Multi-objective optimization

Kevin Duh

Bayes Reading Group

Aug 5, 2011

The Problem

Optimization of K objectives simultaneously:

$$\min_{x}[F_1(x), F_2(x), \ldots, F_K(x)], \ s.t. \ x \in X \tag{1}$$
$$X = \{x \in \mathcal{R}^n | \ g_j(x) \leq 0, j = 1, 2, \ldots, m\}$$

What does it mean to be the optimum? How to compute it?

Examples of multiple objectives

- Engineering design: e.g. facility location, aircraft wing design
- Finance/portfolio: expected return vs. standard deviation
- Machine learning: accuracy vs. speed vs. complexity
- Problems without a single perfect metric: BLEU, TER, RIBES as different approximations to human judgment in machine translation evaluation
- Many decisions in life also involve multiple (possibly conflicting) objectives

Outline

Preference and Utility

Preference: decision makers's opinions on relative importance of each objective F_k

Utility: measures decision maker's satisfaction

Approximating preference with utility functions:

$$\min_x U[F_1(x), F_2(x), \ldots, F_K(x)] , \ s.t. \ x \in X \tag{2}$$

where $U : \mathcal{R}^K \to \mathcal{R}$ is a scalarization.

What sorts of utility functions are appropriate? When does solving Eq. 2 also solve Eq. 1?

Pareto Optimality/Efficiency (1/2)

Definition (Pareto Optimal)

A point $x^* \in X$ is Pareto optimal iff there does not exist another point $x \in X$ such that $F_k(x) \leq F_k(x^*) \forall k$ and $F_k(x) < F_k(x^*)$ for at least one k.

Definition (Weakly Pareto Optimal)

A point $x^* \in X$ is weakly Pareto optimal iff there does not exist another point $x \in X$ such that $F_k(x) < F_k(x^*) \ \forall \ k$

Definition (Properly Pareto Optimal)

A point $x^* \in X$ is properly Pareto optimal iff it's Pareto optimal and there is some $M > 0$ such that for each F_k satsifying $F_k(x) < F_k(x^*)$ there exists at least one F_j such that $F_j(x^*) < F_j(x)$ and $\frac{F_k(x^*) - F_k(x)}{F_j(x) - F_j(x^*)} \leq M$.

Pareto Optimality/Efficiency (2/2)

Some comments:

- Properly Pareto Optimal $\in$ Pareto Optimal $\in$ Weakly Pareto Optimal
- Properly Pareto Optimal means the tradeoff (between F_k and F_j) is bounded; we cannot arbitrarily improve on one objective.
- A Pareto Optimal point has no other point that improves at lease one objective without detriment to another, i.e. it's "not dominated"
- The set of all Pareto Optimal points is known as Pareto Frontier.
- This kind of "something is optimal if there does not exist.." definition is needed since we only have partial ordering in the objective space. More generalized notion of domination have also been proposed.
- The preference of Pareto simply says: for any objective, less is better.

Illustration
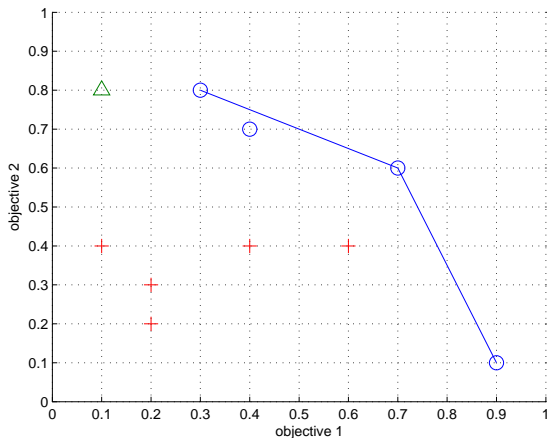
Note: here the plot shows a maximization problem (aiming for higher objectives). Circles are pareto optimal, triangles are weakly pareto optimal.

Another Illustration

Hotel	Stars	Distance from Beach	Price
A	**	0.7	1,175
B	*	1.2	1,237
C	*	0.2	750
D	***	0.2	2,250
E	***	0.5	2,550
F	**	0.5	980

We prefer hotels that are fancier, closer to the beach, and cheaper. Hotels A, B, E can be eliminated (not pareto optimal). Note that Hotel F is not the best in any objective, yet it is pareto optimal.[1]

---

[1]Example from: Godfrey et. al. "Algorithms and analyses for maximal vector computation", VLDB Journal (2007) 16.

