Fondation mathématique FMJH Jacques Hadamard

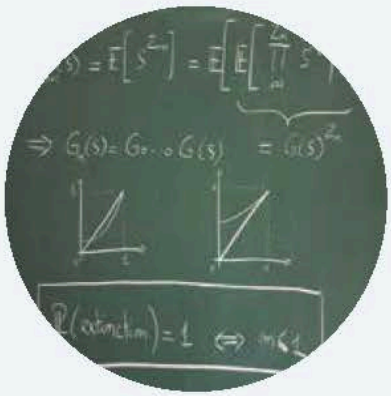
EDF

PGMO-COPI'14

GASPARD MONGE PROGRAM FOR OPTIMIZATION
CONFERENCE ON OPTIMIZATION & PRACTICES IN INDUSTRY

28 - 31 october 2014 Paris-Saclay (Ecole Polytechnique)

PROCEEDINGS



www.fondation-hadamard.fr/fr/PGMO

pgmo.copi.14@fondation-hadamard.fr

Inria · Road · Société Mathématique de France SMF · ENSTA ParisTech · SMAI · X · digiteo · GDR RO · GdR MOA

web

PGMO - COPI'14

Solving Stochastic Dynamic Programs for Multi-reservoir Hydropower Optimization.

Mehdi Towhidi. · Pascal Côté. · Dominique Orban

Abstract

We present a method for solving the Stochastic Dynamic Programming (SDP) "cost-to-go function" when applied to multi-reservoir stochastic optimization. We show how the SDP can be solved efficiently using Sequential Linear Programming (SLP). The SLP algorithm significantly reduces the number of evaluation of the SDP "cost-to-go function". The method is compared with the state-of-the-art interior-point implementation Ipopt and a heuristic approach where some important decision variables are discretized. For all three methods, SDP algorithm is used to solve two stochastic hydropower optimization problems where the first one includes only linear constraints and the second one involves linear and non-linear constraints. Numerical results are obtained using real hydropower systems in Quebec, Canada, manage by Rio Tinto Alcan.

Keywords: Stochastic Dynamic Programming, Sequential Linear Programming, Hydropower Reservoir, Stochastic Optimization.

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

Stochastic Dynamic Programming (SDP) has been intensively used in the past decades for solving multi-reservoir hydropower systems [1-2]. This algorithm aim to compute the so-called "cost-to-go function" (CF) used to compute the operating policy of the reservoirs. The CF depends on the water stored in the reservoirs at the beginning of the period and the water inflow to the installations. The evaluation of the function requires the solution of a small-scale optimization problem where the decision variables are optimized to maximize the current benefits plus the expected future benefits of water stored at the end of the period. So, the SDP algorithm can be viewed as a procedure that breaks down the multi-stage optimization problem into a series of one-stage small-scale optimization problems. But this function cannot be solved for all possible values of reservoir storage and water inflows (which would be impossible), but only for a set of discretization points. However, when the numbers of reservoirs exceed 3 or 4, the computation time of the SDP algorithm becomes very large. Efforts have been made to reduce the computation time of this algorithm. For example, it is possible to reduce the number of discretization points of the state variable [3-4]. With this strategy, the number of optimization problem to solve is reduced but, on the other hand, the CF is defined on a coarser grid. Another strategy consists in solving the optimization problems in a way that the number of evaluation of the objective function is reduced. Like in [6], if we have access to the derivatives of the objective function, an efficient non-linear programming approach can be used to solve the optimization problem. This strategy is used in this project and we present a new approach based on Sequential Linear Programming to solve the CF of SDP.

The aim of this project is to develop a new method to compute the cost-to-go function efficiently and reduce the computation time of the SDP algorithm applies to a hydropower system. The system, which is run by Rio Tinto Alcan Quebec Power Operations Division, consists of 6 generating stations and 3 major reservoirs on Peribonka and Saguenay Rivers in Quebec, Canada, for an installed capacity of 3100 MW. The objective function of the one-stage optimization problem that must be solved to evaluate the CF includes the summation of two terms: the current benefits function and the expected water value function. The current benefits function is given by the hydropower function of the generating stations. At a weekly time step (step for which the SDP is applied), the hydropower function is smoothed to avoid discontinuity or non-differentiability. From an operational point of view, this approximation is good since a short-term optimization problem will be solved to disaggregate the mid-term solution. The expected water value function is actually the CF function at the next period and is obtained

recursively by applying SDP. Thus, this part of the objective function is also smooth. According to these properties, we can approximate the objective function accurately with a linear model. In addition, the models typically have few nonlinear constraints. We thus expect the trust-region Sequential Linear Programming (SLP) paradigm to be efficient. Nonlinear constraints are penalized in the L1 sense and the resulting non-smooth problem is rewritten using elastic variables. Starting from a feasible point, an improved candidate is found by solving a linear sub-problem. We compare SLP with the state-of-the-art interior-point implementation Ipopt and a "heuristic enumerative" approach where some important decision variables are discretized. These approaches are tested on two problems. The first is a sub-problem involving only the last two reservoirs of the system. In this case, the CF sub-problem is composed of 12 variables and 5 linear constraints. The second problem is the entire hydropower system problem including all 6 generating stations and the three reservoirs. Each CF sub-problem problem has 24 variables and 16 constraints (5 non-linear constraints and 11 linear constraints). We apply SDP to both problems with SLP, Ipopt and the heuristic approach. For each run of SDP, SLP gives a significant reduction in computation time without affecting the SDP operating policy.

[1] Labadie, J.W., 2004. Optimal operation of multi-reservoir systems: State-of-the-art review. Journal of Water Resources Planning and Management 130, 93–111.

[2] Celeste, A.B., Billib, M., 2009. Evaluation of Stochastic Reservoir Operation Optimization Models. Advances in Water Resources 1429–1443.

[3] Cervellera, C., Wen, A., Chen, V.C.P., 2007. Neural network and regression spline value function approximations for stochastic dynamic programming. Computers & Operations Research 34, 70–90. doi:10.1016/j.cor.2005.02.043

[4] Stedinger, J.R., Faber, B.A. , Lamontagne, J.R., 2013. Developments in stochastic dynamic programming for reservoir operation optimization. World Environmental and Water Resources Congress 2013: Showcasing the Future; Cincinnati, OH; United States; 19 May 2013 through 23 May 2013; Code 100671

[5] Efi Foufoula-GeorgiouPeter K. Kitanidis, 1988. Gradient dynamic programming for stochastic optimal control of multidimensional water resources systems. Water Resources Research
Volume 24, Issue 8, pages 1345–1359, August 1988.

Mehdi Towhidi
Mathematics and Industrial Engineering Department, École Polytechnique de Montreal, Montreal, QC, H3C 3A7, Canada
E-mail: mehdi.towhidi@gerad.ca

Pascal Côté
Québec Power Operation, Rio Tinto Alcan, Saguenay, QC, G7S 4R5 Canada
E-mail: pascal.cote@riotinto.com

Dominique Orban
Mathematics and Industrial Engineering Department, École Polytechnique de Montreal, Montreal, QC, H3C 3A7, Canada
E-mail: dominique.orban@gerad.ca

PGMO - COPI'14

Short-term unit commitment and loading problem of a multi reservoir hydropower system

Sara Séguin. · Pascal Côté. · Charles Audet. · Alec Mercier.

Abstract

We present a new method for solving the short-term unit commitment and loading problem of a hydropower system. Dynamic programming is used to compute maximum power output generated by a power plant. This information is then used as input of a two-phase optimization process. The first phase solves the relaxation of a nonlinear mixed-integer program in order to obtain the water discharge, reservoir volume and optimal number of units working at each period in the planning horizon. The second stage solves a linear integer problem to determine which combination of turbines to use at each period. The goal is to maximize total power produced over all periods of the planning horizon which consists of a week divided into hourly periods. Start-up of turbines are penalized. Numerical experiments are conducted on the Rio Tinto Alcan power plants in Quebec, Canada.

Keywords Hydro unit commitment, nonlinear programming, linear integer programming.

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

The planning of hydroelectric systems is complex and requires different optimization processes. A good planning allows to produce more power with the same quantity of water, generating substantial savings for the producer, even with a slight computational improvement. Short-term optimization is mandatory to determine how to split the available water volume in an optimal way between the turbines of a plant. Each turbine has a different efficiency curve which means that for the same water discharge the power will differ. The planning horizon is a week divided into hourly periods and the problem consists of finding the optimal water discharge as well as the volume of the reservoir for each plant in order to maximize power production, and penalize turbine start-ups. These optimization problems are difficult to solve since the hydroelectric production functions are non-convex. They are also highly nonlinear and depend on turbine efficiency, net water head which is a nonlinear function of the water discharge and reservoir elevation and finally, water discharge of each unit. Furthermore, turbines have forbidden operating zones, which complicates the problem. Short-term unit commitment and loading problem have been studied in the past and many researches are still undergoing. Many methods have been proposed to solve the short-term unit commitment and loading problem, including dynamic programming [1,2], linear programming [3], nonlinear programming [4,5] or global optimization techniques [6,7].

In this project, we propose a new approach for modeling the short-term unit commitment and loading problem that requires two stages and allows to find a feasible solution at the end of the first stage. The models are then tested on the Saguenay-Lac-St-Jean hydroelectric system which is privately owned by Rio Tinto Alcan in the province of Quebec, Canada. This company operates aluminum plants in that region and can produce 90% of the energy they need to operate them. The installed capacity is of 3100 MW and is composed of 42 turbines divided in five hydroelectric plants. Five reservoirs are available and three of them have a stocking capacity of over 2000 hm^3.

We propose a model with a reasonable number of variables, embedded into a two-stage optimization approach.
A dynamic programming algorithm is used as a pre-process to the optimization method. Total power output that can be generated by a certain combination of active turbines is computed. For every discretization of the water discharge, volume and combinations of active turbines, the algorithm calculates the power output. The maximum power output for every disretization of the water discharge, volume and for each number of active turbines is retained and a maximum output surface is created for each number of active turbines. Smoothing splines are then used to fit the data obtained and the optimization models attempt to maximize the total energy production given by these splines. This allows to solve the models in a very short computational time since the hydroelectric

production functions are calculated independently. Since the goal is also to limit the unit restarts, the optimization method uses two models. The first stage solves the relaxation of a nonlinear mixed-integer program in order to find volume, water discharge and number of active turbines at each period. The objective function maximizes energy production and the constraints are the water balance, choice of one surface of active turbines as well as volume and water discharge limits. The second stage solves a linear integer model to find the exact combination of turbines that maximizes total power but also penalizes start-up of turbines. Constraints are choosing one combination of turbines and the link between the start-up of turbines and the combination used.

Numerical comparisons of our approach versus the real historical values were conducted on two hydropower plants with both five turbines. The planning horizon consisted of one week divided in hourly periods for a total of 168 periods. The first nonlinear model has 1680 real variables and 674 constraints. The linear integer model has 7056 binary variables and 26918 constraints. The computational time to solve the unit commitment model is very low (a few seconds) so that the approach proposed allows us to find a solution in a computational time that is more than satisfying for needs of operation. Thirty test cases were compared. In 27 of the cases, an improvement ranging from 0.002 GWh to 2.145 GWh is observed, with all cases given the same initial starting point. The average improvement for all cases is of 0.4 GWh. We found that our approach is slightly sensitive to the starting point but very little work has been done on selecting the starting point. Multistarts or variable neighborhood searches will be the subject of future research. Also, other developments based on this method will involve using uncertainty related to inflows in order to create a stochastic programming model.

[1] A. Arce, T. Ohishi, and S. Soares. Optimal dispatch of generating units of the itaipu hydroelectric plant. IEEE Transactions on Power Systems, 17(1):154–158, 2002.

[2] J. Yi, J. W. Labadie, and S. Stitt. Dynamic optimal unit commitment and loading in hydropower systems. Journal of Water Resources Planning and Management, 129(5):388–398, 2003.

[3] A. Borghetti, C. D'Ambrosio, A. Lodi, and S. Martello. An milp approach for short-term hydro scheduling and unit commitment with headdependent reservoir. IEEE Transactions on Power Systems, 23(3):1115–1124, 2008.

[4] M. Kadowaki, T. Ohishi, L. S. A. Martins, and S. Soares. Short-term hydropower scheduling via an optimization-simulation decomposition approach. In 2009 IEEE Bucharest PowerTech: Innovative Ideas Toward the Electrical Grid of the Future, 2009.

[5] Finardi, E.C., daSilva, E.L., 2006. Solving the Hydro Unit Commitment Problem via Dual Decomposition and Sequential Quadratic Programming. IEEE Transactions on Power Systems 21, 835–844. doi:10.1109/TPWRS.2006.873121.

[6] S. O. Orero and M.R. Irving. A genetic algorithm modelling framework and solution technique for short term optimal hydrothermal scheduling. Power Systems, IEEE Transactions on, 13(2):501–518, 1998.

[7] C. Ma. Short term hydropower dispatching optimization of cascaded hydropower stations based on two-stage optimization. In 2010 2nd International Conference on Industrial and Information Systems, IIS 2010, volume 1, pages 230–233, 2010.

Sara Séguin
Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, Montréal, QC, H3C 3A7 Canada
E-mail: sara.seguin@polymtl.ca

Pascal Côté
Québec Power Operation, Rio Tinto Alcan, Saguenay, QC, G7S 4R5 Canada
E-mail: pascal.cote@riotinto.com

Charles Audet
Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, Montréal, QC, H3C 3A7 Canada
E-mail: charles.audet@polymtl.ca

Alec Mercier
Québec Power Operation, Rio Tinto Alcan, Saguenay, QC, G7S 4R5 Canada
E-mail: alec.mercier@riotinto.com

PGMO - COPI'14

Finding optimal strategies of almost acyclic simple stochastic games

Pierre Coucheney

The optimal value computation for turned-based stochastic games with reachability objectives, also known as simple stochastic games, is one of the few problems in NP \ coNP which are not known to be in P. However, there are some cases where these games can be easily solved, as for instance when the underlying graph is acyclic. In this work, we try to extend this tractability to several classes of games that can be thought of as "almost" acyclic. We give some fixed parameter tractable or polynomial algorithms in terms of different parameters such as the number of cycles or the size of the minimal feedback vertex set.

Pierre Coucheney
PRiSM
Université de Versailles - St Quentin en Yvelines E-mail: pierre.coucheney@uvsq.fr

PGMO - COPI'14

European energy equilibrium and decomposition

Anes Dallagi

Abstract We consider addressing the issue of numerical resolution of network multistage stochastic optimization problems. We will particularly focus on different decomposition schemes and present different coordination algorithms approximating optimal strategies in a stochastic framework. We apply these schemes to a long-term stochastic model of the European electricity market.

Keywords Stochastic optimization · splitting Algorithm

1 Introduction

Modeling energy transmission is a very important issue for any electric utility in Europe. It is strategic for utility companies to be able to plan the energy exchange evolution up to 30 years ahead. This prospective is used for example in deciding which capacity we need to expand over the next years. The way the prices of commodities evolve and an accurate forecasting of the consumption load are related issues. Due to the time horizon of these problems one should take into account the uncertainty in order to provide reliable decisions. Furthermore, due to computational issues (curse of dimensionality), one can think about computing decentralized strategies for long term investments and daily power production. This turn out to be sub-optimal for most of network configurations and we need elaborated coordination schemes in order to approach optimality.

2 The model

We consider a network with zones and links. Each zone has its own production units and is subject to a unit commitment problem: it has to satisfy a stochastic demand using its hydro and thermal units and eventually importing and exporting using its links.

We denote by $F_a(q_a)$ the cost of producing in Area a while exporting the quantity q_a, $F(q) = \sum_{a \in A} F_a(q_a)$ and $C(q)$ will denote the cost of transmitting through the network the quantities $q = (q_a)_{a \in A}$, where A is the set of modeled areas.

Thus, the problem to be solved can be presented as follow :

Anes Dallagi
EDF R&D, 1 Av Gnral De Gaulle, 92141 Clamart
Tel.: +33-(0)1 47 65 33 87
E-mail: anes.dallagi@edf.fr

$$\min_q \quad F(q) + C(q)$$

3 The Algorithm

Splitting methods aims at finding a primal dual solution to the system:

$$\lambda \in \partial F(q)$$
$$-\lambda \in \partial C(q)$$

where $\partial F(\cdot)$ and $\partial C(\cdot)$ denote the subgradient operator associated to functions $F(\cdot)$ and $C(\cdot)$. We are using now the forward-backward method developped in ([LM79], [Pas79]).

1. Choose arbitrarily $q^{k=0}$
2. Find $\lambda^{k+1} \in \partial F(q^k)$
3. Compute $q^{k+1} = \text{argmin}_q \quad C(q) + < \lambda^{k+1}, q > + \frac{1}{2\rho^k}||q - q^k||^2$
4. Go to step 2 with $k \leftarrow k + 1$

When one of the two functions is additive, say $F(\cdot)$, the problem rewrites:

$$\min_q \quad C(q) + \sum_{a \in A} F_a(q_a)$$

then the resulting first step of the forward-backward algorithm becomes separated. It amounts to finding for all $a \in A$ $\lambda_a^{k+1} \in \partial F_a(q_a^k)$. The second step involving the "coupling function" $C(\cdot)$ will be interpreted as a "coordination step".

In our presentation we will try to answer and illustrate two main questions :

1. How to extend this algorithm when the variables are stochastic?
2. Which interpretation can we make of the original algorithm and the different approximations that will be made.

References

[LM79] P.L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16:964–979, 1979.

[Pas79] G.B Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, (72):383–390, 1979.

PGMO - COPI'14

A Unified View on Tight Formulations for the Unit Commitment Problem with Optimal Transmission Switching

Fabio D'Andreagiovanni · Michela Di Lullo · Giovanni Felici · Fabrizio Lacalandra

Abstract The classical Unit Commitment Problem (UC) consists in determining the production of a set of thermal and hydro power generation units over a given time horizon, in order to meet a forecast energy demand and minimize a cost function [2]. In the last years, in connection with the UC, increasingly attention has been given to active switching, namely the possibility of changing the topology of the transmission network by tripping some of the lines. Active switching has been indeed recognized as an important way to improve capacity and reliability of the network. Modifying the network topology is also very useful in modern electricity systems, where the penetration of non-programmable renewable energy sources is high and tends to be very concentrated in some zones of the network. As a consequence, the interest for automatic and optimized switching procedures is currently spreading [1].

In this paper, we address the UC when complemented by active switching, thus leading to so-called *Unit Commitment with Optimal Transmission Switching* (UCOTS). We formulate the UCOTS as a mixed integer linear program, where we unify tight formulations recently proposed in literature for relevant subfamilies of constraints of the problem (e.g. [4]). Concerning the cost function, in our formulation we approximate the quadratic thermal cost objective function by a piecewise linear function based on perspective cuts [3] and we propose a new simple and effective way to perturb it, with the aim of breaking symmetries induced by optimal transmission switching decision variables. Our original solution approach also exploits specialized branching priorities.

Computational experiments on realistic UCOTS instances highlight that our new framework is able to find optimal and near-optimal solutions in reasonable amount of time by direct use of a state-of-the-art commercial MIP solver.

A preliminary version of this work is presented in [5], to which we refer the reader for details about the optimization model.

Fabio D'Andreagiovanni
Dept. of Optimization, Zuse-Institut Berlin (ZIB), Takustr. 7, 14195 Berlin, Germany
DFG Research Center MATHEON, Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany
Einstein Center for Mathematics Berlin (ECMath), Straße des 17. Juni 135, 10623 Berlin, Germany
E-mail: d.andreagiovanni@zib.de

Michela Di Lullo
Dept. of Computer, Control and Management Engineering, Sapienza Universit di Roma, via Ariosto 25, 00185 Roma, Italy
E-mail: dilullo@dis.uniroma1.it

Giovanni Felici
Institute for Systems Analysis and Computer Science (IASI-CNR) via dei Taurini 19, 00185 Roma, Italy
E-mail: giovanni.felici@iasi.cnr.it

Fabrizio Lacalandra
Quantek s.r.l., via Marconi 3, 40122 Bologna, Italy
E-mail: fabrizio.lacalandra@quantek.it

Keywords Unit Commitment · Transmission Switching · Tight Formulations · Perspective Cuts · Symmetry Breaking

References

1. K.W. Hedman, S.S. Oren, R.P. O'Neill: "A review of transmission switching and network topology optimization". Power and Energy Society General Meeting, 2011 IEEE, pp. 1-7, 2011.
2. M Tahanan, W. van Ackooij, A. Frangioni, F. Lacalandra: "Large-scale Unit Commitment under uncertainty: a literature survey". Technical Report 14-01, Dipartimento di Informatica, Universitá di Pisa, 2014.
3. A. Frangioni, C. Gentile: "Perspective cuts for a class of convex 01 mixed integer programs", Mathematical Programming 106 (2), pp. 225–236, 2006.
4. G. Morales-Espaa, J. M. Latorre, and A. Ramos: "Tight and compact MILP formulation of start-up and shut-down ramping in unit commitment", IEEE Transactions on Power Systems 28 (2), pp. 1288-1296, 2013.
5. S. Bruno, M. Di Lullo, G. Felici, F. Lacalandra, M. La Scala, "Tight Unit Commitment models with Optimal Transmission Switching: Connecting the dots with Perturbed Objective Function", submitted to the 2014 Workshop on Complexity in Engineering 2014, 2014.

