Multi-Objective Optimization of Spectra Using Genetic Algorithms

Neil H. Eklund
eklund@acm.org
Oak Grove Scientific
11A Manchester Drive
Clifton Park, NY 12065

Mark J. Embrechts
embrem@rpi.edu
Department of Decision Sciences and Engineering Systems
Rensselaer Polytechnic Institute
Troy NY 12180

Abstract

This paper applies genetic algorithms (GAs), a powerful general-purpose biologically motivated optimization technique, to the multi-objective problem of spectrum optimization. Two objectives, color and efficiency, are address using real spectra, although the addition of other objectives (e.g., color rendering, color temperature) is relatively straightforward.

The direct application of the method presented is to transform the spectrum of newly developed lighting technologies to have desirable color properties while maximizing efficiency. Other applications of this methodology include the design of a filter for the input of a fiber optic system such that the color at then end of a given length of fiber has particular properties (e.g., appears “white”), while the efficiency of the system is minimally affected.

The principal findings described in this paper are the implementation of an efficient multi-objective fitness function tailored to this problem and a method for speeding convergence of the GA by "smoothing the chromosomes."

An algorithm, data and results from several approaches are presented.

Introduction

It is the goal of lighting manufacturers to produce light sources with maximum luminous efficacy, the ratio of the total luminous flux to total power input (i.e., "amount of light" per Watt). Luminous flux (F) is defined as:

$$F = 683\sum_{380}^{780} P_\lambda V_\lambda \Delta_\lambda$$

where P_λ is the spectral radiant flux (in Watts) of the light source at wavelength λ, V_λ is the photopic luminous efficiency function, and Δ_λ is the wavelength interval over which values of spectral radiant flux are evaluated[1]. Some typical values of luminous efficacy for the lamps considered in this paper are 107 lm/W for High Pressure Sodium (HPS), 107 lm/W for Metal Halide (MH), and 13 lm/W for an incandescent lamp.

However, the desire to maximize luminous efficacy is tempered by the desire for the source to have appropriate color appearance. The most desirable color for "white" light is generally considered to be somewhere on the blackbody locus (the locus of points in color space corresponding to blackbody radiators at different temperatures).

Over time, new light source technologies are developed. To be successful in the marketplace, these sources may need to be tuned to have maximum luminous efficacy while maintaining good color appearance. It is possible to filter a broad-spectrum light in an infinite number of ways so that it is acceptably colored, but most of these will reduce luminous efficacy by an unacceptable amount. Thus, a problem facing the lighting industry is to develop filters (or reflector coatings) that produce acceptable color while at the same time minimizing the reduction of luminous efficacy.

The goal of this research is to develop a general-purpose method for tuning the visible spectrum of any light source to any achievable color while maximizing luminous efficacy. Genetic algorithms are used to solve this multi-objective optimization problem.

Genetic Algorithms

Most optimization methods make strong assumptions about the search space (e.g., the fitness space is approximately quadratic; local minima or maxima are small or non-existent), which allows the optimal solution to be quickly determined. These methods are very powerful (i.e., fast) for the comparatively small set of problems for which their assumptions are known to be correct (or nearly so). However, there are many problems where the character of the fitness space is unknown, or known to be unsuitable for classic optimization techniques. Genetic algorithms (GAs) are a family of general-purpose optimization methodologies based on the theory of natural selection. GAs make no assumptions about the search space, so they can be applied to almost all optimization problems. However, GAs exchange applicability for speed –

although they can be used on a wide variety of problems, they are typically slower to converge to a solution than algorithms designed for a specific problem.

GAs employ a vocabulary borrowed from genetics[2]. A chromosome is the encoding of an individual solution to an optimization problem. A chromosome is composed of an ordered series of genes (i.e., a specific gene always occupies the same position in a chromosome), which each represent a parameter of the problem. Each gene has a set of alleles, which are valid values for that gene. A population is composed of a collection (typically of fixed size) of chromosomes. An iteration of the algorithm during which a new population is produced is known as a generation.

Fitness functions are used to evaluate the "goodness" of a chromosome, and can be either minimized or maximized, depending on the goal of the optimization. For example, consider a minimization problem represented by a chromosome with two genes, [x, y], and the fitness function:

$$fitness = (x - y - 2)^2$$

The fitness of the chromosome with gene values of [9, 3] is 16 and the fitness of chromosome [0, -1] is 1; therefore, [0, -1] is a better (“near optimal”) solution for this minimization problem, although not the optimal solution.

Two operations are performed on the population of chromosomes to explore the search space, mutation and crossover. Mutation is the assignment (with low probability of occurrence) of a random change to the allele value of one gene in a chromosome. Crossover is the exchange of portions genetic material from a pair of 'parent' chromosomes to produce a pair of 'children'. Parents are chosen (through a variety of means) based on fitness - fitter chromosomes are selected more frequently than less fit chromosomes.

A simple method of crossover is single point crossover, where the chromosomes are split at a randomly selected point, and genes to the left of the split for one chromosome are exchanged with genes to the right of the split for the other chromosome, and vice versa. For example, consider these two parent chromosomes:

[*3, 4, 9, 1, 7, 6, 4, 8, 1*]
[**6, 3, 8, 1, 4, 3, 4, 6, 0**]

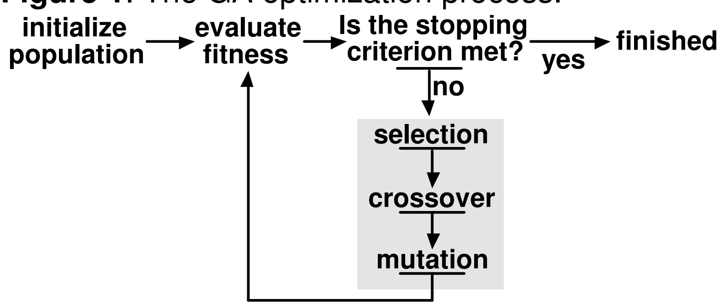
If the crossover point were between the third and fourth gene, these children would result:

[**6, 3, 8**, *1, 7, 6, 4, 8, 1*]
[*3, 4, 9***, 1, 4, 3, 4, 6, 0**]

There are many other methods of crossover that are not described here.

