Estimating Missing Entries of a Partial Mean First Passage Time Matrix

Masterarbeit
am Fachbereich Mathematik und Informatik
der Technischen Universität Berlin

vorgelegt von
Andreas Grever

17.11.2016

Gutachter:

PD Dr. Konstantin Fackeldey
PD Dr. Marcus Weber

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 17.11.2016

Andreas Grever

Danksagung

Ich möchte allen danken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Mein besonderer Dank gilt Marcus Weber für die hilfreichen Gespräche und allgemeine Unterstützung.

Jannes Quer und Marta Bednarczyk möchte ich danken für Feedback und Korrektur.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Frage, wie die Matrix der erwarteten Trefferzeiten (MFP-Matrix) einer Markov Kette abgeschätzt und vervollständigt werden kann, wenn einige der Einträge bereits bekannt sind. Es wird eine Verbindung zur Theorie der $\mathcal{M}$-Matrizen gezeigt, die es ermöglicht obere und untere Schranken für die unbekannten Einträge zu berechnen. Diese Intervalle können durch Intervallarithmetik weiterhin eingegrenzt werden. Schließlich werden zwei Optimierungsansätze vorgestellt, die eine Vervollständigung der partiellen Matrix berechnen.

Die Motivation für diese Arbeit stammt aus dem Bereich der Moleküldynamik. In der Moleküldynamik wird das Verhalten von Molekülen mittels Computersimulationen erforscht. Besonders ist man an Biomolekülen wie Proteinen oder Wirkstoffen für neue Medikamente interessiert. Eine der Herausforderungen besteht in den großen Zeitskalen die überwunden werden müssen, da die Zeitschritte der Simulation in der Größenordnung von Femtosekunden (10^{-15}s) liegen, während viele interessante Vorgänge wie die Faltung von Proteinen einige Mikrosekunden benötigen. Zwar ist der auf Moleküldynamik spezialisierte Supercomputer *Anton 2* von D. E. Shaw Research in New York in der Lage bis zu 85 Mikrosekunden eines Proteins mit 23.558 Atomen an einem Tag zu berechnen, gewöhnliche Maschinen sind allerdings weit entfernt von solchen Leistungen (siehe [41]).

Eine weitere Herausforderung ist die Analyse der hochdimensionalen Simulationsdaten, da jedes Atom drei Freiheitsgrade besitzt. Beispiele für Methoden, die Komplexität reduzieren und gleichzeitig den stochastischen Charakter von Molekülbewegungen berücksichtigen, sind " Markov State Models "(MSM, [36, 40]). In MSMs wird der Zustandsraum in Untermengen aufgeteilt und die Übergangswahrscheinlichkeiten zwischen diesen berechnet. Diese Wahrscheinlichkeiten lassen sich als Übergangsmatrix darstellen, welche wiederum zur Berechnung der Matrix der erwarteten Trefferzeiten (mean first passage times) verwendet werden kann. Die erwartete Trefferzeit m_{ij} gibt an wie lange der Prozess, der in Zustand i gestartet ist, braucht, um Zustand j zu erreichen. Diese Werte sind charakteristische Größen des Prozesses und enthalten wichtige Informationen. Für den Wirkstoffdesign ist es zum Beispiel nicht nur wichtig, dass das Medikament mit hoher Wahrscheinlickeit an sein Ziel kommt, sondern auch, dass es möglichst lange dort bleibt ([32]).

Die Kinetik von Biomolekülen wird in vielen Fällen modelliert mittels zwei Hauptmechanismen. Einerseits strebt das System einen Zustand niedriger Energie an wobei die Energiefunktion viele lokale Minima besitzt. Gleichzeitig bewirkt eine sogenannte *Brown-*

sche Bewegung spontane Änderungen im System. In der Konsequenz hält sich das Molekül für lange Perioden in der Nähe lokaler Minima auf und ändert seine Konformation nur selten. Diese stabilen Zustände werden manchmal auch Metastabilitäten genannt ([36]). In den letzten Jahren wurde vermehrt nach Methoden geforscht, mit denen Simulationen gelenkt werden können, sodass seltene Ereignisse häufiger und mit geringer Varianz auftreten (siehe [11, 35]). Eine kleine Varianz bedeutet, dass ein beobachtetes Ereignis mit hoher Wahrscheinlichkeit in der Nähe des Erwartungswertes liegt. Mit diesen Methoden kann zum Beispiel die durchnittliche Zeit einen bestimmten Zustand zu verlassen schneller und mit weniger Daten berechnet werden.

Die Einträge der Matrix der erwarteten Trefferzeiten (MFP-Matrix) können entweder direkt mittels Simulationsdaten oder indirekt über die Übergangswahrscheinlichkeiten berechnet werden. Das Ziel dieser Arbeit ist es zu zeigen, dass falls die erwarteten Trefferzeiten direkt berechnet werden, es nicht nötig ist alle Zeiten zu berechnen. Schon ein kleiner Teil der Einträge genügt um Intervalle anzugeben in denen sich die unbekannten Werte befinden.

Es stellt sich heraus, dass eine positive Matrix N die MFP-Matrix einer Markov Kette ist, falls die Einträge von N und die Einträge der Inversen der Matrix

$$\begin{bmatrix} N & e \\ e^\top & 0 \end{bmatrix}$$

bestimmte Ungleichungen erfüllen, wobei e der Vektor mit ausschließlich Einsen ist. Die Herleitung dieser Ungleichungen beruht auf der Theorie der $\mathcal{M}$-Matrizen, mit dessen Hilfe es möglich ist nicht nur die unbekannten Einträge der MFP-Matrix abzuschätzen, sondern auch die Übergangswahscheinlichkeiten des zugrunde liegenden Prozesses.

Lineare Systeme für Matrizen und Vektoren deren Einträge durch Intervalle anstatt Zahlen gegeben sind, können mittels Intervallarithmetik beschrieben werden. Methoden aus dem Bereich der Intervallarithmetik können verwendet werden, um die bisherigen Abschätzungen zu verbessern.

Um nicht nur Abschätzungen sondern auch eine konkrete Matrix zu erhalten, müssen Matrizen X, Y gefunden werden sodass neben diversen linearen Bedingungen es gilt, dass X die inverse Matrix zu Y ist. Mit anderen Worten soll $\|XY - I\|$ über einer konvexen Menge minimiert werden. Es werden zwei Ansätze vorgestellt um dieses Optimierungsproblem zu lösen. Zunächst wird gezeigt, dass es sich um ein *d. c. programming* Problem handelt. Anschließend wird eine linearisierte Variante des Problems formuliert und auf zwei Beispiele angewendet.

Die Methoden, die in dieser Arbeit beschrieben werden bedeuten eine deutliche Zeitersparnis im Vergleich zur direkten Berechnung von Trefferzeiten mittels Sampling von Daten in hochdimensionalen Räumen.

Contents

Introduction

The initial motivation for this thesis came from some new ideas in the field of molecular dynamics. In molecular dynamics computer simulations are used to study the behavior of molecules like proteins on an atomic level. One of the challenges in this field is to overcome great gaps in spatial and temporal scale. While the simulation requires the computation of forces between atoms in discrete time steps of femtoseconds (10^{-15}s), the events that are being studied often take place on the level of microseconds (10^{-6}s), like the folding of a protein. Though the specialized supercomputer *Anton 2* built by D. E. Shaw Research in New York simulates up to 85 microseconds per day for a protein with 23,558 atoms (see [41]), ordinary machines are far from such results, which makes methods that reduce simulation time very interesting.

Another challenge of molecular dynamics is to analyze the complex high dimensional simulation data. One approach to overcome complexity is to use coarse grained representations of the domain. An example for such methods that at the same time reflects the nondeterministic character of molecules in motion are Markov State Models ([36]). In Markov State Models the state of the system at a any time is described by a probability distribution that is propagated in time by a so called *transfer operator* ([40]). The domain of the distribution function is partitioned into a finite number of subsets and the time series of the simulation is used to compute the transition probabilities between these subsets. The probabilities can be represented as a transition matrix, which itself can be used to compute the matrix of mean first passage times between each pair of events, i.e., the average time it will take to get from one state to another. Mean first passage times are characteristic quantities of the process and give valuable information about the systems behavior. In the field of drug design it is, for example, not only important that the drug has a high tendency to reach its target, but also that it stays there sufficiently long ([32]).

The kinetics of biomolecules often involve several states in which the molecule resides for a long time (metastabilities, see [36] and references therein). As a result, significant changes occur rarely and on different timescales. Recently, developments have been made in manipulating the simulation to steer the system, such that on the one hand rare events are not rare anymore, and on the other hand reducing the variance of this event (see [11, 35]). A small variance means that the observed event is likely to be close to the expected or average outcome of multiple experiments. With these methods the average time it takes to reach or leave a certain state can be computed faster using less data.

The matrix of mean first passage times (MFP-matrix) can either be sampled directly

using simulations, or indirectly using a transition matrix. The aim of this thesis is to show that if the mean first passage times are being computed directly, it is not necessary to compute all of the matrix entries, but that only a fraction is needed to give intervals in which the missing entries can be found. Furthermore, we present a strategy to complete the matrix such that it defines a unique Markov chain that is at least close to the one that was initially simulated.

The intervals can be utilized in various ways. Apart from defining the feasible set in for matrix completion, large intervals can be used during simulation as an indicator for which states should be sampled next to improve the estimates up to that point. Small intervals can contain information on transition states, e.g. states that have to be visited, to move from one metastability to another. It should be noted that these results are not limited to molecular dynamics but in fact are applicable to any irreducible Markov chain.

The approach of this thesis is somewhat unusual, because in most situations the matrix of transition probabilities is known and can be used to compute the matrix of mean first passage times. This is why the problem of determining whether a positive matrix is the MFP-matrix of a Markov chain will be called the *inverse MFP-matrix problem* (see also [34]). Because some of the entries are assumed to be known the problem at hand is the following.

MFP-matrix completion problem: *Given a partial MFP-matrix, how can the missing entries be chosen such that the completed matrix is the MFP-matrix of a Markov chain?*

It will turn out that a positive matrix N is the MFP-matrix of a Markov chain if the inverse of the matrix

$$\begin{bmatrix} N & e \\ e^\top & 0 \end{bmatrix}$$

and N itself meet several linear conditions, where e is the vector ones. Conditions on the inverse of a matrix are nonlinear and make the problem difficult to solve.

In the first chapter some basics of Markov chains are presented and it is shown how to compute mean first passage times from transition matrices.

The MFP-matrix completion problem is formulated in chapter 2. The problem is explored by drawing a connection between MFP-matrices and a class of matrices called $\mathcal{M}$-matrices.

In chapter 3 the theory of $\mathcal{M}$-matrices is utilized in order find several inequalities not only for mean first passage times but also for the transition probabilities and the stationary distribution, i.e., the unknown entries are replaced by intervals.

Putting matrices specified by intervals into relation with each other leads to the field of interval arithmetics. Methods to improve the information given by linear equations involving intervals are described in chapter 4.

In chapter 5, to find a matrix that not only respects the upper and lower bounds but also is a MFP-matrix of a Markov chain, the problem is formulated as a *d.c.* program and numerical examples are discussed.

Chapter 1

The MFP-Matrix of a Markov Chain

Markov State Models (MSM) are a coarse graining method to analyze the structural changes in molecular dynamics simulation data ([36]) by modeling the process as a memoryless stochastic process, i.e., a Markov chain. This means that the state space is projected onto a finite number of subsets, such that the simulated time series can be viewed as jumps in the network defined by these subsets. The subsets will be referred to as states. The probabilities to jump from one state to another can be described by a stochastic transition matrix. In this first chapter, it will be shown how the mean first passage times of a Markov chain can be computed using simple matrix equations. After that the inverse MFP-matrix problem can be formulated.

The basics from probability theory and Markov chains that will be introduced here serve primarily as a motivation and illustration, because we will discuss the problem of estimation and completion using the properties of the matrices and the equations that relate them with each other.

Definition 1.1 (Notation)**.** *For the whole thesis the following notation will be used:*

(i) The matrix E is the matrix with all entries equal to 1

(ii) The column vector e is the vector with all entries equal to 1

(iii) For a vector v the diagonal matrix, whose diagonal entries are given by v will be denoted by $diag(v)$. For a matrix A the vector, whose entries are the diagonal entries of A will be denoted by $diag(A)$.

(iv) In the context of matrices and vectors the relations $\leq, <, \geq, >$ are to be understood as entrywise inequalities.

1.1. Markov Chains and Transition Probabilities

To understand the fundamental ideas of probability theory that we need in this chapter some basic definition from the field of measure theory will be recalled. For a more detailed introduction to probability theory see e.g. [7].

Definition 1.2. *Let Ω be an arbitrary set and $\mathcal{F}$ a nonempty collection of subsets of Ω.*

If $\mathcal{F}$ satisfies

(i) $A \in \mathcal{F}$ then $A^c := \Omega \setminus A \in \mathcal{F}$, and

(ii) $A_i \in \mathcal{F}$ is a countable sequence of sets then $\cup_i A_i \in \mathcal{F}$

it is called a σ-algebra. The tuple $(\Omega, \mathcal{F})$ is called a measurable space.

Definition 1.3. *A measure $\mu : \mathcal{F} \to \mathbb{R}$ on a measurable space $(\Omega, \mathcal{F})$ is a function with*

(i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$,

(ii) if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu(\cup_i A_i) = \sum_i \mu(A_i). \tag{1.1}$$

If $\mu(\Omega) = 1$, it is a probability measure and will be denoted by $\mathbb{P}$. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space.

Using a discretized time series on n states it would be possible to compute how often a certain state has been visited and divide that number by the length of the time series to obtain the probability that the system is found in that state. Though to describe the dynamic of the system it would be more interesting to know where a system that is residing at a state i will go next and how likely it is that it will go to a state j. Therefore, we need to introduce the conditional probability.

Definition 1.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$. The conditional probability of the event A given the event B is defined by*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cup B)}{\mathbb{P}(B)}.$$

In the following the notation $\mathbb{P}(A, B) := \mathbb{P}(A \cup B)$ will be used. We will now consider a discrete-time stochastic process $X : \mathbb{N}_0 \to 1, \cdots, n$ in a system of n states. The function X is a random variable and can be seen as an evaluation of an experiment, where at every time step k the process takes as a value one of the states, i.e. $X_k = i$ for $k \in \mathbb{N}_0, i \in \{1, \cdots, n\}$.

