1 Introduction

Multidisciplinary design optimization (MDO): a field of engineering that uses numerical optimization to perform the design of systems that involve a number of disciplines or subsystems.

- the best design of a multidisciplinary system can only be found when the interactions between the system’s disciplines are fully considered.

- Considering these interactions in the design process cannot be done in an arbitrary way

- requires a sound mathematical formulation.

By solving the MDO problem early in the design process and taking advantage of advanced computational analysis tools, designers can simultaneously improve the design and reduce the time and cost of the design cycle.
The origins of MDO can be traced back to Schmit [131, 132, 133] and Haftka [55, 56, 58], who extended their experience in structural optimization to include other disciplines.

One of the first applications of MDO was aircraft wing design, where aerodynamics, structures, and controls are three strongly coupled disciplines [51, 52, 100, 99].

Since then, the application of MDO has been extended:

- complete aircraft [88, 102], rotorcraft [45, 49], and spacecraft [23, 27].
- bridges [9] and buildings [32, 47];
- railway cars [64, 42], automobiles [109, 85], and ships [123, 70];
- and even microscopes [124].
Important considerations when implementing MDO:

- How to organize the disciplinary analysis models?
- What optimization software/methods do we choose?
- Do we use approximation (surrogate) models?

**MDO architecture:** Combination of problem formulation and organizational strategy. The MDO architecture defines both how the different models are coupled and how the overall optimization problem is solved.

There are several terms in literature used to describe what we mean by “architecture” [22, 87, 138, 30]:

- “method” [88, 144, 158, 128, 1],
- “methodology” [77, 114, 112],
- “problem formulation” [35, 4, 5],
- “strategy” [161, 59],
- “procedure” [145, 81] and
- “algorithm” [135, 143, 137, 41]
Our preference is for the term “architecture”, because the relationship between problem formulation and solution algorithm is not one-to-one. For example, replacing a particular disciplinary simulation with a surrogate model or reordering the disciplinary simulations do not affect the problem formulation but strongly affect the solution algorithm.

Choosing the most appropriate architecture for the problem can significantly reduce the solution time. These time savings come from

- the methods chosen to solve each discipline;
- the optimization algorithm driving the process;
- the coupling scheme used in the architecture, and;
- the degree to which operations are carried out in parallel.

The latter consideration becomes especially important as the design becomes more detailed and the number of variables and/or constraints increases.
The purpose of this chapter:

- survey the available MDO architectures
- present architectures in a unified notation to facilitate understanding and comparison
- introduce the Extended Design Structure Matrix (XDSM), a diagram to visualize the algorithm of a given MDO architecture
2 Unified Description of MDO Architectures

2.1 Terminology and Mathematical Notation

design variable: a variable in the MDO problem that is always under the explicit control of an optimizer.

local design variable: design variables relevant to a single discipline only → denoted by $x_i$ for discipline $i$

shared design variable: design variables used by several disciplines → denoted by $x_0$

Full set of design variables:

$$x = [x_0^T, x_1^T, \ldots, x_N^T]^T$$

Think of some examples of local and shared design variables in the context of aerostructural optimization
**discipline analysis:** a simulation that models the behavior of one aspect of a multidisciplinary system
  → represented by equations $\mathcal{R}_i = 0$,

**state variables:** set of variables determined by solving a discipline analysis.
  → denoted by $\bar{y}_i$

Give some examples of aerodynamic discipline analyses. What are the corresponding state variables?
**coupling variables**: variables determined by one discipline and that influence another discipline

**response variables**: coupling variables from a specific discipline  
→ denoted by $y_i$ for discipline $i$

**target variables**: values of the response variables that we need to match  
→ denoted by $y^t_i$ for discipline $i$

**consistency constraints**: constraints that ensure the response variables match the values of the target variables  
→ for discipline $i$ we have $c^c_i = y^t_i - y_i$

Choose a multidisciplinary problem. Identify the coupling variables.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>Vector of design variables</td>
</tr>
<tr>
<td>$y^t$</td>
<td>Vector of coupling variable targets (inputs to a discipline analysis)</td>
</tr>
<tr>
<td>$y$</td>
<td>Vector of coupling variable responses (outputs from a discipline analysis)</td>
</tr>
<tr>
<td>$ar{y}$</td>
<td>Vector of state variables (variables used inside only one discipline analysis)</td>
</tr>
<tr>
<td>$f$</td>
<td>Objective function</td>
</tr>
<tr>
<td>$c$</td>
<td>Vector of design constraints</td>
</tr>
<tr>
<td>$c^c$</td>
<td>Vector of consistency constraints</td>
</tr>
<tr>
<td>$\mathcal{R}$</td>
<td>Governing equations of a discipline analysis in residual form</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of disciplines</td>
</tr>
<tr>
<td>$n(\cdot)$</td>
<td>Length of given variable vector</td>
</tr>
<tr>
<td>$m(\cdot)$</td>
<td>Length of given constraint vector</td>
</tr>
<tr>
<td>$(\cdot)_0$</td>
<td>Functions or variables that are shared by more than one discipline</td>
</tr>
<tr>
<td>$(\cdot)_i$</td>
<td>Functions or variables that apply only to discipline $i$</td>
</tr>
<tr>
<td>$(\cdot)^*$</td>
<td>Functions or variables at their optimal value</td>
</tr>
<tr>
<td>$(\cdot)$</td>
<td>Approximation of a given function or vector of functions</td>
</tr>
</tbody>
</table>
2.2 Architecture Diagrams — The Extended Design Structure Matrix

An Extended Design Structure Matrix, or XDSM [89], is a convenient and compact way to describe the sequence of operations in an MDO architecture. The XDSM was developed to simultaneously communicate data dependency and process flow between computational components of the architecture on a single diagram.

The XDSM is based on Design Structure Matrix [146, 25] and follows its basic rules:

- architecture components are placed on main diagonal of the “matrix”
- inputs to a component are placed in the same column
- outputs to a component are placed in the same row
- External inputs and outputs may also be defined and are placed on the outer edges of the diagram
• Thick gray lines are used to show the data flow between components

• A numbering system is used to show the order in which the components are executed (The algorithm starts at component zero and proceeds in numerical order)

• Consecutive components in the algorithm are connected by a thin black line

• Loops are denoted using the notation $j \rightarrow k$ for $k < j$ so that the algorithm must return to step $k$ until a looping condition is satisfied before proceeding
Example 1: Gauss–Seidel MDA

Algorithm 1 Block Gauss–Seidel multidisciplinary analysis algorithm

Input: Design variables $x$

Output: Coupling variables, $y$

0: Initiate MDA iteration loop

repeat

1: Evaluate Analysis 1 and update $y_1$
2: Evaluate Analysis 2 and update $y_2$
3: Evaluate Analysis 3 and update $y_3$

until $4 \rightarrow 1$: MDA has converged
Figure 1: A block Gauss–Seidel multidisciplinary analysis (MDA) process to solve a three-discipline coupled system.
Example 2: Gradient-based Optimization

- objective, constraints, and gradient evaluations are the (3) individual components
Example 2: Gradient-based Optimization

- objective, constraints, and gradient evaluations are the (3) individual components

Figure 2: A gradient-based optimization procedure.
2.3 The All-At-Once (AAO) Problem Statement

Consider the general “all-at-once” (AAO) MDO problem statement:

\[
\begin{align*}
\text{minimize} & \quad f_0(x, y) + \sum_{i=1}^{N} f_i(x_0, x_i, y_i) \\
\text{with respect to} & \quad x, y^t, y, \bar{y} \\
\text{subject to} & \quad c_0(x, y) \geq 0 \\
& \quad c_i(x_0, x_i, y_i) \geq 0 \quad \text{for } i = 1, \ldots, N \\
& \quad c_i^c = y_i^t - y_i = 0 \quad \text{for } i = 1, \ldots, N \\
& \quad R_i(x_0, x_i, y_{j\neq i}^t, \bar{y}_i, y_i) = 0 \quad \text{for } i = 1, \ldots, N.
\end{align*}
\]

- Typically omit the local objective functions \( f_i \) unless necessary

- Some authors refer to AAO as simultaneous analysis and design

- AAO is rarely solved in practice: we can eliminate the \( c_i^c \).
“All-at-once” (AAO) problem

Figure 3: XDSM for solving the AAO problem.
3 Monolithic Architectures

AAO provides a unifying starting point for other architectures; depending on which constraints are eliminated, we can derive different monolithic architectures.

