Bilevel learning of hyper-parameter estimation in image reconstruction problems

M2 internship proposal for spring 2024 (Duration: 5/6 months)

Supervisors: Luca Calatroni (calatroni@i3s.unice.fr), Samuel Vaiter (samuel.vaiter@univ-cotedazur.fr)
Hosting labs: Morpheme group (CNRS, Inria, Sophia-Antipolis, FR) & LJAD (UniCA, Nice, France)

Context

In the framework of regularisation approaches for imaging ill-posed problems, given a (vectorised) imperfect image $y \in \mathbb{R}^m$ possibly affected by blur, noise and under-sampling modelled by the action of a linear operator $A \in \mathbb{R}^{m\times n}$, a reconstructed image $x \in \mathbb{R}^n$ is found by solving the problem:

$$x(\theta) \in \underset{v\in\mathbb{R}^n}{\operatorname{argmin}} \ \left\{ \mathcal{J}_\theta(v;y) := R_{\theta_1}(v) + \Psi_{\theta_2}(Av;y) \right\}, \tag{1}$$

where $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^{N_\theta}, N_\theta > 1$ denotes the vector of hyper-parameters associated to the cost functional, $R_{\theta_1}(\cdot)$ denotes a regularisation functional parametrised by θ_1 and combined with an appropriate data-fidelity term $\Psi_{\theta_2}(A\cdot;y)$ whose form depends on the assumed noise statistics. For good performance, an accurate choice of the terms R, Ψ and of the hyper-parameters θ has to be made. Over the years, increasingly complex hand-crafted regularisation and data fidelity terms R and Ψ suited each for specific classes of imaging problems have been considered and thoroughly studied. With the intent of defining more data-adaptive regularisation models, new approaches aiming at estimating highly-parametrised (hence, flexible) regularisers *from the data* rapidly developed [1]. Standard *a posteriori*, *a priori* and *heuristics* parameter selection strategies can be used. Notable examples are the discrepancy principle, the SURE criterion, the L-curve. In most cases, however, such approaches require either the prior knowledge of additional information on the noise level and/or the use of large matrix/vector manipulations. A different approach for hyper-parameter estimation which combines model- with data-driven approaches in a shallow way consists in the estimation of optimal model hyper-parameters $\hat{\theta}$ by solving a further (upper-level) minimisation task where plausible parameter-dependent solutions are compared with training examples (typically, image patches) by means of sample averages. The nested problem reads:

$$\hat{\theta} \in \underset{\theta\in\Theta}{\operatorname{argmin}} \ \mathcal{L}(\theta) := \frac{1}{K}\sum_{k=1}^{K} \ell(x_k(\theta), \tilde{x}_k) \tag{2}$$

$$\text{s.t.} \quad x_k(\theta) \in \underset{v\in\mathbb{R}^n}{\operatorname{argmin}} \ \mathcal{J}_\theta(v;y_k), \quad k = 1,\dots,K \tag{3}$$

where $\{\tilde{x}_k\} \subset \mathbb{R}^n$ are training examples paired with their imperfect versions $\{y_k\} \subset \mathbb{R}^m$, $\mathcal{Q} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is the upper-level objective assessing, for each $k = 1,\dots,K$ the quality of the reconstruction process $\mathcal{S} : \theta \mapsto x_k(\theta)$ by comparison with $\tilde{x}_k$. Problem (2)-(3) is thus cast in the form of a *bilevel optimisation* problem for which several theoretical results and algorithmic approaches exist for computing its solution [8, 5].

In the context of machine learning, bilevel learning has been used as a powerful tool in a number of applications. Among them, we may cite hyperparameter optimization [3] – similar to the one considered in this internship – implicit deep learning [2], or neural architecture search [7].

Internship objectives

The internship will focus on the study of convergence properties of optimisation algorithms solving (2)-(3) under the assumption that $\mathcal{J}_\theta(\cdot)$ is C^1 with L-Lipschitz gradient and not necessarily convex.

