

Onderwerpen Masterproef Opleiding Wiskunde

Academiejaar 2022-2023

Vakgroep Toegepaste Wiskunde, Informatica en Statistiek

Titel: Studie van q-reeksen en hun toepassingen.

Promotor: prof. Joris Van der Jeugt (Joris.VanderJeugt@UGent.be)

Korte beschrijving:

Een studie van zogenaamde q-reeksen gebeurt meestal in de context van hypergeometrische reeksen. Nochtans kunnen q-reeksen ook onafhankelijk ingevoerd worden, bijvoorbeeld als genererende functies van bepaalde combinatorische grootheden, van partities, enz. Meestal ziet men q-reeksen als een onderdeel van "Speciale functies", en wordt de nadruk gelegd op analytische eigenschappen. Maar q-reeksen treden op in heel veel verschillende wiskundige contexten of toepassingen: in getaltheorie, combinatoriek, computeralgebra, wiskundige fysica,...

Voor deze scriptie wordt van de student verwacht dat hij/zij zich enkele standaardtechnieken van q-reeksen eigen maakt. Vervolgens wordt de keuze gelaten om zich verder te verdiepen in theoretische resultaten (vb. orthogonale polynomen geassocieerd met q-reeksen), of om enkele toepassingen in andere domeinen te bestuderen.

Referentiewerken:

- q-Special functions, a tutorial, T.H. Koornwinder.
- Lecture Notes for an introductory minicourse on q-series, G. Gasper.
- q-Series: their development and application in analysis, number theory, combinatorics, physics and computer algebra, G.E. Andrews.
- A q-product tutorial for a q-series maple package, F. Garvan.

Voor: Studenten uit de master Wiskunde.

Titel: Matrices met alternerende tekens en toepassingen.

Promotor: prof. Joris Van der Jeugt (Joris.VanderJeugt@UGent.be)

Korte beschrijving:

Matrices met alternerende tekens (alternating sign matrices) zijn vierkante matrices met elementen uit $\{0,1,-1\}$ zodanig dat elke rij- en kolomsum 1 is, en zodanig dat de niet-nul elementen in elke rij en kolom van teken wisselen. Over het aantal A_n van zulke matrices van orde $(n \times n)$ werden er in de jaren 1980 een aantal conjecturen gepubliceerd. Al snel bleek dat dit schijnbaar eenvoudig probleem verbanden vertoont met vraagstukken in andere wiskundedomeinen: partities, roosterpaden en vlakpartities, symmetrische functies, hypergeometrische reeksen, ...; maar ook met vraagstukken uit statistische mechanica zoals "vierkant ijs" (square ice). Het uiteindelijke bewijs van de zgn. alternating sign matrix conjecture werd pas in 1995 voltooid.

In het boek Proofs and confirmations wordt op een boeiende manier het verhaal achter dit bewijs verteld. De nadruk ligt op wiskundige experimenten (vb. met behulp van Maple), en op een kennismaking met de verschillende domeinen waarin matrices met alternerende tekens een rol spelen.

Voor deze scriptie wordt van de student verwacht dat hij/zij enkele van de wiskundige experimenten effectief uitvoert, en zich vervolgens verder verdiept in de theoretische resultaten die uiteindelijk hebben geleid tot een bewijs van de conjectuur.

Referentiewerk:

- Proofs and Confirmations: The story of the alternating sign matrix conjecture, D.M. Bressoud.

Voor: Studenten uit de master Wiskunde.

Titel: Afhankelijke kansvariabelen en hun som

Promotor: prof. David Vyncke (david.vyncke@ugent.be)

Korte beschrijving:

Het risico waaraan een financiële of verzekeringsportefeuille blootgesteld is, hangt nauw samen met de verdeling van een som van kansvariabelen. Indien de kansvariabelen onafhankelijk zijn, kan men die som schrijven als een convolutie, maar een dergelijke voorwaarde is in de praktijk zelden voldaan. Rekening houden met de reële afhankelijkheidsstructuur brengt echter heel wat problemen met zich mee. Arbenz et al (2011) ontwierpen een algoritme dat de verdeling snel zou moeten berekenen, maar dat algoritme is beperkt tot positieve kansvariabelen en vereist bovendien dat de volledige copula gekend is. In de praktijk is echter vaak slechts gedeeltelijke informatie over de afhankelijkheid bekend. Door gebruik te maken van dergelijke informatie is het mogelijk om de (verdeling van de) som te begrenzen, zoals geïllustreerd in Bernard et al (2017) en Lux & Papapantoleon (2019). In deze masterproef bestudeert de student verscheidene technieken om de verdeling van een som van afhankelijke kansvariabelen te berekenen en/of te begrenzen.

