

# Dominance Measures for Multi-Objective Simulated Annealing

Kevin Smith    Richard Everson    Jonathan Fieldsend

June 20 2004



# Outline

- 1 Background
  - Dominance and Pareto Optimality
  - Simulated Annealing
  - Multi-Objective SA with Composite Objective Functions
- 2 A Truly Multi-Objective SA
  - A Dominance Based Energy Function
- 3 Increasing Energy Resolution
  - Attainment Surface Sampling
- 4 Realtime Algorithm Parameter Optimisation
  - Parameter Scaling Adjustment
- 5 Results
  - Test Problems
  - CDMA Network Optimisation



# Dominance and Pareto Optimality

- Minimise  $D$  objectives,  $y_i$ , functions of  $P$  decision variables,  $\mathbf{x} = (x_1, x_2, \dots, x_P)$ :

$$y_i = f_i(\mathbf{x}), \quad i = 1, \dots, D$$

- A multi-objective optimisation problem may be expressed as:

$$\text{Minimise } \mathbf{y} = \mathbf{f}(\mathbf{x}) \equiv (f_1(\mathbf{x}), \dots, f_D(\mathbf{x}))$$

- Dominance is used to compare two solutions  $\mathbf{a}$  and  $\mathbf{b}$ . If  $\mathbf{f}(\mathbf{a})$  is no worse for all objectives than  $\mathbf{f}(\mathbf{b})$  and wholly better for at least one objective it is said that  $\mathbf{a}$  *dominates*  $\mathbf{b}$ , written  $\mathbf{a} \prec \mathbf{b}$ .  $\mathbf{a} \prec \mathbf{b}$  iff:

$$f_i(\mathbf{a}) \leq f_i(\mathbf{b}) \quad \forall i = 1, \dots, D \quad \text{and} \\ f_j(\mathbf{a}) < f_j(\mathbf{b}) \quad \text{for at least one } i.$$



# Dominance and Pareto Optimality

- Two solutions are *mutually non-dominating* if neither dominates the other. A set  $F$  of solutions is a non-dominating set if no element of the set dominates any other:

$$\mathbf{a} \not\prec \mathbf{b} \quad \forall \mathbf{a}, \mathbf{b} \in F$$

- A solution is said to be globally non-dominated, or Pareto-optimal, if no other feasible solution dominates it.
- The set of all Pareto-optimal solutions is known as the Pareto-optimal front, or the Pareto set,  $\mathcal{P}$ .



# Simulated Annealing

- Computational analogue of slowly cooling a metal so that it adopts a low-energy, crystalline state.
- Minimisation function: *energy*  $E(\mathbf{x})$ , of the state  $\mathbf{x}$ , with a temperature  $T$  which is reduced to (near-)zero during optimisation.
- A new 'state' (solution) is generated from a perturbation of the current state/solution.
- Acceptance Probability of the new solution:

$$A = \min(1, \exp\{-\delta E(\mathbf{x}', \mathbf{x})/T\})$$

where

$$\delta E(\mathbf{x}', \mathbf{x}) \equiv E(\mathbf{x}') - E(\mathbf{x})$$



# Multi-Objective SA with Composite Objective Functions

- Single-objective annealers can be used on multi-objective problems if there exists a mapping from the multi-dimensional to single-dimensional space



# Multi-Objective SA with Composite Objective Functions

- Single-objective annealers can be used on multi-objective problems if there exists a mapping from the multi-dimensional to single-dimensional space
- The objectives can be added together to form a single objective (possibly with some multiplicative weighting, many variations exist)

# Multi-Objective SA with Composite Objective Functions

- Single-objective annealers can be used on multi-objective problems if there exists a mapping from the multi-dimensional to single-dimensional space
- The objectives can be added together to form a single objective (possibly with some multiplicative weighting, many variations exist)
- This isn't ideal.

# Multi-Objective SA with Composite Objective Functions

- Single-objective annealers can be used on multi-objective problems if there exists a mapping from the multi-dimensional to single-dimensional space
- The objectives can be added together to form a single objective (possibly with some multiplicative weighting, many variations exist)
- This isn't ideal.
- The search concentrates on a single solution, and knowledge of the search space need not be gained.