In this setting, the design of provable converging bilevel optimisation schemes is not trivial: standard approaches based on implicit differentiation require indeed further smoothness (typically, C^2) of the lower-level objective to compute the quantities $\partial\mathcal{L}/\partial\theta_i$ required for designing gradient-based schemes. However, such requirement is often a bit restrictive in imaging contexts. Furthermore, standard numerical solvers solving the lower-level constraints, are typically based on the use of (accelerated) optimisation schemes relying on first-order differential information only, such as, e.g., Nesterov Gradient Descent.

To further reduce the computational burden (typically, quite high) required for solving exactly problem (2)-(3), we will also consider stochastic estimation of the hypergradient, and the use of implicit differentiation, as in [6], and define suitable stopping criteria defining precision values $\{\varepsilon_t\}_t$ so that the lower-level constraint is enforced with a variable (typically, increasing, i.e. $\varepsilon_t \searrow 0$) precision value.

Algorithm 1 Bilevel optimisation algorithm to be studied

Input: $\theta^0, u_k^0, \tau \in (0, \frac{1}{L}], \eta > 0$
while not converging **do**
- solve (3) with precision ε_t:

$$x_k^t(\theta^t) = \texttt{Nesterov\_GD}(u_k^0, \tau, \varepsilon_t; y_k, \nabla \mathcal{J}), \quad \text{for } k = 1, \ldots, K$$

- update θ^{t+1} via:

$$\theta^{t+1} = \theta^t - \eta \widehat{\nabla \mathcal{L}}_\theta(\theta^t)$$

where $\widehat{\nabla \mathcal{L}}(\theta^t)$ is a possibly stochastic approximation of $\nabla \mathcal{L}$ requiring $\{x_k^t(\theta^t)\}$.
end while

Candidate profile

Second year of Master degree in applied mathematics, computer science, engineering or data science with background in optimisation, imaging inverse problems and learning. Good coding skills for numerical simulation (Python and the numerical stack associated, MATLAB, ...).

Practical information

The internship is **funded by the ANR JCJC project TASKABILE** (project description). It will take place within the Inria Morpheme team, a joint research group between Inria centre at Université Côte d'Azur, I3S Lab (Université Côte d'Azur and CNRS) in collaboration with the LJAD mathematics department (Université Côte d'Azur and CNRS). During the internship, the candidate will be provided with the working material (working station, PC and access to local GPU clusters) as well as with a discounted rate for the close-by Inria canteen, together with an internship gratification of approximately 600 euros/month. Depending on the candidate interest and internship outcomes, we will discuss about the opportunities of carrying on the research activities pursued during the internship in a **PhD thesis** funded by the ERC Starting Grant MALIN (PI: L. Calatroni).

Application procedure

Please send your CV, a motivation letter and a copy of your last year transcripts to Luca Calatroni (calatroni@i3s.unice.fr) and Samuel Vaiter (samuel.vaiter@univ-cotedazur.fr).

References

[1] S. Arridge, P. Maass, O. Öktem, and C.- B. Schönlieb. *Solving inverse problems using data-driven models*, Acta Numerica, 28, 2019.
[2] S. Bai, J. Z. Kolter, B. Koltun. *Deep Equilibrium Models*, NeurIPS, 2019.
[3] Y. Bengio. *Gradient-Based Optimization of Hyperparameters*, Neural Computation, 12(8), 2000.
[4] L. Calatroni, A. Lanza, M. Pragliola, F. Sgallari, *A Flexible Space-Variant Anisotropic Regularization for Image Restoration with Automated Parameter Selection*, SIAM Journal on Imaging Sciences, 12(2), 2019.
[5] L. Calatroni, C. Cao, J. C. De los Reyes, C. - B. Schönlieb, and T. Valkonen, *Bilevel approaches for learning of variational imaging models*, in Variational Methods In Imaging and Geometric Control, De Gruyter, 2017.
[6] M. Dagréou, P. Ablin, S. Vaiter, T. Moreau, *A framework for bilevel optimization that enables stochastic and global variance reduction algorithms*, Advances in Neural Information Processing Systems, 35, 2022.
[7] H. Liu, K. Simonyan, Y. Yang. *DARTS: Differentiable Architecture Search*, ICLR, 2018.
[8] K. Kunisch and T. Pock. *A Bilevel Optimization Approach for Parameter Learning in Variational Models*, SIAM Journal on Imaging Sciences, 6(2), 2013.