$ of $[0, \bar{\xi}]$ for $\nu \geq 4$. To achieve this some technical steps are necessary. Pick $\alpha_{0,j}$, $j = 1, 2, \dots, m$ (pseudo) randomly from a standard normal distribution and $\alpha_{1,j}, \beta_{i,j}$, $i = 1, 2, \dots, \nu - 3$, $j = 1, 2, \dots, m$ from a beta distribution according with parameter $(\frac{1}{2}, \frac{1}{2})$ transformed onto the interval $[-1, 1]$, Then, center the values by redefining

$$\alpha_{i,j} \leftarrow \alpha_{i,j} - \frac{1}{m} \sum_{j'=1}^{m} \alpha_{i,j'}, \qquad i = 0, 1,$$

$$\beta_{i,j} \leftarrow \beta_{i,j} - \frac{1}{m} \sum_{j'=1}^{m} \beta_{i,j'}, \qquad i = 1, 2, \dots, \nu - 3.$$

Finally, for some scale parameters $\sigma_{\text{offset}}, \sigma_{\text{skew}}, \sigma_{\text{drift}} > 0$ define $C_1, \dots, C_m$ by

$$\begin{aligned} C_j(x) &= \sigma_{\text{offset}}\alpha_{0,j} + (1 + \sigma_{\text{skew}}\alpha_{1,j})x, && x \in \{\xi_0, \xi_\nu\}, \\ C_j'(x) &= 1 + \sigma_{\text{skew}}\alpha_{1,j}, && x \in \{\xi_0, \xi_\nu\}, \\ C_j''(x) &= 0, && x \in \{\xi_0, \xi_\nu\}, \\ C_j''(\xi_i) &= \sigma_{\text{drift}}\beta_{i,j}, && i = 2, \dots, \nu - 2. \end{aligned}$$

Then, the clock functions are strictly increasing for $\sigma_{\text{skew}}, \sigma_{\text{drift}} \ll 1$ and are correct in average by construction.

For the simulation we choose the default values $\nu = 9$ (or 6 cubic B-spline basis functions), $\sigma_{\text{offset}} = 100$, $\sigma_{\text{skew}} = 10^{-5}$, and $\sigma_{\text{drift}} = 10^{-9}$. Notice that the rate error is dominated by σ_{skew} for shorter time periods, whereas for longer time periods σ_{drift} dominates the rate error. The transition is around $\bar{\xi} = 10^4$.

Random Timestamps

The correct timestamps (without clock deviations and delays) $t_1, t_2, \dots, t_n$ are generated according to the uniform distribution on $[0, \bar{\xi}]$ plus some Gaussian noise with variance much less than $\bar{\xi}/n$. This ensures the theoretical condition that the timestamps are supported on the complete real line. However, concerning applications, the timestamps remain basically bounded. Then, per device, we add to t_k a delay $d_{j,k}$ that is exponentially distributed with mean μ_{delay} and apply the clock function C_j,

$$z_{j,k} := C_j(t_k + d_{j,k}),$$

to yield the final timestamp $z_{j,k}$. For the simulation, the default mean delay μ_{delay} is 10^{-4}.

Censoring

To simulate incomplete log-sets, for $m \geq m_{\text{group}} \geq 1$ consider the system

$$\mathscr{J} = \big\{\{j, j \oplus 1, \dots, j \oplus (m_{\text{group}} - 1)\} \mid 1 \leq j \leq m\big\},$$

where $\oplus$ denotes the addition modulus m in $\{1, 2, \dots, m\}$. For each event $k \in \{1, 2, \dots, n\}$, choose a subset $\mathscr{J}_k \subseteq \mathscr{J}$ to determine the receivers of the

event k; that is, for each $j \notin \bigcup \mathcal{J}_k$ redefine

$$z_{j,k} \leftarrow \infty.$$

We assume that the cardinality $|\mathcal{J}_k|$ is geometric distributed with success probability $\pi_{\text{group}} \in (0, 1]$. That is, the probability is $\pi_{\text{group}}(1 - \pi_{\text{group}})^{l-1}$ for $|\mathcal{J}_k| = l$. The parameter m_{group} is referred to as *group size*, each set in $\mathcal{J}_k$ as a *observer group* of event k, and $|\mathcal{J}_k|$ as *numbers of observer groups* of event k.

In the case of $m_{\text{group}} = m$, the log-sets are complete. In the case of $m_{\text{group}} \ll m$, the synchronization problem is sparse yet does not break into two independent subproblems (compare to Assumption 3.15 on page 51). Notice that this setup reflects the local broadcast network, as nodes in a "neighborhood" share common events. However, the neighborhood is not entirely static, but also has non-deterministic elements. If not stated otherwise, we use $m_{\text{group}} = 5$ and $\pi_{\text{group}} = \frac{1}{2}$ by default.

4.1.2 Conditional Maximum-Likelihood Estimate

The conditional maximum likelihood estimate is computed according to Problem 2.3 on page 19 with a B-spline inverse clock model for each device. For the sake of simplicity, each inverse clock model employs the same dimension $d_{\text{est}} \geq 1$, and encompasses the probability range from $\underline{\tau} \in (0, 1)$ to $\bar{\tau} = 1 - \underline{\tau}$ (compare to Section 3.3 on page 47 and Section 2.4.3 on page 30).

As the clock functions are only twice continuously differentiable, linear B-spline inverse clock models shall be sufficient, which admit the approximation order 2 (compare to Section 2.4.2 on page 27). However, the simulated clock functions are locally polynomial. A higher degree spline estimate may benefit from the local regularity. Thus, we will vary the degree of the estimating splines in the following.

For numerical reasons, the estimate is computed using standardized time-stamps given by

$$\tilde{z}_{j,k} = \frac{z_{j,k} - \mu_{\boldsymbol{z}}}{\sigma_{\boldsymbol{z}}},$$

where

$$\mu_{\boldsymbol{z}} = \frac{1}{m} \sum_{j=1}^{m} \text{mean}\big\{ z_{j,k} \,|\, z_{j,k} < \infty,\, 1 \leq k \leq n \big\}$$

is the average sample mean, and

$$\sigma_{\boldsymbol{z}} = \frac{1}{m}\sum_{j=1}^{m} \operatorname{std}\big\{ z_{j,k} \,|\, z_{j,k} < \infty,\, 1 \le k \le n \big\}$$

is the average sample standard deviation. The parameters $\mu_{\boldsymbol{z}}$ and $\sigma_{\boldsymbol{z}}$ can also be replaced by the average of other location and scale parameters, like sample median and sample inter quantile range. To compensate the standardization, the estimates need to be rescaled too.

As default, we set $\underline{\tau} = 5 \cdot 10^{-3}$ and $d_{\text{est}} = 16$ (2 for the linear term, 14 for the B-spline term). Unlike in Problem 2.3 and Section 3.3, we simply fix $\eta_n = 10^{-12}$.

4.1.3 Measuring Estimation Error

The solution of the synchronization problem is given by the inverse functions of $C_1, C_2, \dots, C_m$. However, we can not compute them explicitly. This makes comparing the estimates $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_m$ with $C_1^{-1}, C_2^{-1}, \dots, C_m^{-1}$ difficult. Instead, for a given estimate $\hat{u}_j$ to the solution C_j^{-1} consider following integral

$$\frac{1}{\bar{\xi}} \int_0^{\bar{\xi}} |\hat{u}_j(C_j(x)) - x| \,\mathrm{d}x = \frac{1}{\bar{\xi}} \int_{C_j(0)}^{C_j(\bar{\xi})} |\hat{u}_j(z) - C_j^{-1}(z)| (C_j^{-1})'(z) \,\mathrm{d}z. \tag{4.1}$$

That is, the left hand side of Equation (4.2) is the weighted $\mathcal{L}^1$ error with weight $(C_j^{-1})'/\bar{\xi}$. For the complete estimate $\hat{\boldsymbol{u}}$, we employ the average $\mathcal{L}^1$ error

$$\frac{1}{\bar{\xi}} \int_0^{\bar{\xi}} \frac{1}{m} \sum_{j=1}^{m} |\hat{u}_j(C_j(x)) - x| \,\mathrm{d}x. \tag{4.2}$$

The average error measures the distance of the estimates to the inverse clock functions. However, in many applications, it is sufficient that the estimates are similar to the inverse clock functions modulo some common global transformation; that is, the following average standard deviation

$$\frac{1}{\bar{\xi}} \int_0^{\bar{\xi}} \sqrt{\frac{1}{m} \sum_{j=1}^{m} \left(\hat{u}_j(C_j(x)) - \frac{1}{m} \sum_{j'=1}^{m} \hat{u}_{j'}(C_{j'}(x)) \right)^2} \,\mathrm{d}x \tag{4.3}$$

is small.

For the simulation results presented in the following, the integrals in Equations (4.2) and (4.3) are approximated using the componsite Simpson's rule with 1024 equidistant subintervals. The quadrature error will be sufficiently small, as the integrands are piecewise smooth according to a much coarser partition, and the higher order derivatives are small in comparison. We need approximate the integrals to high precision, as we are interested in the magnitude of the errors only.