# Multi-Objective SA with Composite Objective Functions

- Single-objective annealers can be used on multi-objective problems if there exists a mapping from the multi-dimensional to single-dimensional space
- The objectives can be added together to form a single objective (possibly with some multiplicative weighting, many variations exist)
- This isn't ideal.
- The search concentrates on a single solution, and knowledge of the search space need not be gained.
- At worst there's a potential for the search to be compromised through rejection of pareto-optimal solutions.



# A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .



## A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .
- In the multi-objective case optimum solutions are only defined in relation to each other.



## A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .
- In the multi-objective case optimum solutions are only defined in relation to each other.
- We can compare the relative quality of  $\mathbf{x}$  and  $\mathbf{x}'$  with the dominance relation...



# A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .
- In the multi-objective case optimum solutions are only defined in relation to each other.
- We can compare the relative quality of  $\mathbf{x}$  and  $\mathbf{x}'$  with the dominance relation...
- But it gives only three values of quality: better, worse, or equal. A continuum is usually given in uni-objective problems.



## A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .
- In the multi-objective case optimum solutions are only defined in relation to each other.
- We can compare the relative quality of  $\mathbf{x}$  and  $\mathbf{x}'$  with the dominance relation...
- But it gives only three values of quality: better, worse, or equal. A continuum is usually given in uni-objective problems.
- If the Pareto front  $\mathcal{P}$  were available, we could define a simple energy of  $\mathbf{x}$  as the measure of the front that dominates  $\mathbf{x}$ .



# A Dominance Based Energy Function - The Perfect Case

- In single objective optimisation problems the energy  $E(\mathbf{x})$  is an absolute measure of the quality of any solution  $\mathbf{x}$ .
- In the multi-objective case optimum solutions are only defined in relation to each other.
- We can compare the relative quality of  $\mathbf{x}$  and  $\mathbf{x}'$  with the dominance relation...
- But it gives only three values of quality: better, worse, or equal. A continuum is usually given in uni-objective problems.
- If the Pareto front  $\mathcal{P}$  were available, we could define a simple energy of  $\mathbf{x}$  as the measure of the front that dominates  $\mathbf{x}$ .

- Let  $\mathcal{P}_{\mathbf{x}}$  be the portion of  $\mathcal{P}$  that dominates  $\mathbf{x}$ :

$$\mathcal{P}_{\mathbf{x}} = \{\mathbf{y} \in \mathcal{P} \mid \mathbf{y} \prec \mathbf{x}\}$$

Then we define  $E(\mathbf{x}) = \mu(\mathcal{P}_{\mathbf{x}})$  where  $\mu$  is a measure defined on  $\mathcal{P}$ .



# A Dominance Based Energy Function - Without the Pareto Front

- We propose an energy function defined in terms of the current estimate of the Pareto front,  $F$ .



## A Dominance Based Energy Function - Without the Pareto Front

- We propose an energy function defined in terms of the current estimate of the Pareto front,  $F$ .
- We define  $\tilde{F}$  to be the union of the  $F$ , the current solution  $\mathbf{x}$  and the proposed solution  $\mathbf{x}'$ , that is  $\tilde{F} = F \cup \mathbf{x} \cup \mathbf{x}'$ .



## A Dominance Based Energy Function - Without the Pareto Front

- We propose an energy function defined in terms of the current estimate of the Pareto front,  $F$ .
- We define  $\tilde{F}$  to be the union of the  $F$ , the current solution  $\mathbf{x}$  and the proposed solution  $\mathbf{x}'$ , that is  $\tilde{F} = F \cup \mathbf{x} \cup \mathbf{x}'$ .
- Then let  $\tilde{F}_{\mathbf{x}}$  be the elements of  $\tilde{F}$  that dominate  $\mathbf{x}$ :

$$\tilde{F}_{\mathbf{x}} = \{\mathbf{y} \in \tilde{F} \mid \mathbf{y} \prec \mathbf{x}\}$$