Figure 1 is a flowchart of the GA optimization process. First an initial population is generated, and the fitness of each chromosome is calculated. A stopping criterion (e.g., a maximum number of generations; a fitness threshold; low population diversity; etc.) is checked: if it is passed, the algorithm is finished, if it fails, selection, crossover, and mutation take place, another generation is produced, and the process returns to the “evaluate fitness” step.

Figure 1. The GA optimization process.



The three items in the shaded box in Figure 1 are where the GA process differs substantially from other optimization methods. Selection favors fitter chromosomes. Selection and crossover tend to explore 'promising' regions of the search space. Mutation helps prevent premature convergence to local optima by sampling new areas of the search space.

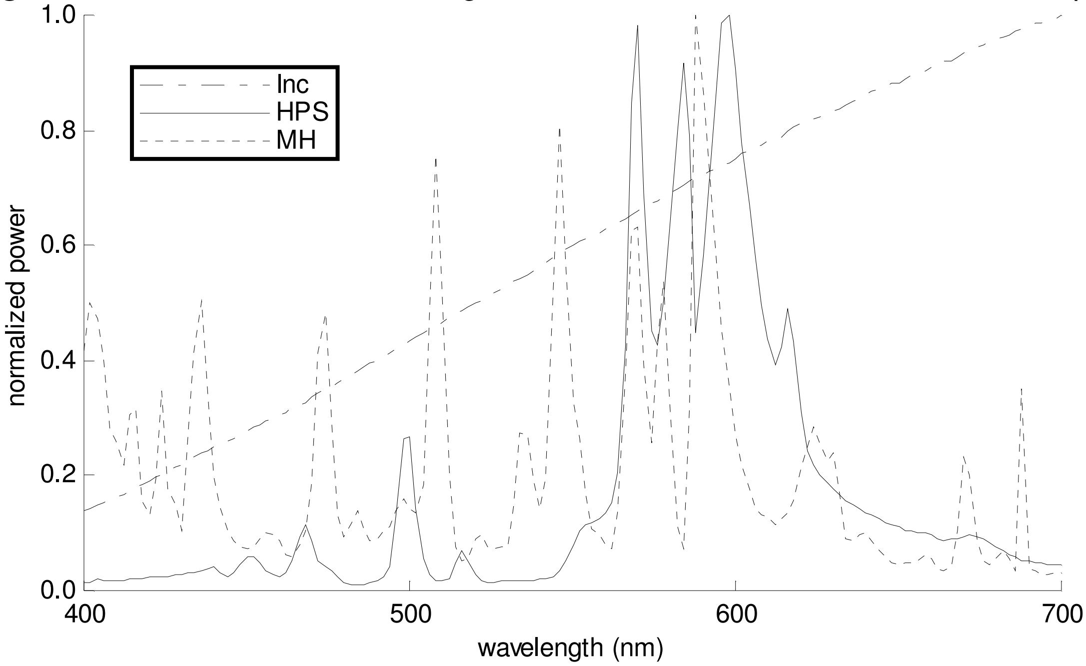
This section provided an overview of genetic algorithms, so that the reader is at least familiar with terms and concepts used later in the paper. However, because of space limitations, only the barest outline was possible. Detailed presentations of genetic algorithms are provided by Goldberg[3], Michalewicz[2], or Mitchell[4]. With the exception of Ashdown[5], GAs have not been widely employed by the lighting research community.

Spectral power distribution & color measurement

The radiant power per unit wavelength as a function of wavelength is known as the spectral power distribution (SPD). Radiometric, photometric, and colorimetric properties of a source can be determined from the SPD. The goal of this work is to take any given SPD and filter it such that it has a specific color (e.g., near the blackbody locus) while maximizing luminous efficacy. The SPDs used in this project are shown in Figure 2.

Color is characterized using a system developed by the CIE (the ISO recognized body for all matters regarding the science and art of lighting). The CIE system[6] measures color in terms of pairs of x and y chromaticity coordinates. The SPD can be converted to x and y following a standard methodology[1]. Brightness is not a factor in the CIE system: an orange (fruit) and a chocolate bar have about the same chromaticity coordinates, but an orange has a much higher reflectance than chocolate. Similarly, two lamps may have different SPDs, but still have the same chromaticity coordinates.

Figure 2. SPD for Metal Halide, High Pressure Sodium and Incandescent lamps.



The *a priori* fitness function

A lamp might be filtered in an infinite number of ways to have a certain set of chromaticity coordinates. However, only one of those ways will filter out the least amount of light, maximizing luminous flux for that lamp at that color. Thus it is desirable for a filter to have at least the following two properties: a) it minimizes the color difference between the desired and the actual light, and b) it maximizes relative efficiency. Moreover, from a practical perspective, it is desirable for filters to be relatively "smooth", with only a few notches. Therefore, a reasonable fitness function to maximize might be:

$$\text{fitness} = \text{efficiency} - \alpha^*\text{color} + \beta^*\text{smoothness}$$

where α and β are weighting factors, efficiency is the ratio of the luminous flux from the filtered lamp to the luminous flux from the unfiltered lamp, color is the distance in CIE space from the desired color, and smoothness is "smoothness" of the spectrum, which could be defined in a number of ways.

A critical part of any GA optimization problem is the proper implementation of the fitness function. This is particularly the case for multi-objective optimization: the objectives need to be balanced such that no single term dominates the others. If one term is much greater than the others near the optimum, the GA may get trapped in a region where that term is acceptable, but the other terms are unacceptable, causing the GA to converge slowly or not at all. In this case, the efficiency term was left as is, and the other terms where scaled as described below. The value of the color term should be low near the objective, and high away from the objective. Since distance from the desired color is always less than one in CIE space, this suggests straight Euclidean distance.

