Optimisation tools beyond QMC

Mariapia Marchi (Numerical Methods Group, Esteco S.p.A, Trieste, Italy)
Outline

• Introductory shopping list: nomenclature, multi-objective optimisation problems and Pareto dominance,…
• Brief overview of genetic algorithms in general
• MOGAII and NSGAII
• Just to play a little bit…
• Few applications: integration of modeFRONTIER in material modelling studies
• Conclusions
Inputs & Outputs

• **Input variables**: quantities that a designer can vary or choices that he/she can make.
  - **Continuous variables**:
    - point coordinates
    - process variables,…
  - **Discrete variables**:
    - components from a catalogue
    - number of components,…

• **Objectives**: response parameters, i.e. the quantities that the designer wish to be MAX or MIN
  - **MAX**
    - Efficiency, performance, …
  - **MIN**
    - Cost, weight, …

  (A MAX problem can always be transformed into a MIN problem)

• **Constraints**: restrictions and limits the designer must meet due to norms, functionalities, etc. They define a feasible region.

  **General constraints**
  - Max admissible stress, deformation, acceleration, …
  - Min performance, …

  **Constraints on variables**
  - Total volume, thickness, weight, …
  - Relations holding between variables, …
Multi-objective optimisation problems: def.

\[
\begin{align*}
\text{min} & \{ f_1(x_1, \ldots, x_n), \ldots, f_k(x_1, \ldots, x_n) \}, \\
& g_i(x) \leq 0 \text{ for some } i \in \mathbb{N}, \\
& h_j(x) \geq 0 \text{ for some } j \in \mathbb{N}, \\
& l_m(x) = 0 \text{ for some } m \in \mathbb{N}.
\end{align*}
\]

**OBJECTIVES**

**CONSTRAINTS**

**OPTIMISATION PROBLEM**

If \( k > 1 \) and conflicting objectives, then **MULTI-OBJECTIVE!!**

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Weighted Function:

n objectives can be added in a single objective using **weights**: 

\[ F(x) = w_1 \cdot \text{Obj}_1 + w_2 \cdot \text{Obj}_2 + w_3 \cdot \text{Obj}_3 \ldots \]

**Pro:**

simple formulation

**Cons:**

weights are problem-dependent and must be empirically defined (which value?)

weights are connected to objective values and might lose significance for different physics (how to sum pears with apples?)

**Multi-objective formulation is generally more accurate and is to be preferred**
Multi-Objective Optimisation Problems: solution

Pareto-Frontier, set of optimal solutions:

Dominated design: exist solutions with better (lower) values of both objectives

Pareto front: no solutions exist with better values for both objectives

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Pareto Dominance:

Design \textbf{a dominates} Design \textbf{b} if:

\[ f_1(a) \leq f_1(b) \text{ and } f_2(a) \leq f_2(b) \text{...and } f_n(a) \leq f_n(b) \]

\textbf{and}

it exists \( i \) in \{1,\ldots,n\} such that \( f_i(a) < f_i(b) \)

- In the Pareto frontier none of the components can be improved without deterioration of at least one of the other components.

- Pareto dominance for one objective coincides with a classical optimization approach.

- Pareto dominance defines a group of efficient solutions: in case of \( n \) objectives, the group of efficient solutions contains at Max \( \infty^{(n-1)} \) points
Robustness vs. Accuracy

**Robustness** is the ability of an optimisation algorithm to reach the absolute extreme of an objective function.

**Accuracy** measures the capability of the optimisation algorithm to find the function’s extreme.

- **Robust algorithms** reach global extremes
- **Non-Robust algorithms** get stuck in local extremes
Genetic Algorithms use the analogy of natural selection and reproduction as an optimisation target.

**PRO:**
high robustness, multi-objective optimisation is possible, independence from structure of constraints and objectives (no derivatives!), ability to locate global optimum

**CON:**
low convergence rate if high accuracy is required, high computational cost
Genetic algorithms loosely parallel biological evolution and are based on Darwin's theory of natural selection.

