Reservoir computing, Regularity structures and learning of Dynamics in Finance

Josef Teichmann

ETH Zürich

May 14, 2018

Challenges in mathematical Finance

- High dimensional stochastic control problems often of a non-standard type (hedging in markets with transaction costs or liquidity constraints).
- High-dimensional inverse problems, where models (PDEs, stochastic processes) have to be selected to explain a given set of market prices optimally.
- High-dimensional prediction tasks (long term investments, portfolio selection).
- High-dimensional feature selection tasks (limit order books).
- Task: understand, quantify and store the information of such non-linear maps (input-output maps).

Challenges in mathematical Finance

- High dimensional stochastic control problems often of a non-standard type (hedging in markets with transaction costs or liquidity constraints).
- High-dimensional inverse problems, where models (PDEs, stochastic processes) have to be selected to explain a given set of market prices optimally.
- High-dimensional prediction tasks (long term investments, portfolio selection).
- High-dimensional feature selection tasks (limit order books).
- Task: understand, quantify and store the information of such non-linear maps (input-output maps).

Challenges in mathematical Finance

- High dimensional stochastic control problems often of a non-standard type (hedging in markets with transaction costs or liquidity constraints).
- High-dimensional inverse problems, where models (PDEs, stochastic processes) have to be selected to explain a given set of market prices optimally.
- High-dimensional prediction tasks (long term investments, portfolio selection).
- High-dimensional feature selection tasks (limit order books).
- Task: understand, quantify and store the information of such non-linear maps (input-output maps).

Challenges in mathematical Finance

- High dimensional stochastic control problems often of a non-standard type (hedging in markets with transaction costs or liquidity constraints).
- High-dimensional inverse problems, where models (PDEs, stochastic processes) have to be selected to explain a given set of market prices optimally.
- High-dimensional prediction tasks (long term investments, portfolio selection).
- High-dimensional feature selection tasks (limit order books).
- Task: understand, quantify and store the information of such non-linear maps (input-output maps).

Challenges in mathematical Finance

- High dimensional stochastic control problems often of a non-standard type (hedging in markets with transaction costs or liquidity constraints).
- High-dimensional inverse problems, where models (PDEs, stochastic processes) have to be selected to explain a given set of market prices optimally.
- High-dimensional prediction tasks (long term investments, portfolio selection).
- High-dimensional feature selection tasks (limit order books).
- Task: understand, quantify and store the information of such non-linear maps (input-output maps).

Approaches from Machine learning:

In order to approximate and store the information of non-linear maps ...

- Basis regressions,
- non-linear regressions of, e.g., neural networks,
- feature extractions

... are performed.

Deep Networks in Finance

Recent ideas to use machine learning in Finance

- Deep pricing: use neural networks to constitute efficient regression bases in, e.g., the Longstaff Schwartz algorithm for pricing call-able products like American options (e.g. recent works of Patrick Cheridito, Calypso Herrera, Arnulf Jentzen, etc)
- Deep hedging: use neural networks to approximate hedging strategies in, e.g., hedging problems in the presence of market frictions (joint work with Hans Bühler, Lukas Gonon, Ben Wood).
- Deep filtering: use neural networks on top of well selected dynamical systems to approximate laws of signals conditional on "noisy" observation (e.g. recent joint projects with Lukas Gonon, Lyudmila Grigoryeva, Juan-Pablo Ortega).
- Deep calibration: use machine learning to perform a solution of inverse problems (model selection) in Finance (joint work with Christa Cuchiero, Andres Hernandez and Wahid Khosrawi-Sardroudi).

Deep Networks in Finance

Recent ideas to use machine learning in Finance

- Deep pricing: use neural networks to constitute efficient regression bases in, e.g., the Longstaff Schwartz algorithm for pricing call-able products like American options (e.g. recent works of Patrick Cheridito, Calypso Herrera, Arnulf Jentzen, etc)
- Deep hedging: use neural networks to approximate hedging strategies in, e.g., hedging problems in the presence of market frictions (joint work with Hans Bühler, Lukas Gonon, Ben Wood).
- Deep filtering: use neural networks on top of well selected dynamical systems to approximate laws of signals conditional on "noisy" observation (e.g. recent joint projects with Lukas Gonon, Lyudmila Grigoryeva, Juan-Pablo Ortega).
- Deep calibration: use machine learning to perform a solution of inverse problems (model selection) in Finance (joint work with Christa Cuchiero, Andres Hernandez and Wahid Khosrawi-Sardroudi).

