Optimization with State Space Approaches I: A System Theoretic View on Optimization Algorithms

G. Steiner and M. Neumayer
Institute of Electrical Measurement and Measurement Signal Processing
Kopernikusgasse 24/IV, A-8010 Graz, Graz University of Technology, Austria
Email: gerald.steiner@tugraz.at

***Abstract*—The number of optimization algorithms has seen a rigorous growth in the past years, leading to a variety of different algorithms. In system and control theory, state space models are a unified basis of further analysis and control of dynamic systems. In this work we show how optimization problems and algorithms can be generically modeled in the unified state space framework. This approach facilitates the comparison among different algorithms and the design of novel hybrid optimization schemes. As a further key advantage it opens way for the application of powerful Bayesian algorithms for the solution of optimization problems. This includes the active consideration of the discretization error as well as additional information about the quality of obtained solutions.**

I. Introduction

Optimization can generally be referred to as solving a mathematical problem of the kind

$$\boldsymbol{x}^* = \arg\min \Psi(\boldsymbol{x}) \quad (1)$$
$$\text{s.t. } \boldsymbol{C}(\boldsymbol{x}) \leq \boldsymbol{0}, \quad (2)$$

where $\Psi : \mathbb{R}^N \rightarrow \mathbb{R}^1$ is called the objective function, the vector $\boldsymbol{x} \in \mathbb{R}^N$ contains the variables of interest and $\boldsymbol{C}(\boldsymbol{x})$ is a vectorial function which expresses possible constraints on $\boldsymbol{x}$. An enormous variety of algorithms for the solution of the optimization problem 2 is available, where methods can be grouped in deterministic and stochastic methods. Deterministic methods most often make use of gradient and curvature information in order to efficiently detect the minimum. Stochastic methods rely on some sort of randomness to sample the parameter space in search for the minimum. Of course, hybrid algorithms have been proposed to combine the advantages of the two classes of algorithms.

Classical first and second order deterministic methods try to minimize Ψ by determining a descent direction out of gradient or gradient and Hessian information of Ψ. The steepest descent method, e.g., calculates consecutive steps using

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - s\boldsymbol{g}(\boldsymbol{x}_k), \quad (3)$$

where $\boldsymbol{g}$ denotes the gradient $\nabla\Psi$ and s is the step width. More elaborate methods calculate Newton steps and make use of line search techniques [1].

With the availability of increased computational power, stochastic optimization methods have become popular. Stochastic methods explore the search space $\mathbb{R}^N$ for a minimum by evaluating Ψ for "randomly" drawn vectors $\boldsymbol{x}_i$. As these methods are computationally heavy, efficient schemes for the generation of candidate solutions are necessary. Several effects from nature and biology were adapted to design such methods. Prominent examples are methods like Particle Swarm Optimization (PSO) [2], Differential Evolution (DE) [3], Genetic Algorithms (GA), and ant colony optimization. In the presence of the enormous variety of differently labeled stochastic algorithms it is often hard to distinguish the differences, and more important, to rate the efficiency and suitability of methods for certain applications. In this paper, which is the first part of two contributions on the use of state space approaches for optimization and the similarities between the two disciplines, we want to present the idea of a system theoretic framework for optimization algorithms as well as the use of ideas of state estimation for optimization. Numerical examples demonstrating the power of this concept and the application of sequential Bayesian estimation algorithms are presented in [4].

II. System theory

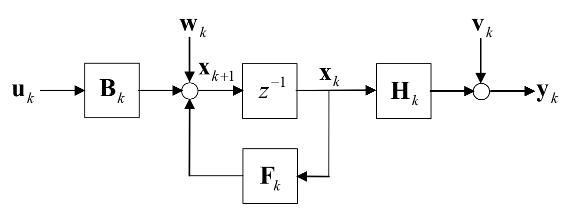

Fig. 1. Block diagram of a nonlinear time variant discrete time state space system.

In system and control theory, dynamic systems are conveniently described using state space models. A state space model can be formulated in continuous or discrete time and consists of internal state variables, measurable output quantities, and optional input quantities. The models are commonly formulated using systems of first order differential or difference equations. A general nonlinear time variant stochastic discrete time state space system is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}_k(\boldsymbol{x}_k) + \boldsymbol{B}_k(\boldsymbol{u}_k) + \boldsymbol{w}_k \quad (4)$$
$$\boldsymbol{y}_k = \boldsymbol{H}_k(\boldsymbol{x}_k) + \boldsymbol{v}_k, \quad (5)$$

where (4) is the state evolution equation with $\boldsymbol{F}$ describing the evolution of the state vector $\boldsymbol{x}$ from time step k towards time step $k+1$. The state is subjected by the input variables $\boldsymbol{u} \in \mathbb{R}^N$ via the input function $\boldsymbol{B} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ and a random noise process $\boldsymbol{w}$ with arbitrary probability distribution. If the system is an autonomous system, the input function $\boldsymbol{B}$ vanishes. In general, the state $\boldsymbol{x}$ is not directly observable. The measurable quantities $\boldsymbol{y}$ are linked to the system state in the measurement equation (5) by the function $\boldsymbol{H}$. The measurement is affected by measurement noise $\boldsymbol{v}$. The block diagram of the state space system is shown in Fig. 1. One of the most important problems in conjunction with state space systems is the estimation of the unknown system state $\boldsymbol{x}$ given the noisy observations $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}, \boldsymbol{y}_k\}$ and available prior knowledge about the system [5]. Thus, it is essentially an optimization problem where some suitable norm of the estimation error $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ with the state estimate $\hat{\boldsymbol{x}}$ is minimized. Powerful state estimation methods have been developed, including the Luenberger observer, Kalman filter, H-infinity filter, unscented filter, particle filter, and Markov chain Monte Carlo methods.

