# The Mathematical Approach to Academic Scheduling:
## The Objective Function

by Richard D. Hamlin[1]

*This article introduces the Objective Function: a key concept to academic scheduling.*

**What is the objective function?** The objective function is the measure for comparing valid schedules. If two schedules have the same objective function value they are considered comparable (i.e., neither is better than the other). Otherwise, the schedule with the better objective function value is the better schedule.

In math terms, academic scheduling is combinatorial optimization. Most generally, the combinatorial optimization problem is to maximize the objective function value so that the state (the schedule) is in the feasible domain. In mathematical notation,

$$\max\ f(\mathbf{x})\ \ni\ \mathbf{x} \in X_{feas}, \tag{1}$$

where $\mathbf{x}$ is a vector, $[x_1, x_2, ..., x_m]$, representing the state (the schedule), $f(\mathbf{x})$ is the objective function, and $X_{feas}$ is the set of all schedules which are feasible (permissible, does not violate any constraints).

Typically, for academic scheduling the objective function primarily measures how well the schedule meets students' needs, and takes the form:

$$f(\mathbf{x}) = \sum_{i=1}^{n} w_i x_i = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n \tag{2}$$

where $x_i$ is a binary variable that indicates whether a particular student is enrolled into a particular class (1 if enrolled, 0 otherwise), $n$ is the number of such variables, and $w_i$ is the weighting factor that indicates how important that enrollment is. Some classes, such as academically required classes, are more important than, say, electives and thus receive a higher weight. Some electives may be more important than others. This is why merely minimizing the number of unsatisfied requests or the number of unsatisfied students is not actually usually desirable.

**Weights are relative** The absolute values of the weighting factors, $w$, do not matter, since a functionally equivalent weighting vector could be achieved by multiplying it by any scalar factor, $s$: $\mathbf{w}' = s\mathbf{w}$. What matters are all the relative weights $w_i/w_j$; e.g. if $w_1 = 2$ and $w_2 = 1$ or $w_1 = 4$ and $w_2 = 2$, they both yield the same ratio $w_1/w_2 = 2$. To assign weights, start with the typical class and nominally assign it a weight of one. All other classes are specified relative to that class. Students may also be given different weights, say for certain grade levels, to receive priority. Then the total weight of each request, $w_i$ has a student and a class factor, $w_{s,i}$ and $w_{c,i}$, respectively.

$$w_i = w_{c,i} w_{s,i} \tag{3}$$

**What is the importance of the objective function?** The objective function is highly important. It represents the utility of a schedule. Any true optimization algorithm must know exactly what to optimize.

In constructing the objective function, setting the appropriate weights, is critical. There are many possible measures of "schedule quality." Some of those only look at a particular desirable quality for a schedule but ignore other desirable qualities. The objective function, properly modeled, represents an appropriately weighted composite of all desired qualities.

**Where is the objective function?!** Not all academic schedulers actually use an objective function.

There are two fundamentally different ways in which to schedule.

1. One is to define a *scheduling problem*, a comprehensive mathematically rigorous criteria for the schedule consisting of an objective function and a set of constraints, and then solve it with
   (a) a complete algorithm (which is guaranteed to find the optimal solution).
   (b) an incomplete/heuristic algorithm.

[1] Richard Hamlin is the Founder and CEO of Automated Scheduling. He has a M.S. degree from UC-Berkeley and a S.B. degree from MIT in Mechanical Engineering.

2. The other is to define a *process* by which the schedule is produced.

In many programs which use the latter approach, you may also specify different priorities, but these may be used (either directly or possibly in conjunction with other factors) to determine the *order* in which classes are scheduled or which students are enrolled into what class sections. So, what is wrong with this? This approach is practically certain to produce suboptimal schedules. Whether or not you choose to specify it, there is an objective function (not necessarily linear) that accurately represents the school's priorities. Finding the global optimum value of that objective function will not be attained by a process that does not even evaluate the objective (except in the rarest of circumstances—really you have a much better chance of winning the lottery, so invest your money there if you are so lucky!). This is clear on a theoretical basis alone: since academic scheduling is an *NP-Complete* problem (a class of many of the most difficult problems). If a linear-time process solved the general academic scheduling optimization problem, it would mean *P=NP* and all the other "difficult" problems like the traveling sales person and cracking certain "strong" encryption could be solved easily by reformulating them as academic scheduling problems. Likewise, in practice, schedules thus produced typically have twice as many unsatisfied requests as the optima, sometimes many more (even 10 times more), and sometimes less.

Simply using an objective function by itself does not guarantee a better result. The quality of the algorithm and how well it works on the problem at hand is critical. By definition, the best results come from running a complete algorithm to completion. Since academic scheduling is such a difficult problem in general, it is not always practical to do that. Within the category of incomplete algorithms, there is a vast spectrum: some algorithms may get you very close to the theoretical optima, whereas in the other extreme, some especially poor heuristics could conceivably, on average, give no better results than techniques not using an objective function. (After all, an algorithm could potentially evaluate an objective function but not make any use of the value, except to report it.) The ability to evaluate the objective function gives even incomplete algorithms a great advantage (if not squandered). If any optimizing search is performed at all (as opposed to simply returning the first solution found), even an incomplete algorithm can be expected to produce better results than a process devoid of any objective function, because an incomplete algorithm can return the best solution from the many solutions it happens to run across.[2]

**Conclusion** Mathematically modeling the scheduling problem is a necessary step to produce optimized schedules. Accurately modeling the objective function is critical to producing the best results. Finding the optimal solution to the wrong problem, using the wrong objective function, will generally yield a solution that is not optimal for the right objective function. Further, if the modeling errors are too egregious, it can produce a bad schedule. So, it is important to take great care to set scheduling weights to appropriately reflect the school's priorities. In a future issue, we will discuss how to go about doing that.

[2] *N.B.:* Not all incomplete algorithms return the best solution uncovered. These should not be used!