Nonlinear Multi-Objective Constrained Optimization: Using Second Order Approximation of Pareto Frontier Local Geometry in Descent-Diffusion Approach

Alexis Pospelov, Fedor Gubarev

DATADVANCE
AN AIRBUS GROUP COMPANY

Outline

Outline

Problem Definition

We consider multi-objective optimization problem

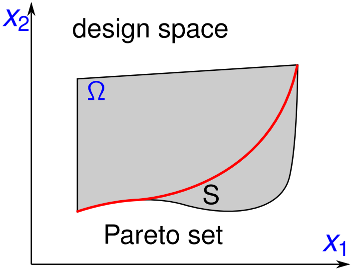
$$\min_x f^i(x) \qquad K > 1 \quad \text{objective functions}$$

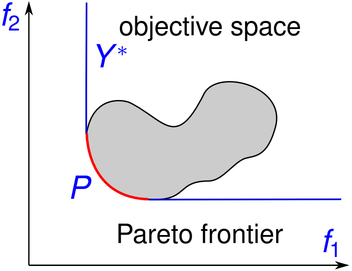
$$c_L^j \leq c^j(x) \leq c_U^j \qquad M \quad \text{generic constraints}$$

$$x_L^k \leq x^k \leq x_U^k \qquad N \quad \text{box bounds}$$

Pareto frontier

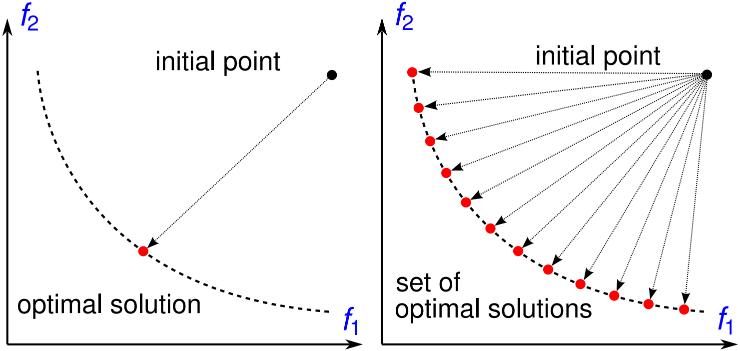
$\Omega = \{x \mid c_L \leq c(x) \leq c_U\}$	Admissible set
$Y = f(\Omega)$	Feasible set
$S = \{x \mid \{\widetilde{x} \in \Omega \mid f(\widetilde{x}) \leq f(x),\ f(\widetilde{x}) \neq f(x)\} = \emptyset\}$	Pareto set
$P = f(S)$	Pareto frontier
$Y^* = P + \mathbb{R}^K_+ = Y + \mathbb{R}^K_+$	Edgeworth-Pareto hull





Two stage of approach

- **Local**: Find a single non-dominated solution (nearest to the initial guess in some sense).
- **Global**: Find a whole variety of non-dominated points



Outline

Finding one Pareto point

Purposes:

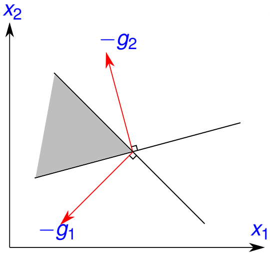
- Universal estimation of optimality of current iterate x_k

  (suitable for single- and multi-objective problems,
  with or without constraints)

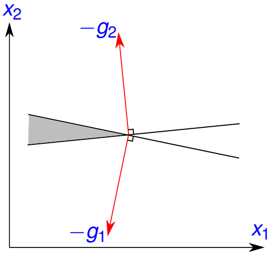
- Obtain the direction of optimal descent

  (direct analog of steepest descent in $K = 1$, $M = 0$ case)

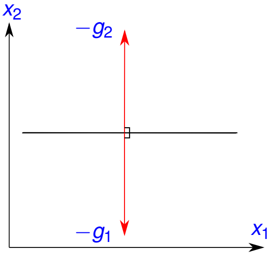
Gradients in multiobjective case



(a) Far from solution



(b) Close to solution



(c) Solution

It follows from Karush - Kuhn - Tucker conditions that zero vector in optimal point can be represented as linear combination of gradients of objective components with positive coefficients.

