Sensitivity to Crossover and Mutation Probabilities in Levee’s Setback Optimization

M. SHAFIEI, O. BOZORG HADDAD
Department of Civil Engineering
Iran University of Science and Technology
Narmak, Tehran
IRAN

Abstract: Floods are one of the major natural disasters that often threaten human lives and cause significant economic losses around the world. The history of mankind is filled with the stories of our struggles with floods to protect human races and to sustain the progression of our civilizations. Flood defense systems are designed and constructed to protect low-lying areas against flooding. Engineering design often is at the final stage for finding technical means to best accomplish the project goals. Over the years, risk-based design and optimization methods have proven to be useful tools to obtain economic design of protection systems. The most common economic framework for floodplain management is minimization of expected annual damages and flood management expenses, structural and nonstructural flood control options. Levee systems have been built for flood protection in numerous rivers, lakes and coasts in the world over the long human history. Economic design of a levee system for flood protection involves balancing costs of levee building (height), the losses of land value sacrificed for floodway expansion (setback) and flood damages from inadequate channel capacity. The application of GAs to water resources problems has been increased in recent years. The study of genetic algorithms (GAs) has developed into a powerful optimization approach. GAs have so far had very little applications in flood defense systems optimization. In this paper some new different approaches to GA formulation are considered, along with a range of sensitivity analysis. The object has been to present GAs as a practical tool in levee design optimization and to examine the potential of different GA formulations for solving the problem. The results obtained indicate that there is potential for application of GAs to levees optimization problems, where the objective function is nonlinear and other optimization techniques may be difficult to apply.

Keywords: Optimization, Levee, Setback, GA, Classified Population, Genetic Operators

1 Introduction

Floods are one of the major natural disasters that often threaten human lives and cause significant economic losses around the world. The history of mankind is filled with the stories of our struggles with floods to protect human races and to sustain the progression of our civilizations. Even with centuries of experiences on flood defense and tremendous amount of progresses have been achieved, flood still appears to enjoy being the main enemy of public in the category of natural disaster. Berz (2000) recently compares flood disasters with earthquakes, storms and other forms of nature disasters in the world. The study indicates that floods contribute to 58% of total death and 33% of economic losses [1]. Great majority of flood, related death and economic losses occurred in developing countries. Flood defense systems are designed and constructed to protect low-lying areas against flooding. Decisions for a flood defense system are multi-dimensional which involve a set of goals and constraints arising from political, economical and engineering aspects. Over the years engineering design concepts have been evolved as the science and technology in dealing with flood issues progress and improve.

There have been several application of GAs to water resources problems. Wang (1991) applied a GA to the calibration of a conceptual rainfall-runoff model. Similar work has been reported by Franchini (1996), who used a GA in combination with sequential quadratic programming to calibrate a conceptual rainfall-runoff model. There have
been also several applications of GAs to pipe network problems. Goldberg (1987), Murphy et al. (1993), Davidson and Goulter (1995) and Dandy et al (1996) used a simple and also improved GA for pipeline, pipe network and water supply network optimization. In ground-water pollution problems, Ritzel al. (1994) and McKinney and Lin (1994) have been made good experiences.

Levee systems have been built for flood protection in numerous rivers, lakes and coasts in the world over the long human history. Early flood levees usually were designed with scant quantitative analysis, relying primarily on occasional observations of flood stages and empirical judgments on required project scales. The achievements in experimental and theoretical hydraulics since the 18th century, rational estimation of storm discharge in the mid 19th century and the emerging of early economic-engineering analysis (Humphreys, 1861) made possible the “modern sense” designs of flood levees. In recent decades, several studies have addressed the economic aspects of flood levee design, usually with benefit-cost analysis and optimization techniques (Tung and Mays, 1981, Wurbs, 1983).

GAs have so far had little applications in flood defense systems optimization. Excellent introductions to GAs are given by Goldberg (1989) and by Michalewicz (1992) and several recent papers give summaries of the essentials (e.g. Oliveira and Loucks, 1997 and Savic and Walters, 1997). Shafiei et al. used genetic algorithms for optimization of levee setback along sensitivity analysis to probability of crossover and mutation operators [12]. They have also presented a new approach for GA formulation in optimization of levee’s setback [12]. In this paper, several different probabilities of genetic operators in GA formulation through classified population are considered, along with a range of sensitivity analysis in the same problem of setback optimization. The object has been to present GAs as a practical tool in levee design optimization and to examine the potential of different GA formulations for solving the problem.

