



Aggregation- och separationsprocesser

En undersökning av molekylära systems asymptotiska egenskaper

Fusion-disaggregation processes

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Lydia Andersson

Karin Furufors

Jessica Gilmsjoe

Emelie Lööf

Aggregation- och separationsprocesser

En undersökning av molekylära systems asymptotiska egenskaper

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet

Lydia Andersson Jessica Gilmsjoe Emelie Lööf

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers

Karin Furufors

Handledare: Sergei Zuyev
Examinator: Ulla Dinger
 Maria Roginskaya

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2020

Förord

Det här är ett kandidatarbete av undersökande karaktär. Arbetet behandlar stokastiska processer, i synnerhet Markovkedjor och dess asymptotiska egenskaper. Vi vill tacka vår handledare Sergei Zuyev för sitt hängivna engagemang och sin uppmuntrande vägledning under arbetets gång.

En individuell tidslogg och dagbok har förts över samtliga författares bidrag. Under rapportskrivandet har alla varit delaktiga i samtliga avsnitt dock har ansvarsområden delats ut. Lydia Andersson har varit ansvarig för den populärvetenskapliga presentationen och avsnittet Inledning. Jessica Gilmsjoe har ansvarat för framställningen av avsnittet Teori. Jessica Gilmsjoe och Karin Furufors har tillsammans ansvarat för avsnittet Modell. Emelie Lööf har med hjälp från Karin Furufors ansvarat för avsnittet Resultat. Emelie Lööf har också ansvarat för simulering och slutligen har Karin Furufors ansvarat för avsnittet Slutsats. Vi vill dock betona att rapporten skrivits gemensamt och att alla har varit involverade i samtliga delar.

Populärvetenskaplig presentation

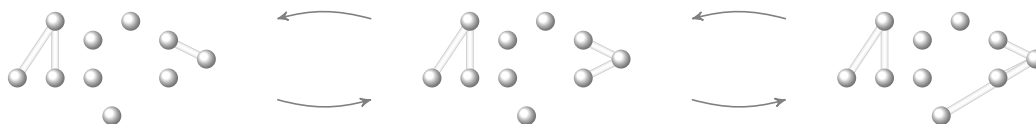
För över 2000 år sedan påstod filosofen Leukippos från Miletos [1] att världen endast bestod av två saker – tomrum och atomer. Det senare menade filosofen var små bitar av materia, så små att de var osynliga för det mänskliga ögat. Dessa atomer bildade olika geometriska mönster och Leukippos tänkte att de på så vis kunde skapa allt i världen, förutom tomrum då förstås.

Över 1500 år senare, under renässansen, började vetenskapsmän använda sig av experimentella undersökningar istället för filosofiskt tänkande [2]. Genom experimentella undersökningar kunde kemister som Robert Boyle lägga grunden till dagens moderna kemi [3]. Med detta kunde John Dalton under 1800 talets början analysera hur syreatomer förenar sig med såväl väte som koppar. Från dessa observationer drog Dalton slutsatsen att vid en kemisk reaktion förenar odelbara atomer sig till en delbar enhet, vilken han kallade en *molekyl* [4].

Sedan John Dalton introducerade sin forskning har en uppsjö resultat inom området tillkommit och för att förstå hur en molekyl fungerar finns idag en rad olika verktyg [5]. Det kan vara intressant att istället för att undersöka en enskild molekyl och dess lokala egenskaper undersöka hur *en samling* atomer och molekyler samverkar över tid i ett *molekylärt system* globalt. Resultaten i rapporten är centrerade kring det senare.

Vid studier av hur molekyler förändras kan de experimentella undersökningarna verka oombärliga men faktum är att strukturen av molekyler går att representera allmängiltigt. Genom att i undersökningen applicera matematik, i synnerhet sannolikhetssteori och stokastiska processer, gör den generella matematiska representationen att den komplexa strukturen inte bara bli mer begriplig, den genererar också intressanta matematiska resultat.

År 1907 publicerade Andrei Markov avhandlingen *Investigation of a noteworthy case of dependent trials* innehållande den fundamentala upptäckten av vad som ligger till grund för Markovkedjor och Markovprocesser [6]. En Markovprocess är en stokastisk process som beskriver hur ett system förflyttar sig mellan olika tillstånd där processens framtid endast är beroende av det nuvarande tillståndet [7]. Markovprocesser lämpar sig utmärkt för undersökningar av molekylära system genom att representera varje kombination av atomer och molekyler som ett tillstånd.



Figur 1: I figuren ses ett molekylärt system med tio atomer i tre olika tillstånd. Ett tillstånd representerar fördelningen av molekylernas olika längd.

I undersökningen representeras det molekylära systemets olika tillstånd som i figur 1. Rapportens resultat undersöker hur systemet rör sig mellan olika tillstånd och betar sig över tid. Speciellt fokuserar rapporten på spridningen av molekylernas storlek i ett stabilt läge. Med ett stabilt läge menas ett läge där systemets fördelning av antalet molekyler i systemet inte förändras med tiden. Resultaten erhålls genom att använda analytiska resultat som exempelvis reversibilitet från stokastiska processer, vilket är en egenskap som innebär att processen betar sig likadant oberoende om processen rör sig framåt eller bakåt i tiden. De analytiska resultaten verifieras med datorsimuleringar.

Det sägs att Andrei Markov ska ha inspirerats av växlingarna mellan konsonanter och vokaler i Pusjkins poesi när han lade grunden för modellen vilken idag kallas Markovkedjor [7]. Det kan verka konstigt att samma modell som är sprungen ur sannolikheten för en viss vokal i ett ord i *Eugene Onegin*, idag kan användas för att undersöka hur molekylära system betar sig över tid. Sa-

ken är den, att Markov genom sina observationer lyckades precisera något på ett matematiskt plan som Leukippos elev Demokritos sägs ha sagt för redan 2000 år sedan, nämligen "*All förändring är endast förening och upplösning av delar*"[8].

Sammanfattning

Undersökningen studerar utvecklingen av slutna linjära molekylära system och dess asymptotiska egenskaper. För att göra detta tittar vi inledningsvis på system med ett litet antal atomer för att sedan generalisera resultaten till att även gälla allmänna system. För att göra detta tillämpar vi metoder för Markovkedjor, vilket är en underliggande teoretisk modell för det molekylära systemet. Det molekylära systemets stationära fördelning erhålls explicit genom att tillämpa den så kallade balansekvationen och Kolmogorovs kriterium för reversibilitet. Frekvensfunktionen för den stationära fördelningen av det totala antalet molekyler fastställs och undersökningen fortgår med att studera det molekylära systemets asymptotiska beteende. Slutligen verifieras de erhållna teoretiska resultaten med datorsimuleringar.

Abstract

The report studies an evolution of a closed linear molecular system and its asymptotic properties. To fix the problematics, we consider initially a system with a small number of atoms and further generalise the results to general systems. To do this, we apply probability methods of Markov chains, which is an underlying theoretical model of the molecular system. The stationary distribution of the system is explicitly found by establishing the reversibility of the model. The method used is based on Kolmogorov's criterion for reversibility and checking the so-called balance equations. The probability mass function for the stationary distribution of the total number of molecules is established and the report continues by studying its asymptotic properties, in particular, the convergence speed to stationarity. Finally, the obtained theoretical results are verified with computer simulations.

Innehåll

1	Inledning	1
1.1	Bakgrund	1
1.2	Avgränsningar	1
1.3	Problembeskrivning	2
1.4	Syfte och frågeställningar	2
2	Teori	3
2.1	Stationäritet	3
2.2	Reversibilitet	5
3	Modell	6
4	Resultat	7
4.1	Mindre system	7
4.2	Generella system	9
4.2.1	Reversibilitet	10
4.2.2	Asymptotisk fördelning	14
4.3	Simulering	15
5	Slutsats	17
A	Bilagor	19
A.1	Vilka tillstånd är möjliga?	19
A.2	Ordlista	20
A.3	Kod	21

1 Inledning

1.1 Bakgrund

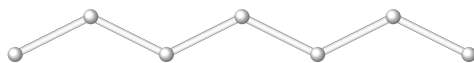
Genom att abstrahera verklighetens ofta komplicerade problem med matematik kan problemens lösning bli av mer generell karaktär och ibland till och med lösa närliggande problem [9]. Det finns också fall då ett matematiskt problem kan preciseras och göras mer begripligt genom att representera de olika delarna av problemet med fenomen från andra discipliner [9]. Undersökningar av dessa slag bör med fördel modelleras så att det finns utrymme för att lösa ett observerbart problem med matematik, samtidigt som det abstraherade problemet säger något meningsfullt om och utvecklar matematiken i sig.

Ett exempel på ett sådant samspel är undersökning av molekyllära system med hjälp av stokastiska processer. Ett molekyllärt system är en samling molekyler bestående av atomer vilka kan aggregera eller separera. En stokastisk process som beskriver hur ett system förflyttar sig mellan olika tillstånd, där processens framtid endast är beroende av det nuvarande tillståndet kallas för en Markovkedja [10]. Genom att använda sig av Markovprocesser kan det molekyllära systemets långsiktiga egenskaper studeras. En möjlig tillämpning av de matematiska resultaten skulle, med viss utveckling, till exempel kunna förutspå olika materials livslängd. För nog skulle det vara lockande för ett elektronikföretag att kunna säga exakt när en kabelstrumpa behöver bytas ut? Detta blir särskilt viktigt för framsteg i utvecklandet av hållbara material.

Frank Kelly bröt ny mark när han år 1979 gav ut boken *Reversibility and stochastic networks* i vilken han beskriver reversibilitet och hur analysen av ett system kan förenklas genom att använda sig av egenskaper som följer av reversibilitet [11]. En tillämpning Kelly tar upp i sin bok berör det molekyllära system vilket också undersöks i denna rapport. Rapporten skiljer sig från Kellys resultat i den bemärkelsen att den utvidgar resonemangen som förs i *Reversibility and stochastic networks* genom att bestämma frekvensfunktionen av den stationära fördelningen och därigenom också kan undersöka asymptotiska egenskaper hos det molekyllära systemet. De asymptotiska egenskaperna verifieras också med simuleringar.

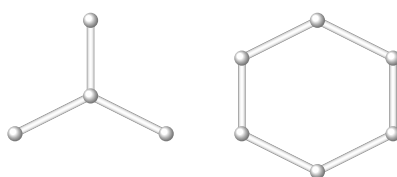
1.2 Avgränsningar

I detta projekt undersöks slutna, molekyllära system där molekyllerna endast består av en sorts atom. Med ett slutet system menas att systemet inte interagerar med omvärlden. Det antas att molekyllerna är av linjär struktur, det vill säga att atomerna högst kan binda två atomer till sig, se figur 2.



Figur 2: Grafisk representation av molekyllstrukturer som kommer att undersökas.

Arbetet avgränsas ytterligare i den bemärkelsen att det antas att molekylen inte är av cirkulär struktur. Med andra ord kan inte en molekyls ändpunkter binda samman. Exempel på molekyllstrukturer vi inte kommer undersöka kan ses i figur 3.



Figur 3: Grafisk representation av molekyler som inte avses studeras.

Därtill undersöks inte specifika material. Detta ger en mer generell inriktning på projektet och leder således till att slutsatser inte kan dras gällande särskilda material. Yttre faktorer så som hur åldrande påverkar materialet är också en parameter som inte tas med i projektet. Notera även att de molekylära systemen, på grund av slutenheten, kommer innehålla en konstant mängd atomer under hela förloppet.

1.3 Problembeskrivning

För att beskriva det molekylära systemet med Markovkedjor definieras systemets olika tillstånd. Varje tillstånd i systemet beskrivs med en unik kombination av molekyler. Genom att studera intensiteterna av övergångarna mellan de olika tillstånden dras slutsatser kring huruvida systemets molekylantal tenderar att röra sig mot en viss fördelning. Studien utförs analytiskt men också med hjälp av simuleringar.

Ett centralt begrepp i rapporten är *stationär fördelning*. En stationär fördelning är en fördelning som Markovkedjan antar för att sedan bli kvar i den för alltid. Rapporten redogör att om systemet uppvisar egenskapen reversibilitet kan den stationära fördelningen erhållas genom att använda resultat från denna egenskap.

