Nonlinear Programming Models of Power-Aware Scheduling and Cloud Computing

Akiyoshi Shioura[1], Natalia V. Shakhlevich[2], and Vitaly A. Strusevich[3]

[1] Tokyo Institute of Technology, Japan,
[2] University of Leeds, United Kingdom,
[3] University of Greenwich, United Kingdom

In the off-line speed scaling problems of power-ware scheduling it is required to determine the processing speed of each job either on a single machine or on parallel machines. The speeds are selected in such a way that (i) the cost of speed changing, often understood as energy needed to maintain a certain speed, is minimized, and (ii) the actual processing time of each job allows its preemptive processing within a given time window. We propose a new methodology for the speed scaling problem based on its link to scheduling with controllable processing times and submodular optimization. We reduce the speed scaling problems to a generic problem of minimizing a convex separable function over a base polyhedron and adapt a decomposition algorithm by Fujishige for its solution. This results in faster algorithms for traditional speed scaling models, characterized by a common speed/energy function. In addition, it handles efficiently the most general models with job-dependent speed/energy functions, which to the best of our knowledge have not been addressed prior to this study. In particular, the running time of the improved algorithm that solves the general version of the single-machine problem depends quadratically on the number of tasks to be scheduled.

Additionally, we address speed scaling problems that arise in cloud computing, where it is becoming a standard service level to start computing a customers task on its arrival, which is possible due to abundant resource of the cloud.