## A Dominance Based Energy Function - Without the Pareto Front

- We propose an energy function defined in terms of the current estimate of the Pareto front,  $F$ .
- We define  $\tilde{F}$  to be the union of the  $F$ , the current solution  $\mathbf{x}$  and the proposed solution  $\mathbf{x}'$ , that is  $\tilde{F} = F \cup \mathbf{x} \cup \mathbf{x}'$ .
- Then let  $\tilde{F}_{\mathbf{x}}$  be the elements of  $\tilde{F}$  that dominate  $\mathbf{x}$ :

$$\tilde{F}_{\mathbf{x}} = \{\mathbf{y} \in \tilde{F} \mid \mathbf{y} \prec \mathbf{x}\}$$

- So we obtain an energy difference between the current and proposed solutions of

$$\delta E(\mathbf{x}, \mathbf{x}') = \frac{1}{|\tilde{F}|} (|\tilde{F}_{\mathbf{x}}| - |\tilde{F}_{\mathbf{x}'}|)$$



# A Dominance Based Energy Function - Without the Pareto Front

- Division by  $|\tilde{F}|$  provides some robustness against fluctuations in the number of solutions in  $F$



## A Dominance Based Energy Function - Without the Pareto Front

- Division by  $|\tilde{F}|$  provides some robustness against fluctuations in the number of solutions in  $F$
- The inclusion of the current solution and the proposal in  $\tilde{F}$  ensures that proposals that move the estimated front towards the true front are always accepted.



# A Dominance Based Energy Function - Without the Pareto Front

- Division by  $|\tilde{F}|$  provides some robustness against fluctuations in the number of solutions in  $F$
- The inclusion of the current solution and the proposal in  $\tilde{F}$  ensures that proposals that move the estimated front towards the true front are always accepted.
- This method has an interesting and beneficial consequence; it promotes exploration of the front by assigning high energies to highly populated regions of the front



## Simulated Annealing - Algorithm

$\{L_k\}_{k=1}^K$  Sequence of epoch durations  
 $\{T_k\}_{k=1}^K$  Sequence temperatures,  $T_{k+1} < T_k$   
 $\mathbf{x}$  Initial feasible solution  
for  $k := 1, \dots, K$   
  for  $i := 1, \dots, L_k$   
     $\mathbf{x}' := \text{perturb}(\mathbf{x})$   
     $\delta E := E(\mathbf{x}') - E(\mathbf{x})$   
     $u := \text{rand}(0, 1)$   
    if  $u < \min(1, \exp(-\delta E/T_k))$   
       $\mathbf{x} := \mathbf{x}'$   
    end  
  end  
end

# Increasing Energy Function Resolution to Counter Sparse Fronts

- A sparse front allows only a small number of potential energies.



# Increasing Energy Function Resolution to Counter Sparse Fronts

- A sparse front allows only a small number of potential energies.
- Increasing the number of points in the estimated front will prevent this behaviour.



# Increasing Energy Function Resolution to Counter Sparse Fronts

- A sparse front allows only a small number of potential energies.
- Increasing the number of points in the estimated front will prevent this behaviour.
- We can sample from the Attainment surface to achieve this.



# Attainment Surface Sampling

- The attainment surface,  $\mathcal{S}_F$ , is an interpolating surface between the elements of  $F$  that has suitable properties.
- $\mathcal{S}_F$ , is a conservative interpolation of the elements of  $F$  so that every point of  $\mathcal{S}_F$  is dominated by an element of  $F$
- $\mathcal{S}_F$  is the boundary of the region in objective space which is dominated by elements of  $F$ .



# Attainment Surface Sampling

- The attainment surface,  $\mathcal{S}_F$ , is an interpolating surface between the elements of  $F$  that has suitable properties.
- $\mathcal{S}_F$ , is a conservative interpolation of the elements of  $F$  so that every point of  $\mathcal{S}_F$  is dominated by an element of  $F$
- $\mathcal{S}_F$  is the boundary of the region in objective space which is dominated by elements of  $F$ .
- We use random samples, uniformly distributed on  $\mathcal{S}_F$  to interpolate  $F$ .



