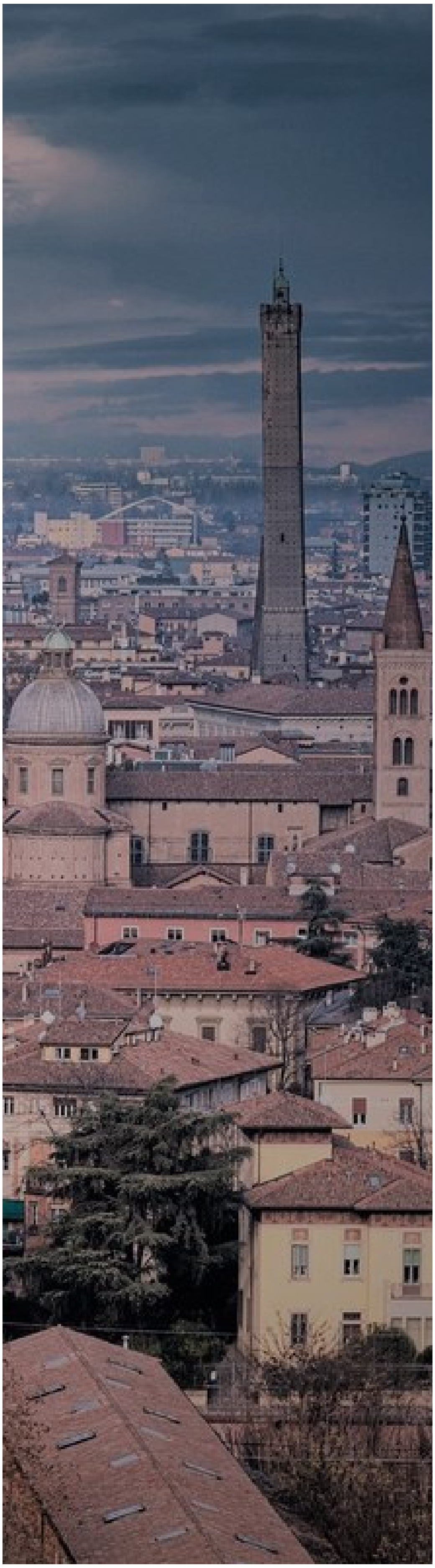
Book of abstract

PRIMO Workshop 2021

PRIMO WORKSHOP



Bologna, October 11-13 2021

Keynote speakers:

Daniela di Serafino

University of Naples Federico II

Silvia Gazzola

University of Bath

Organizing Committee:

Alessandro Buccini University of Cagliari

Pasquale Cascarano University of Bologna

Serena Crisci University of Campania "L. Vanvitelli"

Gianmarco Gurioli SISSA

Monica Pragliola University of Bologna

Marco Viola University of Campania "L. Vanvitelli"

Supported by:

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
DEPARTMENT OF MATHEMATICS

[AM]²

MiVA

https://events.unibo.it/primo-workshop-2021

Sparse Approximations with Interior Point Methods

Daniela di Serafino
Department of Mathematics and Applications, University of Naples Federico II, Italy

Abstract. Large-scale optimization problems that seek sparse solutions have become ubiquitous. They are routinely solved with various specialized first-order methods. Although such methods are often fast, they usually struggle with not-so-well conditioned problems. In this talk, specialized variants of an interior point-proximal method of multipliers are proposed and analyzed for problems of this class. Computational experience on a variety of problems, namely, multi-period portfolio optimization, classification of data coming from functional Magnetic Resonance Imaging, restoration of images corrupted by Poisson noise, and classification via regularized logistic regression, provides substantial evidence that interior point methods, equipped with suitable linear algebra, can offer a noticeable advantage over first-order approaches. This is joint work with V. De Simone and M. Viola (University of Campania 'L. Vanvitelli', Italy) and with J. Gondzio and S. Pougkakiotis (University of Edinburgh, UK).

Hybrid projection methods for large-scale linear inverse problems.

Silvia Gazzola
Department of Mathematical Sciences, University of Bath, UK

Abstract. Inverse problems are ubiquitous in many areas of Science and Engineering and, once discretized, they lead to ill-conditioned linear systems, often of huge dimensions: regularization consists in replacing the original system by a nearby problem with better numerical properties, in order to find a meaningful approximation of its solution. After briefly surveying some standard regularization methods, both iterative (such as many Krylov methods) and direct (such as Tikhonov method), this talk will introduce a recent class hybrid projection methods, which merge an iterative and a direct approach to regularization. In particular, strategies for choosing the regularization parameter and the regularization matrix will be emphasized, eventually leading to the computation of approximate solutions of Tikhonov problems involving a regularization term expressed in a p-norm.