PGMO - COPI'14

Robust multi-objective optimization with surrogate models

Sebastien Da Veiga

Abstract In the robust optimization framework, available methods usually encompass two intertwining stages: an optimization loop on the controllable variables requires several propagations of the uncertain variables to evaluate the criterion to be optimized (e.g. the mean of the output of interest). When several criteria are considered, multi-objective algorithms necessitate even more propagations since they generally involve a higher number of function evaluations. When the objective functions are obtained through a computationally expensive simulator (e.g. resolution of large-scale PDEs), such a straightforward approach is impractical.
A standard technique to overcome this computational burden is to replace expensive calls to the simulator by a so-called surrogate model, which is cheap to evaluate. This approximating model is built with a small number of simulator evaluations and is used as a proxy during propagation and optimization loops. In particular, Gaussian Process (GP) regression is a popular nonparametric model which has been used in many industrial applications for the past decade [11,10]. Besides its interpolation property, it also provides an estimate of the prediction error at any unobserved points. However, the response surface step suffers from non-adaptivity: indeed, the simulations used to build the proxy model are fixed a priori. This leads to a global approximating model, while it should be refined with respect to our goal. The standard way to overcome this limitation is to use a goal-oriented adaptive strategy. For example, the classical Efficient Global Optimization strategy (EGO [5]) builds upon the GP properties to propose an adaptive algorithm for mono-objective optimization: starting from an initial design, additional simulations are added in order to refine the proxy model in promising regions, i.e. where the minimum is likely to be and where the model is not predictive. Interesting regions are defined with the so-called Expected Improvement (EI) criterion, that is a balance between exploration and optimization. Several extensions for multi-objective optimization have also been proposed ([4,6–9,13,12]).
In this paper, we generalize these ideas to the robust multi-objective framework. More precisely, we first recall the extension of EGO for the robust minimization of a mean which was developed by Janusevskis & Le Riche [3]. We also detail the HyperVolume Expected Improvement (HV-EI) criterion, which is the generalization of EI for multiple objectives [2]. In a nutshell, it measures the improvement in hypervolume that could be gained by adding a simulator evaluation to the experimental design used to build the proxy model. For robust multi-objective optimization, we deal with dependent GP models (e.g. the mean of the simulator vs its variance): our main contribution is to generalize and study the extension of HV-EI for correlated objectives. Our methodology is then the following:

Sebastien Da Veiga
Snecma
2 chemin de Viercy
Aerodrome de Villaroche
77019 Melun
E-mail: sebastien.daveiga@snecma.fr

1. Build an initial experimental design (in the space of controllable and uncertain variables) and get the corresponding simulator evaluations
2. Compute the GP surrogate model for each objective (mean, variance, quantile, ...)
3. Estimate the Pareto front by providing the proxy models as objectives to any multi-objective optimization algorithm (e.g. NSGA2 [1])
4. Find the best point with respect to the generalized HV-EI criterion and perform the simulation.
5. Add this new simulation result to the experimental design, update the GP models and go back to step 3.

This loops usually stops when it reaches a criterion specified by the user or the maximum number of simulations allowed.
Finally, we illustrate the potential of our methodology on a synthetic example and on a reservoir engineering problem related to well placement. In practice, the issue is to determine where to drill new wells in an existing oil reservoir in order to maximize the potential production. In this setting, uncertainty comes from the geological parameters of the reservoir model and the goal is to estimate well locations that maximize the expected recovery while minimizing the risk expressed by its variance.

Keywords Robust optimization · Multi-objective optimization · Surrogate models · Gaussian process regression

Mathematics Subject Classification (2000) MSC 90C29 · MSC 62G08

References

1. K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: Nsga-ii. *IEEE Trans. Evol. Comput.*, 6 (2):182–197, 2002.
2. M. Emmerich, A.H. Deutz, and J.W. Klinkenberg. The computation of the expected improvement in dominated hypervolume of pareto front approximations. Technical Report 4-2008, Leiden Institute of Advanced Computer Science, LIACS, 2008.
3. J. Janusevskis and R. Le Riche. Simultaneous kriging-based estimation and optimization of mean response. *J. Glob. Optim*, 55 (2):313–336, 2013.
4. S. Jeong and S. Obayashi. Efficient global optimization (ego) for multi-objective problem and data mining. In et al. Corne, D., editor, *Proc. CEC*, pages 2138–2145, Los Alamitos, 2005. IEEE.
5. D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black-box functions. *J. Glob. Optim*, 13 (4):455–492, 1998.
6. A.J. Keane. Statistical improvement criteria for use in multiobjective design optimization. *AIAA J.*, 44 (4):879–891, 2006.
7. J. Knowles. Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Trans. Evol. Comput.*, 10 (1):50–66, 2006.
8. W. Liu, Q. Zhang, E. Tsang, C. Liu, and B. Virginas. On the performance of metamodel assited moea/d. In L. Kang, Y. Liu, and S. Zeng, editors, *ISICA 2007. LNCS*, volume 4683, pages 547–557. Springer, Heidelberg, 2007.
9. W. Ponweiser, T. Wagner, D. Biermann, and M. Vincze. Multiobjective optimization on a limited amount of evaluations using s-metric selection. In G. Rudolph, T. Jansen, S. Lucas, C. Poloni, and N. Beume, editors, *PPSN 2008. LNCS*, volume 5199, pages 784–794. Springer, Heidelberg, 2008.
10. Carl Edward Rasmussen and CKI Williams. *Gaussian processes for machine learning. 2006*, volume 38. The MIT Press, Cambridge, MA, USA, 2006.
11. Thomas J Santner, Brian J Williams, and William I Notz. *The design and analysis of computer experiments*. Springer Verlag, 2003.
12. T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser. On expected-improvement criteria for model-based multi-objective optimization. In *Parallel Problem Solving from Nature XI*, pages 718–727. Springer, 2011.
13. Q. Zhang, W. Liu, E. Tsang, and B. Virginas. Expensive multiobjective optimization by moea/d with gaussian process model. *IEEE Trans. Evol. Comput.*, 14 (3):456–474, 2010.

PGMO - COPI'14

Decomposition Methods in Multistage Stochastic Optimization

Michel De Lara · Pierre Carpentier · Jean-Philippe Chancelier · Vincent Leclère

Abstract Emerging power systems are becoming more and more complex, with the increase of intermittent and variable renewable energies, the expansion of networks and markets and the penetration of smart devices and storage. It follows that optimizing energy systems becomes more and more difficult. As optimization is challenged by the complexity due to large size, dynamical aspects, and uncertainties, we feel that decomposition approaches may prove particularly adapted. This is why we present, in an unified framework, the main approaches to decompose multistage stochastic optimization problems for numerical resolution. This framework covers both *Stochastic Programming (SP)* (and scenario-based resolution methods) and *Stochastic Optimal Control (SOC)* (and state-based resolution methods like *Stochastic Dynamic Programming (SDP)*), the two most well-known approaches and methods in multistage stochastic optimization. This done, we go in more detail and outline more specific ones like *Progressive Hedging (PH)*, *Stochastic Dual Dynamic Programming (SDDP)*, and *Dual Approximate Dynamic Programming (DADP)*.

Keywords optimization · stochastic · multistage · decomposition

We present, in an unified framework, the main approaches to decompose multistage stochastic optimization problems for numerical resolution.

To fix ideas and simplify the exposition, we present a setting where all variables are parametrized by discrete indexes. For this purpose, suppose given a finite integer horizon $T \geq 1$ (so that the discrete time $t \in \{0, T\}$), a finite probability space $\big(\Omega, \mathcal{F}, \mathbb{P}\big)$, endowed with a filtration $\{\mathcal{F}_t\}_0^{T-1}$, a finite number N of units (space). We consider the multistage stochastic optimization problem

$$\min_{\boldsymbol{X}, \boldsymbol{U}} \quad \sum_{\omega \in \Omega} \sum_{i=1}^{N} \sum_{t=0}^{T-1} \mathbb{P}\big(\{\omega\}\big)\, L_t^i\Big(\boldsymbol{X}_t^i(\omega), \boldsymbol{U}_t^i(\omega), \boldsymbol{W}_t(\omega)\Big) \tag{1a}$$

$$s.t. \quad \boldsymbol{X}_{t+1}^i(\omega) = f_t^i\big(\boldsymbol{X}_t^i(\omega), \boldsymbol{U}_t^i(\omega), \boldsymbol{W}_t(\omega)\big) \qquad \forall t\ ,\ \forall i\ ,\ \forall \omega \tag{1b}$$

$$\sum_{i=1}^{N} \theta_t^i\big(\boldsymbol{X}_t^i(\omega), \boldsymbol{U}_t^i(\omega)\big) = 0 \qquad \forall t,\quad \forall \omega \tag{1c}$$

$$\boldsymbol{U}_t^i \text{ is } \mathcal{F}_t \text{ -measurable} \qquad \forall t\ ,\ \forall i\ , \tag{1d}$$

Michel De Lara, Jean-Philippe Chancelier, Vincent Leclère
Cermics, École des Ponts ParisTech
E-mail: delara@cermics.enpc.fr, E-mail: jpc@cermics.enpc.fr, E-mail: vincent.leclere@gmail.com

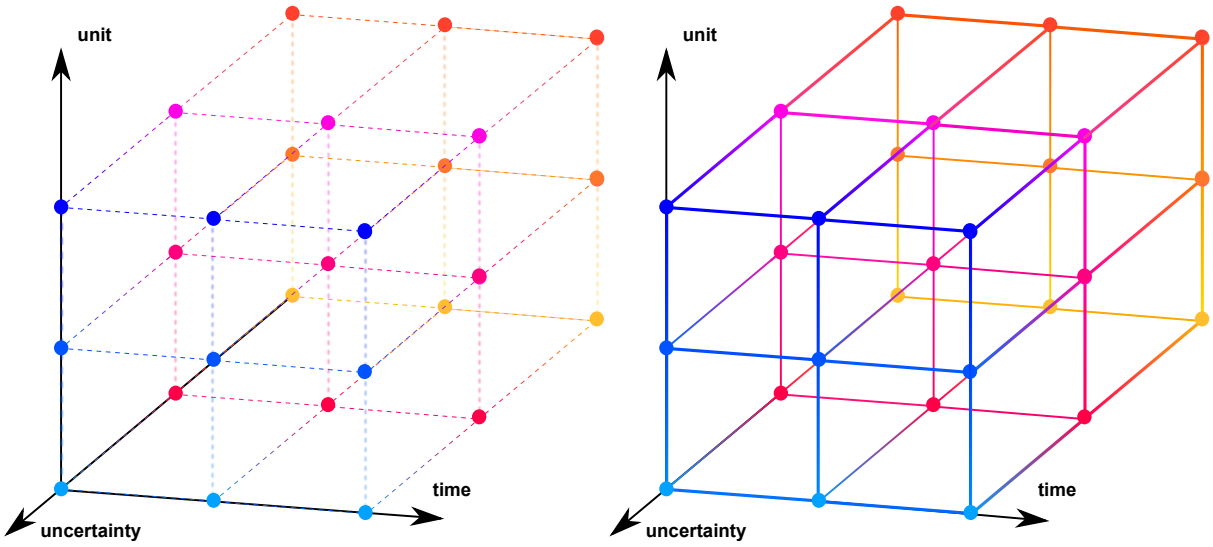
Pierre Carpentier
UMA, ENSTA ParisTech
E-mail: pierre.carpentier@ensta-paristech.fr

where ω is a scenario of uncertainties given by $\omega = \{\omega_t\}_{t=0}^{T-1}$. The constraint (1b) represents the dynamics of each subsystem, the constraint (1c) represents the coupling constraint between the subsystems (also called units), and the constraint (1d) is the non-anticipativity constraint. Constraints function θ_t^i are assumed to have image in $\mathbb{R}^{n_c}$.

As the SOC framing [1,8,9,4,2] includes the SP one [6,3], the above setting applies both to SP and SOC problems.

In Problem (1), we minimize the sum of local costs — depending on step t, uncertainty ω and unit i — over time, uncertainty and space. Were the constraints (1b)-(1d) absent, Problem (1) (illustrated in Figure 1a) would consist in minimizing a sum of independent costs; as the minimum of the sum is the sum of the minima, the optimization problem would be decomposed. However, as illustrated in Figure 1b, the local costs are linked with respect to

- time through the dynamics of the system (e.g. Equation (1b));
- unit through the coupling constraints (e.g. Equation (1c));
- scenario (uncertainty) through the nonanticipativity constraint (e.g. Equation (1d)).



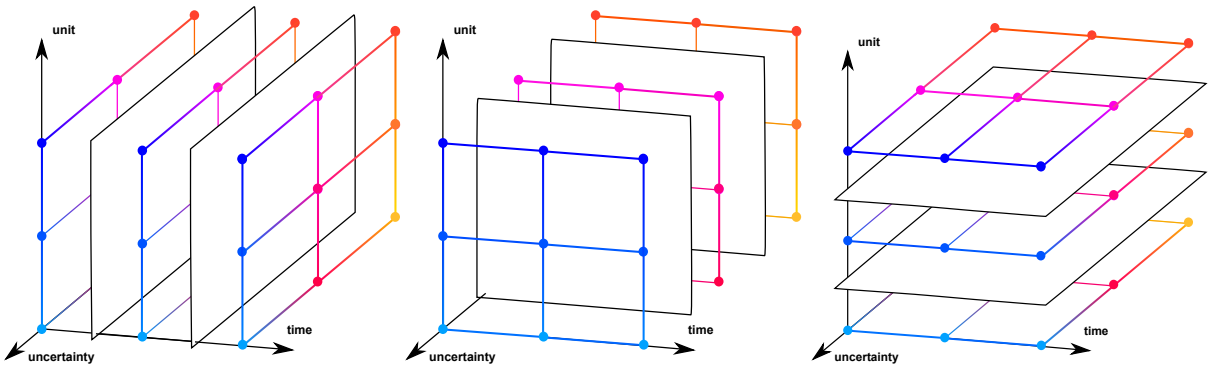
(a) Local costs (b) Local costs linked

Fig. 1: Representation of the local costs depending on time, uncertainty (scenario) and space (unit) and the links induced by the constraints

We lay out different ways to divide the original complex problem into easier to solve subproblems. We propose three angles to decompose the original problem: decomposition in time (step), decomposition in scenario (uncertainty) and decomposition in space (unit), as illustrated in Figure 2.

Moreover, we distinguish two types of decomposition.

- In *chained decomposition*, like Dynamic Programming (see [1,2]), the original problem is solved by means of successive smaller subproblems, solved one after the other (in Dynamic Programming, each subproblem is solved only once). Chained decomposition relies on a specific structure of the coupling constraint, like the flow of time.
- In *parallel decomposition*, like Progressive Hedging (see [5,7]), the original problem is solved by means of parallel smaller subproblems, coordinated and updated by a master algorithm. These subproblems can be obtained by dualizing the constraint, and have to be solved several times before obtaining an optimal solution to the global problem.



(a) Time decomposition (b) Uncertainty decomposition (c) Space decomposition

Fig. 2: Decomposition according to time, uncertainty (scenario) or space (unit). Each plane carries a problem with coupling in only two dimensions.

References

1. R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.
2. D. Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, 1995.
3. Alan J. King and Stein W. Wallace. *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2012.
4. M. L. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.
5. R.T. Rockafellar and R. J-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16(1):119–147, 1991.
6. A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on stochastic programming: modeling and theory*. The society for industrial and applied mathematics and the mathematical programming society, Philadelphia, USA, 2009.
7. Jean-Paul Watson and David L Woodruff. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*, 8(4):355–370, 2011.
8. P. Whittle. *Optimization over Time: Dynamic Programming and Stochastic Control*, volume 1. John Wiley & Sons, New York, 1982.
9. P. Whittle. *Optimization over Time: Dynamic Programming and Stochastic Control*, volume 2. John Wiley & Sons, New York, 1982.

PGMO - COPI'14

BtrPlace: Flexible VM Management in Data Centers

Sophie Demassey · Fabien Hermenier

Abstract Data Centers are typical dynamic packing systems that operate at a large scale. Tens of thousands of Virtual Machines (VMs) have to be deployed over thousands servers to execute clients applications. For an efficient usage of resources (CPU, memory, storage, network bandwidth, etc.), virtualization allows to co-locate several VMs on a single server if their cumulated resource demand does not exceed their host capacity. The dynamicity of the system comes from variations in terms of resource demand, VM arrival and departure, and server availability. Dynamicity comes also from the numerous side constraints occasionally expressed by the users of the data center for different concerns. The assignment of the VMs to the servers according to the resource and side constraints is under the control of the so-called VM manager. This software component periodically reconsiders the current assignment to fit the changes. It then schedules the required transitions, including live migrations, in order to minimize the impact on the service. We present BtrPlace [1], an open-source VM manager taking benefits from the expressivity and flexibility of the Constraint Programming paradigm to handle this optimization problem in a dynamic way. Power saving strategies are discussed for giving an insight into the way the VM manager may help to reduce the energy consumption of data centers.

Keywords VM manager · Constraint Programming · Flexibility · Power Saving

1 VM Repacking and Scheduling Problem

A conceptual model of the optimization problem solved by the VM manager can be stated as follows. Consider a 2-states (initial/final) dynamic system which consists of a set $\mathcal{R}$ of p-dimensional bins with static capacities $B_r \in \mathbb{N}^p$, for all $r \in \mathcal{R}$, and a set $\mathcal{J}$ of items with dynamic initial $b^o_j \in \mathbb{N}^p$ and final $b^f_j \in \mathbb{N}^p$ heights, for all $j \in \mathcal{J}$. The initial state of the system is known and defined as an assignment $s_o : \mathcal{J} \to \mathcal{R}$ satisfying $\sum_{j \in s_o^{-1}(r)} b^o_j \leq B_r$ for each bin $r \in \mathcal{R}$. The system state changes by applying a transition action to each item $j \in \mathcal{J}$. The restricted set of allowed transitions is given as a table $\Delta_j \subseteq \mathcal{T} \times \mathcal{R}$, where each element $\delta = (\tau, r)$ indicates that a transition of type $\tau \in \mathcal{T} = \{S, L, M, U\}$ (standing for *Suspend, Launch, Migrate, Unmoved*) can be applied to item $j \in \mathcal{J}$ to reassign it from bin $s_o(j)$ to bin r. With any transition $\delta \in \Delta_j$ are associated a duration $d_\delta \in \mathbb{N}$ and a weight $w_\delta \in \mathbb{N}$. If

Sophie Demassey
MINES ParisTech, PSL Research University, CMA - Centre de mathématiques appliquées, France
E-mail: sophie.demassey@mines-paristech.fr

Fabien Hermenier
University Nice Sophia Antipolis, CNRS, I3S, France
E-mail: fabien.hermenier@unice.fr

feasible, the problem is to associate with each item $j \in \mathcal{J}$, a transition $\delta(j) = (\tau(j), s_f(j)) \in \Delta_j$ and a time $t_j \in \mathbb{N}$ to start this transition, such that the bin capacities are satisfied at any time

$$\sum_{\substack{j \in s_o^{-1}(r)| \\ t < t_j + d_{\delta(j)}}} b_j^o + \sum_{\substack{j \in s_f^{-1}(r)| \\ t \geq t_j}} b_j^f \leq B_r, \quad \forall r \in \mathcal{R}, \forall t \geq 0, \tag{1}$$

and the weighted sum of the completion times is minimized

$$\sum_{j \in \mathcal{J}} w_{\delta(j)}(t_j + d_{\delta(j)}). \tag{2}$$

In this problem, the transition typecast is determined by the item j itself, its origin $s_o(j)$ and destination $s_f(j)$. Hence, determining a set of transitions δ_j comes to compute a feasible Vector Packing. This problem is NP-complete in the strong sense even in the one-dimensional case ($p = 1$). In turn, determining the times t_j yields to a particular scheduling problem where a migrant VM consumes resources on its origin server from time 0 to the end of its migration $t_j + d_j$, and on its destination server from the start of its migration t_j to the end of the transition plan. This can be viewed as a Resource Constrained Scheduling Problem with no-wait, variable durations and consumer/producer tasks.

2 Handling Side Constraints with Constraint Programming

Another characteristic of the VM management problem is the presence of numerous and heterogeneous side constraints. They are expressed by clients and data center operators to restrict the VM placement according to different concerns (security, energy efficiency, performance, etc.). In practice, a constraint forces or disallows the co-location of some VMs or restricts the cumulated resource usage on some servers. As these requirements change over the time, the constraints must also be handled dynamically.

We developed BtrPlace [4], an autonomous VM manager based on Constraint Programming to handle the dynamicity of the problem and the variety of the side constraints. BtrPlace comes with a high-level language in which the users may express their constraints. Automatically and periodically, BtrPlace retrieves the data (the current configuration and the future needs) and translates the user constraints to generate a Constraint Programming model of the problem. Once a first solution is found, by solving the model using an ad-hoc heuristic, the computed transition plan is applied. BtrPlace provides the awaited flexibility to handle the heterogeneous side constraints. It is currently bundled with more than 20 types of side constraints. Simulation shows its effectiveness for managing datacenters with 5,000 servers and 30,000 VMs [5].

3 Application to Power Saving

Rapid rates of growth in data center electricity use prevailed from 2000 to 2005 and slowed to 56% since, yielding total electricity use by data centers in 2010 of about 1.3% of all electricity use for the world [6]. Virtualization is one main factor contributing to slow down this growth. However, electricity consumption is still growing, and as service demands keep rising, energy efficiency has become a new key metric, besides performance and reliability, of data center management.

A number of energy-aware approaches are proposed in the recent literature (see e.g. [2] for a survey). Acting at the level of the VM manager with a consolidation policy – for minimizing the number of servers hosting VMs and turning off the idle ones – is one effective approach for power saving.

The generic and flexible framework of BtrPlace makes possible to integrate such a policy, both within the objective function and as additional side constraints, while taking into account together the contracted Service Level Agreements, including resource provisionning and individual user requirements, and the energy consumption of the servers, including overheads generated by VM migrations.

References

1. BtrPlace, an open-source VM manager. `http://btrp.inria.fr/`.
2. A. Corradi, M. Fanelli, L. Foschini, Management Infrastructures for Power-Efficient Cloud Computing Architectures, in *Cloud Computing*, Springer, p. 133–152, 2013.
3. C. Dupont et al., An energy aware framework for virtual machine placement in cloud federated data centres. *E-energy 2012: the 3rd International Conference on Future Energy Systems*, 2012.
4. F. Hermenier, S. Demassey, X. Lorca, Bin-Repacking Scheduling in Virtualized Datacenters. *17th International Conference on Principles and Practice of Constraint Programming (CP'11), application track*, LNCS 6876:27–41, 2011.
5. F. Hermenier, J. Lawall, G. Muller. BtrPlace: A Flexible Consolidation Manager for Highly Available Applications, *IEEE Transactions of Dependable and Secure Computing*, 10(5):273–286, 2013.
6. J. Koomey, Growth in Data center electricity use 2005 to 2010, *Analytics Press*, 2011.

Assessment of Operation Policies for Different Levels of Reservoir Aggregation models in the Long Term Hydrothermal Scheduling

Vitor Luiz de Matos. · Paulo Vitor Larroyd. · Erlon Cristian Finardi.

Abstract

The Long-Term Hydrothermal Scheduling (LTHS) problem plays an important role in power systems that rely heavily on hydroelectricity. The purpose of the LTHS problem is to define an optimal operation policy that minimizes the operation costs to meet demand over a long horizon. A popular solution approach to this problem is called Stochastic Dual Dynamic Programming (SDDP). To incorporate the inflow uncertainties, the LTHS problem is modeled as a multi-stage linear stochastic problem. In the Brazilian LTHS problem, some simplifications are made in the hydro power plants modeling in order to reduce the computational burden. As a result, this paper assesses the consequences of those simplifications in the operation policies. We show the results considering the Brazilian hydrothermal power system.