Deep Networks in Finance

Recent ideas to use machine learning in Finance

- Deep pricing: use neural networks to constitute efficient regression bases in, e.g., the Longstaff Schwartz algorithm for pricing call-able products like American options (e.g. recent works of Patrick Cheridito, Calypso Herrera, Arnulf Jentzen, etc)
- Deep hedging: use neural networks to approximate hedging strategies in, e.g., hedging problems in the presence of market frictions (joint work with Hans Bühler, Lukas Gonon, Ben Wood).
- Deep filtering: use neural networks on top of well selected dynamical systems to approximate laws of signals conditional on "noisy" observation (e.g. recent joint projects with Lukas Gonon, Lyudmila Grigoryeva, Juan-Pablo Ortega).
- Deep calibration: use machine learning to perform a solution of inverse problems (model selection) in Finance (joint work with Christa Cuchiero, Andres Hernandez and Wahid Khosrawi-Sardroudi).

Deep Networks in Finance

Recent ideas to use machine learning in Finance

- Deep pricing: use neural networks to constitute efficient regression bases in, e.g., the Longstaff Schwartz algorithm for pricing call-able products like American options (e.g. recent works of Patrick Cheridito, Calypso Herrera, Arnulf Jentzen, etc)
- Deep hedging: use neural networks to approximate hedging strategies in, e.g., hedging problems in the presence of market frictions (joint work with Hans Bühler, Lukas Gonon, Ben Wood).
- Deep filtering: use neural networks on top of well selected dynamical systems to approximate laws of signals conditional on "noisy" observation (e.g. recent joint projects with Lukas Gonon, Lyudmila Grigoryeva, Juan-Pablo Ortega).
- Deep calibration: use machine learning to perform a solution of inverse problems (model selection) in Finance (joint work with Christa Cuchiero, Andres Hernandez and Wahid Khosrawi-Sardroudi).

Neural Networks

Neural networks in their various topological features are frequently used to approximate functions due ubiquitous universal approximation properties. A neural network, as for instance graphically represented in Figure 1,

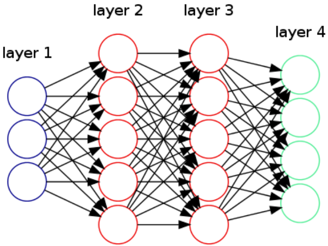

Figure: A 2 hidden layers neural network with 3 input and 4 output dimensions

just encodes a certain concatenation of affine and non-linear functions by composition in a well specified order.

Neural Networks and Universal Approximation

- Neural networks appear in the 1943 seminal work by Warren McCulloch and Walter Pitts inspired by certain functionalities of the human brain aiming for articial intelligence (AI).
- Arnold-Kolmogorov Theorem represents functions on unit cube by sums and uni-variate functions (Hilbert's thirteenth problem), i.e.

$$F(x_1, \dots, x_d) = \sum_{i=0}^{2d} \varphi_i\Big(\sum_{j=1}^{d} \psi_{ij}(x_j)\Big)$$

- Universal Approximation Theorems (George Cybenko, Kurt Hornik, et al.) show that *one hidden layer networks* can *approximate* any continuous function on the unit cube.

Neural Networks and Universal Approximation

- Neural networks appear in the 1943 seminal work by Warren McCulloch and Walter Pitts inspired by certain functionalities of the human brain aiming for articial intelligence (AI).
- Arnold-Kolmogorov Theorem represents functions on unit cube by sums and uni-variate functions (Hilbert's thirteenth problem), i.e.

$$F(x_1, \dots, x_d) = \sum_{i=0}^{2d} \varphi_i \Big(\sum_{j=1}^{d} \psi_{ij}(x_j) \Big)$$

- Universal Approximation Theorems (George Cybenko, Kurt Hornik, et al.) show that *one hidden layer networks* can *approximate* any continuous function on the unit cube.

Neural Networks and Universal Approximation

- Neural networks appeard in the 1943 seminal work by Warren McCulloch and Walter Pitts inspired by certain functionalities of the human brain aiming for articial intelligence (AI).
- Arnold-Kolmogorov Theorem represents functions on unit cube by sums and uni-variate functions (Hilbert's thirteenth problem), i.e.

$$F(x_1, \dots, x_d) = \sum_{i=0}^{2d} \varphi_i\Big(\sum_{j=1}^{d} \psi_{ij}(x_j)\Big)$$

- Universal Approximation Theorems (George Cybenko, Kurt Hornik, et al.) show that *one hidden layer networks* can *approximate* any continuous function on the unit cube.

Neural networks and basis regression

- Deep neural networks easily approximate wavelet basis, whence the theory of all sorts of wavelet approximations of non-linear functions applied. This can be used to explain the *unreasonable effectiveness* of deep neural networks (Helmut Bölcskei, Philipp Grohs et al.)
- However, the approach is static with respect to dimension of the input space.
- In dynamical or in very high dimensional situations the *static* theory of universal approximation appears sometimes too rigid.
- Take for instance time series analysis: there input has changing length, output are time series characteristics.