Referenties:

- Arbenz P., Embrechts P. & Puccetti G. (2011). The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. *Bernoulli* 17(2), 562–591.
 - Bernard C., Rüschemdorf L. & Vanduffel S. (2017). Value-at-risk bounds with variance constraints. *Journal of Risk and Insurance* 84, 923–959.
 - Lux T. & Papapantoleon A. (2019). Model-free bounds on Value-at-Risk using extreme value information and statistical distances. *Insurance: Mathematics and Economics* 86, 73-83.
-

Titel: Een ander onderwerp in de financiële of actuariële wiskunde

Promotor: prof. David Vyncke (david.vyncke@ugent.be)

Korte beschrijving:

Bespreekbaar

Hedgen van basisrisico met behulp van reinforcement learning

Promotor: Prof. dr. M. Vanmaele

Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving

Basisrisico wordt gedefinieerd als het inherente risico dat een handelaar neemt bij het afdekken van een positie door een tegengestelde positie te nemen in een derivaat van het actief, zoals een futurescontract. Dit basisrisico wordt aanvaard in een poging het prijsrisico af te dekken. Bijvoorbeeld, als de huidige spotprijs, of contante marktprijs, van goud 1190 EUR is, en de

prijs van goud in het goudtermijncontract in juni 1195 EUR bedraagt, is de basis, het verschil, 5,00 EUR. Het basisrisico is het risico dat de prijs van de futures niet beweegt in normale, gestage correlatie met de prijs van de onderliggende waarde en dat deze fluctuatie in de basis de effectiviteit van een afdekkingsstrategie (die wordt gebruikt om de blootstelling van een handelaar aan mogelijk verlies te minimaliseren) teniet kan doen. Het verschil tussen de contante prijs en de prijs van de futures kan ofwel groter of kleiner worden tussen het moment waarop een afdekkingspositie wordt ingenomen en het moment waarop die positie wordt gesloten. Er zijn verschillende types basisrisico te onderscheiden namelijk basisrisico verbonden aan de prijs, de locatie, de kalender of de kwaliteit van het product.

De bedoeling van deze masterscriptie is om hedgingstrategieën te bestuderen in de aanwezigheid van basisrisico en deze te vergelijken met een hedgingstrategie gebaseerd op reinforcement learning. Reinforcement learning (RL) is een gebied van machine learning dat zich bezighoudt met hoe intelligente actoren acties moeten ondernemen in een omgeving om een cumulatieve beloning te maximaliseren. Hiervoor kan er voortgebouwd worden op een masterproef van 2020-2021 waar een RL algoritme op basis van "deep deterministic policy gradient" is gebruikt.

Relevante literatuur vormen de volgende artikels:

- Monoyios, M. (2004). Performance of utility-based strategies for hedging basis risk. *Quantitative Finance*, 4(3), 245–255.
- Trottier, D. A., Godin, F., & Hamel, E. (2018). Local hedging of variable annuities in the presence of basis risk. *ASTIN Bulletin*, 48(2), 611–646.
- Axel F.A. Adam, Ingmar Nolte (2011). Cross hedging under multiplicative basis risk. *Journal of Banking & Finance*, 35:2956–2964.
- Broll, U., Welzel, P. & Wong, K. P. (2015) Futures hedging with basis risk and expectation dependence. *International Review of Economics*, 62(3):213–221.
- Sutton, R. & Barto, A.G. (2018) *Reinforcement learning: An introduction*, MIT press
- Xue, X., Zhang, J. & Weng, C. (2019) Mean-variance hedging with basis risk. *Applied Stochastic Models in Business and Industry*, 35:704–716

Risk-sharing en peer-to-peer verzekeringen

Promotor: Prof. dr. M. Vanmaele

Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving

Zoals in Dhaene (2022) te lezen staat, bij peer-to-peer (P2P) insurance wordt een risk-sharing netwerk opgezet door een groep gelijkgezinden met de bedoeling om hun risico's (tot op zekere hoogte) in een pool onder te brengen en zelf te beheren. Een voorbeeld van risk-sharing in de praktijk zijn netwerken van relatief kleine groepen van ondernemers die hun risico op arbeidsongeschiktheid poolen en delen. Risk-sharing mechanismen treffen we ook aan bij uitoefenaars van eenzelfde beroep die een pool vormen om zich te beschermen tegen professionele risico's. Ook catastrofes en grote industriële risico's worden soms gedekt door fondsen die opereren via de techniek van het risk-sharen.

