Scale-free modeling and optimization techniques for control of complex networks
Tehnici scalabile de modelare si optimizare pentru controlul sistemelor complexe de tip retea

(ScaleFreeNet)

Scientific Report 2018
Scalable algorithms for modeling and optimization of complex network systems
(Algoritmi scalabili de modelare si optimizare pentru sisteme complexe de tip retea)

Ion Necoara, Lucian Toma, Tudor Ionescu, Andrei Patrascu, Daniela Lupu

December 2018
Abstract

The main objective of this phase (31 December 2017 - 31 December 2018) was to Develop New Scalable Algorithms For Modeling And Optimization Of Complex Network Systems. Therefore, in this report we briefly present the most important scientific results obtained by the ScaleFreeNet team related to model reduction of highly dimensional systems and optimization of complex systems. The main activities for the second phase were:

A 2.1: Develop modeling algorithms for complex systems
A 2.2: Develop optimization algorithms for complex systems

Expected results in phase II: 2 ISI journal papers.

Achieved results in phase II:

- 2 ISI journal papers accepted and 1 ISI journal paper provisionally accepted
- 4 papers submitted to journals
- 3 papers in conferences and 3 papers submitted to conferences
- 1 Matlab optimization toolbox

More specifically, the research of the ScaleFreeNet team in the second phase (31 December 2017 - 31 December 2018) of the project led to the publication/submission of the following research papers:

Journal papers (ISI) accepted


Journal papers (ISI) submitted


Conference papers accepted/submitted


Toolbox

[T1]: Primal-Dual toolbox for Support Vector Machine (PD-SVM): Matlab code optimization toolbox for solving large-scale Support Vector Machine (SVM) problems (by D. Lupu and I. Necoara). This toolbox is partially based on paper [C1].

Besides the above accepted/submitted papers the ScaleFreeNet team was also working in this second phase at the following papers (hopefully, some of them will be finished at the end of this year):

Papers in preparation


In the next sections we briefly present the main results obtained by our team in the second phase of the project (31 December 2017 - 31 December 2018). For a detailed exposition of the theoretical results from the journal papers (accepted or submitted) you can consult arXiv or the webpage of the project ScaleFreeNet (http://acse.pub.ro/person/ion-necoara/project), from where these papers can be downloaded. You can also contact directly the director of this project, Ion Necoara, at email: ion.necoara@acse.pub.ro.

ScaleFreeNet team: Prof. Ion Necoara, Assoc. prof. Lucian Toma, Assoc. prof. Tudor Ionescu, Dr. Andrei Patrascu, Msc. Daniela Lupu.
## Contents

1. A2.1: Modeling algorithms for complex systems .......................... 5
2. A2.2: Optimization algorithms for complex systems .................. 8
3. Toolbox .............................................................................. 12
4. Papers acknowledging ScaleFreeNet project ........................... 13
   4.1 Papers published in ISI journals ....................................... 13
   4.2 Papers under review in ISI journals ................................. 13
   4.3 Papers accepted/submitted to conferences .......................... 13
   4.4 Papers in preparation ..................................................... 14
   4.5 Toolbox ........................................................................ 14
1 A2.1: Modeling algorithms for complex systems

The research of ScaleFreeNet team led to the following publications in 2018 related to the first activity of the second phase (A2.1-modeling algorithms for complex systems):


Next, we briefly explain the main results obtained by our team in the papers listed above related to activity A2.1.

[J1] & [C1] Mathematical tools yield complex and highly dimensional dynamical models from, e.g., partial-differential equations or networks of interconnected subsystems. Hence, for purposes such as simulation and control design, scientists and engineers need model reduction to render the systems simpler and useful. The main idea of model order reduction is to find a low-order mathematical model that approximates the given highly dimensional dynamical system. The approximation yielded through model reduction is accurate if the approximation error is small and if the most important physical and/or structural properties, such as the stability are preserved.