Definition 1.5. *A discrete-time stochastic process is called a Markov process if it meets the Markov property*

$$\mathbb{P}(X_k = i_k | X_{k-1} = i_{k-1}, \cdots, X_0 = i_0) = \mathbb{P}(X_k = i_k | X_{k-1} = i_{k-1}) \qquad \text{for all } k \in \mathbb{N}. \tag{1.2}$$

It is called a Markov chain, if equation (1.2) does not depend on k.

The Markov property is often referred to as memoryless, because the future only depends on the present state. To justify this assumption in molecular dynamics the time steps and discretization have to be chosen with care.

Because for Markov chains equation (1.2) does not depend on k, the conditional probability to jump from i to j in one step can be simply denoted by p_{ij}, which means that every Markov chain on n states is defined by a transition matrix P.

Definition 1.6. *A matrix $P \in \mathbb{R}^{n\times n}$ with the properties*

(i) $P \geq 0$,

(ii) $Pe{=}e$

will be called transition matrix. A vector $v \in \mathbb{R}^n$ such that $v \geq 0$ and $v^\top e = 1$ is called a probability vector.

If an initial distribution is given by a probability vector $v^{(0)}$, $v_i^{(0)} := \mathbb{P}(X_0 = i)$, the distribution after one step is given by

$$\mathbb{P}(X_1 = i) = \sum_{k=1}^{n} \mathbb{P}(X_1 = i, X_0 = k) = \sum_{k=1}^{n} \mathbb{P}(X_1 = i | X_0 = k)\mathbb{P}(X_0 = k),$$

i.e. $(v^{(1)})^\top = (v^{(0)})^\top P$ and more generally, the distribution at the k-th step is given by $(v^{(k)})^\top = (v^{(0)})^\top P^k$. The matrix P^k is the k-th matrix product of P with itself and its entries $p_{ij}^{(k)}$ are the probabilities to start at i and be in j after k steps. This relationship is summarized in the Chapman-Kolmogorov equation

Corollary 1.7 (Chapman-Kolmogorov equation,[7])**.**

$$\mathbb{P}(X_{k+h} = j | X_0 = i) = \sum_{s=1}^{n} \mathbb{P}(X_{k+h} = j | X_k = s)\mathbb{P}(X_k = s | X_0 = i)$$

Apart from the interpretation of the initial distribution as the probability measure describing the probability that a process can be found in a certain state at time 0, we may also understand it as the given distribution of an ensemble of particles at the beginning of the process. The transition probabilities p_{ij} are then interpreted as the percentage of particles at state i that move to state j in one time step. This interpretation as an ensemble is used especially in thermodynamics and therefore in molecular dynamics from which our initial motivation stems.

In the next section the first passage time τ_{ij} will be introduced, which is the first time a process that started in state i reaches state j. For this time to be a finite number the transition matrix has to be irreducible.

Definition 1.8 ([10])**.** *An $n \times n$ matrix A with $n \geq 2$ is called reducible if there exists a permutation matrix S such that*

$$S^\top A S = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

with A_{11}, A_{22} square. A matrix is called irreducible if it is not reducible.

It is easy to see that if A is reducible then A^k is reducible as well for every $k \geq 1$. From a probabilistic point of view a Markov chain defined by a reducible transition matrix has a subset of states that can't be left, not even after multiple steps. Eigenvalues and eigenvectors can be used to analyze transition matrices and the systems they model (see [5, 42]). A very useful tool in this context is the Perron-Frobenius-Theorem.

Definition 1.9. *For a square matrix A the set of all eigenvalues of A is called the spectrum and is denoted by $\sigma(A)$. The spectral radius of A is defined by*

$$\rho(A) := \max\{|\lambda|, \lambda \in \sigma(A)\}$$

Lemma 1.10 (Perron-Frobenius-Teorem,[10])**.** *Let $A \geq 0$ be an irreducible matrix. Then the following hold:*

(i) The spectral radius $\rho(A)$ is a simple eigenvalue of A

(ii) The corresponding eigenvector is elementwise positive and is the only eigenvector with that property

(iii) If A has h eigenvectors with eigenvalues $\rho(A) = \lambda_1, \cdots, \lambda_h$ such that $|\lambda_i| = \rho(A)$ for $i = 1, \cdots, h$, then these eigenvalues are all simple and solve the equation $\lambda^h - \rho(A)^h = 0$. Moreover the Matrix can be reduced to the cyclic block form

$$\begin{pmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ 0 & 0 & A_{23} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{(h-1)h} \\ A_{h1} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where on the diagonal we have square matrices with zero entries.

In the case of an irreducible transition matrix P it holds that $Pe = e$, which means that e is the eigenvector associated with the eigenvalue $1 = \rho(P)$. Furthermore, the Perron-Frobenius-Theorem guaranties the existence of a left eigenvalue such that

$$\pi^\top P = \pi^\top, \qquad \sum_{i=1}^{n} \pi_i = 1,$$

where the second property can be achieved by scaling.

This positive left eigenvector π is called the stationary distribution of P i.e., if an ensemble starts out at time zero distributed according to π, the distribution will stay unchanged after each time step. The value h in the Perron-Frobenius-Theorem gives another classification of processes. For $h > 1$ the process is called periodic with period h, otherwise the process is called aperiodic. For an aperiodic irreducible transition matrix there exists an additional property regarding the stationary distribution vector π, namely

$$\lim_{k\to\infty} P^k = e\pi^\top.$$

This means that any initial distribution converges to the stationary distribution. The last concept we want to introduce in this chapter is reversibility.

Definition 1.11 ([23])**.** *Let P be an irreducible transition matrix of a Markov chain with stationary distribution π and set $\Pi := diag(\pi)$. The Markov chain is called reversible, if the detailed balance condition*

$$\pi_i p_{ij} = \pi_j p_{ji} \quad \textit{for all } i,j = 1,\dots,n \qquad \Leftrightarrow \qquad \Pi P = P^\top \Pi$$

is fulfilled, otherwise it is called irreversible.

The motivation for the term *reversible* becomes clear if we write $\pi_i p_{ij} = \mathbb{P}(X_0 = i, X_1 = j)$ i.e., a jump from i to j is as likely as the opposite direction. Reversibility is an important property in molecular dynamics on the one hand, because it is the behavior one would expect from the reversible Newtonian equations. The symmetry of ΠP on the other hand is an assumption in analytic methods like spectral clustering.

1.2. From Transition Matrix to MFP-Matrix

The transition matrix in itself gives answers to the question what might happen in the next time step. If we think of a Markov chain as a spatial process making small jumps, the transition probabilities give us only information about the events near in time and near in space.

In molecular dynamics, the system typically stays for a long time in certain so called *metastable states* between which jumps occur very rarely. This raises the question: How long do we have to expect to wait?

Definition 1.12. *(i) The first passage time τ_{ij} of a Markov chain is the random variable describing the time the process X takes to reach the state j if it started in state i, i.e.*

$$\tau_{ij} := \min_k\{k \in \mathbb{N} | X_0 = i, X_k = j, X_s \neq j \ \ s = 1, \cdots, k-1\} \qquad i, j = 1, \cdots, n$$

τ_{ii} is also called the return time to i.

(ii) The mean first passage time (MFP) m_{ij} of a Markov chain is the expected time the process that started in i takes to reach j and is defined by the number

$$m_{ij} := \mathbb{E}[\tau_{ij}] = \sum_{k=1}^{\infty} kP(\tau_{ij} = k)$$

It can be shown that the mean first passage times for an irreducible Markov chain are always finite (see [23]), which is why we can consider the matrix $M = m_{ij} \in \mathbb{R}^{n\times n}$, which we will call the mean first passage time matrix (MFP-matrix). The MFP-matrix has obviously positive entries and we will develop in the following several relationships between M and the transition matrix P. The following lemma is quite intuitive and it will help us to relate the matrices M and P in form of a matrix equation.

Lemma 1.13 ([25])**.** *Let τ_{ij}, $i, j = 1, \cdots, n$ be the first passage time of a Markov chain on n states. Then for $k > 1$ it holds that*

$$\mathbb{P}(\tau_{ij} = k) = \sum_{s\neq j} p_{is}\mathbb{P}(\tau_{sj} = k - 1).$$

Proof. Assuming $k > 1$, we write:

$$\begin{aligned}
\mathbb{P}(\tau_{ij} = k) &= \mathbb{P}(X_k = j, X_l \neq j,\ j = 2, \cdots, k-1 | X_0 = i) \\
&= \frac{\mathbb{P}(X_k = j, X_l \neq j\ j = 1, \cdots, k-1, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{s \neq j} \frac{\mathbb{P}(X_k = j, X_l \neq j\ j = 2, \cdots, k-1, X_1 = s, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{s \neq j} \frac{\mathbb{P}(X_k = j, X_l \neq j\ j = 2, \cdots, k-1 | X_1 = s, X_0 = i)\mathbb{P}(X_1 = s, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{s \neq j} \mathbb{P}(X_k = j, X_l \neq j\ j = 2, \cdots, k-1 | X_1 = s)\mathbb{P}(X_1 = s | X_0 = i) \\
&= \sum_{s \neq j} p_{is}\mathbb{P}(\tau_{sj} = k-1)
\end{aligned}$$

□

Theorem 1.14 ([23])**.** *The MFP-matrix for an irreducible transition matrix P with stationary distribution π is the unique solution of the matrix equation*

$$M = E + P(M - M_{dg})$$

where M_{dg} is the diagonal matrix with the same diagonal entries as M. Furthermore the diagonal is given by $m_{ii} = \pi_i^{-1}$ for $i = 1, \ldots, n$

Proof. We compute M entrywise:

$$\begin{aligned}
m_{ij} := \mathbb{E}[\tau_{ij}] &= \sum_{k=1}^{\infty} kP(\tau_{ij} = k) = p_{ij} + \sum_{k=2}^{\infty} kP(\tau_{ij} = k) \\
&= p_{ij} + \sum_{k=2}^{\infty} k \sum_{s \neq j} p_{is}P(\tau_{sj} = k-1) = p_{ij} + \sum_{s \neq j} p_{is} \sum_{k=1}^{\infty} (k+1)P(\tau_{sj} = k) \\
&= p_{ij} + \sum_{s \neq j} p_{is} \left[\sum_{k=1}^{\infty} kP(\tau_{sj} = k) + \sum_{k=1}^{\infty} P(\tau_{sj} = k) \right] \\
&= p_{ij} + \sum_{s \neq j} p_{is}m_{sj} + \sum_{s \neq j} p_{is} \\
&= 1 + \sum_{s \neq j} p_{is}m_{sj}
\end{aligned}$$

The last line is equivalent to the desired matrix equation

$$M = E + P(M - M_{dg}) \Leftrightarrow (I - P)M = E - PM_{dg}. \tag{1.3}$$

If we multiply the right equation with $\pi^\top$

$$0 = \pi^\top(I-P)M = \pi^\top(E - PM_{dg}) = e^\top - \pi^\top M_{dg}$$

we see that $m_{ii} = \pi_i^{-1}$ has to hold.

Now let us assume, that $M, \tilde{M}$ are solutions of (1.3). Taking the difference and using that $M_{dg} = \tilde{M}_{dg}$ we get

$$M - \tilde{M} = P(M - \tilde{M})$$

and see that from the Perron-Frobenius-Theorem it follows that the columns of $M - \tilde{M}$ are multiples of e. We know that the diagonal entries are 0 which means that $M - \tilde{M} = 0$. □

In the following we will write $M_{dg} := \Pi^{-1}$ with Π being the diagonal matrix with π as entries and set $G := (I - P)$.

The equation $(I-P)M = E - P\Pi^{-1}$ already indicated a connection between the MFP-matrix and the inverse of $(I-P)$, but of course, due to $\pi^\top(I-P) = (I-P)e = 0$, this matrix is singular. With the concept of generalized inverses we will be able to describe the MFP-matrix further and show some interesting properties.

Definition 1.15 ([1]). *Let $A \in \mathbb{R}^{n\times m}$ be any matrix and consider the following conditions for a matrix $X \in \mathbb{R}^{m\times n}$*

(1) $AXA = A$ *(3)* $(AX)^\top = AX$

(2) $XAX = X$ *(4)* $(XA)^\top = XA$.

If X fulfills any of the above conditions it is called a generalized inverse of A. In the case of A, X being square matrices of the same order there is in addition the condition

(5) $AX = XA$.

A generalized inverse X satisfying the set of conditions $\{i_1, \cdots, i_k\} \subset \{1, \cdots, 5\}$ will be called a $\{i_1, \cdots, i_k\}$-inverse.

Probably the most commonly used generalized inverse is the Moore-Penrose-Inverse, which is the uniquely defined $\{1,2,3,4\}$-inverse. It is used in particular to solve singular linear equations. Jeffrey J. Hunter summarized in [15] several generalized inverses and described how to use them to compute the MFP-matrix, though in the theory of Markov chains it is common to use the $\{1,2,5\}$-inverse that goes by the name *generalized group inverse*. Carl D. Meyer for example showed in [30] the role of the generalized group inverse in the theory of Markov chains and described probabilistic interpretations of its entries.

Another advantage of the generalized group inverse is the fact that the spectral properties are very similar to those of a real inverse. The generalized group inverse of a matrix A is typically denoted by $A^\#$.

Corollary 1.16. *For a transition matrix P the generalized group inverse of $G := I - P$ given by*

$$G^{\#} := (I-P)^{\#} = \left[I - P + e\pi^{\top}\right]^{-1} - e\pi^{\top}$$

is uniquely defined and has the following properties:

(i) $G^{\#}e = \pi^{\top}G^{\#} = 0$

(ii) $I - GG^{\#} = e\pi^{\top}$

(iii) If $\sigma(P) = \{1, \lambda_2, \ldots, \lambda_n\}$ *is the spectrum of* P *then* $\sigma(G^{\#}) = \{0, \frac{1}{1-\lambda_2}, \ldots, \frac{1}{1-\lambda_n}\}$

Proof. Following [8] it holds that any matrix that fulfills conditions $(1), (2)$ and (5) is the unique solution. This can be shown by assuming that $A_1^{\#}$ and $A_2^{\#}$ are two different generalized group inverses of the matrix A and simple application of the conditions:

$$\begin{aligned} AA_1^{\#} &= (AA_2^{\#}A)A_1^{\#} = A_2^{\#}(AA_1^{\#}A) = A_2^{\#}A \\ A_1^{\#} &= A_1^{\#}AA_1^{\#} = A_2^{\#}AA_1^{\#} = A_2^{\#}AA_2^{\#} = A_2^{\#} \end{aligned}$$

Next it will be shown that $G^{\#}$ is the generalized inverse of G by first showing (i) and (ii). The vectors e and π are right and left eigenvectors of $\left[I - P + e\pi^{\top}\right]$ to the eigenvalue 1 and thereby eigenvectors of $\left[I - P + e\pi^{\top}\right]^{-1}$ as well. Property (i) follows from $\pi^{\top}e = 1$. With the same argument the product $GG^{\#}$ can be computed:

$$\begin{aligned} GG^{\#} &= (I-P)(\left[I - P + e\pi^{\top}\right]^{-1} - e\pi^{\top}) \\ &= (I - P + e\pi^{\top} - e\pi^{\top})\left[I - P + e\pi^{\top}\right]^{-1} - (I-P)e\pi^{\top} \\ &= I - e\pi^{\top}\left[I - P + e\pi^{\top}\right]^{-1} \\ &= I - e\pi^{\top} \end{aligned}$$

The equation for $G^{\#}G$ works analogously, which shows conditions (5). Condition (1) and (2) we get be multiplying the last equation in property (ii) with G from the right and $G^{\#}$ from the left respectively.

