Research Statement for Jeffrey Larson

I investigate algorithms for mathematical optimization motivated by real-world situations. I seek to prove convergence results and other properties for these algorithms while solving problems in areas including particle accelerator design, quantum computing, vehicle routing, and sports league scheduling. I am continuously looking for domain experts with compelling problems with solutions requiring optimization techniques. Working with these applications inspires future mathematical research, and their solutions contribute valuably to other scientific fields.

Applications of My Research

Particle Accelerators: The operation and design of particle accelerators are the driving forces behind my current research at Argonne National Laboratory. I work with a team of computational physicists who simulate the interactions between charged particles as they are propelled to high speeds by electromagnetic fields inside accelerators. These complex simulations allow scientists to predict properties of an accelerator (for example, the top speed of particles within an accelerator) at a fraction of the cost of actually constructing the device. I am designing optimization algorithms that adjust design or operational parameters in order to find multiple, high-quality, local optima and also run efficiently on world-class high-performance computers. Efficiently finding multiple optima is considerably different from finding a single solution, and therefore of great mathematical interest.

Vehicle Platooning: I also study the application and complexity of network algorithms that optimize routes for platooning vehicles in large-scale networks. Vehicles driving in a line with small intervehicle distances greatly reduce their fuel consumption because of decreased aerodynamic drag. Since a large percentage of the cost of transporting goods comes from fuel, even small savings in fuel use can have a significant financial and environmental impact. This problem is theoretically interesting since most routing algorithms look to decrease congestion, not coordinate the routes of fleets of vehicles as they travel to their destinations. I am the principal investigator on a three-year, Laboratory Directed Research and Development grant for studying platooning in metropolitan road networks.

Quantum Computing: Quantum computation and information present methods (currently in their infancy) for solving difficult problems in computer science. These methods use and work on data that is encoded as the superposition of particles' quantum states, and therefore require entangled quantum-mechanical systems. I am working with theoretical physicists to maximize the entanglement of quantum dots within simulations as a proof of concept for quantum computing. As we find particle arrangements with high entanglement, we turn to experimentalists to validate our results.

Sports Scheduling: As a postdoctoral researcher at KTH, I was presented with an interesting scheduling problem. The top Swedish handball league, Elitserien, wanted to extend their double round-robin schedule by splitting the league into two divisions that would hold an additional single round-robin tournament. I was able to exploit the problem's structure (previously unaddressed in the literature) to efficiently construct the 2013-2014 schedule [4, 5]. We are currently expanding our techniques in order to schedule multiple leagues simultaneously. This work is relevant to the Swedish Handball Federation because many men's and women's teams in Swedish cities share an arena, so a schedule must account for both leagues at the same time.

Active Areas of Optimization Research

A large part of my research focuses on optimizing functions or simulations that are expensive to evaluate, are possibly noisy, and lack accurate derivatives. In these cases, evaluating the objective function at a single point may take seconds to hours, significantly slowing optimization routines that evaluate the function frequently. Equally confounding for many traditional algorithms is the noise often present in such functions; a small perturbation in the input parameters can lead to large changes in the output merely due to the numerical routines required to evaluate the function. Automatic differentiation or finite difference approximations to the gradient are not suitable in such cases. The need for DFO has increased alongside the growth in computational simulation in most scientific fields. My research is specifically focused on model-based trust-region methods, a class of algorithms that builds successive, easily optimized models to approximate the original function.

Finding Multiple Local Minima: For many real-world problems, application scientists desire multiple, high-quality options that can be compared. Common approaches for finding such solutions often involve starting local optimization routines from random points in the domain. Such multistart schemes are successful in locating many local minima, but they are often unnecessarily expensive since many of the minima will be discovered multiple times. In [9], we develop a generalization of the multilevel single linkage algorithm for the derivative-free case with strong theoretical results. We prove that under mild assumptions on the objective function, our multistart algorithm will almost surely locate all local minima while starting only a finite number of local optimization runs. Importantly, an implementation of the method finds many, high-quality minima on synthetic and real-world problems in fewer evaluations than other algorithms use.

