特集論文

スマート社会を支える先進最適化技術

ジョセフ・カッツ* マシュー・ブランド*
ダニエル・ニコフスキー* 猪又憲治*
キーラン・パーソンズ*

Advanced Optimization Technology for Smart Society

Joseph Katz, Daniel Nikovski, Kieran Parsons, Matthew Brand, Kenji Inomata

要　旨

アインシュタインは“世界は私たちの考えに基づいて作り上げられたものでできており，私たちの考え方を変えなければ世界は変わらない”という有名な言葉を残している。将来，スマート社会を実現するためには，今日の私たちが直面する問題を解決・改善する新しい考え方である先進技術の適用が不可欠である。最適化アルゴリズムはそのような重要な技術の１つである。それは単に使い勝手の良いインフラを実現するものではなく，私たち人間がより暮らしやすく，より良い人間社会を実現する力がある。私たちはその信念の下，最適化技術の開発を追求している。

１つ目は，電力系統の最適化である。グローバル最適化をどのように成し遂げるのかを数学的アプローチで解き明かす。２つ目は，環境の変化に強い最適経路選択である。３つ目は，通信の強化である。高次元変調という概念を導入して更なる通信品質の改善を図る。最後は，非常に大規模なシステムにおける最適化であり，計算量をいかに減らし，実用上有効な解をいかに導くかを示す。

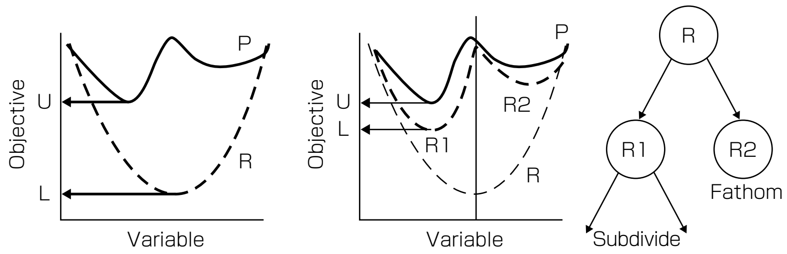
我々が考える技術と社会のシンボル

数学を駆使したアプローチによってスマート社会を底辺で支え，私たち人間がより暮らしやすく，より良い成果を残せるものへと変える。最適化は私たちの暮らしを本質的に変える力がある。

*Mitsubishi Electric Research Laboratories (Ph.D.)

1. Introduction

Future pervasive connectivity among people and machines — referred to by acronyms such as Internet of Things (IoT), Machine to Machine (M2M), and Vehicle to X (V2X) — is inevitable. Such connectivity will be the bedrock of the smart society, but by itself will not suffice to achieve its many promises, including efficient resource utilization and increased safety and security.



(a) Lower and Upper Bounding (b) Domain Subdivision (c) Search tree

Fig. 1 Power flow optimization method

Albert Einstein famously said that the world we have created is a product of our thinking, and it cannot be changed without changing our thinking. The future smart society will need to apply advanced technologies in order to solve or significantly improve the issues facing our world today. Optimization algorithms are one important kind of such technologies that will help translate the connectivity infrastructure into improved outcomes for humans. In this article we describe examples of such algorithms for finding global optimum for power flow (Section 2), achieving robust optimal routing in changing environments (Section 3), enhancing communication throughput of optical fiber links through the use of high-dimensional modulation (Section 4), and achieving practical solutions for very large systems that would otherwise require huge computational resources (Section 5).

2. Globally Optimal Power Flow

We seek to find the optimal operating point for a large electrical network, such that demand at every node of the network is satisfied, and the cost of generated electricity is minimized, subject to voltage and capacity constraints of the power line generators. It is one of the main decision problems in everyday operation of most transmission and distribution systems, and its optimal solution can have a huge economic and societal benefit, especially in situations where demand may outstrip supply.

