

Titel: Veralgemeende Fourier transformaties: discreet versus continu

Promotoren: prof. Hendrik De Bie (Hendrik.DeBie@UGent.be); prof. Joris Van der Jeugt (Joris.VanderJeugt@UGent.be)

Korte beschrijving:

De klassieke Fourier transformatie (FT) kan men linken aan een realisatie in termen van differentiaaloperatoren van de Lie algebra sl_2 . Deze herschrijving laat vervolgens toe om vergaande veralgemeningen van de FT te bekomen, door nieuwe differentiaaloperator realisaties van sl_2 (of zelfs ingewikkeldere Lie (super)algebra's) te construeren. De laatste jaren vormt dit een actief domein van onderzoek, waar zowel continue als discrete transformaties ingevoerd werden. In veel gevallen is er bovendien een interessante kwantummechanische interpretatie.

Doel van deze scriptie is om een aantal dergelijke transformaties in detail te bestuderen, en in het bijzonder een vergelijkende studie te maken van het discrete en continue geval.

Titel: Constant term identities

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

Korte beschrijving:

In de literatuur zijn er tal van intrigerende “constant term identities” te vinden, die dikwijls hun oorsprong vinden in algebra of combinatoriek. Het berekenen of bepalen van de constante term van een rationale functie (van een bepaalde klasse) is niet eenvoudig. Meestal moet men zich wenden tot technieken die dan kunnen geïmplementeerd worden in computeralgebrapakketten.

Voor deze scriptie onderzoekt de student eerst een gebied waarin de berekening van constante termen van belang is, zoals de bepaling van Kronecker coëfficiënten voor symmetrische functies. Hiervoor is kennis van symmetrische functies en karakters van S_n een pluspunt.

Vervolgens onderzoekt de student enkele methodes/algorithmes om constante term berekeningen te doen. Implementatie van zulke algoritmes in Maple (of Sage) maakt ook deel uit van het werk.

Referentiewerken:

- A. Garsia, N. Wallach, G. Xin and M. Zabrocki, Kronecker coefficients via symmetric functions and constant term identities. (zie o.a. [arXiv:0810.0060 \[math.CO\]](https://arxiv.org/abs/0810.0060)).
 - G. Xin, A fast algorithm for MacMahon’s partition analysis ([arXiv:math/0408377 \[math.CO\]](https://arxiv.org/abs/math/0408377)).
-

Titel: Afhankelijke kansvariabelen en hun som

Promotor: prof. David Vyncke

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., Rüschendorf 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.

Een ander onderwerp in de financiële of actuariële wiskunde

Promotor: prof. David Vyncke

Korte beschrijving:

Bespreekbaar

Titel: Alles normaal of niet ?

Promotor: prof. Christophe Ley (christophe.ley@ugent.be)

Korte beschrijving:

De normale verdeling heeft een speciale plaats in kansrekening, statistiek en, algemeen, in wetenschappen. Dit heeft historische en wiskundige redenen. Er zijn heel wat verschillende stellingen die zeggen dat slechts de normale verdeling een speciale eigenschap bezit: maximale entropie, limiet van een som van toevalsveranderlijken, onafhankelijkheid tussen positie en dispersie schatters, etc. Het doel van deze masterproef is een studie van de verschillende stellingen die de normale verdeling karakteriseren, en een vooruitzicht op beperkingen van de normale verdeling.

Titel: Information theory

Promotor: prof. Christophe Ley (christophe.ley@ugent.be)

Korte beschrijving:

De historische werken van Claude Shannon over transport van informatie vormen de basis voor een hoog interessant onderzoeksdomein : information theory. Er zijn verschillende maten van informatie: entropie, wederzijds informatie, Kullback-Leibler divergence, etc. Deze maten zijn verbonden met codetheorie.

Over de laatste jaren hebben onderzoekers meer en meer links gevonden tussen information theory en statistiek. Het doel van deze masterproef is het bestuderen van deze links en dus van een gloednieuw domein.

Referentiewerken:

er zijn veel goede tekstboeken.

Voorbeeld: Cover, T. and Thomas, J. (2006) : Elements of Information Theory. Wiley New York.

Titel: Het prijzen en hedgen van kwetsbare opties

Promotor: Prof. dr. M. Vanmaele

Korte beschrijving:

Bij de bepaling van de prijs van een financieel afgeleid product wordt er doorgaans ondersteld dat de uitgevende instelling financieel gezond is. Onder deze onderstelling mag het falingsrisico van de uitgever verwaarloosd worden. In realiteit is falingsrisico aanwezig ook bij heel grote bedrijven of instellingen zij het eerder in de zogenaamde over-the-counter (OTC) markt omdat de handel daar niet zo georganiseerd en gereguleerd is als in het geval er een clearinghouse actief is. (Een clearinghouse zorgt voor de administratie van transacties, saldeert deze, houdt margeverplichtingen bij en garandeert de uiteindelijke betaling en levering van stukken zoals overeengekomen in de transactie.)Vandaar dat houders vandergelijke contracten kwetsbaar zijn voor falingsrisico. Kwetsbare opties werden voor het eerst bestudeerd door Johnson en Stulz (1987). Sindsdien is het prijzen en hedgen van dergelijke opties onder verschillende modellen voor de onderliggende in de literatuur uitvoerig bestudeerd. De bedoeling van deze masterproef is om bepaalde types (bijv. Europees, Aziatisch) kwetsbare opties voor bepaalde types van modellen (bijv. met stochastische volatiliteit of met sprongen) voor de onderliggende te bestuderen.

Voor deze studie kan onder meer gesteund worden op volgende artikels en hun referenties:

- Johnson, H., and Stulz, R. (1987). The pricing of options with default risk. Journal of Finance, 267–280.
- Hui, C., C. Lo, and H. Lee. 2003. "Pricing Vulnerable Black-Scholes Options with Dynamics Default Barriers." Journal of Derivatives 10: 62–69. doi:10.3905/jod.2003.319206.
- Jeon, J., J. Yoon, and M. Kang. 2016. "Valuing Vulnerable Geometric Asian Options." Computers and Mathematics with Applications 71: 676–691. doi:10.1016/j.camwa.2015.12.038.
- Puneet Pasricha, Anubha Goel, Pricing vulnerable power exchange options in an intensity based framework, Journal of Computational and Applied Mathematics, Volume 355, 2019, Pages 106-115, <https://doi.org/10.1016/j.cam.2019.01.019>.
- Yang, S., M. Lee, and J. Kim. 2014. "Pricing Vulnerable Options under a Stochastic Volatility Model." Applied Mathematics Letters 34: 7–12. doi:10.1016/j.aml.2014.03.007.

Titel: Stein's lemma en toepassingen in portefeuilleselectieproblemen

Promotor: Prof. dr. M. Vanmaele

Korte beschrijving:

Voor een bivariatenormaalverdeelde 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 toe 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. risictheorie (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: Eerlijke validering van verzekeringsverplichtingen

Promotor: Prof. dr. M. Vanmaele

Korte beschrijving:

De validering van financiële producten gebeurt anders dan bij (zuiver) verzekeringsproducten.