Utifrån den stationära fördelningen kan sedan en frekvensfunktion hörande till antalet molekyler definieras. Frekvensfunktionens polynom undersöks för att analysera det molekylära systemets asymptotiska egenskaper. Dessa verifieras sedan via simulering.

1.4 Syfte och frågeställningar

Syftet med denna uppsats är att med hjälp av Markovkedjor undersöka molekylära systems beteende över tid.

Undersökningen ämnar besvara följande frågeställningar:

1. Är Markovkedjan som representerar det molekylära systemet reversibel?
2. Hur ser det molekylära systemets stationära fördelning av antalet molekyler och dess längder ut?
 - (a) Kan den stationära fördelningens frekvensfunktion hörande till antalet molekyler uttryckas?
 - (b) Är frekvensfunktionen hörande till antalet molekyler asymptotiskt normalfördelad?

2 Teori

I detta arbete betraktar vi stokastiska processer $X(t)$, $t \in T$ som kan förflytta sig i ett ändligt tillståndsrum S . Specifikt betraktar vi Markovkedjor där $t \in T$ kan vara diskret eller kontinuerlig tid. I detta avsnitt presenteras de begrepp som ligger till grund för arbetets resultat.

Definition 2.1. En Markovkedja är en stokastisk process $X(t)$ som uppfyller Markovegenskapen:

$$\begin{aligned} P(X(t_{n+1}) = j | X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_n) = i_n) \\ = P(X(t_{n+1}) = j | X(t_n) = i_n) \end{aligned}$$

för $i, j \in S$ och $t_0 < t_1 \cdots < t_n < t_{n+1} \in T$.

Definitionen innebär att sannolikheten att förflytta sig till nästkommande tillstånd i ett tidssteg endast beror på det nuvarande tillståndet. I och med att processen ej beror av dess tidigare tillstånd brukar man tala om att Markovkedjan saknar minne [10].

Om det gäller att $P(X(t + \tau) = i | X(t) = j)$, $t, \tau \in T$ inte beror av t sägs Markovkedjan vara *tidshomogen* [11]. Fortsättningsvis antar vi att Markovkedjan vi behandlar i teorin är tidshomogen.

Om det är möjligt att för något antal steg nå ett tillstånd j från ett tillstånd i , sägs i kommunicera med j . En Markovkedja vars tillstånd alla kommunicerar med varandra kallas *irreducibel* [10].

En markovkedja i diskret tid är periodisk om det existerar ett heltal $\delta > 1$ så att $P(X(t + \tau) = j | X(t) = j) = 0$, såvida τ ej är delbar med δ . I annat fall är kedjan *aperiodisk* [11].

Sannolikheten för en övergång från tillstånd i till tillstånd j för en Markovkedja i diskret tid ges av övergångssannolikheten

$$p(i, j) = P(X(t + 1) = j | X(t) = i).$$

Alla möjliga övergångar representeras i övergångsmatrisen \mathbf{P} och det gäller att $\sum_{j \in S} p(i, j) = 1$, $i \in S$.

För en Markovkedja i kontinuerlig tid kan vi definiera *intensiteten* att ta sig från tillstånd i till tillstånd j enligt följande

$$q(i, j) = \lim_{\tau \rightarrow 0} \frac{P(X(t + \tau) = j | X(t) = i) - \delta_{ij}}{\tau} \quad i \neq j.$$

Tiden Markovkedjan tillbringar i tillstånd i är exponentialfördelad med parameter

$$q(i) = \sum_{j \in S} q(i, j).$$

Att tiden är exponentialfördelad lämpar sig då exponentialfördelningen också saknar minne [10]. Intensiteterna $q(i, j)$ mellan möjliga övergångar tillhörande Markovkedjan utgör elementen i intensitetsmatrisen \mathbf{Q} . Övergångssannolikheten ges av sambandet

$$p(i, j) = \frac{q(i, j)}{\sum_{j \in S} q(i, j)}.$$

2.1 Stationäritet

En irreducibel och aperiodisk Markovkedja i diskret tid kan ha en asymptotisk fördelning. Om tillståndsrummet S dessutom är ändligt existerar den alltid. Den asymptotiska fördelningen är en samling av positiva tal $\pi(j)$, $j \in S$ med $\sum_{j \in S} \pi(j) = 1$ som uppfyller

$$\pi(j) = \sum_{i \in S} \pi(i)p(i, j) \quad j \in S. \quad (2.1)$$

Om vi kan hitta en samling av positiva tal som uppfyller 2.1 vars summa är ändlig, kan samlingen normeras för att bilda en asymptotisk fördelning. Med villkoren irreducibilitet och aperiodicitet är den asymptotiska fördelningen π unik och

$$\lim_{t \rightarrow \infty} P(X(t) = i | X(0) = j) = \pi(i),$$

och π är då också gränsfördelningen. Det betyder att den asymptotiska fördelningen inte beror av det initiala tillståndet. Om det gäller att $P(X(0) = j) = \pi(j)$, $j \in S$ följer det att $P(X(t) = j) = \pi(j)$, $j \in S$ och därmed sammanfaller π med kedjans stationära fördelning [11].

Den stationära fördelningen beskriver Markovkedjans långsiktiga beteende över tid.

Definition 2.2. En Markovkedja med övergångsmatrix P sägs ha stationär fördelning π om det finns samling av positiva tal $\pi(j)$, $j \in S$ med $\sum_{j \in S} \pi(j) = 1$ för alla j , som uppfyller

$$\pi = \pi P. \quad (2.2)$$

Detta innebär att om Markovkedjan startar i fördelningen π kommer kedjan förbli i denna fördelning för alltid.

I fallet då Markovkedjan är tidskontinuerlig är den asymptotiska fördelningen en samling av positiva tal $\pi(j)$, $j \in S$ med $\sum_{j \in S} \pi(j) = 1$ som uppfyller

$$\pi(j) \sum_{i \in S} q(j, i) = \sum_{i \in S} \pi(i)p(i, j) \quad j \in S. \quad (2.3)$$

Om den existerar i fallet då Markovkedjan är irreducibel, är den asymptotiska fördelningen unik och sammanfaller med gränsfördelningen samt den stationära fördelningen [11].

Nedan följer en sats gällande irreducibla och aperiodiska Markovkedjors exponentiella konvergens.

Sats 2.3. Låt P vara övergångsmatrisen för en Markovkedja med ändligt tillståndsrum S . Om P är irreducibel och aperiodisk, med stationär fördelning π , existerar konstanter $0 < \alpha < 1$ och $C > 0$ så att:

$$\max_{x \in S} \|P^t(x, \cdot) - \pi\|_{TV} \leq C\alpha^t, \quad (2.4)$$

där $P(x, \cdot)$ är den x :te raden i P .

Bevis. [12, Sats 4.9] □

Vänsterledet innehåller den totala variansen mellan fördelningen och dess stationära fördelning vilket är ett mått på hur mycket de skiljer sig åt.

Definition 2.4. Den totala variansen mellan två fördelningar p_1, p_2 ges av:

$$\|p_1 - p_2\|_{TV} = \max_{A \in S} |p_1(A) - p_2(A)|, \quad (2.5)$$

där $p(A) = \sum_{x \in A} p(x)$.

Sats 2.3 gäller för diskret tid. I kontinuerlig tid kan villkoret om aperiodicitet frångås och sats 2.4 går att använda för irreducibla Markovkedjor [13, Sats 1.2].

2.2 Reversibilitet

En Markovkedja som är reversibel har egenskapen att kedjans beteende är likadant oberoende av om vi förflyttar oss framåt eller bakåt i tiden. Denna egenskap är central för att kunna bestämma Markovkedjans stationära fördelning och dess asymptotiska egenskaper.

Definition 2.5. *En Markovkedja $X(t)$ i kontinuerlig tid med tillståndsrum S och intensiteter $q(i, j)$ är reversibel om $(X(t_1), \dots, X(t_n))$ har samma fördelning som $(X(t_0 - t_1), \dots, X(t_0 - t_n))$ för alla $t_1, \dots, t_n \in T$.*

Vi kan avgöra om en Markovkedja är reversibel genom att finna en sannolikhetsvektor som uppfyller balansekvationen.

Sats 2.6. *En Markovkedja i kontinuerlig tid är reversibel om och endast om det finns samling av positiva tal $\pi(j)$, $j \in S$ med $\sum_{j \in S} \pi(j) = 1$ för alla j , som uppfyller balansekvationen*

$$\pi(j)q(j, i) = \pi(i)q(i, j) \quad i, j \in S. \quad (2.6)$$

Om det existerar en sådan samling $\pi(j)$, $j \in S$, så är det kedjans asymptotiska fördelning.

Bevis. [11, Sats 1.3] □

$\pi(j)q(j, i)$ betecknar sannolikhetsflödet från tillstånd j till i , och balansekvationen kräver att sannolikhetsflödet mellan ett par av tillstånd är samma. Om vi kan hitta en sannolikhetsvektor som uppfyller balansekvationen så är det Markovkedjans asymptotiska fördelning och därmed dess stationära fördelning [11]. Balansekvationen 2.6 för en Markovkedja i diskret tid motsvaras i att ersätta intensiteterna i 2.6 med kedjans övergångssannolikheter.

Förutom att avgöra reversibilitet genom att uppfylla balansvillkoret, kan vi med Kolmogorovs kriterium fastställa reversibilitet genom att enbart använda intensiteterna.

Sats 2.7 (Kolmogorovs kriterium för reversibilitet). *En Markovkedja i kontinuerlig tid är reversibel om och endast om det för alla cykliska sekvenser $\{i, i_1, i_2, \dots, i_n, i\} \in S$ gäller*

$$q(i, i_1)q(i_1, i_2) \cdots q(i_n, i) = q(i, i_n)q(i_n, i_{n-1}) \cdots q(i_1, i). \quad (2.7)$$

Bevis. [11, Sats 1.7] □

För att Kolmogorovs kriterium ska gälla behöver vi inte välja alla möjliga slutna vägar $i_1, i_2, \dots, i_n, i_1$, utan vi kan välja en enkel och specifik väg. Alltså är det möjligt att välja en enkel väg och Kolmogorovs kriterium följer då även i det generella fallet [11].

2.6 och 2.7 tillåter oss att uttrycka den stationära fördelningen av en reversibel Markovkedja i kontinuerlig tid på formen

$$\pi(j) = \pi(i_0) \frac{q(i_0, i_1)q(i_1, i_2) \cdots q(i_n, j)}{q(i_n, i_{n-1}) \cdots q(i_1, i_0)}, \quad (2.8)$$

för något initialt tillstånd i_0 och en sekvens av tillstånd i, i_1, \dots, i_n, j där produkten av intensiteterna i 2.8 är nollskilda. Det explicita värdet av $\pi(i)$ fås genom normalisering: $\sum_j \pi(j) = 1$. Om $\pi(j)$ uppfyller 2.6 vet vi att Markovkedjan är reversibel och $\pi(j)$ är dess stationära fördelning.

3 Modell

I det här avsnittet preciseras vad som gäller för systemen i rapporten. Modellen avser slutna molekylära system med ändliga antal atomer. Dessa atomer bildar molekyler av linjär typ, i enighet med tidigare representation. Hur systemets sammansättning av molekyler av olika storlek förändras kallar vi dess tidsutveckling. Denna utveckling kan beskrivas med en tidskontinuerlig Markovkedja. Detta är lämpligt då det enda som avgör vilka tillstånd som kan nås och på vilka sätt det kan göras är systemets tidigare tillstånd. Systemet uppfyller alltså Markovegenskapen definierad i avsnitt 2. I varje tidssteg är två skeenden möjliga. Det kan antingen ske en aggregation i vilken två molekyler binds samman eller en separation där bindningen mellan två atomer i en molekyl upplöses. I modellen beror inte sannolikheten för dessa två skeenden på var i tiden kedjan befinner sig och Markovkedjan är följaktligen tidshomogen. Vi antar att tiden kedjan tillbringas i ett tillstånd är exponentialfördelad med parameter λ .

Tillståndet \mathbf{n}_i som systemet befinner sig i kan beskrivas med en vektor, $\mathbf{n}_i = (n_1, n_2, \dots, n_N)$ i tillståndsrummet S_N . Vektorn är av samma längd som antalet atomer N i systemet. Det n_j :te representerar antalet molekyler i systemet med j atomer sammanlänkade i en kedja av $j - 1$ stycken bindningar och det gäller att

$$\sum_{j=1}^N n_j j = N. \quad (3.1)$$

Alltså är antalet atomer i systemet fixt och kan fördelas på molekyler av längd j , där $j \in (1, \dots, N)$.

