EDUCATION

The Education section of *SIAM Review* presents two papers in this issue.

In the first paper Gábor Pataki focuses on "Characterizing Bad Semidefinite Programs: Normal Forms and Short Proofs." Semidefinite programs (SDPs) are optimization problems where the decisions are semidefinite matrices, entering linearly into the objective and constraint functions. These types of optimization problems are widespread and have enjoyed a lot of interest in the last three decades, providing new tools in combinatorial optimization, in polynomial optimization, and in many engineering and economics problems.

Duality theory plays a fundamental role in optimization, and strong duality results are available in semidefinite programming as part of convex optimization with additional structure. However, SDPs exhibit different behavior than linear optimization problems, and the dual pairs, as defined in linear programming, do not behave in the same way. For example, the optimal values of the primal and dual SDPs may differ and may not be attained at the same time. The author calls those SDPs pathological. He points out that pathological SDPs are often difficult or even impossible to solve.

The paper contains a discussion on necessary and sufficient conditions for "good" or "bad" behavior of an SDP pair. The problems are reformulated into normal forms that make their behavior easy to identify. The normal forms are obtained by linear algebra tools such as Gaussian elimination and are akin to the row echelon form of a linear system of equations. The set of positive semidefinite matrices of a given dimension forms a closed convex cone $\mathcal{S}^n_+$, which is frequently referred to as the semidefinite cone. Part of the discussion ties the behavior of the SDPs to the closeness of the image of $\mathcal{S}$ under certain associated linear maps. Next to formal arguments, the discussion is supplemented by examples and geometric illustrations.

The paper also contains literature pointing to many engineering applications, connections to combinatorics and algebraic geometry, as well as further information on SDP duality, stability and sensitivity analysis, and other aspects of semidefinite programming.

The second paper is "Deep Learning: An Introduction for Applied Mathematicians," by Catherine Higham and Desmond Higham. It delves into an area attracting a lot of attention and which is the subject of many papers and lectures and much software development. The availability of high power computing capabilities and the universal character of artificial neural networks has resulted in many success stories. Deep learning is applied in a vast range of fields where image recognition, speech recognition, and natural language processing are necessary. Deep learning techniques entail solving large nonlinear and nonconvex optimization problems for the purpose of parametric or semiparametric model fitting and for online recommendations. Identifying the statistical model is referred to as training the neural net. Then the model is used to predict (classify) future instances.

The paper starts with a simple example which provides a gentle introduction to the structure of artificial neural networks and to the mathematical issues occurring in training them. A more involved example of image classification is discussed as illustration later on. The authors present a formal description of a network with multiple layers. The notion of a cost function is introduced: the objective to be minimized in order to determine the desired parameters and its role is discussed. Other important ingredients of a neural net are the activation functions of the nodes; they determine the node's output. A substantial part of the paper explains the operation of the stochastic subgradient method, including a pseudocode for training a network. The authors point out the danger of statistical overfitting, which occurs when the model fits the noisy training data too well and does not generalize well to new data. The final section of

Downloaded 11/18/19 to 138.195.214.230. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

the article points to various relevant questions which had not been reflected upon in the presentation, including how to decide on the structure and the depth (number of layers) of the neural net, determining when the stochastic subgradient method converges, the use of regularization terms in the cost function, and others.

Original MATLAB code associated with the article is provided as supplementary material. The article also contains references and links to other publicly available code and data.

Darinka Dentcheva
Section Editor
darinka.dentcheva@stevens.edu

Downloaded 11/18/19 to 138.195.214.230. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.