**Monolithic Architectures:** an architecture that solves the MDO problem as a single optimization problem.

We will consider three monolithic architectures:

1. Simultaneous Analysis and Design (SAND);
2. Individual Discipline Feasible (IDF), and;
3. Multidisciplinary Feasible (MDF).

The monolithic architectures are distinguished by how they achieve *multidisciplinary feasibility*. 
3.1 Simultaneous Analysis and Design (SAND)

**Idea:** use a single copy of the coupling variables that is shared between disciplines to eliminate the consistency constraints. This simplification yields the SAND architecture [57]:

\[
\begin{align*}
\text{minimize} & \quad f_0(x, y) \\
\text{with respect to} & \quad x, y, \bar{y} \\
\text{subject to} & \quad c_0(x, y) \geq 0 \\
& \quad c_i(x_0, x_i, y_i) \geq 0 \quad \text{for } i = 1, \ldots, N \\
& \quad R_i(x_0, x_i, y, \bar{y}_i) = 0 \quad \text{for } i = 1, \ldots, N.
\end{align*}
\]
Simultaneous Analysis and Design

Figure 4: Diagram for the SAND architecture.
Several features of the SAND architecture are noteworthy:

- The optimizer is responsible for simultaneously analyzing and designing the system

- At each iteration, the analyses do not need to be solved exactly: $R_i \neq 0$
  $\rightarrow$ inexact solution to analyses: potential to solve optimization quickly
  $\rightarrow$ 5-10 times a single MDA

- Can also be used to solve single discipline optimization problem

- Full-space PDE-constrained optimization is an example

**Question:** Can you think of some disadvantages of SAND?
Disadvantages of the SAND approach:

1. Large problem size: all state variables must be available to the optimizer → existing optimization software cannot handle typical CFD problem

2. Globalization is an issue: converging problem when initial design/state is far from final solution → globalizing analysis codes is a formidable challenge on its own

3. Residual values and (likely) their derivatives must be available to optimization software → Many discipline analysis codes operate like “black-boxes”

4. Quasi-Newton methods may be slow, because of large problem size → need to consider some form of Newton’s method; possible but complicated
3.2 Individual Discipline Feasible (IDF)

**Idea:** Solve the discipline analyses exactly for given $x$ and $y^t$

- eliminates the disciplinary analysis constraints $\mathcal{R}_i(x_0, x_i, y_i, y^t_{j\neq i}, \bar{y}_i) = 0$ from the optimization problem

- invokes the implicit function theorem on the $\mathcal{R}_i$, hence

$$\bar{y}_i = \bar{y}_i(x, y^t)$$

$$y_i = y_i(x, y^t)$$

- IDF also called distributed analysis optimization [4] and optimizer-based decomposition [87]
The problem statement for the IDF architecture is

minimize \( f_0 (x, y (x, y^t)) \)

with respect to \( x, y^t \)

subject to \( c_0 (x, y (x, y^t)) \geq 0 \)

\( c_i (x_0, x_i, y_i (x_0, x_i, y_j^t)) \geq 0 \) \( \text{for } i = 1, \ldots, N \)

\( c_i^c = y_i^t - y_i (x_0, x_i, y_{j \neq i}^t) = 0 \) \( \text{for } i = 1, \ldots, N. \)
Individual Discipline Feasible

Figure 5: Diagram of the IDF architecture.
Some advantages of IDF include the following:

- all state variables and discipline analysis equations are removed from optimization
- discipline analyses can be performed in parallel
- analysis codes are already highly specialized and efficient at solving their respective equations → for example, globalization is simplified

The increased flexibility of IDF comes at a price. List some of the drawbacks of this architecture?
Disadvantages of IDF:

1. if optimization terminates before satisfying the KKT conditions (for example), the intermediate solution is not multidisciplinary feasible → in some sense, this is no better than SAND

2. number of coupling variables may still be large → typically $10^3 - 10^5$ for aerostructural problems

3. gradient computation, if necessary, is expensive → Adjoint-based gradients are critical for efficiency, but requires intrusion into source code
3.3 Multidisciplinary Feasible (MDF)

Idea: Solve the multidisciplinary analysis exactly for given \( x \).

- eliminate both the disciplinary analyses and consistency constraints from optimization problem

- also called Fully Integrated Optimization [4] and Nested Analysis and Design [10]

MDF problem statement:

\[
\begin{align*}
\text{minimize} & \quad f_0(x, y(x, y)) \\
\text{with respect to} & \quad x \\
\text{subject to} & \quad c_0(x, y(x, y)) \geq 0 \\
& \quad c_i(x_0, x_i, y_i(x_0, x_i, y_{j \neq i})) \geq 0 \text{ for } i = 1, \ldots, N.
\end{align*}
\]
Figure 6: Diagram for the MDF architecture with a Gauss–Seidel multidisciplinary analysis.
Advantageous features of MDF:

- optimization algorithm is responsible for the design variables, objective, and design constraints only
  $\rightarrow$ optimization problem is as small as possible

- design is always feasible in a multidisciplinary sense

The MDA can be solved in various ways, but efficiency of optimization is tied to this choice:

- simple, popular choice is a fixed-point iteration like Gauss-Seidel

- Newton-Krylov approach is much more efficient, but requires intrusion

What about disadvantages of MDF?
Drawbacks of using MDF:

1. architecture requires full MDA for each (optimization) iteration
   MDA is, on its own, a challenging task
   → may not be an issue if only one discipline dominates CPU time

2. gradient computation is complex
   → for efficiency, need a coupled adjoint for whole MDA

3. For premature iteration state variables are feasible, but design may not be
   → dependent on choice of optimization algorithm
4 Distributed Architectures

Consider the following problem:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{N} f_i(x_i) \\
\text{with respect to} & \quad x_1, \ldots, x_N \\
\text{subject to} & \quad c_0(x_1, \ldots, x_N) \leq 0 \\
& \quad c_1(x_1) \leq 0, \ldots, c_N(x_N) \leq 0.
\end{align*}
\]

• there are no shared design variables, \( x_0 \)

• the objective function is separable: it can be expressed as a sum of functions, each of which depends only on corresponding local design variables

• constraints depend on more than one set of design variables

**Question:** if \( c_0 \) did not exist, how could we solve this problem?
The previous problem is referred to as a *complicating constraints problem* [33].

Another possibility:

\[
\text{minimize } \sum_{i=1}^{N} f_i(x_0, x_i)
\]

with respect to \( x_0, x_1, \ldots, x_N \)

subject to \( c_1(x_0, x_1) \leq 0, \ldots, c_N(x_0, x_N) \leq 0. \)

This is referred to as a problem with *complicating variables* [33].

- Decomposition would be straightforward if there were no shared design variables, \( x_0 \), and we could solve \( N \) optimization problems independently and in parallel.
Distributed Architectures: an architecture that solves the MDO problem using a set of optimization problems or subproblems.

The primary motivation for decomposing the MDO problem comes from the structure of the engineering design environment.

- Typical industrial practice involves breaking up the design of a large system and distributing aspects of that design to specific engineering groups.

- These groups may be geographically distributed and may only communicate infrequently.

- These groups typically like to retain control of their own design procedures and make use of in-house expertise; they object to a central design authority [87].