De bedoeling van deze masterproef is de eigenschappen van deze verschillende risk-sharing regels te bestuderen aan de hand van de hierna volgende artikels.

- Denuit M., Dhaene J. (2012). 'Convex order and comonotonic conditional mean risk sharing', *Insurance: Mathematics and Economics*, 51(2), 265-270.
- Denuit M., Dhaene J., Robert C. (2021). 'Risk-sharing rules and their properties, with applications to peer-to-peer insurance', *LIDAM Discussion Paper ISBA 2021/37*, pp. 44.
- Denuit M., Robert, C.Y. (2021) 'From risk sharing to pure premium for a large number of heterogeneous losses', *Insurance: Mathematics and Economics*, 96, 116-126.
- Denuit M., Robert, C.Y. (2022). 'Collaborative Insurance with Stop-Loss Protection and Team Partitioning', *North American Actuarial Journal*, 26(1), 143-160.

- Dhaene J. (2022). 'Risk sharing en peer-to-peer verzekering', De Actuaris, AG, 29-3.

Stein's lemma en toepassingen in portefeuilleselectieproblemen

Promotor: Prof. dr. M. Vanmaele

Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving

Voor een bivariate normaalverdeelde vector $(X, Y)^T$, heeft Stein (1973, 1981) aangetoond dat
$$\text{Cov}[h(X), Y] = \text{Cov}[X, Y] E[h'(X)]$$

voor een willekeurige afleidbare functie h waarvoor $E[h'(X)]$ bestaat. Dit lemma heeft zijn nut bewezen in heel wat disciplines, zoals statistiek, waarschijnlijkheid, besliskunde en financiën. Rendementen van aandelen zijn echter niet altijd symmetrisch verdeeld, ze vertonen scheefheid. Deze observatie zette Adcock ertoe aan om in een serie van artikels een type van Stein's lemma op te stellen voor multivariate verdelingen die consistent zijn met Simaan's (1993) modelering van aandelenrendementen. Landsmann en co-auteurs hebben verdere uitbreidingen van Stein's lemma bewezen voor (multivariate) elliptisch verdeelde veranderlijken met toepassingen in o.a. risicotheorie (zie Landsmann & Valdez, 2016). Vanduffel en Yao (2017)

vertrekken van Simaan's setting en stellen een nieuw type van Stein's lemma op voor de multivariate veralgemeende hyperbolische verdeling. Als toepassing, beschouwen ze een portefeuilleselectievraagstuk.

De opdracht van deze masterproef bestaat in het bestuderen van Stein's lemma voor verschillende verdelingen en de toepassingen ervan in portefeuilleselectieproblemen en bij risicomaten.

Voor deze studie kan gesteund worden op de volgende artikels:

- C.J. Adcock (2014). Mean–variance–skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution. *European Journal of Operational Research*, 234:392–401
- Z. Landsmann and E. Valdez (2016). The tail Stein's identity with applications to risk measures. *North American Actuarial Journal*, 20(4): 313-326.
- Y. Simaan (1993). Portfolio selection and asset pricing-three parameter framework. *Management Science*, 39 (5):568–587.
- C.M. Stein (1973). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague symposium on asymptotic statistics* (pp. 345–381).
- C. M. Stein (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9:1135–1151.
- S. Vanduffel and J. Yao (2017). A Stein type lemma for the multivariate generalized hyperbolic distribution. *European Journal of Operational Research*, 261(2):606-612.

Titel: Een ander onderwerp in het domein van de financiële of actuariële wiskunde

Promotor: Prof. dr. M. Vanmaele

Doelgroep: studenten uit de tweede master Wiskunde

Bespreekbaar

Titel: Synthetische data-analyse

Promotor: prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Om redenen van confidentialiteit beloven meer en meer start-up bedrijven dat zij gevoelige data (bvb. uit ziekenhuizen) synthetisch kunnen maken. Dit wil zeggen dat de echte data gemodelleerd worden door iemand die toegang tot de data heeft (het zij via parametrische modellen, hetzij via neural networks), op basis waarvan nieuwe data gesimuleerd worden, dewelke vervolgens vrij aan onderzoekers beschikbaar kunnen worden gesteld voor een data-analyse. Dit doet vragen rijzen wat de kwaliteit is van de resulterende analyse, en hoe men de onnauwkeurigheid erop kan kwantificeren. In deze masterproef zullen we bestuderen hoe synthetische data aangemaakt worden (via zogenaamde generative adversarial networks) en zullen we vervolgens onderzoeken hoe men de onnauwkeurigheid (bvb standaard error) in een synthetische data-analyse kan bepalen. In samenwerking met het UZGent zullen we bovendien vergelijkingen maken tussen de analyse van een echte medische gegevensbank, en synthetische data bekomen op basis van die databank.

