Multi-Objective Optimization in pSeven

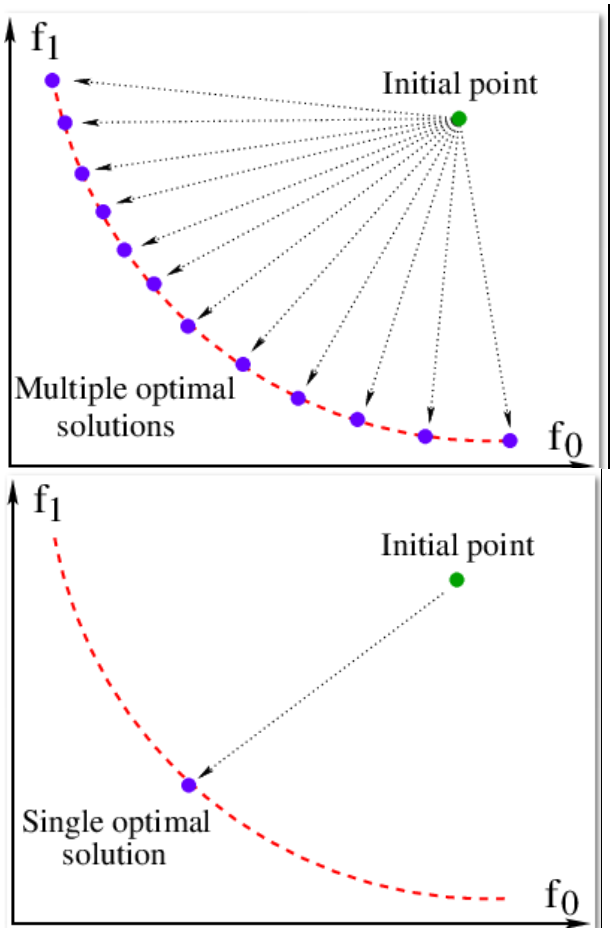
I. Overview

There are two different approaches for multi-objective optimization (MO):

- *Global MO* – finding the variety of non-dominated points (Pareto frontier in objective functions space and respective Pareto set in design space)

- *Local MO* – finding a single non-dominated solution (can be closest to the initial guess).

Pictures below illustrate the key difference between these two approaches:



Methods implemented in pSeven combine local and global interpretations of multi-objective optimization to produce efficient algorithms of Pareto frontier discovery. The need for such combination is explained by one of the main problems in solving multi-objective cases. Typically, a Pareto set is not localized, and thus global methods should be applied to approximate the whole Pareto frontier. This causes the evaluations far away from Pareto variety, which is computationally expensive.

The basic principle of multi-objective optimization in pSeven is that in the majority of cases Pareto optimal set could be described as a union of compact (in objective space)continuous components, all of them allowing differential-geometric description in case of continuous objective functions. Therefore, it is enough to find any optimal point for a single connected component and then perform incremental local spreading from the known optimal solutions in the tangent plane to Pareto set. New points generated this way are not precisely optimal, because Pareto set is not a hyper plane. However, the corresponding distance to optimality is small, so resources required for a subsequent optimization are minimal. As for the multiple connected components, their sequential discovery is already built into the described strategy, provided that a single point optimization always finds the nearest Pareto optimal solution. Due to the negativity of Pareto frontier slopes in space of objectives, the algorithm eventually discovers all the connected components if it had been started from the points minimizing each particular objective (anchor points).

Therefore, there are two primary components of pSeven multi-objective optimization:

- Local algorithm to reach Pareto optimal set, finding the nearest optimal solution.

- Pareto frontier discovery: spreading along optimal manifold (with a new optimization in case of curved Pareto set) with optimal points distribution kept even.

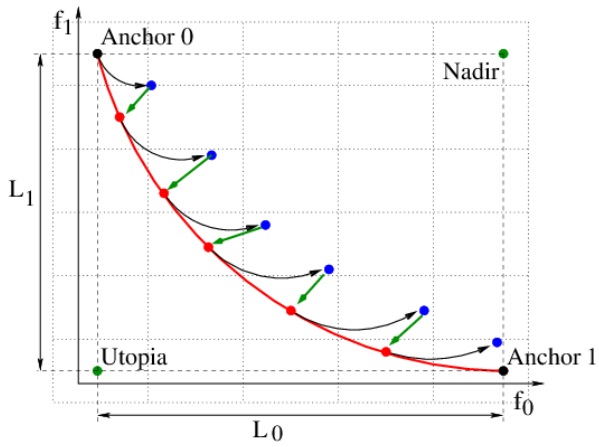
First, multi-objective optimization algorithms of pSeven searches for a small number of strictly Pareto optimal solutions. In theory, any optimal set could be used, but for the reasons explained above anchor points are the most convenient points to start optimization.

