TIES598 Nonlinear Multiobjective Optimization

spring 2017

Jussi Hakanen
firstname.lastname@jyu.fi
General information

- Master (/PhD) level course in mathematical information technology, 5 credits
- Suitable e.g. for master students in computational sciences
- [https://korppi.jyu.fi/kotka/r.jsp?course=198878](https://korppi.jyu.fi/kotka/r.jsp?course=198878)
- Homepage: [http://users.jyu.fi/~jhaka/ties598/](http://users.jyu.fi/~jhaka/ties598/)
- On Tuesdays at 10.15 and on Thursdays at 14.15, March 14th – May 18th, 2017
- Room: AgD 121
- Mailing list: ties598_2017@korppi.jyu.fi
Contents

- Introduction to multiobjective optimization (MOO)
- Problem formulation
- MOO methods
  - Multiple Criteria Decision Making
  - Evolutionary Multiobjective Optimization
- Solving practical MOO problems
- Approximation methods in MOO
- MOO software
- Visualization
Learning outcomes

- Understand why MOO methods are needed
- Understand basic concepts in solving MOO problems
- Understand optimality in MOO
- Understand different approaches to solve multiobjective optimization problems
- Understand basics of choosing and implementing MOO methods
- Know how to find and apply software for solving MOO problems
Passing the course

Course consists of

- Lectures (Tuesdays)
- Discussion sessions (Thursdays)
- Assignments (take home)

Grading

- Assignments: 70%
- Active participation in group discussions: 30%
How demanding the course is?

- 5 credits $\rightarrow$ $5 \times 26 = 130$ hours of work
- Lectures: $9 \times 2h = 18$ hours
- Group discussion: $7 \times 2h = 14$ hours
- Assignments: $7 \times 8h = 56$ hours
- Self study: $130 - 88 = 42$ hours
Study practices in the course

- Lectures are on Tuesdays
- Discussion session on Thursdays
  - Material is given to you beforehand that you should study before the session
  - During the session, a task related to the material is given that you should discuss in your group
  - Every group prepares a presentation (slides) during the session which is then presented to other groups
  - Be active in the group → a part of grading
- Assignments are given from each topic of the course that you should do at home and return a report
  - Major part of grading
Let’s introduce ourselves

- Who are you and what are you studying?
- What is the phase of your studies (MSc/PhD)?
- Previous experiences of (multiobjective) optimization?
- How is this course related to your studies/research?
- What are your expectations about this course?
Outline of this lecture

- Very brief introduction to single objective optimization
  - Traditional methods
  - Evolutionary methods
- What means multiobjective optimization
  - Differences to single objective optimization
- How to characterize optimal solutions
- Solution approaches for multiobjective problems
SINGLE OBJECTIVE OPTIMIZATION
Single objective optimization problem

\[
\begin{align*}
\text{min } & \quad f(x) \\
\text{s.t. } & \quad g_i(x) \leq 0 \text{ for all } i \in I \\
& \quad h_j(x) = 0 \text{ for all } j \in J \\
& \quad x \in \mathbb{R}^n
\end{align*}
\]

- Objective function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \)
- Decision variable vector \( x = (x_1, ..., x_n)^T \)
- Feasible region \( S = \{ x \in \mathbb{R}^n \mid g_i(x) \leq 0 \ \forall \ i \ \& \ h_j(x) = 0 \ \forall \ j \} \subset \mathbb{R}^n \)
- Inequality constraints: \( g_i: \mathbb{R}^n \rightarrow \mathbb{R}, i \in I \)
- Equality constraints: \( h_j: \mathbb{R}^n \rightarrow \mathbb{R}, j \in J \)
Optimal solution to a single objective optimization problem

• Solution $x \in S$ is **optimal**, if there does not exist a $y \in S$ such that $f(y) < f(x)$

• When **no constraints** and objective function twice differentiable, optimal solution must satisfy **gradient zero** and **hessian positive semidefinite**

\[
\nabla f = \frac{\partial f}{\partial x_1} e_1 + \cdots + \frac{\partial f}{\partial x_n} e_n
\]

\[
H(f) = \begin{bmatrix}
\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
\end{bmatrix}
\]
Optimal solution to a single objective optimization problem