Keywords: Stochastic Programming · Long-Term Hydrothermal Scheduling Problem · Reservoir Modeling

Mathematics Subject Classification (2000): MSC 90C15 · MSC 90B36 · MSC 90B05

1 Introduction

The long-term hydrothermal scheduling (LTHS) problem aims to define the dispatch of each power plant in a power system of hydroelectric reservoirs and thermal plants in order to meet the demand for electricity over a long-term planning horizon (of months or years) at minimum expected operation cost (consisting of thermal fuel costs and penalties). In power systems that rely heavily on hydroelectricity and cannot buy energy from neighbor countries, such as those of Brazil and New Zealand, the hydrothermal scheduling problem plays an important role, given that if the water in the reservoirs is not used wisely throughout the operation, the system may present an expensive operation with high shortage risk.

The LTHS problem can be formulated as a stochastic dynamic programming problem, in which reservoirs' inflows are assumed to be random variables with a known continuous probability distribution. As it is impossible to solve the LTHS problem taking into account the continuous probability distribution, it is necessary to simplify the problem by modeling the original continuous distribution as an inflow scenario tree with a finite number of realizations.

In the Brazilian hydrothermal power system, the ISO is responsible for the hydrothermal scheduling of the system and, as a consequence, it uses a chain of optimization models to make its decisions. In this context, the main result of the LTHS problem is the Future Cost Function (FCF), also known as Operation Policy, which represents the system expected operation costs in the future as a function of the stored water and the previous water inflows. This FCF is used as a boundary condition in the subsequent optimization model in the chain ([1], [2]). Due to the size of the scenario tree in a long term problem, the operation policy is obtained by means of the Stochastic Dual Dynamic Programming (SDDP), which was proposed by Pereira and Pinto in 1991[3].

Several models for the Brazilian LTHS problem ([2], [4]), including the official model used by the ISO, aggregates a set of hydro plants in a smaller number of equivalent reservoirs in order to reduce the size of the optimization problem [4]. In this case, the decisions variables are presented in terms of energy instead of water. According to Arvanitidis and Rosing ([5], [6]), the EER is most applicable when the sequence of monthly decisions on the total hydro production is more important than the allocation of this total among the hydro plants.

This modeling was considered in the Brazilian case in order to reduce the computational burden, due to the large number of hydro power plants.

However, the reservoir aggregation tends to be optimistic about the amount of energy stored in the energy reservoir, as inflows to a set of hydro power plant can be stored and/or used in other plants that belong to the same aggregated reservoir. Although the chain of models is supposed to be able to handle these modeling aspects and reduce its consequence, an optimistic operation policy with similar water values for all aggregated hydro plant may lead to a more expensive operation and in some cases to an increased shortage risk without the ISO realizing it.

Therefore, in this paper we assess operation policies with different levels of hydro plants modeling aggregation. We aggregate the hydro power plants in reservoirs that belong to some electrical system in Brazil (official model approach) or hydro plants that are in the same cascade. Naturally, a more detailed modeling requires longer computational times to yield tight policies, when upper and lower bounds of the SDDP algorithm get closer. However, given that the ISO has limited time to compute the policy, it is important to understand the consequences of having a loosen policy in a more detailed modeling or a tighten policy in a simplified modeling. In order to assess the operation policies we consider a set of the Brazilian LTHS problem.

2 Conclusions

In this paper we show that simplifications allow us to build an operation policy that is closer to the optimal policy in terms of the simplified model. However, the simplified operation policy does not yield as good decisions as the policy obtained in a more detailed modeling, even though the optimality gap is bigger in that case. In this paper we analyzed the policies running for a fixed number of iterations and for a fixed time period. It is important to mention that those results indicate that in the Brazilian case we are not interested in monthly decisions of total hydro power production, but we are rather concerned about the water values for each reservoir. As it allows the ISO to understand where the water is more valuable to the system.

1 References

[1] M. V. F. Pereira and L. M. V. G. Pinto, “A Decomposition Approach to the Economic Dispatch of Hydrothermal Systems,” *Power Appar. Syst. IEEE Trans.*, vol. PAS-101, no. 10, pp. 3851–3860, 1982.

[2] M. E. P. Maceira, L. A. Terry, F. S. Costa, J. M. Damázio, and A. C. G. Melo, “Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System,” in *14th Power Systems Computation Conference (PSCC)*, 2002, no. June.

[3] M. V. F. Pereira and L. M. V. G. Pinto, “Multi-stage stochastic optimization applied to energy planning,” *Math. Program.*, vol. 52, pp. 359–375, 1991.

[4] V. L. de Matos and E. C. Finardi, “A computational study of a stochastic optimization model for long term hydrothermal scheduling,” *Int. J. Electr. Power & Energy Syst.*, vol. 43, no. 1, pp. 1443–1452, 2012.

[5] N. V Arvanitidis and J. Rosing, “Composite Representation of a Multireservoir Hydroelectric Power System,” *Power Appar. Syst. IEEE Trans.*, vol. PAS-89, no. 2, pp. 319–326, 1970.

[6] N. V Arvanitidis and J. Rosing, “Optimal Operation of Multireservoir Systems Using a Composite Representation,” *Power Appar. Syst. IEEE Trans.*, vol. PAS-89, no. 2, pp. 327–335, 1970.

Acknowledgements: This work received financial support from Tractebel Energia (in partnership with eight Brazilian companies) through the R&D Project ANEEL (PD-0403-0035-2013).

Vitor Luiz de Matos ✉
Plan4 Engenharia SS.,
Rua Lauro Linhares, 2055, sala 707,
Torre Max - Trindade. Florianópolis, SC,

Brazil, Post Code 88036-003.
Tel.: +55 - (48) 3025.7564
Fax: +55 - (48) 3025.7564
E-mail: vitor@plan4.com.br

Paulo Vitor Larroyd
Laboratório de Planejamento em Sistemas de Energia Elétrica,
Universidade Federal de Santa Catarina,
UFSC - CTC - EEL - LabPlan / Campus Universitário,
Trindade, Florianópolis, SC, Brasil, Post Code 88040-900.
E-mail: paulo.larroyd@posgrad.ufsc.br

Erlon Cristian Finardi
Laboratório de Planejamento em Sistemas de Energia Elétrica,
Universidade Federal de Santa Catarina,
UFSC - CTC - EEL - LabPlan / Campus Universitário,
Trindade, Florianópolis, SC, Brasil, Post Code 88040-900.
E-mail: erlon.finardi@ufsc.br

PGMO - COPI'14

Extended formulations for robust maintenance planning at power plants

Boris Detienne

Abstract We consider the large-scale power plant maintenance scheduling and production planning problem submitted by EDF to the 2010 Euro/Roadef Challenge. Electricity demand must be met by the combined production of nuclear power plants and thermic plants. The latter can operate continuously while the former have to be shut down regularly for refuelling and maintenance, and cannot produce during outage periods. In this model, we introduce uncertainty about the duration of maintenance operations: we assume that the durations of outages can be extended beyond their normal durations. The objective is to minimize the worst-case total power production cost. In the context of a rolling horizon approach, we investigate the impact of different hypotheses on the uncertainty set on the cost of the solutions and on the difficulty to obtain good solutions. Our solver uses an extended formulation of the problem, which is solved by column generation.

Keywords energy ; electricity production ; maintenance planning ; column generation ; robust optimization ; rolling horizon

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Problem

We consider the large-scale power plant maintenance scheduling and production planning problem submitted by EDF to the 2010 Euro/Roadef Challenge [3]. Two types of power plants are used to satisfy a customer demand over a specific time horizon. Thermic plants can operate continuously while nuclear plants have to be shut down regularly for refuelling and maintenance, and cannot produce during outage periods. The decision to be made consists of the dates of outages, the amount of refueling for nuclear plants, and production level for both types of plants.

In the variant of the problem considered here, we assume that the duration of the outages is uncertain. More precisely, the outage of a nuclear plant can be randomly extended within a given range, this change in data being revealed at the expected end of outage. Furthermore, we assume that missing production is counterbalanced by additional production by thermic plants. The cumulative capacity of thermic plants is always sufficient given the presence of an artificial plant that represents outsourcing. Therefore, our model is two-stage: the first-stage decisions are the planning of nuclear plants as well as the refueling levels, while the plannings of thermic plants constitute the second-stage, or recourse, decisions. The

Boris Detienne
Institut de Mathématiques de Bordeaux, Université de Bordeaux, & Team ReAlOpt, INRIA Bordeaux Sud-Ouest E-mail: boris.detienne@math.u-bordeaux1.fr

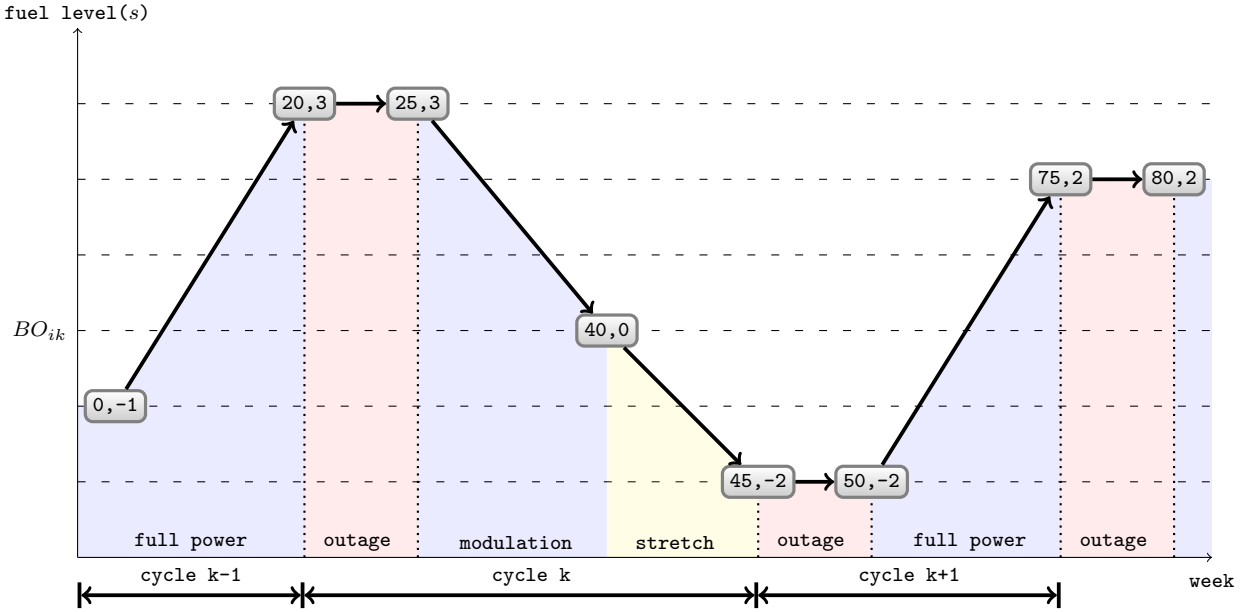


Fig. 1 Path example, node label =(week, fuel level)

objective is to minimize the cost of refueling nuclear plants plus the cost of production by thermic plants.

2 Solution approach

For the first-stage model, we use an extended formulation that we designed for the deterministic version of the problem [2]. The production schedule of nuclear plants is modeled as a path flow in a state network (see Figure 1). Each arc of this network represents a transition from a given pair defined by a time period and a fuel level to another such pair at a later period. To an arc is associated a production vector that defines the level of production of the plant in each of the concerned time periods. Path solutions are generated dynamically as pricing problem solutions in solving the extended formulation by column generation.

In the following Mixed Integer Linear Programming model, simplified for the sake of clarity, d_t is the customer demand at period t, while b_i is zero for all nodes i of the state network, except for source and sink nodes. Variable x_{ij}^k is equal to 1 if arc (i, j) is part of the path for nuclear plant k (first-stage decisions). The amount of production of thermic plant j during t (second-stage decisions) is denoted p_{jt}, while, a_{ij}^{kt} is the amount of production by nuclear plant k during period t, when arc (i, j) is part of the corresponding path. Constraints (2) are the network flow balance constraints, Constraints (3) are the demand satisfaction constraints, and Constraints (4) are various constraints (e.g. minimum spacing between outages...) restricting the set of first-stage variables.

$$(P): \min \sum_{k,i,j} c_{ij}^k x_{ij}^k + \sum_{j,t} C_{jt} p_{jt} \tag{1}$$

$$\sum_{(i,j)\in E^k} x_{ij}^k - \sum_{(j,i)\in E^k} x_{ji}^k = b_i \qquad \forall k, i \tag{2}$$

$$\sum_{k,i,j} a_{ij}^{kt} x_{ij}^k + \sum_{j,t} p_{jt} = d_t \qquad \forall t \tag{3}$$

$$Fx \leq f \tag{4}$$

$$x_{ij}^k \in \{0,1\} \qquad \forall k, i, j \tag{5}$$

$$0 \leq p_{jt} \leq U_{jt} \qquad \forall j, t \tag{6}$$

In the robust extension of the model, the uncertainty on the duration of outages translates into a random modulation of the production vector a_{ij}^k associated to an arc (i, j) in a network k. Assuming that the actual level of production of a plant belongs to an interval between zero and its nominal production capacity, we use Bertsimas and Sim's reformulation [1] to model the simple uncertainty set where, at each time period, not more than Γ outage extensions occur. More complex polyhedral uncertainty sets are adressed using various reformulations of the problem and we show how our optimization algorithms must be modified to take them into account.

We have tested our robust approaches within a rolling horizon framework, to assess the benefits and the relevance of the robust hypotheses.

Acknowledgements Our exprimental work builds on the rolling horizon approach developed by Jinil Han within the ReAlOpt team at INRIA Bordeaux Sud-Ouest. The extended formulation approach for the determinsitic model was initially developed by Cédric Joncour, Ruslan Sadykov and Francois Vanderbeck, in ReAlOpt's work on addressing the ROADEF Challenge. The author thanks the members of ReAlOpt (Francois Clautiaux, Pierre Pesneau, and the above mentionned) for fruitful discussions on this subject. Our problem knowledge arises from Pascale Bendotti, Georgios Petrou and Marc Porcheron, EDF R & D, with whom we have discussed our assumption at length.

References

1. Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
2. Jinil Han, Pascale Bendotti, Boris Detienne, Georgios Petrou, Marc Porcheron, Ruslan Sadykov, and François Vanderbeck. Extended Formulation for Maintenance Planning at Power Plants. In *ROADEF - 15ème congrès annuel de la Société française de recherche opérationnelle et d'aide à la décision*, Bordeaux, France, February 2014.
3. Marc Porcheron, Agnès Gorge, Olivier Juan, Thomas Simovic, and Guillaume Dereu. Challenge roadef/euro 2010: A large-scale energy management problem with varied constraints. *EDF R&D, Clamart, France*, 27, 2010.

PGMO - COPI'14

Optimization algorithms: worst-case behaviour and related conjectures

Antoine Deza

Abstract :

The simplex and primal-dual interior point methods are currently the most computationally successful algorithms for linear optimization. While the simplex methods follow an edge path, the interior point methods follow the central path. Within this framework, the curvature of a polytope, defined as the largest possible total curvature of the associated central path, can be regarded as the continuous analogue of its diameter. In this talk we highlight links between the edge and central paths, and between the diameter and the curvature of a polytope. We recall continuous results of Dedieu-Malajovich-Shub, and discrete results of Holt and Klee and Klee and Walkup, as well as related conjectures such as the Hirsch conjecture, which was disproved by Santos in 2012. We also present analogous results dealing with average and worst-case behaviour of the curvature and diameter of polytopes, including a recent result of Allamigeon, Benchimol, Gaubert, and Joswig who constructed a counterexample to the continuous analogue of the polynomial Hirsch conjecture. Based on joint works with Tamás Terlaky (Lehigh), Feng Xie (Microsoft), and Yuriy Zinchenko (Calgary).

Keywords Linear optimization - central path – interior-point methods – diameter – simplex methods – d-step conjecture

Mathematics Subject Classification (2000) · MSC 90C05 - MSC 90C51 - MSC 90C27 - MSC 52C35

1 Introduction :

Rational decision-making through quantitative modelling and analysis is the guiding principle behind operations research, a field with several far-reaching applications in research and industry. Finding optimal allocations of resources, scheduling tasks, and designing prototypes are a few of the areas operations research is concerned with. In many cases, these problems can be formulated or approximated as linear optimization problems, which involve maximizing or minimizing a linear function over a domain defined by a set of linear inequalities. The simplex and primal-dual interior point methods are currently the most computationally successful algorithms for linear optimization. While the simplex methods follow an edge path, the interior point methods follow the central path. The algorithmic issues are closely related to the combinatorial and geometric structure of the feasible region. Within this framework, the curvature of a polytope, defined as the largest possible total curvature of the associated central path, can be regarded as the continuous analogue of its diameter.

2 Conclusions :

We highlight links between the edge and central paths, and between the diameter and the curvature of a polytope. We recall continuous results of Dedieu, Malajovich, and Shub, and discrete results of Holt-Klee and Klee-Walkup, as well as related conjectures such as the Hirsch conjecture, which was disproved by Santos in 2012. We also present analogous results dealing with average and worst-case behaviour of the curvature and diameter of polytopes, including a recent result of Allamigeon, Benchimol, Gaubert, and Joswig who constructed a counterexample to the continuous analogue of the polynomial Hirsch conjecture

Acknowledgements : This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, and by the Canada Research Chairs program.

Antoine Deza
McMaster University, Ontario, Canada
Tel.: +1-(905) 525 9140 ext. 23750
E-mail: deza@mcmaster.ca

PGMO - COPI'14

The Simplex Algorithm is NP-mighty

Yann Disser · Martin Skutella

Abstract We propose to classify the power of algorithms by the complexity of the problems that they can be used to solve. Instead of restricting to the problem a particular algorithm was designed to solve *explicitly*, however, we include problems that, with polynomial overhead, can be solved '*implicitly*' during the algorithm's execution. For example, we allow to solve a decision problem by suitably transforming the input, executing the algorithm, and observing whether a specific bit in its internal configuration ever switches during the execution.

We show that the Simplex Method, the Network Simplex Method (both with Dantzig's original pivot rule), and the Successive Shortest Path Algorithm are NP-mighty, that is, each of these algorithms can be used to solve any problem in NP. This result casts a more favorable light on these algorithms' exponential worst-case running times. Furthermore, as a consequence of our approach, we obtain several novel hardness results. For example, for a given input to the Simplex Algorithm, deciding whether a given variable ever enters the basis during the algorithm's execution and determining the number of iterations needed are both NP-hard problems. Finally, we close a long-standing open problem in the area of network flows over time by showing that earliest arrival flows are NP-hard to obtain.

Keywords simplex method · network simplex · successive shortest path · NP-mightiness · earliest arrival flow

Mathematics Subject Classification (2000) MSC 68Q25 · MSC 90C05

TU Berlin, Institut für Mathematik,
E-mail: disser@math.tu-berlin.de, E-mail: martin.skutella@tu-berlin.de

PGMO - COPI'14

Towards a Complexity Theory for Black-Box Optimization

Anne Auger · Benjamin Doerr · Carola Doerr · Nikolaus Hansen · Timo Kötzing · Johannes Lengler · Jonathan Rowe

Abstract For most real-world optimization problems, designing efficient algorithms to compute optimal or approximately optimal solutions is far from being feasible. In practice, such problems are therefore typically solved by heuristics. To understand which heuristic is favorable in what situation, we are interested in studying search heuristics from a theoretical perspective.

While analyzing the optimization times of search heuristics is by now one of the well established research streams in the heuristic search community, not much is known on the intrinsic difficulty of optimization problems when tackled by search heuristics. In a series of recent papers, the community has now started to investigate so-called *black-box complexity* notions. The goal of black-box complexity theory is to show that for a given problem a certain effort is necessary to solve it via heuristic search algorithms. That is, black-box complexity focuses on providing lower bounds for the running time of heuristic search algorithms. The difference to classical complexity theory is that in black-box optimization models an algorithm learns about the problem at hand only by sampling and evaluating solution candidates. It does not have any other access to the problem instance. In short, black-box complexity thus asks for how many function evaluations are needed to solve a problem. This notion is well studied under different terms (e.g., query complexity, decision tree complexity) outside of the search heuristics community, though not with a focus of developing a complexity theory for heuristics.

In the context of analyzing search heuristics, several different black-box complexity notions exist. We provide a short overview of the existing models, and show that the study of black-box complexity can inspire the design of new search methods that seem to outperform commonly employed search strategies.

Keywords Black-Box Optimization · Randomized Algorithms · Performance of Search Heuristics

1 Black-Box Complexity

Randomized search heuristics like simulated annealing, evolutionary algorithms, or particle swarm optimization are among the most commonly employed algorithms for real-world optimization problems. They are typically fast to implement, robust with respect to noise, and are observed to compute good solutions in reasonable time.

While randomized search heuristics have been successfully applied to many challenging optimization problems in both industrial and academic contexts, our theoretical understanding of these general-purpose optimizers is much less developed. To date, most research in the theory of randomized search

Affiliations: Anne Auger: INRIA, LRI Orsay · Benjamin Doerr: LIX Palaiseau · Carola Doerr: CNRS, LIP6 (UPMC) · Nikolaus Hansen: INRIA, LRI Orsay · Timo Kötzing: Univ. Jena · Johannes Lengler: ETH Zürich · Jonathan Rowe: Univ. Birmingham

heuristics focuses on computing convergence rates (in continuous domains) or optimization times (in discrete domains) for particular heuristics. Not much is known, however, about the intrinsic complexity of optimization problems for black-box optimization methods like the search heuristics mentioned above. One problem in quantifying this complexity is that the classical (e.g., Turing) complexity models are not suitable to capture this difficulty appropriately: while these models typically assume that the algorithms have full access to the concrete problem instance to be solved, black-box optimization algorithms do not have access to the problem other than by sampling and evaluating search points. Since this observation is relevant also in many other settings in theoretical computer science (e.g., in the learning and machine learning communities), alternative complexity notions, so-called black-box (or query) complexity models, have been developed. In the context of studying randomized search heuristics, however, these models typically fall short in providing reasonable bounds, as they typically allow for highly problem-tailored algorithms. Search heuristics, in contrast, often have very limited memory, use only relative and not absolute function values to guide the search, do not use gradient information etc. For this reason, new complexity notions need to be developed in this context.

We give a very brief sketch of the existing black-box complexity models in the search heuristics context and discuss how we can use insights from black-box complexity theory to design new (and hopefully more powerful) search heuristics.

In this summary, we focus on discrete optimization problems, but all models are meaningful also for continuous optimization problems.