I ett första steg i användningen av Markovkedjor behöver vi ange hur intensiteterna för de olika rörelserna i det molekylära systemets tillståndsrum ska beräknas. System med antal atomer $N \geq 2$ kan skapa molekyler av storlek $n \leq N$ som är linjärt sammanlänkade av bindningar. Varje molekyl av storlek n aggregerar med en annan molekyl av storlek $r \leq N - n$ med intensitet λ och skapar därmed en molekyl av storlek $n + r$. Varje molekylpar aggregerar med en intensitet $2\lambda > 0$ oberoende av dess storlek. Varje bindning bryts med intensitet 1. För molekyler av storlek n gäller därför att de separeras till två mindre molekyler med intensitet $n - 1$.

Vi låter $X(t)$ vara Markovkedjan tillhörande systemet och låter e_k vara koordinatvektorn k i \mathbb{R}^N . Om τ är tiden för den första förändringen av systemets tillstånd, så gäller det att $X(\tau) - X(0)$ antar positiv sannolikhet och kan anta följande fyra värden:

1. $v_{jk} = -e_j - e_k + e_{j+k}$, $j \neq k$, två molekyler av olika storlekar j och k aggregerar. Intensiteten för övergången är $2\lambda n_j n_k$.
2. $v_{kk} = -2e_k + e_{2k}$, två molekyler av samma storlek aggregerar. Intensiteten för övergången är $\lambda n_k (n_k - 1)$.
3. $-v_{ik} = e_j + e_k - e_{j+k}$, $j \neq k$, en molekyl av storlek $k + j$ separeras till två molekyler av olika storlekar. Intensiteten är $2n_{k+j}$.
4. $-v_{kk} = 2e_k - e_{2k}$, en molekyl av storlek $2k$ separeras till två molekyler, båda av storlek k . Intensiteten är n_{kk} .

Alla övergångar är ej möjliga från alla tillstånd då negativa värden ej är tillåtna. Sannolikheten för att en specifik förändring skall ske ges av dess intensitet dividerat med summan av de fyra intensiteterna med avseende på alla möjliga utfall.

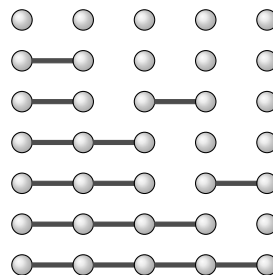
Att alla tillstånd i modellen kommunicerar inses till exempel genom att varje tillstånd med en kombination av steg 3 och 4 kan reduceras till ett system med endast ensamma atomer och därifrån byggs upp med en kombination av steg 1 och 2 till godtyckligt tillstånd. Det följer alltså från avsnitt 2 att Markovkedjan definierad ovan är irreducibel.

4 Resultat

4.1 Mindre system

Innan det molekylära systemet undersöks generellt betraktas ett mindre system. Fördelen med ett mindre system är att dess tillståndsrum S är litet och med enkelhet kan beskrivas uttömmande. I det här avsnittet undersöks ett molekylärt system bestående av fem atomer. Det inses snabbt att det endast finns sju tillstånd som ett sådant system kan befinna sig i: $S = \{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \mathbf{n}_4, \mathbf{n}_5, \mathbf{n}_6, \mathbf{n}_7\}$. Tillståndsrummet beskrivs på två sätt i figur 4 och 5.

$$\begin{aligned}\mathbf{n}_1 &= (5, 0, 0, 0, 0) \\ \mathbf{n}_2 &= (3, 1, 0, 0, 0) \\ \mathbf{n}_3 &= (1, 2, 0, 0, 0) \\ \mathbf{n}_4 &= (2, 0, 1, 0, 0) \\ \mathbf{n}_5 &= (0, 1, 1, 0, 0) \\ \mathbf{n}_6 &= (1, 0, 0, 1, 0) \\ \mathbf{n}_7 &= (0, 0, 0, 0, 1).\end{aligned}$$



Figur 4: Vektorrepresentation av tillståndsrummet för fem atomer.

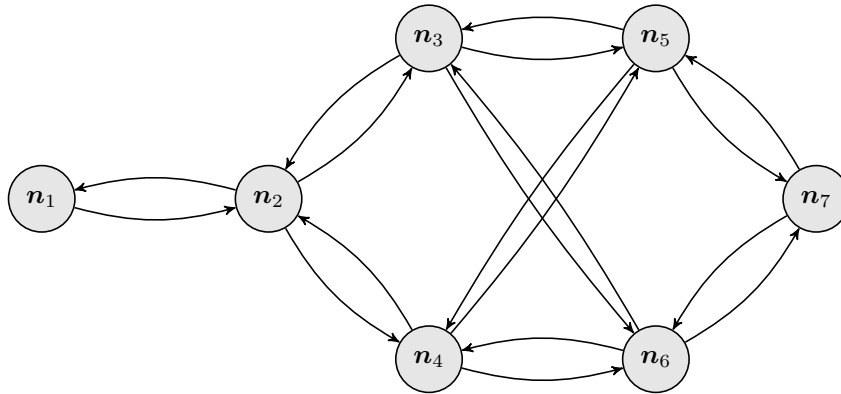
Figur 5: Grafisk representation av tillståndsrummet för fem atomer.

I denna representation är \mathbf{n}_3 , till exempel, tillståndet med två stycken 2-molekyler och en ensam atom. I Markovkedjan sker i varje steg en aggregation eller en separation. För \mathbf{n}_3 innebär en separation att någon av bindingarna mellan de två 2-molekylerna upphör och systemet övergår i tillstånd \mathbf{n}_2 . I händelse av en aggregation kan det antingen ske mellan de två 2-molekylerna eller den ensamma atomen och en 2-molekyl. Aggregationen skulle resultera i en övergång till tillstånd \mathbf{n}_6 alternativt \mathbf{n}_5 . Det finns inga andra tillgängliga tillstånd från \mathbf{n}_3 . För systemen saknar atomernas inbördes ordning betydelse, men kombinatoriskt kan varje tillstånd ses som ett makrotillstånd av alla permutationer av atomer och bindingar som resulterar i precis den kombinationen.

Vi kan beräkna intensiteterna för övergångarna mellan de sju olika tillstånden genom att tillämpa de fyra formler som introducerades i avsnitt 3. Den tidshomogena Markovkedjan får då intensitetsmatrisen \mathbf{Q} , enligt nedan.

$$\mathbf{Q} = \begin{matrix} & \mathbf{n}_1 & \mathbf{n}_2 & \mathbf{n}_3 & \mathbf{n}_4 & \mathbf{n}_5 & \mathbf{n}_6 & \mathbf{n}_7 \\ \mathbf{n}_1 & \left[\begin{array}{ccccccc} 0 & 20\lambda & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 6\lambda & 6\lambda & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 4\lambda & 2\lambda & 0 \\ 0 & 2 & 0 & 0 & 2\lambda & 4\lambda & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 2\lambda \\ 0 & 0 & 1 & 2 & 0 & 0 & 2\lambda \\ 0 & 0 & 0 & 0 & 2 & 2 & 0 \end{array} \right] \\ \mathbf{n}_2 & \\ \mathbf{n}_3 & \\ \mathbf{n}_4 & \\ \mathbf{n}_5 & \\ \mathbf{n}_6 & \\ \mathbf{n}_7 & \end{matrix}$$

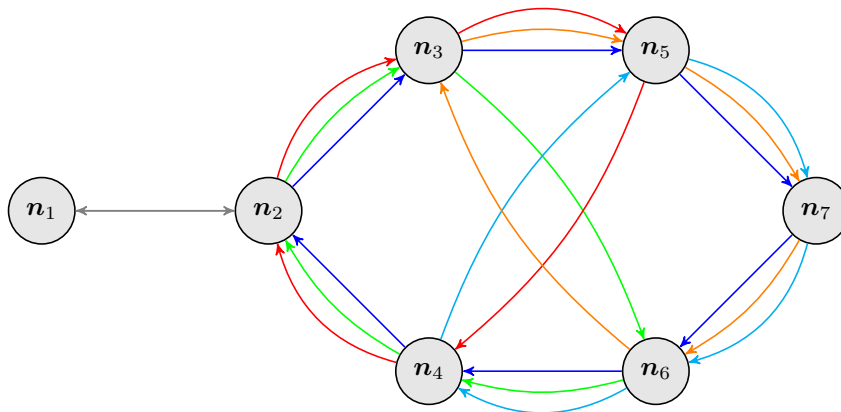
Nollorna i \mathbf{Q} representerar omöjliga övergångar, däribland huvuddiagonalen där varje element innebär en övergång från ett tillstånd tillbaka till sig självt, vilket i denna modell skulle kräva åtminstone två stycken av aggregationer och separationer för att inträffa. Markovkedjan för systemet med fem atomer visualiseras i figur 6 med utsatta intensiteter.



Figur 6: Markovkedja för ett system med fem atomer.

För att undersöka huruvida Markovkedjan är reversibel kan vi i enlighet med Kolmogorovs kriterium kontrollera att samtliga fem cykler är reversibla. Det innebär att om de löps igenom i den ena riktningen så sker det med samma intensitet som om de skulle löpas igenom åt motsatt håll. Trädstrukturen i Markovkedjan mellan n_1 och n_2 behöver inte inkluderas i cyklerna eftersom intensiteterna fram och tillbaka mellan dem båda skulle läggas till en gång i vardera riktningar och följaktligen ta ut sig själva. Totalt finns det fem cykler inuti Markovkedjan: $\{C_1, \dots, C_5\}$. C_1 besöker sex utav de sju tillstånden medan de övriga fyra cyklerna besöker fyra tillstånd vardera. Cyklerna är utskrivna nedan med färgen på den motsvarande cykeln i figur 7 givna inom parentes.

- $C_1 = n_2 \rightarrow n_3 \rightarrow n_5 \rightarrow n_7 \rightarrow n_6 \rightarrow n_4 \rightarrow n_2$ (mörkblå)
- $C_2 = n_2 \rightarrow n_3 \rightarrow n_5 \rightarrow n_4 \rightarrow n_2$ (röd)
- $C_3 = n_2 \rightarrow n_3 \rightarrow n_6 \rightarrow n_4 \rightarrow n_2$ (grön)
- $C_4 = n_3 \rightarrow n_5 \rightarrow n_7 \rightarrow n_6 \rightarrow n_3$ (orange)
- $C_5 = n_4 \rightarrow n_5 \rightarrow n_7 \rightarrow n_6 \rightarrow n_4$ (ljusblå)



Figur 7: Markovkedjan med de fem cyklerna markerade i färg.

Markovkedjan är reversibel om och endast om alla cykler har samma intensitet i båda riktningarna.

Kvoterna av cyklerna visas nedan:

$$\begin{aligned}\frac{C_1^{medsols}}{C_1^{motsols}} &= \frac{6\lambda \cdot 4\lambda \cdot 2\lambda \cdot 2 \cdot 2 \cdot 2}{6\lambda \cdot 4\lambda \cdot 2\lambda \cdot 2 \cdot 2 \cdot 2} = 1 \\ \frac{C_2^{medsols}}{C_2^{motsols}} &= \frac{6\lambda \cdot 4\lambda \cdot 1 \cdot 2}{6\lambda \cdot 2\lambda \cdot 2 \cdot 2} = 1 \\ \frac{C_3^{medsols}}{C_3^{motsols}} &= \frac{6\lambda \cdot 2\lambda \cdot 2 \cdot 2}{6\lambda \cdot 4\lambda \cdot 1 \cdot 2} = 1 \\ \frac{C_4^{medsols}}{C_4^{motsols}} &= \frac{4\lambda \cdot 2\lambda \cdot 2 \cdot 1}{2\lambda \cdot 2\lambda \cdot 2 \cdot 2} = 1 \\ \frac{C_5^{medsols}}{C_5^{motsols}} &= \frac{2\lambda \cdot 2\lambda \cdot 2 \cdot 2}{4\lambda \cdot 2\lambda \cdot 2 \cdot 1} = 1.\end{aligned}$$

Eftersom samtliga kvoter är ett har cyklerna samma intensitet i ena riktningen som i den motsatta och det konstateras att systemet med fem atomer är reversibelt.