We are developing numerous important extensions to our parallel algorithm for finding good local minima. Many intense simulations running on high-performance computers can be evaluated faster with more computational resources. We are extending our algorithm from [9] to adjust the resources given to different simulation evaluations. We aim for an algorithm that can discover how the objective function scales in the initial evaluations and then use that information to allocate resources and asynchronously evaluate points of interest. We are also developing the software so there is efficient communication between concurrent local optimization runs; this reduces the amount of possible wasted computational effort (for example, by preventing two optimization runs from evaluating the same point). We are also developing domain-specific local optimization routines that can be used within our multistart framework, for example the nonsmooth local optimization routine in [8]. While developing these extensions, we will take care to ensure that the original strong theoretical results are maintained.

Stochastic Derivative-Free Optimization: In [3] we prove almost sure convergence of a class of trust region algorithms for optimizing stochastic functions that lack computable derivatives. Convergence results for most existing methods for optimizing stochastic functions involve repeatedly sampling the of function, require the user to define a decaying sequence of step sizes, or evaluate a fixed pattern of points at every iterate. All three of these approaches for removing the stochasticity are undesirable when the function is expensive to evaluate. Repeated sampling requires significant computational cost often at points far from the minimum, and algorithmic performance is sensitive to the sequence of step sizes or pattern of points used by the algorithm. Determining the best sequence or pattern requires knowledge of the problem, something that is often unavailable in practice. Under reasonable assumptions, we prove that the class of algorithms in [3], which builds models of the function using previously evaluated points, converges almost surely to a stationary point of a stochastically noisy function. A prototype of our algorithm is shown to outperform existing algorithms for optimizing stochastic functions.

When building models of the noisy function, current theory requires all points to be generated by an independent process. When minimizing the objective function, the iterates of most algorithms will be naturally biased towards points with smaller function values. In order to prove almost sure convergence, previous function values must be removed from the model-building process. We are developing methods that can retain these points and their biased function values when building new models, but still ensure algorithmic convergence, the number of function evaluations required to converge to a solution will be greatly reduced.

Weighted Regression Models: Many existing model-based DFO algorithms use various interpolation or regression functions, but such schemes consider each point equally important. When building a regression model, it is intuitive to more heavily weight those points with better function values, or those closer to the current best point. In [1], we develop an algorithm using weighted regression quadratic models; this algorithm requires fewer function evaluations than methods using interpolation or regression models require to converge to the local optima on a suite of benchmark problems. We also proved that our algorithm is guaranteed to converge to a second-order stationary point of sufficiently smooth functions, provided the weighting scheme satisfies a simple requirement.

The convergence theory developed in [1] provides a general framework for designing algorithms based on weighted regression. I want to investigate several specific strategies for selecting weights to achieve optimal performance. One strategy would be to choose weights by minimizing a bound on the accuracy of the model function within the trust region derived in [1]. A related method involves choosing the weights to minimize an error bound near a predicted minimizer of the model function. If either of these minimization subproblems is easier to solve than the cost of evaluating the function, it is worthwhile to expend the computational effort to find the best weights at each iterate, especially if the algorithm can progress to an optimum with fewer function evaluations.

I am also interested in applying our weighted regression algorithm to problems where a trade-off exists between the accuracy and the speed of function evaluations. For example, if we wish to minimize a function whose evaluation involves solving partial differential equations, one can increase the mesh size to more quickly calculate a function value, although it may differ considerably from the true function value. Significant improvement is realized if the algorithm and function work in tandem, calling inaccurate, but fast, function evaluations as the algorithm begins, eventually requiring more accurate evaluations closer to a suspected optimum. Using weighted regression trust region models to give less importance to inaccurate function evaluations should produce models that more accurately approximate the underlying function.

Terminating Expensive Optimization: When optimizing expensive, noisy functions, most algorithms proceed until a budget of evaluations is exhausted or as long as improvement is observed in the objective function. However, the improvement observed may be due not to finding a better point but only because of random variation in the noisy function evaluations. If this is the case, a more prudent use of the computational budget would be to stop the algorithm when improvement is close to the magnitude of the noise, and to restart at a different point in the search space. In [10] we propose a variety of solver-independent termination tests, requiring minimal information from the algorithm, that indicate when an algorithm has stagnated or is improving because of variation in the noise.

A useful extension of this work would be to determine when to stop individual, expensive function evaluations running on high-performance computers. Often such simulations take orders of magnitude longer to evaluate when a poor function value will be returned. If we can probabilistically model the likelihood that a long-running evaluation will be sufficiently bad (for example, by observing previous values and their computation time), we could determine when to best halt a function evaluation, and use those resources to evaluate the simulation at a different point.