2 The Different Black-Box Models—An Overview

We present the different black-box notions. To illustrate the differences of the models, we shall always consider the following simple problem, which we call the *Hamming distance problem*. For an unknown bitstring z of length n, each function evaluation returns n minus the Hamming distance $H(x,z) := |\{i \mid x_i \neq z_i\}|$ of the queried search point x and the unknown string z. That is, $f_z(x)$ is the number of positions in which x and z agree. We are interested in maximizing f_z, i.e., we want to identify z.

The *unrestricted black-box complexity* [5] of the Hamming distance problem is simply the smallest number of samples x that have to be evaluated in order to identify z. It is known that this number is of order $n/\log n$ [6]. Typical randomized search heuristics (RSH), however, need $\Omega(n \log n)$ function evaluations to maximize f_z. This and many similar examples therefore seem to suggest that the unrestricted black-box complexity does not reveal the full truth about the difficulty of a problem when optimizing it with randomized search heuristics.

One idea to overcome this discrepancy was already suggested in [5]: instead of allowing the algorithm to store all previously queried search points, allow it to store only a certain number of previous search point and its corresponding function value. This is the *memory-restricted model*. Unfortunately, it turns out that this does not change the complexity of the Hamming distance problem. Even in the most restrictive model, in which an algorithm can store only one previous solution, $O(n/\log n)$ query algorithms are possible [3].

An alternative idea to restrict the algorithms is to request that they should only use relative function values in order to decide on the distribution from which they sample the next solution candidate. That is, the algorithm may not make use of the absolute function values $f_z(x)$ but can merely use the fact that $f_z(x_1) \geq f_z(x_2) \geq \ldots \geq f_z(x_t)$ for the previously sampled search points $x_1, \ldots, x_t$. Interestingly, while this restriction increases the complexity of many problems, the complexity of the Hamming distance problem remains unaffected, i.e., the *ranking-based black-box complexity* is again of order $n/\log n$.

If we combine the memory-restricted model with the ranking-based model, then the complexity of the Hamming distance problem increases to linear in n (if only a constant number of previous samples and their relative function values can be stored). This is the *comparison-based model*. Still it does not explain why most search heuristics need $\Omega(n \log n)$ function evaluations for maximizing f_z.

A breakthrough in the area of black-box complexity research was the so-called *unbiased black-box complexity model*. In essence, it requires that all distributions from which new solution candidates are sampled are symmetric with respect to the bit positions $1, \dots, n$ and the bit entries $0, 1$. In addition, the unbiased model allows to talk about the *arity* of the algorithms. If only one previously sampled search point determines the sampling distribution (the *unary* model), the best algorithm has complexity $\Omega(n \log n)$ on the Hamming distance problem [7]. That is, this complexity matches the runtime of the commonly studied search heuristics. However, if two or more solution candidates can determine the sampling distribution, this complexity drops to at most linear in n [2],[4]. Since many search heuristics do sample from more than one search point, it is therefore an interesting question of whether this complexity can be achieved by RSH. A partial answer to this question is the result described in the next section.

3 Learning from Black-Box Complexity Theory

An inspection of the algorithms yielding the upper bounds mentioned in Section 2 shows that all of them greatly profit from sampling search points that are inferior to the current best solution. This is not what typical search heuristics do—they typically discard inferior solutions immediately. In [1], a new genetic algorithm is presented that experiments with a simple way of exploiting inferior solutions.

The key idea of the algorithm in [1] is to sample from the current best solution x a set of λ new solution candidates that all have the same distance to x. The best one of those (which, if x is already close to an optimal solution, often have smaller function value than x) is used for a recombination with the original *parent* x. More precisely, denoting by y this best of the λ *offspring* solutions, y is recombined with x λ times (using some standard recombination procedures). With reasonably large probability the best one y' of these newly created recombinations has better function value than x. If indeed $f(y') \geq f(x)$, the algorithm proceeds with x being replaced by y', and it does not update x otherwise. This algorithm solves the Hamming distance problem using only $O(n\sqrt{\log n})$ function evaluations. (The tight bound, in fact, is $\Theta(n\sqrt{\log(n) \log\log\log(n) / \log\log(n)})$.) While this is not optimal ($O(n)$ algorithms exist), it is the first example of a non-problem tailored algorithm exhibiting a runtime which is smaller than the classic $\Theta(n \log n)$ bound.

4 Future Work

Black-box complexity models offer a plethora of interesting and challenging open problems. Two examples that we are planning to work on in the PGMO project are

1. Noisy settings: While all the results mentioned above hold in noiseless settings, it would be very interesting to see how the stated complexity bounds change if the algorithms or the function evaluations are subject to noise.
2. Elitist algorithms: typical search heuristics keep only the best solution candidates evaluated so far and discard inferior solutions. To understand the influence of this *elitist selection behavior*, this greedy selection should be reflected in a meaningful black-box complexity model.

References

1. Benjamin Doerr, Carola Doerr, and Franziska Ebel. Lessons from the black-box: Fast crossover-based genetic algorithms. In *Proc. of Genetic and Evolutionary Computation Conference (GECCO'13)*, pages 781–788. ACM, 2013.
2. Benjamin Doerr, Daniel Johannsen, Timo Kötzing, Per Kristian Lehre, Markus Wagner, and Carola Winzen. Faster black-box algorithms through higher arity operators. In *Proc. of Foundations of Genetic Algorithms (FOGA'11)*, pages 163–172. ACM, 2011.
3. Benjamin Doerr and Carola Winzen. Playing Mastermind with constant-size memory. In *Proc. of Symposium on Theoretical Aspects of Computer Science (STACS'12)*, pages 441–452. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.
4. Benjamin Doerr and Carola Winzen. Reducing the arity in unbiased black-box complexity. In *Proc. of Genetic and Evolutionary Computation Conference (GECCO'12)*, pages 1309–1316. ACM, 2012.
5. Stefan Droste, Thomas Jansen, and Ingo Wegener. Upper and lower bounds for randomized search heuristics in black-box optimization. *Theory of Computing Systems*, 39:525–544, 2006.
6. Paul Erdős and Alfréd Rényi. On two problems of information theory. *Magyar Tudományos Akadémia Matematikai Kutató Intézet Közleményei*, 8:229–243, 1963.
7. Per Kristian Lehre and Carsten Witt. Black-box search by unbiased variation. *Algorithmica*, 64:623–642, 2012.

PGMO - COPI'14

A Bayesian subset simulation approach to constrained global optimization of expensive-to-evaluate black-box functions

Paul FELIOT · Julien BECT · Emmanuel VAZQUEZ

Abstract This paper addresses the problem of derivative-free global optimization of a real-valued function under multiple inequality constraints. Both the objective function and the constraint functions are assumed to be smooth, nonlinear, expensive-to-evaluate black-box functions. As a consequence, the number of evaluations that can be used to carry out the optimization is very limited. We focus in this work on the case of strongly constrained problems, where finding a feasible design, using such a limited budget of simulations, is a challenge in itself. The method that we propose to overcome this difficulty has its roots in the recent literature on Gaussian process-based methods for reliability analysis—in particular, the Bayesian Subset Simulation (BSS) algorithm of Li, Bect and Vazquez—and multi-objective optimization. More specifically, we consider a decreasing sequence of nested subsets of the design space, which is defined and explored sequentially using a combination of Sequential Monte Carlo (SMC) techniques and sequential Bayesian design of experiments. The proposed method obtains promising result in a benchmark against state-of-the-art methods on challenging test cases from the literature.

Keywords Optimization · Kriging · Gaussian Process · Subset Simulation · Sequential Monte Carlo

Mathematics Subject Classification (2000) 90C56 · 90C59 · 62K99 · 62L05 · 62P30

1 Context

This paper addresses the problem of derivative-free global optimization of a real-valued function under multiple inequality constraints:

$$\begin{cases} \text{Minimize} & f(x), \\ \text{Subject to} & x \in \mathbb{X} \quad \text{and} \quad c(x) \leq 0, \end{cases}$$

where f is the function to be minimized, $\mathbb{X} \subset \mathbb{R}^d$ is the design space and $c = (c_1, \ldots, c_q)$ is the vector of constraint functions, $c_j : \mathbb{X} \to \mathbb{R}$, $1 \leq j \leq q$.

Both the objective function f and the constraint functions c_j are assumed to be smooth, nonlinear, expensive-to-evaluate black-box functions. More specifically, it is assumed that the values of $f(x)$

Paul Feliot (corresponding author), Julien Bect, Emmanuel Vazquez
Institut de Recherche Technologique SystemX & Supélec
E-mail: firstname.lastname@irt-systemx.fr ou firstname.lastname@supelec.fr
IRT SystemX: 8, avenue de la Vauve, 91120 Palaiseau, France.
Supélec: 3 rue Joliot-Curie, 91192 Gif-sur-Yvette cedex, France.

and $c(x)$, for a given $x \in \mathbb{X}$, are provided simultaneously by a single call to some time-consuming computer program—a setup that typically applies to industrial design problems, where numerical simulations are used to mimic the actual physical behavior of the system to be designed. Such simulations may for instance require fluid dynamic, heat transfer or mechanical deformation computations and can take from a few hours to several days to compute.

The number of runs that can be afforded to carry out the optimization is therefore very limited. We focus in this work on the case of strongly constrained problems, where finding a feasible design, using such a limited budget of simulations, is a challenge in itself.

2 Proposed method

Global optimization methods have been investigated intensively for the last decades. When expensive-to-evaluate functions are involved, cheap-to-evaluate approximations of the objective and constraints functions—often referred to as surrogate models or meta-models—are classically relied upon. We adopt here a Bayesian approach, which provides not-only natural surrogate models for the expensive-to-evaluate functions that we have to deal with, but also an elegant framework to help design efficient optimization algorithms. More precisely, the objective function f and the constraint functions c_j are modeled as (independent) Gaussian processes, following a now classical approach that has been made popular by Jones et al. (1998) for unconstrained optimization problems. Subsequent developments for constrained optimization problems have been proposed by, among others, Schonlau et al. (1998); Sasena et al. (2002); Gramacy and Lee (2011); Parr et al. (2012); Picheny (2014a).

We focus on this paper on the case of strongly constrained problems, where the volume of the feasible space is small compared to the size of the design space. As a consequence, locating even a single feasible point becomes difficult and most existing Bayesian optimization methods, which require at least one feasible point to begin with, fail to be applicable. The method that we propose to overcome this difficulty has its roots in the literature on Bayesian sequential design of experiments for reliability analysis (see, e.g., Bect et al., 2012; Li et al., 2012)—estimating a failure region or a feasible set are very similar problems—and multi-objective optimization (see, e.g., Emmerich et al., 2006; Wagner et al., 2010; Picheny, 2014b). More specifically, we consider a decreasing sequence of nested subsets of the design space, which is defined and explored sequentially using a combination of Sequential Monte Carlo (SMC) techniques and sequential Bayesian design of experiments, in the spirit of Li et al. (2012); Benassi et al. (2012); Li (2012); Benassi (2013).

3 Results and future work

We are able to report good results on challenging test cases from the literature compared with state-of-the-art methods. Future work will include the extension of our method to multi-objective problems and the optimization of various aspects of our algorithm (Sequential Monte Carlo algorithm, sampling criterion...). Simulation failures should also be taken into account as they are inherent to complex industrial simulation codes. The performance of the method will be evaluated on a real-life industrial problem provided by Safran, with the contribution of Cenaero (optimization of the performances of a turbo-machine fan blade under aerodynamic, mechanic and acoustic constraints).

Acknowledgements This research work has been carried out in the frame of the Technological Research Institute SystemX, and therefore granted with public funds within the scope of the French Program “Investissements d’Avenir”.

References

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

R. Benassi. *Nouvel algorithme d'optimisation bayésien utilisant une approche Monte-Carlo séquentielle.* PhD thesis, Supélec, 2013.

R. Benassi, J. Bect, and E. Vazquez. Bayesian optimization using sequential Monte Carlo. In *Learning and Intelligent Optimization. 6th International Conference, LION 6, Paris, France, January 16-20, 2012, Revised Selected Papers*, volume 7219 of *Lecture Notes in Computer Science*, pages 339–342. Springer, 2012.

M. T. M. Emmerich, K. C. Giannakoglou, and B. Naujoks. Single- and multi-objective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006.

R. L. Gramacy and H. Lee. Optimization under unknown constraints. In *Bayesian Statistics 9. Proceedings of the Ninth Valencia International Meeting*, pages 229–256. Oxford University Press, 2011.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

L. Li. *Sequential Design of Experiments to Estimate a Probability of Failure.* PhD thesis, Supélec, 2012.

L. Li, J. Bect, and E. Vazquez. Bayesian Subset Simulation: a kriging-based subset simulation algorithm for the estimation of small probabilities of failure. In *Proceedings of PSAM 11 & ESREL 2012, 25-29 June 2012, Helsinki, Finland*. IAPSAM, 2012.

J. M. Parr, A. J. Keane, Alexander I. J. Forrester, and C. M. E. Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization*, 44(10):1147–1166, 2012.

V. Picheny. A stepwise uncertainty reduction approach to constrained global optimization. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS), 2014, Reykjavik, Iceland.*, volume 33, pages 787–795. JMLR: W&CP, 2014a.

V. Picheny. Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction. arXiv:1310.0732 (to appear in *Statistics and Computing*), 2014b.

M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering optimization*, 34(3):263–278, 2002.

M. Schonlau, W. J. Welch, and D. R. Jones. Global versus local search in constrained optimization of computer models. In *New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Conference*, volume 34 of *IMS Lecture Notes-Monographs Series*, pages 11–25. Institute of Mathematical Statistics, 1998.

T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser. On expected-improvement criteria for model-based multi-objective optimization. In *Parallel Problem Solving from Nature, PPSN XI. 11th International Conference, Krakov, Poland, September 11-15, 2010, Proceedings, Part I*, volume 6238 of *Lecture Notes in Computer Science*, pages 718–727. Springer, 2010.

PGMO - COPI'14

Linear decision rules applied to hydropower scheduling

Stein-Erik Fleten · Ida Grønvik · Ajla Hadziomerovic · Nina Ingvoldstad

Abstract Hydropower scheduling is cast as a multistage stochastic linear program, and we discuss the feasibility of applying linear decision rules for solving this problem. This approach restricts the control variables to be affine functions of current and past observations of the uncertain parameters, and the original problem is transformed into a tractable one with short computing time. This overcomes obstacles in traditional approaches which suffer from a computing effort that grows exponentially with the number of stages and the number of state variables. In our case, a price-taking producer wants to determine a reservoir management strategy over one to two years, aiming to maximize expected profit subject to uncertainty in reservoir inflows and electricity spot market prices. The approach is demonstrated on four Norwegian hydropower plants. We find that the approach is effective at reducing computational complexity, and is well-suited to multistage hydro scheduling.

Keywords Robust optimization · affinely adjustable robust counterpart

Mathematics Subject Classification (2000) MSC 47N10 · MSC 65K05 · MSC 90C15

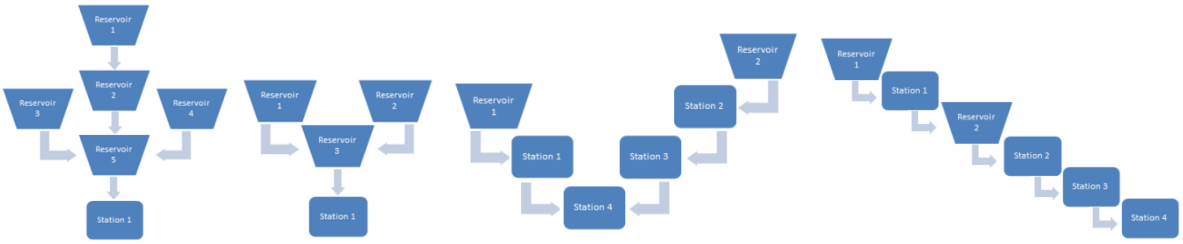
1 Introduction

The challenge of optimally releasing water for power production from hydroelectric reservoirs is a well-known problem both in literature and practice. We focus on scheduling of releases in view of maximizing expected electricity market revenues, under uncertainty of inflows and spot prices. The objective is to investigate the feasibility of using linear decision rules, which is an approximation method for solving stochastic programming problems.

The hydroelectric scheduling problem is often divided in three tasks with increasing degree of detail and different scheduling horizon; long-term, seasonal and short-term scheduling [3,13,1]. Here we consider long-term scheduling, whose aim is optimal allocation of the water resources with a time horizon of one to five years. Long-term scheduling is a form of strategic reservoir management, and gives boundary conditions to the seasonal or short-term tasks [11].

Stein-Erik Fleten
Norwegian University of Science and Technology
Tel.: +47-73591296
E-mail: stein-erik.fleten@iot.ntnu.no

Ida Grønvik, Ajla Hadziomerovic, Nina Ingvoldstad
Norwegian University of Science and Technology
E-mail: {idagronvik,ajla.hadz,nina.ingvoldstad}@gmail.com



(a) Plant 1 (b) Plant 2 (c) Plant 3 (d) Plant 4

Fig. 1: Topology of power plants in this study.

We specify stochastic sub-models for electricity spot prices and river inflows in order to provide input for the long-term scheduling. Parameters for these dynamics are set based on historical spot price time series, forward and futures prices and historical inflow data provided by four Norwegian hydropower producers, cf. Fig. 1.

The classical approach to the reservoir operation problem under uncertainty is stochastic dynamic programming [9]. However, the size of problems grows exponentially with the number of state variables, i.e. the number of reservoirs, inflow trend/level states and price trend/level states. A common step in approaches for solving stochastic hydropower scheduling problems is to approximate the underlying stochastic process of the uncertain parameters to be discrete, with a finite set of values for the random variables to take on. A prevalent approach is Benders decomposition, in particular, stochastic dual dynamic programming [7]. A comparison between stochastic dynamic programming and stochastic dual dynamic programming is in [8]. According to [5], the ability to generate good scenario trees is highly dependent on the knowledge of the underlying stochastic process. As the probability distribution of the uncertain parameters is rarely known, all of the above approaches rely on quite brave simplifying assumptions in order to achieve tractability. The size of scenario-tree-based problems grows exponentially with the number of time stages.

Stochastic programs must be simplified in order to gain computational tractability [6]. The linear decision rules (LDR) approximation involves restricting the recourse decisions associated with the stochastic program to be functions of the realisations of the uncertain parameters. Following [2], the uncertain parameters are assumed to be within an interval around their expected values, in which the size of the interval is given by the level of uncertainty. By applying the LDR approach, the original stochastic program is converted into a semidefinite program. In the case of fixed recourse, provided that the uncertainty set itself is computationally tractable, this derived program is computationally tractable as well. In the case of a polyhedraluncertainty set, the fixed recourse program is equivalent to a linear program. Although the LDR approach is effective at reducing computational complexity, it may incur a considerable loss of optimality. By applying the approximation to both the primal version and the dual version of the original problem, this loss of optimality can be estimated by the gap between the optimal objective values of the two programs [6].

The contribution of this work is the development of a multistage stochastic hydropower scheduling model based on the LDR approximation. Currently, there are few studies in this ([12,4,10]). To evaluate the applicability of the approximation to the long-term scheduling problem, the LDR will be demonstrated on four Norwegian hydropower plants. Focus will also be on how uncertainty in price and inflow affects the problem and how the flexibility of the hydropower plants relates to the reservoir management strategies given by the optimization.

2 Conclusions

We develop a multistage stochastic model for hydropower scheduling, in the framework of an LDR approximation. The random variables are restricted to vary within an interval around their expected values, in which the sizes of the intervals are given parametrically. Using data from four different hydroelectric facilities in Norway, we demonstrate that increasing the interval sizes for inflows causes more restrictive reservoir management strategies. We also show that more restrictive strategies are rational for plants with less flexibility compared to plants with more flexibility. The problem is feasible only up to a certain inflow interval size, depending on the flexibility of the hydropower plants and on the length of the scheduling horizon.

We compare the performance of the LDR approximation to the corresponding deterministic problem, using a ten year simulation for one of the hydropower plants. This is a back-test using the realisations of price and inflow in the years 2002 to 2012, and both models are rerun every fourth week with updated initial reservoir levels. At an interval size of 10 % the LDR model is shown to give an average price per released MWh that is 11.5 NOK higher than the price given by the deterministic model. The main reason for this result is the ability of the LDR model to adjust production to the realisations of price and inflow.

The complexity reduction in the the LDR approximation leads to acceptable computing effort. When solved for 104 time stages, the LDR model has a run time of 268 seconds. This implies that the model can be rerun whenever new information about price and inflow is available. As a result, the model is able to provide updated boundary conditions to a shorter term model more frequently than models with longer computation times. In conclusion, we consider the LDR approximation a promising alternative for multistage stochastic hydropower scheduling problems.

Acknowledgements The authors would like to thank the electricity producers for graciously providing the data. We recognize the Norwegian research centre FME CenSES, Centre for Sustainable Energy Studies (RCN grant 209697), and Fleten acknowledges financial support from the Research Council of Norway through project 228811.

References

1. Ellen K. Aasgard, Gørild S. Andersen, Stein-Erik Fleten, and Daniel Haugstvedt. Evaluating a stochastic-programming-based bidding model for a multireservoir system. *IEEE Trans. Power Systems*, 29(4):1748–1757, 2014.
2. Aharon Ben-Tal, Alexander Goryashko, Elana Guslitzer, and Arkadi Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
3. O. B. Fosso, A. Gjelsvik, A. Haugstad, B. Mo, and I. Wangensteen. Generation scheduling in a deregulated system. The Norwegian case. *IEEE Trans. Power Systems*, 14(1):75–81, 1999.
4. Vincent Guigues. Robust production management. *Optimization and Engineering*, 10(4):505–532, 2009.
5. M. Kaut and S. W. Wallace. Evaluation of scenario generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271, 2007.
6. Daniel Kuhn, Wolfram Wiesemann, and Angelos Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.
7. M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52(2):359–375, 1991.
8. Andy Philpott, Anes Dallagi, and Emmanuel Gallet. On cutting plane algorithms and dynamic programming for hydroelectricity generation. In R. Kovacevic, G. Pflug, and M. T. Vespucci, editors, *Handbook of Risk Management in Energy Production and Trading*, volume 199 of *International Series in Operations Research & Management Science*, pages 105–127. Springer, 2013.
9. G. Pritchard, A.B. Philpott, and P.J. Neame. Hydroelectric reservoir optimization in a pool market. *Mathematical Programming*, 103(3):445–461, 2005.
10. Paula Rocha. *Medium-Term Planning in Deregulated Energy Markets with Decision Rules*. PhD thesis, Imperial College London, 2012.
11. S. Stage and Y. Larsson. Incremental cost of water power. *AIEE Trans. Power Apparatus and Systems*, 80:361–365, 1961.
12. Julien Thénié. *Décisions séquentielles dans lincertain: nouvelles approches par la programmation stochastique et applications*. PhD thesis, Université de Genevé, 2008.
13. S. W. Wallace and S.-E. Fleten. Stochastic programming models in energy. In A. Ruszczynski and A. Shapiro, editors, *Stochastic programming*, pages 637–677. Vol. 10 of *Handbooks in Operations Research and Management Science*. Elsevier Science, 2003.

