

Анализ на данни, математически финанси и управление на процеси

*Огнян Кунчев и Цветомир Цачев
Институт по математика и информатика при БАН*

С настоящия текст бихме искали да запознаем читателя с три от научните направления, по които в Института по математика и информатика при Българската академия на науките са извършвани изследвания през последното десетилетие. Става дума за анализа на данни, за математическите финанси и за оптималното управление на процеси.

Анализ на данни

Съвременният Анализ на Данни обхваща почти всички познати конструктивни методи в математиката, като например в областта на апроксимация на функциите, теория на сплайните и уейвлетите, числения анализ.

През периода 2008-2018, Анализът на Данни намери приложение в различни области, като основната работа беше извършвана както в рамките на следните проекти с ФНИ, така също и в рамките на международни проекти:

1. „Астроинформатика: Обработка и анализ на дигитизирани астрономически данни и web-базирано приложение“ (с ръководител проф. дмн Огнян Кунчев, за периода 2008–2012)
2. „Нови математически и статистически методи за машинно обучение с приложения в съвременните технологии за генетично секвениране“ (с ръководител проф. дмн Евгения Стоименова, за периода 2014–2018)
3. „Съвременни математически методи за анализ на големи данни с приложения“ (с ръководител проф. дмн Огнян Кунчев, за периода 2017–2019)

Изследвания по астроинформатика

По проект с ФНИ, „Астроинформатика: Обработка и анализ на дигитизирани астрономически данни и web-базирано приложение“ (с ръководител проф. дмн Огнян Кунчев, за периода 2008 – 2012), в колектив от над 20 учени и докторанти, от различни области на Математиката, Информатиката и Астрономията, изкристализира новата област, наречена Астроинформатика, [13].

От колектива в Института по Математика и Информатика активни участници са: Огнян Кунчев (ръководител на колектива), доц. Николай Киров, ас. Емил Келеведжиев, докторантите Дамян Калагларски и Васил Колев. Работата беше извършвана в активно сътрудничество с астрономите доц. Милчо Цветков, н.с. Катя Цветкова и други колеги от Института по Астрономия при БАН, както и доц. Димо Димов (ИИКТ–БАН) и проф. Явор Чапанов (НИГГГ–БАН).

Създадена е информационна технология за описание на плаки (над 50 параметъра), като в момента 370 хил. плаки са описани в База Данни от широкобъгълни астрономически плаки (налични за свободен достъп в Интернет, wfpdb.org), която е единствена по рода си в света. Създадена е технология за дигитизация на самите плаки, приложена в няколко обсерватории в Германия и България, като тази технология гарантира по-бърза и по-евтина дигитизация с по-малко възможности за грешки. Разработени са специални методи за компресия на данни, както и средства за съхранение и достъп, което се налага от размера на получените дигитални образи: 0,5 - 1,0 гигабайта за една плака. Бяха разработени нови ИТ методи за анализ на астрономически

изображения. Можем да посочим следните публикации, свързани с развитите технологии и софтуер за дигитизация, обработка, компресия и съхранение на астрономически фотографски плаки: [3], [4], [5], [6], [7], [25], [26]. Изследванията са извършвани в тясно сътрудничество с колеги от Германия, Унгария, Сърбия, Русия, Украйна, Румъния, и др.

Базата данни от астрономически изображения е отразена на една от страниците на проекта, <http://wfpdb.org/ftp/>

Сред математическите методи, които бяха разработвани и/или прилагани паралелно следва да бъдат отбелязани новите класове от уейвлети. Беше развит нов клас от уейвлети с подразбиване (от типа на уейвлетите на Добеши), които са наречени полихармонични субдивизионни уейвлети. Те са гладки многомерни уейвлети, които са несепарабелни, и демонстрират експериментално превъзходство над досега съществуващите, при компресия и анализ на широк клас от изображения. Основната статия по въпроса бе публикувана в едно от най-престижните списания по приложна математика [2] (за повече информация вижте и цитатите в статията). По-нататък, на основата на полихармоничните субдивизионни уейвлети беше приложена успешно схемата на съвременния подход Compressive Sensing за компресия на данни, [16], [17]. От друга страна, е развит нов подход в областта на Compressive Sensing на многомерни сигнали, основан на многомерно обобщение на теорията на Колмогоров-Гелфанд на напречниците (theory of widths). Той използва новооткритите многомерни пространства на Чебишов, [14], [15].