Neural networks and basis regression

- Deep neural networks easily approximate wavelet basis, whence the theory of all sorts of wavelet approximations of non-linear functions applied. This can be used to explain the *unreasonable effectiveness* of deep neural networks (Helmut Bölcskei, Philipp Grohs et al.)
- However, the approach is static with respect to dimension of the input space.
- In dynamical or in very high dimensional situations the *static* theory of universal approximation appears sometimes too rigid.
- Take for instance time series analysis: there input has changing length, output are time series characteristics.

Neural networks and basis regression

- Deep neural networks easily approximate wavelet basis, whence the theory of all sorts of wavelet approximations of non-linear functions applied. This can be used to explain the *unreasonable effectiveness* of deep neural networks (Helmut Bölcskei, Philipp Grohs et al.)
- However, the approach is static with respect to dimension of the input space.
- In dynamical or in very high dimensional situations the *static* theory of universal approximation appears sometimes too rigid.
- Take for instance time series analysis: there input has changing length, output are time series characteristics.

The reservoir computing paradigm

- Goal is to understand an input-output map by splitting it into a composition of two maps. Input dimension can be changing.
- the first map transforms by means of a *generic dynamical system* (often with physical realization and, of course, with some relationship to the input-output map) the input into features. The dynamical system is called the *reservoir*.
- on those features training is performed: in the simplest case just a linear regression.
- Obviously a complicated training is circumvented by using a generic non-linear map – only the last layers needs to be trained.

The reservoir computing paradigm

- Goal is to understand an input-output map by splitting it into a composition of two maps. Input dimension can be changing.
- the first map transforms by means of a *generic dynamical system* (often with physical realization and, of course, with some relationship to the input-output map) the input into features. The dynamical system is called the *reservoir*.
- on those features training is performed: in the simplest case just a linear regression.
- Obviously a complicated training is circumvented by using a generic non-linear map – only the last layers needs to be trained.

The reservoir computing paradigm

- Goal is to understand an input-output map by splitting it into a composition of two maps. Input dimension can be changing.
- the first map transforms by means of a *generic dynamical system* (often with physical realization and, of course, with some relationship to the input-output map) the input into features. The dynamical system is called the *reservoir*.
- on those features training is performed: in the simplest case just a linear regression.
- Obviously a complicated training is circumvented by using a generic non-linear map – only the last layers needs to be trained.

The reservoir computing paradigm

- Goal is to understand an input-output map by splitting it into a composition of two maps. Input dimension can be changing.
- the first map transforms by means of a *generic dynamical system* (often with physical realization and, of course, with some relationship to the input-output map) the input into features. The dynamical system is called the *reservoir*.
- on those features training is performed: in the simplest case just a linear regression.
- Obviously a complicated training is circumvented by using a generic non-linear map – only the last layers needs to be trained.

Goal of this talk

- show mathematical contexts where the paradigm of reservoir computing appears.
- develop a theory of *reservoirs* within Martin Hairer's regularity structures.

Goal of this talk

- show mathematical contexts where the paradigm of reservoir computing appears.
- develop a theory of *reservoirs* within Martin Hairer's regularity structures.

Applications

- Many problems in Finance are of filtering nature, i.e. calculating conditional laws of a true signal X_{t+h}, at some point in time $t+h$, given some noisy observation $(Y_s)_{0 \le s \le t}$.
- Such problems often depend in a complicated, non-robust way on the trajectory of Y, i.e. no Lipschitz dependence on Y: regularizations are suggested by, e.g., the theory of regularity structures, and its predecessor, rough path theory. By lifting input trajectories Y to more complicated objects (later called *models*) one can increase robustness to a satisfactory level (see, e.g., recent works of Juan-Pablo Ortega, Lyudmila Grigoryeva, etc).
- We shall apply the abstract theory of expansions as developed by Martin Hairer in a series of papers to cristallize properties of reservoirs.

Applications

- Many problems in Finance are of filtering nature, i.e. calculating conditional laws of a true signal X_{t+h}, at some point in time $t+h$, given some noisy observation $(Y_s)_{0 \leq s \leq t}$.
- Such problems often depend in a complicated, non-robust way on the trajectory of Y, i.e. no Lipschitz dependence on Y: regularizations are suggested by, e.g., the theory of regularity structures, and its predecessor, rough path theory. By lifting input trajectories Y to more complicated objects (later called *models*) one can increase robustness to a satisfactory level (see, e.g., recent works of Juan-Pablo Ortega, Lyudmila Grigoryeva, etc).
- We shall apply the abstract theory of expansions as developed by Martin Hairer in a series of papers to cristallize properties of reservoirs.

