Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
IDETC/CIE 2017
August 6-9, 2017, Cleveland, Ohio, USA

DETC2017-67676

UNIFIED SCALING OF DYNAMIC OPTIMIZATION DESIGN FORMULATIONS

Daniel R. Herber, James T. Allison
University of Illinois at Urbana-Champaign
Industrial & Enterprise Systems Engineering
Urbana, IL 61801
Email: {herber1,jtalliso}@illinois.edu

ABSTRACT

In this article, we explore scaling in dynamic optimization with a particular focus on how to leverage scaling in design studies. Here scaling refers to the process of suitable change of variables and algebraic manipulations to arrive at equivalent forms. The necessary theory for scaling dynamic optimization formulations is presented and a number of motivating examples are shown. The presented method is particularly useful for combined physical-system and control-system design problems to better understand the relationships between the optimal plant and controller designs. In one of the examples, scaling is used to understand observed results from more complete, higher-fidelity design study. The simpler scaled optimization problem and dimensionless variables provide a number of insights. Scaling can be used to help facilitate finding accurate, generalizable, and intuitive information. The unique structure of dynamic optimization suggests that scaling can be utilized in novel ways to provide better analysis and formulations more favorable for efficiently generating solutions. The mechanics of scaling are fairly straightforward but proper utilization of scaling is heavily reliant on the creativity and intuition of the designer. The combination of existing theory and novel examples provides a fresh perspective on this classical topic in the context of dynamic optimization design formulations.

Nomenclature

Acronyms

DAE	Differential-algebraic equations
DO	Dynamic optimization
ODE	Direct transcription
NLP	Nonlinear program
ODE	Ordinary differential equation
SASA	Strain-actuated solar array

Notation

$\bar{\square}$	Scaled quantity
$\dot{\square}$	Time derivative
$\square'$	Scaled derivative
$\square^*$	Optimal quantity

Variables

α	Linear term in a scaling law
β	Constant term in a scaling law
$\boldsymbol{C}$	Path constraints
ϵ	Constraint tolerance
e_r	Relative error
$\boldsymbol{f}$	State derivative function (dynamics)
H	Hamiltonian
$\mathcal{L}$	Lagrange (running cost) term in Ψ
$\boldsymbol{\lambda}$	Costates or Lagrange multipliers for $\boldsymbol{\xi}$
$\mathcal{M}$	Mayer (terminal cost) term in Ψ
O	Optimality conditions for P
$\boldsymbol{p}$	Time-independent optimization variables
P	Optimization problem
$\boldsymbol{\phi}$	Boundary constraints
$\boldsymbol{\rho}$	Problem parameters
$\boldsymbol{\xi}$	States
t	Time continuum
$\boldsymbol{u}$	Open-loop control variables
$\boldsymbol{x}$	Optimization variables
$\boldsymbol{\nu}$, $\boldsymbol{\mu}$	Lagrange multipliers for $\boldsymbol{\phi}$, $\boldsymbol{C}$

1 Introduction

Dynamics play an increasingly important role in the advancement of many complex engineering systems [1]. The primary goal of design studies is to find solutions and gain a general understanding of the design trade-offs, building design knowledge for the particular system. Here we show how scaling can facilitate finding accurate, generalizable, and intuitive information for the design problem at hand.

At a basic level, scaling is simply the stretching, squeezing, and shifting of the problem elements and this can include the time continuum, design variables, constraints, objective function, etc. For example, if we have the inequality constraint $ax \leq b$ and $b > 0$, then we can arrive at an equivalent scaled constraint $\rho x \leq 1$ where $\rho = a/b$. The mechanics of scaling are fairly straightforward but proper utilization of scaling is heavily reliant on the creativity and intuition of the designer [2]. This barrier may be one of the reasons why scaling is often overlooked, but these manipulations can help define problem formulations that are 1) better suited for analysis and 2) more favorable for solution methods (e.g., higher quality solutions and faster convergence). In this article, we provide the necessary theory to scale dynamic optimization (DO) problems and some examples of how to use scaling in the context of a design study.

First, we review some of the previous uses of scaling focusing on examples relevant to obtaining solutions and the analysis of engineering design problems. Some of these examples are quite well-established, while others are rarely used. The context that all the examples provide is crucial for defining the most useful scaling procedure for a particular design problem. Some authors state that the importance of scaling can only be fully appreciated through examples [3].

1.1 Previous Uses of Scaling

One of the primary uses of scaling is to reduce the number of parameters in a set of equations [2–4]. Under certain conditions, algebraic manipulations can lead to a reduced set of parameters (e.g., consider the example above where now ρ is the only parameter). Buckingham's Pi theorem is one well-known method for reducing the number of parameters, which applies constraints on the mathematical interaction of the fundamental units of the system and does not necessarily need a specific mathematical expression of the system (e.g., the equations of motion) [2, 3, 5]. However, this approach does not necessarily leverage the equations of motion nor does it directly consider the numerical aspects of finding approximate solutions to the system of equations.