2 Economic Design of Flood Defense Systems

Flood defense systems are designed and constructed to protect low-lying areas against flooding. The objective in economic design of a hydraulic structure is to minimize the sum of capital investment cost, the expected flood damage costs and operation and maintenance costs.

Economic design of a levee system for flood protection involves balancing construction costs of levee, the losses of land value sacrificed for floodway expansion (Setback) and flood damages from inadequate channel capacity. The most common economic objective for floodplain management is minimization of expected annual damages and flood management expenses. Under static conditions, the flood frequency distribution is stationary and economic factors, such as the value of damage to properties, construction cost, and floodplain land values, are constant.

The information needed for this process can be categorized into four types.

- Hydrologic/Physiographical Data
- Hydraulic Data
- Structural Data
- Economic Data

2.1 Optimal Tradeoff of Levee Setback and Height

A static model is formulated to minimize the sum of expected flood damage, considering levee construction cost and resultant land value loss due to floodway occupancy. This simple model allows preliminary quantitative examination of the tradeoff between optimal setback and optimal height in designing a new levee. The objective function is:

\[
\text{Max} \frac{B}{C}(X_s, X_h) = \frac{B(X_s, X_h)}{C(X_s)}
\]  

Where \( X_s \) and \( X_h \) are designed levee setback and height respectively. B and C also identify the benefit and cost of levee system.

The land value benefit function \( B(\cdot) \) depends not only on levee setback but also on levee height because the bottom width of levee cross-section may change with levee height. Considering partial derivatives of \( B/C(X_s, X_h) \), with respect to \( X_s \) and \( X_h \) equal zero,

\[
\frac{\partial (B/C)}{X_s} = 0 \quad (2)
\]

\[
\frac{\partial (B/C)}{X_h} = 0 \quad (3)
\]

Given a levee overtopping flow \( Q(X_s, X_h) \), we have

\[
\frac{\partial Q}{\partial X_h} = \frac{\partial Q}{\partial X_s} \quad (4)
\]

Equation (4) holds for the optimal levee height and setback. The optimal levee height and setback can be found by numerically solving combined
equations (2) and (3) and verifying that a minimum has been found, even though the expected total cost function in Equation (1) is not convex [14].

3 Genetic Algorithms

Evolutionary Algorithms can be divided into three main areas of research: Genetic Algorithms (GA), Evolution Strategies (ES) and Evolutionary Programming (EP). Genetic Programming began as a general model for adaptive process but has since become effective at optimization while Evolution Strategies was designed from the beginning for variable optimization. The schematic diagram of these algorithms which are made of the several iterations of basic Evolution Cycle is shown below:

![Schematic Diagram of Evolution Cycle](image)

A GA is a search algorithm based upon the mechanics of natural selection, derived from the theory of natural evolution. GAs simulate mechanisms of population genetics and natural rules of survival in pursuit of the ideas of adaptation, indeed this has led to a vocabulary borrowed from natural genetics [3]. Goldberg (1989) identifies the following as the significant differences between GA and more traditional optimization methods:

- GAs work with a coding of the parameter set, not with the parameters themselves.
- GA search from a population of points, not a single point
- GAs use objective function information, not derivatives or other auxiliary knowledge
- GAs use probabilistic transition rules not deterministic rules

A GA is a robust method for searching the optimum, solution to a complex problem, although it may not necessarily lead to the best possible solution. A GA generally represents a solution using strings (also referred to as chromosomes) of variables that represent the problem. In early GAs (Goldberg and Kuo 1987, Wang 1991) these strings were comprised of binary bits. In binary representation, the bits may encode integers, real numbers, sets or whatever else is appropriate to the problem. Real-value coding is now proving more effective in many problems than binary coding (e.g., Oliveira and Loucks 1997).

Coding components of possible solutions into a chromosome is the first part of a GA formulation. Each chromosome is a potential solution and is comprised of a series of sub-strings or genes, representing components or variables that either form or can be used to evaluate the objective function of the problem. The fitness of a chromosome as a candidate solution to a problem is an expression of the value of the objective function represented by it. It is also a function of the problem constraints and may be modified through the introduction of penalties when constraints are not satisfied.

A GA starts with a population of chromosomes, which are combined through genetic operators to produce successively better chromosomes. The genetic operators used in the reproductive process are selection, crossover and mutation. Chromosomes in the population with high fitness values have a high probability of being selected for combination of other chromosomes of high fitness. Combination is achieved through the crossover of pieces of genetic material between selected chromosomes. Mutation allows for the random mutations of bits of information in individual genes. Through successive generations fitness should progressively improve. Various schemes for selection, crossover, and mutation exist and will be discussed.