Givet reversibiliteten vet vi att systemet har en unik stationär fördelning enligt teorin i avsnitt 2.2. Med denna vetenskap kan vi sedan använda balansekvationen för att finna dess stationära fördelning π explicit. Vi ställer upp och löser ett ekvationssystem där alla sju tillstånd ingår: $\pi(1)q(1,2) = \pi(2)q(2,1)$, $\pi(2)q(2,3) = \pi(3)q(3,2)$, $\pi(2)q(2,4) = \pi(4)q(4,2)$, $\pi(4)q(4,5) = \pi(5)q(5,4)$, $\pi(4)q(4,6) = \pi(6)q(6,4)$ och $\pi(6)q(6,7) = \pi(7)q(7,6)$. Med hjälp av dessa uttrycker vi de övriga variablerna i termer av $\pi(1)$:

$$\begin{aligned}\pi(2) &= \pi(1)20\lambda \\ \pi(3) &= \pi(1)60\lambda^2 \\ \pi(4) &= \pi(3) \\ \pi(5) &= \pi(1)120\lambda^3 \\ \pi(6) &= \pi(5) \\ \pi(7) &= \pi(1)120\lambda^4\end{aligned}$$

Eftersom det är sannolikheter så vet vi att $\sum_{j=1}^7 \pi(j) = 1$, vilket ger:

$$\begin{aligned}\pi(1) + \pi(2) + \pi(3) + \pi(4) + \pi(5) + \pi(6) + \pi(7) &= \pi(1)(1 + 20\lambda + 120\lambda^2 + 240\lambda^3 + 120\lambda^4) = 1 \\ \Rightarrow \pi(1) &= \frac{1}{1 + 20\lambda + 120\lambda^2 + 240\lambda^3 + 120\lambda^4}.\end{aligned}$$

Den stationära fördelningen för systemet med fem atomer är därför:

$$[\pi(1), \pi(2), \pi(3), \pi(4), \pi(5), \pi(6), \pi(7)] = \pi(1)[1, 20\lambda, 60\lambda^2, 60\lambda^2, 120\lambda^3, 120\lambda^3, 240\lambda^4],$$

med $\pi(1)$ enligt ovan.

4.2 Generella system

För att undersöka molekylära systems beteende över tid för olika intensiteter i något verkligt scenario krävs att N , antalet atomer i systemet, blir betydligt större. En av svårigheterna i övergången från mindre till större antal atomer är att avgöra strukturen av ett sådant system. Eftersom antalet möjliga tillstånd växer hastigt med antalet atomer, se bilaga A.1, blir det snabbt mycket svårt att explicit bestämma tillstånden och hur de kommunicerar med varandra på samma sätt som användes för systemet med fem atomer. Det skulle därför vara önskvärt att finna en generell formel för att explicit kunna finna systemets asymptotiska fördelning. Enligt tidigare presenterad teori vet vi att ett tillräckligt villkor för att det molekylära systemet skall ha en asymptotisk fördelning är reversibilitet. Om vi kan visa att kedjan är reversibel för ett allmänt fall kan vi, med hjälp av Kolmogorovs kriterium 2.7, finna ett explicit uttryck för systemets stationära fördelning

och samtidigt kringgå att finna alla möjliga tillstånd och övergångar enligt tillvägagångssättet för systemet med fem atomer. Det visade sig i det föregående avsnittet att ett system med fem atomer är reversibelt. I detta avsnitt undersöks istället om systemet är reversibelt i det allmänna fallet, alltså för ett godtyckligt antal atomer. I samband med detta ska ett uttryck för den stationära fördelningen i det generella fallet tas fram.

4.2.1 Reversibilitet

Vi minns att balansekvationen 2.6 och Kolomogorovs kriterium 2.7 tillåter oss att uttrycka den stationära fördelningen av en reversibel Markovkedja enligt 2.8.

Eftersom egenskap 2.6 kan vara svår att kontrollera kan vi istället utveckla 2.8 och kontrollera om den detaljerade balansekvationen håller för den valda sekvensen i, i_1, \dots, i_n, j . Om så är fallet är Markovkedjan reversibel och $\pi(j)$ är dess stationära fördelning. Det generella systemet genereras med följande resultat.

Sats 4.1. *En Markovkedja som beskriver tidsutvecklingen av systemet med N atomer är reversibel. Dess stationära fördelning uppfyller, för $\mathbf{n}_0 = (N, 0, \dots, 0)$,*

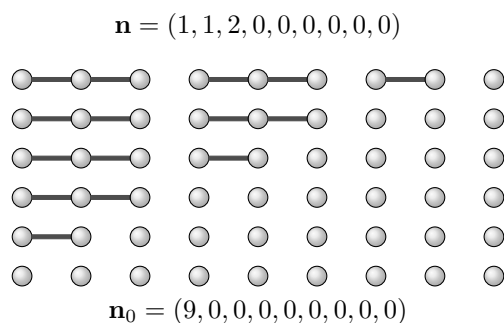
$$\pi(\mathbf{n}) = \pi(\mathbf{n}_0) \lambda^{L(\mathbf{n})} \frac{N!}{\prod_{k \geq 1} n_k!}, \quad (4.1)$$

där $L(\mathbf{n}) = \sum_{k \geq 2} (k-1)n_k$ är antalet länkar i tillstånd \mathbf{n}

Bevis. För att uttrycka intensiteterna mellan tillstånden från \mathbf{n}_0 till $\mathbf{n} = (n_1, \dots, n_N) \neq \mathbf{n}_0$ på formen (2.7) konstruerar vi följande väg. Notera att vi enligt tidigare presenterad teori i avsnitt 2.2 kan vi välja valfri sekvens av tillstånd.

Tag tillståndet $\mathbf{n}_0 = (N, 0, \dots, 0)$, där alla atomerna är separerade. Låt m vara längden av den längsta molekylerna i \mathbf{n} , det vill säga $n_{m+1} = \dots = n_N = 0$.

Först aggregerar atomerna en efter en för att bilda första m -molekylen i \mathbf{n} . Om $n_m > 1$, bildas de andra m -molekylerna en efter en till dess att antalet är m -molekyler är samma som i \mathbf{n} . I nästa steg bildas alla $m-1$ -molekyler genom samma manér som ovan. Sedan, på samma itererande vis, bildas alla molekyler med längd större än två i \mathbf{n} . Notera att den valda sekvensen endast sammanlänkar enskilda atomer med molekyler, det vill säga varje molekyl bildas genom att addera en atom i taget tills respektive storlek uppnås. I figur 8 ges ett illustrativt exempel för $N = 9$.



Figur 8: En illustration av hur vägen skulle kunna konstrueras enligt vår metodik då $N = 9$. Molekylerna är representerade med segment ordnade efter storlek.

Ovan har vi konstruerat vägen och nu behöver vi beräkna intensiteten av övergångarna mellan tillstånden.

Intensiteten för att den första länken skall bildas, det vill säga att två enskilda atomer bildar en 2-molekyl, är $\lambda N(N-1)$ eftersom det finns $\binom{N}{2}$ par av atomer i tillstånd $\mathbf{n}_0 = (N, 0, \dots, 0)$ och varje par bildas med intensitet 2λ . Nästa atom länkar samman med den nyligen bildade 2-molekylen med intensitet $2\lambda(N-2)$, nästa med intensitet $2\lambda(N-3)$ osv. Den sista atomen att slutföra den första m -molekylen länkar samman med intensitet $2\lambda(N-m+1)$.

Nu finns $N-m$ enskilda atomer samt en m -molekyl i systemet. Produkten av intensiteterna för att bilda den första m -molekylen är

$$2^{m-2}\lambda^{m-1}N(N-1)\cdots(N-m+1) = 2^{m-2}\lambda^{m-1}\frac{N!}{(N-m)!}.$$

Om $n_m > 1$ innebär det att intensiteten för nästa m -molekyl ges av uttrycket ovan med skillnaden att N byts ut mot $N-m$. Detta ger att intensiteten av stegen från \mathbf{n}_0 till att den andra m -molekylen har bildats ges av produkten

$$2^{m-2}\lambda^{m-1}\frac{N!}{(N-m)!} \times 2^{m-2}\lambda^{m-1}\frac{(N-m)!}{(N-2m)!} = 2^{2(m-2)}\lambda^{2(m-1)}\frac{N!}{(N-2m)!}.$$

Efter att alla n_m m -molekyler har bildats är produkten av intensiteterna

$$2^{(m-2)n_m}\lambda^{(m-1)n_m}\frac{N!}{(N-mn_m)!}.$$

Nu finns $(N-m)n_m$ atomer i systemet och konstruktionen av de n_{m-1} $(m-1)$ -molekylerna sker på samma sätt som för de n_m m -molekylerna. Proceduren itereras till dess alla molekyler av storlek $k \geq 3$ skapats. När alla k -molekyler för $k \geq 3$ är bildade finns $N' = N - \sum_{k \geq 3} kn_k$ lösa atomer kvar i systemet.

Om $n_2 > 0$ behöver vi beräkna intensiteten för att två atomer aggregerar. Den första 2-molekylen binder två atomer med intensitet $\lambda N'(N'-1)$, den andra med intensitet $\lambda(N'-2)(N'-3)$. Detta fortgår till dess att den sista molekylen av storlek två skapats med intensitet $\lambda(N'-2(n_2-1))(N'-2(n_2-1)-1)$. Produkten av intensiteter för alla övergångar hörande till skapandet av n_2 molekyler av storlek två blir då

$$\lambda^{n_2}\frac{N!}{(N'-2n_2)!} = \lambda^{n_2}\frac{(N - \sum_{k \geq 3} kn_k)!}{(n - \sum_{k \geq 2} kn_k)!} = \lambda^{n_2}\frac{(N - \sum_{k \geq 2} kn_k)!}{n_1!}.$$

Detta ger följande produkt av intensiteter för alla övergångar från \mathbf{n}_0 till \mathbf{n} :

$$2^{\sum_{k \geq 3} (k-2)n_k} \lambda^{\sum_{k \geq 3} (k-1)n_k} \frac{N!}{n_1!} = 2^{I(\mathbf{n})} \lambda^{L(\mathbf{n})} \frac{N!}{n_1!},$$

där $I(\mathbf{n}) = \sum_{k \geq 3} (k-2)n_k$ är antalet inre atomer, det vill säga antalet atomer som är sammanlänkade med två andra atomer i en molekyl och där $L(\mathbf{n}) = \sum_{k \geq 3} (k-1)n_k$ är antalet länkar.

Om vi istället vill beräkna intensiteten att förflytta sig från \mathbf{n} till \mathbf{n}_0 , börjar det med att en molekyl av storlek två separerar. Detta sker med intensitet n_2 eftersom det finns just en länk i varje 2-molekyl. Nästa molekyl av storlek två separerar med intensitet $n_2 - 1$, detta upprepas till dess att det endast finns en molekyl av storlek två kvar. Den sista separerar med intensitet 1. Detta ger att produkten av nedbrytningen av alla 2-molekyler är $n_2!$.

För $n_2 = 0$ är följande första steget. Om $n_3 \geq 1$ är nästa steg att en yttre länk i en molekyl av storlek tre bryts. Detta sker med intensitet $2n_3$, följt av att nästa bindning i molekylen separerar med intensitet 1 (då det är en molekyl molekyl av storlek två). Antalet molekyler av storlek tre i systemet har nu reducerats med en molekyl vilket ger, om $n_3 > 1$, att nästa 3-molekyl bryts ned med intensitet $2(n_3 - 1)$. Produkten av intensiteterna för övergångarna från \mathbf{n} till \mathbf{n}_0 , tillståndet endast bestående av lösa atomer, ges av $n_2!2^{n_3}n_3!$.

Efter att alla molekyler av storlek mindre än k har brutits ned till lösa atomer bryts första länken i en k -molekyl med intensitet $2n_k$. Nästa länk i samma molekyl bryts med intensitet 2 fram tills den sista länken bryts med intensitet 1. Detta är sant då vägen är konstruerad så att det endast är lösa atomer som separerar.