Multiobjective descent

Mathematically, we would like to find (or ensure the absence of) a direction $d \neq 0$ such that:

- d is a descent direction for all objectives:

$$d \cdot \nabla f^i \leq 0 \qquad \forall i$$

- d violates none of imposed bounds in linear approximation

$$c_L^j \leq c^j + d \cdot \nabla c^j \leq c_U^j \qquad \forall j$$
$$x_L^k \leq x^k + d \leq x_U^k \qquad \forall k$$

Solution of the problem give us multiobjective steepest descent.

Based on original work J. Fliege, B. F. Svaiter *Steepest Descent Methods for Multicriteria Optimization* Mathematical Methods of Operations Research, 2000

Multiobjective (quasi-)Newton descent

Problems with optimal descent:

- Slow convergence if used in iterative line-search based algorithms
- Badly scaled search direction (no prediction on optimal step size)

Remedy is to include Hessians information.

Basic equations for $M = 0$:

$$\min_d \max_i \left[d \cdot \nabla f^i + 1/2\, dH^i d\right] \quad \Leftrightarrow \quad \begin{array}{c} \min_{d,t} \ t \\ d \cdot \nabla f^i + 1/2\, dH^i d \ \leq \ t \end{array}$$

- Problem type is QCQP
- Hard to solve, but internal and hence cheap by definition

... with constraints

True formulation in case of constrained problems:

$$\min_{d,t} \; t$$

$$d \cdot \nabla f^i + 1/2\, dH^i d \leq t$$

$$\pm [d \cdot \nabla c^j + 1/2\, dH_c^j d] \leq t \quad j \in \mathcal{A}$$

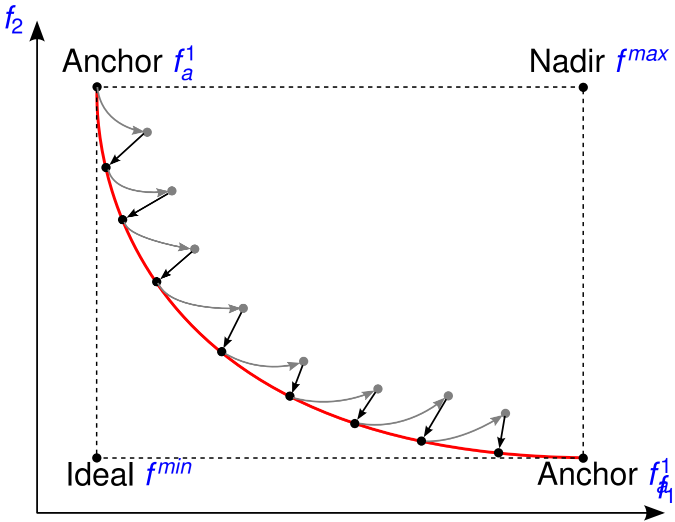
$$k \in \mathcal{A}_b \; : \; d_k \in \begin{cases} \geq 0 & x_k \text{ is lower-active} \\ \leq 0 & x_k \text{ is upper-active} \end{cases}$$

Summary on multiobjective (quasi-)Newton descent

- Allows finding Pareto optimal points
- Has good speed of convergence
- Is suitable for non-convex fronts
- Is locally find nearest Pareto point
- Inherits the smoothness of underline problem

Outline

Finding Pareto frontier



Local geometry of Pareto set

For simplicity let's consider optimal descent in $M = 0$ case.

$$d \cdot \nabla f^i \leq 0 \qquad \forall i$$

In Pareto optimal point

$$\text{rank}(\nabla f^i) \leq K - 1$$

and generically $\text{rank}(\nabla f^i) = K - 1$ ("front dimensionality is $K - 1$")
Moreover, there are $\lambda_i \geq 0$, $\sum \lambda = 1$ such

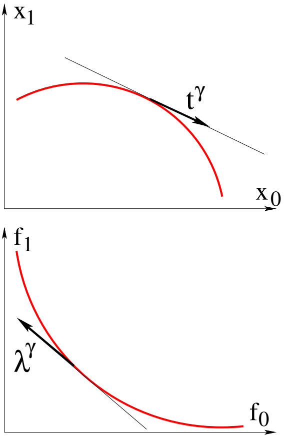
$$\sum_i \lambda_i \nabla f^i = 0.$$

And the

$$\text{Lin}(\nabla f^i)$$

is tangent hyperspace in the design space.

Tangent direction



From the above, it follows that there are the set of vectors

$$t^\gamma, \ \gamma = 1, ..., K - 1,$$

that forms orthonormal basis in Pareto set tangent plane

(Pareto front tangents $\lambda^{(\gamma)}$ could be also identified)

Constrained case

Let $\mathcal{P}_{\mathcal{A}}$ be an orthogonal projector onto the space tangent to active constraints (including box constraints):

$$\mathcal{P}_{\mathcal{A}}^2 = \mathcal{P}_{\mathcal{A}}, \;\; \mathcal{P}_{\mathcal{A}} \nabla c_i = 0, \;\; \forall i \in \mathcal{A}$$

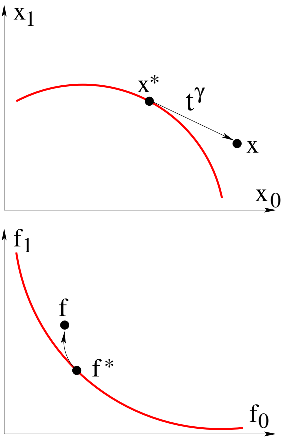
E.g.

$$\mathcal{P}_{\mathcal{A}} = I - J^T (JJ^T) J, \;\; J = (\nabla c_i)$$

Then analysis of Pareto front local geometry goes through with the only change

$$\nabla f \to \mathcal{P}_{\mathcal{A}} \nabla f$$

Diffusion along Pareto Frontier



For infinitesimal shift in Pareto set tangent plane

$$x = x^* + \varepsilon\, t^\gamma$$

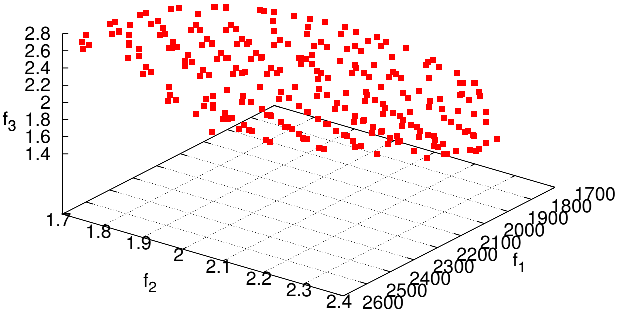
sub-optimality of x is of order $O(\varepsilon)$.

It remains to push x back to optimality which is rather cheap (we're still in the small vicinity of optimal set!)

First example

Ten-dimensional ($N = 10$) three-objective problem

$$\mathrm{FDS} = \begin{cases} f_1 = \frac{1}{N^2} \sum_i i\,(x_i - i)^4 \\ f_2 = \exp\{\sum_i x_i/N\} + |x|^2 \\ f_3 = \frac{1}{N(N+1)} \sum_i i\,(N - i + 1)\, e^{-x_i} \end{cases} \qquad \text{subject to} \quad |x|^2 = 1$$



Problem from J. Fliege, L. M. Grana Drummond, B. F. Svaiter *Newton's Method for Multiobjective Optimization* SIAM J. Optim, 2007

Outline

Second example

Four-dimensional ($N = 4$) two-objective problem

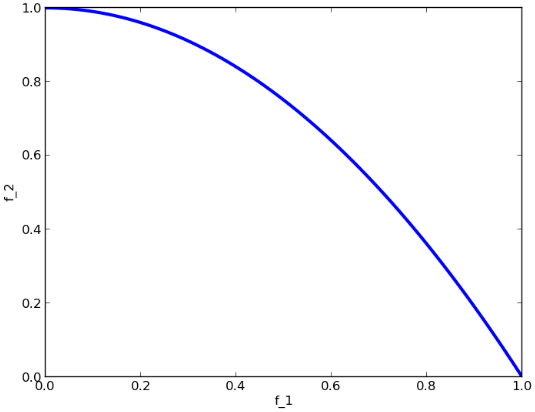
$$\min \left[x_1 + \frac{2}{|J_1|} \sum_{j \in J_1} h_j(x),\ 1 - x_1^2 + \frac{2}{|J_2|} \sum_{j \in J_2} h_j(x) \right]$$

where

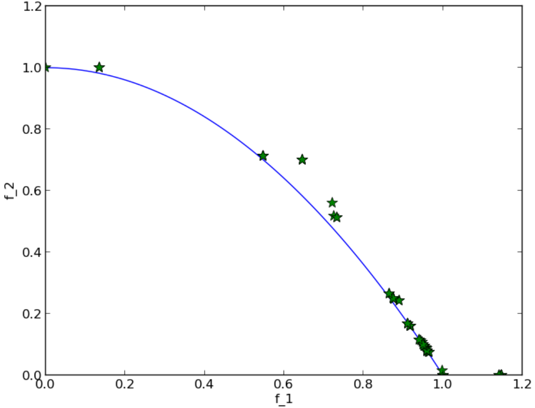
$$h_j(x) = x_j - \sin\left(6\pi x_1 + \frac{j\pi}{N}\right)$$
$$J_1 = \{j \,|\, \text{is odd and } 2 \le j \le N\}$$
$$J_2 = \{j \,|\, \text{is even and } 2 \le j \le N\}$$

Problem is based on Q. Zhang, A. Zhou, S. Zhao†, P. N. Suganthan, W. Liu, S. Tiwari *Multiobjective optimization Test Instances for the CEC 2009 Special Session and Competition*

Analytical solution



Diffusion solution



Singularity of Hessians make first order approximation inefficient.

Idea of second order correction

As before in optimal point we have

$$\sum_k \lambda_k \nabla f^k = 0, \sum \lambda_k = 1, \lambda \geq 0$$

That grants optimality for small step v_ε with the first order.

To move but to keep optimality with the second order:

$$\sum_k \tilde{\lambda}_k \left(H^k v_\varepsilon + \nabla f^k \right) = 0, \sum \tilde{\lambda}_k = 1, \tilde{\lambda} \geq 0$$

Second order correction

Assuming infinitesimal step v_ε

$$\tilde{\lambda}_k = \lambda_k + \mu_k,\ \mu_k = O(v_\varepsilon)$$

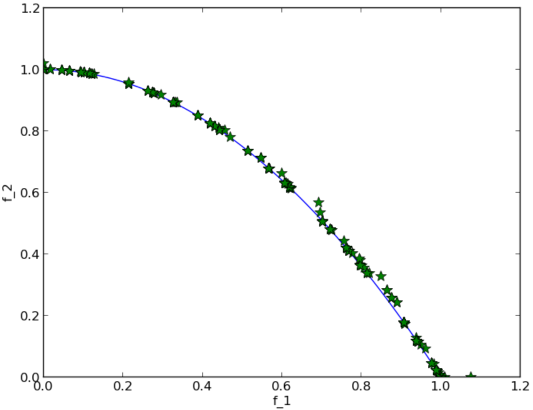
That leads idea to the structure of the second order correction

$$v_\varepsilon = \left(\sum_k \lambda_k H_i\right)^{-1} \sum_k \mu_k \nabla f^k$$

And finally the correction that we use is

$$t_\gamma^c = \left(\sum_k \lambda_k H_i\right)^{-1} t_\gamma$$

Applying second order correction



Conclusion

Presented approach

- Allows discovering Pareto front;
- Finds exact points on Pareto front uniformally;
- Avoids multiple evaluations far from Pareto front;
- Can be used in the constrained case;
- Was successfully implemented in module GT Opt in pSeven;
- Was test and found efficient for large variety of MO problems;
- With additional second order correction it works even for functions with singular Hessians behavior.

Thanks for your attention!

Alexis Pospelov (alexis.pospelov@datadvance.net)

DATADVANCE (datadvance.net)

pSeven (http://datadvance.net/product/pSeven)