It proved to be difficult to obtain reasonable solutions while trying to simultaneously balance the three contributing terms of the proposed fitness function (efficiency, color, and smoothness). In an effort to make the problem easier to solve, smoothness was initially dropped from the fitness function. This proved to be fortunate, because it turned out that smoothness is automatically achieved when the other two criteria are met (provided they are properly balanced). However, some manipulations related to smoothness can speed up the convergence of the GA by half an order of magnitude as explained in the section on smoothing the chromosome.

GA implementation and early results

A floating-point representation was used for this problem. Michalewicz[2] suggests that floating point representations tend to converge faster, reach more consistent results, and provide higher precision than binary representation. The visible spectrum (400 to 701 nm) was partitioned in 151 bins, each 2 nm wide. Each chromosome consisted of 151 genes, where each gene represents the transmittance of the filter over a 2 nm band of the visible spectrum. The 2 nm bin width was chosen as a compromise between smoothness and computational tractability. Valid allele values for each gene could range from zero to one.

Population sizes between 8 and 100 were experimented with. The results presented in this paper are all for a population size of 50. Three mechanisms of crossover were applied simultaneously: (i) single point crossover (described previously); (ii) arithmetic cross-over (produces two complimentary linear combinations of the parents); and (iii) heuristic cross-over (cross-over based on interpolation, moving in the direction of the fitter chromosome). Three mechanisms of mutation were applied: (i) uniform mutation (set a gene to a random value); (ii) non-uniform mutation (change a gene by a random amount, the maximum amount of change decreasing as the maximum number of generation is approached); and (iii) boundary mutation (set the gene to its maximum or minimum value). Because the value of most of the genes could be expected a priori to be near the boundaries, the boundary mutation rate was set relatively high. Details of the crossover and mutation mechanisms are presented in Houck[7].

It is necessary for the fitness function to be well balanced to achieve reasonable results. Figure 3 represents an early solution (i.e., the spectral transmittance of a filter) before a reasonable balance for the fitness function was found: the color is right, but the efficacy is terrible, and the solution is physically infeasible because it is impossible to synthesize a filter with the transmittance properties in Figure 2. The GA can not search around the efficacy space in this case, because the color penalty is too restrictive.

Figure 3. An infeasible solution for the incandescent lamp.

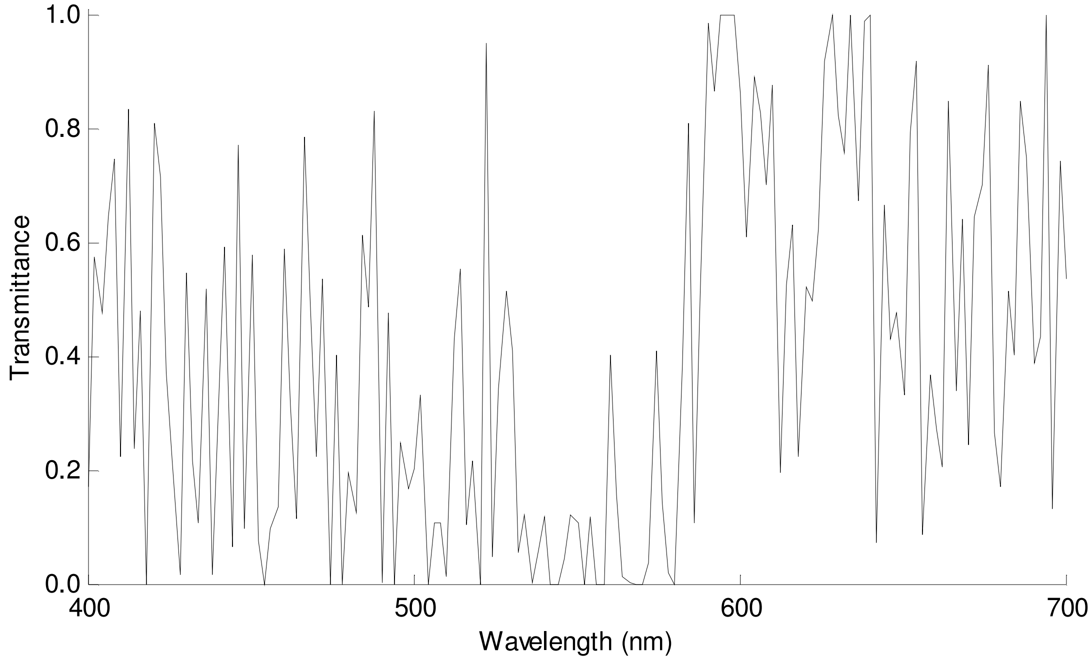
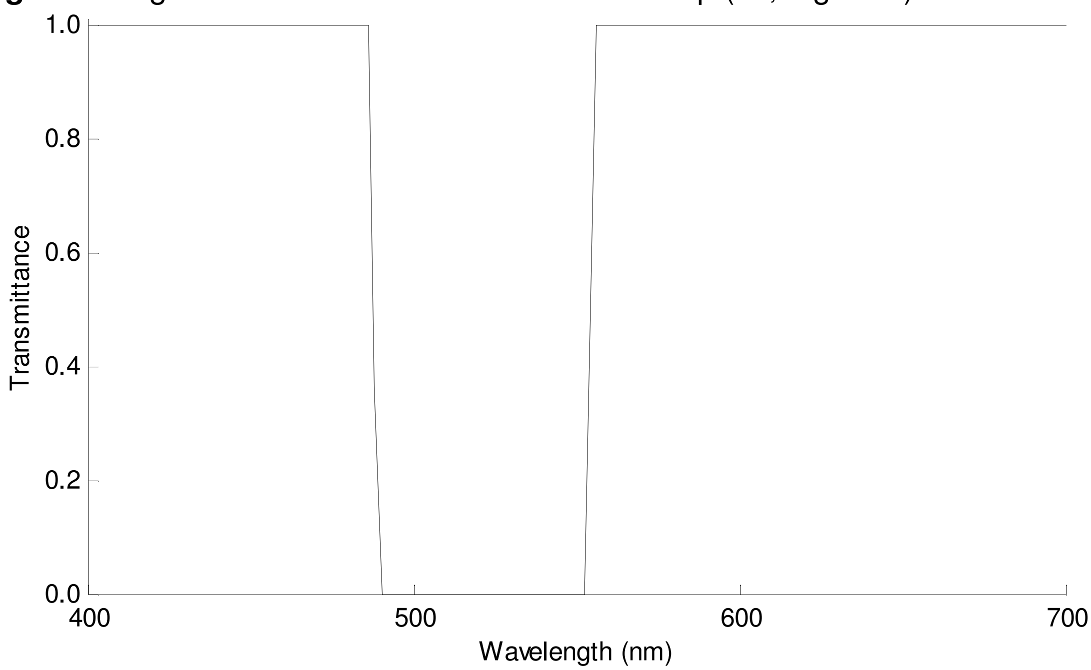