Another common use for scaling is to determine characteristic properties of the system. These properties may be scalars or functions. Characteristic scalars are well studied in many engineering domains such as fluid dynamics and heat transfer [4]. An example of such a scalar is the Reynolds number, a dimensionless constant that relates inertial forces to viscous forces within a fluid [4]. Other examples include intrinsic resonance frequency, length, damping, or time constant. Such characteristic properties can be conceptually easier to understand [2]. These properties may be well understood for classical domains but for others, these characteristic properties may be the key to building the required knowledge of the system.

In dynamic or spatially-defined systems, it can be advantageous to scale entire continuous functions. For example, for an initial value problem, characteristic solutions to the differential equation can be obtained that scale linearly with the initial condition [2]. Under certain conditions, even the solution to optimization problems can be scaled for different parameter values. There are some potentially restrictive issues with directly scaling optimal solutions that are discussed in Ref. [6].

The central tool in many design studies is design optimization. Frequently, the solution of a single optimization problem is not sufficient to address the complicated nature of a design activity. This typically leads to variations on the problem formulation elements (e.g., parametric sweeps of the problem's parameters). Both reducing the number of parameters and obtaining scalable solutions can greatly improve this process.

Scaling can also be used to help decide if certain parts of a model or optimization formulation are small, and therefore negligible in a consistent manner [2,7]. This can help with developing asymptotic solutions to differential equations, reveal multiple-time-scale structures, and simplify simulations, stability analysis, and controller design [7]. With respect to optimization formulations, scaled forms of the constraints can help determine if certain constraints are likely to be active or inactive based on the magnitude of their scaled forms [8].

One common motivation for scaling in optimization is to change the order of magnitude of the variables and constraints to be more favorable for computation [8, 9]. Large order of magnitude differences in either the variables or function values can produce ill-conditioned matrices such as Jacobians and Hessians; thus, algorithmic calculations may become unstable or inefficient [8]. Systematic preconditioning methods have been developed to scale special cases of problem elements such as linear constraints [10, 11]. This has been observed to be especially important in DO to ensure robustness and accuracy [9, 12]. Some automatic scaling procedures have been developed to help alleviate some of the computational issues associated with solving dynamic optimization problems [9].

Scaling has also been directly included in some design studies. In Ref. [6], the author develops a method for assessing the importance of scaling laws in design optimization and approximate similitude metrics for obtaining dynamically similar optimal solutions. There have been a number of interesting examples of controller design that utilizes scaling, including Refs. [13, 14], but are typically limited to linear feedback controllers. Additionally, in Ref. [15], the authors utilized scaling to understand the trends found in the solutions to a more complete optimization formulation (and this example will be discussed in detail in Sec. 3.1).

Some final uses include performing scaled tests [2, 4] and checking for dimensional homogeneity [4]. All of the examples provide a rich history of scaling in engineering.

1.2 Dynamic Optimization

In this article, we consider design optimization problems that are well-posed as the following DO problem:

$$\min_{\boldsymbol{u},\boldsymbol{p}} \quad \int_{t_0}^{t_f} \mathcal{L}(t,\boldsymbol{\xi},\boldsymbol{u},\boldsymbol{p})\, dt + \mathcal{M}\big(\boldsymbol{\xi}(t_0),\boldsymbol{\xi}(t_f),\boldsymbol{u},\boldsymbol{p}\big) \tag{1a}$$

$$\text{subject to:} \quad \dot{\boldsymbol{\xi}} - \boldsymbol{f}(t,\boldsymbol{\xi},\boldsymbol{u},\boldsymbol{p}) = \boldsymbol{0} \tag{1b}$$

$$\boldsymbol{C}(t,\boldsymbol{\xi},\boldsymbol{u},\boldsymbol{p}) \le \boldsymbol{0} \tag{1c}$$

$$\boldsymbol{\phi}\big(\boldsymbol{\xi}(t_0),\boldsymbol{\xi}(t_f),\boldsymbol{p}\big) \le \boldsymbol{0} \tag{1d}$$

where the optimization variables are $\boldsymbol{p}$ (time-independent variables) and $\boldsymbol{u}$ (open-loop control variables), t is the time continuum defined between t_0 and t_f, and $\boldsymbol{\xi}$ are the states. The objective function in Eqn. (1a) is in Bolza form where $\mathcal{L}$ is the Lagrange (running cost) term and $\mathcal{M}$ is the Mayer (terminal cost) term. Equation (1b) enforces the dynamics modeled as a first-order ordinary differential equation (ODE), Eqn. (1c) enforces any time-varying path constraints, and Eqn. (1d) enforces any time-independent constraints.