For a continuously differentiable problem there must exist Karush-Kuhn-Tucker (KKT) multipliers $\mu_i \geq 0$, $\lambda_j$ such that

$$-\nabla f(x) = \sum_{i \in I} \mu_i \nabla g_i(x) + \sum_{j \in J} \lambda_j \nabla h_j(x)$$

and $\mu_i g_i(x) = 0$ for all $i \in I$

Unfortunately, there is no direct way to find all the points that satisfy KKT conditions

This is why one must resort to either iterative or population-based approaches
Choosing the optimization method

- Different optimization methods
  - have different requirements
  - can use different information (e.g., gradients)
  - perform differently with different problems

- No absolute truth can be said about which method to choose for different problems

- Best results can be gained, by combination of optimization methods
Choosing the optimization method: a rule of thumb

- Optimization problems can be located on different dimensions
- Optimization methods can be placed on the same dimensions based on whether they handle them
- Ideally, the optimization method should handle just the amount of complexity, but not any more, because that usually comes with convergence trade-off

<table>
<thead>
<tr>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Convex</td>
</tr>
<tr>
<td>Unimodal</td>
<td>Multimodal</td>
</tr>
<tr>
<td>No constraints</td>
<td>Linear constraints</td>
</tr>
<tr>
<td>Differentiable</td>
<td>Continuous nondifferentiable</td>
</tr>
<tr>
<td>Real-valued variables</td>
<td>Discrete variables</td>
</tr>
<tr>
<td>Deterministic</td>
<td>Stochastic</td>
</tr>
</tbody>
</table>
Examples of optimization methods: the Simplex method

- Listed as one of the 10 most influential algorithms of the 20th century*
- Requires a linear problem with continuous variables
- Performs well, when the number of constraints is relatively low
- Works by moving from one corner point to another

\[
\begin{align*}
g_1(x) &\leq 0 \\
g_2(x) &\leq 0 \\
g_3(x) &\leq 0 \\
g_4(x) &\leq 0 \\
g_5(x) &\leq 0 \\
f &\geq 0
\end{align*}
\]

Sequential quadratic programming (SQP)

- Requires a twice differentiable problem with real-valued variables
- Iteratively approximates the problem with a quadratic optimization subproblem
  - Taylor’s series approximations of the functions
- Quadratic subproblems are solved with plethora of existing methods
Branch and bound

A general algorithm for integer or mixed-integer variables

Repeat until stops:
1. Optimize the continuous relaxation(s) of the problem
2. If the optimal solution with smallest value (lower bound) is integer, then stop
3. If the optimal solution of one relaxation is greater than the value of already found integer solution, then discard that subset
4. Branch the search space into subset(s) so that the integer solutions remain, but the optimal non-integer solution leaves

Optimal solution
Using traditional gradient based methods

\[ f(x) \]

\[ f(x) \]

\[ f(x) \]

\[ f(x) \]
Evolutionary algorithms

**Terminologies**

1. **Individual** - carrier of the genetic information (chromosome). It is characterized by its state in the search space, its fitness (objective function value).

2. **Population** - pool of individuals which allows the application of genetic operators.

3. **Fitness function** - The term “fitness function” is often used as a synonym for objective function.

4. **Generation** - (natural) time unit of the EA, an iteration step of an evolutionary algorithm.
Sample solution
Examples of optimization literature

- P.E. Gill et al., Practical Optimization, 1981
- M.S. Bazaraa et al., Nonlinear Programming: Theory and Algorithms, 1993
- D.P. Bertsekas, Nonlinear Programming, 1995
- J. Nocedal, Numerical Optimization, 1999
- A.R. Conn et al., Introduction to Derivative-Free Optimization, 2009
- M. Hinze et al., Optimization with PDE Constraints, 2009
MULTIOBJECTIVE OPTIMIZATION
Multiple objectives – what is that?

- Multiple objectives to be optimized *simultaneously*
  - Cost (investment vs. operating), quality, safety, profit, reliability, operability, environmental impact, etc.
  - Typically conflicting, i.e., have different optimal solutions

- *Need to find compromise between objectives*
  - How to find the best compromise?
  - In practice compromise can be better than optimal solution for a single objective (cf. optimize only costs/profit)

- Methods of *multiobjective optimization* (MOO) are required
  - [http://en.wikipedia.org/wiki/Multiobjective_optimization](http://en.wikipedia.org/wiki/Multiobjective_optimization)
Example 1