Figure 4. A good solution for the incandescent lamp (cf., Figure 3).



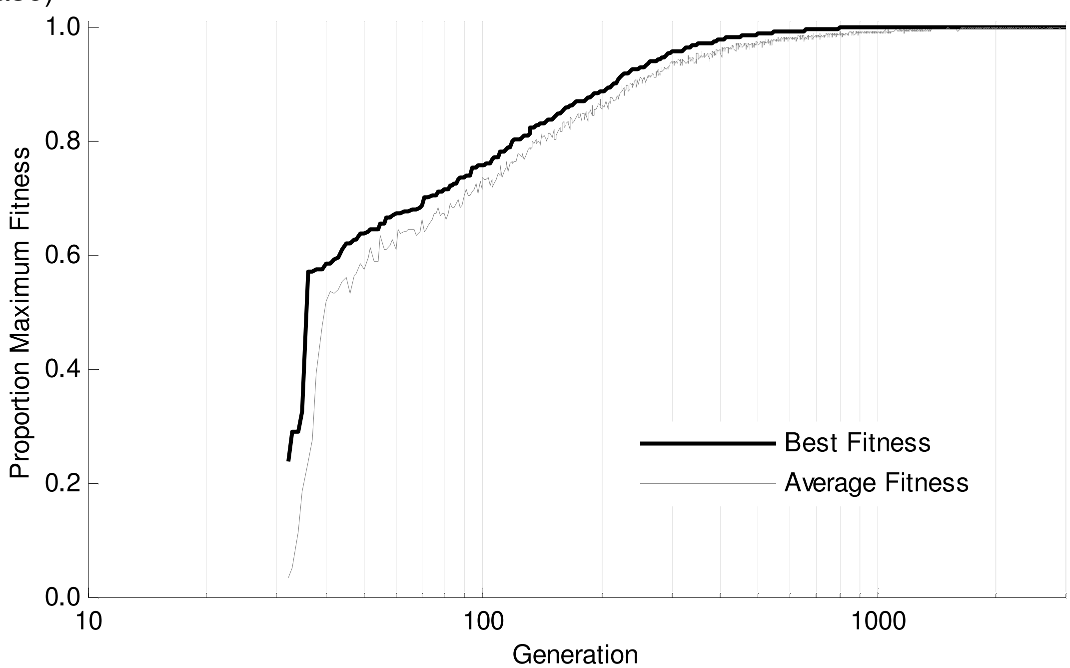
Results like those presented in Figure 3 indicate that color is dominating the fitness function in the region of the search space near the color optimum (i.e., the fitness function was not well balanced). Therefore, the color portion of the fitness function was modified such that it was low in a small region near the objective (to give the GA room to search efficacy space) and quite high away from the objective (encouraging the GA to stay away from unfruitful areas in the search space). Trial and error led to adopting:

$$2.5*(\text{distance}/0.08)^2$$

for the color term of the fitness function. This small area of low penalty around the color objective allows the efficiency to improve (in slight steps), even if color gets slightly poorer (also in small steps). Figure 4 shows a plot of a solution for the incandescent lamp with the color penalty relaxed. The color is good, efficacy is high (compared to the maximum efficiency this source can achieve at this color), and the solution is (in principal) physically realizable.

However, even for the relatively simple problem (because of the smoothness of the SPD) of the incandescent lamp, the GA does not converge quickly. Figure 5 is a plot of the fitness of the best chromosome and the average chromosome in the population by generation number for a typical run on the incandescent lamp. The population is still improving noticeably after 500 generations, and does not achieve 0.995 % of the maximum until about 1350 generations. The GA converges even slower for more difficult problems (e.g., the HPS and MH spectra), taking about 10,000 generations to reach stable results.

Figure 5. Best and average fitness for early version of the GA (incandescent case).



Improved performance by "smoothing the chromosomes"

While the time to run the GA in the above example is relatively small and sufficient to be a practical tool for lamp design, a faster implementation is always desirable. Convergence time of the GA was improved by half an order of magnitude by capitalizing on known properties of good chromosomes – learned in part from experimentation with the system as described so far. Chromosomes for good solutions should exhibit the following two properties: (i) many gene values are exactly at the limits of the allele (i.e., 100% transmission or 0% transmission); and (ii) adjacent wavelength-bins have nearly the same value (smoothness). These two properties were incorporated directly in a revised GA scheme.