By definition, there are exactly K anchors for a problem with K objectives.

k-th anchor is defined as minimizer of *k*-th objective function with no regard to other objectives (all relevant constraints remain imposed). As a byproduct, the set of anchor points also provides raw estimate of Pareto frontier extent along objective axes. Along with user-provided number of optimal points to be finally generated, this gives the minimal objective space distance to be used at spreading stage (see below).

Second, algorithm performs sequential spreading along the optimal manifold from all currently discovered optimal solutions. To this end, Pareto set local geometry is reconstructed, prime purpose being to determine tangent plane at current optimal solution. Then, the algorithm generates a number of new (not optimal in general) points by shifting along basis axes of tangent plane from initial optimal position, until the corresponding objective space separation becomes of required order (see above). If there are already discovered optimal solutions close to the proposed new position, the attempt is discarded. Otherwise, new point is accepted and (if necessary) is pushed back to optimality. Process iterates until no missed optimal solutions remain.

Process illustration



Stages in detail

1. Anchor search stage

Proper determination of anchor points constitutes an essential part of the method. In terms of implementation, it is almost equivalent to a single-objective optimization and hence could be performed with standard pSeven algorithms. Anchors are used to estimate the Pareto frontier's global geometry and as starting points for scattering process.

2. *Re-optimization of scattered candidates*

This stage is performed using either steepest or QN descent algorithms. Due to the sufficiently small deviation from optimal variety, pushing candidate solutions back to optimality is not an expensive procedure, and steepest/QN descent fits well.

3. *Scattering process*

Let us consider the procedure of finding the optimal descent in the unconstrained case *M*=0 to explain what is a local Pareto set geometry and how it is reconstructed.

Constraint $|d|_\infty \leq 1$ could be traded for the term $1/2|d|^2$ in objective function. Then dualized optimal descent problem reads:

$$\min_x |x - x_0|$$
$$c_L^j \le c^j(x) \le c_U^j \quad M \text{generic constraints}$$
$$x_L^k \le x^k \le x_U^k \quad N \text{box bounds}$$

with positive semi-definite Hessian $G_{ij} = (\nabla f^i \cdot \nabla f^j)$. For locally optimal solution $G\lambda^* = 0, \Longrightarrow$ rank$G \le K - 1$, and in general, there is rank$G = K - 1$, confirming that frontier's dimensionality is $K - 1$. Therefore, one of eigenvectors of *G* determines optimal descent direction, which disappears at locally optimal position (corresponding eigenvalue becomes zero as well). What are the remaining eigenvectors $\lambda^{(\gamma)}$ of G?

One can show that $t^{(\gamma)} = \mu_{(\gamma)}^{-1/2} \sum_i \lambda_i^{(\gamma)} \nabla f^i \gamma = 1, \ldots, K-1$ provides an orthonormal basis in Pareto set tangent plane (Pareto frontier tangents are $\lambda^{(\gamma)}$). Here, $\mu_{(\gamma)}$ are the corresponding eigenvalues, $G\lambda^{(\gamma)} = \mu_{(\gamma)}\lambda^{(\gamma)}$.

The above construction generalizes trivially to the case of active constraints. Namely, let $\mathcal{P}_{\mathcal{A}}$ be an orthogonal projector onto the tangent plane to all active constraints (including active box bounds):

$$\mathcal{P}_{\mathcal{A}}^2 = \mathcal{P}_{\mathcal{A}}\mathcal{P}_{\mathcal{A}}\nabla c^i = 0 i \in \mathcal{A}$$

Then the analysis of Pareto frontier's local geometry goes through with the only change $\nabla f^i \to \mathcal{P}_{\mathcal{A}} \nabla f^i$. To proceed, we assume that for the infinitesimal shift in Pareto set tangent plane $x = x^* + \epsilon t^{(\gamma)}$ sub-optimality of x is of order $O(\epsilon)$. Then it only remains to push *x* back to optimal position, which is rather cheap (still close of optimal set!). And it is necessary then to decide how to choose ϵ in order to make the described approach reasonable from a practical standpoint. *Size selection* step in pSeven is implemented with the following strategy:

- End-user is required to provide single number N_f specifying how many points he or she needs to have on the Pareto frontier as a result.
- Method estimates the extent of Pareto frontier L_i along each axis in objective space (anchor search stage)
- Objective space is subdivided into non-overlapping boxes of size $bb^i = \frac{L_i}{N_f i} = 1, \ldots, K$
- Ultimate goal is to have no more than one solution in each box

Hence, the appropriate strategy to select value of ϵ parameter is as follows: objective space points $f(x^*)$ and $f(x^* + \epsilon t^{(\gamma)})$ should belong to the neighboring boxes.