Titel: Factor analyse

Promotor: prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Factor analyse is een zeer populaire techniek om metingen te representeren die moeilijk te meten zijn, zoals intelligentie, levensvoldoening, ... Hiertoe worden meerdere metingen opgevraagd die iets vertellen over intelligentie, levensvoldoening, ... en die worden vervolgens op een slimme manier samengevat in een zogenaamde 'factor score'. In deze masterproef zullen we factor analyse bestuderen en vervolgens ingaan op recent werk waar aangetoond wordt dat dergelijke factor analyses problematisch kunnen zijn wanneer de factor score in verdere analyses verwerkt wordt. Hiertoe werden statistische testen ontwikkeld om na te gaan of het model achterliggend aan een factor analyse de data tegenspreekt. We zullen de performantie van deze testen onderzoeken in simulatiestudies en ze vervolgens toepassen op gegevens rond levensvoldoening bekomen voor bijna 13000 deelnemers aan de Health and Retirement Study. Afhankelijk van de voorkeur van de student kan een doelstelling er verder in bestaan om de voorgestelde testen uit te breiden om ze meer flexibel te maken.

Titel: Statistische besluitvorming na gebruik van machine learning technieken

Promotor: prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Meer en meer statistische analyses maken gebruik van machine learning technieken om bijvoorbeeld het effect van een behandeling of interventie te schatten. Zoals alle statistische analyses, genereren ook machine learning technieken onzekerheid op de resultaten (ten gevolge van steekproefvariatie); omwille van de complexiteit van de technieken wordt dit in de verdere analyse echter meestal genegeerd. Als gevolg hiervan zijn de betrouwbaarheidsintervallen die men bekomt, typisch zwaar vertekend. In deze masterproef zult u betrokken worden in recente ontwikkelingen, alsook onderzoek binnen de onderzoeksgroep van de promotor, om na te gaan hoe men hiermee kan omgaan. De bestudeerde methodes zullen theoretische en/of door middel van simulatiestudies en concrete data-analyses worden geëvalueerd.

Titel: Een ander onderwerp in het domein van de mathematische statistiek en/of toegepaste data-analyse

Promotor: prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Bespreekbaar

Title: Statistische besluitvorming na gebruik van machine learning technieken

Promotor: prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Meer en meer statistische analyses maken gebruik van machine learning technieken om bijvoorbeeld het effect van een behandeling of interventie te schatten. Zoals alle statistische analyses, genereren ook machine learning technieken onzekerheid op de resultaten (ten gevolge van steekproefvariatie); omwille van de complexiteit van de technieken wordt dit in de verdere analyse echter meestal genegeerd. Als gevolg hiervan zijn de betrouwbaarheidsintervallen die men bekomt, typisch zwaar vertekend. In deze masterproef zult u betrokken worden in recente ontwikkelingen, alsook onderzoek binnen de onderzoeksgroep van de promotor, om na te gaan hoe men hiermee kan omgaan. De bestudeerde methodes zullen theoretische en/of door middel van simulatiestudies en concrete data-analyses worden geëvalueerd.

Title: Algoritmen voor het Steiner Tree Problem

Promotor: prof. Veerle Fack (Veerle.Fack@ugent.be)

Korte beschrijving:

In een brede betekenis bestaat het doel van een Steiner Tree Problem erin om de goedkoopste manier te bepalen om een set van objecten te verbinden.

In de meest voorkomende varianten zijn deze objecten ofwel punten in een metrische ruimte ofwel een deelverzameling van de toppen van een netwerk/graaf, en het doel is het bepalen van een boom die ze allemaal verbindt.

Er zijn talloze toepassingen van deze problemen, zoals netwerkoptimalisatie, reconstructie van phylogenetische bomen, ontwerp van computercircuits, multicast-routing in communicatienetwerken.