Decomposition through distributed architectures allows individual design groups to work in isolation, controlling their own sets of design variables, while periodically updating information from other groups to improve their aspect of the overall design. This approach to solving the problem conforms more closely with current industrial design practice than the approach of the monolithic architectures.
4.1 Classification

Previous classifications of MDO architectures:

- based on observations of which constraints were available to the optimizer to control [10, 3].

- Alexandrov and Lewis [3] used the term “closed” to denote when a set of constraints cannot be satisfied by explicit action of the optimizer, and “open” otherwise. For example, the MDF architecture is closed with respect to both analysis and consistency constraints, because their satisfaction is determined through the process of converging the MDA. Similarly, IDF is closed analysis but open consistency since the consistency constraints can be satisfied by the optimizer adjusting the coupling targets and design variables.

- Tosserams et al. [154] expanded on this classification scheme by discussing whether or not distributed architectures used open or closed local design constraints in the system subproblem.
Closure of the constraints is an important consideration when selecting an architecture because most robust optimization software will permit the exploration of infeasible regions of the design space. Such exploration may result in faster solutions via fewer optimizer iterations, but this must be weighed against the increased optimization problem size and the risk of terminating the optimization at an infeasible point.
**New classification for distributed MDO architectures:** a distributed MDO architecture can be classified based on their monolithic analogues: either MDF, IDF, or SAND.

This stems from the different approaches to handling the state and coupling variables in the monolithic architectures.

- similar to the previous classifications in that an equality constraint must be removed from the optimization problem — i.e., closed — for every variable removed from the problem statement.

- However, using a classification based on the monolithic architectures makes it much easier to see the connections between distributed architectures, even when these architectures are developed in isolation from each other.

In many cases, the problem formulations in the distributed architecture can be derived directly from that of the monolithic architecture by adding certain elements to the problem, by making certain assumptions, and by applying a specific decomposition scheme. This classification can also be viewed as a framework in which we can develop new distributed architectures, since the starting point for a distributed architecture is always a monolithic architecture.
Figure 7: Classification of the MDO architectures.
Some notes on the classification diagram:

- Known relationships between the architectures are shown by arrows.

- we have only included the “core” architectures in our diagram. Details on the available variations for each distributed architecture are presented in the relevant sections.

- Note that none of the distributed architectures developed to date have been considered analogues of SAND.

- Our classification scheme does not distinguish between the different solution techniques for the distributed optimization problems. For example, we have not focused on the order in which the distributed problems are solved. Coordination schemes are partially addressed in the Distributed IDF group, where we have classified the architectures as either “penalty” or “multilevel”, based on whether penalty functions or a problem hierarchy is used in the coordination. This grouping follows from the work of de Wit and van Keulen [38].
The following sections introduce the distributed architectures for MDO.

• We prefer to use the term “distributed” as opposed to “hierarchical” or “multilevel” because these architectures do not necessarily create a hierarchy of problems to solve.

• Furthermore, neither the systems being designed nor the design team organization need to be hierarchical in nature for these architectures to be applicable.

• Our focus here is to provide a unified description of these architectures and explain some advantages and disadvantages of each. Along the way, we will point out variations and applications of each architecture that can be found in the literature.

• We also aim to review the state-of-the-art in architectures, since the most recent detailed architecture survey in the literature dates back from more than a decade ago [145]. More recent surveys, such as that of Agte et al. [1], discuss MDO more generally without detailing the architectures themselves.
4.2 Concurrent Subspace Optimization (CSSO)

• CSSO is one of the oldest distributed architectures for large-scale MDO problems.

• The original formulation [139, 18] decomposes the system problem into independent subproblems with disjoint sets of variables.

• Global sensitivity information is calculated at each iteration to give each subproblem a linear approximation to a multidisciplinary analysis, improving the convergence behavior.

• At the system level, a coordination problem is solved to recompute the “responsibility”, “tradeoff”, and “switch” coefficients assigned to each discipline to provide information on design variable preferences for nonlocal constraint satisfaction. Using these coefficients gives each discipline a certain degree of autonomy within the system as a whole.
The version we consider here, due to Sellar et al. [135], uses metamodel representations of each disciplinary analysis to efficiently model multidisciplinary interactions. Using our unified notation, the CSSO system subproblem is given by

\[
\begin{align*}
\text{minimize} & \quad f_0(x, \tilde{y}(x, \tilde{y})) \\
\text{with respect to} & \quad x \\
\text{subject to} & \quad c_0(x, \tilde{y}(x, \tilde{y})) \geq 0 \\
& \quad c_i(x_0, x_i, \tilde{y}_i(x_0, x_i, \tilde{y}_{j \neq i})) \geq 0 \text{ for } i = 1, \ldots, N
\end{align*}
\]

and the discipline \( i \) subproblem is given by

\[
\begin{align*}
\text{minimize} & \quad f_0(x, y_i(x_i, \tilde{y}_{j \neq i}), \tilde{y}_{j \neq i}) \\
\text{with respect to} & \quad x_0, x_i \\
\text{subject to} & \quad c_0(x, \tilde{y}(x, \tilde{y})) \geq 0 \\
& \quad c_i(x_0, x_i, y_i(x_0, x_i, \tilde{y}_{j \neq i})) \geq 0 \\
& \quad c_j(x_0, \tilde{y}_j(x_0, \tilde{y})) \geq 0 \quad \text{for } j = 1, \ldots, N \ j \neq i.
\end{align*}
\]
Figure 8: Diagram for the CSSO architecture.
Algorithm 2 CSSO

Input: Initial design variables $x$

Output: Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate main CSSO iteration
repeat
1: Initiate a design of experiments (DOE) to generate design points
for Each DOE point do
2: Initiate an MDA that uses exact disciplinary information
repeat
3: Evaluate discipline analyses
4: Update coupling variables $y$
until 4 → 3: MDA has converged
5: Update the disciplinary metamodels with the latest design
end for 6 → 2
7: Initiate independent disciplinary optimizations (in parallel)
for Each discipline $i$ do
repeat
8: Initiate an MDA with exact coupling variables for discipline $i$ and approximate coupling variables for the other disciplines
repeat
9: Evaluate discipline $i$ outputs $\tilde{y}_i$, and metamodels for the other disciplines, $\tilde{y}_j \neq i$
until 10 → 9: MDA has converged
11: Compute objective $f_0$ and constraint functions $c$ using current data
until 12 → 8: Disciplinary optimization $i$ has converged
end for
13: Initiate a DOE that uses the subproblem solutions as sample points
for Each subproblem solution $i$ do
14: Initiate an MDA that uses exact disciplinary information
repeat
15: Evaluate discipline analyses.
until 16 → 15 MDA has converged
17: Update the disciplinary metamodels with the newest design
end for 18 → 14
19: Initiate system-level optimization
repeat
20: Initiate an MDA that uses only metamodel information
repeat
21: Evaluate disciplinary metamodels
until 22 → 21: MDA has converged
until 25 → 1: CSSO has converged
A potential pitfall of CSSO architecture is the necessity of including all design variables in the system subproblem. For industrial-scale design problems, this may not always be possible or practical.

There have been some benchmarks comparing CSSO with other MDO architectures. Perez et al., [121] Yi et al. [162], and Tedford and Martins [150] all show CSSO requiring many more analysis calls than other architectures to converge to an optimal design. The results of de Wit and van Keulen [37] showed that CSSO was unable to reach the optimal solution of even a simple minimum-weight two-bar truss problem. Thus, CSSO seems to be largely ineffective when compared with newer MDO architectures.
4.3 Collaborative Optimization (CO)

- In CO, the disciplinary optimization problems are formulated to be independent of each other by using target values of the coupling and shared design variables [20, 21].

- These target values are then shared with all disciplines during every iteration of the solution procedure.