When the chromosome is passed to the function that evaluates the fitness, two versions of the chromosome are evaluated: the version directly passed to the evaluation function, and a version where the chromosome is modified based on the two properties outlined above. The fitness function is compared for both versions and the fitter chromosome is passed back to the population. The smoothing is done as follows:

```
If {10% of the time}
  Genes greater than 0.5 are increased by 10%
  Genes less than 0.5 are decreased by 10%
                                                    [push towards boundaries]
Elseif {90% of the time}
  Genes greater than 0.96 are set to 1.0
  Genes less than 0.04 are set to 0.0
                                                    [push towards boundaries]
  If {gene not at the boundary (i.e. not 1 or 0)}
    Gene[i]=0.2*Gene[i-1]+0.6*Gene[i]+0.2*Gene[i+1]
                                                                     [smooth]
  Elseif {gene at the boundary}
    If {Gene[i-2] != Gene[i] != Gene[i+2] & 20% of the time}
      Gene[i]= 0.1*Gene[i-1] + 0.8*Gene[i]+0.1*Gene[i+1]
                                [detach single points "stuck" at the boundary]
    Endif
  Endif
Endif
```

Smoothing the chromosome has a remarkable effect on the convergence rate. The solution for the incandescent lamp looks similar to the one presented in Figure 4 and differs by a small amount over a few nm right at the edges of the notch. However, the time to arrive at this solution differs dramatically as illustrated by Figure 6, a plot of the fitness of the best chromosome and the average chromosome in the population against generation number for a typical run of the incandescent lamp using the smoothed GA. The population is improving noticeably up to 400 generations, and achieves 0.995% of the maximum after 385 generations (cf. 1350 for the unsmoothed algorithm).

For ease of comparison, the fitness of the best chromosome is plotted against generation for the smoothed and unsmoothed versions in Figure 7. The smoothed version is essentially done at 500 generations, while the unsmoothed version needs many additional generations achieve the maximum fitness. The improved performance of the smoothed GA is similar for the tougher MH spectrum, and the still tougher HPS spectrum. The smoothed GA converges after 1000 generations for the MH and HPS spectra.

Figure 6. Best and average fitness for the "smoothed GA" (incandescent case).

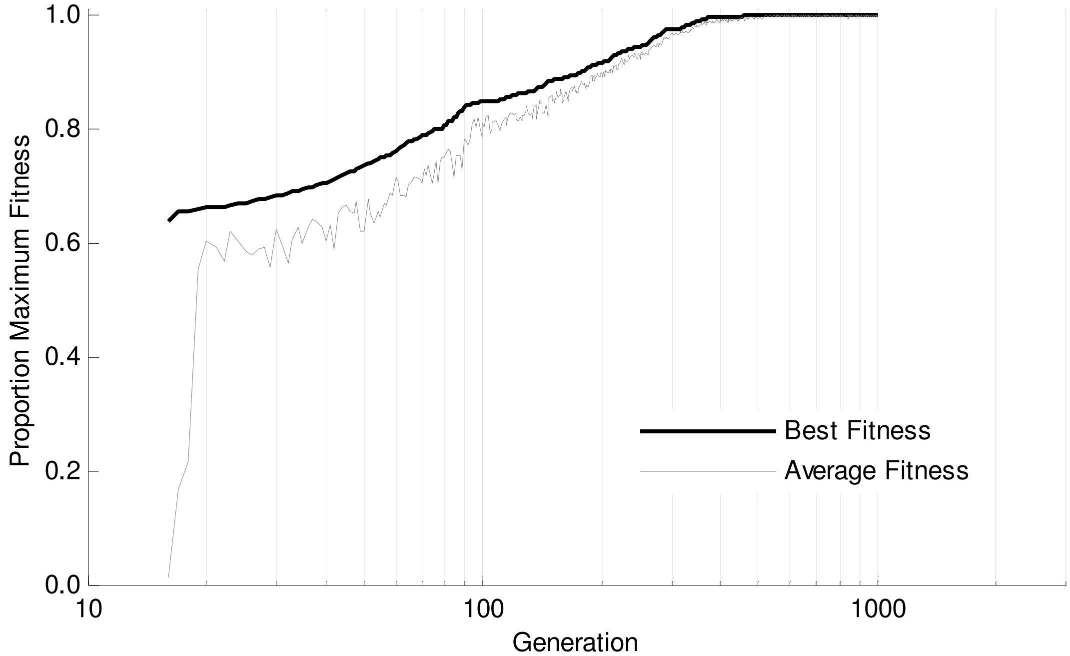
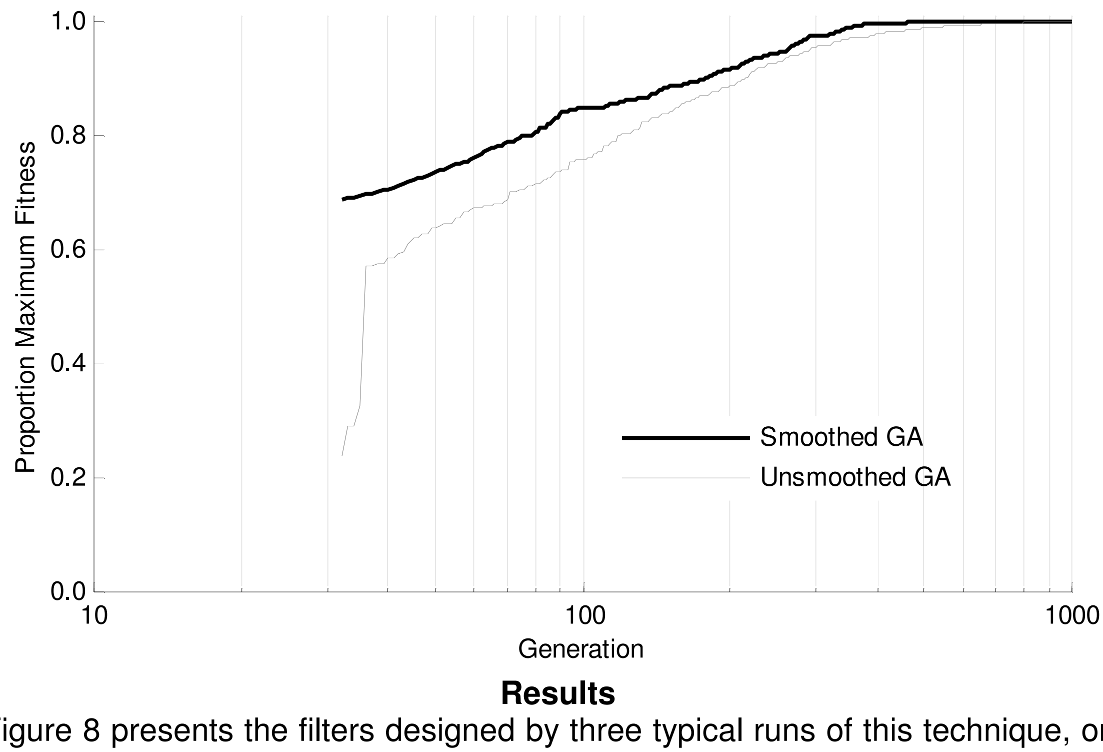