В областта на Числения Анализ са разработени нови кубатурни формули, които са точни за полихармонични функции от зададен ред. Поради това те са наречени кубатурни формули от полихармоничен тип. Едно много важно тяхно свойство е, че по аналогия с едномерните квадратурни формули, и за тях е получена оценка за грешката. Тази оценка на грешката е получена в съответни пространства от многомерни функции с определена регулярност, като са въведени пространства на Харди, [18], [19].

Нови интересни резултати в Многомерната интерполация бяха получени в областта на Операторите на Бернщайн, които обобщават едномерните такива. Фундаментални резултати за тези оператори бяха публикувани в статията [1], където са получени важни свойства на едномерни Бернщайнови оператори построени чрез експоненциални полиноми.

Нови математически методи за машинно обучение и за анализ на Big Data

В рамките на проектите „Нови математически и статистически методи за машинно обучение с приложения в съвременните технологии за генетично секвениране“ (с ръководител проф. дмн Евгения Стоименова, за периода 2014 – 2018) и „Съвременни математически методи за анализ на големи данни с приложения“ (с ръководител проф. дмн Огнян Кунчев, периода 2017 – 2019), бяха изследвани Разнообразни математически методи за анализ на Големи данни.

Нов метод основан на Спектрална теория на графите (мрежите) се използва за клъстериране на големи масиви от данни в областта на генетиката (РНК, ДНК анализ на експресиите), [29]. Характерното за него е, че отчита наличието на априорна информация под формата на матрица на корелациите, която е формирана на основата на огромно количество исторически наблюдения върху експресиите на гените. От друга страна, в областта на Уейвлети върху мрежи и клъстериране на данни върху мрежи, са получени нови методи за автоматично клъстериране на големи данни върху мрежи.

В областта на непараметричните регресии за анализ на Big Data, проучванията са отразени в статиите [8], [9], [10], където по аналогия с едномерните изглаждащи сплайни, са използвани изглаждащи полисплайни. Друг клас от непараметрични регресии са разгледани в статиите [22], [23].

Получени са нови резултати в областта на интерполация с така наречените експоненциални полиноми, които са съществени за многомерната интерполация с полихармонични функции и за полисплайнови регресии. Те са публикувани в статиите [11], [12].

В статията [24], бяха получени интересни резултати, показващи способността на уейвлетовата регресия да разделя йоносферната компонента от компонентата на кръговия ток в геомагнитните данни. Изследванията са проведени върху геомагнитни данни. Интересното наблюдение е, че определени нискочестотни компоненти на геомагнитните данни са предизвикани от устойчиви явления в йоносферата, които най-вероятно могат да бъдат свързани с така наречените солитони в йоносферата. Последните възникват в резултат на слънчевия терминатор и са изследвани от руския учен В. Белашов в серия от работи от 1991 до 2015.

Резултатите бяха представени с постер на реномираната конференция на EGU (European Geophysical Union), който е наличен онлайн на адрес https://presentations.copernicus.org/EGU2018-18289_presentation.pdf

Математически финанси

В областта на математическите финанси са развити методи за оценяване на конвертируеми облигации (бондове), като е разработен модел на бинарното дърво за оценяване на конвертируеми бондове, които са подложени на кредитен риск, [20]. Кредитният риск се основава на хипотезата за специфична зависимост между интензивността на фалит и оценката на активите на фирмата. Този дискретен модел е сходящ към популярния модел с непрекъснато време Ayache–Forsyth–Vetzal за оценка на конвертируеми бондове, който датира от 2003 г. Посочени са редица проблеми, като например вероятности на прехода и критични гранични стойности, като стъпката на дървото и цената на акциите. Предложеният модел дава алтернативен метод за развиването на ценовата динамика, предложена от модела на Ayache–Forsyth–Vetzal, като се взима предвид и кредитният риск.

В същата област, на конвертируеми бондове, е построен модел за оценяване на така наречените Коко облигации, които са едно обобщение на Конвертируемите облигации, породено от банковата криза през 2008 г., [21]. Тук моделите са основани на частните диференциални уравнения, където пространствената променлива отразява акцията. Чрез предложения подход може да се пресмятат делта, гама, а също дюрацията и изпъкналостта за Коко бондовете, вкл. колбек свойството. В тази статия е преразгледано решението на Spiegeleer and Schoutens (2011), дадено в явна формула, като новото решение е приложимо към всички практически проблеми. Този подход е подходящ за бенчмарк за проверка на числените методи. Развити са два типа методи. Първият се базира преди всичко на първичното пазарно допускане на Блек-Шоулс, докато вторият включва моделирането на кредитния риск чрез скокообразно дефолтиране на акцията.

