

Kwantificering van de accuratesse in evenwichtsberekeningen:

Monte Carlo studie van de foutenpropagatie bij de potentiometrische bepaling van stabiliteitsconstanten

Proefschrift voorgelegd tot het bekomen van de graad van Doctor in de Wetenschappen Richting Scheikunde aan het Limburgs Universitair Centrum te verdedigen door

WANDA GUEDENS

Promotor: Prof. Dr. L.C. Van Poucke Copromotor: Dr. E. Pauwels (KU Leuven)

Diepenbeek 1999



541.12.

991201



2 9 SEP. 1999

541.12 GUED 1999

.luc.luc.luc.





FACULTEIT WETENSCHAPPEN

Kwantificering van de accuratesse in evenwichtsberekeningen:

Monte Carlo studie van de foutenpropagatie bij de potentiometrische bepaling van stabiliteitsconstanten

991201

Proefschrift voorgelegd tot het bekomen van de graad van

Doctor in de Wetenschappen Richting Scheikunde aan het Limburgs Universitair Centrum te verdedigen door

WANDA GUEDENS



29 SEP. 1999

Promotor: Prof. Dr. L.C. Van Poucke Copromotor: Dr. E. Pauwels (KU Leuven)

Diepenbeek 1999

Schilderij omslag: "Uit chaos ontstaat structuur"

KOPTLEVEN

Lic. Scheikunde

Vormgeving omslag: Idefix, Wellen

Druk: Paesen, Opglabbeek

© LUC Diepenbeek, 1999

Aan mijn vader, *1932 - †1999



Dank

Het vinden van de put ('pit search') was het belangrijkste doelwit in deze thesis; dank aan allen die mij ervoor behoed hebben er in te vallen...

Prof. dr. Lucien Van Poucke, promotor, voor het aanbrengen van het onderwerp, de geboden vrijheid tijdens mijn onderzoek en zijn professioneel wetenschappelijk advies, Dr. Eric Pauwels (ESAT, KU Leuven), copromotor, voor de verhelderende discussies en zijn aanstekelijke gedrevenheid tot fundamentele perfectie,

Prof. dr. Jules Mullens en Prof. dr. Jan Yperman, die er steeds voor te vinden waren om mij de zo noodzakelijke feedback te geven,

Alle overige ZAP-leden van de Vakgroep Scheikunde, die de ruimte en kansen boden om dit proefschrift binnen een redelijke termijn te realiseren,

Alle (ex-)collega's, vrienden en kennissen, die mijn inspanningen steeds naar waarde hebben geschat, waaruit ik moed putte om dit werk tot een goed einde te brengen,

Het administratief en technisch personeel van het secretariaat SBG en WNI, didactieken research Scheikunde, wier sympathie voor mij een hart onder de riem betekende, Monique Reynders, mijn steun en toeverlaat vanaf het eerste uur, voor haar kritische reflecties op woensdagen,

Viviane Mebis, voor haar excellent typewerk waarvoor ze menig weekend opofferde, Magda Ieven en Mieke Daenen, die met kennis van zaken een gedeelte van de grafische voorstellingen verzorgden,

Mijn familie, heel in het bijzonder, mijn moeder, mijn echtgenoot en mijn zoontje, voor hun niet aflatend geduld en enthousiasme en hun hulp op alle mogelijke vlakken. Zij kregen gedurende de periode van dit doctoraat niet de aandacht die ze van mij verdienden.

3 november, 1999

Wanda



Inhoudsopgave

prant	AUCIAC			
Hoofd	stuk 1	Algemene inleiding	5	
	Refere	enties	8	
Hoofd	stuk 2	Lineaire foutenpropagatie en grenzen voor linearisering	13	
2.1	Inleidi	ng	13	
2.2	Foutenpropagatie bij functies van meerdere variabelen			
	2.2.1	Algemene beschouwingen	17	
	2.2.2	Toepassing: standaardformules voor foutenpropagatie	18	
	2.2.3	Concrete voorbeelden van foutenpropagatie in de chemie	20	
2.3	Fouter	apropagatie in gesimuleerde experimenten	28	
	2.3.1	Bepalen van de grenzen voor lineaire foutenpropagatie	28	
	Refere	nties	43	
Deel ling v	II N van st	Iodelleren van potentiometrische data: de bepa abiliteitsconstanten	45	
Deel ling v Hoofds	II N van st stuk 3	Iodelleren van potentiometrische data: de bepa abiliteitsconstanten Potentiometrische experimenten	- 45 49	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding	- 45 49 49	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse	- 45 49 49 49	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte	- 49 49 49 49 50	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie	- 49 49 49 49 50 54	
Deel ling v Hoofd: 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie	- 49 49 49 50 54 56	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie tentiometrische techniek	- 49 49 49 50 54 56 60	
Deel ling v Hoofds 3.1 3.2	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding	- 49 49 49 50 54 56 60 60	
Deel ling v Hoofds 3.1 3.2	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen	- 49 49 50 54 56 60 60 60 60	
Deel ling v Hoofds 3.1 3.2	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2 3.2.3	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen Het indifferent elektrolyt	- 49 49 49 50 54 56 60 60 60 60 60 60	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2 3.2.3 3.2.4	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen Het indifferent elektrolyt	- 45 49 49 50 54 56 60 60 60 60 64 64	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5	Jodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen Het indifferent elektrolyt Ijking van de meetopstelling	- 45 49 49 50 54 56 60 60 60 60 64 64 72	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6	Iodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen Het indifferent elektrolyt De elektroden Ijking van de meetopstelling De klassieke analysemethode voor het bepalen van stabiliteitscon-	- 45 49 49 50 54 56 60 60 60 60 64 64 72	
Deel ling v Hoofds 3.1	II N van st stuk 3 Theor 3.1.1 3.1.2 3.1.3 3.1.4 De po 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6	Jodelleren van potentiometrische data: de beparabiliteitsconstanten Potentiometrische experimenten etische beschouwingen: inleiding Potentiometrie in de evenwichtsanalyse De methode van de constante hoge ionensterkte De voordelen van de potentiometrie Ontwikkelingen in de potentiometrie Inleiding Potentiaalmetingen Het indifferent elektrolyt De klassieke analysemethode voor het bepalen van stabiliteitsconstanten	- 49 49 49 50 54 56 60 60 60 60 64 64 72 80	

Hoofd	stuk 4	Kwalitatieve en kwantitatieve analyse van potentiometrisc	he		
data n	net beh	ulp van optimalisatieprogramma's	91		
4.1	Inleidi	ng	91		
4.2	Het optimalisatieprogramma SUPERQUAD voor het berekenen van sta-				
	biliteit	sconstanten	92		
	4.2.1	Inleiding	92		
	4.2.2	De modelfunctie en de kleinste kwadraten benadering	94		
	4.2.3	Computationele methode	98		
	4.2.4	Implementatie van SUPERQUAD in MatLab ^{TM}	104		
	4.2.5	Validatie van het MatLab TM programma	109		
4.3	Interpretatie van de resultaten				
	4.3.1	Modelselectie	114		
	4.3.2	Foutenanalyse	117		
	Refere	$nties \ldots \ldots$	122		
Hoofd	stuk 5	De verhoudingsmethode	127		
5.1	Bereke	eningen met het standaard SUPERQUAD programma	127		
	5.1.1	Binaire systemen	127		
	5.1.2	Ternaire systemen	135		
5.2	Berekeningen met de MatLab ^{TM} implementatie van SUPERQUAD .		141		
	5.2.1	Binaire systemen	141		
	5.2.2	Ternaire systemen	142		
	Refere	nties	143		

Deel III Optimaliseren van "synthetische" potentiometrische data 147

Hoofdstuk 6 Monte Carlo studie van de foutenpropagatie bij de bepaling van stabiliteitsconstanten met behulp van SUPERQUAD 151 6.1 151 6.2 De invloed van een afleesfout op de pH in elk titratiepunt . . . 153 6.2.1 6.2.2 De invloed van perturbaties in het toegevoegd volume base per 6.2.3 De invloed van perturbaties op het initieel volume in de recipiënt 162 6.2.4 6.3.1

6.3.2	Analyse van modelsysteem 'sys 1'	. 168
6.3.3	Analyse van modelsysteem 'sys 2'	. 182
Refere	nties	. 202
Hoofdstuk 7	Samenvatting en algemeen besluit	203

Deel IV Appendices

Appen	dix A Relevante begrippen en notaties uit de kanstheorie en	de
statisti	iek	211
A.1	Stochastische variabele	. 211
A.2	Verwachting en variantie van stochastische variabelen	. 217
A.3	Covariantie en onafhankelijkheid van stochastische variabelen	. 218
A.4	Functies van stochastische variabelen	. 219
	A.4.1 Rekenregels voor verwachtingen	. 219
	A.4.2 Rekenregels voor varianties	. 222
A.5	Het normaal - kwantiel - diagram, normaliteitsdiagram of 'normal prob	-
	ability plot'	. 224
A.6	De ongelijkheid van Jensen	. 229
A.7	De Centrale Limietstelling	. 230
A.8	Multivariate lineaire regressie	. 234
A.9	Verwachtingswaarde van de som der kwadratische residuelen $\ .$. 235
	Referenties	. 238
Appen	dix B Partiële afgeleiden van de totale naar de vrije concentra	tie
van de	componenten berekend in de MatLab TM functie	
'jacobi	an_free.m'	239
Appen	dix C Partiële afgeleiden van de totale concentratie naar de j	pa-
ramete	ers berekend in de MatLab TM functie 'jacobian_beta.m'	243
Acade	mische loopbaan	245



Deel I

Statistische foutenpropagatie in de chemie: een praktische benadering



Hoofdstuk 1 Algemene inleiding

Hoofdstuk 2 Lineaire foutenpropagatie en grenzen voor linearisering

2.1	Inleid	ing					
2.2	Foutenpropagatie bij functies van meerdere variabelen						
	2.2.1	Algemene beschouwingen					
	2.2.2	Toepassing: standaardformules voor foutenpropagatie					
	2.2.3	Concrete voorbeelden van foutenpropagatie in de chemie					
		2.2.3.1 Volumetrie					
		2.2.3.2 Calorimetrie					
		2.2.3.3 Reactiekinetiek					
2.3	Foute	npropagatie in gesimuleerde experimenten					
	2.3.1	Bepalen van de grenzen voor lineaire foutenpropagatie					
	Referenties						



Hoofdstuk 1

Algemene inleiding

Men kan zonder overdrijven stellen dat moderne natuurwetenschap slechts mogelijk is indien men in staat is de fenomenen die men bestudeert, nauwkeurig te meten. Immers, niet alleen leveren deze metingen ons de noodzakelijke gegevens om (wiskundige) modellen van de natuur te construeren, maar bovendien stellen zij ons in staat de op die modellen gebaseerde voorspellingen te controleren. Dit maakt het mogelijk hiaten in ons begrijpen van de fysische en chemische werkelijkheid op te sporen en helpt ons een beter inzicht te verkrijgen in de werking van de natuurwetten.

Nu is het echter zo dat men bij herhaaldelijk opmeten van eenzelfde grootheid haast nooit exact dezelfde waarden bekomt. Daarom is het van het allergrootste belang dat bij elk resultaat aangegeven wordt hoe nauwkeurig het is. Dit zal ons toelaten om bij de interpretatie van het experiment uit te maken of er al dan niet een werkelijk verschil is tussen de resultaten. Merk op dat de resultaten waarover hier sprake is niet noodzakelijk rechtstreeks uit metingen volgen. In vele gevallen zal het nodig zijn op de ruwe gegevens allerlei wiskundige bewerkingen te verrichten alvorens we tot de gewenste uitkomst komen. Het zal dus ook van belang zijn te weten op welke wijze de onnauwkeurigheden op de individuele metingen bijdragen tot de onzekerheid op het eindresultaat, met andere woorden een foutenpropagatiestudie dringt zich op.

Tijdens mijn onderzoek dat gestart is met een praktische benadering van foutenanalyse toegepast op volumetrische, calorimetrische en kinetische experimenten in hoofdstuk 2 [1], [2], groeide de interesse in een foutenpropagatiestudie bij de potentiometrische bepaling van stabiliteitsconstanten van Ag(I)-complexeringsevenwichten. Daarenboven kon ik voortbouwen op een lange "potentiometrische" traditie binnen de onderzoeksgroep Anorganische en Fysische Scheikunde van het Limburgs Universitair Centrum [3], [4], [5], [6], [7], [8] op gebied van zowel experimentele als computationele expertise. In hoofdstuk 3 verduidelijken we enkele principes verbonden aan het gebruik van de potentiometrische techniek. We leggen de experimentele voorwaarden vast voor de bepaling van globale stabiliteitsconstanten van Ag(I) met lange - alifatische - keten diamines, waarbij in eerste instantie de zuurconstanten van deze diamines en de zilverhydrolyseconstanten bepaald worden. Tevens geven we een gedetailleerde beschrijving van de gebruikte reagentia en de geautomatiseerde meetopstelling met aangegeven meetnauwkeurigheid.

Bij de volgende stap – de evaluatie en analyse van evenwichtsmetingen – vermelden we het gebruik van grafische methoden die tot in de jaren zestig als belangrijkste instrument aanzien werden. Het is bijvoorbeeld essentieel dat men over goede startwaarden beschikt voor de constanten om te voorkomen dat tijdens de iteratieprocedure de optimalisatie vastloopt in een lokaal minimum. Ten tweede wil men een gevoel voor de betekenis en grootteorde van de constanten krijgen [9]. Ten derde is het mogelijk om vast te stellen in welk pM of pH gebied het eenvoudige model faalt. Tenslotte is het resultaat van een grafische analyse een indicatie voor het basismodel indien men met verschillende en onafhankelijke technieken gelijkwaardige resultaten kan bekomen. Aangezien de evaluatie van grafische methoden niet het onderwerp van deze thesis is, gaan we hier niet dieper op in.

Nieuwe ontwikkelingen in wiskundige algoritmen en de opkomst van krachtige computerprogramma's (o.a. SUPERQUAD) voor de analyse van potentiometrische data – tabel 1.1 geeft een volledig overzicht – hebben ervoor gezorgd dat de complexchemie tijdens de voorbije dertig jaar een extra stimulans kreeg [10], [11], [12], [13].

Recent baseert men zich meer op chemometrische methoden op basis van niet-lineaire modelleringstechnieken [50] om de accuratesse van evenwichtsmodellen en hun corresponderende stabiliteitsconstanten aan te tonen [51], [52], [53]. Bij de meeste optimalisatieprogramma's ontbreekt echter een statistisch verantwoorde foutenanalyse. Het is dan ook een uitdaging om hiertoe een bijdrage te leveren.

Daar MatLabTM tegenwoordig de programmeertaal bij uitstek is voor vele chemometrici wordt het algoritme van SUPERQUAD (standaard geprogrammeerd in Fortran 77) volledig geïmplementeerd in MatLabTM versie 5.2.1.1420. Deze implemen-

tatie wordt gedetailleerd beschreven in hoofdstuk 4 en uitvoerig gevalideerd met behulp van synthetische potentiometrische data. Een vergelijking met de protonerings- en stabiliteitsconstanten bekomen via de standaardversie van SUPERQUAD op basis van de experimentele data van de bestudeerde binaire en ternaire systemen, moet in hoofdstuk 5 nogmaals de degelijkheid van de MatLabTM versie onderstrepen.

Programma	Ref.	Datatype ^a	Programma	Ref.	Datatype ^a
ACBA	[14]	V	MUCOMP	[33]	V
ACREF3AM	[15]	v	MUPROT	[34]	v
AP	[16]	v	NONLIN15	[35]	V, A
BEST	[17]	v	Papanastasiou and Zogas	[36]	v
BSTAC	[18]	v	PKAS	[37]	v
DALSFEK	[19]	V, A	POLET	[38]	v
DHMINOPT	[20]	v	PROTAF	[39]	v
ECORM	[21]	v	PSEQUAD	[40]	V, A
ESAB2M	[22]	v	SCOGS2a	[41]	v
HYPERQUAD	[23], [24]	V, A	SCOGS2b	[42]	v
KATCOM	[25]	v	Sidrak and Aboul-Seoud	[43]	v
LETAGROP-VRID	[26]	v	STBLTY	[44]	V
LOGMIN	[27]	v	STEW	[45]	V
MAXIPOT-F	[28]	v	SUPERQUAD	[46]	v
MINIGLASS	[29]	v	TITAN	[47]	v
MINIQUAD	[30]	v	TITFIT	[48]	v
MINIQUAD75	[31]	v	TITRA	[49]	V
MINUITS	[32]	v			

Tabel 1.1 Computerprogramma's voor het berekenen van evenwichtsconstanten uitgaande van potentiometrische data (V)

^a Bijkomende spectrofotometrische data werden in de berekeningen gebruikt (A)

In hoofdstuk 6 tenslotte, maken we met behulp van de ontwikkelde MatLabTM programmatuur gebruik van een Monte Carlo techniek om de precisie waarmee stabiliteitsconstanten bepaald worden, te kwantificeren. In de literatuur zijn hierover slechts weinig voorbeelden te vinden [54]. De invloed van zowel *experimentele* (toevallige) als systematische fouten in verschillende titratieparameters wordt nagegaan, zodat aan de hand van de geselecteerde parameterset ontegensprekelijk het resultaat met de hoogst mogelijke accuratesse volgt en discrepanties in gepubliceerde waarden van vormingsconstanten verklaard kunnen worden. Tevens worden via deze Monte Carlo studie aan de hand van *structurele* fouten in het chemisch model de potentiële implicaties voor modelselectie onderzocht.

Referenties

- W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 776, (1993)
- [2] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 838, (1993)
- [3] J. Yperman, Potentiometrisch en calorimetrisch onderzoek van het zuurbase gedrag en van de complexvorming met Ag(I) van een aantal Nmethylgesubstitueerde-diethyleentriaminen, proefschrift, U.I.A., (1981)
- [4] J. Vliegen, Potentiometrisch, calorimetrisch en densimetrisch onderzoek van het auto-associatiegedrag van enkele α-L-aminozuren in waterig midden, proefschrift, L.U.C., (1982)
- [5] J. Vanhees, Thermodynamische studie van het asymmetrische elektrolytmengsel Indiumperchloraatperchloorzuur-water, proefschrift, L.U.C., (1983)
- [6] P. Luts, Potentiometrische bepaling van de stabiliteitsconstanten van de sulfatocomplexen van koper (II) en cadmium (II) in 1 M NaClO₄, proefschrift, L.U.C., (1991)
- [7] R. Garner, Een potentiometrische studie van het complexeringsgedrag van Ag(I) met alifatisch gesubstitueerde mono- en diaminen in 1 M nitraat, proefschrift, L.U.C., (1993)
- [8] K. Brassinne, Potentiometrische studie van de oxalatocomplexen van Cu²⁺ en Y³⁺ in KNO₃-midden, proefschrift, L.U.C., (1997)
- [9] H. Rossotti, The Study of Ionic Equilibria: An Introduction, Longman, New York, (1978)
- [10] D.J. Legett, Computational Methods for the Determination of Formation Constants, Plenum Press, New York, (1985)
- [11] A.E. Martell, R.J. Motekaitis, Determination and Use of Stability Constants, VCH publishers, Weinheim, (1988)

- [12] M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria: A guide to Methods in Potentiometry, Extraction and Spectrophotometry, Ellis Horwood Ltd, Chichester, (1988)
- [13] M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria, Ellis Horwood Ltd, Chichester, (1994)
- [14] G. Arena, R. Rizzarelli, S. Sammartano, C. Rigano, Talanta, 26, 1, (1979)
- [15] M. Cromer-Morin, J.P. Scharff, R.P. Martin, Analusis, 10, 92, (1982)
- [16] P. Decock, B. Sarker, Can. J. Chem., 33, 531, (1986)
- [17] R.J. Motekaitis, A.E. Martell, Can. J. Chem., 60, 168, (1982)
- [18] C. De Stefano, P. Mineo, C. Rigano, S. Sammartano, Ann. Chim., 83, 243, (1993)
- [19] R.A. Alcock, F.R. Hartley, D.E. Rogers, J. Chem. Soc., Dalton Trans., 115, (1978)
- [20] M. Meloun, M. Javúrek, J. Militký, Mikrochim. Acta, 109, 221, (1992)
- [21] E. Bottari, T. Coccitto, G. Curzio, M.R. Festa, R. Jasionowska, Ann. Chim., 78, 635, (1988)
- [22] C. De Stefano, P. Princi, C. Rigano, S. Sammartano, Ann. Chim., 77, 643, (1987)
- [23] A. Sabatini, A. Vacca, P. Gans, Coord. Chem. Rev., 120, 389, (1992)
- [24] P. Gans, A. Sabatini, A. Vacca, Talanta, 43, 1739, (1996)
- [25] V. Gadiokov, Kh. Mikhailova, Comput. Chem., 13, 325, (1989)
- [26] L.G. Sillén, Acta Chem. Scand., 18, 1085, (1964)
- [27] A. Poznajlo, J. Radioanal. Nucl. Chem., 134, 97, (1989)
- [28] F. Gaizer, I.I. Kiss, Talanta, 41, 419, (1994)
- [29] A. Izquierdo, J.L. Beltrán, Anal. Chim. Acta, 181, 87, (1986)
- [30] A. Sabatini, A. Vacca, P. Gans, Talanta, 21, 53, (1974)

- [31] P. Gans, A. Sabatini, A. Vacca, Inorg. Chim. Acta, 18, 239, (1976)
- [32] T. Michalowski, Talanta, 39, 1127, (1992)
- [33] M. Wozniak, J. Canonne, G. Nowogrocki, J. Chem. Soc., Dalton Trans., 2419, (1981)
- [34] M. Wozniak, G. Nowogrocki, Talanta, 25, 633, (1979)
- [35] P.D. Taylor, I.E.G. Morrison, R.C. Hider, Talanta, 35, 507, (1988)
- [36] G. Papanastasiou, I. Zogas, Talanta, 42, 827, (1995)
- [37] R.J. Motekaitis, A.E. Martell, Can. J. Chem., 60, 3403, (1982)
- [38] J. Havel, M. Meloun, Talanta, 33, 525, (1986)
- [39] R. Fournaise, C. Pettifaux, Talanta, 34, 385, (1987)
- [40] L. Zekany, I. Nagypal in D.J. Legett (Ed.), Computational Methods for the Determination of Formation Constants, Plenum Press, New York, 291, (1985)
- [41] J.P. Chandler, R.E. Thomson, H.O. Spivey, E.L.F. Li, Anal. Chim. Acta, 162, 399, (1984)
- [42] D.D. Perrin, H. Stunzi in D.J. Legett (Ed.), Computational Methods for the Determination of Formation Constants, Plenum Press, New York, 71, (1985)
- [43] Y.L. Sidrak, A. Aboul-Seoud, J. Comput. Chem., 8, 575, (1987)
- [44] A. Avdeef in D.J. Legett (Ed.), Computational Methods for the Determination of Formation Constants, Plenum Press, New York, 355, (1985)
- [45] P. Gans, A. Vacca, Talanta, 21, 45, (1974)
- [46] P. Gans, A. Sabatini, A. Vacca, J. Chem. Soc., Dalton Trans., 1195, (1985)
- [47] W.E. Gordon, Anal. Chem., 54, 1595, (1982)
- [48] A. Zuberbühler, T.A. Kaden, Talanta, 29, 201, (1982)

- [49] A. Johansson, S. Johansson, Analyst, 104, 601, (1979)
- [50] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Data Handling in Science and Technology, Volume 2, Chemometrics: A Textbook, Elsevier, A'dam, (1990)
- [51] S.D. Dimitrov, D.I. Kamenski, Computers Chem., 21, 167, (1997)
- [52] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, J. Coord. Chem., 47, 17, (1999)
- [53] M.K. Hartnett, M. Bos, W.E. van der Linden, D. Diamond, Anal. Chim. Acta, 316, 347, (1995)
- [54] P.M. May, K. Murray, Talanta, 35, 933, (1988)



Hoofdstuk 2

Lineaire foutenpropagatie en grenzen voor linearisering

2.1 Inleiding

Met de opkomst van krachtige computers hebben wetenschap en technologie een nieuwe wending genomen: men spreekt van "digitale wetenschap" [1], [2]. Deze computers laten immers toe de fysische wereld te vervangen door een digitale simulatie in de vorm van een numeriek model. Getransformeerd naar beelden op het computerscherm, geeft het model een nauwkeurige kopie van de essentiële kenmerken van natuurfenomenen weer. Het voordeel van simulaties is ten eerste dat de onderzoeker het experiment zo dikwijls kan herhalen als hij maar wil, terwijl de meeste natuurlijke fenomenen slechts eenmalig of enkele malen kunnen geobserveerd worden. Ten tweede kan een ganse verzameling van oplossingen bestudeerd worden die elk hetzelfde fenomeen beschrijven maar met variabelen die verschillende waarden aannemen. Tenslotte kunnen de waarden van alle fysische parameters tegelijk bepaald worden omdat ze variëren zowel in tijd als in ruimte. De twee traditionele manieren om wetenschap te benaderen — experiment en theorie - blijken dus heden ten dage gecomplementeerd te kunnen worden door een derde en moderne methodologie: de computationele benadering [3]. Deze heeft tot doel een numerieke oplossing te bieden voor de mathematische modellen van de theoreticus in al hun complexiteit. Dit in tegenstelling tot theoretische modellen waar men vaak allerlei benaderingen dient in te voeren om de modellen te kunnen oplossen (cfr. linearisering voor foutenpropagatie). In het verder verloop van dit werk wordt daarom uitvoerig gebruik gemaakt van de computationele benaderingswijze in de vorm van simulaties. Dit moet ons toelaten om de invloed na te gaan van statistische fluctuaties gedurende het ganse meetproces op het eindresultaat. Gezien de complexiteit

van vele experimentele processen is het immers dikwijls onmogelijk om deze onzekerheden theoretisch te voorspellen. Simulatie door middel van computermodellen is dan de enige haalbare optie.

Het voorgaande onderstreept het belang van formele modellen voor het kwantificeren en simuleren van statistische onzekerheid. Vandaar dat we voor de geïnteresseerde chemicus in **appendix A** een aantal relevante begrippen opfrissen uit de statistiek en de kanstheorie die in het vervolg van deze thesis veelvuldig zullen gebruikt worden.

In de meeste gevallen zal het onderliggend model redelijk expliciet in termen van stochastische variabelen beschreven worden. Dit is belangrijk omdat het ons toelaat om het voorgestelde model op een nauwkeurige en eenduidige manier te vertalen naar een simulatie, die dan het vertrekpunt kan zijn voor een studie van de inherente statistische variabiliteit. We zullen heel regelmatig — in een poging de lezer te overtuigen van het nut van simulaties (voor zover dat nodig zou zijn) — welbekende formules herhalen en illustreren aan de hand van simulaties. Dit zal de lezer voorbereiden op het laatste gedeelte van deze thesis waar de complexiteit van de vraagstelling betekent dat we alleen op deze computationele mogelijkheid kunnen terugvallen.

Elementaire terminologie

Bij een kwantitatieve meting van bijvoorbeeld de pH van een oplossing van een zuur waren de meetresultaten de volgende:

$$x_1 = 3.75$$

 $x_2 = 3.82$
 $x_3 = 3.67$

Deze data geven ten eerste aan dat een potentiometer gebruikt wordt waarvan de precisie gelijk is aan 0.01/2 = 0.005 eenheden en ten tweede dat de onzekerheid op het resultaat veel groter is dan de precisie omdat de individuele metingen onderling ongeveer 0.2 eenheden verschillen.

De spreiding van de pH waarden is het cumulatieve resultaat van een aantal controleerbare of oncontroleerbare fluctuaties in de condities die de metingen beïnvloeden, bijvoorbeeld veranderingen in temperatuur, verlichting, ... Omdat deze fluctuaties zich toevallig voordoen baseert men zich op statistische theorieën om de resultaten te analyseren.

Volgende notatie wordt afgesproken:

- ξ : onbekende "echte" waarde van een fysische of chemische parameter (wordt aangeduid met een Griekse letter), bijvoorbeeld de "echte" pH waarde van de beschouwde oplossing of het "echte" gemiddelde μ ;
- X : een stochastische variabele die het resultaat van een meetproces voorstelt (Appendix A);
- x: het resultaat van een individuele meting.

Binnen het kader van het experimenteel model mogen we meestal veronderstellen dat de waarden van de stochastische variabele dicht in de buurt van de "echte" waarde liggen. We modelleren dit wiskundig door te stellen dat:

$$X = \xi + \varepsilon$$

waarbij ε een stochastische variabele is waarvan de waarden relatief klein zijn ten opzichte van ξ . Bovendien gaan we ervan uit dat ε toevallige fouten modelleert:

$$E(\varepsilon) = 0$$

Als gevolg hiervan zien we dat de gemiddelde waarde van de stochast X samenvalt met de "echte" waarde ξ :

$$E(X) = E(\xi + \varepsilon) = \xi + E(\varepsilon) = \xi$$

Foutenpropagatie is meestal gebaseerd op de veronderstelling dat de onderliggende foutenverdeling gaussiaans is; dit hoeft niet zo te zijn. Een algemeen gevolg van de "Centrale Limietstelling" (Appendix A), [4] geeft immers aan dat onafhankelijk van de verdeling van X_i , het steekproefgemiddelde $\overline{X} = \frac{1}{n} \sum X_i$ ongeveer normaal verdeeld is. Deze benadering zal bovendien nauwkeuriger worden naarmate de steekproefomvang, n toeneemt.

We kunnen m.a.w. bij benadering stellen dat:

$$\overline{X} \sim N(\mu_{\overline{X}}, \sigma_{\overline{X}}^2) \tag{2.1}$$

waarbij:

en

$$\mu_{\overline{X}} = E(\frac{1}{n}\sum X_i) = \frac{1}{n}\sum X_i = \xi$$

$$\sigma_{\overline{X}}^2 = \operatorname{Var}(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\sum \operatorname{Var}X_i = \frac{\sigma^2}{n}$$

Meer bepaald betekent dit dat we bij benadering kunnen stellen dat:

$$P[\overline{X} - 3\frac{\sigma}{\sqrt{n}} \le \xi \le \overline{X} + 3\frac{\sigma}{\sqrt{n}}] \approx 0.997$$
(2.2)

Vergelijking (2.2) is gebaseerd op de regels dat voor de normale populatie ongeveer 68% van de data in het interval $\mu \pm \sigma$ ligt, 95.5% binnen $\mu \pm 2\sigma$ ligt en 99.7% binnen $\mu \pm 3\sigma$ ligt. Vergelijking (2.2) is enkel bij benadering juist omdat in de meeste gevallen \overline{X} slechts bij benadering normaal verdeeld is. Bovendien wanneer σ moet geschat worden, wat dikwijls het geval is, wordt de student-t [5] verdeling gebruikt in plaats van de normale verdeling. Maar in de meerderheid van de gevallen zijn de afwijkingen verwaarloosbaar klein zodat we ons hierom in wat volgt niet zullen bekommeren. Samengevat komt het er in de praktijk op neer dat men er zeker van mag zijn dat de "echte" waarde ξ (met $\xi = \mu$) van de gemeten grootheid zich situeert in het interval:

$$[\overline{X} - 3\sigma_{\overline{X}}, \overline{X} + 3\sigma_{\overline{X}}] \tag{2.3}$$

met $3\sigma_{\overline{X}}$ de theoretische absolute (statistische) fout, AF(x), op het resultaat.

Theoretisch wordt hier gebruikt omdat in de praktijk $Var(\overline{X})$ niet gekend is maar moet geschat worden uitgaande van de experimentele data en dit als volgt:

$$S^{2} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i} - \overline{X})^{2}$$
(2.4)

met S^2 een schatter voor $Var(X_i) = \sigma^2$. Daar uit het voorgaande blijkt¹

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} \tag{2.5}$$

zodat $\sigma_{\overline{x}}^2$ kan geschat worden:

$$\frac{1}{n(n-1)}\sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{S^2}{n}$$
(2.6)

Uitgaande van deze beschouwingen wordt in de volgende samenvatting de definitie voor de absolute fout AF(x) vastgelegd.

Samenvatting: wanneer x_1, \ldots, x_n de waarnemingen zijn van een onbekende grootheid x, dan wordt de beste schatting voor deze grootheid gegeven door

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.7}$$

¹In hetgeen volgt zijn de gelijkheden enkel bij benadering juist voor algemene stochastische variabelen. De gelijkheden gelden strikt wanneer de stochastische variabelen gaussiaans verdeeld zijn.

en de absolute (statistische) fout, AF(x), op het resultaat

$$AF(x) = 3\sqrt{\frac{s^2}{n}} = 3\sqrt{\frac{\sum(x_i - \overline{x})^2}{n(n-1)}}$$
(2.8)

wat (op de factor 3 na) ook de "standard error of the mean" (MSE) genoemd wordt. De relatieve fout, RF(x) wordt gedefinieerd als

$$RF(x) = \frac{AF(x)}{x} \tag{2.9}$$

2.2 Foutenpropagatie bij functies van meerdere variabelen

2.2.1 Algemene beschouwingen

Zoals hierboven werd uiteengezet is de systematische benadering van foutenpropagatie gebaseerd op de berekening van de variantie van de overeenkomstige stochastische variabelen. Veronderstel bijvoorbeeld dat y een functie is van n variabelen:

$$y = f(x_1, \dots, x_n) \tag{2.10}$$

Dan geldt voor de overeenkomstige stochastische variabelen dat in eerste benadering:

$$\operatorname{Var}(Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_{i}}\right) \left(\frac{\partial f}{\partial x_{j}}\right) \operatorname{Cov}(X_{i}, X_{j})$$
(2.11)

of nog

$$\operatorname{Var}(Y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \operatorname{Var}(X_i) + 2\sum_{i < j} \sum_{i < j} \left(\frac{\partial f}{\partial x_i}\right) \left(\frac{\partial f}{\partial x_j}\right) \operatorname{Cov}(X_i, X_j)$$
(2.12)

Men weet echter dat in de meeste toepassingen de opmetingen X_i en X_j onafhankelijk zijn, zodat de covariantieterm verdwijnt:

$$\operatorname{Var}(Y) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2 \operatorname{Var}(X_i)$$
(2.13)

Een concreet voorbeeld hiervan is de volgende eenvoudige toepassing:

$$AF(f(x,y)) = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 AF^2(x) + \left(\frac{\partial f}{\partial y}\right)^2 AF^2(y)}$$
(2.14)

2.2.2 Toepassing: standaardformules voor foutenpropagatie

1. Som en verschil

Beschouw $f(x_1, \ldots, x_n) = \sum \alpha_i x_i$ met $\alpha_i = \pm 1$, dan is $\frac{\partial f}{\partial x_i} = \alpha_i$

en dus

 $\left(\frac{\partial f}{\partial x_i}\right)^2 = 1 \tag{2.15}$

Toepassen van vergelijking (2.14) geeft

$$AF^{2}\left(\sum \alpha_{i}x_{i}\right) = \sum AF^{2}(x_{i})$$

$$(2.16)$$

en

$$AF(x_1 \pm x_2) = \sqrt{AF^2(x_1) + AF^2(x_2)}$$
(2.17)

2. Product en quotiënt met "foutloze" constante

i) Wanneer $f(x) = \alpha x$ met α een "foutloze" constante, dan is

$$\frac{\partial f}{\partial x} = \alpha$$
 (2.18)

en

$$\operatorname{Var} f(X) = \alpha^2 \operatorname{Var}(X) \tag{2.19}$$

vandaar

$$AF(\alpha x) = |\alpha|AF(x) \tag{2.20}$$

ii) Wanneer $f(x) = \frac{x}{\beta}$ met β een "foutloze" constante, dan is

$$\frac{\partial f}{\partial x} = \frac{1}{\beta} \tag{2.21}$$

en

$$\operatorname{Var} f(X) = \left(\frac{1}{\beta}\right)^2 \operatorname{Var}(X)$$
 (2.22)

vandaar

$$AF\left(\frac{x}{\beta}\right) = \frac{1}{|\beta|}AF(x) \tag{2.23}$$

3. Product en quotiënt, macht en wortel

Beschouw

$$f(x,y,z) = \frac{\alpha^q x^n y^m}{z^k} \tag{2.24}$$

waarbij $q,\,n,\,m$ en k positieve "foutloze" exponenten zijn en α een "foutloze" constante is.

Partieel afleiden van de functie naar x, y en z geeft:

$$\frac{\partial f}{\partial x} = \frac{\alpha^q n x^{n-1} y^m}{z^k} = \frac{n}{x} \frac{\alpha^q x^n y^m}{z^k} = \frac{n}{x} f(x, y, z)$$

$$\frac{\partial f}{\partial y} = \frac{\alpha^q x^n m y^{m-1}}{z^k} = \frac{m}{y} \frac{\alpha^q x^n y^m}{z^k} = \frac{m}{y} f(x, y, z)$$

$$\frac{\partial f}{\partial z} = -k \alpha^q x^n y^m z^{-(k+1)} = \frac{-k}{z} \frac{\alpha^q x^n y^m}{z^k} = \frac{-k}{z} f(x, y, z)$$
(2.25)

Ingevuld in vergelijking (2.14) geeft dit:

$$AF^{2}(f(x,y,z)) = \frac{n^{2}}{x^{2}}f^{2}AF^{2}(x) + \frac{m^{2}}{y^{2}}f^{2}AF^{2}(y) + \frac{(-k)^{2}}{z^{2}}f^{2}AF^{2}(z)$$
(2.26)

delen door f^2 , wat de verkorte schrijfwijze is voor $f^2(x, y, z)$, geeft:

$$RF^{2}(f(x, y, z)) = n^{2} \left(\frac{AF(x)}{x}\right)^{2} + m^{2} \left(\frac{AF(y)}{y}\right)^{2} + k^{2} \left(\frac{AF(z)}{z}\right)^{2}$$

$$= n^{2}RF^{2}(x) + m^{2}RF^{2}(y) + k^{2}RF^{2}(z)$$
(2.27)

Vandaar

$$RF(f(x, y, z)) = \sqrt{n^2 R F^2(x) + m^2 R F^2(y) + k^2 R F^2(z)}$$
(2.28)

4. Exponentiële en logaritmische functies

i) Beschouw de volgende exponentiële functie:

$$y = f(x) = e^x \tag{2.29}$$

daar $\frac{\partial f}{\partial x} = f(x)$ en

$$\operatorname{Var}(Y) = \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var}(X)$$
 (2.30)

 $= (e^x)^2 \operatorname{Var}(X)$

delen door Y^2 geeft:

$$\frac{\operatorname{Var}(Y)}{Y^2} = \operatorname{Var}(X) \tag{2.31}$$

en dus

$$RF(e^x) = AF(x) \tag{2.32}$$

ii) Beschouw de natuurlijke logaritmische functie:

$$y = f(x) = \ln x \tag{2.33}$$

dan is

$$\operatorname{Var}(Y) = \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var}(X) = \left(\frac{1}{x}\right)^2 \operatorname{Var}(X)$$
(2.34)

of

$$AF(\ln x) = RF(x) \tag{2.35}$$

iii) Analoog voor de tiendelige logaritmische functie:

$$y = f(x) = \log_{10} x = 0.4343 \ln x$$

$$\operatorname{Var}(Y) = \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var}(X) = \left(\frac{0.4343}{x}\right)^2 \operatorname{Var}(X)$$
(2.36)

of

$$AF(\log_{10} x) = 0.4343RF(x) \tag{2.37}$$

2.2.3 Concrete voorbeelden van foutenpropagatie in de chemie

2.2.3.1 Volumetrie: bepaling van de fout op de onbekende concentratie van een oplossing

Bij het bepalen van de onbekende titer N_A van een oplossing A met een bekende titer N_B van oplossing B worden een aantal meetresultaten gecombineerd [6], [7]. De lineaire foutenpropagatieformules (paragraaf 2.2.2) worden toegepast in de veronderstelling dat deze meetresultaten onafhankelijk zijn. Bovendien hebben de individuele metingen een

niet al te grote variantie (zie verder paragraaf 2.3). Uitgaande van de titerformule:

$$N_A = \frac{N_B V_B}{V_A} \tag{2.38}$$

worden de varianties op de afzonderlijke factoren, N_B en V_A , berekend alvorens de variantie op N_A te berekenen. In de praktijk zijn tenminste vier waarden (brutotitratie en drie precieze titraties) beschikbaar voor V_A , bepaald door het toegevoegd volume uit een buret met een precisie van 0.05 ml. In termen van figuur A.3 (Appendix A) vormen de waarden voor V_A (samen met hun gemiddelde) de eerste rij data. In principe kan een tweede rij, een derde rij enz. opgebouwd worden door de metingen te herhalen onder dezelfde omstandigheden en het gemiddelde van elke rij in de vierde kolom te schrijven. Op die manier kan de variantie van het gemiddelde volume \overline{V} geschat worden door de standaarddeviatie $\sigma_{\overline{V}}$ van de laatste kolom te berekenen.

Deze procedure is uiteraard onpraktisch, men schat $\sigma_{\overline{V}}$ daarom door de standaarddeviatie van de actuele metingen te berekenen en deze te delen door het aantal observaties, m.a.w. men schat de variantie van de laatste kolom (het gemiddelde) door de variantie van de eerste rij te berekenen.

De precieze titraties leveren volgend resultaat voor V_A (in ml):

$$25.21$$
 25.25 25.23

met het gemiddelde \overline{V}_A en de fout:

$$\overline{V}_A = 25.23 \text{ ml} \tag{2.39}$$

$$\sigma(\overline{V}_A) = \sqrt{\frac{8.10^{-4}}{3(3-1)}} = 0.01 \text{ ml}$$
(2.40)

zodat

$$AF(\overline{V}_A) = 3\sigma(\overline{V}_A) = 0.03 \text{ ml}$$
(2.41)

De variantie op de normaliteit N_B wordt bepaald via de relatieve fout, daar:

$$N_B = \frac{\left(\frac{3.041}{M_r/2}\right)}{0.5000} = \frac{\left(\frac{3.041}{126.066/2}\right)}{0.5000} = 0.09649 \text{ geq } 1^{-1}$$
(2.42)

Wanneer bijvoorbeeld 3.041 g (met absolute fout 0.001/2 g) gehydrateerd oxaalzuur opgelost wordt in water en aangelengd wordt tot exact 500.00 ml (maatkolf klasse A

met absolute fout 0.25 ml, de molaire massa wordt als foutloze constante beschouwd), volgt:

$$RF(N_B) = \sqrt{\left(\frac{0.5}{3041}\right)^2 + \left(\frac{0.25}{500.00}\right)^2} = 0.000526 \tag{2.43}$$

en

$$AF(N_B) = 0.000526 \times 0.09649 = 0.00005 \text{ geq } l^{-1}$$
 (2.44)

Tenslotte is de normaliteit N_A en de fout:

$$N_A = \frac{25.00 \times 0.09649}{25.23} = 0.09561 \text{ geq } 1^{-1}$$
(2.45)

De absolute fout op de 25 ml-pipet (volume gepipetteerd V_B) is 0.03 ml (klasse A glaswerk), zodat:

$$RF(N_A) = \sqrt{\left(\frac{0.03}{25.23}\right)^2 + \left(\frac{0.03}{25.00}\right)^2 + \left(\frac{0.00005}{0.09649}\right)^2} = 1.8 \times 10^{-3}$$
(2.46)

en

$$AF(N_A) = 1.8 \times 10^{-3} \times 0.09561$$

= 0.00017 geg 1⁻¹ (2.47)

De normaliteit N_A wordt dan gegeven door:

$$N_A = (95.61 \pm 0.17) \times 10^{-3} \text{ geq } l^{-1}$$
(2.48)

Merk op dat wanneer de onderzoeker een experiment bijvoorbeeld in drievoud uitvoert hij niet de twee "beste" resultaten, dit wil zeggen deze die het dichtst bij elkaar liggen mag selecteren, daar dit kan resulteren in een systematische fout. Dit wordt duidelijk in figuur 2.1.





2.2.3.2 Calorimetrie: bepaling van de fout op de verbrandingsenthalpie van een alcohol

De lineaire foutenpropagatieformules worden toegepast, dezelfde benaderingen in acht genomen, als in het volumetrisch voorbeeld [6], [7] (zie verder paragraaf 2.3).

De enthalpieverandering te wijten aan de verbranding van een exact gekende hoeveelheid van bijvoorbeeld ethanol volgt uit:

$$\Delta H^{\circ} = -C\Delta T \frac{M_{r \ ethanol}}{\Delta g} \tag{2.49}$$

met C, de warmtecapaciteit van de calorimeter bepaald door een gekende hoeveelheid van bijvoorbeeld 1-propanol te verbranden, ΔT , het temperatuurincrement te wijten aan de verbranding en Δg , het gewichtsverschil ten gevolge van de verbranding. De variantie op C volgt uit:

$$C = -\frac{\Delta H^{\circ} \times \Delta g}{M_{r \ 1-propanol} \times \Delta T} \tag{2.50}$$

met gekende $\Delta H^{\circ} = (-2010.4 \pm 0.1) \text{ kJ mol}^{-1}$ voor de verbranding van 1-propanol. Een in duplo uitgevoerd experiment met 1-propanol levert bijvoorbeeld de volgende resultaten:

$$C_1 = \frac{2010.4 \times 0.764}{60.033 \times 8.3} = 3.0825 \cong 3.1 \text{ kJ K}^{-1}$$
(2.51)

en

$$C_2 = \frac{2010.4 \times 0.809}{60.033 \times 8.9} = 3.0440 \cong 3.0 \text{ kJ K}^{-1}$$
(2.52)

Hieruit volgt dat $\overline{C} = 3.0633 \text{ kJ K}^{-1}$

en

$$RF(C_1) = \sqrt{\left(\frac{0.7}{764}\right)^2 + \left(\frac{0.07}{8.3}\right)^2 + \left(\frac{0.1}{-2010.4}\right)^2} = 0.0085$$

$$RF(C_2) = \sqrt{\left(\frac{0.7}{809}\right)^2 + \left(\frac{0.07}{8.9}\right)^2 + \left(\frac{0.1}{-2010.4}\right)^2} = 0.0079$$
(2.53)
(absolute fout op ΔT en Δg berekend volgens vergelijking 2.17)

$$AF(C_1) = 0.0085 \times 3.1 = 0.026 \cong 0.03 \text{ kJ K}^{-1}$$

$$AF(C_2) = 0.0079 \times 3.0 = 0.024 \cong 0.02 \text{ kJ K}^{-1}$$
(2.54)

zodat

$$AF(\overline{C}) = AF\left(\frac{C_1 + C_2}{2}\right) = \frac{1}{2}\sqrt{(0.03)^2 + (0.02)^2} = 0.018$$

$$\cong 0.02 \text{ kJ K}^{-1}$$
(2.55)

De warmtecapaciteit van de calorimeter wordt gegeven door:

$$C = (3.06 \pm 0.02) \text{ kJ K}^{-1}$$
 (2.56)

Merk op dat in dit geval de voorkeur gegeven wordt aan de accuratesse zoals bepaald door de statistische foutenanalyse niettegenstaande het aantal beduidende cijfers een grotere marge van inaccuratesse aangeeft. De reden waarom we slechts twee beduidende cijfers na het decimaal teken weerhouden in het eindresultaat van C moet gezocht worden in de bepaling van T (8.3 K en 8.9 K) wat dicht genoeg bij 10.0 K ligt om een resultaat met drie beduidende cijfers te aanvaarden.

De resulterende variantie op ΔH° volgt uit (experiment met ethanol in duplo uitgevoerd):

$$\Delta H_1^{\circ} = -3.06 \times 8.8 \times \frac{46.021}{0.942} = -1315.6 \text{ kJ mol}^{-1}$$
(2.57)

en

$$\Delta H_2^{\circ} = -3.06 \times 9.1 \times \frac{46.021}{0.898} = -1427.1 \text{ kJ mol}^{-1}$$
(2.58)

De overeenstemmende fouten zijn:

$$RF(\Delta H_1^{\circ}) = \sqrt{\left(\frac{0.02}{3.06}\right)^2 + \left(\frac{0.07}{8.8}\right)^2 + \left(\frac{0.7}{942}\right)^2} = 0.0103$$
(2.59)

en

$$RF(\Delta H_2^{\circ}) = \sqrt{\left(\frac{0.02}{3.06}\right)^2 + \left(\frac{0.07}{9.1}\right)^2 + \left(\frac{0.7}{898}\right)^2} = 0.0101$$
(2.60)

zodat

$$AF(\Delta H_1^\circ) = 0.0103 \times (-1315.6) \cong -14 \text{ kJ mol}^{-1}$$

$$AF(\Delta H_2^{\circ}) = 0.0101 \times (-1427.1) \cong -14 \text{ kJ mol}^{-1}$$

en

$$\overline{\Delta H^{\circ}}$$
 = -1371.4 kJ mol⁻¹

$$AF(\overline{\Delta H^{\circ}}) = \frac{1}{2}\sqrt{(-14)^2 + (-14)^2} \cong 10 \text{ kJ mol}^{-1}$$

De verbrandingsenthalpie van ethanol wordt dan gegeven door:

$$\Delta H^{\circ} = (-1371 \pm 10) \text{ kJ mol}^{-1}$$
(2.62)

2.2.3.3 Reactiekinetiek: bepaling van de fout op de reactiesnelheidsconstante

De volgende reactie:

$$S_2O_8^{2-}(aq) + 2I^-(aq) \to 2SO_4^{2-}(aq) + I_2(aq)$$

is een pseudo-eerste orde reactie wanneer de I^- concentratie in overmaat is:

$$v = k'[S_2 O_8^{2-}] \tag{2.63}$$

met $k' = k[I^-]$.

In tabel 2.1 zijn de data weergegeven van een kinetische studie van de vermelde reactie bij 25 °C.

Men start met een gekende concentratie aan $K_2S_2O_8$ en volgt de afname van de concentratie in functie van de tijd [8]. Een maat voor deze afname is de vorming van I_2 . I_2 wordt gedoseerd door titratie met $Na_2S_2O_3$ met zetmeel als indicator (zie tabel 2.1).