Figure 7. Fitness of the fittest chromosome for "smoothed" and "unsmoothed GA".



Results

Figure 8 presents the filters designed by three typical runs of this technique, one on each of the three spectra (MH, HPS, and incandescent) plotted in Figure 2. The chromaticity coordinates of the target color was (x = 0.48; y = 0.32). This point was selected to be about equidistant from the unfiltered spectra of the three lamps considered, and achievable (albeit at low efficiency) by all three lamps. Table 1 gives the chromaticity coordinates and efficiency for these solutions. The color is correct, and the efficiency is the maximum achievable for those lamps at that color.

Table 1. Chromaticity coordinates and efficiency of the filtered spectra.

Lamp	Unfiltered x	Unfiltered y	Filtered x	Filtered y	Filtered Efficiency
Incandescent	0.418	0.397	0.479	0.321	0.650
HPS	0.525	0.414	0.480	0.321	0.230
MH	0.369	0.382	0.479	0.321	0.510

Figure 8. Optimal filter transmittance for the three spectra (for the given problem).

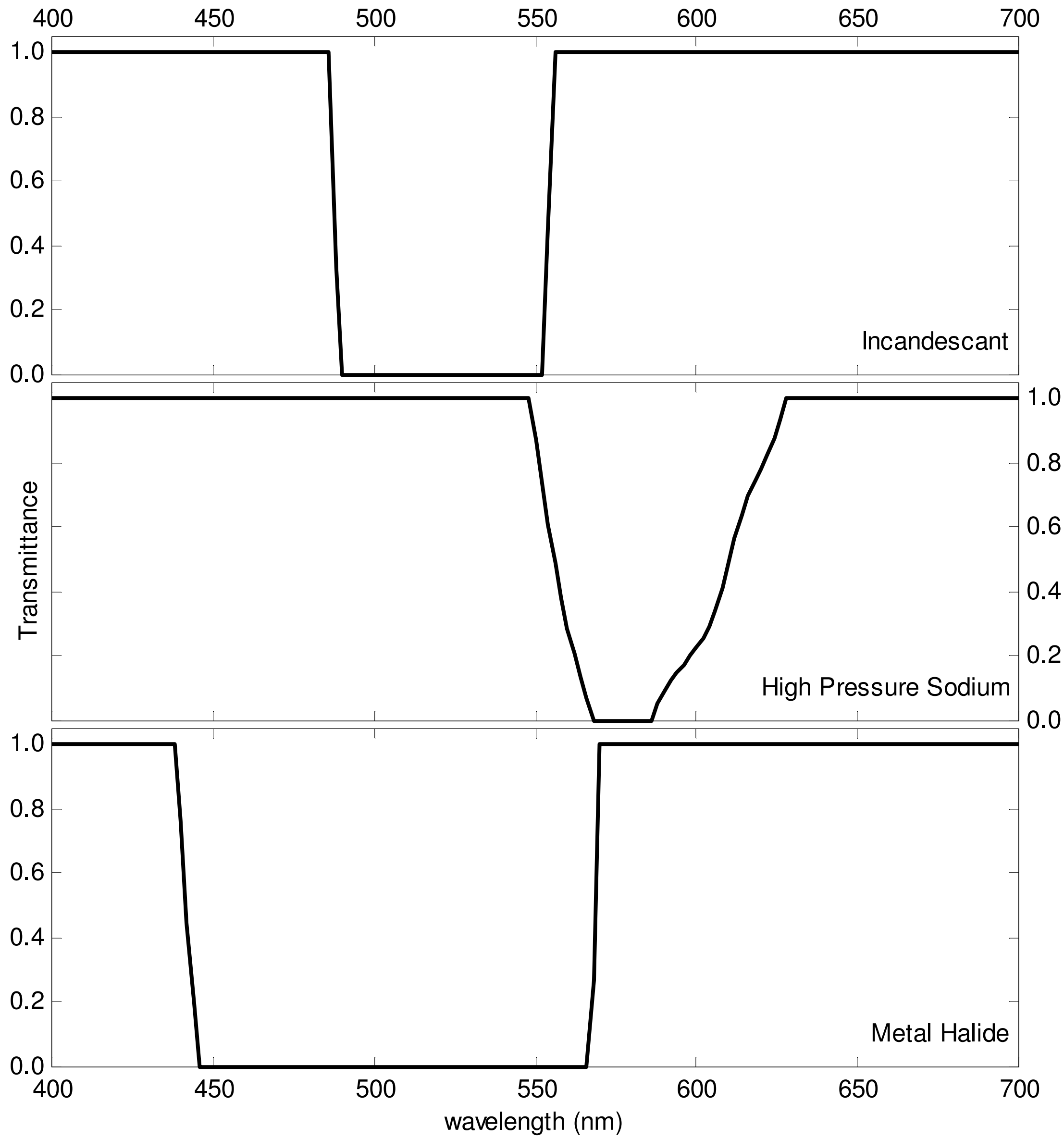


Figure 9 shows the progress of the GAs through color space. The filled symbol shows the chromaticity coordinates of the unfiltered lamp. The line leading from each symbol is the chromaticity coordinate of the lamp filtered by the filter encoded in the best chromosome each time a new best solution is generated. The inset figure in the lower left corner shows the CIE chromaticity coordinates plotted (not to the scale of the main figure) in the x-y plane. The box in the inset figure shows the area plotted in the main portion of the figure. About 260 filtered spectra are plotted for each lamp. Note that only approximately the first 20 are distinguishable outside the "blob" at (.48, .32).