PGMO - COPI'14

Handling Inexactness in Nondifferentiable Optimization: a Computational Comparison

Rosa Maria Videira de Figueiredo · Antonio Frangioni · Claudia Sagastizábal

Abstract Lagrangian relaxation is a powerful technique for deriving strong lower bounds for hard combinatorial problems. However, its use typically brings to difficult decisions: on one hand, the Lagrangian subproblems need be "hard" for the bound to be good, but on the other hand they have to be solved many times, and therefore one has to do this quickly. It is therefore helpful in many applications to be able to solve the subproblems only approximately, at least in some (hopefully, most) iterations of the approach; this is indeed possible, and the convergence theory of inexact approaches, in particular of the well-known Bundle type, has been recently perfected. The convergence proofs leave ample scope for different implementation; in particular, there are different ways in which inexact data can be used in Bundle algorithms, and the parameter that dictates the degree of accuracy in each solution of the Lagrangian problem can be handled in several different ways. Our aim is to explore a large set of these variants on a significant set of hard optimization problems with different characteristics, in order to develop guidelines for choosing the best strategy (the one providing the best solution in the shortest possible running time) in accordance with the features of the problem at hand.

Keywords Langrangian Relaxation · Inexact Bundle Methods · Computational Experience

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Introduction

Most optimization problems arising in applications, such as logistics, telecommunications, and energy, are very hard to solve; this is typically the combination of different factors, among which the sheer scale of the problem (number of variables and constraints) and the presence of nonconvex features, most often under the guise of integer variables. While some problems have both these features, often one of them alone is enough to make the problem intractable computationally; for instance, two-stage stochastic linear programs would theoretically be "easy" ones, but their size explodes with the number of scenarios, that has to be kept large for the results to be statistically significant. On the other hand, relatively small-scale integer programs can nonetheless be extremely hard to solve.

Rosa Maria Videira de Figueiredo
Institute of Mathematics and Statistics, Rio de Janeiro State University, 20550-900 RJ, Brazil, E-mail: rosamaria.figueiredo@gmail.com

Antonio Frangioni
Department of Computer Science, University of Pisa, 56125 Pisa, Italy, E-mail: frangio@di.unipi.it

Claudia Sagastizábal
IMPA visiting researcher, Brazil, E-mail: sagastiz@impa.br

Linear programming based techniques have made tremendous improvements in the last 10 or 20 years alone; this, together with advances in strengthening the continuous relaxation of the formulations, e.g. by valid inequalities, makes it possible nowadays to routinely solve problems that would have been intractable only a few years ago. However, there still comes a threshold where problems can only be tackled by the use of specialized techniques that exploit valuable structure information. This is, for instance, very true for many problems arising in energy optimization, see e.g. [10][11, §3.3].

Lagrangian relaxation (or its "primal cousin", Benders' decomposition) is widely used approaches because it only relies on very general structural properties for the problem at hand, namely the presence of "complicating constraints" (variables) whose removal (fixing) leaves a much easier problem to solve. This is often in large part due to the fact that the problem decomposes into a (large) set of independent (much) smaller problems, that may then either be below the size threshold where general-purpose techniques work well, or allow specialized approaches (see e.g. [3] for one of the very many cases). However, the standard theory of Lagrangian relaxation for combinatorial optimization problems (e.g. [2]) reveals an unpleasant truth: a "no free lunch" principle applies, where in order for the bound to improve, e.g. w.r.t. that of the ordinary continuous relaxation, the subproblems need necessarily be "difficult". The algorithm designer hence has to confront a difficult decision in order to rightly balance the difficulty of the subproblems, and therefore the quality of the obtained solution, with the total running time devoted to their solution.

One helpful idea in many applications is to solve the subproblems *only approximately*. For Lagrangian relaxation of combinatorial problems this "just" amounts at early terminating the solution process of the Lagrangian subproblem, most often of a Branch&Something type; this provides the best solution found so far, which yields a linearization of the Lagrangian function, as well as an estimate of the gap with the optimal one. For stochastic programs, instead, one possibility is to artificially restrict the subproblem to a small subset of the original large set of scenarios [9]. Both cases result in an approximate computation of the objective function, that has been shown to be possible, in particular for solution methods belonging to the class of Bundle approaches [1,4,5]; lately, convergence theory has been perfected [7,8].

The recent version of the convergence analysis shows that there is ample scope for different implementations; so far, however, little is known about the practical impact of these different choice. Aim of this work is to explore a large set of this variant on a significant set of hard optimization problems with different characteristics, in order to develop guidelines for choosing the best strategy (the one providing the best solution in the shortest possible running time) in accordance with the features of the problem at hand.

2 Testing Inexact Bundle Methods

In terms of the general convergence theory [8], the inexact oracles we will consider, coming from either Lagrangian relaxation of integer programs or from two-stage stochastic programs, cover only some of the possibilities; in particular, they are all both *lower* (in the sense that the linearizations provide reliable underestimates of the objective function) and *controllable* (in the sense that in principle the accuracy can be brought to zero, albeit possibly at a very large cost). Bundle methods are known to work even under less stringent assumptions. Furthermore, as already discussed Lagrangian oracles may produce, as a natural by-product of the solution process (i.e., at no extra cost), *explicit estimates of the obtained accuracy*. However, even with these properties, there are several different ways in which inexact data can be used in Bundle algorithms.

For a start, there are two different approaches about the handling of inexact information in Bundle methods; in particular, one may or may not use the information about the attained accuracy in the solution of the Lagrangian problem, when this is available. Furthermore, the degree of accuracy in each solution of the Lagrangian problem can be handled in several different ways, depending e.g. on:

- whether the accuracy is controlled by the Bundle algorithm, or by the oracle;

- whether or not the parameter that dictates the accuracy is set according to the current estimate produced by the Bundle algorithm of the distance between the current solution and the optimal one;
- whether the accuracy parameter is continuously reduced, or this is only done when the algorithm detects that it would otherwise "stall";
- whether the accuracy parameter is reduced according to the number of iterations of the algorithm, or only of the "successful" ones (the so-called "serious steps");
- whether the accuracy parameter is monotonically non increasing or it is "reset" after successful iterates as proposed in [6].

All these issues produce algorithmic parameters, whose tuning will provide insight on what the best strategies are. In order to reach conclusions with a good degree of generality, our tests will be performed on a variety of different problems, such as:

- capacitated lot-sizing problems;
- cutting stock problems;
- multicommodity network design problem;
- unit commitment problem;
- two-stage stochastic linear problems;
- vehicle routing problems.

Each of these problems has a specific combination of the main relevant factors, such as the number of Lagrangian multipliers and whether or not they are constrained in sign, the number of subproblems arising from the decomposition and their "difficulty". By testing on such a diverse set we will be able to start drawing guidelines for choosing the best strategy (the one providing the best solution in the shortest possible running time) in accordance with the features of the problem at hand. These will hopefully show that inexact Bundle methods, when the accuracy of the subproblem solution is properly dynamically handled, can be a competitive option for the solution of large-scale and/or difficult optimization problems.

Acknowledgements We gratefully acknowledge the financial support of the Gaspard Monge program for Optimization and Operations Research under the project "Consistent Dual Signals and Optimal Primal Solutions".

References

1. C. Fábián (2000) "Bundle-type methods for inexact data" *Central European Journal of Operations Research* 8, 35–55
2. A. Frangioni "About Lagrangian Methods in Integer Optimization" *Annals of Operations Research* 139, 163–193, 2005
3. A. Frangioni, C. Gentile (2006) "Solving Nonlinear Single-Unit Commitment Problems with Ramping Constraints" *Operations Research* 54(4), 767–775
4. M. HintermuÏller (2001) " A proximal Bundle method based on approximate subgradients" *Computational Optimization and Applications* 20, 245–266
5. K.C. Kiwiel (2006) "A proximal Bundle method with approximate subgradient linearizations" *SIAM Journal on Optimization* 16(4), 1007–1023
6. J. Malick, W. de Oliveira, S. Zaourar (2013) "Nonsmooth Optimization Using Uncontrolled Inexact Information" available at http://www.optimization-online.org/DB_HTML/2013/05/3892.html
7. W. de Oliveira, C. Sagastizábal (2012) "Level Bundle methods for oracles with on-demand accuracy" available at http://www.optimization-online.org/DB_HTML/2012/03/3390.html
8. W. de Oliveira, C. Sagastizábal and C. Lemaréchal (2013) "Bundle methods in depth: a unified analysis for inexact oracles" available at http://www.optimization-online.org/DB_HTML/2013/02/3792.html
9. W. de Oliveira, C. Sagastizábal and S. Scheimberg (2011) "Inexact Bundle Methods For Two-Stage Stochastic Programming" *SIAM Journal on Optimization* 21(2), 517–544
10. C. Sagastizábal (2012) "Divide to conquer: decomposition methods for energy optimization" *Mathematical Programming*, 187–222
11. M. Tahanan, W. van Ackooij, A. Frangioni and F. Lacalandra "Large-scale Unit Commitment under uncertainty: a literature survey" *Technical Report* 14-01, Dipartimento di Informatica, Universit di Pisa, 2014

PGMO - COPI'14

Implementation of an augmented Lagrangian algorithm for solving convex quadratic optimization problems, which can deal with infeasibility

A. Chiche · J. Ch. Gilbert · É. Joannopoulos

Abstract When a solver of convex quadratic optimization problem (QP) is used within the SQP algorithm (a Newton-like method to solve nonlinear optimization problems), it is important that it deals appropriately with the special QPs that can be generated by the SQP algorithm, those that are *infeasible* or *unbounded*. The goal of this talk is twofold. First, it describes how the augmented Lagrangian (AL) algorithm can deal with these special convex QPs. Second, it presents the numerical results obtained with `C++`/`Matlab` implementations of that algorithm in codes called `Oqla`/`Qpalm`.

Keywords Augmented Lagrangian algorithm · augmentation parameter update · closest feasible problem · convex quadratic optimization · feasible shift · global linear convergence · infeasible problem · proximal point algorithm · quasi-global error bound · shifted constraint.

Mathematics Subject Classification (2000) 49M27 · 49M29 · 65K05 · 90C05 · 90C06 · 90C20 · 90C25

1 Introduction

To simplify the presentation, we consider a convex quadratic optimization problem written as follows

$$\inf_x \{q(x) : l \leqslant Ax \leqslant u\}. \tag{1}$$

In that problem, the objective function

$$q : x \in \mathbb{R}^n \mapsto q(x) = g^\mathsf{T} x + \frac{1}{2} x^\mathsf{T} H x$$

is convex quadratic ($g \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n\times n}$ is positive semidefinite), the constraints are defined by a matrix $A \in \mathbb{R}^{m\times n}$ and bounds l and $u \in \overline{\mathbb{R}}^m$ that must satisfy $l < u$, and the sign "$^\mathsf{T}$" denotes transposition. We introduce the interval $[l, u] := \{y \in \mathbb{R}^m : l \leqslant y \leqslant u\}$. Since H may vanish, the problem encompasses linear optimization.

A. Chiche
Artelys France, 12 rue du Quatre Septembre, 75002 Paris (France)
E-mail: Alice.Chiche@artelys.com

J. Ch. Gilbert
INRIA Paris-Rocquencourt, Pomdapi team, BP 105, F-78153 Le Chesnay Cedex (France)
E-mail: Jean-Charles.Gilbert@inria.fr

É. Joannopoulos
INRIA Paris-Rocquencourt, Pomdapi team, BP 105, F-78153 Le Chesnay Cedex (France)
E-mail: Emilie.Joannopoulos@inria.fr

Solving a problem like (1) may look quite easy, in particular due to its assumed convexity; one may believe that the problem has been fully explored in the 20th century. However, if the standard QP solver `Quadprog` of `Matlab` (version `R2012a (7.14.0.739) 64-bit (maci64)`) is run on the bound constraint convex QP with the data

$$g = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{pmatrix}, \qquad \text{the single constraints} \qquad x \geqslant \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix},$$

and the option `'LargeScale'` set to `'off'` (this one triggers an active-set method; the default is `'on'` and launches the reflective trust-region method of [3]), it returns a message containing the lines

```
Exiting: the solution is unbounded and at infinity;
Function value:  3.20000e+33
```

while the problem is uniquely solved by $x = (-1, -1, 2)$ with the optimal value -1.5. This is a benign flaw, probably due to the singularity of the positive semidefinite Hessian and rounding errors, since if one adds εI ($\varepsilon =$ `eps` is the machine epsilon) to the Hessian, `Quadprog` with the same options finds a solution very close to the one of the unperturbed problem. Nevertheless, like many other QP solvers, in case of an infeasible or unbounded problem, `Quadprog` does not give other information than an infeasibility or unboundedness flag, which is unfortunate since there is much more to say in these circumstances.

On the other hand, the more recent interior point solvers `Ooqp` [9] and `Qpb` fail to solve more than 50 % and 20 %, respectively, of the 135 convex QPs in the `CUTEst` [12] collection in a reasonable time (and solve the other ones very quickly), which shows that these problems are not that trivial.

The goal of this work, based on [4,2] and further detailed in [10], is to explore the numerical abilities of an AL algorithm to solve the convex QP (1). This approach is rarely used ([7] is an exception), despite some of its nice features, which should guarantee it with the application niche that is described below.

2 The AL algorithm for a convex QP

The AL algorithm is defined by first introducing an auxiliary vector of variables $y \in \mathbb{R}^m$ and by rewriting (1) as follows

$$\inf_{(x,y)} \{q(x) : Ax = y,\ l \leqslant y \leqslant u\}. \tag{2}$$

Given an *augmentation parameter* $r \geqslant 0$, the *AL function* $\ell_r : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is then defined at $(x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by

$$\ell_r(x, y, \lambda) = q(x) + \lambda^T(Ax - y) + \frac{r}{2}\|Ax - y\|^2,$$

where $\|\cdot\|$ denotes the ℓ_2-norm. For $r = 0$, one recovers the usual Lagrangian function, relaxing the equality constraints of (2) thanks to the *multiplier* or *dual variable* λ. The *AL algorithm* generates a sequence of dual variables $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^m$, which are aimed at converging to a dual solution to (1), as follows. Knowing $r_k > 0$ and $\lambda_k \in \mathbb{R}^m$, the next dual iterate λ_{k+1} is computed by

$$\begin{aligned} (x_{k+1}, y_{k+1}) &\in \arg\min \{\ell_{r_k}(x, y, \lambda_k) : (x, y) \in \mathbb{R}^n \times [l, u]\}, \\ \lambda_{k+1} &:= \lambda_k + r_k(Ax_{k+1} - y_{k+1}), \end{aligned} \tag{3}$$

where "arg min" denotes the set of minimizers of the problem on which it applies. Next, r_k is updated by a rule that depends on the implementation and to which we pay much attention in [4,2,10]. The QP in (3) is called the *AL subproblem*.

The salient features of the AL algorithm just described, which specify the contour of its application niche, are

- it does not require any matrix factorization, so that it can be used for large scale problems,

- it uses an active set method on the bound constrained AL subproblems (3) that it generates, so that it is likely to be more efficient when an estimate of the active constraints at the solution is known, like for the QPs generated by the SQP algorithm [1],
- it can provide precious information when the considered problem is infeasible or unbounded; this information can be useful when the QP solver is viewed as a tool in nonlinear optimization.

3 Global linear convergence results

This work on the behavior of the AL algorithm on an infeasible QP can be viewed as a continuation of the one initiated in [4], in which the *global* linear convergence of the constraint norm to zero is established, when (1) has a solution. More specifically, it is shown in [4] that in this case:

$$\begin{array}{c} \forall\, \beta > 0, \quad \exists\, L > 0, \quad \mathrm{dist}(\lambda_0, \mathcal{S}_{\mathrm{D}}) \leqslant \beta \quad \text{implies that} \\ \forall\, k \geqslant 1, \quad \|Ax_{k+1} - y_{k+1}\| \leqslant \frac{L}{r_k} \|Ax_k - y_k\|, \end{array} \tag{4}$$

where $\mathcal{S}_{\mathrm{D}}$ denotes the necessarily nonempty set of optimal multipliers associated with the equality constraints of (2) and the operator "dist" denotes the Euclidean distance. Computationally, this result is interesting because it allows the AL algorithm to tune the augmentation parameter r_k on the basis of the observed behavior of the constraint norm ratio $\|Ax_{k+1} - y_{k+1}\| / \|Ax_k - y_k\|$, from the very first iteration. For example, r_k can be increased when this ratio is larger than a desired rate of convergence (this rate is easier to prescribe by the solver user than r_k). Now, when the problem is *infeasible*, the constraint norm cannot, of course, tend to zero and the just described rule for tuning r_k makes the augmentation parameters blow up.

In [2], we give more properties on the AL algorithm when problem (1) is *infeasible*, with the aim at improving its stopping criterion and augmentation parameter update rule. Since, the AL algorithm is equivalent to the *proximal algorithm* on the dual function, these results are related to those on the behavior of the proximal method on monotone inclusion problems without solution (Bruck, Eckstein, Reich, Rockafellar, Silva, Spingarn, and others), but it goes a little further, by taking advantage of the special structure of the *quadratic* optimization problem (1). The result makes use of the notion of *feasible shift*, which is a vector s that makes $l \leqslant Ax + s \leqslant u$ feasible for x. The nonempty set of feasible shifts is denoted by $\mathcal{S}$ and the *smallest feasible shift* is denoted by $\overline{s} := \arg\min\{\|s\| : s \in \mathcal{S}\}$. Of course, $\overline{s} = 0$ if and only if the QP is feasible. The *closest feasible problem* is the feasible problem defined by

$$\inf_x \{q(x) : l \leqslant Ax + \overline{s} \leqslant u\}. \tag{5}$$

It is shown in [2] that, provided (5) has a solution, the following extension of (4) holds:

$$\begin{array}{c} \forall\, \beta > 0, \quad \exists\, L > 0, \quad \mathrm{dist}(\lambda_0, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \beta \quad \text{implies that} \\ \forall\, k \geqslant 1, \quad \|s_{k+1} - \overline{s}\| \leqslant \frac{L}{r_k} \|s_k - \overline{s}\|, \end{array} \tag{6}$$

where $\tilde{\mathcal{S}}_{\mathrm{D}}$ is the set of dual solutions to (5). This is quite similar to (4), except that $s_k := y_k - Ax_k$ no longer converges to zero but to $\overline{s}$. A difficulty occurs, however, when one tries to use the estimate (6) to tune the augmentation parameter r_k, since $\overline{s}$ is not known before solving problem (1). A bypass can be realized by observing the behavior of $s'_k := s_{k+1} - s_k$, which converges to zero, instead of that of s_k, yielding an update rule for the augmentation parameter r_k, which maintains it bounded, even when the problem is infeasible.

4 Numerical implementation

The second part of the talk will present the numerical results obtained with two implementations of the above AL algorithm for solving convex quadratic optimization problems [10]. The code `Oqla` is implemented in `C++`, which results in a fast and flexible solver. Flexibility means here that the solver is able to deal with various data representations, such as the dense, sparse, or ℓ-BFGS formats, in a manner similar to [9]. The code `Qpalm` is implemented in `Matlab`, which is slower but offers the possibility to

have a fair comparison with the `Matlab` solver `Quadprog` (`Oqla` and `Qpalm` are comparable in terms of iteration counters) and to make easier the experimentation of new ideas and features. Performance profiles [6] will compare the efficiency of `Oqla`/`Qpalm` to active set solvers (like `Quadprog` or `Qpa` from `Galahad` [11]) and interior point solvers (like `Ooqp` [9] or `Qpb` from `Galahad` [11]).

The `Oqla`/`Qpalm` solvers have also the nice features of terminating with one of the two following exclusive possibilities: either (i) the closest feasible problem (5) is unbounded, in which case a direction of unboundedness is provided, or (ii) a solution to the closest feasible problem (5) is found, in which case $\overline{s}$ is also determined. A direction of unboundedness is a direction $d \in \mathbb{R}^n$ such that

$$g^\mathsf{T} d < 0, \qquad Hd = 0, \qquad \text{and} \qquad Ad \in [l, u]^\infty, \tag{7}$$

where $[l, u]^\infty$ denotes the asymptotic cone of $[l, u]$. Such a direction always exists when the closest feasible problem is unbounded. This termination data is rarely provided by the other QP solvers, while it is very useful, for instance, to give a precise description of the QPs generated by the SQP algorithm [1].

Let us conclude by giving some features of `Oqla`/`Qpalm`:

- the implementation language is `C++`/`Matlab`, respectively;
- they can solve *convex* quadratic optimization problems;
- the augmented Lagrangian (AL) algorithm is used to solve the QPs;
- the AL subproblems (3) are solved by an active-set method with Rosen's inactivation technique [13, 8] to avoid exploring completely the unlikely optimal activated faces;
- the linear systems are solved by the preconditioned conjugate gradient algorithm; the preconditioners take into account the activation/inactivation of the bounds and inequality constraints; they can be diagonal or Cholesky-like (in each case the singularity of the augmented Hessians is dealt with);
- when the problem is unbounded, the solvers provide a direction of unboundedness d, satisfying (7);
- when the problem is infeasible, the solvers return a solution to the closest feasible problem (5);
- `Oqla` is currently distributed under the GPL license and `Qpalm` under the QPL license.

According to our experience in [5], these features make `Oqla`/`Qpalm` well adapted to an insertion into an SQP solver.