3.1 Genetic Algorithm Process

Canonical form of GA approach can be summarized as:
1. Define the objective function which is appropriate to conditions of problem.
2. Present the possible solutions (phenotype) as real value (genotype or chromosome).
3. Generate a random population of specific size. The population size affects the efficiency and performance of GA.
4. Evaluate the fitness of every solution over the objective function. There are many methods to evaluate fitness and assign a real number to each chromosome.
5. Select a population of chromosomes of the same size of initial population for mating, by a random selection method.
6. Apply crossover operation on selected pairs if they have been chosen for crossover.
7. Replace the parent population with new generation.
8. Applying mutation operator based on the probability of mutation. At this point the process of producing a pair of offspring from two selected parents is finished.
9. Go through steps 4 to 8 until the termination criteria met.

3.2 Representation Schemes
Traditionally GAs have used binary coding, in which a chromosome is represented by a string of binary bits that can, encode integers, real numbers, or anything else appropriate to a problem. In a real-value representation, individual genes of a chromosome are initially allocated values randomly within feasible limits of the variable represented, with a sufficiently large population of chromosomes adequate representation will be achieved. There is a significant advantage in not wasting computer time on decoding for objective function evaluation, although a more careful approach to mutation is required. In real-value coding there is no discretization of the decision variable space. This is another advantage of this approach.

3.2.1 Selection Approaches
Selection is the procedure by which chromosomes are chosen for participation in the reproduction process. A popular approach has been fitness proportionate selection (Goldberg 1989), in which the probability $P$ of an individual $k$ being selected is given by:

$$p_k = \frac{f_k}{\sum_j f_j} \quad (5)$$

where $f$ is fitness of individuals along the population.

Various rank selection schemes are in use (Michalewicz 1992 [6]) that tend to ensure that good chromosomes have higher chances of being selected for the next generation. Ranking schemes operate by sorting the population on the basis of fitness values and then assigning a probability of selection based upon the rank. The roulette wheel approach is one of ranking schemes of selection.

A constant selection differential is thus maintained between the best and the worst individuals in the population. A drawback is that information on the relative fitness of the individuals is not used. Goldberg and Deb (1990) have compared various selection schemes, and indicated a preference for the tournament selection scheme. In tournament selection a group of individuals are chosen at random from the population, and the individual with the highest fitness is selected for inclusion in the next generation. The procedure is repeated until the appropriate number of individual are selected for the new generation. The approach had originally been developed with groups of two individuals and was called binary tournament selection, but larger groups lead to greater diversity and a smoother progression to a solution. Tournament selection was used by some authors in the ground-water monitoring problem.

3.2.2 Crossover Approaches
The general theory behind the crossover operation is that, by exchanging important building blocks between two strings that perform well, the GA attempts to create new strings that preserve the best material from two parent strings. The number of strings in which material is exchanged is controlled by the crossover probability forming part of the parametric data. Goldberg (1989) and Michalewicz (1992) describe the following methods of crossover (I) one-point crossover; (2) two- point crossover and (3) uniform crossover [3], [6].

Crossover occurs between two selected chromosomes with some specified probability. In one-point crossover, a crossover point is selected at random at some point $C$ in the chromosome length $L$ and two new individuals are created by swapping all genes between positions $C$ and $L$. In two-point crossover, genetic material between two positions chosen at random along the length of the chromosomes, $C_1$ and $C_2$, is exchanged. Uniform crossover operates on individual genes of the selected chromosomes, rather than on blocks of genetic maternal, and each gene is considered in turn for crossover or exchange.
3.2.3 Mutation Approaches

Mutation is an important process that permits new genetic material to be introduced to a population. A mutation probability is specified that permits random mutations to be made to individual genes. The two basic approaches to mutation for real-value representations are uniform mutation and non-uniform mutation [6]. Uniform mutation permits the value of a gene to be mutated randomly within its feasible range of values; possibly resulting in significant modification of otherwise good solutions. Modified uniform mutation permits modification of a gene by a specified amount, which may be either positive or negative. In non-uniform mutation, the amount by which genes are mutated can be reduced as a run progresses, and can therefore help in the later generations to fine tune the solutions. This operator is particularly suited to problems where high precision is required.