Outline

1. Basic Concepts: Preference and Pareto optimality

2. **Methods with a priori articulation of preferences**

3. Methods with a posteriori articulation of preferences

4. Discussion

A Priori Articulation vs. A Posteriori Articulation of Preference

	A Priori	A Posteriori
When is preference expressed	Before optimization	After optimization
Output result	one solution	a set of solutions
Advantages	Use single-objective tools	User chooses solution, not formulation
Main Issues	Eliciting preference Necessary/sufficient cond.	Computation Even sampling
General approach	Scalarization	Genetic algorithm, Iterative methods

A simple method for a priori preference articulation

Weighted Sum Method ($U[\cdot]$ as linear combination):

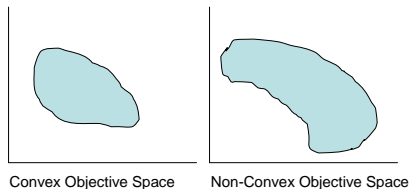$$\min_x \sum_{k=1}^{K} w_k F_k(x) \ , \ s.t. \ x \in X \tag{3}$$

where $\sum_{k=1}^{K} w_k = 1$ and $w_k > 0 \forall k$.
Sufficient Condition:

- If x^* is a optimal solution to Eq. 3, then it's Weakly Pareto Optimal.
- Further if x^* is unique, then it's Pareto Optimal.
- Assuming convex F_k, x^* is also Properly Pareto Optimal.

But doesn't provide Necessary Condition if objective space is non-convex!

Convex vs. Non-convex objective space

Convex Objective Space   Non-Convex Objective Space

An improvement: Weighted Exponential Sum method

Weighted Exponential Sum:

$$\min_x \sum_{k=1}^{K} w_k F_k(x)^p \ , \ s.t. \ x \in X \tag{4}$$

where $\sum_{k=1}^{K} w_k = 1$ and $w_k > 0 \forall k$ and $p > 0$.

- The designer selects both $w_k \forall k$ and p
- p can be thought of as a compensation parameter: high p means one prefers solutions with both very high and very low objective values rather those with averaged values.
- Necessary condition is satisfied. But p may need to be very large to capture Pareto points in non-convex regions.

Lexicographic Method

Arrange the objective functions by importance, then solve repeatedly:

$$\min_{x \in X} F_i(x)$$
$$s.t. \quad F_j(x) \leq F_j(x_j^*) \, , \; j = 1, 2, \ldots, i - 1 \tag{5}$$

where x_j^* is the optimum found in the j-th iteration.

- The most important objective is solved first
- In practice, one can introduce slacks to the constraints (e.g. $F_j(x) \leq F_j(x_j^*) + \delta_j$).
- Varying δ_j generates different Weakly Pareto optimal points.

Tchebycheff Method

Idea: Minimize "distance" to an ideal "utopia" point y:

$$\min_{x \in X} \{ \sum_{k=1}^{K} w_k^p [F_k(x) - y_k]^p \}^{\frac{1}{p}} , \ s.t. \ x \in X \tag{6}$$

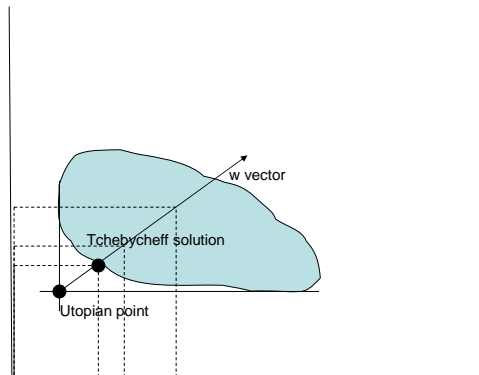Note Eq. 6's similarity to Weighted Exponential Sum. Let $p \rightarrow \infty$:

$$\min_{x \in X} \max_k w_k [F_k(x) - y_k] , \ s.t. \ x \in X \tag{7}$$

which can be solved by:

$$\min_{x \in X, \lambda} \lambda, \ s.t. \ w_k [F_k(x) - y_k] - \lambda \le 0 \forall k \tag{8}$$

This method provides necessary condition for Pareto optimality and sufficient condition for Weakly Pareto Optimality.

Illustration of Tchebycheff

Summary of A Priori Methods[2]

Methods differ by scalarization, e.g.:

- Weighted Sum: $\sum_{k=1}^{K} w_k F_k(x)$
- Weighted Exponential Sum: $\sum_{k=1}^{K} w_k F_k(x)^p$
- Tchebycheff: $\max_k w_k [F_k(x) - y_k]$
- Lexicographic: $F_i(x)$ by importance order

which leads to different necessary/sufficient conditions and properties.