В статиите [27], [28] са изследвани стохастични модели на деривати, които могат да претърпят дефолт. Данните, които се изследват с такива деривати са по същество големи.

Управление на процеси

Няколко уводни думи

Ако можем да въздействаме на даден процес по време на неговото протичане, ние се стремим да му въздействаме така, че той да протече по оптимален за нас начин. Използването на математически техники за изучаването на процеси от различни области (инженерните науки, икономиката, обществените науки), както и за тяхното управление, има богата история. Започва се с описанието на процеса на математически език с помощта на еволюционни уравнения. Според свойствата на различните им елементи, въпросните уравнения се класифицират като:

- в дискретно време (диференчни уравнения) или в непрекъснато време;
- обикновени или частни диференциални уравнения;
- уравнения с дробни производни;
- уравнения с локална динамика или с нелокална динамика;
- интегрални уравнения, и т. н.

След като процесът е описан чрез съответния математически модел, трябва да бъдат идентифицирани променливите, съответстващи на параметрите, чрез които ние можем да въздействаме на процеса (управляващите променливи, управлението), както и променливите, съответстващи на параметрите, описващи моментното състояние на променящата се система (фазовите променливи). Задачите за управление на процеса, формулирани на базата на неговото математическо описание, се различават по отношение на поставената цел на управлението. Един клас такива задачи имат за цел да повлияят на процеса така, че той да протича в стабилен режим през дълъг период от време, това обикновено се постига чрез т. нар. (стабилизираща) обратна връзка. Други задачи се фокусират върху търсенето на такова въздействие върху процеса, че той да протече по оптимален за нас (според даден критерий) начин. Вторият клас задачи формира математическата област на оптималното управление, възникнала на основата на вариационното смятане преди около шестдесет години.

През шестте десетилетия, изминали от средата на петдесетте години на миналия век, бяха решени огромно количество задачи на оптималното управление – както теоретични, така и такива, имащи практическа насоченост. Важността на приложенията на математическите техники в управлението на реални процеси привлече вниманието на много изследователи към тази област. Решените задачи включват управлението на летящи изделия, подобряване качествата на продукцията в металолееенето, оптималното потушаване на вибрации, проектиране на ядрени реактори и много други.

Към настоящия момент теорията на оптималното управление на крайно-мерни системи (чието моментно състояние се описва с краен брой числа) е добре разработена и, повече или по-малко, завършена. В същото време, има много нерешени задачи от областта на оптималното управление на безкрайно-мерни системи. Те често биват наричани „системи с разпределени параметри“ и се описват чрез частни диференциални уравнения, диференциални уравнения със закъснение, интегрални уравнения, комбинация от изброените типове уравнения, и т. н. Голямо е разнообразието на реалните процеси, моделирани чрез въпросните системи. Тук можем да изброим динамика на флуидите, растеж на кристали, ядрени реакции, електромагнитни процеси и много други.

Оптимално управление на хетерогенни системи

В Института по математика и информатика при БАН изследвания по оптимално управление се правят от първата половина на седемдесетте години на миналия век. В последните години тези изследвания се фокусират основно върху хетерогенни динамични системи. Те се извършват в тясно сътрудничество с учени от Техническия университет във Виена, Австрия.

Една динамична система се нарича хетерогенна ако всеки от нейните елементи (агенти) има собствена динамика. Една от разпределените променливи в такива системи удовлетворява квазилинейно частно диференциално уравнение от първи ред, като динамиката зависи нелокално от управлението и от фазовите променливи. Граничните условия могат да бъдат ендеогенни (да зависят от начина, по който описваният процес протича). Нелокалността се дължи на факта, че динамиката на всеки от агентите зависи от колективното поведение на останалите агенти (от агрегираното състояние). Това се моделира с една или повече интегрални фазови променливи, получени чрез интегриране на съответната разпределена фазова променлива по параметъра, описващ хетерогенността.