Dit probleem was ook het onderwerp van een [DIMACS Implementation Challenge](#) in 2014, waarbij o.m.

een collectie benchmark data opgesteld werd, naast het uitwerken van gevorderde algoritmen voor het aanpakken van dit probleem.

Met deze Implementation Challenge als startpunt, is het de bedoeling van deze masterproef om bestaande algoritmen voor het probleem te bestuderen en te analyseren, evenals eigen benaderingen uit te werken en te toetsen tegenover de bestaande methodes

Title: Debaised machine learning for the average treatment effect on the treated under endogeneity.

Promotor: dr. Oliver Dukes (Oliver.Dukes@UGent.be) en prof. Stijn Vansteelandt (Stijn.Vansteelandt@UGent.be);

This project will consider inference for the average treatment effect on the treated (ATT) in a setting with a valid instrumental variable. This is a variable that has no direct effect on the outcome, is itself not subject to unmeasured confounding, and is predictive of the exposure. If an additional homogeneity condition holds, it can be used to infer the ATT even when the exposure-outcome relationship is distorted by unmeasured confounding. So far, most of the inferential procedures for the ATT have relied on a correctly specified (semi)parametric model. In this project, we will instead take a model-free perspective. We will begin by obtaining efficiency bounds under differing homogeneity conditions, and then understanding the asymptotic properties of corresponding 'debaised machine learning' estimators. The considered methods will be evaluated theoretically and/or by means of simulation studies and concrete data analyses.

Title: Predictive intervals for individual treatment effects based on machine learning

Promotor: dr. Oliver Dukes (Oliver.Dukes@UGent.be) en prof. Stijn Vansteelandt (Stijn.Vansteelandt@UGent.be);

A key task of personalized medicine is to learn about the effect of an intervention at the patient-specific level. Although it is not usually possible to identify or estimate individual treatment effects (ITEs) directly, one may nevertheless be able to construct prediction intervals for the ITE. This can be done by leveraging recent work on conformal prediction/inference, a framework that takes a prediction model and calibrates it to produce a prediction set with marginal coverage guarantees. An appeal of this framework is that the predictions can be obtained using flexible machine learning tools, as well as parametric regression methods.

In this project, we will compare two recent competing proposals for generating prediction intervals for the ITE, both in terms of theoretical properties as well as empirical performance via simulations.

Title: Het optimaliseren van klinische onderzoeken door het incorporeren van baseline variabelen

Promotor: dr. Kelly Van Lancker (kelly.vanlancker@ugent.be) en prof. Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Tijd is altijd cruciaal als het gaat om de evaluatie van potentiële geneesmiddelen en vaccins, maar de laatste jaren meer dan ooit met het oog op de COVID-19-pandemie. 'Covariate adjustment' (d.w.z. het in rekening brengen voor vooraf gespecificeerde, prognostische variabelen die gemeten worden vóór randomisatie) is een statistische methode met een groot potentieel om de precisie te verbeteren en de vereiste steekproefgrootte te verkleinen voor veel klinische studies. In de meeste klinische studies worden op het moment van rekrutering gegevens verzameld over 10-tallen tot zelfs meer dan 100 kenmerken van de patiënten. In studies waarin behandelingen voor COVID-19 worden vergeleken, kan men bijvoorbeeld informatie vastleggen over leeftijd, ernst van de ziekte en BMI. Dergelijke variabelen zijn inderdaad nuttig omdat ze informatie bevatten over de prognose voor de patiënt in zowel de behandelings- als de controle-arm. Hoewel 'covariate adjustment' wordt aanbevolen door de Amerikaanse Food and Drug Administration en het Europees Geneesmiddelenbureau, maken veel onderzoeken geen of beperkt gebruik van de beschikbare informatie in deze patiënt karakteristieken. Het verminderen van het aantal patiënten en daarmee het verminderen van de duur en kosten van deze onderzoeken is van cruciaal belang. Dit wordt ook benadrukt in een richtlijn die door de Amerikaanse Food and Drug Administration voor de industrie is vrijgegeven over 'covariate adjustment'.

Het doel van dit project is om het nut van 'covariate adjustment' in verschillende ziektegebieden aan te tonen door middel van echte datasets, dit kan zowel door data-analyse als door simulaties. Afhankelijk van de interesses van de student, kunnen we dieper ingaan op praktische of theoretische kwesties, dit omvat onder meer "hoe rekening te houden met de vermindering van de steekproefomvang bij het bepalen van de steekproefgrootte?", "hoe te bepalen welke variabelen leiden tot de meeste efficiëntiewinst?", ... Lopende samenwerkingen met de vakgroep Biostatistiek aan Johns Hopkins University, de Amerikaanse Food and Drug Administration en het farmaceutische bedrijf Novartis zal de bruikbaarheid van deze methoden in de praktijk ten goede komen.