- Best method depends on, e.g. how much preference information one has and assumptions of Pareto set.

---

[2]Note: I'm skipping a large portion of the literature (e.g. physical programming, interactive multi-objective programming) which deals with how to elicit preferences and utility from human users (i.e. setting the weights w_k).

Outline

1. Basic Concepts: Preference and Pareto optimality

2. Methods with a priori articulation of preferences

3. Methods with a posteriori articulation of preferences

4. Discussion

Motivation for A Posteriori Approaches

It is sometimes difficult to express preferences of objective functions, but it is possible to choose from a set of Pareto solutions.

Hotel	Stars	Distance from Beach	Price
C	*	0.2	750
D	***	0.2	2,250
F	**	0.5	980

These approaches are called: "cafeteria" or "generate-first-choose-later".

Challenge of running a cafeteria

Basic approach (vary weights and re-run):

1. for $t = 1, \ldots, T$
2.   randomly determine weights w_k
3. $x^* = \arg\min_{x \in X} \sum_{k=1}^{K} w_k F_k(x)$
4.   Add x^* to ParetoFrontierSet
5. end

Challenges:

- Can we sample the entire Pareto Front as $T \rightarrow \infty$?
- Can we reduce wasteful work (different $\{w_k\}$ generating the same solution)?
- Can we generate diverse Pareto points with few samples?

Normal Boundary Intersection (NBI) Method

1. Compute minima of individual objectives. Form convex hull Φ where columns are these minimum.

2. Sample on the convex hull with randomized $\mathbf{w}$.

3. Project these points onto Pareto Front by solving:

$$\max_{\lambda \in \mathcal{R}^+} \lambda \ , \ s.t. \ \Phi\mathbf{w} + \lambda\mathbf{n} = F(x), x \in X \tag{9}$$

where $\mathbf{n}$ is the normal vector (pointing towards origin).
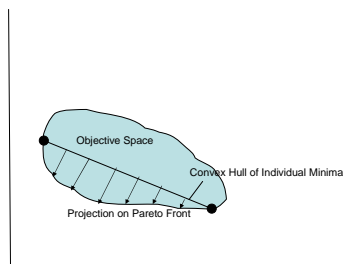
Genetic Algorithm solution

Many methods have been proposed. One general idea:

1. Initial Population at iteration t
2. Individual Mutation, etc.
3. Compute multiple objectives for each individual
4. Selection of individuals based on Pareto/Non-dominance "ranks" i.e. Pareto points have highest fitness; points which become nondominated after deleting Pareto points get second-highest fitness, etc.
5. Return to Step 1

After T iterations, the population approximates the Pareto Frontier.

Checking for Pareto optimality

NBI and GA do not guarantee all solutions are Pareto. A filter is needed:

- Brute-force: compare all pairs and check dominance $O(N^2)$, $N =$ number of points in candidate set
- Borrow ideas from Maximum Vector Computation: Divide-and-conquer gives $O(N \log N)$
- Solve for each point x' and check if δ_k's are 0:
  $\min_{x \in X, \delta_k \geq 0} \sum_{k=1}^{K} \delta_k$, $s.t.$ $F_k(x) + \delta_k = F_k(x')$ $\forall k$

Outline

1. Basic Concepts: Preference and Pareto optimality

2. Methods with a priori articulation of preferences

3. Methods with a posteriori articulation of preferences

4. **Discussion**

Discussion

Is it possible to view your research problem as multi-objective? e.g.,

- Connections to regularization path [Hastie, JMLR04]
- Creating multiple objectives to jump out of local optima in single objective problem
- Multi-objective reinforcement learning [Barrett, ICML2008]
- From scalar optimization to vector optimization

Good starting-point references

- Engineering optimization perspective: Marler & Arora, "Survey of multi-objective optimization methods for engineering", Structural and Multidisciplinary Optimization, 2004, vol 26.

- Examples in machine learning: Jin & Sendhoff, "Pareto-based multiobjective machine learning: an overview and case studies", IEEE Transactions on Systems, Man, and Cybernetics, 2008, vol. 38(3).

- Max vector computation (Pareto filter): Godfrey, Shipley, & Gryz, "Algorithms and analyses for maximal vector computation", VLDB Journal, 2007 vol 16(5).

- Simple geometrical understanding: Stueur, "An overview in graphs on multiple objective programming" http://www.terry.uga.edu/~rsteuer/PDF_Links/Overview.pdf