4.2 Linear Estimates, a Reference Point

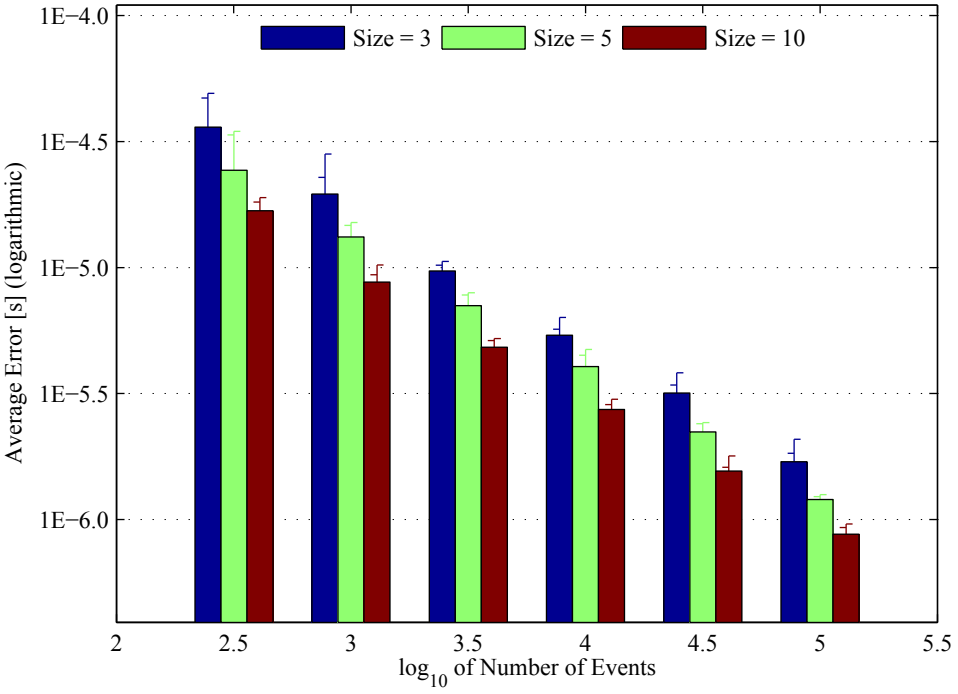
A reliable and accurate synchronization method was presented in [22] using an affine linear inverse clock model for timestamps in a 1200 seconds range. As the setup of [22] is slightly different from the one in this thesis, we provide some results for the sake of comparison. Notably, the *normalization* is handled differently. In [22], the estimates $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_m$ of the inverse clock functions are normalized such that

$$\frac{1}{m}\sum_{j=1}^{m} \hat{u}_j = \mathrm{id}_{\mathbb{R}}$$

holds, without using the sample quantiles, unlike in this thesis (compare to Problem 2.1 on page 17). Generalizing that approach to arbitrary inverse clock models is not straightforward, and it is not clear how to establish consistency in that setting.

Different from the default setup (Section 4.1), let the clock functions and the estimates be affine linear, and let $\bar{\xi} = 20 \cdot 60$. Figure 4.1 on the next page shows the estimation error for different group sizes (m_{group}). The results for $m_{\text{group}} = 5$ and 10 are very comparable to the results in [22]. The estimator presented there stagnates pretty fast. Whereas the estimator presented in this thesis improves further with order $\frac{1}{2}$ beyond the cited estimator. Remember that the order $\frac{1}{2}$ decrease was anticipated and is typical for most sample average approximation techniques.

Figure 4.1: Group Size Dependence (Linear Clocks and Estimates)



4.3 Piecewise Linear Estimates

To handle non-linear clock deviations, we could divide the log-sets in smaller subsets, each within shorter time periods. Obviously, that makes the estimates discontinuous. Also, each subinterval will not benefit from the data in other subintervals. However, that approach provides a useful baseline for comparison. We concentrate on the default case of $m_{\text{group}} = 5$, on a longer interval, say with $\bar{\xi} = 8 \cdot 60 \cdot 60$. In the following, the case of 1 subinterval reduces to an affine linear inverse clock model and is given as reference. For the case of $k > 1$ subintervals, the probability range $[\underline{\tau}, \bar{\tau}]$ is divided into k subintervals of equal length. That is, the inverse clock estimates are (not necessarily continuous) piecewise linear polynomials.

Let us consider linear clock functions first with piecewise linear estimates. Figure 4.2 on the following page shows higher estimation errors compared to Figure 4.1, even in the globally linear case. This does not surprise, as the timestamps have absolutely higher values. However, the error does not scale

Figure 4.2: Linear Clocks and Piecewise Linear Estimates

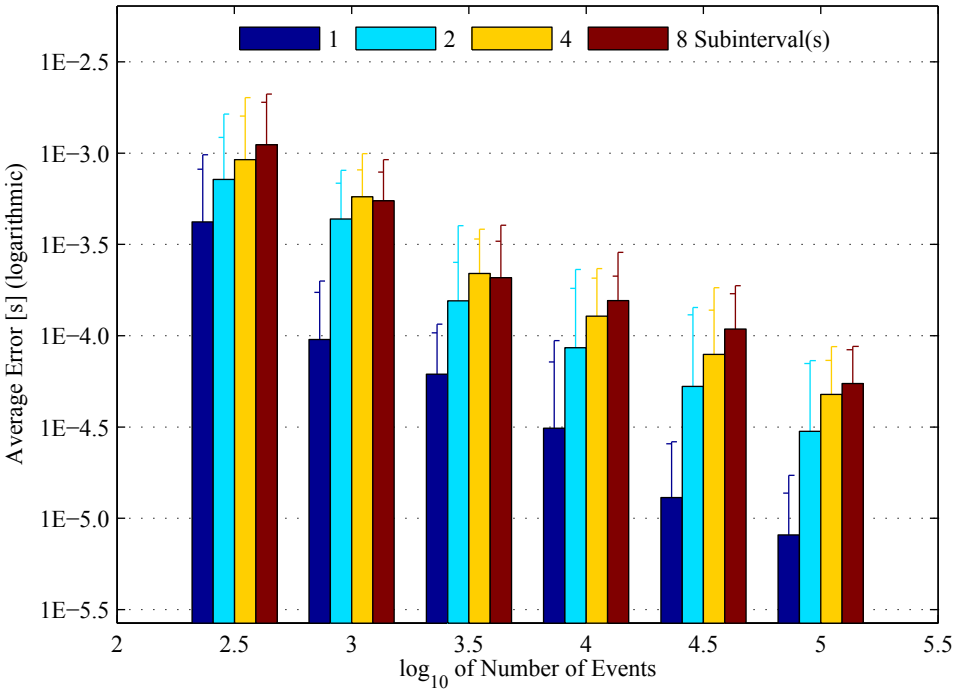
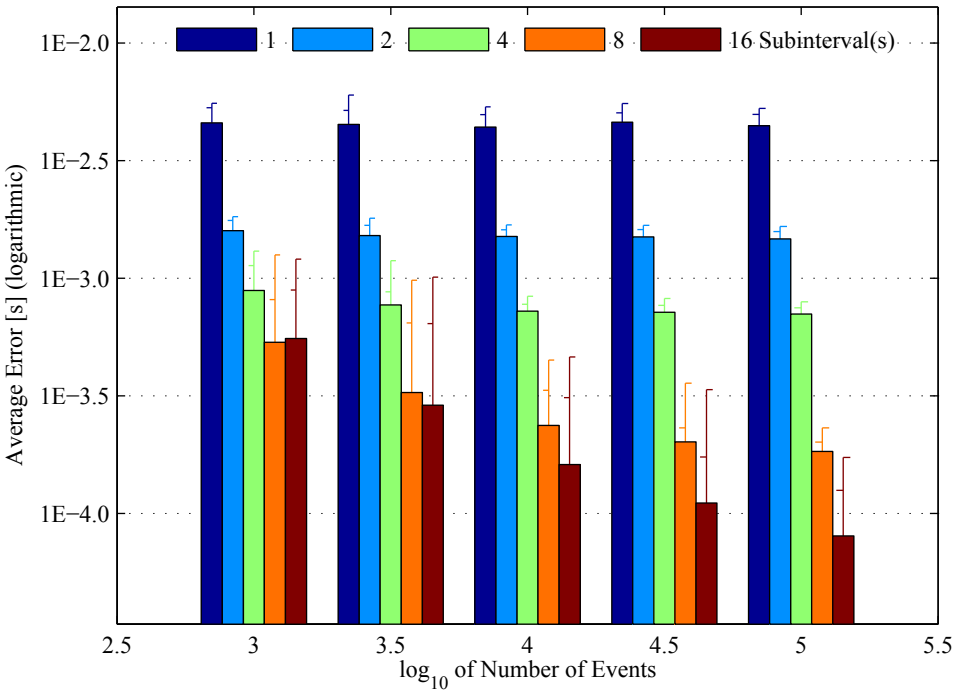


Figure 4.3: Non-Linear Clocks and Piecewise Linear Estimates



proportional to the interval length $\bar{\xi}$, which increases by factor 24, whereas the estimation error increases roughly by factor 5 only. Unfortunately, the error of the piecewise linear cases is much higher than the globally linear case, particularly visible in the case of 10^5 events. This stark difference can not be explained by the lower number of events per subinterval alone. In fact, the piecewise linear estimates decrease slightly slower than the global one. Notice that overfitting is a common problem of regression techniques. Fortunately, the effect is less prominent if we consider the piecewise linear estimates alone.

Finally, we introduce non-linear clock deviations into the simulation. Figure 4.3 on page 60 shows that the approximation error dominates the estimation error. The globally linear and piecewise linear estimates perform much worse than in the case of linear clock deviations. Yet, the piecewise linear estimates perform much better than the globally linear one, particularly in the case of many subintervals. In the case of 4 and less subintervals, the estimation error hardly improve for increasing event numbers. This behavior is expected, as the clock functions are cubic splines on 9 subintervals. In the case of 8 and more subintervals, the estimation error decreases with order $\frac{1}{2}$ at the beginning and then slows down.

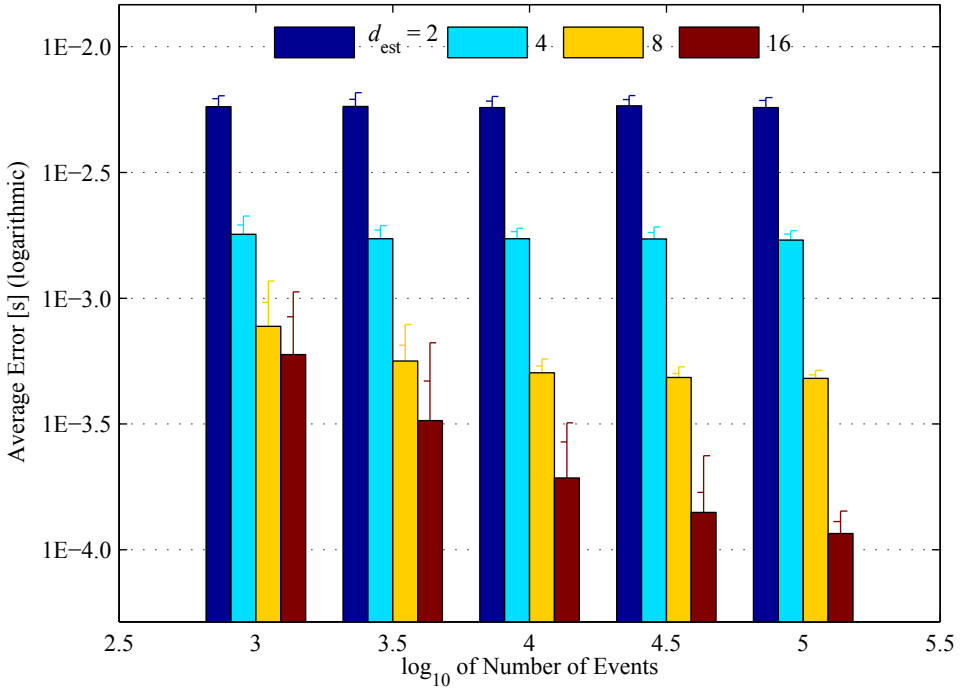
As piecewise linear polynomials have the approximation order 2, we may also expect that the estimation error decreases with order 1.5 when the length of subintervals decreases. That is, the error shall decrease by a factor

$$\frac{1}{2\sqrt{2}} \approx 10^{-0.45}$$

as the length of subintervals halves. Figure 4.3 exhibits clearly that behavior. Compare the estimation error for different numbers of subintervals when they stagnate. In the most cases the error decreases roughly by a factor $10^{-0.5}$ when the number of subintervals doubles. The only anomaly is the case when 2 subintervals doubles to 4. Noteworthy, we need more events to realize the higher approximation ability. For example, the 16 subintervals case improves substantially over the 8 subintervals case not before 10^4 events.