Applications

- Many problems in Finance are of filtering nature, i.e. calculating conditional laws of a true signal X_{t+h}, at some point in time $t+h$, given some noisy observation $(Y_s)_{0 \leq s \leq t}$.
- Such problems often depend in a complicated, non-robust way on the trajectory of Y, i.e. no Lipschitz dependence on Y: regularizations are suggested by, e.g., the theory of regularity structures, and its predecessor, rough path theory. By lifting input trajectories Y to more complicated objects (later called *models*) one can increase robustness to a satisfactory level (see, e.g., recent works of Juan-Pablo Ortega, Lyudmila Grigoryeva, etc).
- We shall apply the abstract theory of expansions as developed by Martin Hairer in a series of papers to cristallize properties of reservoirs.

A guiding mathematical example:

Consider a stochastic differential equation of diffusion type with smooth vector fields

$$dX_t = \sum_{i=0}^{d} V_i(X_t) \circ dB_t^i \,,\; X_0 \in \mathbb{R}^d$$

Rough path theory says that ...

- X_t can be written as a non-linear Lipschitz function of the Brownian path $(B_s)_{0 \leq s \leq t}$ together with its (one-step) iterated integral $\int_0^t dB_s \otimes dB_s$ (Lyons' universal limit theorem).
- X_t can be (almost) written as a linear map on the input signal's full signature up to time t, i.e. the collection of *all* iterated integrals.
- Whence the signature can be considered a reservoir on which the equation's solution can be learned.

A guiding mathematical example:

Consider a stochastic differential equation of diffusion type with smooth vector fields

$$dX_t = \sum_{i=0}^{d} V_i(X_t) \circ dB^i_t\,,\ X_0 \in \mathbb{R}^d$$

Rough path theory says that ...

- X_t can be written as a non-linear Lipschitz function of the Brownian path $(B_s)_{0 \le s \le t}$ together with its (one-step) iterated integral $\int_0^t dB_s \otimes dB_s$ (Lyons' universal limit theorem).
- X_t can be (almost) written as a linear map on the input signal's full signature up to time t, i.e. the collection of *all* iterated integrals.
- Whence the signature can be considered a reservoir on which the equation's solution can be learned.

A guiding mathematical example:

Consider a stochastic differential equation of diffusion type with smooth vector fields

$$dX_t = \sum_{i=0}^{d} V_i(X_t) \circ dB_t^i \,,\ X_0 \in \mathbb{R}^d$$

Rough path theory says that ...

- X_t can be written as a non-linear Lipschitz function of the Brownian path $(B_s)_{0 \leq s \leq t}$ together with its (one-step) iterated integral $\int_0^t dB_s \otimes dB_s$ (Lyons' universal limit theorem).
- X_t can be (almost) written as a linear map on the input signal's full signature up to time t, i.e. the collection of *all* iterated integrals.
- Whence the signature can be considered a reservoir on which the equation's solution can be learned.

The model space

Let $A \subset \mathbb{R}$ be an index set, bounded from below and without accumulation point, and let $T = \bigoplus_{\alpha \in A} T_\alpha$ be a direct sum of Banach spaces T_α graded by A. Let furthermore G be a group of linear operators on T such that, for every $\alpha \in A$, every $\Gamma \in G$, and every $\tau \in T_\alpha$, one has $\Gamma\tau - \tau \in \bigoplus_{\beta<\alpha} T_\beta$.

The triple $\mathcal{T} = (A, T, G)$ is called a *regularity structure* with *model space* T and *structure group* G.

Given $\tau \in T$, we will write $\|\tau\|_\alpha$ for the norm of its T_α-projection.

Meaning: $\mathcal{T}$ represent abstractly expansions of "functions" at some space-time point in terms of "model functions" of regularity α.

Some notation

Given a test function φ on $\mathbb{R}^d$, we write φ_x^λ as a shorthand for

$$\varphi_x^\lambda(y) = \lambda^{-d}\varphi\big(\lambda^{-1}(y-x)\big)\ .$$

Given $r > 0$, we denote by B_r the set of all functions $\varphi\colon \mathbb{R}^d \to \mathbb{R}$ with $\varphi \in C^r$, its norm $\|\varphi\|_{C^r} \leq 1$ and supported in the unit ball around the origin.

