A Compositional Framework for First-Order Optimization

Tyler Hanks   Matthew Klawonn   Matthew Hale   Evan Patterson
James Fairbanks

Decomposition methods such as primal and dual decomposition are fundamental tools to develop distributed solution algorithms for large scale optimization problems arising in machine learning [1], optimal control [2], and operations research [3]. These methods generally work by splitting a large problem into several simpler subproblems and repeatedly solving these to arive at a solution to the original problem. As such, *decomposition* methods are most naturally applicable to problems which themselves are *composites* of subproblems, for some appropriate notion of composition. We say that such problems have *compositional structure*.

This talk, based on our arXiv preprint [4], introduces a category theoretic framework which unifies various first-order optimization decomposition methods under the single abstraction of a *morphism of operad algebras*. Specifically, we use algebras on the operad of undirected wiring diagrams (UWDs) [5, 6] to model the compositional structure of various classes of optimization problems. The central idea of our framework can then be summarized as follows. If a first-order optimization algorithm defines an algebra morphism from a UWD-algebra of optimization problems to a UWD-algebra of dynamical systems, that algorithm decomposes problems defined on arbitrary UWDs. Furthermore, applying such a morphism to a problem generates a dynamical system to solve the problem in a distributed fashion using message passing semantics. Our main contribution is to construct several instances of this general pattern, shown in Figure, and demonstrate how these can be used to recover primal and dual decomposition algorithms for problems with arbitrary UWD structure.

This novel perspective on decomposition algorithms allows us to derive a new sufficient

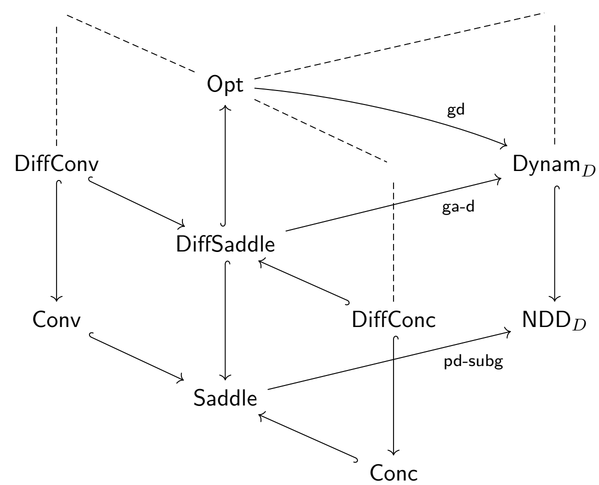
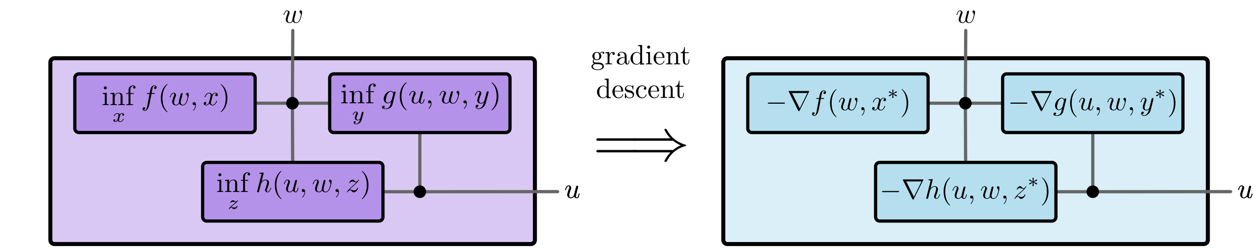


Figure 1: The hierarchy of results presented in our paper. Nodes represent the various UWD-algebras developed including those for composing saddle problems, convex problems, concave problems, all with and without differentiability assumptions, as well as composing deterministic and non-deterministic dynamical systems. Hooked arrows indicate that there is an inclusion of one algebra into another. Non-hooked arrows are the algebra morphisms including gradient descent (gd), gradient ascent-descent (ga-d) and the primal-dual subgradient method (pd-subg). Composing the inclusions with the gradient algebra morphisms yields (sub)gradient descent for convex problems and (super)gradient ascent for concave problems.



$$\underset{u,w}{\text{minimize}}\left(\inf_{x,y,z} f(w,x) + g(u,w,y) + h(u,w,z)\right)$$

$$x^* = \arg\min_x f(w,x)$$
$$y^* = \arg\min_y g(u,w,y)$$
$$z^* = \arg\min_z h(u,w,z)$$

Figure 2: A graphical depiction of primal decomposition for a given problem in our framework. The left diagram is an Opt-UWD while the right diagram is a Dynam-UWD. Gradient descent gives a structure preserving map.

condition for when a problem is decomposable: namely, if the translation of problem data into the associated optimization problem furnishes an algebra morphism from a UWD-algebra of data to one of our algebras of optimization problems. We refer to this as the *compositional data condition*. To demonstrate the use of this sufficient condition, we show that the minimum cost network flow (MCNF) problem defines an algebra morphism from a UWD-algebra of flow networks to the UWD-algebra of unconstrained concave optimization problems, which when composed with the gradient ascent morphsim recovers a generalization of the standard dual decomposition algorithm for solving MCNF which respects arbitrary hierarchical decompositions of flow networks. We briefly discuss our implementation of this framework in the Julia programming language and provide experiments showing that exploiting hierarchical compositional structure yields faster solution algorithms.

References

[1] Stephen Boyd. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2010.

[2] Pontus Giselsson, Minh Dang Doan, Tamás Keviczky, Bart De Schutter, and Anders Rantzer. Accelerated gradient methods and dual decomposition in distributed model predictive control. *Automatica*, 49(3):829–833, March 2013.

[3] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numer. Math.*, 4(1):238–252, dec 1962.

[4] Tyler Hanks, Matthew Klawonn, Evan Patterson, Matthew Hale, and James Fairbanks. A compositional framework for first-order optimization. *arXiv preprint arXiv:2403.05711*, 2024.

[5] David I. Spivak. The operad of wiring diagrams: formalizing a graphical language for databases, recursion, and plug-and-play circuits, May 2013. arXiv:1305.0297 [cs, math].

[6] Brendan Fong and David I. Spivak. Hypergraph Categories, January 2019. arXiv:1806.08304 [cs, math].