Data-Driven Regularization by Projection

Andrea Aspri
Department of Mathematics, University of Pavia, Italy

Abstract. In this talk I will speak about some recent results on the study of linear inverse problems under the premise that the forward operator is not at hand but given indirectly through some input-output training pairs. We show that regularisation by projection and variational regularisation can be formulated by using the training data only and without making use of the forward operator. I will provide some information regarding convergence and stability of the regularised solutions. Moreover, we show, analytically and numerically, that regularisation by projection is indeed capable of learning linear operators, such as the Radon transform. This is a joint work with Yury Korolev (University of Cambridge) and Otmar Scherzer (University of Vienna and RICAM).

A Local–Global Graph Approach for Coloured Image Segmentation

Alessandro Benfenati
Department of politicy and environmental science, University of Milan, Italy

Abstract. Image segmentation is an essential component in several computer vision areas, such as image analysis, pattern recognition, robotic systems. Colour based segmentation offers more significant extraction of information as compared to gray–intensity based segmentation. This work propose a new method which combines a random walk based model [1] with a direct label assignment computed by employing a suitable color distance. This approach requires user interaction for the initialization of the segmentation process, hence it falls under the umbrella of semi–automatic techniques. The random walk part involves a combinatorial Dirichlet problem for a weighted graph, where the nodes are the pixels of the image, whilst the positive weights are related to the distances between pixels: a novel color distance for computing such weights is employed for computing such distances. In the random walker model a probability is assigned to each pixel: such probability quantifies the likelihood that the node belongs to some subregion. The computation of the color distance is pursued by employing the coordinates of a pixel in a suitable color space (RGB, Lab, YCbCr) and of the ones in its neighbourhood. The segmentation process hence consists in an optimization problem involving both the probabilities from the random walker approach and the similarity with respect the labelled pixels. We explore also some Machine Learning strategies for learning suitable weights to be used in the computation of the color distance. WBC [3] and GrabCut [4] datasets are employed to assess the performance of the proposed method, which is compared with state–of–the art approaches, such as NRW, NLRW [2]. The experimental results show that the proposed approach achieves remarkable results wrt to the other algorithms. Moreover, it reveals to be very robust with respect to the presence of noise and to the choice of the colorspace.

References

[1] Grady, L. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Ma chine Intelligence, 28, 1768–1783, 2006.

[2] Bampis, C.G., Maragos, P., Bovik, A.C. Graph-Driven Diffusion and Random Walk Schemes for Image Segmentation. IEEE Transactions on Image Processing, 26, 35–50, 2017.

[3] Zheng, X., Wang, Y.l, Wang, G., Liu, J. Fast and Robust Segmentation of White Blood Cell Images by Self-supervised Learning. Micron, 107, 55–71, 2018.

A Tensor-Train Dictionary Learning algorithm based on spectral proximal alternating linearized minimization

Domitilla Brandoni
Departmente of Mathematics, University of Bologna, Bologna, Italy

Abstract. Dictionary Learning (DL) is one of the leading sparsity promoting techniques in the context of image classification, where the 'dictionary' matrix D of images and the sparse matrix X are determined so as to represent a redundant image dataset Y. The resulting constrained optimization problem $\min_{D,X} \|Y - DX\|_F$ is nonconvex, non-smooth and NP-hard, providing several computational challenges for its solution (see e.g. [1]). To preserve multidimensional data features, various tensor DL formulations have been introduced, adding to the problem complexity (see e.g. [2]). Unfortunately all the tensor-based DL methods in the literature are not supported with theoretical convergence analysis. We propose a new tensor formulation of the DL problem using a Tensor-Train decomposition ([3]) of the multi-dimensional dictionary, together with a new alternating algorithm for its solution. The new method belongs to the Proximal Alternating Linearized Minimization (PALM) algorithmic family (see e.g. [4]), with the inclusion of second order information to enhance efficiency. We discuss a rigorous convergence analysis, and report on the new method performance on the image classification of several benchmark datasets. This talk is based on [5].