Een financiële prijsbepaling steunt op het principe van geen-arbitragemogelijkheden terwijl een actuariële prijsbepaling gebaseerd is op diversificatie. Bij verzekeringsverplichtingen zijn beide echter verweven. De vraag naar een eerlijke prijs dringt zich dan ook op. In Solvency II staat te lezen 'If (part of) the cash flows of an insurance liability can be replicated, then the value of the (part of the) cash flows is determined on the basis of the market value of these financial instruments. Otherwise, the value is equal to the sum of the best estimate and a risk margin.'

Een moderne actuaaris zoekt dus een antwoord op de vraag: "Voor welke prijs is een andere partij bereid om een verplichting over te nemen, rekening houdend met de indekkingsmogelijkheden in de financiële markt?"

Het doel van deze masterproef is deze eerlijke validering te bestuderen aan de hand van de volgende drie artikels:

- Dhaene, Stassen, Barigou, Linders & Chen (2017). Fair valuation of insurance liabilities: merging actuarial judgement and market-consistency. Insurance : Mathematics & Economics 76: 14-27.
- Barigou & Dhaene (2019). Fair valuation of insurance liabilities via mean-variance hedging in a multi-period setting. Scandinavian Actuarial Journal, 2019(2):163-187.
- Barigou, Chen & Dhaene (2019). Fair dynamic valuation of insurance liabilities: merging actuarial judgement with market- and time-consistency. Insurance: Mathematics and Economics 88:19-29.

Titel: Filtering, Markov chain Monte Carlo (MCMC) met toepassingen in financiële wiskunde

Promotor: Prof. dr. M. Vanmaele

Korte beschrijving:

Bij het prijzen van financiële en verzekeringsproducten gaat men ervan uit dat het mogelijk is om alle bronnen van onzekerheid te observeren. Dit is een onrealistische onderstelling. Bijvoorbeeld de stochastische volatiliteit van aandelenprijzen waarop opties geschreven zijn, is vaak maar gedeeltelijk waarneembaar. Of een belegger wil een afgeleid product prijzen waarvan de onderliggende zelf niet continu waarneembaar is maar slechts op discrete tijdstippen. Ook al hebben beleggers toegang tot de volledige markt, moeten ze vaak beslissingen nemen op basis van partiële informatie. Bij levensverzekeringsproducten kunnen verzekeraars de individuele sterfte van verzekeringsnemers waarnemen maar niet de zogenaamde "mortality rate". Dus ook hier is er sprake van partiële informatie. Bij het prijzen kan dan gebruik gemaakt worden van filters zoals de Kalman-filter in het geval van een lineair, Gaussisch model. Gaat het echter om een niet-lineair, niet-Gaussisch model dan kan "particle filtering" of "sequential Monte Carlo" gebruikt worden.

Het doel van deze masterproef is het bestuderen van deze filtertechniek(en) en de toepassingen ervan in het bijzonder bij het prijzen van financiële producten of bij portefeuille-optimalisatie.

Voor deze studie kan gesteund worden op volgende artikels:

- Creal, D. (2008). Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models. *Computational Statistics & Data Analysis* 52:2863–2876.
- Desai, R., Lele, T., Viens, F. (2003). A Monte-Carlo method for portfolio optimization under partially observed stochastic volatility. 2003 IEEE International Conference on Computational Intelligence for Financial Engineering. *Proceedings*, pp. 257-263.
- Fouqué, J.-P., Papanicolaou, A., Sircar, R. (2015). Filtering and portfolio optimization with stochastic unobserved drift in asset returns. *Communications in Mathematical Sciences* 13(4):935–953
- Lindström, E., Ströjby, J., Brodén, M., Wiktorsson, M., Holst, J. (2008). Sequential calibration of options. *Computational Statistics & Data Analysis* 52:2877–2891.
- Rambharat, B.R. (2012). American option valuation with particle filters. In: R.A. Carmona et al. (eds.), *Numerical Methods in Finance*, Springer Proceedings in Mathematics 12, pp. 51-82.

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

Promotor: Prof. dr. M. Vanmaele

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: Prespecified statistical analysis

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

Korte beschrijving:

Het kunnen vastleggen van een statistisch analyseplan, alvorens de data verzameld werden, is van cruciaal belang om bewuste of onbewuste bijsturing van de resultaten te vermijden. Het is de norm in experimentele klinische studies, en tevens een vereiste van de regulator (bvb. FDA), maar moeilijk in observationele studies. De reden is dat er in dergelijke studies een nood is om te corrigeren voor confounding, en het is moeilijk om op voorhand vast te leggen volgens welke modellen dat zal gebeuren. In deze thesis zullen we daarom recente ontwikkelingen bestuderen die erop gericht zijn om een analyse volledig op voorhand vast te leggen, en deze toepassen in een concrete context.

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

Titel: Schrödinger-vergelijking. Titel: Een vergelijking van algoritmes voor de numerieke oplossing van de twee-dimensionale tijdsonafhankelijke

Begeleider(s): Promotor: prof Marnix Van Daele; Begeleider: Toon Baeyens

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, and I. 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

Begeleider(s): Promotor: prof Marnix Van Daele; Begeleider: Toon Baeyens

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 tijdsafhankelijke 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 tijdsafhankelijke 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: Metric learning for domain adaptive segmentation:

Begeleider(s): Promotor: Yvan Saeys; Begeleider: Joris Roels

Image segmentation is the process of classifying each pixel of an image in a particular class. To improve generalization, it is desired that the developed classifiers are easily adaptable to new image domains: e.g. a segmentation algorithm that works fine for a particular dataset should not require lots of adjustments to work on a new dataset. This problem is also called domain adaptive segmentation and is far from straightforward with the state-of-the-art approaches.

In this thesis, we will use metric learning approaches to find similarities and differences between different domains and adjust the classifier accordingly to generalize to new domains with as little supervision as possible. To do this, the student will start with a literature study in domain adaptation and metric learning, experiment with deep learning frameworks (e.g. Pytorch or Tensorflow) and develop a new domain

adaptive segmentation technique for large-scale biological datasets. The student can make use of a large annotated database of 3D microscopy datasets for training and validation.

Titel: Unsupervised feature extraction and visualization of single-cell data:

Begeleider(s): Promotor: Yvan Saeys; Begeleider: Joris Roels

Feature extraction and visualization is a challenging subject, especially for single-cell datasets that consist of many dimensions (thousands of genes) and many instances (thousands of cells). It is crucial for biomedical research to gain more insights in these datasets by extracting interpretable and semantically meaningful features of these datasets.

Recent approaches in unsupervised feature extraction are based on autoencoder architectures that attempt to reconstruct the data, while reducing their dimensionality at the same time. To increase interpretability of such networks, additional constraints such as sparsity and hierarchical representations are typically included. In this thesis, the student will look into state-of-the-art feature extractors for high-dimensional single-cell datasets. The goal is to develop an algorithm that is able to find structure (e.g. identify different cell types) in this data without manual intervention.

Titel: Automated lesion scoring of medical image data:

Promotor: Yvan Saeys; Begeleider : Joris Roels

Sacroiliac joint pain is experienced in the lower back and caused by damage or injury to the joint between the spine and the hip. Typically, MRI imaging is required to identify inflammatory and structural lesions (e.g. sclerosis, erosions, etc.). Moreover, early diagnosis is crucial in order to rapidly respond with a suitable treatment. Lesion scoring involves time-consuming and tedious manual analysis of different regions in the MRI data.