Models

Given a regularity structure $\mathcal{T}$ and an integer $d \geq 1$, a *model* for $\mathcal{T}$ on $\mathbb{R}^d$ consists of maps

$$\begin{aligned} \Pi\colon \mathbb{R}^d &\to L\big(T, \mathcal{D}'(\mathbb{R}^d)\big) \\ x &\mapsto \Pi_x \end{aligned} \qquad\qquad \begin{aligned} \Gamma\colon \mathbb{R}^d \times \mathbb{R}^d &\to G \\ (x,y) &\mapsto \Gamma_{xy} \end{aligned}$$

such that $\Gamma_{xx} = \mathrm{id}$, $\Gamma_{xy}\Gamma_{yz} = \Gamma_{xz}$ and $\Pi_x \Gamma_{xy} = \Pi_y$ for all x, y, z. Furthermore, given $r > |\inf A|$, for any compact set $K \subset \mathbb{R}^d$ and constant $\gamma > 0$, there exists a constant C such that the inequalities

$$\big|\big(\Pi_x \tau\big)(\varphi_x^\lambda)\big| \leq C\lambda^{|\tau|}\|\tau\|_\alpha\,, \qquad \|\Gamma_{xy}\tau\|_\beta \leq C|x-y|^{\alpha-\beta}\|\tau\|_\alpha\,,$$

hold uniformly over $\varphi \in B_r$, $(x,y) \in K$, $\lambda \in]0,1]$, $\tau \in T_\alpha$ with $\alpha \leq \gamma$, and $\beta < \alpha$.

Modeled distributions

The regularity structure allows to speak about abstract expansions at some space time point. We can now introduce spaces of functions taking values in abstract expansion (functions of expansion coefficients with respect to some basis) and ask under which conditions we can actually associate a generalized function to such coefficient functions.

Given a regularity structure $\mathcal{T}$ equipped with a model (Π, Γ) over $\mathbb{R}^d$, the space $\mathcal{D}^\gamma = \mathcal{D}^\gamma(\mathcal{T}, \Gamma)$ is given by the set of functions $f \colon \mathbb{R}^d \to \bigoplus_{\alpha<\gamma} T_\alpha$ such that, for every compact set K and every $\alpha < \gamma$, the exists a constant C with

$$\|f(x) - \Gamma_{xy} f(y)\|_\alpha \leq C|x-y|^{\gamma-\alpha}$$

uniformly over $x, y \in K$.

The polynomial regularity structure

We choose T the polynomial ring in d (commutative) variables $e_1, \ldots, e_d$, and $A = \mathbb{N}$ the natural grading. Abstract expansion are just Taylor expansions at a point $x \in \mathbb{R}^d$ *if* we interpret the abstract polynomials as

$$\Pi_x \mathbf{e}^{\mathbf{k}}(y) := (y - x)^{\mathbf{k}}$$

for a multi-index $k \in \mathbb{N}^d$ and $x, y \in \mathbb{R}^d$. Of course $\Gamma_h \mathbf{e}^{\mathbf{k}} := \big(\mathbf{e} + h\big)^{\mathbf{k}}$ and $\Gamma_{x,y} := \Gamma_{y-x}$.

With these definitions modeled distributions $f \in D^\gamma$ (for $\gamma \geq 0$ correspond to Hölder functions of order γ (in the appropriate sense). Notice that this is a derivative free definition of Hölder functions.

Rough paths as models

Consider the following abstract situation: let $0 < \gamma \leq 1$ be a degree of roughness and E a (Banach) space where the rough path will take values (think of $\mathbb{R}^d$ for simplicity). The index set A will be $\gamma\mathbb{N}$.

- We define $T_{k\gamma} = (E^*)^{\otimes k}$, i.e. the k fold tensor product of the space of linear functionals on E.
- The model space $T = \bigoplus_{\alpha \in A} T_\alpha$.
- We also consider the predual T_*, i.e. the space of infinite series $T((E))$ in non-commutative variables from E.
- This pairing comes with several intriguing algebraic structures: T_* are non-commutative infinite series in varaibles from E, whence an algebra. We can consider the Lie algebra $\mathfrak{g}$ generated by E in this algebra and look at its exponential image $G := \exp(\mathfrak{g})$.

Rough paths as models

- Analytically speaking this corresponds to the collection of all infinite series of the type

$$\sum_{k\geq 0} \sum_{(i_1,\ldots,i_k)\in\{1,\ldots,d\}^k} \int_{0\leq t_1\leq\cdots\leq t_k\leq t} d\omega_{i_1}(t_1)\ldots d\omega_{i_k}(t_k) e_{i_1}\ldots e_{i_k}$$

  for all possible finite variation curves ω taking values in $E = \mathbb{R}^d$.

- For every $a \in G$ we define $\langle a^{-1} \otimes c, b\rangle = \langle c, \Gamma_a b\rangle$, whence G acting on T via adjoing left multiplication by the inverse.

- Algebraically T_* is a (formal closure of a) free algebra, T is a commutative algebra with the shuffle product. By duality this yields a Hopf algebra structure.

- Elements from T may be considered as linear functionals on $G \subset T_*$. The shuffle product of two such linear functionals just calculates the unique linear functional whose restriction on G coincides with the product!