References

[1] Mairal, J., Bach, F.,Ponce, J. Sparse Modeling for Image and Vision Processing. Found. Trends. Comput. Graph. Vis. 8(2-3), 85–283, 2014.

[2] Dantas, C. F., Cohen, J. E., Gribonval, R. Learning tensor-structured dictionaries with application to hyperspectral image denoising. In 27th European Signal Processing Conference (EUSIPCO) pp. 1-5, IEEE, 2019.

[3] Oseledets, I. V. Tensor-train decomposition. pp. 2295–2317, SIAM J. Sci. Comput., 33, 2011.

[4] Bolte, J., Sabach, S., Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Prog., 146(1), 459–494, 2014.

[5] Brandoni, D., Porcelli, M., Simoncini, V. A Tensor-Train Dictionary Learning algorithm based on Spectral Proximal Alternating Linearized Minimization. arXiv: 2107.11644, 2021.

Regularization graphs for inverse problems

Marcello Carioni
Department of Mathematics, University of Cambridge, UK

Abstract. In this talk we introduce a mathematical framework for a broad class of regularization functionals for ill-posed inverse problems: Regularization Graphs. Regularization graphs allow to construct functionals using as building blocks linear operators and convex functionals, assembled by means of operators that can be seen as generalizations of classical infimal convolution operators. This class of functionals exhaustively covers existing regularization approaches and it is flexible enough to craft new ones in a simple and constructive way. We provide well-posedness and convergence results with the proposed class of functionals in a general setting. Further, we consider a bilevel optimization approach to learn optimal weights for such regularization graphs from training data. We demonstrate that this approach is capable of optimizing the structure and the complexity of a regularization graph, allowing, for example, to automatically select a combination of regularizers that is optimal for given training data.

Regularizing ill-posed Inverse problems in imaging using Deep Learning

Francesco Colibazzi
Department of Mathematics, Unviersity of Bologna, Italy

Most inverse problems in imaging are ill-posed and require appropriate regularization treatment for recovering meaningful solutions. Traditional methods minimize a cost function that consists of a data-fit term, which measures how well the reconstructed image matches the observations, and a regularizer, which reflects prior knowledge and promotes images with desirable properties. Recently, data-driven methods using deep learning techniques demonstrated to significantly outperform these classical solution methods. A key distinction between many types of learned inverse problem solvers is what is known about the forward model. In this talk, we review deep-learning regularization techniques for solving inverse problems in imaging: in particular, we focus on how to combine data-driven and model-based methods to strengthen the stability of the solution.

Discontinuous Neural Networks and Discontinuity Learning

Francesco Della Santa
Polytechnic University of Turin, Turin, Italy

Abstract. In this work, we introduce a new typology of layers for Neural Networks (NNs); the nov elty is given by the introduction of discontinuities in the function characterizing the layer's action, obtaining a discontinuous Neural Network. In the framework of NNs, discontinuities were involved in the first mathematical models [3, 6, 8], because activation functions of the NN units were mainly inspired by the mechanisms of the biological neurons; then, the Heaviside step function, or suitable variants, were used. Nonetheless, in the following decades, the activation functions evolved into the continuous (and often smooth) ones used nowadays in almost all the deep learning algorithms. To the best of the authors' knowledge, in recent literature do not exist examples of discontinuous NN layers or discontinuous NNs, with the exception of [7] and [4] that study the theoretical consequences for the approximation task of using the floor and the Heaviside functions, respectively, as activation functions. We are interested in introducing discontinuities in NN learning models to approximate discontinuous functions and, at the same time, detect the discontinuity interfaces. This lat ter problem is a challenging task, especially for functions with a high-dimensional domain and, moreover, the information can be quite relevant in several applications (e.g., numerical methods for stochastic collocation in the framework of uncertainty quantification). Then, knowing the discontinuity interfaces and the domain regions where the function is smooth, can be of paramount importance. Nonetheless, the discontinuity interfaces detection prob lem is a very hard task and the existing methods (such as [1, 2]) obtain good results only for $n = 1, 2$ and $m = 1$ or, using locally adaptive approaches and under particular assumptions on the function, for $n \geq 2$ and $m = 1$. The existing approximation methods based on NNs are not suitable to effectively tackle general discontinuous functions while simultaneously detecting their discontinuity inter faces. Indeed, approximating a discontinuous function with a NN is quite a simple task [5], but the function represented by the NN will actually be continuous, if "classic" contin uous activation functions are used. The present work aims at building new discontinuous NNs able to approximate discontinuous functions with other discontinuous functions, whose discontinuity interfaces are relatively easy to be detected.

References

[1] Archibald, R., Gelb, A., Saxena, R., Xiu, D. Discontinuity detection in multivariate space for stochastic simulations. Journal of Computational Physics, 228, pp. 2676-2689, 2009.

[2] Jakeman, J. D., Archibald, R., Xiu, D. Characterization of discontinuities in high dimensional stochastic problems on adaptive sparse grids. Journal of Computational Physics, 230, pp. 3977-3997, 2011.

[3] McCulloch, W. S., Pitts, W. H. A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, 5, pp. 115-133, 1943.

[4] Park, S., Yun, C., Lee, J., Shin, J. Minimum Width for Universal Approximation. In: International Conference on Learning Representations, 2021.

[5] Petersen, P., Voigtlaender, F. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. Neural Networks, 108, pp. 296-330, 2018.

[6] Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, 65, pp. 386-408, 1958.

[7] Shen, Z., Yang, H., Zhang, S. Deep Network With Approximation Error Being Reciprocal of Width to Power of Square Root of Depth. Neural Computation, 33, pp. 1005-1036, 2021.

[8] Widrow, B., Hoff, M. E. Adaptive switching circuits. IREWESCON Convention Record, 4, pp. 96-104, 1960.

Identifying a conductive and permeable sphere by exponential sums

Patricia Diaz de Alba
Gran Sasso Scientific Institute, L'Aquila, Italy

Abstract. The aim of this talk is to present and compare two different numerical approaches to identify the radius, the magnetic permeability, and the electrical conductivity of a conductive and permeable sphere from its time-domain electromagnetic response. The starting point is a model mathematically represented by an exponential time series whose coefficients and exponents depend on the three parameters that we want to recover. The first technique, which has been already applied to this problem, used to approximate these coefficients and exponents is the Prony method whereas the second one, never experimented in this context, is a matrix pencil method. Finally, by solving a system of nonlinear equations involving the recovered coefficients, the three physical parameters of the sphere are identified by Newton's method. Numerical experiments verify the effectiveness of both methods showing a better stability of the second approach.

All You Can Embed: Natural Language based Vehicle Retrieval with Spatio-Temporal Transformers

Giorgia Franchini
Department of Mathematics, University of Ferrara, Ferrara, Italy

Abstract. Combining Natural Language with Vision represents a unique and interesting challenge in the domain of Artificial Intelligence. This work focuses on the problem of combining visual and textual information, applied to a smart-city use case. In this talk, I present All You Can Embed (AYCE), a modular solution to correlate single vehicle tracking sequences with natural language. The main building blocks of the proposed architecture are (i) BERT to provide an embedding of the textual descriptions, (ii) a convolutional backbone along with a Transformer model to embed the visual information. For the training of the retrieval model, a variation of the Triplet Margin Loss is proposed to learn a distance measure between the visual and language embeddings.

Recent Trends in Adaptive Regularisation under Inexact Evaluations for Nonconvex Optimisation and Machine Learning Applications

Gianmarco Gurioli

Abstract. Within the context of nonconvex unconstrained and inexpensively-constrained optimisation, a class of adaptive regularisation methods under inexact function and derivatives evaluations is presented. At variance with the ARC framework, the underlying algorithm is not limited to refer to the cubic model, allowing for the use of potentially higher degrees to search for arbitrary order optimality points. At each iteration, it features an adaptive mechanism for determining the inexactness which is needed to compute objective function values and derivatives, in order to preserve the complexity results of its counterpart with exact evaluations. Sharp global evaluation complexity bounds, assuming that the right accuracy level in function and derivatives estimates is deterministically achievable, are derived and hold for any model degree and any order of optimality, thereby generalising known results for first and second-order versions of the method. High probability and stochastic complexity bounds are also shown. For lower orders, preliminary numerical tests are finally reported.

Proximal algorithms in variable exponents Lebesgue spaces

Marta Lazzaretti
University of Genoa, Genoa, Italy

Abstract. We consider a convex optimisation problem in variable exponents Lebesgue spaces written in the form

$$\arg \min_{x \in L^{p(\cdot)}(\Omega)} \phi(x) := f(x) + g(x) \tag{P}$$

where

- Ω, $n \in \mathbb{N}$, $n \geq 2$ is a Lebesgue measurable subset with positive measure, $p(\cdot) : \Omega \to [1, +\infty]$ is a Lebesgue measurable function which characterizes the variable exponents Lebesgue space $L_p^{(\cdot)}(\Omega)$;
- $f : L^{p(\cdot)}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous, proper, convex, differentiable;
- $g : L^{p(\cdot)}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous, proper, convex.

The setting of the addressed problem is unusual: the functions are not defined on Hilbert spaces, as it is traditionally considered, but on $L^{p(\cdot)}(\Omega)$, which is a Banach space and, under certain assumptions on the exponent function $p(\cdot)$, is also reflexive, strictly convex and smooth.

An effective way to tackle this minimization problem in a Hilbert space setting is the forward-backward splitting algorithm [2], which enables to exploit the differentiability of the smooth function f decoupling the contributions of f and g in a forward step defined in terms of the gradient of f $x^k \mapsto x^k - \tau_k \nabla f(x^k)$ and in a backward step defined in terms of g and a proximal operator $x^{k+1} = \text{prox}_{\tau_k g}(x^k - \tau_k \nabla f(x^k))$ with $\text{prox}_{\gamma g}(u) \arg\min_{x \in \mathcal{H}} \frac{1}{2}\|x - u\|_{\mathcal{H}}^2 + \gamma g(x)$ for $\gamma > 0$.

The generalisation of this algorithm to a Banach space $\mathcal{X}$ is more complicated because the Riesz representation theorem does not hold and thus the dual space $\mathcal{X}^*$ cannot be identified with the space $\mathcal{X}$ itself. This affects the applicability of the forward-backward algorithm: the gradient is an element of the dual space and thus the forward step cannot be performed directly on x^k. To do so, it is necessary to introduce the duality mappings, which link primal and dual space, allowing to perform the gradient step in the dual. In [1, 3, 4], forward-backward splitting algorithms have been proposed to solve minimization problems of the sum of two functions in a reflexive, strictly convex and smooth Banach space $\mathcal{X}$. They involve duality mappings and define the backward step in terms of the Bregman distance in [3, 4]. In [1], forward and backward steps are no longer performed separately, but a different type of iteration is performed, taking into account both f and g at the same time. It is important to point out that duality mappings are non-linear and this introduces new challenges in the definition of a backward step and in the convergence analysis of the algorithms.

These algorithms have been proposed for reflexive, strictly convex and smooth Banach spaces and thus a priori they can be used also in $L^{p(\cdot)}(\Omega)$ if certain assumptions on the variable exponent $p(\cdot)$ are met. However, the definition of the

norm $\|\cdot\|_{p(\cdot)}$ in $L^{p(\cdot)}(\Omega)$ makes their use impracticable, since the norm (and the duality mapping as well) in this setting are non-separable. In many other Banach spaces, this is not an issue as the norm is usually separable. For instance, in $L^p(\Omega)$ spaces with $p > 1$ both the norm and the duality mappings are separable and this algorithms can be easily implemented, studying the one-dimensional problem and then generalising to the multi-dimensional scenario by taking advantage of the separability of all the involved objects. In $L^{p(\cdot)}(\Omega)$ spaces, this strategy fails due to the non-separability. Inspired by the previous works of Bredies [1] and Guan and Song [3], we propose an iterative forward-backward splitting procedure where the role of the norm in $L^{p(\cdot)}(\Omega)$ spaces is the replaced by the modular, that is a separable function with similar to the norm properties, defined by

$$\bar{\rho}_{p(\cdot)}(x) = \int_\Omega \frac{1}{p(t)}|x(t)|^{p(t)}dt\,, \quad x \in L^{p(\cdot)}(\Omega)\,. \tag{1}$$

We will present some basic properties of Banach spaces in general and of variable exponents Lebesgue spaces, focusing in particular on the non differentiability of the norm and duality mappings. We will detail the proposed algorithm and discuss its convergence properties. To conclude, we point out that the use of variable exponents Lebesgue spaces is motivated by their space variant properties which are interesting to consider in the regularisation of inverse problems.

References

[1] Bredies,K. A forward–backward splitting algorithm for the minimization of non smooth convex functionals in banach space. Inverse Problems, 25(1):015005, Nov 2008.

[2] Combettes, P. L., Wajs, V. R. Signal recovery by proximal forward backward splitting. Multiscale Modeling & Simulation, 4(4):1168–1200, 2005.