To automate this approach, the student will develop novel algorithms that predict inflammatory and structural scores of MRI datasets. To do this, the student can employ state-of-the-art supervised regression techniques and high-quality data of real patients, annotated by experts of the Ghent University Hospital.

Titel: Bias en variabiliteit in de data-analyse van genexpressiestudies met herhaalde metingen

Promotoren: prof.Lieven Clement Clement (lieven.clement@ugent.be); prof.Stijn Vansteelandt (stijn.vansteelandt@ugent.be)

Korte beschrijving:

Dankzij hoge-doorvoer sequencerings technologie kan genexpressie (RNA-seq) simultaan worden geëvalueerd op het niveau van duizenden individuele transcripten. De data in RNA-seq studies zijn telprofielen waarbij de tellingen voor een bepaald gen een proxy zijn voor de concentratie in een staal. De teldata worden typisch gemodelleerd aan de hand van negatief binomiale (NB) modellen aangezien de tellingen in biologische studies overdispers zijn t.o.v. een Poisson distributie omwille van de biologische variabiliteit. In de literatuur werden reeds flexibele software-tools ontwikkeld voor het modelleren van

data afkomstig van complexe studiedesigns waarbij de parallelle datastructuur van genomische studies wordt uitgebuit om betere schattingen te bekomen van variantie parameters door informatie te ontleen over de genen heen. De huidige RNA-seq tools zijn echter niet ontwikkeld om correlatiestructuren te modelleren in experimenten met herhaalde metingen en longitudinale designs.

In de RNA-seq literatuur worden subject-specifieke effecten vaak gemodelleerd door het introduceren van additionele vaste effecten in de lineaire predictor van het NB model. In de econometrische literatuur is echter beschreven dat het gebruik van de volledige set van individuele dummy variabelen in NB modellen kan leiden tot inconsistente schatters voor de overige vaste factoren. Daarenboven modelleert de gerandomiseerde compleet blok (RCB) analyse in conventionele RNA-seq tools enkel de binnen-subject error. Hierdoor kunnen deze tools sowieso niet worden ingezet voor inferentie over effecten die zich manifesteren tussen subjecten aangezien de huidige implementatie van de statistische toetsen de variabiliteit tussen subjecten niet in rekening kan brengen.

Het doel van deze masterproef is om inzicht te krijgen in het negatief binomiaal regressiemodel. Hierbij zal eerst de bias en de controle van type I fouten worden geëvalueerd van de RCB analyse met standaard RNA-seq tools waarbij zowel effecten binnen als tussen subjecten zullen worden bestudeerd. We zullen gebruik maken van zowel gesimuleerde alsook echte RNA-seq data. In het tweede deel van de thesis zal worden geëxploreerd hoe kan worden gecorrigeerd voor de eventuele bias en voor de variabiliteit binnen en tussen subjecten in de statistische toetsen die worden gebruikt in RNA-seq studies.

Dankzij hoge-doorvoer sequeneringstechnologie kan genexpressie (RNA-seq) simultaan worden geëvalueerd op het niveau van duizenden individuele transcripten. De data in RNA-seq studies zijn telprofielen waarbij de tellingen voor een bepaald gen een proxy zijn voor de concentratie in een staal. De teldata worden typisch gemodelleerd aan de hand van negatief binomiale (NB) modellen aangezien de tellingen in biologische studies overdispers zijn t.o.v. een Poisson distributie omwille van de biologische variabiliteit. In de literatuur werden reeds flexibele software-tools ontwikkeld voor het modelleren van data afkomstig van complexe studiedesigns waarbij de parallelle datastructuur van genomische studies wordt uitgebuit om betere schattingen te bekomen van variantie parameters door informatie te ontleen over de genen heen. De huidige RNA-seq tools zijn echter niet ontwikkeld om correlatiestructuren te modelleren in experimenten met herhaalde metingen en longitudinale designs.

In de RNA-seq literatuur worden subject-specifieke effecten vaak gemodelleerd door het introduceren van additionele vaste effecten in de lineaire predictor van het NB model. In de econometrische literatuur is echter beschreven dat het gebruik van de volledige set van individuele dummy variabelen in NB modellen kan leiden tot inconsistente schatters voor de overige vaste factoren. Daarenboven modelleert de gerandomiseerde compleet blok (RCB) analyse in conventionele RNA-seq tools enkel de binnen-subject error. Hierdoor kunnen deze tools sowieso niet worden ingezet voor inferentie over effecten die zich manifesteren tussen subjecten aangezien de huidige implementatie van de statistische toetsen de variabiliteit tussen subjecten niet in rekening kan brengen.

Het doel van deze masterproef is om inzicht te krijgen in het negatief binomiaal regressiemodel. Hierbij zal eerst de bias en de controle van type I fouten worden geëvalueerd van de RCB analyse met standaard RNA-seq tools waarbij zowel effecten binnen als tussen subjecten zullen worden bestudeerd. We zullen gebruik maken van zowel gesimuleerde alsook echte RNA-seq data. In het tweede deel van de thesis zal worden geëxploreerd hoe kan worden gecorrigeerd voor de eventuele bias en voor de variabiliteit binnen en tussen subjecten in de statistische toetsen die worden gebruikt in RNA-seq studies.

Titel: Splines voor het schatten van niet-lineaire trends in longitudinale proteomics experimenten

Promotoren: prof. Lieven Clement (lieven.clement@ugent.be); prof. Lennart Martens (lennart.martens@ugent.be)

Korte beschrijving:

Vele biologische processen zijn dynamisch en tijdreeks 'omics experimenten waarbij de expressie van duizenden genen of proteïnes worden gevolgd over de tijd zijn essentieel om deze processen verder te ontrafelen. De analyse van high-throughput tijdreeks data is erg uitdagend gezien de temporele effecten van stimuli op de expressie van proteïnes typisch niet-lineair zijn. Daarom is regularisatie nodig om het onderliggende continue signaal te schatten uit de geobserveerde expressie profielen die typisch sterk variabel zijn. Splines zijn populair binnen de context van 'omics tijdreeks experimenten. Ze kunnen niet-lineaire temporele expressiepatronen schatten op een datagedreven manier, maar splines zijn nog niet goed geïntroduceerd in het veld van de label-vrije differentiele proteomics data-analyse.

In massa-spectrometrie gebaseerde proteomics neemt de kwantificatie plaats op het niveau van de peptides (stukjes van de proteïnes) en de peptide intensiteiten moeten worden geaggregeerd op het niveau van de proteïnen. Recent hebben Goeminne et al. (2016, doi: 10.1074/mcp.M115.055897) MSqRob ontwikkeld, een robuust lineair model voor het detecteren van differentiële abundante proteïnen dat onmiddellijk peptide intensiteiten modelleert. De methode aggregeert de peptide data binnen het model en heeft een hogere sensitiviteit en specificiteit dan de hangbare methoden in de literatuur. MSqRob laat echter nog niet toe om niet-lineaire trends te schatten in tijdreeks data. De software-tool is ingebed binnen het raamwerk van gemengde modellen waarbij de data kan worden gemodelleerd met vaste effecten (temporele trend, effect van de stimulus,...) en random effecten (sample en subject-specifieke effecten). Het doel van deze masterproef is (a) om inzicht te krijgen in schattings- en inferentie methoden voor splines, (b) hoe splines binnen het raamwerk van gemengde modellen kunnen worden ingepast, (c) om de methode van Goeminne et al. uit te breiden voor de analyse van longitudinale studies waarbij het proteoom van patiënten of celculturen wordt gevolgd over de tijd en (d) om de uitbreiding te implementeren binnen het software pakket MSqRob (<https://github.com/statOmics/MSqRob>). Hierbij ligt de nadruk zowel op het schatten van niet-lineaire trends binnen een biologische conditie en om verschillen in temporele expressie-profielen op te pikken tussen biologische condities (e.g. ziek vs gezond, behandeld vs onbehandeld, ...). Het uitgangspunt is het boek "Semiparametric regression" van D. Ruppert, M. Wand en R. Carroll, Cambridge series in statistical and probabilistic mathematics (H1-H6).

