Xuran Meng and Yi Li's contribution to the Discussion of "On optimal linear prediction" by I. Helland

Xuran Meng* and Yi Li†

Congratulations to Professor Helland on the elegant development of optimal partial least squares (PLS) regression, which reconceptualizes model reduction through the geometry of group actions and Hilbert spaces. Within this framework, PLS identifies a lower-dimensional subspace of functions that is provably optimal among all m-dimensional reductions under specified covariance or spectral conditions. This rigorous notion of optimality naturally raises the question: can an analogous concept be established for domain adaptation [Ganin et al., 2016, He et al., 2024, Luo et al., 2025], a technique that has recently attracted significant attention in modern machine learning?

To make the context of domain adaptation concrete, consider a simple example. Suppose we want to classify images of cats and dogs. In the *source domain* S, we observe clean images; in the *target domain* T, the same cats and dogs appear, but all images are rotated by arbitrary angles. A classifier trained on S may fail badly in T, because the marginal distributions of pixels, say, f_S and f_T, differ drastically. Yet the domains share invariant predictive features, such as those extracted by convolutional pooling layers that ignore rotations. If we can find a representation map $h : \mathcal{X} \to \mathcal{Z}$ that captures these invariants, then prediction becomes possible across domains.

This rotation example highlights the central challenge of domain adaptation: although f_S and f_T differ, a representation h may capture outcome-relevant invariants across domains. The key question is whether one can rigorously define an *optimal representation*, balancing predictive accuracy with invariance, analogous to how PLS identifies optimal subspaces in regression. Moreover, what does it mean for a representation h to be "optimal" for domain adaptation. This matters because, without a formal notion of optimality, existing methods such as adversarial training can only propose candidate solutions without offering a principled benchmark. The challenge arises from the fact that the distributions of $\mathbf{x}$ may differ arbitrarily across domains, making direct transfer of models infeasible and requiring careful extraction of predictive features that remain invariant. Helland's symmetry-based framework provides precisely the right tools for this task: group actions, Hilbert spaces, and invariance measures that together define optimality in rigorous geometric terms.

Extending Helland's work, we view domain shifts as group actions. Suppose the target domain arises from the source via a measurable transformation g, so that $f_T = g_\# f_S$, where $g_\#$ denotes the pushforward of measures. For a representation $h : \mathcal{X} \to \mathcal{Z}$, extended to $(\mathbf{x}, y) \mapsto (h(\mathbf{x}), y)$, the induced distribution is $h_\# f_S$. Here,

$$h_\# f_S(A \times B) = f_S\big(\{(\mathbf{x}, y) \in \mathcal{X} \times \mathcal{Y} : (h(\mathbf{x}), y) \in A \times B\}\big), \quad A \subseteq \mathcal{Z}, B \subseteq \mathcal{Y}.$$

If $h_\# g_\# f_S = h_\# f_S$, then h quotients out g and achieves invariance. On such a space, one may define a transitive group G acting on $\mathcal{Z} \times \mathcal{Y}$, with trivial isotropy and a left-invariant measure ν. This generalizes Helland's notion of "related" distributions to the representation-learning setting.

*Biostatistics Department, University of Michigan; e-mail: xuranm@umich.edu
†Biostatistics Department, University of Michigan; e-mail: yili@umich.edu

Given that f_S and f_T are different, we seek a representation h such that the push forward distributions $h_\# f_S$ and $h_\# f_T$ are close or even identical. This motivates introducing both predictor $\mathrm{Pred} : \mathcal{Z} \to \mathcal{Y}$ and invariance constraint. Here, $\mathrm{Pred} \in \mathcal{P}$ where $\mathcal{P}$ is a function class of predictors such as linear functions. We consider $h_\# f_S$ and $h_\# f_T$ embedding into the Hilbert space $L^2(\nu)$ by the Radon–Nikodym derivatives $p_S = \frac{d(h_\# f_S)}{d\nu}$, $p_T = \frac{d(h_\# f_T)}{d\nu}$. The prediction risk is thus

$$R(\mathrm{Pred}, h) = \mathbb{E}_{(\mathbf{z},y)\sim h_\# f_S}\left[\ell(\mathrm{Pred}(\mathbf{z}), y)\right] = \int \ell(\mathrm{Pred}(\mathbf{z}), y)\, p_S(\mathbf{z}, y) d\nu(\mathbf{z}, y),$$

and the distributional invariance is given by

$$\|p_T - p_S\|^2_{L^2(\nu)} = \int \left(p_T(\mathbf{z}, y) - p_S(\mathbf{z}, y)\right)^2 d\nu(\mathbf{z}, y).$$

We then define optimality of h as

$$\min_{\mathrm{Pred}\in\mathcal{P},\, h\in\mathcal{H}} R(\mathrm{Pred}, h) \quad \text{s.t.} \quad \|h_\# f_T - h_\# f_S\|^2_{L^2(\nu)} \le \gamma, \tag{1}$$

where γ quantifies the tolerance for domain discrepancy. This mirrors Helland's framework: both risk and invariance are embedded in the same Hilbert space $L^2(\nu)$, yielding a precise optimization-based notion of optimality. Here, the notion of *optimality* primarily concerns the representation h, with "Pred" adapting accordingly within its class $\mathcal{P}$.

In practice, adversarial training [Ganin et al., 2016] enforces $h_\# f_S \approx h_\# f_T$ by jointly training a domain discriminator to distinguish between $h_\# f_S$ and $h_\# f_T$ and a representation map h to both confuse the discriminator and minimize prediction loss on the source domain; however, optimality was not addressed. Our formulation (1), grounded in Helland's geometry, provides the missing rigor by defining "optimal" invariant representations as solutions to constrained optimization problems, in the same way that PLS characterizes optimal subspaces.

In sum, the challenge in domain adaptation is that source and target domains differ in their raw covariates $\mathbf{x}$, making direct transfer infeasible and necessitating extraction of invariant predictive features $\mathbf{z}$. By embedding this problem into Helland's symmetry-based framework, we obtain a principled notion of optimality, that is, minimizing prediction risk subject to distributional invariance. This new formulation replaces heuristic criteria with a precise benchmark for optimality, clarifying why some representations are preferable and enabling fair comparison across methods. This also demonstrates the novel applicability of Helland's framework to modern machine learning.

References

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Yifei He, Haoxiang Wang, Bo Li, and Han Zhao. Gradual domain adaptation: Theory and algorithms. *Journal of Machine Learning Research*, 25(361):1–40, 2024.

Weibin Luo, Mingye Chen, Jian Gao, Yanping Zhu, and Fang Wang. Multi-view affinity-based projection alignment for unsupervised domain adaptation via locality preserving optimization. *Scientific Reports*, 15(1):20452, 2025. doi: 10.1038/s41598-025-05331-3.