Author: Gaida, Daniel
Title: Dynamic real-time substrate feed optimization of anaerobic co-digestion plants
Issue Date: 2014-10-22

Chapter 3

Multi-Objective Optimization Algorithms

Multi-objective optimization algorithms try to find the Pareto front and the corresponding Pareto optimal set of a multi-objective optimization problem. As in case of continuous function optimization a Pareto front could contain infinite many elements these algorithms in general cannot find the complete Pareto front. Therefore they try to find solutions which approximate the form of the Pareto front best possible.

In this thesis a multi-objective optimization algorithm is used to solve the MONMPC problem stated in Chapter 2. In the simulation and optimization studies in Chapter 9 two multi-objective optimization methods are compared. They are SMS-EMOA (Emmerich et al., 2005, Beume et al., 2007) and SMS-EGO (Wagner et al., 2007, Ponweiser et al., 2008). Both methods are briefly introduced in the following two sections.

Next to the two methods there are also other famous multi-objective optimization methods. Examples are NSGA-II (Deb et al., 2002), SPEA2 (Zitzler et al., 2001), ϵ-MOEA (Deb et al., 2003) and ParEGO (Knowles, 2006). Various publications comparing these different multi-objective optimization methods reveal that both SMS-EMOA and SMS-EGO belong to the best methods of their kind, e.g. (Ponweiser et al., 2008, Wagner et al., 2007, 2010).

3.1 Hypervolume-based Evolutionary Algorithm

An evolutionary algorithm is a stochastic optimization technique inspired by the principles of natural evolution (cf. Alba and Cotta (2006)). It describes a class of different optimization methods which all have the following in common. The key feature of evolutionary algorithms (EAs) is that in each iteration of the algorithm a collection of potential solutions, the so-called population, of the optimization problem is evaluated in parallel. The elements of the population are called individuals. The performance of each individual is measured by the objective function. In each iteration the population may change, what means that new solutions are created and already existing solutions are discarded from the population to keep the number of individuals inside the population constant. As EAs typically return a set of solutions (the population) in one call, they are especially suited to solve multi-objective optimization problems, compared to methods, which only return one solution at a time.

The purpose of a hypervolume-based EA is to maximize a scalar criterion, which is named the hypervolume indicator (or $\mathcal{S}$-metric, Zitzler and Thiele (1998)), see Definition 3.1. This criterion is a property of a set and describes the size of a space covered by this set. Below it is shown that the hypervolume indicator of the Pareto front $\mathcal{PF}^*$ is maximal for a given optimization problem. Therefore, by maximizing the hypervolume indicator the algorithm tries to find the best approximation (with a finite number of elements) of the true Pareto front $\mathcal{PF}^*$. Note that the multi objectives are mapped onto one objective, so that in general each single objective optimization method can be used to solve a multi-objective optimization problem using the hypervolume indicator (Fleischer, 2003, Knowles et al., 2003).

Definition 3.1 (Hypervolume indicator, Custódio et al. (2012))**:** The hypervolume indicator for some set $\mathcal{A} \subset \mathbb{R}^{n_\mathrm{o}}$ and a reference point $\boldsymbol{r} \in \mathbb{R}^{n_\mathrm{o}}$ that is dominated by all the points in $\mathcal{A}$ is defined as:

$$I_\mathrm{H}(\mathcal{A}) := Vol \left\{ \boldsymbol{b}_\mathcal{B} \in \mathbb{R}^{n_\mathrm{o}} \,|\boldsymbol{b}_\mathcal{B} \leq \boldsymbol{r} \wedge \exists \boldsymbol{a}_\mathcal{A} \in \mathcal{A} : \boldsymbol{a}_\mathcal{A} \leq \boldsymbol{b}_\mathcal{B} \right\} = Vol \left(\bigcup_{\boldsymbol{a}_\mathcal{A} \in \mathcal{A}} [\boldsymbol{a}_\mathcal{A}, \boldsymbol{r}] \right)$$

Here Vol denotes the Lebesgue measure of a n_o-dimensional set of points, and $[\boldsymbol{a}_\mathcal{A}, \boldsymbol{r}]$ denotes the interval box with lower corner $\boldsymbol{a}_\mathcal{A}$ and upper corner $\boldsymbol{r}$.