Een ander 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: Fuzzy Rough Set Regression

Begeleider(s): Promotor: prof. Chris Cornelis; Begeleider: Oliver Lenz

Korte omschrijving:

Fuzzy rough set theory has been applied in a number of ways in machine learning. Fuzzy rough nearest neighbour classification (FRNN) is a lazy learner that associates an upper and a lower approximation with each decision class and classifies test instances according to their membership in these. This has a

transparent interpretation: upper approximation membership encodes to what extent a test instance is similar to the training instances of a class, whereas lower approximation membership encodes to what extent a test instance is not similar to the training instances of other classes.

Can we also apply fuzzy rough set theory to regression? In principle, this could make fuller use of fuzzy rough set theory. Upper and lower approximations can be defined for sets that are themselves fuzzy, but we don't make use of this in classification, because decision classes are classical sets.

The thesis should consist of two parts. In the first part, the student should study the limited work done on fuzzy rough set regression so far (Jensen & Cornelis 2011, An et al 2014), and develop their own proposal.

One of the central questions is how the numerical data of a regression dataset should be represented with fuzzy sets. Fuzzy sets express class membership, but what is the relevant class expressed by e.g. a temperature measurement? How do we deal with the fact that a fuzzy set has a bounded membership between 0 and 1, whereas numerical attributes may be unbounded? Should we perhaps use more than one fuzzy set to model a single numerical attribute?

In the second part of the thesis, the student should construct a collection of publicly available datasets with varying properties, and systematically evaluate the performance of fuzzy rough set regression and, for comparison, a number of other regressors. Fuzzy rough set theory requires a number of choices of fuzzy operators. How do these choices affect the performance of fuzzy rough set regression?

References:

Jensen, R., & Cornelis, C. (2011). Fuzzy-rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42), 5871-5884.

An, S., Shi, H., Hu, Q., Li, X., & Dang, J. (2014). Fuzzy rough regression with application to wind speed prediction. *Information Sciences*, 282, 388-400.

Title: Usage of Local Interpretable Model-agnostic Explanations (LIME) in Interpretable Machine Learning.

Begeleider(s): Promotor: prof. Chris Cornelis; Begeleider: Marko Palangetic

The interpretation of Machine Learning (ML) models is becoming a very active research field. Many machine learning models have strong predictive capabilities and they provide great results in practice. A well-known example are neural networks, the main tool used in computer vision, natural language processing, sentiment analysis and many other fields. However, neural networks are complex structures by their nature, and we do not know too much how they learn from data. Sometimes that is not important (we do not care how face identification on our smartphone works) but there are cases where that may have a significant impact.

The first case is when we are interested in knowledge generated by the machine learning algorithm. High predictive capability means that the algorithm learned a lot about the relationship between data, detected some trends, found impact of different factors on the output etc. So, we would like to extract all that important information from the algorithm for which we need reliable interpretation methods.

For example, if a client is rejected for a loan in a bank based on ML algorithm, he/she would like to know why that happened.

The second case may be in purpose of debugging of an ML algorithm. Next to valuable knowledge, the algorithm may learn things which do not correspond to the reality of the problem. That happens mainly due to the poor data quality or if the model is not robust to noise or outliers. We want to avoid making predictions based on fake knowledge, even if predictive capability is high on the first impression. To debug the algorithm, we need to “unpack” it, to see what it learned and to fix the possible issues.

The third case is when a wrong prediction may have significant impact. The example are self-driving cars. For example, bad image segmentation done by a neural network in such cars may have fatal consequences. We already have some examples where self-driving cars injured bike riders. So, we have to know how ML algorithms used in such sensitive environment work to be able to avoid negative consequences. There are also many other cases why interpretation is important, but we just listed some known examples.

Due to all these cases, we have to find a proper mathematical tool which can extract the knowledge out of ML algorithms. One popular tool nowadays is Local Interpretable Model-agnostic Explanations (LIME) approach [1]. It is a meta-algorithm which can be integrated with simple and interpretable ML models (linear regression, decision trees etc.) in order to interpret particular decisions of black box ML models. More precisely, it measures the impact of features on a single prediction. For example, it can identify the financial parameters which caused that a client is rejected for a loan in a bank.

The goal of the thesis is to investigate different possibilities for using of the LIME approach for understanding machine learning models. That includes an integration of the LIME approach with different interpretable methods in order to obtain good explanations for the decisions of a black-box ML algorithm. All different integrations will be tested on various datasets from different areas. That will mainly include computer vision, natural language processing and tabular data from different domains (medicine, engineering etc.). The thesis will also discuss if obtained results are meaningful in practice and how much new knowledge they provide. As a conclusion, the thesis should answer in which cases the LIME approach should be used, and which cases it should not.

The work on this thesis requires a decent pre-knowledge in Machine Learning. The student should be already familiar with concepts of supervised learning, prediction ML models (linear models, support vector machines, decision trees, random forests, neural networks etc.) and some statistical aspects of predictions (overfitting, curse of dimensionality, train-test split, cross validation etc.). Even if some datasets from computer vision and natural language processing will be used, a specific knowledge from those areas is not necessary. He/she should be familiar with Python programming and packages numpy, pandas and scikit-learn. Knowledge in some package for deep learning would be useful but not necessary.

References:

[1] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144)

Titel: Transitiviteit van vaagrelaties voor het modelleren van benaderende gelijkheid met toepassing in machine learning.

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

Korte beschrijving

De klassieke gelijkheidsrelatie in een universum U is een equivalentierelatie: inderdaad, elk element van U is gelijk aan zichzelf (reflexief), als x gelijk is aan y , is y ook gelijk aan x (symmetrisch) en uit de gelijkheid van x en y samen met die van y en z volgt automatisch dat x en z gelijk zijn (transitief).

Als we toelaten dat elementen op elkaar kunnen lijken in een bepaalde mate, spreken we van benaderende gelijkheid. In deze masterproef modelleren we dit a.d.h.v. een vaagrelatie R in U , d.w.z. een afbeelding van U^2 naar $[0,1]$. Meestal voeren we enkele beperkingen in voor de relatie R , zoals:

- $R(x,x) = 1$ (reflexief)
- $R(x,y) = R(y,x)$ (symmetrisch)
- $R(x,z) \geq \min(R(x,y), R(y,z))$ (min-transitief)

Ook varianten hierop zijn mogelijk, zo kan men bijvoorbeeld het minimum in de transitiviteitsvoorwaarde vervangen door het product of door een andere operator. Soms laat men deze voorwaarde zelfs gewoon weg, vanuit de gedachte dat benaderende gelijkheid in het algemeen niet transitief is (wat reeds werd opgemerkt door de bekende wiskundige Henri Poincaré).