In papers [J1] & [C1] we write families of reduced order models that match a prescribed set of $\nu$ moments of a highly dimensional linear time-invariant system. First, we fully parametrize the models in the interpolation points and in the free parameters, and then we fix the set of interpolation points and parametrize the models only in the free parameters. Based on these parametrizations and using as objective function the $H_2$-norm of the approximation error we derive nonconvex optimization problems, i.e., we search for the optimal free parameters and eventually the interpolation points to determine the approximation model yielding the minimal $H_2$-norm error. For these optimization problems we compute the necessary first-order optimality conditions given in terms of the controllability and the observability Gramians of a minimal realization of the error system. Furthermore, using these optimality conditions, we propose several gradient-type methods for solving our optimization problems, with mathematical guarantees on their convergence. We also provide convex SDP relaxations for these nonconvex optimization problems and analyze when the relaxations are exact. More precisely:
We first formulate a general model reduction problem with reduced models from the family of models matching \( \nu \) moments parameterized in the interpolation points and in the free parameters. A corresponding optimization formulation is derived, where the objective function is the \( H_2 \)-norm of the approximation error, written explicitly in terms of the controllability and observability Gramians of a minimal realization of the error system. We also write the necessary first-order optimality conditions (KKT system) of this optimization problem, based on these Gramians.

For this general model reduction problem we propose several numerical optimization algorithms. The first method is using a gradient update for solving the KKT system, leading to a simple iteration involving matrix multiplications. However, with this update the stability of the approximation is achieved asymptotically. The second solution is based on a partial minimization approach. We show that for the evaluation of the gradient of the objective function we need to solve two Lyapunov equations yielding the Gramians, but the gradient is Lipschitz continuous. Therefore, a gradient-based algorithm is developed, ensuring convergence due to the smoothness of the objective function. Although the gradient evaluation is expensive, each iteration provides a stable reduced order model, whereas the first method yields a stable reduced order model only asymptotically. Finally, we propose a convex SDP relaxation of the original optimization problem and write sufficient conditions when the relaxation is exact. Note that the interpolation points obtained are the spectrum of a squared matrix computed by each of these algorithms.

We also consider a relaxed version of the general model reduction problem, searching only for the free parameters that yield the optimal reduced order model from the family of models matching \( \nu \) moments at fixed interpolation points. Optimization formulations for this particular problem are also derived and subsequently the previous numerical optimization algorithms can be applied to solve this relaxed problem as well, with similar convergence guarantees. At this point, we want to emphasize that our optimization approach is very flexible, allowing to easily incorporate in the optimization formulations physical and/or structural constraints. Finally, we illustrate the efficiency of our results on several test problems.

[\textit{J2}] & [\textit{C2}] Finite-impulse response filters (FIR) often appear in a wide variety of systems and control applications. They are present, among other things, in dead-time compensators for processes with dead-time, in particular as a part of the finite-spectrum assignment control law or modified Smith predictor.

In papers [\textit{J2}] & [\textit{C2}] we present two moment matching-based solutions to the problem of finding a rational approximation of a FIR system: a direct moment matching approximation performed on the FIR system, and a sequential moment matching approximation performed on a block split model of a FIR system. The first approach stems from recent time domain moment matching results. Imposing moment matching conditions yields families of parametrized, low-order, rational approximations of the FIR filters, with guaranteed low frequency behavior. The parameters of the resulting families of models may be tuned such that stability is achieved. Furthermore, we show how to select the free parameters and calculate the (unique) model that achieves moment matching at both zero and first order derivatives of the transfer function of the given system, which is a member of the family of models that match \( \nu \) moments of the given FIR system. Note that matching first order derivatives is used for first order optimality conditions of the optimal \( H_2 \)-norm model order reduction problem. The second approach is performed on a block split model of the FIR system. We perform component-wise moment matching, such that the FIR may be approximated by a chain of \( N \) linear, time-invariant systems of order \( \nu \).
We show that existing rational implementations are particular instances of the results yielded by the proposed sequential moment matching approach. The theory is illustrated with an example of a FIR system.