The specific mechanics of the algorithms involve the language of microbiology and, in developing new potential solutions, mimic genetic operations.

- A population represents a group of potential solution points.
- A generation represents an algorithmic iteration.
- A chromosome is comparable to a design point, and a gene is comparable to a component of the design vector.
Genetic Algorithms: basics

One generic design (individual)

<table>
<thead>
<tr>
<th>Design Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>220</td>
</tr>
<tr>
<td>2</td>
<td>1.34</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>0.978</td>
</tr>
</tbody>
</table>

Initial screening of designs starting population

Combination between designs (input parameters):

**CROSSOVER**, **MUTATION**, ...

Encoding problem:
- Binary representation
- Continuous operators

\[ x = a + \frac{b - a}{2^l - 1} DV(s) \]

- **SELECTION** of best designs
  - **fitness** criteria for S.O. cases
  - **dominance** criteria for M.O. cases

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
• **Individual**: a configuration with associated variables and objectives

\[ f: \mathbb{R}^n \Rightarrow \mathbb{R}^m \]

\[ i^{th} \text{ individual} \]

\[ (x_1^i, \ldots, x_n^i) \Rightarrow (f_1^i, \ldots, f_m^i) \]

• **Main operators**
  1) Selection
  2) Cross-over
  3) Mutation

**Problem Definition**

**Definition of objectives**

**First Generation** definition (D.O.E.)

**Evaluation** of designs

**Parent set** definition (dominance)

**Genetic operations to create children**

**Creation of the next generation**

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Scope:

- **transfer the best characteristics** to the next generation by selecting design vectors from the current generation to be used in the next generation.
- **probability of selection** proportional to **fitness function (single objective)** or **domination criteria (multi-objective)**

**Roulette Selection**

- F1 > F2 > F3 > F4

**Local selection**

\[
\begin{array}{ccc}
& i-k & \\ i-l & i & i+1 \\
& i+k & \\
\end{array}
\]

\[k = \sqrt{\text{nind}}\]
**Scope:**

- Define a new individual for next generations, keeping best information of parents

Binary crossover

\[ v_i^* = v_{i1} + \text{sign}(F1-F2)*(v_{i1} - v_{i2}) + T*\text{sign}(F1-F3)*(v_{i1} - v_{i3}) \]

Directional crossover

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Scope:

- Operator used to avoid premature convergence, it makes a random modification of one variable
### DOE & Schedulers

#### Design of Experiments
- User Sequence
- Random Sequence
- Constraint Satisfaction Problem
- Sobol
- Latin Hypercube - Monte Carlo
- Full Factorial
- Reduced Factorial
- Cubic Face Centered
- Box-Behnken
- Latin Square
- Taguchi Matrix
- Plackett Burman
- D-Optimal
- Incremental Space Filler
- Inscribed Composite Design
- Uniform Reducer
- Dataset Reducer

#### Schedulers and Optimizers

**DOE Sequence**
- DOE Sequence
- Multi-variate Adaptive Kriging (MACK)

**Basic Optimizers**
- Simplex
- B-BFGS
- Levenberg-Marquardt
- Simulated Annealing (SA)
- Multi-Objective Genetic Algorithm (MOGA-II)
- ARMOGA

**Advanced Optimizers**
- Multi-Objective Simulated Annealing (MOSA)
- NSGA-II
- Multi-Objective Game Theory (MOGT)
- FMOGA-II
- FSimplex
- Lipschitz Sampling
- Multi-Objective Particle Swarm Optimization (MOPSO)

#### Evolution Strategies
- 1P1-ES
- Derandomized Evolution Strategy (DES)
- Multi-Membered Evolution Strategy (MMES)
- Evolution Strategy

#### Sequential Quadratic Programming
- NLPQLP
- NBI-NLPQLP

#### Robust Design Methods
- Robust Design
- Polynomial Chaos for MORDO
Definition of **parent set** at each generation – **Elitism** enabled

- i-th generation ($n$ individuals)

Not dominated-designs from all Design Space (ONLY IF **ELITISM** ENABLED)

- Parent set ready for genetic operations

- Parent set randomly reduced to size $n$

- Put two sets together

---

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Adaptive penalty is scaled in function of the range of variation of the function.