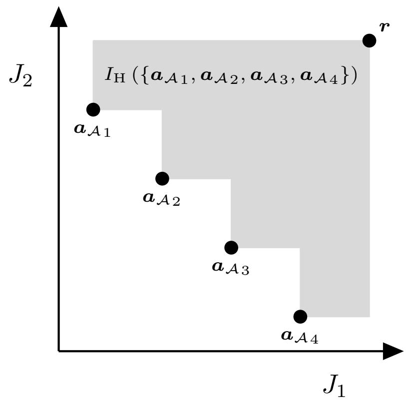
Figure 3.1a shows an example of the hypervolume indicator for a set $\mathcal{A}$ in a two-dimensional and Figure 3.1b in a three-dimensional space. To be able to find the approximation of the Pareto front we first have to be able to compare two different approximate Pareto fronts and to decide which one is better.

Definition 3.2 (Custódio et al. (2012))**:** Given two nondominated sets $\mathcal{A}$ and $\mathcal{B}$. $\mathcal{A}$ is better than $\mathcal{B}$, which is represented by $\mathcal{A} \prec \mathcal{B}$, if and only if

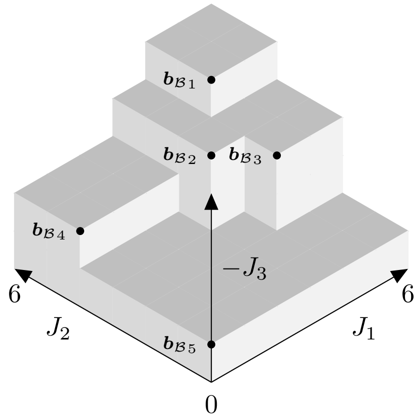
$$\forall \boldsymbol{b}_\mathcal{B} \in \mathcal{B} : \exists \boldsymbol{a}_\mathcal{A} \in \mathcal{A} : \boldsymbol{a}_\mathcal{A} \leq \boldsymbol{b}_\mathcal{B} \text{ and } \exists \boldsymbol{b}_\mathcal{B} \in \mathcal{B} : \exists \boldsymbol{a}_\mathcal{A} \in \mathcal{A} : \boldsymbol{a}_\mathcal{A} \prec \boldsymbol{b}_\mathcal{B}$$

Now the hypervolume indicator is used to compare two Pareto front approximations. In Zitzler et al. (2003) it was shown, that if a certain property holds, the better nondominated set has a higher hypervolume indicator, see Lemma 3.1.

Lemma 3.1 (Custódio et al. (2012), Zitzler et al. (2003))**:** Let $\mathcal{A}$ and $\mathcal{B}$ be two nondominated sets with the properties $\mathcal{A} \prec \mathcal{B}$ and $\forall \boldsymbol{\varphi} \in \mathcal{A} \cup \mathcal{B} : \boldsymbol{\varphi} \prec \boldsymbol{r}$, where $\boldsymbol{r}$



(a) The hypervolume indicator I_{H} for the set $\mathcal{A}$ is shaded in grey, cf. Custódio et al. (2012).



(b) The hypervolume indicator I_{H} for the set $\mathcal{B}$ with $\boldsymbol{r} = (6, 6, 0)^T$, cf. Custódio et al. (2012).

Figure 3.1: Hypervolume indicator for a set $\mathcal{A} := \{\boldsymbol{a}_{\mathcal{A}1}, \boldsymbol{a}_{\mathcal{A}2}, \boldsymbol{a}_{\mathcal{A}3}, \boldsymbol{a}_{\mathcal{A}4}\} \subset \mathbb{R}^2$ in a two-dimensional space and a set $\mathcal{B} := \{\boldsymbol{b}_{\mathcal{B}1}, \ldots, \boldsymbol{b}_{\mathcal{B}5}\} \subset \mathbb{R}^3$ in a three-dimensional space.

is the reference point used in the hypervolume computations. Then $I_{\mathrm{H}}(\mathcal{A}) > I_{\mathrm{H}}(\mathcal{B})$.

This means that the hypervolume indicator of the true Pareto front is maximal, because it is always better than or equal to any other possible nondominated set, and therefore $I_{\mathrm{H}}(\mathcal{PF}^*) \geq I_{\mathrm{H}}(\mathcal{A})$ for any nondominated set $\mathcal{A}$. Knowing that, it is obviously of interest to maximize the hypervolume indicator, so that the best possible approximation of the Pareto front can be found.
Furthermore in Zitzler et al. (2003) the following Lemma was shown.

Lemma 3.2 (Custódio et al. (2012), Zitzler et al. (2003))**:** Let $\prec$ be defined as in Definition 3.2, and $\mathcal{A}$ and $\mathcal{B}$ be two nondominated sets with the property $\forall \boldsymbol{\varphi} \in \mathcal{A} \cup \mathcal{B} : \boldsymbol{\varphi} \prec \boldsymbol{r}$, where $\boldsymbol{r}$ is the reference point used in the hypervolume computations. If $I_{\mathrm{H}}(\mathcal{A}) > I_{\mathrm{H}}(\mathcal{B})$ then $\mathcal{B} \not\prec \mathcal{A}$.

This means, that if an algorithm exists, which provably never decreases the hypervolume indicator of the current approximation of the Pareto front, then the approximation will never be worse than the approximation of the previous iteration.
The $\mathcal{S}$ metric selection evolutionary multi-objective optimization algorithm (SMS-EMOA) is such a method, which is presented in the following.

3.1.1 SMS-EMOA

SMS-EMOA is initialized with an initial population $\mathcal{P}_0$ of size μ. In each iteration of the algorithm one solution candidate $\boldsymbol{\varphi}$ is created out of the current population $\mathcal{P}_\kappa$ using variation. If the new solution improves the quality of the current population it is kept and another solution is deleted, else it is discarded. The SMS-EMOA algorithm is shown in Algorithm 3.1.

Algorithm 3.1 A SMS-EMOA algorithm (Beume et al., 2008)

Input: $\mathcal{P}_0 \leftarrow$ init
Input: $\kappa \leftarrow 0$
1:
2: **repeat**
3: $\quad \boldsymbol{\varphi} \leftarrow \text{variation}\left(\mathcal{P}_\kappa\right)$
4: $\quad \mathcal{D} \leftarrow \text{dominated\_individuals}\left(\mathcal{P}_\kappa \cup \{\boldsymbol{\varphi}\}\right)$
5: $\quad$ **if** $\mathcal{D} \neq \emptyset$ **then**
6: $\quad\quad \boldsymbol{\phi}^* \leftarrow \arg\max_{\boldsymbol{\phi} \in \mathcal{D}} d_\mathrm{n}\left(\boldsymbol{\phi}, \mathcal{P}_\kappa \cup \{\boldsymbol{\varphi}\}\right)$
7: $\quad$ **else**
8: $\quad\quad \boldsymbol{\phi}^* \leftarrow \arg\min_{\boldsymbol{\phi} \in (\mathcal{P}_\kappa \cup \{\boldsymbol{\varphi}\})} \Delta I_\mathrm{H}\left(\boldsymbol{J_x}\left(\boldsymbol{\phi}\right), \mathcal{PF}_\kappa \cup \{\boldsymbol{J_x}\left(\boldsymbol{\varphi}\right)\}\right)$
9: $\quad$ **end if**
10: $\quad \mathcal{P}_{\kappa+1} \leftarrow \left(\mathcal{P}_\kappa \cup \{\boldsymbol{\varphi}\}\right) \setminus \{\boldsymbol{\phi}^*\}$
11: $\quad \kappa \leftarrow \kappa + 1$
12: **until** some stopping criterion

In Algorithm 3.1 the number of dominating points d_n (card $(\mathcal{A})$ calculates the cardinality of the set $\mathcal{A}$)

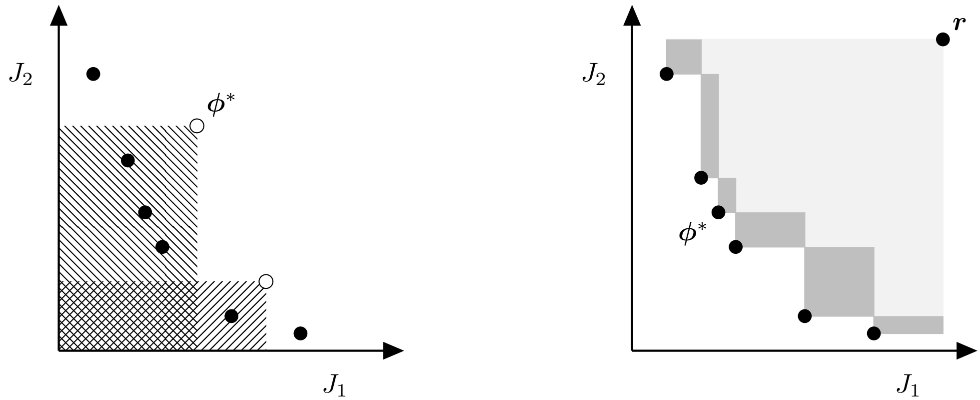
$$d_\mathrm{n}\left(\boldsymbol{\phi}, \mathcal{A}\right) := \mathrm{card}\left(\{\boldsymbol{a}_\mathcal{A} \in \mathcal{A} \,|\, \boldsymbol{a}_\mathcal{A} \prec \boldsymbol{\phi}\}\right) \tag{3.1}$$

and the contributing hypervolume ΔI_H

$$\begin{aligned} \Delta I_\mathrm{H}\left(\boldsymbol{a}_\mathcal{A}, \mathcal{A}\right) := I_\mathrm{H}\left(\mathcal{A}\right) - I_\mathrm{H}\left(\mathcal{A} \setminus \{\boldsymbol{a}_\mathcal{A}\}\right) \\ \text{with } \boldsymbol{a}_\mathcal{A} \in \mathcal{A} \end{aligned} \tag{3.2}$$

are used, cf. (Beume et al., 2007). In Algorithm 3.1 all dominated solutions are collected in the set $\mathcal{D}$ which are returned by the function dominated_individuals, called in line 4 of the algorithm. The population of solution candidates in iteration κ is symbolized by $\mathcal{P}_\kappa$, which is the current approximation of the Pareto optimal set. The corresponding approximation of the Pareto front is given by $\mathcal{PF}_\kappa$.

In Figure 3.2a the concept of the number of dominating points d_n is visualized. If there are dominated solutions, visualized as the two white circles in Figure 3.2a, then d_n specifies the number of solutions that dominate each dominated solution. One of the solutions with the largest d_n is deleted from the current population, here $\boldsymbol{\phi}^*$. If there is no dominated solution, thus all solutions are non-dominated within the population,



(a) The two dominated solutions (white circles) are dominated by solutions which lie in the shaded areas. The point ϕ^* has the higher dominance number, which is three compared to one.

(b) The dark shaded areas visualize ΔI_{H} of the points, whereas the area of ϕ^* is the smallest.

Figure 3.2: The solutions of a two-dimensional optimization problem are shown. The worst solution ϕ^* will be deleted from the current population, cf. Beume et al. (2008).

then the solution with the smallest contributing hypervolume ΔI_{H} is deleted, see Figure 3.2b. As by discarding the solution with the smallest contribution always a subset of size μ with largest hypervolume is selected, the hypervolume indicator I_{H} will never decrease. Either the new solution is directly deleted from the population, which leaves the hypervolume indicator unchanged or the new solution increases the hypervolume indicator.

3.2 SMS-EGO

$\mathcal{S}$-metric selection-based efficient global optimization (SMS-EGO) was first introduced in Ponweiser et al. (2008). SMS-EGO is a multi-objective variant of so called Efficient Global Optimization Algorithms (EGO) (Jones et al., 1998), which were earlier known as Statistical Global Optimization (Cox and John, 1997, Mockus et al., 1978).

In SMS-EGO a meta-model is used to predict objective function evaluations, that are assumed to be expensive. The meta-model is learned from previous exact evaluations. Based on the meta-model it is decided which point is evaluated next using the exact objective function.

The general idea of SMS-EGO is to replace during optimization the original objective function $\boldsymbol{J}$ with the meta-model generated one $\widehat{\boldsymbol{J}}$. Thus, an optimization method solves the optimization problem given by $\widehat{\boldsymbol{J}}$. The returned optimal solution is evaluated by the original objective function $\boldsymbol{J}$ and this result is used to update the meta-model restarting the optimization process again.

For each component J_{i_o} of the objective function a separate meta-model is created. As meta-model a DACE stochastic process model is used, where DACE is short for 'Design and Analysis of Computer Experiments' (Jones et al., 1998). Each such DACE model returns an estimate of the objective function $\hat{J}_{i_\mathrm{o}} \in \mathbb{R}$ and a standard deviation $\hat{s}_{J_{i_\mathrm{o}}} \in \mathbb{R}$ representing the uncertainty in the estimation. Both values are collected in the vectors $\widehat{\boldsymbol{J}} := \left(\hat{J}_1, \ldots, \hat{J}_{i_\mathrm{o}}, \ldots, \hat{J}_{n_\mathrm{o}}\right)^T \in \mathbb{R}^{n_\mathrm{o}}$ and $\hat{\boldsymbol{s}}_{\boldsymbol{J}} := \left(\hat{s}_{J_1}, \ldots, \hat{s}_{J_{i_\mathrm{o}}}, \ldots, \hat{s}_{J_{n_\mathrm{o}}}\right)^T \in \mathbb{R}^{n_\mathrm{o}}$.

As the meta-models also return the estimated uncertainty $\hat{\boldsymbol{s}}_{\boldsymbol{J}}$ the lower confidence bound $\widehat{\boldsymbol{J}}_\mathrm{pot} := \widehat{\boldsymbol{J}} - \alpha_\mathrm{LCB} \cdot \hat{\boldsymbol{s}}_{\boldsymbol{J}}$, with $\alpha_\mathrm{LCB} = -\Phi^{-1}_\mathrm{CDF}\left(0.5 \cdot \sqrt[n_\mathrm{o}]{0.5}\right)$ (Wagner et al., 2010, Emmerich et al., 2006), is used as the objective of some evaluated solution and not just $\widehat{\boldsymbol{J}}$. Here, $\Phi_\mathrm{CDF} : \mathbb{R} \to (0, 1)$ is the cumulative normal distribution function.

Each evaluated $\widehat{\boldsymbol{J}}$ is validated by a measure named additive ϵ-dominance, defined in Def. 3.3.

Definition 3.3 (cf. Zitzler et al. (2003))**:** Given two vectors of optimization variables $\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2 \in \mathcal{U}_\mathcal{F}$, it is said that $\underline{\boldsymbol{u}}_1$ ϵ-dominates $\underline{\boldsymbol{u}}_2$, being represented by $\underline{\boldsymbol{u}}_1 \preceq_{\epsilon+} \underline{\boldsymbol{u}}_2$, iff for some $\epsilon \in \mathbb{R}^+$ $\forall i_\mathrm{o} \in \{1, \ldots, n_\mathrm{o}\} : J_{\boldsymbol{x}, i_\mathrm{o}}\left(\underline{\boldsymbol{u}}_1\right) \leq J_{\boldsymbol{x}, i_\mathrm{o}}\left(\underline{\boldsymbol{u}}_2\right) + \epsilon$.

The single-objective function, which tries to find the optimum of $\widehat{\boldsymbol{J}}$ uses additive ϵ-dominance. It distinguishes between two kinds of solution candidates: ϵ-dominated and non-ϵ-dominated solution candidates, see Figure 3.3. Non-ϵ-dominated candidates $\boldsymbol{\varphi}_\mathrm{pot}$ yielding $\widehat{\boldsymbol{J}}_\mathrm{pot}$ are evaluated based on the negative value of their additional hypervolume contribution: $I_\mathrm{H}\left(\mathcal{PF}_\kappa\right) - I_\mathrm{H}\left(\mathcal{PF}_\kappa \cup \left\{\widehat{\boldsymbol{J}}_\mathrm{pot}\right\}\right)$, whereas $\mathcal{PF}_\kappa$ is the current approximation of the Pareto front of $\boldsymbol{J}$. However, ϵ-dominated solutions are penalized by a penalty given in equation (3.3), with $\mathcal{P}_\kappa$ being the current approximation of the Pareto optimal set (only containing non-dominated points), (Wagner et al., 2010).

$$p_\epsilon := \begin{cases} \max_{\boldsymbol{\varphi} \in \mathcal{P}_\kappa} \left[-1 + \prod_{i_\mathrm{o}=1}^{n_\mathrm{o}} \left(1 + \max\left(\widehat{J}_{\mathrm{pot}, i_\mathrm{o}} - J_{\boldsymbol{x}, i_\mathrm{o}}\left(\boldsymbol{\varphi}\right), 0\right)\right)\right] & \text{if } \boldsymbol{\varphi} \preceq_{\epsilon+} \boldsymbol{\varphi}_\mathrm{pot} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

In equation (3.4) the single-objective function is shown, that is used to find an optimal solution candidate to be evaluated at the original objective function $\boldsymbol{J}$.

$$f_\mathrm{EGO} := \begin{cases} I_\mathrm{H}\left(\mathcal{PF}_\kappa\right) - I_\mathrm{H}\left(\mathcal{PF}_\kappa \cup \left\{\widehat{\boldsymbol{J}}_\mathrm{pot}\right\}\right) & \text{non-}\epsilon\text{-dominated } \widehat{\boldsymbol{J}}_\mathrm{pot} \\ p_\epsilon & \epsilon\text{-dominated } \widehat{\boldsymbol{J}}_\mathrm{pot} \end{cases} \tag{3.4}$$

In SMS-EGO this objective function is minimized using an interior point method.

The value for ϵ is calculated as in Ponweiser et al. (2008) using $\epsilon = \frac{\Delta \mathcal{PF}_\kappa}{\mathrm{card}(\mathcal{PF}_\kappa) + c \cdot n_\mathrm{left}}$.

There, $\Delta\mathcal{PF}_\kappa := \max(\mathcal{PF}_\kappa) - \min(\mathcal{PF}_\kappa)$, where

$$\max(\mathcal{PF}_\kappa) := \left(\max_{\boldsymbol{J}_{\boldsymbol{x}} \in \mathcal{PF}_\kappa} J_{\boldsymbol{x},1}, \ldots, \max_{\boldsymbol{J}_{\boldsymbol{x}} \in \mathcal{PF}_\kappa} J_{\boldsymbol{x},i_\mathrm{o}}, \ldots, \max_{\boldsymbol{J}_{\boldsymbol{x}} \in \mathcal{PF}_\kappa} J_{\boldsymbol{x},n_\mathrm{o}} \right)^T,$$

likewise $\min(\mathcal{PF}_\kappa)$. Furthermore, $c = 1 - \frac{1}{2^{n_\mathrm{o}}}$ is a correction factor and n_left is the number of remaining evaluations (Ponweiser et al., 2008).

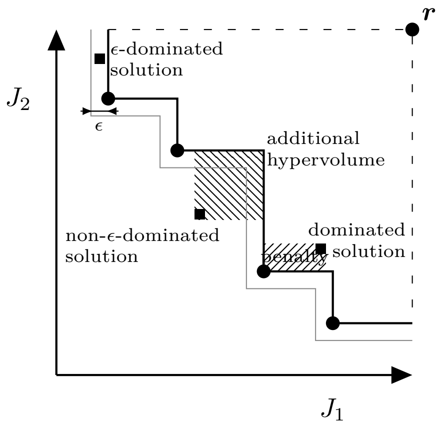


Figure 3.3: Graphical explanation of the concept of ϵ-dominance used in SMS-EGO, cf. Ponweiser et al. (2008).