A large number of engineering design problems can be posed in this form [1, 16, 17]. Some problems involve designing only plant design variables ($\boldsymbol{p}$ in the formulation above) such as tuning an automotive suspension [16, 18]. Other problems are control-only design (e.g., spacecraft trajectories [12] or batch reactor control [19]). Problems with both plant and control design are referred to as co-design problems [1, 16, 17, 20, 21] and have included the design of wave energy converters [16], active suspensions [21], and horizontal axis wind turbines [22].

There are two approaches for finding solutions to Prob. (1). Indirect methods utilize the optimality conditions of the infinite-dimensional DO problem [12, 19, 23]. It can be quite challenging (or impossible) to solve these equations analytically. Therefore, numerical methods are often employed that provide approximate solutions to the original problem. The alternative solution methods are known as direct methods [12, 19]. Instead of stating the optimality conditions, the control and/or state are parametrized using function approximation and the objective function is approximated using numerical quadrature. This creates a discrete, finite-dimensional problem that then is optimized using large-scale nonlinear programming (NLP) solvers [9, 12, 16, 19]. Scaling can be advantageousness for both methods.

The optimality conditions will be important when scaling DO formulations. The Hamiltonian for Prob. (1) is:

$$H = \mathcal{L} + \boldsymbol{\lambda}^{\mathsf{T}}\boldsymbol{f} + \boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{C} \tag{2}$$

where $\boldsymbol{\lambda}(t)$ are the costates (multipliers for the state dynamics) and $\boldsymbol{\mu}(t)$ are the multipliers for $\boldsymbol{C}$. The necessary conditions are stated as:

$$\dot{\boldsymbol{\lambda}}^* = -\left[\frac{\partial H}{\partial \boldsymbol{\xi}}\right]^* \tag{3a}$$

$$\boldsymbol{0} = \left[\frac{\partial H}{\partial \boldsymbol{u}}\right]^* \tag{3b}$$

$$0 = [\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{C}]^*, \quad 0 = [\boldsymbol{\nu}^{\mathsf{T}}\boldsymbol{\phi}]^*, \quad \boldsymbol{\mu}^* \ge \boldsymbol{0}, \quad \boldsymbol{\nu}^* \ge \boldsymbol{0} \tag{3c}$$

$$\boldsymbol{0} = \left[\boldsymbol{\lambda} + \frac{\partial \mathcal{M}}{\partial \boldsymbol{\xi}} + \boldsymbol{\nu}^{\mathsf{T}}\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\xi}}\right]^*_{t_0}, \quad \boldsymbol{0} = \left[\boldsymbol{\lambda} - \frac{\partial \mathcal{M}}{\partial \boldsymbol{\xi}} - \boldsymbol{\nu}^{\mathsf{T}}\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\xi}}\right]^*_{t_f} \tag{3d}$$

$$\boldsymbol{0} = \left[\frac{\partial \mathcal{M}}{\partial \boldsymbol{p}} + \boldsymbol{\nu}^{\mathsf{T}}\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{p}}\right]^*_{t_0} + \int_{t_0}^{t_f} \left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{p}} + \boldsymbol{\lambda}^{\mathsf{T}}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}} + \boldsymbol{\mu}^{\mathsf{T}}\frac{\partial \boldsymbol{C}}{\partial \boldsymbol{p}}\right]^* dt \tag{3e}$$

where $\boldsymbol{\nu}$ are the Lagrange multipliers for $\boldsymbol{\phi}$, Eqn. (3a) is the costate dynamics, Eqn. (3b) is the control stationarity condition, Eqn. (3c) are the complementary slackness conditions and the dual feasibility conditions, and Eqn. (3d) is the initial and final time transversality conditions, and Eqn. (3e) is the parameter stationary condition. Of course, the conditions in Eqn. (3) are in addition to the constraints Eqns. (1b)–(1d). Please refer to Ref. [17] for derivation and discussion of these conditions (including the variable horizon case).

2 Theory of Scaling Dynamic Optimization Formulations

Now the theory of scaling DO formulations is presented with examples to help illustrate the concepts. The basics are described first, which are applicable to sets of differential-algebraic equations (DAEs). Second is scaling in the context of optimization formulations.

2.1 Scaling Basics

The first step when scaling a set of equations is to introduce a change of variables. Here we consider linear scaling with:

$$x = \alpha_x \bar{x} + \beta_x \tag{4a}$$

$$y(x) = \alpha_y \bar{y}(x) + \beta_y \tag{4b}$$

where x is an independent variable, y is a dependent variable, $\{\bar{x}, \bar{y}\}$ are the new dimensionless variables, and $\{\alpha_x, \beta_x, \alpha_y, \beta_y\}$ are the user-defined scaling constants. The only restriction on the scaling constants is $\alpha \ne 0$ to avoid an ill-defined mapping between the scaled and original variables. Other types of scaling are possible [24] but the linear scaling rule often proves to be suitable for DO.