Figure 9. Path of the GA through the color space (see text for explanation).

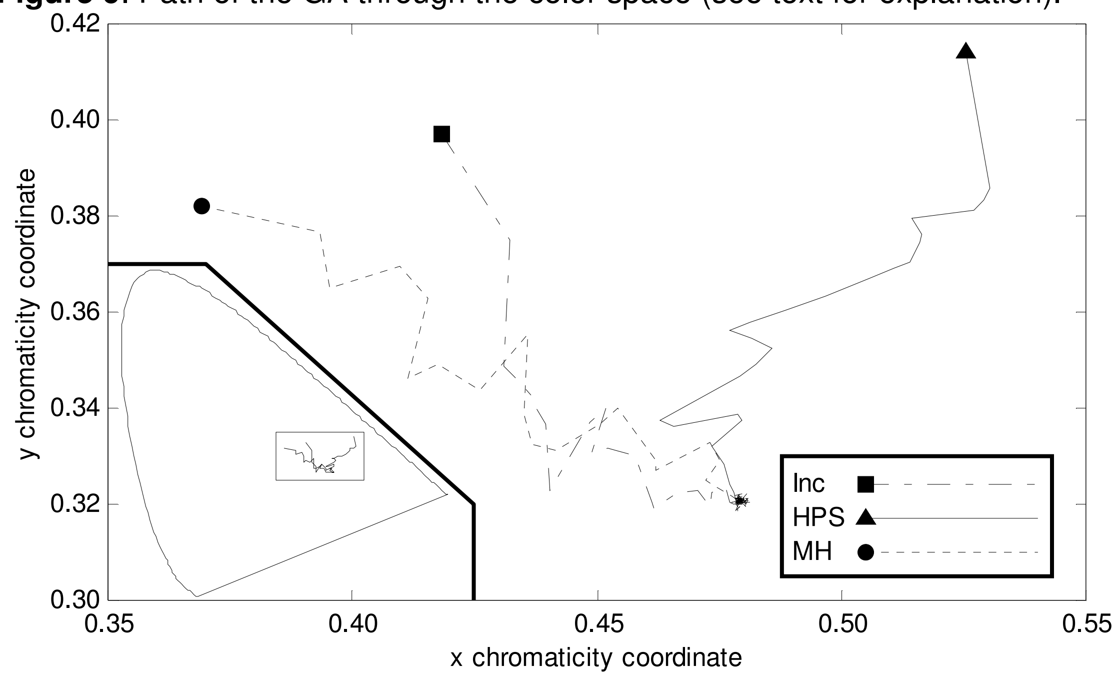
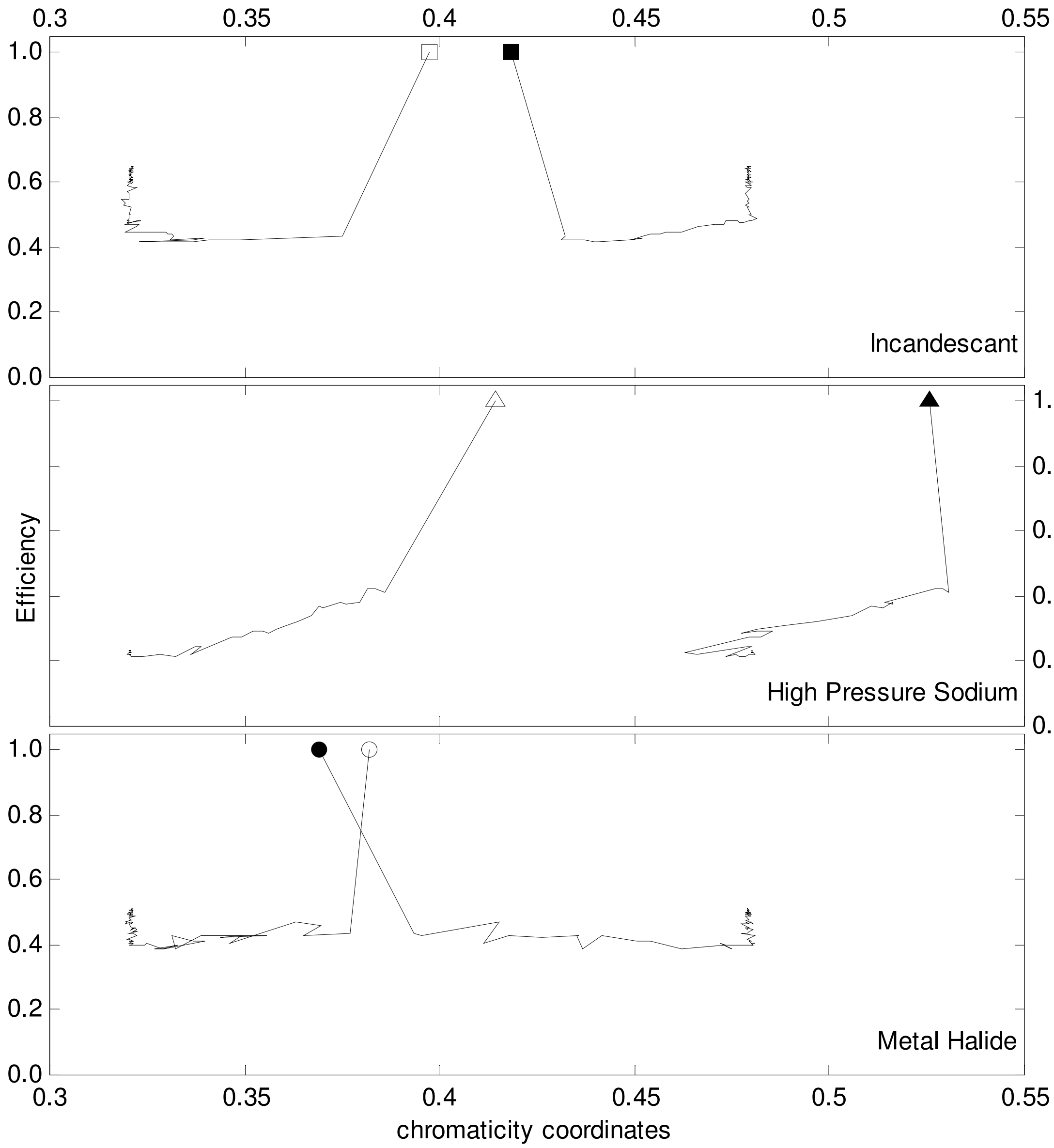