References

1. J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2006). *Numerical Optimization – Theoretical and Practical Aspects* (second edition). Universitext. Springer Verlag, Berlin. [authors] [editor] [google books]. 3, 4
2. A. Chiche, J.Ch. Gilbert (2014). How the augmented Lagrangian algorithm deals with an infeasible convex quadratic optimization problem. Research Report 8583, INRIA, BP 105, 78153 Le Chesnay, France. [pdf]. 2, 3
3. T.F. Coleman, Y. Li (1996). A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM Journal on Optimization*, 6(4), 1040–1058. [doi]. 2
4. F. Delbos, J.Ch. Gilbert (2005). Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. *Journal of Convex Analysis*, 12, 45–69. [preprint] [editor]. 2, 3
5. F. Delbos, J.Ch. Gilbert, R. Glowinski, D. Sinoquet (2006). Constrained optimization in seismic reflection tomography: a Gauss-Newton augmented Lagrangian approach. *Geophysical Journal International*, 164, 670–684. [doi]. 4
6. E.D. Dolan, J.J. Moré (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91, 201–213. [doi]. 4
7. Z. Dostál (2005). Inexact semimonotonic augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming. *SIAM Journal on Numerical Analysis*, 43, 96–115. [doi]. 2
8. D.-Z. Du, X.-S. Zhang (1989). Global convergence of Rosen's gradient projection method. *Mathematical Programming*, 44, 357–366. [doi]. 4
9. E.M. Gertz, S. Wright (2003). Object-oriented software for quadratic programming. *ACM Transactions on Mathematical Software*, 29, 58–81. [doi]. 2, 3, 4
10. J.Ch. Gilbert, É. Joannopoulos (2014). OQLA/QPALM - Convex quadratic optimization solvers using the augmented Lagrangian approach, which can deal with infeasibility. Technical report, INRIA, BP 105, 78153 Le Chesnay, France. (to appear). 2, 3
11. N.I.M. Gould, D. Orban, Ph.L. Toint (2003). GALAHAD: a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. Technical report, Rutherford Appleton Laboratory, Oxfordshire OX11 0QX. 4
12. N.I.M. Gould, D. Orban, Ph.L. Toint (2013). Cutest: a constrained and unconstrained testing environment with safe threads. Technical report, Rutherford Appleton Laboratory, Didcot, Oxfordshire OX11 0QX. 2
13. J.B. Rosen (1960). The gradient projection method for nonlinear programming; part I: linear constraints. *Journal of the Society for Industrial and Applied Mathematics*, 8, 181–217. [doi]. 4

PGMO - COPI'14

Optimizing the benefits of power infrastructure investments

Alice Chiche · Pierre Girardeau · Peio Lahirigoyen · Arnaud Renaud

Abstract We focus on showing practical examples of the use of numerical optimization for assessing the value of power interconnections between countries on a long-term horizon. The talk is application-oriented, with concrete examples of recent studies led on the European power system for transmission system operators, regulators, or national agencies. We also explain the current optimization models that are used and the numerical challenges faced for these types of analysis.

Keywords Investment planning · Cost-benefit analysis ·Interconnections

Mathematics Subject Classification (2000) MSC 91B25 · MSC 91B32 · MSC 49M27

Context

Power transmission projects are of particular importance for countries in the European Union, in an energy context including more and more renewable energies and extending the coupling of its power markets. In particular, interconnections allow for:

- ensuring better security of supply between countries and more generally between geographical areas;
- a facilitated access to the electricity market to various stakeholders;
- an increased value of economic surplus on the European perimeter;
- a more efficient integration of intermittent renewable energy, allowing them better expansion and larger distribution of the energy they produce.

To evaluate the cost of these projects, most EU countries, through their network operators, have developed their own calculation methods. Moreover, ENTSO-E[1], for the realization of its Ten Year Network Development Plan (TYNDP), has developed and published a common methodology among TSOs.
The proposed cost-benefit analysis consists in evaluating the total economic surplus generated over Europe, with or without a new piece of infrastructure. The evaluation of this surplus means solving an optimal dispatch problem, where power mixes in countries are represented by technology fleets in zones, connected by power transmission capacities. It is already used by TSOs and regulators to evaluate the interest of new transmission capacities between zones in Europe.

Artelys

[1] European Network of Transmission System Operators for Electricity

Artelys is a French independent company which specializes in numerical optimization, applied to various fields: energy, transport, defense, finance. Artelys often delivers studies to industrials, regulators or transmission system operators on the topic of techno-economical valuation of interconnection projects. For this purpose, Artelys has developed for almost ten years a simulation platform, called Artelys Crystal, dedicated to the energy systems.

The talk

The presentation will give a flavor of the newest developments introduced in our simulation platform: the capability to optimize the capacities of the energy mix (transmission, storage, production) jointly with the solving of the optimal dispatch problem on an hourly basis over the year. Within the POST project, a 4M€ "Investissements d'avenir" project funded by ADEME[2] and leaded by Artelys, we collaborate with INRIA in the development of cutting-edge algorithms for solving capacity expansion problems on a large-scale (European-wide) while keeping the underlying power dispatch problem intact (simulation of the production-demand equilibrium on an hourly basis).

This talk aims at explaining how optimization models are used in order to evaluate and rank interconnection projects together. There is a strong emphasis on the indicators used to perform the study. Not only production and investment costs are to be considered, but also more difficult constraints like security of supply concerns for countries. We base it on a concrete example of a recent study led at Artelys on the topic.

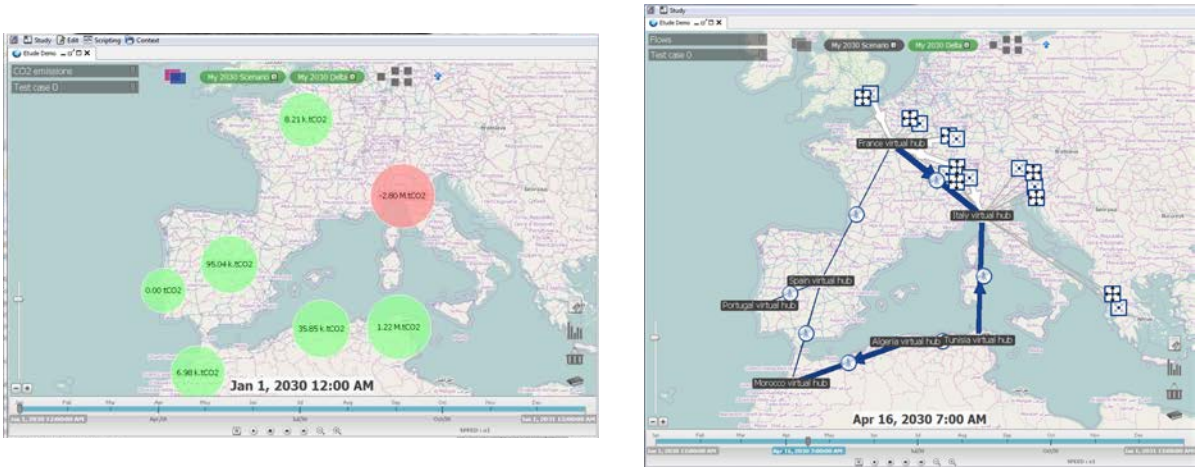


Figure 1- Example of a recently led study on the valuation of new transmission capacities across the Mediterranean

Alice Chiche, Pierre Girardeau, Peio Lahirigoyen, Arnaud Renaud
12, rue du Quatre Septembre
Tel.: +33-(0)1 44 77 89 00
Fax: +33-(0)1 42 96 22 61
E-mail: alice.chiche@artelys.com, pierre.girardeau@artelys.com, peio.lahirigoyen@artelys.com, arnaud.renaud@artelys.com

[2] French Environment and Energy Management Agency

PGMO - COPI'14

Natural Evolution Strategies for Direct Search

Tobias Glasmachers

Abstract Keywords evolutionary optimization, information geometry, direct search, randomized, rank-based

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Introduction

Randomized direct search heuristics are a class of optimization methods well suited for black-box optimization. Consider the general problem $\min f(x)$ for $x \in X$ where f is a black-box: its analytical form is unknown. It can be thought of either as the result of a complex computer simulation with parameters x to be tuned to a goal encoded by f, or even as a measurement obtained from an actual physical experiment. Each simulation run or experiment is costly (e.g., time intensive), and derivatives are not available. The objective function may be non-differentiable and even discontinuous, and evaluations may even be noisy.

2 Evolutionary Algorithms

Evolutionary Algorithms (EAs) are a sub-class of randomized direct search algorithms that mimic principles of Darwinian evolution theory. They operate on a population (set) of individuals (solutions) $x_1, \ldots, x_\mu \in X$. At all times the algorithm maintains a parent population. In each generation (iteration) the algorithm creates an offspring population $x'_1, \ldots, x'_\lambda \in X$ by means of randomized variation operators, often by first applying a recombination (or "crossover") operation mixing or merging properties of multiple parents, followed by a mutation step. This process of building new solutions from (relatively successful) current ones realizes the Darwinian principles of inheritance and variation. The final step is selection ("survival of the fittest"): since an EA usually aims for a fixed size memory footprint it reduces the two populations formed by parents and offspring to the size of the original parent population. This step is where EAs differ considerably from their biological counterpart. Selection pressure is not created by a dynamically changing environment with limited resources. Instead the selection process is based

Tobias Glasmachers
Institut für Neuroinformatik
Ruhr-Universität Bochum, Germany
Tel.: +49-(0)234-32-25558
E-mail: tobias.glasmachers@ini.rub.de

on the (relative) objective function values of the individuals. Hence in an EA the objective function is often called fitness function. Individuals with better fitness are granted higher changes to survive and to reproduce, driving the population as a whole towards better regions of the search space.

From an optimization perspective this process implements randomized direct search. The variation operators introduce random, undirected variations. They explore new, hopefully better solutions. The selection operator prefers better solutions, which amounts to exploitation. In contrast to many other classes of optimization algorithms EAs usually do not make direct use of objective function values. These values are solely used for pairwise comparisons and for ranking of solutions. Thus, a rank-preserving (strictly monotonically increasing) transformation of objective values does not impact the optimization. In other words, the objective function is "perceived" only in terms of its level sets. This is a strong invariance property of rank-based algorithms.

3 Evolution Strategies

Evolution strategies (ESs) were introduced by Rechenberg in the 1970s [7]. This class of evolutionary algorithms is specialized to optimization on the domain $X = \mathbb{R}^d$. ESs are characterized by stringent selection schemes (keeping the best μ individuals in each generation, either from the offspring or from the union of parents and offspring), as well as by their emphasis on mutation (as opposed to crossover) as a variation operator. Mutation is realized by adding a Gaussian random vector $z \in \mathbb{R}^d$, $z \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$. A key property of ESs is step size control: the variance σ^2 of the mutation operator is adapted online during the optimization run. This property allows for linear convergence on all scale invariant problems (where the level sets are scale-invariant around the optimum, e.g., convex quadratic functions and hence in arbitrarily good approximation local optima of $\mathcal{C}^2$ functions).

The modern default ES is the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [5]. This algorithm does not only adapt the overall variance but instead the full covariance matrix of the mutation distribution $\mathcal{N}(0, C)$, $C \in \mathbb{R}^{d \times d}$ symmetric and positive definite. Intuitively this is equivalent to online learning of a suitable linear transformation of the search domain $X = \mathbb{R}^d$ such that the problem becomes as well-conditioned as possible. Hence CMA-ES converges on all convex quadratic problems not only at a linear rate, but rather (nearly) with the best such rate over the whole problem class. CMA-ES and many variants thereof has been benchmarked on a large testbed of problems against a large set of competitors within the Black-Box Optimization Benchmarking (BBOB) competition. It has shown excellent performance over a wide range of problems [1].

4 Information Geometry

The classic perspective on an EA is that (the key part of) its algorithm state is given by the population $x_1, \ldots, x_\mu$. A more modern perspective is to consider the distribution $\mathcal{N}(m, C)$ from which offspring are drawn (m is a weighted mean of the parent population, which is the result of recombination). Hence the algorithm state is a distribution, which evolves iteratively within the statistical manifold of Gaussian distributions with parameters $\theta = (m, C)$.

With θ being the new algorithm state it makes sense to consider optimization of θ instead of x, for example by minimizing

$$\min_\theta \quad W_f(\theta) = \mathbb{E}_{x \sim \mathcal{N}(\theta)} \Big[w\big(f(x)\big) \Big]$$

where $w : \mathbb{R} \to \mathbb{R}$ is a monotonic weight function [6] (in the simplest case the identity). Under weak assumptions the objective function W_f is continuous and nearly everywhere differentiable. It can be minimized with gradient methods. The gradient involved integration of f and is hence intractable in the black-box setting, however, it can be efficiently approximated with a Monte Carlo estimator. The

estimated gradient is randomized, hence stochastic gradient descent (SGD) becomes applicable. This procedure alone does not always work well. In addition the intrinsic geometry of the statistical manifold of distributions, its information geometry, must be taken into account: the parameter manifold is equipped with a canonical Riemannian geometry with metric tensor given by the Fisher information matrix. The resulting gradient w.r.t. the intrinsic geometry is called the natural gradient [3].

5 Natural Evolution Strategies

A straightforward optimization strategy is to follow the natural gradient flow on the parameters of the statistical manifold of search distributions. Doing so by means of stochastic natural gradient descent indeed gives rise to an evolution strategy. In each generation the offspring population serves as a Monte Carlo sample for the estimation of the natural gradient. The resulting class of algorithm is called *Natural Evolution Strategies (NES)* [8], with the xNES (exponential NES) algorithm [4] marking the current state of the art. This ES is conceptually interesting in its own right. Importantly, it was shown to be closely related to the highly efficient CMA-ES [2]. This connection gives new insights into the working principles of modern evolution strategies.

6 Conclusions

Evolutionary Algorithms are randomized direct search methods. They are well suited for black-box optimization. Modern algorithms such as the CMA-ES demonstrate high efficiency on a large class of benchmark problems.

Modern evolution strategies have an additional interpretation as stochastic gradient descent algorithms on the space of search distributions. This property greatly enhances our understanding of the evolutionary search process.

References

1. http://coco.gforge.inria.fr/doku.php?id=bbob-2013-results.
2. Y. Akimoto, Y. Nagata, I. Ono, and S. Kobayashi. Bidirectional relation between CMA evolution strategies and natural evolution strategies. In *Parallel Problem Solving from Nature (PPSN)*, 2010.
3. S. Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10(2):251–276, 1998.
4. T. Glasmachers, T. Schaul, Y. Sun, D. Wierstra, and J. Schmidhuber. Exponential Natural Evolution Strategies. In *Genetic and Evolutionary Computation Conference (GECCO)*, Portland, OR, 2010.
5. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
6. Y. Ollivier, A. Auger, N. Hansen, and L. Arnold. Information-geometric optimization algorithms: a unifying picture via invariance principles. Technical Report arXiv preprint arXiv:1106.3708, arXiv.org, 2011.
7. I. Rechenberg. *Evolutionsstrategie: Optimierung technisher Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog, 1973.
8. D. Wierstra, T. Schaul, J. Peters, and J. Schmidhuber. Natural Evolution Strategies. In *Proceedings of the Congress on Evolutionary Computation (CEC08), Hongkong*. IEEE Press, 2008.

PGMO - COPI'14

CMA-ES: A Function Value Free Second Order Optimization Method

Nikolaus Hansen

Abstract We give a bird's-eye view introduction to the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) and emphasize relevant design aspects of the algorithm, namely its invariance properties. While CMA-ES is gradient and function value free, we show that using the gradient in CMA-ES is possible and can reduce the number of iterations on unimodal, smooth functions.

Keywords Optimization · Invariance · Evolution Strategies · CMA

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Introduction

We consider the problem to minimize an objective function

$$f : \mathbb{R}^n \to \mathbb{R} \ . \tag{1}$$

We do not assume to have any particular knowledge on the structure of f, thereby considering f as a black box. The *cost* of search is given to be the number of calls to the "black-box" f. In this context, we define a parameter vector $\boldsymbol{\theta}$ and a generic optimization procedure taking three steps

1. propose one or several new candidate solutions, depending on $\boldsymbol{\theta}$
2. evaluate the candidate solution(s) on f
3. update $\boldsymbol{\theta}$

This framework covers essentially any optimization procedure. We focus here on algorithms that sample from a probability distribution, and more specifically, sample in each iteration the same number of i.i.d. candidate solutions. In the continuous search space, a (multi-variate) normal distribution suggests itself as sample distribution, because it has maximum entropy (given that variances exist) and it is in a natural way detached from the given coordinate system. If the covariance matrix is a multiple of the identity, the distribution is isotropic.

Nikolaus Hansen
Inria
Research centre Saclay – Île-de-France
E-mail: nikolaus.hansen AT inria.fr

2 A Second Order Method

We consider a moderate search space dimensionality n, that is, $n \not\ll 10$ and $n \not\gg 100$. In this case, utilizing second-order information seems indispensable to achieve a competitive method. With a multivariate normal distribution, this can be naturally achieved by using a full covariance matrix to parametrize the distribution. We have $\boldsymbol{\theta}_t = (\mathbf{m}_t, \sigma_t, \mathbf{C}_t) \in \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{R}^{n \times n}$, and at iteration t, for $i = 1, \ldots, \lambda$, new candidate solutions obey

$$\mathbf{x}_i = \mathcal{N}(\mathbf{m}_t, \sigma_t^2 \mathbf{C}_t) = \mathbf{m}_t + \sigma_t \, \mathcal{N}(\mathbf{0}, \mathbf{C}_t) \ , \tag{2}$$

where $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. Choosing $\mathbf{C}_t$ mimics in effect the linear coordinate system transformation $\mathbf{C}_t^{-1/2}$, because (2) is equivalent to

$$\mathbf{C}_t^{-1/2} \mathbf{x}_i = \mathbf{C}_t^{-1/2} \mathbf{m} + \sigma_t \, \mathcal{N}(\mathbf{0}, \mathbf{I}) \ . \tag{3}$$

On convex-quadratic functions, $\mathbf{C}$ resembles in the ideal case the inverse Hessian matrix.

3 Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

The CMA-ES method prescribes the updates of $\mathbf{m}_t, \sigma_t, \mathbf{C}_t$ and of some additional hidden variables (evolution paths). The updates of $\mathbf{m}$ and $\mathbf{C}$ can be (to a large extend) derived from information geometry, as they follow the natural gradient in $\boldsymbol{\theta}$-space [7]. The updates can be motivated equally well from the maximum likelihood principle [2]. In contrast, the overall step-size σ_t is updated with the aim to achieve $\mathbf{C}^{-1}$-conjugate (orthogonal) movements of the mean $\mathbf{m}$. The update procedure is detailed in [1].

The main **governing design principle** of the CMA-ES algorithm is invariance (see also the contribution of A. Auger in the present volume). Namely, we find for CMA-ES the same invariance properties as for the Nelder-Mead Simplex Downhill method [6]: invariance to affine transformations of the search space (including translations and rotations) and invariance to order-preserving transformations of the f-value. The former is a natural consequence of a "full" second order method. The latter sets the two methods apart from other derivative-free or gradient-based methods to be in essence *function value free*.

Invariances mean generalization of behavior on single functions to the entire set of functions belonging to an respective equivalence class [2], thereby making previous observations meaningful for prediction.

Most **internal parameters** of CMA-ES are not left to the users choice, because reasonable settings do not depend on the given problem f (in contrast to optimal settings). However, the user must choose a suitable representation (scaling or transformation of parameters used in f) and corresponding values for $\mathbf{m}_0$ and σ_0. The population size λ is an optional parameter to modify, as well as termination settings.

4 Current Developments

We describe two recent developments in the CMA-ES method. First, in implementing the idea of so-called active CMA [5], i.e. using also *negative* weights for the update of the covariance matrix, we choose slowly decreasing weights, proportional to

$$-\log k + \log \frac{\lambda + 1}{2} \tag{4}$$

for $k = \lceil (\lambda+1)/2 \rceil, \ldots, \lambda$. Positive weights are by default set with the same equation for $k = 1, \ldots, \lfloor \lambda/2 \rfloor$. For $\lambda = 15$, the effective parent number becomes 4.5 for positive and 5.9 for negative update. Further-

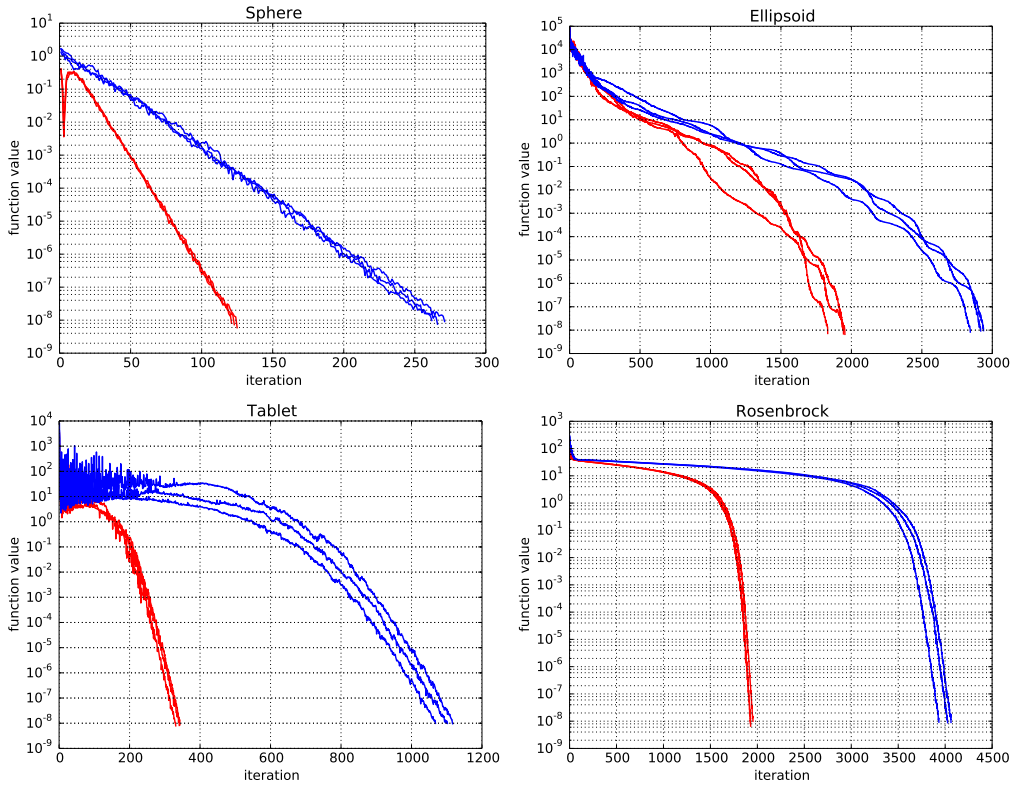


Fig. 1 Three runs in dimension $n = 40$ with default setting (blue) and with injected gradient in replacement of one sample (red). Initial values are $m_0 = -0.2$ in each component and $\sigma_0 = 0.1$, and $\lambda = 15$. The gradient reduces the number of iterations by a factor of between 1.5 and 3. The spike at the early stage on the Sphere function shows a possibly much faster convergence rate, then prevented by the slow decrement of the step-size.

more, the steps used with negative weights are normalized to constant length [4] and positive definiteness is guarantied by modulating the learning rate accordingly.