Rough paths as models

- Whence we have constructed a regularity structure (A, G, T): the meaning of elements of T is to be abstract expansions of solutions of controlled ODEs with respect to the control ω.
- Models of the regularity structure are nothing else than $\Gamma_{st} = \Gamma_{X_{st}}$, i.e. group valued functions of two variables satisfying certain algebraic relations, and $\Pi_s a(t) := \langle X_{st}, a \rangle$, for times s, t, and hence precisely *geometric rough paths*.
- Notice that linear functionals on G, i.e. all possible elements from T, form a dense subset of all continuous functions on G (in an appropriate topological sense).

Reconstruction operator

The most fundamental result in the theory of regularity structures then states that given a coefficient function $f \in \mathcal{D}^\gamma$ with $\gamma > 0$, there exists a *unique* distribution $\mathcal{R}f$ on $\mathbb{R}^d$ such that, for every $x \in \mathbb{R}^d$, $\mathcal{R}f$ equals $\Pi_x f(x)$ near x up to order γ. More precisely, one has the following reconstruction theorem, whose proof relies on results from wavelet analysis (multi-resolution analysis, see Martin Hairer's Inventiones article).

Let $\mathcal{T}$ be a regularity structure and let (Π, Γ) be a model for $\mathcal{T}$ on $\mathbb{R}^d$. Then, there exists a unique linear map $\mathcal{R}\colon \mathcal{D}^\gamma \to \mathcal{D}'(\mathbb{R}^d)$ such that

$$\big|\big(\mathcal{R}f - \Pi_x f(x)\big)(\varphi_x^\lambda)\big| \lesssim \lambda^\gamma \|f\|_{\mathcal{D}^\gamma}, \tag{1}$$

uniformly over $\varphi \in B_r$ and $\lambda \in]0,1]$, and locally uniformly in $x \in \mathbb{R}^d$. Additionally the reconstruction operator is Lipschitz with respect to the dependence on models (Π, Γ). Notice that models form a non-linear space.

Besov models and Reconstruction on Besov spaces

All the previous constructions and results can be found in Martin Hairer's seminal paper 'Theory of regularity structures' (2014).

Several extensions are of interest from a point of view of stochastic analysis:

- Hölder spaces are a special case of Besov spaces $\mathcal{B}_{p,q}$ (i.e. Hölder spaces correspond to the case $p = q = \infty$). From a point of view of stochastic analysis, but also from a point of view of approximation theory, Besov spaces with finite p, q are preferable (UMD properties, reflexivity, etc).
- Hairer-Labbé (2016) and Prömel-Teichmann (2016) developed the Besov theory for modeled distributions for $\gamma > 0$. The equally important case for $\gamma < 0$ is treated in Liu-Prömel-Teichmann (2018) (important for explicit model construction).

Besov models and Reconstruction on Besov spaces

All the previous constructions and results can be found in Martin Hairer's seminal paper 'Theory of regularity structures' (2014).

Several extensions are of interest from a point of view of stochastic analysis:

- Hölder spaces are a special case of Besov spaces $\mathcal{B}_{p,q}$ (i.e. Hölder spaces correspond to the case $p = q = \infty$). From a point of view of stochastic analysis, but also from a point of view of approximation theory, Besov spaces with finite p, q are preferable (UMD properties, reflexivity, etc).
- Hairer-Labbé (2016) and Prömel-Teichmann (2016) developed the Besov theory for modeled distributions for $\gamma > 0$. The equally important case for $\gamma < 0$ is treated in Liu-Prömel-Teichmann (2018) (important for explicit model construction).

Besov models and Reconstruction on Besov spaces

- Liu-Prömel-Teichmann (2018ab) provide furthermore discrete characterizations of Besov rough paths, polish space valued Besov maps and prove a universal limit theorem for Besov rough paths.
- A full theory of Besov models and Besov modeled distributions is at sight now. Advantages: insights from stochastic integration and regularity structures can be benefitially combined.
- All proofs rely in wavelet analysis and generalizations of it, surprisingly even in the polish space valued case.

Besov models and Reconstruction on Besov spaces

- Liu-Prömel-Teichmann (2018ab) provide furthermore discrete characterizations of Besov rough paths, polish space valued Besov maps and prove a universal limit theorem for Besov rough paths.
- A full theory of Besov models and Besov modeled distributions is at sight now. Advantages: insights from stochastic integration and regularity structures can be benefitially combined.
- All proofs rely in wavelet analysis and generalizations of it, surprisingly even in the polish space valued case.