# Attainment Surface Sampling

- The attainment surface,  $\mathcal{S}_F$ , is an interpolating surface between the elements of  $F$  that has suitable properties.
- $\mathcal{S}_F$ , is a conservative interpolation of the elements of  $F$  so that every point of  $\mathcal{S}_F$  is dominated by an element of  $F$
- $\mathcal{S}_F$  is the boundary of the region in objective space which is dominated by elements of  $F$ .
- We use random samples, uniformly distributed on  $\mathcal{S}_F$  to interpolate  $F$ .
- Sampling from  $\mathcal{S}_F$  is performed by sampling a point from a uniform distribution on the axis-parallel hyper-rectangle bounding  $F$  and then restricting one coordinate so that the point is dominated by an element of  $F$ .



# Automatic Configuration of Model Parameters

- The algorithm should be able to statistically determine the best parameters to use within the model.



# Automatic Configuration of Model Parameters

- The algorithm should be able to statistically determine the best parameters to use within the model.
  - Initial Temperature



# Automatic Configuration of Model Parameters

- The algorithm should be able to statistically determine the best parameters to use within the model.
  - Initial Temperature
    - Set so initial acceptance rate of dominated moves is 50%



# Automatic Configuration of Model Parameters

- The algorithm should be able to statistically determine the best parameters to use within the model.
  - Initial Temperature
    - Set so initial acceptance rate of dominated moves is 50%
  - Perturbation Sizes



# Adjusting the Scale of Parameter Perturbations

- There are two distinct types of movement in objective space:
  - Traversal Scalings; These move along a front
  - Location Scalings; These move towards or away from fronts.
- These need to be treated differently

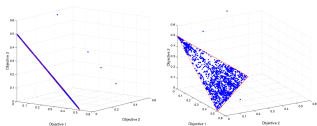


# Experiments

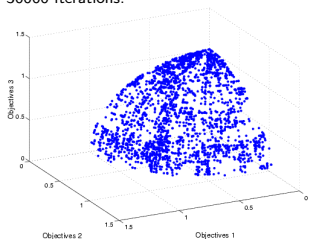
- Test Problems
  - DTLZ Suite
- Industrial Problem
  - CDMA Network Optimisation



# DTLZ Test Problems



The annealer archive on test problem DTLZ1 after 30000 iterations.

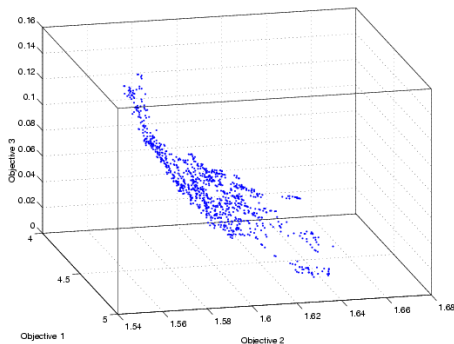


The annealer archive on test problem DTLZ3 after 30000 iterations.

- Test problems DTLZ1, and DTLZ3 are minimisation problems, used here in their standard 3-objective formulation.
- The true front of DTLZ1 lies on the plane where the sum of all objectives equals 0.5.
- The front of DTLZ3 lies on the spherical front where the sum of the squared objectives equals 1.0.

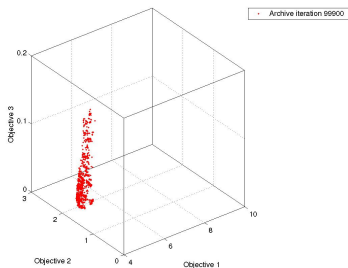


# CDMA Network Optimisation



- Mobile phone network parameter optimisation.
- Adjust the *Pilot Powers* of the network.
- Three objectives, evaluations of overall network quality.

# A Movie - A day in the life of a Pareto front.



Formation of an estimated Pareto front for CDMA network optimisation.

# Conclusions