Obviously, the number of events per subinterval decreases, if the number of subintervals increases. Thus, we can not employ arbitrarily many subintervals. Also, there are twice many unknowns as subintervals for each

Figure 4.4: Inverse Clock Dimension Dependence (Degree 1 Spline Estimates)



inverse clock estimate. Thus, increasing the polynomial degree further is not a favorable option. In summary, to handle non-linear deviations, we shall investigate low dimensional estimates with high approximation order.

4.4 Spline Estimates

Similarly to the previous two sections, this section generates log-sets randomly according to the default setting with spline estimates (Section 4.1 on page 53). Again, we change the value of some specific parameters, like the degrees of freedom or the delay mean, to gauge their influence.

4.4.1 Approximation Error

Consider degree 1 splines first, which are continuous piecewise linear polynomials. We expect that the results are very comparable to those of the piecewise linear estimates in the last section. In particular, the estimation error should decreases with order $\frac{1}{2}$ for increasing events and with order 1.5

for increasing subintervals.

Figure 4.4 on page 62 shows the simulation results for different degrees of freedom d_{est}. The case $d_{\text{est}} = 2$ reduces to the linear inverse clock model and is given as reference. For the cases $d_{\text{est}} > 2$, the probability range $[\underline{\tau}, \bar{\tau}]$ is divided equidistantly into $d_{\text{est}} - 1$ subintervals. That is, with each bar, the length of the subintervals is roughly halved. Figure 4.4 on page 62 confirms our expectation. Noteworthy, the degree 1 spline estimate achieves similar results with $d_{\text{est}} = 16$ as the piecewise linear estimate on 16 subintervals with 32 unknowns. In other words, we can roughly half the estimate error with the same computational effort.

Do the approximation order of the inverse clock model always transfer to the estimation error? To obtain empirical evidence for that conjecture, we repeat the simulation with cubic spline estimates. Although the clock functions are globally twice continuously differentiable only, they are locally smooth. Thus, we shall achieve the approximation order 3.5. In other words, the estimation error decreases roughly by a factor 10^{-1} if the length of the subintervals halves.

The cubic spline estimator performs as expected (Figure 4.5 on the next page). The case of $d_{\text{est}} = 8$ is notably better than the linear spline estimator. Despite the immense amount of events, we see no benefit of $d_{\text{est}} = 32$ over $d_{\text{est}} = 16$. Putting that into perspective, 10^5 events in 8 hours means about 4 events per second. The sample quantile functions offer a possible explanation, as they determine the average of the clock estimates, and as they converges with an order lower than $\frac{1}{2}$ in general. In fact, the average standard deviation shows that the inverse clock estimates are similar to each other (Figure 4.6 on the following page). Thus, just their mean is off. This indicates that the estimation accuracy is limited by the sample quantile functions and not by the approximation ability of the splines. Noteworthy, the presented method reduces the clock deviations below 10 microseconds in an average over 8 hours, or about 0.3 nanoseconds per second, a stability comparable to a GPS disciplined oscillator.

Figure 4.5: Inverse Clock Dimension Dependence (Degree 3 Spline Estimates)

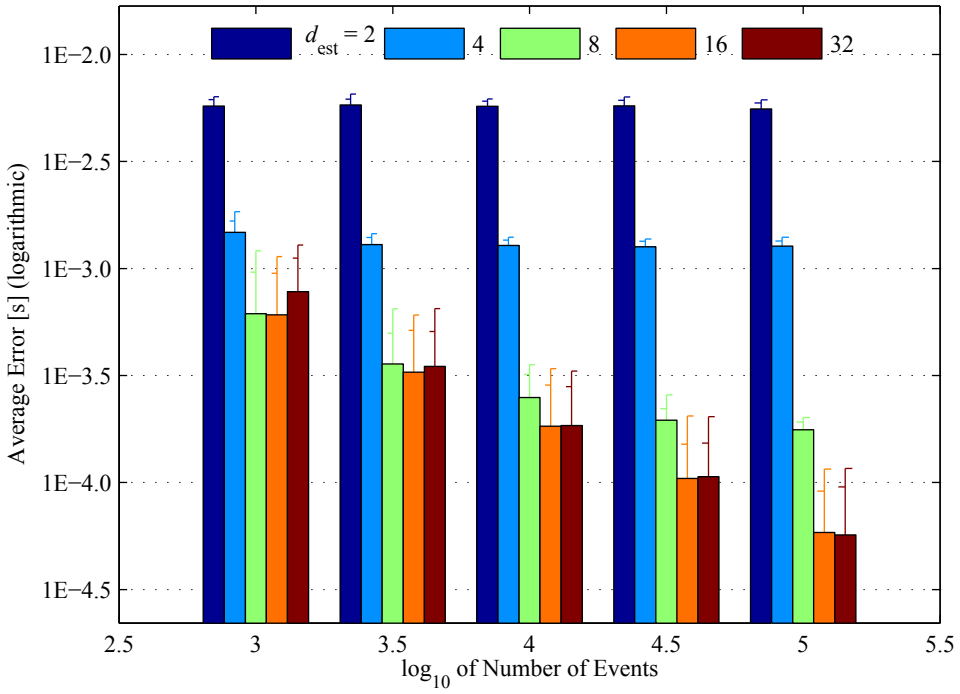


Figure 4.6: Inverse Clock Dimension Dependence (Degree 3 Spline Estimates)

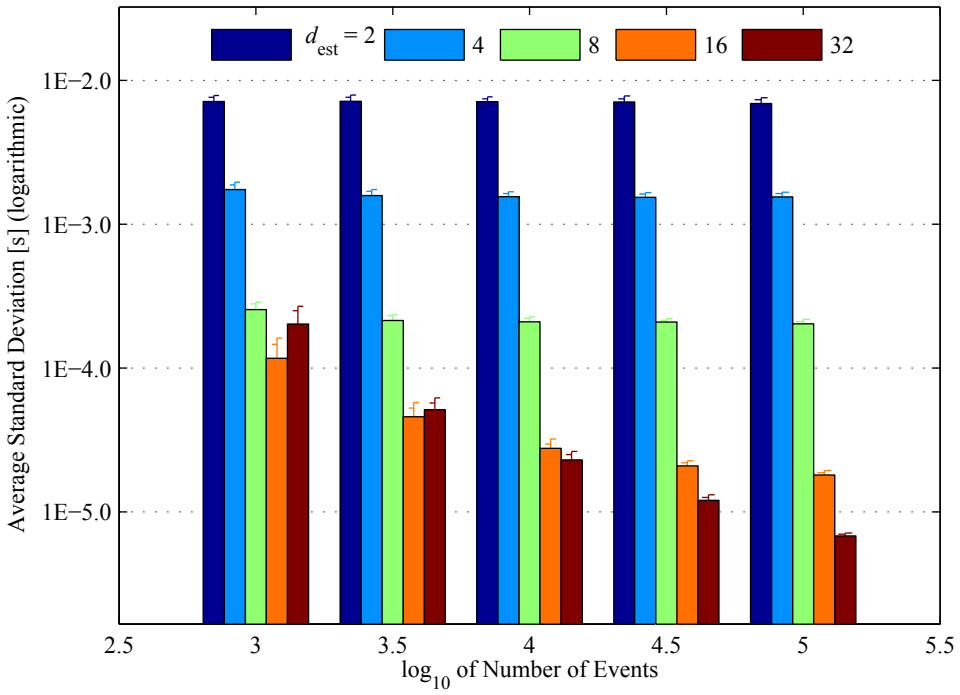
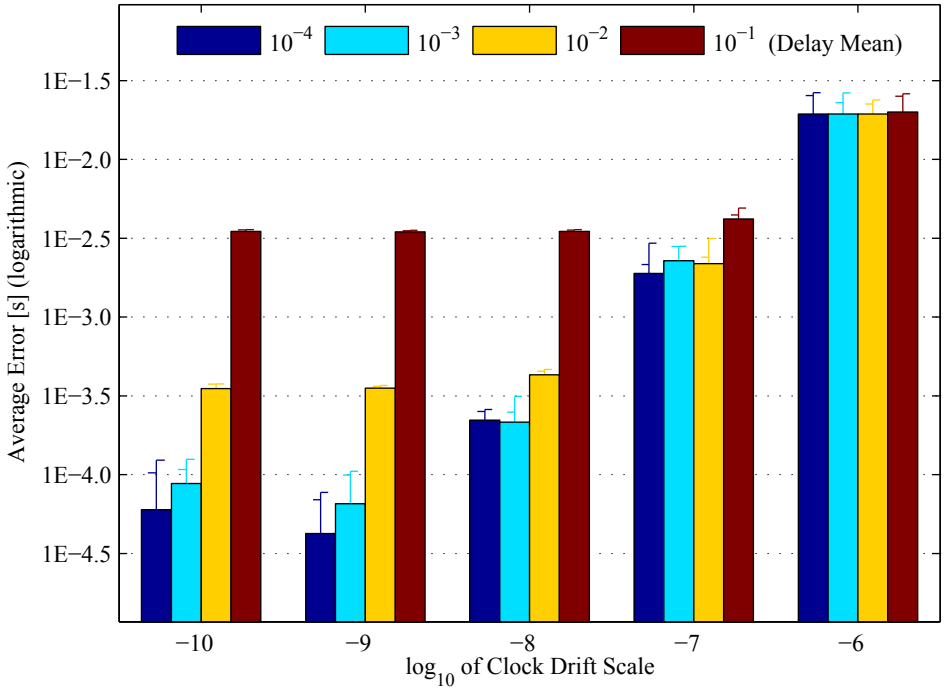


Figure 4.7: Clock Drifts vs Delays



4.4.2 Delays versus Clock Drifts

Reassessing the cubic estimates (Figure 4.5 on page 64), the approximation error plays a minor part in the estimation error for the case $d_{\text{est}} = 16$ independent of the event numbers up to 10^5. Using this, we can analyze the influence of the clock drifts and the delays and simulate the setup for various values of drift scale parameter σ_{drift} and the delay mean parameter μ_{delay}.