Het doel van deze masterproef bestaat erin om in te zoomen op de transitiviteitsvoorwaarde. Afhankelijk van de interesse van de student kan hierbij een meer theoretische richting gevolgd worden door na te gaan:

- welke verschillende vormen van afgezwakte transitiviteit werden voorgesteld in de literatuur
- welke transitiviteitseigenschappen voldaan zijn voor de meest courante benaderende gelijkheidsmaten
- welke constructiemethodes er bestaan voor benaderende gelijkheidsmaten met gegeven transitiviteitseigenschappen,

of er kan ook worden toegespitst op machine learning, waarbij het de bedoeling is om na te gaan via een experimentele studie of het opleggen van afgezwakte transitiviteit de nauwkeurigheid van similariteitsgebaseerde classificatie-algoritmen al of niet kan verbeteren.

Er kan voor deze opdracht vertrokken worden van een eerdere masterproef die het onderwerp binnen een ruimer kader bestudeerde.

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: Hate speech detection from tweets

Begeleider(s): Promotor: prof. Chris Cornelis; Begeleider: Olha Kaminska

Korte beschrijving:

Hate speech is usually defined as communication that humiliates one person or a group of people based on some characteristics. The main source of opinions nowadays is the Internet, which contains large collections of textual data in social media. An important task is to detect hate speech in social media to be able to limit it and fight against sexism, racism, xenophobia, etc.

In this thesis, we provide the student with data about hate speech in the form of tweets in English which disparage immigrants and women. The target will be to build a classification model that can detect hate in text.

The goal will be to build a binary classification system which has to predict whether a proposed tweet in English (with a given target either women or immigrants) is hateful or not.

This task was provided at the SemEval 2019 as Task 5, which is the source of data and more details. One of the participating teams to the competition was the Language and Translation Technology Team (LT3) of Ghent University. Therefore, we expect the student to familiarize themselves with their approach, and to propose another solution, using a classification model of your choice (our group has a background in machine learning, so we can guide you with that).

References:

- SemEval competition - <https://competitions.codalab.org/competitions/19935>
- Paper by competition organizers - <https://www.aclweb.org/anthology/S19-2007.pdf>

- Paper by LT3 group - <https://biblio.ugent.be/publication/8604488/file/8640071.pdf>
-

Titel: Technieken voor experimentele algoritmiëk

Promotor: prof. Veerle Fack

Korte beschrijving:

Onderzoek over algoritmen heeft twee fundamentele doelstellingen: het ontwerpen van algoritmen en het analyseren van algoritmen. Het analyseren van algoritmen behelst het voorspellen hoe goed een algoritme zal presteren in een gegeven situatie met gegeven voorwaarden en veronderstellingen. Het ontwerpen van algoritmen behelst het bouwen van betere algoritmen, zoals snellere algoritmen of algoritmen die een goede (bij voorkeur optimale) oplossing voor een probleem bekomen.

Het gebied van de experimentele algoritmiëk combineert het theoretische domein van algoritmiëk met de praktijk. Traditionele technieken voor de analyse van algoritmen volgen een theoretisch model (RAM-model met kost 1 per bewerking) voor het bestuderen van het gedrag van algoritmen. Voor meer accurate inschattingen van het gedrag van een algoritme zijn meer verfijnde modellen nodig, die ook ondersteund worden door experimenten op actuele computersystemen.

De kern van dergelijk 'algorithm engineering' is een cyclus van ontwerp, analyse, implementatie en experimenten.

Doelstelling:

Deze masterscriptie heeft tot doel de student vertrouwd te maken met deze cyclus van algoritme-ontwerp, via een reeks van case studies, gaande van enkele eenvoudige algoritmische problemen (zoals zoeken in gesorteerde lijsten en sorteren) tot meer realistische problemen (zoals routeplanning).

Title: Assessment of treatment effect heterogeneity using modern machine learning techniques.

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

Korte beschrijving:

In the era of personalized medicine, there is much interest in understanding how the effect to a given treatment depends on an individual's characteristics. Traditionally this has been done using simple parametric statistical models, but in recent years some proposals have been developed to learn about effect heterogeneity in a data-driven way, based on the advances in machine learning. Surprisingly, several of these proposals also facilitate statistical inference on the conditional treatment effect. In this project, you will review and compare the different frameworks, initially from a theoretical perspective, to understand their strengths and weaknesses. Further comparison can then be made via running simulation studies and conducting data analyses.

Similariteitsrelaties voor multi-instance data

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

Situering

In een traditionele classificatie-dataset wordt elke observatie beschreven aan de hand van zijn attribuutwaarden en een bijhorend klasselabel. Tabel 1 bevat enkele observaties uit de bekende iris-dataset. De observaties i_1 , i_2 en i_3 worden beschreven met vier attributen (sepalLength, sepalWidth, petalLength, petalWidth) die eigenschappen van een irisplant voorstellen. Er zijn drie mogelijke klasselabels (soorten iris): setosa, versicolor en virginica.

ID	sepalLength	sepalWidth	petalLength	petalWidth	Class
i_1	5.1	3.5	1.4	0.2	setosa
i_2	6.0	2.2	4.0	1.0	versicolor
i_3	6.2	3.4	5.4	2.3	virginica

Tabel 1: Enkele observaties uit de dataset Iris.

In multi-instance data komt elke observatie overeen met een groep elementen. Zo'n groep wordt een *bag* genoemd, en het aantal elementen kan verschillen van bag tot bag. Elk element kan worden voorgesteld met een attribuutvector, maar heeft geen afzonderlijk klasselabel. Het label wordt toegekend aan de gehele bag. Tabel 2 toont twee observaties uit de multi-instance dataset Musk, waarbij elke bag diverse conformaties van eenzelfde molecule bevat. De eerste bag behoort tot de positieve klasse (muskusgeur), de tweede tot de negatieve klasse (geen muskusgeur).

Bag ID	$\langle f_1, f_2, \dots, f_{m-1}, f_m \rangle$	Class
MUSK-jf59	$\langle 52, -110, \dots, -60, -29 \rangle$	Positive
	$\langle 49, -98, \dots, -13, -12 \rangle$	
	$\langle 23, -113, \dots, -9, 90 \rangle$	
	$\langle 47, -110, \dots, -5, -8 \rangle$	
	$\langle 9, -114, \dots, -28, 112 \rangle$	
NON-MUSK-334	$\langle 7, -197, \dots, 34, 55 \rangle$	Negative
	$\langle 25, -198, \dots, 20, -8 \rangle$	

Tabel 2: Enkele observaties uit de multi-instance dataset Musk.

Probleemstelling

Een cruciale component van veel classificatie-algoritmen is het gebruik van afstand of similariteit. Nearest neighbor methoden bijvoorbeeld voorspellen het klasselabel van een nieuwe observatie door diens dichtste buren in de dataset op te sporen, en het meest voorkomende klasselabel onder de buren te kiezen. In de Iris-dataset kan dit zoekproces gebeuren door een metriek (bvb. Euclidische afstand) te evalueren op de attribuutvectoren van observaties.