Alltså bidrar den första nedbrutna k -molekylen med $2^{k-2}n_k$ till produkten av intensiteter. På samma sätt bidrar den andra k -molekylen med $2^{k-2}(n_k - 1)$ till produkten av intensiteter och den sista k -molekylen bidrar med $2^{k-2}(n_k - (n_k) + 1)$. Produkten av alla intensiteter för alla steg från \mathbf{n} tillbaka till \mathbf{n}_0 ges av

$$\prod_{k \geq 2} 2^{n_k(k-2)} n_k! = 2^{I(\mathbf{n})} \prod_{k \geq 2} n_k!. \quad (4.2)$$

Detta tillåter oss att definiera uttrycket

$$\pi(\mathbf{n}) = \pi(\mathbf{n}_0) \frac{2^{I(\mathbf{n})} \lambda^{L(\mathbf{n})} \frac{N!}{n_1!}}{2^{I(\mathbf{n})} \prod_{k \geq 2} n_k!} = \pi(\mathbf{n}_0) \lambda^{L(\mathbf{n})} \frac{N!}{\prod_k n_k!}. \quad (4.3)$$

Nästa steg består av att verifiera att (4.3) uppfyller balansekvationen för övergångarna i systemet.

För övergången mellan \mathbf{n} och $\mathbf{n} + v_{jk}$, $j \neq k$, ges balansekvationen av

$$\pi(\mathbf{n}) 2\lambda n_j n_k = \pi(\mathbf{n} + v_{jk}) 2(n_{j+k} + 1).$$

Eftersom $L(\mathbf{n} + v_{jk}) = L(\mathbf{n}) + 1$ kan vi göra omskrivningen

$$\lambda^{L(\mathbf{n})} \frac{N!}{\prod_k n_k!} 2\lambda n_j n_k = \lambda^{L(\mathbf{n})+1} \frac{n_j n_k N!}{(n_{j+k} \prod_k n_k!)} 2(n_{j+k} + 1). \quad (4.4)$$

Vi konstaterar att likheten ovan håller. På samma sätt kan vi verifiera att balansekvationen för övergången mellan \mathbf{n} och $\mathbf{n} + v_{kk}$, $j = k$

$$\pi(\mathbf{n}) \lambda n_k (n_k - 1) = \pi(\mathbf{n} + v_{kk}) (n_{2k} + 1)$$

också håller. Därav följer, enligt (2.6), att Markovkedjan är reversibel och dess stationära fördelning ges av (4.1). \square

Från dessa resultat följer det att vi kan finna ett uttryck för frekvensfunktionen hörande till antal molekyler. Låt det totala antalet molekyler (inklusive enskilda atomer) i tillstånd \mathbf{n} ges av $M(\mathbf{n}) = \sum_k n_k$. Under den stationära fördelningen π ges frekvensfunktionen hörande till M av

$$\pi(M(\mathbf{n}) = m) = Z^{-1}(\lambda, N) \frac{\lambda^{-m}}{m!} \binom{N-1}{m-1}, \quad m = 1, \dots, N \quad (4.5)$$

där den normaliserande konstanten ges av

$$Z(\lambda, N) = \sum_{m=1}^N \frac{\lambda^{-m}}{m!} \binom{N-1}{m-1} = \frac{1}{\lambda N} L_{N-1}^1(-1/\lambda), \quad (4.6)$$

där $L_n^\alpha(x)$ är det generaliserade Laguerrepolynom^{*} av grad n . Fördelningen är unimodal^{*} med kärnan i punkten

$$m^* = \left\lceil \frac{-(1+\lambda) + \sqrt{(1+\lambda)^2 + 4}}{2\lambda} \right\rceil \sim O(\sqrt{N/\lambda}) \text{ då } N \rightarrow \infty. \quad (4.7)$$

Dess Laplacetransformen^{*} ges av

$$\mathcal{L}_M(t) = \mathbf{E}[e^{tM}] = \frac{Z(e^t \lambda, N)}{e^t Z(\lambda, N)} = \frac{e^{-t} L_{N-1}^1(-e^{-t}/\lambda)}{L_{N-1}^1(-1/\lambda)}, \quad (4.8)$$

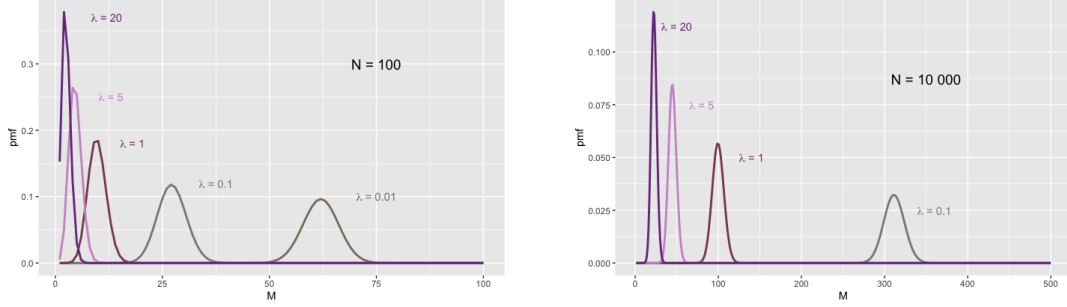
^{*}För definition se ordlista i avsnitt A.2

^{*}För definition se ordlista i avsnitt A.2

de två första momenten ges av

$$\mathbf{E}[M] = 1 + \frac{l_2}{\lambda l_1}; \quad \mathbf{var}[M] = \frac{l_2}{\lambda l_1} + \frac{l_1 l_3 - l_2^2}{\lambda^2 l_1^2}, \quad (4.9)$$

där $l_k = L_{N-k}^k(-1/\lambda)$, $k = 1, 2, 3$. Frekvensfunktionen av M antal molekyler under den stationära fördelningen π visas i Figur 9.



(a) Frekvensfunktionen för antal molekyler då $N=100$. (b) Frekvensfunktionen för antal molekyler då $N=10.000$.

Figur 9: Figureerna visar frekvensfunktionerna för två olika antal atomer samt för olika värden på λ .

Uttrycket (4.5), för frekvensfunktionen, kan motiveras enligt följande. Eftersom $\sum_k kn_k = N$ ger (4.1) att

$$\pi(M(\mathbf{n}) = m) = \pi(\mathbf{n}_0) \lambda^{N-m} \sum_{\substack{\mathbf{n}: \sum_k n_k = m \\ \sum_k kn_k = N}} \frac{N!}{\prod_k n_k!}. \quad (4.10)$$

Antag att systemet innehåller exakt m molekyler. Numrera dem från 1 till m och låt (l_1, \dots, l_m) vara deras storlekar. Om vi representerar varje k -molekyl med ett sammanlänkat linjärt segment av längd $k - 1$, då kan vi placera molekylerna från höger till vänster på samma sätt som i Figur 6 med skillnaden att de inte är ordade i storlekordning, utan efter deras numrering.

Eftersom $\sum_k kn_k = N$, kommer alla punkter innehållna i $[1, N]$ vara täckta och det finns exakt $m - 1$ tommasegment på formen $(i, i + 1)$ som separerar molekyler. Därför kan alla sekvenser (l_1, \dots, l_m) av molekylernas storlek av ordnade molekyler fås genom att extrahera $m - 1$ "öppnasegment av totalt $N - 1$ av $[1, N] \setminus \mathbb{N}$, alltså är det $N - 1 - 1$ av dessa.

Alternativt, kan sekvenserna (l_1, \dots, l_m) , $l_i \geq 1$ som uppfyller $\sum_{i=1}^m l_i = N$ fås av $\mathbf{n} = (n_1, n_2, \dots)$ som uppfyller $\sum_k n_k = m$ samt $\sum_k kn_k = N$ genom att tillskriva n_1 ettor, n_2 tvåor, etc. för platser 1, ..., m så att varje \mathbf{n} genererar det multinomiala talet $m! / \prod_k n_k!$ av sådana distinkta sekvenser.

Därför har vi att

$$\sum_{\substack{\mathbf{n}: \sum_k n_k = m \\ \sum_k kn_k = N}} \frac{m!}{\prod_k n_k!} = \binom{N-1}{m-1},$$

vilket ger

$$\pi(M(\mathbf{n}) = m) = \pi(\mathbf{n}_0) \lambda^{N-m} \frac{N!}{m!} \binom{N-1}{m-1}.$$

Det normaliserande villkoret $\sum_{m=1}^N \pi(M(\mathbf{n}) = m)$ tillåter oss att skriva

$$\pi(\mathbf{n}_0) = \left[\sum_{m=1}^N \lambda^{N-m} \frac{N!}{m!} \binom{N-1}{m-1} \right]^{-1}, \quad (4.11)$$

vilket ger, för $M(n)$, (4.7) efter att ha strukit $N!\lambda^N$.

Notera att

$$\pi(M(\mathbf{n}) = m + 1) = \frac{N - m}{\lambda m(m + 1)} \pi(M(\mathbf{n}) = m).$$

Den multiplikativa koefficienten som ges av bråket är monotont avtagande och blir mindre än 1 då $m = m^*$ som i (4.7). Därav är fördelningen unimodal. Resonemanget som förts ovan motiverar följande sats.

Sats 4.2. *Den stationära fördelningen π ges av*

$$\pi(\mathbf{n}) = Z^{-1}(\lambda, N) \frac{\lambda^{-M(\mathbf{n})}}{\prod_k n_k!} \quad (4.12)$$

med $Z(\lambda, N)$ som i (4.6). Den villkorliga fördelningen av molekylernas längder givet deras totala antal är

$$\pi(\mathbf{n} | M(\mathbf{n}) = m) = \frac{m!}{\binom{N-1}{m-1} \prod_k n_k!} = \frac{\binom{m}{n_1, \dots, n_N}}{\binom{N-1}{m-1}} \quad (4.13)$$

vilket är den symmetriska multinomiala fördelningen $\mathbf{Mult}(m; N^{-1}, \dots, N^{-1})$ villkorligt på mängden $\{\mathbf{n} : \sum_k n_k = N\}$.

4.2.2 Asymptotisk fördelning

Tidigare har vi uttryckt Laplacetransformen av den stationära fördelningen av det totala antalet $M = M_n$ molekyler i systemet med N atomer i termer av Laguerrepolytom. I [14], erhöles följande asymptotiska fördelning i n för fixa parametrar α och z

$$L_n^\alpha(-z) = \frac{e^{-z/2}}{2\sqrt{\pi}} \frac{e^{2\text{sqr}t(n+1)z}}{z^{1/4+n/2}(n+1)^{1/4-\alpha/2}}. \quad (4.14)$$

Substitution av detta uttryck till 4.8 med $n = N - 1$ och $z = 1/\lambda$, ger

$$\mathcal{L}_{M_s}(t) = \exp\left\{-\frac{t}{4} - \frac{z}{2}(e^{-t} - 1)2\sqrt{Nz}(e^{-t/2} - 1)\right\} \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right). \quad (4.15)$$

Vi introducerar $a_N = \sqrt{Nz} = \sqrt{N/\lambda}$ och låter

$$\mu_N = \frac{M_N - a_N}{\sqrt{a_N/2}}$$

För Laplacetransform av μ_N har vi identiteten

$$\mathcal{L}_{\mu_N}(t) = e^{t\sqrt{2a_N}} \mathcal{L}_{m_N}\left(\frac{t}{\sqrt{a_N/2}}\right) \quad (4.16)$$

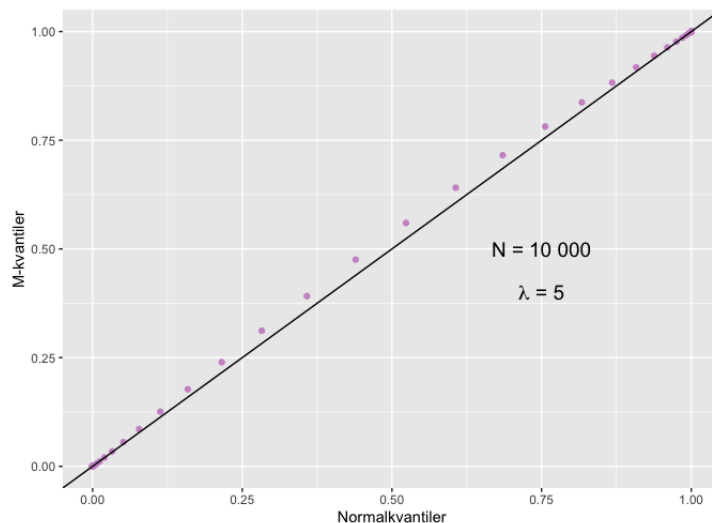
vidare blir det enkelt att kontrollera, via 4.14, att

$$\lim_{N \rightarrow \infty} \mathcal{L}_{\mu_N}(t) = t^2/2.$$

Således har vi bevisat den följande Centrala gränsvärdesatsen.