In the direct moment matching approximation approach we establish a relation between the moments of the FIR system and the unique solutions of Sylvester equations. Imposing moment matching conditions yields families of parametrized, low-order, rational approximations of the FIR filters, with guaranteed low frequency behavior. We also compute the rational approximations. Although the approximant is delay-free, the effect of the delays is indirectly present in the moment matching property. Furthermore, we show explicitly how to select the free parameters to obtain the (unique) model that achieves moment matching at both zero and first order derivatives of the transfer function of the given FIR system, which is a member of the family of models that match the moments of the given FIR system. Note that the idea of matching first order derivatives is used for first order optimality conditions of the optimal 2-norm model order reduction problem, i.e., interpolate the mirror images of the poles of the approximant, as is the case of the Iterative Rational Krylov Algorithm (IRKA). From a computational point of view, the solutions of the Sylvester equations, required for the explicit formulation of the free parameters when derivatives are matched, can be obtained from the moments, i.e., evaluations of the given system in the chosen interpolation points. We give an illustrative example, comparing (in terms of the 2/1-norms of the approximation error) the proposed moment matching-based approximation with the Padé approximation-based method, the rational implementation and the rational approximation yielded by the Loewner framework.

In the sequential moment matching approximation approach we perform on a block split model of the FIR system. We perform component-wise moment matching, such that the FIR may be approximated by a chain of $N$ linear, time-invariant systems of order $\nu$. We show that the previous implementations from the literature are particular instances of the results yielded by the proposed sequential moment matching approach. Finally, we illustrate our theoretical results on a numerical example.

\[P1\] This paper is in preparation and we hope to finish it by the end of this year. In this paper we extend the framework developed in the journal paper \[J1\] (detailed above) to linear network systems. In particular, we pose an optimal $H_2$-norm model reduction problem using moment matching parameterizations for linear network systems. We prove that we can obtain a reduced order model having the same network structure as the original network system by solving (non)convex optimization problems with convex constraints to impose the network structure. Consequently, we either need to use a projected gradient algorithm or we need to solve a semidefinite programming problem. We also illustrate the efficiency of this approach on positive network systems.
2 A2.2: Optimization algorithms for complex systems

The extensive research on optimization made by the ScaleFreeNet team in the second phase of the project also led to the following publications related to second activity (A2.2-optimization algorithms for complex systems):


Next, we briefly explain the main results obtained by our team in the papers listed above and related to activity A2.2.

[J1] The standard assumption for proving linear convergence of first order methods for smooth convex optimization is the strong convexity of the objective function, an assumption which does not hold for many practical applications. In this paper, we derive linear convergence rates of several first order methods for solving smooth non-strongly convex constrained optimization problems, i.e. involving an objective function with a Lipschitz continuous gradient that satisfies some relaxed strong convexity condition. In particular, in the case of smooth constrained convex optimization, we provide several relaxations of the strong convexity conditions and prove that they are sufficient for getting linear convergence for several first order methods such as projected gradient, fast gradient and feasible descent methods. We also provide examples of functional classes that satisfy our proposed relaxations of strong convexity.
conditions. Finally, we show that the proposed relaxed strong convexity conditions cover important applications ranging from solving linear systems, Linear Programming, and dual formulations of linearly constrained convex problems.

[J2] A popular approach for solving stochastic optimization problems is the stochastic gradient descent (SGD) method. Although the SGD iteration is computationally cheap and its practical performance may be satisfactory under certain circumstances, there is recent evidence of its convergence difficulties and instability for unappropiate choice of parameters. To avoid some of the drawbacks of SGD, stochastic proximal point (SPP) algorithms have been recently considered. We introduce a new variant of the SPP method for solving stochastic convex problems subject to (in)finite intersection of constraints satisfying a linear regularity condition. While most of the existing papers from the stochastic optimization literature consider convex models without constraints or simple (easy projection onto) constraints, in this paper we consider stochastic convex optimization problems subject to (in)finite intersection of constraints satisfying a linear regularity type condition. It turns out that many practical applications, including those from machine learning, fits into this framework: e.g. classification problems, regression problems, finite sum problems, portfolio optimization problems, convex feasibility problems, optimal control problem, etc. For this general stochastic optimization model we introduce a new stochastic proximal point (SPP) algorithm, for which we prove new nonasymptotic sublinear convergence results depending on the properties that we impose on the objective function: convex vrs. strongly convex objective function, objective function with bounded (sub)gradients vrs. Lipschitz continuous gradients. It is important to note that the derived rates of convergence do not contain any exponential term as it is the case for the SGD scheme, which makes SPP more robust than SGD even in the constrained case. Then, since the best complexity of our basic SPP scheme can be attained only under some natural restrictions on the initial stepsize, we also introduce a restarting stochastic proximal point algorithm that overcomes these difficulties and derive the corresponding nonasymptotic convergence rates. The main advantage of this restarted variant of SPP algorithm is that it is parameter-free and thus it is easily implementable in practice. Numerical evidence supports the effectiveness of our methods in real problems.