Multi-instance data zijn duidelijker complexer dan klassieke data. Hoe bepalen we bijvoorbeeld de afstand of mate van gelijkenis tussen twee bags? Er bestaan in de wetenschappelijke literatuur diverse voorstellen om dit probleem op te lossen. In deze masterproef zullen we de mate van gelijkenis tussen multi-instance bags voorstellen aan de hand van *vaagrelaties*, d.w.z. afbeeldingen van $\mathcal{B}$ naar $[0, 1]$ waarbij $\mathcal{B}$ de klasse der bags uit het gestelde classificatieprobleem voorstelt.

Doelstelling

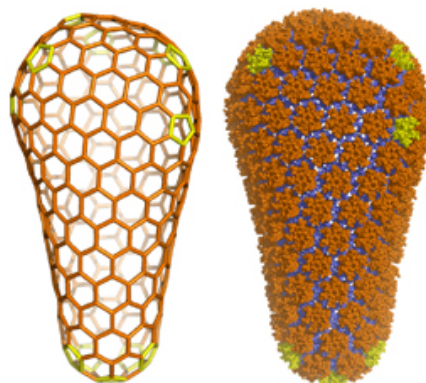
Het doel van deze masterproef omvat:

1. Een grondige en kritische literatuurstudie van de bestaande voorstellen, gezien als vaagrelaties en ingedeeld volgens hun constructiemethode en de eigenschappen (bvb. diverse vormen van transitiviteit) die ze vervullen.
2. Afhankelijk van je interesse, ofwel (a) een experimentele studie waarbij je de belangrijkste methoden vergelijkt in een classificatie-experiment, ofwel (b) een theoretische studie waarbij je zelf één of meerdere nieuwe vaagrelaties voorstelt en onderzoekt, of (c) een combinatie van beide.

Clusters in Fullerenen en HIV virussen

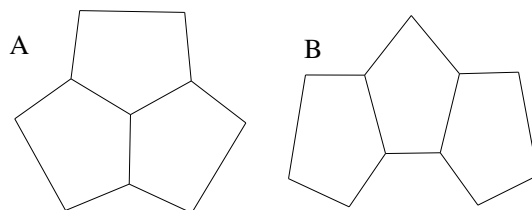
Promotor: Gunnar Brinkmann

Fullerenen en HIV virussen hebben vanuit een wiskundig oogpunt een gelijkaardige structuur: het zijn 3-reguliere vlakke grafen met alleen maar 5- en 6-hoeken.



Een pentagoncluster in een Fullereen of HIV virus is een maximale groep van vijfhoeken die samenhangen. Er is maar één mogelijke cluster met één of twee vijfhoeken, er zijn twee clusters met drie vijfhoeken, vier clusters met vier vijfhoeken, etc.

In een Fullereen of HIV virus moet de som van de groottes van de clusters altijd 12 zijn, maar als de som van de groottes 12 is, betekent dat niet dat er ook een Fullereen of HIV virus met deze clusters bestaat – zelfs dan niet als je alleen maar een theoretisch mogelijke structuur zoekt en geen rekening houdt met scheikundige of biologische beperkingen. Als één van de clusters heel groot is, is het gemakkelijk gevallen te vinden waar een Fullereen met die clusters niet kan bestaan. Voor kleine clusters ligt het vaak niet voor de hand welke combinaties wel dan niet mogelijk zijn. Is er bv. een Fullereen of HIV virus die drie clusters van het type A en één cluster van het Type B uit de volgende afbeelding bevat?



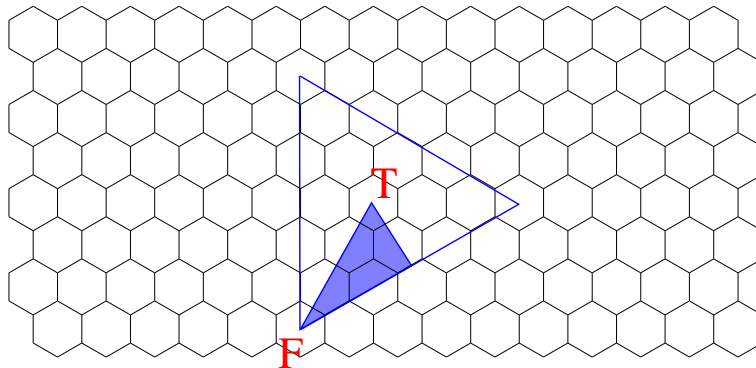
In het kader van deze thesis zal door een combinatie van computationele en theoretische technieken onder andere bepaald worden welke combinaties van pentagonclusters in Fullerenen kunnen voorkomen.

Meer uitleg kan natuurlijk aan de promotor gevraagd worden. Het is zinvol het vak *algoritmische grafentheorie* gevolgd te hebben.

Goldbergoperaties voor clusters

Promotor: Gunnar Brinkmann

Michael Goldberg heeft in 1934 een operatie gedefinieerd om 3-reguliere vlakke grafen met icosahedrale symmetrie en alleen maar 5-hoeken en 6-hoeken te construeren. Later werd een constructie Goldberg-operatie genoemd en heel populair omdat sommige virussen en vooral de in 1985 ontdekte Fullerenen een dergelijke structuur hebben. Inderdaad werd deze Goldberg-operatie door de biologen Caspar en Klug geïntroduceerd, wij noemen die dus beter Caspar-Klug-operaties. Zowel door middel van (de echte) Goldbergoperaties en de Caspar-Klug-operaties kan je van een Fullereen een groter Fullereen maken met dezelfde symmetriegroep. Goldberg gebruikt daarbij een driehoek gebied uit de hexagonale tralie. Deze Goldberg driehoek voor de parameters $(3, 3)$ zie je in het blauw in de afbeelding. Als je meerdere kopieën van deze driehoek (en de gespiegelde driehoek) uitknipt en op de juiste manier in de vlakken van een Fullereen plakt, dan kan je een Fullereen met 27 keer meer toppen krijgen.



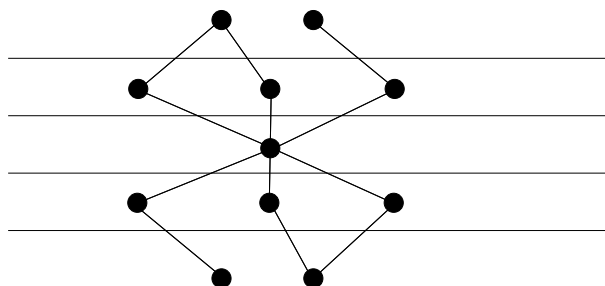
In het resultaat zullen de 12 vijfhoeken altijd geïsoleerd zijn van elkaar. Het zou mooi zijn ook een (op de Goldbergoperaties opbouwende) operatie te hebben die clusters van driehoeken behoudt – waar er dus als er bv. in het oorspronkelijke Fullereen drie vijfhoeken rond een gemeenschappelijke top zitten dat ook in het grotere Fullereen zo is. Dat kan voor sommige (maar niet alle) Goldberg operaties inderdaad door achteraf delen van het Fullereen te vervangen. In het kader van deze thesis worden de details van deze nieuwe operaties uitgewerkt en wordt bewezen dat de nieuwe operaties de gewenste eigenschappen hebben.