Title: Een vergelijking van algoritmes voor de numerieke oplossing van de twee-dimensionale tijdonafhankelijke Schrödinger

Promotor: prof Marnix Van Daele (marnix.vandaele@ugent.be); Begeleider: Toon Baeyens (toon.baeyens@ugent.be)

Korte beschrijving:

De Schrödinger-vergelijking is een partiële differentiaalvergelijking die de toestand beschrijft van een kwantummechanisch systeem. Van deze vergelijking, oorspronkelijk opgesteld door de Oostenrijkse natuurkundige Erwin Schrödinger, bestaat een tijdsafhankelijke vorm en een tijdsonafhankelijke vorm.

Bij dit thesisonderwerp focussen we ons op de tijdsonafhankelijke vorm in twee ruimtelijke dimensies. Er bestaan verschillende numerieke standaardtechnieken om dit probleem op te lossen (bvb. matrixmethoden op basis van eindige-differenties of eindige-elementen), maar er bestaan ook meer gespecialiseerde methoden zoals matslise2D, wat ontwikkeld werd door Toon Baeyens in het kader van zijn doctoraatsonderzoek. Hierbij wordt gebruik gemaakt van principes en methoden die gebruikt worden in de matlab toolbox Matslise (cfr. Wiskundige modellering). De bedoeling is een vergelijking te maken van de verschillende methoden.

Startpunt/referenties:

voor deze vergelijkende studie kan je starten met enkele bekende methodes te vergelijken. Het is niet de bedoeling dat je zelf grote algoritmen implementeert.

- Ixaru, L. Gr. "New Numerical Method for the Eigenvalue Problem of the 2D Schrödinger Equation." *Computer Physics Communications* 181, no. 10 (October 1, 2010): 1738–42. <https://doi.org/10.1016/j.cpc.2010.06.031>.
- Braun, M., S. A. Sofianos, D. G. Papageorgiou, E. Lagaris. "An Efficient Chebyshev–Lanczos Method for Obtaining Eigensolutions of the Schrödinger Equation on a Grid." *Journal of Computational Physics* 126, no. 2 (July 1, 1996): 315–27. <https://doi.org/10.1006/jcph.1996.0140>.

Titel: De numerieke oplossingen van tijdsafhankelijke Schrödinger-vergelijkingen in één en twee dimensies

Promotor: prof Marnix Van Daele (marnix.vandaele@ugent.be); Begeleider: Toon Baeyens (toon.baeyens@ugent.be)

Korte beschrijving:

De Schrödinger-vergelijking is een partiële differentiaalvergelijking die de toestand beschrijft van een kwantummechanisch systeem. Van deze vergelijking, oorspronkelijk opgesteld door de Oostenrijkse natuurkundige Erwin Schrödinger, bestaat een tijdsafhankelijke vorm en een tijdsonafhankelijke vorm. Bij dit thesisonderwerp focussen we ons op de tijdsafhankelijke vorm in een of twee ruimtelijke dimensies.

Er bestaan verschillende numerieke standaardtechnieken om dit probleem op te lossen (bvb. discretisatie via eindige differentie-methodes voor de ruimtelijke veranderlijke en de method of lines voor de tijdsintegratie), maar er bestaan ook meer gespecialiseerde methoden. In de onderzoeksgroep Numerieke wiskunde wordt een methode ontwikkeld die gebaseerd is op principes/methoden van Matslise (cfr. Wiskundige modellering).

De bedoeling van dit thesisonderwerp is om met behulp van reeds bestaande methoden voor de numerieke oplossing van het tijdsonafhankelijke probleem een nieuwe implementatie te bouwen om ook het tijdsafhankelijk probleem op te lossen.

Startpunt/referenties:

- je kan vertrekken vanaf de geziene technieken in 'Gevorderde numerieke methoden' om gewone differentiaalvergelijkingen op te lossen. Je zal zelf een implementatie bouwen in Sage/Python (of indien je dat wenst C++).
- Ledoux, V., and M. Van Daele. "The Accurate Numerical Solution of the Schrödinger Equation with an Explicitly Time-Dependent Hamiltonian." *Computer Physics Communications* 185, no. 6 (June 1, 2014): 1589–94. <https://doi.org/10.1016/j.cpc.2014.02.023>.