Let v be an eigenvalue of a matrix A for the non zero eigenvalue λ. Using condition (1) and (5) for generalized inverses we see that

$$\lambda v = Av = AA^{\#}Av = A^{\#}A^2v = \lambda^2 A^{\#}v \quad \Leftrightarrow \quad A^{\#}v = \frac{1}{\lambda}v.$$

If λ is an eigenvalue of P, $1 - \lambda$ is an eigenvalue of G, which shows (iii). □

The formula for $G^{\#}$ presented in the corollary can be interpreted as first adding the vectors spanning the kernel of $I - P$ as a rank one matrix to create a proper inverse and afterwards subtracting that rank one matrix such that G and $G^{\#}$ have the same kernel.

It will be useful to introduce in addition to the MFP-matrix M the matrix

$$N := M - \Pi^{-1},$$

which is just the matrix M with zeros on the diagonal and therefore will also be referred to as MFP-matrix. Plugging N into equation (1.3) and using the last theorem shows that there is a direct connection between $G^{\#}$ and N:

$$\begin{aligned}(I-P)N = E - \Pi^{-1} = (e\pi^\top - I)\Pi^{-1} = -(I-P)G^{\#}\Pi^{-1} \\ \Leftrightarrow (I-P)(G^{\#} + N\Pi)\Pi^{-1} = 0 \\ \Leftrightarrow (I-P)(G^{\#} + N\Pi) = 0\end{aligned}$$

Because e is a simple eigenvalue of P the columns of $G^{\#} + N\Pi$ must be either zero or another multiple of e. We write $N\Pi = ed^\top - G$ with some vector d. The diagonal entries of N are zero, which means that the vector d consists of the diagonal entries of G. We summarize this relationship with a theorem.

Theorem 1.17 ([15])**.** *Let P be the irreducible transition matrix of a Markov chain, π the corresponding stationary vector, $G^{\#}$ the generalized group inverse of $(I-P)$. Then the MFP-matrix M is given by*

$$m_{ij} = \begin{cases} \dfrac{g^{\#}_{jj} - g^{\#}_{ij}}{\pi_j} & i \neq j \\ \dfrac{1}{\pi_i} & i = j. \end{cases}$$

The next lemma shows what effect the addition of rank one matrices have on the spectrum if eigenvectors are involved.

Lemma 1.18 ([17])**.** *Let A be a square matrix with spectrum $\sigma(A) = \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ and u a right eigenvector of A associated with the eigenvalue λ_1. For a vector v the spectrum of the perturbed matrix $A + uv^\top$ is given by $\sigma(A + uv^\top) = \{\lambda_1 + v^\top u, \lambda_2, \cdots, \lambda_n\}$*

The property of the trace operator that $trace(A) = \sum_i \lambda_i$ where the λ_i are the eigenvalues of A yields the following.

Corollary 1.19. *The spectrum $\sigma(N\Pi)$ is determined by the spectrum of the transition matrix $\sigma(P) = \{1, \lambda_2, \ldots, \lambda_n\}$ and is given by*

$$\sigma(N\Pi) = \{\sum_{i=2}^{n} \frac{1}{1-\lambda_i}, -\frac{1}{1-\lambda_2}, \ldots, -\frac{1}{1-\lambda_n}\}$$

Proof. We write $N\Pi$ as rank 1 perturbation of the generalized inverse

$$N\Pi = ed^\top - G^{\#}$$

where d is the vector of the diagonal entries of $G^{\#}$. The eigenvalues of $-G^{\#}$ are the same as those oft $G^{\#}$ with reversed signs and the same eigenvectors. The vector e is an eigenvector of $G^{\#}$ coresponding to the eigenvalue 0 so we can use the last lemma to get

$$\sigma(N\Pi) = \{d^\top e, -\frac{1}{1-\lambda_2}, \ldots, -\frac{1}{1-\lambda_n}\}.$$

Because $d^\top e = trace(G^{\#})$ the corollary is shown. □

The eigenvalue $d^\top e = \sum_{i=2}^{n} \frac{1}{1-\lambda_i}$ is the spectral radius of $N\Pi$ and is connected to the Kemeny constant.

Definition 1.20. *For a transition matrix P with spectrum $\sigma(P) = \{1, \lambda_2, \cdots, \lambda_n\}$ the Kemeny constant is defined by*

$$\tilde{\kappa} := \tilde{\kappa}(P) := \sum_{i=2}^{n} \frac{1}{1-\lambda_i} + 1$$

and $\kappa = \tilde{\kappa} - 1$ will be called the reduced Kemeny constant.

Remark 1.21. *It was argued in [16] that the Kemeny constant could be interpreted as a measure for the expected mixing time,i.e., the expected time until a process that started in an arbitrary state is close to the stationary distribution. From Corollary 1.19 it immediately follows that*

$$M\pi = \tilde{\kappa}e \qquad N\pi = \kappa e,$$

which states a direct connection between MFP matrix and Kemeny constant. This also means that, if the matrix N is known, we can compute the stationary distribution by first solving $Nx = e$ and then setting $\kappa = \sum_i x_i$. The vector π is given by $\pi = \kappa^{-1}x$.

In [23] Kemeny and Snell introduced the matrix

$$Z = \left[I - P + e\pi^\top\right]^{-1} = G^{\#} + e\pi^\top$$

and called it the fundamental matrix, because many of the interesting properties of the process can be deduced from it. Carl D. Meyer has argued in [30], that everywhere the matrix Z is used it can be replaced by the group inverse $G^{\#}$ and showed that its computation is cheaper.

It was shown that the MFP-matrices M and N are not only directly connected to the group inverse, but hold also information on the stationary vector, the Kemeny constant and the spectrum of P. This shows how useful the MFP-matrix can be for the analysis of Markov chains and would deserve the title "fundamental matrix" as well.

After stating the basic properties and connections between transition matrices and MFP-matrices, we are ready to address the problem setting of this thesis.

Chapter 2

Two inverse Problems

So far we have seen how to compute the MFP-matrix given the corresponding transition matrix P, but the main question of this thesis is connected to the inverse problem. We assume that some of the mean first passage times have been computed directly, so the setting from which we start is of the following kind.

Definition 2.1. *A matrix, in which some entries are specified by real numbers while the remaining entries are free to be chosen will be called a partial matrix.*

The topic of this thesis can then be described in the following way.

MFP-matrix completion problem: *Given a partial MFP-matrix, how can the missing entries be chosen such that the completed matrix is the MFP-matrix of a Markov chain?*

This chapter is based on the ideas described by M.Neumann and Nung-Sing Sze in [34]. In their paper they showed a connection between the so called *inverse MFP-problem* and the *inverse $\mathcal{M}$-matrix problem*. The properties of $\mathcal{M}$-matrices will be utilized in the next chapter to find estimates on the missing mean first passage times.

2.1. inverse MFP-matrix problem

The *inverse MFP-problem* for a nonnegative, square matrix N will be understood as the question under what conditions there exists a Markov chain, such that N is the MFP-matrix of that Markov chain.

There are several reasons why we will work with the matrix N instead of the matrix M. The diagonal entries of M can be recovered from N by applying Remark 1.21 and many of the formulas take a simpler form if N is used. Furthermore, if it comes to molecular dynamics the methods to speed up the computation of mean first passage times are not necessarily applicable to the return times, i.e., the diagonal entries. Another advantage of the matrix N is that from Corollary 1.19 it is clear that the inverse of N exists, because all eigenvalues of $N\Pi$ are non zero, so we can write

$$(I-P)N = E - \Pi^{-1} \quad \Leftrightarrow \quad \Pi(I-P) = (\pi e^\top - I)N^{-1}. \tag{2.1}$$

The next lemma states conditions for N^{-1} such that N is an MFP-matrix.

Lemma 2.2 ([34]). *Suppose that N is a nonnegative invertible matrix with diagonal entries equal to zero. Let $N^{-1} := (z_{ij})$ and $\Pi = diag(\pi)$. Then $N = M - \Pi^{-1}$ for some MFP-matrix M of an irreducible transition matrix P with stationary distribution π if and only if*

$$\begin{cases} \sum_{k=1}^{n} z_{ik} > 0 & \text{for all } i = 1, \cdots, n \\ z_{ij} \geq \dfrac{\sum_{k=1}^{n} z_{ik} \sum_{k=1}^{n} z_{kj}}{\sum_{k,l=1}^{n} s_{lk}} & \text{for all } i \neq j, \quad i, j = 1, \cdots, n \\ z_{ii} \dfrac{\sum_{k,l=1}^{n} z_{lk}}{\sum_{k=1}^{n} z_{ik}} - \sum_{k=1}^{n} z_{ki} \geq -1 & \text{for all } i = 1, \cdots, n \end{cases} \tag{2.2}$$

and any matrix $A = a_{ij}$ whose off-diagonal entries are given by

$$a_{ij} = z_{ij} - \frac{\sum_{k=1}^{n} z_{ik} \sum_{k=1}^{n} z_{kj}}{\sum_{k,l=1}^{n} s_{lk}} \qquad \text{for all } i \neq j, \quad i, j = 1, \cdots, n \tag{2.3}$$

is irreducible.

Proof. Let's assume that N is a nonnegative invertible matrix with diagonal entries equal to zero and N^{-1} such that (2.2) and (2.3) hold. Define the vector

$$\pi_i = \frac{\sum_{k=1}^{n} z_{ik}}{\sum_{k,l=1}^{n} z_{lk}} > 0 \qquad \text{for all } i = 1, \cdots, n$$

and the diagonal matrix $\Pi := diag(\pi)$. It needs to be shown that $P = I + (\Pi^{-1} - E)N^{-1}$ is an irreducible transition matrix. The entries of $(\Pi^{-1} - E)N^{-1}$ are given by

$$z_{ij} \frac{\sum_{k=1}^{n} z_{kj}}{\sum_{k,l=1}^{n} z_{lk}} - \sum_{k=1}^{n} z_{kj} \qquad \text{for all } i, j = 1, \cdots, n$$

which due to condition (2.3) means that P is irreducible. From (2.2) follows that P is nonnegative. We chose π such that it is the staionary vector of P, i. e. $\pi^\top P = \pi^\top > 0$ and $\sum_{k=1}^{n} \pi_k = 1$ as well as $(\Pi^{-1} - E)N^{-1}e = 0 \Rightarrow Pe = e$.

If on the other hand M is the MFP-matrix of an irreducible transition matrix P with stationary distribution π and $N = M - \Pi^{-1}$, N^{-1} is given by equation (2.1) and the conditions (2.2) and (2.3) hold naturally. □

The conditions in Lemma 2.2 can be rewritten as

(i) $N^{-1}e > 0$ (ii) $\Pi(I - P) = (\pi e^\top - I)N^{-1} \leq \Pi$

if $\pi = \frac{N^{-1}e}{N^{-1}e}$. Let's consider the block matrix

$$\mathcal{N} := \begin{bmatrix} N & e \\ e^\top & 0 \end{bmatrix}.$$

Taking the inverse of this block matrix leads to the following expression:

$$\begin{bmatrix} N & e \\ e^\top & 0 \end{bmatrix}^{-1} = \begin{bmatrix} N^{-1} - \dfrac{N^{-1}ee^\top N^{-1}}{e^\top N^{-1}e} & \dfrac{N^{-1}e}{e^\top N^{-1}e} \\ \dfrac{e^\top N^{-1}}{e^\top N^{-1}e} & -\dfrac{1}{e^\top N^{-1}e} \end{bmatrix} = \begin{bmatrix} -\Pi(I-P) & \pi \\ e^\top N^{-1}\kappa & -\kappa \end{bmatrix} \tag{2.4}$$

where $\kappa = \frac{1}{e^\top N^{-1}e}$ is the reduced Kemeny constant. In fact, the first equality is always true as long as the inverse N^{-1} exists and $e^\top N^{-1}e = 0$.

Corollary 2.3. *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix with $diag(A) = 0$ and otherwise positive entries such that the inverse*

$$\begin{bmatrix} A & e \\ e^\top & 0 \end{bmatrix}^{-1} =: \begin{bmatrix} U & v \\ w & z \end{bmatrix} \tag{2.5}$$

exists. The matrix A is the MFP-matrix of an irreducible Markov chain on n states with stationary distribution v if the following hold:

(i) U is irreducible *(ii) $v > 0$*

(iii) $diag(v)(E-I) \geq U \geq -diag(v) > -I$.

Proof. Because of (2.12) it holds that

$$Ue = e^\top U = 0, \qquad we = e^\top v = 1. \tag{2.6}$$

Define the matrix $P := I + diag(v)^{-1}U$. From condition *(iii)* it follows that $P \geq 0$ and with (2.6) it is easy to see that $v^\top P = v^\top$ and $Pe = e$. This means that P is a transition matrix with stationary distribution $v > 0$. The irreducibility of U means that P is irreducible. By setting $\pi = v$ we see that A is the MFP-matrix N from the equations (2.4) and (2.1). □

This corollary makes it possible to decide whether a matrix is a MFP-matrix by checking whether the entries of $\mathcal{N}$ are inside certain upper and lower bounds and U is irreducible. Tough checking the irreducibility of a matrix is not easily done. Next we will see a second inverse relationship which will lead us to the theory of $\mathcal{M}$-matrices.

2.2. inverse $\mathcal{M}$-matrix problem

The term $\mathcal{M}$-matrix is not to be confused with the MFP-Matrix M. The $\mathcal{M}$-matrix was named after the mathematician Minkowski by Alexander Ostrowski (see [43]).

Definition 2.4. *Any matrix A of the form*

$$A = sI - T$$

with $s > 0, T \geq 0$ for which $s \geq \rho(T)$ is called an $\mathcal{M}$-Matrix. If $s = \rho(T)$, A is a singular $\mathcal{M}$-Matrix, otherwise it is nonsingular.