[3] Guan, W., Song, W. The generalized forward-backward splitting method for the minimization of the sum of two functions in banach spaces. Numerical Functional Analysis and Optimization, 36(7):867–886, 2015.

[4] Guan, W., Song, W. The forward–backward splitting method and its conver gence rate for the minimization of the sum of two functions in banach spaces. Optimization Letters, 15:1735–1758, 2021.

Parameter-Robust Preconditioning for Oseen Iteration Applied to Navier–Stokes Control

Santolo Leveque
Departmente of Mathematics, Unviersity of Edinburgh, Edinburgh, UK

Abstract. In this talk I will present novel, fast, and parameter-robust preconditioned iterative methods for Navier–Stokes control problems. The key ingredients of the solver are a saddle-point type approximation for the linear systems, an inner iteration for the (1, 1)-block accelerated by a preconditioner for convection–diffusion control, and an approximation to the Schur complement based on a potent commutator argument applied to an appropriate block matrix. A range of numerical experiments validate the effectiveness of our new approach.

Iterative regularization with convex regularizers

Cesare Molinari
Italian Institue of Technology, Genoa, Italy

Abstract. In this talk, we present iterative regularization for linear models, when the bias is convex but not necessarily strongly convex. We characterize the stability properties of a primal-dual gradient based approach, analyzing its convergence in the presence of worst case deterministic noise. As a main example, we specialize and illustrate the results for the problem of robust sparse recovery. Key to our analysis is a combination of ideas from regularization theory and optimization in the presence of errors. Theoretical results are complemented by experiments showing that state-of-the-art performances are achieved with considerable computational speed-ups.

Unifying convergence analysis for degenerate preconditioned proximal point algorithms

Emanuele Naldi
Department of Mathematics, Technische Universität Braunschweig, Braunschweig Germany

Abstract. In this talk we describe a systematic procedure to tackle degenerate preconditioned proximal point algorithms, used to solve monotone inclusion problems. A usual hypothesis for the preconditioner is to be linear, bounded, self-adjoint and strongly positive definite. We show that it is possible to relax the last condition and make use of preconditioners which are only positive semi-definite, allowing in this way degenerate preconditioners with a possibly large kernel. We establish weak convergence results under mild assumptions that can be easily employed in the context of splitting methods for monotone inclusion and convex minimization problems. Moreover, we show that the degeneracy of the preconditioner allows for a reduction of the variables involved in the iteration updates. We show the strength

of the proposed framework, giving a new perspective on existing splitting schemes such as Douglas-Rachford, Chambolle-Pock and Forward Douglas-Rachford, simplifying the analysis of convergence, and paving the way for new algorithms that can also allow tackling objective functions composed by many terms.

Regularization of a system of Fredholm integral equations of the first kind

Federica Pes
Department of Mathematics, University of Cagliari, Cagliari, Italy

Abstract. Systems of first kind integral equations arise in many applications. It is well-known that Fredholm integral equations of the first kind are often ill-posed problems. When the right-hand side is only known at a finite set of points, e.g., when it consists of experimental measurements, the difficulties related to ill-posedness are enforced, as the problem has infinitely many solutions. We propose a numerical method to compute the minimal-norm solution of systems of integral equations of the first kind in the presence of boundary constraints. The problem is solved in a reproducing kernel Hilbert space (RKHS), by using the Riesz Representation Theorem. Indeed, the minimal-norm solution is written as a linear combination of the Riesz representers. Since the resulting linear system is strongly ill-conditioned, we construct a regularization method based on a truncated expansion of the minimal-norm solution in terms of the singular functions of the integral operator. Numerical experiments are presented to illustrate the excellent performance of the method. This is a joint work with Patricia Diaz de Alba, Luisa Fermo, and Giuseppe Rodriguez.

Variational Physics Informed Neural Networks: the role of quadratures and test functions

Moreno Pintore
Polytechnic University of Turin, Turin, Italy

Abstract. Physics Informed Neural Networks (PINNs) are neural networks used to approximate the solution of differential equations exploiting the equation itself or some known solutions. Since it has been observed that they can efficiently approximate the solution of nonlinear, high-dimensional or parametric problems, their popularity is increasing and different types of networks have been developed; one of them is the Variational Physics Informed Neural Network (VPINN). The main difference between PINNS and VPINNs is the fact that the former are trained minimizing the strong form of the equation, while the latter focus on its variational formulation. We analyzed the relation between the VPINN accuracy and the way in which it is trained while solving second order elliptic boundary-value problems. In particular, the VPINN is trained minimizing the integral residuals against a predefined set of test functions on a fixed mesh. We prove that, under suitable assumptions, the error decay rate with respect to the mesh size h is of order

$O(h^{k_{int}})$, where $k_{int} = q - k_{test} + 2$. Here q is the order of the chosen quadrature rule and k_{test} the one of the test functions. Such an estimate suggests that, in order to increase the convergence rate, it is convenient to raise the quadrature rule order as much as possible and using piecewise linear test functions.