4 Methodology

4.1 Hydrologic Analysis

In this paper, the HEC-HMS was applied to fulfill hydrologic analysis. The resulting peak flows of 25-year and 100-year hydrograph is to extract from the outputs of the model.

4.2 Geometry Model

In this paper, the digital topographic maps of study area were used to make the required DEM of study area and the cross-sections were defined and extracted to use in next steps.

4.3 Hydraulic modeling

In this paper, the steady uniform current is applied along the river. The design discharge for optimizing the setback of levees of study area was selected as the peak flow of hydrograph with return period of 100 year.

For accomplishing hydraulic modeling and analysis in this study, HEC-RAS model, HEC-RAS version 3.0.1, as developed by the Hydrologic Engineering Center, U.S. Army Corps of Engineers, was used as a basis model to verify water surface profiles based on the flood discharges which are excluded from a visual basic code determining water level of each section. This code make the calculations having the characteristics of the cross-sections such as ground points elevations and maning value and also discharge of current with unit of cubic meter per second. This part of analysis is done to make the hydraulic simulator able to participate in the iterative process of optimization.

4.4 Economic Analysis And Optimization

In this stage, we continue to use a visual basic code written to analyze the cost and benefit of flood defense system and determine the optimum levees setback with the maximum economic benefits. Genetic algorithm is applied to achieve best solution. Substituting the hydrologic and hydraulic study outputs in the optimization analysis and refining the construction cost estimate based on developing knowledge of cost-sensitive features such as water level, we make it possible to determine the optimum design.

4.4.1 Genetic Algorithm Formulation

Since the objective function is based on maximizing ratio between benefit and cost of flood levees, levee height in each section should be the decision variable on which the GA is based. Construction cost is calculated by using volume of constructed levee and unit cost of it. Benefit is calculated from the value of protected land due to constructed hydraulic structure (Levee). With 10
sections and one levee height, there are thus 10 discrete variables to be represented in the GA. Each of these may be considered to be a gene. Elevation values are to be considered as non-integer quantities with precision of 0.1 meter. This is toward defining the problem and is not a limitation for GAs.

An alternative approach to formulation of the GA is to use a representation appropriate to the components of the problem. Here, Real-value chromosomes have been used with success where individual genes of a chromosome are initiated by randomly within feasible zone. With a sufficiently large population good representation will be achieved.

In this paper, the structure of real-value formulation of GA, presented by Shafiei et al., is applied and a series of sensitivity analysis to the corresponding genetic operators would be done [12]. In this new structure, called classified population, the population size is 80. After initial simulation and corresponding calculation of fitness, in selection step, first the chromosome with highest fitness is forced to be selected for new generation and then, other chromosomes will be selected through roulette-wheel selection method till the size of new population overcomes a quarter of initial population. Then, crossover will be applied through selected chromosomes and another quarter of new population is made of changed chromosomes. In this step, 1-point cut and 2-point cut approaches are applied. Finally, mutation operator is going to be applied based on the probability of mutation through the first half of new generation and remained half of new generation will be produced thorough this process. The probabilities between 1 and 3 genes per chromosome would be applied for mutation. This process is repeated until the termination criteria met. The diagram of new structure to GA is shown below:

For generating a random population, the population size is 80 chromosomes for making the sensitivity analysis and 400 chromosome for final optimization analysis.

5 Study Area

The Ajichai Watershed is located in Tabriz. It flows from the southern part of the Sabalan Mountain in a westerly direction across the urbanized Herris and Sarab County and through the city of Tabriz to its confluence with the Uroumieh lake.

The Ajichai River natural valley flood plain averages about 600 meters wide while the main channel averages about 50 meters wide through the study reach.

The Ajichai catchment elevation ranges from 1458 m to 3883 m above sea level and the annual average precipitation is approximately 300 mm.

Bank-full discharge corresponding to an event with the 4-percent chance of exceedance (25-year) is about 400 cubic meters per second and the 1-percent chance of exceedance (100-year) event is about 700 cubic meters per second. The average bed slope through the project reach is on the order of 20 centimeters per 100 meters.

A subset area was selected along the river with a length of 2 kilometers and used as study area during the hydraulic and optimization process. Such a subset is just large enough to represent the river and the surrounding surfaces so that the computing time is reduced to minimum in the optimization algorithm.