Хетерогенността играе съществена роля в еволюцията на икономическите системи, в популационната динамика, в разпространението на заразни заболявания и други. Типични

представители на хетерогенните системи са системите, структурирани по възраст. Оптималното управление на такива системи се прилага в различни области, например при промишления риболов, противодействието на разпространението на заразни заболявания, вземането на инвестиционни решения, противодействието на замърсяването на околната среда и т. н. Хетерогенните системи включват също така системите, структурирани по размери, системите, структурирани по възраст и по продължителност и някои други класове. Подробно описание на различни класове хетерогенни системи в непрекъснато време както и необходими условия за оптималност в задачите за оптимално управление на такива системи са дадени във

Veliov, V., *Optimal Control of Heterogeneous Systems: Basic Theory*, J. Math. Anal. Appl., 346 (2008), 227–242.

Важността на хетерогенните динамични системи може да бъде илюстрирана с два примера. Единият от тях описва връзката между демографията и икономиката – в повечето индустриални държави стопанската дейност ще се осъществява от намаляващо и застаряващо икономически активно население. Европа, и в частност България, не правят изключение от тази тенденция. От ключово значение е да се идентифицират последствията от нея за икономическата продуктивност на обществото с цел да се разработят оптимални политики в областта на инвестициите в човешки капитал и в пенсионната система. Вторият пример е свързан с опазването на околната среда – математическите модели на инвестиции в хетерогенен (с различни производствени и замърсяващи характеристики) производствен капитал са подходящо средство за изучаване на отрицателните (по-високи разходи) и положителните (по-бързо подновяване на производствените мощности) последствия за икономическия растеж от усилията за намаляване на вредните емисии. Хетерогенността в тези модели се задава от моралното и физическо остаряване (vintage) на производствения капитал – по-новите машини и съоръжения произвеждат повече и замърсяват по-малко, но са по-скъпи от по-старите.

През последните десет години изследователи от ИМИ–БАН и от Техническия университет във Виена работиха съвместно върху задачи на оптималното управление за различни хетерогенни системи.

В [34] се изучава какви трябва да бъдат оптималните инвестиции в образование на макро ниво при наличие на хетерогенна по отношение на възраст и на квалифицираност работна сила. Приемат се за дадени еластичността на заместване по възраст и квалификация, влияеща на търсенето на работна ръка, и демографските тенденции, влияещи на нейното предлагане. Показано е, че в случая на пълна заменяемост по възраст и квалификация оптималните инвестиции за образование на различните възрастови групи не зависят от демографските тенденции. В случая на непълна заменяемост демографските тенденции влияят (както в краткосрочен, така и в дългосрочен план) на оптималната образователна политика.

Пазарите на квоти за емисии на парникови газове се изучават в [31] и [32]. Използва се модел с хетерогенен производствен капитал – всяка фирма има машини и съоръжения, различаващи се по своята производителност и по степента, с която замърсяват околната среда. Намаляване на замърсяването се постига чрез промяна на структурата на производствената база – по-старите машини се заменят с по-нови, като се държи сметка за цената на тази замяна, която намалява печалбата на фирмата. Посочено е, че при определени обстоятелства пазарът на квоти за емисии не функционира. Също така, показано е, че е по-подходящо ограничението за парникови емисии да се налага да е в рамките на определени периоди (а не във всеки един момент). Така се постига функциониране на пазара и се намалява неговата волатилност.

В [30], [35] и [36] се разглежда модел на макроикономически растеж с ендегенно (зависещо от развитието на процеса) подобряване на върховите технологии. Тук хетерогенността се изразява в наличието на различни по производителност технологии. Създаването на все по-производителни върхови технологии е обусловено от количеството на инвестициите в изследователска дейност. Полученото необходимо условие за оптималност от типа на принципа за максимума на Л. С. Понтрягин позволява моделът да се изследва както аналитично, така и числено.

Оценки за разпространението на инфекциозно заболяване в хетерогенна популация са представени в [37]. Популацията се дели на незаразени (податливи на инфекцията) и заразени в случая на SI (susceptible – infected) модели и на незаразени (податливи на инфекцията), заразени и оздравели (счита се, че вече са си създали имунитет към инфекцията) в случая на SIR (susceptible – infected – recovered) модели. Хетерогенността на популацията се състои в това, че отделните индивиди имат различна податливост на инфекцията. Предложеният в статията подход е състоянието на популацията в поредица от дискретни моменти да се характеризира чрез принадлежност към определено чрез формулираната математическа задача множество.