- The complete independence of disciplinary subproblems combined with the simplicity of the data-sharing protocol makes this architecture attractive for problems with a small amount of shared data.
Figure 9: Diagram for the CO architecture.
Braun [20] formulated two versions of the CO architecture: CO$_1$ and CO$_2$. CO$_2$ is the most frequently used of these two original formulations so it will be the focus of our discussion. The CO$_2$ system subproblem is given by:

minimize $f_0 (x_0, \hat{x}_1, \ldots, \hat{x}_N, y^t)$

with respect to $x_0, \hat{x}_1, \ldots, \hat{x}_N, y^t$

subject to $c_0 (x_0, \hat{x}_1, \ldots, \hat{x}_N, y^t) \geq 0$

\[ J^*_i = ||\hat{x}_{0i} - x_0||_2^2 + ||\hat{x}_i - x_i||_2^2 + \\
||y^t_i - y_i (\hat{x}_{0i}, x_i, y^t_{j \neq i})||_2^2 = 0 \text{ for } i = 1, \ldots, N \]

where $\hat{x}_{0i}$ are copies of the global design variables passed to discipline $i$ and $\hat{x}_i$ are copies of the local design variables passed to the system subproblem.
• Note that copies of the local design variables are only made if those variables directly influence the objective.

• In CO$_1$, the quadratic equality constraints are replaced with linear equality constraints for each target-response pair. In either case, post-optimality sensitivity analysis, i.e. computing derivatives with respect to an optimized function, is required to evaluate the derivatives of the consistency constraints $J_i^*$.

The discipline $i$ subproblem in both CO$_1$ and CO$_2$ is

$$\text{minimize} \quad J_i \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_{j \neq i}^t \right) \right)$$

with respect to $\hat{x}_{0i}, x_i$

subject to $c_i \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_{j \neq i}^t \right) \right) \geq 0.$

$$\text{(4)}$$

• The system-level problem is responsible for minimizing the design objective

• The discipline level problems minimize system inconsistency.

• Braun [20] showed that the CO problem statement is mathematically equivalent to the original AAO MDO problem.
Algorithm 3 Collaborative optimization

Input: Initial design variables $x$

Output: Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate system optimization iteration

repeat
  1: Compute system subproblem objectives and constraints
    for Each discipline $i$ (in parallel) do
      1.0: Initiate disciplinary subproblem optimization
        repeat
          1.1: Evaluate disciplinary analysis
          1.2: Compute disciplinary subproblem objective and constraints
          1.3: Compute new disciplinary subproblem design point and $J_i$
        until 1.3 $\rightarrow$ 1.1: Optimization $i$ has converged
    end for
  2: Compute a new system subproblem design point
until 2 $\rightarrow$ 1: System optimization has converged
In spite of the organizational advantage of having fully separate disciplinary subproblems, CO has major weaknesses in the mathematical formulation that lead to poor performance in practice [4, 40].

- In particular, the system problem in CO$_1$ has more equality constraints than variables, so if the system cannot be made fully consistent, the system subproblem is infeasible. This can also happen in CO$_2$, but it is not the most problematic issue.

- The most significant difficulty with CO$_2$ is that the constraint gradients of the system problem at an optimal solution are all zero vectors. This represents a breakdown in the constraint qualification of the Karush–Kuhn–Tucker optimality conditions, which slows down convergence for most gradient-based optimization software [4]. In the worst case, the CO$_2$ formulation may not converge at all.

These difficulties with the original formulations of CO have inspired several researchers to develop modifications to improve the behavior of the architecture.
In a few cases, problems have been solved with CO and a gradient-free optimizer, such as a genetic algorithm [164], or a gradient-based optimizer that does not use the Lagrange multipliers in the termination condition [96] to handle the troublesome constraints. While such approaches do avoid the obvious problems with CO, they bring other issues.

- Gradient-free optimizers are computationally expensive and can become the bottleneck within the CO architecture.

- Gradient-based optimizers that do not terminate based on Lagrange multiplier values, such as feasible direction methods, often fail in nonconvex feasible regions. As pointed out by DeMiguel [40], the CO system subproblem is set-constrained, i.e., nonconvex, because of the need to satisfy optimality in the disciplinary subproblems.
The approach taken by DeMiguel and Murray [40] to fix the problems with CO is to relax the troublesome constraints using an $L_1$ exact penalty function with a fixed penalty parameter value and add elastic variables to preserve the smoothness of the problem. This revised approach is called Modified Collaborative Optimization (MCO).

- This approach satisfies the requirement of mathematical rigor, as algorithms using the penalty function formulation are known to converge to an optimal solution under mild assumptions [44, 115].

- However, the test results of Brown and Olds [24] show strange behavior in a practical design problem. In light of their findings, the authors rejected MCO from further testing.
Another idea, proposed by Sobieski and Kroo [138], uses surrogate models, also known as metamodels, to approximate the post-optimality behavior of the disciplinary subproblems in the system subproblem.

- This both eliminates the post-optimality sensitivity calculation and improves the treatment of the consistency constraints.

- While the approach does seem to be effective for the problems they solve, to our knowledge, it has not been adopted by any other researchers to date.
The simplest and most effective known fix for the difficulties of CO involves relaxing the system subproblem equality constraints to inequalities with a relaxation tolerance, which was originally proposed by Braun et al. [21].

- This approach was also successful in other test problems [120, 103], where the choice of tolerance is a small fixed number, usually $10^{-6}$.

- The effectiveness of this approach stems from the fact that a positive inconsistency value causes the gradient of the constraint to be nonzero if the constraint is active, eliminating the constraint qualification issue.

- Nonzero inconsistency is not an issue in a practical design setting provided the inconsistency is small enough such that other errors in the computational model dominate at the final solution.

Li et al. [93] build on this approach by adaptively choosing the tolerance during the solution procedure so that the system-level problem remains feasible at each iteration. This approach appears to work when applied to the test problems in [4], but has yet to be verified on larger test problems.
Despite the numerical issues, CO has been widely implemented on a number of MDO problems. Most of applications are in the design of aerospace systems. Examples include

1. the design of launch vehicles [23],
2. rocket engines [27],
3. satellite constellations [26],
4. flight trajectories [22, 91],
5. flight control systems [122],
6. preliminary design of complete aircraft [88, 102], and
7. aircraft family design [7].

Outside aerospace engineering, CO has been applied to problems involving automobile engines [109], bridge design [9], railway cars [42], and even the design of a scanning optical microscope [124].
Adaptations of the CO architecture have also been developed for multiobjective, robust, and multifidelity MDO problems. Multiobjective formulations of CO were first described by Tappeta and Renaud [149]. McAllister et al. [110] present a multiobjective approach using linear physical programming. Available robust design formulations incorporate the decision-based models of Gu et al. [53] and McAllister and Simpson [109], the implicit uncertainty propagation method of Gu et al. [54], and the fuzzy computing models of Huang et al. [68]. Multiple model fidelities were integrated into CO for an aircraft design problem by Zadeh and Toropov [163].
The most recent version of CO — Enhanced Collaborative Optimization (ECO) — was developed by Roth and Kroo [129, 128]. Figure 10 shows the XDSM corresponding to this architecture. The problem formulation of ECO, while still being derived from the same basic problem as the original CO architecture, is radically different and therefore deserves additional attention. In a sense, the roles of the system and discipline optimization have been reversed in ECO when compared to CO. In ECO the system subproblem minimizes system infeasibility, while the disciplinary subproblems minimize the system objective. The system subproblem is

$$\text{minimize} \quad J_0 = \sum_{i=1}^{N} \| \hat{x}_{0i} - x_0 \|_2^2 + \| y^t_i - y_i (x_0, x_i, y^t_{j \neq i}) \|_2^2$$

with respect to \( x_0, y^t \). 