25

(2.61)

Tabel 2.1

tijd	[I2]	$[S_2 0_8^{2-}]$	$\ln[S_2 O_8^{2-}]$
(\min)	$(10^{-3} \text{ mol } l^{-1})$	$(10^{-3} \text{ mol } l^{-1})$	· . · · · ·
2.11	1.473	6.045	-5.1085
5.03	2.723	4.795	-5.3402
7.03	3.404	4.114	-5.4934
12.08	4.724	2.794	-5.8803
20.04	5.989	1.529	-6.4831
30.08	6.812	0.706	-7.2559
40.07	7.213	0.305	-8.0952
55.12	7.427	0.091	-9.3047

Kinetische studie bij 25 °C van de reactie $S_20_8^{2-}(aq) + 2I^-(aq) \rightarrow 2SO_4^{2-}(aq) + I_2(aq)$

Uit het feit dat men te maken heeft met een eerste-orde reactie volgt:

$$[S_2 O_8^{2-}] = \alpha e^{-kt} \tag{2.64}$$

met $\alpha = [S_2 O_8^{2-}]_0$ en

$$\ln[S_2 O_8^{2-}] = \beta - kt \tag{2.65}$$

met $\beta = \ln \alpha$. Figuur 2.2 toont het verloop van $\ln [K_2 S_2 O_8]$ in functie van de tijd, t.

Een regressieanalyse van de data [9], [10] aan de hand van de formules vermeld in paragraaf A.8 (Appendix A), levert de volgende schatting op voor de helling en intercept van de rechte met hun respectievelijke standaardafwijking:

- intercept = β = -4.9185 (σ = 0.0143)
- helling = $k = 0.07914 \text{ min}^{-1}$ ($\sigma = 0.00052 \text{ min}^{-1}$)

De standaardafwijking σ is een maat voor de variantie indien de coëfficiënten berekend zijn voor een oneindige herhaling van het experiment. Vandaar dat in overeenstemming met de definitie van de theoretische absolute (statistische) fout, drie maal deze standaardafwijking de absolute fout op de coëfficiënt oplevert.



Figuur 2.2: Verloop van ln $[K_2S_2O_8]$ versus t

Voor de absolute fout op de helling, k geeft dit:

$$AF(k) = 3\sigma(k) = 3 \times 0.00052 \cong 0.002 \text{ min}^{-1}$$
(2.66)

zodat

$$k = (0.079 \pm 0.002) \min^{-1}$$
 (2.67)

De resultaten van de regressie analyse laten tevens toe de fout op de initiële concentratie, $\alpha = [S_2O_8^{2-}]_0$ te schatten [6], [7].

De begin
concentratie $\alpha=e^{\beta}=7.310\times 10^{-3} {\rm mol}\ {\rm l}^{-1},$ zodat uit vergelijking 2.32 volgt dat

$$RF(\alpha) = AF(\beta) \tag{2.68}$$

Uit de definitie van de theoretische absolute fout volgt:

$$AF(\beta) = 3 \times 0.0143 = 0.04 \tag{2.69}$$

zodat

$$AF(\alpha) = 0.04 \times 7.310 \times 10^{-3} = 0.3 \times 10^{-3} \text{ mol } 1^{-1}$$
(2.70)

waaruit

$$[S_2 O_8^{2-}]_0 = (7.3 \pm 0.3) \times 10^{-3} \text{ mol } l^{-1}$$
(2.71)

2.3 Foutenpropagatie in gesimuleerde experimenten

2.3.1 Bepalen van de grenzen voor lineaire foutenpropagatie

In de vorige secties werd uiteengezet hoe de techniek van *linearisering* toelaat om benaderende waarden te vinden voor de fout op relatief ingewikkelde functies f van onnauwkeurig gekende grootheden x_1, \ldots, x_n . Zoals gezegd is deze aanpak gebaseerd op de onderstelling dat lineaire benaderingen van de overeenkomstige functies voldoende nauwkeurig zijn om het gedrag binnen de te verwachten intervallen te beschrijven. In deze sectie zal met behulp van simulaties nagegaan worden binnen welke grenzen de lineaire foutenpropagatieformules vermeld in paragraaf 2.2.2 geldig zijn.

In eerste instantie worden problemen gebaseerd op zogenaamde gesloten formules van de vorm $y = f(x_1, x_2, \ldots, x_n)$, geanalyseerd. Vervolgens wordt de foutenpropagatie gesimuleerd in een voorbeeld waarbij de experimentele waarde y gedefinieerd wordt aan de hand van de impliciete karakterisatie:

$$y = \min_{x} f(x_1, x_2, \dots, x_n)$$
 (2.72)

De gevolgde strategie bestaat er in de experimenten vermeld in paragrafen 2.2.3.1, 2.2.3.2 en 2.2.3.3 te simuleren met behulp van MatLabTM functies (= M-files): de foutenpropagatie op basis van linearisering wordt vergeleken met de fout bekomen door het experiment een "groot" aantal maal te herhalen. Zoals in Appendix A paragraaf A.1 gesteld wordt, resulteert dit in een aantal kolommen $X_1, X_2, X_3, ..., X_n$ met in elke kolom de resultaten van respectievelijk de eerste, de tweede, de derde, ..., *n*-de opmeting van elke reeks.

Volumetrie

De gesloten formule in deze toepassing (M-file 'volumetrics.m': programma op aanvraag te verkrijgen) is van de vorm $N_A = N_B V_B / V_A$ met N_B de normaliteit van de bekende oplossing B, V_B het volume van B bepaald door de pipetinhoud, V_A het volume van A bepaald door het toegevoegd volume uit een buret en N_A de onbekende normaliteit van de oplossing A.

Als input worden ingegeven: N_B , V_B , V_A , AF_{N_B} , AF_{V_B} , AF_{V_A} , aantal experimenten.

De output van 'volumetrics.m' bevat:

- AF_{N_A-lin} op basis van lineaire foutenpropagatieformules;
- vector $N_{A-simul}$ met lengte = aantal experimenten, $\overline{N_{A-simul}}$ en $AF_{N_A-simul}$ op basis van simulaties met respectievelijk 100, 1000, 5000 en 10000 experimenten waarbij aan iedere factor in de gesloten formule een random, normaalverdeling wordt gekoppeld ~ $N(\mu, \sigma^2)$ met μ de inputwaarde van de factor en σ de absolute fout op de betrokken factor.

Tabel 2.2 geeft een overzicht van de fout bekomen in gesimuleerde volumetrische experimenten in vergelijking met deze op basis van lineaire foutenpropagatie.

De tabel toont aan dat AF_{N_A-lin} en $AF_{N_A-simul}$ goed overeenkomen naarmate het aantal experimenten in de simulatie stijgt.

Tabel 2.2 Vergelijking lineaire foutenpropagatieformules met foutenpropagatie in gesimuleerde volumetrische experimenten

Formule : $N_A = N_B V_B / V_A$

ь. -		lineaire fouten- propagatieformules	_		gesim foutenp		_			
$\begin{array}{l} AF_{V_B} = AF_{V_A} \\ (\mathrm{ml}) \end{array}$	AF_{N_B} (×10 ⁻³ geq l ⁻¹)	$\begin{array}{c} AF_{N_A-lin} \\ (\times 10^{-3} \ \mathrm{geq} \ \mathrm{l}^{-1}) \end{array}$	$(\overline{N_A} - \pm AF_{N_A} \times 10^{-3})$ 100 expe	simul -simul) geq 1 ⁻¹ rimenten	$(\overline{N_A} - \pm AF_{N_A} \times 10^{-3}$ 1000 expe	simul (-simul) geq l ⁻¹ erimenten	$(\overline{N_A} - \pm AF_{N_A} \times 10^{-3}$ 5000 expe	simul -simul) geq 1 ⁻¹ erimenten	$(\overline{N_A} - \pm AF_{N_A})$ $\times 10^{-3}$ 10000 exp	simul (a-simul) geq 1 ⁻¹ perimenten
0.03	0.05	0.17	95.62	0.17	95.61	0.17	95.61	0.17	95.61	0.17
0.3	0.5	1.7	95.6	1.9	95.7	1.8	95.6	1.7	95.6	1.7
2	3	8.4	95.8	8.8	95.7	8.7	95.6	8.5	95.6	8.4
3	5	17	96	18	96	17	96	17	96	17
0.003	0.005	0.017	95.611	0.016	95.610	0.017	95.610	0.017	95.610	0.017
0.0006	0.001	0.0034	95.6104	0.0035	95.6104	0.0033	95.6104	0.0034	95.6104	0.0034
0.0003	0.0005	0.0017	95.6104	0.0016	95.6104	0.0017	95.6104	0.0017	95.6104	0.0017





Figuur 2.3(a-d) toont de normaliteitsplot van de kolomvector $N_{A-simul}$ voor respectievelijk 100, 1000, 5000 en 10000 experimenten in de simulatie. Deze plots illustreren dat het normale karakter van de verdeling meer geaccentueerd wordt naarmate het aantal experimenten in de simulatie stijgt. Bij 10000 simulaties is het verloop van de curve nagenoeg lineair (zie Appendix A).

Calorimetrie

Op analoge wijze wordt het calorimetrisch experiment uit paragraaf 2.2.3.2 gesimuleerd in het MatLabTM - programma 'enthalpy.m' (programma op aanvraag te verkrijgen). De gehanteerde formule is $\Delta H^{\circ} = -C\Delta T M_r/\Delta g$. Als input worden ingegeven: relatieve molaire massa M_r , warmtecapaciteit C, temperatuurincrement ΔT , gewichtsverschil Δg ten gevolge van verbranding, AF_C , $AF_{\Delta T}$, $AF_{\Delta g}$, aantal experimenten. Als output bekomt men:

- AF_{ΔH°-lin} op basis van lineaire foutenpropagatieformules;
- vector $\Delta H^{\circ} simul$ met lengte = aantal experimenten in de simulatie, $\overline{\Delta H^{\circ} - simul}$ en $AF_{\Delta H^{\circ} - simul}$ op basis van simulaties met respectievelijk 100, 1000, 5000 en 10000 experimenten waarbij aan iedere factor in de formule een random, normaalverdeling wordt gekoppeld $\sim N(\mu, \sigma^2)$ met μ de inputwaarde van de factor en σ de absolute fout op de betrokken factor.

In tabel 2.3 wordt een schematisch overzicht gegeven van de fout bekomen in gesimuleerde calorimetrische experimenten in vergelijking met deze op basis van lineaire foutenpropagatie. De tabel toont eveneens een goede overeenkomst tussen $AF_{\Delta H^{\circ}-lin}$ en $AF_{\Delta H^{\circ}-simul}$. Merken we op dat $\overline{\Delta H^{\circ}-simul}$ ongeveer 0.5% verschilt van $\overline{\Delta H^{\circ}}$ berekend op basis van het in duplo uitgevoerde experiment.

De normaliteitsplots van de kolomvector $\Delta H^{\circ} - simul$ voor respectievelijk 100, 1000, 5000 en 10000 simulaties (figuur 2.4 (a-d)) accentueren het normale karakter van de data naarmate het aantal experimenten in de simulatie stijgt.

Tabel 2.3 Vergelijking lineaire joutenpropagatiejormules met joutenpropagatie in gesimuleerue calorimetrische exp	erde calorimetrische experimente	gesimuleerde	foutenpropagatie in	foutenpropagatieformules r	lineaire	Vergelijking	Tabel 2.3
---	----------------------------------	--------------	---------------------	----------------------------	----------	--------------	-----------

Formule : $\Delta H^{\circ} = -C\Delta T M_{\tau}/\Delta g$

	lineaire fouten- propagatieformules						gesin foutenp	uleerde propagatie			
$AF_C(kJ K^{-1})$	$AF_{\Delta T}(\mathbf{K})$	$\begin{array}{c} AF_{\Delta g} \\ (\times 10^{-2} \text{ g}) \end{array}$	$AF_{\Delta H^\circ - lin}$ (×10 ² kJ mol ⁻¹)	$(\overline{\Delta H^\circ} - \pm AF_{\Delta H^\circ})$ $\times 10^2 \text{ kJ}$ 100 exper	simul -simul) mol ⁻¹ imenten	$(\overline{\Delta H^\circ} - \pm AF_{\Delta H^\circ})$ $\pm AF_{\Delta H^\circ}$ $\times 10^2 \text{ kJ}$ 1000 expe	- <i>simul</i> 9- <i>simul</i>) 1 mol ⁻¹ erimenten	$(\overline{\Delta H^\circ} - \pm AF_{\Delta H^\circ} + 10^2 \text{ kJ}$ 5000 expe	simul -simul) mol ⁻¹ rimenten	$(\overline{\Delta H^{\circ}} - \pm AF_{\Delta H'} \times 10^2 \text{ kJ})$ 10000 exp	- simul °-simul) I mol ⁻¹ erimenten
0.02	0.07	0.07	0.14	-13.63	0.15	-13.62	0.15	-13.62	0.14	-13.62	0.14
0.2	0.7	0.7	1.4	-13.64	1.4	-13.62	1.4	-13.61	1.4	-13.63	1.4
1	4	4	7.0	-13.67	7.6	-13.74	6.9	-13.61	7.0	-13.62	7.0
0.002	0.007	0.007	0.014	-13.623	0.014	-13.623	0.015	-13.623	0.014	-13.623	0.014
0.0004	0.001	0.001	0.0028	-13.6232	0.0028	-13.6232	0.0028	-13.6232	0.0028	-13.6232	0.0028
0.0002	0.0007	0.0007	0.0014	-13.6232	0.0015	-13.6232	0.0014	-13.6232	0.0014	-13.6232	0.0014

a)

b)



a) 100 experimenten in de simulatie, b) 1000 experimenten



c) 5000 experimenten, d) 10000 experimenten

Kinetica

In het MatLabTM - programma 'kinetics.m' wordt het kinetisch experiment uit paragraaf 2.2.3.3 gesimuleerd.

De waarde voor de reactiesnelheidsconstante, k wordt gevonden via de impliciete karakterisatie $k = f_{min}(C, t)$ met C de concentratie en t de tijd. Het programma vereist als input het aantal experimenten, AF_t , AF_C , C_0 (7.310.10⁻³ mol 1⁻¹), $k_{initieel}$ (0.070 min⁻¹) en geeft als output de kolomvector k - simul met lengte = aantal experimenten in de simulatie, $\overline{k - simul}$, AF_{k-lin} en $AF_{k-simul}$. Op dezelfde wijze als in de hoger vermelde simulaties wordt het experiment 100, 1000, 5000 en 10000 maal gesimuleerd waarbij aan elke factor in de impliciete vergelijking een randomnormale foutenverdeling gekoppeld wordt.

Tabel 2.4 geeft schematisch de vergelijking weer tussen de lineaire foutenpropagatie met de foutenpropagatie in het gesimuleerde kinetische experiment. Het valt op dat AF_{k-lin} en $AF_{k-simul}$ zelfs bij een groot aantal experimenten in de simulatie niet meer overeenkomen. De normaliteitsplot van de vector k - simul (figuur 2.5(a-d)) vertoont een afwijkend (rechtsscheef) lineair gedrag (zie Appendix A) zelfs bij een groot aantal (10000) experimenten in de simulatie, zodat de data in k - simul niet normaal verdeeld zijn. Bovendien blijkt uit deze simulatiestudie dat de initiële schatter voor k vertekend is, hetgeen aangetoond kan worden vertrekkende van het foutenmodel [11]:

$$C = \gamma + \varepsilon \tag{2.73}$$

met C de experimenteel gemeten concentratie, γ de "echte" concentratie en de fout $\varepsilon \sim N(0, \sigma^2), \sigma^2$ onafhankelijk van C. De impliciete vergelijking voor k: $C = C_0 e^{-kt}$ of $\ln C = \ln C_0 - kt$ kan dus geschreven worden als:

$$\ln(\gamma + \varepsilon) = \ln C_0 - kt \tag{2.74}$$

of

$$\ln(\gamma + \varepsilon) = \ln\left(\gamma\left(1 + \frac{\varepsilon}{\gamma}\right)\right)$$
$$= \ln\gamma + \ln\left(1 + \frac{\varepsilon}{\gamma}\right) \quad \left|\frac{\varepsilon}{\gamma}\right| \ll 1 \quad (2.75)$$
$$\cong \ln\gamma + \frac{\varepsilon}{\gamma} = \ln\gamma + \delta$$

Tabel 2.4	Vergelijking	lineaire .	fout enpropagatie formules	met	foutenpropagatie	in	gesimuleerde	kinetische	experimenten
-----------	--------------	------------	----------------------------	-----	------------------	----	--------------	------------	--------------

		lineaire foutenpropagatie- formules		gesimuleero	le fouten	propagatie				
$\begin{array}{c} AF_t \\ (\times 10^{-3} \text{ min}) \end{array}$	$\begin{array}{c} AF_C \\ (\times 10^{-3} \text{ mol } l^{-1}) \end{array}$	$AF_{k-lin}(\min^{-1})$	$(\overline{k} - \pm AF_{k-s})$ 100 exp	- simul _{imul}) min ⁻¹ perimenten	$(\overline{k - simul})$ $\pm AF_{k-simul}) \min^{-1}$ 1000 experimenten		$(\overline{k-simul})$ $\pm AF_{k-simul}) \min^{-1}$ 5000 experimenten		$(\overline{k-simul})$ $\pm AF_{k-simul}) \min^{-1}$ 10000 experimenten	
1.1	0.1	0.0020	0.0705	0.0148	0.0705	0.0154	0.0708	0.0150	0.0708	0.0149







waarbij $\delta = \frac{\varepsilon}{\gamma}$, vandaar $\delta \sim N\left(0, \frac{\sigma^2}{\gamma^2}\right)$ of de fout wordt groter naarmate de concentratie daalt (figuur 2.6).

Figuur 2.6 toont de simulatie van een kinetisch experiment (5000 simulaties). De betrouwbaarheidsintervallen zijn drie maal de berekende standaardafwijking en worden duidelijk groter naarmate C daalt of t stijgt.

Daar de logaritmische functie een convexe functie is, geldt (zie Appendix A):

$$E(\ln X) \le \ln E(X) \tag{2.76}$$

indien X = C





Figuur 2.6: Gesimuleerd kinetisch experiment (5000 simulaties) met betrouwbaarheidsintervallen drie maal de berekende standaardafwijking

en verder volgens vergelijking (Appendix A.25) met $f(x) = \ln(x)$ en dus $f''(x) = -\frac{1}{x^2}$:

$$E(\ln C) = E(\ln(\gamma + \varepsilon))$$

$$\cong \ln(E(\gamma + \varepsilon)) - \frac{1}{2} \frac{\operatorname{Var}(\varepsilon)}{\gamma^2}$$

$$\cong \ln \gamma - \frac{1}{2\gamma^2} \sigma^2$$
(2.78)

met $C \sim N(\gamma, \sigma^2)$ zodat systematisch een te lage concentratie of te hoge k-waarde wordt gevonden (tabel 2.5).

Een extra controle van vergelijking (2.78) aan de hand van MatlabTM (10000 simulaties) geeft voor $E(\ln C) = -5.0772$ op t = 2.11 min en $E(\ln C) = -7.7591$ op t = 40.07 min.

Dit resultaat stemt overeen met de berekende waarde voor $E(\ln C)$ aan de hand van vergelijking (2.78):

• op
$$t = 2.11 \text{ min}, \gamma = 6.247.10^{-3} \text{ mol } 1^{-1}, \sigma^2 = 10^{-8}$$

 $E(\ln C) = \ln(6.247.10^{-3}) - \frac{1}{2} \frac{10^{-8}}{(6.247.10^{-3})^2} = -5.0758$

• op $t = 40.07 \text{ min}, \gamma = 0.438.10^{-3} \text{ mol } l^{-1}, \sigma^2 = 10^{-8}$

$$E(\ln C) = \ln(0.438.10^{-3}) - \frac{1}{2} \frac{10^{-8}}{(0.438.10^{-3})^2} = -7.7594$$

In tabel 2.5 wordt voor respectievelijk 1000, 5000 en 10000 simulaties de berekende k-waarde (= gemiddelde waarde over het totaal aantal experimenten) telkens vergeleken met een initiële schatter voor k; hieruit blijkt de kleine maar systematische trend k-berekend > k-initieel.

Tabel 2.5 Vergelijking berekende k met initiële schatter voor k in gesimuleerde kinetische experimenten

k-initial (min ⁻¹)	k-berekend (min ⁻¹) bij 1000 simulaties	k-berekend (min ⁻¹) bij 5000 simulaties	k-berekend (min ⁻¹) bij 10000 simulaties
0.0600	0.0601	0.0603	0.0603
0.0650	0.0653	0.0655	0.0656
0.0700	0.0705	0.0707	0.0708
0.0750	0.0757	0.0759	0.0759

Referenties

- [1] J. Couzin, Science, 281, 762, (1998)
- [2] W.J. Kaufmann III, L. Smarr, Supercomputing and the Transformation of Science, Chapter 1, Scientific American Library, (1993)
- [3] R.Y. Rubinstein, B. Melamed, Modern Simulation and Modeling, Chapter 1, John Wiley & Sons, New York, (1998)
- [4] J. Klaessens, H. Van Leeuwen, Praktische statistiek voor het laboratorium, hoofdstuk 2, ten Hagen & Stam, Den Haag, (1996)
- [5] D.S. Moore, G.P. McCabe, Statistick in de Praktijk, hoofdstuk 7, 2e herziene druk, Academic Service, Schoonhoven, (1997)
- [6] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 776, (1993)
- [7] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 838, (1993)
- [8] M.R.K. Hemalatha, I. NoorBatcha, J. Chem. Educ., 74, 972, (1997)
- [9] MATLAB, Reference Guide, the MATHWORKS Inc., Natick, (1995)
- [10] MATLAB, Statistics Toolbox User's Guide, the MATHWORKS Inc., Natick, (1996)
- [11] T.J. Zielinski, R.D. Allendoerfer, J. Chem. Educ., 74, 1001, (1997)



Deel II

Modelleren van potentiometrische data: de bepaling van stabiliteitsconstanten



Hoofdstuk 3 Potentiometrische experimenten

3.1	Theor	etische beschouwingen: inleiding
	3.1.1	Potentiometrie in de evenwichtsanalyse
	3.1.2	De methode van de constante hoge ionensterkte
	3.1.3	De voordelen van de potentiometrie
	3.1.4	Ontwikkelingen in de potentiometrie
3.2	De po	tentiometrische techniek
	3.2.1	Inleiding
	3.2.2	Potentiaalmetingen
		3.2.2.1 Evenwichtscriteria
		3.2.2.2 De geautomatiseerde meetopstelling en datatransmissie .
	3.2.3	Het indifferent elektrolyt
	3.2.4	De elektroden
		3.2.4.1 De Ag/AgCl referentieelektrode
		3.2.4.2 De zoutbrug
		3.2.4.3 De indicatorelektroden
	3.2.5	Ijking van de meetopstelling
		3.2.5.1 Ijking van de pipetten en de automatisch gestuurde buretten
		3.2.5.2 Ijking van de glaselektrode en bepaling van de opera-
		tionele pK_W
		3.2.5.3 Ijking van de zilverelektrode
	3.2.6	De klassieke analysemethode voor het bepalen van stabiliteitscon-
		stanten
		3.2.6.1 Binaire systemen: de bepaling van de zuurconstanten van
		de alifatische diaminen 1,4-diaminobutaan en 1,5-diaminopentaan
		en de zilverhydrolyse in 1.0 M KNO3
		3.2.6.2 Ternaire systemen: de bepaling van de stabiliteitscon-
		stanten van de Ag(I)-complexen van 1,4-diaminobutaan en
		1,5-diaminopentaan in 1.0 M KNO ₃
	Refere	nties

Hoofdstuk 4 Kwalitatieve en kwantitatieve analyse van potentiometrische data met behulp van optimalisatieprogramma's

4.1	Inleidi	ng
4.2	Het op	otimalisatieprogramma SUPERQUAD
	4.2.1	Inleiding
	4.2.2	De modelfunctie en de kleinste kwadraten benadering
	4.2.3	Computationele methode
		4.2.3.1 Berekening van de vrije concentraties
		4.2.3.2 Optimalisatie van stabiliteitsconstanten
		4.2.3.3 De gewichtenmatrix
		4.2.3.4 Convergentiecriteria
	4.2.4	Implementatie van SUPERQUAD in MatLab TM
	4.2.5	Validatie van het MatLab TM programma
4.3	Interp	retatie van de resultaten
	4.3.1	Modelselectie
		4.3.1.1 Criteria
		4.3.1.2 Stroomdiagram
	4.3.2	Foutenanalyse
	Refere	enties

Hoofdstuk 5 De verhoudingsmethode

5.1	Berel	keningen met het standaard SUPERQUAD programma
	5.1.1	Binaire systemen
		5.1.1.1 De zuurconstanten van 1,4-diaminobutaan en 1,5-diaminopentaan:
		resultaten en discussie
		5.1.1.2 De hydrolyse constanten van Ag(I): resultaten en discussie
	5.1.2	Ternaire systemen
		$5.1.2.1$ De stabiliteits constanten van de ${\rm Ag}({\rm I}){\rm -complexen}$ van
		1,4-diamin obutaan en 1,5-diaminopentaan: resultaten en discussie
5.2	Berel	keningen met de MatLab TM implementatie van SUPERQUAD
	5.2.1	Binaire systemen
	5.2.2	Ternaire systemen
	Refere	enties

Hoofdstuk 3

Potentiometrische experimenten

3.1 Theoretische beschouwingen: inleiding

Zoals gesteld in de algemene inleiding heeft dit onderzoek tot doel op een pragmatische manier de propagatie van (toevallige) fouten bij de potentiometrische bepaling van stabiliteitsconstanten te kwantificeren. Hierbij is een systematische analyse van enerzijds de gebruikte geautomatiseerde meettechniek — de potentiometrie — en anderzijds de computationele methode — de niet-lineaire kleinste kwadraten methode toegepast in optimalisatieprogramma's zoals SUPERQUAD (zie hoofdstuk 4) — [1], [2], [3], [4] essentieel. De betrouwbaarheid van de berekeningen, m.a.w. beschrijven de berekeningen het chemisch systeem voldoende adequaat, wordt nagegaan op verschillende niveau's. Om dit doel te realiseren werd het volgende systeem potentiometrisch onderzocht: de protonerings- en Ag(I)-complexeringsevenwichten van 1,4-diaminobutaan en 1,5-diaminopentaan in 1.0 M KNO_3 .

3.1.1 Potentiometrie in de evenwichtsanalyse

Volgens de IUPAC–conventie (Stockholm 1953) wordt een elektrochemische cel voorgesteld als [5] :

+

referentiehalfcel | (|) meetoplossing | halfcel voor metaal, ligand of proton E_{ref} (E_j) E_{ind}

Een indicatorelektrode en een referentieelektrode, eventueel gecombineerd via een zoutbrug, vormen samen het elektrodekoppel dat reageert op de activiteit van één van de vrije deeltjes A volgens de algemene Nernstvergelijking [6], [7]:

$$E = E_{ind} - E_{ref} + E_j = E^{0'} + \frac{RT}{nF} \ln a_A + E_j$$
(3.1)

met a_A de activiteit van het deeltje A, E_j de diffusiepotentiaal van het elektrodekoppel, F de Faraday constante (96485.3 C mol⁻¹), n het aantal mol uitgewisselde elektronen, R de algemene gasconstante (8.31451 JK⁻¹ mol⁻¹) en T de absolute temperatuur uitgedrukt in kelvin.

De term $E^{0'}$ omvat de constante halfcelpotentiaal van de referentieelektrode, de standaardpotentiaal E^{0} van de indicatorelektrode en de asymmetriepotentiaal.

Bij gebruik van een grote hoeveelheid niet complexerend indifferent elektrolyt blijven de activiteitscoëfficiënten constant (zie verder onder 3.2.3), waardoor de Nernstvergelijking kan herschreven worden als:

$$E = E^{0''} + \frac{RT}{nF} \ln[A] + E_j$$
(3.2)

met

$$E^{0''} = E^{0'} + \frac{RT}{nF} \ln \gamma_A$$
 (3.3)

waarbij γ_A de activiteitscoëfficiënt van deeltje A voorstelt-

Rekening houdend met de diffusiepotentiaal kan men, na ijking van de meetopstelling, de vrije concentratie van deeltje A rechtstreeks bepalen.

Een beschrijving van de karakteristieken van de gebruikte elektroden en de ijking ervan wordt gegeven in paragrafen 3.2.4 en 3.2.5.

3.1.2 De methode van de constante hoge ionensterkte

Een stabiliteitsconstante is een thermodynamische grootheid die in het geval van een ternair systeem, zoals we in hoofdstuk 4 en 5 zullen tegenkomen, verwijst naar de vorming van een complex $M_pL_qH_r$ waarbij M het metaal voorstelt, L het ligand en H een proton.

Wiskundig beschrijft de stabiliteitsconstante bij constante druk en temperatuur het volgende evenwicht:

$$pM^{m+} + qL^{\ell-} + rH^+ \rightleftharpoons (M_p L_q M_r)^{pm+r-q\ell}$$

$$(3.4)$$

met

$$\beta_{pqr}^{0} = \frac{(a_{M_{p}L_{q}H_{r}})^{pm+r-q\ell}}{(a_{M}^{m+})^{p}(a_{L}^{\ell-})^{q}(a_{H}^{+})^{r}}$$
(3.5)

 β_{pqr}^0 is de thermodynamische evenwichtsconstante van het complex $M_p L_q H_r$ waarbij a_i de activiteit van het deeltje *i* bij evenwicht voorstelt en de coëfficiënten *p* en *q* positieve gehele getallen zijn. Is *r* positief dan worden geprotoneerde complexen gevormd, terwijl een negatieve *r* op de aanwezigheid van hydroxocomplexen duidt.

Met de stechiometrische coëfficiënten p, q en r kan men in een ternair systeem elk deeltje in oplossing beschrijven als $M_pL_qH_r$.

In tabel 3.1 worden voorbeelden van de "p, q, r" notatie weergegeven.

Tabel 3.1 De "p,q,r" notatie

р	q	r	Deeltje	Benaming van de constante
0	1	1	HL	de eerste protoneringsconstante van een ligand
1	0	-1	MOH	een hydrolyseconstante van een metaal
1	1	1	MLH	de constante van een geprotoneerd complex
1	1	-1	MLOH	de constante van een hydroxocomplex
1	2	0	ML_2	de constante van een binair mononucleair complex
2	2	2	$M_2L_2H_2$	de constante van een ternair polynucleair complex

De standaard Gibbs-vrije energieverandering ΔG^0 voor de vorming van het complex $M_p L_q H_r$ wordt gegeven door:

$$\Delta G^0 = -RT \ln \beta_{pqr}^0 \tag{3.6}$$

met

$$\beta_{pqr}^{0} = \frac{a_{M_{p}L_{q}H_{r}}}{(a_{M})^{p}(a_{L})^{q}(a_{H})^{r}}$$
(3.7)

Voor de eenvoud worden in deze en de volgende formules de ladingen weggelaten. De activiteit van een deeltje i kan in functie van de molaire concentratie geschreven worden als:

$$a_i = c_i \gamma_i$$
 (3.8)

waarbij c_i de molaire concentratie en γ_i de activiteitscoëfficiënt van het deeltje *i* zijn. In sterk verdunde oplossingen kan men met goede benadering activiteiten gelijk stellen

aan concentraties. Verschillende technieken om stabiliteitsconstanten te bepalen, waaronder spectroscopie, calorimetrie, polarografie en potentiometrie maken hiervan gebruik. De minimalisatieprogramma's om stabiliteitsconstanten te berekenen die verder in dit werk geëvalueerd worden zijn eveneens gebaseerd op deze benadering.

In geconcentreerde oplossingen echter zijn de concentraties van de aanwezige species meestal van die aard dat deze veronderstelling niet meer toegestaan is. In dit geval moet men de activiteitscoëfficiënten γ_i kennen om de activiteiten uit de bekende concentratie c_i of om de concentraties uit de gemeten activiteiten te berekenen. In het algemeen zijn de activiteitscoëfficiënten moeilijk te bepalen. Ze zijn sterk afhankelijk van de aard en de concentraties van alle deeltjes in oplossing. Pogingen om de coëfficiënten te berekenen leiden tot grove benaderingen.

Wanneer reacties echter uitgevoerd worden in een medium met hoge ionensterkte, wordt het probleem omzeild daar de activiteitscoëfficiënten dan nagenoeg constant zijn, zodat de globale stabiliteitsconstante van het complex $M_pL_qH_r$ eerder als een stechiometrische constante kan gedefinieerd worden:

$$\beta_{pqr} = \beta_{pqr}^0 / \left(\frac{\gamma_{M_p L_q H_r}}{(\gamma_M)^p (\gamma_L)^q (\gamma_H)^r} \right) = \frac{[M_p L_q H_r]}{[M]^p [L]^q [H]^r}$$
(3.9)

[M], [L] en [H] verwijzen naar de vrije concentraties van respectievelijk het metaalion, het ligand en het proton.

De activiteitscoëfficiënten en de globale stabiliteitsconstante refereren naar dezelfde molaire concentratieschaal. De standaardtoestand voor β_{pqr} wordt gedefinieerd als de hypothetische molaire oplossing van de deeltjes M, L en H in het gebruikte ionaire midden [6], zodat:

$$\Delta G_c = -RT \ln \beta_{pgr} \tag{3.10}$$

De methode van de constante hoge ionensterkte is gebaseerd op het principe van Brønsted van een constante ionenomringing [8], [9]. Volgens dit principe is de activiteitscoëfficiënt van elk aanwezig deeltje constant indien dit slechts een kleine fractie van de totale concentratie aan elektrolyt uitmaakt. Volgens Sillén blijft het principe van de constante ionenomringing geldig zolang de reagentia minder dan 10% van het medium uitmaken [10]. Hoewel het gebruik van een indifferent elektrolyt al teruggaat tot 1905 [11], hebben Sillén en Biedermann het principe pas in de vijftiger jaren systematisch uitgewerkt [12]. Zij stellen vast dat een ideaal inert elektrolyt aan welbepaalde eisen moet voldoen:

- het indifferent elektrolyt moet een sterk elektrolyt zijn;
- het anion mag niet associëren met het metaalion of met het complex, het kation mag niet associëren met het ligand of met het complex;
- er mogen geen (redox-)reacties optreden tussen het elektrolyt, het ligand, het metaalion of de complexen;
- het indifferent elektrolyt moet goed oplosbaar zijn in het gebruikte solvent;
- het indifferent elektrolyt moet een verwaarloosbare bijdrage hebben tot de te meten fysische of chemische eigenschappen.

Volgende binaire elektrolyten zijn bijzonder populair als indifferent elektrolyt: KCl, KNO_3 , $NaNO_3$, $NaClO_4$, NaCl. Meestal worden ionensterkten van 0.1, 0.2, 0.5, 1.0 en 3.0 M toegepast. In dit werk kiezen we voor KNO_3 als indifferent elektrolyt met een ionensterkte van 1.0 M. Verschillende auteurs duiden deze ionensterkte aan als het meest interessante milieu om stabiliteitsconstanten te bepalen [13], [14]. Het is tevens een compromis tussen de noodzaak om de activiteitscoëfficiënten constant te houden en de nadelen, verbonden met het toevoegen van een grote hoeveelheid indifferent elektrolyt. Bij hoge concentraties aan deze laatste stellen zich immers problemen op het vlak van de zuiverheid en/of oplosbaarheid van de zouten. Zwakke interacties tussen het indifferent elektrolyt en de metaalionen en/of het ligand worden bovendien belangrijker naarmate de concentratie van het indifferent elektrolyt toeneemt.

Tijdens de complexeringstituaties (zie verder onder 3.2.6.2) kunnen we de K^+ - of de NO_3^- -concentratie constant houden. Gezien de diaminen als geprotoneerde nitraatzouten in de oplossing gebracht worden en $AgNO_3$, commercieel verkrijgbaar is in pro analyse toestand, kiezen we voor een constante NO_3^- -concentratie in de oplossingen.

Wanneer complexen van de vorm $M_pL_qH_r$ gevormd worden, kunnen de massabalansen voor een drie-componenten systeem geschreven worden als:

$$C_M = [M] + \sum_{p,q,r} p[M_p L_q H_r] = [M] + \sum_{p,q,r} p\beta_{pqr} [M]^p [L]^q [H]^r$$
(3.11)

$$C_L = [L] + \sum_{p,q,r} q[M_p L_q H_r] = [L] + \sum_{p,q,r} q\beta_{pqr} [M]^p [L]^q [H]^r$$
(3.12)

$$C_H = [H] + \sum_{p,q,r} r[M_p L_q H_r] = [H] + \sum_{p,q,r} r\beta_{pqr} [M]^p [L]^q [H]^r$$
(3.13)

 C_M , C_L en C_H zijn de totale analytische concentraties van respectievelijk het metaalion, het ligand en het proton. De bepaling van de stabiliteitsconstanten voor een gegeven model wordt herleid tot het oplossen van een stelsel van bovenstaande vergelijkingen. Deze zijn uiteraard alleen geldig bij de gebruikte ionensterkte. Het ligt voor de hand dat het systeem nauwkeuriger bepaald kan worden naarmate er meer vrije concentraties van metaalionen, ligand, proton of complexen gekend zijn. In de massabalansen zijn de binaire complexen eveneens opgenomen. Hun vormingsconstanten kunnen vooraf bepaald worden in afzonderlijke experimenten.

De complexeringstitraties worden in een zo groot mogelijk concentratiegebied uitgevoerd zodat alle mogelijke complexen gevormd kunnen worden. Een voordeel van potentiometrische titraties is dat, na toevoegen van een kleine hoeveelheid oplossing aan een systeem in evenwicht, het systeem relatief snel een nieuwe evenwichtstoestand bereikt. De keuze van de titratietechniek is afhankelijk van de vrije concentraties die experimenteel bepaald kunnen worden en van de oplosbaarheid van de gevormde complexen.

3.1.3 De voordelen van de potentiometrie

Complexvorming leidt tot veranderingen in verschillende fysicochemische eigenschappen van een oplossing. In principe biedt de meting van elke eigenschap of parameter die verandert in functie van de complexeringsgraad de mogelijkheid om de samenstelling van de complexen te bepalen en de overeenkomstige vormingsconstanten te berekenen. In de praktijk moet echter aan een aantal welbepaalde voorwaarden voldaan zijn:

- de correlatie tussen de complexeringsgraad en de gemeten parameter(s) moet exact gekend zijn;
- de verandering in de parameter(s) ten gevolge van de complexvorming moet beduidend groter zijn dan de experimentele fout;
- de experimenten moeten uitgevoerd worden in een concentratiegebied waar de vorming van de complexen bestudeerd kan worden zonder dat de species reeds volledig gevormd zijn.

Een techniek is pas bruikbaar wanneer men het verband tussen de gemeten eigenschap en de concentratie van minstens één van de deeltjes van het onderzocht systeem kent. Het is daarbij van groot belang dat de analytische concentraties van alle componenten in de oplossing bekend zijn. Slechts wanneer uiterst nauwkeurige metingen uitgevoerd worden in een groot concentratiegebied kunnen accurate stabiliteitsconstanten berekend worden bij de aangegeven ionensterkte.

Voor een overzicht van gebruikte technieken en de problemen inherent aan een bepaalde techniek, verwijzen we naar de standaardwerken van Rossotti & Rossotti [7], Beck [15], [16] en naar het overzichtsartikel van Jameson [17].

Beck vergelijkt enkele technieken aan de hand van K_{min} en K_{max} . Dit zijn respectievelijk de kleinste en grootste stabiliteitsconstante die men kan bepalen via één of meerdere fysicochemische eigenschappen van het complex. Tabel 3.2 geeft een overzicht van de haalbare grenzen. De EMF-metingen hebben duidelijk het grootste meetbereik.

Bestudeerde eigenschap	K_{min}	Kmax
Colligatieve eigenschappen	10	10^{4}
Absorbanties	1	10^{6}
EMF	1	10^{22}

Tabel 3.2 Meetbereik van enkele analytische technieken

Wanneer men de keuze heeft tussen een techniek waarbij de experimentele resultaten afhankelijk zijn van de concentratie van een individueel deeltje, zoals bij potentiometrie, of een techniek waarbij de output afhankelijk is van de eigenschappen van verschillende deeltjes, zoals spectrofotometrie en conductometrie, is de eerste methode te verkiezen omwille van de meer eenvoudige interpretatie van de resultaten. Hoewel spectrofotometrische experimenten moeilijk interpreteerbaar kunnen zijn omdat de absorbantie van de oplossing afhankelijk is van de concentratie en de molaire extinctiecoëfficiënt van elk deeltje in oplossing, is deze techniek bij de vorming van één enkel complex bij uitzonderlijk hoge pH of in agressieve solventen vaak de enige optie. In de literatuur zijn meer dan 90% van de constanten elektrochemisch of spectrofotometrisch bepaald. Volgens Rossotti [18] kunnen stabiliteitsconstanten potentiometrisch accurater bepaald worden dan bij gebruik van andere technieken. Bovendien zijn potentiometrische metingen relatief eenvoudig en goedkoop uit te voeren.

3.1.4 Ontwikkelingen in de potentiometrie

Sinds Bjerrum [19] en Leden [20] in 1941 als eersten de wereld van de potentiometrie hebben verkend bij de bepaling van stabiliteitsconstanten van monodentaat liganden, heeft dit type van onderzoek een geweldige vaart genomen. Bovendien heeft het fundamenteel onderzoek van de Scandinavische school met als onderzoekers Sillén en Biedermann en de Zwitserse school met Schwarzenbach en Anderegg een belangrijke bijdrage geleverd [21]. De Scandinavische school legde zich toe op de studie van anorganische liganden, terwijl de Zwitsers zich meer specialiseerden in het gedrag van organische liganden. Het verschil tussen beide scholen is de hoeveelheid indifferent elektrolyt dat aan de oplossingen toegevoegd wordt (zie verder). De Zwitserse school verweet de Scandinavische school van bijna smelten te gebruiken (3 M $NaClO_4$), terwijl omgekeerd de Scandinavische school de Zwitsers verweet dat de activiteitscoëfficiënten niet constant blijven tijdens titraties in 0.1 M $NaNO_3$.

Historisch gezien is het Sillén die, in samenwerking met Ingri en Dyrssen, het eerste computerprogramma schreef dat stabiliteitsconstanten van complexen berekent aan de hand van evenwichtsmetingen: LETAGROP [22], [23]. De Italiaanse school onder leiding van Paoletti, Barbucci en Ciampolinni, leverde op thermodynamisch vlak baanbrekend werk [24]. De successvolle analyseprogramma's LEAST en MINIQUAD van Sabatini en Vacca [25] zijn eveneens een Italiaanse verdienste, alhoewel de auteurs op het formalisme van LETAGROP en SCOGS [26], [27] konden voortbouwen.

In de jaren '50 maakte de ontwikkeling van exact algebraïsche vergelijkingen en alternatieve hulpfuncties voor de berekening van stabiliteitsconstanten, de benaderingen van de Bjerrum-methode overbodig. Deze wiskundige ontwikkeling en de relatief goedkope instrumentele vereisten voor de uitvoering van potentiometrische experimenten veroorzaakten een explosieve groei in dit onderzoeksdomein [28]. In de jaren '60 en '70 werden bovendien nieuwe ion-selectieve elektroden ontwikkeld, zodat nieuwe evenwichten bestudeerd konden worden [29].

3.1 Theoretische beschouwingen: inleiding

Mede dankzij de snelle verspreiding van publicaties wordt de bepaling van stabiliteitsconstanten heden ten dage als routine aanzien, waardoor het prestigieus karakter van dit onderzoeksdomein vermindert. Verder wordt het negatief imago aangedikt zowel door het verschijnen van een aantal zwakke publicaties afkomstig van ontwikkelingslanden en ex-oostbloklanden, als door het meten van gemengde stabiliteitsconstanten in gemengde oplosmiddelen [28]. Rossotti publiceerde nochtans een aantal richtlijnen waaraan een onderzoeker zich stipt moet houden wil hij accurate constanten bepalen [6]. Martell berekende dat, gegeven het groot aantal beschikbare liganden, het meten van gemengde stabiliteitsconstanten oneindig lang kan doorgaan zonder enige vooruitgang te boeken in de complexchemie [28].

Er was dus duidelijk behoefte aan een internationale overeenkomst in verband met de experimentele vereisten, nodig voor de bepaling van accurate constanten. Bovendien drong een kritische compilatie van gekende vormingsconstanten zich op. Halfweg de jaren '70 kwam het 'Nickel-Glycine-project' tot stand [30]. De vormingsconstanten van de complexen van Ni(II) met het glycinaation werden potentiometrisch bepaald in zeven verschillende universitaire laboratoria. De bedoeling was de betrouwbaarheid van de constanten te evalueren aan de hand van de zelf vastgelegde experimentele omstandigheden en rekentechnieken. De algemene reactievoorwaarden zoals temperatuur, ionensterkte, de aard van het indifferent elektrolyt enzovoort, werden onderling afgesproken, zodat het onderzoek zich kon toespitsen op die parameters die de bepaling van vormingsconstanten verder konden beïnvloeden. Een IUPAC - commissie kon aan de hand van de bekomen resultaten een lijst opstellen van ideale experimentele voorwaarden en rekenprocedures [31], [32]. Op basis van de experimentele gegevens kon men eveneens realistische schattingen maken van de fouten op de constanten [33], [34], [35]. Vermeldenswaard hierbij is dat de IUPAC - commissie voorstelt de compilatietabellen van stabiliteitsconstanten in te delen in een viertal categorieën [36], [37]. Zo worden de constanten ingedeeld naargelang de overeenkomst tussen willekeurige publicaties over eenzelfde constante en de indicaties, aanwezig in de respectievelijke publicaties, voor het volgen van geschikte experimentele en rekenkundige procedures. Wanneer alle voorzorgsmaatregelen getroffen zijn om betrouwbare constanten te bepalen en het verschil tussen de experimenteel bepaalde constanten en deze uit de literatuur kleiner is dan 0.05 log-eenheden [36], vallen de berekende constanten in de eerste categorie 'R' ('recommended'). De andere categorieën zijn 'T' ('tentative'), 'D' ('doubtful') en 'R_i' ('rejected'). Een dergelijk onderscheid is enkel mogelijk indien er voldoende gegevens gepubliceerd zijn.

De jaren '80 en '90 worden voornamelijk gekenmerkt door de ontwikkeling van computerinfrastructuur en de mogelijkheid tot automatische sturing van titraties [38], [39], zodat de meeste onderzoekscentra de mogelijkheid hebben om grote thermodynamische databanken en simulatieprogramma's op te bouwen. Daardoor kunnen — 50 jaar na de eerste berekening van stabiliteitsconstanten — belangrijke geologische, ecologische en biologische problemen thermodynamisch behandeld of benaderd worden [40], [41], [42]. Bovendien is er ook een evolutie waarneembaar in de optimalisatieprogramma's voor de analyse van potentiometrische gegevens. Naast de gewogen of niet-gewogen kleinste kwadraten methode [25], [43], gaat er momenteel meer belangstelling uit naar een statistisch verantwoorde foutenanalyse en zoekt men naar meer objectieve criteria voor modelselectie [2], [44] and [45]. In hoofdstuk 6 komen we hierop uitgebreid terug.

Hierna volgt een bondig overzicht van de belangrijkste naslagwerken die handelen over de bepaling van stabiliteitsconstanten en van de belangrijkste overzichtstabellen van stabiliteitsconstanten, omdat ze van groot belang zijn bij het plannen van potentiometrische experimenten.

- L.G. Sillén, A.E. Martell, Stability Constants of Metal Ion Complexes, Special Publication, 17, The Chemical Society, London, (1964)
- D.D. Perrin, Dissociation Constants of Organic Bases in Aqueous Solution, IU-PAC Chemical Data Series, 12, Pergamon Press, Oxford, (1965)
- L.G. Sillén, A.E. Martell, *Stability Constants of Metal Ion Complexes*, Special Publication, **25**, The Chemical Society, London, (1971)
- D.D. Perrin, Dissociation Constants of Inorganic Acids and Bases in Aqueous Solution, IUPAC Chemical Data Series, 19, Pergamon Press, Oxford, (1972)
- E.P. Sergeant, B. Dempsey, Ionisation Constants of Organic Acids in Aqueous Solution, IUPAC Chemical Data Series, 23, Pergamon Press, Oxford, (1979)
- D.D. Perrin, Stability Constants of Metal Ion Complexes. Part B. Organic Ligands, IUPAC Chemical Data Series, 22, Pergamon Press, Oxford, (1979)
- E. Högfeldt, Stability Constants of Metal Ion Complexes. Part A. Inorganic Ligands, IUPAC Chemical Data Series, 21, Pergamon Press, Oxford, (1982)

- A.E. Martell, R.M. Smith, Critical Stability Constants, Volume 1: Amino Acids (1974), Volume 2: Amines (1975), Volume 3: Other Ligands (1977), Volume 4: Inorganic Complexes (1976), Volume 5: First Supplement (1982), Volume 6: Second Supplement (1989), Plenum Press, New York
- NIST Standard Reference Database 46 Version 2.0, NIST Critically Selected Stability Constants of Metal Complexes, (1995).

Een vollediger overzicht van de compilatiewerken over stabiliteitsconstanten, uitgegeven door IUPAC en andere organisaties, vindt men in het werk van Duffield et al. [14]. De experimentele bepaling en de berekening van stabiliteitsconstanten worden uitvoerig besproken in de volgende werken:

- F.J.C. Rossotti, H. Rossotti, The Determination of Stability Constants, McGraw-Hill, New York, (1961)
- H.L. Schläfer, Komplexbildung in Lösung, Springer-Verlag, Berlin, (1961)
- H.B. Jonassen, A. Weissberger, Technique of Inorganic Chemistry, Volume 1: Hoofdstuk 1: Fronœus: Determination of Formation Constants of Complexes, John Wiley & Sons, New York, (1963)
- P.W. Linder, R.G. Torrington, D.R. Williams, Analysis using Glass Electrodes, Open University Press, Milton Keynes, (1984)
- D.J. Legett, Computational Methods for the Determination of Formation Constants, Plenum Press, New York, (1985)
- A.E. Martell, R.J. Motekaitis, Determination and Use of Stability Constants, VCH Publishers, Weinheim, (1988)
- M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria: A Guide to Methods in Potentiometry, Extraction and Spectrometry, Ellis Horwood Ltd., Chichester, (1988)
- M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria, Ellis Horwood Ltd., Chichester, (1994)
• M.T. Beck, I. Nagypál, *Chemistry of Complex Equilibria*, Ellis Horwood Ltd., Chichester, (1990)

3.2 De potentiometrische techniek

3.2.1 Inleiding

Rond de experimentele vereisten voor het bepalen van stabiliteitsconstanten zijn heel wat richtlijnen gepubliceerd [6], [17], [31], [32]. Volgens IUPAC dienen in elk onderzoek de volgende punten expliciet vermeld te worden:

- een nauwkeurige beschrijving van de ijking van zowel de elektroden als de buretten;
- een accurate beschrijving van de experimentele opstelling en de gebruikte procedures;
- een goede temperatuurcontrole;
- een verantwoorde keuze van het ionair midden;
- de vermelding van de gebruikte pH-schaal;
- de vermelding van de zuiverheid van de reagentia.

Daar er gewerkt wordt in een overmaat indifferent elektrolyt ($1.0 \text{ M} \text{ KNO}_3$) kunnen de indicatorelektroden geijkt worden als concentratiesensoren, zodat deze gegevens rechtstreeks in de massabalansen kunnen ingevoerd worden om de stabiliteitsconstanten te berekenen.

In deze studie worden de pK_W , de zuurconstanten van ethyleendiamine en 1,3-propyleendiamine en hun vormingsconstanten met Ag(I) als referentiesysteem gebruikt in 1 M KNO_3 [46].

3.2.2 Potentiaalmetingen

3.2.2.1 Evenwichtscriteria

De respons van een elektrodekoppel op een systeem in evenwicht volgt ongeveer een asymptotisch verloop in functie van de tijd. Volgens Chang [47] kan het bereiken van

3.2 De potentiometrische techniek

het evenwicht beschreven worden via een eerste orde relatie. Algemeen wordt echter aangenomen dat de respons van een elektrodekoppel veel complexer is [48], [49], [50]. Door het asymptotisch verloop van de respons is er nood aan een arbitrair evenwichtscriterium. Op deze wijze wordt de evenwichtspotentiaal in elk titratiepunt op dezelfde manier bepaald. Dit criterium moet een compromis vormen tussen de tijd die nodig is om evenwicht in de oplossing in te stellen nadat er een verandering in de massabalans is opgetreden en het vermijden van tijdsafhankelijke fenomenen zoals verdampingseffecten of veranderingen in de asymmetriepotentiaal van de elektroden. Voor de glas- en zilverelektrode wordt volgend criterium vastgelegd: na toevoegen van een welbepaalde hoeveelheid titratievloeistof wachten we twintig seconden vooraleer er gemeten wordt. Met een tussentijd van telkens twee seconden worden twintig metingen van de potentiaal E_i uitgevoerd. Van deze metingen wordt het gemiddelde \overline{E}_i berekend. Na twintig seconden wachttijd wordt deze procedure herhaald, een nieuw gemiddelde \overline{E}_{i+1} wordt berekend en vergeleken met de waarde voor \overline{E}_i . Deze cyclus wordt herhaald tot:

$$|\overline{E}_{i+1} - \overline{E}_i| < 0.05 \text{ mV}$$

Wanneer het proces langer dan vijf minuten duurt en er nog steeds geen evenwicht is ingetreden, wordt het titratiepunt niet in rekening gebracht en wordt een nieuwe hoeveelheid oplossing toegevoegd.

3.2.2.2 De geautomatiseerde meetopstelling en datatransmissie

De geautomatiseerde opstelling (figuur 3.1) bestaat uit een dubbelwandige recipiënt in een bij (25.0 ± 0.2) °C gethermostatiseerde kast. De recipiënt wordt afzonderlijk gethermostatiseerd op (25.00 ± 0.02) °C door middel van een thermostaat van het type HAAKE F3. De temperatuur in de kast en in het thermostaatbad wordt regelmatig gecontroleerd met een digitale thermometer Yellow Springs Instrument Model 49 TA, voorzien van een YSI 400 probe. Dit toestel heeft in het temperatuurinterval 20-43 °C een accuraatheid van 0.01 °C. Zowel de oplossingen als de elektroden worden in de gethermostatiseerde kast bewaard.

De titratie gebeurt met behulp van computergestuurde buretten die voorzien zijn van een diffusieloze buretpunt om diffusie van de oplossing naar de titratievloeistof en/of omgekeerd te vermijden. Het deksel van de recipiënt wordt aangepast zodat het lekvrij op de beker kan geplaatst worden. In het deksel zijn boringen met een diameter B14 aangebracht voor de antidiffusieburetpunten, de elektroden en de gastoevoer. Tijdens

de titraties zijn de boringen lekvrij afgesloten. Voor en na elke titratie wordt de indicatorelektrode extern geijkt in een tweede dubbelwandige beker die zich eveneens in de kast bevindt. Er wordt voldoende geroerd met een in teflon gehulde metaalstaaf, aangedreven door een magnetische roerder. Deze houdt de oplossing homogeen en vermindert daardoor de responstijd. Tussen de titratiebeker en de roerplaat zit een polystyreen plaatje om mogelijke warmtetransfer te vermijden.

Het gebruik van twee potentiometers van het type Radiometer pHM 84 (resolutie 0.1 mV) en Knick pH Meter 764 multicalimatic, met meetnauwkeurigheid van 0.2 mV [51], [52], laat toe respectievelijk de vrije protonenconcentratie en de vrije zilverconcentratie te meten. De bruikbaarheid van de potentiometers wordt getest door met

een spanningsbron (Knick) een geijkte spanning van 500.00 mV aan te leggen over een

weerstand van 1 M Ω . Bij geen van beide meters is een spanningsverlies merkbaar. Voor het meten van de H^+ -concentratie werd gekozen voor een INGOLD glaselektrode (type pH 0-14 HA 265-S7/120) (zie verder), terwijl de Ag^+ -concentratie bepaald wordt met een zilverelektrode (ORION model 94-16) (zie verder). Als referentieelektrode wordt geopteerd voor een Argenthal TM elektrode (INGOLD type 363-S7) (zie verder). Een zoutbrug (Schott Geräte B512) minimaliseert de diffusiepotentiaal die ontstaat door het mobiliteitsverschil van de verschillende ionen (zie verder).

Tijdens elke meting borrelt er boven en in de oplossing N_2 -gas (L'Air Liquide A₅₂ of A₂₈). Dit gas wordt doorheen natronkalk, 2 M NaOH, 6 M H_2SO_4 en 1 M KNO_3 gestuurd om CO_2 en andere zure of basische resten te verwijderen en bovendien dezelfde dampspanning te verkrijgen als de meetoplossing. Het debiet van het gas wordt met fijnschroefregelaars gecontroleerd. Voor het opstarten van de titratie wordt gedurende een twintigtal minuten CO_2 -vrij N_2 -gas boven en door de meetoplossing gestuurd.

Door het werken met oplossingen met een hoge en constante nitraatconcentratie van 1.0 M blijven de activiteitscoëfficiënten van de deeltjes in oplossing constant (zie verder). Na ijking van de elektrodesystemen kunnen hierdoor rechtstreeks concentraties worden gemeten. Deze concentraties worden in de massabalansen verwerkt. De stabiliteitsconstanten kunnen zo op een relatief eenvoudige manier berekend worden. De opstelling wordt gestuurd met behulp van een Apple II GS microcomputer. Het sturingsprogramma is geschreven in UCSD-Pascal met ingebouwde Assembler subroutines die via I/O 32 kaarten de automatische buretten sturen en de potentiometers uitlezen. De gegevens worden op diskette bewaard en geanalyseerd met behulp van het optimalisatieprogramma SUPERQUAD op een DEC/Alpha 3000/800X/AXP com-

puter (hoofdstuk 5).



Figuur 3.1: De geautomatiseerde potentiometrische opstelling

_

3.2.3 Het indifferent elektrolyt

De hoeveelheid indifferent elektrolyt wordt meestal zodanig aangepast dat men tijdens de potentiometrische titraties een constante ionensterkte I behoudt, gegeven door:

$$I = 1/2\sum_{i} c_i Z_i^2$$
(3.14)

waarbij c_i en Z_i respectievelijk de molaire concentratie en de lading van het ion i voorstellen.

Sommige onderzoekers verkiezen te werken bij een constante totale concentratie aan ionenequivalenten:

$$I = 1/2\sum_{i} c_i |Z_i| \tag{3.15}$$

Op dit ogenblik bestaat er geen enkele theoretische basis om aan één van beide methoden de voorkeur te geven. Zoals gezegd zal in dit onderzoek gewerkt worden bij een constante nitraatconcentratie van 1.0 M. Bij hogere concentraties aan indifferent elektrolyt stellen zich immers problemen van zuiverheid en/of oplosbaarheid. Bovendien zullen de zwakke interacties tussen het elektrolyt, het metaalion en ligand nog meer uitgesproken zijn. Indien gewerkt wordt binnen het pH-gebied 2-12 zullen de H^+ respectievelijk OH^- ionen maximaal 1% van de ionensterkte uitmaken in het grensgebied. Hierdoor zijn de invloeden op de diffusiepotentiaal en op de glaselektrode minimaal. Omdat millimolaire oplossingen beschouwd worden als de limietconcentraties voor betrouwbare metingen, vereisen deze veranderingen een minimale ionensterkte van 1 M [13].

3.2.4 De elektroden

3.2.4.1 De Ag/AgCl referentieelektrode

De keuze van de referentieelektrode bepaalt in hoge mate de werking van de indicatorelektrode. Tijdens de meting zou de potentiaal van de referentieelektrode zo constant mogelijk moeten blijven. Doorgang van kleine hoeveelheden elektrische stroom $(10^{-9} \text{ Å of minder})$, nodig voor de aandrijving van de potentiometers, alsmede veranderingen in samenstelling van het systeem zouden zeker geen invloed mogen hebben op de potentiaal van de referentieelektrode. Alleen onder deze ideale omstandigheden kan de indicatorelektrode ondubbelzinnige informatie geven over de ionactiviteit in de cel. De ArgenthalTM elektrode (INGOLD type 363-S7) voorgesteld in figuur 3.2 voldoet aan de vooropgestelde eisen. Deze elektrode bestaat uit een zilverdraad bedekt met een zilverchloride laagje. Een verzadigde KCl-oplossing rond de zilverdraad vermijdt verdampingseffecten tijdens de metingen. Vaste AgCl-korreltjes verzekeren verzadiging en voorkomen het oplossen van AgCl aan het oppervlak van de elektrode. Hierdoor is de ArgenthalTM elektrode beter beschermd tegen hysterese effecten [53] als deze waaraan de SCE (standaard kalomelelektrode) gevoelig is.



Figuur 3.2: De ArgenthalTM elektrode

De bekomen halfcel kan als volgt voorgesteld worden:

$$AgCl(v)/Ag(v)/Cl^{-}(xM)$$

De evenwichten die de potentiaal van deze elektrode bepalen zijn:

$$AgCl_v = Ag_{opl}^+ + Cl_{opl}^- \tag{3.16}$$

met

$$K_S = a_{Ag^+} \cdot a_{Cl^-}$$
 (3.17)

$$Ag_{opl}^{+} + e^{-} \rightleftharpoons Ag_{v} \tag{3.18}$$

Bij een grote overmaat aan Cl^{-} -ionen kan $AgCl_{2}^{-}$ gevormd worden:

$$Ag_{opl}^{+} + 2Cl_{opl}^{-} \rightleftharpoons AgCl_{2(opl)}^{-}$$

$$(3.19)$$

waarbij:

$$\beta_{AgCl_2^-} = \frac{a_{AgCl_2^-}}{a_{Ag+.a_{Cl-}^-}} \tag{3.20}$$

De totale elektrodereactie is:

$$2AgCl_v + e^- \rightleftharpoons Ag_v + AgCl_{2(opl)}^- \tag{3.21}$$

Volgens de ideale Nernstvergelijking is de potentiaal gelijk aan:

$$E_{AgCl/Ag/AgCl_{2}^{-}} = E^{0} - \frac{RT}{F} \ln(K_{S}\beta_{AgCl_{2}^{-}}) - \frac{RT}{F} \ln(a_{Cl_{opl}^{-}})$$
(3.22)

Daar $a_{Cl_{opt}^-}$ een constante is zal de potentiaal van de cel bij constante temperatuur constant blijven. In een verzadigde KCl-oplossing en bij 25.0 °C bedraagt de potentiaal van een dergelijke Ag/AgCl-elektrode ten opzichte van standaard waterstofelektrode 198.9 mV [54].

Doch de ArgenthalTM elektrode heeft ook nadelen. Ten eerste veroorzaakt de doorgang van ionen door het keramisch membraan van de elektrode Cl^- -lekken, waardoor de studie van de Ag(I)-complexering ernstig verstoord wordt door de vorming van AgClneerslag en complexen. Het debiet van het verlies aan elektrolytoplossing wordt geschat op ca. 1 ml per 24 uur bij 25 °C voor een elektrode met keramische fit [55]. Het gebruik van een zoutbrug elimineert dit nadeel (zie volgende paragraaf). Ten tweede kan er een diffusiepotentiaal ontstaan aan het grensoppervlak tussen de meetoplossing en de ArgenthalTM elektrode. Dit is het gevolg van het verschil in concentratie en aard van de ionen in beide oplossingen. Deze diffusie is rechtstreeks afhankelijk van de concentratiegradiënt en de migratiesnelheid van de ionen en kan enkel geminimaliseerd worden maar niet volledig worden verholpen.

3.2.4.2 De zoutbrug

Potentiaalmetingen kunnen aan accuraatheid inboeten als geen rekening gehouden wordt met diffusiepotentiaalfenomenen. In het grensgebied tussen de meetoplossing en de ArgenthalTM elektrode zijn de kationen en anionen ongelijkmatig verdeeld. De verschillend geladen deeltjes diffunderen doorheen de oplossing met een verschillende snelheid wat resulteert in een diffusiepotentiaal. In een ideale zoutbrug zullen de

aanwezige ionen de mobiliteitsverschillen tussen de ionen in de oplossingen opheffen. Daarom moet de zoutbrug gevuld zijn met een elektrolyt waarvan de tegengesteld geladen ionen quasi dezelfde snelheid hebben. De effectiviteit van de zoutbrug vermeerdert tevens als de concentratie van het zout in de brug toeneemt.

De potentiaal aan het grensoppervlak tussen oplossing A aan de negatieve pool en oplossing B aan de positieve pool wordt bij verwaarlozing van het solventtransport gegeven door [7]:

$$E_j = E_B - E_A = -\frac{RT}{F} \int_A^B \sum \frac{t_i}{Z_i} dlna_i$$
(3.23)

waarbij t_i , Z_i en a_i respectievelijk het transportgetal, de lading en de activiteit van het ion *i* voorstellen. De diffusiepotentiaal wordt groter naarmate beide oplossingen ionen bevatten die grotere verschillen vertonen in concentratie, mobiliteit en lading. Bij een constante diffusiepotentiaal kan voor deze term gecorrigeerd worden door het elektrodekoppel vóór en na elke titratie te ijken. De diffusiepotentiaal blijkt echter afhankelijk te zijn van de constructie van de elektroden en experimenteel weinig reproduceerbaar te zijn [54]. Wanneer gewerkt wordt bij constante ionensterkte blijkt de diffusiepotentiaal voornamelijk af te hangen van de proton- en de hydroxideconcentratie. Dit kan verklaard worden door de hoge mobiliteit van H^+ en OH^- (tabel 3.3).

Tabel 3.3 Mobiliteit μ (10⁻⁴ cm²s⁻¹V⁻¹) van enkele ionen bij 25 °C en oneindige verdunning [56]

Kation	μ	Anion	μ
H+	36.23	OH-	20.64
K ⁺	7.62	Cl-	7.91
Na ⁺	5.19	NO_3^-	7.40
Cu^{2+}	5.46		
Ag ⁺	6.42		

Er wordt in eerste instantie aangenomen dat E_j kan beschreven worden als een lineaire functie van H^+ en OH^- [12], [57]:

$$E_j = j_H[H^+] + j_{OH}[OH^-]$$
(3.24)

De bepaling van de diffusiepotentiaalconstanten j_H en j_{OH} wordt in paragraaf 3.2.5 besproken. Bij gebruik van een zoutbrug worden in feite twee diffusiepotentialen opgebouwd: een eerste tussen de referentieelektrode en de zoutbrugoplossing, een tweede tussen de zoutbrugoplossing en de meetoplossing. De vuloplossing van de zoutbrug moet zo gekozen worden dat de som van beide potentialen minimaal is. De ideale zoutbrugoplossing voldoet aan volgende eisen [58]:

- de oplossing bestaat uit een sterk binair elektrolyt, opgelost in hetzelfde solvent als de meetoplossing;
- de concentratie van het elektrolyt in de zoutbrugoplossing moet hoger zijn dan de elektrolytconcentratie in de meetoplossing;
- het binair elektrolyt moet equitransferent zijn. Dit betekent dat het kationtransportgetal gelijk is aan het aniontransportgetal, of ook dat beide ionen dezelfde mobiliteit hebben.

In dit onderzoek gebruiken we een Schott Geräte B512 zoutbrug met keramische junctie. De zoutbrug wordt gevuld met een waterige 1.5 M KNO_3 -oplossing waarbij de mobiliteit van respectievelijk de K^+ -ionen en de NO_3^- -ionen de diffusiepotentiaal verwaarloosbaar klein maakt. Uit een onderzoek van Luts [59], waarbij zoutbrugoplossingen variërend van 1.0 tot 3.0 M KNO_3 getest werden, blijkt immers dat de beste resultaten verkregen worden met een vuloplossing van 1.5 M KNO_3 .

3.2.4.3 De indicatorelektroden

De glaselektrode

De glaselektrode, hier gebruikt als pH-elektrode, bestaat uit een dun H^+ -ion gevoelig glasmembraan met een interne referentie Ag/AgCl- elektrode met een constante interne H^+ -concentratie.

Schematisch geeft dit:

referentieelektrode//zoutbrugopl.// $H^+(a_j)$ /glasmembraan/ $HCl(a_i = \text{cte})/Ag_{(v)}/AgCl_{(v)}$ OPLOSSING GLASELEKTRODE

+

Een voordeel van de glaselektrode is dat er een verwaarloosbaar elektronentransport is, zodat de werking ervan niet beïnvloed wordt door oxidantia of reductantia in de

oplossing. Een groot nadeel is echter dat de elektrode een zeer grote weerstand heeft (van 1 tot 500 M Ω) [53]. Om het meetsignaal af te schermen van elektrische storingen wordt het spanningssignaal via een coaxiale kabel met aangepaste schroefdop naar de potentiometer gebracht. De hoge weerstand vereist een potentiometer met hoge ingangsimpedantie. Zonder compensatiecircuit zou het potentiaalverschil aanleiding geven tot een elektrische stroom doorheen de oplossing. De gemeten concentraties zouden hierdoor veranderen, hetgeen tot niet verwaarloosbare fouten zou leiden.

Het glasmembraan is opgebouwd uit silicaatgroepen waaraan een metaalion gekoppeld is. Elk zuurstofatoom wordt bovendien gedeeld door twee silicaatgroepen. Het membraan bestaat in feite uit drie delen: een interne en een externe actieve gellaag en een droge middenlaag. Ondergedompeld in een waterige oplossing zullen aan het contactoppervlak de kationen van de buitenste gellaag worden vervangen door protonen zodat een grensfasepotentiaal ontstaat waarbij een gehydrateerd silicatenlaagje van ongeveer 50 nm dikte wordt gevormd [60]. De dikte van het laagje wordt bepaald door de samenstelling van het glas. De glaselektrode reageert volgens de Nernstvergelijking wat betreft de activiteit van de waterstofionen:

$$E = E^{0} + \frac{RT}{F} \ln \frac{a_{H^{+}_{opl}}}{a_{H^{+}_{oel}}}$$
(3.25)

Bij extreme pH-waarden kunnen er afwijkingen optreden. Als de concentratie aan alkali-ionen groot is en deze van protonen klein dan wordt het evenwicht aan het elektrodeoppervlak verstoord. Het evenwicht:

$$-SiO^{-}M^{+} + H^{+} \rightleftharpoons -SiO^{-}H^{+} + M^{+}$$

$$(3.26)$$

verschuift naar links. De elektrode gedraagt zich alsof er meer H^+ -ionen in oplossing aanwezig zijn dan in de realiteit het geval is. Dit resulteert in een lagere pH-waarde dan verwacht. Een uitbreiding van de Nernstvergelijking houdt hiermee rekening [61]:

$$E = E^{0} + \frac{RT}{F} \ln(a_{H^{+}} + k_{H,M^{n+}} (a_{M^{n+}})^{1/n})$$
(3.27)

In deze aangepaste Nernstvergelijking bepaalt de selectiviteitscoëfficiënt, $k_{H,M^{n+}}$ de grootte van de 'alkalifout'. Deze selectiviteitscoëfficiënt is beduidend groter bij alkalimetaalionen in het bijzonder bij Na^+ -ionen waar de alkalifout dan ook het meeste zal doorwegen.

In het zure pH-gebied heeft de glaselektrode last van een 'zuurfout' [53]. De geconjugeerde basen van het zuur kunnen terechtkomen in de gellaag en/of de activiteit van

het water in de gellaag kan veranderen [62]. Het langzaam oplossen van de gehydrateerde gellaag aan de buitenkant van de elektrode kan het instellen van het evenwicht beïnvloeden. De meeste onderzoekers raden dan ook aan om de metingen in waterig milieu enkel uit te voeren in het pH-gebied 2 tot 12.

De glaselektrode die gebruikt wordt in dit onderzoek (INGOLD type pH 0-14 HA 265-S7/120) (figuur 3.3) is ontworpen om o.a. metingen uit te voeren in een breed pH-interval, 0 < pH < 14, en zou een zeer kleine alkalifout vertonen [60].





Figuur 3.4: Een zilverelektrode

Tenslotte kan de glaselektrode ook onderhevig zijn aan een asymmetriepotentiaal. Bij de vorming van het glasmembraan kwam de buitenkant in contact met een vlam. Het is dus mogelijk dat het extern en intern contactoppervlak een andere structuur kregen. Hierdoor kan het aantal sites voor de uitwisseling van H^+ -ionen verschillend zijn in de interne en externe gellaag. Tevens is de externe gellaag onderhevig aan mechanische en chemische aanvallen tijdens het gebruik. Een verandering in hydratatie is hierdoor mogelijk zodat de asymmetriepotentiaal niet constant blijft. Als de elektrode niet gebruikt wordt is het daarom nodig ze te bewaren in water om dehydratatie te vermijden en alzo de asymmetriepotentiaal zo klein mogelijk en constant mogelijk te houden. De grootte van de asymmetriepotentiaal is moeilijk te voorspellen zodat de elektrode regelmatig moet geijkt worden.

Volgens de leverancier [63] van het gebruikte type elektrode heeft het 'HA glas' een weerstand van ongeveer 500 M Ω bij 25 °C, is de hoofdcomponent in deze glassoort Lithium en voldoet de elektrode over het volledige *pH*-gebied (2.5 - 12) bij een temperatuur van 0 °C tot 130 °C.

De zilverelektrode

De zilverelektrode (ORION model 94-16) (figuur 3.4) is opgebouwd uit een weinig oplosbaar Ag₂S ($K_S = 10^{-51}$) membraan gegoten in een inert epoxy omhulsel. Een zilverdraad, ingebed in het membraan, wordt verbonden met een voltmeter. De epoxyhars bevat fijnverdeeld zuiver zilver om het elektrisch contact te verzekeren [64]. Er ontstaat een evenwicht tussen het zilver en de Ag^+ -ionen in het membraan:

$$Ag^+(v) + e^- \rightleftharpoons Ag(v) \tag{3.28}$$

De activiteit van het zuivere zilver is 1. Dit legt de activiteit van de Ag^+ -ionen aan de interne oppervlakte van het membraan vast. Het grote voordeel van een dergelijke 'all solid state' elektrode is dat ze geen inwendige oplossing bevat die kan lekken of verdampen. Een verandering in de concentratie van de inwendige oplossing van een elektrode geeft immers aanleiding tot onbetrouwbare resultaten. Wanneer de elektrode inactief is wordt ze droog bewaard bij 25 °C.

De Ag^+ -ionen migreren in het membraan als er in een oplossing Ag^+ -ionen aanwezig zijn. Er wordt een grenspotentiaal opgebouwd die perfect voldoet aan de Nernstvergelijking :

$$E = E^{0'} + \frac{RT}{F} \ln(a_{Ag^+} + k_{Ag,M^{n+}} (a_{M^{n+}})^{1/n})$$
(3.29)

Het potentiaalverschil kan worden gemeten t.o.v. de ArgenthalTM elektrode. Uit de gemeten potentiaal kan de vrije Ag^+ -ionenconcentratie worden berekend op voorwaarde dat er geen andere metaalionen zoals Cu^{2+} , Hg^{2+} , enz. aanwezig zijn. Daar de elektrode enkel reageert op vrije Ag^+ -ionen zal de gemeten potentiaal dalen als er complexvorming optreedt met de Ag^+ -ionen. Dit maakt het dan ook mogelijk het complexeringsgedrag van de Ag^+ -ionen met een grote nauwkeurigheid te volgen. Tabel 3.4 vat de belangrijkste karakteristieken van de zilverelektrode samen. De gegevens zijn afkomstig van de fabrikant [65].

Tabel 3.4 Karakteristieken van de Orion 94-16 zilverelektrode

Concentratiegebied:	1 tot 10^{-7} M totaal zilver
	$1 \text{ tot } 10^{-23} \text{ M}$ vrij zilver
pH-gebied:	0 tot 14
Temperatuurgebied:	0 tot 80 °C
Elektrode omhulsel:	zuur, base en solvent inert epoxy
Elektrode weerstand:	$< 1 M\Omega$

Tenslotte is er weinig geweten over het responsmechanisme van deze elektrode en de reacties die aan het oppervlak ervan plaatsvinden. Men neemt aan dat wanneer een zilverelektrode in een zilveroplossing gedompeld wordt, er een uitwisseling van ionen tussen de oplossing en de kationensites van het membraan ontstaat en dat er een grenspotentiaal opgebouwd wordt door een ruimteladingmechanisme. De Marco et al. hebben echter waargenomen dat de buitenste oppervlaktelaag van het membraan gehydrateerd is en dat gehydrateerde ionen deze laag niet kunnen binnendringen [66]. Hierdoor kan geen eenvoudig uitwisselingsmechanisme optreden aan het Ag_2S oppervlak. Waarschijnlijk spelen redoxreacties een cruciale rol in het responsmechanisme.

3.2.5 Ijking van de meetopstelling

3.2.5.1 Ijking van de pipetten en de automatisch gestuurde buretten

Rond de experimentele vereisten om stabiliteitsconstanten zo accuraat mogelijk te bepalen zijn tal van richtlijnen gepubliceerd [6], [31] en [32]. Naast een gedetailleerde bespreking van de opstelling en de gebruikte experimentele procedures, dient volgens IUPAC een nauwkeurige beschrijving van de ijking van de meetopstelling worden gegeven.

In deze sectie geven we een overzicht van de resultaten van de ijking van de automatisch gestuurde buretten en de pipetten, alsook een beschrijving van de calibratie van de glas- en zilverelektrode. Tenslotte wordt de operationele pK_W en de E^0 van de glaselektrode bepaald met behulp van het programma E0TIT [46].

Ijking van de pipetten

Bij de titraties maken we enkel gebruik van klasse A pipetten, geijkt bij 25 °C. De pipetten worden regelmatig herijkt met MilliporeTM water bij 25 °C. De standaardafwijking op de gebruikte pipetten wordt samengevat in tabel 3.5.

3.2 De potentiometrische techniek

Inhoud pipet (ml)	Exp. volume (ml)	σ_{volume} (ml)
10	9.99	0.01
20	20.01	0.03
25	25.00	0.02
50	49.97	0.02

Tabel 3.5 Ijking van de pipetten bij 25 °C

Ijking van de automatisch gestuurde buretten

De automatisch gestuurde Schott T100 buretten van 5, 10, 20 en 50 ml worden geijkt bij 25 °C om tijdens de berekeningen te beschikken over nauwkeurige gegevens en tevens rekening te houden met experimentele fouten op het toegevoegd volume sterke base. Als gevolg van de alkaliaantasting van het glazen oppervlak kan het volume van de glazen behouder per jaar tussen 0.2 en 0.6 % vergroten [67]. In eerste instantie zijn deze volumevergrotingen niet zo dramatisch omdat men meestal dezelfde buret gebruikt bij standaardisatie van de titratievloeistof. Het is echter vanzelfsprekend dat men dergelijke systematische fouten beter vermijdt, vandaar dat regelmatig herijken aangewezen is. De betrouwbaarheid van de buretten wordt gecontroleerd door de hoeveelheid uitgestoten MilliporeTM water af te wegen en het overeenkomstige volume te berekenen met behulp van de dichtheid van water bij 25 °C ($\rho_{H_{20}}^{298.15} = 0.997075$ g/cm³). De bekomen standaardafwijkingen zijn van dezelfde grootteorde als het laatste cijfer op de display van de buretopzet. De resultaten van de ijking zijn samengevat in tabel 3.6. Andere onderzoekers vinden waarden van vergelijkbare grootteorde [68], [69], [70], [71], [72], [73], [74].

Inhoud buret (ml)	Exp. volume (ml)	σ_{volume} (ml)
5	5.000	0.002
10	10.00	0.01
20	20.00	0.01
50	50.00	0.01

Tabel	3.6	Ijking	van	de	automatisch	gestuurde	buretten	bij	25	C
	~ ~ ~	- 1.00.09	0.0016	000	0000011000000000	y contant we	o wr coucre	001	20	0

3.2.5.2 Ijking van de glaselektrode en bepaling van de operationele pK_W

Tijdens de titratie van een standaard HNO_3 -oplossing en een standaard, CO_2 -vrije KOH-oplossing is de vrije concentratie aan protonen en hydroxide-ionen in elk punt van de titratiecurve gekend. Bij het volgen van deze zuur-base titraties met een

glas/ArgenthalTM-elektrodekoppel kunnen de karakteristieken van de glaselektrode $E^{0'}$, de diffusiepotentiaal E_i en het nernstiaans gedrag bepaald worden. De relatie tussen de potentiaal en deze parameters wordt beschreven door de algemene potentiaalvergelijking (3.1). De potentiaal/volume gegevens kunnen op verschillende manieren geanalyseerd worden. De meest gebruikte technieken zijn: Gran-plot, E^0 -analyse en multiparameterfit. Het programma E0TIT- ontwikkeld binnen de onderzoeksgroep Anorganische en Fysische Scheikunde van het LUC — combineert deze drie analyses [46].

Voor een gedetailleerde bespreking van de multiparameterfit techniek verwijzen we naar het werk van Garner, Luts en Brassinne [46], [59], [75].

Als belangrijkste conclusies uit deze studies leren we ten eerste dat bij 25 °C in het base gebied de term $j_{OH}[OH^{-}]$ mag verwaarloosd worden bij een pH kleiner dan 12. Anderzijds werd de diffusiepotentiaal
constante j_H bepaald in een E^0 -analyse bij 25 °C in het zure gebied in vergelijkbare omstandigheden als deze in dit onderzoek met als resultaat: (-39 ± 2) mV/M. Ten derde neemt de diffusiepotentiaal sterk af in functie van de pH en wordt verwaarloosbaar bij pH = 2.5. Rossotti en Rossotti [76] ontdekten een lineair verband tussen deze diffusiepotentiaal E_i en de proton- en de hydroxideconcentratie, dat kan beschreven worden door vergelijking 3.24.

Om de diffusiepotentiaal bij de ijking van de elektroden te minimaliseren wordt het glas/ArgenthalTM - elektrodekoppel geijkt tijdens de titratie van (0.005 M HNO_3 + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃). Aangezien bij een pH kleiner dan 12 geldt:

$$E_j = j_H[H^+] (3.30)$$

bedraagt de diffusiepotentiaal in deze oplossing maximaal 0.20 mV hetgeen vergelijkbaar is met de accuraatheid van de potentiometer, zodat we de E_j tijdens deze titraties inderdaad mogen verwaarlozen. De data van deze ijktitraties werden geanalyseerd m.b.v. het programma E0TIT via de optie 'Gran-plot' en de optie ' E^0 -analyse'. We verwijzen naar de literatuur voor een afleiding van de relevante vergelijkingen van de Gran-plot techniek [78]. Deze gevoelige techniek vermijdt een minder nauwkeurige bepaling van het eindpunt van een titratie bij gebruik van de eerste of tweede afgeleide in het buigpunt van een kleine potentiaalsprong. De oorspronkelijke logaritmische functie van E in functie van het toegevoegd volume wordt in deze methode gelineariseerd. Vóór en na het equivalentiepunt worden twee rechten bekomen die mekaar in het ideale

geval op de volumeas snijden. Vermeldenswaard hierbij zijn volgende definities:

$$\varphi_1 = (V+v)10^{EF/2.303RT} \tag{3.32}$$

en

$$\rho_2 = (V+v)10^{-EF/2.303RT} \tag{3.33}$$

met V het initieel volume sterk zuur, v het toegevoegd volume sterke base, E de gemeten potentiaal, F de Faraday constante, T de temperatuur en R de gasconstante. In figuur 3.5 is een typische Gran-plot van de ijktitraties weergegeven. De gevoeligheid van deze techniek weerspiegelt zich nogmaals in het opsporen van afwijkingen in de lineariteit van φ_1 en φ_2 in sterk zuur of basisch midden, waarschijnlijk te wijten aan een sterk veranderende diffusiepotentiaal. Deze krommingen werden ook door Rossotti et al. [77] bestudeerd. Figuur 3.6 toont de resultaten van een residuelenanalyse. De procentuele residuelen worden berekend volgens:

$$\% a f wijking = \frac{(\varphi^{exp} - \varphi^{calc})}{\varphi^{exp}} 100$$

Hieruit blijkt dat voor de vermelde titratie (figuur 3.5) de afwijking van de residuelen zeer klein is en de kans op systematische fouten eveneens klein is.



Figuur 3.5: Gran-plot van een titratie van (0.005 M HNO₃ + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃)

Via de Gran-plot methode kan tevens de operationele pK_W en de carbonaatcontaminatie — waaraan de berekende K_W gevoelig is — bepaald worden. Niettegenstaande de 0.1 M KOH-oplossingen bereid worden door ' CO_2 -free' KOH-ampullen van 0.1 M aan te lengen tot 1 liter met CO_2 - vrij water, blijkt het experimenteel bijna onmogelijk om perfect CO_2 -vrije baseoplossingen te maken. Volgens Martell et al. bevatten dergelijke ampullen tussen de 0.5 en 1.5% carbonaat [28].

Onze experimentele gegevens tonen echter aan dat wanneer de oplossingen bereid worden uitgaande van Merck Titrisol KOH-ampullen pro analyse een carbonaatgehalte van 0.2% haalbaar is. Door de opname van CO_2 tijdens de bereiding van baseoplossingen enerzijds en anderzijds een opname door lekken in het titratiesysteem, worden oplossingen ouder dan 14 dagen niet meer gebruikt. De carbonaatcontaminatie overschrijdt dan immers de 0.5%. Bij carbonaatgehaltes lager dan 0.5% is een aanpassing van de protonmassabalans overbodig [46].

In tabel 3.7 zijn de resultaten bekomen met E0TIT (opties 'Gran-plot' en ' E^0 -analyse') van de titratie van (0.005 M $HNO_3 + 0.995$ M KNO_3) met (0.1 M KOH + 1.0 M KNO_3) samengevat. Voor een bespreking van de theorie achter de E^0 -analyse die vertrekt van de algemene potentiaalbetrekking voor de glaselektrode:

$$E = E^{0'} + sl_{exp}log[H^+] + E_j$$

verwijzen we eveneens naar de literatuur [46], [75].

Belangrijk hierbij is dat indien de zuur- en baseconcentraties perfect gekend zijn, een grafiek van $(E - sl_{exp}log[H^+])$ vs. $[H^+]$ in het zure gebied een rechte oplevert met intercept $E^{0'}$ en helling j_{H} . Een analoge vergelijking bekomt men voor de basische kant via K_W (figuren 3.7 en 3.8).

De voornaamste bekommernis echter bij zulk een analyse is de controle van het nernstiaans gedrag van de glaselektrode, met name bij 25 °C moet $sl_{exp} = 59.16 \text{ mV/decade}$ zijn. Uit figuren 3.7 en 3.8 blijkt dat over het gehele pH-gebied geen junctieverschijnselen bestaan, zodat de potentiaal E uitgezet kan worden in functie van $\log[H^+]$ of $\log[OH^-]$ en nagegaan kan worden of wel degelijk aan het nernstiaans gedrag voldaan is.

Tabel 3.7 Gran-plot- en E^{0} - analyse m.b.v. E0TIT van een titratie van (0.005 M HNO₃ + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃)

parameter	Gran-plot	E^{0} -analyse	
$E^0 (mV)$	_a	420.02 ± 0.2	
$C_B \pmod{l}$	0.1001^{b}	0.1001 ± 0.0002	
pK_W	13.776^{b}	13.775 ^b	
$\% (CO_3^{2-})$	0.07^{b}	0.4^{b}	

^a wordt niet berekend

 b standaardafwijkingen worden voor deze parameters niet berekend

Vermits de analyse van de experimentele gegevens aantoont dat de gebruikte glaselektrode een nernstiaans gedrag vertoont, wordt de standaardpotentiaal ervan nagegaan vóór en na elke titratie in (0.005 M $HNO_3 + 0.995$ M KNO_3). Deze potentiaal correleert perfect met de E^0 -waarden uit de zuur-base analyse en laat toe te compenseren voor kleine verschuivingen in E^0 . Bovendien geeft deze manier van werken voldoende zekerheid over de betrouwbaarheid van de elektrodeparameters, waardoor een dagelijkse sterk zuur - sterke base titratie en een analyse van de gegevens met E0TIT overbodig wordt. De nieuw aangemaakte zuur- en/of base-oplossingen worden wel steeds op zuiverheid en concentratie nagegaan m.b.v. een Gran-plot en een E^0 -analyse.



Figuur 3.6: Systematische afwijkingen in de φ -waarden tijdens de titratie van (0.005 M HNO₃ + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃)



Figuur 3.7: E^0 -analyse in het zure gebied van een titratie van (0.005 M HNO₃ + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃)



Figuur 3.8: E⁰-analyse in het base gebied van een titratie van (0.005 M HNO₃ + 0.995 M KNO₃) met (0.1 M KOH + 1.0 M KNO₃)

3.2.5.3 Ijking van de zilverelektrode

De zilverelektrode wordt hier gecalibreerd via de 'batch'- methode. Ag^+ -oplossingen met een concentratie variërend van 5.10^{-6} M tot 10^{-2} M worden aangemaakt. De nitraatconcentratie wordt constant gehouden op 1.0 M. In figuur 3.9 wordt de gemeten potentiaal uitgezet in functie van $\log[Ag^+]$. De helling van de rechte bedraagt 59.2 mV/decade, wat beantwoordt aan de Nernstvergelijking. Rekening houdend met een experimentele fout van 0.2 mV kan men besluiten dat de zilverelektrode zich in het bestudeerde concentratiegebied nernstiaans gedraagt.

Zoals bij de glaselektrode wordt de invloed van de diffusiepotentiaal op de respons van de zilverelektrode bij hogere protonconcentraties nagegaan. Hiervoor wordt 25 ml van een oplossing van (0.001 M $AgNO_3 + 0.999$ M KNO_3) getitreerd met 125 ml van een oplossing met verhoogde protonconcentratie: (0.001 M $AgNO_3 + 0.04$ M $HNO_3 + 0.959$ M KNO_3).

Niettegenstaande deze hoge protonconcentraties op het einde van de titratie is de verandering in potentiaal slechts 0.40 mV, wat niet veel groter is dan de experimentele fout, zodat deze zilverelektrode ongevoelig is voor de pH van de oplossing.

Tot slot wordt de zilverelektrode vóór en na iedere complexeringstitratie nog eens extern geijkt in $(0.005 \text{ M} AgNO_3 + 0.995 \text{ M} KNO_3)$.



Figuur 3.9: Ijking van de ORION zilverelektrode in 1.0 M KNO₃ bij 25 °C

3.2.6 De klassieke analysemethode voor het bepalen van stabiliteitsconstanten

3.2.6.1 Binaire systemen: de bepaling van de zuurconstanten van de alifatische diaminen 1,4-diaminobutaan en 1,5-diaminopentaan en de zilverhydrolyse in 1.0 M KNO_3

Reagentia en experimentele omstandigheden

• Reagentia

Alle oplossingen worden bereid met water dat gezuiverd is met het MilliporeTM systeem. De maatkolven die gebruikt worden voor de bereiding van de oplossingen zijn van het type A en zijn geijkt bij 25 °C. Alle oplossingen worden gethermostatiseerd in een kast op (25.0 \pm 0.2) °C en worden enkel bij deze temperatuur stapsgewijs aangelengd tot aan de ijkstreep.

De ionensterkte wordt constant op 1.0 M gehouden met KNO_3 (Merck p.a.). Uit literatuurgegevens blijkt dat KNO_3 -oplossingen licht zuur zijn door de natuurlijke oplosbaarheid van atmosferisch CO₂ in waterige oplossingen, waarbij HCO_3^- gevormd wordt ($p_{CO_2} = 33 \ Pa \ [79]$). Baes schat de HCO_3^- -concentratie in 1.0 M KNO_3 -oplossingen op 4.10⁻⁵ M [80].

Stockoplossingen van KOH worden bekomen uitgaande van Titrisol ampullen 0.1 M KOH (Merck p.a.). Tijdens de bereiding van deze oplossingen worden alle voorzorgsmaatregelen genomen om CO_2 -opname te vermijden. Zo worden de baseoplossingen bereid onder N_2 atmosfeer met bigedeïoniseerd water dat gedurende twee uur gekookt heeft en snel afgekoeld wordt in ijs. De vers bereide KOH-oplossing wordt bewaard in een polyethyleen fles die afgesloten is van de lucht door een buisje gevuld met natronkalk (Merck p.a.). Enkel die KOH-oplossingen met een CO_3^{2-} -gehalte kleiner dan 0.5% worden tijdens de titraties gedurende maximaal twee weken gebruikt. Na deze periode overschrijdt de carbonaatcontaminatie de 0.5% zodat ze niet meer geschikt zijn voor gebruik.

Stockoplossingen van HNO_3 worden bereid uitgaande van Titrisol ampullen 0.1 M HNO_3 (Merck p.a.).

De diamines: 1,4-diaminobutaan (Fluka > 99%) en 1,5-diaminopentaan (Fluka > 97%) worden geleverd in de aminevorm. Ze worden tot hun overeenkomstige dinitraatzouten omgezet door een techniek die beschreven is in de literatuur [81], [82]. Een lichte overmaat van twee equivalenten HNO_3 (Merck p.a. 65%) wordt onder sterk roeren en druppelsgewijs toegevoegd aan de verdunde diamineoplossing bij 0 °C. De gevormde dinitraatzouten worden vast bij -18 °C en worden gezuiverd tijdens herhaalde omkristallisatie in absolute ethanol en droge diethylether.

Tenslotte worden de zeer hygroscopische kristallen gedroogd op P_4O_{10} in een vacuüm exsiccator. Dit levert mooie witte kristallen op in het geval van 1,4-diaminobutaan. Daarentegen worden in het geval van 1,5-diaminopentaan na behandeling van de nitraatoplossing met actieve kool (Merck p.a.), een lang droogproces en afscherming van de lucht, geel-bruine kristallen bekomen. De stechiometrie van de dinitraatzouten wordt door titratie met KOH getest, wat in beide gevallen een zuiverheid oplevert groter dan 99%.

De nitraatoplossingen worden van het licht afgeschermd met aluminiumfolie.

De $AgNO_3$ -oplossingen worden bereid uitgaande van kristallijn $AgNO_3$ (Merck p.a.) en water dat gezuiverd is met behulp van het MilliporeTM systeem.

Experimentele omstandigheden

De zuurconstanten van respectievelijk 1,4-diaminobutaan en 1,5-diaminopentaan worden bepaald met behulp van het glas/ArgentalTM elektrodekoppel in 1 M NO_3^- . Een startoplossing van (0.01 M $LH_2(NO_3)_2 + 0.98$ M KNO_3) wordt getitreerd met (0.1 M KOH + 1.0 M KNO_3) tot pH = 11.500. $LH_2(NO_3)_2$ staat voor het diammoniumnitraatzout van het overeenkomstige diamine. De karakteristieken van de glaselektrode worden gecontroleerd voor en na iedere titratie in (0.005M $HNO_3 +$ 0.995 M KNO_3). Tijdens de bepaling van de zuurconstanten bedraagt de minimale pH van de oplossing 2.50, zodat de afwijking ten gevolge van E_j verwaarloosbaar is. Elke titratie wordt minstens tweemaal uitgevoerd. Het aantal bruikbare titratiepunten per titratie varieert tussen 70 en 100.

Voor een analyse en discussie van de resultaten van deze zwak zuur-sterke base titraties ter bepaling van de zuurconstanten van de respectievelijke diamines verwijzen we naar paragraaf 5.1.1.1.

Aangezien er nogal wat discrepanties voorkomen in gepubliceerde waarden voor de vormingsconstanten van de deeltjes AgOH en $Ag(OH)_2^-$ in vergelijkbare omstandigheden (25 °C), [83], [84], [85], [86], [87], [88] bepalen we deze nogmaals. Het optimale titratiegebied voor zilverhydrolyse werd in een vorig onderzoek afgebakend [46]. Uit simulaties is gebleken dat het gebied tussen 10^{-5} en 10^{-6} M C_{Ag} het meest interes-

sante is. Vandaar dat oplossingen van $(0.002 \text{ M } HNO_3 \text{ en respectievelijk } 10^{-5}, 2.10^{-5}, 3.10^{-5}, 5.10^{-5} \text{ M } AgNO_3 + 0.998 \text{ M } KNO_3)$ getitreerd worden met $(0.1 \text{ M } KOH + 1.0 \text{ M } KNO_3)$ totdat de potentiaal van de zilverelektrode onstabiel wordt of tot neerslagvorming visueel wordt waargenomen. Daar de gepubliceerde waarden van pK_s^* van Ag_2O in 1.0 M variëren van 7.18 (1 M KNO_3) tot 7.27 (1 M $NaClO_4$) [89] worden titratiepunten waarbij het ionenproduct $[Ag^+][OH^-] > 10^{-7.3}$ (arbitrair lichtjes hogere pK_s^*) niet in rekening gebracht bij de analyse van de gegevens (paragraaf 5.1.1.2). Deze relatief hoge waarde van pK_s^* , het oplosbaarheidsproduct van onderstaande reactie, bemoeilijkt bovendien de bepaling van hogervernoemde vormingsconstanten:

$$1/2Ag_2O_{(v)} + 1/2H_2O \rightleftharpoons Ag_{(opl)}^+ + OH_{(opl)}^-$$

3.2.6.2 Ternaire systemen: de bepaling van de stabiliteits
constanten van de Ag(I)-complexen van 1,4-diaminobutaan en 1,5-diaminopentaan in 1.0 M
 KNO_3

Reagentia en experimentele omstandigheden van de verhoudingsmethode

• Reagentia

De $AgNO_3$ -oplossingen worden bereid zoals hierboven vermeld.

De overige oplossingen zoals de KOH-oplossing en de KNO_3 -oplossingen komen overeen met deze beschreven in de vorige paragrafen.

De maatkolven en pipetten die gebruikt worden om de startoplossingen te bereiden zijn van het type A en zijn geijkt bij 25 °C.

• Experimentele omstandigheden van de verhoudingsmethode

Zoals bij de zuur-base titraties worden de karakteristieken van de glaselektrode gecontroleerd voor en na iedere titratie in $(0.005 \text{ M } HNO_3 + 0.995 \text{ M } KNO_3)$.

De externe ijking van de zilverelektrode gebeurt eveneens voor en na iedere titratie. Om een zo groot mogelijk complexeringsgebied te bestrijken worden verschillende C_{Ag}/C_L verhoudingen onderzocht, waarbij C_{Ag} en C_L respectievelijk de totale concentratie van zilver en ligand voorstellen. Er wordt geopteerd om de verhoudingen: 2/1, 1/1, 1/2, 1/4, 1/8 te testen [46].

De te titreren startoplossingen worden bereid uitgaande van de volgende twee stockoplossingen:

- de ligandoplossing (lo): 0.01 M $LH_2(NO_3)_2$ + 0.98 M KNO_3

- de zilver of metaaloplossing (mo): 0.01 M $AgNO_3$ + 0.99 M KNO_3

De hoeveelheid ligand in de oplossingen met C_{Ag}/C_L verhoudingen 1/1, 1/2, 1/4, 1/8, 2/1 blijft constant en bedraagt 0.2 mmol. Het toegevoegde volume metaaloplossing varieert om tot de gewenste metaal/ligand verhouding te komen. Het totale beginvolume wordt constant gehouden door 1.0 M KNO_3 aan de startoplossing toe te voegen. De verhouding 2/1 wordt verkregen door 20 ml ligandoplossing te mengen met 40 ml metaaloplossing. In tabel 3.8 wordt een overzicht gegeven van de samenstelling van de verschillende startoplossingen.

Tabel 3.8 Samenstelling van de startoplossingen bij de studie van de Ag(I)-diamine complexen in 1.0 M KNO₃

Verhouding C_{Ag}/C_L	Sar	nens	stelling van d	le st	artoplossing(ml)
1/8	20 ml lo	+	2.5 ml mo	+	17.5 ml 1.0 M KNO ₃
1/4	20 ml lo	+	5 ml mo	+	15 ml 1.0 M KNO ₃
1/2	20 ml lo	+	10 ml mo	+	10 ml 1.0 M KNO ₃
1/1			20 ml lo +	- 20	ml mo
2/1			20 ml lo +	- 40	ml mo

De verkregen oplossingen worden minstens in tweevoud getitreerd met $(0.1 \text{ M KOH} + 1.0 \text{ M KNO}_3)$ tot beide protonen van het diammoniumzout geneutraliseerd zijn of tot precipitatie visueel of door onstabiele elektroderespons vastgesteld wordt.

Bij de berekeningen worden alle titratiepunten waarbij het ionenproduct $[Ag^+][OH^-]$ > $10^{-7.3}$ verwijderd om er zeker van te zijn dat geen vorming van Ag_2O optreedt. Voor een analyse en discussie van de resultaten van deze complexometrische titraties verwijzen we naar hoofdstuk 5.

Referenties

- [1] P. Gans, A. Sabatini, A. Vacca, Inorg. Chim. Acta, 79, 219, (1983)
- [2] P. Gans, A. Sabatini, A. Vacca, J. Chem. Soc. Dalton Trans., 1195 (1985)
- [3] P. Gans, Manual SUPERQUAD, (1989)
- [4] A. Sabatini, A. Vacca, P. Gans, Coord. Chem. Rev., 120, 389, (1992)
- [5] D.A. Skoog, D.M. West, Fundamentals of Analytical Chemistry, 6 th Ed., Saunders College Publishing, Fort Worth, (1992)
- [6] H. Rossotti, Talanta, 21, 809, (1974)
- [7] F.J.C. Rossotti, H. Rossotti, The Determination of Stability Constants, McGraw-Hill, New York, (1961)
- [8] J.N. Brønsted, J. Amer. Chem. Soc., 44, 877, (1922)
- [9] J.N. Brønsted, J. Amer. Chem. Soc., 45, 2898, (1923)
- [10] L.G. Sillén in A.E. Martell, *Coordination Chemistry*, Volume 1, ACS Monograph, Van Nostrand Reinhold Company, New York, (1971)
- [11] H. Grossman, Z. Anorg. Chem., 43, 356, (1905)
- [12] G. Biedermann, L.G. Sillén, Arkiv Kemi, 5, 425, (1953)
- [13] G. Anderegg, Pure & Appl. Chem., 54, 2693, (1982)
- [14] J.R. Duffield, F. Marsicano, Polyhedron, 10, 1105, (1991)
- [15] M.T. Beck in E. Wänninen, Essays on Analytical Chemistry, Williams Clowes & sons, London, (1977)
- [16] M.T. Beck, I. Nagypál, Chemistry of Complex Equilibria, Ellis Horwood Ltd., Chichester, (1990)
- [17] R.F. Jameson, Nato Adv. Study Inst. Ser. C, 55, 145, (1980)

- [18] H. Rossotti, The study of Ionic Equilibria. An Introduction, Longman, London, (1978)
- [19] J. Bjerrum, Metal Amine Formation in Aqueous Solution, proefschrift, Copenhagen, 1941, Herdruk : P. Haase & Sons, Copenhagen, (1957)
- [20] I. Leden, Z. Phys. Chem., A188, 160, (1941)
- [21] L.G. Sillén in W. Schneider, G. Anderegg, Experimentia Suplementum IX: Essays in Coordination Chemistry, Birkhauser Verlag, Basel, (1964)
- [22] D. Dyrssen, N. Ingri, L.G. Sillén, Acta Chem. Scand., 15, 694 (1961)
- [23] N. Ingri, L.G. Sillén, Arkiv Kemi, 23(10), 97 (1964)
- [24] P. Paoletti, L. Fabrizzi, Inorg. Chim. Acta Rev., 7, 43, (1973)
- [25] A. Sabatini, A. Vacca, Talanta, 21, 54, (1974)
- [26] I.G. Sayce, Talanta, 15, 1397, (1968)
- [27] I.G. Sayce, Talanta, 18, 653, (1971)
- [28] A.E. Martell, R.J. Motekaitis, Determination and Use of Stability Constants, VCH publishers, Weinheim, (1988)
- [29] H. Freiser, Ion-selective Electrodes in Analytical Chemistry, Volume 1 en 2, Plenum Press, New York, (1978)
- [30] E. Bottari, A. Braibanti, L. Ciavatta, A.M. Corrie, P.G. Daniele, F. Dallavalle, M. Grimaldi, A. Mastroianni, G. Mori, G. Ostacoli, P. Paoletti, E. Rizarelli, S. Sammartano, C. Severini, A. Vacca, D.R. Williams, *Anal. Chim. (Rome)*, 68, 813, (1978)
- [31] G.H. Nancollas, M.B. Tomson, Pure & Appl. Chem., 54, 2675, (1982)
- [32] A. Braibanti, G. Ostacolli, Pure & Appl. Chem., 59, 1721, (1987)
- [33] A. Braibanti, F. Dallavalle, G. Mori, B. Veroni, Talanta, 29, 725, (1982)

- [34] A. Braibanti, F. Dallavalle, G. Mori, M. Pasquali, Gazzetta Chim. Ital., 113, 407, (1983)
- [35] P.M. May, K. Murray, Talanta, 35, 933, (1988)
- [36] IUPAC, Coord. Chem. Rev., 17, 358, (1975)
- [37] I. Sóvágó, T. Kiss, A. Gerely, Pure & Appl. Chem., 65, 1029, (1993)
- [38] L. Pehrsson, F. Ingman, Talanta, 24, 79, (1977)
- [39] A. Avdeef, J.J. Bucher, Anal. Chem., 50, 2137, (1978)
- [40] J.B. Butler, Marine Chem., 38, 251, (1992)
- [41] P.M. May, K. Murray, Talanta, 38, 1409, (1991)
- [42] P.M. May, K. Murray, Talanta, 38, 1419, (1991)
- [43] G. Arena, C. Rizarelli, Talanta, 26, 1, (1979)
- [44] P.M. May, K. Murray, Talanta, 35, 825, (1988)
- [45] R. Garner, J. Yperman, J. Mullens, L.C. Van Poucke, Anal. Chim. Acta, 282, 471, (1993)
- [46] R. Garner, Een potentiometrische studie van het complexeringsgedrag van Ag(I) met alifatisch gesubstitueerde mono- en diaminen in 1 M nitraat, proefschrift, L.U.C., (1993)
- [47] O. Chang, J. Chem. Educ., 64, 91, (1987)
- [48] E. Lindner, K. Tóth, E. Pungor, Anal. Chem., 48, 1071, (1976)
- [49] E. Pungor, Y. Umezawa, Anal. Chem. 55, 1432, (1983)
- [50] E. Lindner, K. Tóth, E. Pungor, Pure & Appl. Chem., 58, 469, (1986)
- [51] Instruction manual, Radiometer PHM 84, Radiometer, Copenhagen
- [52] Mikroprozessor-pH-meter, Knick, Berlin

- [53] H.H. Bauer, G.D. Christian, J.E. O'Reilly, *Instrumental Analysis*, Allyn and Bacon, Inc., Boston, (1978)
- [54] G. Mattock, D.M. Band, Glass Electrodes for Hydrogen and other Cations: Principles and Practice, G. Eisenman, Chapter 2, Marcel Dekker Inc., New York, (1967)
- [55] Dr. W. Ingold, Electrode Catalogue, Edition ELE-6-CH, Urdorf, Switzerland, (1985)
- [56] P.W. Atkins, *Physical Chemistry*, 4 th Ed., Oxford University Press, Oxford, (1990)
- [57] H.S. Dunsmore, D. Midgley, Anal. Chim. Acta, 61, 115, (1972)
- [58] T. Mussini, J. Chem. Educ., 65, 242, (1988)
- [59] P. Luts, Potentiometrische bepaling van de stabiliteitsconstanten van de sulfatocomplexen van koper (II) en cadmium (II) in 1 M NaClO₄, proefschrift, L.U.C., (1991)
- [60] G. Johansson, B. Karlberg, A. Wikby, Talanta, 22, 953, (1975)
- [61] W.E. Morf, Talanta, 26, 719, (1979)
- [62] R.A. Durst, J. Chem. Educ., 44, 175, (1967)
- [63] Elscolab N.V., Kruibeke, Belgium
- [64] A. Evans, Potentiometry and Ion-Selective Electrodes, John Wiley & Sons, Chichester, (1987)
- [65] Instruction Manual Orion, U.S.A., (1986)
- [66] R. De Marco, R.A. Cattrall, J. Liesegang, G.L. Nyberg, I.C. Hamilton, Anal. Chem., 62, 2339, (1990)
- [67] P.A. Johansson, U. Stefansson, Anal. Chim. Acta, 223, 461, (1989)
- [68] W.R. Harris, C.J. Carrano, S.R. Cooper, S.R. Sofen, A.E. Avdeef, J.V. McArdle, K.N. Raymond, J. Am. Chem. Soc., 101(20), 6067, (1979)

- [69] E. Casassas, R. Tauler, M. Filella, Anal. Chim. Acta, 191, 399, (1986)
- [70] M. Meloun, M. Bartos, Mikrochim. Acta, 108, 227, (1992)
- [71] D.M. Wambeke, D. Van De Vondel, E. Claeys, G.G. Herman, A.M. Goeminne, J. Chem. Soc. Dalton Trans., 829, (1992)
- [72] R. Garner, J. Yperman, J. Mullens, L.C. Van Poucke, Bull. Soc. Chim. Belg., 102(1), 3, (1993)
- [73] C. De Stefano, S. Sammartano, A. Gianguzza, Talanta, 40(5), 629, (1993)
- [74] M.A. Olazabal, G. Borge, R. Castano, N. Etxebarria, J.M. Madariaga, J. Sol. Chem., 22(9), 825, (1993)
- [75] K. Brassinne, Potentiometrische studie van de oxalatocomplexen van Cu²⁺ en Y³⁺ in KNO₃-midden, proefschrift, L.U.C., (1997)
- [76] F.J.C. Rossotti, H. Rossotti, J. Phys. Chem., 68, 3773, (1964)
- [77] F.J.C. Rossotti, H. Rossotti, J. Chem. Educ., 42, 375, (1965)
- [78] G. Gran, Analyst, 77, 661, (1952)
- [79] M. Macca, J. Chem. Educ., 63, 691, (1986)
- [80] C.F. Baes, N.J. Meyer, Inorg. Chem., 1, 780, (1962)
- [81] H. Creyf, Studie van de Cd(II)-complexen van enkele mono- en diaminen, proefschrift, R.U.G., (1974)
- [82] J. Yperman, Potentiometrisch en Calorimetrisch onderzoek van het zuurbase gedrag en van de complexvorming met Ag(I) van een aantal Nmethylgesubstitueerde-diethyleentriaminen, proefschrift, U.I.A., (1981)
- [83] A.O. Gubeli, J. Ste-Marie, Can. J. Chem., 45, 827, (1967)
- [84] J.S. Redinha, M.L.P. Leitão, Rev. Port. Quim., 16, 14, (1974)
- [85] C.F. Baes, R.E. Mesmer, The Hydrolysis of Cations, John Wiley & Sons, New York, (1976)

- [86] R. Keim, Gmelins Handbuch der Anorganischen Chemie, Silber, Teil B1, 122, Verlag Chemie, Weinheim, (1971)
- [87] P.J. Antikainen, D. Dyrssen, Acta Chem. Scand., 14, 86, (1960)
- [88] G. Biedermann, L.G. Sillén, Acta Chem. Scand., 14, 717, (1960)
- [89] R. Näsänen, P. Meriläinen, Suomen Kem, 11, B197, (1960)

90

Hoofdstuk 3

Hoofdstuk 4

Kwalitatieve en kwantitatieve analyse van potentiometrische data met behulp van optimalisatieprogramma's

4.1 Inleiding

In een ternair evenwichtssysteem in oplossing, zoals gedefinieerd in hoofdstuk 3, met componenten M (metaalion), L (ligand) en H (proton) wordt de concentratie van elk gevormd complex gegeven door :

$$[M_p L_q H_r] = \beta_{pqr} [M]^p [L]^q [H]^r$$

$$\tag{4.1}$$

met [M], [L] en [H] de vrije concentratie van elke component en β_{pqr} , de globale stabiliteitsconstante. De totale concentratie, C van elke component beantwoordt aan de massabalansvergelijking:

$$C_{M} = [M] + \sum_{p,q,r} p\beta_{pqr}[M]^{p}[L]^{q}[H]^{r} \quad (= [M] + \sum_{p,q,r} p[M_{p}L_{q}H_{r}])$$

$$C_{L} = [L] + \sum_{p,q,r} q\beta_{pqr}[M]^{p}[L]^{q}[H]^{r} \quad (= [L] + \sum_{p,q,r} q[M_{p}L_{q}H_{r}])$$

$$C_{H} = [H] + \sum_{p,q,r} r\beta_{pqr}[M]^{p}[L]^{q}[H]^{r} \quad (= [H] + \sum_{p,q,r} r[M_{p}L_{q}H_{r}])$$
(4.2)

het doel van de berekeningen is in eerste instantie de bepaling van de stabiliteitsconstanten. In vele gevallen echter is dit eerder een deelaspect en is de bepaling van een correct chemisch model van het systeem cruciaal.

De experimentele data bestaan uit paren (E_i, v_i) met E_i de gemeten potentiaal en v_i het

toegevoegd volume base. De waarnemingen E_i zijn impliciet functie van de onafhankelijke variabelen, v_i . De totale concentratie van elke component X op titratiepunt i wordt gegeven door:

$$C_{X_i} = \frac{(\text{initiële hoeveelheid})_X + (\text{buretconcentratie})_X \times v}{v_0 + v}$$
(4.3)

waarbij de initiële hoeveelheid van de component in mmol gegeven wordt indien het toegevoegd volume, v en het initieel volume, v_0 in ml uitgedrukt worden. Indien de evenwichtsconstanten bekend zijn, worden [M], [L] en [H] berekend door de drie vergelijkingen voor de massabalans (4.2) op te lossen. Bovendien kan het kleinste kwadraten probleem gesteld worden: de waarnemingen bestaan uit paren van (E, v)behorende bij een set van initiële hoeveelheden van elke component, waarbij de evenwichtsconstanten de parameters zijn. De onbekenden [M], [L] en [H] in elk titratiepunt moeten uiteindelijk voldoen aan de massabalans.

Sillén en Ingri [1],[2],[3],[4] hebben in 1962 als eersten een computerprogramma gepubliceerd dat complexe evenwichten kan analyseren op basis van een kleinste kwadraten benadering: LETAGROP. Het programma minimaliseert de residuelen in de potentiaal. Alhoewel LETAGROP oorspronkelijk enkel bedoeld was als aanvulling voor grafische methoden is deze supplementaire status nu volledig achterhaald. De laatste decennia worden minimalisatieberekeningen als volwaardig en onafhankelijk instrument beschouwd om complexe evenwichten te beschrijven. Toch mag niet ontkend worden dat de selectie van het model dat het meest consistent is met de data enig risico inhoudt. Experimentele tekorten kunnen met andere woorden niet gecompenseerd worden door de berekeningen, zodat de data extreem betrouwbaar moeten zijn. Een overzicht en een uitvoerige beschrijving van de specificaties van de belangrijkste potentiometrische analyseprogramma's ontwikkeld gedurende de laatste dertig jaar, worden gegeven in de publicaties van Rossotti [5], Gaizer [6], Gans [7], Sabatini [8] en in de referentiewerken van Meloun [9], Hartley [10] en Legett [11].

4.2 Het optimalisatieprogramma SUPERQUAD voor het berekenen van stabiliteitsconstanten

4.2.1 Inleiding

Gans, Sabatini en Vacca (tabel 4.1) ontwikkelden een aantal programma's om stabiliteitsconstanten te berekenen aan de hand van potentiometrische gegevens.

SUPERQUAD is één van de meest recente in die reeks van gelijkaardige programma's. Een historisch overzicht wordt gegeven in tabel 4.1.

Tabel 4.1 Historische ontwikkeling van SUPERQUAD

Programma	Jaar publicatie	Auteurs			
LEAST GN/NR	1972	Sabatini en Vacca [12]			
$STEW\beta$	1974	Gans en Vacca [13]			
MINIQUAD	1974	Gans, Sabatini en Vacca [14]			
MINIQUAD75a/b	1976	Gans, Sabatini en Vacca [15]			
MINIQUAD82	1985	Vacca en Sabatini [16]			
MIQUV	1985	Vacca en Sabatini [16]			
SUPERQUAD	1983-1989	Gans, Sabatini en Vacca [17], [18], [19]			
HYPERQUAD 1992		Gans, Sabatini en Vacca [8], [20]			

SUPERQUAD is ontstaan als een logisch gevolg van de voorgaande programma's en bevat als dusdanig verbeterde rekentechnieken.

De kritiek op MINIQUAD (MINImiQUADrati = kleinste kwadraten) slaat op de volgende aspecten:

• in MINIQUAD worden de residuelen in de massabalans geoptimaliseerd (of beter: $C_H - [H]$, respectievelijk $C_M - [M]$ en $C_L - [L]$).

De toepassing van hierop gebaseerde statistische tests is moeilijk omdat de berekende statistische grootheden niet rechtstreeks in verband staan met experimentele fouten. Vandaar dat het logischer is om experimenteel bepaalde grootheden zoals de potentiaal te minimaliseren. Een voordeel is echter dat de massabalans een expliciete functie is, terwijl pH(pM) dit niet zijn;

 in MINIQUAD krijgen alle meetpunten een eenheidsgewicht. Dit is in tegenstelling met de experimentele vaststelling dat meetpunten in het buffergebied accurater en preciezer bepaald kunnen worden dan punten in de buurt van het equivalentiepunt.

Het is dan ook statistisch verantwoord en essentieel in een kleinste kwadraten benadering om meetpunten in het buffergebied een hoger gewicht toe te kennen dan punten in de potentiaalsprong. SUPERQUAD (SUPERcede miniQUAD) bevat een aantal uitbreidingen in vergelijking tot MINIQUAD:

- de mogelijkheid bestaat om de totale beginconcentraties en de elektrodeparameters simultaan te verfijnen. De auteurs hebben deze mogelijkheid ingebouwd omdat sommige liganden, voornamelijk deze van biologische oorsprong, soms moeilijk in uiterst zuivere toestand te bekomen zijn;
- in SUPERQUAD is een keuze in gewichten ingebouwd. Men kan kiezen tussen eenheidsgewichten, gewichten berekend op basis van de experimentele titratiecurven of gewichten bekomen op basis van de theoretische modellen. Gans en Sabatini [18] raden het gebruik van de tweede methode aan;
- nieuw is het betrekken van de experimentele fout zowel op het titervolume als op de potentialen tijdens de berekening van de stabiliteitsconstanten en hun standaardafwijkingen;
- vermeldenswaard is tenslotte de automatische, zij het vrij rudimentaire modelselectieprocedure die ingebouwd is in het programma (zie verder onder 4.3.1).

HYPERQUAD (extensie van SUPERQUAD) biedt de mogelijkheid om potentiometrische en spectrofotometrische gegevens tegelijkertijd te verwerken. HYPERQUAD gebruikt dezelfde geoptimaliseerde rekentechnieken als SUPERQUAD.

4.2.2 De modelfunctie en de kleinste kwadraten benadering

Vertrekkende van een set van waarnemingen is het vaak de bedoeling deze samen te vatten door ze te fitten aan een zogenaamde functie of "model".

Veronderstel dat er n waarnemingen zijn. Met elke waargenomen f_i^{exp} stemt een berekende waarde $f_i(v_i, p_1, \ldots, p_m)$ overeen, die functie is van de onafhankelijke variabele v in punt i en een set van m parameters p_1, \ldots, p_m (zie Appendix A.9).

Het model voorspelt de relatie tussen de gemeten afhankelijke en de onafhankelijke variabelen. Wat het model ook is, het hoofddoel blijft de bepaling van de parameterset p_1, \ldots, p_m door de modelfunctie zo goed mogelijk te fitten doorheen de experimentele data.

Uiteindelijk moet "zo goed mogelijk" resulteren in de beste "fit", gebaseerd op een degelijke wiskundige basis, een zogenaamd "goodness-of-fit" criterium.

De verschillen tussen de geobserveerde grootheden en deze voorspeld door het model – de berekende grootheden– worden gewoonlijk aangeduid als de residuelen, r_i en worden gedefinieerd als:

$$r_i = f_i^{exp} - f_i(v_i, p_1, \dots, p_m)$$
(4.4)

of1

$$\mathbf{r} = [r_1, \dots, r_n]^T \tag{4.5}$$

 f_i^{exp} staat voor de gemeten potentiaal in het *i*-de titratiepunt, f_i voor de berekende potentiaal in het *i*-de titratiepunt, v_i voor het toegevoegd volume in het *i*-de titratiepunt en de parameters p_1, \ldots, p_m voor de berekende stabiliteitsconstanten, β_1, \ldots, β_m .

We definiëren, U de gewogen sommatie van de producten der residuelen:

$$U = \sum_{h=1}^{n} \sum_{i=1}^{n} W_{hi} r_h r_i$$
(4.6)

of

$$U = \mathbf{r}^T \mathbf{W} \mathbf{r} \tag{4.7}$$

De methode van de kleinste kwadraten bestaat erin om de waarde van de *m*-vector, **p** van de parameters te berekenen die U minimaliseert. De gewichtenmatrix $\mathbf{W} = (W_{hi})$ laat toe om het relatief belang van de residu's onderling te laten verschillen, waarbij in de meeste gevallen \mathbf{W} gelijk wordt gesteld aan de inverse van de covariantie matrix van de waarnemingen. We komen hier later op terug.

Merk op dat U via f_i een functie is van de parameters p_1, \ldots, p_m . Het minimum van U wordt gevonden door de partiële afgeleide naar iedere parameter gelijk te stellen aan 0:

$$\frac{\partial U}{\partial p_j} = -\sum_{h=1}^n \sum_{i=1}^n W_{hi} \left(r_h \frac{\partial f_i}{\partial p_j} + r_i \frac{\partial f_h}{\partial p_j} \right) = 0 \tag{4.8}$$

Om de notatie te vereenvoudigen introduceren we de gradiëntvector, $\mathbf{g} = \left(\frac{\partial U}{\partial p_1} \dots \frac{\partial U}{\partial p_m}\right)$ en de Jacobiaan, J met $J_{ij} = \left(\frac{\partial f_i}{\partial p_j}\right)$ zodat:

$$\mathbf{g} = -2\mathbf{J}^T \mathbf{W} \mathbf{r} = 0 \tag{4.9}$$

¹Matrices en vectoren worden in het vet aangeduid, T staat voor de getransponeerde vector of matrix
In een lineair model zijn de berekende waarden lineaire functies van de parameters: f = Jp zodat de residuelen gelijk zijn aan:

$$\mathbf{r} = \mathbf{f}^{exp} - \mathbf{J}\mathbf{p} \tag{4.10}$$

Substitutie van vergelijking (4.10) in (4.9) levert de normaalvergelijkingen op:

$$-2\mathbf{J}^T \mathbf{W} (\mathbf{f}^{exp} - \mathbf{J}\mathbf{p}) = 0 \tag{4.11}$$

of

$$\mathbf{J}^T \mathbf{W} \mathbf{J} \mathbf{p} = \mathbf{J}^T \mathbf{W} \mathbf{f}^{exp} \tag{4.12}$$

waaruit we de waarde van de parameter expliciet kunnen oplossen:

$$\mathbf{p} = (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} (\mathbf{J}^T \mathbf{W} \mathbf{f}^{exp})$$
(4.13)

In evenwichtssystemen zijn de modellen echter niet lineair, zodat een methode van opeenvolgende benaderingen moet gebruikt worden. In eerste instantie is het nodig een startwaarde voor de parameters te schatten. De vector van de geschatte waarden wordt voorgesteld door p⁰. In het ideale geval benadert de geschatte waarde de optimale waarde en ligt ze met andere woorden dicht bij die waarden waar de gradiëntvector gelijk is aan nul. Elk element van de gradiëntvector wordt daarom ontwikkeld als een eerste orde Taylor-ontwikkeling rond p⁰, zodat we bij benadering kunnen stellen:

$$g_j = g_j^0 + \sum_{k=1}^m \frac{\partial g_j}{\partial p_k} (p_k - p_k^0)$$
(4.14)

of

$$\mathbf{g} = \mathbf{g}^0 + \mathbf{H}\mathbf{s} \tag{4.15}$$

De vector **s** wordt de shiftvector genoemd, **H** is de Hessiaan. Men veronderstelt dat in het nieuwe punt $\mathbf{g} \approx 0$, zodat men in eerste benadering $\mathbf{g} = 0$ kan stellen en vergelijking (4.15) kan oplossen:

$$\mathbf{Hs} = -\mathbf{g}^0 \tag{4.16}$$

De normaalvergelijking is dan wegens (4.9) gelijk aan:

$$\mathbf{Hs} = 2\mathbf{J}^T \mathbf{Wr} \tag{4.17}$$

De elementen van de Hessiaan worden gevonden door vergelijking (4.8) af te leiden:

$$\frac{\partial g_j}{\partial p_k} = -\sum_{h=1}^n \sum_{i=1}^n W_{hi} \left(r_h \frac{\partial^2 f_i}{\partial p_j \partial p_k} + r_i \frac{\partial^2 f_h}{\partial p_j \partial p_k} - \frac{\partial f_h}{\partial p_k} \frac{\partial f_i}{\partial p_j} - \frac{\partial f_i}{\partial p_k} \frac{\partial f_h}{\partial p_j} \right)$$
(4.18)

Bij invoeren van de notatie:

$$C_{jk} = \sum_{h=1}^{n} \sum_{i=1}^{n} W_{hi} \left(r_h \frac{\partial^2 f_i}{\partial p_j \partial p_k} + r_i \frac{\partial^2 f_h}{\partial p_j \partial p_k} \right)$$
(4.19)

wordt de Hessiaan:

$$\mathbf{H} = 2\mathbf{J}^T \mathbf{W} \mathbf{J} - 2\mathbf{C} \tag{4.20}$$

zodat substitutie van vergelijking (4.20) in de normaalvergelijkingen de Newton-Raphson vergelijkingen levert:

$$(\mathbf{J}^T \mathbf{W} \mathbf{J} - \mathbf{C})\mathbf{s} = \mathbf{J}^T \mathbf{W} \mathbf{r}$$
(4.21)

De vector van de residuelen, \mathbf{r} is nu het verschil tussen de waarnemingen en de waarden berekend met de initiële geschatte parameterwaarden.

In de Gauss-Newton methode [21], [22] wordt de matrix C gelijkgesteld aan de nulmatrix (zoals in lineaire systemen), zodat men de volgende normaalvergelijkingen bekomt:

$$\mathbf{J}^T \mathbf{W} \mathbf{J} \mathbf{s} = \mathbf{J}^T \mathbf{W} \mathbf{r} \tag{4.22}$$

Oplossen van deze vergelijkingen levert de shiftvector op:

$$\mathbf{s} = (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} (\mathbf{J}^T \mathbf{W} \mathbf{r}) \tag{4.23}$$

Indien het systeem lineair is, zodat de eerste orde Taylor-ontwikkeling exact is, worden de parameterwaarden waarbij de gradiënt gelijk is aan nul bekomen door de shiftvector op te tellen bij de initiële geschatte waarden.

De som $\mathbf{p}^0 + \mathbf{s}$ wordt als een nieuwe schatting gebruikt bij niet-lineaire systemen waar de Taylor ontwikkeling slechts een benadering is. In dit punt heeft de *U*functie haar benaderende minimumwaarde, U_{\min} bereikt.

Is convergentie bereikt, dan wordt de variantie-covariantie matrix, M(p) van de parameters, berekend volgens de foutenpropagatieregels [22], [23]: uit het

onderliggend model voor lineaire regressie, $y = \mathbf{X}p + \varepsilon$, volgt immers (vergelijking A.53)

$$\hat{p} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T y) \tag{4.24}$$

zodat voor gewogen data, volgens Appendix A vergelijkingen (A.11), (A.54) en (A.55) volgt:

$$\mathbf{M}(\mathbf{p}) = \frac{U_{\min}}{n-m} (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1}$$
(4.25)

De standaardafwijkingen σ_i op de parameters en de correlatie
coëfficiënten ρ_{ij} tussen de parameters kunnen dan respectievelijk berekend worden als:

$$\tau_i = (M_{ii})^{1/2} \tag{4.26}$$

$$\rho_{ij} = \left(\frac{M_{ij}}{M_{ii}M_{jj}}\right)^{1/2} \tag{4.27}$$

4.2.3 Computationele methode

4.2.3.1 Berekening van de vrije concentraties

De originele versie van SUPERQUAD is geprogrammeerd in standaard Fortran (Fortran 77). De werking ervan is gebaseerd op een aantal veronderstellingen waarvan het belangrijk is om ze expliciet te vermelden [8], [17], [18], [19], [22]:

• voor elk complex $M_pL_q...$ in de evenwichtsoplossing bestaat een vormingsconstante:

$$\beta_{pq\dots} = \frac{[M_p L_q \dots]}{[M]^p [L]^q \dots}$$
(4.28)

Door de vormingsconstanten in functie van de concentraties uit te drukken gaat men er impliciet van uit dat de activiteitscoëfficiënten tijdens de titraties constant gehouden worden. Door de titraties in oplossingen met een hoge ionensterkte van een inert elektrolyt uit te voeren is aan deze voorwaarde voldaan;

• men veronderstelt dat een proton- respectievelijk metaalgevoelige ionselectieve elektrode een pseudo-nernstiaans gedrag vertoont:

$$E_H = E^{0'} + sl_{\exp}\log[H] + E_j$$

$$E_M = E^{0'} + sl_{\exp}\log[M] + E_j$$
(4.29)

 sl_{exp} is de helling van de ijkcurve E (mV) vs. $\log[H]$ (resp. $\log[M]$). In het ideale geval, voor een eenwaardig ion en bij 25 °C is deze gelijk aan 59.2 mV/decade;

- voor elk chemisch systeem in evenwicht bestaat een model dat de experimentele waarnemingen op een adequate manier beschrijft;
- de totale concentratie C_i verandert tijdens de titratie volgens:

$$C_{i} = \frac{\text{mmol}_{0} + buret concentratie(M) \cdot V(\text{ml})}{V_{0}(\text{ml}) + V(\text{ml})}$$
(4.30)

Voor gekende waarden van de parameters β_{pqr} en de totale concentraties C_i kunnen de numerieke waarden voor de vrije concentraties [M], [L] en [H] in ieder *i*-de titratiepunt worden berekend. Dit gebeurt door de massabalansen (4.2) op te lossen via een iteratieve Newton-Raphson procedure: een initiële schatting voor de vrije concentraties levert een initiële waarde C_i^{ber} voor elk van de totale concentraties op. Door deze te vergelijken met de gekende totale concentraties C_i^{exp} vinden we $\Delta C_i = C_i^{exp} - C_i^{ber}$, waaruit we dan de verandering in elk van de vrije concentraties kunnen berekenen door middel van:

$$\begin{pmatrix} [M]\frac{\partial C_M}{\partial [M]} & [L]\frac{\partial C_M}{\partial [L]} & [H]\frac{\partial C_M}{\partial [H]} \\ [M]\frac{\partial C_L}{\partial [M]} & [L]\frac{\partial C_L}{\partial [L]} & [H]\frac{\partial C_L}{\partial [H]} \\ [M]\frac{\partial C_H}{\partial [M]} & [L]\frac{\partial C_H}{\partial [L]} & [H]\frac{\partial C_H}{\partial [H]} \end{pmatrix} \begin{pmatrix} \frac{\Delta [M]}{[M]} \\ \frac{\Delta [L]}{[L]} \\ \frac{\Delta [H]}{[H]} \end{pmatrix} = \begin{pmatrix} \Delta C_M \\ \Delta C_L \\ \Delta C_H \end{pmatrix}$$
(4.31)

De partiële afgeleiden in het linkerlid kunnen expliciet berekend worden op basis van de gekende concentraties via 4.33 en 4.34. De reden voor het berekenen van de relatieve shifts moet gezocht worden in het feit dat de elementen van de coëfficiëntenmatrix veelvouden zijn van de concentraties en dat de matrix symmetrisch is bijvoorbeeld: $[L] \left(\frac{\partial C_M}{\partial [L]} \right) = [M] \left(\frac{\partial C_L}{\partial [M]} \right).$

4.2.3.2 Optimalisatie van stabiliteitsconstanten

Omwille van de massabalans weten we dat elke totale concentratie C_i afhangt van de constanten β_{pqr} en de vrije concentraties [M], [L] en [H]. Zoals gezegd is het doel dat SUPERQUAD zich stelt de waarden van de constanten β_{pqr} te bepalen op basis van de gemeten waarden voor pH (en eventueel pM) of de overeenkomstige potentialen E_H (resp. E_M). Hiervoor dienen we de gradiënten $\partial[H]/\partial\beta$ (resp. $\partial[M]/\partial\beta$) te kennen. We beschikken echter niet over een expliciete uitdrukking van [H] ([M]) in functie van de β 's en daarom moeten we overgaan tot impliciete differentiatie. Het impliciete verband tussen voorgenoemde variabelen wordt immers geregeld via de massabalans (4.2): aangezien de totale concentratie in elk punt een experimenteel gegeven is, zullen wijzigingen in de β 's gecompenseerd moeten worden door veranderingen in de vrije concentraties zodat het totale rechterlid opnieuw gelijk wordt aan het gekende linkerlid. Wiskundig wordt dit vertaald in termen van impliciete differentiatie welke het impliciet verband tussen $d\beta$ en d[X] ([X] = [H], resp. [M], [L]) aangeeft omdat C_H (resp. C_M , C_L) constant is (figuur 4.1):



Figuur 4.1: Impliciet verband tussen een verandering in de te optimaliseren parameter, β en een verandering in de vrije concentratie van een component, [X]

$$\frac{dC_M}{d\beta} = 0 = \frac{\partial C_M}{\partial \beta} + \frac{\partial C_M}{\partial [M]} \frac{\partial [M]}{\partial \beta} + \frac{\partial C_M}{\partial [L]} \frac{\partial [L]}{\partial \beta} + \frac{\partial C_M}{\partial [H]} \frac{\partial [H]}{\partial \beta}$$

Vandaar

$$-\frac{\partial C_M}{\partial \beta} = \frac{\partial C_M}{\partial [M]} \frac{\partial [M]}{\partial \beta} + \frac{\partial C_M}{\partial [L]} \frac{\partial [L]}{\partial \beta} + \frac{\partial C_M}{\partial [H]} \frac{\partial [H]}{\partial \beta}$$

In navolging van Nagypál et al. [24], wordt dit opgeschreven onder de vorm van een lineair stelsel: $\frac{\partial[H]}{\partial\beta}$ (resp. $\frac{\partial[M]}{\partial\beta}$) kan berekend worden door volgend stelsel op te

lossen:

$$\begin{pmatrix} [M]\frac{\partial C_{M}}{\partial [M]} & [L]\frac{\partial C_{M}}{\partial [L]} & [H]\frac{\partial C_{M}}{\partial [H]} \\ [M]\frac{\partial C_{L}}{\partial [M]} & [L]\frac{\partial C_{L}}{\partial [L]} & [H]\frac{\partial C_{L}}{\partial [H]} \\ [M]\frac{\partial C_{H}}{\partial [M]} & [L]\frac{\partial C_{H}}{\partial [L]} & [H]\frac{\partial C_{H}}{\partial [H]} \end{pmatrix} \begin{pmatrix} \frac{\beta}{[M]}\frac{\partial [M]}{\partial \beta} \\ \frac{\beta}{[L]}\frac{\partial [L]}{\partial \beta} \\ \frac{\beta}{[H]}\frac{\partial [H]}{\partial \beta} \end{pmatrix} = \begin{pmatrix} -\beta\frac{\partial C_{M}}{\partial \beta} \\ -\beta\frac{\partial C_{L}}{\partial \beta} \\ -\beta\frac{\partial C_{H}}{\partial \beta} \end{pmatrix}$$
(4.32)

Zowel de coëfficiënten uit het linkerlid als het rechterlid kunnen uit de massabalansen (4.2) worden afgeleid. Meer bepaald:

$$[M]\frac{\partial C_M}{\partial [M]} = [M] + \sum p^2 \beta_{pqr} [M]^p [L]^q [H]^r$$
(4.33)

$$[L]\frac{\partial C_M}{\partial [L]} = \sum pq\beta_{pqr}[M]^p[L]^q[H]^r$$
(4.34)

$$\beta_{pqr} \frac{\partial C_M}{\partial \beta_{pqr}} = \sum p \beta_{pqr} [M]^p [L]^q [H]^r$$
(4.35)

Wenst men een andere parameter, zoals de initiële hoeveelheid of de concentratie van een buretoplossing te optimaliseren, dan kan de afgeleide berekend worden uit de relevante vergelijking, in dit geval vergelijking (4.30).

4.2.3.3 De gewichtenmatrix

In SUPERQUAD veronderstelt men dat zowel de titervolumes als de potentiaalmetingen onderhevig zijn aan willekeurige fouten. SUPERQUAD is het eerste programma dat hiermee rekening houdt. De functie U, door SUPERQUAD geminimaliseerd, wordt gegeven door:

$$U = \sum_{h=1}^{n} \sum_{i=1}^{n} W_{hi} (E_h^{\exp} - E_h) (E_i^{\exp} - E_i)$$
(4.36)

De gewichten worden, in navolging van Deming [25], berekend via een kleinste kwadraten minimalisatie, waarbij fouten zowel in de afhankelijke als in de onafhankelijke variabele beschouwd worden [22]. De elektrodepotentialen worden gekozen als afhankelijke variabelen, het toegevoegd volume als een onafhankelijke veranderlijke. Er worden constante fouten σ_E aan de potentiaal en σ_V aan het toegevoegd volume toegekend. Is slechts één ionselectieve elektrode aanwezig en veronderstellen we dat er geen correlatie bestaat tussen de fouten in de verschillende titratiepunten, dan wordt de covariantiematrix herleid tot een diagonaalmatrix met als elementen:

$$\sigma_i^2 = \sigma_E^2 + \left(\frac{\partial E}{\partial V}\right)_i^2 \sigma_V^2 \tag{4.37}$$

en kunnen de gewichten bepaald worden als:

$$W_{ii} = \frac{1}{\sigma_i^2} \tag{4.38}$$

In dat geval wordt de functie, U (vergelijking 4.6) vereenvoudigd tot de gewogen sommatie van de kwadraten van de residuelen:

$$U = \sum_{i=1}^{n} W_{ii} r_i^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2} r_i^2$$
(4.39)

Zijn er meerdere ionselectieve elektroden aanwezig zijn en veronderstellen we opnieuw dat er geen correlatie is tussen de fouten van de verschillende titratiepunten, dan is de covariantie matrix een blok-diagonaalmatrix. We verwijzen hier naar vergelijking (A.11) van appendix A voor de berekening van de variantie op een functie van meerdere variabelen. De grootte van elk blok is afhankelijk van het aantal gebruikte elektroden. Wanneer twee ionselectieve elektroden gebruikt worden zal ieder blok de vorm hebben:

$$\begin{pmatrix} \sigma_{E_1}^2 + \left(\frac{\partial E_1}{\partial V}\right)_i^2 \sigma_V^2 & \left(\frac{\partial E_1}{\partial V}\right)_i \left(\frac{\partial E_2}{\partial V}\right)_i \sigma_V^2 \\ \left(\frac{\partial E_1}{\partial V}\right)_i \left(\frac{\partial E_2}{\partial V}\right)_i \sigma_V^2 & \sigma_{E_2}^2 + \left(\frac{\partial E_2}{\partial V}\right)_i^2 \sigma_V^2 \end{pmatrix}$$
(4.40)

waarbij E_1 en E_2 de twee elektrodepotentialen voorstellen. De afgeleide $\frac{\partial E}{\partial V}$ wordt numeriek berekend, bijvoorbeeld door een kubische polynoom doorheen vijf opeenvolgende titratiepunten te fitten. Vergelijking (4.22) wordt in SUPERQUAD uitgebreid met de Marquardt parameter, λ als protectie tegen divergentie. Deze uitbreiding werd reeds in MINIQUAD75 [15] geïmplementeerd en uitgetest:

$$(\mathbf{J}^T \mathbf{W} \mathbf{J} + \lambda \mathbf{D})\mathbf{s} = \mathbf{J}^T \mathbf{W} \mathbf{r}$$
(4.41)

D bevat de diagonale elementen van $\mathbf{J}^T \mathbf{W} \mathbf{J}$. Het nut van deze parameter bestaat er in dat men op een continue wijze kan variëren tussen de voorgestelde kleinste kwadraten methode en de meer robuste maar langzamere methode van "gradient descent" of

"steilste daling". Voor grote waarden van λ zullen de elementen van $\lambda \mathbf{D}$ domineren over de elementen van $\mathbf{J}^T \mathbf{W} \mathbf{J}$ zodat de shiftvector s vooral in de richting van de sterkste daling wijst; de lengte van de shiftvector zal omgekeerd evenredig verminderen met de grootte van λ :

$$\mathbf{s} = \frac{1}{\lambda} \mathbf{D}^{-1} \mathbf{J}^T \mathbf{W} \mathbf{r} \tag{4.42}$$

Gedurende de iteratie wordt de Marquardt parameter stelselmatig verkleind zodat wanneer men in de buurt van het minimum komt, λ convergeert naar 0. Hierdoor wordt de optimalisatie gewoon een kleinste kwadraten verfijning.

4.2.3.4 Convergentiecriteria

Statistisch kan men aantonen dat wanneer de U-functie rechtstreeks gebaseerd is op experimentele grootheden en de gewichtenmatrix W gelijk is aan het inverse van de variantie-covariantie matrix van de geobserveerde grootheden, $\left(\frac{U}{n-m}\right)^{1/2}$ een verwachtingswaarde gelijk aan 1 heeft (Appendix A, paragraaf A.9). Is het model correct en zijn er geen systematische fouten in de gegevens aanwezig, dan is deze waarde onafhankelijk van de waarschijnlijkheidsdistributiefunctie van de willekeurige experimentele fouten. De S-statistiek (of sample standaardafwijking) wordt gedefinieerd als:

$$S = \left(\frac{\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}}{n-m}\right)^{1/2} = \left(\frac{U_{\min}}{n-m}\right)^{1/2} \tag{4.43}$$

met ε de vector van de potentiaalresiduelen, *n* het aantal observaties en *m* het aantal te verfijnen parameters. De *S*-waarde wordt gebruikt om de 'goodness-of-fit' van de individuele modellen te vergelijken. Cruciaal in deze berekening is een goede schatting van σ_E en σ_V . Verschillende auteurs [26], [27], [28], [29] vermelden dat met een goede potentiometrische opstelling σ_E gelijk aan 0.1 - 0.2 mV haalbaar is. Een schatting van σ_V is afhankelijk van de buret die men gebruikt. Voor een buret met een resolutie van 0.001 ml variëren de schattingen van 0.002 [26] tot 0.005 ml [28]. Het is aan te raden zowel σ_E als σ_V experimenteel te bepalen. Meestal wordt een *S*-waarde tussen 1 en 3 beschouwd als aanvaardbaar voor een goed model.

Een tweede statistiek, de χ^2 -test [29], wordt eveneens door SUPERQUAD berekend. De bedoeling van deze Pearson-test is te controleren of de residuelen normaal verdeeld zijn

(in de MatLab implementatie van SUPERQUAD gebeurt dit op basis van normaliteitsplots: zie hoofdstuk 6). De residuelen worden verdeeld in acht klassen, bepaald door de grenzen:

$$-\infty; -1.15S_r; -0.675S_r; -0.319S_r; 0.0; 0.319S_r; 0.675S_r; 1.15S_r; +\infty$$
(4.44)

met S_r de residuele standaardafwijking. De verwachte populatie voor iedere groep is 1/8 van de totale populatie. Voor iedere *j*-de groep wordt eerst de werkelijke populatie en een partiële χ_j^2 berekend. Daarna worden deze χ_j^2 opgeteld om de totale χ^2 te bekomen:

$$\chi_j^2 = \frac{(\# \text{ aanwezige meetpunten - } (\text{totaal } \# \text{ meetpunten}/8))^2}{\text{totaal } \# \text{ meetpunten}/8}$$

$$\chi^2 = \sum_{j=1}^8 \chi_j^2$$

De totale χ^2 kan men vergelijken met getabellariseerde waarden. Vermits de residuele standaardafwijking berekend is, is het aantal vrijheidsgraden van de totale χ^2 gelijk aan 5(= 8(residuelen)-1(gemiddelde)-2(intercept,helling)). Voor een betrouwbaarheidsinterval van 95% vindt men dat de verdeling als normaal beschouwd mag worden indien de totale χ^2 kleiner is dan 12.6. Indien de *S*-waarde aanvaardbaar is maar χ^2 groot (> 12.6) zou dit op systematische tendensen in de residuelen duiden en een aanwijzing kunnen zijn voor het bestaan van systematische fouten in de gegevens of in de keuze van het model. SUPERQUAD heeft een optie om de residuelen grafisch voor te stellen zodat inspectie ervan mogelijk is.

4.2.4 Implementatie van SUPERQUAD in MatLabTM

Het algoritme van SUPERQUAD werd geïmplementeerd in een MatLabTM programma 'SQ_simul.m' dat modulair is opgebouwd volgens een hoofdprogramma gekoppeld aan verschillende deelprogramma's (functies in MatLabTM) vergelijkbaar met de "ESTA" (Equilibrium Simulation for Titration Analysis) programma's van May et al. [30]. Hierna volgt een gedetailleerde beschrijving van elke functie ('M-file') die wordt opgeroepen in 'SQ_simul.m' [31], [32], [33].

Daar we voor ons verder werk omtrent de beoordeling van de te verwachten fout op de stabiliteitsconstanten beroep zullen doen op Monte Carlo simulatiestudies hebben we naast het eigenlijke SUPERQUAD-algoritme ook MatLab-routines geschreven voor

4.2 SUPERQUAD

het genereren van gesimuleerde "experimentele" gegevens die kunnen dienen als input voor 'SQ.m' (zie verder). (Al deze routines zijn gecombineerd in het simulatiepakket 'SQ_simul.m' dat op aanvraag kan verkregen worden²). Op die manier kunnen we het gehele verloop van een potentiometrisch experiment simuleren en het uiteindelijke resultaat vergelijken met de gekende ground truth. Dit laat ons toe om te onderzoeken hoe onnauwkeurigheden in de verschillende experimentele stadia zich propageren tot in het eindresultaat.

In onderstaand schema wordt de algemene structuur van 'SQ_simul.m' met de verschillende datastructuren en functies weergegeven.

Data

- 1. nr_titr_pts: aantal titratiepunten
- 2. nr_species: aantal complexen
- 3. nr_comp: aantal componenten
- 4. <u>P:</u> modelmatrix van het systeem met dimensie $(nr_comp \times nr_species)$ met P_{kj} = exponent van voorkomen van de k-de component in het j-de complex
- 5. β_j (j = 1...nr_species): stabiliteits constante van het j-de gevormde complex
- 6. C_k $(k = 1...nr_comp)$: totale concentratie van de k-de component
- 7. X_k (k = 1... nr_comp): vrije concentratie van de k-de component.

Overzicht van de programma's

De schematische opbouw van het hoofdprogramma 'SQ_simul.m' door middel van de verschillende subroutines is als volgt:

'SQ_simul.m'

- 1. Genereren van "synthetische" (experimentele) data:
 - 'gen_exp_data.m': genereert op basis van het totaal beginvolume in de recipiënt, het toegevoegd volume base over het gehele titratiegebied, de beginconcentratie base en de totale beginconcentratie van elke component de waarden voor het totaal volume in de recipiënt en de totale concentratie van elke component in elke titratiestap;

²E-mail: wanda.guedens@luc.ac.be

- 'equil.m': genereert op basis van de output van 'gen_exp_data.m', de modelmatrix en "gekende" β's de waarden voor de vrije concentratie van de componenten en de complexen [34];
- 2. Optimaliseren van β op basis van de experimentele (synthetisch of reëel) gegevens:
 - 'SQ.m' (gedetailleerde beschrijving zie hieronder)
 - 'jacobian_free.m'
 - 'jacobian_beta.m'

Programmadetails

Eenmaal de vrije concentraties van de componenten gegenereerd zijn, wordt in de functie 'SQ.m' – het eigenlijke SUPERQUAD-gedeelte van 'SQ_simul.m' waar de optimalisatie van β 's gebeurt – de modelfunctie, U (vergelijking 4.36) geminimaliseerd met behulp van de functies 'jacobian_free.m' en 'jacobian_beta.m'.

'SQ.m':

- Input: P, startwaarden voor β_j, C_k, gegenereerde waarden voor X_k, aantal titratiepunten i;
- Output: geoptimaliseerde β_j 's per titratie;
- Methode:
 - i) 'equil.m': input = C_k, startwaarden voor β_j, P, startwaarden voor X_k; output = nieuwe waarden voor X_k; methode = minimalisatie van ∑_k ΔC²_k met ΔC_k = C^{exp}_k - C^{ber}_k(X_k, β_j), de residuelen in massabalans;
 - ii) 'jacobian_free.m': input = gegenereerde X_k , initiële geschatte β_j 's; output = $\frac{\partial C_k}{\partial X_l}$, m.a.w. de coëfficiënten van het linkerlid van vergelijking (4.32). Deze partiële afgeleiden worden berekend volgens de relevante vergelijkingen vermeld in Appendix B afhankelijk van het model;
 - iii) 'jacobian_beta.m': input = gegenereerde X_k , initiële geschatte β_j 's; output = $\frac{\partial C_k}{\partial \beta_j}$, de coëfficiënten van het rechterlid van vergelijking (4.32). De relevante vergelijkingen zijn weergegeven in Appendix C afhankelijk van het model;

 iv) oplossen van stelsel 4.32 via de output van 'jacobian_free.m' (F) en 'jacobian_beta.m' (G) (coëfficiënten van respectievelijk linker-en rechterlid van stelsel):

$$\mathbf{X} = \mathbf{F}^{-1}\mathbf{G} \tag{4.45}$$

waarbij X (dimensies $k \times j$) de elementen $\frac{\partial X_k}{\partial \beta_j}$ bevat;

v) opstellen van de normaalvergelijkingen volgens (4.22):

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{s} = \mathbf{X}^T \mathbf{W} \mathbf{r} \tag{4.46}$$

en volgens vergelijking (4.9)

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{s} = \mathbf{g} \mathbf{W} \tag{4.47}$$

Oplossen van vergelijking (4.47) levert de shiftvector, s op, waarbij de inverse van de matrix, $\mathbf{X}^T \mathbf{X}$ berekend wordt volgens de 'singuliere waarde decompositie' techniek [21], [31]. De waarde van de shiftvector wordt opgeteld bij de initiële schatting voor β en de som wordt gebruikt als nieuwe schatting voor verdere optimalisatie.

Uitgaande van de hierboven besproken functies en variabelen wordt in volgend organigram (figuur 4.2) de opzet voor de foutenpropagatiestudie bij de bepaling van stabiliteitsconstanten (hoofdstuk 6), vastgelegd. De inputgegevens kunnen experimenteel $(pM (E_M), pH (E_H))$ of synthetisch van oorsprong zijn:

- (a) in de synthetische route worden via het MatLab programma 'equil.m' vrije concentraties X_k^{exp} gegenereerd uitgaande van de "echte" waarden voor β ($\beta_{j_{echt}}$), Pen C_k (de zogenaamde "ground truth");
- (b) alternatief: X_k^{exp} op basis van experimentele pM (E_M) en pH (E_H) (vergelijking (4.29);
- (c) MatLab programma 'SQ.m' berekent nieuwe vrije concentraties, X_k^{ber} en geoptimaliseerde β_j's (β_{jgeoptimaliseerd}) uitgaande van startwaarden voor de β_j's, P en C_k. Cruciaal in dit opzet is de overeenkomst tussen β_{jgeoptimaliseerd} en β_{jecht}. In hoofdstuk 6 gaan we de invloed na van perturbaties op verschillende stappen in dit organigram.