- \( G(x) \) is sum of constraint violations

Note: always normalise constraints! Otherwise, \( G(x) \) is influenced by higher scale ones.
NSGA-II = Non-dominated Sorting Genetic Algorithm II developed by K. Deb – KanGAL.

- It is a fast and elitist multi-objective evolutionary algorithm.

References:
NSGA-II

Main features:

• A fast and clever non-dominated sorting procedure is implemented: cost is \(O(MN^2)\), \(M=\text{#obj.}\) and \(N=\text{pop.}\).

• It implements elitism for multi-objective search, using an elitism-preserving approach.

• A parameter-less diversity preservation mechanism is adopted. Diversity and spread of solutions is guaranteed without use of sharing parameters, adopting a suitable parameter-less niching approach.

• The constraint handling method does not make use of penalty parameters.

• NSGA-II allows both continuous ("real-coded") and discrete ("binary-coded") design variables.
NSGA II: Parent set & Elitism

Definition of **parents** at each generation - **Elitism**

(i-1)-th generation \((n)\) individuals

i-th generation \((n)\) individuals

Parent set originally made by last 2 generations \((2*n)\) individuals

**Ranking** definition (next slide)

Keep in parent set only the first best \(n\) designs

Parents ready for genetic operations

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
NSGA II: Ranking procedure and selection fitness

Ranking procedure for parent set definition and selection fitness

- **First, define front order:**
  - 4 is better than 5 because it dominates 5.
  - 1, 2, 3 are better than 4 because they dominate it.

- **Second, crowding distance parameter:**
  order the points of same front maximising their mean reciprocal distance:
  - 1 is the best point since it is located in the less crowded region,
  - 3 is the worst.
NSGA II: constraint handling

- **No penalty function is defined**, but constraints are taken into account directly during ranking procedure.

- **If A and B are both feasible**, ranking is defined as before (first front order, then crowding distance).

- **If A is feasible and B is unfeasible**, \( \text{rank}(A) \) is always higher than \( \text{rank}(B) \).

- **If A and B are both unfeasible**, higher ranking is assigned to designs with lower sum of constraints violation.
Applications

- Just to play a little bit
- Applications to material modelling
Energy as a functional of $r_s$ and $\zeta$

\[
E(r_s, \zeta) = \epsilon_0(r_s, \zeta) + e_x(r_s, \zeta) + e_{\text{corr}}(r_s, \zeta),
\]
\[
e_x(r_s, \zeta) = -\frac{4}{3\pi r_s} \frac{1}{(1 + \zeta)^{3/2} + (1 - \zeta)^{3/2}}.
\]
\[
\epsilon_0(r_s, \zeta) = \frac{1 + \zeta^2}{2r_s^2}.
\]
\[
e_{\text{corr}}(r_s, \zeta) = (e^{-\beta r_s} - 1)e_x^{(6)} + \alpha_0(r_s) + \alpha_1(r_s)\zeta^2 + \alpha_2(r_s)\zeta^4,
\]
\[
\alpha_i(r_s) = A_i + (B_i r_s^2 + C_i r_s^2 + D_i r_s^3) \cdot \ln \left(1 + \frac{1}{E_i r_s^3 + F_i r_s^{3/2} + G_i r_s^2 + H_i r_s^3}\right).
\]


1 valley:
A couple of tests…

TEST 1:
Minimisation of $E(r_s, \zeta)$ for 2valley case (single obj.)
Minimum found for $r_s = 0.97$ and $\zeta = 0$ (as expected).