В ИМИ–БАН са изследвани и задачи на оптималното управление на възрастово-структурирани системи, включващи поточкови терминални ограничения за разпределена фазова променлива, мотивирани от желанието да се получи за тях необходимо условие за оптималност от типа на принципа за максимума на Понтрягин. В [33] беше получен такъв резултат като следствие на по-общ абстрактен резултат. Абстрактният резултат беше получен чрез подобряване на техниката от

X.J. Li, Jiongmin Yong, *Optimal control theory for infinite dimensional systems*, Basel, Birkhäuser, 1995.

Подобриеното позволи да се изостави едно от двете основни предположения за получаването на съответния резултат на Li & Yong и съществено да се отслаби второто. Около двадесет години се е считало, че този резултат на Li & Yong не може да бъде доказан без въпросните две предположения.

Литература

- [1] Aldaz, J. M., O. Kounchev, H. Render, Bernstein operators for extended Chebyshev systems, *Applied Mathematics and Computation*, 217 (2010), No 2, 790–800
- [2] Dyn, N., O. Kounchev, D. Levin, H. Render, Regularity of generalized Daubechies wavelets reproducing exponential polynomials, *Applied and Computational Harmonic Analysis*, 37 (2014), No 2, 288–306
- [3] Kirov, N., M. Tsvetkov, K. Tsvetkova, Technology for digitization of astronomical photographic plates, In: *Proceedings of the 8th Annual International Conference on Computer Science and Education in Computer Science*, 5–8 July 2012, Boston, USA, 109–114
- [4] Kirov N., M. Tsvetkov, K. Tsvetkova, Software Tools for Digitization of Astronomical Photographic Plates, *Serdica Journal of Computing*, 6 (2012), No 1, 67–76
- [5] Kirov N., M. Tsvetkov, K. Tsvetkova, D. Kalaglarsky, Brashear Plate Catalogues in the Wide-Field Plate Database, In: *Proceedings of the VII Bulgarian-Serbian Astronomical Conference (VII BSAC)*, Chepelare, Bulgaria, June 1-4, 2010, Editors: M. K. Tsvetkov, M. S. Dimitrijević, K. Tsvetkova, O. Kounchev, Ž. Mijajlović, *Publ. Astron. Soc. “Rudjer Bošković”*, No 11, 2012, 147–152
- [6] Kirov, N., *Astroinformatics and Digitization of Astronomical Heritage*, Review of the National Center for Digitization, 19 (2011), 7–10
- [7] Kirov, N., O. Kounchev, M. Tsvetkov, *Astroinformatics: Image Processing and Analysis of Digitized Astronomical Data with WEB-based Implementation*, *Proceedings of the 6th Annual International Conference on Computer Science and Education in Computer Science*, 26-27 June 2010 in Fulda and 28-29 June 2010 in Munich, Germany, 176–181
- [8] Kounchev, O., Ts. Tsachev, Interpolation and Smoothing L splines, *Fast Algorithms*, C.r. acad. Bulg. Sci., 70 (2017), No 10, 1347–1354
- [9] Kounchev, O., H. Render, Ts. Tsachev, L-spline smoothing for differential operators with constant coefficients of order 4, submitted
- [10] Kounchev, O., H. Render, Ts. Tsachev, Fast algorithms for interpolation and smoothing with L-splines of order 4, in preparation