III. Optimization with state space models

From a system theoretic view, iterative optimization algorithms may be regarded as time discrete dynamical systems. At every iteration or time step, respectively, a new candidate solution or population of candidate solutions is generated based on calculations, heuristics, or randomness.

Deterministic optimization algorithms in state space form can be derived by introducing feedback from the output to the input with a suitable control law. In this case a linear state equation with $\boldsymbol{F}_k(\boldsymbol{x}_k) = \boldsymbol{I}\boldsymbol{x}_k$, $\boldsymbol{B}_k(\boldsymbol{u}_k) = \boldsymbol{I}\boldsymbol{u}_k$, and $\boldsymbol{w}_k = \boldsymbol{0}$, with the identity matrix $\boldsymbol{I}$ can be defined. It should be noted that the stochastic noise contribution vanishes for deterministic algorithms. The measurement equation is composed of the objective function and a noise contribution that can be attributed to discretization and measurement noise,

$$\boldsymbol{y}_k = \Psi(\boldsymbol{x}_k) + \boldsymbol{v}_k, \qquad (6)$$

Using the feedback law $\boldsymbol{u}_k = -s\boldsymbol{g}(\boldsymbol{x}_k)$, the steepest descent method is easily obtained. Similarly, other optimization algorithms can be formulated as a dynamic system control problem.

Stochastic optimization can be formulated in state space by setting the deterministic input vector $\boldsymbol{u}_k = \boldsymbol{0}$ and the process noise $\boldsymbol{w}_k \sim p(\boldsymbol{0}, \boldsymbol{Q}_k)$ with a suitably chosen probability distribution p, leading to an autonomous system. Please note that this input wiring is converse to the deterministic case. Now the optimization task is essentially a state estimation problem. The hidden states of the system given the optimal and actually observed objective function values need be recovered from noisy observations. Both single point and population-based methods can be modeled with this approach. An advantage of a state space formulation is that it enables the use of powerful stochastic state estimation algorithms like the Kalman filter, particle filter, or sequential Monte Carlo methods.

With the two orthogonal state space formulations for deterministic and stochastic optimization methods, the design of well-balanced hybrid optimization algorithms is facilitated due to the simple system structure. In addition, two major aspects build a general difference between classical optimization and state space approaches:

- Measurement Noise is taken explicitly into concern.
- Prior knowledge is extensively used.

While prior knowledge can be directly interpreted as a form of constraints, measurement noise is a less used term with respect to optimization. In this context, it can be interpreted as modeling or discretization error of the computer model. Recently the use of reduced order models has gained in interest for optimization [6]. However, neglecting this error may result in deviated results for the real problem. Prior information about the measurement/model error can be used to overcome adverse effects caused by reduced models, e.g. by applying the enhanced error model [7]. This allows to take advantage of the decreased computation time of simplified models, but without losing quality in the result.

IV. Conclusion

It was demonstrated that optimization algorithms, both deterministic and stochastic, can be formulated within a system theoretic state space approach. While deterministic algorithms correspond to deterministic feedback control, stochastic algorithms are related to state estimation problems. Both approaches can be combined to obtain hybrid optimization algorithms. Key advantages of the state space formulations are the explicit consideration of modeling and measurement errors, the possibility to make use of available prior knowledge, as well as the access to powerful control and state estimation algorithms. The mentioned issues will be elaborated and advantages discussed in more detail in the full paper. The practical application of the presented framework will be presented in a companion paper [4] by means of sample algorithms and an optimization example.

References

[1] R. Fletcher, *Practical Methods of Optimization; (2nd Ed.)*, Wiley-Interscience, New York, USA, 1987.

[2] L. dos Santos Coelho and P. Alotto, *Global Optimization of Electromagnetic Devices Using an Exponential Quantum-Behaved Particle Swarm Optimizer*, IEEE Transactions on Magnetics, vol.44, no.6, pp.1074-1077, June 2008.

[3] R. Storn and K. Price, *Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces*, Journal of Global Optimization 11: pp.341-359, 1997.

[4] M. Neumayer and G. Steiner, *Optimization with State Space Approaches II: System Theoretic Concepts per Example*, submitted to 18^{th} International Conference on the Computation of Electromagnetic Fields, COMPUMAG 2011.

[5] M.S. Grewal and A.P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB; (2nd Ed.)*, John Wiley-Interscience, New York, USA, 2001.

[6] Albunni, M.N.; Rischmuller, V.; Fritzsche, T.; Lohmann, B.; , *Multiobjective Optimization of the Design of Nonlinear Electromagnetic Systems Using Parametric Reduced Order Models*, IEEE Transactions on Magnetics, vol.45, no.3, pp.1474-1477, March 2009.

[7] J. P. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, New York: Applied Mathematical Sciences, Springer, 2004.