Figure 10 is a way to visualize the path of the GA through the color and efficiency space. This figure plots efficiency against x and y chromaticity coordinates (individually), with each spectrum in a different subplot. The symbols represent the unfiltered chromaticity coordinates. The filled symbols represent the x chromaticity coordinate; the empty symbols represent the y chromaticity coordinate. The line leading from each symbol is the efficiency of the lamp filtered by the filter encoded in the best chromosome each time a new best solution is generated, plotted against the respective x or y chromaticity coordinate.

Figures 9 and 10 plot the path of the GA through color and efficiency space over time. Early on, relatively large steps toward the correct color are made, as shown in the region of Figure 9 where the individual points are distinguishable and the region of relatively horizontal movement in Figure 10 (following the first big step from the unfiltered lamp). After the correct color is found, efficiency is optimized in many small steps (the predominantly vertical, “squiggly” portion of Figure 10, at the opposite end of the line from the symbol). The small horizontal movements in this portion of the line represent tradeoffs between color and efficiency (i.e., a small deviation from the ideal color for a small gain in efficiency).

Figure 10. Path of the GA through the search space (see text for explanation).



The two portions of the path of the GA through the color and efficiency space described above are a feature of the solution space (rather than a result of the way color and efficiency are scaled, as one might expect). It is easier to find the right color, because there are *many* different solutions to that portion of the problem. Meanwhile, it is quite difficult to find the most efficient solution because there are very few (probably only one) “most efficient” solution, and comparatively few solutions that are “very efficient”. It is a tribute to the power of genetic algorithms that they are able to find an efficient solution at all.

Conclusions

This study showed that a genetic algorithm approach proved to be very effective for the problem of filter optimization. The "smoothed GA" modification resulted in faster convergence of the algorithm by about half an order of magnitude, by incorporating the knowledge that in order to achieve a smooth spectrum adjacent genes should have similar values.

An interesting result of this study is the insight into the trade-off between multiple objectives. The balance between objectives is not very important early on; however, as the population approaches the optimum, changes in any one gene should produce a comparable magnitude change in all of the objectives. Thus changes in the transmittance at any one wavelength should affect the color and the efficiency terms by very small, and comparable, amounts, allowing tradeoffs to be effectively made (i.e., moving toward worse color for a small gain in efficiency, or vice versa).

If the researcher (or manufacturer) were willing to accept a simple notch as a solution (for reasons of manufacturing cost, for example), a chromosome with only two genes, one for position and one for width, might be a better encoding of the problem. Moreover, a deterministic method of solving this modified problem is also suggested by the results: draw a ray from the starting point of the lamp on a CIE (x,y)-chromaticity diagram to the desired filtered color, and continue this ray to the far spectrum locus. Use the intersection of the ray and the spectrum locus as a starting point for a simple hill-climbing algorithm.

However, because GAs are so adaptable, the application of GAs to spectrum tuning goes well beyond the scope of this problem. For example, this technique can provide rather than one particular solution (i.e., maximum efficiency at a particular color, as outlined here) the set of Pareto-optimal[8] solutions to a multi-objective spectrum optimization problem (e.g., the surface of all color and efficiency solutions possible, or the CRI and efficacy front suggested by Opstelten[9]). This technique could be used to design a filter for a light source such that the light, after passing through a given length of fiber optic cable, appears white (or any other color). Other objectives can easily be added, such as determining solutions with a minimum color rendering index (CRI) or a target correlated color temperature (CCT).

Acknowledgement

The authors would like to thank Andrew Bierman (at the Lighting Research Center) for providing lamp spectra and technical assistance with this project. The GA Optimization Toolbox[7] (GAOT) was used to implement the GAs for this research.

References

1. IESNA (1993). *Lighting Handbook* (8th Ed.). Illuminating Engineering Society of North America, New York.
2. Michalewicz, Z. (1996). *Genetic Algorithms + Data Structures = Evolution Programs* (3rd. Ed). Springer-Verlag, Berlin.
3. Goldberg, D. (1989). *Genetic Algorithms in Search Optimization & Machine Learning*. Addison Wesley Longman, Inc., Reading, Massachusetts.
4. Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, Massachusetts.
5. Ashdown, I. (1994). Non-Imaging Optics Design Using Genetic Algorithms. *Journal of the Illuminating Engineering Society*. 23(1), 12.
6. CIE (1971). Colorimetry (Official Recommendations of the International Commission on Illumination). CIE Publication No. 15 (E-1.3.1), Bureau Central de la CIE, Paris.
7. Houck, C., Joines, J., and Kay, M. (1995). A Genetic Algorithm for Function Optimization: A MATLAB Implementation. North Carolina State University - Industrial Engineering Technical Report 95-09.
8. Winston, W. (1994). *Operations Research: Applications and Algorithms*. Duxbury Press, Belmont, California.
9. Opstelten, J., Radielovic, D., and Verstegen, J. (1975). Optimum spectra for light sources. *Philips Technical Review*, 35(11/12), 361.