Titel: De numerieke oplossing van het 3D Schrödinger-probleem

Promotor: prof. Marnix Van Daele (marnix.vandaele@ugent.be); Begeleider: Toon Baeyens (toon.baeyens@ugent.be)

Korte beschrijving:

De Schrödingervergelijking is een lineaire tweede orde partiële differentiaalvergelijking, ze kent haar oorsprong in de kwantummechanica. Ze beschrijft de onderliggende golfvergelijking van een kwantummechanisch systeem. Zoals de naam 'golfvergelijking' doet vermoeden, zijn oplossingen van de Schrödingervergelijking zeer oscillatorisch. Dit geeft numerieke benaderingsmethoden een extra uitdaging.

In deze masterproef zoeken we numerieke oplossingen van het drie-dimensionale tijdsafhankelijk Schrödingerprobleem $-\Delta \psi + V(x,y,z) \psi = E \psi$ over (een deel van) \mathbb{R}^3 . Aan de rand van dit gebied veronderstellen we dat $\psi(x,y,z) = 0$. Dit is een eigenwaarde- en eigenfunctieprobleem.

De techniek die we zullen gebruiken om dit probleem op te lossen, is een uitbreiding van een techniek die we momenteel aan het optimaliseren zijn voor het equivalente twee-dimensionale probleem. Bij deze techniek gebruiken we een discretisatie in twee dimensies. Voor de overige dimensie stellen we een goed gekozen basis op, gebaseerd op de één-dimensionale Schrödinger-vergelijking. Er bestaan, dankzij Matslise, zeer snelle routines om deze basis te construeren.

Om de tweede orde afgeleide te bepalen in de gediscretiseerde assen maken we gebruik van klassieke eindige differentieformules, zoals $h^2 f''(x) = f(x-h) - 2f(x) + f(x+h)$. Om het oscillatorisch gedrag van de eigenfuncties nauwkeurig te benaderen zullen we hoge orde eindige differentie-formules gebruiken.

In deze masterproef zal de student

- zichzelf vertrouwd maken met deze methode in twee dimensies.
- de uitdaging identificeren en oplossingen ervoor voorstellen.
- een implementatie bouwen van deze methode.
- numerieke experimenten uitvoeren om de sterktes en zwaktes van deze techniek te illustreren.

Afhankelijk van de interesses van de studenten kunnen we ons toespitsen op ofwel de wiskundige achtergrond, ofwel een efficiënte implementatie. Toon Baeyens zal instaan voor de begeleiding..

Titel: Een ander onderwerp in het domein van de numerieke wiskunde

Promotor: prof. Marnix Van Daele (marnix.vandaele@ugent.be)

Korte beschrijving:

Bespreekbaar

Titel: Een onderwerp in het domein van machinaal leren

Promotor: Yvan Saeys (yvan.saeys@ugent.be)

Korte inhoud: bespreekbaar met geïnteresseerde studenten

Titel: Een onderwerp in het domein van de statistische bioinformatica en/of toegepaste data-analyse

Promotor: prof. Lieven Clement (lieven.clement@ugent.be)

Korte beschrijving: Bespreekbaar

Titel: Reductie in coveringgebaseerde ruwverzamelingenleer: een vergelijkende studie

Promotor: prof. Chris Cornelis (chris.cornelis@ugent.be)

Korte beschrijving:

Ruwverzamelingenleer (Eng., rough set theory) is een wiskundige theorie waarbij men deelverzamelingen A van een universum U benadert m.b.v. een equivalentierelatie R over U , of gelijkwaardig hiermee, een partitie P van U . Meer bepaald omvat de onderbenadering van A alle equivalentieklassen van R die volledig bevat zitten in A en de bovenbenadering van A alle klassen die een niet-ledige doorsnede hebben met A . Deze theorie kent uitgebreide toepassing binnen data-analyse, waar ze onder meer gebruikt wordt voor de reductie van classificatietabellen door het selecteren van de meest informatieve objectkenmerken.

Wanneer men de eis dat P een partitie van U vormt minder strikt maakt door slechts te eisen dat P een bedekking of covering vormt, spreekt men over coveringgebaseerde ruwverzamelingen. De definitie van onder- en bovenbenadering is in dit geval niet langer eenduidig bepaald. Zo werden er in de literatuur tientallen verschillende paren van benaderingsoperatoren gebaseerd op een covering voorgesteld. De studie van de verbanden tussen verschillende definities en de eigenschappen die ze vervullen is een relevant en actueel onderzoeksthema.

In deze masterproef dient de student de beschikbare literatuur over reductie van classificatietabellen met coveringgebaseerde ruwverzamelingen kritisch te bestuderen en de verbanden tussen verschillende definities en aanpakken in kaart te brengen. Via dit onderwerp kan de student een rechtstreekse bijdrage leveren tot het actueel wetenschappelijk onderzoek in de wiskunde.

Titel: Explainable methods for irony detection from tweets

Promotor: prof. Chris Cornelis (chris.cornelis@ugent.be); **Begeleider:** Olha Kaminska (olha.kaminska@ugent.be)

Korte beschrijving:

The branch of data science that works with text and learns how to extract the knowledge from it is called Natural Language Processing (NLP). Over the past decades, NLP has made significant headway in the field of sentiment analysis for social media, for instance in the identification of cyberbullying, or in the detection of different emotions expressed by tweets. A driving force in this evolution has been the use of deep learning approaches in the representation and classification of textual data. Deep learning can solve the aforementioned problems with remarkable accuracy, but also has an important drawback: as a black-box model, it does not provide any insight into how it came to a particular conclusion. Explainable AI (XAI) is an emerging field in machine learning that aims to address how AI systems make decisions. It refers to AI methods and techniques that produce human comprehensible solutions. The latter can either be achieved by adding an additional layer of interpretability to black box methods, or by using intrinsically interpretable machine learning methods.

In this thesis, we focus on irony detection from tweets, a particularly challenging task, sometimes even for humans. For example, the sentence "Great. Another rainy day. How wonderful." could easily be mistaken by a machine to express positive sentiment. Ideally, we would like to obtain a system that accurately predicts whether a given tweet is ironic, and at the same time provides the motivation for its decisions (for example, "rainy days" are typically not "great" and "wonderful").

We focus on the the solution of SemEval 2018 Task 3, "Irony detection in English tweets". The goal will be to build a binary classification system, predicting whether a proposed tweet in English is ironic. By contrast to the original SemEval competition, the quality of the solution will be evaluated not only based on its accuracy, but also on its ability to explain predictions. Therefore, a suitable trade-off between both characteristics needs to be sought.

The preferred language of programming for this task is Python.

Titel: Semi-gesuperviseerde classificatie op basis van vaagruwverzamelingen

Promotor: prof. Chris Cornelis (chris.cornelis@ugent.be)

Korte beschrijving:

Bij klassieke classificatie is het de bedoeling om op basis van een gelabelde training set een model te leren dat kan worden geëvalueerd op nieuwe, ongelabelde datapunten. Semi-gesuperviseerde classificatie vormt hierop een uitbreiding waarbij naast een (typisch kleine) gelabelde verzameling L van datapunten ook een (typisch veel grotere) ongelabelde verzameling U als deel van de training set wordt aangeboden. Het idee hierachter is dat het voor veel reële classificatieproblemen moeilijk of duur is om gelabelde data te bekomen, maar dat er wel heel veel ongelabelde data ter beschikking is. Semi-gesuperviseerde algoritmen trachten dan om de classificatie op testdata te verbeteren door niet enkel L maar ook U te gebruiken.

Een belangrijke deelklasse van zulke algoritmen werkt op basis van het self-labelling principe: d.w.z., het algoritme tracht eerst de datapunten uit U van een label te voorzien, en gebruikt deze labels (eventueel aangevuld met een confidence score die aanduidt hoe zeker we zijn over het gekozen label) samen met die uit L nadien voor de classificatie van nieuwe datapunten.

De bedoeling van deze masterproef is om na te gaan hoe goed classificatiemethoden op basis van vaagruwverzamelingen (fuzzy rough sets) zich lenen tot dit probleem. Er kan daarbij worden voortgebouwd op een bestaande experimentele studie uit (Vluymans, 2018); na het doorgronden van de aanpak en de conclusies van deze studie, dient de student zelfstandig één of meerdere alternatieven voor te stellen en hun werking te evalueren op benchmark data.

Er is voor dit onderwerp geen specifieke voorkennis omtrent vaagverzamelingen of ruwverzamelingen vereist, maar wel een gezonde dosis interesse in machine learning.

Referentie:

Sarah Vluymans, Dealing with Imbalanced and Weakly Labelled Data in Machine Learning using Fuzzy and Rough Set Methods, PhD thesis, Ghent University, 2018.