Figure 4.7 shows that the estimation error decreases linearly with the drift scale parameter and also with the delay mean until it stagnates. In fact, the clock drifts and the delays are interacting with and limiting each other. Notice that the difference between drift scale parameter 10^{-10} and 10^{-9} is rather insignificant. Also, delay mean lower than 10^{-3} shows little effect.

4.5 Incorrect Models

Some employed assumptions are overly simplified and not realistic. For example, in a complex device like a computer, various components cause delays.

Thus, the delay in a timestamp is likely not exponentially distributed. Here, we also consider gamma distribution, which includes the exponential distribution as special case. Also, network devices of various types and from various manufacturers are used in larger experiments. Thus, the delays are not identically distributed in general.

Further, the censoring and the values of the timestamps are not necessarily dependent. For example, a device traveling through a tunnel and temporarily losing the connection would result in incomplete communications and incomplete logs. Such kind of gap is time dependent and is not modeled by our timestamp model. Obviously, a large gap might alter the sample distribution gravely. In particular, we will not be able to use the sample quantile function for synchronization. Thus, we will alter the simulation setup to gauge the robustness of the estimator with respect to these modeling errors.

4.5.1 Gamma or Non Identically Distributed Delays

Figure 4.8 on the next page shows the simulation result with gamma distributed delays according to different parameters. They are chosen such that the delay mean remains $\mu_{\text{delay}} = 10^{-4}$, yet the variance σ_{delay} ranges between 10^{-12} and 10^{-4}. The $\sigma_{\text{delay}} = 10^{-8}$ case reduces to the exponential distribution.

The $\sigma_{\text{delay}} \leq 10^{-8}$ cases show surprisingly few difference, and each of them also exhibits order $\frac{1}{2}$ decrease. A possible explanation is that the delays can be viewed as sums of many independent exponentially distributed delays (with even lower mean). On the other hand, the $\sigma_{\text{delay}} > 10^{-8}$ cases do not have such convenient interpretation. Although, the $\sigma_{\text{delay}} = 10^{-6}$ case shows similar asymptotic, the $\sigma_{\text{delay}} = 10^{-4}$ case is significantly different.

As for non identically distributed delays, we choose the delay mean, device dependent, randomly according to the gamma distribution with mean $\mu_{\text{delay}} = 10^{-4}$ and different variance, which we call *delay mean variation* for the sake of clarity. Figure 4.9 on the facing page shows that the estimation error exhibit much stronger fluctuations for higher delay mean variations. However, lower variations have little influence and the asymptotic is preserved.

Figure 4.8: Variance Dependence (Gamma Distributed Delays)

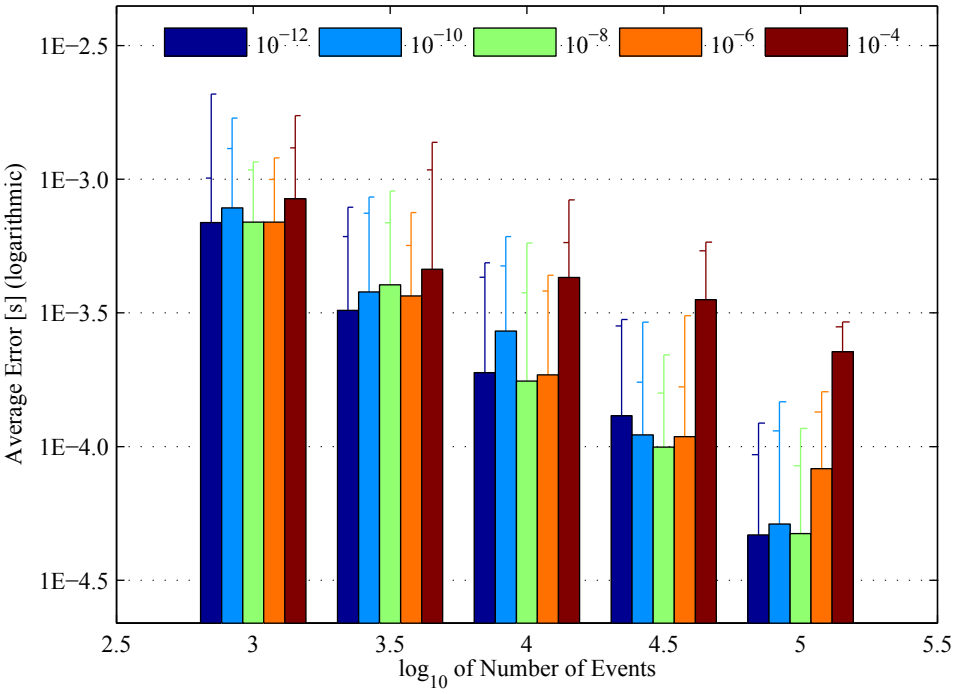
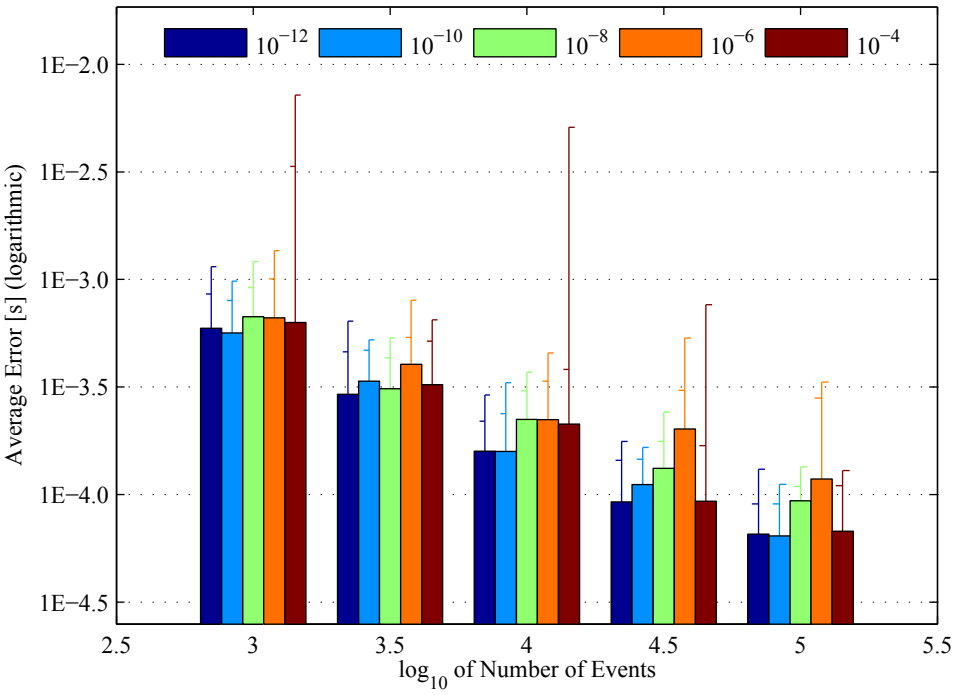


Figure 4.9: Delay Mean Variation



In summary, the simulations indicate that the presented approach might remain valid for particularly (non necessarily identically) gamma distributed delays.

4.5.2 Dependent Censoring

Remember, the set of devices is divided into groups by the system $\mathscr{J}$, and an event is observed by groups randomly chosen from $\mathscr{J}$ where the number of groups per event is geometric distributed. For sake of simplicity, we drop the geometric distribution and assume that each event has exactly a group of observers. Thus, the probability for a group $J \in \mathscr{J}$ observes an particular event is constantly $\frac{1}{m}$ over the time. Now, we alter that probability time dependently.

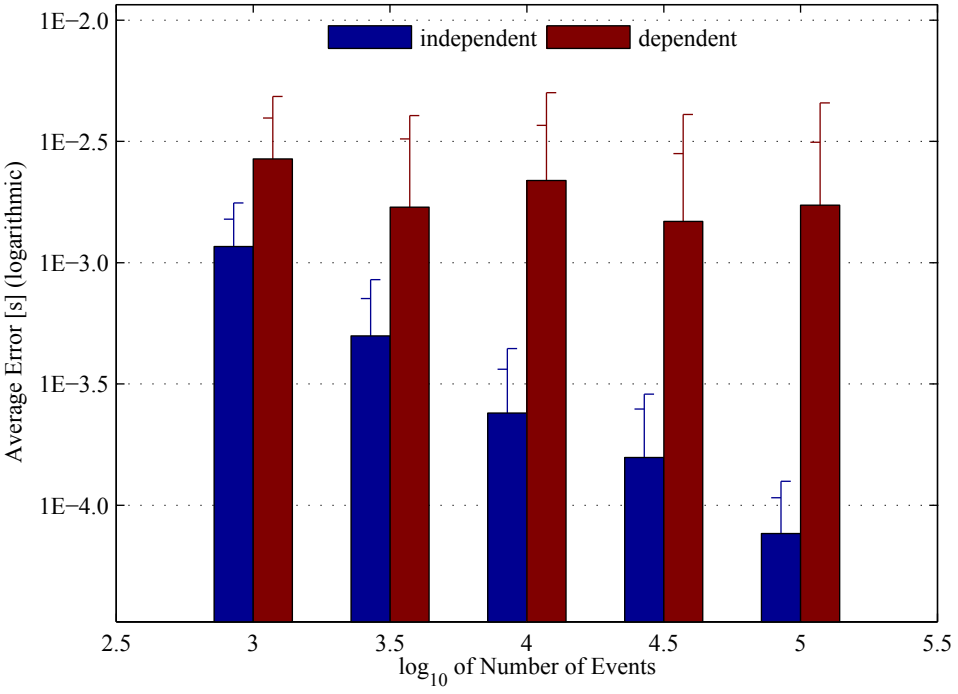
Assume that the number of groups is even and divide $\mathscr{J}$ into two disjoint subsystems $\mathscr{J}_1$ and $\mathscr{J}_2$ of the same size. To be more precise, let