We note that the previously introduced $G = I - P$ is an example of a singular $\mathcal{M}$-matrix. The following theorem will tell us more about principal submatrices of G. Recall that a principal submatrix of a matrix A is the submatrix one gets by deleting the same rows and columns of A.

Lemma 2.5 ([3])**.** *(i) Let A be a singular, irreducible $\mathcal{M}$-matrix. Then each principal submatrix of A other than A itself is a nonsingular $\mathcal{M}$-matrix.*

(ii) Let A be a nonsingular $\mathcal{M}$-matrix. Then each principal minor of A is a nonsingular $\mathcal{M}$-matrix.

A nonsingular matrix, whose inverse is an $\mathcal{M}$-matrix will be called $\mathcal{IM}$-matrix (inverse $\mathcal{M}$-matrix). In the next chapter we will find out more about the structure of $\mathcal{M}$-matrices and their inverses, but to proof the main theorem of this chapter, we need some information about the signs of their entries. Let A be a nonsingular $\mathcal{M}$-matrix, $A = sI - T = s(I - s^{-1}T)$. Because $\rho(s^{-1}T) < 1$ we can form the inverse by a Neumann series

$$A^{-1} = \frac{1}{s}\left(I - \frac{T}{s}\right)^{-1} = \frac{1}{s}\sum_{k=0}^{\infty}\left(\frac{T}{s}\right)^{k} \geq 0$$

and see that $\mathcal{IM}$-matrices are nonnegative. This is not true for the generalized group inverse of a singular $\mathcal{M}$-matrix because we have seen that $G^{\#}e = 0$ in the previous chapter. It is obvious from the definition that the off diagonal entries of an $\mathcal{M}$-matrix are nonpositive but we can also infer that the diagonal entries of both an $\mathcal{M}$-matrix and its inverse are positive by writing

$$\sum_{k=1}^{n} a_{ik}a_{ki}^{(-)} = 1 \quad \Leftrightarrow \quad a_{ii}a_{ii}^{(-)} = 1 - \sum_{k\neq i} a_{ik}a_{ki}^{(-)} \geq 1 \quad i = 1, \ldots, n$$

Indeed we have the following characterization.

Lemma 2.6 ([29])**.** *Suppose A is a real matrix with nonpositive offdiagonal elements. Then A is a nonsingular $\mathcal{M}$-matrix if and only if $A^{-1} \geq 0$*

Using this Lemma we can define the inverse $\mathcal{M}$-matrix problem as the question under what conditions a nonnegative, nonsingular matrix is the inverse of an $\mathcal{M}$-matrix. To connect this problem to the *inverse MFP-matrix problem* we introduce the $(n-1) \times n$ matrices

$$S^{(k)} := [e_1, \cdots, e_{k-1}, -e, e_k, \cdots, e_{n-1}]$$

where the e_i are the standard base vectors in $\mathbb{R}^{n-1}$. In the following we will use the notation $A_{k,j}$ for the $(n-1)\times(n-1)$ submatrix of an $n\times n$ matrix A that we get by removing the k-th row and j-th column and A_k to indicate that only the k-th row has been deleted.

Lemma 2.7. *Suppose that P is the transition matrix of a Markov chain on n states with the MFP-matrix M and the stationary distribution vector π. Let $G = I - P$ and $\Pi = diag(\pi_1, \cdots, \pi_n)$. For $k = 1, \cdots, n$ define*

$$H^{(k)} := -S^{(k)} N (S^{(k)})^\top$$

Then

$$\Pi_{k,k} G_{k,k} H^{(k)} = I$$

Proof. First we note that

$$\Pi_{k,k} G_{k,k} e = (\pi_1 p_{1,k}, \cdots, \pi_{k-1} p_{k-1,k}, \pi_{k+1} p_{k+1,k}, \cdots, \pi_n p_{n,k1})^\top \tag{2.7}$$

which is the k-th column of $\Pi_k G_k$. It follows that

$$\Pi_{k,k} G_{k,k} S^{(k)} = (\Pi G)_k$$

Similarly using that $\Pi G N = \Pi E - I$ we see that

$$\Pi_{k,k} G_{k,k} S^{(k)} N = (\Pi G)_k N = (\Pi E - I)_k.$$

Finally, using that $S^{(k)} e = 0$ and that $S^{(k)} I_k^\top = I$ we get that

$$\Pi G S^{(k)} N S^{(k)\top} = -I \quad \Leftrightarrow \quad \Pi_{k,k} G_{k,k} H^{(k)} = I$$

□

The entries of $H^{(n)}$ are given by

$$h_{ij}^{(n)} = \begin{cases} m_{in} + m_{nj} - m_{ij} & i \neq j \\ m_{in} + m_{ni} & i = j. \end{cases}$$

The notation for $k \neq n$ can get quite confusing:

$$\begin{aligned} H^{(k)} &= -\begin{bmatrix} I & -e & 0 \\ 0 & -e & I \end{bmatrix} \begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & 0 & N_{23} \\ N_{31} & N_{32} & N_{33} \end{bmatrix} \begin{bmatrix} I & 0 \\ -e^\top & -e^\top \\ 0 & I \end{bmatrix} \\ &= -\begin{bmatrix} N_{11} - eN_{21} - N_{12}e^\top & N_{13} - eN_{23} - N_{12}e^\top \\ N_{31} - eN_{21} - N_{32}e^\top & N_{33} - eN_{23} - N_{32}e^\top \end{bmatrix} \end{aligned}$$

such that the entries are given by

$$h_{ij}^{(k)} = \begin{cases} m_{\tilde{i}k} + m_{k\tilde{j}} - m_{\tilde{i}\tilde{j}} & i \neq j \\ m_{\tilde{i}k} + m_{k\tilde{i}} & i = j \end{cases} \qquad \text{where} \qquad \tilde{i} = \begin{cases} i & i < k \\ i+1 & i \geq k. \end{cases} \tag{2.8}$$

Because this notation is hard to read and the relations between N and H is less clear, we will often only describe the case $k = n$ and write $H := H^{(n)}$. Nevertheless all the properties that hold for $H^{(n)}$ also hold for the other $H^{(k)}$.

From the matrices $H^{(k)}$ and N the transition matrix P can be recovered with the row and column sum relations as in (2.7) and similarly

$$e^\top (H^k)^{-1} = (\pi_k p_{k,1}, \cdots, \pi_k p_{k,k-1}, \pi_k p_{k,k+1}, \cdots, \pi_k p_{k,n}).$$

The stationary distribution vector π can be computed from N with Remark 1.21. The next theorem gives a relation between the inverse MFP-matrix problem and the inverse $\mathcal{M}$-matrix problem.

Theorem 2.8 ([34])**.** *Let $H \in \mathbb{R}^{(n-1)\times(n-1)}$. Then the following are equivalent:*

(i) H is nonsingular, H^{-1} is a row and column diagonally dominant $\mathcal{M}$-matrix, and

$$\mathit{trace}((I+E)H^{-1}) \leq 1$$

(ii) There exists an irreducible Markov chain on n states with a transition matrix $P \in \mathbb{R}^{n\times n}$ and a stationary distribution vector π such that

$$\Pi_{n,n} G_{n,n} H = I.$$

where $G = I - P$, $\Pi = \mathit{diag}(\pi)$.

(iii) There exists an MFP-matrix M of a Markov chain on n states, such that

$$H = -S^{(n)} N (S^{(n)})^\top.$$

where N is the matrix M with zero entries on the diagonal.

Proof. From Lemma 2.7 and the uniqueness of the inverse it is clear that (ii) and (iii) are equivalent. We will show the equivalence of (i) and (ii).

Suppose first that (ii) holds. As described in the proof of lemma 2.7, the row and column sums of $H^{-1} = \Pi_{n,n} G_{n,n}$ are given by the nonnegative vectors

$$H^{-1} e = (\pi_1 p_{1,n}, \cdots, \pi_{n-1} p_{n-1,n})^\top \tag{2.9}$$

$$e^\top H^{-1} = (\pi_n p_{n,1}, \cdots, \pi_n p_{n,n-1}) \tag{2.10}$$

$$\Rightarrow e^\top H^{-1} e = \pi_n (1 - p_{nn}). \tag{2.11}$$

Furthermore it holds, that

$$\begin{aligned} trace((I+E)H^{-1}) &= trace(H^{-1}) + trace(EH^{-1}) \\ &= \sum_{k=1}^{n-1} \pi_k(1-p_{kk}) + \sum_{k=1}^{n-1} \pi_n p_{nk} \\ &= \sum_{k=1}^{n-1} \pi_k - \sum_{k=1}^{n-1} \pi_k p_{kk} + \pi_n(1-p_{nn}) \\ &= 1 - \sum_{k=1}^{n} \pi_k p_{kk} \\ &\leq 1 \end{aligned}$$

To see that H^{-1} is an $\mathcal{M}$ matrix recall that the principal submatrices of G are nonsingular $\mathcal{M}$-matrices. From Lemma 2.6 it follows that the set of $\mathcal{M}$-matrices is closed under multiplication with positive diagonal matrices, which makes $\Pi_{n,n}G_{n,n}$ an $\mathcal{M}$-matrix as well.

Now assume, that (i) holds. The diagonal of EH^{-1} is given by $eH^{-1} \geq 0$. Because H is nonsingular $eH^{-1} \neq 0$. This means that

$$\text{trace}(H^{-1}) < \text{trace}(H^{-1}) + \text{trace}(EH^{-1}) = \text{trace}((I+E)H^{-1}) \leq 1.$$

Let $d := \{d_1, \ldots, d_{n-1}\}$ be the vector of diagonal entries of H^{-1}, which are strictly positive because H^{-1} is a nonsingular $\mathcal{M}$-matrix. Because $\text{trace}(H^{-1}) \leq 1$ we can choose a vector $\pi = \pi_1, \ldots, \pi_n$ such that $\pi_i \geq d_i, i = 1, \ldots, n-1$ and $\sum_{i=1}^{n} \pi_i = 1$. Set $\Pi := \text{diag}(\pi)$ and

$$P := I - \Pi^{-1} \begin{bmatrix} H^{-1} & -H^{-1}e \\ -e^\top H^{-1} & e^\top H^{-1} e \end{bmatrix}$$

The diagonal dominance together with the sign structure of the entries of an $\mathcal{M}$-matrix guarantees that the off diagonal entries of P are nonnegative and π was chosen such that the diagonal entries of P are nonnegative. Furthermore we have $\pi^\top P = \pi^\top$ and $Pe = e$, which makes P a transition matrix with stationary distribution π.

The matrix P is irreducible if and only if the block matrix

$$\tilde{H}^{-1} := \begin{bmatrix} H^{-1} & -H^{-1}e \\ -e^\top H^{-1} & e^\top H^{-1} e \end{bmatrix}$$

is irreducible. Because $H^{-1}e \neq 0$ and $e^\top H^{-1} \neq 0$ the block matrix is only reducible, if H^{-1} is reducible. Assume, that H^{-1} is reducible and Q a permutation matrix such that $QH^{-1}Q^\top$ is block upper triangular as in Definition 1.8. The equation

$$\begin{aligned} \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} H^{-1} & -H^{-1}e \\ -e^\top H^{-1} & e^\top H^{-1} e \end{bmatrix} \begin{bmatrix} Q^T & 0 \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} QH^{-1}Q^\top & -QH^{-1}e \\ -e^\top H^{-1}Q^\top & e^\top H^{-1} e \end{bmatrix} \\ &= \begin{bmatrix} QH^{-1}Q^\top & -QH^{-1}Q^\top e \\ -e^\top QH^{-1}Q^\top & e^\top QH^{-1}Q^\top e \end{bmatrix} \end{aligned}$$

shows that we can without loss of generality assume that H^{-1} itself is already in block upper triangular form and therefore $Q = I$. We write

$$H^{-1} = \begin{bmatrix} (H^{-1})_{1,1} & (H^{-1})_{1,2} \\ 0 & (H^{-1})_{2,2} \end{bmatrix}$$

and see that $\tilde{H}^{-1}$ is only reducible if $e^\top (H^{-1})_{1,1} = 0$, but $(H^{-1})_{1,1}$ is nonsingular which means that $\tilde{H}^{-1}$ is irreducible. □

From the last proof we can give properties of H and H^{-1} for some special cases.

Remark 2.9. *(i) If the transition matrix P is reversible i.e., $\Pi P = (\Pi P)^\top$ the matrices H and H^{-1} are symmetric.*

(ii) If the diagonal entries of the transition matrix P are all equal to zero the trace inequality becomes an equality i.e.,

$$trace((I + E)H^{-1}) = 1.$$

Such a case is not uncommon and is given for so called Embedded Markov Chains.

(iii) In [24] one can find a probabilistic interpretation of the entries of H. Let N_j^{ik} be the expected number of visits to j on a path from i to k, then

$$h_{ij} := \frac{N_j^{ik}}{\pi_j}.$$

While in the proof of Theorem 2.8 the choice of the stationary distribution was arbitrary as long as its entries were greater than the diagonal entries of $\tilde{H}^{-1}$, in practice the stationary distribution can be computed using equation (2.1) by writing

$$\begin{bmatrix} H^{-1} & -H^{-1}e \\ -e^\top H^{-1} & e^\top H^{-1} e \end{bmatrix} N = \pi e^\top - I$$

An advantage of Theorem 2.8 is that the conditions given in (i) are sometimes easier to check than the irreducibility in (ii), which can be applied to Corollary 2.3. Note that for $H := H^{(k)}$ for some $k = 1, \ldots, n$, H^{-1} is the principal submatrix of $\mathcal{N}^{-1}$ that we get by deleting the k-th and $n+1$-st rows and columns. The inequalities given in Corollary 2.3 make sure H^{-1} has the sign pattern of an $\mathcal{M}$-matrix. If its inverse that is given by $H = -S^{(k)} A (S^{(k)})^\top$ has nonnegative entries, H^{-1} is an $\mathcal{M}$ matrix and the transition matrix $I + diag(v)^{-1}U$ is irreducible because of Theorem 2.8. This means that the condition "U is irreducible" can be replaced by $H = -S^{(k)} A (S^{(k)})^\top \geq 0$. This property will be used to update Corollary 2.3.

Corollary 2.10. *Let $A \in \mathbb{R}^{n\times n}$ be a nonsingular matrix with $diag(A) = 0$ and otherwise positive entries such that the inverse*

$$\begin{bmatrix} A & e \\ e^\top & 0 \end{bmatrix}^{-1} =: \begin{bmatrix} U & v \\ w & z \end{bmatrix} \tag{2.12}$$

exists. The matrix A is the MFP-matrix of an irreducible Markov chain on n states with stationary distribution v if the following hold:

(i) $S^{(k)}A(S^{(k)})^\top \leq 0$ *(ii)* $v > 0$

(iii) $diag(v)(E-I) \geq U \geq -diag(v) > -I$.