Besov models and Reconstruction on Besov spaces

- Liu-Prömel-Teichmann (2018ab) provide furthermore discrete characterizations of Besov rough paths, polish space valued Besov maps and prove a universal limit theorem for Besov rough paths.
- A full theory of Besov models and Besov modeled distributions is at sight now. Advantages: insights from stochastic integration and regularity structures can be benefitially combined.
- All proofs rely in wavelet analysis and generalizations of it, surprisingly even in the polish space valued case.

- Many solutions of problems in stochastics can be translated to solving fixed point equation on modelled distributions.
- By applying the reconstruction operator the modeled distribution is translated to a real world object, which then depends – by inspecting precisely its continuities – in an at least Lipschitz way on the underlying model, i.e. stochastic inputs.
- The theory of regularity structures tells precisely how 'models' have to be specified such that stochastic inputs actually constitute models: this yields a way to construct reservoirs.
- Supervised learning: by creating training data (in appropriate input format!) one can learn the input-output map.
- Applications: solutions of stochastic differential equations (Friz, Lyons, Victoir, etc), solutions of correlated filtering problems (Crisan, Friz, etc), solutions of sub-critical stochastic partial differential equations (Gubinelli, Hairer, Perkowski, etc), solutions of stochastic optimization problems and stochastic games.

- Many solutions of problems in stochastics can be translated to solving fixed point equation on modelled distributions.
- By applying the reconstruction operator the modeled distribution is translated to a real world object, which then depends – by inspecting precisely its continuities – in an at least Lipschitz way on the underlying model, i.e. stochastic inputs.
- The theory of regularity structures tells precisely how 'models' have to be specified such that stochastic inputs actually constitute models: this yields a way to construct reservoirs.
- Supervised learning: by creating training data (in appropriate input format!) one can learn the input-output map.
- Applications: solutions of stochastic differential equations (Friz, Lyons, Victoir, etc), solutions of correlated filtering problems (Crisan, Friz, etc), solutions of sub-critical stochastic partial differential equations (Gubinelli, Hairer, Perkowski, etc), solutions of stochastic optimization problems and stochastic games.

- Many solutions of problems in stochastics can be translated to solving fixed point equation on modelled distributions.
- By applying the reconstruction operator the modeled distribution is translated to a real world object, which then depends – by inspecting precisely its continuities – in an at least Lipschitz way on the underlying model, i.e. stochastic inputs.
- The theory of regularity structures tells precisely how 'models' have to be specified such that stochastic inputs actually constitute models: this yields a way to construct reservoirs.
- Supervised learning: by creating training data (in appropriate input format!) one can learn the input-output map.
- Applications: solutions of stochastic differential equations (Friz, Lyons, Victoir, etc), solutions of correlated filtering problems (Crisan, Friz, etc), solutions of sub-critical stochastic partial differential equations (Gubinelli, Hairer, Perkowski, etc), solutions of stochastic optimization problems and stochastic games.

- Many solutions of problems in stochastics can be translated to solving fixed point equation on modelled distributions.
- By applying the reconstruction operator the modeled distribution is translated to a real world object, which then depends – by inspecting precisely its continuities – in an at least Lipschitz way on the underlying model, i.e. stochastic inputs.
- The theory of regularity structures tells precisely how 'models' have to be specified such that stochastic inputs actually constitute models: this yields a way to construct reservoirs.
- Supervised learning: by creating training data (in appropriate input format!) one can learn the input-output map.
- Applications: solutions of stochastic differential equations (Friz, Lyons, Victoir, etc), solutions of correlated filtering problems (Crisan, Friz, etc), solutions of sub-critical stochastic partial differential equations (Gubinelli, Hairer, Perkowski, etc), solutions of stochastic optimization problems and stochastic games.

- Many solutions of problems in stochastics can be translated to solving fixed point equation on modelled distributions.
- By applying the reconstruction operator the modeled distribution is translated to a real world object, which then depends – by inspecting precisely its continuities – in an at least Lipschitz way on the underlying model, i.e. stochastic inputs.
- The theory of regularity structures tells precisely how 'models' have to be specified such that stochastic inputs actually constitute models: this yields a way to construct reservoirs.
- Supervised learning: by creating training data (in appropriate input format!) one can learn the input-output map.
- Applications: solutions of stochastic differential equations (Friz, Lyons, Victoir, etc), solutions of correlated filtering problems (Crisan, Friz, etc), solutions of sub-critical stochastic partial differential equations (Gubinelli, Hairer, Perkowski, etc), solutions of stochastic optimization problems and stochastic games.

Structure of non-linear solution maps

Consider a regularity structure (A, T, G) and a solution map (related to a fixed problem, e.g. a prediction task) $f : \mathcal{M} \to \mathcal{D}^\gamma$ for some $\gamma > 0$. The space $\mathcal{M}$ is *a metric space of models* on which f depends in a continuous way (Notice that we are not dealing with re-normalization procedures here where non-continuities of the solution map would have to be re-normalized).