Note that this subproblem is unconstrained. Also, unlike CO, post-optimality sensitivities are not required by the system subproblem because the disciplinary responses are treated as parameters. The system subproblem chooses the shared design variables by averaging all disciplinary preferences.
Figure 10: XDSM for the ECO architecture
The $i^{th}$ disciplinary subproblem in ECO is

\[
\text{minimize} \quad J_i = \tilde{f}_0(\hat{x}_0i, y_i(\hat{x}_0i, x_i, y_{j \neq i})) + \\
\quad w_{Ci} \left( ||\hat{x}_0i - x_0||_2^2 + ||y_i^t - y_i(\hat{x}_0i, x_i, y_{j \neq i})||_2^2 \right) + \\
\quad w_{Fi} \sum_{j=1, j \neq i}^{N} \sum_{k=1}^{n_s} s_{jk}
\]

with respect to $\hat{x}_0i, x_i, s_{j \neq i}$

subject to $c_i(\hat{x}_0i, x_i, y_i(\hat{x}_0i, x_i, y_{j \neq i})) \geq 0$

$\tilde{c}_{j \neq i}(\hat{x}_0i) - s_{j \neq i} \geq 0 \quad j = 1, \ldots, N$

$s_{j \neq i} \geq 0 \quad j = 1, \ldots, N,$

(6)

where $w_{Ci}$ and $w_{Fi}$ are penalty weights for the consistency and nonlocal design constraints, and $s$ is a local set of elastic variables for the constraint models. The $w_{Fi}$ penalty weights are chosen to be larger than the largest Lagrange multiplier, while the $w_{Ci}$ weights are chosen to guide the optimization toward a consistent solution.
The main new idea introduced in ECO is to include linear models of nonlocal constraints, represented by $c_{j \neq i}$, and a quadratic model of the system objective function in each disciplinary subproblem, represented by $\tilde{f}_0$. This is meant to increase each discipline’s “awareness” of their influence on other disciplines and the global objective as a whole. The constraint models for each discipline are constructed by first solving the optimization problem that minimizes the constraint violation with respect to local elastic and design variables.

$$\begin{align*}
\text{minimize} & \quad \sum_{k=1}^{n_s} s_{ik} \\
\text{with respect to} & \quad x_i, s_i \\
\text{subject to} & \quad c_i (x_0, x_i, y_i (x_0, x_i, y_{j \neq i}^t)) + s_i \geq 0 \\
& \quad s_i \geq 0.
\end{align*}$$

(7)
Algorithm 4 Enhanced collaborative optimization

**Input:** Initial design variables $x$

**Output:** Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate ECO iteration

repeat

for Each discipline $i$ do

1: Create linear constraint model

2: Initiate disciplinary subproblem optimization

repeat

3: Interrogate nonlocal constraint models with local copies of shared variables

3.0: Evaluate disciplinary analysis

3.1: Compute disciplinary subproblem objective and constraints

4: Compute new disciplinary subproblem design point and $J_i$

until 4 $\rightarrow$ 3: Disciplinary optimization subproblem has converged

end for

5: Initiate system optimization

repeat

6: Compute $J_0$

7: Compute updated values of $x_0$ and $y^t$.

until 7 $\rightarrow$ 6: System optimization has converged

until 8 $\rightarrow$ 1: The $J_0$ is below specified tolerance
Based on the Roth’s results [129, 128], ECO is effective in reducing the number of discipline analyses compared to CO. The trade-off is in the additional time required to build and update the models for each discipline, weighed against the simplified solution to the decomposed optimization problems. The results also show that ECO compares favorably with the Analytical Target Cascading architecture (which we will describe in Section 4.5).

While ECO seems to be effective, CO tends to be an inefficient architecture for solving MDO problems.

- Without any of the fixes discussed in this section, the architecture always requires a disproportionately large number of function and discipline evaluations [79, 83, 37, 162], assuming it converges at all.

- When the system-level equality constraints are relaxed, the results from CO are more competitive with other distributed architectures [121, 103, 150] but still compare poorly with the results of monolithic architectures.
4.4 Bilevel Integrated System Synthesis (BLISS)

The BLISS architecture [143], like CSSO, is a method for decomposing the MDF problem along disciplinary lines.

- Unlike CSSO, however, BLISS assigns local design variables to disciplinary subproblems and shared design variables to the system subproblem.

- The basic approach of the architecture is to form a path in the design space using a series of linear approximations to the original design problem, with user-defined bounds on the design variable steps, to prevent the design point from moving so far away that the approximations are too inaccurate.

This is an idea similar to that of trust-region methods [34]. These approximations are constructed at each iteration using global sensitivity information.
The BLISS system level subproblem is formulated as

\[
\text{minimize } (f^*_0)_0 + \left( \frac{df^*_0}{dx_0} \right) \Delta x_0 \\
\text{with respect to } \Delta x_0 \\
\text{subject to } (c^*_0)_0 + \left( \frac{dc^*_0}{dx_0} \right) \Delta x_0 \geq 0 \\
(c^*_i)_0 + \left( \frac{dc^*_i}{dx_0} \right) \Delta x_0 \geq 0 \text{ for } i = 1, \ldots, N \\
\Delta x_{0L} \leq \Delta x_0 \leq \Delta x_{0U}.
\]
The BLISS discipline $i$ subproblem is given by

minimize $$(f_0)_0 + \left( \frac{df_0}{dx_i} \right) \Delta x_i$$

with respect to $\Delta x_i$

subject to $$(c_0)_0 + \left( \frac{dc_0}{dx_i} \right) \Delta x_i \geq 0 \quad (9)$$

$$(c_i)_0 + \left( \frac{dc_i}{dx_i} \right) \Delta x_i \geq 0$$

$$\Delta x_{iL} \leq \Delta x_i \leq \Delta x_{iU}.$$ 

Note the extra set of constraints in both system and discipline subproblems denoting the design variables bounds.
Figure 11: Diagram for the BLISS architecture
Algorithm 5 BLISS

Input: Initial design variables $x$

Output: Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate system optimization

repeat
  1: Initiate MDA

  repeat
    2: Evaluate discipline analyses
    3: Update coupling variables
  until $3 \rightarrow 2$: MDA has converged

4: Initiate parallel discipline optimizations

for Each discipline $i$ do
  5: Evaluate discipline analysis
  6: Compute objective and constraint function values and derivatives with respect to local design variables
  7: Compute the optimal solutions for the disciplinary subproblem
end for

8: Initiate system optimization

9: Compute objective and constraint function values and derivatives with respect to shared design variables using post-optimality sensitivity analysis

10: Compute optimal solution to system subproblem

until $11 \rightarrow 1$: System optimization has converged
In order to prevent violation of the disciplinary constraints by changes in the shared design variables, post-optimality sensitivity information is required to solve the system subproblem.

For this step, Sobieski [143] presents two methods:

**BLISS/A:** using a generalized version of the Global Sensitivity Equations [140], and

**BLISS/B:** using the “pricing” interpretation of local Lagrange multipliers.

Other variations use response surface approximations to compute post-optimality sensitivity data [81, 73].
Some notes on the BLISS algorithm:

- due to the linear nature of the optimization problems under consideration, repeated interrogation of the objective and constraint functions is not necessary once gradient information is available.

- However, this reliance on linear approximations is not without difficulties. If the underlying problem is highly nonlinear, the algorithm may converge slowly. The presence of user-defined variable bounds may help the convergence if these bounds are properly chosen, such as through a trust region framework.

- Detailed knowledge of the design space can also help, but this increases the overhead cost of implementation.
Two other adaptations of the original BLISS architecture are known in the literature.

1. The first is Ahn and Kwon’s proBLISS [2], an architecture for reliability-based MDO. Their results show that the architecture is competitive with reliability-based adaptations of MDF and IDF.

2. The second is LeGresley and Alonso’s BLISS/POD [92], an architecture that integrates a reduced-order modeling technique called Proper Orthogonal Decomposition [13] to reduce the cost of the multidisciplinary analysis and sensitivity analysis steps. Their results show a significant improvement in the performance of BLISS, to the point where it is almost competitive with MDF.
As an enhancement of the original BLISS, a radically different formulation called **BLISS-2000** was developed by Sobieski et al. [144].

- BLISS-2000 does not require a multidisciplinary analysis to restore feasibility of the design, so we have separated it from other BLISS variants in the classification tree shown earlier.

- like other IDF-derived architectures, BLISS-2000 uses coupling variable targets to enforce consistency at the optimum. Information exchange between system and discipline subproblems is completed through surrogate models of the disciplinary optima.
The BLISS-2000 system subproblem is given by

\[
\begin{align*}
\text{minimize} & \quad f_0 \left( x, \tilde{y} \left( x, y^t \right) \right) \\
\text{with respect to} & \quad x_0, y^t, w \\
\text{subject to} & \quad c_0 \left( x, \tilde{y} \left( x, y^t, w \right) \right) \geq 0 \\
\quad & \quad y_i^t - \tilde{y}_i \left( x_0, x_i, y_{j \neq i}^t, w_i \right) = 0 \quad \text{for } i = 1, \ldots, N.
\end{align*}
\] (10)

The BLISS-2000 discipline \( i \) subproblem is

\[
\begin{align*}
\text{minimize} & \quad w_i^T y_i \\
\text{with respect to} & \quad x_i \\
\text{subject to} & \quad c_i \left( x_0, x_i, y_i \left( x_0, x_i, y_{j \neq i}^t \right) \right) \geq 0.
\end{align*}
\] (11)

A unique aspect of this architecture is the use of a vector of weighting coefficients, \( w_i \), attached to the disciplinary states. These weighting coefficients give the user a measure of control over state variable preferences. Generally speaking, the coefficients should be chosen based on the structure of the global objective to allow disciplinary subproblems to find an optimum more quickly. How much the choice of coefficients affects convergence has yet to be determined.
Figure 12: Diagram for the BLISS-2000 architecture
Algorithm 6 BLISS-2000

**Input:** Initial design variables $x$

**Output:** Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate system optimization

repeat
  for Each discipline $i$ do
    1: Initiate a DOE
    for Each DOE point do
      2: Initiate discipline subproblem optimization
    repeat
      3: Evaluate discipline analysis
      4: Compute discipline objective $f_i$ and constraint functions $c_i$
      5: Update the local design variables $x_i$
    until $5 \rightarrow 3$: Disciplinary optimization subproblem has converged
    6: Update metamodel of optimized disciplinary subproblem with new solution
  end for $\rightarrow 1$

end for

8: Initiate system subproblem optimization

repeat
  9: Interrogate metamodels with current values of system variables
  10: Compute system objective and constraint function values
  11: Compute a new system design point
until $11 \rightarrow 9$ System subproblem has converged

until $12 \rightarrow 1$: System optimization has converged
A unique aspect of this architecture is the use of a vector of weighting coefficients, $w_i$, attached to the disciplinary states. These weighting coefficients give the user a measure of control over state variable preferences.

- The coefficients should be chosen based on the structure of the global objective to allow disciplinary subproblems to find an optimum more quickly.

- How much the choice of coefficients affects convergence has yet to be determined.
BLISS-2000 possesses several advantages over the original BLISS architecture.

1. the solution procedure is much easier to understand.

2. the decomposed problem formulation of BLISS-2000 is equivalent to the AAO problem we want to solve [144].

3. by using metamodels for each discipline, rather than for the whole system, the calculations for BLISS-2000 can be run in parallel with minimal communication between disciplines.

4. BLISS-2000 seems to be more flexible than its predecessor.

Recently, Sobieski detailed an extension of BLISS-2000 to handle multilevel, system-of-systems problems [142]. In spite of these advantages, it appears that BLISS-2000 has not been used nearly as frequently as the original BLISS formulation.
4.5 Analytical Target Cascading (ATC)

The ATC architecture was not initially developed as an MDO architecture, but as a method to propagate system targets — i.e., requirements or desirable properties — through a hierarchical system to achieve a feasible system design satisfying these targets [74, 77].

- If the system targets were unattainable, the ATC architecture would return a design point minimizing the inattainability.

- Effectively, the ATC architecture is no different from an MDO architecture with a system objective of minimizing the squared difference between a set of system targets and model responses.

- By simply changing the objective function, we can solve general MDO problems using ATC.
The ATC problem formulation that we present here is due to Tosserams et al. [153]:

\[
\text{minimize } f_0 (x, y^t) + \sum_{i=1}^{N} \Phi_i \left( \hat{x}_{0i} - x_0, y_i^t - y_i \left( x_0, x_i, y^t \right) \right) + \\
\Phi_0 \left( c_0 \left( x, y^t \right) \right)
\]

with respect to \( x_0, y^t \),

(12)

where \( \Phi_0 \) is a penalty relaxation of the global design constraints and \( \Phi_i \) is a penalty relaxation of the discipline \( i \) consistency constraints. The \( i^{th} \) discipline subproblem is:

\[
\text{minimize } f_0 \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_j^t \neq i \right), y_j^t \neq i \right) + f_i \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_j^t \neq i \right) \right) + \\
\Phi_i \left( y_i^t - y_i \left( \hat{x}_{0i}, x_i, y_j^t \neq i \right), \hat{x}_{0i} - x_0 \right) + \\
\Phi_0 \left( c_0 \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_j^t \neq i \right), y_j^t \neq i \right) \right)
\]

with respect to \( \hat{x}_{0i}, x_i \)

subject to \( c_i \left( \hat{x}_{0i}, x_i, y_i \left( \hat{x}_{0i}, x_i, y_j^t \neq i \right) \right) \geq 0. \)
Figure 13: Diagram for the ATC architecture
Algorithm 7 ATC

Input: Initial design variables $x$

Output: Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate main ATC iteration

repeat

for Each discipline $i$ do

1: Initiate discipline optimizer

repeat

2: Evaluate disciplinary analysis

3: Compute discipline objective and constraint functions and penalty function values

4: Update discipline design variables

until 4 $\rightarrow$ 2: Discipline optimization has converged

end for

5: Initiate system optimizer

repeat

6: Compute system objective, constraints, and all penalty functions

7: Update system design variables and coupling targets.

until 7 $\rightarrow$ 6: System optimization has converged

8: Update penalty weights

until 8 $\rightarrow$ 1: Penalty weights are large enough
Note that ATC can be applied to a multilevel hierarchy of systems just as well as a discipline-based non-hierarchic system.

- In the multilevel case, the penalty functions are applied to all constraints that combine local information with information from the levels immediately above or below the current one.

- Also note that post-optimality sensitivity data is not needed in any of the subproblems as nonlocal data are always treated as fixed values in the current subproblem.
The most common penalty functions in ATC are quadratic penalty functions.

- The proper selection of the penalty weights is important for both final inconsistency in the discipline models and convergence of the algorithm.

- Michalek and Papalambros [112] present an effective weight update method that is especially useful when unattainable targets have been set in a traditional ATC process.

- Michelena et al. [114] present several coordination algorithms using ATC with quadratic penalty functions and demonstrate the convergence for all of them.

Note that, as with penalty methods for general optimization problems, (see, e.g., Nocedal and Wright [115, chap. 17]) the solution of the MDO problem must be computed to reasonable accuracy before the penalty weights are updated. However, because we are now dealing with a distributed set of subproblems, the whole hierarchy of subproblems must be solved for a given set of weights. This is due to the nonseparable nature of the quadratic penalty function.
Several other penalty function choices and associated coordination approaches have also been devised for ATC.

1. Kim et al. [75] outline a version of ATC that uses Lagrangian relaxation and a sub-gradient method to update the multiplier estimates.