The introduction of the $\mathcal{M}$-matrix has given us a new perspective on the inverse MFP-matrix problem. In the next chapter we will develop some inequalities for the entries of M based on both, the properties of mean first passage times and the class of $\mathcal{M}$-matrices.

Chapter 3

Inequalities for Mean First Passage Times

Before we attempt to complete a partial MFP-matrix we will develop upper and lower bounds for the missing entries and the transition probabilities. As it turns out determining entries such that the completed matrix is an MFP-matrix requires that certain inverse matrices have the correct sign pattern. Having bounds on the inverse makes the problem very hard to solve but even if a solution has been found it is not necessarily clear whether this solution is close to the original process. With the bounds presented in this chapter it will be possible to give first estimates, to narrow down the set of possible solutions, and evaluate uncertainties in the solutions.

Mean first passage times can be understood as a distance measure for a random walk on a graph that is weighted by the transition matrix P. Though this distance is not a metric because it is not symmetric, we have the following triangle inequality.

Theorem 3.1 ([26])**.** *Let M be the MFP-matrix of an irreducible Markov chain on n states. Then for $i, j, k = 1, \ldots, n$ it holds that*

$$m_{ik} + m_{kj} \geq m_{ij}.$$

Equality holds if and only if j is distinct from both i and k, and in addition, every path from state i to state k passes through state j.

Note that these triangle inequalities are basically the entries of the matrices $H^{(k)}$ from the last chapter, which shows again that $H^{(k)} \geq 0$. We illustrate this theorem with an example.

Example 3.2. *Figure 3.1 shows a transition network with three subsets (blue, red and green) with each containing three states. The probabilities to stay in a group is much higher than to leave it. Between the groups only jumps in one direction are possible which creates a directed cycle between the groups. Any path from blue to green has to pass through red. We therefore have for example $m_{36} = m_{69} = 30$ and $m_{39} = 60$. In section 5.2 we will explore this example further.*

The triangle inequality already yields upper and lower bounds for an unknown entry m_{ij}:

$$\max\{m_{ik} - m_{jk}, m_{kj} - m_{ki}, 1\} \leq m_{ij} \leq m_{ik} + m_{kj} \quad k = 1, \ldots, n$$

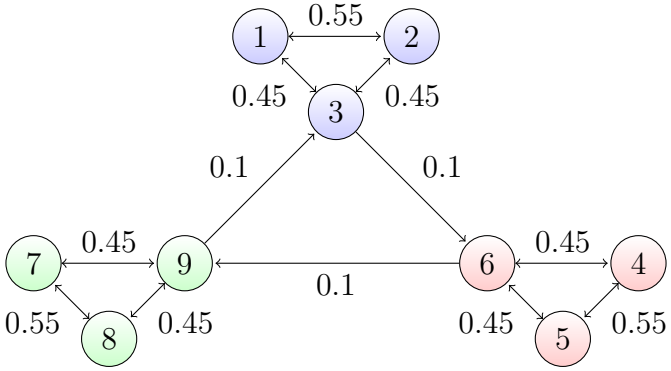


Figure 3.1: A cyclic process between three sets

Defining a digraph using a partial MFP-matrix with edges between states if their mean first passage time is known, the triangle inequality gives upper and lower bounds for each entry if the graph is strongly connected, i.e., if every state can reach any other state. In other words, the time m_{ij} can be approximated if there is a path $i_1, \dots, i_s$ such that $m_{i_1 i_2}, \dots, m_{i_{s-1} i_s}$ are known, then $m_{ij} \leq \sum_{k=1}^{s-1} m_{i_k i_{k+1}}$. In the following we will improve these bounds by using the $\mathcal{IM}$-matrices $H^{(k)}$.

Recall that a nonsingular $\mathcal{M}$-matrix A is defined by a matrix $T \geq 0$ and a value $s > \rho(T)$ such that $A = sI - T$. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of T with $|\lambda_i| < s$. It follows from T being real valued, that $\det(A) = \sum_{i=1}^{n}(s - \lambda_i) > 0$.

There are various parallels between $\mathcal{M}$ and $\mathcal{IM}$-matrices. In Theorem 2.5 it was stated that each principal submatrix of a nonsingular $\mathcal{M}$-matrix is itself a nonsingular $\mathcal{M}$-matrix. The same holds for $\mathcal{IM}$-matrices.

Lemma 3.3 ([18])**.** *Let A be a nonsingular $\mathcal{IM}$-matrix. Then each principal submatrix of A is a nonsingular $\mathcal{IM}$-matrix.*

Proof. It follows from Lemma 2.6, that for every permutation matrix Q the matrices QAQ and $QA^{-1}Q$ are $\mathcal{IM}$ and $\mathcal{M}$-matrices respectively. This means, it is enough to show that the leading principal submatrices of A are $\mathcal{IM}$-matrices.

Let $B := A^{-1}$ and using that the principal minors of B are positive we write

$$\begin{aligned} B &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \\ &= \begin{pmatrix} B_{11} - B_{12}B_{22}^{-1}B_{21} & B_{12}B_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ B_{21} & B_{22} \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ B_{21}B_{11}^{-1} & B_{22} - B_{21}B_{11}^{-1}B_{12} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ 0 & I \end{pmatrix} \end{aligned}$$

to see that $\det(B) = \det(B_{11})\det(B_{22} - B_{21}B_{11}^{-1}B_{12}) = \det(B_{22})\det(B_{11} - B_{12}B_{22}^{-1}B_{21}) > 0$

and A can therefore be written using the Schur-complement

$$\begin{aligned} A &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\ &= \begin{pmatrix} [B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} & -B_{11}^{-1}B_{12}[B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} \\ -[B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1}B_{21}B_{11}^{-1} & [B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} \end{pmatrix} \\ &= \begin{pmatrix} [B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} & -[B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1}B_{12}B_{22}^{-1} \\ -B_{22}^{-1}B_{21}[B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} & [B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} \end{pmatrix}. \end{aligned}$$

Because B_{11}, B_{22} are $\mathcal{M}$-matrices and B_{12}, B_{21} are nonpositive, $A_{11}^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}$ has only nonpositive offdiagonal entries and it follows with Lemma 2.6 that A_{12} is an $\mathcal{IM}$-matrix.The same holds for A_{22} and in fact any principal submatrix. □

In the notation of the last proof, let A_{11} be the leading principal submatrix of order $n-1$. The matrix $A_{11} - A_{12}A_{22}^{-1}A_{21}$ is then an $\mathcal{IM}$-matrix as well and for $i, j = 1, \ldots, n-1$ it holds that

$$\frac{a_{in}a_{nj}}{a_n} \leq a_{ij}.$$

As it turns out, this inequality is just a special case.

Definition 3.4. *Let $A \geq 0$ a matrix of order n with positive diagonal entries. We call A a path product (PP) matrix if for every triple $i, j, k = 1, \ldots, n$ we have*

$$\frac{a_{ik}a_{kj}}{a_{kk}} \leq a_{ij}.$$

If for $i = j$ the inequality is strict, we call A a strict path product matrix (SPP-matrix).

Lemma 3.5 ([20])**.** *Every $\mathcal{IM}$-matrix is an SPP-matrix.*

Proof. Let A be the 2×2 principal submatrix given by i and k. We know from Theorem 3.3 that A is a nonsingular $\mathcal{IM}$-matrix. and therefore

$$0 < \det(A) = a_{ii}a_{kk} - a_{ik}a_{ki}$$

which is the strict inequality needed for SPP-matrices. Now let A be the 3×3 principal submatrix given by the distinct coordinates i, j, k

$$A = \begin{pmatrix} a_{ii} & a_{ik} & a_{ij} \\ a_{ki} & a_{kk} & a_{kj} \\ a_{ji} & a_{jk} & a_{jj} \end{pmatrix}$$

The adjugate $adj(A)$ of a matrix A is given via

$$adj(A)_{ij} = (-1)^{i+j} \det(A\{j, i\})$$

and if A is nonsingular its inverse is given via

$$A^{-1} = \frac{1}{\det(A)} adj(A).$$

Applying this formula to the $\mathcal{M}$-matrix $C := A^{-1}$ and taking advantage of its sign pattern gives

$$\begin{aligned} & c_{31} = \frac{a_{ik}a_{kj} - a_{ij}a_{kk}}{\det(A)} \leq 0 \\ \Rightarrow \quad & a_{ik}a_{kj} - a_{ij}a_{kk} \leq 0 \\ \Leftrightarrow \quad & \frac{a_{ik}a_{kj}}{a_{kk}} \leq a_{ij} \end{aligned}$$

□

The SPP-property gives lower bounds on the inverse of an $\mathcal{IM}$-matrix.

Theorem 3.6. *For an $\mathcal{M}$-matrix $A = \{a_{ij}\}_{i,j=1,\dots,n}$ and its inverse $A^{-1} := B = \{b_{ij}\}_{i,j=1,\dots,n}$ the following inequalities hold:*

(i) $a_{ij} \geq -\dfrac{b_{ij}}{b_{ii}b_{jj} - b_{ij}b_{ji}}$

(ii) $a_{ii} \geq \dfrac{b_{jj}}{b_{ii}b_{jj} - b_{ij}b_{ji}}$ *for all* $j \neq i$

Proof. Inequality (i) and (ii) are a result of applying the SPP-property in the following way:

$$\begin{aligned} 0 = \sum_{k=1}^{n} a_{ik}b_{kj} &= a_{ij}b_{jj}b_{ii} + a_{ii}b_{ij}b_{ii} + \sum_{k\neq i,j} a_{ik}b_{kj}b_{ii} \\ &\leq a_{ij}b_{jj}b_{ii} + a_{ii}b_{ij}b_{ii} + \sum_{k\neq i,j} a_{ik}b_{ki}b_{ij} \\ &= a_{ij}b_{jj}b_{ii} + a_{ii}b_{ij}b_{ii} + (1 - a_{ij}b_{ji} - a_{ii}b_{ii})b_{ij} \\ &= b_{ij} + a_{ij}(b_{jj}b_{ii} - b_{ij}b_{ji}) \\ \Rightarrow a_{ij} &\geq -\frac{b_{ij}}{b_{ii}b_{jj} - b_{ij}b_{ji}} \end{aligned}$$

$$\begin{aligned} 0 = \sum_{k=1}^{n} a_{ik}b_{kj} &= a_{ij}b_{jj}b_{ji} + a_{ii}b_{ij}b_{ji} + \sum_{k\neq i,j} a_{ik}b_{kj}b_{ji} \\ &\geq a_{ij}b_{jj}b_{ji} + a_{ii}b_{ij}b_{ji} + \sum_{k\neq i,j} a_{ik}b_{ki}b_{jj} \\ &= a_{ij}b_{jj}b_{ji} + a_{ii}b_{ij}b_{ji} + (1 - a_{ij}b_{ji} - a_{ii}b_{ii})b_{jj} \\ &= b_{jj} - a_{ii}(b_{jj}b_{ii} - b_{ij}b_{ji}) \\ \Rightarrow a_{ii} &\geq \frac{b_{jj}}{b_{ii}b_{jj} - b_{ij}b_{ji}} \end{aligned}$$

□

In particular the inequality (i) implies, that $b_{ij} = 0 \Rightarrow a_{ij} = 0$. We should note that the SPP-property for $\mathcal{IM}$-matrices and Theorem 3.6 can also be seen as special cases of more general results. The proof of Lemma 3.5 indicates that similar, but more complex inequalities can be found for higher dimensional submatrices and applying the following lemma for $|\alpha| = 2$ results in Theorem 3.6.

Lemma 3.7 ([21]). *Let A be an $\mathcal{M}$- or $\mathcal{IM}$-matrix and $\emptyset \neq \alpha \subset \{1, \ldots, n\}$. If the notation $A[\alpha]$ refers to the submatrix of A with respect to the set α, then*

(i) $(A^{-1}[\alpha])^{-1} \leq A[\alpha]$

(ii) $A[\alpha]^{-1} \leq A^{-1}[\alpha]$

Let A be an $\mathcal{M}$-matrix, B its inverse and $\alpha = \{i, j\} \subset \{1, \ldots, n\}$, $i \neq j$. The inequalities from Theorem 3.6 follow from

$$A[\alpha]^{-1} := \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}^{-1} = \begin{bmatrix} a_{jj} & -a_{ij} \\ -a_{ji} & a_{ii} \end{bmatrix} \frac{1}{a_{ii}a_{jj} - a_{ji}a_{ij}} \leq \begin{bmatrix} b_{ii} & b_{ij} \\ b_{ji} & b_{jj} \end{bmatrix} = A^{-1}[\alpha].$$

Plugging the entries of the $\mathcal{IM}$-matrices $H^{(k)}$ into the SPP-property improves the triangle inequalities from Theorem 3.1. For that we introduce the notation

$$q_{ij}^k = \begin{cases} m_{ik} + m_{kj} - m_{ij} & i \neq j \\ m_{ik} + m_{ki} & i = j. \end{cases} \qquad k \neq i, j.$$

This notation makes it easier to see the relationship between $H^{(k)}$ and the mean first passage times.

Theorem 3.8. *Let M be the MFP-matrix of an irreducible Markov chain on n states and a_{ij}^k given as above. For fixed states $k, s = 1, \ldots, n$, $s \neq k$ and every $i, j \neq k$, the following inequalities hold:*

$$m_{ij} \leq m_{ik} + m_{kj} - \frac{q_{is}^k q_{sj}^k}{q_{ss}^k} \tag{3.1}$$

$$m_{ij} \geq m_{ik} + m_{kj} - \frac{q_{ii}^k q_{sj}^k}{q_{si}^k} \qquad \text{if} \qquad q_{si}^k \neq 0 \tag{3.2}$$

$$m_{ij} \geq m_{ik} + m_{kj} - \frac{q_{jj}^k q_{is}^k}{q_{js}^k} \qquad \text{if} \qquad q_{js}^k \neq 0 \tag{3.3}$$

More generally, for distinct i, j, k, s it holds that

$$m_{ij} \leq m_{ik} + m_{kj} - \frac{(m_{ik} + m_{ks} - m_{is})(m_{sk} + m_{kj} - m_{sj})}{(m_{sk} + m_{ks})} \tag{3.4}$$

The inequalities are just a result of reordering the SPP-property for the matrices $H^{(k)}$. The formulation implies, that the k-th and s-th rows and columns are known.