Sats 4.3. *Den stationära fördelningen av M_N antal molekyler i ett system med N atomer, då systemet är korrekt centrerat och skalat, konvergerar svagt till den standardiserade normalfördelningen*

$$\frac{M_N - a_N}{\sqrt{a_N/2}} \rightarrow \mathcal{N}(0, 1). \quad (4.17)$$



Figur 10: En QQ-plot med frekvensfunktionen för ett system där $N = 10\,000$ och $\lambda = 5$ tillsammans med en normalfördelning med medelvärde a_N och standardavvikelse $\sqrt{a_N/2}$.

Då punkterna i figur 10 efterliknar en rät linje stärker det tidigare resultatet att den stationära fördelningen av M_N antal molekyler är normalfördelad.

4.3 Simulering

Resultaten som har presenterats visar att fördelningen av ett, exempelvis molekyllärt, system av sammanlänkade kedjor av varierande storlek förväntas följa ett visst förändringsmönster. Detta givet att systemet förändras enligt tidigare introducerade egenskaper gällande aggregation och separation. För att uppnå någon grad av verklighetsförankring kring detta samt att i någon mån kunna 'testa' våra teoretiska resultat utförs simuleringar i R.

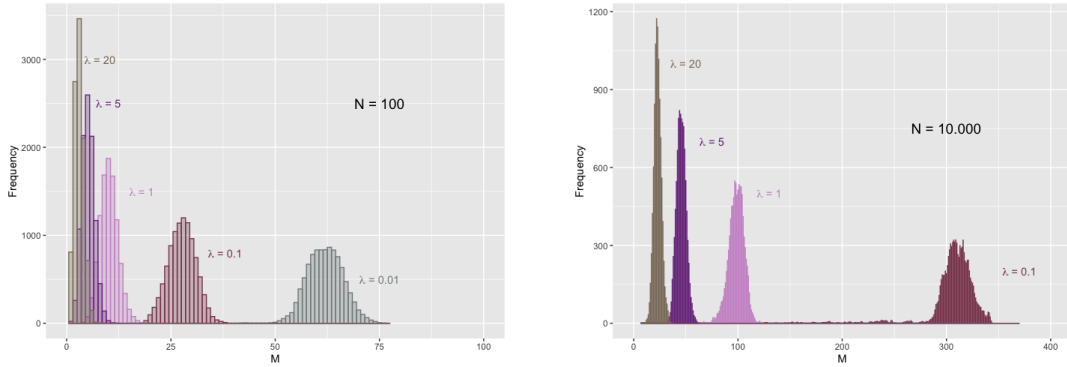
Initialt skapas en vektor, `system`, som har N , systemets atomantal, antal platser. Vidare simuleras ett initialt system genom att slumpmässigt välja ett index `i_1` uniformt fördelat mellan 1 och N . Vidare ökar plats `i_1` i `system` med ett. Systemet har nu en molekyl av storlek `i_1`. Nu dras ett nytt index `i_2` mellan 1 och $N - i_1$ och nu ökar plats `i_2` i `system` med ett. Nu har systemet två molekyler, en av storlek `i_1` samt en av storlek `i_2`. Denna processen fortsätter tills det sammanlagda antalet 'använda' atomer är lika med N . Nu har vi skapat ett initialt system.

Simuleringen av systemets beteende gällande aggregation och separation kommer ske med avseende på $\lambda = 0.01, 0.1, 1, 5, 20$. Dessa värden är samlade i vektorn `lambda`. Vidare används först en `while`-loop och sedan en `for`-loop över dessa värden på `lambda`. I varje iteration över `lambda` utförs en simulering för att uppnå systemets stationära läge. Simuleringen inleds med en `while`-loop. Inledningsvis beräknas intensiteten för aggregation enligt

$$\text{prob_agg} = \frac{2 * \text{lambda}[i] * (\text{Antal möjliga molekyllpar})}{2 * \text{lambda}[i] * (\text{Antal möjliga molekyllpar}) + (\text{Antalet länkar i systemet})}$$

För att besluta om aggregation eller separation skall ske slumpas ytterligare ett tal, `rand`, uniformt fördelat mellan 0 och 1. Om `rand < prob_agg` sker aggregation och analogt sker separation om `rand ≥ prob_agg`. Vid aggregation väljer vi två molekyler, `i` och `j`, slumpmässigt och binder samman dessa genom att `system[i+j]` ökar med ett samt att både `system[i]` och `system[j]` minskar med ett. Om separation skall ske väljer vi en molekyl slumpmässigt, säg en molekyl av storlek `k`, och sedan en av dess länkar, säg länk `j ∈ (1, ..., k-1)` med lika stor sannolikhet och bryter den. Detta gör att `system[k]` minskar med ett samt att `system[j-1]` och `system[k-j]` ökar med ett. När aggregation alternativt separation har skett sparas det nuvarande antalet molekyler i systemet tillsammans med det totala medelvärdet av molekyllernas längd över alla iterationer. Denna loop

körs fram tills dess att en ny iteration inte förändrar medelvärdet av molekylernas längd mer än en vald liten skillnad. Systemet anses nu ha uppnått, eller åtminstone vara nära sitt stationära läge. Nu körs `for`-loopen enligt exakt samma procedur som för den tidigare loopen med skillnaden att vi istället sparar antal molekyler i systemet för varje iteration. Sedan visualiseras hur många gånger systemet har innehållit ett visst antal molekyler med hjälp av ett histogram, se figur 11.

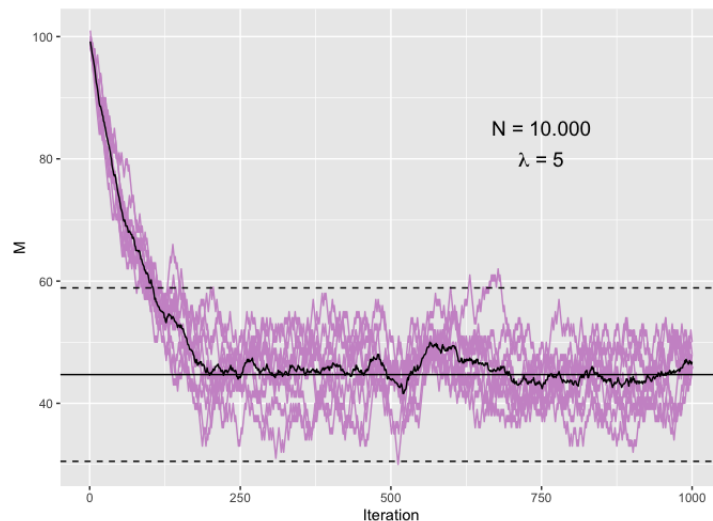


(a) Frekvens av antal molekyler då $N = 100$.

(b) Frekvens av antal molekyler då $N = 10000$.

Figur 11: Figurerna visar frekvenserna för hur många gånger systemet har innehållit ett visst antal atomer för två olika värden på N .

Vi kommer ihåg figur 9 och jämför med figur 11 och noterar att simuleringarna efterliknar frekvensfunktionerna genererade via teorin i avsnitt 4.2.1. Vidare stärker detta även de teoretiska resultaten kring att systemet har en asymptotisk normalfördelning. Nedan, i figur 12, visualiseras hur utvecklingen av antalet molekyler i ett system kan se ut. I figur 12 består systemet av $N = 10.000$ atomer och har initialt 100 molekyler, $\lambda = 5$. Vi noterar att när systemet har mellan lite knappt 40 och lite drygt 50 molekyler har det nått sitt stationära läge. Vi har även att $a_N = \sqrt{N/\lambda} \approx 44,7$ vilket är medelvärdet för frekvensfunktionen hörande till ett system med $N = 10000$ atomer och $\lambda = 5$. Standardavvikelsen hörande till detta fall är $\sqrt{a_N} \approx 4.7$. Vi noterar även att visualiseringen i figur 12 följer i enlighet med detta.



Figur 12: Figuren visar utvecklingen för 20 system med initialt 10 000 atomer fördelat över 100 molekyler över 20 iterationer där $\lambda = 5$. Den svarta grafen visar medelvärdet av de 20 iterationerna. Den svarta heldragna horisontella linjen motsvarar medelvärdet för frekvensfunktionerna. De streckade linjerna motsvarar tre standardavvikelser från medelvärdet.

5 Slutsats

Syftet med detta arbete var att med hjälp av Markovteori studera ett molekylärt systems utveckling på lång sikt. Sergei Zuyev som handledde arbetet hade en föraning om att systemet som definierat i avsnitt 3 skulle vara reversibelt. Detta arbete bekräftar denna hypotes. Genom konstruktionen av en väg mellan olika tillstånd i avsnitt 4.2.1 kunde en stationär fördelning som uppfyllde balansekvationen, sats 2.6, uttryckas. Därmed är systemet reversibelt och den första frågeställningen besvarad.

Den andra frågeställningen berör egenskaper hos systemet efter att den stationära fördelningen inställt sig. Utifrån den stationära fördelningen undersöktes två asymptotiska egenskaper. Den första är molekylernas längd som visade sig kunna beskrivas med den symmetriska multinomiala fördelningen givet deras antal. Den andra är fördelningen av antalet molekyler som med den centrala gränsvärdessatsen visade sig konvergera svagt mot den standardiserade normalfördelningen. Med detta är även den andra frågeställningen besvarad.

Simuleringarna som genomfördes under arbetet ger stöd åt att antalet molekyler är normalfördelat genom att visa på en liknande fördelning efter lång tid. Detta kan utläsas från jämförelsen av den teoretiska frekvensfunktionen i figur 9 med de simulerade frekvenserna i figur 11. Utöver det finns en formell jämförelse av kvantilerna för frekvensfunktion med kvantilerna för den standardiserade normalfördelningen i QQ-plotten i figur 10. Det är värt att notera att fördelningarna för antal och längd på molekylerna är beroende av att den stationära fördelningen inträffat. För detta ändamål ges i figur 12 utvecklingen av systemet, sett i antalet molekyler. När frekvenserna för dessa stämmer överens med den teoretiska fördelningen ger det en indikation på att stationäritet uppnåtts.

Läsaren har hänvisats till bilden av ett molekylärt system eftersom det var därifrån inspirationen hämtades. Även om det är där tillämpningen är tydligast kan det finnas andra system med liknande dynamik där denna modell är tillämpbar. Resultaten av detta arbete är därmed generella i bemärkelsen att systemet är öppet för tolkning samtidigt som de är strikt begränsade av modellformuleringen.

Mot slutet av arbetet blev vi medvetna om att problemformulering hade mycket gemensamt med ett system som behandlas i boken *Reversibility and Stochastic Networks* av matematikern Frank Kelly från 1979 [11]. Kelly gör samma koppling till molekylära system men med en mer generell modell som även tillåter förgreningar av molekyler likt exemplet som presenteras på vänster sida i figur 3. Vidare presenterar han dess stationära fördelning. Det som detta arbete tillför är ett explicit uttryck för denna modell samt en undersökning av dess asymptotiska egenskaper. Detta sker genom analysen av molekylernas längd och antal vid stationäritet, samt de simuleringar som ger stöd åt dessa.

Avslutningsvis vill vi nämna några fördjupningsmöjligheter för att gå vidare med detta arbete. Det hade varit intressant att uppnå skarpare resultat gällande antalet molekylers konvergens mot en normalfördelning. Detta visades med den centrala gränsvärdessatsen men det hade varit önskvärt att även använda *local limit theorem*. Det hade också varit intressant att kvantifiera tiden det tar för Markovkedjan att närma sig sin stationära fördelning, det vill säga dess *mixing time*. Studien hade även kunnat utvidgas genom att anpassa den idealiserade modellen efter verkliga förutsättningar. Förslagsvis hade den kunnat tillåta utomstående påverkan i form av atomer som tillkommer och bortfaller. Alternativt hade modellen kunnat utvecklas för att tillåta ett tidsberoende hos intensiteterna.