[J3] In this paper we analyze new equivalent stochastic reformulations of the convex feasibility problem that are governed by an arbitrary discrete or continuous variable over some probability space. In particular, our reformulations can be seen as stochastic (non)smooth optimization problem, stochastic fixed point problem or stochastic intersection problem. By combining these reformulations with certain regularity assumptions on the individual sets we extend the concept of condition number from convex optimization to convex feasibility. Based on these reformulations and the new characterization of the conditioning parameters we introduce a general random batch projection algorithmic framework with an extrapolated stepsize, which generates new algorithms or extends to random settings many existing projection schemes. Our method allows also to project simultaneously on several sets, thus providing great flexibility in matching the implementation of the algorithm on the parallel architecture at hand. Finally, we derive asymptotic convergence results and also (sub)linear convergence rates for this general algorithmic framework that depend explicitly on the conditioning parameters and the number of projections computed at each iteration. Moreover, our convergence rates explain when extrapolation works. Preliminary numerical results also show a better performance of our extrapolated stepsize scheme over its constant stepsize counterpart.
In this paper we introduce new global and local inexact oracle concepts for a wide class of convex functions in composite convex minimization. Such inexact oracles naturally come from primal-dual framework, barrier smoothing, inexact computations of gradients and Hessian, and many other situations. We also provide examples showing that the class of convex functions equipped with the newly inexact second-order oracles is larger than standard self-concordant as well as Lipschitz gradient function classes. Further, we investigate several properties of convex and/or self-concordant functions under the inexact second-order oracles which are useful for algorithm development. Next, we apply our theory to develop inexact proximal Newton-type schemes for minimizing general composite convex minimization problems equipped with such inexact oracles. Our theoretical results consist of new optimization algorithms, accompanied with global convergence guarantees to solve a wide class of composite convex optimization problems. When the first objective term is additionally self-concordant, we establish different local convergence results for our method. In particular, we prove that depending on the choice of accuracy levels of the inexact second-order oracles, we obtain different local convergence rates ranging from linear and superlinear to quadratic. In special cases, where convergence bounds are known, our theory recovers the best known rates. We also apply our settings to derive a new primal-dual method for composite convex minimization problems. Finally, we present some representative numerical examples to illustrate the benefit of our new algorithms.

In this paper we consider large-scale smooth optimization problems with multiple linear coupled constraints. Due to the non-separability of the constraints, arbitrary random sketching would not be guaranteed to work. Thus, we first investigate necessary and sufficient conditions for the sketch sampling to have well-defined algorithms. Based on these sampling conditions we developed new sketch descent methods for solving general smooth linearly constrained problems, in particular, random sketch descent and accelerated random sketch descent methods. From our knowledge, this is the first convergence analysis of random sketch descent algorithms for optimization problems with multiple non-separable linear constraints. For the general case, when the objective function is smooth and non-convex, we prove for the non-accelerated variant sublinear rate in expectation for an appropriate optimality measure. In the smooth convex case, we derive for both algorithms, non-accelerated and accelerated random sketch descent, sublinear convergence rates in the expected values of the objective function. Additionally, if the objective function satisfies a strong convexity type condition, both algorithms converge linearly in expectation. In special cases, where complexity bounds are known for some particular sketching algorithms, such as coordinate descent methods for optimization problems with a single linear coupled constraint, our theory recovers the best-known bounds. We also show that when random sketch is sketching the coordinate directions randomly produces better results than the fixed selection rule. Finally, we present some numerical examples to illustrate the performances of our new algorithms.