6 Results

As discussed above, consideration of GA formulation has been given to real-value coding with some approaches of crossover and mutation through classified population. Performance of each of this different operator probabilities is discussed below.
6.1 Sensitivity analysis

A series of sensitivity analyses were carried out to establish appropriate parameter settings under real-value coding and alternative operators. In many practical problems, GA results are found to be sensitive to crossover and mutation probabilities. This is because genetic material lost at the stage of a run, through either crossover or mutation, may be needed to improve fitness. Sensitivity to crossover and mutation approaches is discussed below for real-value coding scheme in GA formulation through classified populations.

6.1.1 Sensitivity to Crossover and Mutation Approaches

In this study, sensitivity to crossover and mutation approaches was carried out using a population size of 80. The roulette wheel selection approach was adopted with 1-point and 2-point cut, and a uniform mutation operator with probability of 1 to 3 gene per chromosome were considered through runs with a fixed length of 500 generations. Fig. 7 and 8 show the sensitivity of the achieved fitness to crossover probability for each of the schemes considered. Fitness is expressed as maximum and average values of objective function, as the ratio between benefit and cost of levee problem.

Table 1 include the information such as minimum, average and maximum of obtained results through different operators of crossover and mutation.

<table>
<thead>
<tr>
<th>Crossover Prob.</th>
<th>Mutation Prob.</th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
<th>Standard Deviation</th>
<th>Coefficient Of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.005</td>
<td>1.077</td>
<td>1.148</td>
<td>1.188</td>
<td>0.035</td>
<td>0.030</td>
</tr>
<tr>
<td>0.2</td>
<td>0.01</td>
<td>1.146</td>
<td>1.169</td>
<td>1.192</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td>0.2</td>
<td>0.02</td>
<td>1.109</td>
<td>1.127</td>
<td>1.153</td>
<td>0.019</td>
<td>0.017</td>
</tr>
<tr>
<td>0.4</td>
<td>0.005</td>
<td>1.103</td>
<td>1.146</td>
<td>1.188</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td>0.4</td>
<td>0.01</td>
<td>1.057</td>
<td>1.126</td>
<td>1.163</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td>0.4</td>
<td>0.02</td>
<td>1.097</td>
<td>1.130</td>
<td>1.158</td>
<td>0.019</td>
<td>0.017</td>
</tr>
<tr>
<td>0.6</td>
<td>0.005</td>
<td>1.081</td>
<td>1.136</td>
<td>1.185</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td>0.6</td>
<td>0.01</td>
<td>1.125</td>
<td>1.154</td>
<td>1.175</td>
<td>0.015</td>
<td>0.013</td>
</tr>
<tr>
<td>0.6</td>
<td>0.02</td>
<td>1.072</td>
<td>1.125</td>
<td>1.173</td>
<td>0.025</td>
<td>0.022</td>
</tr>
</tbody>
</table>

The results demonstrate clearly that GAs are robust, with reasonable results being obtained by 1-point cut crossover and mutation probability of 1 gene per chromosome. As it is seen, by this structure, it would be possible to obtain better results than previous works (Shafiei et al., 2005b). Figure 9 illustrates objective function value through generations along selected approach.

![Fig. 7: sensitivity to crossover and mutation probability, maximum of objective function values](image)

![Fig. 8: sensitivity to crossover and mutation probability, average of objective function values](image)

![Fig. 9: objective function value through new GA formulation](image)

Table 2 includes comparing values of benefit to cost of flood in through previous studies and selected approach of new formulation of GA. The optimization process has been done through 4000 iterations with population size of 400 for both approaches.

<table>
<thead>
<tr>
<th>Description</th>
<th>Maximum of Objective Function Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous Studies</td>
<td>1.175</td>
</tr>
<tr>
<td>New Approach, Classified Population</td>
<td>1.252</td>
</tr>
</tbody>
</table>

It is obviously seen that in addition to type of genetic operators, the selected formulation in comparison with previous studies, gives better solution for the problem of levees setback optimization.
7 Conclusion

It has been demonstrated that GAs provide robust and acceptable solutions to the levees setback optimization problem. Several possible formulations have been considered, along with their sensitivity to various parameters. It is included that in real-value representation scheme, incorporating roulette wheel selection will operate most efficiently and produce the best results through 1-point cut crossover and uniform mutation with low probability.

Developing a new formulation of GA having different approaches for crossover and different probabilities for mutation, 1-point cut crossover and mutation probability of 1 gene per chromosome are appropriate for the problem presented here. For the levee problem, a more precise solution can be achieved within 4000 generations with a population of 400.

The results achieved indicate that there is potential for the application of GAs to large rivers levees optimization problems, where the objective function is complex and other techniques are difficult to apply.

References: