



## A Novel Approach to Configuration Redesign: Using Multiobjective Monotonicity Analysis to Alter the Pareto-set

Sigurdarson, Nökkvi S.; Eifler, Tobias; Ebro, Martin; Papalambros, Panos Y.

*Published in:*  
Journal of Mechanical Design

*Link to article, DOI:*  
[10.1115/1.4053524](https://doi.org/10.1115/1.4053524)

*Publication date:*  
2022

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Sigurdarson, N. S., Eifler, T., Ebro, M., & Papalambros, P. Y. (2022). A Novel Approach to Configuration Redesign: Using Multiobjective Monotonicity Analysis to Alter the Pareto-set. *Journal of Mechanical Design*, 144(6), Article 061704. <https://doi.org/10.1115/1.4053524>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# A Novel Approach to Configuration Redesign: Using Multiobjective Monotonicity Analysis to Alter the Pareto-set

**Nökkvi S. Sigurdarson\***  
 Mechanical Engineering,  
 Technical University of Denmark,  
 Kgs. Lyngby, Denmark,  
 noksig@mek.dtu.dk

**Tobias Eifler**  
 Mechanical Engineering,  
 Technical University of Denmark,  
 Kgs. Lyngby, Denmark,  
 tobeif@mek.dtu.dk

**Martin Ebro**  
 Device R&D,  
 Novo Nordisk A/S,  
 Hillerød, Denmark,  
 mixe@novonordisk.com

**Panos Y. Papalambros**  
 Mechanical Engineering,  
 University of Michigan,  
 Ann Arbor, MI 48109,  
 pyp@umich.edu

*Configuration (or topology or embodiment) design remains a ubiquitous challenge in product design optimization and in design automation, meaning configuration design is largely driven by experience in industrial practice. In this article, we introduce a novel configuration redesign process founded on the interaction of the designer with results from rigorous multiobjective monotonicity analysis. Guided by Pareto-set dependencies, the designer seeks to reduce trade-offs among objectives or improve optimality overall, deriving redesigns that eliminate dependencies or relax active constraints. The method is demonstrated on an ingestible medical device for oral drug delivery, currently in early concept development.*

## Nomenclature

$\mathcal{A}$  Attainable set  
 $\mathcal{C}$  Pareto Set  
 $\mathbf{c}$  Vector of bound objectives in the upper bound problem  
 $D_s$  Indices of the constraint functions that depend on a shared variable  $x_i$   
 $f$  Primary objective function in the upper bound problem  
 $f(x^+)$  A function increasing monotonically w.r.t.  $x$   
 $f(x^-)$  A function decreasing monotonically w.r.t.  $x$   
 $F^0$  The utopia point  
 $\mathbf{g}(\mathbf{x})$  Vector of inequality constraints of the design problem

$\mathbf{h}(\mathbf{x})$  Vector of equality constraints of the design problem  
 $j$  Number of computational iterations  $\epsilon$  is sampled over  
 $k$  Number of objectives  
 $p$  Number of redesign iterations  
 $\mathcal{X}$  The set constraint  
 $\mathbf{x}$  vector of design variables  
 $\underline{x}$  Argument of the infimum of the design problem  
 $\bar{x}$  Argument of the supremum of the design problem  
 $\bar{x}$  Trade-off variable  
 $\bar{\bar{x}}$  A monotonically decreasing harmonious variable  
 $\underline{\underline{x}}$  A monotonically increasing harmonious variable  
 $\epsilon$  A  $k-1$  dimensional vector of upper-bound parameters  
 $\epsilon_i$  Upper-bound parameter for the  $i$ th bound objective  
 $\epsilon_L$  Lower limit of objective bounds  
 $\epsilon_U$  Upper limit of objective bounds  
 $\tilde{\epsilon}_i$  Reduced-objective variable

## 1 Introduction

In the 'double-diamond' model of the design process [1], the first diamond results in the generation of a design concept in a functional form as starting point for the generation of a particular embodiment design in the second diamond. Product design optimization typically iterates on the particular embodiment instantiation to achieve optimality with respect to given objectives while satisfying a set of given constraints. In practice, starting with a particular design and

\*Corresponding Author

finding an optimum iteratively is in fact a design improvement or redesign process. Such improvements are associated with (i) proportional changes, where design variables are resized; (ii) parametric changes, where parameters, e.g. material properties or production processes are modified allowing for relaxation of design constraints; and (iii) configuration changes where the embodiment of functions or distribution of sub-functions is modified.

Proportional (or size) design has been facilitated by advancements in computational modeling, computing power, and optimization techniques. More robust and higher fidelity numerical solutions in size optimization have reduced risk in constraint relaxation for increasingly larger and more complex problems. Parametric design with gradual constraint relaxation has been a primary approach for product performance improvement in industry. Examples of such include new or improved materials, new production processes, and a deeper understanding of failure phenomena leading to reduced design margins, as seen in the design of engines [2].

Configuration design, also referred to as embodiment design [3] and layout design [4], is the process of creating actual 'buildable' instantiations of a design concept, transitioning from an initial abstract functionality to a more geometric realisation. Conceptual functionality may be implemented in various ways, and so many alternative configurations may be created, one of which is selected as "the best" for further development. Critical decisions that determine how a system fits together, how its parts interact with each other, and how the system interacts with its surroundings are made in the embodiment/configuration design stage. Examples include the layout and distribution of parts and sub-functions, force paths, mechanism design, and assembly sequence [3, 5].

In design optimization, configuration design remains elusive. The fundamental challenge is the lack of appropriate mathematical modeling capabilities: different configurations require different mathematical problem formulations. A notable exception is topology optimization [6], the success of which lies in the introduction of a unified mathematical model across configurations using a tensor field representation. Some success has also been achieved through combinatorial methods such as grammars [7] and graph-based models [8, 9]. These use a predefined set of system elements that are combined by some algorithm. Both approaches have limitations in the context of early iterations in configuration design. Topology optimization relies on a predefined set of boundary conditions and loads, while combinatorial techniques can only capture configurations accounted for in the model. For further reference on computational synthesis see Antonsson & Cagan [10] or Chakrabarti et al [11].

In product design, configuration methods mostly involve heuristics (e.g., axiomatic design [12] and TRIZ [13]), error avoidance [3, 5], or prescription of specific characteristics under the 'Design for X' moniker [14–16]. Configuration improvement is often prescribed by simply avoiding dependencies among design objectives [3, 12]. Such prescriptions tend to disregard the effect of active (binding) constraints, which often link objectives indirectly through their activity, a situation common in mechanical design.

Dependencies and active constraints are inevitable in practice. Organizations generally remain competitive by integrating more features while improving performance in each new product generation [17], preferably with as little production complexity as possible. Increased dependency comes with increased integration, leading to more trade-offs between objectives [3] and a need for more design iterations [18], ultimately influencing development time and cost.

Formal design optimization techniques offer a rigorous route to the study of trade-offs. Their utility in the design process beyond the identification of the optimum is often touted [19–21]. Yet, they provide little systematic support in identifying when and how to improve the problem formulation by changing the configuration design, as they mostly deal with proportional or parametric changes. In this context, the application of Monotonicity Analysis (MA) [22] might further such understanding. MA is an opportunistic approach used to identify active constraints and allow model reduction by revealing dependencies in the design that are unique to the optimum. Although some work has been done on the application of MA in aiding configuration design (c.f. [21, 23]), its potential remains relatively unexplored.

In view of the above limitations, successful configuration design in practice is largely dependent on the proficiency and experience of the designer [24, 25]. Management of trade-offs [24] and constraints [26, 27] is a key differentiator between novice and experienced designers. Ahmed et al. [24] found that experienced designers were intuitively aware of the trade-offs they need to deal with and were focused on doing so upfront, while novice designers were more prone to trial and error. Without an understanding of how to configure a system in a way that limits trade-offs, inexperienced designers are thus left at a disadvantage. We hypothesize that the methodical approach presented here can provide such insights and guide configuration redesign.

This article presents a systematic, analytical foundation for configuration redesign using *Pareto-set Dependency Analysis* [28]. This analysis relies on multiobjective monotonicity analysis to identify dependencies in the Pareto-set that cause trade-offs while also accounting for the constraints. It derives relationships necessary at optimality and thus helps the designer identify configuration design changes that yield performance improvements beyond what could be reached by mere size and parameter changes. Section 2 presents the theoretical foundations of this work, followed by a description of the methodology in Section 3. Section 4 presents the case study; the Self-Orienting Millimeter-Scale Applicator (SOMA), which is a mechanical pill for the oral delivery of insulin and other drugs [29]. Finally, a discussion and conclusion is offered in Sections 5 and 6, respectively.

## 2 Theoretical Foundation

*Multiobjective design optimization problems* are stated in negative-null form as ([19]):

$$\min \mathbf{f}(\mathbf{x}) \quad (1)$$

$$\begin{aligned} \text{subject to } \mathbf{g}(\mathbf{x}) &\leq 0 & (2) \\ \mathbf{h}(\mathbf{x}) &= 0 & (3) \\ \mathbf{x} &\in \mathcal{X} & (4) \end{aligned}$$

where  $\mathbf{f}(\mathbf{x})$  is a vector of objectives  $f_i$ ,  $i = [1, 2, \dots, k]^T$  to be minimised,  $\mathbf{x}$  is a vector of real-valued design variables,  $\mathbf{h}(\mathbf{x})$ ,  $\mathbf{g}(\mathbf{x})$  are the equality and inequality constraints respectively, and  $\mathcal{X}$  is the set constraint that may include additional restrictions besides those of Eq. (2) and (3). If the set constraint is in just the real space, then  $\mathcal{X}$  denotes the feasible domain consisting of the  $\mathbf{x}$  values fulfilling (2) and (3). The attainable set  $\mathcal{A}$  contains all feasible values of  $\mathbf{f}(\mathbf{x})$ . A point  $\mathbf{f}(\mathbf{x}^*) \in \mathcal{A}$  is said to be Pareto-optimal if and only if there exists no other point  $\mathbf{f}(\mathbf{x}) \in \mathcal{A}$  such that  $\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{x}^*) \wedge f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$ . The Pareto set  $\mathcal{C}$  containing all Pareto-optimal points lies on the boundary of  $\mathcal{A}$  facing the origin, hence it is also referred to as the Pareto frontier. The utopia point  $F^0$  is a  $k$ -dimensional point consisting of all the single-objective minima, which lies outside of  $\mathcal{A}$  and is often used as a reference for evaluating Pareto points.

The *upper-bound formulation* is one way of identifying the Pareto set (see e.g. [30]). It creates a scalar substitute problem by splitting  $\mathbf{f}(\mathbf{x})$  into a single objective function  $f(\mathbf{x})$  and a vector of (upper) bound objectives,  $\mathbf{c}(\mathbf{x}; \boldsymbol{\varepsilon})$ :

$$\min \quad f(\mathbf{x}) \quad (5)$$

$$\text{s.t.} \quad \mathbf{c}(\mathbf{x}; \boldsymbol{\varepsilon}) \leq 0 \quad (6)$$

$$\mathbf{g}(\mathbf{x}) \leq 0 \quad (7)$$

$$\mathbf{h}(\mathbf{x}) = 0 \quad (8)$$

$$\mathbf{x} \in \mathcal{X}, \boldsymbol{\varepsilon} \in \mathcal{R}^{k-1}, \quad (9)$$

where  $\mathbf{c}$  is a  $k-1$  dimensional vector of bound objectives expressed in the form  $c_i(\mathbf{x}; \boldsymbol{\varepsilon}_i) = f_{i+1}(\mathbf{x}) - \boldsymbol{\varepsilon}_i \leq 0$  or  $c_i(\mathbf{x}; \boldsymbol{\varepsilon}_i) = \boldsymbol{\varepsilon}_i - f_{i+1}(\mathbf{x}) \leq 0$ ,  $i = [1, 2, \dots, (k-1)]$ ;  $\boldsymbol{\varepsilon}$  is a vector of parameters  $\boldsymbol{\varepsilon}_i$  in the real space  $\mathcal{R}^{k-1}$  for the bound objectives. When  $f(\mathbf{x})$  is minimised for given values of  $\boldsymbol{\varepsilon}_i$ , then the solution  $\mathbf{x}^*$  is Pareto optimal if all of the bound objectives are active with non-zero Lagrange multipliers. Pareto points are thus identified by varying  $\boldsymbol{\varepsilon}$  systematically between lower and upper limits  $\boldsymbol{\varepsilon}_L$ . See [30, 31] for an overview of the upper bound formulation (also known as  $\boldsymbol{\varepsilon}$ -constraint), the underlying mathematics, and how to define limits for  $\boldsymbol{\varepsilon}$ . As discussed in [28], this formulation has some computational limitations but its use here benefits the proposed analysis.

*Monotonicity Analysis* [22] leverages any monotonic behaviour in objective- and constraint functions to reduce optimization models and check their boundedness. A scalar function is monotonically increasing with respect to a variable  $x$ , if it holds that  $f(x_2) > f(x_1)$  for any  $x_2 > x_1$ , denoted as  $f(x^+)$ , and is said to be monotonically decreasing w.r.t.  $x$ , if it holds that  $f(x_2) < f(x_1)$ , denoted as  $f(x^-)$ . In the presence of monotonicity, the following principles [19] can be used to find active inequality constraints without needing to find the optimum first and prior to any computation:

*First Monotonicity Principle (MP1)*: In a well-constrained minimization problem, every increasing variable is bounded below by at least one non-increasing active constraint.

*Second Monotonicity Principle (MP2)*: In a well-constrained minimization problem every nonobjective variable is bounded both below by at least one non-increasing semi-active constraint and above by at least one non-decreasing semi-active constraint.

If an active constraint can be identified, it can be used to 'partially minimize' the model; namely, solve the constraint function with respect to one of its dependent variables and back-substitute the solution for that variable into the objective and remaining constraint functions, thus reducing the model dimensionality. This process reveals the relationships that *necessarily* exist at the optimum as a consequence of the constraint activity. For simplification and consistency with typical design situations, it is customary in monotonicity analysis to assume that the set constraint is the strictly positive real space  $\mathcal{P}$ . Hence, from here on, we assume that the feasible domain  $\mathcal{X}$  is a subset of  $\mathcal{P}$ .

In [28], these principles were applied to derive the Multiobjective Monotonicity Analysis (MOMA) process for rigorous identification of the dependencies that cause trade-offs between objectives in problems of the form shown in Eqs. 6-9. In well-bounded problems, a set of theorems and definitions specific to MOMA can be used to simultaneously reduce multiple objectives so that the degrees of freedom remaining in the Pareto set, and the constraints that shape it, are revealed. For the analysis in the present paper, we need the following definition and theorem:

#### Definition 1 Trade-off and Harmonious Variables

If an objective pair  $f$  and  $c_i$  have a variable  $x_1$  in common, but differ in monotonicity w.r.t.  $x_1$ , e.g.,  $f(x_1^+)$  and  $c_i(x_1^-)$ , then  $x_1$  is said to be a trade-off variable, denoted  $\bar{x}_1$ . Correspondingly, an objective pair of like monotonicity w.r.t. a common variable, indicates that the variable is harmonious and can be used to partially minimise both simultaneously.

#### Theorem 1 Influence of Monotonic Trade-off Variables

In the presence of monotonic trade-off variables, no dominant minimum exists, resulting in a Pareto set. The proof for this is a corollary to MP1.

*Proof.* Let  $f_1$  be monotonically increasing w.r.t.  $x \in \mathcal{P}$  and  $f_2$  monotonically decreasing, and let  $x$  be well bounded from above and below. Then by MP1,  $\arg \min f_1(x) = \underline{x}$ , and  $\arg \min f_2(x) = \bar{x}$ , meaning that the minimizers for the two objectives are respectively defined by the *greatest lower bound* (glb) and the *least upper bound* (lub) of  $x$ . Hence any feasible value of  $x$  will yield a unique Pareto point. ■

Using this basic insight, along with other theorems that ensure correct model reduction [28], MOMA allows identification of the conditions under which the bound objectives are active, i.e., the values of  $\boldsymbol{\varepsilon}$  that affect the feasible domain of  $\mathbf{x}$ , as illustrated in Fig. 1. This in turn allows reduction of multiobjective problems to reveal dependencies existing at the Pareto set. Furthermore, determining



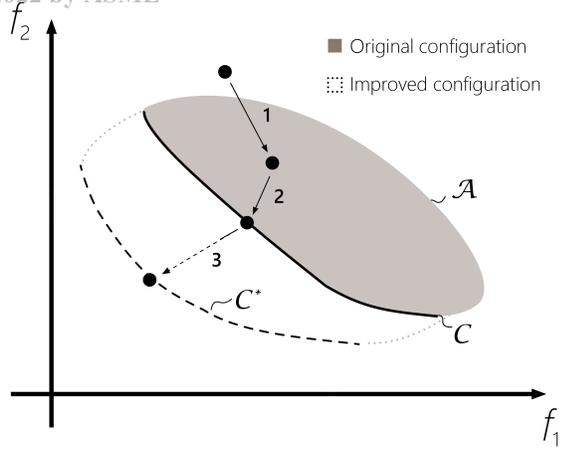


Fig. 2: The difference between constraint satisfaction (1), optimization (2), and configuration improvement (3). In this example the meta Pareto set is identical to  $C^*$

### Theorem 2 Existence of Pareto set

If no trade-off variables exist (globally or regionally) after back-substitution of active constraints, then the optimum is a point  $F^*$ , rather than a set. Therefore, a Pareto set cannot exist without trade-off variables.

*Proof.* If  $x_i \notin \bar{x}$ , for any design variable  $i$ , then following Theorem 1 and MP1,  $\text{argmin} f_i(\mathbf{x}) = \text{argmin} f_j(\mathbf{x})$  for  $i \wedge j = 1..k, i \neq j$ . Hence,  $\dim(\mathbf{x}^*) = [n, 1]$ , meaning a single dominant optimum exists. ■

Thus, the more trade-off variables exist in a design problem, the larger the distance between the utopia point  $F^0$ , and  $C$ . Furthermore, recall from Theorem 1 that the optimum of each objective dependent on a trade-off variable  $\bar{x}_i$ , exists at either  $\underline{x}_i$  or  $\bar{x}_i$  depending on the objective's monotonicity, with any other feasible value of  $\bar{x}_i$  yielding a Pareto point. As a result, the size of the feasible domain of trade-off variables contributes to the size of the Pareto set.

Finally, being oppositely monotonic, the partial derivatives of an objective pair in trade-off w.r.t  $\bar{x}_i$  will have an opposite sign across the entire Pareto-set. The larger the difference between these, the larger the slope of the frontier. Thus, the impact of  $\bar{x}_i$  on the trade-offs between objectives can be worsened by multipliers and divisors.

In summary, the trade-off variables in an optimization problem cause the existence of the Pareto set, determining its span and to some extent its shape. From this observation, we can derive several corollaries which reveal design principles to guide generation of an improved Pareto set.

### Corollary 2.1 Separation of Trade-off Variables

If the trade-off variable  $\bar{x}_1$  affecting the objectives  $f_i(x_1^+)$  and  $f_j(x_1^-)$ , is substituted through design change in one objective by a new variable  $x_2$ , such that  $\hat{f}_i(x_2^+), x_2 \notin f_j$ , and  $\min \hat{f}_i \leq \min f_i$ , then  $\text{argmin} \hat{\mathbf{f}}(\mathbf{x}) = \{\bar{x}_1, x_2\}$  while  $\text{argmin} \mathbf{f}(x_1) = x_1 \in \mathcal{X}$ . As a result  $\hat{\mathbf{f}}(\mathbf{x}) < \mathbf{f}(\bar{x}_1)$  for any value of  $x_1$ .

The same would also apply if  $x_2$  were substituted into the problem such that  $\hat{f}_j(x_2^-)$ , or if the influence of  $x_1$  upon one of the objectives in a multivariate problem was simply eliminated without the introduction of another variable. In such problems, it follows that such changes would result in a new Pareto set  $C_2$  that at least weakly dominates the original Pareto set,  $C_1$ . The term separation here is used in the same spirit as in TRiZ [13], reflecting the removal of a dependency.

Correspondingly, the same would occur if a design change is introduced that makes  $x_1$  a harmonious variable without otherwise affecting the objective functions. In fact, any modification of the design problem which eliminates or reduces the influence of the trade-off variable on one objective may result in a new, weakly dominant Pareto set. This is stated formally with the following corollaries.

### Corollary 2.2 Flipping Trade-off Variables

If the monotonicity of a trade-off variable  $\bar{x}_n$  affecting the objectives  $f_i(x_n^+)$  and  $f_j(x_n^-)$ , is flipped in one objective through design change, such that  $\hat{f}_i(x_n^-)$ , and  $\min \hat{f}_i \leq \min f_i$ , then  $\text{argmin} \hat{\mathbf{f}}(\mathbf{x}) = \bar{x}_n$  whereas  $\text{argmin} \mathbf{f}(x_n) = x_n \in \mathcal{X}$ . As a result  $\hat{\mathbf{f}}(x_n) < \mathbf{f}(\bar{x}_n)$ .

### Corollary 2.3 Scaling Trade-off Variables

If the influence of a trade-off variable  $\bar{x}_n$  affecting the objectives  $f_i(x_n^+)$  and  $f_j(x_n^-)$  is scaled through the introduction of an independent variable  $x_{n1}$  in  $f_i$  such that  $\partial \hat{f}_i / \partial x_n < \partial f_i / \partial x_n$  then  $\min \hat{f}_i < \min f_i$  for any value of  $x_n$ , reducing the trade-off between  $f_i$  and  $f_j$ . Correspondingly, if  $\partial \hat{f}_j / \partial x_n > \partial f_j / \partial x_n$  then  $\min \hat{f}_j < \min f_j$ .

As mentioned, harmonious variables are shared between objectives of like monotonicity, denoted  $\bar{x}$  when the objectives are monotonically decreasing, and  $\underline{x}$  when they are increasing. For such variables, the glb (for  $\underline{x}$ ) or lub (for  $\bar{x}$ ) is active at all Pareto points. They might be optimized out using MOMA if globally active constraints are identified, or remain in the model if there are regionally active constraints. While trade-off variables create the Pareto set, identifying harmonious variables reveals other useful information:

### Theorem 3 Position of Pareto Set $C$

Harmonious variables,  $\underline{x}$  and  $\bar{x}$ , affect the position of the Pareto set  $C$  relative to the origin. Thus design changes that widen their feasible domains in an improving direction, yield a new strongly dominant Pareto set,  $C_{i+1} < C_i$ .

*Proof.* Let  $f_i$  and  $f_j$  depend on  $\bar{x}_1, \underline{x}_2, \bar{x}_3$ , i.e.,  $f_i(x_1^-, x_2^+, x_3^+)$ ,  $f_j(x_1^-, x_2^+, x_3^-)$  where  $x_1, x_2, x_3 \in \mathcal{P}$ . If the problem is well bounded, then by MP1 and Theorem 1,  $\text{argmin} f_i(\mathbf{x}) = \{\bar{x}_1, \underline{x}_2, \bar{x}_3\}$  and  $\text{argmin} f_j(\mathbf{x}) = \{\bar{x}_1, \underline{x}_2, \bar{x}_3\}$ . If the active constraints are modified or relaxed, such that  $\bar{x}_1 < \hat{\bar{x}}_1$  and/or  $\hat{x}_2 < \underline{x}_2$ , then  $\hat{\mathbf{f}}(\mathbf{x}) < \mathbf{f}^*(\mathbf{x})$  for any value of  $\bar{x}_3$ , given the monotonicity of  $f_i$  and  $f_j$ . Hence,  $\hat{C} = C^*$ . The reverse is true if the active constraints are tightened, such that  $\bar{x}_1 > \hat{\bar{x}}_1$  and/or  $\hat{x}_1 > \underline{x}_1$ . Hence, the harmonious variables and their bounds influence the position of  $C$ . ■

Whereas changing the bounds of trade-off variables only enlarges the Pareto set and moves its utopia point, relax-

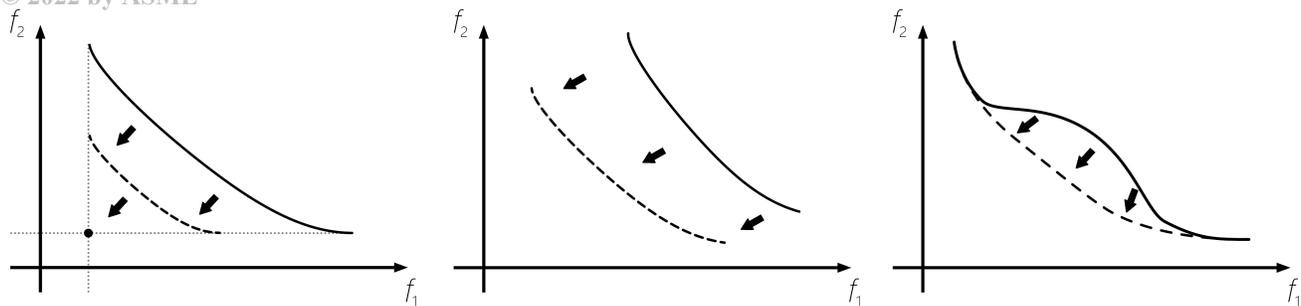


Fig. 3: The relation between analysis outputs and the Pareto set. (Left): Trade-off variables result in an optimum that is a set rather than a point; eliminating the underlying dependency brings the set closer to utopia. (Center): Harmonious Variables affect the position of the Pareto set; relaxing their constraints shift the set. (Right:) Pareto constraints introduce regional relationships that affect the Pareto set; eliminating or relaxing them changes the shape of the set

ing the constraints of harmonious variables results in an improved Pareto set. Furthermore the slope of  $C$  will be affected as well, unless  $x_1$  and  $x_2$  influence  $f_i$  and  $f_j$  equally.

Lastly, Pareto-constraints are a consequence of the systematic reduction of multiobjective problems modelled in an upper bound formulation. When globally active, i.e., for any  $\tilde{\epsilon}_L \leq \tilde{\epsilon} \leq \tilde{\epsilon}_U$ , Pareto constraints allow the derivation of terms of the form  $\tilde{\epsilon}_i(\mathbf{x}, \tilde{\epsilon})$ , revealing additional trade-off variables while describing the relationship that exist between the objectives at the Pareto-set. In this case, they are merely a representation of trade-off variables, albeit one which may significantly impact the Pareto set. When they are regionally active, however, Pareto constraints reveal regional trade-off variables. This occurs when some of the constraints in the non-reduced model become active for specific values of  $\epsilon$ , causing discontinuous trade-offs. In higher dimensional problems,  $k \geq 3$ , regionally active Pareto constraints might cause a Pareto frontier between an objective pair. Such situations can be studied through a case analysis procedure described in [28]. Thus, Pareto constraints may, when studied, reveal additional trade-off variables or discontinuous relationships such as variables that are in trade-off in specific regions of the Pareto-set, thereby affecting its shape.

In summary, Pareto-set dependency analysis helps explain the relationship between the design problem and the shape of the Pareto set. Thus, we can utilise these theorems and corollaries to derive a set of redesign principles.

### 3.2 Configuration Redesign Principles

Insights into the relationship between the design problem and the shape of the Pareto set is of substantial value in the synthesis and improvement of configuration designs. As discussed in [28], the Pareto set is created by variables and constraints that are shared between objectives. Some shared variables can be used to improve upon several objectives simultaneously, while others cannot. To a large extent, these relationships are determined by decisions made in conceptual and configuration design. Designers hence need to identify and manage global (i.e. shared variables) and regional dependencies (i.e., shared active constraints) at an early stage to reach *good* configuration designs. We posit

that experienced designers apply tacit knowledge of constraints [26] and trade-offs [24] to synthesise and improve configurations. They use this knowledge to configure the components of a system in a way that leverages harmonious variables to achieve a high performance, e.g., placing rotating components as far inside an assembly as possible and load-bearing components as far outside. Similarly, they will attempt to avoid trade-off variables or obviate them through design changes.

Reaching the required insights is not trivial, especially in highly interdependent systems. Pareto set dependency analysis bridges this gap, providing a causal link between optimality and configuration design limitations. This understanding allows more informed and deliberate identification, prioritisation, and handling of the dependencies that cause trade-offs. The introduced theorems, proofs, and corollaries demonstrate how certain types of model transformation based on the results of this analysis lead to an improved Pareto set. Translating these transformations into specific design changes would mitigate the dependencies that create the Pareto set and relax the constraints that position it, just as experienced designers do through tacit knowledge.

In this spirit, we state four reconfiguration design principles, illustrated in Figs. 4-6, stemming from the theorems and corollaries presented in Section 3.1. When employed in configuration redesign, these principles lead to an associated improvement of the Pareto set. Within each principle, we state a number of more specific strategies stemming from basic model transformations that result in an improved Pareto set. Each represents an alternative way of implementing the principles and corresponds to certain forms of design change. The figures illustrate each of the four principles and the available strategies within each principle:

1. **Align Trade-off Variables.** Reduce or eliminate the effect of trade-off variables on the objectives without impacting their single-objective optima, thereby improving their alignment and the Pareto set, c.f. Theorem 2. This involves eliminating the dependency, making the variable harmonious, or scaling it, c.f. Corollaries 2.1-2.3.
2. **Leverage Harmonious Variables.** Widening the feasi-

# Align Trade-off Variables

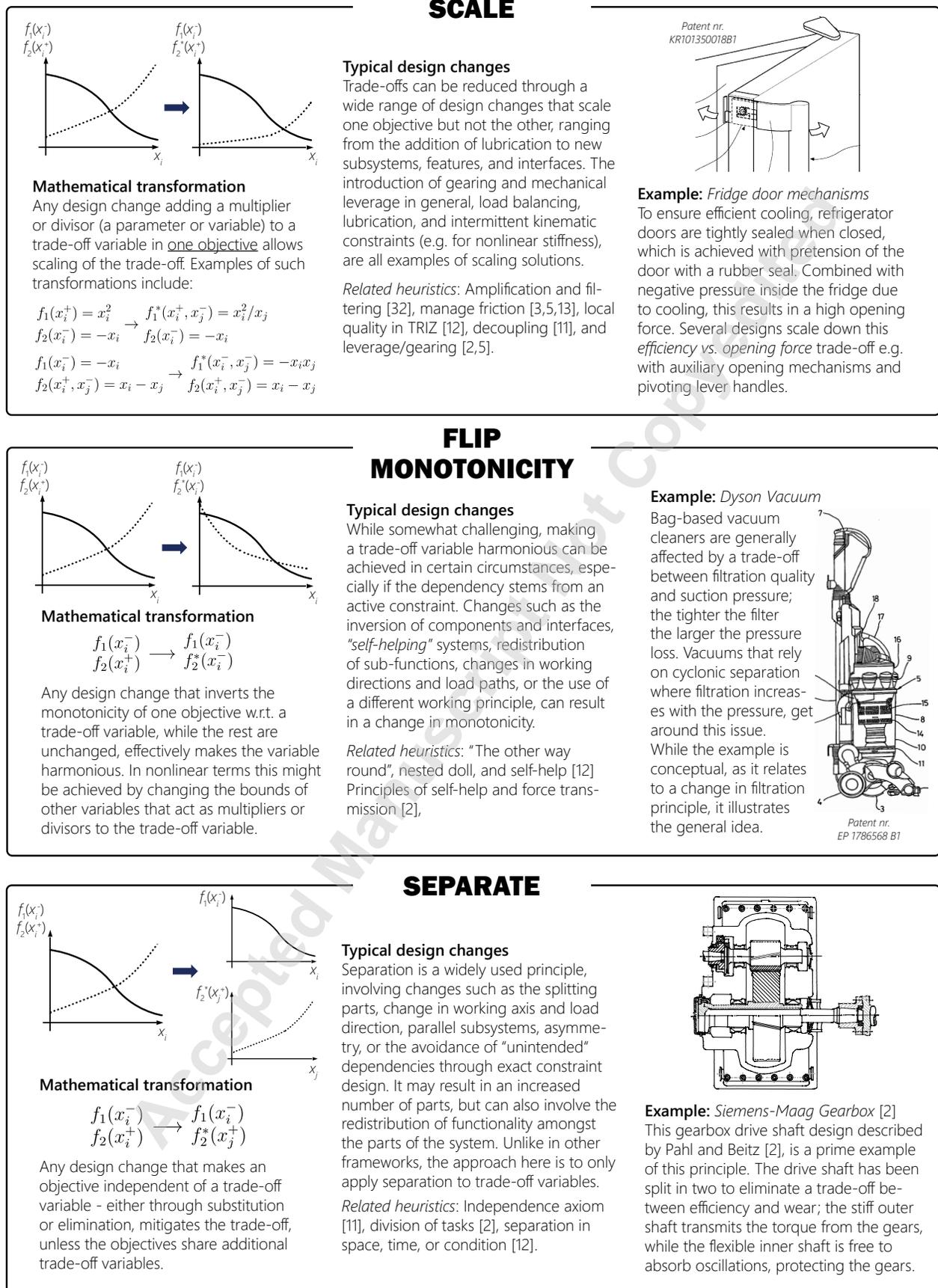
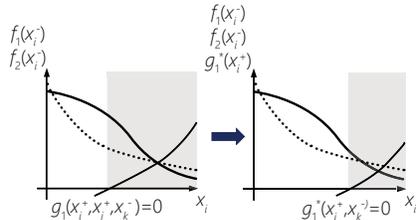


Fig. 4: The strategies within Principle 1: Reduce or eliminate the impact of a trade-off variable upon an objective pair

# Leverage Harmonious Variables

## SHIFTED BOUNDS



### Mathematical transformation

Any design change that shifts the *glb/lub* of a harmonious variable, improves the optimum of its dependant objectives. This involves eliminating increasing contributors, introducing additional decreasing contributors, or scaling parts of the constraint. This can also scale a trade-off, when the harmonious variable is a multiplier or divisor of a trade-off variable.

### Typical design change

In configuration design terms, these changes are specific to the type of constraint. Generally speaking this is oft matter of positioning components in an assembly in the most beneficial way - e.g. locating parts with decreasing variables as far *inside* an assembly as possible and increasing variables on the outside. Further, it involves designing to avoid unnecessary contributors to the active constraint, e.g. stress concentrations and associated loads in structural constraints.

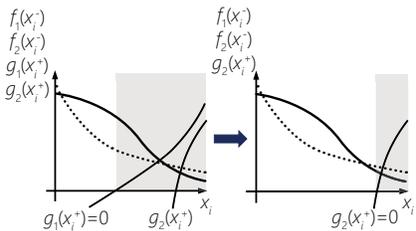
*Related heuristics:* Reduce information content [12], Principle of balanced forces [3], Minimise tolerance paths [5]



### Example: Mazda Skyactiv-G [34]

In the design of combustion engines, thermal efficiency increases with the compression ratio. Yet, this ratio cannot be increased beyond a point where knocking occurs, which is in part driven by residual gas after combustion. Most petrol engines hence have a ratio between 8:1-12:1. In the Skyactiv engine, Mazda pushed this ratio 14:1, using a longer exhaust manifold, increasing gas scavenging, and shifting the knocking constraint.

## CONSTRAINT RELAXATION



### Mathematical transformation

When a harmonious variable is actively (but not critically) constrained, we might try to change the configuration design in a way that eliminates the active constraint. This shifts the *glb* or *lub* of the variable to the next constraint, improving the optimum of all its dependent objectives.

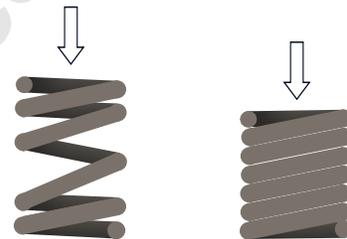
### Typical design changes

As with shifted bounds, the changes required to eliminate a constraint, are contextual. Examples include changes aimed at redirection of force paths to eliminate a load case, a new part structure to avoid certain parts being bound by limiting geometric constrains (e.g. one part inside another), a change in assembly sequence to avoid some alignment constraint.

*Related heuristics:* Vary the structure of main elements [15,], redirect load path [5], merge parts [5,14], shielding [34].

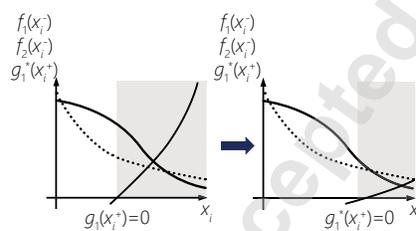
### Example: Spring strength at block

A well known example of load path redirection, compression springs are self-



reinforcing when deformed to their block length. In design applications where a maximum load resistance is desired, a spring design that is deformed to its' block length rather than to its' elastic yield limit is far stronger. Introducing such a change to a design, is equivalent to eliminating the yield constraint driven by shear stress.

## NEW FUNCTIONAL FORM



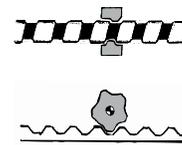
### Mathematical transformation

A complete change in functional form of an active constraint, may yield a widened feasible domain. This is distinct from *shifted bounds*, in that it involves the entire function, and may hence result in changed constraints, monotonicity, exponents, and so on.

### Typical design changes

Such a drastic model change will most likely probably require substantial design change, e.g. a change in components, working principles, and/or the physics of the problem. Examples of such include a change in production process, the separation or combination of parts, a change in the realisation of a given sub-function, a change in load type and distribution, and so on.

*Related heuristics:* Design for pure compression and tension [5], Select rotary over linear motion [3,5] Self-help [3,13]



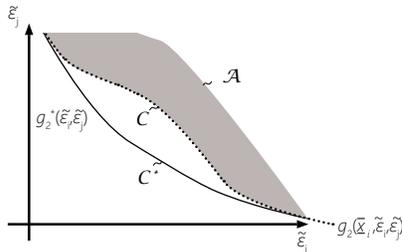
### Example: Rotary to linear movement

A rack and pinion and a lead screw fundamentally meet the same functional purpose - to convert rotation into linear motion, or vice versa. Yet, what is superior, depends on the objectives, primarily due to quite different constraints involved in their design. For instance, the rack for instance only slides, and as a result the mechanical stress expressions are quite different, compared to the rotating screw, which is why they are commonly used for high load applications.

Fig. 5: The strategies within Principle 2: Increase the influence of harmonious variables

# Relax Pareto Constraints

## SHIFTED PARETO CONSTRAINT



**Mathematical transformation**  
 The elimination of increasing contributions to Pareto constraints shifts the frontier regionally or globally. Transformations include the elimination of parametric-, variable-, or objective contributions, e.g.:

$$\begin{aligned} g(\mathbf{x}, \tilde{\epsilon}_i^+, \tilde{\epsilon}_j^-) &\rightarrow g^*(\mathbf{x}, \tilde{\epsilon}_j^-) \\ g(x_1^+, x_2^-, \tilde{\epsilon}_i^+) &\rightarrow g^*(x_2^-, \tilde{\epsilon}_i^+) \\ g(\mathbf{x}, \tilde{\epsilon}_i^+; P^+) &\rightarrow g^*(\mathbf{x}, \tilde{\epsilon}_i^+) \end{aligned}$$

### Typical design changes

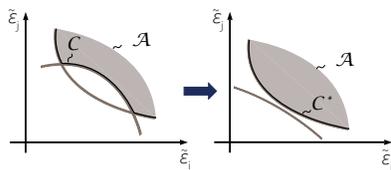
In many ways, the design changes and model transformations involved here, resemble those of *Leverage Harmonious variables*. Shifting Pareto constraints is tantamount to reducing the equilibrium that exists between the objectives in certain (or all) regions of the Pareto set, due to the activity of constraints. Thus, this might involve rearrangement of parts, change in load distribution, and so forth.

*Related heuristics:* Reduce information content [12], principle of balanced forces [3], vary the structure of elements [15], redirect load path [5]

### Example - Additive Manufacturing and Topology Optimization (TO)

In industrial practice, TO efforts are usually actively constrained by material and the manufacturing constraints. In this context, the utility of additive manufacturing is broadly cited, as it essentially shifts several manufacturing constraints, e.g. allowing hollow geometry and undercuts, and shaping not being limited by tooling directions. This has allows increasingly light load bearing components, reducing the trade-off between stiffness and mass. While this is more a process change than a design change, it serves to illustrate the model transformation.

## DEPENDENCY REDUCTION



**Mathematical transformation**  
 Reduction of the dependencies between Pareto constraints, that either result in trade-off variables, variables with empty feasible domains beyond the Pareto set (i.e. *two-sided failure*), or regional bounds for  $\tilde{\epsilon}$ , will change the optimal set. An example of such a transformation is:

$$\begin{aligned} g_1(x_1^+, \tilde{\epsilon}_i^-) &\rightarrow g_1(x_1^+, \tilde{\epsilon}_i^+) \\ g_2(x_1^-, \tilde{\epsilon}_j^+) &\rightarrow g_2^*(x_2^-, \tilde{\epsilon}_j^-) \end{aligned}$$

### Typical design changes

The design changes and model transformations involved here, resemble those of *Align trade-off variables*. The difference is that these might be regional trade-off variables. Hence, it is equally impactful to introduce changes to the eliminated active constraints that contribute to the Pareto constraint, creating the dependencies. Examples include inverting components and interfaces, eliminating load cases, change in working axis and load direction.

**Example: FlexTouch Safety Mechanism**  
 Insulin pens cannot be dialled to a dose setting beyond the amount of insulin left. An "end of content" locking mechanism prevents the user from receiving a smaller dose than has been set. Such locks need



to withstand substantial loads when users unknowingly attempt to get beyond this limit. Ultimately, this affects the achievable combination of device size and dose setting torque (which is important to users with limited dexterity). In the FlexTouch<sup>®</sup> device, the dial is connected to the dose setting mechanism via a flexible spline connection, which disengages if the user attempts to set a dose beyond what is left. No load is transferred, protecting the pen from overloading, and eliminating a dependency between the size and torque caused by the yield constraint.

# Eliminate Parasitic Contributions

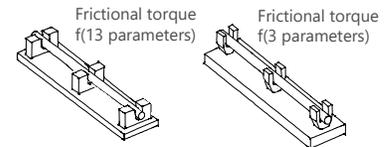
### Mathematical transformation

This category is too broad to provide a universal mathematical transformation, but all the sub-types of transformations are well known. These are the removal of parametric and scalar contributions which increase the optimal value of one or more objectives (without decreasing any), and the elimination of harmonious variables that are bound in such a way that they cannot be *leveraged*. These may be unintended contributions (i.e. from design error), or contributions involving active constraints with little room to introduce further relaxation (e.g. a wall thickness constraint).

### Typical design changes

There are many types of contributions that are parasitic. Examples the negative impact of undesired vibrations, electro-magnetic fields, heat, parasitic loads, friction, unintended contact points, and manufacturing and assembly features. Design changes mostly involve efforts to remove these effects from performance critical part geometries or locations in the assembly. E.g. relocating assembly features to another cross section.

*Related heuristics:* Exact constraint design [3], reduce information content [12], avoid associated loads [14], shielding [34].



### Example: Overconstrained axle [34]

An oft cited example in kinematic design, over-constrained axles cause major issues w.r.t. production and efficiency. The typical design error is to increase system stiffness by introducing more radial bearings, resulting in statical indeterminacy. This issue can be reduced or resolved entirely by using fewer or different bearings.

Fig. 6: The strategies within Principles 3 & 4: Reduce regional trade-offs and eliminate parasitic contributions

ble domains of harmonious variables in the improving direction, per Theorem 3. This involves design changes that modify or delete the constraints that bound harmonious variables, striving towards letting  $\bar{x} \rightarrow \infty \wedge \underline{x} \rightarrow 0$ .

3. **Relax Pareto Constraints.** Relax globally active Pareto constraints, thereby aligning trade-off variables. Relax regionally active Pareto constraints, or eliminate the inconsistencies that exist between them beyond the Pareto set (i.e., in the infeasible region). This might change or eliminate the Pareto frontiers between certain objectives.
4. **Eliminate Parasitic Contributors.** Consider situations where it is not possible to widen the feasible domain of harmonious and independent variables; e.g. when their bounds represent unmodelled objectives or physical phenomena that cannot be circumvented. Such situations can introduce parametric or scalar contributions to the objectives that worsen their optima. Therefore, it may be better to eliminate the influence of these variables on the objectives rather than relax the constraint.

These principles and associated strategies relate to specific variables and constraints. They can be applied recursively to improve a configuration beyond the identified optima. Recall that Pareto constraints are representations of trade-off variables. Furthermore, parasitic contributions are harmonious variables that are bound in a manner that prevents them from being leveraged to move the Pareto set. As such, Principles 3 and 4 are special cases of Principles 1 and 2 respectively.

The strategies are very general in that they apply to any design that has been studied through Pareto set dependency analysis. Their specificity becomes evident when considering a specific design problem. Just as design 'goodness' is contextual to the objectives at hand, so are the design improvements. Hence, the strategies are not intended for use in initial configuration design; rather, they motivate designers to identify improvements after careful analysis. Thus, the process of optimization becomes a driver for redesign.

As Figs.4-6 illustrate, the forms of design change involved are common in product design. The redesign strategies are related to well-known design heuristics, albeit with key differences. First, they are opportunistic but have a rigorous foundation and are hence valid independent of context. Second, they are applied following Pareto-set dependency analysis, letting designers rely on analysis results rather than intuition to identify which heuristic to apply where. Heuristics such as *separation* [13] and *independence* [12] or *division of tasks* [3] for instance, prescribe avoidance of dependency. As Section 3.1. shows, this is actually only relevant for trade-off variables when aiming to improve performance.

### 3.2.1 Sample Problem

In [28], we used a sample problem to demonstrate the application of MOMA to reveal hidden trade-off variables:

$$\min. \quad f_1(x_1^+, x_2^-, x_3^+) = x_1^2 - x_2 + x_3 \quad (14)$$

$$\text{s.j.t} \quad c_1(x_2^-, x_4^-, x_5^+; \epsilon_1^-) = \frac{1}{x_2} - x_4^2 + 2x_5 - \epsilon_1 \leq 0 \quad (15)$$

$$g_1(x_1^-, x_4^+) = 2x_4 - x_1 \leq 0 \quad (16)$$

$$g_2(x_2^+, x_3^-) = x_2^2 + 4x_2 - 2x_3 \leq 0 \quad (17)$$

$$g_3(x_2^+, x_4^+) = x_2^3 + 2x_4 - P_1 \leq 0 \quad (18)$$

$$g_4(x_5^-) = 10 - x_5^2 - 3x_5 \leq 0 \quad (19)$$

$$\epsilon_L \leq \epsilon_1 \leq \epsilon_U \quad (20)$$

$$\mathbf{x}, \epsilon_1 \in \mathbf{P} \quad (21)$$

This problem was reduced using MP1 and Theorem 1, revealing that all of the degrees of freedom are trade-off variables, due to  $g_1$ ,  $g_2$ , and  $g_4$  being critical. The resulting back-substitution of  $x_1^* = 2x_4$ ,  $x_3^* = \frac{1}{2}x_2^2 + 2x_2$ , and  $x_5^* = 2$ , yields:

$$\min. \quad f_1(x_2^+, x_4^+) = 4x_4^2 + \frac{1}{2}x_2^2 + x_2 \quad (22)$$

$$\text{s.j.t} \quad c_1(x_2^-, x_4^-; \epsilon_1^-) = \frac{1}{x_2} - x_4^2 + 4 - \epsilon_1 \leq 0 \quad (23)$$

$$g_3(x_2^+, x_4^+) = x_2^3 + 2x_4 - P_1 \leq 0 \quad (24)$$

$$\epsilon_L \leq \epsilon_1 \leq \epsilon_U \quad (25)$$

We can use this same problem to illustrate some of the underlying model transformations involved in the redesign principles. In this problem it is clear that the span of the Pareto set between will be defined by  $\bar{x}_2$ ,  $\bar{x}_4$   $g_3$  and  $\epsilon_L$ . While  $x_2$  and  $x_4$  are trade-off variables,  $x_1$ ,  $x_3$  and  $x_5$  are harmonious variables, albeit not shared between objectives. The activity of  $g_1$  makes  $x_4$  a trade-off variable, the activity of  $g_2$  makes  $x_2$  a trade-off variable, while the activity of  $g_4$  introduces a parasitic contribution of +4 to  $c_1$ . Thus, following the optimality principles, there are different routes improvement to be explored in a design change process:

$\underline{x}_4$  As it is caused by  $g_1$ , we could either try to apply scale, flip and separate principles to  $x_4$  in  $g_1$ , or to  $x_1$  in  $f_1$ .

$\underline{x}_2$  As it is caused by  $g_2$ , we could either try to apply the scale, flip and separate principles to  $x_2$  in  $g_2$ , or to  $x_3$  and  $x_2$  in  $f_1$ .

$\underline{x}_5$  We can either remove its influence on  $c_1$  or relax  $g_4$ .

There are many options available and, in a design context, some of these would be more practical than others. If we imagine that we were able to identify design changes that substitute  $x_2$  in  $g_2$  with a new variable,  $x_6$ , and substitute  $x_4$  in  $g_1$  with  $x_5$ , the resulting reduced problem becomes:

$$\min. \quad f_1(x_2^-) = 16 - x_2 + \frac{1}{2}x_6^2 + 2x_6 \quad (26)$$

$$\text{s.j.t} \quad c_1(x_2^-, x_4^-; \epsilon_1^-) = \frac{1}{x_2} - x_4^2 + 4 - \epsilon_1 \leq 0 \quad (27)$$

$$g_3(x_2^+, x_4^+) = x_2^3 + 2x_4 - P_1 \leq 0 \quad (28)$$

$$\epsilon_L \leq \epsilon_1 \leq \epsilon_U \quad (29)$$

With these changes, the problem is now poorly bounded, meaning a constraint bounding  $x_6$  needs to be introduced. When this is done, it is simple to evaluate whether the

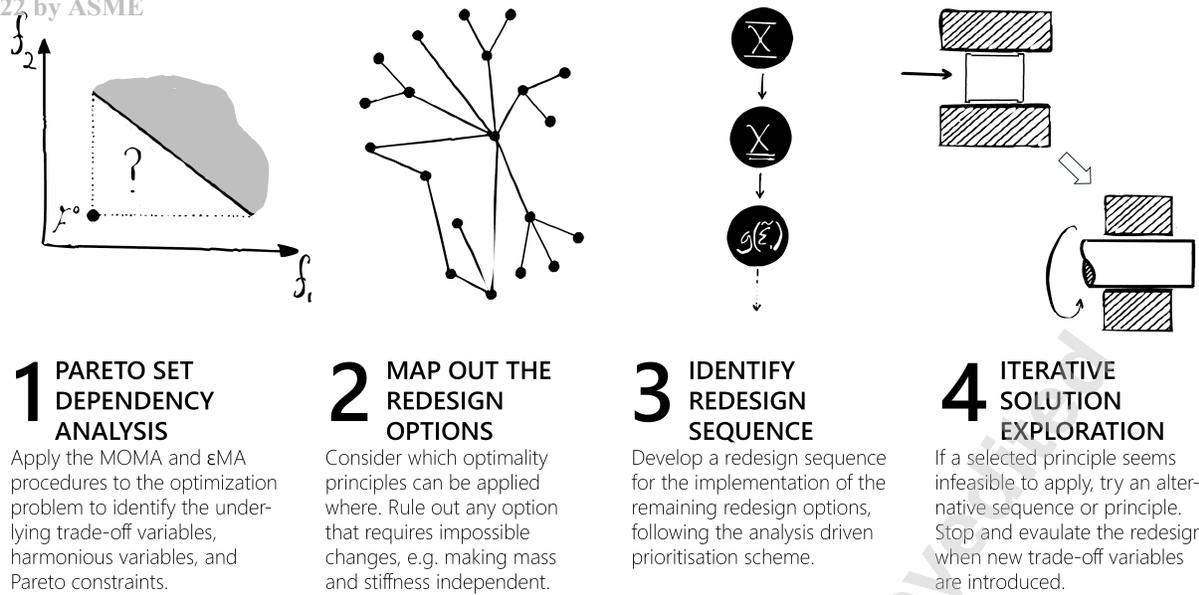


Fig. 7: Configuration Redesign Process - A redesign procedure supported by rigorous analysis

changes are an improvement by inspecting the bounds. So long as  $16 \leq 4x_4^2$ , and  $x_6 \leq x_2$ , not only is the optimum of the new design a dominant minimum rather than a set, it also dominates the Pareto set of the old design. This illustrates how we utilise the configuration redesign principles to steer the exploration of design changes and the output of the original MOMA to evaluate the effect of said changes, namely, the concurrent evaluation and exploration of design changes.

### 3.3 Configuration Redesign Process

The previous sections may seem excessively formal compared to many design frameworks. However, without the insights MOMA and  $\epsilon$ MA provide, one might introduce changes to eliminate dependencies or relax constraints that have no bearing on the Pareto set or even accidentally worsen the set. Ultimately, the above principles come down to a more targeted approach for dependency reduction and constraint relaxation. These are design practices that are already widely advocated [2, 3, 12, 13, 18, 26].

As alluded to, the strategies within the same principle are mutually exclusive. For example, we cannot make an objective function independent of a trade-off variable through separation while also scaling the same variable. Depending on the problem, some design changes are also more influential or easier to implement than others. Thus, it is beneficial to map out all options for improvement after analysis and select the most promising ones, rather than randomly applying the principles. While the strategies and underlying principles have a quantitative foundation, the designer must still determine which principle to apply to each variable and constraint, and in which sequence. As summarized in Fig. 7, we thus propose a systematic configuration redesign procedure involving said mapping and prioritization steps between

analysis and design change. A critical element in this is the use of a prioritization scheme in Step 3 to identify a redesign sequence. We suggest the following scheme, which is determined by two factors; the magnitude of the potential influence of the change and the ease of implementation:

1. Eliminate parasitic influences.
2. Leverage the harmonious variables, attempting *relaxation* rather than *shifted bounds* when possible. Only leverage the variables that:
  - influence multiple objectives,
  - have a multiplying effect on a trade-off variable.
  - are bound by a constraint with a comparatively high Lagrange multiplier.
  - are actively constrained in a manner that introduces a new trade-off variable or a contribution to a Pareto constraint
3. Relax Pareto constraints that depend on more than one  $\tilde{\epsilon}$  variable and/or are globally active
4. Align trade-off variables in an order based on the number of influenced objectives and on the relationship between  $F^*$  and  $\bar{\mathbf{x}}^*$ . To avoid increasing system complexity, apply flipped monotonicity over the other strategies, and separate over scale unless separation only is possible through the introduction of new variables.
5. Leverage remaining harmonious variables and relax remaining Pareto constraints

The underlying logic behind this scheme is to ensure that independent issues and design errors (i.e., parasitic contributions) are handled first, followed by the changes that result in the largest improvement to the Pareto set. The step order is defined based on the observation that the globally active Pareto constraints and the harmonious variables in Step 2 will, in most cases, exceed the influence of single trade-off variables. Alternatively, one could base the sequence on

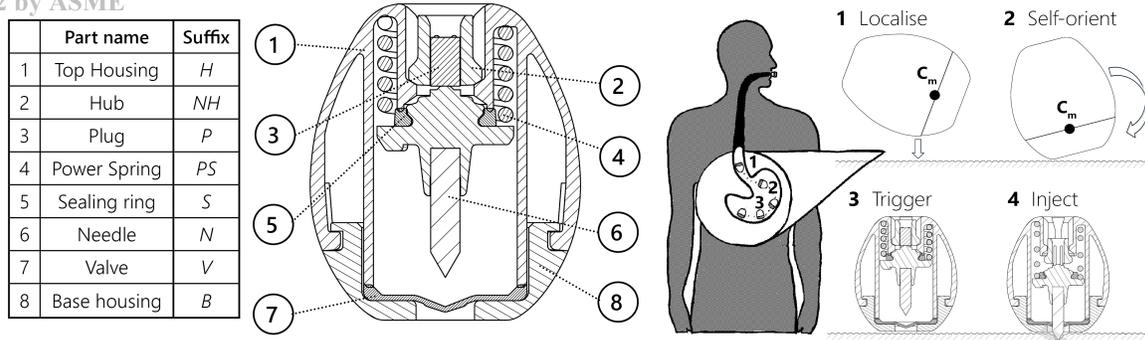


Fig. 8: An overview of the components and functionality of the SOMA device (adapted from [29])

objective weighting. The overall process is suggested to increase the likelihood of successful redesigns. It can be used iteratively and requires a combination of analysis, qualitative reasoning, engineering judgement, and creativity.

#### 4 Case - Design of the SOMA Device

The Self Orienting Millimeter-Scale Applicator (SOMA), is a medical device for oral delivery of pharmaceutical protein compounds such as insulin. Such compounds cannot otherwise be administered orally as the gastric system breaks down large proteins by its nature, and are thus administered using needle-based injection devices today. First described by Abramson et al. [29], SOMA was still in early development at the time of the present study. When swallowed, SOMA falls into the stomach, where it self-orientates into a stable position on the lining of the stomach thanks to its shape and mass distribution. Once oriented, a compression spring (4 in Fig. 8) injects a milipost of pure insulin (6) or another Active Pharmaceutical Ingredient (API), penetrating into a deep enough tissue layer to reach capillaries, resulting in systemic uptake as the milipost dissolves. This injection is triggered by a plug (3) which dissolves upon contact with liquid, allowing the compliant snap features on the hub component (2) to pass through a ratchet interface on the top housing (1).

The SOMA device presents a number of interesting design challenges. It must reliably deliver a large enough amount of API to meet the dosage needs of patients without compromising the self-orientation performance or injection depth, while being small enough to be swallowed without discomfort. Hence, SOMA was used in [28] to demonstrate the use of Pareto set dependency analysis, applying it to a 4-objective optimization model in upper-bound form:

$$\min f_1(\mathbf{x}) = -\frac{\sum_{p=1}^{p=8} m_p \cdot (C_p + Z_p)}{(l_{t1} + l_{t2} + l_{b1}) \cdot \sum_{p=1}^{p=8} m_p} \quad (30)$$

$$\text{s.t. } c_1(\mathbf{x}; \epsilon_1) = d_{t1} - \epsilon_1 \leq 0 \quad (31)$$

$$c_2(\mathbf{x}; \epsilon_2) = \epsilon_2 - \rho \frac{\pi}{4} d_{n1}^2 \left( l_{n1} + \frac{1}{3} \cdot l_{n2} \right) \leq 0 \quad (32)$$

$$c_3(\mathbf{x}; \epsilon_3) = \epsilon_3 - \sqrt{2 \left( g + \frac{F_s}{m_{acc}} \right) z_{acc}} \leq 0 \quad (33)$$

$$\mathbf{g}(\mathbf{x}) \leq 0 \quad (34)$$

$$\mathbf{h}(\mathbf{x}) = 0 \quad (35)$$

$$\mathbf{x}, \epsilon \in \mathbb{P} \quad (36)$$

where  $f_1$  is a self-orientation objective, maximising the normalized distance,  $Z_{cm}$ , between the top of the device and the system centre of mass,  $C_m$ . This contains intermediate functions;  $m_p$  describing the mass of each part in the device,  $C_p$  the centre of mass in each part, and  $Z_p$  the vertical position of each part.  $c_1$  is the bound size objective, minimizing  $d_{t1}$ , as pill swallowability is proportional with their minor diameter [36].  $c_2$  is the bound API capacity objective, maximizing the mass of the needle. Finally,  $c_3$  is the bound velocity objective, maximising the velocity of impact between needle and tissue. Here,  $F_s$  is a nonlinear expression for the accelerating force,  $m_{acc}$  the mass that is accelerated,  $z_{acc}$  the stroke between the initial position of the needle tip and the gastric tissue, and  $g$  is gravity. Constraints such as geometric fits, manufacturability, and structural load cases, are represented by  $\mathbf{g}(\mathbf{x})$ , while  $\mathbf{h}(\mathbf{x})$  mostly accounts for the shape of the device. In this paper, we demonstrate the application of the redesign methodology, using the analysis results presented in [28], which are described briefly for the sake of exposition.

#### 4.1 Results of Pareto-set Dependency Analysis

The optimization results [28] (shown in Fig. 11) revealed the trade-offs involved in the design of SOMA. The size-related trade-offs are of special importance, as the US FDA recommends that the minor diameters of capsules do not exceed 8.35mm to avoid medical complications [36]. While this is infeasible in the current design, it is possible to stay below  $\varnothing 9.91\text{mm}$ , the largest standard capsule size. However, this comes at the cost of self-orientation and impact velocity. Applying MOMA revealed several trade-off variables, see Table 1, such as the spring wire diameter ( $d_{ps2}$ ), needle length ( $l_{n1}$ ), position of the split between

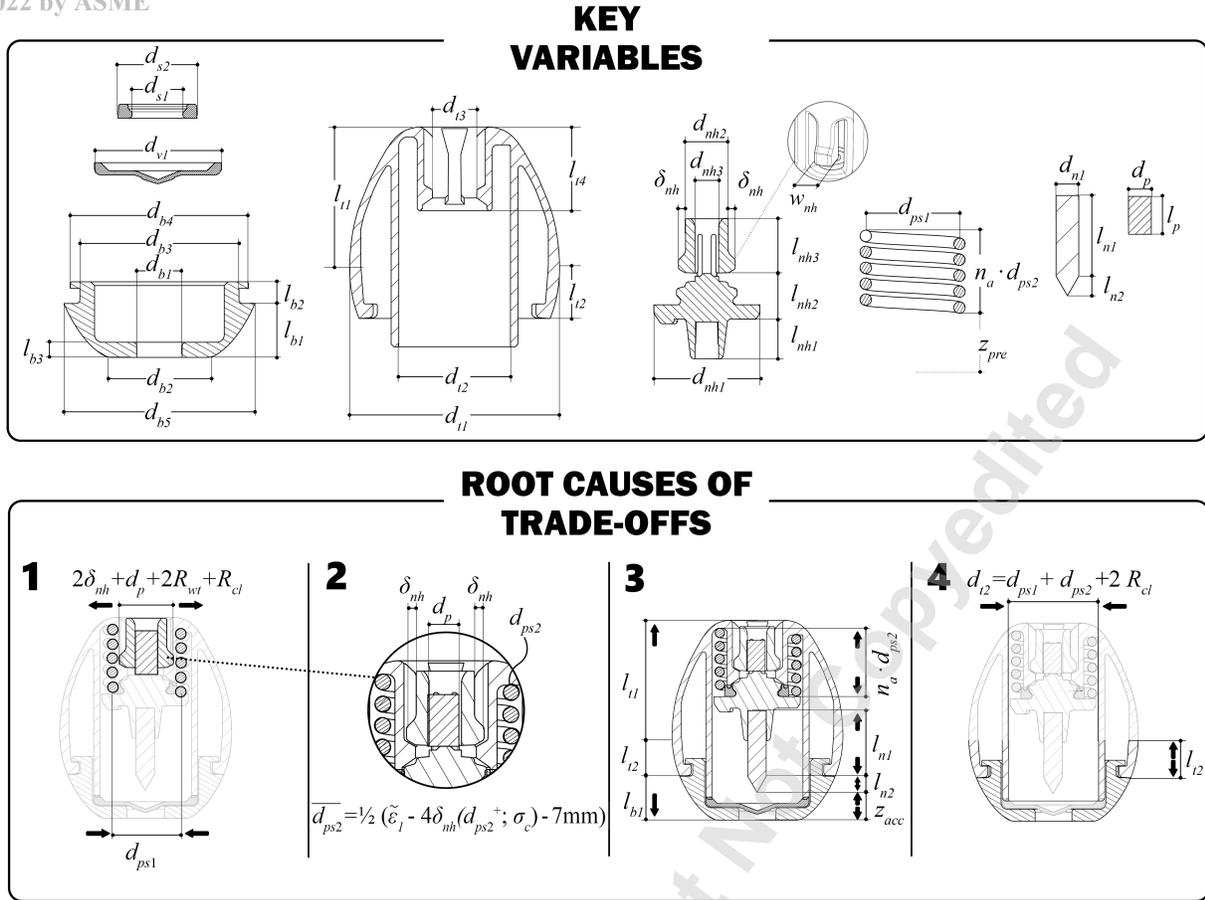


Fig. 9: An overview of the design variables in the SOMA and the key drivers of trade-off identified through analysis in [28]

top housing and base ( $l_{t2}$ ), and device diameter ( $d_{t1}$ ). Variables such as spring coiling diameter ( $d_{ps1}$ ), number of active windings ( $n_a$ ), and trigger arm overlap ( $\delta_{nh}$ ) were found to be harmonious. Combined with  $\epsilon$ MA post computation, the causes of the trade-offs (also shown in Fig. 9) become clear:

1. **Radial fit:** The glb of spring coiling diameter,  $d_{ps1}$  is determined by a constraint stemming from that the spring needs to fit around the trigger,  $g_5 = d_p + 2\delta_{nh} + d_{ps2} - d_{ps1} + 6R_{cl} + 4R_{wt} \leq 0$ . Here,  $R_{cl}$  and  $R_{wt}$  are radial clearance and wall thickness parameters respectively. Back-substituting this glb and the bound size objective,  $c_1 = d_{t1} - \epsilon_1 \leq 0$ , yields:

$$\min. \quad f_1(l_{t2}^-, d_{t3}^+, d_{ps2}^+, \delta_{nh}^+, d_p^+, \epsilon_1^-) \quad (37)$$

$$\text{s.j.t.} \quad c_3(l_{t2}^-, d_{ps2}^+, d_p^+, \delta_{nh}^+, \epsilon_1^-; \epsilon_3) \quad (38)$$

$$g_1(\epsilon_1^-, l_{t2}^+, d_{ps2}^+, d_p^+, \delta_{nh}^+) = 2d_{ps2} + 2\delta_{nh} + d_p$$

$$7\text{mm} - \sqrt{\frac{2(C_T \epsilon_1 - l_{t2}) \epsilon_1}{C_T} - \frac{(C_T \epsilon_1 - l_{t2})^2}{C_T^2}} \leq 0 \quad (39)$$

where  $\epsilon_1$  is the device diameter objective-variable and  $C_t$  is the device height-width ratio parameter. The radial fit constraint  $g_1$  which stems from the fits between the upper and lower housing, is now a Pareto constraint. When  $g_1$  is active,

any increase in spring wire diameter results in an increase in device size, a reduction in trigger overlap, or a reduction of  $l_{t2}$ . In [28],  $g_1$  was indeed found active at the 3-objective Pareto frontier between  $f_1$ ,  $c_1$ , and  $c_3$ , and violated in the region between the frontier and the utopia point. The constraint essentially prevents the device size from being minimized without reducing the space available for the spring and trigger, or shifting the housing snap upward to the widest point of the device meaning  $l_{t2} = \bar{l}_{t2} = 0$ , which in turn moves the centre of mass upwards, worsening self-orientation.

2. **Boundedness of the trigger:** Prior to computation, the harmonious variables  $d_p$  and  $\delta_{nh}$  were bound by a conditionally critical set of constraints. Looking at constraint activity at the bi-objective Pareto frontier between size and velocity, revealed a locally active glb of  $d_p$ ,  $g_{10} = 2\delta_{nh} - d_p - R_{cl} \leq 0$ , which prevents the trigger arms from colliding with each other thereby jamming the device. Further, a stress criterion for the trigger interface,  $g_{11}(d_{ps2}^+, \delta_{nh}^-)$ , is globally active and is now critical w.r.t.  $\delta_{nh}$ . As it has no closed-form solution, its implicit solution and back-substitution into Eq.39 yields

$$g_1(\epsilon_1^-, l_{t2}^+, d_{ps2}^+) = 2d_{ps2} + 4\delta_{nh}(d_{ps2}^+; \sigma_{IF}) + 7\text{mm}$$

$$- \sqrt{\frac{2(C_T \epsilon_1 - l_{t2}) \epsilon_1}{C_T} - \frac{(C_T \epsilon_1 - l_{t2})^2}{C_T^2}} \leq 0 \quad (40)$$

where  $\sigma_{IF}$  is the allowable interface stress in the trigger. Inserting this into the glb of the spring coiling diameter yields  $d_{ps1} = d_{ps2} + 4\delta_{nh}(d_{ps2}^+; \sigma_{IF}) + 6R_{cl} + 4R_{wt}$ . This implies that the activity of  $g_{10}$  and  $g_{11}$  multiplies the influence of  $d_{ps2}$  upon the trade-off of velocity against size and self-orientation, as any increase in  $d_{ps2}$  results in an increase in both  $d_p$  and  $\delta_{nh}$ , all contributing to size and mass. Further, any decrease in  $d_{ps1}$  also decreases the size of the load-bearing area in the trigger. These dependencies, specific to the Pareto-set, mean that the spring force can only be increased to a certain point for a given device size. Beyond this point, the device would fail due to high static interface stress or simply not fit together radially.

**3. Vertical fit of internal components:** The impact velocity is determined by the force profile exerted by the spring, system mass and frictional resistance, and the acceleration stroke distance between the tip and tissue ( $z_{acc}$ ). Given that the internal parts in the SOMA device are mounted in a vertical series,  $z_{acc}$  is involved in the following constraints:

$$\begin{aligned} h_8 &= R_{wt} + (n_a + n_d)d_{ps2} + l_{nh2} + l_{nh1} + l_{n1} + l_{n2} \\ &\quad + z_{acc} - l_{b1} - l_{t2} - l_{t1} = 0 \quad (41) \\ g_{20} &= l_{b3} + l_{v1} + P_{tol} - z_{acc} \leq 0 \end{aligned}$$

Through MOMA and computation, several constraints were found to be active, meaning that  $l_{nh1} = 1.5\text{mm}$ ,  $l_{nh2} = 5/2R_{wt}$ ,  $l_{t1} = C_t d_{t1}$ ,  $n_d = 1.5$ . Following the reductions introduced previously,  $n_a$  is bound from below by a spring yield limit criterion, meaning that  $n_a = n_a(d_{ps2}^+, z_{pre}^+; \sigma_{ps})$ , where  $\sigma_{ps}$  is the spring's yield limit. Solving  $h_8$  for  $z_{acc}$  and back-substituting these terms into the expression yields:

$$\begin{aligned} z_{acc} &= C_t d_{t1} + l_{t2} + l_{b1} - (n_a(d_{ps2}^+, z_{pre}^+; \sigma_{ps}) + 1.5)d_{ps2} \\ &\quad - l_{n1} - l_{n2} - 1.5\text{mm} - 7/2R_{wt} \quad (42) \end{aligned}$$

If one were to disregard the effect of constraints,  $z_{acc}$  might have seemed to be an independent variable to be used to optimize the impact velocity. However, these constraint activities have resulted in an expression that introduces trade-off variables into  $c_3$  upon back substitution, namely  $l_{n1}$  and  $l_{n2}$ , and increases the trade-off through  $d_{t1}$  and  $d_{ps2}$ . This also means that  $g_{20}$ , a locally active constraint that prevents the needle from protruding through the valve before injection, contributes to the trade-offs involving all four objectives.

**4. Assembly Features:** The housing snap ( $d_{b4} - d_{b3} = 2R_{ov} = 1.2\text{mm}$ ) and the cylinder on the top housing which seals the valve against the base, are assembly features that result in parasitic contributions that detrimentally affect the Pareto set. These contribute to the trade-offs between self orientation, velocity, and device diameter, given that they affect the relationship between the achievable device size, and the position of the housing split  $l_{t2}$  (through  $g_1$ ). Similarly, the needle attachment ( $l_{nh1}$ ) affects the set through the axial constraints ( $g_{20}$  and  $h_8$ ). Hence, any design change that eliminates these contributions would improve the Pareto-set.

	$d_{t1}$	$l_{t2}$	$l_{n1}$	$d_{n1}$	$d_{ps1}$	$d_{ps2}$	$n_a$	$d_p$	$\delta_{nh}$
$f_1$	-	-	+	+	+	+	+	+	+
$c_1$	+	(+)	(+)		(+)	(+)	(+)	(+)	(+)
$c_2$	(-)		-	-		(+)	(+)		
$c_3$	-		+	+	+	-	+	+	+

Table 1: The monotonicities of the objectives w.r.t. variables of interest, before  $\epsilon$ MA. Parentheses indicate a local dependency caused by constraint activity.

## 4.2 Redesign Mapping, Sequencing, and Exploration

Clearly, several dependencies cause trade-offs and worsen the Pareto set, most notably that the spring fits around the trigger. Any increase in spring force via  $d_{ps2}$  results in a larger trigger system to increase the load-bearing area, increasing the device's size. Further,  $d_{ps1}$ , an otherwise influential harmonious variable, affects the trade-offs given that reducing the coiling diameter reduces the space available for load-bearing geometry. This worsens the influence of  $d_{ps2}$  on size and self-orientation. As the optimal device diameter,  $\tilde{\epsilon}_1$ , is determined by a radial fit constraint, it also seems inopportune that the spring force is absorbed over an area in the radial direction. As such, a trade-off will always exist between velocity and size unless we find a more space-efficient way of distributing the load while making the  $d_{ps1}$  less dependant on the trigger design and vice versa.

The serial vertical arrangement of the internal components results in several trade-off variables; it also causes a trade-off between velocity and API payload, as the needle length  $l_{n1}$  cannot be increased at the optimum without reducing the acceleration stroke  $z_{acc}$  or the spring length (and thereby  $d_{ps2}$ ). Further,  $n_a(d_{ps2}^+, z_{pre}^+; \sigma_{ps})$  multiplies the negative influence of  $d_{ps2}$  on self orientation, as the spring mass is mounted at the top of the device. Finally, the parasitic contributions introduced by the assembly features in crucial cross-sections detrimentally affects the Pareto set. After mapping out the redesign options to solving these issues, we used the redesign sequence procedure along with the Lagrange multipliers and variable-objective plots to identify the following sequence of changes:

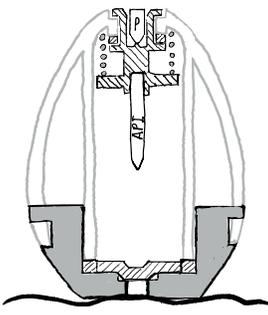
- 1. Eliminate parasitic contributions:** Reduce or eliminate the parasitic contributions of the housing snap and needle-hub interface upon the radial and axial fits, and upon the objective functions. Reduce the volume/mass of the plastic components when possible. Explore alternative linear guides for the needle hub and sealing principles for the valve component.
- 2. Shifted bounds:** Shift  $d_{ps1}$ , leveraging its harmonious influence, which is to the third power w.r.t. velocity.
- 3. Pareto constraint dependency reduction:** In some activity cases, the trigger interface stress becomes a Pareto constraint of the form  $g_{11}(\tilde{\epsilon}_1^-, \tilde{\epsilon}_3^+; \sigma_{IF})$ . Reduce the geometric dependency between the spring and trigger,

### 1 FLIPPED TRIGGER

**Design changes**  
**Trigger:** The ratchet arms have been inverted to work in tension rather than compression.  
**Assembly:** The interface between top and base housings has been changed to a female-male snap feature. The interface between needle and hub has also been changed from a shaft to a hole.

**Optimality Strategy**  
 $d_{ps1}^*$  - The plug diameter,  $d_p$ , and two wall thicknesses  $2R_{wt}$  have been removed from its GLB (*bound shift*).  
 $w_{rh}$  - With the trigger arms flipped, the LUB of their width is no longer determined by a mold tool constraint (*constraint relaxation*)  
 $g_1^*(\tilde{\epsilon}_1, \bar{d}_{ps2}^+, \bar{l}_{T2}^+, \delta_{rh}^+)$  - The contribution of  $d_p$ ,  $2R_{wt}$ ,  $2R_{ov}$  has been removed (*Shift Pareto constraint*)

**Impact**  
 $d_{ps2}$  and  $l_{t2}$  are some of the driving trade-off variables between self-orientation, impact velocity, and size. By relaxing the multiobjective Pareto constraint, and by shifting the bound of their shared harmonious variable,  $d_{ps1}$ , the optimum of all three is improved. However, the sealing ring diameter,  $d_{sr}$ , is now a part of the GLB of  $d_{ps1}$  and  $\epsilon_1$ .

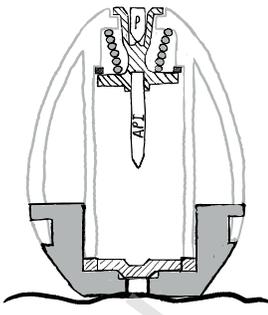


### 2 FLIPPED SEAL

**Design changes**  
**Sealing ring:** The layout of the top housing, sealing ring, and spring, has been changed. Instead of creating a seal against the top housing inside the spring, the sealing ring fits around the spring, allowing a stiffer conical spring.

**Optimality Strategy**  
 $d_{ps1}^*$  - The sealing ring diameter,  $d_{sr}$ , has been removed from the GLB (*bound shift*). The new conical spring coil has the smallest possible coiling diameter beyond placing the trigger arms around the spring, which would cause molding issues.

**Impact**  
 Combined with the changes made in the first iteration, this iteration has reduced the trade-off between velocity and size, and between velocity and self-orientation. The bounds on two important harmonious variables,  $d_{ps1}$  and  $Z_{acc}$  have been shifted. In doing so, two wall layers became redundant, meaning two Pareto constraints -  $g_1$  (radial fit) and  $g_{20}$  (axial fit) - are shifted by the removal of parametric contributions. The reduced trade-offs still exist however, especially due to the activity of the trigger interface load constraint, which increases the size of the trigger overlap, affecting device size as the spring force is increased.

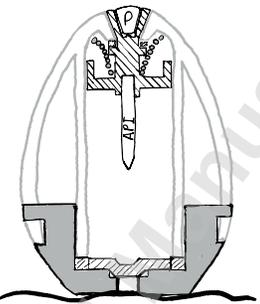


### 3 WEDGE TRIGGER

**Design changes**  
**Trigger redesign:** Building on the changes made in iteration 1, the ratchet based triggering mechanism has been replaced with a wedge design which is also loaded in tension. Hence the spring force is distributed over a larger surface than before.

**Optimality Strategy**  
 $g_1^*(\tilde{\epsilon}_1, \bar{d}_{ps2}^+, \bar{l}_{T2}^+)$  - The radial fit constraint is now independent of  $\delta_{rh}$ , which is critically constrained by the load constraint  $g_{42}(\bar{d}_{ps2}^+, \delta_{rh}^+)$  at velocity Pareto frontiers.  $g_{42}$  is in effect a Pareto constraint, as it depends on  $\bar{d}_{ps2}$ . Unless another radial-fit constraint actively bounds the outer diameter,  $d_{iv}$ , the spring force and the load bearing surface in the trigger can now be increased, without increasing the size of the device beyond the contribution of  $d_{ps2}$  (*Pareto constraint dependency reduction*)

**Impact**  
 After the 2<sup>nd</sup> iteration, the combination of size and velocity comes down to how much stress the loaded plastic components can withstand under long term static loading. Here, this issue has drastically been reduced, as the load is distributed over an additional dimension. In introducing a new degree of freedom affecting the size of this load bearing area, we can effectively decrease the size of the device further, without compromising the shelf life of the device.



### 4 FLIPPED ACTUATOR

**Design changes**  
**Actuation:** The spring has been replaced with a telescopic tension spring to allow the spring coil to pass through itself. As this spring is self-centering, the cylindrical guide can be removed, by using the valve to prevent tilt, ensuring that the needle exits the device.

**Optimality Strategy**  
 $d_{ps2}^*$  - The wire's influence on  $f_1$  is multiplied by  $n_a$  and  $d_{ps1}$  and its mounting height. Inverting the spring scales its negative influence and made it geometrically independent of the trigger. (*scale trade-off variable and bound shift*).  
 $Z_{acc}$  - This inversion also allows the elimination of the contribution of  $n_a \cdot d_{ps2}$  from  $h_g$ , which determined the achievable acceleration stroke,  $Z_{acc}$ . Hence, we have leveraged a harmonious variable and shifted  $g_{20}$ , which causes a trade-off between velocity and API payload. (*bound shift*)

**Impact**  
 This iteration addresses the trade-off between self orientation and velocity. Previously, we have primarily leveraged harmonious variables. Here, we aligned a trade-off variable, by shifting the spring mass downward. We have also radically shifted the upper limit of the acceleration stroke which is mass-less. Yet, this comes at the cost of a new dependency between the amount of spring material and the device diameter.

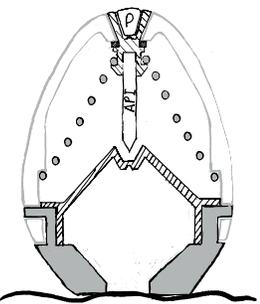
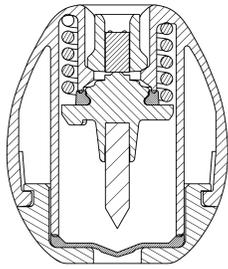


Fig. 10: Redesign iterations supported by the systematic application of the Principles of Optimality Improvement. Note that these are only principle sketches and do not reflect relative sizing.

### Original Design



### 2<sup>nd</sup> Redesign

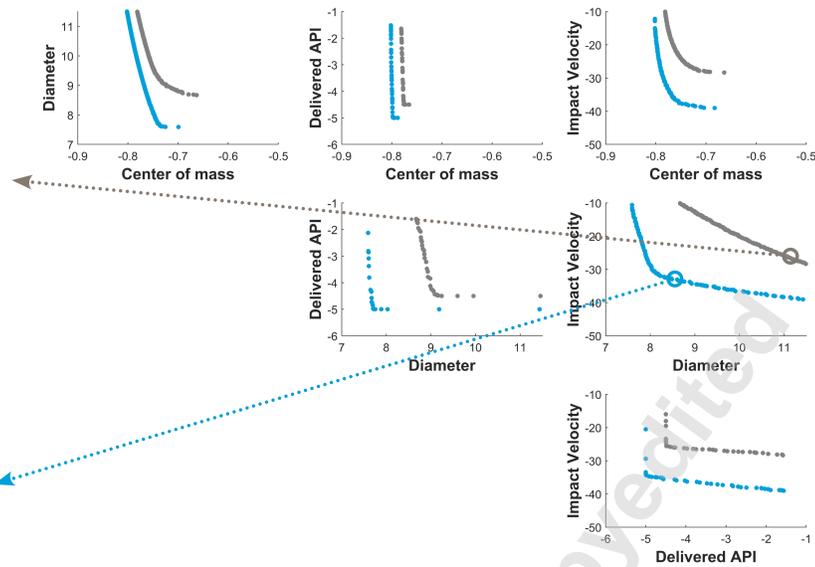
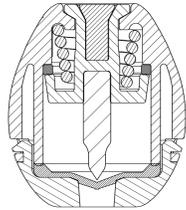


Fig. 11: A head-to-head comparison of the original configuration (grey) against the 2nd redesign (blue), the *Flipped Seal*. The 4D Pareto-set is shown with a 2D projection showing the bi-objective Pareto frontiers between each objective pair, which shows how the redesign is a clear improvement on all accounts. The relative size of the two designs is to scale.

making the radial fit constraint  $g_1(\tilde{\epsilon}_1^-, \overline{d_{ps2}^+}, \overline{l_{t2}^+})$ , and the  $g_{11}$  less interdependent. Attempt this by changing the working direction of the trigger interface to add additional degrees of freedom, resulting in a new interface stress criterion (which is currently globally active).

4. *Scale Trade-off Variable*: Reduce the influence of  $d_{ps2}$  upon size and self-orientation by moving the spring closer to the centre of mass - e.g. using a tension spring.
5. *Shifted bounds*: Eliminate the contributors to  $h_8$  that reduce  $\overline{z_{acc}}$ .
6. *Eliminate Parasitic Contributions*: Reduce the volume/mass of the plastic components when possible. Explore alternative linear guides for the needle hub and sealing principles for the valve component.

The application of this sequence of redesign principles led to a series of redesigns shown and explained in Figure 10. The design changes are relatively simple, and are essentially analogous to well known redesign heuristics, namely, *inversion* [3, 5, 13, 15] and *change in working direction/load path* [3, 5, 13], and *contributor reduction* [3, 12].

In the 1st iteration shown in Fig 9.1, the trigger arms are inverted to work in tension. The plug can now move upward, meaning the spring coil can in part be shaped independently of the plug as the trigger arms are flexible. This allows a slimmer, stiffer spring and an increased  $\overline{z_{acc}}$ , and eliminates a mold tool constraint which limits the achievable trigger arm width. The housing snap and needle-hub interfaces have also been changed from overlaps to holes. The 2nd iteration shifts the glb of  $d_{ps1}$  further by moving the seal outside the spring. A stiffer conical spring can thus be used. Now,  $\overline{\delta_{nh}}$  is determined by the outer shape of the device and the length of the trigger arms, as they can flex inside the spring during in-

jection. The 3rd iteration replaces the trigger with a conical wedge-like interface. This change in working direction adds a degree of freedom to the trigger interface design, changing the primary loading direction from radial to axial, allowing the spring to be stiffened without reducing the load-bearing area. In the 4th iteration, the spring is replaced by a tension spring, making the trigger and spring geometrically independent. This also increases  $\overline{z_{acc}}$  by eliminating the dependency between spring length and needle length, all the while reducing the impact of the spring's mass upon self-orientation. Interestingly, none of these redesign iterations involve drastic changes, such as additional components or a change in working principle. The use of analysis has guided the identification of simple changes that will substantially impact the optimization model and hence the Pareto set.

### 4.3 Redesign Evaluation

We built an optimization model to compare a redesign with the original SOMA as a form of validation of improvement according to Definition 3. For brevity, we limited the comparison to a single design selecting the 2<sup>nd</sup> redesign. It required relatively few modelling changes while still embodying influential changes of the configuration design.

The original optimization model was rebuilt with new constraint functions to reflect the new part fits, updated expressions for mass distribution to reflect the changes in geometry, and changes in spring equations to reflect the conical shape. The model structure, governing equations, and level of detail remained unchanged. This model was run with 200.000 iterations, with  $\epsilon_L = [7\text{mm}; 1.5\text{mg}; 10\text{m/s}]$  and  $\epsilon_U = [11.5\text{mm}; 5\text{mg}; 45\text{m/s}]$ . The results in Fig. 11 show the new Pareto set lying beyond the original one. For the union

of the Pareto sets,  $C_{U1} = C_0 \cup C_2$  the meta Pareto-set was found to only consist of solutions from the 2<sup>nd</sup> redesign, i.e.,  $\tilde{C} = C_2$ , and the single-objective optima of self-orientation has been improved by 2.63%, the size by 12.41 %, API payload by 11.11%, and velocity by 37.68%. We can thus conclude that the redesign is, in fact, a design improvement, as it meets the criteria in Definition 3. For the subsequent redesigns, it is likely that the achievable combination of impact velocity and self-orientation is improved even further, as the design changes are aimed at increasing the load-bearing area in the trigger system and shifting the centre of mass downward while increasing the acceleration stroke.

## 5 Discussion

With the methodology presented here, a degree of rigour is brought into the iterative design process, allowing the designer to utilize optimization to qualify the introduction of design changes. Thus, it is not singularly a configuration design or design optimization methodology - it is both. As seen with the SOMA device, the actual changes required to achieve an improved Pareto set can be relatively simple. Inversion, change of working direction, and changes to how the components fit together. Still, the impact on the Pareto set is substantial, as seen in the dominance of the *Flipped Seal* redesign over the original design.

One could argue that the presented methodology is a formalization of how experienced designers work, with their decisions largely being based on knowledge of trade-offs [13, 24] and active constraints [2, 26]. They often exhibit a degree of opportunism in identifying and solving issues by obviating dependencies that negatively affect performance and feasibility. This may allow them to synthesize designs that are easier to optimize, merging compatible functionality in certain subsystems/parts while separating functionality that is not. The proposed method shows its benefits by forcing designers to identify, understand, and mitigate the weaknesses of their configuration designs, potentially breaking fixation in the process. This might be especially useful for design tasks not met previously.

The methodology has its limitations. The most obvious one is that the analysis involved would seem onerous to most designers. Here a cost-benefit mindset comes in: if the benefit gained through redesign is accrued over a production volume counted in millions or billions (as is the potential with SOMA), then the cost of analysis becomes almost trivial. Another limitation is that the methodology's success depends entirely on whether all objectives and constraints of importance have been taken into account in the model. Therefore, the importance of a restrained approach to applying the redesign principles cannot be understated. If there is some tacit constraint or objective involved or one which the model simply does not consider, we must keep those in mind when introducing design changes.

The redesign procedure is opportunistic, as it is based on monotonicity analysis. Hence, it may not always be applicable. For designs with non-monotonic objective and constraint functions, the effort involved in understanding the

changes in monotonicity across the design space can be prohibitively time consuming or inconclusive. As noted already, cases such as SOMA may warrant the effort. Occasionally, one might perform the model reductions in MOMA and  $\epsilon$ MA using numerical data (i.e., post-optimality) instead of formal monotonicity analysis, as discussed in [28]. One might also handle more complex problems by applying the redesign principles on a different level of abstraction, e.g., at the functional architectural level, looking for redesign opportunities that redistribute functionality across the subsystems and parts. Alternatively, one might also use the methodology to explore a part of the system, e.g., to understand a trade-off issue between a pair of essential objectives.

When introducing design changes, situations might arise where new trade-off variables are introduced, such as in the flipped actuator redesign of SOMA. Here, the new spring design introduces new trade-off variables between impact velocity and device size. Certain changes might also be incompatible, resulting in trade-offs between design changes. In such situations, it would be necessary to quantify the influence of these changes.

Finally, when introducing configuration design changes iteratively, there will likely be a degree of path dependency. Despite it relying on the outputs of analysis and optimization, our sequencing approach is heuristic, and the final redesigns will depend on the early iterations. This can be overcome in part by exploring solutions before implementing them, but this still does not guarantee compatibility. A pragmatic approach of letting the relative importance of the objectives affect the sequence can be worthwhile. For example, if there were no swallowable designs in the Pareto set of the original SOMA, then it would have made little sense to start mitigating trade-offs between self-orientation and impact velocity.

## 6 Conclusion

The question of systematic configuration design is a challenge for design and optimization research. In practice, it is mainly driven by the skill and experience of the designer rather than the application of a clear design theory. In this contribution, we have expanded upon previous work on multiobjective monotonicity analysis to demonstrate its application in configuration redesign. The result is a rigorously founded methodology that enables the designer to identify design changes to improve on all objectives and reduce or eliminate trade-offs represented in the Pareto set. When applied systematically, the result is performance beyond what is achievable through proportional or parametric optimization alone. We demonstrated this capability the SOMA device. A new optimization model comparing the 2<sup>nd</sup> redesign iteration with the original design was built to show the method's validity. This revealed a substantial size reduction and increase in impact velocity without worsening self-orientation or API payload. Such analysis and redesign methodology empowers designers to explore better configuration designs systematically.

## Acknowledgements

The authors would like to thank the support of the Danish Innovations Fund and the Novo Nordisk STAR-programme (grant no. 7038-00221B), and the University of Michigan Donald C. Graham Endowment. We are grateful to Novo Nordisk for sharing design information and data, to Chris McMahon of the University of Bristol, for his advice and input, and Giovanni Traverso and his colleagues at MIT for their helpful comments and input. The opinions presented here are solely those of the authors.

## References

- [1] Design Council. *Eleven lessons: managing design in eleven global companies*. Tech. rep. 272099. 2007.
- [2] McMahon, C. A. “Observations on Modes of Incremental Change in Design”. In: 5.3 (1994). DOI: 10.1080/09544829408907883.
- [3] Pahl, G. and Beitz, W. *Engineering design — A systematic approach*. 1999. DOI: 10.1016/0261-3069(96)84970-3.
- [4] Ullman, D. G., Dietterich, T. G., and Stauffer, L. A. “A model of the mechanical design process based on empirical data”. In: *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing* 2.1 (1988), pp. 33–52. DOI: 10.1017/S0890060400000536.
- [5] French, M. J. *Conceptual Design for Engineers*. 1985. DOI: 10.1007/978-3-662-11364-6.
- [6] Bendsøe, M. P. and Kikuchi, N. “Generating optimal topologies in structural design using a homogenization method”. In: *Computer Methods in Applied Mechanics and Engineering* (1988). DOI: 10.1016/0045-7825(88)90086-2.
- [7] Schmidt, L. C. and Cagan, J. “Optimal Configuration Design: An Integrated Approach Using Grammars”. In: *Journal of Mechanical Design* 120.1 (Mar. 1998), pp. 2–9. DOI: 10.1115/1.2826672.
- [8] Bayrak, A. E., Kang, N., and Papalambros, P. Y. “Decomposition-Based Design Optimization of Hybrid Electric Powertrain Architectures: Simultaneous Configuration and Sizing Design”. In: *Journal of Mechanical Design, Transactions of the ASME* 138.7 (2016), pp. 1–9. DOI: 10.1115/1.4033655.
- [9] Bayrak, A. E., Ren, Y., and Papalambros, P. Y. “Topology Generation for Hybrid Electric Vehicle Architecture Design”. In: *Journal of Mechanical Design, Transactions of the ASME* 138.8 (2016). DOI: 10.1115/1.4033656.
- [10] Antonsson, E. K. and Cagan, J., eds. *Formal Engineering Design Synthesis*. Cambridge University Press, Nov. 2001. DOI: 10.1017/CBO9780511529627.
- [11] Chakrabarti, A. et al. “Computer-based design synthesis research: An overview”. In: *Journal of Computing and Information Science in Engineering* 11.2 (2011). DOI: 10.1115/1.3593409.
- [12] Suh, N. P. “Axiomatic Design Theory for Systems”. In: *Research in Engineering Design - Theory, Applications, and Concurrent Engineering* 10.4 (1998), pp. 189–209. DOI: 10.1007/s001639870001.
- [13] Altshüller, G. *Creativity As an Exact Science*. Gordon and Breach, 1984. DOI: 10.1201/9781466593442.
- [14] Matthiassen, B. “Design for Robustness and Reliability - Improving the Quality Consciousness in Engineering Design”. PhD thesis. Technical University of Denmark, 1997.
- [15] Tjalve, E. “Form Design - a Systematic Approach.” In: *Schriftenreihe WDK (Workshop Design - Konstruktion)*. 1981, pp. 559–571.
- [16] Boothroyd, G., Dewhurst, P., and Knight, W. A. *Product Design for Manufacture and Assembly*. 2010. DOI: 10.1201/9781420089288.
- [17] Arthur, W. B. “Why Do Things Become More Complex?” In: *Scientific American* 268.5 (1993), pp. 144–144. DOI: 10.1038/scientificamerican0593-144.
- [18] Wynn, D. C. and Eckert, C. M. *Perspectives on iteration in design and development*. Vol. 28. 2. Springer London, 2017, pp. 153–184. DOI: 10.1007/s00163-016-0226-3.
- [19] Papalambros, P. Y. and Wilde, D. J. *Principles of Optimal Design*. Cambridge University Press, Jan. 2017. DOI: 10.1017/9781316451038.
- [20] Sobek, D. K., Ward, A. C., and Liker, J. K. “Toyota’s Principles of Set-Based Concurrent Engineering”. In: *Sloan Management Review* 40.2 (1999), pp. 67–83.
- [21] Cagan, J. and Agogino, A. M. *Innovative design of mechanical structures from first principles*. 1987. DOI: 10.1017/S0890060400000275.
- [22] Papalambros, P. and Wilde, D. J. “Global Non-iterative Design Optimization Using Monotonicity Analysis”. In: *Journal of Mechanical Design, Transactions of the ASME* 78 -WA/DE-17 (1978).
- [23] Jain, P. and Agogino, A. M. “Theory of design: An optimization perspective”. In: *Mechanism and Machine Theory* 25.3 (1990), pp. 287–303. DOI: 10.1016/0094-114X(90)90030-N.
- [24] Ahmed, S., Wallace, K. M., and Blessing, L. T. “Understanding the differences between how novice and experienced designers approach design tasks”. In: *Research in Engineering Design* 14.1 (2003), pp. 1–11. DOI: 10.1007/s00163-002-0023-z.
- [25] Cross, N. “Expertise in design: An overview”. In: *Design Studies* 25.5 (2004), pp. 427–441. DOI: 10.1016/j.destud.2004.06.002.
- [26] Eckert, C. M. and Stacey, M. K. “Constraints and Conditions: Drivers for Design Processes”. In: *An Anthology of Theories and Models of Design: Philosophy, Approaches and Empirical Explorations*. Ed. by Chakrabarti, A. and Blessing, L. T. M. London: Springer London, 2014, pp. 395–415. DOI: 10.1007/978-1-4471-6338-1\_{\\_}19.

- [27] Onarheim, B. “Creativity from constraints in engineering design: Lessons learned at Coloplast”. In: *Journal of Engineering Design* 23.4 (2012), pp. 323–336. DOI: 10.1080/09544828.2011.631904.
- [28] Sigurdarson, N. S., Eifler, T., Ebro, M., and Papalambros, P. Y. “Multiobjective Monotonicity Analysis: Pareto Set Dependency and Tradeoffs Causality in Configuration Design”. In: *Journal of Mechanical Design* (2021), pp. 1–18. DOI: 10.1115/1.4052444.
- [29] Abramson, A. et al. “An ingestible self-orienting system for oral delivery of macromolecules”. In: *Science* 363.6427 (2019). DOI: 10.1126/science.aau2277.
- [30] Marler, R. T. and Arora, J. S. “Survey of multi-objective optimization methods for engineering”. In: *Structural and Multidisciplinary Optimization* 26.6 (2004), pp. 369–395. DOI: 10.1007/s00158-003-0368-6.
- [31] Carmichael, D. “Computation of Pareto Optima in Structural Design”. In: *International Journal for Numerical Methods in Engineering* 15 (1980), pp. 925–952. DOI: 10.1017/S0022029900029393.
- [32] Athan, T. W. and Papalambros, P. Y. “A quasi-Monte Carlo method for multicriteria design optimization”. In: *Engineering Optimization* 27.3 (1996), pp. 177–198. DOI: 10.1080/03052159608941405.
- [33] Jugulum, R. and Frey, D. D. “Toward a taxonomy of concept designs for improved robustness”. In: *Journal of Engineering Design* 18.2 (2007), pp. 139–156. DOI: 10.1080/09544820600731496.
- [34] Goto, T., Isobe, R., Yamakawa, M., and Nishida, M. “The New Mazda Gasoline Engine Skyactiv-G”. In: *MTZ worldwide eMagazine* 72.6 (2011), pp. 40–47. DOI: 10.1365/s38313-011-0063-8.
- [35] Ebro, M. and Howard, T. J. “Robust design principles for reducing variation in functional performance”. In: *Journal of Engineering Design* 27.1-3 (2016), pp. 75–92. DOI: 10.1080/09544828.2015.1103844.
- [36] U.S. Department of Health and Human Services Food and Drug Administration (CDER). “Guidance for Industry: Size, Shape and Other Physical Attributes of Generic Tablets and Capsules”. In: *Pharmaceutical Quality/CMC* December (2013), pp. 1–11.