Machine learning tools are become recently very popular for solving real applications from many areas. Most of the learning problems are formulated as optimization problems with simple objective function but large number of constraints of order the number of training data. When considering the dual formulation, usually the objective function is difficult to minimize but the constraints are simple. One relevant application that fits into this pattern is the support vector machine (SVM). A popular approach for solving the primal SVM problem is based on first order methods due to their superior
empirical performance. When considering the dual SVM formulation, which has simple constraints, coordinate descent schemes are typically the method of choice in practice due to their cheap iteration. In this paper we present a comparative study of several first order methods for solving primal or dual SVM problems. Numerical evidence on support vector machine classification for automatic detection of driver fatigue supports the effectiveness of such first order methods in real-world problems.

[C3] In this paper we derive gradient methods with random projections for solving constrained convex optimization problems, where the constraints are described as the intersection of a finite number of convex sets. Each set of the intersection is assumed to be simple, that is the projection onto each set can be computed efficiently. Our algorithms can easily address streaming settings, where the whole feasible set is not known in advance, but it is rather learned in time through observations, as in machine learning applications. Also, the algorithms are of interest for optimization problems having a large number of given constraints, as in constrained predictive control. We analyze the convergence behavior of the proposed algorithms for the case when the objective function is strongly convex and with Lipschitz continuous gradients. We prove linear convergence rate into a noise dominated region for the expected quadratic distance of the iterates from the optimal set for constant stepsize. However, our analysis allows to also provide global sublinear rates for diminishing stepsize.
3 Toolbox

The research made by the ScaleFreeNet team in the second phase of the project also led to a toolbox:


Description of the toolbox [T]: In learning applications the optimization algorithms involve numerical computation of parameters for a system designed to make decisions based on large amount of data. In particular, in linear SVM the goal is to find a hyperplane that best divides a dataset into two classes. This leads to an optimization problem with simple objective function, e.g. a quadratic expression with diagonal Hessian, but a large number of linear constraints, equal the number of training data. When considering the dual formulation, usually the objective function is difficult to minimize, but the constraints are simple. The recent success of some first order optimization methods for SVM has motivated increasingly great efforts into developments of new numerical algorithms or into efficient implementation of existing ones.

The optimization algorithms we implement in this toolbox, called (PD-SVM), use first order information combined with random choice of the sets or of the coordinates. More precisely, for solving the primal SVM problem we use conditional gradient. Moreover, instead of dealing with the whole set of constraints at each iteration as conditional gradient does, we also implement a stochastic gradient variant that uses only one constraint randomly per iteration. When considering the dual SVM formulation, which has simple constraints, we first implement projected gradient with an efficient routine for solving the projection and also a two coordinate gradient descent algorithm. More details on these primal or dual first order schemes can be found in the paper D. Lupu, I. Necoara, Primal and dual first order methods for SVM: applications to driver fatigue monitoring, in Proceedings of International Conference on System Theory, Control and Computing, 2018, or in the documentation of the toolbox which can be downloaded freely from: http://acse.pub.ro/person/ion-necoara.

The toolbox PD-SVM has a friendly interface created in Matlab2016. In the input section, load data button asks the user to defines his data of the problem through a file of type .mat. All the optimization algorithms depend on 3 parameters (maximum number of iterations, accuracy of the solution, and the penalty parameter) which the user can keep at some already defined default values or change them accordingly. After giving all the input data, the user can run all the four primal/dual first order methods described above by selecting the corresponding button. For all methods, after finding the optimal solution, the user is asked if he wants to save it and where. At the same time, in the outputs panel, some information is displayed: the number of iterations and the running time.
4 Papers acknowledging ScaleFreeNet project

All the journal papers (accepted or submitted) can be downloaded from arXiv and also from the webpage of the project ScaleFreeNet (http://acse.pub.ro/person/ion-necoara/project). The toolbox can be also downloaded from the webpage of the project (http://acse.pub.ro/person/ion-necoara/project).

4.1 Papers published in ISI journals


4.2 Papers under review in ISI journals


4.3 Papers accepted/submitted to conferences


4.4 Papers in preparation


4.5 Toolbox