Figuur 4.2: Organigram voor de Monte Carlo simulatiestudie van de foutenpropagatie bij de bepaling van stabiliteitsconstanten

4.2.5 Validatie van het MatLabTM programma

Synthetische data zonder perturbatie

Met behulp van synthetische potentiometrische datasets gegenereerd via het eerste gedeelte van het MatLab programma wordt 'SQ_simul.m' gevalideerd aan de hand van een vergelijking met de standaardversie van SUPERQUAD in fortran 77.

Tabel 4.2 Samenstelling en globale stabiliteitsconstanten (β) in log eenheden voor twee artificiële modelsystemen 'sys1' en 'sys2'

	'sys1'		'sys2'
deeltje	$\log \beta$	deeltje	$\log \beta$
ML	8	MLH	12
ML_2	14	ML_2	6
HL	9	ML	3
H_2L	11	HL	10
		H_2L	17
		OH^-	-14

Twee artificiële modelsystemen 'sys1' en 'sys2', weergegeven in tabel 4.2, worden geanalyseerd. De synthetische data, aangemaakt op basis van de "echte" waarden voor de stabiliteitsconstanten (tabel 4.2), worden ingegeven in het originele fortran programma. De initiële geschatte waarden voor de constanten in de fortran versie (f77) zijn precies dezelfde als deze in de MatLab versie. De optimalisatie in beide versies gebeurt met deze initiële startwaarden. Voor de simulaties in functie van de twee modelsystemen (tabel 4.3) wijken de initiële schattingen respectievelijk 50% (experiment 1) en 70% (experiment 2) af van de "echte" waarden (arbitrair). Zoals gesteld, veronderstelt SUPERQUAD dat zowel de titervolumes als de potentiaalmetingen onderhevig zijn aan experimentele fouten. De functie U, door SUPERQUAD geminimaliseerd, wordt gegeven door vergelijking (4.36), waarbij de gewichten, in navolging van Deming [25], berekend worden via de techniek van de 'rigoureuze' kleinste kwadraten minimalisatie. Er moeten dus constante fouten σ_E aan de potentiaal en σ_V aan het toegevoegd volume base worden toegekend. Bij een goede experimentele set-up is een waarde voor σ_E van 0.1 mV of 0.002 pH-eenheden aanvaardbaar [18]. De simulaties in deze paragraaf veronderstellen σ_V gelijk aan 0, daar een schatting voor één van beide termen in vergelijking 4.37 voldoende blijkt.

		Gen (0.1 - 17 - 0 - 1)	Mattab (idam)
'sys1'		$177 \ (\sigma_E = 0.1 \ mv, \ \sigma_V = 0 \ mu)$	Mathab $(o_E, o_V, \operatorname{idem})$
experiment	$\beta_{ m geschat}$	$\beta_{\text{geoptimaliseerd}}$	$\beta_{ m geoptimaliseerd}$
1	$8.17609^{a}(8), 13.69897^{b}(14)$	$8.00015^a, 14.00027^b$	$8.00000^a, 14.00000^b$
	$9.17609^{c}(9), 10.69897^{d}(11)$	$9.00030^c, 11.00070^d$	$9.00000^c, 11.00000^d$
2	$8.23045^{a}(8), 13.47712^{b}(14)$	$8.00026^a, 14.00034^b$	7.99999 ^a , 13.99998 ^b
	$9.23045^{c}(9), 10.47712^{d}(11)$	$8.99968^c, 10.99989^d$	$8.99998^c, 11.00000^d$
'sys2'		f77 ($\sigma_E = 0.1 \ mV, \ \sigma_V = 0 \ ml$)	MatLab (σ_E, σ_V : idem)
experiment	$\beta_{ m geschat}$	$\beta_{\rm geoptimaliseerd}$	$\beta_{\rm geoptimaliseerd}$
1	$12.17609^{a}(12), 5.69897^{b}(6)$	$12.00249^a, 6.00027^b$	$12.00000^a, 6.00000^b$
	$3.17609^{c}(3), 9.69897^{d}(10)$	3.00110^c , 10.00026^d	3.00000^c , 10.00000^d
	$17.17609^e(17),-14.30103^f(-14)$	$17.00066^e, \ -14.00056^f$	$17.00000^e, -14.00000^f$
2	$12.23045^{a}(12), 5.47712^{b}(6)$	$12.00312^a, 6.00030^b$	12.00001 ^a , 6.00000 ^b
	$3.23045^{c}(3), 9.47712^{d}(10)$	3.00355 ^c , 10.00063 ^d	3.00002^c , 10.00000^d
	$17.23045^{e}(17) - 14.52288^{f}(-14)$	17.00084^{e} , -14.00079^{f}	$17.00001^e, -13.99999^f$

114

'sys2' $a = \beta_{111}, \ b = \beta_{120}, \ c = \beta_{110}, \ a = \beta_{011}, \ c = \beta_{012}, \ r = \beta_{00-1}$

De resultaten van een kwantitatieve validatieanalyse voor beide SUPERQUAD versies aan de hand van 'sys1'en 'sys2' worden weergegeven in tabel 4.3. Tabel 4.3 illustreert dat de geoptimaliseerde parameterwaarden in beide programma's goed overeenkomen, zodat het MatLab programma kan concurreren met de fortran versie. Bovendien convergeren de geoptimaliseerde constanten in beide versies naar hun "echte" waarde (tabel 4.2).

Synthetische data met perturbatie in pH en pM

Om de experimente le situatie beter te benaderen wordt de matrix met de synthetische vrije concentraties van M, L en H geperturbe erd met een random, gaussiaans verdeelde fout, ε :

$\mathbf{pMpLpH_error} = \mathbf{pMpLpH_exp} + \varepsilon$

met $\varepsilon \sim N(0, 0.002 \ pH)$. Deze geperturbeerde data (enkel pH_{error}) worden als input ingegeven in de originele fortran versie. De initiële geschatte parameterwaarden in de fortran versie (f77) zijn precies dezelfde als deze in de MatLab versie. De optimalisatie zowel in de fortran versie als in het MatLab programma gebeurt met deze initiële startwaarden. Er worden ook hier constante fouten σ_E aan de potentiaal en σ_V aan het toegevoegd volume base toegekend in beide SUPERQUAD versies: $\sigma_E = 0.1 \text{ mV}$ en $\sigma_V = 0 \text{ ml}$, daar de perturbatie enkel in de afhankelijke variabele (de pH) gebeurt.

De geschatte startparameters in tabel 4.4 zijn arbitrair gekozen en wijken voor de eerste drie simulatie-experimenten 50% af van de "echte" parameterwaarde (tabel 4.2). In experiment 4 en 5 bedraagt de afwijking 70%. Het resultaat van een kwantitatieve validatie aan de hand van 'sys1' voor beide SUPERQUAD versies wordt samengevat in tabel 4.4. Uit tabel 4.4 blijkt opnieuw dat de geoptimaliseerde parameterwaarden goed overeenkomen in beide versies en tevens convergeren naar de "echte" waarde van de parameter. Er dient opgemerkt te worden dat bij simulaties aan de hand van 'sys1' met σ_E in de fortran versie verwaarloosbaar klein t.o.v. 0.1 mV (niet in de tabel vermeld) wel convergentie naar de "echte" waarde van de parameter bereikt werd, maar dat de S-statistiek (model fit) zeer slecht was, hetgeen verwacht kon worden.

Een validatie aan de hand van 'sys2' voor beide SUPERQUAD versies wordt samengevat in tabel 4.5. De geschatte parameterwaarden zijn eveneens arbitrair gekozen en wijken voor de experimenten met nummer 1, 2 en 3, 50% af van de "echte" waarde (tabel 4.2). In de experimenten met nummer 4 en 5 bedraagt de afwijking 70%.

'sys1'		f77 ($\sigma_E = 0.1 \ mV, \ \sigma_V = 0 \ ml$)	MatLab (σ_E, σ_V : idem)	
experiment	$\beta_{ m geschat}$	$\beta_{\text{geoptimaliseerd}}$	$\beta_{geoptimaliseerd}$	
1	8.17609 ^a (8), 13.69897 ^b (14)	8.01484 ^a , 14.01305 ^b	8.01389 ^a , 14.01204 ^b	
	$9.17609^{c}(9), 10.69897^{d}(11)$	$9.00198^{\circ}, 11.00622^{d}$	9.00165 ^c , 11.00475 ^d	
2	$8.17609^{a}(8), 13.69897^{b}(14)$	8.00099 ^a , 14.00034 ^b	7.99995 ^a , 13.99925 ^b	
	$9.17609^{c}(9), 10.69897^{d}(11)$	8.99942 ^c , 10.99699 ^d	8.99908 ^c , 10.99541 ^d	
3	$8.17609^{a}(8), 13.69897^{b}(14)$	7.99994^a , 14.00064^b	$7.99972^a, 14.00030^b$	
	$9.17609^{c}(9), 10.69897^{d}(11)$	$9.00022^c, 11.00115^d$	8.99990 ^c , 11.00048 ^d	
4	$8.23045^{a}(8), 13.47712^{b}(14)$	$8.00035^a, 14.00128^b$	8.00013 ^a , 14.00095 ^b	
	$9.23045^{\circ}(9), 10.47712^{d}(11)$	$9.00039^c, 11.00147^d$	9.00009 ^c , 11.00078 ^d	
5	$8.23045^{a}(8), 13.47712^{b}(14)$	8.00024^a , 14.00036^b	7.99999 ^a ,14.00000 ^b	
	$9.23045^{c}(9), 10.47712^{d}(11)$	$9.00030^{\circ}, 11.00075^{d}$	9.00000^c , 11.00000^d	
gem.		$8.00327^{a'}, 14.00313^{b'}$	$8.00274^{a'}, 14.00251^{b'}$	
$\pm st.dev.$		$\pm 0.00648 \pm 0.00556$	$\pm 0.00624 \pm 0.00536$	
		$9.00046^{c'}, 11.00132^{d'}$	$9.00015^{c'}, 11.00028^{d'}$	
		$\pm 0.00093 \pm 0.00328$	$\pm 0.00093 \pm 0.00321$	
$^{a}=eta_{110},^{b}=eta_{120},^{c}=eta_{011},^{d}=eta_{012}$				
$=\beta_{110}(gem.), \ b'=\beta_{120}(gem.), \ c'=\beta_{011}(gem.), \ d'=\beta_{012}(gem.)$				

Tabel 4.4 Validatie van 'SQ_simul.m' (MatLab) in vergelijking met SUPERQUAD (f77) a.h.v. 'sys1'. Gesimuleerde pH-data met $\varepsilon \sim N(0, 0.002 \text{ pH}), \beta$'s in log eenheden, "echte" β 's tussen haakjes

Hoofdstuk 4

112

Analyse van potentiometrische data

'sys2'	f77 ($\sigma_E = 0.1 \ mV, \ \sigma_V = 0 \ ml$)	MatLab (σ_E , σ_V : idem)
$\beta_{ ext{geschat}}$	$\beta_{\text{geoptimaliseerd}}$	$\beta_{\text{geoptimaliseerd}}$
exp.1) $12.17609^{a}(12), 5.69897^{b}(6)$	negatief! ^a , 5.67556 ^b	$11.83563^a, 5.95968^b$
$3.17609^{c}(3), 9.69897^{d}(10)$	2.46930^c , 9.99321^d	$2.94068^c, 10.00000^d$
$17.17609^{e}(17), -14.30103^{f}(-14)$	$16.93124^e, -14.00638^f$	$16.99080^e, -14.00174^f$
exp.2) 12.17609 ^a (12), $5.69897^{b}(6)$	$12.04460^a, 6.05701^b$	$11.96534^a, 6.03543^b$
$3.17609^{c}(3), 9.69897^{d}(10)$	3.01284^c , 10.00526^d	$2.97186^c, 10.00500^d$
$17.17609^e(17), -14.30103^f(-14)$	$17.00787^e, -14.00529^f$	$17.00212^e, -14.00631^f$
exp.3) 12.17609 ^a (12), 5.69897 ^b (6)	11.95511^a , 5.94045^b	11.96972 ^a , 5.94712 ^b
$3.17609^{c}(3), 9.69897^{d}(10)$	2.99178^c , 9.99256^d	3.00061^c , 9.99285^d
$17.17609^{e}(17), -14.30103^{f}(-14)$	$16.99211^e, -14.00161^f$	$16.99306^e, -14.00022^f$
exp.4) $12.23045^{a}(12), 5.47712^{b}(6)$	negatief! ^a , 5.65632 ^b	11.98538 ^a , 6.00243 ^b
$3.23045^{c}(3), 9.47712^{d}(10)$	2.40960^c , 9.99096^d	$2.99698^c, 10.00017^d$
$17.23045^{e}(17), -14.52288^{f}(-14)$	$16.92530^e, -14.01341^f$	$16.99878^e, -14.00462^f$
exp.5) $12.23045^{a}(12), 5.47712^{b}(6)$	$12.04600^a, 6.05703^b$	$11.96535^a, 6.03540^b$
$3.23045^{c}(3), 9.47712^{d}(10)$	3.01286^c , 10.00003^d	$2.97180^c, 10.00497^d$
$17.23045^{e}(17), -14.52288^{f}(-14)$	$17.00403^e,-14.00300^f$	$17.00210^e,\ -14.00630^f$
gemiddelde	$12.01477^{a'}, 5.87727^{b'}$	$11.94428^{a'}, 5.99601^{b'}$
$\pm { m st.dev}.$	$\pm 0.05166 \ \pm 0.19882$	$\pm.06130 \pm 0.04140$
	$2.77928^{c'}, 9.99640^{d'}$	$2.97639^{e'}, 10.00060^{d'}$
	$\pm 0.31105 \pm 0.00605$	$\pm 0.02412 \pm 0.00498$
	$16.97211^{e'}, -14.00594^{f'}$	$16.99737^{e'}, -14.00384^{f'}$
	$\pm 0.04049 \pm 0.00458$	$\pm 0.00521 \ \pm 0.00275$
$a = \beta_{111}, b = \beta_{120}, c = \beta_{110}, d = \beta_{011}, e = \beta_{012}, f = \beta_{00-1}$		
$a' = \beta_{111}(gem.), b' = \beta_{120}(gem.), c' = \beta_{110}(gem.), d' = \beta_{011}(gem.), e' = \beta_{012}(gem.), f' = \beta_{00-1}(gem.)$		

Tabel 4.5 Validatie van 'SQ_simul.m' (MatLab) in vergelijking met SUPERQUAD (f77) a.h.v.'sys2'. Gesimuleerde pH-data met $\varepsilon \sim N(0, 0.002 \text{ pH}), \beta$'s in log eenheden, "echte" β 's tussen haakjes

113

4.2 SUPERQUAD

In het geval van 'sys2' dient ten eerste opgemerkt te worden dat niettegenstaande het hoger aantal vrijheidsgraden de convergentie naar de exacte of "echte" waarde van β moeilijker bereikt wordt in beide SUPERQUAD versies. Dit wordt eveneens vastgesteld door Lomozik et al. [35], [36], [37] in een kwalitatieve performantie-analyse van SUPERQUAD. Verder toont tabel 4.5 aan dat de MatLab versie systematisch een beter optimum geeft: het complex *MLH* wordt in de f77 versie in twee experimenten zelfs verworpen (negatieve waarde voor β).

Voor een validatie aan de hand van de experimentele data verwijzen we naar hoofdstuk 5.

4.3 Interpretatie van de resultaten

4.3.1 Modelselectie

4.3.1.1 Criteria

Het vinden van het model dat de experimentele situatie het nauwkeurigst beschrijft is één van de belangrijkste maar ook één van de moeilijkste taken bij evenwichtsberekeningen gebaseerd op een kleinste kwadraten analyse. Men moet dus een strategie ontwikkelen om het "beste" model te achterhalen.

Het startmodel moet een zinvolle weergave zijn van de deeltjes die tijdens de titraties gevormd kunnen worden. Een eerste schatting van de β -waarden haalt men uit vergelijkbare systemen in de literatuur of uit een grafische analyse van de titratiegegevens. Dit model wordt verfijnd met het minimalisatieprogramma. De verfijnde waarden van de β 's worden opnieuw gebruikt, samen met een mogelijk nieuw deeltje waar men een zo goed mogelijk geschatte β -waarde voor heeft. Heeft men een beduidend beter minimum gevonden, is de fout op de stabiliteitsconstanten kleiner dan een vooropgestelde norm en is het percentage van voorkomen van het nieuwe deeltje voldoende relevant, dan mag men veronderstellen dat het nieuwe model een betere beschrijving van de experimentele data geeft. Hierna kan men weer een nieuw deeltje toevoegen en de procedure herhalen. Wordt het deeltje verworpen, dan wordt het door een ander complex vervangen en verfijnt men opnieuw. Deze werkwijze wordt herhaald totdat alle zinvolle chemische combinaties uitgetest zijn. Normaal zal het model met het laagste minimum tevens het best de samenstelling van de oplossing weergeven. Dit model omvat

het 'beste' stel deeltjes, een verzameling die het best overeenstemt met de nauwkeurig gemeten eigenschappen van een reëel systeem.

De eigenlijke modelselectie in SUPERQUAD is gebaseerd op een startmodel waarbij aan iedere vormingsconstante een soort code wordt toegekend. Deze kan 1 (stabiliteitsconstante wordt geoptimaliseerd), 0 (constante blijft ongewijzigd) of -1 (constante wordt genegeerd) zijn. Het beste model moet de laagste S-waarde hebben en mag bovendien geen slecht gedefinieerde constanten bevatten. Een vormingsconstante is slecht gedefinieerd wanneer deze negatief is of een standaardafwijking groter dan 33% van zijn eigen waarde heeft [22]. Tijdens de verfijning zelf mogen de constanten negatieve waarden aannemen. Dit levert natuurlijk problemen op bij de berekening van de vrije concentraties, maar SUPERQUAD gebruikt verschillende technieken om dit op te vangen. Indien de verfijning convergeert tot een punt waar een of meerdere vormingsconstanten negatief zijn, betekent dit dat dit model slecht is en verworpen moet worden. Elk nieuw model gebruikt de vorige beste waarden voor een nieuwe optimalisatie.

Is het model succesvol verfijnd, dan worden de correlatiecoëfficiënten berekend volgens vergelijking (4.27). De waarde van deze coëfficiënten ligt tussen 1 en -1. Ze zijn een maat voor de afhankelijkheid van de ene parameter ten opzichte van de andere, als de overige parameters constant gehouden worden. Waarden dicht bij 1 of -1 geven aan dat de experimentele gegevens niet toelaten om de parameter afzonderlijk te definiëren. Het is mogelijk een zeer goede fit te bekomen met de gegevens, terwijl één van de parameters volledig foutief geschat is. Deze fout kan immers gecompenseerd worden door andere foutief geschatte parameters. Soms is dit inherent aan het te onderzoeken systeem en daardoor praktisch onmogelijk te vermijden.

4.3.1.2 Stroomdiagram

Een vereenvoudigd schema van de modelselectie in SUPERQUAD is weergegeven in figuur 4.3.



Figuur 4.3: Vereenvoudigd schema van de modelselectie in SUPERQUAD

- i) inlezen van de data en een basisset van vormingsconstanten. Met elke constante worden de stechiometrische coëfficiënten a, b, ... en de code 1, 0 of -1 geassocieerd.
- ii) bij het inlezen van een nieuw model wordt de code aangepast. Op deze manier wordt een reeks van kandidaatmodellen geëvalueerd.
- iii) een model wordt maar 1 maal verfijnd.
- iv) de data (vormingsconstanten) die overeenkomen met het "beste" model worden bewaard.

- v) elk nieuw onderzocht model gebruikt als startwaarden de constanten uit het beste model en wordt daarna verfijnd (vi).
- vii) in het geval van computerstoringen wordt onmiddellijk een nieuw model gegenereerd.
- viii) bij convergentie worden de waarden van vormingsconstanten, de standaardafwijkingen, de correlatiecoëfficiënten, en een χ^2 statistiek weergegeven in een tussentijdse output.
- ix) controle voor slecht gedefinieerde constanten.
- x) het huidig model wordt verworpen bij het vinden van slecht gedefinieerde constanten door de codes op -1 te zetten. Automatisch wordt een nieuw model verfijnd.
- xi) indien alle constanten goed gedefinieerd zijn wordt de volledige output, inclusief plots van de residuelen en de verdeling van de verschillende species weergegeven.

4.3.2 Foutenanalyse

In het licht van de foutenpropagatiestudie (hoofdstuk 6) willen we wijzen op twee belangrijke problemen die zich stellen bij het verklaren van de resultaten verkregen met behulp van SUPERQUAD: enerzijds de interpretatie van de berekende standaardafwijkingen en anderzijds de analyse van de gewogen residuelen.

Zoals bij een kleinste kwadraten benadering dikwijls het geval is zijn de berekende standaardafwijkingen idealistisch en is het niet ongewoon dat voor eenvoudige systemen standaardafwijkingen op de stabiliteitsconstanten van 0.001 of zelfs 0.0001 log-eenheden worden bekomen. Dit impliceert dat de constanten met een hogere precisie bepaald zouden zijn dan de haalbare accuratesse in de potentiometrie [26]. Indien alle mogelijke foutenbronnen geëlimineerd zouden zijn, komt de nauwkeurigheid van de glaselektrode, $\sigma_{E_{glas}} = 0.2 \text{ mV}$, overeen met een systematische fout van 0.003 pH-eenheden. Bates schat de accuratesse van een glaselektrode tussen twee of meer metingen op 0.01 - 0.02 pH-eenheden [38], hetgeen een fout van 0.01 - 0.02 pK_A-eenheden oplevert [39], [40], [41]. Volgens Avdeef et al. [42] kunnen bij de routinematige bereiding van stock-oplossingen zelfs fouten van 1% op de massabalansen optreden.

Variantieanalysen (ANOVA) uitgevoerd door Braibanti et al. op de gegevens van het Nikkel-Glycine-project [43], [44], [45], [46] en door Casassas en Tauler [47] op het Cu(II)-ammoniak systeem geven aan dat de log β 's op slechts twee cijfers na de komma betrouwbaar zijn.

De variantie van stabiliteitsconstanten kan bepaald worden door de som van verschillende varianties [43], [48]:

$$\sigma_{pqr}^2 = \sigma_i^2 + \sigma_{tit}^2 + \sigma_{lab}^2 + \sum$$
 covarianties

waarbij:

σ_{pqr}^2	=	de werkelijke variantie op de constante
σ_i^2	-	de variantie binnen een titratie (intratitratie)
σ_{tit}^2	=	de variantie tussen de titratie (intertitratie)
$\sigma_{1,1}^2$	-	de variantie gemeten tussen de verschillende laboratoria

Braibanti [43] bewees aan de hand van de gegevens van het Nikkel-Glycine-project dat $\sigma_i^2 \ll \sigma_{tit}^2$ en $\sigma_{lab}^2 \ll \sigma_{pqr}^2 \approx \sigma_{tit}^2$. Binnen de zeven labo's die aan het project deelnamen is $\sigma_{lab}^2 \approx 0$. Dit resultaat is niet zo verrassend gezien de experimentele omstandigheden op voorhand nauwkeurig afgesproken werden.

De systematische fouten tussen de verschillende titraties moeten beschouwd worden als een willekeurig verdeelde populatie. Om σ_{tit}^2 te bepalen stelde Braibanti onderstaande analyse voor. Men behandelt eerst alle titratiecurven samen en berekent de constanten en hun standaardafwijkingen. Vervolgens analyseert men iedere titratiecurve afzonderlijk en berekent men de stabiliteitsconstanten en hun standaardafwijking voor de individuele curve. Achteraf berekent men het gemiddelde en de standaardafwijking op de individueel bepaalde stabiliteitsconstanten. De alzo bekomen afwijking σ_{gmd} is een realistische weergave van de fout op de bepaalde constante.

Tegenover deze werkwijze staat de bedenking van Casassas et al. [47] dat op deze manier een gemiddelde bepaald wordt van stabiliteitsconstanten van deeltjes die in sommige titratiecurven dominant aanwezig kunnen zijn, terwijl ze in andere curven slecht gedefinieerd kunnen zijn (laag procentueel voorkomen). Dit is zeker het geval indien de individuele titratiecurven uit verschillende metaal/ligand verhoudingen bestaan en wanneer men de vorming van polynucleaire complexen verwacht. Bij het uitrekenen

4.3 Interpretatie van de resultaten

van de individuele titratiecurven kan het zelfs voorkomen dat sommige constanten verworpen worden of slechte waarden opleveren, terwijl ze in een andere titratie tot het beste model zouden behoren. Een andere optie om realistischer standaardafwijkingen te berekenen werd o.a. door Cabaniss [49] uitgewerkt via Monte Carlo simulaties. May et al. [50] gebruikte in één van de ESTA-subroutines [30] eveneens de Monte Carlo techniek om gegevens afkomstig van het 'Nickel-Glycine project' [46] te testen. Het blijkt dat met deze techniek standaardafwijkingen van de orde 10^{-2} gevonden worden. May kon met de gegevens zelfs aantonen dat de resultaten van één van de deelnemende laboratoria systematische afwijkingen vertoonden in de standaardisatie van de reagentia.

Desondanks zijn in de literatuur weinig voorbeelden te vinden van het gebruik van de Monte Carlo techniek bij foutenpropagatiestudies in evenwichtsberekeningen.

Het tweede probleem doet zich voor bij de analyse van de gewogen residuelen. Zelfs bij perfect gekende complexeringssystemen vertonen deze residuelen in de meerderheid van de gevallen geen normaalverdeling [18], [22], [30], [47]. Bij eenvoudige systemen, zoals de bepaling van zuurconstanten, kan Avdeef [51] de gewichtenprocedure zo aanpassen dat men toch een normaalverdeling krijgt, maar voor de meerderheid van de systemen zijn dergelijke aanpassingen niet uitgewerkt. Uit een studie van de grafische voorstelling van de gewogen residuelen blijkt dat de systematische fouten voornamelijk optreden bij de begin- en de eindpunten van de titraties en bij grote potentiaalsprongen (slecht gebufferde systemen). De aanwezigheid van systematische fouten in de titratiegegevens is de hoofdoorzaak van dergelijke afwijkingen van de normaalverdeling. Mogelijke foutenbronnen zijn [47]:

- fouten tijdens de ijking van de elektrode;
- tijdsafhankelijke fenomenen bij de elektroden tijdens en tussen de titraties;
- fouten tijdens de bereiding, de standaardisatie en de verdamping van de stockoplossingen en reagentia;
- veranderingen in de ionensterkte waardoor de activiteiten van de species in de oplossing beïnvloed kunnen worden;
- afwijking in de selectiviteit van de elektroden;
- verkeerde berekening van de gewichten.

De oorsprong van systematische tendensen blijft speculatief, omdat het zeer moeilijk is ze op te sporen [18].

De keuze van geschikte meetpunten is eveneens van primordiaal belang. Gans et al. [18], [19] raden aan enkel die gegevens te gebruiken waar significante complexering merkbaar is. We kunnen bijvoorbeeld enkel datapunten gebruiken waar minstens 3% van de totale metaalionconcentratie gecomplexeerd is.

De meeste onderzoekers vermoeden dat potentiometrische gegevens niet normaal verdeeld zijn [8], [17], [25], [52]. Daarom wordt er in dit hoofdstuk voorlopig geen belang gehecht aan de Pearson-test die in SUPERQUAD voorhanden is.

Een voorbeeld waarin beide hogervernoemde problemen aan bod komen wordt beschreven in de literatuur. We verwijzen in dit verband naar het oorspronkelijk artikel van Gans et al. [18] waar de stabiliteitsconstanten van Ni(II) met glycine bepaald werden aan de hand van twee pH-titraties.

De resultaten worden weergegeven in tabel 4.6.

Tabel 4.6 De resultaten van Gans et al. [18] van complexeringstitraties van Ni(II) met glycine uitgerekend met SUPERQUAD

curve	$\log \beta_{NiL}$	$\log \beta_{NiL_2}$	$\log \beta_{NiL_3}$	S	χ^2 -test	
1	5.627(1)	10.434(1)	13.992(1)	1,4	8.8	
2	5.634(2)	10.400(2)	13.899(3)	2.1	12.6	
$(1+2)^{a}$	5.630(5)	10.422(4)	13.963(4)	7.2	32.6	
$(1+2)^{b}$	5.629(5)	10.40(1)	13.85(4)	2.1	16.9	
$(1+2)^{c}$	5.630(5)	10.42(2)	13.94(7)	-	-	
Literatuur [43]	5.63(1)	10.40(1)	13.91(2)	-		

Eerst werden de gegevens van de twee titratiecurven afzonderlijk uitgerekend. Opmerkelijk hierbij zijn de lage S- en χ^2 -waarden. Vervolgens werden beide curven samen uitgerekend $(1+2)^a$. Hier worden S- en χ^2 -waarden gevonden die zelfs groter zijn dan de som van de individuele curven. Indien we de S-regel ($S \leq 3$) van Gans zouden aannemen, is de fit onvoldoende. Pas bij het tegelijk optimaliseren van de 'dangerous' sterke-base-parameter vonden Gans et al. [18] een aanvaardbare fit met nog tendensen in de residuelen $(1+2)^b$. Deze sterke-base-parameter optimalisatie is enigszins te ver-

antwoorden daar het model op voorhand gekend is. In curve $(1+2)^c$ werd de Braibantianalyse [46] uitgerekend om te kunnen vergelijken met curve $(1+2)^a$ en curve $(1+2)^b$.

Ondanks de bovengenoemde beperkingen is SUPERQUAD momenteel één van de meest gebruikte optimalisatieprogramma's, zowel voor de berekening van zuurconstanten als voor de bepaling van stabiliteitsconstanten [53], [54], [55].

Referenties

- [1] L.G. Sillén, Acta Chem. Scand., 16, 159, (1962)
- [2] N. Ingri, L.G. Sillén, Acta Chem. Scand., 16, 173, (1962)
- [3] L.G. Sillén, Acta Chem. Scand., 18, 1085, (1964)
- [4] N. Ingri, W. Kakolowicz, L.G. Sillén, B. Warnquist, Talanta, 14, 1261, (1967)
- [5] F.J.C. Rossotti, H. Rossotti, R.J. Whewell, J. Inorg. Nucl. Chem., 33, 2051, (1971)
- [6] F. Gaizer, Coord. Chem. Rev., 27, 195, (1979)
- [7] P. Gans, Adv. Molec. Relax. Inter. Proc., 18, 139, (1980)
- [8] A. Sabatini, A. Vacca, P. Gans, Coord. Chem. Rev., 120, 389, (1992)
- M. Meloun, J. Havel, E. Högfeldt, Computation of Solution Equilibria: A guide to Methods in Potentiometry, Extraction and Spectrophotometry, Ellis Horwood Ltd., Chichester, (1988)
- [10] F.R. Hartley, C. Burgess, Solution Equilibria, Ellis Horwood Ltd., Chichester, (1980)
- [11] D.J. Legett, Computational Methods for the Determination of Formation Constants, Plenum Press, New York, (1985)
- [12] A. Sabatini, A. Vacca, J. Chem. Soc. Dalton Trans., 1693, (1972)
- [13] P. Gans, A. Vacca, Talanta, 21, 45, (1974)
- [14] A. Sabatini, A. Vacca, P. Gans, Talanta, 21, 5, (1974)
- [15] P. Gans, A. Sabatini, A. Vacca, Inorg. Chim. Acta, 18, 237, (1976)
- [16] A. Vacca, A. Sabatini, in D.J. Leggett, Computational Methods for the Determination of Formation Constants, Chapter 5, Plenum Press, New York, (1985)

- [17] P. Gans, A. Sabatini, A. Vacca, Inorg. Chim. Acta, 79, 219, (1983)
- [18] P. Gans, A. Sabatini, A. Vacca, J. Chem. Soc. Dalton Trans., 1195, (1985)
- [19] P. Gans, Manual SUPERQUAD, (1989)
- [20] P. Gans, A. Sabatini, A. Vacca, Talanta, 43, 1739, (1996)
- [21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in Fortran, Second Edition, Cambridge University Press, Cambridge, (1992)
- [22] P. Gans, Data Fitting in the Chemical Sciences, J. Wiley, Chichester, (1992)
- [23] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Third Edition, Prentice-Hall International, New York, (1992)
- [24] I. Nagypál, I. Paka, Talanta, 25, 549, (1978)
- [25] W.E. Deming, Statistical Adjustment of Data, J. Wiley, New York, (1944)
- [26] E. Casassas, R. Tauler, M. Filella, Anal. Chim. Acta, 191, 399, (1986)
- [27] L. Bologni, A. Sabatini, A. Vacca, Inorg. Chim. Acta, 69, 7, (1983)
- [28] P. Djurdjevic, D. Radanovic, Talanta, 38, 455, (1991)
- [29] P. Decock, B. Sakar, Can. J. Chem., 65, 2804, (1987)
- [30] P.M. May, K. Murray, D.R. Williams, Talanta, 35, 825, (1988)
- [31] MATLAB, Reference Guide, the MATHWORKS Inc., Natick, (1995)
- [32] MATLAB, Statistics Toolbox User's Guide, the MATHWORKS Inc., Natick, (1996)
- [33] A. Biran, M. Breiner, MATLAB for Engineers, Addison-Wesley, Wokingham, (1995)
- [34] I. Ting-Po, G.H. Nancollas, Anal. Chem., 44, 1940, (1972)

- [35] L. Lomozik, M. Jaskólski, A. Wojciechowska, Polish J. Chem., 65, 1797, (1991)
- [36] L. Lomozik, M. Jaskólski, A. Gasowska, J. Chem. Educ., 72, 27, (1995)
- [37] L. Lomozik, Chem. Anal. (Warsaw), 40, 953, (1995)
- [38] R.G. Bates, Determination of pH, Theory and Practice, Wiley, New York, (1964)
- [39] G. Anderegg, Acta Chem. Scand., A33, 74, (1979)
- [40] Y.H. Lee, Acta Chem. Scand., A33, 474, (1979)
- [41] G. Anderegg, Acta Chem. Scand., A33, 474, (1980)
- [42] A. Avdeef, D.L. Kearney, J.A. Brown, A.R. Chemotti, Anal. Chem., 54, 2322, (1982)
- [43] A. Braibanti, F. Dallavalle, G. Mori, B. Veroni, Talanta, 29, 725, (1982)
- [44] A. Braibanti, F. Dallavalle, G. Mori, M. Pasquali, Gass. Chim. Ital., 113, 407, (1983)
- [45] A. Braibanti, G. Ostacoli, P. Paoletti, L.D. Petit, S. Sammartano, Pure & Appl. Chem., 59, 1721, (1987)
- [46] E. Bottari, A. Braibanti, L. Ciavatta, A.M. Corrie, P.G. Daniele, F. Dallavalle, M. Grimaldi, A. Mastroianni, G. Mori, G. Ostacoli, P. Paoletti, E. Rizzarelli, S. Sammartano, C. Severini, A. Vacca, D.R. Williams, Anal. Chim. (Rome), 68, 813-885, (1978)
- [47] E. Casassas, R. Tauler, J. Chim. Phys., 82, 1067, (1985)
- [48] M. Meloun, M. Bartos, Talanta, 108, 227, (1992)
- [49] S.E. Cabaniss, Anal. Chem., 69, 3658, (1997)
- [50] P.M. May, K. Murray, Talanta, 35, 933, (1988)
- [51] A. Avdeef, Anal. Chim. Acta, 148, 237, (1983)

- [52] L.C. Van Poucke, Talanta, 23, 161, (1976)
- [53] H.S. Dunsmore, L.G. Sillén, Acta Chem. Scand., 17, 2657, (1963)
- [54] D.M. Wambeke, W. Lippens, Polyhedron, 11, 1305, (1992)
- [55] L.D. Petit, I. Steel, J. Chem. Soc. Dalton Trans., 1201, (1985)



Hoofdstuk 5

De verhoudingsmethode

5.1 Berekeningen met het standaard SUPERQUAD programma

5.1.1 Binaire systemen

5.1.1.1 De zuurconstanten van 1,4-diaminobutaan en 1,5-diaminopentaan: resultaten en discussie

In het eerste gedeelte van dit hoofdstuk gebeurt de bepaling van de constanten van de protonerings- en Ag(I)-complexeringsevenwichten van 1,4-diaminobutaan en 1,5diaminopentaan in 1.0 M KNO_3 via een analyse met de standaard fortran versie van SUPERQUAD. Hierbij gaan we tevens na wat het effect is van de lengte van de koolstofketen in de alifatische diaminen op de complexering [1]. De complexvorming van Ag(I) met lange-alifatische-keten diamines, $H_2N - (CH_2)_n - NH_2$, is nauwelijks onderzocht. In de literatuur worden slechts een aantal indicatieve waarden voor de vormingsconstanten vermeld [2], [3], [4], [5], [6], [7]. Recent echter is men meer en meer geïnteresseerd in het complexeringsgedrag van biogene amines zoals 1,4-diaminobutaan, welke voorkomen in de cellen van levende organismen en een rol spelen in de genetische informatieoverdracht [8]. In de chemotherapie is het geweten dat deze amines van cruciaal belang zijn als liganden in dinucleaire platinacomplexen, waarbij een ander spectrum van antitumor activiteit wordt vastgesteld dan bij cis-platina [9].

Alvorens de stabiliteitsconstanten van Ag(I) met de hogervermelde diamines bepaald kunnen worden dienen we eerst over betrouwbare waarden voor de zuurconstanten te beschikken, zodat deze in de berekening van de stabiliteitsconstanten niet mee geoptimaliseerd dienen te worden. Naderhand, in paragraaf 5.2, worden de geoptimaliseerde parameterwaarden vergeleken met deze bekomen via berekeningen met de MatLabTM implementatie van SUPERQUAD.

Voor een alifatisch diamine worden de volgende protoneringsevenwichten beschouwd in waterige oplossing:

$$L + H_2 O \rightleftharpoons H L^+ + O H^- \tag{5.1}$$

of vereenvoudigd:

$$L + H^+ \rightleftharpoons HL^+ \tag{5.2}$$

met L het overeenkomstige diamine 1,4-diaminobutaan (BN) of 1,5-diaminopentaan (PN).

 H_2L^{2+} wordt gevormd door stapsgewijze protonatie van HL^+ :

$$HL^+ + H^+ \leftrightarrows H_2 L^{2+} \tag{5.3}$$

of door globale protonatie van L:

$$L + 2H^+ \leftrightarrows H_2 L^{2+} \tag{5.4}$$

Met elk van deze evenwichten komt een vormingsconstante overeen. Voor het eerste protoneringsevenwicht wordt dit:

$$K_{011} = \beta_{011} = \frac{[HL^+]}{[L][H^+]}$$
(5.5)

zodat

$$K_{011} = \frac{1}{K_{A_2}} \tag{5.6}$$

Voor het evenwicht (5.3) is de stapsgewijze vormingsconstante gedefinieerd als:

$$K_{012} = \frac{[H_2 L^{2+}]}{[HL^+][H^+]} \tag{5.7}$$

zodat

$$K_{012} = \frac{1}{K_{A_1}} \tag{5.8}$$

De globale vormingsconstante voor reactie (5.4) wordt gedefinieerd als:

$$\beta_{012} = \frac{[H_2 L^{2+}]}{[L][H^+]^2} \tag{5.9}$$

De relatie tussen K_{012} , K_{011} en β_{012} kan geschreven worden als:

$$\beta_{012} = K_{011} K_{012} \tag{5.10}$$

Tenslotte wordt de autodissociatie van water beschreven door:

$$2H_2O \rightleftharpoons H_3O^+ + OH^- \tag{5.11}$$

met

$$K_W = [H_30^+][OH^-] \tag{5.12}$$

De respectievelijke vormingsconstanten β_{011} en β_{012} worden berekend aan de hand van de *pH*-gegevens via de gewogen optie in SUPERQUAD [10], [11], [12], [13] met $\sigma_{E_{glaselektrode}} = 0.2 \text{ mV}$ en $\sigma_{V} = 0.002 \text{ ml}$ (bepaald in paragraaf 3.2.5.1 en 3.2.5.2).

We opteren om de literatuurwaarden vermeld in tabel 5.1 (in vergelijkbare omstandigheden) te gebruiken als startwaarden voor de optimalisering van de protoneringsconstanten van BN. Vergelijkbare pK_A -waarden voor PN in 1 M nitraatmilieu worden in de literatuur niet gerapporteerd. Daarom worden de startwaarden in dit geval geschat. Het blijkt immers dat zelfs met zeer slechte startwaarden (afwijkingen 4 log-eenheden) hetzelfde minimum wordt bereikt. Tijdens de berekeningen, uitgevoerd op een mainframe DEC Alpha 3000/800X/AXP computer, worden de massabalansen, de elektrodeparameters en de pK_W (bepaald onder 3.2.5.2) constant gehouden.

Merken we op dat deze zuurconstanten eveneens kunnen berekend worden met de lineaire kleinste kwadraten methode via de volgende ionenbalans en massabalansen:

$$2(H_2L^{2+}) + (HL^+) + (H^+) + (K^+) = (OH^-) + (NO_3^-)$$

$$C_L = (L) + (HL^+) + (H_2L^{2+})$$

$$C_B = (OH^-) - (H^+) + 2(L) + (HL^+)$$

met C_L de totale concentratie aan ligand en C_B de totale concentratie sterke base. Hieruit kan men een lineaire vergelijking in de constanten halen, waarbij de coëfficiënten functies zijn van (H^+) .

Tabel 5.1 Literatuurwaarden voor de zuurconstanten (in log eenheden) van 1,4-diaminobutaan (BN) en 1,5-diaminopentaan (PN) refererend naar vgl.(5.6) en vgl.(5.8)

BN	ref.[2]	ref.[3]	ref.[4]	ref.[5]	ref. [6]	ref.[14]	ref.[15]	ref.[16]
temperatuur (°C)	25	25	25	20	20	25	25	25
ionensterkte	$1M NO_3^-$	3M LiClO ₄	$1M NO_3^-$?	$0.1 M NO_3^-$?	$0.1 \mathrm{M} \ \mathrm{NO}_3^-$	$0.5M NO_3^-$
$\log K_{A_1}$	-	10.39	9.91 ± 0.03	9.35 ± 0.01	9.61 ± 0.01	9.20 ± 0.02	9.842	9.634 ± 0.003
			$9.44 \pm 0.05 \ (0.1 \mathrm{M})$					
			$9.69 \pm 0.06 \ (0.5 M)$					
$\log K_{A_2}$	11.08 ± 3	11.05	11.11 ± 0.04	10.80 ± 0.06	10.82 ± 0.02	10.65 ± 0.02	11.063	10.804 ± 0.002
			$10.72 \pm 0.07 \ (0.1 \mathrm{M})$					
			$10.87 \pm 0.07 \ (0.5M)$					
		2						
PN	ref.[2]	ref.[6]	ref.[15]	ref.[16]				
temperatuur (°C)	25	20	25	25				
ionensterkte	$0.1 M NO_3^-$	$0.1 M NO_3^-$	$0.1 \mathrm{M} \mathrm{NO}_3^-$	$0.5M \text{ NO}_3^-$				
$\log K_{A_1}$	9.85	10.02 ± 0.02	10.195	10.050 ± 0.008				
	10.15 (0.5M)							
$\log K_{A_2}$	10.78	10.96 ± 0.02	11.200	10.916 ± 0.004				
	10.92 (0.5M)							

130

Hoofdstuk 5

De verhoudingsmethode

Tabel 5.2 De protonerings- en zuurconstanten van 1,4-diaminobutaan en 1,5-diaminopentaan (in log eenheden) bepaald in dit werk via de gewogen optie in Superquad aangevuld met literatuurwaarden in vergelijkbare omstandigheden voor enkele andere diamines. De standaardafwijkingen en de S-statistiek worden aangegeven

	$\log\beta_{012}$	$\log \beta_{011}$	$\log K_{A_1}$	$\log K_{A_2}$	$\log K_{A_2}/K_{A_1}$ $(\Delta p K_A)$	S
$ethyleendiamine^{a}$	$17.660\ \pm < 0.001$	$10.170\ \pm < 0.001$	7.490 ± 0.001	10.170 ± 0.001	2.680	1.40
1, 3-propyleendiamine ^a	$19.827 \pm < 0.001$	$10.736\ \pm < 0.001$	9.091 ± 0.001	10.736 ± 0.001	1.645	0.20
1,4-diaminobutaan ^a	20.720 ± 0.002	10.933 ± 0.002	9.787 ± 0.002	10.933 ± 0.002	1.146	1.42
1,5-diaminopentaan ^a	21.483 ± 0.003	11.221 ± 0.004	10.262 ± 0.003	11.221 ± 0.004	0.959	1.65
1,6-diaminohexaan ^{b,c}	21.677	11.241	10.436	11.241	0.805	-
1,7-diaminoheptaan ^{b,c}	21.807	11.272	10.535	11.272	0.737	-
1,8-diaminooctaan ^{b,c}	21.899	11.303	10.596	11.303	0.707	-
1,9-diaminononaan ^{b,c}	21.909	11.294	10.615	11.294	0.679	-
1,10-diaminodecaan ^{b,c}	21.940	11.299	10.641	11.299	0.658	-

^a 1 M NO₃⁻

^b 0.1 M NO₃⁻

^c st.afw. niet vermeld in de literatuur
In tabel 5.2 zijn de protoneringsconstanten en de omgerekende pK_A -waarden voor BN en PN bepaald in dit werk (1.0 M KNO_3 bij 25 °C) samengevat. De tabel wordt aangevuld met literatuurwaarden in vergelijkbare omstandigheden voor ethyleendiamine, 1,3-propyleendiamine [17]; vanaf 6 methyleengroepen werden de pK_A -literatuurwaarden bepaald in 0.1 M NO_3^- [15].

Een vergelijking met de literatuurwaarden in identieke omstandigheden voor BN (1 M NO_3^-) (tabel 5.1) met deze in tabel 5.2 geeft een verschil aan van 0.1 pK_{A_1} -eenheden. Voor pK_{A_2} bedraagt het verschil zelfs 0.2 pK_A -eenheden. Men moet zich echter realiseren dat zulke vergelijkingen slechts opgaan indien de werkwijze – vooral het ijken van de glaselektrode – analoog is. Uit tabel 5.1 blijkt eveneens dat zwakkere ionaire middens bij 25 °C systematisch een lagere pK_A -waarde (zowel pK_{A_1} als pK_{A_2}) opleveren dan deze bepaald in dit werk, hetgeen verwacht kan worden [18].

Opvallend voor PN in tabel 5.2 is de relatief hogere waarde van de standaarddeviatie voor zowel pK_{A_1} als pK_{A_2} . Dit is waarschijnlijk te wijten aan occlusie van onzuiverheden in het dinitraatzout tijdens het kristallisatieproces [1], [19].

Tabel 5.2 toont tevens aan dat de constanten bepaald in dit werk mooi passen in de tendens van de literatuurwaarden [15], m.a.w. de pK_{A_1} neemt toe, maar naarmate de ketenlengte n in het molecule $H_2N - (CH_2)_n - NH_2$ stijgt wordt het increment kleiner. De pK_{A_2} -waarden vertonen dit nog duidelijker.

Volgens Clark en Perrin [20] heeft het toevoegen van een methyleengroep in monoamines een verwaarloosbaar effect op de pK-waarden, zodat het verhogen van de pK-waarden in diamines te wijten is aan de aanwezigheid van de andere $-NH_3^+$ of $-NH_2$ groep in het molecule [21].

De $\Delta p K_A$ -waarde (= $p K_{A_2} - p K_{A_1}$) kan beschouwd worden als een maat voor de elektrostatische interacties tussen de terminale ammoniumgroepen. Figuur 5.1 toont duidelijk dat $\Delta p K_A$ snel afneemt met stijgend aantal methyleengroepen en ongeveer constant blijft bij n = 8. Dit kan verklaard worden doordat de afstand tussen de twee positieve centra in het diamine toeneemt bij stijgende n hetgeen resulteert in een zwakkere repulsie.



Figuur 5.1: ΔpK_A in functie van het aantal methyleengroepen

5.1.1.2 De hydrolyseconstanten van Ag(I): resultaten en discussie

De potentiaal/volume data afkomstig van de zilverhydrolyse titraties worden eveneens geanalyseerd gebruikmakend van de gewichtenoptie in SUPERQUAD (($\sigma_{E_{glasslektrode}} = 0.2 \text{ mV}, \sigma_{E_{Ag_2S}} = 0.1 \text{ mV}$ en $\sigma_V = 0.002 \text{ ml}$). Hierbij worden titratiepunten waar het ionenproduct $[Ag^+][OH^-] > 10^{-7.3}$ niet in rekening gebracht (zie 3.2.6.1). De eerste titratiepunten (in zuur milieu) kunnen bovendien gebruikt worden voor de interne ijking van de zilverelektrode. Een reeks modellen met o.a. de deeltjes AgOH, $Ag(OH)_2^-$, $Ag(OH)_3^{2-}$, Ag_2OH^+ en $Ag_2(OH)_2$ werden getest. Er werd geopteerd om literatuurwaarden te gebruiken als startwaarden voor verdere optimalisatie van de constanten [17]. Een volledige opsomming van alle geteste modellen lijkt ons hier niet relevant. In tabel 5.3 worden enkel deze verfijnde modellen vermeld die het best beantwoorden aan de in volgende paragraaf vernoemde modelselectiecriteria, daar deze criteria uiteraard ook gelden voor de complexeringstitraties.