Whereas direct current optimal power flow (OPF) is a convex optimization problem, in many cases the alternating current OPF problem is non-convex. Most of the widely used non-linear optimization algorithms are not guaranteed to find the true global optimum, but are rather likely to converge to one of the many local minima. Recent experimental studies demonstrated that local solutions could be more expensive by over 120%, resulting in major waste for electrical utilities. This led to active research in advanced methods for finding the true global optimum. Many of these methods are based on the branch-and-bound algorithm that operates on the principle of subdividing the range of decision variables, such as the voltages of generators, into sub-intervals, and eliminating some of these sub-intervals if it can be shown that the best possible solution in them is no better than a known solution in another intervals (Fig. 1). One effective way to show this is if a lower bound on the cost over such an interval can be computed; the tighter the bound, the more efficiently the algorithm can eliminate large regions without sub-dividing them further.

Although existing branch-and-bound algorithms have been effective in finding global OPF minima in small problems, they have not been able to scale up to the large problems many actual customers have, which include hundreds of network nodes and decision variables. The main reason for this is that the lower bounds computed by these algorithms have been quite loose, and the algorithm ends up exploring the entire decision space. Recently, we have proposed two novel methods for computing very tight lower bounds, one based on Lagrangian Relaxation (LR), and another one based on Semi-Definite Programming (SDP). We have demonstrated[1] that networks with hundreds of nodes, such as the IEEE118 benchmark with 118 nodes can be solved in less than 4 seconds on a modern PC, and are working towards solving networks with thousands of nodes.

3. Robust Dynamic Routing

For several years, researchers at Mitsubishi Electric Research Laboratories (MERL) have studied routing problems in shared networks, for example, routing cars through roadway traffic, routing data packets in ad-hoc wireless networks, and routing passengers in people-moving systems such as elevators and public transportation. The goal is to find a minimum-cost route for an individual and/or a maximum flow for a population.

Classical algorithms can be used when one assumes that each link of a network has a fixed cost or delay. In the real world, however, link costs are unknown or uncertain. Traffic makes it impossible to precisely predict travel times, so the cost for any sequence of routing decisions must be described with probability distributions. We focus on this case.

Generally, it is necessary to balance two criteria to select a route: expected performance and reliability. Expected performance is calculated by integrating a utility function over all possible costs according to their probabilities. For example, the utility of a data packet may decrease exponentially as delivery time increases. Reliability is usually defined as a limit on the probability of catastrophically expensive outcomes, for example, arriving at an airport after one's plane has departed. It is possible to have very good expected performance and

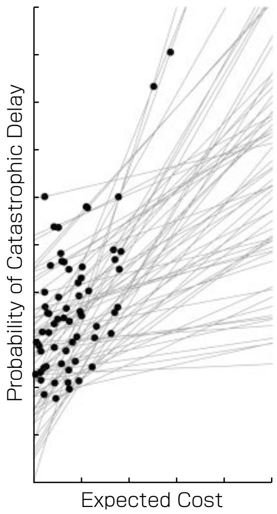


Fig. 2　Expected cost of a route vs. probability of catastrophic delay（see Ref. 2）