Meer uitleg kan natuurlijk aan de promotor gevraagd worden. Het is zinvol het vak *algoritmische grafentheorie* gevolgd te hebben.

Een constructief algoritme voor zelfduale posets

Promotor: Gunnar Brinkmann

Een poset met n punten is een partiële ordening van n elementen – dus een ordening $\leq$ die transitief, reflexief en antisymmetrisch is, maar waar niet noodzakelijk alle elementen met elkaar vergeleken kunnen worden.



Er bestaat een programma (ontwikkeld door de promotor en Brendan McKay) dat – op isomorfie na – alle posets met een gegeven aantal elementen kan opsommen. Een zelfduale poset is een poset dat een isomorfe poset oplevert als je elke relatie $a \leq b$ vervangt door $b \leq a$. In een email schreef Neil Sloan:

```
Brendan, Gunnar,  
Look at this sequence in the OEIS (http://oeis.org/A133983):  
%I A133983  
%S A133983 1,2,3,8,15,50,119,493  
%N A133983 Number of nonisomorphic n-element self-dual posets (or partially ordered sets).  
%C A133983 Posets whose Hasse diagram looks the same if it is turned upside down. -  
      _N. J. A. Sloane_, Mar 22 2013  
%D A133983 R. P. Stanley, Enumerative Combinatorics I, 2nd. ed., p. 247.  
%Y A133983 Cf. A000112.  
%K A133983 nonn,more,changed  
%O A133983 1,2  
%A A133983 _Richard Stanley_, Jan 08 2008  
Perhaps one night when you cannot fall asleep you could compute a few more terms for me!  
Best regards  
Neil
```

Hij heeft gelijk: door het bestaande programma te gebruiken en misschien een beetje aan te passen, zou deze reeks duidelijk uitgebreid kunnen worden. Dat is de eerste taak van deze thesis. Maar omdat de zelfduale posets maar een heel klein deel van alle posets vormen, zou het beter zijn, een gespecialiseerd algoritme te ontwikkelen. Dat vormt de hoofdtak van de thesis. De getallen uit de eerste taak zijn daarbij heel belangrijk om het nieuwe algoritme te testen.

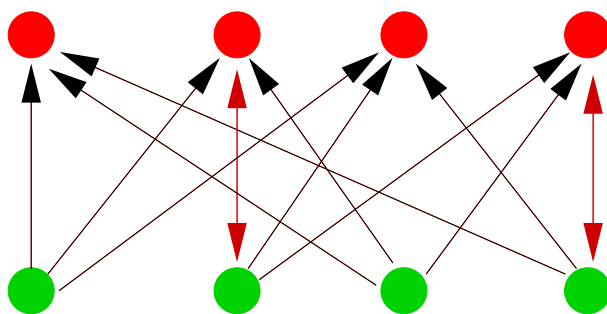
Meer uitleg kan natuurlijk aan de promotor gevraagd worden. Het is zinvol het vak *algoritmische grafentheorie* gevolgd te hebben.

De methode van gesloten deelgrafen

Promotoren: Gunnar Brinkmann en Nico Van Cleemput

Als je volledige lijsten van grafen wilt genereren – bv. alle 3-reguliere grafen met 22 toppen – dan wil je meestal elke graaf op isomorfie na maar één keer. Het eerste idee zou zijn om alle gegenereerde grafen in een lijst bij te houden en elke nieuwe graaf met de al gegenereerde grafen te vergelijken – maar dat is heel traag en de lijsten worden snel veel te groot.

Om dit probleem te vermijden, werden al vroeg verschillende technieken ontwikkeld. Er zijn de methode van *orderly generation*, de *canonical construction path* methode, het *homomorphismeprincipe* en de *double coset-methode*. Deze methoden staan niet los van elkaar en moeten soms in combinatie met elkaar gebruikt worden om efficiënte algoritmen te krijgen.



Niet lang geleden werd er een nieuwe methode gebruikt om reguliere gerichte grafen te genereren, die verwant lijkt aan het homomorphisme principe. De afbeelding toont een situatie tijdens de generatie van deze grafen. Eenvoudig gezegd gaat het erom, vroegtijdig te bepalen dat als je een deelgraaf uitbreidt, sommige automorfismen niet mogelijk zullen zijn – maar dat is **heel** vaag en voldoende details kunnen hier op één pagina niet gegeven worden.

De taak van de thesis is dit principe algemener uit te werken en ook eens toe te passen voor een generatieprobleem waar het niet voor de hand ligt hoe van deze methode gebruik gemaakt kan worden.

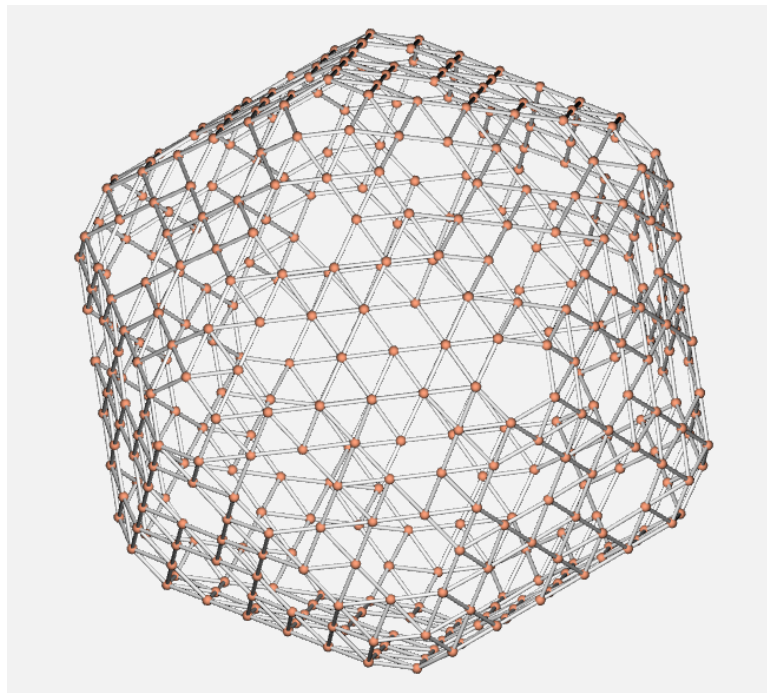
Meer uitleg kan natuurlijk aan een promotor gevraagd worden. Het is zinvol het vak *algoritmische grafentheorie* gevolgd te hebben.

Op zoek naar gelijkzijdige veelvlakken met symmetrie

Promotor: K. Coolsaet (kris.coolsaet@ugent.be)

In een lopend onderzoeksproject zijn we op zoek naar veelvlakken waarin alle zijden gelijke lengte hebben en die voldoen aan een bepaalde symmetrie.

Hiervoor moeten we numerieke oplossingen vinden van grote stelsels van tweede- en derdegraadsvergelijkingen (met duizenden vergelijkingen). Op dit moment gebruiken we hiervoor een variant van de Gauss-Newton-methode, geprogrammeerd in Java. Deze techniek zou echter een heel stuk kunnen verbeterd (versneld) worden als er ook rekening werd gehouden met de inherente symmetrie van het probleem. (In sommige gevallen wordt het stelsel daarvoor zelfs 120 keer kleiner, wat zou kunnen resulteren in een tienduizendvoudige versnelling.) Dit vraagt echter een grondige en intelligente aanpassing van de programmatuur met weinig bestaande literatuur over dit onderwerp.



