A principle for minimizers of Tikhonov-regularized problems

Alessandro Scagliotti, Jona Klemenc, Massimo Fornasier

Abstract

Regularized functionals are ubiquitous in various fields of Applied Mathematics, as Inverse Problems and Machine Learning. In this framework, the prototype functional to be minimized has the form:

$$H_\alpha(\theta) := \ell\big((X,Y),\Phi_\theta\big) + \alpha\,\mathrm{Reg}(\theta) \to \min,$$

where $\theta \in \Theta$ denotes the trainable parameters of the model Φ_θ (e.g., Φ_θ can be a deep neural network), $\ell\big((X,Y),\Phi_\theta\big)$ represents the data-fidelity term (we may thing, e.g., (X,Y) to be a labelled training dataset), $\mathrm{Reg}(\theta)$ is the regularization term, and $\alpha > 0$ is the hyper-parameter that tunes the Tikhonov regularization. In situations arising from real-world problems, the functional $\theta \mapsto H_\alpha(\theta)$ is often highly nonconvex, and it is natural to expect that $\arg\min H_\alpha$ contains more than one element. Nevertheless, it turns out that, *for typical values of* α, we have

$$\ell\big((X,Y),\Phi_{\theta_1}\big) = \ell\big((X,Y),\Phi_{\theta_2}\big), \quad \mathrm{Reg}(\theta_1) = \mathrm{Reg}(\theta_2) \qquad \forall \theta_1,\theta_2 \in \arg\min H_\alpha.$$

In other words, besides the automatic identity $H_\alpha(\theta_1) = H_\alpha(\theta_2)$ due to the fact that $\theta_1,\theta_2 \in \arg\min H_\alpha$, we have in addition that, *for typical values of* α, the performance of θ_1 and θ_2 do coincide also on the data-fidelity term and on the regularizer. As we shall see, this curious fact descends from a natural monotonicity property.

More in general, let us consider a set $U \neq \emptyset$ (without further structure) and two functions $F,G\colon U \to \mathbb{R}$ (without further requirements). Let us define the regularized functional $H_\alpha := F + \alpha G$ with $\alpha \in [0,\infty)$, and let us assume that there exist $0 \le a < b$ such that $H_\alpha^\star := \arg\min_U H_\alpha \neq \emptyset$ for every $\alpha \in [a,b]$. If we consider $\alpha_1,\alpha_2 \in [a,b]$ with $\alpha_1 < \alpha_2$, arguing by contradiction we get the inequality:

$$\inf_{H^\star_{\alpha_1}} G \ge \sup_{H^\star_{\alpha_2}} G.$$

This sounds rather expected, since $\alpha_1 < \alpha_2$ suggests that, when setting the regularization coefficient equal to α_2 in the construction of the regularized functional, we are weighting more the regularizer G than when opting for α_1. This fact further implies that $\inf_{H^\star_{\alpha_1}} G \ge \inf_{H^\star_{\alpha_2}} G$, i.e., that the mapping $\alpha \mapsto \inf_{H^\star_\alpha} G$ is monotonically non-increasing. From this monotonicity, it follows that $\alpha \mapsto \inf_{H^\star_\alpha} G$ admits at most countably many discontinuity points in $[a,b]$. Assuming that $\bar\alpha \in [a,b]$ is such that there exist $u_1^\star \neq u_2^\star$ satisfying

$$H_{\bar\alpha}(u_1^\star) = H_{\bar\alpha}(u_2^\star) = \min_U H_{\bar\alpha}, \qquad F(u_1^\star) \neq F(u_2^\star), \qquad G(u_1^\star) \neq G(u_2^\star),$$

then it is possible to show that $\alpha \mapsto \inf_{H^\star_\alpha} G$ is discontinuous at $\bar\alpha$. However, as recalled above, this can happen *at most for countably many exceptional values of* $\bar\alpha \in [a,b]$. We refer to this phenomenon as the "Trade-off Invariance Principle".

A careful analysis allows us to drop the assumption of existence of minimizers for the functional H_α, unlocking the Trade-off Invariance Principle for minimizing sequences. Namely, for every $\alpha \in [0,\infty)$ we introduce the family of minimizing sequences

$$\mathcal{S}_\alpha := \left\{ (u_i)_{i\ge1} : \lim_{i\to\infty} H_\alpha(u_i) = \inf_U H_\alpha \right\},$$

and we define

$$G_\alpha^+ := \sup_{(u_i)_i \in \mathcal{S}_\alpha} \limsup_{i\to\infty} G(u_i), \qquad G_\alpha^- := \inf_{(u_i)_i \in \mathcal{S}_\alpha} \liminf_{i\to\infty} G(u_i).$$

In analogy with what observed before, if we consider any $\alpha_1 < \alpha_2$, we obtain that

$$G_{\alpha_1}^- \geq G_{\alpha_2}^+,$$

which, in combination with $G_{\alpha_2}^+ \geq G_{\alpha_2}^-$, implies that the function $\alpha \mapsto G_\alpha^-$ is monotonically non-increasing. Hence, it is possible to show that, *excluding at most countably many unfortunate values of* α, we have that, if $\inf_U H_\alpha > -\infty$, there exist constants $F_\alpha, G_\alpha \in [-\infty, +\infty]$ such that for every minimizing sequence $(u_i)_{i\geq 1} \in \mathcal{S}_\alpha$:

$$\lim_{i\to\infty} G(u_i) = G_\alpha, \qquad \lim_{i\to\infty} F(u_i) = F_\alpha.$$

The proof leverages on similar arguments as in the case of H_α admitting minimizers. The intriguing aspects of the previous result are twofold. On the one hand, it is not obvious a priori that the functions F, G have limit *along every minimizing sequence*. On the other hand, this limit does not indeed depend on the minimizing sequence itself.

Finally, we want to emphasize on a corollary of the Trade-off Invariance Principle for minimizing sequences, that is relevant when using the Direct Method in Calculus of Variations. Here, we assume the domain U to be a Banach space equipped with a uniformly convex norm $\|\cdot\|_U$, and let us assume that $G(u) = \|u\|_U^p$, for any $p > 0$. As before, we set $H_\alpha = F + \alpha G$. *Excluding at most countably many unfortunate values of* α, if $(u_i)_{i\geq 1} \in \mathcal{S}_\alpha$ is such that $u_i \rightharpoonup_U u^\star$ (i.e., it is weakly convergent) with $u^\star \in \arg\min_U H_\alpha$, then we obtain for free that $\lim_{i\to\infty} \|u_i - u^\star\|_U = 0$, i.e., *the minimizing sequence is strongly convergent to* $u^\star$.

References

[1] M. Fornasier, J. Klemenc, A. Scagliotti. Trade-off Invariance Principle for minimizers of regularized functionals. *arXiv preprint: 2411.11639* (2024).