De hydrolyse
constanten $\beta_{p0-\tau}$ en $\beta*_{p0-\tau}$ verwijzen respectie
velijk naar volgende reacties:

$$pAg^+ + rH_2O \rightleftharpoons Ag_p(OH)_r^{(p-r)} + rH^+$$

$$(5.13)$$

en

$$pAg^{+} + rOH^{-} \rightleftharpoons Ag_{p}(OH)_{r}^{(p-r)}$$

$$(5.14)$$

met

$$\beta_{p0-r} = \beta *_{p0-r} (\beta_{00-1})^r \tag{5.15}$$

Tabel 5.3 De geoptimaliseerde constanten β_{10-1} , β_{10-2} en de berekende constanten β_{*10-1} , β_{*10-2} (in log-eenheden) bekomen met SUPERQUAD via de gewichtenoptie voor het Ag - OH systeem in 1.0 M KNO₃. Het maximaal % voorkomen van de hydrolysedeeltjes, de standaardafwijking en de S-waarde worden vermeld

Deeltje	model 1	max.% voork.	model 2	max.% voork.
AgOH	${\rm log}\beta_{10-1}=-11.95\pm0.01$	36	$\log\!\beta_{10-1} = -12.03 \pm 0.01$	25
	${\rm log}\beta*_{10-1}=1.83\pm0.01$		${\rm log}\beta*_{10-1} = 1.75 \pm 0.01$	
	S = 2.99			
$Ag(OH)_2^-$			$\mathrm{log}\beta_{10-2} = -24.02 \pm 0.04$	9
			${\rm log}\beta*_{10-2}=3.53\pm0.04$	
			S= 2.70	

De verfijnde constanten van model 2 zijn vergelijkbaar met literatuurwaarden voor verdunde oplossingen: $\log \beta_{10-1} = -12.0$ en $\log \beta_{10-2} = -24.0$ [22], [23], [24]. Deze constanten zullen dan ook gebruikt worden om te corrigeren voor de zilverhydrolyse die kan plaatsgrijpen tijdens de complexering van de alifatische diaminen met Ag(I).

5.1.2 Ternaire systemen

5.1.2.1 De stabiliteits constanten van de Ag(I)-complexen van 1,4 diaminobutaan en 1,5 diaminopentaan: resultaten en discussie

De gecombineerde pH/pAg data (760 geanalyseerde titratiepunten, 1500 punten in totaal) worden geanalyseerd met SUPERQUAD, waarbij verschillende complexeringsmodellen worden getest met de gewogen optie ($\sigma_{E_{glaselektrode}} = 0.2 \text{ mV}, \sigma_{E_{Ag_2S}} = 0.1 \text{ mV}$ en $\sigma_V = 0.002 \text{ ml}$). Volgende deeltjes worden beschouwd als mogelijk aanwezige complexen: AgL, AgLH, AgL_2 , AgL_2H , AgL_2H_2 , Ag_2L , Ag_2L_2 , Ag_2L_2H , Ag_2L_3 , AgLOHen AgL_2OH . Het uiteindelijke model wordt geselecteerd op basis van volgende criteria (zie eveneens hoofdstuk 4):

- de berekende fit-statistiek S moet tussen 1 en 3 liggen;
- de berekende standaarddeviatie van een constante mag niet groter zijn dan 0.1 log-eenheden (arbitrair);
- het deeltje moet voor minimaal 3% voorkomen (arbitrair) wil men enige zekerheid hebben over zijn aanwezigheid in de beschouwde omstandigheden;
- het aantal significante titratiepunten is minimaal 5% (arbitrair) van het totaal;
- toevoeging van een nieuw deeltje aan het model moet een beduidende verlaging van het minimum tot gevolg hebben.

Als startwaarden worden de stabiliteitsconstanten van Ag(I) met ethyleendiamine en 1,3-propyleendiamine uit de literatuur gebruikt [17], [25], [26], [27], [28].

De protoneringsconstanten, de pK_W , de massabalansen en de elektrodeparameters worden constant gehouden tijdens de optimalisatie.

De beste modellen voor de complexering van Ag(I) met BN en PN zijn weergegeven in tabel 5.4. Volledigheidshalve worden ook de stabiliteitsconstanten van Ag(I) met ethyleendiamine (EN) en 1,3-propyleendiamine (TN) [17] vermeld.

Figuren 5.2 a-b-c tonen een typisch distributiediagram in functie van respectievelijk het ganse pH-, pAg- en pL-gebied voor het systeem Ag(I) - PN. De verhouding C_{Ag}/C_{PN} is gelijk aan 1/4.

Tabel 5.4 Een overzicht van de globale stabiliteitsconstanten (in log eenheden), de standaardafwijking en het maximaal % voorkomen over het ganse pH-gebied en voor alle C_{Ag}/C_L-verhoudingen van de complexen van ethyleendiamine (EN), 1,3-propyleendiamine (TN),
1,4-diaminobutaan (BN) en 1,5-diaminopentaan (PN) met Ag(I) berekend via de gewogen optie in SUPERQUAD. De ladingen zijn weggelaten voor de eenvoud

Deeltje	EN	max. %	TN	max. %	BN	max. %	PN	max. %
		voorkomen		voorkomen		voorkomen		voorkomen
AgL					5.45 ± 0.04	29	$6.22 \pm < 0.01$	84
AgL_2	$7.67 \pm < 0.01$	45	7.81 ± 0.01	16			8.33 ± 0.09	9
Ag_2L					7.24 ± 0.09	4		
Ag_2L_2	$13.41 \pm < 0.01$	93	$14.86 \pm < 0.01$	98	14.45 ± 0.02	35	15.05 ± 0.01	48
AgLH	$12.63 \pm < 0.01$	24	$13.75 \pm < 0.01$	20	14.20 ± 0.02	19	14.55 ± 0.04	14
AgL_2H	16.59 ± 0.05	4					20.08 ± 0.02	33
AgL_2H_2	$25.39 \pm < 0.01$	16	$27.72 \pm < 0.01$	20	28.36 ± 0.04	10	30.02 ± 0.02	22
Ag ₂ L ₂ H	20.20 ± 0.02	4					23.58 ± 0.05	4
S-statistiek	2.88		2.07		2.33		2.41	

De verhoudingsmethode



Figuur 5.2 a-b: Deeltjesdistributie i.f.v. pH(a) en pAg(b) van Ag(I)-PN-complexen voor $C_{Ag} = 0.00125 \text{ M}$ en $C_L = 0.00500 \text{ M}$. Het totaal beginvolume bedraagt 40 ml



Figuur 5.2 c: Deeltjesdistributie i.f.v. pL(c) van Ag(I)-PN-complexen voor $C_{Ag} = 0.00125 M$ en $C_L = 0.00500 M$. Het totaal beginvolume bedraagt 40 ml

Om de discussie te verduidelijken definiëren we de volgende evenwichten met hun respectievelijke vormingsconstante:

A

$$Ag^+ + LH^+ \rightleftharpoons AgLH^{2+} \qquad K_{111} = \beta_{111}/\beta_{011}$$
 (5.16)

$$gLH^{2+} + LH^+ \rightleftharpoons AgL_2H_2^{3+} \qquad K_{122} = \beta_{122}/\beta_{111}.\beta_{011} \qquad (5.17)$$

$$AgL_2H_2^{3+} \rightleftharpoons AgL_2^+ + 2H^+ \qquad K_{120} = \beta_{120}/\beta_{122}$$
 (5.18)

$$AgL_2H_2^{3+} + Ag^+ + 2L \rightleftharpoons Ag_2L_2^{2+} + 2LH^+ \qquad K_{220} = \beta_{220}.\beta_{011}^2/\beta_{122} \qquad (5.19)$$

$$2AqL^+ \rightleftharpoons Aq_2L_2^{2+} \qquad K_D = \beta_{220}/\beta_{110}^2$$
 (5.20)

In feite is deze reeks vergelijkingen ook een geldigheidstest. Mochten hier anomalieën optreden dan kunnen grote twijfels ontstaan over het model. Anderzijds wanneer alles "normaal" is heeft men een gerust gevoel.

In tabel 5.5 zijn deze vormingsconstanten uitgerekend voor BN en PN. De waarden

voor EN en TN zijn literatuurwaarden [17] in vergelijkbare omstandigheden.

Tabel 5.5 De berekende vormingsconstanten (in log-eenheden) corresponderend met de evenwichten 5.16 tot 5.20 voor verschillende Ag(I)-diamine systemen

	EN	TN	BN	PN
K_{111}	2.46	3.01	3.28	3.33
K_{122}	2.59	3.23	3.23	4.25
K_{120}	-17.62	-19.80	-	-21.69
K_{220}	8.36	8.61	7.95	7.47
K_D	-	-	3.55	2.61

Uitgaande van de resultaten vermeld in tabel 5.4 en de modelselectiecriteria in acht genomen kunnen we stellen dat de meest betrouwbare stabiliteitsconstanten deze voor de deeltjes Ag_2L_2 , AgLH en AgL_2H_2 zijn.¹ De beschrijving van de complexen wordt bovendien bemoeilijkt door het feit dat een alifatisch diamine steeds twee posities bezit waarmee het Ag(I)-ion kan complexeren. Voor symmetrische liganden zoals BN en PN zijn deze posities gelijkwaardig.

De niet geprotoneerde mononucleaire en polynucleaire complexen: AgL, AgL_2 en Ag_2L_2

Het valt op dat in tegenstelling tot EN en TN, het AgL complex wel gedetecteerd wordt voor BN en PN (tabel 5.4). De respectievelijke constanten 5.45 \pm 0.04 en $6.22 \pm < 0.01$ zijn relatief hoog in vergelijking met constanten voor Ag(I)-complexen met monoamines [17]. Mogelijk treedt hier chelaatvorming op (8-ring). Toch gaat men uit van de veronderstelling dat de N - Ag - N het meest stabiel is als ze lineair is. Van de andere kant geeft de dimerisatieconstante, K_D (5.20) informatie over het simultaan voorkomen van de deeltjes AgL en Ag_2L_2 . De berekende waarden (in logeenheden): log $K_{D_{(BN)}} = 3.55$ en log $K_{D_{(FN)}} = 2.61$ tonen duidelijk het overwicht van het dimeer Ag_2L_2 aan. Bovendien wordt het Ag_2L_2 complex stabieler naarmate de ketenlengte in het alifatisch diamine toeneemt (tabel 5.4). Dit kan verklaard worden indien men aanneemt dat Ag_2L_2 een cyclische structuur heeft, zodat volgens Cassol [29], [30] een 16-armige ring (in het geval van PN) meer gelijkenis vertoont met de

¹Voor de eenvoud worden de ladingen voortaan weggelaten

lineaire N - Ag - N configuratie en bijgevolg stabieler is dan een 10 (EN) -, 12 (TN) - of 14 (BN) - armige ring.

Verdere speculaties over mogelijke structuren voor Ag(I)-alifatische diaminecomplexen zijn niet het onderwerp van deze thesis. We verwijzen hiervoor naar het werk van Garner [17]. Het complex AgL_2 dat gevormd wordt met EN, TN en PN (tabel 5.4), wordt niet gedetecteerd bij BN en vanwege de hoge standaarddeviatie lijkt de waarde van de stabiliteitsconstante onbetrouwbaar in het geval van PN. Een verklaring hiervoor vertrekt van twee mogelijke hypothesen:

- de eerste hypothese verwijst naar het evenwicht (5.19) met vormingsconstante K_{220} . Het is mogelijk dat door toenemende stabiliteit van het Ag_2L_2 complex en van het zuur LH de vorming van AgL_2 uitgaande van AgL_2H_2 verhinderd wordt. In dat geval zou de waarde van K_{220} moeten toenemen in functie van de ketenlengte, hetgeen niet tot uiting komt in de data (tabel 5.5);
- een tweede mogelijke verklaring verwijst naar het evenwicht (5.18) met vormingsconstante K_{120} . Uit tabel 5.5 blijkt dat met stijgende ketenlengte het zuur AgL_2H_2 zeer zwak is bij de pH waar AgL_2 gevormd wordt. Deze pHwaarde situeert zich waarschijnlijk in het neerslaggebied zodat AgL_2 niet kan gedetecteerd worden. Deze hypothese is de meest logische.

De geprotoneerde mononucleaire complexen: AgLH en AgL₂H₂

Naast Ag_2L_2 komen deze complexen bij de 4 diaminen vermeld in tabel 5.4 voor, maar nooit dominant. Dit wordt duidelijk in figuur 5.2 waar het voorkomen van Ag(I)-PNcomplexen met C_{Ag}/C_L verhouding gelijk aan 1/4 is uitgezet in functie van de pH, pAgen pL. Zoals gesteld in de literatuur [26] kunnen deze deeltjes in feite beschouwd worden als alifatische monoamine complexen van het type AgL' en AgL'_2 , waarbij LH zich gedraagt als een monodentaat ligand L'. Tevens tonen de berekende waarden voor K_{111} en K_{122} (tabel 5.5) duidelijk aan dat met stijgende ketenlengte in het diamine AgLHen AgL_2H_2 stabieler worden. Dit kan verklaard worden door de afnemende $Ag^+ \leftrightarrow$ LH^+ repulsie die reeds door Ohtaki [3], [31] en Van Poucke [32], [33] werd gesuggereerd. Als algemene opmerking tenslotte, geldt dat voor al de onderzochte modellen de berekeningen geen evidentie geven voor het bestaan van hydroxocomplexen van Ag(I) met BN en PN. Toevoeging van zilver-ligand-hydroxy-complexen aan de geteste modellen verbetert de fit niet. Deze deeltjes worden eerder verworpen dan aanvaard.

5.2 Berekeningen met de MatLabTM implementatie van SUPERQUAD

5.2.1 Binaire systemen

In hoofdstuk 4 werd de MatLab implementatie gevalideerd aan de hand van synthetische data met en zonder perturbatie in pH en pM. Een analoge studie wordt uitgevoerd met dit verschil dat de **experimentele** pH waarden voor de binaire systemen BN-Hen PN-H als input ingebracht worden in het MatLab programma (= experimentele route organigram figuur 4.2).

De protoneringsconstanten β_{011} en β_{012} worden geoptimaliseerd vertrekkende van identieke initiële parameterwaarden als deze in het standaard fortran programma. In functie van het ingegeven model, worden de partiële afgeleiden van de totale concentratie naar de vrije concentratie van elke component berekend in de functie 'jacobian_free.m' (hoofdstuk 4) volgens de relevante vergelijkingen voor de binaire systemen (Appendix B).

De partiële afgeleiden van de totale concentratie naar de te optimaliseren protoneringsconstanten worden berekend in de functie 'jacobian_beta.m' volgens de vergelijkingen weergegeven in Appendix C (binaire systemen).

Tabel 5.6 geeft een overzicht van de protoneringsconstanten berekend volgens beide SUPERQUAD versies.

Tabel 5.6 Overzicht van de protoneringsconstanten (in log eenheden) berekend met 'SQ_simul.m' (MatLab) in vergelijking met SUPERQUAD f77 (gewogen optie) aan de hand van experimentele pH data over het ganse pH gebied voor de binaire systemen BN-H en PN-H

BN-H	f77 $(\sigma_{E_{glas}} = 0.2 \ mV, \sigma_V = 0.002 \ ml)$ geoptimaliseerde constante	$egin{array}{llllllllllllllllllllllllllllllllllll$
$\log\beta_{011}$	10.933	10.931
$\log \beta_{012}$ $PN\text{-}H$	20.720	20.720
$\log\beta_{011}$	11.221	11.220
$\log\beta_{012}$	21.483	21.482

Tabel 5.6 illustreert dat de geoptimaliseerde protoneringsconstanten in beide SUPER-QUAD versies goed overeenkomen, zodat is aangetoond dat we het MatLab programma als equivalent van de fortran versie kunnen beschouwen.

5.2.2 Ternaire systemen

Als toepassing van een ternair systeem worden de experimentele pH- en pAg-waarden voor Ag-BN-H als input ingevoerd in het MatLab programma. De stabiliteitsconstanten β_{110} , β_{210} , β_{220} , β_{111} en β_{122} worden geoptimaliseerd vertrekkende van identieke initiële parameterwaarden als deze in het standaard fortran programma. In functie van de ingegeven modelmatrix worden de partiële afgeleiden van de totale concentratie naar de vrije concentratie van elke component berekend in de functie 'jacobian_free.m' (hoofdstuk 4) volgens de relevante vergelijkingen voor het ternair systeem (Appendix B).

De partiële afgeleiden van de totale concentratie naar de te optimaliseren stabiliteitsconstanten worden berekend in de functie 'jacobian_beta.m' volgens de vergelijkingen weergegeven in Appendix C (ternair systeem). Tabel 5.7 geeft een overzicht van de globale stabiliteitsconstanten berekend volgens beide SUPERQUAD versies.

Tabel 5.7 Overzicht van de globale stabiliteitsconstanten (in log eenheden) berekend met 'SQ_simul.m' (MatLab) in vergelijking met SUPERQUAD f77 (gewogen optie) aan de hand van experimentele pH en pAg data over het ganse pH en pAg gebied en voor alle C_{Ag}/C_{BN} verhoudingen van de complexen van BN met Ag(I)

	f77 ($\sigma_{E_{glas}} = 0.2 \ mV, \ \sigma_V = 0.002 \ ml$	MatLab ($\sigma_{E_{glas}}, \sigma_V$: idem	
Ag-BN-H	$\sigma_{E_{Ag_2S}} = 0.1 \ mV)$ geoptimaliseerde constante	$\sigma_{E_{Ag_2s}}$: 0) geoptimaliseerde constante	
$\log \beta_{110}$	5.45	5.40	
$\log \beta_{210}$	7.24	7.24	
$\log\beta_{220}$	14.45	14.44	
$\log\beta_{111}$	14.20	14.18	
$\log\beta_{122}$	28.36	28.32	

Tabel 5.7 toont opnieuw een goede overeenkomst tussen beide SUPERQUAD versies. Bovendien selecteren beide versies een analoog complexeringsmodel.

Referenties

- W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, J. Coord. Chem., 47, 17, (1999)
- [2] NIST Standard Reference Database 46 Version 2.0, NIST Critically Selected Stability Constants of Metal Complexes, (1995)
- [3] H. Ohtaki and K. Cho, Bull. Chem. Soc. Japan, 50, 2674, (1977)
- [4] A.E. Martell and R.M. Smith, *Critical Stability Constants*, Volume 2 : Amines (1975) and Volume 5 : First Supplement, Plenum Press, (1982)
- [5] C.R. Bertsch, W.C. Fernelius and B.P. Block, J. Phys. Chem., 62, 444, (1958)
- [6] G. Schwarzenbach, B. Maissen und H. Ackermann, Helv. Chim. Acta, 35, 2333, (1952)
- [7] Y. Hojo, Y. Sugiura and H. Tanaka, J. Inorg. Nucl. Chem., 39, 715, (1977)
- [8] L. Lomozik and A. Wojciechowska, Polyhedron, 8, 2645, (1989)
- [9] S. Borman, Chem. Eng. News, 4, 39, (1995)
- [10] P. Gans and A. Sabatini, J. Chem. Soc. Dalton Trans., 1195, (1985)
- [11] P. Gans and A. Sabatini, Inorg. Chim. Acta, 79, 219, (1983)
- [12] P. Gans, Data Fitting in the Chemical Sciences, Wiley, Chichester, Chapter 10, (1992)
- [13] A. Sabatini, A. Vacca and P. Gans, Coord. Chem. Reviews, 120, 389, (1992)
- [14] J.J. Christensen, R.M. Izatt, D.P. Wrathall and L.D. Hansen, J. Chem. Soc., A, 1212, (1969)
- [15] J. Delrue, Potentiometrische en kalorimetrische studie van de protonering van alifatische diamines, proefschrift, R.U.G., (1972)

- [16] R. Barbucci, P. Paoletti and A. Vacca, J. Chem. Soc., A, 2202, (1970)
- [17] R. Garner, Een potentiometrische studie van het complexeringsgedrag van Ag(I) met alifatisch gesubstitueerde mono-en diaminen in 1 M nitraat, proefschrift, L.U.C., (1993)
- [18] K. Brassinne, Potentiometrische studie van de oxalatocomplexen van Cu²⁺ en Y³⁺ in KNO₃-midden, proefschrift, L.U.C., (1997)
- [19] R. Garner, J. Yperman, J. Mullens and L.C. Van Poucke, Bull. Soc. Chim. Belg., 102, 3, (1993)
- [20] J. Clark, D.D. Perrin, Quart. Rev., 16, 295, (1964)
- [21] P. Paoletti, R. Barbucci, A. Vacca, J. Chem. Soc. (A), 310, (1971)
- [22] A.E. Martell, R.J. Motekaitis, The Determination and Use of Stability Constants, VCH publishers, New York, (1988)
- [23] C.F. Baes, R.E. Mesmer, Hydrolysis of Cations, p. 124, Wiley, New York, (1976)
- [24] R. Keim, A. Bohne-Neuber, Gmelins Handbuch der Anorganischen Chemie: Silber, Teil B1, p. 122-124, Springer Verlag, New york, (1975)
- [25] R. Garner, J. Yperman, J. Mullens and L.C. Van Poucke, J. Coord. Chem., 30, 151, (1993)
- [26] R. Garner, J. Yperman, J. Mullens and L.C. Van Poucke, Inorg. Chim. Acta, 224, 97, (1994)
- [27] R. Garner, J. Yperman, J. Mullens and L.C. Van Poucke, J. Coord. Chem., 31, 117, (1994)
- [28] L.C. Van Poucke, Talanta, 23, 161, (1976)
- [29] A. Cassol, P. Di Bernardo, J. Chem. Soc. Dalton Trans., 1781, (1988)
- [30] A. Cassol, P. Di Bernardo, J. Chem. Soc. Dalton Trans. I, 2445, (1989)
- [31] H. Ohtaki and Y. Ito, J. Coord. Chem., 3, 131, (1973)

- [32] L.C. Van Poucke, Z. Eeckhaut, Bull. Soc. Chim. Belg., 81, 363, (1972)
- [33] L.C. Van Poucke, H.F. De Brabander and Z. Eeckhaut, Bull. Soc. Chim. Belg., 78, 131, (1969)



Deel III

Optimaliseren van "synthetische" potentiometrische data



Hoofdstuk 6 Monte Carlo studie van de foutenpropagatie bij de bepaling van stabiliteitsconstanten met behulp van SUPERQUAD

6.1	Inleiding					
6.2	Kwantificering aan de hand van experimentele fouten					
	6.2.1	De invloed van een affeesfout op de pH in elk titratiepunt \ldots				
	6.2.2	De invloed van perturbaties in het toegevoegd volume base per ti-				
		tratiestap en in de totale concentratie van het metaalion, het ligand				
		en het proton				
	6.2.3	De invloed van perturbaties op het initieel volume in de recipiënt .				
	6.2.4	Conclusie				
6.3	Kwantificering aan de hand van "structurele" fouten					
	6.3.1	Inleiding				
	6.3.2	Analyse van modelsysteem 'sys 1'				
		6.3.2.1 Uitbreiding van het gedefinieerde startmodel: resultaten en				
		discussie				
		6.3.2.2 Beperking van het gedefinieerde startmodel: resultaten en				
		discussie				
		6.3.2.3 Gecombineerde modellen: resultaten en discussie				
		6.3.2.4 Conclusie				
	6.3.3	Analyse van modelsysteem 'sys 2'				
		6.3.3.1 Uitbreiding van het gedefinieerde startmodel: resultaten en				
		discussie				
		6.3.3.2 Beperking van het gedefinieerde startmodel: resultaten en				
		discussie				
		6.3.3.3 Gecombineerde modellen: resultaten en discussie				
		6.3.3.4 Conclusie				
	Referenties					



Hoofdstuk 6

Monte Carlo studie van de foutenpropagatie bij de bepaling van stabiliteitsconstanten met behulp van SUPERQUAD

6.1 Inleiding

In hoofdstuk 2 werden met behulp van MatLabTM-routines experimenten gesimuleerd met als doel na te gaan binnen welke grenzen de lineaire foutenpropagatieformules geldig zijn. Op analoge wijze, maar in n – dimensies (met n het aantal te optimaliseren parameters), zal in dit hoofdstuk aan de hand van simulaties met het MatLabTM-programma 'SQ_simul.m', beschreven in hoofdstuk 4, een foutenpropagatiestudie uitgevoerd worden bij de bepaling van stabiliteitsconstanten.

In de recente literatuur vinden we slechts een aantal voorbeelden van foutenpropagatiestudies bij evenwichtsberekeningen [1], [2], [3], [4], [5], [6]. Algemeen kan uit deze studies geconcludeerd worden dat als gevolg van *experimentele* fouten (toevallig of systematisch) in de input – zelfs voor perfect gekende complexeringssystemen – de residuelen in pHof in *e.m.f.* in bepaalde gebieden van de titratie niet gaussiaans verdeeld zijn of systematische tendensen vertonen, wat mogelijk leidt tot een onbetrouwbare optimalisatie van onbekende constanten. De hoofdoorzaak van afwijkingen van de normaalverdeling bij een gekend model wordt meestal gezocht in de aanwezigheid van systematische fouten in de gegevens. Er werd voorheen, met betrekking op de modelselectie, reeds gesteld dat dergelijke experimentele tekorten niet kunnen gecompenseerd worden door de optimalisatieberekeningen. Daarenboven zijn in de meeste gevallen de berekende standaardafwijkingen op de geoptimaliseerde parameters weinig realistisch van aard, wat impliceert dat de constanten met een hogere precisie bepaald zouden zijn dan de haalbare accuratesse van de meetapparatuur. Bij de interpretatie van de resultaten bekomen met SUPERQUAD zijn bovenvernoemde problemen eveneens aan de orde (zie hoofdstuk 4). Het effect van fouten in de experimentele data op de resulterende fout in de geoptimaliseerde parameters bij complexeringsevenwichten is volgens Kropotov bovendien sterk afhankelijk van de waarde van de constante in de betrokken reactie [7], [8].

Een onderzoek van Lomozik [9], [10] beschrijft een evaluatie van de betrouwbaarheid van de computerberekeningen aan de hand van zogenaamde *structurele* fouten, m.a.w. wijzigingen in het chemisch model. Daar de studie van Lomozik kwalitatief van aard is, zal in dit hoofdstuk een kwantitatieve analyse van zowel *experimentele* als *structurele* fouten uitgevoerd worden, met als doel te komen tot een aantal alternatieve en/of meer objectieve criteria voor modelselectie.

6.2 Kwantificering aan de hand van experimentele fouten

Het is de bedoeling in deze sectie, gegeven het chemisch model, te achterhalen wat de impact is van perturbaties in één van de inputparameters op het uiteindelijke resultaat, de geoptimaliseerde stabiliteitsconstanten. Een Monte Carlo analyse moet ons toelaten de effecten van deze experimentele perturbaties, al of niet systematisch van aard, te kwantificeren: welke stappen in de experimentele keten zijn belangrijk in het genereren van de fout op de geoptimaliseerde constante en indien de grootte van de fout gekend is, wat is dan de grootte-orde van de te verwachten fout op de constante; kan dit gebruikt worden als criterium voor modelselectie? Het invoeren van ruis op het chemisch model situeert zich in fase 1 van het organigram (figuur 4.2) bij het genereren van de data en op het resultaat van fase 1, de gegenereerde vrije concentraties van de componenten.

De keuze van geschikte datapunten is bij de optimalisatie van primordiaal belang. Gans

et al. [11], [12] raden aan enkel die gegevens te gebruiken waar significante complexering merkbaar is. Men kan zodoende enkel datapunten gebruiken waar minstens 3% (arbitrair) van de totale metaalionconcentratie gecomplexeerd is.

In eerste instantie zal de invloed van de precisie van enkele inputparameters op de waarde van de geoptimaliseerde constante nagegaan worden aan de hand van:

- perturbaties in de evenwichtsconcentratie van het proton in elk titratiepunt;
- perturbaties in het toegevoegd volume base per titratiestap;
- perturbaties in de totale concentratie van M, L en H in elk titratiepunt;
- perturbaties in het totaal initieel volume in het titratievat.

Hiervoor zullen "synthetische" potentiometrische datasets, gegenereerd via het $MatLab^{TM}$ programma 'SQ_simul.m' (zie hoofdstuk 4), geanalyseerd worden.

6.2.1 De invloed van een afleesfout op de pH in elk titratiepunt

De vrije concentraties pM_{ideaal} , pL_{ideaal} en pH_{ideaal} in elk titratiepunt worden gegenereerd in het eerste gedeelte van het programma 'SQ_simul.m' (zie hoofdstuk 4). Met perturbatie in de evenwichtsconcentratie van H wordt een afleesfout bedoeld op de pH-meter. We modelleren deze afleesfout in ieder titratiepunt i d.m.v. een gaussiaans verdeelde stochastische variabele ($\varepsilon_i \sim N(0, \sigma^2)$), waarbij we σ^2 constant veronderstellen over het ganse titratiegebied (\pm 80 punten) en we verder veronderstellen dat er geen correlatie is tussen de fouten in de verschillende titratiepunten.

In het programma 'SQ_simul.m' (hoofdstuk 4) houden we rekening met deze afleesfout, ε . Vermits bij de data-analyse de residuelen in de pH worden geminimaliseerd, geldt:

$$pH_{error}(i) = pH_{ideaal}(i) + \varepsilon_i \tag{6.1}$$

met pH_{error} de onnauwkeurige (experimentele) opmeting van de pH en pH_{ideaal} de ideale (berekende) pH.

 σ^2 verwijst naar σ_E^2 (aflees fout op pH) in formule (4.37), zodat de data als volgt gewogen



Figuur 6.1: Typische random perturbatie in de kurve van het aantal titratiepunten versus pH. De volle lijn stelt het ideale verloop van de pH voor, terwijl de punten de "experimentele" (onnauwkeurige) opmetingen weergeven wanneer de "echte" pH-waarde wordt afgelezen met een afleesfout $\varepsilon_i \sim N(0, \sigma^2)$. In deze illustratie is $\sigma = 0.5$ pH-eenheden

worden:

$$\mathbf{W} = \frac{1}{\sigma_E^2} \mathbf{I_n} \tag{6.2}$$

met W de gewichtenmatrix en I_n de overeenkomstige eenheidsmatrix. Aangezien de invloed van de fout op het toegevoegd volume base σ_V in de volgende paragraaf geëvalueerd wordt, stellen we in dit geval $\sigma_V = 0$. Figuur 6.1 toont een typische random perturbatie in de pH kurve. De volle lijn geeft het ideale onderliggend verloop van het aantal titratiepunten in functie van de pH weer. De punten stemmen overeen met de "experimentele" opmetingen wanneer de "echte" pH-waarde wordt afgelezen met een afleesfout $\varepsilon_i \sim N(0, \sigma^2)$. In deze illustratie is $\sigma = 0.5 pH$ -eenheden.

In hoofdstuk 4 werd het MatLab simulatieprogramma 'SQ_simul.m' gevalideerd a.h.v een vergelijking met het originele SUPERQUAD programma in fortran 77.

Om de invloed van de precisie in de pH op de numerieke waarde van de geoptimaliseerde stabiliteitsconstante β na te gaan, worden gesimuleerde data uitgaande van de modelsyste-

men 'sys1' en 'sys2' (tabel 4.2) geanalyseerd. In tegenstelling met de validatie in hoofdstuk 4 – waar slechts één vaste waarde van 0.002 pH-eenheden toegekend werd aan de grootte van de afleesfout σ (vergelijking (6.1): $\varepsilon_i \sim N(0, \sigma^2)$) – laten we in dit geval σ variëren binnen het interval $[10^{-5}, 0.02]$ pH-eenheden. Bij een goede experimentele opzet mag men deze laatste afwijking zeker niet meer als realistisch beschouwen. In deze simulatiestudie echter werd dit groot bereik van σ onderzocht om de trend in de foutenpropagatie te ontdekken.

In het programma wordt in elke Monte Carlo cyclus een nieuwe waarde voor σ binnen het beschouwde interval aangesproken en dit voor een gekozen aantal titraties. De output geeft de gemiddelde, geoptimaliseerde waarde voor β van elk deeltje in het model weer over het ingegeven aantal titraties (n = 10) en over alle Monte Carlo cycli (=15) evenals de corresponderende standaardafwijkingen.

Figuur 6.2 toont voor de afzonderlijke stabiliteitsconstanten uit systeem 1 het verloop in functie van σ (log eenheden). De horizontale lijn geeft de "echte" waarde van de constante aan (tabel 4.2). De 'error bars' worden weergegeven en zijn gelijk aan de 'standard error' (s.e.).

We brengen in herinnering dat als $\hat{\beta}$ een onvertekende schatter is van ("de echte") β er bij benadering geldt (zie hoofdstuk 2):

$$P[\hat{\beta} - 3s.e.(\hat{\beta}) \le \beta \le \hat{\beta} + 3s.e.(\hat{\beta})] \approx 0.997$$
(6.3)

zodat we er van mogen uitgaan dat de "echte" waarde van β met quasi zekerheid tussen de 3 s.e.-grenzen ligt.

Een analoge analyse werd uitgevoerd voor systeem 2 (tabel 4.2). De resultaten hiervan zijn weergegeven in figuur 6.3.

Beide reeksen figuren illustreren dat, zoals verwacht, naarmate de afleesfout groter wordt, de betrouwbaarheid op de numerieke waarde van de geoptimaliseerde parameter daalt en dit voor elke constante zowel uit systeem 1 als systeem 2. Desalniettemin blijft de "echte" waarde van de constante meestal tussen de 3 s.e.-grenzen liggen, hetgeen een indicatie is dat in het geval van relatief eenvoudige modellen, de toegelaten meetprecisie zelfs lager mag zijn dan deze gepostuleerd door de auteurs van SUPERQUAD [11]. Bovendien blij-

ven de geoptimaliseerde parameterwaarden eerder schommelen rond hun "echte" waarde zonder systematische afbuigingen ervan te vertonen.

Een regressieanalyse van de afleesfout op de pH in functie van de 'standard error' op de individuele geoptimaliseerde parameters in systeem 1 geeft een lineair verband aan in figuur 6.4. Een analoge analyse, uitgevoerd voor de individuele parameters in systeem 2, is weergegeven in figuur 6.5. Hier wordt in het geval van β_{111} en β_{110} eerder een kwadratisch verband waargenomen, terwijl we voor β_{00-1} een exponentieel verband zien. Het effect van een afleesfout in de pH op de precisie van de geoptimaliseerde parameter blijkt dus sterk afhankelijk te zijn van de waarde van de individuele parameter en van het gekozen modelsysteem.

6.2.2 De invloed van perturbaties in het toegevoegd volume base per titratiestap

In de volgende reeks simulaties gaan we na wat het effect is van de precisie van het toegevoegd volume base op de numerieke waarde van de geoptimaliseerde constante. Daartoe wordt in 'SQ_simul.m' het volume-increment, d_{vol} geperturbeerd met een random, gaussiaans verdeelde fout, ε :

$$vol_{recipient}(k) = vol_{recipient}(0) + \sum_{i=1}^{k} (d_{vol}(i) + \varepsilon_i)$$
 (6.4)

met $\varepsilon_i \sim N(0, \sigma^2)$ en onafhankelijk, waarbij σ varieert tussen 10^{-5} tot 5.10^{-3} ml. $Vol_{recipient}(k)$ en $vol_{recipient}(0)$ stellen respectievelijk het totaal volume in het titratievat voor in het k-de punt en het beginpunt van de titratie. σ werd bij de calibratie van de automatische pistonburetten in paragraaf 3.2.5.1 bepaald op 0.002 ml, zodat we 5.10^{-3} ml als een relatief grote afwijking kunnen beschouwen.

Merken we op dat σ^2 in deze sectie verwijst naar de term $\left(\frac{\partial E}{\partial V}\right)^2 \sigma_V^2$ uit formule (4.37), zodat de data als volgt gewogen worden:

$$\mathbf{W} = \frac{1}{\left(\frac{\partial E}{\partial V}\right)^2 \sigma_V^2} \mathbf{I_n} \tag{6.5}$$



Figuur 6.2: Gemiddelde β_{110} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{011} (=K3) en gemiddelde β_{012} (=K4) (onder) voor systeem 1 in functie van een afleesfout op de pH (log eenheden) voor tien titraties en vijftien σ -waarden. De 'error bars' zijn gelijk aan de 'standard error' op β . De constante lijn is de "echte" β -waarde. (K1 = 10⁸, $K2 = 10^{14}, K3 = 10^9, K4 = 10^{11})$



Figure 6.3: Gemiddelde β_{111} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{110} (=K3) en gemiddelde β_{011} (=K4) (midden). Gemiddelde β_{012} (=K5) en gemiddelde β_{00-1} (=K6) (onder) voor systeem 2 in functie van een afleesfout op de pH (log eenheden) voor tien titraties en vijftien σ -waarden. De 'error bars' zijn gelijk aan de 'standard error' op β . De constante lijn is de "echte" β -waarde. (K1 = 10¹², K2 = 10⁶, K3 = 10³, K4 = 10¹⁰, K5 = 10¹⁷, K6 = 10⁻¹⁴)



Figuur 6.4: Verloop van de s.e. op gemiddelde β_{110} (=K1) en van de s.e. op gemiddelde β_{120} (=K2) (boven). Verloop van de s.e. op gemiddelde β_{011} (=K3) en van de s.e. op gemiddelde β_{012} (=K4) (onder) voor systeem 1 in functie van de afleesfout op de pH. S.e. wordt telkens beschouwd over een groep van tien simulaties en vijftien σ -waarden



Figuur 6.5: Verloop van de s.e. op gemiddelde β_{111} (=K1) en van de s.e. op gemiddelde β_{120} (=K2) (boven). Verloop van de s.e. op gemiddelde β_{110} (=K3) en van de s.e. op gemiddelde β_{011} (=K4) (midden). Verloop van de s.e. op gemiddelde β_{012} (=K5) en van de s.e. op gemiddelde β_{00-1} (=K6) (onder) voor systeem 2 in functie van de afleesfout op pH. S.e. wordt telkens beschouwd over een groep van tien simulaties en vijftien σ -waarden

6.2 Kwantificering aan de hand van experimentele fouten



Figuur 6.6: Verloop van het ideaal toegevoegd volume base (volle lijn) in vergelijking met het random geperturbeerde toegevoegd volume (punten) versus de pH

met W de gewichtenmatrix en I_n de eenheidsmatrix. σ_E stellen we in dit geval gelijk aan 0. De afgeleide $\frac{\partial E}{\partial V}$ wordt berekend uitgaande van een fit doorheen drie punten in de gesimuleerde pH kurve met aanpassingen voor begin- en eindpunt.

Figuur 6.6 toont een typische verdeling van de random perturbatie in het toegevoegd volume base in functie van de pH voor twee mogelijke titraties. De volle lijn geeft telkens de ideale titratiekurve weer. De punten stemmen overeen met de "experimentele" opmetingen wanneer het "echte" toegevoegde volume wordt afgelezen met een afleesfout $\varepsilon_i \sim N(0, \sigma_V^2)$. In beide illustraties is $\sigma_V = 0.03$ ml. Belangrijk is hierbij op te merken dat deze figuren aantonen dat bij het uitvoeren van slechts één enkele titratie een systematische fout kan geïntroduceerd worden!

Merken we tevens op dat de perturbatie in het volume-increment wordt ingevoerd bij het genereren van de data, (hoofdstuk 4: programma 'gen_exp_data.m'), zodat als gevolg van deze perturbatie de gegenereerde totale en vrije concentraties van de componenten eveneens geperturbeerd zijn.

Voor een welbepaald aantal titraties (n=10) wordt per Monte Carlo cyclus in het programma 'SQ_simul.m' iedere waarde voor σ binnen het interval aangesproken. De output geeft de gemiddelde, geoptimaliseerde β van elk deeltje in het modelsysteem weer over het

aantal σ -waarden (=5) en de 'standard error' op β . Figuur 6.7 toont voor de afzonderlijke stabiliteitsconstanten uit systeem 1 (tabel 4.2) het verloop in functie van σ (log eenheden). De horizontale lijn geeft de "echte" waarde van de constante aan. De 'error bars' worden aangegeven. Een analoge analyse werd uitgevoerd voor systeem 2 (tabel 4.2). De resultaten hiervan zijn weergegeven in figuur 6.8.

Deze twee reeksen figuren illustreren weerom dat de betrouwbaarheid op de numerieke waarde van de geoptimaliseerde constante omgekeerd evenredig is met de fout op het toegevoegd volume en dit voor elke parameter zowel uit modelsysteem 1 als modelsysteem 2.

6.2.3 De invloed van perturbaties op het initieel volume in de recipiënt

In deze paragraaf worden gesimuleerde data geanalyseerd met een gaussiaans verdeelde perturbatie, ε op het beginvolume in de recipiënt, $vol_{recipient}(0)$. Deze fout, die binnen één enkele titratie systematisch van aard is en ingevoerd wordt bij het genereren van de data (programma 'gen_exp_data.m' hoofdstuk 4), propageert eveneens naar de gegenereerde totale en vrije concentratie van de componenten:

$$vol_{recipient}(k) = (vol_{recipient}(0) + \varepsilon) + \sum_{i=1}^{k} (d_{vol}(i))$$
 (6.6)

met $\varepsilon \sim N(0, \sigma^2)$, waarbij σ varieert binnen het arbitrair gekozen interval $[10^{-5}, 5.10^{-2}]$ ml.

Figuur 6.9 geldt voor systeem 1. Figuur 6.10 toont een analoge analyse voor systeem 2. Beide reeksen figuren geven aan dat voor deze "systematische" perturbatie, de "echte" waarde van de constante nog steeds binnen de 3 s.e.-grenzen blijft liggen.



Figuur 6.7: Gemiddelde β_{110} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{011} (=K3) en gemiddelde β_{012} (=K4) (onder) voor systeem 1 in functie van de perturbatie in het toegevoegd volume base. S.e. wordt beschouwd over een groep van tien experimenten en vijf σ -waarden. De 'error bars' zijn gelijk aan de 'standard error'. De constante lijn is de "echte" β -waarde



Figuur 6.8: Gemiddelde β_{111} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{110} (=K3) en gemiddelde β_{011} (=K4) (midden). Gemiddelde β_{012} (=K5) en gemiddelde β_{00-1} (=K6) (onder) voor systeem 2 in functie van de perturbatie in het toegevoegd volume base. S.e. wordt beschouwd over een groep van tien experimenten en vijf σ -waarden. De 'error bars' zijn gelijk aan de 'standard error'. De constante lijn is de "echte" β -waarde

1.41



Figuur 6.9: Gemiddelde β_{110} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{011} (=K3) en gemiddelde β_{012} (=K4) (onder) voor systeem 1 in functie van de perturbatie op het initieel volume in de recipiënt. S.e. wordt beschouwd over een groep van vijf experimenten en vijf σ -waarden. De 'error bars' zijn gelijk aan de 'standard error'. De constante lijn is de "echte" β -waarde



Figuur 6.10: Gemiddelde β_{111} (=K1) en gemiddelde β_{120} (=K2) (boven). Gemiddelde β_{110} (=K3) en gemiddelde β_{011} (=K4) (midden). Gemiddelde β_{012} (=K5) en gemiddelde β_{00-1} (=K6) (onder) voor systeem 2 in functie van de perturbatie op het initieel volume in de recipiënt. S.e. wordt beschouwd over een groep van vijf experimenten en vijf σ -waarden. De 'error bars' zijn gelijk aan de 'standard error'. De constante lijn is de "echte" β -waarde

6.2.4 Conclusie

In de voorgaande paragrafen werd de sensitiviteit nagegaan van de geoptimaliseerde constante voor de precisie van de input in verschillende fasen van de experimentele keten. Zowel afleesfouten op de pH, als fouten op het toegevoegd volume en fouten op het initieel volume in de recipiënt werden gemodelleerd d.m.v gaussiaans verdeelde stochasten.

Uit figuren 6.4 en 6.5 kan nu rechtstreeks de grootte-orde van de fout op de geoptimaliseerde constante afgeleid worden in functie van een gekende afleesfout op de pH. Een kwantificering aan de hand van perturbaties in het toegevoegd volume en perturbaties in het initieel volume kan analoog gebeuren; respectievelijk via figuren 6.7 en 6.8 en figuren 6.9 en 6.10.

Bij nader toezien tonen de figuren (6.4, 6.5, 6.7, 6.8, 6.9 en 6.10) aan dat voor vergelijkbare grootte (σ) van de fout in de geanalyseerde inputparameters, de precisie (procentueel uitgedrukt) van de geoptimaliseerde constante eveneens vergelijkbaar is in beide modelsystemen.

Om na te gaan voor welke inputparameter de geoptimaliseerde constante het meest gevoelig is, moeten we ons realiseren dat de grootte-orde van een realistische fout in het initieel volume echter gemiddeld een factor 10 groter is dan een realistische afleesfout op de pH, terwijl de grootte-orde van een realistische fout in het toegevoegd volume en een afleesfout op de pH vergelijkbaar is. Deze laatste vaststelling kan een indicatie zijn dat afleesfouten in belangrijke mate bepalend zijn voor de uiteindelijke precisie van de geoptimaliseerde constante.
6.3 Kwantificering aan de hand van "structurele" fouten

6.3.1 Inleiding

De betrouwbaarheid van de computeranalyse zal in deze sectie geëvalueerd worden met behulp van de twee modelsystemen gedefinieerd in hoofdstuk 4. Met structurele fout bedoelen we een wijziging in het chemisch model. Deze wijziging situeert zich in fase 3 van het organigram (figuur 4.2), in de *P*-matrix. De vraag die we ons hier stellen is de volgende: "is de verandering van de minimalisatiefunctie (gewogen som der kwadratische residuelen), U (hoofdstuk 4) significant wanneer het model wijzigt?". Volgens een kwalitatieve studie van Lomozik et al. in dit verband [9], [10], [17], [18] is de kans dat in SUPERQUAD een verkeerd model aanvaard wordt, klein. Anderzijds, volgens dezelfde auteur, zal bij het verwerpen van een juiste hypothese de optimalisatieprocedure vastlopen, hetzij door divergentie, hetzij door het berekenen van te grote standaardafwijkingen op de parameters. De optimalisatieprocedure zal hier geëvalueerd worden met behulp van gesimuleerde potentiometrische data, gegenereerd op basis van de gedefinieerde modellen, 'sys1' en 'sys 2' uit hoofdstuk 4. We zullen nieuwe modellen formuleren en deze vergelijken met de gedefinieerde startmodellen.

6.3.2 Analyse van modelsysteem 'sys 1'

6.3.2.1 Uitbreiding van het gedefinieerde startmodel

Het gedefinieerde modelsysteem 'sys 1' (hoofdstuk 4) bestaat uit volgende species met hun overeenkomstige stabiliteitsconstanten, β_{pqr} :

deeltje	$\log \beta_{pqr}$
ML	8
ML_2	14
HL	9
H_2L	11

Tabel 6.1 Gemiddelde waarde en 'standard error' (s.e.) over tien experimenten voor de minimalisatiefunctie, U ('sample standaardafwijking', S tussen haakjes) en de geoptimaliseerde constanten, β_{pqr} na convergentie, wanneer modelsysteem 1 wordt uitgebreid met het specie MLH met respectievelijke initiële

schatting voor β_{111} : $10^3, 10^5, 10^7, 10^8, 10^9, 10^{11}$

β_{111}	$ \overline{U} \pm s.e. $ (\overline{S})	$\overline{\beta_{110}} \pm s.e.$	$\overline{eta_{120}} \pm s.e.$	$\overline{\beta_{011}} \pm s.e.$	$\overline{\beta_{012}} \pm s.e.$	$\overline{\beta_{111}} \pm s.e.$
10 ³	$\begin{array}{c} 2.131010^3 \pm 3.990610^2 \\ (5.33) \end{array}$	$1.001410^8 \pm 1.395810^5$	$1.001810^{14} \pm 1.260910^{11}$	$1.000310^9 \pm 1.590810^5$	$1.000410^{11} \pm 2.724010^7$	$3.054710^1 \pm 4.7982$
10 ⁵	$\begin{array}{c} 4.579510^2 \pm 5.3519 \\ (2.47) \end{array}$	$1.009310^8 \pm 2.986110^5$	$1.008810^{14}\pm 3.010910^{11}$	$1.000910^9 \pm 5.441310^5$	$1.002010^{11}\pm 6.072610^{7}$	$8.009910^{-3} \pm 1.185810^{-4}$
107	$\begin{array}{c} 4.895710^2 \pm 1.8292 \\ (2.55) \end{array}$	$1.010010^8 \pm 2.871410^5$	$1.010210^{14}\pm 2.698710^{11}$	$1.000410^9 \pm 1.627810^5$	$1.002410^{11}\pm 5.811110^7$	$1.567710^{-2} \pm 5.188810^{-4}$
10 ⁸	$\begin{array}{c} 6.332810^2 \pm 9.9028 \\ (2.90) \end{array}$	$1.004610^8 \pm 1.670710^5$	$1.001210^{14}\pm7.079810^{10}$	$1.004210^9 \pm 1.609710^6$	$1.006010^{11}\pm 2.560010^8$	$2.231510^1 \pm 1.7873$
10 ⁹	$7.527710^3 \pm 9.399410^1 \\ (10.02)$	$1.006110^8 \pm 2.046410^5$	$1.004410^{14} \pm 1.917710^{11}$	$1.003810^9 \pm 2.301310^6$	$1.003410^{11}\pm 2.315610^8$	$2.870810^3 \pm 1.498710^2$
1011	$\begin{array}{c} 3.184710^4 \pm 7.688210^2 \\ (20.61) \end{array}$	$1.003410^8 \pm 1.499910^5$	$1.005410^{14}\pm 2.569110^{11}$	$1.003210^9 \pm 1.841310^6$	$1.000910^{11}\pm 4.964110^{7}$	$5.226910^8 \pm 9.269210^6$

170

In tabel 6.1 zijn de resultaten samengevat van de verschillende optimalisatiecycli waarbij het gedefinieerde model wordt uitgebreid met het complex MLH en de overeenkomstige β_{111} achtereenvolgens geschat wordt op 10³, 10⁵, 10⁷, 10⁸, 10⁹ en 10¹¹. De simulatie loopt telkens over tien titraties met een constante afleesfout in pH van 0.002 eenheden. De initiële geschatte waarden voor β_{110} , β_{120} , β_{011} en β_{012} wijken minstens 50% in positieve of negatieve zin af van hun "echte" waarde. \overline{U} is de gemiddelde kwadratische som der gewogen residuelen (*U*-functie), bij convergentie, over tien experimenten. De gemiddelde 'sample standaardafwijking' \overline{S} (hoofdstuk 4) wordt tussen haakjes weergegeven. $\overline{\beta_{pqr}}$ geeft de gemiddelde waarde van elke geoptimaliseerde constante, bij convergentie, weer. S.e. is de 'standard error'.

De data in tabel 6.1 illustreren dat in alle gevallen de parameters β_{110} , β_{120} , β_{011} en β_{012} convergeren naar hun "echte" waarde, hetgeen in eerste instantie kan betekenen dat het afwijkend model aanvaard wordt.