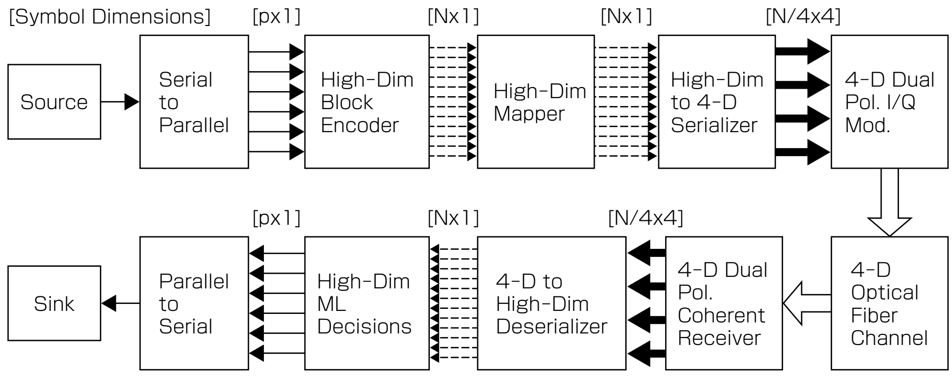


Fig. 3　Block diagram of HDM transceiver

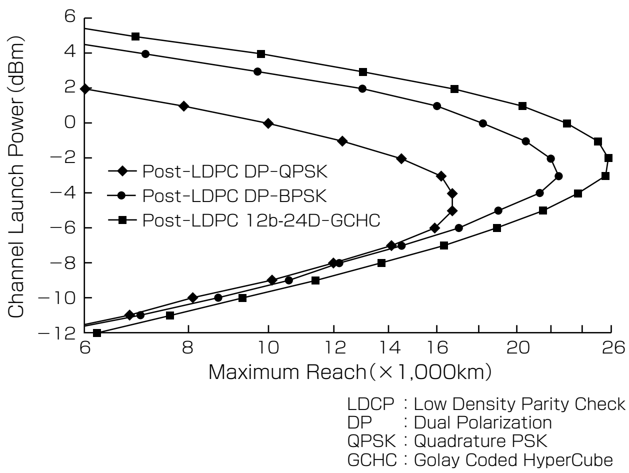


Fig. 4　Lab experiment of transmission distances for various modulation schemes

extremely poor reliability, and vice versa.

We have developed methods to optimize logistically useful combinations and constraints on these criteria. In many cases, it can be proven that the amount of computation needed to find the exact optimal path, or even calculate the exact value of single path, can grow exponentially with the size of the network. For those cases, we have developed fast and efficient approximation algorithms that optimize lower bounds on performance and reliability, so that, for example, it is possible to quickly identify a commuting route to work that offers a 99% probability of avoiding traffic delays, and is within a few seconds of being the fastest-on-average route with such a guarantee(see Fig. 2). These methods also provide routing policies ― real-time decision-making algorithms ― that optimally respond to current and predicted changes in traffic. This has also motivated work on predicting reliability from historical data plus occasional new measurements.

4. High-Dimensional Modulation for Optical Fiber Communication

Supporting the smart society requires a huge transfer of information and content. Cisco predicts that annual global IP traffic will reach 10^{21} bytes in 2016, and will continue to grow at more than 20% annually. Much of this traffic travels on long-distance terrestrial and trans-oceanic fiber optic links. Since optical fibers are very expensive to install on land or sea, maximizing the capacity of the fiber is a key requirement to support the growth in network traffic.

For many years, fiber capacity was improved by increasing the signal rate of each channel, but going beyond 10Gbps per channel proved too challenging due to optical impairments of chromatic and polarization mode dispersion. This limit was overcome using coherent optical systems, which allowed multi-level modulation formats to be used over all four physical dimensions, In-phase and Quadrature (IQ) over X and Y polarizations), with sophisticated digital signal processing(DSP) algorithms to reduce impairments. This has allowed 100Gbps commercial systems to be deployed, with 200 and 400Gbps in trials. However, these new higher rates come with a penalty of lower reach between expensive regenerators due to noise and fiber nonlinearity.

Recently, optimized four dimensional modulation formats have been developed, which provide some gain[(6)]. MERL is at the forefront of research into the use of high dimensional modulation(HDM) formats(>4D) and codes which offer even higher levels of performance. By increasing the dimensionality of the signal space it is possible to more efficiently pack the signal constellation and so improve the tolerance to noise and nonlinearity.

Fig. 3 shows a simplified transceiver block diagram using HDM. In many cases, only the encoder/mapper and soft detector need to be modified, allowing the majority of the coherent transceiver design to be reused. If needed, the error correction code can be optimized for the exact HDM format characteristics to further improve performance.

Fig. 4 shows the result of laboratory experiments using a recirculating optical fiber loop using Corning SMF-28 Ultra Low Loss fiber[(7)]. It can be seen that the HDM method(using 24 dimensions to encode 12 bits) provides a 15% increase compared with the state-of-the-art dual-polarization binary phase shift keying (DP-BPSK) method while providing the same capacity, allowing a transmission distance of more than 25,000km (not including additional system margin).