- [11] Kounchev, O., H. Render, On Polyharmonic Interpolation and Interpolation with One-dimensional Exponential Polynomials, C.R. Acad. Bulg. Sci., 70 (2017), No 11, 1493–1500
- [12] Kounchev O., H. Render, Study of non-negative exponential polynomials, приета за публикуване в Докл. БАН
- [13] Kounchev, O., M. Tsvetkov, et al., Astrominformatics: A Synthesis between Astronomical Imaging and Information & Communication Technologies, In: Modern Trends in Mathematics and Physics ed. S.S. Tinchev, Heron Press, Sofia, 2009, 60–69
- [14] Kounchev, O., Infinite-dimensional generalization of Kolmogorov widths, CONSTRUCTIVE THEORY OF FUNCTIONS, Sozopol 2010, G. Nikolov and R. Uluhev editors, Prof. Marin Drinov Acad. Publishing House, Sofia, 2012
- [15] Kounchev, O., Multidimensional Chebyshev spaces, hierarchy of infinite-dimensional spaces and Kolmogorov-Gelfand widths, electronic publication, arXiv:1105.0786
- [16] Kounchev, O., D. Kalaglarsky, Quantization Opportunities for Polyharmonic Subdivision Wavelets Applied to Astronomical Images, ACM, Proceeding CompSysTech'11 Proceedings of the 12th International Conference on Computer Systems and Technologies, Pages 626–635
- [17] Kounchev, O., D. Kalaglarsky, Compressive Sensing for Polyharmonic Subdivision Wavelets With Applications to Image Analysis, ACM, Proceeding CompSysTech '12 Proceedings of the 13th International Conference on Computer Systems and Technologies, 407–414
- [18] Kounchev, O., H. Render, Polyharmonic Hardy Spaces on the Klein-Dirac Quadric with Application to Polyharmonic Interpolation and Cubature Formulas, Pliska Stud. Math. Bulgar., 21 (2012), 147–176
- [19] Kounchev, O., H. Render, Polyharmonic Hardy Spaces on the Complexified Annulus and Error Estimates of Cubature Formulas, Results in Mathematics, 62 (2012), No 3 & 4, 377–403
- [20] K. Milanov, O. Kounchev, Frank J. Fabozzi, Young Shin Kim and Svetlozar T. Rachev, A Binomial-Tree Model for Convertible Bond Pricing, The Journal of Fixed Income Winter 2013, 22 (3) 79-94.
- [21] K. Milanov, O. Kounchev, CoCo Bonds Assessment, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2667160
- [22] Nikolov, N. I., E. Stoimenova, EM Estimation of the Parameters in Latent Mallows' Models, Studies in Computational Intelligence, Springer Series
- [23] Nikolov N. I., E. Stoimenova, Asymptotic properties of Lee distance, submitted
- [24] Srebrov, B., L. Pashova, O. Kounchev, Study of local manifestations of G5 – extreme geomagnetic storms (29–31 October, 2003) in mid-latitudes using geomagnetic data by continuous wavelet transform, Доклади на БАН, 71 (2018), No. 6, 803–811
- [25] Tsvetkova, K., M. Tsvetkov, N. Kirov, P. Boehm, M. Steinmetz, R. Arlt, H. Enke, R. von Berlepsch, Potsdam Astronomical Photographic Plate Library, Serdica Journal of Computing, 6 (2012), No 1, 35–46
- [26] Tsvetkov, M., K. Tsvetkova, N. Kirov, Technology for scanning of astronomical photographic plates, Serdica Journal of Computing, 6 (2012), No 1, 77–88
- [27] T. S. Zaeviski, O. Kounchev, Stochastic times and defaultable derivatives, публикувана в Conference Proceedings and Abstracts, XVIII-та Международна лятна конференция по теория на вероятностите и Математическа статистика, 2018; http://www.math.bas.bg/~statlab/ISCPS2018/ISCPS%202018_files/ISCPS_book_2018.pdf
- [28] Zaeviski, T., O. Kounchev, A jump moment as a stopping time and defaultable derivatives, приета в Докл. БАН.