Remark 3.9. *(i) The upper bound (3.1) is clearly a strict improvement to the triangle inequality. To see that the lower bounds (3.2) and (3.3) improve the estimates assume that $q^k_{js} \neq 0$ and that $m_{ij} \geq m_{is} - m_{js} \geq m_{ik} - m_{jk}$.*

$$\begin{aligned} m_{ij} &\geq m_{ik} + m_{kj} - \frac{q^k_{jj} q^k_{is}}{q^k_{js}} \pm q^k_{jj} \\ &= m_{ik} - m_{jk} + \frac{q^k_{jj}}{q^k_{js}}(q^k_{js} - q^k_{is}) \pm (m_{is} - m_{js}) \\ &= m_{is} - m_{js} + (m_{is} - m_{js} - (m_{ik} - m_{jk}))\left(\frac{q^k_{jj}}{q^k_{js}} - 1\right) \\ &= m_{is} - m_{js} + (m_{is} - m_{js} - (m_{ik} - m_{jk}))\frac{q^j_{ns}}{q^k_{js}} \\ &\geq m_{is} - m_{js} \end{aligned}$$

An analogues proof works for the inequality (ii) and if $m_{ij} \geq m_{ik} - m_{jk} \geq m_{is} - m_{js}$ we only need to swap k and s.

(ii) Note that

$$q^k_{ij} q^k_{ss} - q^k_{is} q^k_{sj} = q^s_{ij} q^s_{kk} - q^s_{ik} q^s_{kj}.$$

This means that we can swap s and k in (3.1).

(iii) For (3.1) to (3.3) we did not assume, that $i \neq j$. In particular

$$\frac{1}{\pi_i} = m_{ii} \leq q^k_{ii} - \frac{q^k_{is} q^k_{si}}{q^k_{ss}}$$

follows from Corollary 3.11.

Theorem 3.1 said, that $m_{ij} = m_{ik} + m_{kj}$ if and only if i can reach j only by passing k. With the SPP-property we can give conditions under which this is the case.

Corollary 3.10. *Let M be the MFP-matrix of an irreducible Markov chain on n states. Then for distinct states $i, j, k = 1, \ldots, n$ we have*

$$m_{ik} + m_{kj} = m_{ij}.$$

if one of the following two conditions hold:

(i) There is a state s such that

$$m_{sk} + m_{kj} = m_{sj} \qquad \wedge \qquad m_{sk} + m_{ki} > m_{si}$$

(ii) There is a state s such that

$$m_{ik} + m_{ks} = m_{is} \qquad \wedge \qquad m_{jk} + m_{ks} > m_{js}$$

Proof. We will again use the notation q_{ij}^k:

$$q_{si}^k q_{ij}^k \leq q_{ii}^k q_{sj}^k$$

and

$$q_{ij}^k q_{js}^k \leq q_{jj}^k q_{is}^k$$

With $q_{ii}^k, q_{jj}^k > 0$, the left hand sides are zero if $q_{is}^k = q_{sj}^k = 0$ i.e., $q_{sj}^k = 0$ and $q_{si}^k \neq 0$ implies $q_{ij}^k = 0$ and analogously for the second equation. □

The inequalities of Theorem 3.6 applied to $(H^{(k)})^{-1} = (\Pi(I - P))_{k,k}$ result in the following corollary.

Corollary 3.11. *Let $P = \{p_{ij}\}_{i,j=1,\ldots,n}$ be an irreducible transition matrix with stationary distribution $\pi = \{\pi_i\}_{i=1,\ldots,n}$ and MFP-matrix $M = \{m_{ij}\}_{i,j=1,\ldots,n}$. For $H = H^{(n)} = P = \{p_{ij}\}_{i,j=1,\ldots,n-1}$ and $i < n$ it holds that*

$$\pi_i \geq \pi_i(1 - p_{ii}) \geq \frac{h_{jj}}{h_{ii}h_{jj} - h_{ij}h_{ji}} \geq \frac{h_{ij}}{h_{ii}h_{jj} - h_{ij}h_{ji}} \geq \pi_i p_{ij}$$

More generally, for distinct coordinates k, i, j the following inequalities hold:

$$\pi_i(1 - p_{ii}) \geq \frac{m_{jk} + m_{kj}}{(m_{ik} + m_{ki})(m_{jk} + m_{kj}) - (m_{ik} + m_{kj} - m_{ij})(m_{jk} + m_{ki} - m_{ji})}$$

$$\pi_i p_{ij} \leq \frac{m_{ik} + m_{kj} - m_{ij}}{(m_{ik} + m_{ki})(m_{jk} + m_{kj}) - (m_{ik} + m_{kj} - m_{ij})(m_{jk} + m_{ki} - m_{ji})}$$

These inequalities give a lower bound on H^{-1}. An upper bound is given by the identity matrix, because $\pi_i(1 - p_{ii}) \leq 1$ and $\pi_i p_{ij} > 0$. These inequalities guarantee the same sign pattern as an $\mathcal{M}$-matrix, but only those matrices inside these intervals, whose inverses are non-negative are in fact $\mathcal{M}$-matrices. In the same way, only those matrices H inside the intervals specified by equations (3.1)-(3.3) are $\mathcal{IM}$-matrices whose inverses have non-positive off-diagonal entries. In other words, we have two matrices whose entries are only given by intervals, but we know that these matrices must be inverse to each other. In the next chapter it is shown how the solutions of linear systems given by intervals can be approached.

Chapter 4

Interval Arithmetic

In the last chapter lower and upper bounds for the matrices $H^{(k)}, (H^{(k)})^{-1}$ and M, have been developed. This means that for every matrix entry we know an interval that includes the correct value. Interval arithmetic is a tool to formalize computations using intervals instead of numbers. In this chapter we want to give a short introduction to interval arithmetic and it will be shown how the estimates can be improved. For the definitions and notation we will follow [33].

The set of all real intervals will be denoted by $\mathbb{IR}$. Every interval $x \in \mathbb{IR}$ is defined by numbers $\underline{x}, \overline{x} \in \mathbb{R}$ such that

$$x := [\underline{x}, \overline{x}] := \{\tilde{x} \in \mathbb{R} | \underline{x} \leq \tilde{x} \leq \overline{x}\}$$

where $\underline{x} = \inf(x), \overline{x} := \sup(x)$. If $\underline{x} = \overline{x}$ it follows that $x \in \mathbb{R}$. Analogously we can define an $X \in \mathbb{IR}^{n \times m}$ if there are matrices $\underline{X}, \overline{X} \in \mathbb{R}^{n \times m}$ such that X is given by

$$X := [\underline{X}, \overline{X}] := \{\tilde{X} \in \mathbb{R}^{n \times m} | \underline{x}_{ij} \leq \tilde{x}_{ij} \leq \overline{x_{ij}}, i = 1, \ldots, n; j = 1, \ldots, m\}.$$

Instead of the infima and suprema an interval can also be determined by its midpoint and radius

$$\check{A} := mid(A) := 0.5(\overline{A} + \underline{A}) \qquad rad(A) := 0.5(\overline{A} - \underline{A})$$

For a bounded set $S \subset \mathbb{R}^{m \times n}$ the smallest interval that contains S is called the hull of S and is denoted by $\Box S$ such that

$$\Box S = [\inf(S), \sup(S)].$$

Intervals can be compared by

$$x \geq y \Leftrightarrow \underline{x} \geq \overline{y} \qquad\qquad x \leq y \Leftrightarrow \overline{x} \leq \underline{y}.$$

The basic arithmetic rules for intervals are quite intuitive. Addition and subtraction are given by

$$x + y = [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$$
$$x - y = [\underline{x} - \overline{y}, \overline{x} + \underline{y}]$$

xy	$x \geq 0$	$0 \in x$	$x \leq 0$
$x \geq 0$	$[\underline{x}\underline{y}, \overline{x}\overline{y}]$	$[\overline{x}\underline{y}, \overline{x}\overline{y}]$	$[\overline{x}\underline{y}, \underline{x}\overline{y}]$
$0 \in x$	$[\underline{x}\overline{y}, \overline{x}\overline{y}]$	$[\min\{\underline{x}\overline{y}, \overline{x}\underline{y}\}, \max\{\underline{x}\underline{y}, \overline{x}\overline{y}\}]$	$[\overline{x}\underline{y}, \underline{x}\underline{y}]$
$x \leq 0$	$[\underline{x}\overline{y}, \overline{x}\underline{y}]$	$[\underline{x}\overline{y}, \underline{x}\underline{y}]$	$[\overline{x}\overline{y}, \underline{x}\underline{y}]$

Table 4.1: multiplication of intervals

x/y	$y > 0$	$y < 0$
$x \geq 0$	$[\underline{x}/\overline{y}, \overline{x}/\underline{y}]$	$[\overline{x}/\overline{y}, \underline{x}/\underline{y}]$
$0 \in x$	$[\underline{x}/\underline{y}, \overline{x}/\underline{y}]$	$[\overline{x}/\overline{y}, \underline{x}/\overline{y}]$
$x \leq 0$	$[\underline{x}/\underline{y}, \overline{x}/\overline{y}]$	$[\overline{x}/\underline{y}, \underline{x}/\overline{y}]$

Table 4.2: division of intervals

while multiplication and division are summarized in table 4.1 and 4.2. More complex operations like matrix products inherit these rules.

Applying the lower and upper bounds from chapter 3 to the inverse problems in chapter 2 means that the *MFP-matrix completion problem* can be stated in the form "for interval matrices A, B find $X \in A, Y \in B$ such that $XY = I$". The columnwise computation of the intervals for the inverse matrix then takes the form

$$Ax = b \qquad A \in \mathbb{IR}^{m \times n}; b \in \mathbb{IR}^m. \tag{4.1}$$

where b is one of the standard basis vectors. Using equations like $N\pi = \kappa e$ can be used to compute intervals for the stationary distribution. The solution set for (4.1) is defined by

$$\Sigma(A, b) := \{\tilde{x} \in \mathbb{R}^n | \tilde{A}\tilde{x} = \tilde{b}, \quad \tilde{A} \in A, \tilde{b} \in b\}.$$

Theorem 4.1. *Let $A \in \mathbb{IR}^{m \times n}$, $b \in \mathbb{IR}^m$. Then*

$$\Sigma(A, b) = \{\tilde{x} \in \mathbb{R}^n | A\tilde{x} \cap b \neq 0\} = \{\tilde{x} \in \mathbb{R}^n | 0 \in A\tilde{x} - b\}$$

Corollary 4.2. *Let $A \in \mathbb{IR}^{m \times n}$, $b \in \mathbb{IR}^m$. Then*

$$\tilde{x} \in \Sigma(A, b) \Leftrightarrow |\check{A}\tilde{x} - \check{b}| \leq rad(A)|\tilde{x}| + rad(b)$$

The inequality in the last corollary implies the existence of a diagonal matrix D, such that $D\tilde{x} = |\tilde{x}|$ and

$$(\check{A} - rad(A)D)\tilde{x} \leq \overline{b} \qquad\qquad (\check{A} + rad(A)D)\tilde{x} \geq \underline{b}.$$

In general only the hull $\square\Sigma(A,b)$ or an enclosure of it can be computed. There are classes of matrices for which these methods work better than for others.

Definition 4.3. *Let $A \in \mathbb{IR}^{n\times n}$ and $\langle A\rangle$ be given by*

$$\langle A\rangle_{ij} := \begin{cases} \min(|\tilde{A}_{ii}| \in A_{ii}) & i = j \\ -\max(|\tilde{A}_{ij}| \in A_{ij}) & i \neq j \end{cases}$$

then

(i) A is called regular, if every $\tilde{A} \in A$ is nonsingular.

(ii) A is called strongly regular, if $\check{A}^{-1}A$ is regular.

(iii) A is called an H-matrix if $\langle A\rangle$ is an $\mathcal{M}$-matrix.

Lemma 4.4. *Let $A \in \mathbb{IR}^{n\times n}$ and $\check{A}$ nonsingular. Then the following are equivalent:*

(i) A is called strongly regular.

(ii) $\rho(|\check{A}^{-1}|rad(A)) < 1$.

(iii) $\check{A}^{-1}A$ is an H-matrix.

Standard linear equations are typically solved using the LU-decomposition followed by forward and backward substitution. In some cases this method is also applicable to linear interval equations, though the relations are given by

$$A \subseteq LU, \qquad Ly \supseteq b, \qquad Ux \supseteq y.$$

If an LU-decomposition exists, A is regular and if x^G is the solution of the forward and backward substitution, it holds that $\Sigma(A,b) \subseteq x^G$.

Theorem 4.5. *Let $A \in \mathbb{IR}^{n\times n}$ be an H-matrix. Then A has a LU-decomposition (L,U). If $\langle A\rangle = \underline{A}$ then $\langle A\rangle = \langle L\rangle\langle U\rangle$*

Even if A is not strongly regular, the matrix interval $\check{A}^{-1}A$ should be close to the identity matrix I, which is why $\check{A}$ is used for preconditioning.

If, either from some general condition or as result of a method like Gauss elimination, there is a known enclosure $x^0 \supseteq \Sigma(A,b)$, an iterative method like interval Gauss-Seidel might be used to tighten the bounds via

$$x_i^{j+1} := \frac{b_i - \sum_{k<i} A_{ik}x_k^{j+1} - \sum_{k>i} A_{ik}x_k^j}{A_{ii}} \cap x_i^j \qquad i = 1,\ldots,n.$$

In [33] several special cases to compute the inverse interval matrix are discussed, but it should be noted that it was shown in [4] that it is NP-hard to compute the exact bounds on the elements of an inverse interval matrix.

It cannot be guaranteed that any of the interval matrices from the previous chapters are strongly regular or H-matrices, though numerical experiments have shown that interval arithmetics yield improved bounds. Implementations of interval arithmetics for MATLAB/Octave are available in the software package INTLAB (see [39]).

Chapter 5

Completing a Partial MFP-Matrix

The matrix completion problem received broader attention in the last years due to the "Netflix Prize", which was a competition to improve a system of movie recommendations based on a matrix that consisted of movie ratings given by users (see [2]). A key assumption was that the matrix in question had a low rank, because users with similar taste would give similar ratings. A common assumption for molecular systems with, say, m metastabilities is that the spectrum of the transition matrix shows a spectral gap after the m-th eigenvalue. The spectral gap could be used as a motivation for a low rank assumption on N, but to retrieve the transition matrix P we need N and H to be nonsingular. Moreover, the inequalities developed so far already give estimates on the entries, that can be used to analyze the system. We will focus on the problem of finding a real MFP-matrix.