TEST 2:
Minimisation of $E_{1v}(r_s, \zeta)$ and $E_{2v}(r_s, \zeta)$ (multi-obj).
A Pareto front is found.
Optimal solutions have $\zeta = 0$ and $r_s$ close to the extreme value for 2v system.
Concepts behind modeFRONTIER

• Input Variables:
  Entities defining the design space

• Output Variables:
  Measures from the system

• A Black Box:
  Generates the outputs according to the inputs

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
The black box can be:

- A set of solvers that model and solve numerically the design problem (e.g. CAD/CAE tools)
- A set of experiments that produce data
The following two examples are taken from:

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Example 1: Composite materials

- MOTIVATION: Effective **stiffness** of composites consisting of a polymer and two fillers of different particle size **not predicted correctly by additive approach**.


- Particular case: **GLASS BEAD COMPOSITES**
  - Easier processing
  - Better surface properties (scratch, abrasion resistance)
  - Good dimensional stability
  - Cheaper than fibres
  - But worse mechanical properties
  - What is the **best compromise**?

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
15% beads: small reduction in strength

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
DIGIMAT-MF integration in modeFrontier

- Variation of glass bead fraction
- Minimize thermal anisotropy.
- Maximize stress at full load as indicator of strength.
Results

Future: Integration of atomistic level?

Bubble size indicates amount of glass beads

QMC in the Apuan Alps VII, 28 July – 3 August 2012, Vallico di Sotto (Italy)
Example 2: atomistic modelling in engineering

Tipical procedure in ATOMISTIC MODELLING WORKFLOWS:

- **Assign Forcefield**
  - Pre-determined
  - Usually kept fixed.

- **Equilibrate Structure**
  - Procedure depends on user preferences
  - “Black art”

- **Simulate Properties**
  - If no good, try a different forcefield, or change the equilibration protocol.

ENGINEERING-OPTIMISED workflow:

- **Parametric models**
  - Forcefield and Equilibration procedure

- **Objectives and Constraints**
  - Target properties

**Multi-objective optimization**

**Optimal trade-off Solution**
Multi-objective optimization in atomistic modelling of adsorbents

Multi-objective optimization of Specialized Force-Fields for gas adsorption in zeolite adsorbents

Frédéric Bellencontre  Pluton Pullumbi
Air Liquide R&D
Centre de Recherche CLAUDE-DELORME
78354 Jouy-en-Josas, FRANCE
Characteristics of this study

- Molecular simulation approach allows shortening time and effort to generate physical properties used in process development.
- Specialised Force-Fields that realistically describe gas adsorption (pure compounds and mixtures) in well defined zeolite adsorbents are needed.
- modeFRONTIER is used to obtain optimised Force-Fields.
Pseudo-experimental data for gas adsorption in zeolites

Molecular Simulation of gas adsorption in Zeolites

Understand parameters that control adsorption at a microscopic level
GCMC simulations (Towhee code)

Reproduce experimentally observed trends for pure compounds
- Adsorption isotherms
- Henry constants
- Isosteric heats

Predict multi-component adsorption of gas mixtures

The characterization of multi-component gas equilibria in Zeolite adsorbents requires considerable time and experimental effort.
• Zeolite structure from experimental data (mostly)
• Zeolite-sorbate and sorbate-sorbate interactions in terms of a Lennard-Jones interaction potential.
Workflow for force-field optimization

1. Initial FF param.
   Atomistic LJ param.
   Partial charges

2. GCMC simulations
   Towhee code
   Compare to Exp. Ads.Iso.
   Single component
   Ads. Isotherms
   3 different Temp.

3. Single component
   Param. Validation
   Multi component
   Ads. Isotherms
   Co-Ads. bench

4. Parameter Validation with Co-Ads data
   Co-adsorption Data generation

5. ModeFRONTIER.
   Multi-objective FF optimization
**Forcefield optimisation with modeFrontier**

FF optimization protocol

- Input variables: parameters of the FF
- Objectives: minimization of difference between exp. and sim. values for each gas

- Objective function:
  \[ F(X) = \sum_{i=1}^{n} W_i \left( \frac{f_{i,\text{exp}} - f_{i,\text{sim}}(x)}{f_{i,\text{exp}}} \right)^2 \]

- DOE: Sobol (80 design initial population)
- Optimisation: MOGAII (20 generations)
Validated workflow