2. Tosserams et al. [151] use augmented Lagrangian relaxation with Bertsekas’ method of multipliers [14] and alternating direction method of multipliers [15] to update the penalty weights. They also group this variant of ATC into a larger class of coordination algorithms known as Augmented Lagrangian Coordination [153].

3. Li et al. [94] apply the diagonal quadratic approximation approach of Ruszcynski [130] to the augmented Lagrangian to eliminate subproblem coupling through the quadratic terms and further parallelize the architecture.

4. Han and Papalambros [62] propose a version of ATC based on sequential linear programming [11, 28], where inconsistency is penalized using infinity norms. They later presented a convergence proof of this approach in a short note [61].
For each of the above penalty function choices, ATC was able to produce the same design solutions as the monolithic architectures.
Despite having been developed relatively recently, the ATC architecture has been widely used. By far, ATC has been most frequently applied to design problems in the field for which it was developed, the automotive industry [84, 76, 78, 19, 85, 147, 29, 63].

- However, the ATC approach has also proven to be useful in aircraft design [8, 7, 158] and building design [32].

- ATC has also found applications outside of strict engineering design problems, including manufacturing decisions [95], supply chain management [67], and marketing decisions in product design [111].

- Huang et al. [66] have developed an ATC-specific web portal to solve optimization problems via the ATC architecture. Etman et al. [43] discuss the automatic implementation of coordination procedures, using ATC as an example architecture.

- There are ATC formulations that can handle integer variables [113] and probabilistic design problems [86, 98].

- Another important adaptation of ATC applies to problems with
block-separable linking constraints [155]. In this class of problems, $c_0$ consists of constraints which are sums of functions depending on only shared variables and the local variables of one discipline.
The performance of ATC compared with other architectures is not well known because only one result is available.

- In de Wit and Van Keulen’s architecture comparison [37], ATC is competitive with all other benchmarked distributed architectures, including standard versions of CSSO, CO, and BLISS.

- However, ATC and the other distributed architectures are not competitive with a monolithic architecture in terms of the number of function and discipline evaluations.

More commonly, different versions of ATC are benchmarked against each other.

- Tosserams et al. [151] compared the augmented Lagrangian penalty approach with the quadratic penalty approach and found much improved results with the alternating direction method of multipliers.

- Surprisingly, de Wit and van Keulen [37] found the augmented Lagrangian version performed worse than the quadratic penalty method for their test problem.
• Han and Papalambros [62] compared their sequential linear programming version of ATC to several other approaches and found a significant reduction in the number of function evaluations. However, they note that the coordination overhead is large compared to other ATC versions and still needs to be addressed.
4.6 Exact and Inexact Penalty Decomposition (EPD and IPD)

If there are no system-wide constraints or objectives, i.e., if neither $f_0$ and $c_0$ exist, the Exact or Inexact Penalty Decompositions (EPD or IPD) [39, 41] may be employed. Both formulations rely on solving the disciplinary subproblem

$$\text{minimize} \quad f_i (\hat{x}_0, x_i, y_i (\hat{x}_0, x_i, y_{j \neq i})) + \Phi_i (\hat{x}_0 - x_0, y^t_i - y_i (\hat{x}_0, x_i, y_{j \neq i}))$$

with respect to $\hat{x}_0, x_i$

subject to $c_i (\hat{x}_0, x_i, y_i (x_0, x_i, y_{j \neq i})) \geq 0.$

(14)

Here, $\Phi_i$ denotes the penalty function associated with the inconsistency between the $i^{th}$ disciplinary information and the system information. In EPD, $\Phi_i$ is an $L_1$ penalty function with additional variables and constraints added to ensure smoothness. In IPD, $\Phi_i$ is a quadratic penalty function with appropriate penalty weights. The notation $\hat{x}_0$ denotes a local copy of the shared design variables in discipline $i$, while $x_0$ denotes the system copy.
Figure 14: Diagram for the penalty decomposition architectures EPD and IPD
At the system level, the subproblem is an unconstrained minimization with respect to the target variables. The objective function is the sum of the optimized disciplinary penalty terms, denoted as $\Phi_i^*$.

$$\text{minimize} \quad \sum_{i=1}^{N} \Phi_i^* (x_0, y^t) = \sum_{i=1}^{N} \Phi_i \left( \hat{x}_{0i} - x_0, y_i^t - y_i \left( \hat{x}_{0i}, x_i, y_{j \neq i}^t \right) \right)$$

with respect to $x_0, y^t$

(15)

The penalty weights are updated upon solution of the system problem. Figure 14 shows the XDSM for this architecture, where $w$ represents the penalty weights. The sequence of operations in this architecture is detailed in Algorithm 8.
Algorithm 8 EPD and IPD

**Input:** Initial design variables $x$

**Output:** Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate main iteration

repeat

for Each discipline $i$ do

repeat

1: Initiate discipline optimizer
2: Evaluate discipline analysis
3: Compute discipline objective and constraint functions, and penalty function values
4: Update discipline design variables

until 4 $\rightarrow$ 2: Discipline optimization has converged

end for

5: Initiate system optimizer

repeat

6: Compute all penalty functions
7: Update system design variables and coupling targets

until 7 $\rightarrow$ 6: System optimization has converged

8: Update penalty weights.

until 8 $\rightarrow$ 1: Penalty weights are large enough
Both EPD and IPD have mathematically provable convergence under the linear independence constraint qualification and with mild assumptions on the update strategy for the penalty weights [41].

- In particular, the penalty weight in IPD must monotonically increase until the inconsistency is sufficiently small, similar to other quadratic penalty methods [115].

- For EPD, the penalty weight must be larger than the largest Lagrange multiplier, following established theory of the $L_1$ penalty function [115], while the barrier parameter must monotonically decrease like in an interior point method [160].

- If other penalty functions are employed, the parameter values are selected and updated according to the corresponding mathematical theory.

Under these conditions, the solution obtained under EPD and IPD will also be a solution to the desired AAO MDO problem.
Only once in the literature has either penalty decomposition architecture been tested against any others.

• The results of Tosserams et al. [152] suggest that performance depends on the choice of penalty function employed.

• A comparison between IPD with a quadratic penalty function and IPD with an augmented Lagrangian penalty function showed that the latter significantly outperformed the former in terms of both time and number of function evaluations on several test problems.
4.7 MDO of Independent Subspaces (MDOIS)

If the problem contains no system-wide constraints or objectives, i.e., if neither $f_0$ and $c_0$ exist, and the problem does not include shared design variables, i.e., if $x_0$ does not exist, then the MDO of independent subspaces (MDOIS) architecture [137] applies. In this case, the discipline subproblems are fully separable (aside from the coupled state variables) and given by

\[
\begin{align*}
\text{minimize} & \quad f_i (x_i, y_i (x_i, y_j^{t \neq i})) \\
\text{with respect to} & \quad x_i \\
\text{subject to} & \quad c_i (x_i, y_i (x_i, y_j^{t \neq i})) \geq 0.
\end{align*}
\]
Figure 15: Diagram for the MDOIS architecture
Algorithm 9 MDOIS

Input: Initial design variables $x$

Output: Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate main iteration

repeat
  repeat
    1: Initiate MDA
    2: Evaluate discipline analyses
    3: Update coupling variables
  until $3 \rightarrow 2$: MDA has converged

for Each discipline $i$ do
  4: Initiate disciplinary optimization
  repeat
    5: Evaluate discipline analysis
    6: Compute discipline objectives and constraints
    7: Compute a new discipline design point
  until $7 \rightarrow 5$: Discipline optimization has converged
end for

until $8 \rightarrow 1$ Main iteration has converged
Some notes on MDOIS:

- The targets are just local copies of system state information.

- Upon solution of the disciplinary problems, which can access the output of individual disciplinary analysis codes, a full multidisciplinary analysis is completed to update all target values. Thus, rather than a system subproblem used by other architectures, the MDA is used to guide the disciplinary subproblems to a design solution.