$$\mathscr{J}_1 = \Big\{ \big\{ j, j \oplus 1, \dots, j \oplus (m_{\text{group}} - 1) \big\} \mid 1 \leq j \leq \tfrac{m}{2} \Big\},$$
$$\mathscr{J}_2 = \Big\{ \big\{ j, j \oplus 1, \dots, j \oplus (m_{\text{group}} - 1) \big\} \mid \tfrac{m}{2} \leq j \leq m \Big\},$$

where $\oplus$ again denotes the addition modulus m in $\{1, 2, \dots, m\}$. For each event, choose randomly a subset $J \in \mathscr{J}$ as group of observers where the subsets in $\mathscr{J}_1$ have the probability π_{censor}/m and the subsets in $\mathscr{J}_2$ have the probability $(1 - \pi_{\text{censor}})/m$ for some $\pi_{\text{censor}} \in (0, 1)$. The experiment is divide into two periods of the same length over the time. In the first period we set $\pi_{\text{censor}} = \frac{1}{4}$ and in the second period $\pi_{\text{censor}} = \frac{3}{4}$. Thus, there are more observed events for devices in $\bigcup \mathscr{J}_1$ in the second half of the simulation. Therefor, the sample median of the timestamps of these devices is higher compared to the independent censoring case. A similar argument applies to other sample quantiles. By a symmetric argument the sample quantiles of the devices in $\bigcup \mathscr{J}_2$ are lower compared to the independent censoring case. In particular, we can not longer rely on the sample quantiles to fix the average of the clock estimates. Figure 4.10 on the next page confirms our expectation and shows that our method performs poorly in case of dependent censoring.

In that particular case, in which the probability π_{censor} changes suddenly, we can try to separate the events before and after the change by analyzing the relative frequency of each group. However, such an ad-hoc fix does not

Figure 4.10: Independent vs Dependent Censoring



apply if the probability π_{censor} changes continuously with the time. A proper censor model is still an open problem.

Chapter 5

A Bottom Line and Beyond

This thesis provides a flexible, efficient, and consistent framework to approximate the solution of the offline synchronization problem using maximum-likelihood estimation techniques. Numerical simulations display much higher accuracy than indicated by the theoretical results. Thus, the presented approach is not only sound in theory but also efficient in applications.

Although designed with a generic spline inverse clock model in mind, it can be instantly replaced by a more sophisticated inverse clock model to include additional prior knowledge. Although estimating non-linear functions, the shown estimates are just solutions of large linear programs. Although using the conditional probability of the delays only, the estimator is strongly consistent.

There is more to do however. Practical applications need to be able to handle incomplete log-sets with dependent censoring. We also like to adaptively refine the inverse clock model. Finally, we like to establish a convergence order for the estimates. As for the last item, the framework of sample average approximation seems to be promising. As for adaptive refinement of the model, the B-spline inverse clock model shall be sufficiently flexible. As for dependent censoring, the framework of time series may be useful. For these open issues, the method shown in this thesis provides a solid starting point.

5 A Bottom Line and Beyond

Appendix A

Basics

A.1 Dominated Convergence

A.1 Theorem (Compare to [8]):
Let P be a probability measure on $\mathbb{R}^m$.

- *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions from $\mathbb{R}^m$ into $\mathbb{R}$ which converges P-almost everywhere pointwise to $f : \mathbb{R}^m \to \mathbb{R}$.*
- *Let $(g_n)_{n\in\mathbb{N}}$ be a sequence of $\mathcal{L}^1(P,\mathbb{R})$ functions which converges to $g \in \mathcal{L}^1(P,\mathbb{R})$ in $\mathcal{L}^1$.*
- *For every $n \in \mathbb{N}$ let $|f_n| \leq g_n$ hold P-almost everywhere.*

Then, f_n, $n \in \mathbb{N}$, and f are $\mathcal{L}^1$ with $f_n \to f$ in $\mathcal{L}^1$.

A.2 Absolutely Continuous Functions

A.2 Definition: *Let $I \subseteq \mathbb{R}$ be an interval.*

- *A function $f : I \to \mathbb{R}$ is called* absolutely continuous*, if for every $\varepsilon > 0$ there is some $\delta > 0$ such that for every $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n \in I$ with $a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n$ and*

$$\sum_{i=1}^{n} b_i - a_i < \delta$$

follows

$$\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \varepsilon.$$

- *A function $f : I \to \mathbb{R}$ is called* locally absolutely continuous*, if for every compact subinterval $K \subseteq I$ the restriction $f|_K$ is absolutely continuous.*

Notice that a locally absolutely continuous function is continuous, and if it is absolutely continuous, then it is uniformly continuous. Further, a (locally) Lipschitz continuous function is (locally) absolutely continuous.

A.3 Theorem (Fundamental Theorem of Calculus, Compare to [24]):
Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$. Then, f is (locally) absolutely continuous, if and only if f is almost everywhere differentiable, f' is (locally) integrable, and for every $a, b \in I$ it follows

$$f(b) - f(a) = \int_a^b f'(t)\,\mathrm{d}t.$$

A.4 Theorem (Integration by Parts, Compare to [24]):
Let $I = [a, b] \subseteq \mathbb{R}$ be a compact interval and $f, g : I \to \mathbb{R}$ be absolutely continuous. Then, it follows

$$\int_a^b f'(t)g(t)\,\mathrm{d}t = f(b)g(b) - f(a)g(a) - \int_a^b f(t)g'(t)\,\mathrm{d}t.$$

A.5 Theorem (Integration by Substitution, Compare to [24]):
Let $I, J \subseteq \mathbb{R}$ be compact intervals, let $\phi : I \to J$ be absolutely continuous and strictly monotone, and let $f : J \to \mathbb{R}$ be integrable. Then, $f \circ \phi \cdot \phi'$ is integrable, and for each $\alpha, \beta \in I$ it follows

$$\int_{\phi(\alpha)}^{\phi(\beta)} f(t)\,\mathrm{d}t = \int_\alpha^\beta f(\phi(\tau))\phi'(\tau)\,\mathrm{d}\tau.$$

A.2.1 About the Inverse Function

A.6 Theorem (Compare to [28]):
Let $I \subseteq \mathbb{R}$ be a compact interval and $f : I \to \mathbb{R}$ be absolutely continuous. Assume that f' is almost everywhere positive (or negative). Then, f is strictly increasing (or decreasing), and its inverse function $f^{-1} : f(I) \to I$ is absolutely continuous.

A.2.2 Differentiation of Integrals

There are many differentiation of integral type theorems. The most well-known variants base on dominated convergence theorem (compare to the literature, like [8]). If we assume slightly stronger regularity, we can replace the dominated convergence argument with an Fubini type argument and obtain following variant here, which is more suitable for this thesis.

A.7 Theorem:
Let $I \subseteq \mathbb{R}$ be a non-trivial interval, and let P be probability measure on $\mathbb{R}^m$. Further, let $f : I \times \mathbb{R}^m \to \mathbb{R}$ be such that for every $t \in I$ the function $f(t, \cdot\,)$ is P-integrable, and such that for P-almost every $\boldsymbol{x} \in \mathbb{R}^m$ the function $f(\,\cdot\,, \boldsymbol{x})$ is absolutely continuous. Assume that $t_0 \in I$ and $g \in \mathcal{L}^1(P, \mathbb{R})$ satisfy

$$\lim_{t \to t_0} \int |\partial_t f(t, \cdot\,) - g| \, \mathrm{d}P = 0.$$

Then, the function $F : I \to \mathbb{R}$ defined by

$$F(t) = \int f(t, \cdot\,) \, \mathrm{d}P,$$

is differentiable at t_0 with the derivative

$$F'(t_0) = \int g \, \mathrm{d}P.$$

Appendix B

Extremum Estimator and Stochastic Programming

In statistical inference, an *extremum estimate* arises as the optimizer (or rather the set of optimizers) of an optimization problem whose objective function and feasible set depend on random data. Like in the deterministic case, the theory of convex programs is much easier, and it is fortunately also sufficient for the application in this thesis.

Consider the following setting: Let X be a closed convex subset of some Banach space such that X as topological subspace is locally compact. For each $n \in \mathbb{N} \cup \{0\}$ let f_n be a random convex and continuous function from X into $\mathbb{R}$, let S_n be a random closed convex subset of X, and let f_n^* denote the infimum and S_n^* denote the set of minimizers of the problem

$$\text{minimize} \quad f_n(x) \quad \text{subject to} \quad x \in S_n. \tag{P$_n$}$$

Assume that (f_n, S_n) converges to (f_0, S_0) in some to be specified sense, that (f_0, S_0) is deterministic, also that S_0^* is non-empty and compact. Does (f_n^*, S_n^*) converge to (f_0^*, S_0^*)? This chapter provides a positive answer under reasonable conditions. The result is proven for the deterministic case first and then extended to a stochastic setting.

Notice that the subject is rather standard in statistical inference and is covered by many authors, for example by [23]. As the treatment there is less specific, we elaborate the result here specifically for our application. In

particular, the approach to the feasible set S_n is less standard.

B.1 About Notations

In the following, each topological notion is to be understood relatively to X. For example, $S \subseteq X$ is open if S is open in X, or the interior of S is the largest subset of S which is open in X. Also, for $n \in \mathbb{N} \cup \{0\}$ and $\varepsilon > 0$ denote the set of ε-suboptimizers of f_n on S_n by

$$S_n^*(\varepsilon) = \{x \in S_n \mid f_n(x) < f_n^* + \varepsilon\},$$

and for $\delta > 0$ denote the δ-neighborhood of S_0^* by

$$U_\delta = \{x \in X \mid \exists S_0^* \in S_0^* : \|x - S_0^*\| < \delta\}.$$

Further, the following generic notations make the discourse more concise:

- For a metric space (M, d_M), some $a \in M$, and subsets $U, S, T \subseteq M$ let

$$\operatorname{dist}(a, S) = \operatorname{dist}(S, a) = \inf_{y \in S} d_M(a, y) \in [0, \infty],$$
$$\operatorname{dev}(S, T) = \sup_{x \in S} \operatorname{dist}(x, T) \in \{-\infty\} \cup [0, \infty],$$
$$\operatorname{dev}_U(S, T) = \operatorname{dev}(U \cap S, U \cap T),$$
$$\operatorname{dist}(S, T) = \max\{\operatorname{dev}(S, T), \operatorname{dev}(T, S)\}.$$

- For metric spaces (M, d_M), (N, d_N), $f : M \to N$, we denote the *(global) modulus of continuity* of f by $\omega_f : [0, \infty] \to [0, \infty]$, defined by

$$\omega_f(\delta) := \sup\{d_N(f(x), f(\tilde{x})) \mid x, \tilde{x} \in M,\ d_M(x, \tilde{x}) \leq \delta\}.$$

  Further, for any $S \subseteq M$, let

$$\omega_f(\delta, S) := \omega_{f|_S}(\delta).$$

- For normed spaces E, F, $S' \subseteq S \subseteq E$, a function $f : S \to F$, we denote the *sup-norm* of f restricted to S' by

$$\|f\|_{S'} := \|f|_{S'}\|_\infty := \sup_{x \in S'} \|f(x)\| \in [0, \infty].$$

- Finally, the abbreviations *a.s.* for almost surely and *a.e.* for almost every(where) are used extensively.