Another approach to matrix completion is determinant maximization [28], which has been applied to the *$\mathcal{IM}$-matrix completion problem* in the past. In [19] and [12] conditions are given under which a partial matrix can be completed to an $\mathcal{IM}$-matrix. The authors show that if a partial matrix can be completed to an $\mathcal{IM}$-matrix a solution is given by replacing the unknown entries with zeros. Johnson and Smith show in [19] even that this solution is maximizing the determinant. Setting missing entries of the $\mathcal{IM}$-matrix H to zero would in most cases contradict the lower bounds, that are already known, so these approaches don't seem to find the solutions we are looking for.

Instead, we will use a point of view from interval arithmetics. For the MFP-matrix N we again define the $(n+1) \times (n+1)$ dimensional matrices

$$\mathcal{N} := \begin{bmatrix} N & e \\ e^\top & 0 \end{bmatrix}, \qquad \mathcal{N}^{-1} = \begin{bmatrix} -\Pi(I-P) & \pi \\ e^\top N^{-1}\kappa & -\kappa \end{bmatrix}. \tag{5.1}$$

and $H := H^{(k)}$ for some $k = 1, \ldots, n$. Recall that $H = S^{(k)} N (S^{(k)})^\top$ and H^{-1} is the principal submatrix of $\mathcal{N}^{-1}$ that we get by deleting the k-th and $n+1$-st row and column. Let $X_{\mathcal{N}}$, $Y_{\mathcal{N}^{-1}}$, X_H, $Y_{H^{-1}}$ be the interval matrices such that $\mathcal{N} \in X_{\mathcal{N}}$, $\mathcal{N}^{-1} \in Y_{\mathcal{N}^{-1}}$, $H \in X_H$, $H^{-1} \in Y_{H^{-1}}$. First the two problems that we developed to solve the *MFP-completion problem* will be reformulated. Following the ideas from section 2.1 and Corollary 2.10 the problem of finding the matrices $\mathcal{N}$ and $\mathcal{N}^{-1}$ can be formulated as:

Problem 1: *Let the matrices* $X,\ Y \in \mathbb{R}^{(n+1)\times(n+1)}$ *be given by*

$$X := \begin{bmatrix} X_{11} & e \\ e^\top & 0 \end{bmatrix}, \qquad Y := \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$$

where the blocks have the same dimensions as the blocks of $\mathcal{N}$, $\mathcal{N}^{-1}$ *in (5.1). Then the matrices* X, Y *define a solution to the* MFP-matrix completion problem *if they meet the conditions*

(1.i) $XY = I$

(1.ii) $X \in X_{\mathcal{N}}, Y \in Y_{\mathcal{N}^{-1}}$

(1.iii) $Y_{12} \geq -\mathrm{diag}(Y_{11})$

(1.iv) $S^{(k)} X_{11} (S^{(k)})^\top \leq 0.$

Similarly, following section 2.2 the problem of finding the matrices H and H^{-1} can be formulated as:

Problem 2: *The matrices* $X,\ Y \in \mathbb{R}^{(n-1)\times(n-1)}$ *define a solution to the* MFP-matrix completion problem *if they meet the conditions*

(2.i) $XY = I$

(2.ii) $X \in X_H, Y \in Y_{H^{-1}}$

(2.iii) $Ye \geq 0,\ e^\top Y \geq 0$

(2.iv) $\mathrm{trace}((I+E)Y) \leq 1.$

The main problem is the equation $XY = I$, because for both problems all remaining conditions are linear. A condition for a vector x is linear if it has the form $Ax \leq b$ for a matrix A and a vector b. To see that the conditions (ii)-(iv) in Problem 1 and Problem 2 can be formulated as linear conditions we introduce the following notation.

Definition 5.1 ([13])**.** *For arbitrary matrices* $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{s\times t}$ *the Kronecker product* $C := A \otimes B$ *is defined by the matrix*

$$C := \begin{bmatrix} a_{11}B & \cdots & a_{1m}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nm}B \end{bmatrix}.$$

For any matrix $A \in \mathbb{R}^{n\times m}$ *the vectorization* $vec(A) : \mathbb{R}^{n\times m} \to \mathbb{R}^{nm}$ *is the column vector obtained by stacking the columns of* A *on top of another.*

The following lemma shows how matrix equations can be formulated as linear equations.

Lemma 5.2 ([13])**.** *For matrices* $A \in \mathbb{R}^{n\times m}$, $X \in \mathbb{R}^{m\times s}$ $B \in \mathbb{R}^{s\times t}$, *and* $C \in \mathbb{R}^{n\times t}$ *it holds that*

$$AXB = C \qquad \Leftrightarrow \qquad (B^\top \otimes A)vec(X) = vec(C).$$

There is a unique solution X *to this matrix equation if and only if* A *and* B *are nonsingular.*

How can it be achieved that $XY = I$ under the given conditions? In the following it will be shown that the MFP-problem might be solved using a difference of convex functions programming (d.c. programming) approach.

5.1. Formulation as a D.C. Programming Problem

First some facts about convex and concave functions will be stated.

Definition 5.3 ([14]). *A set C is called convex if for all $x, y \in C$ and every $\alpha \in (0,1)$ it holds that $\alpha x + (1-\alpha)y \in C$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if for every $\alpha \in (0,1)$ it holds that*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

Similarly f is called concave if for every $\alpha \in (0,1)$ it holds that

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y).$$

Lemma 5.4 ([14]). *Let f and g be convex functions, then it holds that*

(i) $-f$ *is a concave function*

(ii) $f + g$ *and* $\max(f, g)$ *are convex functions*

(iii) if f is differentiable and ∇f is its gradient, then

$$f(x) \geq f(y) + (x-y)^\top \nabla f(y)$$

(iv) the epigraph of f

$$epi(f) := \{(x,t) | t \in \mathbb{R}, f(x) \leq t\}$$

is a convex set.

Definition 5.5 ([14]). *A real-valued function f defined on $\mathbb{R}^n$ is called* d.c. *if, for all $x \in \mathbb{R}^n$, f can be expressed in the form*

$$f(x) = p(x) - q(x)$$

where p, q are convex functions. A global optimization problem is called a d.c. *programming problem or a* d.c *program if it has the form*

$$\begin{array}{ll} \min & f(x) \\ s.t. & x \in C \\ & g_j(x) \leq 0 \qquad j = 1, \ldots, m \end{array}$$

where C is a closed convex subset of $\mathbb{R}^n$ and all functions f and g_j are d.c.

We will now show that the condition $XY = I$ can be achieved by minimizing a *d.c.* function.

Lemma 5.6. *The function*

$$F : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$$
$$F(X,Y) = \|XY - I\|^2,$$

where $\|\cdot\|$ *is the Frobenius norm, is a* d.c. *function.*

Proof. Let $x_1, \ldots, x_n \in \mathbb{R}^n$ be the rows of the matrix X, $y_1, \ldots, y_n \in \mathbb{R}^n$ the columns of the matrix Y and

$$\delta_{ij} = \begin{cases} 1 & if \quad i = j \\ 0 & else. \end{cases}$$

Then F can be written as

$$\begin{aligned} F(X,Y) = \|XY - I\|^2 &= \sum_{i,j=1}^{n} (x_i^\top y_j - \delta_{i,j})^2 \\ &= \sum_{i,j=1}^{n} \left[\frac{1}{4}\|x_i + y_j\|^2 - \frac{1}{4}\|x_i - y_j\|^2 - \delta_{i,j}\right]^2 \\ &= \sum_{i,j=1}^{n} 2\left[\frac{1}{16}\|x_i + y_j\|^4 + \left(\frac{1}{4}\|x_i - y_j\|^2 + \delta_{ij}\right)^2\right] \\ &\quad - \left[\frac{1}{4}\|x_i + y_j\|^4 + \frac{1}{4}\|x_i - y_j\|^2 + \delta_{ij}\right]^2 \\ &=: p(X,Y) - q(X,Y) \end{aligned}$$

Because $(x,y) \to \|x+y\|^2$ is convex and the square of a positive convex function is convex it follows that p and q are convex and F is a *d.c.* function. □

The convex function p and q can be simplified further by

$$\begin{aligned} q(X,Y) &= \sum_{i,j=1}^{n} \left[\frac{1}{4}\left(\|x_i + y_j\|^2 + \|x_i - y_j\|^2\right) + \delta_{ij}\right]^2 \\ &= \sum_{i,j=1}^{n} \left[\frac{1}{4}\left(2\|x_i\|^2 + 2\|y_j\|^2\right) + \delta_{ij}\right]^2 \\ &= \sum_{i,j=1}^{n} \frac{1}{4}\left(\|x_i\|^2 + \|y_j\|^2\right)^2 + \sum_{i=1}^{n} \|x_i\|^2 + \|y_i\|^2 + n \\ &= \sum_{i,j=1}^{n} \frac{1}{4}\left(\|x_i\|^2 + \|y_j\|^2\right)^2 + \|X\|^2 + \|Y\|^2 + n. \end{aligned}$$

and

$$\begin{aligned} p(X,Y) &= \sum_{i,j=1}^{n} 2\left[\frac{1}{16}\|x_i+y_j\|^4 + \left(\frac{1}{4}\|x_i-y_j\|^2+\delta_{ij}\right)^2\right] \\ &= \sum_{i,j=1}^{n} \frac{1}{8}\left(\|x_i+y_j\|^4 + \|x_i-y_j\|^4\right) + \sum_{i=1}^{n}\|x_i-y_i\|^2 + 2n \\ &= \sum_{i,j=1}^{n} \frac{1}{8}\left(\|x_i+y_j\|^4 + \|x_i-y_j\|^4\right) + \|X-Y\|^2 + 2n \end{aligned}$$

As described in [14] the *d.c.* program can be formulated as a concave program over a convex set using an additional variable $t \in \mathbb{R}$:

$$\begin{aligned} \min \quad & t - q(X,Y) && (5.2)\\ s.t. \quad & p(X,Y) \le t && (5.3)\\ & (X,Y,t) \in \mathcal{C} && (5.4) \end{aligned}$$

Because p is convex the set of all X, Y, t satisfying (5.3) and (5.4) is convex as well. It is a standard result of global optimization, that a concave function over a convex set attains its minimum on the boundary of that set. In the above problem this property is very intuitive, because the variable t can be decreased until $p(X,Y) = t$ and the minimum is attained if $q(X,Y) = t$.

The set $\mathcal{C}$ of feasible solutions is of the form

$$\mathcal{C} = \{(x,t) \in \mathbb{R}^{2n^2} \times \mathbb{R} \mid \underline{x} \le x \le \overline{x},\ \underline{t} \le t \le \overline{t},\ \mathcal{A}x \le b\},$$

where $\mathcal{A}$ and b describe the linear conditions stated in the beginning of this chapter and $x = \operatorname{vec}\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right)$. We can give two additional types of linear conditions. Because p and q are convex, it holds for any fixed $x_0 \in \mathbb{R}^{2n^2}$ by Lemma 5.4 that

$$\begin{aligned} t \ge p(x) \ge p(x_0) + (x-x_0)^\top \nabla p(x_0) \ &\Leftrightarrow\ \begin{bmatrix} x^\top & t \end{bmatrix} \begin{bmatrix} \nabla p(x_0) \\ -1 \end{bmatrix} \le x_0^\top \nabla p(x_0) - p(x_0) \\ t \ge q(x) \ge q(x_0) + (x-x_0)^\top \nabla q(x_0) \ &\Leftrightarrow\ \begin{bmatrix} x^\top & t \end{bmatrix} \begin{bmatrix} \nabla q(x_0) \\ -1 \end{bmatrix} \le x_0^\top \nabla q(x_0) - q(x_0). \end{aligned}$$

From the chapter on interval arithmetic Corollary 4.2 can be utilized in the form

$$\begin{aligned} (\check{X}_H - rad(X_H)D_i)y_i \le e_i \qquad & (\check{X}_H + rad(X_H)D_i)y_i \ge e_i \\ (\check{X}_\mathcal{N} - rad(X_\mathcal{N})D_i)y_i \le e_i \qquad & (\check{X}_\mathcal{N} + rad(X_\mathcal{N})D_i)y_i \ge e_i \end{aligned} \tag{5.5}$$

where the y_i are again the columns of Y respectively and D_i is the diagonal matrix such that $|y_i| = D_i y_i$. The matrix D_i is known, because all the signs of the entries of H^{-1} and $\mathcal{N}^{-1}$ are known except those for $e^\top N^{-1}$ but the last row and column of $\mathcal{N}$ are known.

Algorithm 1 Alternating Minimization

Input: initial matrix X_0; feasible set $\mathcal{C}$; tolerance tol; maximal number of iterations K
Output: matrices X, Y
for $k = 1$ **to** K **do**
  $Y_k := \min_Y \quad \|X_{k-1}Y - I\|$ s.t. $[X_{k-1}, Y] \in \mathcal{C}$
  $X_k := \min_X \quad \|XY_k - I\|$ s.t. $[X, Y_k] \in \mathcal{C}$
  if $\|X_kY_k - I\| < tol$ **then**
    break
  end if
end for

This means that for $X_{\mathcal{N}}$ the last row and column of $rad(X_{\mathcal{N}})$ are zero so that the signs of $e^\top N^{-1}$ do not matter. Because H and $\mathcal{N}$ both have only nonnegative entries, it also follows from Corollary 4.2 that

$$\begin{array}{lll} \underline{Y_{H^{-1}}}X \leq I & \overline{Y_{H^{-1}}}X \geq I & \text{for } X \in X_H \\ \underline{Y_{\mathcal{N}^{-1}}}X \leq I & \overline{Y_{\mathcal{N}^{-1}}}X \geq I & \text{for } X \in X_{\mathcal{N}}. \end{array} \tag{5.6}$$

The inequalities (5.5) and (5.6) further narrow down the feasible set $\mathcal{C}$.

The *d.c.* program (5.2)-(5.4) can be solved using so called *Branch-and-Bound*-Algorithms (see [14], chapter 4.6). In these methods the set of feasible solutions is relaxed and partitioned into subsets, in which a solution can be found easier. Each subset, whose optimal solution is worse than the best solution that has been found up to that point, will be cut from the set of feasible solutions. For this thesis such an algorithm has not been implemented. This is partly due to a lack of time, but also because the focus of this thesis was to create a theoretical basis with which the cost of simulations for sampling mean first passage times can be reduced and not the algorithmic treatment of global optimization problems. Instead, a linearization has been formulated and applied.

5.2. Linearization and Numerical Examples

As a preprocessing step we solved for each variable easier linear programs of the form

$$\begin{array}{rl} \min / \max & x_i \\ s.t. & x \in \mathcal{C} \end{array} \tag{5.7}$$

to find lower and upper bounds that at the same time meet the linear conditions specified by the feasible set $\mathcal{C}$.