Doelstelling: Herimplementeren van het numeriek gedeelte in een bestaand programma (Equi - zie <http://twizz.ugent.be/Equi/>) zodat nu ook symmetrie gebruikt wordt.

Hiervoor moet je

- Begrijpen hoe (in het programma) lineaire transformaties worden gebruikt om symmetrie voor te stellen
- De traditionele manieren kennen om stelsels van veeltermvergelijkingen numeriek op te lossen (methode van Newton en Gauss-Newton)

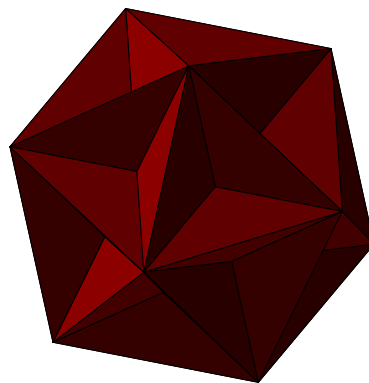
- Ontwikkelen van een algoritme dat deze symmetrie gebruikt bij het numeriek oplossen van het probleem. Hierbij zal hoofdzakelijk gebruik gemaakt worden van algoritmes uit het symbolisch rekenen om de stelsels te vereenvoudigen.
- Incorporeren van dit algoritme in het bestaande programma. (Geschreven in Java.)

Visualisatie van zelfdoorsnijdende veelvlakken

Promotor: K. Coolsaet (kris.coolsaet@ugent.be)

Er bestaan heel wat goede programma's waarmee drie-dimensionale veelvlakken in twee dimensies op een scherm of op papier kunnen worden gevisualiseerd — voornamelijk dankzij de game- en filmindustrie. De meeste van deze programma's veronderstellen echter dat we met 'brave' veelvlakken te doen hebben: dat één zijvlak nooit een ander zijvlak snijdt en dat zijvlakken convex zijn.

Wat wij echter nodig hebben, is een programma dat veelvlakken (die wiskundig zijn geconstrueerd) kan afbeelden op scherm of op papier waarbij zijvlakken elkaar mogen snijden en ook niet noodzakelijk convex zijn. Onderstaande afbeelding is bijvoorbeeld een dergelijk veelvlak waarvan de 12 zijvlakken regelmatige pentagrammen zijn. Dit veelvlak heeft dus niet alleen zijvlakken die elkaar snijden maar ook de zijden in elk veelvlak snijden elkaar.



Doelstelling: Implementatie in Java van een programma dat van dergelijke veelvlakken een 2-dimensionale afbeelding maakt in verschillende formaten. (Een efficiënt programma kan dan ook een roterend veelvlak op het scherm tonen.)

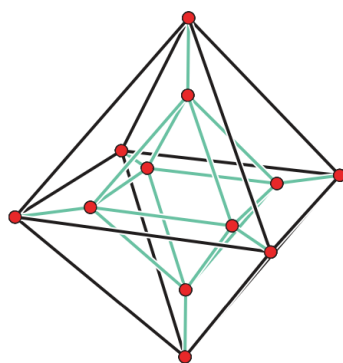
Hiervoor moet je

- Literatuur raadplegen over bestaande algoritmes waarmee veelvlakken worden gevisualiseerd, waarmee niet-convexe veelhoeken in driehoeken worden verdeeld, enz.
- Deze basisalgoritmes implementeren in Java en vergelijken in efficiëntie.
- Deze algoritmes en programma's combineren tot het gevraagde programma.

Polyhedrale grafen met een niet-hamiltoniaans prisma

Promotoren: Jan Goedgebeur en Carol T. Zamfirescu

Dit project gaat over de hamiltoniaanse eigenschappen van het cartesisch product van het 1-skelet van een 3-polytoop (equivalent, een planaire 3-samenhangende graaf) met K_2 (de complete graaf met 2 toppen), d.w.z. zijn *prisma*. Figuur 1 toont de graaf van het prisma van een octahedron. Wanneer we hier spreken over een d -polytoop (d.w.z. een d -dimensionale polytoop), hebben we het altijd over zijn 1-skelet (dus een graaf). Een d -polytoop heet *simpel* of *eenvoudig* als elk van zijn toppen incident is met exact d bogen. Door een stelling van Balinski weten we dat een d -polytoop d -samenhangend moet zijn, dus dat alle toppen minimaal graad d hebben.



Figuur 1: Het prisma van een octahedron.

Rosenfeld en Barnette hebben bewezen dat de 4-kleuren-stelling impliceert dat eenvoudige 3-polytopen hamiltoniaanse prisma's hebben. Fleischner heeft een bewijs gevonden om het gebruik van de 4-kleuren-stelling te vermijden. Paulraja versterkte dit door aan te tonen dat dezelfde conclusie geldt, zelfs als de planariteit niet wordt aangenomen, d.w.z. dat het prisma van een 3-samenhangende kubische graaf hamiltoniaans is. Voor een recente samenvatting van verdere resultaten op het gebied van prisma-hamiltoniciteit, verwijzen we naar [1].

In 1973 formuleerden Rosenfeld en Barnette het vermoeden dat elk 3-polytoop een hamiltoniaans prisma heeft. Špacapan [1] heeft dit vermoeden recentelijk weerlegd. Maar de vraag of alle simpele 4-polytopen hamiltoniaans zijn, blijft open. Deze vraag is de titel van een artikel van Rosenfeld uit 1983. Daarin construeert hij uitgebreide families van niet-hamiltoniaanse 4-reguliere 4-samenhangende grafen waarvan geen enkele het 1-skelet van een 4-polytoop is. Al in 1970 formuleerde Barnette het vermoeden dat het antwoord op deze vraag positief is. Aan de andere kant denkt Mohar dat er voor elke gehele $d \geq 3$ een simpele d -polytoop bestaat die niet hamiltoniaans is. Voor $d = 3$ werd dit bewezen door Tutte in 1946. We merken op dat als men de simpelheid laat vallen, niet-hamiltoniaanse d -polytopen vrij gemakkelijk te beschrijven zijn (bijvoorbeeld met behulp van Kleetopes).

We zien dat in dit probleem, zoals in veel vragen met betrekking tot de hamiltoniaanse eigenschappen van grafen, de graden van toppen een cruciale rol spelen. In een recente samenwerking met Daiki en Maezawa hebben we daarom de oplossing van Špacapan [1] aangepast om enkele natuurlijke vragen aan te pakken, in het bijzonder: hoe laag kunnen we gaan met de maximale graad? Hoeveel toppen van graad 4 kunnen we forceren? Kunnen we de conclusie versterken dat de familie van de 3-polytopen niet prisma-hamiltoniaans is?

De bedoeling van deze thesis is om dit soort vragen vanuit zowel een theoretisch als een computationeel oogpunt aan te pakken. In het bijzonder is het de bedoeling om algoritmes te ontwerpen en te implementeren om dit soort grafen te construeren en hun hamiltoniaanse eigenschappen te bestuderen.

Voor meer uitleg, contacteer: Jan.Goedgebeur@UGent.be en Carol.Zamfirescu@UGent.be

Referenties

- [1] S. Špacapan. A counterexample to prism-hamiltonicity of 3-connected planar graphs. arXiv:1906.06683.