Wanneer bovendien in figuur 6.11 (links) de initiële schatting voor β_{111} (K111) uitgezet wordt ten opzichte van de finale waarde van U bij convergentie (gemiddelde waarde over tien titraties met $\sigma_E = 0.002 \ pH$ -eenheden) blijkt dat U aanvaardbaar blijft binnen het interval $\beta_{111} = [10^5, 10^8]$ (U = 372 voor het gedefinieerde startmodel). In hoofdstuk 4 werd de maximale grens van S voor het aanvaarden van een model arbitrair vastgelegd op 3, dit betekent dat de maximum toelaatbare waarde voor U ($S = \left(\frac{U}{n-m}\right)^{1/2}$) rond 700 ligt (n, aantal titratiepunten = 80 en m, aantal parameters = 5.) Op basis van dit criterium zou het afwijkend model in vernoemd interval, eveneens aanvaard kunnen worden! Wordt β_{111} echter initieel geschat op 10⁹ en hoger, dan stijgt U spectaculair, zodat het toevoegen van een "verdacht" en zeer stabiel deeltje aan het gedefinieerde startmodel in dit gebied geen problemen oplevert bij de selectie van het juiste model. Een lichte verhoging van Uis eveneens merkbaar voor $\beta_{111} = 10^3$, waarschijnlijk te wijten aan het vastlopen van de optimalisatie in een lokaal minimum.

Figuur 6.11 (rechts) toont het verloop van de initiële schatting voor β_{111} in functie van de waarde voor β_{111} bij convergentie. Hieruit kunnen we afleiden dat wanneer het startmodel wordt uitgebreid met het complex *MLH* met schatting voor β_{111} binnen het interval [10⁵, 10⁷] deze waarden bij convergentie quasi uitdoven, hetgeen een indicatie is om ook in dit gebied het toegevoegde complex te verwerpen of het voorgestelde model niet te aanvaarden.

6.3 Kwantificering aan de hand van "structurele" fouten

Deze laatste vaststelling wordt nogmaals verduidelijkt in figuur 6.12; links is het concentratieverloop (evenwichtsconcentraties) van de afzonderlijke componenten en alle voorkomende species in het startmodel systeem 1 uitgezet, terwijl rechts het concentratieverloop voor het afwijkend model waaraan het deeltje *MLH* werd toegevoegd, wordt geschetst. β_{111} wordt hierbij initieel geschat op 22; dit is de gemiddelde waarde die bekomen werd na optimalisatie vertrekkende van een schatting voor β_{111} van 10⁸ (zie figuur 6.11 en data in tabel 6.1 voor $\beta_{111} = 10^8$). Uit figuur 6.12 leiden we af dat er fundamenteel niets verandert aan het concentratieverloop van de componenten en de aanwezige species in het startmodel en dat het toegevoegde complex dus kan worden verworpen op grond van nietaanwezigheid.

Een analoge concentratiekurve wordt uitgezet voor $\beta_{111} = 5.10^8$ in figuur 6.13 (rechts); dit is de gemiddelde waarde die werd bekomen na optimalisatie vertrekkende van een initiële schatting voor β_{111} van 10^{11} . Het verschil in concentratieverloop tussen deze kurve en de kurve voor het startmodel systeem 1 (figuur 6.13 (links)) wordt aangegeven in figuur 6.14, waarbij het titratiepuntnummer een maat is voor de *pH*. Aangezien *MLH* enkel in het zure gebied voorkomt (hoofdstuk 5) stellen we inderdaad vast dat dit deeltje in het basisch gebied geen invloed meer heeft op de concentratie van de andere aanwezige species in het startmodel.

Het voorgaande verklaart tenslotte nog de vaststelling dat de geoptimaliseerde waarden voor de parameters in het startmodel zeer goed hun "echte" waarde benaderen (tabel 6.1); temeer daar de berekende concentraties aan de hand van de modeluitbreiding minimaal zijn.

Uit figuur 6.15 kan geen rechtstreeks verband afgeleid worden tussen de som der kwadratische relatieve fouten ($\beta_{geopt.} - \beta_{echt}/\beta_{echt}$) op de geoptimaliseerde β 's voor de respectievelijke initiële schattingen van β_{111} en de 'sample standaarafwijking', S na convergentie. Normalerwijze verwacht men een rechtstreeks verband tussen deze relatieve fout en S.

6.3.2.2 Beperking van het gedefinieerde startmodel

In tabel 6.2 worden de resultaten samengevat voor de optimalisatie van een aantal afwijkende –één of meerdere species ontbreken– modellen van het startmodel 'sys 1'. De gesimuleerde data, afkomstig van het gedefinieerde startmodel, zijn geperturbeerd met een



Figuur 6.11: Verloop van de initiële schatting voor β_{111} (links), toegevoegd aan het model (in log-eenheden) versus de minimalisatiefunctie, U (na convergentie), telkens gemiddeld over tien titraties met $\sigma_E = 0.002 \text{ pH}$ -eenheden. De 'errorbars' zijn gelijk aan de 'standard error'. Verloop van de initiële schatting voor β_{111} (in log-eenheden) versus de geoptimaliseerde waarde voor β_{111} na convergentie (rechts), telkens gemiddeld over tien titraties met $\sigma_E = 0.002 \text{ pH}$ -eenheden. De 'error bars' zijn gelijk aan de 'standard error'



Figuur 6.12: Concentratieverloop (evenwichtsconcentraties) van de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 1 (links). Overeenkomstig concentratieverloop wanneer het deeltje MLH ($\beta_{111} = 22$) toegevoegd wordt aan het model (rechts)



Figuur 6.13: Concentratieverloop (evenwichtsconcentraties) van de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 1 (links). Overeenkomstig concentratieverloop wanneer het deeltje MLH ($\beta_{111} = 5.10^8$) toegevoegd wordt aan het model (rechts)



Figuur 6.14: Verschil in concentratieverloop tussen de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 1 en afwijkend model waaraan het deeltje MLH ($\beta_{111} = 5.10^8$) toegevoegd wordt



Figuur 6.15: Som der kwadraten van de relatieve fout op geoptimaliseerde β 's in modelsysteem 1 versus S ('sample standaardafwijking') voor tien experimenten en $\sigma_E = 0.002$ pH-eenheden

constante afleesfout in pH van 0.002 eenheden. De geoptimaliseerde parameterwaarden vermeld in de tabel zijn goed reproduceerbaar (de optimalisatieprocedure werd een aantal maal herhaald onder analoge startvoorwaarden); dit geldt eveneens voor de waarde van U (S) bij convergentie.

Tabel 6.2 Resultaten voor de optimalisatie van een aantal afwijkende -één of meerdere

species ontbreken- modellen van het gedefinieerde startmodel 'sys 1'. De "echte" parameterwaarden staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens vermeld

		log/pqr	U (S)
_	startwaarde	geoptimaliseerde waarde	bij convergentie
ML	8.17609 (8)	8.45327	5.35.10 ⁹ (8.23.10 ³
ML	8.17609 (8)	7.77895	$2.84.10^9$ (6.03.10 ³)
ML_2	13.69897 (14)	-5.87324	
HL	9.17609 (9)	9.11257	2.79.10 ⁹ (5.98.10 ³)
H_2L	10.69897 (11)	-3.27768	
ML_2	13.69897 (14)	13.60352	$8.29.10^7 (1.04.10^3)$
HL	9.17609 (9)	7.27167	
H_2L	10.69897 (11)	10.47695	
ML	8.17609 (8)	7.56829	$3.52.10^7 (6.76.10^2)$
HL	9.17609 (9)	7.27180	
H_2L	10.69897 (11)	10.32212	
	ML ML ML ₂ HL H ₂ L ML ₂ HL H ₂ L ML HL H ₂ L	$\begin{array}{rrrr} ML & 8.17609 \ (8) \\ ML & 8.17609 \ (8) \\ ML_2 & 13.69897 \ (14) \\ HL & 9.17609 \ (9) \\ H_2L & 10.69897 \ (11) \\ ML_2 & 13.69897 \ (14) \\ HL & 9.17609 \ (9) \\ H_2L & 10.69897 \ (11) \\ \end{array}$	ML 8.17609 (8)8.45327 ML 8.17609 (8)7.77895 ML_2 13.69897 (14)-5.87324 HL 9.17609 (9)9.11257 H_2L 10.69897 (11)-3.27768 ML_2 13.69897 (14)13.60352 HL 9.17609 (9)7.27167 H_2L 10.69897 (11)10.47695 ML 8.17609 (8)7.56829 HL 9.17609 (9)7.27180 H_2L 10.69897 (11)10.32212

Zoals tabel 6.2 aangeeft, leidt de selectie van alle geteste modellen waarbij één of meerdere species ontbreken tot onaanvaardbare resultaten: de waarde van U(S) bij convergentie is overal veel groter dan de maximaal aanvaardbare waarde van ongeveer 700 (3). Bovendien leidt inspectie van de residuelen in pH tot dezelfde conclusie, zoals verder zal

blijken.

We brengen in herinnering dat de afleesfout in elk titratiepunt gemodelleerd wordt d.m.v. een gaussiaans verdeelde stochastische variabele $\varepsilon_i \sim N(0, \sigma^2)$.

De twee karakteristieken, met name het normale karakter en de onafhankelijkheid van de data in de verdeling worden respectievelijk gecontroleerd aan de hand van een normaliteitsplot en een autocorrelatieplot. Een lineaire normaliteitsplot wijst op een symmetrische normale verdeling (Appendix A), hetgeen vastgesteld wordt in figuur 6.16 (boven, rechts) voor het startmodel systeem 1 met aanvaardbare U(S) van 372 (2.21). Voor het afwijkend model waarbij ML en ML_2 ontbreken (figuur 6.16 (onder, rechts) is de normaliteitsplot niet lineair (hoge U(S)).

Anderzijds geeft de autocorrelatieplot informatie over het onafhankelijk karakter van de data: de correlatie tussen de geordende rij van de residuelen en zichzelf na verschuiving over een bepaalde afstand (='lag') wordt nagegaan. De terugval naar 37% (in logaritmische schaal 1 eenheid (1/e)) is typerend voor de structuur van de ruis of afleesfout en wordt de autocorrelatielengte genoemd.

Figuur 6.17 toont de autocorrelatie van de residus voor het startmodel systeem 1 (links) en voor het afwijkend model waarbij ML en ML_2 ontbreken (rechts). Hieruit blijkt dat in het geval van het goede model de autocorrelatielengte zeer klein is, zodat de residus in pH geen systematische trends vertonen. Voor het afwijkend model daarentegen is de autocorrelatielengte ongeveer 10, hetgeen wijst op een zekere structuur in de residus.

Het uitzetten van de grootte van de ruis (log eenheden) in functie van de 'sample standaardafwijking' S voor het goede startmodel systeem 1 (figuur 6.18) geeft tenslotte aan dat zowel in het geval van "goede" als "slechte" data, de S-waarde beneden de grens van 3 blijft.



Figuur 6.16: Plot van de residuelen in pH versus titratiepuntnummer en overeenkomstige normaliteitsplot van de residuelen voor het gedefinieerde startmodel systeem 1 (boven) met $\sigma_E = 0.002 \ pH$ -eenheden. Plot van de residuelen in pH voor afwijkend model van systeem 1 waarbij de deeltjes ML en ML_2 ontbreken en overeenkomstige normaliteitsplot van de residuelen (onder) met $\sigma_E = 0.002 \text{ pH-eenheden}$



Figuur 6.17: Autocorrelatie van de residuelen voor startmodel systeem 1 (links), autocorrelatie voor afwijkend model van systeem 1: ML en ML_2 ontbreken (rechts)

6.3.2.3 Gecombineerde modellen

Tabel 6.3 Resultaten voor de optimalisatie van een afwijkend (gecombineerd) model van het gedefinieerde startmodel 'sys 1'. De "echte" parameterwaarden staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens vermeld

$\log \beta_{pqr}$	$\log \beta_{pqr}$	U(S)
startwaarde	geoptimaliseerde waarde	bij convergentie
13.69897 (14)	13.77711	$8.28.10^7 (1.04.10^3)$
9.17609 (9)	7.27163	
10.69897 (11)	10.69900	
5.17609	3.45180	
	$\frac{\log \beta_{pqr}}{\text{startwaarde}}$ 13.69897 (14) 9.17609 (9) 10.69897 (11) 5.17609	$\log \beta_{pqr}$ $\log \beta_{pqr}$ startwaardegeoptimaliseerde waarde13.69897 (14)13.777119.17609 (9)7.2716310.69897 (11)10.699005.176093.45180

Tabel 6.3 toont de resultaten van de optimalisatie van een gecombineerd model van systeem 1. Niettegenstaande het optimaliseren naar nieuwe parameterwaarden die niet zo sterk afwijken van de de "echte" waarde, wordt het gecombineerd model van systeem 1 (tabel 6.3) niet aanvaard op grond van een te hoge U(S) waarde.





Figuur 6.18: Verloop van de grootte van de afleesfout in pH (σ_E) (log eenheden) in functie van S ('sample standaardafwijking') voor modelsysteem 1, gemiddeld over vijf experimenten. De 'error bars' zijn gelijk aan de 'standard error'

6.3.2.4 Conclusie

De analyse van de modelselectieprocedure aan de hand van modelsysteem 1 toont aan dat voor stabiele en relatief eenvoudige modellen (in dit geval vier species) de 'sample standaardafwijking' S of de waarde voor de minimalisatiefunctie U als criterium voor modelselectie volstaan. Een controle van het concentratieverloop van de species, van de verdeling en de autocorrelatie van de residuelen zijn aan te bevelen, maar geen noodzakelijke criteria voor selectie van het juiste model.

6.3.3 Analyse van modelsysteem 'sys 2'

6.3.3.1 Uitbreiding van het gedefinieerde startmodel

Het gedefinieerde modelsysteem 'sys 2' (hoofdstuk 4) bestaat uit volgende species met hun overeenkomstige stabiliteitsconstanten, β_{pqr} :

deeltje	$\log \beta_{pqr}$
MLH	12
ML_2	6
ML	3
HL	10
H_2L	17
OH^-	-14

Voor het gedefinieerde startmodel 'sys2' is, na convergentie, de waarde voor $U = 4.77.10^2$. Tabel 6.4 vat de resultaten samen voor de optimalisatie van modelsysteem 2, uitgebreid met het complex ML_2H met initiële schatting voor $\beta_{121} = 1.5.10^5$.

Tabel 6.4 Resultaten voor de optimalisatie van een afwijkend model van het gedefinieerde
startmodel 'sys 2' (het complex ML_2H wordt toegevoegd). De "echte" parameterwaarden
staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens vermeld

Model	$\log \beta_{pqr}$	$\log \beta_{pqr}$	U(S)	
	startwaarde	geoptimaliseerde waarde	bij convergentie	
MLH	12.17609 (12)	12.17615	$8.28.10^7 (1.07.10^3)$	
ML_2	5.77085 (6)	5.77084		
ML	3.17609 (3)	3.17612		
HL	9.69897 (10)	9.69900		
H_2L	17.15229 (17)	17.15226		
OH^-	-14.22915 (-14)	-14.22914		
ML_2H	5.17609	7.95654		

De data in tabel 6.4 illustreren dat niettegenstaande alle parameters geoptimaliseerd worden tot aanvaardbare waarden, het afwijkend model toch wordt verworpen op grond van een te hoge U(S)-waarde.

6.3.3.2 Beperking van het gedefinieerde startmodel

Tabel 6.5 vat de resultaten samen voor de optimalisatie van afwijkende modellen van modelsysteem 2 waarbij ofwel het MLH ofwel het ML_2 ontbreekt.

Uit tabel 6.5 blijkt dat indien het deeltje ML_2 ontbreekt in het model, de waarde van U slechts een factor 10 afwijkt van deze wanneer het "echte" model geoptimaliseerd wordt. Ontbreekt het MLH complex in het model dan is U ongeveer een factor 100 groter in vergelijking met het "echte" model. In beide gevallen echter worden alle parameters geoptimaliseerd tot aanvaardbare waarden!

Inspectie van de residus lijkt hier aangewezen. De residuelen in pH worden uitgezet over het ganse titratiegebied (80 punten), zowel voor het "goede" modelsysteem 2 (figuur 6.19 (boven)) als voor de modellen waarin MLH (figuur 6.19) (midden) of ML_2 (figuur 6.19) (onder) ontbreekt. De overeenkomstige normaliteitsplots van de residuelen zijn telkens rechts weergegegeven in figuur 6.19. Het is duidelijk dat de afwijkende modellen systematische trends in de residus vertonen. Tabel 6.5 Resultaten voor de optimalisatie van een aantal afwijkende -één complex ontbreekt- modellen van het gedefinieerde startmodel 'sys 2'. De "echte" parameterwaarden staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens

vermeld

	Model	$\log \beta_{pqr}$	$\log \beta_{pqr}$	U(S)
		startwaarde	geoptimaliseerde waarde	bij convergentie
1)	ML_2	5.77085 (6)	5.85104	$6.20.10^4$ (28.75)
	ML	3.17609 (3)	2.73477	
	HL	9.69897 (10)	10.00169	
	H_2L	17.15229 (17)	16.96720	
	OH^{-}	-14.22915 (-14)	-13.99952	
	MLH ontbreekt			
2)	MLH	12.17609 (12)	11.98612	6.09.10 ³ (9.01)
	ML	3.17609 (3)	2.65716	
	HL	9.69897 (10)	9.80500	
	H_2L	17.15229 (17)	16.82029	
	OH^-	-14.22915 (-14)	-14.09657	
	ML_2 ontbreekt			

Een controle van de autocorrelatie bevestigt dit: voor het startmodel vertoont figuur 6.20 (boven) een zeer kleine autocorrelatielengte, terwijl voor de afwijkende modellen (midden en onder) de autocorrelatielengte beduidend groter is. Merken we tevens op dat voor het afwijkend model met de hoogste U(S)-waarde de grootste autocorrelatielengte wordt bekomen.

Bovendien geven de residuelenplots – het model is uiteraard gekend – correcte informatie over het ontbreken van het complex MLH enerzijds en het ML_2 anderzijds. In het zure gebied waar het MLH normaliter voorkomt, vertonen de residus systematische tendensen, terwijl in het basisch gebied de berekende pH overeenkomt met de experimentele pH (figuur 6.19 (midden)). Voor het model waarbij het ML_2 (komt voor in het basisch gebied) ontbreekt, nemen we in het zure gebied geen afwijkingen in de residuelen waar (figuur 6.19 (onder)).

Controle van het concentratieverloop

Kunnen de resultaten in tabel 6.5 tevens gecorreleerd worden met het concentratieverloop van de afzonderlijke componenten en de verschillende species in de verschillende modellen?

In figuur 6.21 (links) is het concentratieverloop (evenwichtsconcentraties) van de afzonderlijke componenten en alle voorkomende species in het startmodel systeem 2 over het gehele titratiegebied uitgezet. In dezelfde figuur (rechts) is het concentratieverloop van de componenten en de species over het ganse titratiegebied uitgezet voor een afwijkend model van systeem 2 waarbij het deeltje MLH ontbreekt. Figuur 6.22 toont het concentratieverloop voor het startmodel systeem 2 over het ganse titratiegebied in vergelijking met het concentratieverloop voor het afwijkend model van systeem 2 waarbij het deeltje ML_2 ontbreekt (rechts). Beide figuren bevestigen dat bij ontbreken van een deeltje in het te optimaliseren model het concentratieverloop van de aanwezige species en componenten in het startmodel niet wordt beïnvloed, zodat op grond hiervan het voorgestelde model kan verworpen worden.

Wordt – zoals voor modelsysteem 1 – de grootte van de ruis (σ) uitgezet versus de waarde van S voor modelsysteem 2 (figuur 6.23), dan blijkt ook hier dat zelfs voor onrealistische waarden van σ de grens van S = 3 niet wordt overschreden. Voor de afwijkende modellen van systeem 2 wordt S gunstiger naarmate de grootte van de afleesfout toeneemt (figuur 6.24), hetgeen we verwachten daar een betere fit kan bekomen worden met een grotere marge in de ruis.

Figuur 6.25 toont bovendien aan dat voor de afwijkende modellen vermeld in tabel 6.5 de waarde van S quasi ongevoelig is voor een toenemende (relatieve) fout in de geoptimaliseerde parameter.



Figuur 6.19: Plot van de residuelen in pH versus titratiepuntnummer voor het gedefinieerde startmodel systeem 2 en overeenkomstige normaliteitsplot (boven). Plot van de residuelen in pH wanneer het complex MLH ontbreekt en overeenkomstige normaliteitsplot (midden). Plot van de residuelen in pH wanneer het complex ML_2 ontbreekt in systeem 2 en overeenkomstige normaliteitsplot (onder); $\sigma_E = 0.002 \text{ pH-eenheden in de drie gevallen}$



Figuur 6.20: Autocorrelatie van de residuelen voor startmodel systeem 2 (boven); afwijkend model van systeem 2 waarin het complex MLH ontbreekt (midden); afwijkend model van systeem 2 waarin het complex ML_2 ontbreekt (onder)



Figuur 6.21: Concentratieverloop van de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 2 (links). Overeenkomstig concentratieverloop wanneer het complex MLH ontbreekt in het model (rechts)



Figuur 6.22: Concentratieverloop van de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 2 (links). Overeenkomstig concentratieverloop wanneer het complex ML₂ ontbreekt in het model (rechts)



Figuur 6.23: Verloop van de grootte van de afleesfout in pH (σ_E) (log eenheden) in functie van S ('sample standaardafwijking') voor modelsysteem 2, gemiddeld over vijf experimenten. De 'error bars' zijn gelijk aan de 'standard error'



6.3 Kwantificering aan de hand van "structurele" fouten

Figuur 6.24: Verloop van de grootte van de afleesfout in pH (σ_E) (log eenheden) in functie van S ('sample standaardafwijking') (log eenheden) voor afwijkend model van systeem 2 waarbij het complex ML_2 ontbreekt (links) en afwijkend model waarbij MLH ontbreekt (rechts), gemiddeld over vijf experimenten. De 'error bars' zijn gelijk aan de 'standard error'



Monte Carlo studie van de foutenpropagatie

Figuur 6.25: Som der kwadraten van de relatieve fouten op de geoptimaliseerde β 's in functie van S ('sample standaardafwijking') voor afwijkend model van systeem 2 waarbij het complex ML_2 ontbreekt (links) en afwijkend model waarbij MLH ontbreekt (rechts) voor vijf afzonderlijke experimenten met $\sigma_E = 0.002 \text{ pH-eenheden}$

Hoofdstuk 6

6.3.3.3 Gecombineerde modellen

Tabel 6.6 vat de resultaten samen voor de optimalisatie van een gecombineerd model van systeem 2, waarbij MLH ontbreekt en ML_2H is toegevoegd. De optimalisatie vertrekt in beide simulaties van geperturbeerde data ($\sigma_E = 0.002 \ pH$ -eenheden). In het eerste experiment echter worden alle parameters geoptimaliseerd, terwijl in de tweede simulatie de pK_W en de pK_A 's constant worden gehouden.

Tabel 6.6 Resultaten voor de optimalisatie van een aantal afwijkende (gecombineerde) modellen van het gedefinieerde startmodel 'sys 2' vertrekkende van geperturbeerde data ($\sigma_E = 0.002 \text{ pH-eenheden}$). De "echte" parameterwaarden staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens vermeld

	Model	$\log \beta_{pqr}$ startwaarde	$\log \beta_{pqr}$	U(S)
1)	ML_2	5.77085 (6)	6.03878	$1.64.10^2$ (1.49)
	ML	3.17609 (3)	1.76893	
	HL	9.69897 (10)	10.02018	
	H_2L	17.15229 (17)	16.98589	
	OH^-	-14.22915 (-14)	-13.99346	
	$+ML_2H$	5.17609	15.12163	
	-MLH			
2)	ML_2	5.77085 (6)	6.03257	2.99.10 ⁶ (197)
	ML	3.17609 (3)	2.56745	
	HL	9.69897 (10)	9.69897	
	H_2L	17.15229 (17)	17.15229	
	OH^-	-14.22915 (-14)	-14.22915	
	$+ML_2H$ -MLH	5.17609	15.13518	

Het valt onmiddellijk op in tabel 6.6 dat er hier problemen rijzen bij de selectie van het juiste model; alle parameters worden geoptimaliseerd tot aanvaardbare waarden en bovendien is de U(S)-waarde zelfs kleiner dan deze voor het "goede" model! Het is duidelijk

dat de selectie op basis van enkel het "S-criterium" niet volstaat. Controle van het concentratieverloop en inspectie van de residuelen is dus niet enkel aan te bevelen, maar een noodzaak. Wanneer het concentratieverloop voor het afwijkend model in tabel 6.6 vergeleken wordt met het concentratieverloop in het startmodel systeem 2 (figuur 6.26) blijkt in eerste instantie dat het verdacht complex ML_2H slechts in zeer geringe mate aanwezig is. Bovendien wordt door het ontbreken van MLH de concentratie van de aanwezige species in het startmodel niet beïnvloed, zodat op basis hiervan het afwijkend model kan verworpen worden! Figuur 6.27 (boven en midden) illustreert een grote afwijking van het normale karakter in de residuelen vertrekkende van geperturbeerde data. Bovendien is de autocorrelatielengte groot (figuur 6.28 boven en midden), zowel in het geval dat alle parameters worden geoptimaliseerd als in het geval waarbij de pK_W en de pK_A 's constant worden gehouden. Indien de pK_W en de pK_A 's niet mee verfijnd worden is het logisch dat U(S) veel groter is - het systeem heeft immers minder vrijheidsgraden om tot een goede fit te komen en op basis hiervan kan het afwijkend model verworpen worden. In tabel 6.7 vertrekt de optimalisatie van niet geperturbeerde data voor gecombineerde modellen van systeem 2, waarbij MLH ontbreekt en ML_2H is toegevoegd. In de eerste simulatie worden alle parameters verfijnd; in de tweede simulatie worden de pK_W en de pK_A 's constant gehouden. De lichtjes verhoogde U(S) in het eerste experiment (tabel 6.7) kan verklaard worden door het feit dat er geen ruis op de data zit, waardoor het minder gemakkelijk wordt om de berekende data te fitten aan de experimentele (synthetische) data. De veel hogere U(S)in de tweede simulatie kan op analoge manier verklaard worden als in tabel 6.6. Zowel de residuelenplot (figuur 6.27 onder) als de autocorrelatieplot (figuur 6.28 onder) bevestigen de resultaten van de optimalisatieprocedures vertrekkende van niet geperturbeerde data in tabel 6.7.

Residuelen experimentele studie: Ag/1,4-diaminobutaan/H

Verificatie van de residuelen in E_{Ag} voor het in hoofdstuk 5 bestudeerde experimenteel ternair systeem Ag/1, 4 - diaminobutaan/H geeft voor het model dat het best overeenstemt met de experimentele data (S = 2.33) een mooie lineaire trend aan in de normaliteitsplot (figuur 6.29 onder) en illustreert bovendien de kleine autocorrelatielengte (figuur 6.30 onder). De residuelen in pH (figuur 6.29 boven) vertonen een quasi lineair gedrag, terwijl de grootte van de autocorrelatielengte (figuur 6.30 boven) kan wijzen op een zekere afhankelijkheid van de residus onderling.

.



ŝ

-55

0 10

40

nummer titrateount

50 60

20 30

Figuur 6.26: Concentratieverloop van de afzonderlijke componenten en alle species over het ganse titratiegebied voor startmodel systeem 2 (links). Overeenkomstig concentratieverloop wanneer het complex MLH ontbreekt en ML₂H wordt toegevoegd aan modelsysteem 2 (rechts). De data zijn geperturbeerd ($\sigma_E = 0.002 \text{ pH-eenheden}$)

70

R

-55

0 10 20 30 40 50 60 70 80

nummer Strabepunt

Tabel 6.7 Resultaten voor de optimalisatie van een aantal afwijkende (gecombineerde) modellen van het gedefinieerde startmodel 'sys 2' vertrekkende van **niet geperturbeerde** data. De "echte" parameterwaarden staan tussen haakjes. De U-waarde (S tussen haakjes) wordt eveneens vermeld

	Model	$\log eta_{pqr}$	$\log \beta_{pqr}$	U(S)
		startwaarde	geoptimaliseerde waarde	bij convergentie
1)	ML_2	5.77085 (6)	5.79801	$2.37.10^3$ (5.66)
	ML	3.17609 (3)	2.14550	
	HL	9.69897 (10)	9.79830	
	H_2L	17.15229 (17)	16.76659	
	OH^-	-14.22915 (-14)	-14.09423	
	$+ML_2H$	5.17609	7.95655	
	-MLH			
2)	ML_2	5.77085 (6)	5.79034	2.98.10 ⁶ (197)
	ML	3.17609 (3)	1.39657	
	HL	9.69897 (10)	9.69897	
	H_2L	17.15229 (17)	17.15229	
	OH^-	-14.22915 (-14)	-14.22915	
	$+ML_2H$	5.17609	7.68443	
	-MLH			

6.3.3.4 Conclusie

De analyse van complexe en 'onstabiele' modelsystemen (modelsysteem 2 met zes species en constanten van uiteenlopende grootte-orde) toont aan dat de resultaten van de optimalisatieprocedure met de nodige voorzichtigheid geïnterpreteerd moeten worden. Het verwerpen van "overbodige" species blijkt in de meeste gevallen geen moeilijkheden te geven : de S-waarde is in alle gevallen zeer groot.

Ook het opsporen van "ontbrekende" deeltjes in het model lijkt geen problemen op te leveren: inspectie van de residuelenverdeling en de autocorrelatie in de residuelen brengt systematische afwijkingen in de data aan het licht, waarvoor de berekeningen niet kunnen compenseren. Wanneer echter gecombineerde modellen worden geoptimaliseerd kan een



Figuur 6.27: Plots van de residuelen in pH versus titratiepuntnummer voor gecombineerde modellen (MLH ontbreekt, ML_2H is toegevoegd) van het startmodel systeem 2 en overeenkomstige normaliteitsplots. De data zijn geperturbeerd met een afleesfout van 0.002 pH- eenheden (boven); geperturbeerde data, maar pK_W en pK_A's worden niet geoptimaliseerd (midden); de data zijn niet geperturbeerd en alle parameters worden geoptimaliseerd (onder)



Figuur 6.28: Autocorrelatie van de residuelen voor gecombineerd model van systeem 2: geperturbeerde data met alle parameters geoptimaliseerd (boven); zelfde model maar pK_W en pK_A 's niet geoptimaliseerd (midden); zelfde model maar niet geperturbeerde data en alle parameters worden geoptimaliseerd (onder)



Figuur 6.29: Plot van de residuelen in E_H over het ganse titratiegebied, voor alle C_{Ag}/C_L verhoudingen en overeenkomstige normaliteitsplot voor het experimenteel ternair systeem Ag - 1, 4 - diaminobutaan - H (boven). Plot van de residuelen in E_{Ag} over het ganse titratiegebied, voor alle C_{Ag}/C_L verhoudingen en overeenkomstige normaliteitsplot voor het $experimenteel \ ternair \ systeem \ Ag - 1, 4 - diaminobutaan - H \ (onder)$



Figuur 6.30: Autocorrelatieplot van de residuelen in E_H voor het experimenteel ternair systeem Ag - 1, 4 - diaminobutaan - H (alle C_{Ag}/C_L verhoudingen) (boven). Autocorrelatieplot van de residuelen in E_{Ag} voor het experimenteel ternair systeem Ag - 1, 4 diaminobutaan - H (alle C_{Ag}/C_L verhoudingen) (onder)

6.3 Kwantificering aan de hand van "structurele" fouten

waarde van de 'sample standaardafwijking' S beneden de 3 misleidend zijn voor het aanvaarden van een verkeerd model! In dit geval is het van belang dat zowel de S-waarde, als de residuelenverdeling en het concentratieverloop gecontroleerd worden alvorens men de beslissing neemt om het model te aanvaarden. De voorgaande conclusies zijn uiteraard gebaseerd op het feit dat men het "echte" model kent. Bij het optimaliseren van onbekende modellen, zoals de in hoofdstuk 5 bestudeerde Ag(I)-complexeringsevenwichten met de SUPERQUAD fortran versie, is onze ervaring dat de χ^2 -test het meest controversiële en minst overtuigende criterium is voor modelselectie: een hoge waarde voor χ^2 (> 12.60) wordt meestal genegeerd als de S-waarde gunstig is. Een controle van de residuelenverdeling en de autocorrelatie in de residus voor de experimenteel onbekende modellen (figuren 6.29 en 6.30) laat een overwegend lineaire trend in de normaliteitsplot en een kleine autocorrelatielengte zien, hetgeen dus aantoont dat er praktisch geen systematische tendensen in de data voorkomen, terwijl de berekende χ^2 gelijk was aan 40.54.

Referenties

- [1] P.G. Potvin, Anal. Chim. Acta, 299, 43, (1994)
- [2] A.C. Olivieri, G.M. Escandar, Anal. Letters, 30, 1967, (1997)
- [3] S.E. Cabaniss, Anal. Chem., 69, 3658, (1997)
- [4] P.M. May, K. Murray, Talanta, 35, 933, (1988)
- [5] P.M. May, K. Murray, Talanta, 38, 1409, (1991)
- [6] J. Texter, Anal. Chim. Acta, 248, 603, (1991)
- [7] V.A. Kropotov, J. Anal. Chem., 54, 134, (1999)
- [8] V.A. Kropotov, J. Anal. Chem., 54, 19, (1999)
- [9] L. Lomozik, M. Jaskólski, A. Gasowska, J. Chem. Educ., 72, 27, (1995)
- [10] L. Lomozik, Chem. Anal. (Warsaw), 44, 89, (1999)
- [11] P. Gans, A. Sabatini, A. Vacca, J. Chem. Soc., Dalton Trans., 1195, (1985)
- [12] P. Gans, Manual SUPERQUAD, (1989)
- [13] MATLAB, Reference Guide, the MATHWORKS Inc., Natick, (1995)
- [14] MATLAB, Statistics Toolbox User's Guide, the MATHWORKS Inc., Natick, (1996)
- [15] A. Biran, M. Breiner, MATLAB for Engineers, Addison-Wesley, Wokingham, (1995)
- [16] W.E. Deming, Statistical Adjustment of Data, J. Wiley, New York, (1944)
- [17] L. Lomozik, M. Jaskólski, A. Wojciechowska, Polish J. Chem., 65, 1797, (1991)
- [18] L. Lomozik, Chem. Anal. (Warsaw), 40, 953, (1995)
- [19] MATLAB, Using MatLab Graphics Version 5, the MATHWORKS Inc., Natick, (1996)

Hoofdstuk 7

Samenvatting en algemeen besluit

Het onderzoek waarover in dit werk gerapporteerd wordt, concentreert zich op het gebruik van numerieke methoden voor de studie van de foutenpropagatie in diverse toepassingen en in het bijzonder in evenwichtsberekeningen.

In het eerste gedeelte van het proefschrift worden volumetrische, calorimetrische en kinetische data geanalyseerd, waarbij we aantonen hoe de techniek van *linearisering* toelaat om benaderende waarden te vinden voor de fout op relatief ingewikkelde functies f van onnauwkeurig gekende grootheden x_1, \ldots, x_n .

Computersimulaties vormen daarbij een moderne en zinvolle aanvulling voor de klassieke "experiment-theorie" benadering: uit een vergelijking van de statistische foutenpropagatie aan de hand van de standaardformules met de resultaten bekomen via simulaties blijkt inderdaad dat de lineaire foutenpropagatieregels enkel gelden wanneer de functie benaderd kan worden door een lineaire functie binnen het interval $[-3\sigma, 3\sigma]$. Er zijn relatief eenvoudige voorbeelden uit de chemische praktijk, zoals in de kinetische toepassing, waarbij systematische trends aan het licht komen en de benaderingen dus niet meer gerechtvaardigd zijn.

In het tweede gedeelte worden geautomatiseerde potentiometrische experimenten gedetailleerd besproken.

De karakteristieken van optimalisatieprogramma's voor het berekenen van stabiliteitsconstanten op basis van niet-lineaire regressie – in het bijzonder SUPERQUAD – worden aangetoond in een experimentele potentiometrische studie van de complexering van Ag(I) met een aantal lange-alifatische-keten diaminen in 1 M KNO_3 . We gebruiken hiervoor de verhoudingsmethode waarbij de onderzochte metaal/ligand verhoudingen variëren van 2/1 tot 1/8. In de literatuur zijn slechts een aantal indicatieve waarden
voor de vormingsconstanten van dergelijke evenwichten voorhanden. In eerste instantie worden de binaire systemen geanalyseerd, waarbij de operationele pK_W , de zuurconstanten van 1,4-diaminobutaan en 1,5-diaminopentaan en de zilverhydrolyseconstanten worden bepaald.

De zuurconstanten berekend in dit werk, passen mooi in de literatuurtendensen: pK_{A_1} stijgt, maar naarmate de ketenlengte n in het molecule $H_2N - (CH_2)_n - NH_2$ toeneemt, daalt het increment. De pK_{A_2} -waarden vertonen dit nog duidelijker. De ΔpK_A -waarde, een maat voor de elektrostatische interactie tussen de terminale ammoniumgroepen, neemt snel af met stijgend aantal methyleengroepen.

De zilverhydrolyseconstanten komen goed overeen met de literatuur, tenminste wanneer rekening wordt gehouden met de overeenkomstige pK_W .

Het complexeringsgedrag van Ag(I) met vernoemde liganden kan verklaard worden door de vorming van zowel niet geprotoneerde mono- en polynucleaire complexen als door de vorming van geprotoneerde mononucleaire species. De berekeningen geven geen evidentie voor het bestaan van gemengde ligand-hydroxy-complexen.

Om vervolgens de statistische onzekerheid op de geoptimaliseerde stabiliteitsconstanten te kwantificeren, zijn we, vanwege de complexiteit van de bestudeerde systemen genoodzaakt terug te vallen op simulaties. Om die reden wordt het algoritme van SU-PERQUAD (standaard geprogrammeerd in Fortran 77) geïmplementeerd in MatLabTM. De computationele methode, alsook de MatLabTM routines voor het genereren van de data worden uitvoerig beschreven. Een validiteitsstudie van deze implementatie die vertrekt van synthetisch gegenereerde (geperturbeerde en niet-geperturbeerde) data afkomstig van artificiële modelsystemen, wijst uit dat de geoptimaliseerde parameterwaarden in beide versies goed overeenkomen. Bovendien convergeren de geoptimaliseerde constanten in beide programmaversies naar hun "echte" waarde in het modelsysteem. Een analyse aan de hand van de experimentele data afkomstig van de bestudeerde Ag(I)-complexeringsevenwichten toont nogmaals de degelijkheid van het MatLabTM programma aan: een analoog "beste model" wordt geselecteerd. Tenslotte leren we uit deze studie dat indien divergentie optreedt er iets mis is met de modelkeuze of dat er een inconsistentie in de data aanwezig is.

In het laatste deel maken we met behulp van de ontwikkelde $MatLab^{TM}$ programmatuur gebruik van een Monte Carlo techniek die ons in eerste instantie toelaat de effecten van *experimentele* perturbaties, al dan niet systematisch van aard, te kwantifi-

Hoofdstuk 7 Samenvatting en algemeen besluit

ceren. Deze perturbaties worden gemodelleerd door middel van gaussiaans verdeelde stochasten. De sensitiviteit van de geoptimaliseerde parameters voor de precisie van de input wordt nagegaan aan de hand van een afleesfout op de pH, perturbaties in het initieel volume in de recipiënt en perturbaties in het toegevoegd volume en in de totale concentratie van de componenten aanwezig in het model. De simulatiestudie laat toe te voorspellen wat de grootteorde van de te verwachten fout op de geoptimaliseerde constante is, gegeven een gekende fout in één van de hogervernoemde inputparameters. Er zijn indicaties dat afleesfouten in belangrijke mate bepalend zijn voor de uiteindelijke precisie van de geoptimaliseerde constante in de geanalyseerde modelsystemen. Vastgesteld wordt bovendien, dat de kans dat een systematische fout geïntroduceerd wordt bij het uitvoeren van te weinig experimenten (één of twee), bijzonder groot is.

De betrouwbaarheid van de berekeningen of de eigenlijke optimalisatieprocedure, wordt anderzijds geëvalueerd via *structurele* fouten in gekende, artificiële modelsystemen. Analyse van "stabiele" en relatief eenvoudige modellen toont aan dat een waarde van de 'sample standaardafwijking', $1 \le S \le 3$ als criterium voor modelselectie volstaat. Een controle van het concentratieverloop van de afzonderlijke componenten en de aanwezige species, alsook inspectie van de verdeling en de autocorrelatie van de residuelen zijn aan te bevelen.

De analyseresultaten van complexe modelsystemen daarentegen moeten met de nodige voorzichtigheid geïnterpreteerd worden. Het verwerpen van "overbodige" species in het model lijkt op het eerste zicht geen problemen te geven: de S-waarde is zeer hoog $(\gg 3)$.

Zo ook kan bij het opsporen van "ontbrekende" deeltjes in het model, een afwijking van de lineariteit in de normaliteitsplot en een hoge autocorrelatielengte, systematische afwijkingen in de data aan het licht brengen, waarvoor de berekeningen niet kunnen compenseren.

Worden echter gecombineerde modellen van het startmodel geoptimaliseerd, dan kan een S-waarde beneden de 3 misleidend zijn voor het aanvaarden van een verkeerd model! In dit geval is het van belang dat zowel de S-waarde, als de residuelenverdeling, de autocorrelatie in de residuelen en het concentratieverloop van alle aanwezige species gecontroleerd worden alvorens men de beslissing neemt het model te aanvaarden. Verificatie van de verdeling en de autocorrelatie der residuelen afkomstig van de optimalisatie van de bestudeerde Ag(I)-complexeringsevenwichten, toont tenslotte aan dat de χ^2 -test het meest controversiële en minst overtuigende criterium is voor model-

selectie: een hoge waarde voor $\chi^2~(>12.60)$ wordt meestal genegeerd als de S-waarde gunstig is.

In dit werk worden een aantal nieuwe en bijkomende criteria voor modelselectie geformuleerd; daarbij stellen we vast dat het niet zozeer de beperkingen van de computationele methode zijn, maar wel de kwaliteit van het experiment, die bepalend is om een correct beeld van een reëel systeem te verkrijgen. De betrouwbaarheid van de geoptimaliseerde parameters kan waarschijnlijk significant verhoogd worden wanneer naast de aangehaalde statistische tests nog onafhankelijke meettechnieken betrokken worden. Hier bestaat echter het gevaar dat men "andere" evenwichtsconstanten bepaalt, zoals in het geval van bijvoorbeeld asymmetrisch gesubstitueerde diaminen: de spectrofotometrisch bepaalde zuurconstante is verschillend van de potentiometrisch bepaalde. Aangezien we nu een idee hebben over de manier waarop fouten in de input propageren naar het eindresultaat, zijn de planning van het experiment en de implementatie van "global optimization" technieken voor de bepaling van stabiliteitsconstanten, beloftevolle wegen die zeker moeten bewandeld worden...

Deel IV Appendices



Appendix A Relevante begrippen en notaties uit de kanstheorie en de statistiek

A.1	Stochastische variabele						
A.2	Verwachting en variantie van stochastische variabelen						
A.3	Covariantie en onafhankelijkheid van stochastische variabelen						
A.4	Functies van stochastische variabelen						
	A.4.1 Rekenregels voor verwachtingen						
	A.4.2 Rekenregels voor varianties						
A.5	Het normaal - kwantiel - diagram, normaliteitsdiagram of 'normal prob-						
	ability plot'						
A.6	De ongelijkheid van Jensen						
A.7	De Centrale Limietstelling						
A.8	Multivariate lineaire regressie						
A.9	Verwachtingswaarde van de som der kwadratische residuelen						
	Referenties						

Appendix B Partiële afgeleiden van de totale naar de vrije concentratie van de componenten berekend in de MatLabTM functie 'jacobian_free.m'

Appendix C Partiële afgeleiden van de totale concentratie naar de parameters berekend in de $MatLab^{TM}$ functie 'jacobian_beta.m'



Appendix A

Relevante begrippen en notaties uit de kanstheorie en de statistiek

A.1 Stochastische variabele

Bij herhaalde metingen van éénzelfde grootheid in een experimenteel opzet vindt men doorgaans telkens een ietwat andere uitkomst. De gebruikelijke verklaring hiervoor is dat de eigenlijke ideale waarde in elk experiment een beetje gecontamineerd wordt door toevallige, experimentele afwijkingen. Hierdoor levert elk individueel experiment een ander resultaat op.

Om dit te modelleren voert men in de wiskunde het begrip *kans-variabele* (random variable) of *stochastische variabele* in. Hiermee bedoelt men een afbeelding:

$$X: \Omega \to I\!\!R \tag{A.1}$$

met $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ en $X(\omega_1) = x_1, X(\omega_2) = x_2$ enz. waarbij x_1, x_2, \dots de respectievelijke waarden van de variabele $X(\omega_1), X(\omega_2), \dots$ voorstellen.

De idee hierbij is dat de verzameling Ω alle mogelijke experimenten voorstelt. Elk experiment komt dus overeen met de selectie ("at random") van een ω_i uit Ω , en de meting zelf levert de waarde $x_i = X(\omega_i)$. Omdat het op voorhand onmogelijk te voorspellen is welke ω_i zal geselecteerd worden, is binnen dit model niet zozeer de waarde x_i belangrijk, maar eerder de verdeling van deze waarden.

Intuïtief (zij het enigszins onnauwkeurig¹) is het het gemakkelijkst om over een stochastische variabele X te denken als over een kolom van waarden (figuur A.1): het i^{de} element in de kolom stelt een waarde $x_i = X(\omega_i)$ voor. Wat van belang is, is niet zozeer de individuele waarde van elke x_i maar vooral wat de frequentie van voorkomen van elke

¹We verwaarlozen hierbij immers het verschil tussen aftelbare en overaftelbare verzamelingen

 x_i is. Deze informatie ligt vervat in de zogenaamde kansverdeling of dichtheidsfunctie van X die specificeert welke de frequentie van voorkomen is van elk individueel resultaat x_i in de kolom X.

Om de verdere discussie duidelijk te houden dienen we een onderscheid te maken tussen discrete en continue stochastische variabelen [1], [2]. In het eerste geval zijn slechts een aantal mogelijke waarden toegelaten (bijvoorbeeld het aantal β -deeltjes dat gedurende een seconde wordt uitgezonden door een radioactief staal), terwijl voor continue variabelen alle waarden tussen twee grenzen zijn toegelaten (bijvoorbeeld de pH van een bepaalde oplossing).



Figuur A.1: De verdeling van x_i in twee onafhankelijke stochastische variabelen met X_2 een permutatie van X_1

Discrete stochastische variabele

Een discrete stochastische variabele X neemt een discreet (eindig of oneindig) aantal waarden $x_1, x_2, \ldots, x_k, \ldots$ aan. Een kansmodel voor X wordt gegeven door aan deze uitkomsten kansen p_i toe te kennen:

$$P(X = x_i) = p_i \tag{A.2}$$

De kansen p_i moeten voldoen aan:

- 1. $0 \le p_i \le 1$ voor elke i
- 2. $\sum p_i = 1$

De kans $P(X \in A)$ op een willekeurige gebeurtenis wordt gevonden door sommatie van de kansen p_i van de uitkomsten x_i waaruit de gebeurtenis A is samengesteld:

$$P(X \in A) = \sum_{x_i \in A} P(X = x_i) = \sum_{x_i \in A} p_i$$
(A.3)

Continue stochastische variabele

Een continue stochastische variabele X neemt alle waarden aan in een interval van reële getallen. Een kansmodel voor X wordt gekarakteriseerd door de dichtheidsfunctie p(x) die voldoet aan (figuur A.2):

1.
$$p(x) \ge 0$$
 voor alle x.

2. De totale oppervlakte onder de grafiek van p(x) is gelijk aan 1: $\int_{-\infty}^{+\infty} p(x) dx = 1$

3.





Het best gekende en meest belangrijke voorbeeld van een continue stochastische variabele is de (1-dimensionale) **normale** of **gaussiaanse** verdeling waarvan de kansverdeling wordt gegeven door:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/\sigma^2}$$

met μ , het gemiddelde of verwachting en σ , de standaardafwijking (paragraaf A.2).

213

(A.4)

Onafhankelijke stochasten

Statistische analyse van een dataset is meestal gebaseerd op het concept van een *steekproef.* Hieronder verstaat men dat men een uitspraak probeert te doen over de ganse dataset op basis van een kleine deelverzameling (de zogenaamde steekproef). De voornaamste eigenschap van een steekproef is dat ze representatief is voor de ganse dataset, m.a.w. de opmetingen in de steekproef hebben alle dezelfde kansverdeling als de oorspronkelijke dataset en zijn bovendien onafhankelijk van elkaar. Ware dit laatste niet het geval dan kan men zich voorstellen dat de eerste meting de tweede beïnvloedt waardoor de steekproef niet langer representatief voor de ganse dataset zou zijn.

Dit brengt ons bij het begrip *identisch verdeelde onafhankelijke stochastische variabelen* (zie verder paragraaf A.3). In termen van onze kolom-metafoor voor stochastische variabelen is dit gemakkelijk te interpreteren. Immers, bij een stochast X is de verdeling van de uitkomsten x_i belangrijk (intuïtief het aantal x_i in kolom X_1 (figuur A.1)). Het is dan ook evident dat een tweede stochast X_2 (figuur A.1) dezelfde verdeling heeft als X_1 indien X_2 een permutatie is van X_1 . Nochtans zijn X_1 en X_2 duidelijk verschillend aangezien:

$$X_1(\omega_1) = x_1 \qquad \qquad X_2(\omega_1) = x_3$$

en $x_1 \neq x_3$

Ze zijn bovendien onafhankelijk omdat kennis over de waarde x_i in de eerste kolom X_1 ons niets nieuws vertelt over de overeenkomstige waarde in de tweede kolom X_2 . Merk op dat X_1 en X_2 afhankelijk zouden zijn indien bijvoorbeeld telkens bij een bepaalde waarde x_i in de kolom X_1 een bepaalde waarde x_j in X_2 zou voorkomen.

Model voor een steekproef

Het voorgaande is ook van toepassing op meer ingewikkelde opzetten. Gewoonlijk stopt men een experiment wanneer men bijvoorbeeld drie waarden zijnde x_1 , x_2 en x_3 verzameld heeft, maar men kan zich voorstellen dat de procedure (onder dezelfde experimentele omstandigheden) herhaald wordt. Het resultaat van de eerste metingen van de zogenaamde tweede reeks wordt genoteerd als $x_1^{(2)}$ en op dezelfde wijze bekomt men $x_2^{(2)}$ en $x_3^{(2)}$.