Second, we use the gradient of the function to generate one of the samples along $\mathbf{C}\,\nabla f(\mathbf{m}_t)$ [4]. Figure 1 shows results on four different functions [3]. Using the gradient reduces the number of iterations by a factor between 1.5 and 3. The effect of using the gradient is more pronounced with active CMA, and on the Tablet function, where active CMA is most relevant. Using the gradient also reduces the effect of λ on the number of iterations (not shown).

References

1. N. Hansen. The CMA evolution strategy: a comparing review. In J.A. Lozano, P. Larranaga, I. Inza, and E. Bengoetxea, editors, *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, pages 75–102. Springer, 2006.
2. N. Hansen and A. Auger. Principled design of continuous stochastic search: From theory to practice. In Yossi Borenstein and Alberto Moraglio, editors, *Theory and Principled Methods for the Design of Metaheuristics*, Natural Computing Series, pages 145–180. Springer Berlin Heidelberg, 2013.
3. N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
4. Nikolaus Hansen. Injecting external solutions into CMA-ES. *ArXiv e-prints*, arXiv:1110.4181, 2011.
5. G. A. Jastrebski and D. V. Arnold. Improving evolution strategies through active covariance matrix adaptation. In *IEEE Congress on Evolutionary Computation (CEC 2006), proceedings*, pages 2814–2821. IEEE Press, 2006.
6. John Ashworth Nelder and R Mead. A simplex method for function minimization. *The Computer Journal*, pages 308–313, 1965.
7. Y. Ollivier, L. Arnold, A. Auger, and N. Hansen. Information-geometric optimization algorithms: A unifying picture via invariance principles (2011v1; 2013v2). *ArXiv e-prints*, arXiv:1106.3708v2, 2013.

PGMO - COPI'14

Strategy iteration is strongly polynomial for 2-player turn-based stochastic games with a constant discount factor

Thomas Dueholm Hansen · Peter Bro Miltersen · Uri Zwick

Abstract Ye showed recently that the *simplex method* with Dantzig's pivoting rule, as well as Howard's *policy iteration* algorithm, solve discounted *Markov decision processes* (MDPs), with a constant discount factor, in *strongly polynomial* time. More precisely, Ye showed that both algorithms terminate after at most $O\big(\frac{mn}{1-\gamma}\log\frac{n}{1-\gamma}\big)$ iterations, where n is the number of states, m is the total number of actions in the MDP, and $0<\gamma<1$ is the discount factor. We improve Ye's analysis in two respects. First, we improve the bound given by Ye and show that Howard's policy iteration algorithm actually terminates after at most $O\big(\frac{m}{1-\gamma}\log\frac{n}{1-\gamma}\big)$ iterations. Second, and more importantly, we show that the same bound applies to the number of iterations performed by the *strategy iteration* (or *strategy improvement*) algorithm, a generalization of Howard's policy iteration algorithm used for solving 2-player turn-based *stochastic games* with discounted zero-sum rewards. This provides the first strongly polynomial algorithm for solving these games, solving a long standing open problem. Combined with other recent results, this provides a complete characterization of the complexity of the strategy iteration algorithm for 2-player turn-based stochastic games; it is strongly polynomial for a fixed discount factor, and exponential otherwise.

This work appeared in *Journal of the ACM*, 60(1):1-16, 2013.

Keywords Markov Decision Processes · Turn-Based Stochastic Games · Strategy Iteration · Policy Iteration · Strongly polynomial algorithms

Acknowledgements Work of first two authors supported by the Center for Algorithmic Game Theory, funded by the Carlsberg Foundation; and by the Sino-Danish Center for the Theory of Interactive Computation, funded by the Danish National Research Foundation and the National Science Foundation of China (under the grant 61061130540); and by the Center for research in the Foundations of Electronic Markets (CFEM), supported by the Danish Strategic Research Council. Thomas Dueholm Hansen is a recipient of the Google Europe Fellowship in Game Theory, and this research is supported in part by this Google Fellowship. Work of third author supported by grant 1306/08 of the Israeli Science Foundation.

Thomas Dueholm Hansen is currently supported by The Danish Council for Independent Research | Natural Sciences (grant no. 12-126512).

Thomas Dueholm Hansen
Stanford University
E-mail: tdh@cs.au.dk

Peter Bro Miltersen
Aarhus University
E-mail: bromille@cs.au.dk

Uri Zwick
Tel Aviv University
E-mail: zwick@tau.ac.il

Methodology for management of power system emergency situations

F. Héliodore, S. Poullain, B. Ismail, A. Nakib

Abstract: We are concerned by tools able to manage the risks of widespread disturbance, primarily addressing needs in operations and operational planning. Self-Organized Criticality allows introducing a new condition to manage power grid analysis.

Keywords: Self Organized Criticality · Statistical Power Flow Model

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Introduction

We are concerned by tools able to manage the risks of widespread disturbance, primarily addressing needs in operations and operational planning. Literature on blackout shows that large-scale disturbances occur much more frequently than deterministic risk assessment would suggest. The focus of this development is to provide:

- ✓ An awareness system that flags when the power system is approaching a condition where it is vulnerable to widespread disturbance,
- ✓ Ability to assess the risks associated with operational planning, for example day-ahead scheduling or maintenance scheduling,
- ✓ Provide users with guidance information that enables a prompt and targeted response to risks as they emerge.

It is important that the tools are available such that high impact/low probability events can be managed in the control room and in planning to reduce the system's exposure to risk. It is beneficial to have a real-time view of risk in power systems to alert the user to the occurrence and characteristics of a particular risk issue. It is also useful to quantify risk in relation to the operational planning timeframe, to aid decision making by highlighting critical system states and elements in the power system that are vulnerable. With this information, targeted guidance reports enable operators to move away from high risk states, with a measurement-based approach being used to validate the operational action and providing notification to operators that the system has moved to a lower risk state as a result of the action.
It is widely recognized that the probability of large events is not as scarce as we think, as their Cumulative Distribution Function (CDF) revealed power law regime (figure 1) and analogy with "Self Organized Criticality" regime was shown by Dobson & al.

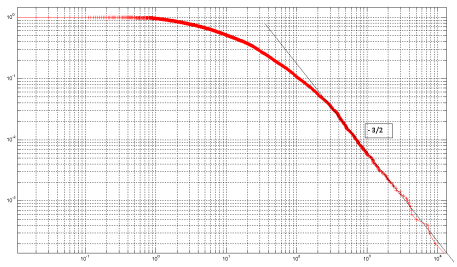


Figure 1: CDF of Energy not supplied

The analogy with the dynamics of sandpile is presented as below:

Power system	Variables	Sandpile

fractional overloads	system state	gradient profile
load increase	driving force	addition of sand
line improvements	relaxing force	gravity
line limit or outage	event	sand topples
cascading lines	cascade	avalanche

The term “self-organized critical” contains two ideas that are:

- ✓ "Critical", meaning that the system is correlated over large distances and long times. A system can be critical without being self-organized.
- ✓ "Self-organized" means that the system has a critical behaviour without the need for an operator to fine tune a control parameter. The system tends to dynamically adjust the parameter. This leads to a steady state and dynamic attractor of evolution. This is the case of sandpile which, from its critical slope, undergoes avalanche to return to a dynamically stable slope, depending on the size of the avalanche.

Considering an electrical network under self-organized criticality regime is therefore needed when one takes into account the grid in its environment, i.e. we consider the necessary reaction (or feedback reaction) for its operation. “feedback” reaction to any dysfunction can be operational policy control (control room), human intervention, maintenance operations, planning policy ... and can be quantify. The grid is then a dynamic system, managed by two opposing forces (load plan and "response to incident"), in the critical regime or not (subcritical, critical, super-critical). The power law behaviour observed experimentally finds its origin in this competition (universal behaviour). The SPFM (Statistical Power Flow Model) is based on a DC/AC Power Flow resolution with, as variables of interest:

- ✓ Evolution of the load (nodes) and generation (nodes),
- ✓ Improvement of the network (lines).

Failures or external events are randomly generated (Gaussian case or not) and two dynamics are represented, the slow dynamics representative of the evolution of the load and fast dynamic representative of the avalanche phenomenon on the lines. Temporal resolution Δt is 1 day and the time horizon may be years or even more. The indexing time can be reassessed because it is a sequence of events.
This means that SOC regime in power grid is equivalent to put the power grid under maximum stress where any random event can produce a minor failure or a major failure all over the network. This is the very significance of this new condition which states that SOC regime introduces new physical limitations as are introduced thermal, voltage or stability limitations (figure 2).

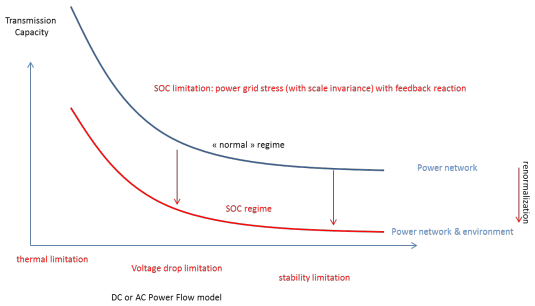


Figure 2: SOC Condition

2 Conclusions

As a matter of fact, a probabilistic approach seems to be accurate to describe the real behaviour of the power grid. Variables of interest (localized distribution, distributed generation, load demand with demand response, regulatory rules …) will be described from a statistical point of view. From the feedback reaction and the variables of interest, the absolute criterion would be to get out of the SOC regime (Gaussian regime). From the variables of interest and the definition of all the constraints (economical, technical, regulatory rules …) the best compromise should be reached (figure 3).

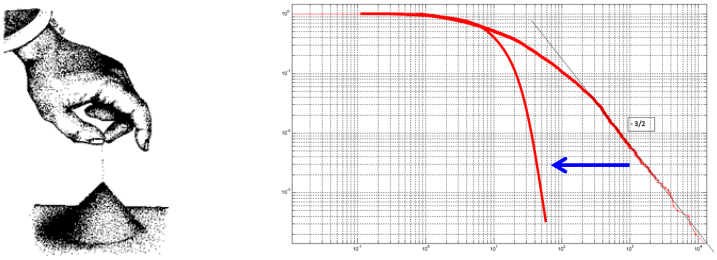


Figure 3: Self Organized Criticality model (sandpile) & Expected Optimization

The technical architecture of the project is organized around the Statistical Power Flow Model (SPFM). We shall address the question and the definition of the optimization process.

Frédéric Héliodore
Immeuble Galilée, 51 Esplanade du Général de Gaulle, 92907 Paris La Défense Cedex
Tel.: +33-(0)6 73 74 40 11
E-mail: Frederic.heliodore@alstom.com

Serge Poullain
Immeuble Galilée, 51 Esplanade du Général de Gaulle, 92907 Paris La Défense Cedex
Tel.: +33-(0)6 73 74 24 42
E-mail: serge.poullain@alstom.com

Boussaad Ismail
Immeuble Galilée, 51 Esplanade du Général de Gaulle, 92907 Paris La Défense Cedex
Tel.: +33-(0)6 48 55 95 39
E-mail: Boussaad.ismail@alstom.com

Amir Nakib
LISSI, University of Paris Est, Domaine Chérioux, 122 rue Paul Armangot, 94400 Vitry sur Seine
Tel.: +33-(0)6 61 83 96 58
E-mail: nakib@u-pec.fr

PGMO - COPI'14

Stochastic dynamic programming applied to microgrid management

Benjamin Heymann

Abstract Energy Management Systems (EMS) are usually based on the integer programming framework. A novel method based on stochastic control and dynamic programming is proposed. It allows for a precise modeling of complex constraints such as the minimal working power of a diesel motor. The EMS proposed here is based on a rolling horizon strategy. It takes into account historical data and randomness as well as the aging of the battery and the motor working constraints. A semi-lagrangian scheme was implemented within BOCOP - a software developed at INRIA - for the numerical simulations. We will compare such method with the one that is currently implemented in a real microgrid in northern Chile. We will also analyse the performance of the software implementation. A practical implementation of the EMS is envisioned. This project is the product of a collaboration between the COMMADS (INRIA, France) and Centro de Energia teams (Universidad de Chile, Chile).

Keywords Dynamic Programming · Stochastic Control

Mathematics Subject Classification (2000) MSC 93E20 · MSC 90B05 · MSC 90C39

1 Introduction

A microgrid is an electrical network that is disconnected from the main network and is operated autonomously. The model described in the following is based on a real microgrid operated in a village in northern Chile. The village is very isolated and relies completely on the microgrid for its electricity supply. The grid uses a mix of fuel and renewable energy. Compared with technics such as interger programming, the use of stochastic control allows for a dramatic reduction of the computational cost. The grid we are considering consists of a solar panel, a wind turbine, a diesel motor and a battery. Our model aims at finding the optimal operation of those elements that meets the demand. The north of Chile is characterized by a sunny and regular climat. We consider in the model that there is no uncertainty from the production side. On the other hand, the demand will be modeled with a stochastic process. The battery state of health depends on the way it is operated. That's why we will take into account the aging of the battery within the model.

Benjamin Heymann
Centre de Mathmatiques Appliques
Tel.: +33-(0)6 62 39 58 04
E-mail: benjamin.heymann@inria.fr

PGMO - COPI'14

Generic uniqueness of the bias vector of mean-payoff zero-sum games

Marianne Akian · Stéphane Gaubert · Antoine Hochart

Abstract Zero-sum mean payoff games can be studied by means of a nonlinear spectral problem. When the state space is finite, the latter consists in finding an eigenpair (u, λ) solution of $T(u) = \lambda \mathbf{1} + u$ where $T : \mathbb{R}^n \to \mathbb{R}^n$ is the Shapley (dynamic programming) operator, λ is a scalar, $\mathbf{1}$ is the unit vector, and $u \in \mathbb{R}^n$. The scalar λ yields the mean payoff per time unit, and the vector u, called the *bias*, allows one to determine optimal stationary strategies. The existence of the eigenpair (u, λ) for all state-dependent perturbations of the payments, i.e. for all the Shapley operators $T + v$ where v is any vector in $\mathbb{R}^n$, is related to an ergodicity condition which depends only on the support of the transition probabilities. A basic issue is to understand for which classes of games the bias vector is unique (up to an additive constant). Here, we consider a perfect information zero-sum game with finite state and action spaces, whose the transition payments are perturbed by additive parameters which depend only on the state. We assume that the spectral problem is solvable for all values of the perturbations and we show that the bias vector, thought of as a function of the perturbation vector, is generically unique (up to an additive constant). The proof uses techniques of nonlinear Perron-Frobenius theory.

Keywords Optimal control · Game theory · Stochastic optimal control

Mathematics Subject Classification (2000) MSC 47H25 · MSC 47H05 · MSC 91A20

1 Mean payoff zero-sum games

Zero-sum repeated games describe long term interactions between two agents, called players, with opposite interests. Here we consider perfect information games with finite state and action spaces. A game is defined by a finite state space $[n] := \{1, \dots, n\}$; a nonempty finite action space A_i for the first player (MIN), depending on the current state $i \in [n]$; a nonempty finite action space $B_{i,a}$ for the second player (MAX), depending on the current state $i \in [n]$ and the action $a \in A_i$ selected by player MIN; a transition payment $r_i^{ab} \in \mathbb{R}$ paid by player MIN to player MAX at each stage and depending on the current state i and the actions a, b selected by the players; a transition probability $P_i^{ab} \in \Delta_{n-1}$, depending on the same data, where Δ_{n-1} is the set of probabilities over $[n]$. A play follows the subsequent rules: at each stage (or time step), if the current state is i, player MIN selects an action $a \in A_i$ then, knowing this choice, player MAX selects an action $b \in B_{i,a}$; this gives rise to a transition payment r_i^{ab} and the next state is chosen

Marianne Akian · Stéphane Gaubert · Antoine Hochart
INRIA Saclay-Ile-de-France and CMAP, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau Cedex
E-mail: marianne.akian@inria.fr (M. Akian)
E-mail: stephane.gaubert@inria.fr (S. Gaubert)
E-mail: hochart@cmap.polytechnique.fr (A. Hochart)

according to the probability P_i^{ab}. The payoff of a play is defined as the sum of the transition payments, that player MIN intends to minimize and player MAX intends to maximize. Given an initial state i, the game in horizon k is known to have a value, denoted by $v_i^k \in \mathbb{R}$. The value vector $v^k = (v_i^k)_{i\in[n]}$ follows a dynamic programing recursion that involves the *Shapley operator* of the game, $T : \mathbb{R}^n \to \mathbb{R}^n$, given by

$$[T(x)]_i = \min_{a\in A_i} \max_{b\in B_{i,a}} \Big(r_i^{ab} + \sum_{j\in[n]} [P_i^{ab}]_j x_j\Big), \tag{1}$$

Then, we have $v^k = T(v^{k-1})$ and $v^0 = 0$. We will be interested in the *mean payoff vector* $\chi(T) := \lim_{k\to\infty} v^k/k = \lim_{k\to\infty} T^k(0)/k$.

The question of the existence of the mean payoff vector is a difficult problem in general, studied by Bewley, Kohlberg, Mertens, Neyman, Rosenberg, Sorin, Vigeral [6,15,16,18]. However, in our settings, T appear as a piecewise affine operator, and a result by Kohlberg [14] implies the existence of $\chi(T)$. It is also guaranteed if there is a pair $(u, \lambda) \in \mathbb{R}^n \times \mathbb{R}$ solution of the so-called *ergodic equation*

$$T(u) = \lambda\mathbf{1} + u, \tag{2}$$

where $\mathbf{1}$ is the unit vector of $\mathbb{R}^n$. In that case, we have $\chi(T) = \lambda\mathbf{1}$. The scalar λ is called the *eigenvalue* of T and it gives the mean payoff per time unit. The vector u, called *bias vector* and defined up to an additive constant, gives optimal stationary strategies.

2 Bias vectors

The structure of the bias vectors is a fundamental issue. In the one-player case, i.e. for optimal control problems, the ergodic equation (2) has been much studied and the representation of bias vectors is well understood. In the deterministic case, the analysis relies on max-plus spectral theory [5]. The set of bias vectors has the structure of a max-plus cone and it has a unique minimal generating family, consisting of certain "extreme" generators, which can be identified by looking at the support of the maximizing measures in the linear programming formulation of the optimal control problem, or at the "recurrence points" of infinite optimal trajectories. A combinatorial interpretation of some of these results, in terms of polyhedral fans, has been recently given by Sturmfels and Tran [17]. The eigenproblem has also been studied for an infinite dimensional state space in the context of infinite dimensional max-plus spectral theory by Akian, Gaubert and Walsh [4], and also in the setting of weak KAM theory, see Fathi [9]. In the stochastic case, i.e. for Markov decision processes, the ergodic equation (2) is known as the "average optimality equation" [11]. The structure of the set of bias vectors in the case of a finite state space is still known [2].

In the two player case, the structure of the set of bias vectors is less well known. The uniqueness of the bias vector (up to an additive constant) is a particularly important matter for algorithmic purposes. Indeed, the nonuniqueness of the bias typically leads to numerical instabilities or degeneracies. In particular, the standard Hoffman and Karp policy iteration algorithm [12] may fail to converge in such a situation. Some refinements of the Hoffman and Karp scheme have been proposed by Akian, Cochet-Terrasson, Detournay and Gaubert [1] and by Bourque and Raghavan [7], allowing one to circumvent such degeneracies at the price of a complexification of the algorithm. Hence, it is of interest to understand when such technicalities can be avoided.

3 Generic uniqueness of the bias

The solvability of the ergodic equation (2) is related to ergodicity conditions. In the zero-player case, i.e. for discrete time Markov chains, such conditions are classical, one of them being the uniqueness of the invariant measure [13]. In [3], the authors have extended the notion of ergodicity of a Markov

chain to zero-sum two-player repeated games with finite state space. In particular, based on results of Gaubert and Gunawardena [10] involving the recession function of the Shapley operator T, defined by $\widehat{T}(x) := \lim_{\rho\to+\infty} T(\rho x)/\rho$, they have shown the following.

Theorem 1 (Ergodicity of zero-sum games [3]) *Let us fix a state space $[n]$ and the nonempty actions spaces A_i and $B_{i,a}$ of the two players. Let r be a bounded transition payment, let P be a transition probability, and let T be the Shapley operator of this game. Then, the following properties are equivalent:*

1. *the recession function $\widehat{T}$ has only trivial fixed points, i.e. proportional to $\mathbf{1}$;*
2. *the mean payoff vector of the game does exist and is constant for all state-dependent additive perturbations v of the transition payment, meaning that the transition payments are $r_i^{ab} + v_i^{ab} = r_i^{ab} + v_i$, for all $i \in [n]$, $a \in A_i$ and $b \in B_{i,a}$;*
3. *the ergodic equation $v + T(u) = \lambda\mathbf{1} + u$ is solvable for all vectors $v \in \mathbb{R}^n$.*

Moreover, these conditions depend only on the support of P, defined as the set of points at which the function $(i, a, b, j) \mapsto [P_i^{ab}]_j$ takes nonzero values.

Here, we assume that the ergodicity condition is satisfied for the game introduced in Section 1 and we think of all the state-dependent perturbations of the payment as a space of potentials, denoted by $\mathcal{V} := \mathbb{R}^n$. We prove the following.

Theorem 2 *The space of potentials $\mathcal{V}$ is covered by polyhedral complex such that for each potential $v \in \mathcal{V}$ in the interior of a full-dimensional cell, the set of bias vectors of the Shapley operator $T + v$ is unique, up to an additive constant.*

A first ingredient in the proof is the nonlinear spectral theorem in the stochastic case [2]. A second ingredient is a general result, showing that the fixed point set of a nonexpansive self-map of $\mathbb{R}^n$ is a retract of $\mathbb{R}^n$ [8]. This allows us to infer the uniqueness of the bias vector of a Shapley operator from the uniqueness of the bias vector of the reduced Shapley operators obtained by fixing the strategy of one player.