B.2 The Deterministic Case

In this section, we additionally assume that f_n and S_n are deterministic for each $n \in \mathbb{N}$. This assumption will not be explicitly mentioned again.

B.1 Theorem:

Under the established setting: Assume, for each $\delta > 0$,

$$\liminf_{n\to\infty} \inf_{S_n \setminus U_\delta} f_n > f_0^*,$$

and

$$\limsup_{n\to\infty} f_n^* \leq f_0^*.$$

Then, for each zero sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ of positive numbers, it follows

$$\lim_{n\to\infty} \operatorname{dev}(S_n^*(\varepsilon_n), S_0^*) = 0.$$

In particular, if S_0^* is a singleton, then it also follows

$$\lim_{n\to\infty} \operatorname{dist}(S_n^*(\varepsilon_n), S_0^*) = 0.$$

Proof: Fix $\delta > 0$. Notice that f_n^* is finite for sufficiently large n, as f_n is bounded from below by f_0^* on $S_n \setminus U_\delta$, and f_n is continuous on the closure of $S_n \cap U_\delta$, which is compact. Thus, S_n^* and $S_n^*(\varepsilon_n)$ are non-empty.

By assumption, there are $\varepsilon > 0$ and $n_1 \in \mathbb{N}$ with

$$\forall n \geq n_1 : f_0^* + \varepsilon \geq f_n^* > -\infty$$

and

$$\forall n \geq n_1 : \inf_{S_n \setminus U_\delta} f_n \geq f_0^* + 2\varepsilon \geq f_n^* + \varepsilon.$$

Thus, for each $x \in S_n \setminus U_\delta$, it follows $f_n(x) \geq f_n^* + \varepsilon$ and $x \in S_n \setminus S_n^*(\varepsilon)$. Now, let $n_2 \in \mathbb{N}$ with

$$\forall n \geq n_2 : \varepsilon_n < \varepsilon.$$

Then, for all $n \geq n_0 := \max\{n_1, n_2\}$, we have

$$S_n^*(\varepsilon_n) \subseteq S_n^*(\varepsilon) \subseteq U_\delta$$

and

$$\operatorname{dev}(S_n^*(\varepsilon_n), S_0^*) < \delta. \qquad \square$$

B.2.1 Limit of the Minima

B.2 Theorem:

Under the established setting: Assume that

- f_n *converges uniformly to* f_0 *on a compact neighborhood* $\bar{U}$ *of* S_0^*,
- $\mathrm{dev}_{\bar{U}}(S_n, S_0)$ *converges to* 0,
- *there exists a convergent sequence* $(\bar{x}_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} S_n$ *with limit* $\bar{x}_0 \in S_0^*$.

Then, it follows

1. $\liminf_{n\to\infty} \inf_{S_n\setminus U_\delta} f_n > f_0^*$ *for each* $\delta > 0$,
2. $\liminf_{(\delta,n)\to(0,\infty)} \inf_{U_\delta} f_n \geq f_0^*$,
3. $\limsup_{n\to\infty} f_n^* \leq f_0^*$, *and in particular,*
4. $\lim_{n\to\infty} f_n^* = f_0^*$.

Concerning the assumptions: Since S_0^* is compact, there exists some $\bar{\delta} > 0$ with $S_0^* \subseteq U_{\bar{\delta}} \subseteq \bar{U}$. Further, $\mathrm{dev}_{\bar{U}}(S_n, S_0)$ converges to 0 if and only if for each infinite subset $K \subseteq \mathbb{N}$ and convergent sequence $(x_n)_{n\in K} \in \prod_{n\in K} S_n \cap \bar{U}$ it follows $\lim_{n\to\infty} x_n \in S_0$.

Proof:

Ad Item 1: Fix some $\delta > 0$. Without loss of generality assume $U_\delta \subseteq \bar{U}$. Then, for each $x \in S_0 \setminus U_\delta$ there exists some $\alpha \in [0,1]$ with

$$\tilde{x} = (1-\alpha)\bar{x}_0 + \alpha x \in S := S_0 \cap (\bar{U} \setminus U_\delta).$$

Since S is compact and f_0 is continuous, it follows

$$f_0^* < \inf_S f_0 \leq \inf_{S_0\setminus U_\delta} \max\{f_0(\bar{x}_0), f_0\} \leq \max\left\{f_0^*, \inf_{S_0\setminus U_\delta} f_0\right\}$$

and

$$\varepsilon := \frac{1}{4}\left(\inf_{S_0\setminus U_\delta} f_0 - f_0^*\right) > 0.$$

Now, let $n_1 \in \mathbb{N}$ with

$$\forall n \geq n_1 : [\bar{x}_n \in \bar{U},\ f_0(\bar{x}_n) < f_0^* + \varepsilon,\ \|f_n - f_0\|_{\bar{U}} < \varepsilon].$$

For $n \geq n_1$ let $x_n \in S_n \setminus U_\delta$. Due to convexity, there exists some $\alpha_n \in [0,1]$ with

$$\tilde{x}_n := (1-\alpha_n)\bar{x}_n + \alpha_n x_n \in S_n \cap (\bar{U} \setminus U_\delta).$$

Since $\bar{U}$ is compact, assume $\tilde{x}_n \to \tilde{x}_0 \in S_0 \setminus U_\delta$ without loss of generality. Let $n_2 \in \mathbb{N}$ such that for each $n \geq n_2$ it follows $f_0(\tilde{x}_n) \geq f_0(\tilde{x}_0) - \varepsilon$. Thus, for each $n \geq n_0 := \max\{n_1, n_2\}$ it follows

$$\begin{aligned}\max\{f_n(\bar{x}_n), f_n(x_n)\} &\geq f_n(\tilde{x}_n) \\ &> f_0(\tilde{x}_n) - \varepsilon \geq f_0(\tilde{x}_0) - 2\varepsilon \geq f_0^* + 2\varepsilon.\end{aligned}$$

Notice that $f_n(\bar{x}_n) < f_0(\bar{x}_n) + \varepsilon < f_0^* + 2\varepsilon$ excludes the case $f_n(x_n) \leq f_n(\bar{x}_n)$. Thus, we have $f_n(x_n) > f_0^* + 2\varepsilon$. As $x_n \in S_n \setminus U_\delta$ was arbitrary, we obtain

$$\inf_{S_n \setminus U_\delta} f_n > f_0^* + 2\varepsilon$$

for $n \geq n_0$.

Ad Item 2: Fix $\delta > 0$ and $x \in U_\delta$. Then, there exists some $\bar{x} \in S_0^*$ with $\|x - \bar{x}\| < \delta$. Thus, for each $n \in \mathbb{N}$ it follows

$$\begin{aligned}f_n(x) &= f_0(\bar{x}) + f_0(x) - f_0(\bar{x}) + f_n(x) - f_0(x) \\ &\geq f_0^* - \omega_{f_0}(\delta, U_\delta) - \|f_n - f_0\|_{U_\delta}.\end{aligned}$$

For sufficiently small $\delta > 0$ we have $U_\delta \subseteq \bar{U}$ and

$$\inf_{U_\delta} f_n \geq f_0^* - \omega_{f_0}(\delta, \bar{U}) - \|f_n - f_0\|_{\bar{U}},$$

where the right-hand side converges to f_0^* for $(\delta, n) \to (0, \infty)$.

Ad Item 3: From $\bar{x}_n \to \bar{x}_0 \in S_0^*$ we have $\bar{x}_n \in \bar{U}$ for sufficiently large n. Thus, for every $n \in \mathbb{N}$ it follows

$$f_n^* \leq f_n(\bar{x}_n) = f_n(\bar{x}_n) - f_0(\bar{x}_n) + f_0(\bar{x}_n) \to f_0(\bar{x}_0) = f_0^*. \qquad \square$$

B.2.2 Limit of the Feasible Set

B.3 Assumption: *For $n \in \mathbb{N} \cup \{0\}$ let X_n be a closed convex subset of X, and let Y_n be a non-empty compact subset of some complete metric space Y. Let*

$g : X \times Y \to \mathbb{R}$ be continuous. Let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers with $\eta_n \to \eta_0 := 0$. For $n \in \mathbb{N} \cup \{0\}$ define $g_n : X \to \mathbb{R}$ by

$$g_n(x) = \max_{Y_n} g(x; \cdot)$$

and let

$$S_n = \{x \in X_n \mid g_n(x) \leq \eta_n\}.$$

Assume that

1. *there exists a totally bounded subset $U \subseteq X_0$ with $S_0 \cap U \neq \emptyset$ such that g is uniformly continuous on $U \times Y$;*

2. *and the modulus of continuity ω_g of g satisfies*

$$\omega_g(\mathrm{dist}(Y_n, Y_0), U \times Y) = o(\eta_n).$$

B.4 Lemma: *Under Assumption B.3: The function g_0 is uniformly continuous on U, the moduli of continuity satisfy*

$$\forall \delta > 0 : \omega_{g_0}(\delta, U) \leq \omega_g(\delta, U \times Y),$$

and for each $n \in \mathbb{N}$ it follows

$$\|g_n - g_0\|_U \leq \omega_g(\mathrm{dist}(Y_n, Y_0), U \times Y).$$

Proof: Concerning the uniform continuity of g_0: Fix $\delta > 0$, and fix $x, \tilde{x} \in U$ with $\|x - \tilde{x}\| < \delta$. Let $y_0 \in Y_0$ with $g(x; y_0) = g_0(x)$. Then, it follows

$$g_0(x) - g_0(\tilde{x}) \leq g(x; y_0) - g(\tilde{x}; y_0) \leq \omega_g(\|x - \tilde{x}\|, U \times Y).$$