To derive such an a priori error estimate, we considered a k_{int}-order piecewise interpolant of the VPINN to satisfy the required inf-sup condition. Nevertheless, even if the a priori estimate has been proved only when the VPINN is interpolated, we observed almost identical convergence rates for non interpolated VPINNs. We also empirically show that such a condition is necessary to avoid spurious modes. The error estimate is proved for classical feed-forward fully connected neural networks, the only required assumption is on its dimension, which has to be large enough to ensure good approximation properties. Several numerical experiments have been performed to confirm the theoretical prediction on different test cases.

A piecewise conservative method for unconstrained convex optimization

Alessandro Scagliotti
Department of Mathematics, SISSA, Trieste, Italy

Abstract. In the last years Theory of Dynamical Systems has been fruitfully applied to study existing accelerated optimization methods and to develop new ones. Typically, continuous-time models for accelerated methods are mechanical systems with damping. In this talk we present an optimization method based on a conservative mechanical system, where the objective function plays the role of the potential energy. Due to the absence of damping, the convergence of this method completely relies on the restart strategy:

- the initial velocity is set equal to zero;
- by the conservation of the mechanical energy, part of the initial potential energy is transformed into kinetic energy;
- when a proper restart condition is met, the velocity is reset to zero and the kinetic energy at the restart time is instantly dissipated.

We prove the convergence result both for the continuous-time method and for the discrete-time version. Finally, we discuss some possible extensions to the nonsmooth case, with particular focus on l1 regularization. Based on a joint work with Prof. Piero Colli Franzone.

Continuous Injective Generative Neural Network

SIlvia Sciutto
Department of Mathematics, University of Genoa, Genoa, Italy

Abstract. We present a Continuous Generative Neural Network (CGNN), showing stability results for inverse problems in infinite dimensional spaces. Inspired by the U-Net architecture, a widely used discrete neural network, we choose to use a generator that involves a fully connected layer and many convolutional layers. We find out that there is a relation between the convolutional layers stride, in the neural network world, and the scale, concept related to the scaling function spaces of wavelet analysis. This ensures a sort of equivalence between the presented continuous model and the discrete one. We discuss the possible conditions under which the CGNN is globally injective, in order to guarantee the uniqueness in the representation of the generated functions or images. Moreover, injectivity is essential to obtain Lipschitz stability properties of our algorithm for solving inverse problems. These conditions involves both the convolutional filters com ponents, the non linearity and the fully connected layer. We explain how we apply the generator idea in our algorithm to solve ill-posed inverse problems. This approach consists of a learning part, that involves only the training of our generator, and an iterative part, that uses a Landweber-type algorithm on a low dimensional latent space (i.e. the generator input space). We also show, as example of how our algorithm works on this kind of problems, the Calderón problem of electrical impedance tomography (EIT), that consists in reconstructing the conductivity of a material.

Constrained and unconstrained Deep Image Prior optimization models with automatic regularization

Andrea Sebastiani
Department of Mathematics, University of Bologna, Bologna, Italy

Abstract. Deep Image Prior (DIP), proposed in [1], is one of the most useful unsupervised deep learning methods for imaging inverse problems. This framework use a Convolutional Neural Network (CNN) as implicit prior to model and represent all the images. Combined with classical and novel regularizers, DIP has been shown to be a very e ective unsupervised approach. However, all the regularized methods, proposed up to now, require a good estimation of the regularization parameter in order to balance the contribute of the regularizer. The goal of this work is to overcome this drawback, proposing two different models formulated as an unconstrained and a constrained optimization problems. In particular, the unconstrained model, introduced in [2], combines DIP with a space-variant Total Variation regularizer with an automatic estimation of the local regularization parameters. Conversely, the constrained formulation it is derived from the Morozov's discrepancy principle considering an estimation of the noise level in the acquired image. This approach avoids to properly set the regularization parameter and it also allows to consider di erent regularizers by making the framework very general and more versatile. The

arising minimization problems are both solved via the exible Alternating Direction Method of Multipliers (ADMM). Moreover, since the proposed methods can be viewed as proximal gradient descent-ascent (GDA) algorithms, we show that it is possible to provide a convergence result upon suitable assumptions. The promising performances of the proposed approaches are assessed, in terms of PSNR and SSIM values, by several experiments both on the denoising and on the deblur of simulated as well as real natural and medical corrupted images.