Met de nodige tijd en middelen zou een toegewijd onderzoeker de procedure een "oneindig" maal kunnen herhalen, wat zou resulteren in drie kolommen, X_1 , X_2 , X_3 met in elke kolom de resultaten van respectievelijk de eerste, de tweede en de derde

opmeting van elke reeks (figuur A.3), met de X_i 's onafhankelijke identisch verdeelde stochastische variabelen [3], [4].

Voor elke reeks is het gemiddelde \overline{x} van drie metingen:

$$\overline{x} = \frac{x_1 + x_2 + x_3}{3} \tag{A.5}$$

en het resultaat wordt opgeslagen in een aparte kolom, \overline{X} , het zogenaamde steekproefgemiddelde.

x_1	X2	<i>X</i> ₃	• • •	Xn	x	
x, ⁽¹⁾	X2 ⁽¹⁾	X3 ⁽¹⁾		x _n ⁽¹⁾	<i>x</i> ⁽¹⁾	een bepaalde steekproef
:	÷	:		:	:	
X1 ^(k)	X2 ^(k)	x ₃ ^(A)		x, ^(A)	x ⁽⁴⁾	een andere steekproef

Figuur A.3: Getabelleerde waarden voor $X_j^{(i)}$.

Bij de meeste experimenten kent X een gaussiaans of normaal verloop. Indien men dus een oneindig gedetailleerd histogram van een oneindig aantal individuele meetresultaten maakt, bekomt men de volgende dichtheidsfunctie (figuur A.4), die wordt beschreven door:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$
(A.6)



Figuur A.4: De normale waarschijnlijkheidsverdeling

De waarschijnlijkheid dat X een waarde aanneemt tussen x en x + dx is f(x)dx:

$$P(x \le X \le x + dx) = f(x)dx \tag{A.7}$$

waarbij f(x) moet worden geïnterpreteerd als een soort van genormaliseerd "oneindig" gedetailleerd histogram.

De dichtheidsfunctie van de normale wordt volledig bepaald door de parameters μ en σ^2 , respectievelijk de verwachting, E(X) en de variantie, Var(X) (zie verder).

Er treedt een systematische fout op in de meting indien het gemiddelde $E(X) = \mu$ niet samenvalt met de parameter (waarde), ξ waarin men geïnteresseerd is.

Het opsporen van systematische fouten is helemaal niet triviaal omdat ξ uiteraard niet gekend is: men kan enkel hopen zulk een discrepantie te ontdekken door zowel de resultaten als de procedure zeer kritisch op consistentie te onderzoeken.

A.2 Verwachting en variantie van stochastische variabelen

De verwachting van een stochastische variabele, genoteerd met E(X) (E van Expectation) of ook μ_X wordt gedefinieerd als:

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x p(x) dx \quad \text{(continue stochastische variabele)}$$
(A.8)

 $\mu_X = E(X) = \sum x_k p_k$ (discrete stochastische variabele)

De verwachtingswaarde E(X) is bijgevolg het gemiddelde over de mogelijke waarden van X, waarbij elke mogelijke uitkomst gewogen wordt met zijn kans van voorkomen [1], [2], [5]. Merk op dat in de intuïtieve voorstelling van X als een (oneindig lange) kolom van de uitkomsten, E(X) zou corresponderen met het "gewone" rekenkundige gemiddelde (d.w.z. de uitkomsten zonder weging). Dit is natuurlijk een gevolg van het feit dat uitkomsten x met een hoge kans p(x) ook proportioneel meer in de kolom X voorkomen, waardoor het rekenkundig gemiddelde automatisch juist gewogen wordt. Voorgaande interpretatie laat ons ook onmiddellijk toe deze formule uit te breiden. Onderstel immers dat we een tweede stochastische variabele Y definiëren die een functie is van de eerste:

Y = g(X)

In termen van de "kolom-representatie" betekent dit dat men de Y-kolom krijgt door in de X-kolom elke getal x te vervangen door g(x). Er volgt dan onmiddellijk dat:

$$E(Y) \equiv E(g(X)) = \int g(x)p(x) \, dx$$

(met een analoge voorstelling voor een discrete stochast).

Als een eenvoudige toepassing van dit algemeen principe kijken we naar de definitie van *variantie*. Daar waar de verwachting een maat is voor het centrum van een verdeling, geeft de variantie een maat voor de *kwadratische spreiding* rond het gemiddelde. Voor elke uitkomst kijken we naar de kwadratische afwijking ten opzichte van het gemiddelde en daarvan berekenen we de verwachte waarde (genoteerd met Var(X) of ook σ_X^2):

$$\operatorname{Var}(X) = E[(X - E(X))^2] = \int (x - \mu_X)^2 p(x) dx \quad \text{(continue stoch. variabele)}$$

$$\operatorname{Var}(X) = E[(X - E(X))^2] = \sum_k (x_k - \mu_X)^2 p_k \quad \text{(discrete stoch. variabele)}$$
(A.9)

Merken we op dat als y = Ax met A een $n \times k$ -matrix dan geldt:

$$E(y) = AE(x) \tag{A.10}$$

of:

$$Var(y) = AVar(x)A^{T}$$
(A.11)

A.3 Covariantie en onafhankelijkheid van stochastische variabelen

Als men twee numerieke grootheden gemeten heeft kan men zich afvragen in hoeverre variatie van de ene grootheid samengaat met verandering in de andere grootheid. Om te meten hoe twee variabelen X en Y samen variëren vertrekt men van X - E(X) en Y - E(Y). De covariantie van X en Y wordt dan uitgedrukt als:

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

= $E[(X - \mu_X)(Y - \mu_Y)]$ (A.12)

Wanneer we de covariantie standaardiseren ten opzichte van de varianties spreekt men van correlatie:

$$\operatorname{Cor}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$
(A.13)

Verder is de volgende definitie belangrijk:

Definitie 1 De stochastische variabelen X en Y noemt men onafhankelijk indien

$$p_{XY}(x,y) = p_X(x)p_Y(y)$$

Deze definitie is echter moeilijk te controleren, vandaar dat verwezen wordt naar een verwant begrip:

$$\operatorname{Cov}(X,Y) = 0 \tag{A.14}$$

Het omgekeerde is echter *niet* waar tenzij X en Y bovendien *normaal verdeeld* zijn. We vatten alles nog eens samen in de volgende propositie:

ZijX en Ytwee stochastische variabelen, dan geldt er

- indien X en Y onafhankelijk zijn dat Cov(X, Y) = 0;
- indien Cov(X, Y) = 0 en X en Y normaal verdeeld zijn, dat ze ook onafhankelijk zijn.

A.4 Functies van stochastische variabelen

Zij X_1, X_2, \ldots, X_n een aantal stochastische variabelen dan kan men voor een willekeurige functie $Y = f(X_1, X_2, \ldots, X_n)$ de gebruikelijke karakteristieken zoals de verwachtingswaarde enz. berekenen. Meer in het bijzonder hebben we:

i) Y = f(X) dat

$$E(Y) = E(f(X)) = \sum_{x} f(x)p(x)$$
 (A.15)

ii) $Y = f(X_1, X_2)$ dat

$$E(Y) = E(f(X_1, X_2)) = \sum_{x_1} \sum_{x_2} f(x_1, x_2) p(x_1, x_2)$$
(A.16)

(Een gelijkaardige uitdrukking kan men schrijven voor de integralen.) Onmiddellijk rijst natuurlijk de vraag of we de karakteristieken van Y rechtstreeks uit die van X kunnen berekenen. Voor lineaire functies gaat dit exact, voor niet-lineaire slechts bij benadering. Dit is het onderwerp van de volgende paragraaf.

A.4.1 Rekenregels voor verwachtingen [1], [2]

Lineaire functies

Indien X en Y twee willekeurige stochastische variabelen zijn geldt, aangezien de verwachtingswaarde E een lineaire operator is, dat voor de gewogen som:

$$E(aX + bY) = aE(X) + bE(Y)$$
(A.17)

of meer algemeen:

$$E(\sum a_i X_i) = \sum a_i E(X_i) \tag{A.18}$$

Als toepassing van de algemene uitdrukking vermelden we het steekproefgemiddelde, $\overline{X} = 1/n \sum_{i} X_{i}$. Aangezien \overline{X} een lineaire combinatie is van stochastische variabelen is het zelf een stochastische variabele, zodat volgens vergelijking (A.18)

$$E(\overline{X}) = \frac{1}{n} [E(X_1) + E(X_2) + \ldots + E(X_n)]$$
(A.19)

en vermits elke waarneming X beantwoordt aan de dichtheidsverdeling p(x)met gemiddelde μ of

$$E(X_1) = E(X_2) = \dots = \mu$$
 (A.20)

geldt

$$E(\overline{X}) = \frac{1}{n}[\mu + \mu + \ldots + \mu] = \frac{1}{n}[n\mu] = \mu$$
 (A.21)

Een bewijs voor vergelijking (A.17) vinden we in de literatuur [1].

Algemene functies

Voor een willekeurige functie Y = f(X) weten we dat in het algemeen

 $E(Y) \neq f(E(X))$. Dit is wellicht het eenvoudigst in te zien als we denken aan de standaard normale $X \sim N(0, 1)$ waarvoor E(X) = 0 omdat de verdeling perfect symmetrisch rond 0 is. Het is echter duidelijk dat $Y = X^2$ enkel maar positieve waarden kan aannemen en dus onmogelijk een gemiddelde gelijk aan nul kan hebben: E(Y) > 0. In het kader van foutenanalyse waar de variantie dikwijls veel kleiner is dan het gemiddelde, geldt echter de onderstaande nuttige benadering. Hiervoor drukken we de stochastische variabele X uit als een variatie ten opzichte van zijn gemiddelde $\xi = E(X)$:

$$X = \xi + (X - \xi)$$

waarbij we veronderstellen dat de afwijking van de waarneming X t.o.v. van de echte waarde ξ klein is. Dit laat ons toe om de functie Y = f(X) te ontwikkelen rond het punt ξ zodat:

$$Y = f(\xi + (X - \xi))$$

= $f(\xi) + f'(\xi)(X - \xi) + \frac{1}{2}f''(\xi)(X - \xi)^2 + \dots$ (A.22)

en bijgevolg

$$E(Y) = f(\xi) + f(\xi)E(X - \xi) + \frac{1}{2}f''(\xi)E(X - E(X))^2 + \dots$$

$$= f(\xi) + f(\xi)(E(X) - \xi) + \frac{1}{2}f''(\xi)\operatorname{Var}(X) + \dots$$
(A.23)

of bij benadering:

$$E(Y) = f(\xi) + \frac{1}{2}f''(\xi)\operatorname{Var}(X)$$
(A.24)

zodat

$$E(f(X)) \approx f(E(X)) + \frac{1}{2}f''(\xi)\operatorname{Var}(X)$$
 (A.25)

Vergelijking (A.25) wordt gecontroleerd aan de hand van een simulatie in MatLabTM (versie 4.2c.1 voor Macintosh) [6], [7], [8] van de functie $Y = e^X$ met X random en $\sim N(0, \sigma^2)$, waarbij σ respectievelijk de waarde 0.05; 0.1; 0.15 en 0.2 aanneemt. Voor elke waarde van σ worden 10000 experimenten uitgevoerd.

De berekende waarde voor E(Y) aan de hand van vergelijking (A.24) met $\sigma^2 = 0.01$ geeft:

$$E(Y) = 1 + \frac{1}{2}0.01 = 1.005$$

In figuur A.5 is het verloop van $(1 + \frac{1}{2}\sigma^2)$ uitgezet in functie van σ . De simulatie geeft als resultaat:

E(Y) = 1.0015 voor $\sigma = 0.05$; E(Y) = 1.0053 voor $\sigma = 0.1$; E(Y) = 1.0108 voor $\sigma = 0.15$ en E(Y) = 1.0220 voor $\sigma = 0.2$.

Deze resultaten – aangeduid met • – komen goed overeen met het theoretisch verloop van E(Y) (figuur A.5), zodat vergelijking (A.25) gerechtvaardigd is.



Figuur A.5: E(Y) versus σ

A.4.2 Rekenregels voor varianties

Ook hier maken we de opdeling in *lineaire* functies waarvoor er exacte resultaten bekend zijn en *algemene* functies waarvoor slechts benaderingen bestaan [1], [2].

Lineaire functies

Indien X en Y twee willekeurige stochastische variabelen zijn, geldt algemeen:

$$\operatorname{Var}(aX + bY) = a^{2}\operatorname{Var}(X) + b^{2}\operatorname{Var}(Y) + 2ab\operatorname{Cov}(X, Y)$$
(A.26)

Voor een bewijs van vergelijking (A.26) verwijzen we naar de literatuur [1]. In het bijzonder, indien X en Y onafhankelijke stochastische variabelen zijn, geldt:

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$
(A.27)

Voor de algemene lineaire uitdrukking volgt:

$$\operatorname{Var}(\sum a_i X_i) = \sum a_i^2 \operatorname{Var}(X_i) \tag{A.28}$$

met als toepassing het steekproefgemiddelde, \overline{X} :

$$\operatorname{Var}(\overline{X}) = \frac{\operatorname{Var}(X)}{n} = \frac{\sigma^2}{n}$$
 (A.29)

Willekeurige functies

Stel dat men geïnteresseerd is in een functie f(x, y) van twee grootheden x en y waarvan men een aantal meetresultaten, respectievelijk x_1, \ldots, x_n en y_1, \ldots, y_n heeft verkregen en die men modelleert met de stochastische variabelen X en Y.

Hierbij mag weer aangenomen worden dat de afwijking van de waarneming X (respectievelijk Y) t.o.v. de gemiddelde waarde ξ (respectievelijk η) klein is, zodat volgens de stelling van Taylor [9]:

$$f(X,Y) = f(\xi + (X - \xi), \eta + (Y - \eta))$$

$$= f(\xi,\eta) + \left(\frac{\partial f}{\partial x}\right)\Big|_{(\xi,\eta)} (X - \xi) + \left(\frac{\partial f}{\partial y}\right)\Big|_{(\xi,\eta)} (Y - \eta) + \dots$$
(A.30)

Verwaarlozen van hogere orde termen en gebruik makend van vergelijking (A.26) geeft dit voor de variantie op de functie:

$$\operatorname{Var} f(X,Y) = \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var}(X) + \left(\frac{\partial f}{\partial y}\right)^2 \operatorname{Var}(Y) + 2\left(\frac{\partial f}{\partial x}\right) \left(\frac{\partial f}{\partial y}\right) \operatorname{Cov}(X,Y)$$
(A.31)

met volgende elementaire veronderstellingen:

- · $\operatorname{Var} f(\xi, \eta) = 0$ omdat ξ en η constant zijn
- · $Var(X \xi) = Var(X)$ en $Var(Y \eta) = Var(Y)$ omdat *translatie* van de data geen invloed heeft op de variantie.

In de meeste gevallen zijn daarenboven de metingen X en Y onafhankelijk van elkaar, wat voor gevolg heeft dat de covariantieterm Cov(X, Y) verdwijnt en vergelijking (A.31) kan geschreven worden als:

$$\operatorname{Var} f(X, Y) = \left(\frac{\partial f}{\partial x}\right)^2 \operatorname{Var}(X) + \left(\frac{\partial f}{\partial y}\right)^2 \operatorname{Var}(Y) \tag{A.32}$$

Illustratief tegenvoorbeeld Een voorbeeld waar de covariantieterm niet nul is komt voor wanneer men de variantie op het nettogewicht (N) wil kennen vertrekkende van brutogewicht (B) en tarragewicht (T).

$$B = T + N \tag{A.33}$$

Stel

$$\operatorname{Var}(B) = \operatorname{Var}(T) + \operatorname{Var}(N) + \underbrace{2\operatorname{Cov}(T, N)}_{=0}$$
(A.34)

zodat

$$Var(B) = Var(T) + Var(N)$$
(A.35)

en

$$Var(N) = Var(B) - Var(T)$$
(A.36)

Hier dient echter een onderscheid te worden gemaakt in de wijze waarop het experiment is opgezet: T en N zijn slechts onafhankelijk (Cov(T, N) = 0) indien men in een lege recipiënt met gekend gewicht het nettogewicht toevoegt (cfr. het absoluut wegen). T en N zijn daarentegen afhankelijk indien het gewicht van de lege recipiënt (T) getarreerd wordt en daarna de netto toegevoegde hoeveelheid wordt bepaald (cfr. wegen met tarreren). Het is instructief op te merken dat we het resultaat (A.36) ook op een andere manier hadden kunnen bekomen:

$$N = B - T \tag{A.37}$$

$$Var(N) = Var(B) + Var(T) - 2Cov(B,T)$$
(A.38)

Indien B, T onafhankelijk zouden zijn dan Cov(B, T) = 0 en dus

$$Var(N) = Var(B) + Var(T)$$
(A.39)

Dit resultaat is niet correct, een bewijs hiervoor vertrekt van:

$$Cov(B,T) = Cov(T + N,T)$$

= Cov(T,T) + Cov(N,T)
= Var(T) (A.40)

Dus:

$$Var(N) = Var(B - T)$$

= Var(B) + Var(T) - 2Cov(B,T)
= Var(B) + Var(T) - 2Var(T)
Var(N) = Var(B) - Var(T) (A.41)

Dit resultaat stemt wel overeen met vergelijking (A.36)

A.5 Het normaal - kwantiel - diagram, normaliteitsdiagram of 'normal probability plot'

De beslissing om een verdeling door een normaal model te beschrijven kan bepalend zijn voor de verdere stappen in de analyse van data.

Een histogram kan kenmerken van een verdeling vertonen die duidelijk niet-normaal zijn, zoals bijvoorbeeld uitschieters of uitgesproken scheefheid.

Een preciezere beoordeling van het passend zijn van het normale model voor een gegevensverzameling wordt verschaft door een normaal - kwantiel - diagram. Kwantielen zijn percentielen² die als fracties zijn uitgedrukt in plaats van als percentages.

De grondidee voor het maken van een normaal - kwantiel - diagram bestaat erin elke

²De spreiding of variabiliteit van een verdeling kan worden aangeduid door verschillende percentielen te geven. Het *p*-de percentiel van een verdeling is de waarde, zodanig dat *p* procent van de waarnemingen lager uitvalt of eraan gelijk is. Zo is bijvoorbeeld de mediaan het 50ste percentiel of het 0.50-kwantiel van een verdeling. De mediaan geeft het centrum van een verdeling aan.

waarneming x uit te zetten tegen het corresponderende kwantiel van de standaardnormale verdeling³. In MatLabTM [6], [7] bevat de "normal probability plot" drie elementen:

- het +-teken geeft de empirische waarschijnlijkheid weer versus de data voor elk punt van de dataset;
- de volle lijn verbindt het 25ste en 75ste percentiel van de data volgens een robuste lineaire fit, zodat deze niet gevoelig is voor extreme waarden uit de dataset;
- de stippellijn is een extrapolatie naar de uiteinden toe.

De schaal van de y-as is niet uniform; de waarden geven kansen weer van 0 tot 1. De afstand tussen twee markeringen op de y-as komt overeen met de afstand tussen de kwantielen van een normale verdeling. Deze kwantielen liggen dichtbij elkaar in de buurt van de mediaan (kans = 0.5) en strekken zich symmetrisch uit weg van de mediaan. Vandaar dat de vertikale afstand van de onderzijde van de grafiek tot aan kans 0.25 groter is dan de afstand tussen 0.25 en 0.50, zo ook is de afstand van de bovenkant van de grafiek tot 0.75 groter dan de afstand van 0.75 tot 0.50.

Als de punten van een normaal - kwantiel - diagram dicht bij een rechte liggen, dan geeft de figuur een aanwijzing voor het normaal zijn van de data. Systematische afwijkingen van de rechte lijn wijzen in de richting van een niet-normale verdeling. Uitschieters verschijnen als punten die ver verwijderd zijn van het globale patroon van de figuur. In figuur A.6 (a-e) zijn vijf normaal - kwantiel - diagrammen weergegeven (de data xzijn horizontaal getekend, de corresponderende waarschijnlijkheid ('probability') is vertikaal getekend) met hun overeenkomstige verdelingen. Figuur A.6 (a) is een normaal - kwantiel - diagram afkomstig van een symmetrische normale verdeling. De opname van spoorelementen in functie van het gezondheidsniveau bijvoorbeeld volgt een symmetrisch verloopt.

Het merendeel van de punten ligt dicht bij een rechte lijn, hetgeen erop wijst dat een normaal model goed past.

³Alle normale verdelingen zijn identiek als de metingen worden gedaan in eenheden σ rondom de verwachting. Deze heten gestandaardiseerde metingen. Als $X \sim N(\mu, \sigma)$, dan heeft de gestandaardiseerde variabele $Z = \frac{(X - \mu)}{\sigma}$ de standaardnormale verdeling N(0, 1). Dit is een toepassing van de rekenregels voor R = a + bX.

Een verdeling kan ook niet-symmetrisch zijn; men spreekt dan van een scheve verdeling. Een dergelijke vorm van verdeling kan scheef naar rechts of positief scheef zijn (figuur A.6(b)i). De distributiecurve van de moleculaire kinetische energie in een gas, de zogenaamde Boltzmann verdeling is rechtsscheef. De verdeling heeft een lange, dunne staart naar rechts, de asymmetrie neemt toe bij stijgende temperatuur. In een positief scheve verdeling zal het gemiddelde altijd groter zijn dan de modus⁴ en de mediaan.

De grootste waarnemingen in het normaal - kwantiel - diagram vallen duidelijk aan de rechterzijde van een lijn die door de hoofdmassa van punten is getrokken (figuur A.6(b)ii).

Een verdeling kan ook scheef naar links of negatief scheef zijn. In een dergelijke verdeling is het gemiddelde kleiner dan de mediaan en de modus (figuur A.6(c)i)).

Linksscheefheid is evident als de kleinste waarnemingen aan de linkerzijde van de lijn vallen die door de hoofdmassa van de punten loopt (figuur A.6(c)ii)).

Figuren A.6(d) en (e) geven respectievelijk een gepiekte [10] en een uniforme verdeling weer met hun overeenkomstig normaliteitsdiagram.



Figuur A.6(a): *i) Symmetrische normale verdeling ii) Normaal - kwantiel - diagram van een symmetrische normale verdeling*

⁴De modus is de meest voorkomende uitslag in een reeks waarnemingen.







Figuur A.6(c): i) Negatief scheve verdeling ii) Normaal - kwantiel - diagram van een linksscheve verdeling



Figuur A.6(e): i) Uniforme verdeling ii) Normaal - kwantiel - diagram van een uniforme verdeling

A.6 De ongelijkheid van Jensen

Een functie $f : \mathbb{R} \to \mathbb{R}$ is convex indien $\forall x, y \in \mathbb{R}, \forall \lambda \in [0, 1]$ geldt [11]:

$$f(\lambda x + (1 - \lambda)y) \le \lambda f(x) + (1 - \lambda)f(y) \tag{A.42}$$

m.a.w. voor elke twee punten x en y ligt de koorde tussen x en y volledig boven de grafiek van f. Nog anders uitgedrukt: voor elk punt x is er een rechte door (x, f(x)) zodat de grafiek van f volledig boven de rechte ligt (zie figuur A.7). Voor een convexe functie geldt eveneens:

$$f(E(X)) \le Ef(X) \tag{A.43}$$



Figuur A.7: Een convexe functie

Indien een convexe functie f tweemaal differentieerbaar is geldt steeds

$$f''(x) \ge 0 \tag{A.44}$$

en volgens vergelijking (A.25):

$$Ef(X) = f(E(X)) + \frac{1}{2}f''(x)\operatorname{Var}(X) + \dots$$
 (A.45)

indien f convex is en $Var(X) \ge 0$ zodat

$$Ef(X) \ge f(E(X)) \tag{A.46}$$



Figuur A.8: Een concave functie

Voor een concave functie f geldt omgekeerd

$$f(E(X)) \ge Ef(X) \tag{A.47}$$

aangezien f is concaaf $\Leftrightarrow -f$ is convex, m.a.w. de koorde ligt steeds onder de functie (figuur A.8).

A.7 De Centrale Limietstelling

Er is in deze appendix al meermaals op gewezen dat de normale verdeling allerlei speciale eigenschappen bezit die in het algemeen door geen van de andere mogelijke kansverdelingen gedeeld worden. Zo weten we ondermeer dat een *lineaire combinatie* van onafhankelijke normale stochastische variabelen eveneens normaal is. Meer bepaald: zij X_1, X_2, \ldots, X_n n onafhankelijke normaal verdeelde variabelen met gemiddelde μ_i en variantie σ_i^2 ($X_i \sim N(\mu_i, \sigma_i^2)$) dan geldt:

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right) \tag{A.48}$$

Wanneer echter X_1, X_2, \ldots, X_n willekeurige stochastische variabelen zijn met gemiddelde μ en variantie σ^2 , dan weten we nog altijd wat het gemiddelde en de variantie van de som is aangezien wegens vergelijking (A.18) en vergelijking (A.28) geldt dat

$$\left(\sum_{i=1}^{n} X_i\right) \sim N(n\mu, n\sigma^2) \tag{A.49}$$

maar deze som zal niet langer normaal verdeeld zijn. De Centrale Limietstelling garandeert ons echter wel dat wanneer het aantal termen n maar voldoende groot is $(n \rightarrow \infty)$ er bij benadering geldt dat de som normaal verdeeld is [2], [12], dus:

$$\sum_{i=1}^{n} X_{i} \sim N(n\mu, n\sigma^{2}) \qquad \text{bij benadering voor } n \longrightarrow \infty.$$

Een belangrijke toepassing hiervan heeft te maken met de verdeling van het steekproefgemiddelde:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(met $E(X_i) = \mu$ en $Var(X_i) = \sigma^2$) waarvoor we kunnen besluiten dat ongeacht de oorspronkelijke verdeling van de variabelen X_i de verdeling van het steekproefgemiddelde \overline{X} bij benadering normaal gaat zijn met volgende karakteristieken:

 $\overline{X} \sim N(\mu, \sigma^2/n)$ (bij benadering voor $n \longrightarrow \infty$).

Een bewijs van deze belangrijke stelling zou ons te ver voeren maar we hebben gepoogd om dit resultaat te illustreren aan de hand van een MatLabTM [6], [7] simulatie (zie figuur A.9 (a-d)).

Ondanks het feit dat X_i exponentieel verdeeld is (zie normaliteitsdiagram figuur A.9(a)) wordt door sommatie van n exponentiëlen toch een quasi normale verdeling bekomen bij n = 20 (normaliteitsdiagram figuur A.9(d)). Figuren A.9(b) en (c) tonen normaliteitsdiagrammen voor respectievelijk n = 3 en n = 10. Merk hierbij op dat n tamelijk groot moet zijn vooraleer de verdeling normaal wordt omdat vertrokken wordt van een exponentiële verdeling.



Figuur A.9: Normaliteits diagram voor a) X_i exponentieel verdeelde stochast b) n = 3





d)



A.8 Multivariate lineaire regressie

De statistische methode om uitgaande van een verzameling van waarden van onafhankelijke variabelen x, waarden te voorspellen voor één of meerdere afhankelijke variabelen y, noemt men regressieanalyse.

Het klassieke model voor multivariate lineaire regressie kan voor n onafhankelijke waarnemingen als volgt beschreven worden (matrix notatie):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{A.50}$$

met $E(\varepsilon) = 0$ en $Cov(\varepsilon) = \sigma^2 I$. I is de eenheidsmatrix, β en σ^2 zijn de onbekende parameters, X is de *design* matrix.

Om de waarden voor y te kunnen voorspellen is het dus nodig het model (A.50) te *fitten* aan de hand van de geobserveerde y_j corresponderend met de gekende waarden $1, x_{j1}, \ldots, x_{jr},$ m.a.w. de regressiecoëfficiënten β en de variantie σ^2 van de fout moeten bepaald worden.

De kleinste kwadraten methode selecteert nu die waarde voor β , verder $\hat{\beta}$ genoemd, zodanig dat de gewogen som der kwadratische residuelen $y_j - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \ldots - \hat{\beta}_r x_{jr}$ = $\hat{\epsilon}$ minimaal is:

$$S(\hat{\boldsymbol{\beta}}) = \sum_{j=1}^{n} W_{j}(y_{j} - \hat{\beta}_{0} - \hat{\beta}_{1}x_{j1} - \dots - \hat{\beta}_{r}x_{jr})^{2}$$

$$= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{T}\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= \mathbf{y}^{T}\mathbf{W}\mathbf{y} - \mathbf{y}^{T}\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}}$$

$$= \mathbf{y}^{T}\mathbf{W}\mathbf{y} - 2\hat{\boldsymbol{\beta}}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{y} + \hat{\boldsymbol{\beta}}^{T}\mathbf{X}^{T}\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}}$$

(A.51)

Het minimum van S vinden we door $S(\hat{\beta})$ af te leiden naar $\hat{\beta}$:

$$\frac{\partial S}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}^T \mathbf{W} \mathbf{y} + 2\mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}$$
(A.52)

Vandaar $\frac{\partial S}{\partial \hat{\boldsymbol{\beta}}} = 0$ indien $\mathbf{X}^T \mathbf{W} \mathbf{y} = \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}$. Of:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{W} \mathbf{y})$$
(A.53)

A.9 Verwachtingswaarde van de som der kwadratische residuelen

Verder is volgens vergelijking (A.11) voor $\mathbf{W} = \mathbf{I}, E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \operatorname{Var}(\mathbf{X}^T \mathbf{y}) \left((\mathbf{X}^T \mathbf{X})^{-1} \right)^T$$
(A.54)

waarbij $\operatorname{Var}(\mathbf{X}^T \mathbf{y}) = \mathbf{X}^T \operatorname{Var}(\mathbf{y}) \mathbf{X} = \mathbf{X}^T \sigma^2 \mathbf{I} \mathbf{X} = \sigma^2 (\mathbf{X}^T \mathbf{X})$ zodat:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2 (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$
(A.55)

A.9 Verwachtingswaarde van de som der kwadratische residuelen

Uit de vorige paragraaf volgt:

 $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$

met $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}.$

Verder is volgens (A.53), waarbij we voor de eenvoud W = I beschouwen:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{I} - \mathbf{H}) \mathbf{y}$$
(A.56)

met $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, de "hat" matrix.

Noteer dat

$$HH = X(X^{T}X)^{-1}X^{T}X(X^{T}X)^{-1}X^{T}$$

= X(X^TX)⁻¹X^T
= H (A.57)

H is dus idempotent en symmetrisch. Vandaar:

$$\hat{\boldsymbol{\varepsilon}}^{T} \hat{\boldsymbol{\varepsilon}} = \mathbf{y}^{T} (\mathbf{I} - \mathbf{H})^{T} (\mathbf{I} - \mathbf{H}) \mathbf{y}$$

$$= \mathbf{y}^{T} (\mathbf{I} - 2\mathbf{H} + \mathbf{H}^{2}) \mathbf{y}$$

$$= \mathbf{y}^{T} (\mathbf{I} - \mathbf{H}) \mathbf{y}$$
 (A.58)

Om nu de verwachtingswaarde, $E(\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}})$ te berekenen gaan we als volgt te werk:

$$\hat{\boldsymbol{\varepsilon}}^{T}\hat{\boldsymbol{\varepsilon}} = (\hat{\boldsymbol{\varepsilon}}^{T} + \hat{\boldsymbol{\beta}}^{T} \mathbf{X}^{T})(\mathbf{I} - \mathbf{H})(\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}})
= \hat{\boldsymbol{\beta}}^{T} \mathbf{X}^{T}(\mathbf{I} - \mathbf{H})\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}}^{T}(\mathbf{I} - \mathbf{H})\hat{\boldsymbol{\varepsilon}} + \hat{\boldsymbol{\varepsilon}}^{T}(\mathbf{I} - \mathbf{H})\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}^{T} \mathbf{X}^{T}(\mathbf{I} - \mathbf{H})\hat{\boldsymbol{\varepsilon}}$$
(A.59)

(i) De eerste term verdwijnt aangezien:

$$\hat{\boldsymbol{\beta}}^T \mathbf{X}^T (\mathbf{I} - \mathbf{H}) \mathbf{X} \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^T [\mathbf{X}^T (\mathbf{I} - \mathbf{H}) \mathbf{X}] \hat{\boldsymbol{\beta}}$$

met

$$\begin{aligned} \mathbf{X}^{T}(\mathbf{I} - \mathbf{H})\mathbf{X} &= \mathbf{X}^{T}\mathbf{X} - \mathbf{X}^{T}\mathbf{H}\mathbf{X} \\ &= \mathbf{X}^{T}\mathbf{X} - \mathbf{X}^{T}(\mathbf{X}(\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T})\mathbf{X} \\ &= \mathbf{X}^{T}\mathbf{X} - \mathbf{X}^{T}\mathbf{X} = \mathbf{0} \end{aligned}$$

(ii) Ook de term in $\hat{\boldsymbol{\varepsilon}}^T$ uit (A.59) verdwijnt, omdat:

$$\begin{split} \hat{\boldsymbol{\varepsilon}}^T (\mathbf{I} - \mathbf{H}) \mathbf{X} \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{\varepsilon}}^T \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\varepsilon}}^T \mathbf{H} \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= \hat{\boldsymbol{\varepsilon}}^T \mathbf{X} \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\varepsilon}}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} \\ &= 0 \end{split}$$

Om dezelfde reden verdwijnt de term in $\hat{\pmb{\varepsilon}}$

(iii) Hierdoor blijft enkel de tweede term van (A.59) over: herinner dat $\mathbf{H}^2 = \mathbf{H}$ en \mathbf{H} symmetrisch, zodat zowel \mathbf{H} als $\mathbf{I} - \mathbf{H}$ orthogonale projecties zijn.

Er bestaat dus een orthogonale matrix \mathbf{U} en een diagonaal matrix \mathbf{D} van de vorm

$$\mathbf{D} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \qquad (1 \le k \le n)$$

zodat

$$\mathbf{P} = (\mathbf{I} - \mathbf{H}) = \mathbf{U}^T \mathbf{D} \mathbf{U}$$

Bijgevolg:

$$\hat{\boldsymbol{\varepsilon}}^{T}(\mathbf{I} - \mathbf{H})\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}}^{T}\mathbf{U}^{T}\mathbf{D}\mathbf{U}\hat{\boldsymbol{\varepsilon}}$$

= $\boldsymbol{\eta}^{T}\mathbf{D}\boldsymbol{\eta}$ (A.60)

met $\eta = U\hat{\varepsilon}$ en normaal verdeeld. Bovendien

$$E(\boldsymbol{\eta}) = \mathbf{U}E(\hat{\boldsymbol{\varepsilon}}) = 0$$
$$\operatorname{Var}(\boldsymbol{\eta}) = \mathbf{U}\operatorname{Var}(\hat{\boldsymbol{\varepsilon}})\mathbf{U}^{T} = \sigma^{2}\mathbf{I}$$

zodat $\boldsymbol{\eta} \sim N(0, \sigma^2 \mathbf{I}).$

Omwille van (A.60) kunnen we schrijven:

$$\hat{\boldsymbol{\varepsilon}}^{T}(\mathbf{I}-\mathbf{H})\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\eta}^{T} \left(egin{array}{cc} \mathbf{I}_{k} & 0 \\ 0 & 0 \end{array}
ight) \boldsymbol{\eta} = \sum_{i=1}^{k} \boldsymbol{\eta}_{i}^{2}$$

en dus zien we dat het linkerlid een som is van k
 onafhankelijke normaal verdeelde $(N(0, \sigma^2))$ stochastische variabelen, zo
dat

$$E\left(\sum_{i=1}^{k} \boldsymbol{\eta}_{i}^{2}\right) = \sum_{i=1}^{k} E(\boldsymbol{\eta}_{i}^{2}) = \sum_{i=1}^{k} \operatorname{Var}(\boldsymbol{\eta}_{i}) = k\sigma^{2}$$
(A.61)

met

$$k = rk D (rk = rang)$$
$$= n - m$$

aantal datapunten min aantal geschatte parameters.

Verder definiëren we:

$$S^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-m} \tag{A.62}$$

met S de 'sample standaardafwijking'. Zodat uit vergelijking (A.61) volgt:

$$E(S^2) = \sigma^2 \tag{A.63}$$

en bijgevolg is S^2 een schatter voor σ^2 .

Een schatting voor $\operatorname{Var}(\hat{\boldsymbol{\beta}})$ wordt dus gegeven door:

$$\operatorname{Var}(\hat{\boldsymbol{\beta}}) = S^2 (\mathbf{X}^T \mathbf{X})^{-1} = \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{n-m} (\mathbf{X}^T \mathbf{X})^{-1}$$
(A.64)

Referenties

- T.H. Wonnacott, R.J. Wonnacott, Introductory Statistics for Business and Economics, Chapter 4, John Wiley & Sons Inc., New York, (1990)
- [2] D.S. Moore, G.P. McCabe, Statistick in de Praktijk, hoofdstuk 4, 2e herziene druk, Academic Service, Schoonhoven, (1997)
- [3] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 776, (1993)
- [4] W.J. Guedens, J. Yperman, J. Mullens, L.C. Van Poucke, E.J. Pauwels, J. Chem. Educ., 70, 838, (1993)
- [5] M.L.
 Wijvekate, Verklarende Statistiek, hoofdstuk 8, Het Spectrum, Utrecht/A'pen, (1976)
- [6] MATLAB, Reference Guide, the MATHWORKS Inc., Natick, (1995)
- [7] MATLAB, Statistics Toolbox User's Guide, the MATHWORKS Inc., Natick, (1996)
- [8] A. Biran, M. Breiner, MATLAB for Engineers, Addison-Wesley, Wokingham, (1995)
- [9] T.M. Apostol, Calculus, 2 nd Ed., Wiley and Sons, New York, (1966)
- [10] A. Slotboom, Statistiek in woorden, p. 127, Wolters-Noordhoff, Groningen, (1987)
- [11] A. Mood, F. Graybill, D. Boes, Introduction to the Theory of Statistics, Chapter 4, 3rd Ed., Mc Graw-Hill, New York, (1974)
- [12] J. Klaessens, H. Van Leeuwen, Praktische statistiek voor het laboratorium, hoofdstuk 6, ten Hagen & Stam, Den Haag, (1996)

Appendix B

Partiële afgeleiden van de totale naar de vrije concentratie van de componenten berekend in de MatLabTM functie 'jacobian_free.m'

De MatLabTM functie 'jacobian_free.m' genereert de partiële afgeleiden van de totale concentratie naar de vrije concentratie van de componenten voor een willekeurig gekozen model. Volledigheidshalve wordt de berekening van de massabalansen voor het metaalion M, het ligand L en het proton $H(T_M, T_L \text{ en } T_H)$ in functie van de te optimaliseren parameters en de vrije concentraties voor een willekeurig model weergegeven. K(i) stelt de stabiliteitsconstante van het *i*-de complex in het betrokken systeem voor, free(k) de vrije concentratie van de k-de component en P de modelmatrix.

n = length(K); T_M = free(1); for i = 1:n T_M = T_M + K(i)*P(1,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*free(3)^(P(3,i)); end

T_M;

T_L = free(2);
 for i=1:n
```
T_L = T_L + K(i)*P(2,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*free(3)^(P(3,i
end
T_L;
T_H = free(3);
    for i=1:n
    T_H = T_H + K(i)*P(3,i)*free(1)^{(P(1,i))*free(2)^{(P(2,i))*free(3)^{(P(3,i)})}
end
T_H;
Hierna volgt de berekening van het linkerlid van het stelsel (4.32):
function [F] = jacobian_free(free,K,P)
n = length(K);
   dT_MM = 1;
    for i=1:n
    dT_MM = dT_MM + K(i)*P(1,i)*P(1,i)*free(1)^(P(1,i)-1)*free(2)^(P(2,i))*
             free(3)^(P(3,i));
end
dT_MM;
    dT_ML = 0;
    for i=1:n
    dT_ML = dT_ML + K(i)*P(1,i)*free(1)^(P(1,i))*P(2,i)*free(2)^(P(2,i)-1)*
             free(3)^(P(3,i));
end
dT_ML;
    dT_MH = 0;
    for i=1:n
    dT_MH = dT_MH + K(i)*P(1,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*P(3,i)*
```

free(3)^(P(3,i)-1);

end

	Appendix B	241
dT_MH;		
$dT_LM = 0;$		
for i=1:n		
$dT_LM = dT_LM + K($	i)*P(2,i)*P(1,i)*free(1)^(P(1,i)-	1)*free(2)^(P(2,i))*
free(3)^(P	(3,i));	
end		
dT_LM;		
$dT_LL = 1;$		
for i=1:n		
$dT_LL = dT_LL + K($	i)*P(2,i)*free(1)^(P(1,i))*P(2,i)	*free(2)^(P(2,i)-1)*
free(3)^(P	(3,i));	
end		
dT_LL;		
$dT_LH = 0;$		
for i=1:n		
$dT_LH = dT_LH + K($	i)*P(2,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*P(3,i)*
free(3)^(P	(3,i)-1);	
end		
dT_LH;		
$dT_HM = 0;$		
for i= 1:n		
$dT_HM = dT_HM + K($	i)*P(3,i)*P(1,i)*free(1)^(P(1,i)-	1)*free(2)^(P(2,i))*
free(3)^(P	(3,i));	
and		
IT_HM;		
$dT_HL = 0;$		
for i= 1:n		
$dT_HL = dT_HL + K($	i)*P(3,i)*free(1)^(P(1,i))*P(2,i)	*free(2)^(P(2,i)-1)*
free(3)^(P	(3,i));	
and		

```
dT_HL;
```

242

end

dT_HH;

```
F = [ free(1)*dT_MM free(2)*dT_ML free(3)*dT_MH;
    free(1)*dT_LM free(2)*dT_LL free(3)*dT_LH;
    free(1)*dT_HM free(2)*dT_HL free(3)*dT_HH];
```

De output van de functie 'jacobian_free.m' komt overeen met de matrix \mathbf{F} in vergelijking (4.45).

Appendix C

Partiële afgeleiden van de totale concentratie naar de parameters berekend in de MatLabTM functie 'jacobian_beta.m'

De MatLabTM functie 'jacobian_beta.m' genereert de partiële afgeleiden van de totale concentratie naar de te optimaliseren stabiliteitsconstanten voor een willekeurig gekozen model. De berekening van de massabalansen $(T_M, T_L \text{ en } T_H)$ is analoog met Appendix B. K(i) stelt de stabiliteitsconstante van het *i*-de complex in het betrokken systeem voor, free(k) de vrije concentratie van de *k*-de component en *P* de modelmatrix.

Hierna volgt de berekening van het rechterlid van het stelsel (4.32):

```
function [G] = jacobian_beta(free,K,P)
n = length(K);
    dT_MK = zeros(1,length(K));
    dT_LK = zeros(1,length(K));
    dT_HK = zeros(1,length(K));
    for i=1:n
    dT_MK(i) = P(1,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*free(3)^(P(3,i));
end
```

```
for i =1:n
dT_LK(i) = P(2,i)*free(1)^(P(1,i))*free(2)^(P(2,i))*free(3)^(P(3,i));
```

end

```
for i=1:n
```

 $dT_HK(i) = P(3,i)*free(1)^{(P(1,i))*free(2)^{(P(2,i))*free(3)^{(P(3,i))}};$ end

G = [K.*dT_MK; K.*dT_LK; K.*dT_HK];

De output van de functie 'jacobian_beta.m' komt overeen met de matrix G in vergelijking (4.45).



Academische loopbaan

1980: Kandidaat in de Wetenschappen - groep Scheikunde, Limburgs Universitair Centrum, Diepenbeek

1982: Licentiaat in de Wetenschappen - groep Scheikunde, richting Biochemie, Universitaire Instelling Antwerpen

1983: Lector aan de Koninklijke Militaire School, Brussel

1984: Geaggregeerde Hoger Secundair Onderwijs, Katholieke Universiteit Leuven

1984 tot heden: didactisch navorser, Vakgroep Scheikunde, Faculteiten Wetenschappen en Geneeskunde met volgend profiel:

- onderwijsopdrachten in het eerste kandidaatsjaar Scheikunde-Biologie en Arts, en in het tweede kandidaatsjaar Scheikunde
- ontwikkeling van nieuwe practica, demonstraties, computerondersteunde lessen en cursusteksten
- medewerking/participatie aan interuniversitaire projecten: IPTO (1985-1987) en ISSAD (1986-1989)
- medewerking aan projectaanvragen in samenwerking met Hogescholen: STIHO (1997, 1998, 1999)
- begeleiding van eindwerken Industrieel Ingenieur
- overgangsproblematiek Secundair Onderwijs Hoger Onderwijs: remediëring
- coördinatie van promotieactiviteiten, dienstverlening en bijscholingsactiviteiten voor leerkrachten natuurwetenschappen S.O.
- 1991: deeltijds lector Biochemie a.i., Provinciaal Instituut voor Verpleegkunde Hasselt
- 1994 tot heden: ombudsvrouw voor het eerste kandidaatsjaar Scheikunde-Biologie
- 1995-1997: secretaris van het departement Scheikunde-Biologie-Geologie

Publicaties

 "Photon and fluorescence correlation spectroscopy and light scattering of eye lens proteins at moderate concentrations", W.J. Guedens et al., *Biophysical Journal*, 43, 345, (1983)

- "Automatisatie van metingen in het didactisch lab scheikunde", W.J. Guedens, M. Reynders, D. Brouwers, J. Yperman, J. Mullens, E. Flerackers en L.C. Van Poucke, Proceedings Euregio Conference: Naturwissenschaften und Unterricht band 18, Enschede, (1992)
- "Statistical Analysis of Errors: A Practical Approach for an undergraduate Chemistry Lab- Part I", W.J. Guedens, J. Yperman, J. Mullens, E.J. Pauwels and L.C. Van Poucke, J. Chem. Educ., 70, 776-779, (1993)
- "Statistical Analysis of Errors: A Practical Approach for an undergraduate Chemistry Lab- Part II", W.J. Guedens, J. Yperman, J. Mullens, E.J. Pauwels and L.C. Van Poucke, J. Chem. Educ., 70, 838-841, (1993)
- "Computeranimatie en simulatie in de anorganische chemie", W.J. Guedens, L. Cuppers, J. Yperman, J. Mullens en L.C. Van Poucke, NVOX, 9, 447-450, (1994)
- "Silver(I) Complex Formation with 1,4-diaminobutane and 1,5-diaminopentane: A Potentiometric Study in 1.0 M KNO₃", W.J. Guedens, J. Yperman, J. Mullens and L.C. Van Poucke, J. Coord. Chem., 47, 17-29, (1999)

Posters

- "Chemistry in the daily Life: A Movable Exhibition to Bring Chemistry to the Non-Scientific Citizin", W. J. Guedens, E. Onkelinx and J. Put, Chemie en Materialen, KNCV-VCV, Maastricht, (1989)
- "Protonerings-en Ag(I)-Complexeringsevenwichten van 1,4-diaminobutaan en 1,5diaminopentaan", W.J. Guedens, J. Yperman, J. Mullens en L.C. Van Poucke, 4de Vlaams Jongeren Congres van de Chemie, Leuven, (1998)

Deelname aan congressen

- "De Computer in het Chemieonderwijs", Brussel, (1986)
- "Computers in de Chemie", Gent, (1987)
- "Contactgroep Research in Wetenschappelijk Onderwijs (CRWO)", Nijmegen, (1987)

- "Computers: biomedische, farmaceutische toepassingen in research, onderwijs en praktijk", Antwerpen, (1986 en 1988)
- "Onderwijs Research Dagen", Leuven, (1988)
- "40ste Congres van Natuurwetenschappen, Hasselt LUC Diepenbeek, (1988)
- "Chemie en Materialen, KNCV-VCV", Maastricht, (1989)
- "International Scientific Film Association Congres", Utrecht, (1990)
- "Euregio Conferentie Chemiedidactiek", Enschede, (1992)
- "Seminaries: Leerstoel ererector Prof. Dr. L. Verhaegen", LUC Diepenbeek, (1994, 1995, 1996, 1997, 1998)
- "1st Chemometrics Symposium", Solvay, Brussel, (1995)
- "Symposium Chemometrics in Belgium", KVCV-SRC, UIA, (1995)
- "2nd Chemometrics Symposium", Leuven, (1996)
- "Symposium 25 jaar Werkgroep en Navorming Natuurwetenschappen, LUC, Diepenbeek, (1996)
- "Online Educa Berlin International Conference on Technology and Distance Learning", Berlijn, (1996)
- "Derde Vlaams Forum voor Onderwijsonderzoek", LUC Diepenbeek, (1998)
- "4de Vlaams Jongeren Congres van de Chemie", Leuven, (1998)
- "4th Chemometrics Symposium, Tervuren, (1998)
- "1st European Conference in Chemical Education", Budapest, (1998)

Mededelingen op congressen

 "De toets als vorm van individuele studiebegeleiding", L. Cuppers, W.J. Guedens, M. Maelstaf, J. Mullens, M. Thoelen, J. Yperman en L.C. Van Poucke, CRWOstudiedag, Nijmegen, (1987) • "De computer in het chemieonderwijs", W.J. Guedens, 40ste Congres van Natuurwetenschappen, LUC Diepenbeek, (1988)

.

- "De microcomputer: animator en simulator voor practica chemie", W.J. Guedens, L. Cuppers, H. Meuwis, J. Mullens, J. Yperman en L.C. Van Poucke, Chemieonderwijs en vakdidactiek in de Euregio, Rijkhoven-LUC Diepenbeek, (1989)
- "Automatisatie van potentiometrische, spectrofotometrische en conductiviteitsmetingen in het didactisch lab scheikunde", W.J. Guedens, M. Reynders, D. Brouwers, J. Yperman, J. Mullens, E. Flerackers en L.C. Van Poucke, Euregio Conference: Current Research in Chemical Education, Enschede, (1992)
- "Geautomatiseerde metingen in het didactisch lab scheikunde", W.J. Guedens, Symposium 25 jaar Werkgroep en Navorming Natuurwetenschappen, LUC Diepenbeek, (1996)
- "Integrating LabViewTM Software in Classical Chemistry Laboratory Experiments", W.J. Guedens, J. Yperman, J. Mullens and L.C. Van Poucke, 1st European Conference in Chemical Education, Budapest, (1998)

248

Aantekeningen



"To err is human, to forgive divine-but to include errors in your design is statistical." Leslie Kish