Vehicle Platooning: In the work first proposed [6] and later extended in [7], we show that coordinating routes of vehicles to increase platooning opportunities can yield significant savings in fuel, with the added benefit of reducing carbon emissions. Our simulations of the German Autobahn network show that a few thousand vehicles coordinating their routes can reduce their fuel consumption by over 5%. We extend the theory of platooning vehicles in [11] by formally defining the vehicle platooning problem. We show the general platooning problem to be NP-hard, but we provide formulations that can be solved in real time for hundreds of vehicles. Previous formulations could be solved exactly only for at most four vehicles.

This work has been an essential contribution to many of the work packages in the COMPANION Project, a European Commission grant of which I wrote significant portions. The 3 Million Euro project is a collaborative effort between KTH and many industrial partners, including Scania and Volkswagen, to develop cooperative mobility technologies for supervised vehicle platooning. The work in [7] is being extended to develop a real-time coordination system that will compute routes for vehicles in order to dynamically create, maintain, and dissolve platoons of vehicles to maximize fuel savings.

Since the large-scale coordinated platooning of vehicles has only recently become viable, many mathematically interesting (and practical) research questions must be addressed. I am the principal investigator on an Laboratory Directed Research and Development grant that is studying many of these issues. We are working to find strategies for optimally routing platooning and non-platooning vehicles within a large metropolitan road network.

Funding Opportunities

I welcome opportunities to use my expertise in optimization to help domain scientists solve novel application problems. My work is of interest to funding agencies as evidenced by the funding I have obtained for vehicle platooning research. Similarly, the recent recommendations of the Department of Energy Advance Scientific Computing Research Working Group [2] highlight the priority funding agencies will place on research intersecting high-performance computing and numerical optimization. The solution to these important computational problems motivates new mathematical research that is relevant to the scientific community.

References

[1] Stephen C Billups, Jeffrey Larson, and Peter Graf. Derivative-Free Optimization of Expensive Functions with Computational Error Using Weighted Regression. *SIAM Journal on Optimization*, 23(1):27–53, January 2013.

[2] Jack Dongarra, Jeffrey Hittinger, John Bell, Luis Chacón, Robert Falgout, Michael Heroux, Paul Hovland, Esmond Ng, Clayton Webster, and Stefan Wild. Applied Mathematics Research for Exascale Computing. http://www.osti.gov/scitech/biblio/1149042, 2014.

[3] Jeffrey Larson and Stephen C Billups. Stochastic Derivative-free Optimization using a Trust Region Framework. Under review at *Computational Optimization and Applications*, 2014.

[4] Jeffrey Larson and Mikael Johansson. Constructing Schedules for Sports Leagues with Divisional and Round-robin Tournaments. *Journal of Quantitative Analysis in Sports*, 10(2):119–129, January 2014.

[5] Jeffrey Larson, Mikael Johansson, and Mats Carlsson. An Integrated Constraint Programming Approach to Scheduling Sports Leagues with Divisional and Round-Robin Tournaments. In Helmut Simonis, editor, *Integration of AI and OR Techniques in Constraint Programming*, volume 8451 of *Lecture Notes in Computer Science*, pages 144–158. Springer International Publishing, 2014.

[6] Jeffrey Larson, Christoph Kammer, Kuo-yun Liang, and Karl Henrik Johansson. Coordinated Route Optimization for Heavy Duty Vehicle Platoons. In *16th International IEEE Conference on Intelligent Transportation Systems*, pages 1196–1202. IEEE, October 2013.

[7] Jeffrey Larson, Kuo-yun Liang, and Karl H Johansson. A Distributed Framework for Coordinated Heavy-Duty Vehicle Platooning. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–11, 2014.

[8] Jeffrey Larson, Matt Menickelly, and Stefan Wild. Manifold Sampling for L1 Nonconvex Optimization. Under review at *SIAM Journal on Optimization*, 2015.

[9] Jeffrey Larson and Stefan Wild. A Batch, Derivative-free Algorithm for Finding Multiple Local Minima. *Optimization and Engineering* (to appear), 2015.

[10] Jeffrey Larson and Stefan M. Wild. Non-intrusive Termination of Noisy Optimization. *Optimization Methods and Software*, 28(5):993–1011, October 2013.

[11] Erik Larsson, Gustav Sennton, and Jeffrey Larson. The Vehicle Platooning Problem: Computational Complexity and Heuristics. *Transportation Research Part C*, 60:258–277, November 2015.