Symmetrically, we have $g_0(\tilde{x}) - g_0(x) \leq \omega_g(\|x - \tilde{x}\|, U \times Y)$. That is

$$|g_0(x) - g_0(\tilde{x})| \leq \omega_g(\|x - \tilde{x}\|, U \times Y).$$

In particular, we have

$$\omega_{g_0}(\delta, U) \leq \omega_g(\delta, U \times Y).$$

Concerning the uniform convergence: Fix $x \in U$. Let $y_0 \in Y_0$ with $g(x; y_0) = g_0(x)$. For $n \in \mathbb{N}$ let $y_n \in Y_n$ with $d_Y(y_n, y_0) \leq \operatorname{dist}(Y_n, Y_0)$. Then, it follows

$$g_0(x) - g_n(x) \leq g(x; y_0) - g(x; y_n) \leq \omega_g(\operatorname{dist}(Y_n, Y_0), U \times Y).$$

Now, for $n \in \mathbb{N}$ let $z_n \in Y_n$ with $g(x; z_n) = g_n(x)$ and $\tilde{z}_n \in Y_0$ with $d_Y(z_n, \tilde{z}_n) \leq \operatorname{dist}(Y_n, Y_0)$. Then, we have

$$g_0(x) - g_n(x) \geq g(x; \tilde{z}_n) - g(x; z_n) \geq -\omega_g(\operatorname{dist}(Y_n, Y_0), U \times Y). \qquad \square$$

B.5 Theorem:
Under Assumption B.3:

1. *Then, $S_n \cap U$ is non-empty for sufficiently large $n \in \mathbb{N}$, and it follows $d_n := \operatorname{dev}_U(S_n, S_0) \to 0$.*
2. *Let $x_0 \in S_0$ with $\omega_g(\operatorname{dist}(X_n, x_0), U \times Y) = o(\eta_n)$. Then, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in X and $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ it follows $x_n \in S_n$.*

Proof (Item 1 of Theorem B.5): Assume that there is an infinite subset $K \subseteq \mathbb{N}$ with

$$\forall n \in K : S_n \cap U = \emptyset.$$

Without loss of generality assume $K = \mathbb{N}$. Thus, for each $n \in \mathbb{N}$ and $x \in U$ it follows

$$g_n(x) > \eta_n.$$

On the other hand, for $x \in S_0 \cap U$ we have

$$0 = \liminf_{n\to\infty} \frac{g_0(x)}{\eta_n} \geq \underbrace{\liminf_{n\to\infty} \frac{g_n(x)}{\eta_n}}_{\geq 1} + \underbrace{\lim_{n\to\infty} \frac{g_0(x) - g_n(x)}{\eta_n}}_{=0} \geq 1,$$

a contradiction. Thus, $S_n \cap U$ is non-empty for sufficiently large n.

Now, as U is totally bounded, we have $d_n \in [0, \infty)$ for sufficiently large n and

$$d_0 := \limsup_{n\to\infty} d_n \in [0, \infty].$$

Without loss of generality assume $d_n \to d_0$. For $n \in \mathbb{N}$ let $x_n \in S_n \cap U$ be such that $d_n - \operatorname{dist}(x_n, S_0 \cap U) \to 0$. Without loss of generality assume $x_n \to x_0 \in X$. By Lemma B.4, we have

$$g_0(x_0) = \underbrace{g_n(x_n)}_{\leq \eta_n} + \underbrace{g_0(x_0) - g_0(x_n)}_{\to 0} + \underbrace{g_0(x_n) - g_n(x_n)}_{\to 0} \to 0$$

and

$$d_0 = \lim_{n\to\infty} d_n = \lim_{n\to\infty} \operatorname{dist}(x_n, S_0 \cap U) = \operatorname{dist}(x_0, S_0 \cap U) = 0. \qquad \square$$

Proof (Item 2 of Theorem B.5): Let $(x_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} X_n$ with

$$\|x_n - x_0\| = \mathcal{O}(\operatorname{dist}(X_n, x_0)).$$

Then, it follows

$$g_n(x_n) \leq \underbrace{g_0(x_n) - g_0(x_0)}_{\mathcal{O}(\omega_g(\operatorname{dist}(X_n, x_0), U\times Y))} + \underbrace{\|g_n - g_0\|_U}_{\mathcal{O}(\omega_g(\operatorname{dist}(Y_n, Y_0), U\times Y))} = o(\eta_n).$$

That is, for $n \in \mathbb{N}$ sufficiently large it follows $x_n \in S_n$. $\square$

B.3 A Uniform Law of Large Numbers

For this section, fix a measurable space $(Y, \mathcal{Y})$. For a function $f : X \times Y \to \mathbb{R}$ and a random (including deterministic) probability P on $(Y, \mathcal{Y})$, define the random function Pf from X into $\mathbb{R}$ by

$$Pf(x) = \int f(x; y) \, \mathrm{d}P(y),$$

as long as the right hand side is well-defined. Even if $Pf(x)$ is well-defined, $Pf(x)$ need not be a random number. However, if $y \mapsto f(x; y)$ is $\mathcal{Y}$-measurable and P is some sample distribution, then $Pf(x)$ is a random number.

B.6 Assumption: *Let P_0 be a probability measure on $(Y, \mathcal{Y})$. For $n \in \mathbb{N}$ let P_n be a sample distribution of P_0 of size n. Let $f : X \times Y \to \mathbb{R}$ be such that $f(x; \cdot)$ is $\mathcal{Y}$-measurable for each $x \in X$. Assume that*

1. *$f(x;\,\cdot\,)$ is P_0-integrable for each $x \in X$;*

2. *there exists a function $\bar{f} : (0,\infty) \times Y \to \mathbb{R}$ and a totally bounded subset $U \subseteq X$ such that for each $\delta > 0$ it follows $\bar{f}(\delta,\,\cdot\,)$ is P_0-integrable,*

$$\forall y \in Y : \omega_f(\delta, U \times \{y\}) \leq \bar{f}(\delta, y),$$

and

$$\lim_{\delta \to 0} P_0 \bar{f}(\delta) = 0.$$

B.7 Theorem:
Under Assumption B.6: $P_0 f$ is uniformly continuous on U and $P_n f$ converges uniformly to $P_0 f$ on U almost surely.

Proof: For each $\delta > 0$ it follows

$$\omega_{P_0 f}(\delta, U) \leq \int \omega_f(\delta, U \times \{y\})\, \mathrm{d}P_0(y) \leq P_0 \bar{f}(\delta) \to 0.$$

Without loss of generality let $P_0 f|_U = 0$. Fix $\varepsilon > 0$ and $\delta > 0$ with $P_0 \bar{f}(\delta) < \varepsilon$. Since U is totally bounded, there exist $x_1, x_2, \ldots, x_{l_\delta} \in U$ with

$$U \subseteq \bigcup_{l=1}^{l_\delta} \{x \in U \mid \|x - x_l\| < \delta\}.$$

Notice that $P_n f(x_l)$ converges to $P_0 f(x_l) = 0$ almost surely by the Strong Law of Large Numbers for each $l \in \{1, 2, \ldots, l_\delta\}$. Putting these together, we have

$$\sup_{x \in U} |P_n f(x)| \leq \underbrace{\sup_{x \in U} \min_{1 \leq l \leq l_\delta} |P_n f(x) - P_n f(x_l)|}_{\leq P_n \bar{f}(\delta) \to P_0 \bar{f}(\delta) \text{ a.s.}} + \underbrace{\max_{1 \leq l \leq l_\delta} |P_n f(x_l)|}_{\to 0 \text{ a.s.}}.$$

Thus, it follows almost surely

$$\limsup_{n \to \infty} \|P_n f\|_U \leq P_0 \bar{f}(\delta) < \varepsilon.$$

That is, $P_n f$ converges uniformly to 0 on U almost surely. □

B.4 The Stochastic Case

B.8 Theorem:

Let X be a closed, convex, and locally compact subset of some Banach space. Let Y be a closed subset of $\mathbb{R}^m$. Let P_0 be a probability measure on Y. For $n \in \mathbb{N}$ let P_n be a sample distribution of size n given by independently sampling P_0. For $n \in \mathbb{N} \cup \{0\}$ let X_n be some random non-empty closed convex subset of X, let Y_n be some random non-empty compact subset of Y, and let $\eta_n > 0$. Let $f, g : X \times Y \to \mathbb{R}$. For $n \in \mathbb{N} \cup \{0\}$ define

$$f_n^* = \inf_{S_n} P_n f \qquad \text{with} \qquad S_n = \left\{ x \in X_n \mid \max_{y \in Y_n} g(x; y) \leq \eta_n \right\}.$$

Assume that

1. *X_0, Y_0, and η_0 are deterministic with $\eta_0 = 0$;*
2. *$f(x; \cdot)$ is P_0-integrable for each $x \in X$;*
3. *$P_0 f$ has exactly one minimizer $\bar{x}_0$ on S_0;*
4. *there exists a totally bounded neighborhood U of $\bar{x}_0$ with*

$$\lim_{\delta \to 0} \int \omega_f(\delta, U \times \{y\}) \, \mathrm{d}P_0(y) = 0;$$

5. *and there exists a neighborhood V of Y_0 with almost surely*

$$\omega_g(\operatorname{dist}(X_n, \bar{x}_0), U \times V) = \mathcal{O}(\eta_n),$$

and almost surely

$$\omega_g(\operatorname{dist}(Y_n, Y_0), U \times V) = \mathcal{O}(\eta_n).$$

Then, the following statements are true:

- *The optimal value f_n^* converges almost surely to f_0^*.*
- *For each $n \in \mathbb{N}$ let $\varepsilon_n > 0$ and x_n be a X-valued random variable with a.s.*

$$P_n f(x_n) < f_n^* + \varepsilon_n.$$

If ε_n converges to 0, then x_n converges a.s. to $\bar{x}_0$.

