

Supervisor's Foreword

The problem of *sparse* optimization has gathered a lot of attention lately. The reason is simple: sparsity is a fundamental structural characteristic of much of the data we encounter. Indeed, one may claim that the structure in these data is an *expression* of sparsity. The sparsity may manifest in different ways. Often the data themselves are sparse, in that the majority of their components are zero-valued. More commonly, the data may simply be restricted in the values they can take or the patterns they may follow. Here, the structure in the data can often be characterized as sparsity in a transformed domain. For instance, the data may be restricted to only inhabit a restricted set of subspaces. In this case descriptions of the data in terms of their projections on these subspaces will be sparse. This sparsity can be exploited for a variety of purposes, e.g., compressed sensing techniques exploit sparsity in signals to characterize them using far fewer measurements than would otherwise be required, RADAR and SONAR applications exploit the spatial sparsity of sources for better detection and localization of sources, etc.

At other times, sparsity may be imputed to characterizations of various aspects of the data, in an attempt to bring out the structure in it. Thus, statistical analyses and various machine learning techniques often attempt to fit sparse models to data, enable better predictions, identify important variables, etc. At yet other times, sparsity may be enforced simply to compensate for paucity of data to learn richer or more detailed models.

In all cases, one ends up having to estimate the sparsest solution that minimizes a loss function of some kind, i.e., with an instance of the aforementioned sparse-optimization problem. The specifics vary chiefly in the loss function minimized. For instance, compressed sensing attempts to minimize the squared error between observations of the data and observations that might be engendered by the sparse solution, machine learning techniques attempt to minimize the negative log probability of the observed data, as predicted by the model, and so on.

Obtaining sparse solutions, however, is not trivial. Sparsity is defined through the ℓ_0 norm—the number of nonzero components—of the variable being optimized. To obtain a sparse solution, this norm must hence be directly minimized or, alternately, imposed as a constraint on the optimization problem. Unfortunately,

optimization problems involving the ℓ_0 norm require determination of the optimal set of components to be assigned nonzero values and are hence combinatorial in nature and are generally computationally intractable. As a result, one must either employ greedy algorithms to obtain a solution or employ proxies that are relaxations of the ℓ_0 norm. Both of these approaches have yielded highly effective algorithms for optimization, when the loss function is quadratic or, more generally, convex in nature.

For more generic classes of loss functions, however, the situation is not so clear. Proxies to the ℓ_0 norm which can be shown to result in optimally sparse solutions for quadratic or convex loss functions are no longer guaranteed to provide optimal solutions for other loss functions. It is similarly unclear whether greedy algorithms that are effective for well-behaved loss functions will be equally effective in the most general case.

This is the problem space that Sohail tackles in this monograph. In an outstanding series of results, he develops and analyzes a greedy framework for sparsity-constrained optimization of a wide class of loss functions, shows how it may be applied to various problems, and finally extends it to handle the case where the solutions are not merely sparse, but restricted to lie in specified subspaces.

GraSP is the proposed greedy framework for sparse optimization of loss functions. Through rigorous analysis, Sohail demonstrates that it imposes far fewer constraints on the loss function, only requiring it to be convex on sparse subspaces, and converges linearly to the optimal solution. As an illustrative application he applies GraSP to the problem of *feature selection* through sparse optimization of logistic functions, and demonstrates that it results in significantly better solutions than current methods. *One-bit* compressive sensing is the problem of reconstructing a signal from a series of one-bit measurements, a challenging but exciting problem. Sohail demonstrates that GraSP-based solutions can result in greatly improved signal recovery over all other current methods.

Subsequently, he develops a solution to deal with *model-based* sparsity: problems where the solutions are not only required to be sparse, but are further restricted to lie on only specific subspaces. Such problems frequently arise, for instance, when additional information is available about the interdependence between the location of nonzero values in the estimated variables.

Finally he reverses gear and addresses a more philosophical problem—that of identifying the best proxy for gradient-based algorithms for sparsity-constrained least-squares optimization—and arrives at the remarkable result that the optimal proxy is the ℓ_0 norm itself.

Together, the contributions of this monograph lay a solid foundation of techniques and results for any aspiring or established researcher wishing to work on the problem of sparse optimization of difficult-to-optimize loss functions. As such, I believe that this monograph is a mandatory inclusion in the library of anybody working on the topic.

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, USA

Prof. Bhiksha Raj

Algorithms for Sparsity-Constrained Optimization

Bahmani, S.

2014, XXI, 107 p. 13 illus., 12 illus. in color., Hardcover

ISBN: 978-3-319-01880-5