Referenser

- [1] Berryman, S. Leucippus, The Stanford Encyclopedia of Philosophy [Internet] Stanford: Stanford University; 2016 [uppdaterad 02-12-2016; citerad 2020-05-06]. Hämtad från:<https://plato.stanford.edu/archives/win2016/entries/leucippus/>.
- [2] Principe, L. The Scientific Revolution: A Very Short Introduction. Oxford: Oxford University Press; 2011
- [3] Lundgren, A. Atomlära [Internet], Stockholm: Nationalencyklopedin; [citerad 2020-05-06]. Hämtad från:<http://www.ne.se/uppslagsverk/encyklopedi/lång/atomlära>
- [4] Tansjö L. John Dalton [Internet], Stockholm: Nationalencyklopedin; [citerad 2020-05-06]. Hämtad från: <http://www.ne.se/uppslagsverk/encyklopedi/lång/john-dalton>
- [5] Tansjö L. Klassisk kemi [Internet], Stockholm: Nationalencyklopedin; [citerad 2020-05-06]. Hämtad från: <http://www.ne.se/uppslagsverk/encyklopedi/lång/kemi/klassisk-kemi>
- [6] Markov, AA. Investigation of a noteworthy case of dependent trials. *Izv Akad Nauk Ser Biol* 1; 1907
- [7] Gut, A. Markovkedja [Internet], Stockholm: Nationalencyklopedin [citerad 2020-03-02]. Hämtad från: <http://www.ne.se/uppslagsverk/encyklopedi/lång/markovkedja>
- [8] Lange, F. Materialismens historia: jämte en kritik av dess betydelse i våra dagar. Stockholm: Bonnier; 1913
- [9] Helenius, O. Mouwitz, L. Matematiken – var finns den? [Internet] Göteborg: Nationellt centrum för matematikutbildning; 2009 [citerad 2020-05-06]. Hämtad från:http://ncm.gu.se/media/ncm/dokument/matematik_var finns den.pdf
- [10] Grimmett, G. Stirzaker, D. Probability and Random Processes. 3:e upplagan. New York: Oxford University Press. 2001.
- [11] Kelly F.P. Reversibility and Stochastic Networks. Cambridge: Cambridge University Press. 1979.
- [12] Freedman, A. Convergence theorem for finite time Markov chains [Internet]; [citerad 2020-05-14] Hämtad från: <https://math.uchicago.edu/~may/REU2017/REUPapers/Freedman.pdf>
- [13] P. Diaconis and L. Miclo. On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse Math.* (6), 24(4):973–1016, 2015.
- [14] D. Borwein, M. Borwein, and R.E. Crandall. Effective Laguerre asymptotics. *SIAM J. Numer. Anal.*, 46(6):3285–3312, 2008.
- [15] Sequence Database. Sequence dps4an1f0bv5f [Internet]; [citerad 2020-05-11] Hämtad från: <http://sequencedb.net/s/dps4an1f0bv5f>
- [16] Rusev, P. Classical orthogonal polynomials and their associated functions in complex domain. Sofia: Mann Drinov Academic Publishing House. 2005.
- [17] Encyclopedia of math. Unimodal fördelning [Internet]; [citerad 2020-05-11] Hämtad från: https://www.encyclopediaofmath.org/index.php/Unimodal_distribution
- [18] Tandf online. Laplace transformation [Internet]; [citerad 2020-05-11] Hämtad från: <https://www.tandfonline.com/doi/pdf/10.1080/0020739720030104?needAccess=true>

A Bilagor

A.1 Vilka tillstånd är möjliga?

En av utmaningarna med att behandla dessa system av molekyler är att de tycks växa hastigt i storlek. Beräkningar för ett litet system som det med fem atomer kan göras för hand. Utökas systemet med några atomer blir det betydligt mindre överskådligt. För att kunna rita upp Markovkedjan måste de möjliga tillstånden först definieras. Därefter behöver det avgöras mellan vilka av tillstånden det finns giltiga övergångar. För att sedan kontrollera reversibilitet med Kolmogorovs kriterium måste cyklerna kartläggas vilket gör det komplicerat. För varje system med N atomer gäller att det kan representeras med en vektor med antalet molekyler av respektive storlek som element. De två extremfallen är en vektor med N ensamma atomer och en vektor med en enda molekyl av längd N .

En annan representation av systemet är som en vektor av längd $N - 1$ med nollor och ettor. Det j :te elementet i vektorn avgör kopplingen mellan de numrerade atomerna j och $j + 1$, där en etta anger en länk och en nolla avsaknaden av länk. Till exempel skulle tillståndet $\mathbf{n}_4 = (2, 0, 1, 0, 0)$ för ett system med fem atomer kunna skrivas $(1, 1, 0, 0)$, $(0, 0, 1, 1)$ eller $(0, 1, 1, 0)$. Atomerna är i det här fallet numrerade vilket tillåter flera möjliga sätt att representera ett och samma system. Vi kan enkelt erhålla antalet permutationer av systemet enligt $p = 2^{N-1}$, eftersom vektorn har $N - 1$ element som vardera har två möjligheter: etta eller nolla.

I systemet som behandlas i rapporten är endast antalet kombinationer intressant eftersom atomerna betraktas som onummerade. När en aggregation sker är storlekarna på de två molekylerna som binds samman det enda av intresse. Det visar sig att antalet tillstånd som ett system av denna typ kan finna sig i är ekvivalent med antalet heltalspartitioner i talteori. Där anges hur många sätt ett tal $N \in \mathbb{Z}^+$ kan skrivas som en summa av positiva heltal utan att ta hänsyn till termernas inbördes ordning. Till exempel kan talet 5 partitioneras på sju sätt:

$$5 = 1 + 1 + 1 + 1 + 1$$

$$5 = 2 + 1 + 1 + 1$$

$$5 = 2 + 2 + 1$$

$$5 = 3 + 1 + 1$$

$$5 = 3 + 2$$

$$5 = 4 + 1$$

$$5 = 5.$$

Dessa sju partitioner motsvarar precis vårt tillståndsrum $S_{N=5}$, där $(2 + 1 + 1 + 1)$ kan utläsas som en molekyl med två atomer och tre ensamma atomer. Det finns en färdig talföljd (A000041 i OEIS) från vilken vi kan utläsa att ett system med 13 atomer har 101 möjliga tillstånd medan ett system med 100 atomer har drygt 190 miljoner tillstånd [15]. Eftersom 100 är ett förhållandevis litet tal i kemisammanhang skulle övergångsmatriserna, som har lika många element som antalet tillstånd i kvadrat, bli ohanterligt stora för detta arbete.

A.2 Ordlista

Generaliserade Laguerrepolyinom

$$L_n^\alpha(x) = \sum_{i=0}^n (-1)^i \binom{n+\alpha}{n-i} \frac{x^i}{i!},$$

med paramater α som är ett reellt tal, $n = 0, 1, 2, \dots$, och $0 \leq x < \infty$ [16].

Unimodal fördelning

En sannolikhetsfördelning som endast har ett lokalt maximum [17].

Laplacetransform

Låt X vara en icke-negativ absolut kontinuerlig slumpvariabel med pdf f . Laplacetransformen \mathcal{L}_X av X definieras av

$$\mathcal{L}_X(s)E[e^{-sX}] = \int_0^\infty e^{-sx} f(x) dx, \quad s \geq 0.$$

[18]

Svag konvergens

Konvergens i en fördelning kan benämnas som svag konvergens och kan definieras som $E[f(X_n)] \rightarrow E[f(x)]$ för alla begränsade kontinuerliga funktioner f [10].

A.3 Kod

Kod för simulering.

```
N<-10000 #Number of atoms
lambda<-c(0.1,1,5,20)
loopsize<-10000
#creating a system, place 1= number of single atoms, place 2= number of 2-molecules etc..
system<- replicate(N,0)
n<-N
while(n>0){
  size<-round(runif(1,1,n))
  system[size]<-system[size]+1
  n<-n-size
}

mean_matrix<-matrix(0,nrow=loopsize,ncol = length(lambda))
#updating the system
for (l in 1:4) {
  system_copy<-system
  num_of_molecules<-sum(system_copy)
  mean_dif<-num_of_molecules
  i<-1
  while (mean_dif>=num_of_molecules*0.2) {#a loop for finding stationarity isch
    rand<-runif(1,0,1)
    prob_agg<-2*lambda[l]*sum(1:(sum(system_copy)-1))/(2*lambda[l]*sum(1:(sum(system_copy)-1))+
      Num_Of_Links(system_copy))

    if(length(which(system_copy!=0))==1 && which(system_copy!=0)==N){
      prob_agg<-0
    }
    if(rand<prob_agg){#aggregation
      nonzero_places<-which(system_copy!=0)
      if(length(nonzero_places)==1 && nonzero_places==1){
        system_copy[1]<-system_copy[1]-2
        system_copy[2]<-system_copy[2]+1
      }
      if(length(nonzero_places)==1 && nonzero_places!=1){
        system_copy[N]<-system_copy[N]+1
        system_copy[nonzero_places]<-0
      }
      if(length(nonzero_places)>1){
        #possible<-which(system_copy!=0)
        rand_to_bind<-sample(GetSampleVec(nonzero_places,system_copy),2)
        system_copy[rand_to_bind[1]]<-system_copy[rand_to_bind[1]]-1
        system_copy[rand_to_bind[2]]<-system_copy[rand_to_bind[2]]-1
        system_copy[sum(rand_to_bind)]<-system_copy[sum(rand_to_bind)]+1
      }
    }
  }
  if(rand>=prob_agg){#breakage
    possible<-GetSampleVec(which(system_copy!=0),system_copy)
    if(length(possible)==1){
      rand_to_break<-possible
    }
  }
}
```

```

    if(length(possible)>1){
      rand_to_break<-sample(possible,1)
    }
    while(rand_to_break==1){
      rand_to_break<-sample(possible,1)
    }
    break_size<-round(runif(1,1,rand_to_break-1))
    system_copy[break_size]<-system_copy[break_size]+1
    system_copy[rand_to_break-(break_size)]<-system_copy[rand_to_break-(break_size)]+1
    system_copy[rand_to_break]<-system_copy[rand_to_break]-1
  }
  new_num_of_molecules<-sum(system_copy)
  mean_dif<-abs(num_of_molecules-new_num_of_molecules)
  num_of_molecules<-new_num_of_molecules
}

for (i in 1:loopszie) {#a loop for collecting means
  rand<-runif(1,0,1)
  prob_agg<-2*lambda[1]*sum(1:(sum(system_copy)-1))/(2*lambda[1]*sum(1:(sum(system_copy)-1))+
  Num_Of_Links(system_copy))

  if(length(which(system_copy!=0))==1 && which(system_copy!=0)==N){
    prob_agg<-0
  }
  if(rand<prob_agg){#aggregation
    nonzero_places<-which(system_copy!=0)
    if(length(nonzero_places)==1 && nonzero_places==1){
      system_copy[1]<-system_copy[1]-2
      system_copy[2]<-system_copy[2]+1
    }
    if(length(nonzero_places)==1 && nonzero_places!=1){
      system_copy[N]<-system_copy[N]+1
      system_copy[nonzero_places]<-0
    }
    if(length(nonzero_places)>1){
      rand_to_bind<-sample(GetSampleVec(nonzero_places,system_copy),2)
      system_copy[rand_to_bind[1]]<-system_copy[rand_to_bind[1]]-1
      system_copy[rand_to_bind[2]]<-system_copy[rand_to_bind[2]]-1
      system_copy[sum(rand_to_bind)]<-system_copy[sum(rand_to_bind)]+1
    }
  }
  if(rand>=prob_agg){#breakage
    possible<-GetSampleVec(which(system_copy!=0),system_copy)
    if(length(possible)==1){
      rand_to_break<-possible
    }
    if(length(possible)>1){
      rand_to_break<-sample(possible,1)
    }
    while(rand_to_break==1){
      rand_to_break<-sample(possible,1)
    }
    break_size<-round(runif(1,1,rand_to_break-1))
    system_copy[break_size]<-system_copy[break_size]+1
  }
}