- Shin and Park [137] show that under the given problem assumptions an optimal design is found using this architecture.

Benchmarking results are available comparing MDOIS to some of the older architectures. These results are given by Yi et al. [162].

- In many cases, MDOIS requires fewer analysis calls than MDF while still being able to reach the optimal solution.

- However, MDOIS still does not converge as fast as IDF and the restrictive
definition of the problem means that the architecture is not nearly as flexible as MDF.

- A practical problem that can be solved using MDOIS is the belt-integrated seat problem of Shin et al. [136]. However, the results using MDOIS have not been compared to results obtained using other architectures.
4.8 Quasiseparable Decomposition (QSD)

Haftka and Watson [59] developed the QSD architecture to solve *quasiseparable* optimization problems.

- In a quasiseparable problem, the system objective and constraint functions are assumed to be dependent only on global variables (i.e., the shared design and coupling variables).

- This type of problem may be thought of as identical to the complicating variables problems.

We have not come across any practical design problems that satisfy this property. However, if required by the problem, we can easily transform the general (AAO) MDO problem into a quasiseparable problem.

- This is accomplished by duplicating the relevant local variables, and forcing the global objective to depend on the target copies of local variables.

- The resulting quasiseparable problem and decomposition is mathematically equivalent to the original problem.
The system subproblem is given by

\begin{align}
\text{minimize} & \quad f_0(x_0, y^t) + \sum_{i=1}^{N} b_i \\
\text{with respect to} & \quad x_0, y^t, b \\
\text{subject to} & \quad c_0(x_0, y^t) \geq 0 \\
& \quad s^*_i(x_0, x_i, y_i(x_0, x_i, y^t_{j \neq i}), b_i) \geq 0 \quad \text{for } i = 1, \ldots, N.
\end{align}

(17)

where \(s_i\) is the constraint margin for discipline \(i\) and \(b_i\) is the “budget” assigned to each disciplinary objective. The discipline \(i\) subproblem becomes

\begin{align}
\text{minimize} & \quad -s_i \\
\text{with respect to} & \quad x_i, s_i \\
\text{subject to} & \quad c_i(x_0, x_i, y_i(x_0, x_i, y^t_{j \neq i})) - s_i \geq 0 \\
& \quad f_i(x_0, x_i, y_i(x_0, x_i, y^t_{j \neq i})) - b_i - s_i \geq 0 \\
& \quad y^t_i - y_i(x_0, x_i, y^t_{j \neq i}) = 0
\end{align}

(18)

where \(k\) is an element of the constraint vector \(c_i\).
Algorithm 10 QSD

**Input:** Initial design variables $x$

**Output:** Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate system optimization

repeat

1: Compute system objectives and constraints

for Each discipline $i$ do

1.0: Initiate discipline optimization

repeat

1.1: Evaluate discipline analysis

1.2: Compute discipline objective and constraints

1.3: Update discipline design point

until 1.3 $\rightarrow$ 1.1: Discipline optimization has converged

end for

2: Compute a new system design point

until 2 $\rightarrow$ 1: System problem has converged
Figure 16: Diagram for the QSD architecture
Some notes on QSD:

• Due to the use of target copies, we classify this architecture as distributed IDF.

• This is a bilevel architecture where the solutions of disciplinary subproblems are constraints in the system subproblem. Therefore, post-optimality sensitivities or surrogate model approximations of optimized disciplinary subproblems are required to solve the system subproblem.

• Haftka and Watson have extended the theory behind QSD to solve problems with a combination of discrete and continuous variables [60].

• Liu et al. [97] successfully applied QSD with surrogate models to a structural optimization problem. However, they made no comparison of the performance to other architectures, not even QSD without the surrogates.

• A version of QSD without surrogate models was benchmarked by de Wit and van Keulen [37]. Unfortunately, this architecture was the worst of all the architectures tested in terms of disciplinary evaluations.
A version of QSD using surrogate models should yield improved performance, due to the smoothness introduced by the model, but this version has not been benchmarked to our knowledge.
4.9 Asymmetric Subspace Optimization (ASO)

The ASO architecture [31] is a new distributed-MDF architecture.

- It was motivated by the case of high-fidelity aerostructural optimization, where the aerodynamic analysis typically requires an order of magnitude more time to complete than the structural analysis [105].

- To reduce the number of expensive aerodynamic analyses, the structural analysis is coupled with a structural optimization inside the MDA.

- This idea can be readily generalized to any problem where there is a wide discrepancy between discipline analysis times.
The system subproblem in ASO is

\[
\text{minimize} \quad f_0(x, y(x, y)) + \sum_k f_k(x_0, x_k, y_k(x_0, x_k, y_j \neq k))
\]

with respect to \( x_0, x_k \)
subject to \( c_0(x, y(x, y)) \geq 0 \)
\[
c_k(x_0, x_k, y_k(x_0, x_k, y_j \neq k)) \geq 0 \quad \text{for all } k,
\]

(19)

where subscript \( k \) denotes disciplinary information that remains outside of the MDA. The disciplinary problem for discipline \( i \), which is resolved inside the MDA, is

\[
\text{minimize} \quad f_0(x, y(x, y)) + f_i(x_0, x_i, y_i(x_0, x_i, y_j \neq i))
\]

with respect to \( x_i \)
subject to \( c_i(x_0, x_i, y_i(x_0, x_i, y_j \neq i)) \geq 0. \)
Figure 17: Diagram for the ASO architecture
Algorithm 11 ASO

**Input:** Initial design variables $x$

**Output:** Optimal variables $x^*$, objective function $f^*$, and constraint values $c^*$

0: Initiate system optimization

repeat

1: Initiate MDA

repeat

2: Evaluate Analysis 1
3: Evaluate Analysis 2
4: Initiate optimization of Discipline 3

repeat

5: Evaluate Analysis 3
6: Compute discipline 3 objectives and constraints
7: Update local design variables

until 7 $\rightarrow$ 5: Discipline 3 optimization has converged

8: Update coupling variables

until 8 $\rightarrow$ 2 MDA has converged

9: Compute objective and constraint function values for all disciplines 1 and 2

10: Update design variables

until 10 $\rightarrow$ 1: System optimization has converged
Notes on ASO:

- The optimality of the final solution is preserved by using the coupled post-optimality sensitivity (CPOS) equations, developed by Chittick and Martins [31], to calculate gradients at the system level.

- CPOS represents the extension of the coupled sensitivity equations [141, 106] to include the optimality conditions.

- ASO was later implemented using a coupled-adjoint approach as well [30].

- Kennedy et al. [72] present alternative strategies for computing the disciplinary subproblem optima and the post-optimality sensitivity analysis.

- As with other bilevel MDO architectures, the post-optimality analysis is necessary to ensure convergence to an optimal design of the original monolithic problem.
Results of ASO show a substantial reduction in the number of calls to the aerodynamics analysis, and even a slight reduction in the number of calls to the structural analysis [31].

- However, the total time for the optimization routine is only competitive with MDF if the aerodynamic analysis is substantially more costly than the structural analysis.

- If the two analyses take roughly equal time, MDF is still much faster.

- Furthermore, the use of CPOS increases the complexity of the sensitivity analysis compared to a normal coupled adjoint, which adds to the overall solution time. As a result, this architecture may only appeal to practitioners solving MDO problems with widely varying computational cost in the discipline analysis.
References


and ISSMO Symposium on Multidisciplinary Analysis and Optimization, 1996.


[62] J. Han and P. Y. Papalambros. A Sequential Linear Programming


[119] S. Parashar and C. L. Bloebaum. Robust Multi-Objective Genetic Algorithm Concurrent Subspace Optimization (R-MOGACSSO) for


[135] R. S. Sellar, S. M. Batill, and J. E. Renaud. Response Surface Based,