Adsorption of multi-component mixtures in a given zeolite can be predicted with good accuracy, based on the proposed workflow:

1. **Atomistic simulation of single gas adsorption**
2. **Exper. adsorption isotherms: several gases, range of T,p**
3. **Frontier optimised, specialized forcefield: match single component and few multicomponent experiments**
4. **Prediction of multicomponent mixture adsorption**
Conclusions

• Genetic algorithm are robust optimisation tools which can deal efficiently with multi-objective optimisation problems.
• Useful applications in material modelling, combined together with mF workflow
• I would like to investigate other physical applications related to QMC…
A genetic algorithm for the 1D electron gas

César O. Stoico\textsuperscript{a}, Danilo G. Renzi\textsuperscript{b}, Fernando Vericat\textsuperscript{c,d,*}

\textsuperscript{a} Área Física, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Argentina
\textsuperscript{b} Facultad de Ciencias Veterinarias, Universidad Nacional de Rosario, Casilda, Argentina
\textsuperscript{c} Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB), CONICET-UNLP, c.c. 365, 1900 La Plata, Argentina
\textsuperscript{d} Grupo de Aplicaciones Matemáticas y Estadísticas de la Facultad de Ingeniería (GAMEFI), UNLP, La Plata, Argentina

Received 26 October 2006; received in revised form 9 March 2007
Available online 11 September 2007

Computation of nodal surfaces in fixed-node diffusion Monte Carlo calculations using a genetic algorithm

Jordan A. Ramlowski* and David Farrelly*

Received 27 April 2010, Accepted 22nd July 2010
DOI: 10.1039/c0cp00373e
Aknowlegments

• Mike (for his invitation)
• My chief (for letting me go 😊!)
• The marketing group (for help in finding material)
• Danilo Di Stefano (for useful discussions)
• Gerhard Goldbeck (for his slide material)

• THANK YOU FOR YOUR ATTENTION!
Enable your imagination

Thank you!

OMG in the Apuan Alps VII; 28 July – 3 August 2012, Vallico di Sotto (Italy)
Structure of the Zeolite

- Framework model (structure from X-ray diffraction)
- Si/Al ratio (from NMR spectra)
- The nature of non framework cations (from XPS...)
- Configuration of non framework cations (from Neutron diffraction)
- Accessible volume (from simulation)

Interaction Potential

- Atom-based pair interaction
- Van der Waals + electrostatic interactions
- Zeolite-Sorbate
- Sorbate-Sorbate

\[ E_{ij} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \]

Repulsion  Dispersion  Electrostatic

Each atom of the zeolite or of the sorbate is characterized by its LJ parameters and its charge. For gas molecules point charges could be localized at different positions in order to correctly represent the multipole property of the sorbate molecule.
Experimental data and simulations

Single component adsorption isotherms
- List of sorbates (A, B, C...)
- Measurements at 3 to 5 Temperatures
- Large Pressure range (0-50 bars)
- Zeolite adsorbent structure (known)
- Zeolite composition (known)

For example: NaLSX zeolite
- Framework structure: FAU
- Unit Cell composition:
  - 384O, 96Si, 96Al, 96Na
- Atomic positions: known (X-ray diffr.)

Grand Canonical Monte Carlo (GCMC) simulation of adsorption
- Towhee code as implemented in MAPS (Scienomics)
- Gas fugacities from EOS (SAFT or PC-SAFT)
- Configurational-biased MC settings selected to provide correct chemical potentials
- Zeolite framework structure fixed; Cation position fixed. Sorbates rigid.
- Free volume analysis for location of hard spheres blocking inaccessible sites
- Ewald summation for electrostatic interactions
- Initial FF from literature
  - For NaLSX zeolite: 4 type of atoms = 4*3 =12 parameters
  - For each gas molecule: 3*n_type_of_atoms, + geometric parameters.

- Python script for running GCMC simulations for 6 to 10 Pressures and 3 different Temperatures for each gas selected from experimental data.