In order to compute a suboptimal solution, a simplified problem has been solved. For a fixed matrix X_0 the objective function $F(Y) = \|X_0Y - I\|^2$ can be written as a linear

program via

$$\begin{aligned} \min \quad & \varepsilon && (5.8)\\ s.t. \quad & X_0Y - \varepsilon E \leq I && (5.9)\\ & X_0Y - \varepsilon E \geq I && (5.10) \end{aligned}$$

where ε is an additional slack variable. Algorithm 1 has been inspired by some elements of the solutions of the Netflix Prize (see [27]). The Algorithm takes some initial matrix X_0, fixes alternately either X or Y and computes the optimal solution relative to the fixed matrix. In each iteration it holds that $\|X_kY_k - I\| \leq \|X_{k-1}Y_{k-1} - I\|$ because

$$\begin{aligned} \|X_kY_k - I\| &= \min_X \quad \|XY_k - I\| \\ &\leq \quad \|X_{k-1}Y_k - I\| \\ &= \min_Y \quad \|X_{k-1}Y - I\| \\ &\leq \|X_{k-1}Y_{k-1} - I\|. \end{aligned}$$

Though these linear problems are relatively easy to solve, there is no guarantee that this approach will lead to an optimal solution, i. e., matrices that are inverse to each other. The sequence $a_k := \|X_kY_k - I\|$ is a bounded monotonically decreasing sequence and therefore will converge, but we do not know the rate of convergence. It is also not clear whether the

In the first application we will return to Example 3.2 and see how well the estimates work if the graph corresponding to the transition matrix has a certain structure. In the second example we will look at a data set from the field of molecular dynamics, from where the initial motivation for this thesis came. In both cases th knowledge of certain rows and columns was assumed. The work flow was the same as the structure of this thesis implied, i. e.

- The inequalities of chapter 3 have been applied to attain upper and lower bounds on $H, \mathcal{N}$, and $\mathcal{N}^{-1}$.
- The interval Gauss-Seidel has been used to tighten these bounds
- The preprocessing (5.7 has been solved or a $\mathcal{C}$ that contained apart from the conditions (ii)-(iv) in Problem 1 and Problem 2 also the inequalities (5.5) and (5.6)
- As a start value for Algorithm 1 the midpoint matrices $\check{X}_{\mathcal{N}}$ and $\check{X}_H$ have been chosen

Cycle of three Cluster As described in Example 3.2 the graph in figure 3.1 describes a process with three sets of states between which jumps occur relatively rare. Every jump between the main sets involves the three nodes $3, 6$ and 9. We assumed that the rows and columns of N for the states $3, 6$ and 9 are known, which corresponds to 42 of the overall

81 mean first passage times. After the preprocessing step for $H := H^{(9)}$ we are left with the radii of the two interval matrices X_H and $Y_{H^{-1}}$

$$rad(X_H) = \begin{bmatrix} 0 & 11.9032 & 0 & 0 & 0 & 0 & 0 & 0 \\ 11.9032 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 11.9032 & 0 & 0 & 0 \\ 0 & 0 & 0 & 11.9032 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 11.9032 \\ 0 & 0 & 0 & 0 & 0 & 0 & 11.9032 & 0 \end{bmatrix}$$

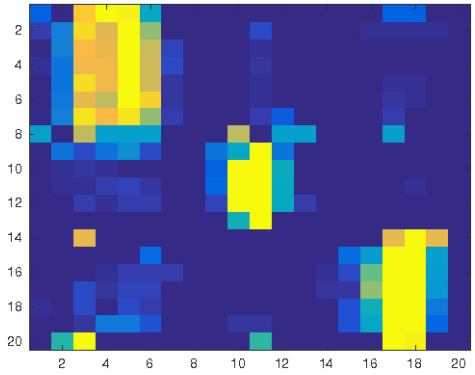
$$rad(Y_{H^{-1}}) = \begin{bmatrix} 0.0487 & 0.0570 & 0.0302 & 0 & 0 & 0 & 0 & 0 \\ 0.0570 & 0.0487 & 0.0302 & 0 & 0 & 0 & 0 & 0 \\ 0.0626 & 0.0626 & 0.0194 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0487 & 0.0570 & 0.0302 & 0 & 0 \\ 0 & 0 & 0 & 0.0570 & 0.0487 & 0.0302 & 0 & 0 \\ 0 & 0 & 0 & 0.0626 & 0.0626 & 0.0194 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0487 & 0.0570 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.0570 & 0.0487 \end{bmatrix}.$$

We see that the number of free variables in H and therefore in N has been reduced to 6. Also already a lot of zero values in H^{-1} were found, which correspond to zero values in P. The remaining entries in H^{-1} have been narrowed down to intervals with the approximate length of 0.1. In fact, when we added the condition $p_{ii} = 0$ as it is indicated in Figure 3.1, the matrix $\check{Y}_{H^{-1}}$ is the same matrix with respect to the transition probabilities in the graph, i.e., the preprocessing not only already solved the problem but even found the correct solution. The preprocessing for $\mathcal{N}$ and $\mathcal{N}^{-1}$ yields the same results.

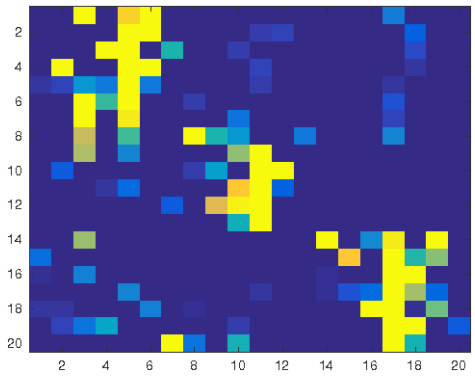
Butane The initial motivation to estimate mean first passage times and transition probabilities from an incomplete MFP-matrix came from molecular dynamics, which is why we want to look at a data set from a simulation of a small molecule.

The transition matrix P shown in figure 5.1a has been constructed from a time series of the dihedral angle that is defined by the four carbon atoms of butane. The three blocks of P are related to the three macroscopic states in which butane spends most of its time. They are called metastabilities. In the last example, to estimate the missing entries we chose those states that needed to be visited on most paths in the graph. With the same motivation in this case we chose one state from each metastability, i.e., the states $5, 11$, and 17 in figure 5.1a. After upper and lower bounds have been computed, Algorithm 1 was applied for $X_0 = \check{X}_{\mathcal{N}}$ and $K = 100$. The transition matrix that was extracted from the resulting estimate of $\mathcal{N}^{-1}$ had a similar block structure as the original matrix and can be seen in figure 5.1b.

The *PCCA+* algorithm is a kind of spectral clustering algorithm that can be used to partition the set of states of a transition matrix such that the probability to stay in a certain partition is maximized. The identification of such a partitioning is a common analysis tool in molecular dynamics (see [5, 6]). Applying *PCCA+* to the two transition matrices yields the same partitioning which supports the impression that the solution describes the dynamic of the system well.



(a) original transition matrix



(b) estimation based on MFPs of the states $5, 11$, and 17 and subsequent application of Algorithm 1.

Figure 5.1: transition matrix based on simulation data of butane in water.

Chapter 6

Conclusion

In molecular dynamics the computation of mean first passage times needs a lot of time, energy and data, so we started out with the question whether it is possible to attain information about the full system from a smaller data set. In the representation of mean first passage times as a matrix the question was how to determine the missing entries of a partial matrix such that the completed matrix is the MFP-matrix for an irreducible transition matrix.

M.Neumann and Nung-Sing Sze ([34]) have worked on the general problem of deciding whether a positive matrix is a MFP-matrix and showed a connection between MFP-matrices and $\mathcal{M}$-matrices. We used this relationship to develop lower and upper bounds not only for mean first passage times but also for the transition probabilities and the stationary distribution based on the knowledge of a subset of states. On the one hand these intervals can be used to compute an estimation using for example a *d.c.* programming approach as described in chapter 4. The inverse MFP-matrix problem is only solved if a global minimum of the optimization problem is found.

On the other hand if the bounds are too large we can use the length of the intervals as an indicator which mean first passage time should be computed next to improve the estimates.

Another interesting application of such estimates would be, if there is a complete MFP-matrix describing a system and either the system itself (say a molecule) or some simulation parameter is changed. We assume that the graph associated with the transition matrix of the altered system is similar to the graph of the initial system. Then the states that yield good estimates on the initial system, should also approximate the altered system well.

As of now the linear algorithm does not take much advantage of the particular problem and future work should be invested in a more efficient implementation. Also a thorough application of the general formulas indicated in Theorem 3.7 would probably lead to a further improvement of the results.

Bibliography

[1] A. Ben-Israel and T. N. Greville. *Generalized Inverses: Theory and Applications (CMS Books in Mathematics)*. Springer, 2003.

[2] J. Bennett and S. Lanning. The netflix prize. In *Proceedings of KDD Cup and Workshop*, 2007.

[3] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences (Classics in Applied Mathematics)*. Society for Industrial and Applied Mathematics, 1987.

[4] G. E. Coxson. Computing exact bounds on elements of an inverse interval matrix is np-hard. *Reliable Computing*, 5(2):137–142, 1999.

[5] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled markov chains. *Linear Algebra and its Applications*, 315(1-3):39 – 59, 2000.

[6] P. Deuflhard and M. Weber. Robust perron cluster analysis in conformation dynamics. *Linear Algebra and its Applications*, 398:161 – 184, 2005. Special Issue on Matrices and Mathematical Biology.

[7] R. Durrett. *Probability: Theory and Examples (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, 2010.

[8] I. Erdelyi. On the matrix equation Ax = ABx. *,Journal of Mathematical Analysis and Applications 17*, 1967.

[9] M. Fiedler and V. Ptak. On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Mathematical Journal*, 1962.

[10] F. R. Gantmacher. *Matrix Theory, Vol. 2*. American Mathematical Society, 2000.

[11] C. Hartmann, R. Banisch, M. Sarich, T. Badowski, and C. Schütte. Characterization of rare events in molecular dynamics. *Entropy (Special Issue)*, 16(1):350 – 376, 2013.

[12] L. Hogben. Completion of inverse M-Matrix patterns. *Linear Algebra and its Applications*, 282:145–160, 1998.

[13] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University

Press, 1994.

[14] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization (Nonconvex Optimization and Its Applications)*. Springer, 1995.

[15] J. J. Hunter. Simple procedures for finding mean first passage times in markov chains. *Research Letters in the Information and Mathematical Sciences Vol.8*, 2005.

[16] J. J. Hunter. Mixing times with applications to perturbed markov chains. *Linear Algebra and its Applications*, 417(1):108 – 123, 2006.

[17] A. Z. Jiu Ding. Eigenvalues of rank-one updated matrices with some applications. *Applied Mathematics Letters 20*, 2007.

[18] C. R. Johnson. Inverse M-matrices. *Linear Algebra and its Applications 47:195-216*, 1982.

[19] C. R. Johnson and R. L. Smith. The completion problem for M-Matrices and Inverse M-Matrices. *Linear Algebra and its Applications*, 241-243:655–667, 1996.

[20] C. R. Johnson and R. L. Smith. Path product matrices. *Linear and Multilinear Algebra*, 46(3):177–191, 1999.

[21] C. R. Johnson and R. L. Smith. Almost principal minors of inverse M-matrices. *Linear Algebra and its Applications*, 337(1):253 – 265, 2001.

[22] C. R. Johnson and R. L. Smith. Inverse M-matrices, ii. *Linear Algebra and its Applications 435: 953-983*, 2011.

[23] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. D. Van Nostrand, 1960.

[24] J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov Chains: with a chapter of Markov Random Fields by David Griffeath (Graduate Texts in Mathematics)*. Springer, 2012.

[25] M. Kijima. *Markov Processes for Stochastic Modeling (Stochastic Modeling Series)*. Chapman and Hall/CRC, 1997.

[26] S. J. Kirkland and M. Neumann. *Group Inverses of M-Matrices and Their Applications (Chapman & Hall/CRC Applied Mathematics & Nonlinear Science)*. Chapman and Hall/CRC, 2012.

[27] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems, 2009.

[28] S. B. Lieven Vandenberghe and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM. J. Matrix Anal. and Appl.*, 19(2):499–533, 1998.

[29] T. L. Markham. Nonnegative matrices whose inverses are M-matrices. *Proc. Amer. Math. Soc.*, 36:326–330, 1972.

[30] C. D. Meyer. The role of the group generalized inverse in the theory of finite markov chains. *Society for Industrial and Applied Mathematics Vol.17, No.3*, 2005.

[31] C. D. Meyer and G. E. Cho. Markov chain sensitivity measured by mean first passage times. *Linear Algebra and its Applications 316:21-28*, 2000.

[32] L. Mollica, S. Decherchi, S. R. Zia, R. Gaspari, A. Cavalli, and W. Rocchia. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.*, 5:11539, jun 2015.

[33] A. Neumaier. *Interval Methods for Systems of Equations (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, 1990.

[34] M. Neumann and N.-S. Sze. On the inverse mean first passage matrix problem and the inverse M-Matrix problem. *Linear Algebra and its Applications 434*, 2011.

[35] D. Perez, B. P. Uberuaga, Y. Shim, J. G. Amar, and A. F. Voter. Chapter 4 accelerated molecular dynamics methods: Introduction and recent developments. volume 5 of *Annual Reports in Computational Chemistry*, pages 79 – 98. Elsevier, 2009.

[36] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics*, 134(17), 2011.

[37] J. Rohn. An algorithm for solving the absolute value equation. *Electronic Journal of Linear Algebra*, 18:589–599, 2009.

[38] J. Rohn and R. Farhadsefat. Inverse interval matrix: A survey. *Electronic Journal of Linear Algebra*, 22(1), jan 2011.

[39] S. Rump. INTLAB - INTerval LABoratory. In T. Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. http://www.ti3.tuhh.de/rump/.

[40] C. Schütte, W. Huisinga, and P. Deuflhard. Transfer operator approach to conformational dynamics in biomolecular systems. Technical Report SC-99-36, ZIB, Takustr.7, 14195 Berlin, 1999.

[41] D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L.-S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. Ben Schafer, N. Siddique, C. W. Snyder, J. Spen-

gler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang, and C. Young. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '14, pages 41–53, Piscataway, NJ, USA, 2014. IEEE Press.

[42] U. von Luxburg. Tutorial on spectral clustering. *Max Planck Institute for Biological Cybernetics, Technical Report No. 149, Tüubingen*, 2006.

[43] G. Windisch. *M-matrices in Numerical Analysis (Teubner-Texte zur Mathematik) (German Edition)*. Vieweg+Teubner Verlag, 1989.