- [29] Zaeovski, T., O. Kounchev, D. Palejev, E. Stoimenova, Spectral clustering of multidimensional genetic data. *Annual of Sofia University St. Climent Ohridski*, 104, Publishing House of Sofia University, 2017
- [30] Belyakov, A., Ts. Tsachev and V. Veliov, Optimal control of heterogeneous systems with endogenous domain of herogeneity, *Applied Mathematics & Optimization*, 64 (2011), 287–311, DOI: 10.1007/s00245-011-9140-2
- [31] Bréchet, Th., Ts. Tsachev and V.M. Veliov, Prices versus Quantities in a Vintage Capital Model, in *Dynamic Systems, Economic Growth, and the Environment*, J. Crespo Cuaresma, T.K. Palokangas, A. Tarasyev (Eds.), *Dynamic Modeling and Econometrics in Economics and Finance, Vol. 12*, Springer 2010, 141 - 159, Proceedings of ECG-2008, November 7-8, 2008, IIASA, Laxenburg, Austria
- [32] Bréchet, Th., Ts. Tsachev and V.M. Veliov, Markets for Emission Permits with Free Endowment: a Vintage Capital Analysis, *Optimal Control, Applications and Methods*, 33 (2012), 214–231, DOI: 10.1002/oca.988
- [33] Krastanov, M.I., N.K. Ribarska and Ts. Tsachev, A Pontryagin maximum principle for infinite-dimensional problems, *SIAM Journal on Control and Optimization*, 49 (2011), No 5, 2155–2182, DOI: 10.1137/100799009
- [34] Prskawetz, A., Ts. Tsachev and V. Veliov, Optimal education in an age-structured model under changing labor demand and supply, *Macroeconomic Dynamics*, 16 (2012), 159–183, DOI:10.1017/S1365100510000465
- [35] Skritek, B., Ts. Tsachev, V.M. Veliov, Optimality conditions and the Hamiltonian for a distributed optimal control problem on controlled domain, *Applied Mathematics & Optimization*, 70 (2014), No 1, 141–164, DOI: 10.1007/s00245-014-9237-5
- [36] Skritek, B., Ts. Tsachev, V.M. Veliov, Pontryagin's type optimality conditions for a distributed control problem arising in endogenous growth theory, Proceedings of the 9-th International Conference on “Large-Scale Scientific Computations”, June 3-7, 2013, Sozopol, Bulgaria: Large-Scale Scientific Computing, I. Lirkov, S. Margenov, J. Wasniewski (Eds.), *Lecture Notes in Computer Science*, 8353: 143-151, Springer, 2014
- [37] Tsachev, Ts., V.M. Veliov, A. Widder, Set-membership estimations for the evolution of infectious diseases in heterogeneous populations, *Journal of Mathematical Biology*, 74 (2017), No 5, 1081–1106

Data Analysis, Mathematical Finance and Control of Processes

Ognyan Kounchev and Tsvetomir Tsachev

ABSTRACT: We provide a short presentation of the activities in three research directions at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, carried out mostly during the last decade, namely, Data Analysis, Mathematical Finance, and Optimal Control.

In the area of Data Analysis one has to mention the Astroinformatics project with Bulgarian NSF in the period 2008-2012 (coordinated by Prof. Ognyan Kounchev) with a team of more than 20 participants from Astronomy, Mathematics, Computer Science, and Engineering, which has produced an essential impact in establishing the new area of Astroinformatics not only at national, but also at European and world scale. The project "New mathematical and statistical methods for Machine Learning with applications in the modern technologies for gene sequencing" headed by Prof. Evgenia Stoimenova, during the period 2014-2018, has joined the efforts of a team for developing new methods of Machine Learning with applications in the modern genetics. Similar in spirit is the project "Modern mathematical methods for analysis of Big Data and applications", \$2016-2019\$ (coordinated by Prof. O. Kounchev). The two

projects have developed and are developing new methods for analysis of Big Data, in the modern framework of Machine Learning, as e.g. polyharmonic wavelet analysis (spline and subdivision ones), multivariate interpolation and cubature formulas with error estimates, spectral graph methods for clustering based on a priori information, non-parametric regression with L-splines and polysplines, Bernstein polynomials and splines. A companion project (partially funded by the above) is the one on Mathematical Finance. It has been devoted to application of constructive methods in mathematics and stochastics to yield

curve building, as well as to pricing of Convertible and Coco bonds in modern finance. Dr. Krasimir Milanov has defended his PhD thesis in this area, and the novel models developed have received a significant recognition in terms of journal publications as well as from the financial industry on Wallstreet.

As far as the optimal control is concerned, it is about influencing a process as it evolves in time, so that it evolves in the best possible for us way. One has to first describe the development of the process by evolution equations. Then one identifies the variables corresponding to parameters with which the evolution can be influenced (the control variables) and the variables that describe the state of the system at each moment in time (the state variables). The control problems that are then formulated vary widely with respect to the goal that is set.

Research in optimal control has been done at IMI--BAS since the early 1970-ies. In the recent years this research, carried out in close collaboration with scientists from Vienna University of Technology, has been focused mainly on heterogeneous dynamical systems. A dynamical system is called heterogeneous if each of its elements has its own dynamics, the latter being interdependent. Heterogeneity plays a substantial role in the evolution of economic systems, of populations, of epidemic diseases, etc.

Проф. дмн Огнян Кунчев,
проф. д-р Цветомир Цачев,
1113 София, ул. „Акад. Г. Бончев“, бл. 8
Институт по математика и информатика при БАН