Asymmetric compressive learning guarantees with applications to quantized sketches

Vincent Schellekens and Laurent Jacques
UCLouvain, Belgium
vschellekens95@gmail.com, laurent.jacques@uclouvain.be

The compressive learning framework reduces the computational cost of training on large-scale datasets. In a sketching phase, a dataset $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^n \subset \mathbb{R}^d$ with n samples is first compressed to a lightweight, m-dimensional sketch vector $\boldsymbol{z}_{\Psi,\mathcal{X}}$, obtained by mapping the data samples through a well-chosen feature map Ψ, and averaging those contributions (see Fig. 1). In a learning phase, the desired model parameters are then extracted from this sketch by solving an optimization problem, which also involves a feature map Φ. When the feature map is identical during the sketching and learning phases ($\Psi = \Phi$), formal statistical guarantees (excess risk bounds) have been proven [1, 2].

However, the desirable properties of the feature map are different during sketching and learning (such as quantized outputs, and differentiability, respectively). We thus study the relaxation where this map is allowed to be different for each phase ($\Psi \neq \Phi$). First, we prove that the existing guarantees carry over to this asymmetric scheme, up to a controlled error term, provided some Limited Projected Distortion (LPD) property holds. We then instantiate this framework to the setting of quantized sketches, by proving that the LPD indeed holds for binary sketch contributions. Finally, we further validate the approach with numerical simulations, including a large-scale application in audio event classification.

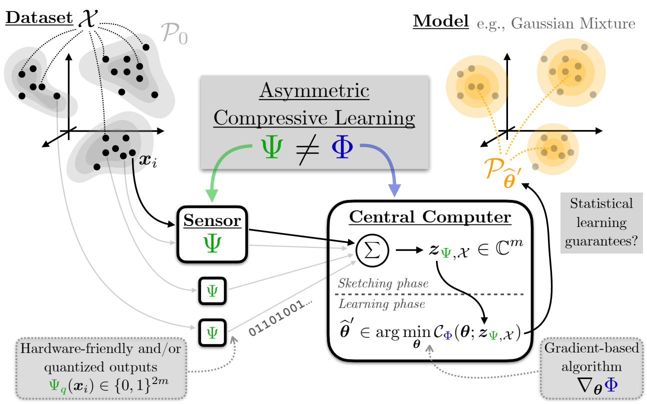

Figure 1: *Our Asymmetric Compressive Learning (ACL) scheme: a dataset $\mathcal{X}$ of n examples $\boldsymbol{x}_i$ (sampled i.i.d. from $\mathcal{P}_0$) is first compressed as a lightweight vector—the* sketch*—by averaging data features $\Psi(\boldsymbol{x}_i)$. This operation can be performed in parallel by a sensor network, which benefits greatly from hardware-friendliness and quantization. A model $\widehat{\boldsymbol{\theta}}'$ is then learned from the sketch $\boldsymbol{z}_{\Psi,\mathcal{X}}$ by solving a CL optimization procedure (minimizing a task-specific cost function $\mathcal{C}_\Phi$ between the sketch and the model) that uses a different, differentiable map $\Phi \neq \Psi$. Our goal is to prove statistical learning guarantees (w.r.t. $\mathcal{P}_0$) for the model $\widehat{\boldsymbol{\theta}}'$.*

References

[1] Rémi Gribonval, Gilles Blanchard, Nicolas Keriven, and Yann Traonmilin. Compressive statistical learning with random feature moments. *Mathematical Statistics and Learning*, 3(2):113–164, 2021.

[2] Rémi Gribonval, Antoine Chatalic, Nicolas Keriven, Vincent Schellekens, Laurent Jacques, and Philip Schniter. Sketching data sets for large-scale learning: Keeping only what you need. *IEEE Signal Processing Magazine*, 38(5):12–36, 2021.