- Chemical separation process based on chromatography
- Applied e.g. in sugar, petrochemical, and pharmaceutical industries
- Typically a profit function is optimized
  - Formulation of a profit is not easy
- Multiobjective formulation
  - max throughput
  - min desorbent consumption
  - max purity of the product
  - max recovery of the product

Hakanen et al. *(Control & Cybernetics, 2007)*
Example 2

- Wastewater treatment plant design and operation
- Challenges: operational requirements, economical efficiency, operational reliability
- Multiobjective formulation
  - min total amount of nitrogen (quality)
  - min operational costs
    - min aeration energy
    - min methanol use
    - min excess sludge produced
    - max biogas production

Hakanen, Miettinen, Sahlstedt (Decision Support Systems, 2011)
Hakanen, Sahlstedt, Miettinen (Environmental Modelling & Software, 2013)
Multiobjective decision making process

- Need for optimization
- Optimization problem formulation
- Modelling of the problem (numerical simulation)
- Optimization & decision making
- Implementation & validation of the best solution found
Optimization problem formulation

- By optimizing only one criterion, the rest are not considered
- Objective vs. constraint
- Summation of the objectives
  - adding apples and oranges
- Converting the objectives (e.g. as costs)
  - not easy, includes uncertainties
- Multiobjective formulation reveals interdependences between the objectives
Mathematical formulation of MOO problem

- Vector valued objective function
  \[ f(x) = (f_1(x), \ldots, f_k(x))^T \]
  - \textit{k objective functions to be optimized}
- Decision variable vector \( x = (x_1, \ldots, x_n)^T \)
- Feasible region \( S \subset \mathbb{R}^n \)
- Objective vector \( z = f(x) \)
- Image of the feasible region \( Z = f(S) \)
Optimality for multiple objectives: Pareto optimality

- All the objectives don’t have the same optimal solution → optimality needs to be modified
- **Pareto optimality (PO)**
  - A solution is Pareto optimal if none of the objectives can be improved without impairing at least one of the others

  A decision vector \( \mathbf{x}^* \in S \) is Pareto optimal, if there does not exist another feasible decision vector \( \mathbf{x} \in S \) s.t. \( f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*) \) for all \( i = 1, \ldots, k \) and at least one of the inequalities is strict. An objective vector \( \mathbf{z}^* = f(\mathbf{x}^*) \) is PO, if the corresponding decision vector \( \mathbf{x}^* \in S \) is PO.
Illustration of PO: objective space

$Z = f(S)$

Best values are located down and left

$f_1 \text{ min}$

$f_2 \text{ min}$
Pareto optimal solutions

- Other terms used: *efficient/compromise/nondominated solution*

- *Pareto optimal set*: all PO solutions in the decision space

- *Pareto front*: all PO solutions in the objective space

- There can exist infinitely many PO solutions, all of them mathematically incomparable
  - *How to choose the best one?*
Some concepts

- **Decision maker (DM)**
  - Person expert in the application area (e.g. design engineer, plant operator)
  - Able to express preferences related to objectives, e.g., able to compare PO solutions
  - No need for expertize in optimization
  - Helps in finding the most preferred PO solution

- **Ranges for PO set (useful information in decision making)**
  - Ideal objective vector $z^*$: best values for each objective in PO set
  - Nadir objective vector $z^{nad}$: worst values for each objective in PO set

- **Reference point $\bar{z}$**: consists of aspiration levels (desired values for the objectives)
  - one way for the DM to express preferences
SOLVING MOO PROBLEMS
What means solving a multiobjective problem?

- Means different things for different people

1. Find all PO solutions
   - Theoretical approach, not feasible in practice

2. Approximate PO set
   - Good diversity (representatives in all parts of PO set)
   - Can also be an approximation of some part of PO set

3. Find the most preferred PO solution
   - Requires preferences from DM
Scalarizing MOO problem

- Often MOO problem is in some way converted into a single objective one
  - Single objective optimization methods can be used
  - Produce usually one PO solution at a time
  - Include some parameters

- This is called *scalarization*
  - Can be done in a good way or in a bad way

- Good method should
  - Find only and all PO solutions
  - Have parameters meaningful for DM
Solution approaches

- Plenty of methods developed for MOO
- We concentrate on methods that aim at finding the most preferred PO solution
- MOO methods can be categorized based on the role of DM
  - No-preference methods (no DM)
  - A priori methods
  - A posteriori methods
  - Interactive methods
Examples of MOO literature

Examples of MOO literature

Material for the discussion on March 16th


– Read the preface