Proof: The assumptions of Theorem B.5 are satisfied almost surely, and the assumptions of Theorem B.7 are satisfied. Both together imply that the assumptions of Theorem B.2 are satisfied almost surely, which yields the almost sure convergence of f_n^* to f_0^*. Further, Theorem B.1 yields the almost sure convergence of x_n to $\bar{x}_0$. □

Appendix C

Uniform Law of the Iterated Logarithm for Sample Quantiles

The Law of the Iterated Logarithm about sum of random numbers is a well studied subject. The particular case about sum of independent and identically distributed random numbers $X_1, X_2, ...$ with finite variances is well-known [15]:

$$X_1 + X_2 + \cdots + X_n - n\,\mathbf{E}[X_1] = \mathcal{O}\left(\sqrt{n \log\log n}\right) \quad \text{almost surely.}$$

Applying it pointwise to the sample distribution function

$$nF_n(x) = \sum_{k=1}^{n} \mathbb{1}_{(-\infty,x]}(X_n)$$

yields only a pointwise version of Law of the Iterated Logarithm for F_n. However, Chung showed in [3] that the law is in fact valid uniformly, if the distribution function F_0 of X_1 is continuous. That is

$$\|F_n - F_0\|_\infty = \mathcal{O}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{almost surely.}$$

Using that result, we can elementarily obtain a locally uniform Law of the Iterated Logarithm for the sample quantile function, if F_0 is strictly increasing, and F_0^{-1} is locally lipschitz continuous. Notice that the Bahadur representation of the sample quantiles yields a more precise asymptotic (compare to

[2, 14]) and also allows us to weaken the independent assumption (compare to [20]). However, such type of arguments usually requires stronger regularity on F_0, like twice continuous differentiability.

C.1 Assumption: *For $n \in \mathbb{N} \cup \{0\}$ let F_n be a random increasing function from $\mathbb{R}$ into $[0,1]$ and let Q_n be a random increasing function from $(0,1)$ into $\mathbb{R}$. Let $K = [\underline{\tau}, \bar{\tau}] \subseteq (0,1)$ be a compact interval. For each $\delta > 0$ let $K_\delta = [\underline{\tau} - \delta, \bar{\tau} + \delta]$. Assume that*

- *F_0 and Q_0 are deterministic, strictly increasing, and continuous;*
- *$F_0 = Q_0^{-1}$;*
- *$\|F_n - F_0\|_\infty$ and $\|F_n \circ Q_n - \mathrm{id}_{(0,1)}\|_\infty$ converge to 0 almost surely.*

C.2 Lemma: *Additionally to Assumption C.1, assume F_n and Q_n being deterministic for each $n \in \mathbb{N}$.*

1. *For each $\tau \in (0,1)$ it follows $Q_n(\tau) \to Q_0(\tau)$.*
2. *For each $\delta > 0$ there exists some $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ and $\tau \in K$ it follows*
$$F_0(Q_n(\tau)), F_n(Q_n(\tau)) \in K_\delta.$$
3. *For each $\delta \in (0, \min\{\underline{\tau}, 1 - \bar{\tau}\})$ it follows*
$$\|Q_n - Q_0 \circ F_n \circ Q_n\|_{\infty,K} = \mathcal{O}(\omega_{Q_0}(\|F_n - F_0\|_\infty, K_\delta).$$
4. *In particular, for each $\delta \in (0, \min\{\underline{\tau}, 1 - \bar{\tau}\})$ it follows*
$$\|Q_n - Q_0\|_{\infty,K} = \mathcal{O}(\omega_{Q_0}(\|F_n - F_0\|_\infty + \|F_n \circ Q_n - \mathrm{id}_{(0,1)}\|_\infty, K_\delta)).$$

Proof: Ad Item 1: By assumption, we have

$$\|F_0 \circ Q_n - \mathrm{id}_{(0,1)}\|_\infty \leq \|F_0 \circ Q_n - F_n \circ Q_n\|_\infty + \|F_n \circ Q_n - \mathrm{id}_{(0,1)}\|_\infty \to 0.$$

By the continuity of Q_0, it follows that $Q_n(\tau)$ converges to $Q_0(\tau)$

Ad Item 2: Fix $\delta > 0$. Then, there exists some $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ it follows

$$\|F_0 - F_n\|_\infty < \frac{\delta}{3}$$

and

$$\|F_n \circ Q_n - \mathrm{id}_{(0,1)}\|_\infty < \frac{\delta}{3}.$$

Thus, for every $\tau \in K$, we obtain

$$\underline{\tau} - \delta\tfrac{2}{3} < F_0(Q_n(\underline{\tau})) \leq F_0(Q_n(\tau)) \leq F_0(Q_n(\bar{\tau})) < \bar{\tau} + \delta\tfrac{2}{3}$$

and

$$\underline{\tau} - \delta < F_n(Q_n(\tau)) < \bar{\tau} + \delta.$$

Ad Item 3: Fix $\delta \in (0, \min\{\underline{\tau}, 1 - \bar{\tau}\})$. There exists some $n_0 \in \mathbb{N}$ such that for each $n \geq n_0$ and $\tau \in K$ it follows

$$F_0(Q_n(\tau)), F_n(Q_n(\tau)) \in K_\delta.$$

Particularly, we have

$$|Q_0(F_0(Q_n(\tau))) - Q_0(F_n(Q_n(\tau)))| \leq \omega_{Q_0}(\|F_0 - F_n\|_\infty, K_\delta). \qquad \square$$

C.3 Theorem:
Additionally to Assumption C.1: For $n \in \mathbb{N}$ assume that F_n is a sample distribution function of sample size n given by an independent sampling of F_0 and that Q_n is the sample quantile function to F_n. Then, for each $\delta \in (0, \min\{\underline{\tau}, 1 - \bar{\tau}\})$ it follows almost surely

$$\|Q_n - Q_0\|_{\infty,K} = \mathcal{O}(\omega_{Q_0}(c_n + \tfrac{1}{n}, K_\delta))$$

with

$$c_n := \sqrt{\frac{\log\log n}{2n}}.$$

In particular, if Q_0 is Lipschitz continuous on a neighborhood of K, then we have almost surely

$$\|Q_n - Q_0\|_{\infty,K} = \mathcal{O}(c_n).$$

Proof: By Chung's Uniform Law of Iterated Logarithms[3] we have almost surely

$$\|F_n - F_0\|_\infty = \mathcal{O}(c_n).$$

Let $\zeta_1, \zeta_2, \ldots, \zeta_n$ denote the samples corresponding to F_n. By definition, F_n is a right-continuous step function with jumps at the samples $\zeta_1, \zeta_2, \ldots, \zeta_n$

and with jumps of size $\frac{1}{n}$. Further, Q_n is a left-continuous step function with jumps at $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$ and the values $\zeta_1, \zeta_2, \ldots, \zeta_n$. Thus, for each $k \in \{1, 2, \ldots, n\}$ and $\tau \in (\frac{k-1}{n}, \frac{k}{n}]$ it follows

$$F_n(Q_n(\tau)) = \frac{k}{n}.$$

In particular, we have

$$\sup_{\tau \in (0,1)} |F_n(Q_n(\tau)) - \tau| \leq \frac{1}{n}.$$

Together, these yield the statement. □

Glossary

finite dimensional subset A subset of some linear space whose linear span is finite dimensional. 17

homeomorphism A continuous and bijective function with continuous inverse. 3

integrable random vector The expectation of the norm is finite. 6

proper domain The subset of the domain on which an extended real valued function is finite. 11

C Glossary

Index

Bibliography

[1] H.W. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer-Lehrbuch Masterclass. Springer, 2006.

[2] R Raj Bahadur. A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37(3):577–580, 1966.

[3] Kai-Lai Chung. An estimate concerning the kolmogroff limit distribution. *Transactions of the American Mathematical Society*, 67(1):pp. 36–50, 1949.

[4] M. G. Cox. The numerical evaluation of b-splines. *IMA Journal of Applied Mathematics*, 10(2):134–149, 1972.

[5] Carl de Boor. On calculating with b-splines. *Journal of Approximation Theory*, 6(1):50–62, 1972.

[6] Carl de Boor. Good approximation by splines with variable knots. In A. Meir and A. Sharma, editors, *Spline Functions and Approximation Theory*, volume 21 of *ISNM International Series of Numerical Mathematics / Internationale Schriftenreihe zur Numerischen Mathematik / Série Internationale d'Analyse Numérique*, pages 57–72. Birkhäuser Basel, 1973.

[7] Carl de Boor. Splines as linear combinations of b-splines. a survey, 1976.

[8] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Grundwissen Mathematik. Springer, 2007.

[9] Thomas S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Taylor & Francis, 1996.

[10] Robert Fourer and Sanjay Mehrotra. Solving symmetric indefinite systems in an interior-point method for linear programming. *Mathematical Programming*, 62(1-3):15–39, 1993.

[11] Florian Jarre, Wolfang Kiess, Martin Mauve, Magnus Roos, and Björn Scheuermann. Least square timestamp synchronization for local broadcast networks. *Optimization and Engineering*, 11(1):107–123, 2010.

[12] Florian Jarre and Josef Stoer. *Optimierung*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2003.

[13] Samuel. Karlin. *Total Positivity*, volume 1 of *Total Positivity*. Stanford University Press, 1968.

[14] J Kiefer. On bahadur's representation of sample quantiles. *The Annals of Mathematical Statistics*, pages 1323–1342, 1967.

[15] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springerlehrbuch Masterclass. Springer, 2008.

[16] I. Lustig, R. Marsten, and D. Shanno. On implementing mehrotra's predictor–corrector interior-point method for linear programming. *SIAM Journal on Optimization*, 2(3):435–449, 1992.

[17] Sanjay Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[18] D.L. Mills. *Computer Network Time Synchronization: The Network Time Protocol*. Taylor & Francis, 2006.

[19] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

[20] CR Rao and LC Zhao. Law of the iterated logarithm for empirical cumulative quantile regression functions. *Statistica Sinica*, 6:693–702, 1996.

[21] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):pp. 832–837, 1956.

[22] B. Scheuermann, W. Kiess, M. Roos, F. Jarre, and M. Mauve. On the time synchronization of distributed log files in networks with local broadcast media. *IEEE/ACM Transactions on Networking*, 17(2):431–444, 2009.

[23] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2009.

[24] K.R. Stromberg. *Introduction to classical real analysis*. Wadsworth international mathematics series. Wadsworth International Group, 1981.

[25] Jos F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653, 1999.

[26] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

[27] AW. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, 1996.

[28] Wolfgang Walter. *Analysis 2:*. Analysis. Springer Berlin Heidelberg, 2002.

[29] J.G. Webster and H. Eren. *Measurement, Instrumentation, and Sensors Handbook, Second Edition: Electromagnetic, Optical, Radiation, Chemical, and Biomedical Measurement*. Measurement, instrumentation, and sensors handbook. Taylor & Francis, 2014.