High-dimensional modulation formats will enable significant increases in transmission distance of fiber-optic links, or allow higher capacity for the same distance, thus supporting the rapid growth in global traffic demands.

5. Making Algorithms Run Faster

As the world continues to automate, an increasingly large part of the value proposition of machines and services comes from software, particularly optimization methods from data mining, machine learning, and control theory. Unfortunately, these optimization methods scale very poorly: a problem that takes N bits to describe may require solution times on the order of N^3, N^6, or even 2^N. Problems sizes are growing rapidly but CPUs stopped getting faster roughly 10 years ago, which, for example, is why most of the functionality and value of a smartphone actually resides in massive compute centers thousands of kilometers away. Of course, for many products the computation cannot be performed remotely by a supercomputing cluster, so it is necessary to develop approximation algorithms that produce near-optimal solutions in linear or linear-log time.

This has been a major focus at MERL since its founding. We give two examples here. One of MERL's most widely used technologies is an accelerated approximation algorithm for computing the thin singular value decomposition (SVD), which factors a data matrix into smaller matrices with many useful properties. Typically SVD reveals that most of the variation in the data can be explained with a small number of degrees of freedom, called its rank. Therefore it is widely used in data compression, data mining, and system identification. In current applications the data matrix may have a million rows and billions of columns. Since SVD is fundamentally an N^3 optimization, some approximation is needed to enable real-time computation. We developed a linear-time approximation algorithm that can deliver an exact factorization if the data is indeed low-rank. The algorithm is widely used in the research community, in some Mitsubishi Electric Corp. (MELCO) products, and, via licensing, in industries where MELCO does not participate in the market.

Data is usually corrupted with Gaussian noise, consequently many optimization problems involve fitting a parameterized model by minimizing a "soft" squared-error term subject to some "hard" constraints. If these constraints can be written as linear equations, the optimization problem is called a quadratic program (QP). QPs are the core of industrially significant technologies including model-predictive control, support-vector prediction machines, and financial portfolio optimization. We have developed a very fast parallel approximation algorithm that solves a significant class of QPs in linear time. The algorithm is unusually fast and simple, enabling the use of advanced control and optimization technologies in new settings where compute power is limited, notable, optimal control of fast-moving machinery.

6. Conclusion

This paper presented several examples highlighting the potential of smart algorithms to help realize the promise of the future smart society.

References

(1) Gopalakrishnan, A., et al. : Global optimization of optimal power flow using a branch & bound algorithm, Proc. Allerton Conf. Comm., Control & Computing (2012)

(2) Brand M., et al. : Routing with probabilistic delay guarantees in wireless ad-hoc networks, IEEE GLOBECOM 2008 (2008)

(3) Nishiuma, N., et al. : Travel Time Prediction using Singular Value Decomposition, J. Soc. Instr. & Control Engrs Jpn (2008)

(4) Nikolova E., et al. : Stochastic Shortest Paths via Quasi-Convex Maximization, Proc. European Symp. Algorithms (ESA) (2006)

(5) Nikolova, E., et al. : Optimal Route Planning under Uncertainty, ICAPS 2006 (2006)

(6) Agrell E., et al. : Power-efficient modulation formats in coherent transmission systems, IEEE J. Lightwave Tech., **27**, 5115～5126 (2009)

(7) Millar, D.S., et al. : High-dimensional modulation for coherent optical communications systems, Optics Express, **22**, No. 7, 8798～8812 (2014)

(8) Brand, M. : Fast Low-Rank Modifications of the Thin Singular Value Decomposition, Linear Algebra and Its Applications, **415**, No. 1, 20～30 (2006)

(9) Brand, M., et al. : Parallel Quadratic Programming for Image Processing, IEEE Imt'l Conf. Image Processing (ICIP), 2261～2264 (2011)

(10) Di Cairano S., et al. : Projection-free Parallel Quadratic Programming for Linear Model Predictive Control, Int'l J. Control (2013)