```

```

    system_copy[rand_to_break-(break_size)]<-system_copy[rand_to_break-(break_size)]+1
    system_copy[rand_to_break]<-system_copy[rand_to_break]-1
  }
  mean_matrix[i,1]<-sum(system_copy)
}
}

newdf<-as.data.frame(mean_matrix)

ggplot() + coord_cartesian(xlim = c(0,400))+
  #geom_histogram(data = newdf,aes(x=V1) , color = "azure4", fill=("azure4"),
    binwidth = 1, alpha=.3) +
  geom_histogram(data = newdf, aes(x = V1) ,color = "palevioletred4", fill=("palevioletred4"),
    binwidth = 0.1, alpha=.3) +
  geom_histogram(data = newdf, aes(x = V2) ,color = "plum3", fill=("plum3"),
    binwidth = 0.1, alpha=.3) +
  geom_histogram(data = newdf, aes(x = V3) ,color = "mediumorchid4", fill=("mediumorchid4"),
    binwidth = 0.1, alpha=.3)+
  geom_histogram(data = newdf, aes(x = V4) ,color = "bisque4",fill=("bisque4"),
    binwidth = 0.1, alpha=.3)+
  #annotate("text", x = 75, y = 500, label = expression(paste(lambda," = 0.01")),color="azure4")+
  annotate("text", x = 370, y = 200, label = expression(paste(lambda," = 0.1")),
    color="palevioletred4") +
  annotate("text", x = 130, y = 500, label = expression(paste(lambda," = 1")),
    color="plum3") +
  annotate("text", x = 75, y = 700, label = expression(paste(lambda," = 5")),
    color="mediumorchid4") +
  annotate("text", x = 50, y = 1000, label = expression(paste(lambda," = 20")),color="bisque4") +
  annotate("text", x = 300, y = 750, label = "N = 10.000",size=5) +
  xlab("M") +
  ylab("Frequency")

#####FUNCTIONS#####
Num_Of_Links<-function(system){
  links<-0
  for (i in 1:length(system)) {
    links<-links+system[i]*(i-1)
  }
  return(links)
}

GetSampleVec<-function(nonzero_places, system_copy){
  sample_vec<-0
  for (i in 1:length(nonzero_places)) {
    sample_vec<-append(sample_vec,replicate(system_copy[nonzero_places[i]],nonzero_places[i]))
  }
  return(sample_vec[-1])
}

GetData<-function(mean_matrix){
  dataframe<-matrix(0,nrow = nrow(mean_matrix),ncol = 2*ncol(mean_matrix))

```



```

i<-1
j<-1
while (i<ncol(dataframe)) {
  prob<-Get_Prob(mean_matrix[,j])
  dataframe[,i]<-prob[,1]
  dataframe[,i+1]<-prob[,2]
  i<-i+2
  j<-j+1
}
df<-as.data.frame(dataframe)
return(df)
}

Get_Prob<-function(column){
  mat<-matrix(0,nrow = length(column), ncol = 2)
  column<-round(column,digits = 3)
  prob<-replicate(length(column), 0)
  for (i in 1:length(column)) {
    prob[i]<-length(which(column==column[i]))/length(column)
  }
  mat[,1]<-column
  mat[,2]<-prob
  return(mat)
}

Create_Start_System<-function(system, rand_start){
  molecules<-replicate(rand_start,0)
  for (i in 1:length(system)) {
    rand_place<-sample(1:rand_start,1)
    molecules[rand_place]<-molecules[rand_place]+1
  }
  zeros<-which(molecules==0)
  if(length(zeros)>0){
    for (k in 1:length(zeros)) {
      giver<-sample(which(molecules>1),1)
      molecules[zeros[k]]<-molecules[zeros[k]]+1
      molecules[giver]<-molecules[giver]-1
    }
  }
  for (j in 1:length(molecules)) {
    system[molecules[j]]<-system[molecules[j]]+1
  }
  return(system)
}

Get_Mean<-function(mat){
  add<-replicate(nrow(mat),0)
  for (r in 1:nrow(mat)) {
    sum<-0
    for (c in 1:ncol(mat)) {
      sum<-sum+mat[r,c]
    }
    add[r]<-sum/ncol(mat)
  }
}

```

```

mat_new<-cbind(mat,add)
return(mat_new)
}

#####for plotting expected-time#####
set.seed(20)
N=10000
ntimes<-10
length_matrix<-matrix(0,nrow = 1000,ncol =ntimes )
system<-replicate(N,0)
rand_start<-100
system<-Create_Start_System(system,rand_start)
for (l in 1:ntimes) {
  system_copy<-system
  #system_copy[1]<-N
  #length_matrix[1,1]<-N
  for (i in 1:nrow(length_matrix)){
    rand<-runif(1,0,1)
    prob_agg<-2*lambda[3]*sum(1:(sum(system_copy)-1))/(2*lambda[3]*sum(1:(sum(system_copy)-1))+
      Num_Of_Links(system_copy))

    if(length(which(system_copy!=0))==1 && which(system_copy!=0)==N){
      prob_agg<-0
    }
    if(rand<prob_agg){#aggregation
      nonzero_places<-which(system_copy!=0)
      if(length(nonzero_places)==1 && nonzero_places==1){
        system_copy[1]<-system_copy[1]-2
        system_copy[2]<-system_copy[2]+1
      }
      if(length(nonzero_places)==1 && nonzero_places!=1){
        system_copy[N]<-system_copy[N]+1
        system_copy[nonzero_places]<-0
      }
      if(length(nonzero_places)>1){
        #possible<-which(system_copy!=0)
        rand_to_bind<-sample(GetSampleVec(nonzero_places,system_copy),2)
        system_copy[rand_to_bind[1]]<-system_copy[rand_to_bind[1]]-1
        system_copy[rand_to_bind[2]]<-system_copy[rand_to_bind[2]]-1
        system_copy[sum(rand_to_bind)]<-system_copy[sum(rand_to_bind)]+1
      }
    }
  }
  if(rand>=prob_agg){#breakage
    possible<-GetSampleVec(which(system_copy!=0),system_copy)
    if(length(possible)==1){
      rand_to_break<-possible
    }
    if(length(possible)>1){
      rand_to_break<-sample(possible,1)
    }
    while(rand_to_break==1){
      rand_to_break<-sample(possible,1)
    }
    break_size<-round(runif(1,1,rand_to_break-1))
    system_copy[break_size]<-system_copy[break_size]+1
  }
}

```

```

        system_copy[rand_to_break-(break_size)]<-system_copy[rand_to_break-(break_size)]+1
        system_copy[rand_to_break]<-system_copy[rand_to_break] -1
    }
    length_matrix[i,1]<-sum(system_copy)
}
}
added_matrix<-Get_Mean(length_matrix)
lengthdf<-as.data.frame(added_matrix)
ggplot()+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V1),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V2),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V3),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V4),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V5),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V6),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V7),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V8),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V9),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=V10),color="plum3")+
  geom_line(data=lengthdf, aes(x=c(1:1000),y=add),color="black")+
  annotate("text", x = 750, y = 80, label = expression(paste(lambda," = 5")),size=5) +
  annotate("text", x = 750, y = 85, label = "N = 10.000",size=5) +
  xlab("Iteration") +
  ylab("M")

```

```
#####
```

Kod för att visualisera frekvensfunktionerna.

```

    Z <- function(lam,N)
    as.function(laguerre(N-1,1))(-1/lam)/lam/N

modm <- function(lam,N) # mode of the distribution
    ceiling((sqrt((1+lam)^2+4*lam*N)-1-lam)/2/lam)

prob<-function(x,lambda, N){
  m <- modm(lambda,N)
  pmf <- rep(0,N)
  pmf[m] <- 1
  i <- 0
  while(m-i>1 | m+i <N)
  {
    i <- i+1
    if(m-i>=1)
      pmf[m-i] <- pmf[m-i+1]*lambda*(m-i+1)*(m-i)/(N-m +i)
    if(m+i<=N)
      pmf[m+i] <- pmf[m+i-1]*(N-m-i+1)/(m+i-1)/(m+i)/lambda
  }
  pmf <- pmf/sum(pmf)
  return(pmf[x])
}

N=10000
lam <- c(0.01,0.1,1,5,20)
cols <- c(1:4,"cyan3")

```

```

eps=10(-3)
# qqplot(c(0,N), c(0,0.4),xlab="m",ylab="pmf")
datamat<- matrix(0,nrow = N,ncol = 5)
for (i in 1:length(lam)){
  datamat[,i] <- prob(1:N,lam[i],N)
  sup <- (1:N)[datamat>eps]
  datamat[sup,i]<-0
  # lines(sup,pmf[sup],lty=2,col=cols[i])
  # points(sup,pmf[sup],pch=20,col=cols[i])
  # abline(h=0)
}
df<-data.frame("0.01"=datamat[,1],"0.1"=datamat[,2],"1"=datamat[,3],"5"=datamat[,4],
              "20"=datamat[,5])

ggplot() +
  geom_line(data = df, aes(x = c(1:N), y = X0.01),size=1 ,
            color = "bisque4") +
  geom_line(data = df, aes(x = c(1:N), y = X0.1), size=1 ,
            color = "azure4") +
  geom_line(data = df, aes(x = c(1:N), y = X1), size=1 ,
            color = "palevioletred4") +
  geom_line(data = df, aes(x = c(1:N), y = X5), size=1 ,
            color = "plum3") +
  geom_line(data = df, aes(x = c(1:N), y = X20), size=1 ,
            color = "mediumorchid4")+
  # annotate("text", x = 75, y = 0.1, label = expression(paste(lambda," = 0.01")),
            color="bisque4") +
  # annotate("text", x = 37.5, y = 0.12, label = expression(paste(lambda," = 0.1")),
            color="azure4") +
  # annotate("text", x = 18, y = 0.18, label = expression(paste(lambda," = 1")),
            color="palevioletred4") +
  # annotate("text", x = 13, y = 0.25, label = expression(paste(lambda," = 5")),
            color="plum3") +
  # annotate("text", x = 12, y = 0.37, label = expression(paste(lambda," = 20")),
            color="mediumorchid4") +
  # annotate("text", x = 75, y = 0.3, label = "N = 100",size=5) +
  xlab("M") +
  ylab("pmf")

df<-df[1:500,]
ggplot() +
  geom_line(data = df, aes(x = c(1:500), y = X0.1), size=1 ,color = "azure4") +
  geom_line(data = df, aes(x = c(1:500), y = X1), size=1 ,color = "palevioletred4") +
  geom_line(data = df, aes(x = c(1:500), y = X5), size=1 ,color = "plum3") +
  geom_line(data = df, aes(x = c(1:500), y = X20), size=1 ,color = "mediumorchid4")+
  annotate("text", x = 360, y = 0.025, label = expression(paste(lambda," = 0.1")),
            color="azure4") +
  annotate("text", x = 140, y = 0.05, label = expression(paste(lambda," = 1")),
            color="palevioletred4") +
  annotate("text", x = 80, y = 0.075, label = expression(paste(lambda," = 5")),
            color="plum3") +
  annotate("text", x = 50, y = 0.112, label = expression(paste(lambda," = 20")),
            color="mediumorchid4") +
  annotate("text", x = 350, y = 0.087, label = "N = 10 000",size=5) +
  xlab("M") +
  ylab("pmf")

```

```

# qqdata<-matrix(0,nrow = 500,ncol = 4)
# for (i in 1:4) {
#   meen<-sqrt(N/lam[i+1])
#   for (j in 1:500) {
#     newdata<-((df[j,i+1]*N-meen)/(sqrt(meen/2)))
#     qqdata[j,i]<-newdata
#   }
# }
# qqdf<- as.data.frame(qqdata)
# ggplot()+
#   stat_qq(data = qqdf, aes(sample=V1), color="azure4", distribution = stats::qnorm)+
#   geom_qq(data = qqdf, aes(sample=V2), color="palevioletred4")+
#   geom_qq(data = qqdf, aes(sample=V3), color="plum3")+
#   geom_qq(data = qqdf, aes(sample=V4), color="mediumorchid4")

N=10000
lam=5
r=1:N
cs=cumsum(pm(r,lam,N))
meen=sqrt(N/lam)
x<-pnorm(r,meen,sqrt(meen/2))
y<-cs

df<-as.data.frame(t(rbind(x,y)))
ggplot(data=df,aes(x=x,y=y))+geom_point(color="plum3")+geom_abline(intercept = 0,slope = 1)+
  xlab("Normalkvantiler")+
  ylab("M-kvantiler")+theme_gray()+
  annotate("text", x = 0.75, y = 0.5, label = "N = 10 000",size=5)+
  annotate("text", x = 0.75, y = 0.4, label = expression(paste(lambda," = 5")),size=5)

```