References

1. M. Akian, J. Cochet-Terrasson, S. Detournay, and S. Gaubert. Policy iteration algorithm for zero-sum multichain stochastic games with mean payoff and perfect information.
2. M. Akian and S. Gaubert. Spectral theorem for convex monotone homogeneous maps, and ergodic control. *Nonlinear Anal.*, 52(2):637–679, 2003.
3. M. Akian, S. Gaubert, and A. Hochart. Ergodicity conditions for zero-sum games.
4. M. Akian, S. Gaubert, and C. Walsh. The max-plus Martin boundary. *Doc. Math.*, 14:195–240, 2009.
5. F. Baccelli, G. Cohen, G. Olsder, and J.-P. Quadrat. *Synchronization and linearity*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, 1992.
6. T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Math. Oper. Res.*, 1(3):197–208, 1976.
7. M. Bourque and T.E.S. Raghavan. Policy improvement for perfect information and additive reward additive transition stochastic games with discounted and average payoff. Preprint available from the author's web page, 2013.
8. Jr. Bruck, R. E. Properties of fixed-point sets of nonexpansive mappings in banach spaces. *Transactions of the American Mathematical Society*, 179:pp. 251–262, 1973.
9. A. Fathi. Weak Kam theorem in Lagrangian dynamics. to appear, 2014.
10. S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Trans. Amer. Math. Soc.*, 356(12):4931–4950 (electronic), 2004.
11. O. Hernández-Lerma and J. B. Lasserre. *Further topics on discrete-time Markov control processes*, volume 42 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.
12. A. J. Hoffman and R. M. Karp. On nonterminating stochastic games. *Management Sci.*, 12:359–370, 1966.
13. J. G. Kemeny and J. L. Snell. *Finite Markov chains*. Springer-Verlag, New York-Heidelberg, 1976. Reprinting of the 1960 original, Undergraduate Texts in Mathematics.
14. E. Kohlberg. Invariant half-lines of nonexpansive piecewise-linear transformations. *Math. Oper. Res.*, 5(3):366–372, 1980.
15. J.-F. Mertens and A. Neyman. Stochastic games. *Internat. J. Game Theory*, 10(2):53–66, 1981.
16. D. Rosenberg and S. Sorin. An operator approach to zero-sum repeated games. *Israel J. Math.*, 121:221–246, 2001.
17. B. Sturmfels and N. M. Tran. Combinatorial types of tropical eigenvectors. *Bull. Lond. Math. Soc.*, 45(1):27–36, 2013.
18. G. Vigeral. A zero-zum stochastic game with compact action sets and no asymptotic value. *Dyn. Games Appl.*, 3(2):172–186, 2013.

PGMO - COPI'14

The operational planning of interventions at ERDF

Christophe BATTU (ERDF), Olivier JUAN (EDF R&D), Bayram KADDOUR (EDF R&D), Eric THOMAS (INFOTEL), Riadh ZORGATI (EDF R&D)

Abstract : ERDF (Electricité Réseau Distribution France) is the distribution system operator (DSO) in France. Planning the routings of its agents involves up to 11,000,000 interventions, 20,000 agents (or technicians), 225,000,000 kilometers, and 17,000 vehicles. The planning phase is performed in several steps described below .

(1) Strategic planning consists in determining the location of the sites.
(2) Tactical planning consists in partitioning the intervention zones.
(3) Operational planning consists in daily building the routings of the agents of each site.
(4) On-line planning consists in adjusting the above scheduled routings according to random events (e.g., cancellation of a client, weather conditions, traffic jam, etc.).

The operational planning is an asynchronous variant of travelling salesman problem (TSP) with time-window constraint and multiple agents. We will describe in this presentation the objectives of this problem and its different constraints.

Keywords : scheduling, travelling salesman problem with time window, mixed integer programming

Christophe BATTU
ERDF, Nanterre, France
Christophe.battu@erdf.fr

Bayram KADDOUR
EDF R&D, Clamart, France
Bayram.kaddour@edf.fr

PGMO - COPI'14

Dual Approximate Dynamic Programming: Theoretical and Practical Questions.

P. Carpentier · V. Leclère

Abstract Multistage stochastic optimization problems are large scale optimization problems and thus numerically hard to solve. One way to tackle a large scale problem is to use a decompostion method that consists in constructing a number of problems of smaller size that can be solved independently. Those subproblems are solved iteratively and adapted until we are able to synthetize an optimal solution of the global problem from the optimal solution of the subproblems.

Spatial decomposition methods are well known in a deterministic setting. However their direct counterparts in a stochastic setting lead to a number of difficulties. We present a stochastic spatial decomposition method that solves an approximation of the original problem. The approximation consists in relaxing an almost sure coupling constraint into its conditional expectation. We discuss theoretical and practical questions raised by this algorithm. The algorithm is illustrated on the problem of managing a chain of hydroelectric dams.

Keywords Multistage Stochastic Optimization · Decomposition Method · DADP

Mathematics Subject Classification (2000) MSC 90C39 · 90C15 · MSC 49M27

1 Introduction

We consider N stochastic dynamic systems coupled by almost sure equality constraints. The global cost to be minimized is the expectation of a sum over the N systems of the sum over time of local costs. The problem considered is detailed in §2.

The price decomposition scheme consists in dualizing the coupling constraints, fixing the multiplier, and obtaining N uncoupled subproblems. From the solution of all subproblems we update the multiplier before iterating. However, we show in §2.1 that this price decomposition scheme leads to subproblems which are too difficult to solve by Dynamic Programming (dimension of the state too large). Thus, we propose an approximation method called Dual Approximate Dynamic Programming (DADP) and based on the main following ideas:

Pierre Carpentier
UMA (ENSTA)
E-mail: Pierre.Carpentier@ensta-paristech.fr

Vincent Leclère
Université Paris-Est, CERMICS (ENPC), 6-8 Avenue Blaise Pascal, Cité Descartes , F-77455 Marne-la-Vallée
E-mail: leclerev@cermics.enpc.fr

– relaxing the almost sure coupling equality constraints into conditional expectation constraint,
– using a price decomposition scheme to obtain subproblems,
– solving the subproblems through methods like Dynamic Programming.

2 Presentation of DADP

We are interested in a production problem involving N units of internal state $\boldsymbol{X}_t^i$. The system is affected by time-independent exogenous noise $\{\boldsymbol{W}_t\}_0^{T-1}$.

We consider the following optimization problem

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \sum_{i=1}^{N} \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t)\Big] \tag{1a}$$

$$\boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) \tag{1b}$$

$$\sum_{i=1}^{N} \theta_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) = 0 \tag{1c}$$

where $\boldsymbol{U}_t^i$ is assumed to be non-anticipative.

2.1 First Idea: Price Decomposition Scheme

If it were not for constraint (1c), Problem (1) would lead to a sum of independent subproblems, that could be optimized independently. Hence, we dualize the coupling constraints (1c), where the multiplier $\boldsymbol{\lambda}$ is a (adapted) stochastic process $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_t\}_{t=0}^{T}$. We solve the maximization part of the dual problem using a gradient-like algorithm on $\boldsymbol{\lambda}$. Thus, for a fixed multiplier process $\boldsymbol{\lambda}^{(k)}$, we have to solve N independent problems of smaller size

$$\begin{aligned} \min_{\boldsymbol{x}^i,\boldsymbol{U}^i} \quad & \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) + \boldsymbol{\lambda}_t^{(k)} \cdot \theta_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t)\Big] \\ & \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) \end{aligned} \tag{2}$$

Problem (2) has a physical state $\boldsymbol{X}_t^i$ of smaller dimension than the state of Problem (1). Unfortunately, the multiplier process $\{\boldsymbol{\lambda}_t^{(k)}\}_{t=0}^{T-1}$ can be seen as a noise that is not time independent. Hence, Dynamic Programming is not numerically tractable: the information state is a priori the past realizations of $\boldsymbol{W}$.

2.2 Second Idea: Constraint Relaxation

We have seen in the previous section that, if we apply a price decomposition scheme to Problem (1) the subproblems (2) cannot be solved numerically by the Dynamic Programming approach because of the curse of dimensionality. Thus, we approximate Problem (1) by relaxing the almost sure constraints, in order to obtain subproblems with a smaller dimension state.

We consider a stochastic process $\{\boldsymbol{Y}_t\}_{t=0}^{T-1}$, called an *information process*, that follows a dynamic

$$\forall t \in [\![0, T-1]\!], \qquad \boldsymbol{Y}_{t+1} = \tilde{f}_t(\boldsymbol{Y}_{t-1}, \boldsymbol{W}_t)\ , \tag{3}$$

where $\tilde{f}_t$ are known deterministic functions. The choice of the information process is arbitrary, but determines the quality of the method.

We replace, in Problem (1), Constraint (1c) by its conditional expectation w.r.t the information process (see Constraint (4c)):

$$\min_{\boldsymbol{X},\boldsymbol{U}} \quad \mathbb{E}\Big[\sum_{i=1}^{N}\sum_{t=0}^{T} L_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t)\Big] \tag{4a}$$

$$\boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) \tag{4b}$$

$$\mathbb{E}\Big[\sum_{i=1}^{N} \theta_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) \;\Big|\; \boldsymbol{Y}_t\Big] = 0\,. \tag{4c}$$

We easily see that we can restrict ourselves to multiplier processes $\boldsymbol{\mu}$, such that $\boldsymbol{\mu}_t$ is measurable w.r.t $\boldsymbol{Y}_t$. The multiplier can thus be seen as a function of $\boldsymbol{Y}_t$, and the decomposed problem (equivalent to Problem (2)) reads, for a given multiplier $\mu_t^{(k)}$:

$$\begin{aligned} \min_{\boldsymbol{X}^i,\boldsymbol{U}^i} \quad & \mathbb{E}\Big[\sum_{t=0}^{T} L_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) + \mu_t^{(k)}(\boldsymbol{Y}_t)\cdot\theta_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t)\Big] \\ & \boldsymbol{X}_{t+1}^i = f_t^i(\boldsymbol{X}_t^i, \boldsymbol{U}_t^i, \boldsymbol{W}_t) \\ & \boldsymbol{Y}_{t+1} = \tilde{f}_t(\boldsymbol{Y}_t, \boldsymbol{W}_t) \end{aligned} \tag{5}$$

which can be solved by Dynamic Programming with the state $\big(\boldsymbol{X}_t^i, \boldsymbol{Y}_t\big)$.

2.3 General Scheme

Data: Information process, evolution functions $\tilde{f}_t$, starting point y_0 and initial multipliers $\boldsymbol{\mu}_t^{(0)}$;
Result: optimal multipliers $\mu_t^\sharp$, admissible feedback ;
repeat
  forall the $i \in [\![1, N]\!]$ **do**
    Solve Problem (5) ;
  forall the $t \in [\![0, T-1]\!]$ **do**
    Estimate $\mathbb{E}\big[\boldsymbol{\Delta}_t^k \mid \boldsymbol{Y}_t\big]$;
    Update the multiplier : $\boldsymbol{\mu}_t^{(k+1)} = \boldsymbol{\mu}_t^{(k)} + \rho \cdot \mathbb{E}\big[\boldsymbol{\Delta}_t^k \mid \boldsymbol{Y}_t\big]$;
until $\mathbb{E}\big[\boldsymbol{\Delta}_t^k \mid \boldsymbol{Y}_t\big] \simeq 0$;
Compute admissible feedbacks ;

Algorithm 1: General Scheme of DADP

3 Conclusion

In the talk we present theoretical questions (existence of multiplier, convergence of the master problem of the decomposition methods) and practical difficulties (choice of information process, update of multiplier) of the Dual Approximate Dynamic Programming algorithm.

Acknowledgements We would like to thanks the PGMO programm for its financial support of this work.

PGMO - COPI'14

Feature selection at scale: the case study of online advertising

Rodolphe Jenatton · Clément Calauzènes · Huang Xu · Olivier Chapelle · Nicolas Le Roux

Abstract While feature selection are ubiquitous, they are often slow or inaccurate, depending on whether the number of datapoints of the number of features is large. We propose a hybrid feature selection technique which scales well to datasets of several billion datapoints and several thousands of features. We also theoretically prove the effectiveness of filtering as a preprocessing step to an embedded method. Further, we analyze the influence of every part of the system.

Keywords Feature selection · Large-scale

Mathematics Subject Classification (2000) MSC 49M37 · MSC 65K05 · MSC 90C15

1 Introduction

Optimal feature selection is known to be an NP-hard problem in general [1], where the complexity grows at least exponentially with the number of features considered. The feature selection methods are usually classified into three types [2,4]: *Filter*, *Wrapper* and *Embedded* methods. Filter methods select features as a preprocessing step; they are cheap but can be suboptimal. Wrapper methods use the prediction method as a black box to score subsets of features; they are good but very expensive. Embedded methods [6], which belong to a middle ground perform feature selection as part of the training process of the prediction method. Linear classifiers that use ℓ_1 regularization on the weights [9] and Recursive Feature Elimination (RFE) [5], a backward elimination method that uses smallness of weights to decide feature removal, fall in this class.

Filter methods are often preferred in practice because they are cheap and more practical. They typically rank features according to some correlation score between the feature and the target. In case of categorical features – the focus of this paper – a widely popular method is to use the Mutual Information [7] as

Rodolphe Jenatton
E-mail: rodolphe.jenatton@gmail.com

Clément Calauzènes
E-mail: c.calauzenes@criteo.com

Huang Xu
E-mail: h.xu@criteo.com

Olivier Chapelle
E-mail: o.chapelle@criteo.com

Nicolas Le Roux
E-mail: n.leroux@criteo.com

a correlation score. This method has some limitations, including its reliance on empirical probability estimates to compute the mutual information. One way to address this was studied in [12] by considering approximate distributions over the mutual information value instead of a point estimate.

In this paper, the categorical features are encoded as sparse binary vectors. This means that one feature associated with several parameters, one per element of the vector. Our goal will be to select whole features rather than individual parameters. Indeed, rejecting a feature allows us to get rid of all the costly extraction and logging of the feature in real-time. ℓ_1 regularization for instance is an effective way of reducing the number of parameters in the system but it does not necessarily suppress all the parameters of a given feature and cannot therefore be used for *feature selection* in our context. For this, a possible regularization is the so-called ℓ_1/ℓ_2 regularization or group lasso [8]. This type of regularization has been used for feature selection in the context of multi-task learning [10] and multi-class classification [3].

2 Conclusions

While feature selection can have multiple purposes, it is usually applied for three main objectives [4], namely to avoid overfitting and improve model performance, as measured by the supervised task at hand, to provide cost-effective models and to summarize the description of the data to make it more interpretable and understandable. In our context, we are primarily interested in the first two aspects above, in the sense of bringing as much performance as possible while being able to maintain a reasonable number of selected features. Indeed, the number of features in the model directly impacts the learning time of our models, the storage requirements of our dataset and, most importantly, the latency at prediction time.

This general objective is made more challenging because of the specifics of our setting:

- **Categorical variables:** When dealing with categorical variables, it is important to distinguish the number of parameters—in our case p, from the number of features, denoted here by $|\mathcal{F}|$ with $p \gg |\mathcal{F}|$.In our setting, a single parameter thus corresponds to a pair (f, m) with $f \in \mathcal{F}$ and $m \in \mathcal{M}_f$, and removing this sole parameter would not permit to perform feature selection. As a result, we need to design an approach capable of selecting (or discarding) the whole set of modalities $\mathcal{M}_f$ for a given feature f, i.e., by considering jointly the set of parameters related to *all* the modalities $\mathcal{M}_f$ of f.
- **Conjunction features:** Our prediction system is based on linear models. One way of bringing non-linearity in our models and improving their performance is to introduce *conjunctions* of categorical features. Although the number of *base* features is relatively low in practice, considering conjunctions with multiple orders can lead to a combinatorial growth of the size of the set of features.
- **Scale:** The practical setting we are interested in implies large scales in all the dimensions of the problem: The parameter space has a size in the order of 10^7, the data set is comprised of about 10^8 observations, while the total number of features $|\mathcal{F}|$ can be up to 10^5 when exploring the space of conjunctions. Hence, the need for scalability is at the core of the approach proposed. In particular, it is worth mentioning that we want our method to leverage a parallel computation environment, which will notably explain some choices made in the optimization.

We propose an approach based on the combination of two steps:

- **Forward:** A univariate screening phase built from the technique developed in [11, **?**].
- **Backward:** A multivariate step relying on a convex, group-sparsity inducing technique.

The rationale of our strategy is quite simple and intuitive: Since our initial set of features $\mathcal{F}$ can be quite large and most likely contains poorly predictive features, we want to be able to quickly and cheaply discard them. This is the role of the forward screening step. On the other hand, as the set of selected

features after the first phase tends to be too rough, the backward step makes it possible to refine the selection.

We show, both theoretically and empirically, that this procedure leads to efficient feature selection, even for very large datasets.

References

1. Edoardo Amaldi and Viggo Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209(1):237–260, 1998.
2. Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1):245–271, 1997.
3. Olivier Chapelle and S Sathiya Keerthi. Multi-class feature selection with support vector machines. In *Proceedings of the American statistical association*, 2008.
4. I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.
5. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.
6. T. N. Lal, O. Chapelle, J. Weston, and A. Elisseeff. Embedded methods. In I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, editors, *Feature Extraction: Foundations and Applications*, Studies in Fuzziness and Soft Computing ; 207, pages 137–165. Springer, 2006.
7. Dennis V Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27:986–1005, 1956.
8. Lukas Meier, Sara Van De Geer, and Peter Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.
9. Andrew Y Ng. Feature selection, l_1 vs. l_2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, 2004.
10. Guillaume Obozinski, Ben Taskar, and Michael I Jordan. Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing*, 20(2):231–252, 2010.
11. Rómer Rosales and Olivier Chapelle. Attribute selection by measuring information on reference distributions. In *Tech Pulse Conference, Yahoo!*, 2011.
12. Marco Zaffalon and Marcus Hutter. Robust feature selection by mutual information distributions. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, pages 577–584, 2002.

PGMO - COPI'14

Elliptically distributed joint probabilistic constraints

Abdel Lisser · Jianqiang Cheng · Michal Houda.

Abstract This paper studies chance constraints problems where the random constraint matrix rows are dependent. The dependence is handled by the means of copulas, namely Archimedean copulas. Convexity issues are addressed, and conic approximations are given. Numerical experiments are performed on a set of randomly generated instances.

Keywords chance constrained programming · Archimedean copulas · elliptical distributions · second-order cone programming

Mathematics Subject Classification (2000) MSC 90C59 · MSC 90C15

1 Introduction

Consider a *linear optimization problem with uncertainty*

$$\min c^T x \text{ s.t. } Tx \leq h,\ x \in X, \tag{1}$$

where $X \subset \mathbb{R}^n$ is a deterministic closed convex set, $c \in \mathbb{R}^n$, $h = (h_1, \ldots, h_K)^T \in \mathbb{R}^K$ deterministic vectors, and $T \in \mathbb{R}^K \times \mathbb{R}^n$ ia an uncertain (unknown) matrix with rows $t_1^T, \ldots, t_K^T$; n, K are structural components of the optimization problem (1) under consideration. If X is polyhedral, and a realization

Abdel Lisser
Université Paris Sud
91405 Orsay Cedex, France
Tel.: +33-(0)69156621
Fax: +33-(0)69156573
E-mail: lisser@lri.fr.fr

Jianqing Cheng
Université Paris Sud
91405 Orsay Cedex, France
Tel.: +33-(0)69156621
Fax: +33-(0)69156573
E-mail: cheng@lri.fr.fr

Michal Houda
Institute of Information Theory and Automation
Pod Vodarenskou ve(z 4, CZ-182 08, Praha 8, Czech Republic
Tel. : +420 26605 2417
Fax: +420 26605 2417
E-mail: houda@ef.jcu.cz

of the data matrix T is known and fixed in advance, (1) is a standard linear programming problem. In order to deal with uncertainty of the data matrix, several approaches were developed in the literature, e.g., classical sensitivity analysis, parametric programming, or robust optimization methods. In this talk, we concentrate on *stochastic chance-constrained programming* approach. We assume that T is a random matrix with a known distribution. Additionally, the uncertain constraints of the problem (1) are required to be satisfied with a known fixed probability level $p \in [0;1]$. The *linear chance-constrained problem with random matrix* can be written as

$$\min c^T x \text{ s.t. } \Pr\{Tx \leq h\} \geq p,\ x \in X. \tag{2}$$

2 Dependence and copulas

Dependence structure of random vector has been described for years in probability theory and mathematical statistics. In this paper, this notion will be used in order to characterize the dependence between constraint rows of the problems (2).

In this section, some basic facts about copulas needed for our following investigation are mentioned. We refer the readers to [5] for a complete information on this theory.

Definition 1 A *copula* is the distribution function $C : [0;1]^K \to [0;1]$ of some K-dimensional random vector whose marginals are uniformly distributed on [0;1].

Proposition 1 (Sklar's theorem) *For any K-dimensional distribution function $F : \mathbb{R}^K \to [0;1]$ with marginals $F_1, \ldots, F_K$, there exists a copula C such that*

$$\forall z \in \mathbb{R}^K \quad F(z) = C(F_1(z_1), \ldots, F_K(z_K)). \tag{3}$$

If, moreover, F_k are continuous, then C is uniquely given by

$$C(u) = F(F_1^{(-1)}(u_1), \ldots, F_K^{(-1)}(u_K)). \tag{4}$$

Otherwise, C is uniquely determined on range $F_1 \times \cdots \times$ range F_K.

This theorem provides a direct link between a copula and the joint distribution function of a given random vector. Thanks to Sklar's theorem, the copula represents a convenient tool to describe the dependence structure of the random vector considered. Moreover, this description separates from description of the marginal distribution functions of the vector.

3 Our results

In this paper, we investigate the problem of linear joint probabilistic constraints (2). We assume that the rows of the constraint matrix are dependent, and the dependence is handled by a convenient Archimedean copula [4]. Further, we assume the distribution of the constraint rows to be elliptically distributed, covering normal, t, or Laplace distributions [3]. Under these and some additional conditions, we prove the convexity of the investigated set of feasible solutions. We also develop two approximation schemes for this class of stochastic programming problems based on second-order cone programming, which provides lower and upper bounds [1,2]. Finally, a numerical study on randomly generated data is given to illustrate the tightness of these bounds.

4 Conclusions

This paper shows the efficiency of using copulas for dealing with the random matrix rows dependence. The conic approximations allow to come up with numerical results with applications perspectives in many topics including energy management. Further research will extend this work to other copulas, and to more efficient conic approximations, namely copositive reformulations together with tight relaxations.

Acknowledgements This research benefited from the support of the FMJH Program Gaspard Monge in optimization and operation research, and from the support to this program from EDF.

References

1. Jianqiang Cheng and Abdel Lisser. A second-order cone programming approach for linear programs with joint probabilistic constraints. *Operations Research Letters*, 40(5):325–328, 2012.
2. Jianqiang Cheng and Abdel Lisser. A completely positive representation of 0-1 linear programs with joint probabilistic constraints. *Operations Research Letters*, 41(6):597–601, 2013.
3. M. Houda. Role of dependence in chance-constrained and robust programming. *Acta Oeconomica Pragensia*, 15(4):111–120, 2007.
4. Michal Houda and Abdel Lisser. Second-order cone programming approach for elliptically distributed joint probabilistic constraints with dependent rows. Research Report 1566, Laboratoire de Recherche en Informatique, Université Paris Sud XI, October 2013.
5. Roger B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2nd edition, 2006.