References:

[1] Ulyanov, D. et al. Deep image prior , in Proceedings for IEEE Conference on Computer Vision and Pattern Recognition. pp. 9446 9454, 2018.

[2] P. Cascarano et al. Combining Weighted Total Variation and Deep Image Prior for natural and medical image restoration via ADMM . arXiv preprint arXiv:2009.11380, 2020.

Solving systems of nonlinear equations via spectral residual methods

Cristina Sgattoni
CNR, Italy

Abstract. This talk addresses the numerical solution of systems of nonlinear equations via spectral residual methods. Spectral residual methods are iterative procedures, they use the residual vector as search direction and a spectral steplength, i.e., a steplength that is related to the spectrum of the average matrices associated to the Jacobian matrix of the system. Such procedures are derivative-free and low-cost per iteration. The first aim of the seminar is to analyze the properties of the spectral residual steplengths and study how they affect the performance of the methods. The second purpose is to propose a variant of the derivative-free spectral residual method used in the first part and obtain a general scheme globally convergent under more general conditions. The robustness of the new method is potentially improved with respect to the previous version. Numerical experiments are conducted both on the problems arising in rolling contact models and on a set of problems commonly used for testing solvers for nonlinear systems.

Stochastic quasi-Newton methods for training deep neural networks

Mahsa Yousefi
Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy

Abstract. Many large scale problems, such as image recognition, can be solved by a leading technique of machine learning (ML) called deep learning (DL). Training of a deep neural network (DNN) is often formulated as the solution of a empirical risk minimization problem resulting in a nonlinear and non-convex optimization problem whose objective function is a sum of N loss functions, one for each data

sample. In DL applications, both the number of samples N and the number of variables of the objective function (depending on the type and depth of the network) can be extremely large, so that computing the gradient is expensive while the Hessian can not be computed in practice. Stochastic optimization methods try to overcome this challenge. Over the last decade, stochastic first-order op timization methods (e.g. stochastic gradient descent and its variants) have emerged as the canonical tool for training DNNs. These methods are particularly appealing due to their low runtime and memory costs. However, they come with their own limitations, due to their undesirable effect of not escaping saddle points and the need of exhaustive trial and error to fine-tune their hyper-parameters. There has been a lot of effort to find ways to incorporate second-order information and limited memory quasi-Newton methods have recently attracted much attention; see e.g. [2]. In their stochastic variants, suitable sampling strategies [1] should be employed to avoid instabilities in the computation of gradient differences thus allowing for better Hessian approximations. In this talk, we will describe practical stochastic quasi-Newton algorithms relying on a stable sampling strategy for computing the curvature pairs needed for two well-known limited memory quasi-Newton updates, L-BFGS and L-SR1, in a trust region framework (TR). We will show the results obtained in the training of several different DNNs architec tures, ranging from a shallow LeNet-like network, convolutional neural networks (CNNs) with and without batch normalization layers, up to a modern residual network (ResNet 20), for image classification problems. We will include results showing the influence of the limited memory parameter and the batch size on the achievable training and testing accuracies and training time. We will analyze whether faster training can be achieved by using more stable positive definite rank-two BFGS updates or cheaper symmetric rank-one (SR1) updates that, allowing for indefinite Hessian approximations, could hopefully better capture the curvature of the true Hessian.

References

[1] Berahas, A. S., Takáč, M. A robust multi-batch L-BFGS method for machine learning. Optimization Methods and Software, 35(1):191–219, 2020.

[2] Erway, J. B., Griffin, J., Marcia, R. F., Omheni, R.. Trust-region algo rithms for training responses: machine learning methods using indefinite hessian approxima tions. Optimization Methods and Software, 35(3):460–487, 2020.

[3] Martìnez, A., Yousefi, M. Deep neural network training with stochastic Quasi-Newton methods: an experimental study. In preparation.