By the previous results and their respective proofs we know that the re-constructed solution has the form

$$Z \mapsto \mathcal{R}(f(Z)) = \sum_{n,x} \langle \Pi_x f(Z)(x), \psi_x^n \rangle \psi_x^n$$

where convergence is understood in a respective Besov space topology. The essential point is now to expand $Z \mapsto \langle \Pi_x f(Z)(x), \psi_x^n \rangle$ with respect to a *reservoir*, i.e. a dense subspace in the space of continuous functions on $\mathcal{M}$.

Reservoir

Definition

A set of continuous functions $(R_n)_{n\geq 0}$ such that

- the linear span of $(R_n)_{n\geq 0}$ is an algebra of continuous functions Res,
- for every compact set K there is a dense subset $L \subset K$ such that $\mathrm{Res}\,|_L$ is point separating on L,

is called a *reservoir*.

From Hambly-Lyons (2006) the signature of a rough path of order γ, i.e. the set of all iterated integrals of order $\lfloor \frac{1}{\gamma} \rfloor$ up to time t, spans a point separating algebra on smooth paths without excursions (which are dense with respect to the rough path topology) up to time t.

Whence signature is a reservoir for the previously introduced regularity structure of rough paths.

Existence of Reservoirs I

Theorem (Martin Hairer (2014))

Let $\mathcal{H}$ be a $\mathbb{Z}_+^d$-graded, commutative Hopf algebra with product $\star$, coproduct $\circ$ and $\alpha = (\alpha_i) \in \mathbb{R}_{>0}^d$ some coefficients, then

$$A := \{\langle \alpha, k \rangle \mid k \in \mathbb{Z}^d\}$$

is a discrete set, $T_\beta = \bigoplus_{\beta = \langle k, \alpha \rangle} \mathcal{H}_k$, for $\gamma \in A$, defines a regularity structure with structure group $G = \exp(\mathfrak{g})$, where $\mathfrak{g}$ is the Lie algebra of primitive elements in $\mathcal{H}_*$, acting via on T (as algebra morphisms)

$$\langle l \circ g, f \rangle = \langle l, \Gamma_g f \rangle$$

for $g \in G$, $l \in \mathcal{H}_*$ and $f \in \mathcal{H}$ (Notice that $\mathcal{H}_*$ is the universal enveloping algebra of $\mathfrak{g}$). Then the maps $g \mapsto \langle l, \Gamma_g f \rangle$ on G are closed under multiplication, i.e. generate an algebra.

Existence of Reservoirs II

The previous theorem and obvious ramifications beyond Hopf algebras of it allow to write (non-linear) polynomial functions on G as restrictions of linear ones, which in turn leads to an easily described algebra of functions on models. The phenomenon appears when the elements of T have word character with a not-necessarily globally defined commutative product $\star$ and a compatible co-product Δ such that the acting group is a group of primitive elements.

Since models will appear as maps $x, y \mapsto g_x \circ g_y^{-1}$ the point separating property translates to the question whether a map (in a large class of maps) of type $x \mapsto g_x inG$ is already characterized by a finite number of points on the respective graph.

Existence of Reservoirs III

Theorem

Let (A, T, G) be a regularity structure and T_* a pre-dual space of T, which is additionally an algebra (with product $\circ$), whose group of unitals contains G as adjoints of left translations.
If the adjoint Δ of $\star$ satisfies

- $\Delta(g) = g \otimes g$,
- $\Delta(g_1 \circ g_2) = \Delta(g_1) \circ \Delta(g_2)$

for unitals $g, g_1, g_2 \in G$, then

$$\langle I \circ g, f_1 \rangle \langle I \circ g, f_2 \rangle = \langle I \circ g, f_1 \star f_2 \rangle$$

for all $I, g \in G$ and $f_i \in T$. Take (Π, Γ) a model of the regularity structure and fix $x, y \in \mathbb{R}^d$, then the span of $I \mapsto \langle I \circ \Gamma_{xy}, f \rangle$ for $I \in G$, $f \in T$ form an algebra.

Prediction Tasks

... in the future could be seen as follows:

- consider certain noisy observations of a true signal and model their structure by a corresponding regularity structure (this might be necessary in since there is no reason why non-linear functions of noisy objects should be well defined at all).
- construct solutions of the optimal filter by solving a fixed point equation on modelled distributions.
- reconstruct the real world filter by the reconstruction operator, which yields – under appropriate regularity conditions – a non-linear, Lipschitz map from the space of observations (the 'models') to the optimal filter.
- construct a reservoir and an efficient way how to compute its values on the noisy signal.
- learn the optimal filter by regressing on the reservoir.

References

- M. Hairer:
  Theory of regularity structures, Inventiones mathematicae, 2014.