To substitute higher-order derivatives properly, we need to consider the chain rule and linearity of differentiation [25]:

$$\frac{d^n y}{dx^n} = \frac{\alpha_y}{\alpha_x^n}\frac{d^n \bar{y}}{d\bar{x}^n} \tag{5}$$

where n is the order of the derivative. If integrals are present, we can use integration by substitution when changing the integral limits [25]:

$$\int_{x_f}^{x_0} f(x)dx = \alpha_x \int_{\bar{x}_f}^{\bar{x}_0} f(\alpha_x \bar{x} + \beta_x) d\bar{x} \tag{6}$$

where $f(x)$ is some function and the integration limits have been shifted from $\{x_0, x_f\}$ to $\{\bar{x}_0, \bar{x}_f\}$. With this small set of formulas, we can apply scaling to all the problem elements of Prob. (1).

The original system of DAEs can have problem parameters, denoted $\boldsymbol{\rho}$. Through suitable choices of scaling variables, every system of DAEs with dimensional homogeneity (i.e., the dimensions on the left and right sides are the same) can be transformed into a dimensionless set of DAEs [2]. This will lead to the creation of dimensionless parameters, denoted $\bar{\boldsymbol{\rho}}$, and will prove quite important as they are critical to many of the uses discussed in Sec. 1.1. These dimensionless quantities can be the same as the dimensionless quantities derived by using Buckingham's Pi theorem [2,3,5]. Here we assume that all equations, inequalities, and inequations have dimensional homogeneity.

2.1.1 Example: Spring-Mass System

To illustrate the concepts of the previous section, we apply scaling to a simple spring-mass system. The first step is to write down the equations and assumptions:

$$m\ddot{y}(t) + ky(t) = 0 \tag{7a}$$

$$y(t_0) = y_0, \quad \dot{y}(t_0) = v_0 \tag{7b}$$

$$m, k > 0, \quad y_0 \neq 0 \tag{7c}$$

In this system, the independent variable is t and the dependent variable is y. Now consider the following change of variables based on simple scaling in Eqn. (4):

$$t = \alpha_t \bar{t} + \beta_t, \quad y(t) = \alpha_y \bar{y}(t) \tag{8}$$

The higher-order derivatives, using Eqn. (5), are then:

$$\frac{dy(t)}{dt} = \frac{\alpha_y}{\alpha_t}\frac{d\bar{y}(\bar{t})}{d\bar{t}}, \quad \frac{d^2y(t)}{dt^2} = \frac{\alpha_y}{\alpha_t^2}\frac{d^2\bar{y}(\bar{t})}{d\bar{t}^2} \tag{9}$$

We are free to choose the scaling constants so let's first look at the system of DAEs with the substitutions applied:

$$m\frac{\alpha_y}{\alpha_t^2}\frac{d^2\bar{y}(\bar{t})}{d\bar{t}^2} + k\alpha_y\bar{y}(\bar{t}) = 0 \tag{10a}$$

$$\alpha_y\bar{y}\left(\frac{t_0 - \beta_t}{\alpha_t}\right) = y_0, \quad \frac{\alpha_y}{c_t}\frac{d\bar{y}}{d\bar{t}}\left(\frac{t_0 - \beta_t}{\alpha_t}\right) = v_0 \tag{10b}$$

We can perform some algebraic manipulations to make the left-hand side of the equations have unity coefficients:

$$\frac{d^2\bar{y}(\bar{t})}{d\bar{t}^2} = -\frac{k\alpha_t^2}{m}\bar{y}(\bar{t}) \tag{11a}$$

$$\bar{y}\left(\frac{t_0 - \beta_t}{\alpha_t}\right) = \frac{y_0}{\alpha_y}, \quad \frac{d\bar{y}}{d\bar{t}}\left(\frac{t_0 - \beta_t}{\alpha_t}\right) = \frac{\alpha_t v_0}{\alpha_y} \tag{11b}$$

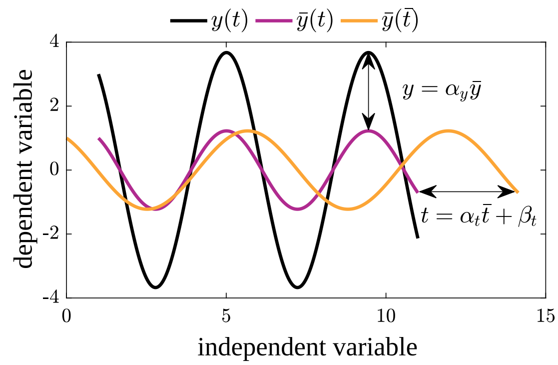


FIGURE 1: Spring-mass system with parameter values $k = 4$, $m = 2$, $y_0 = 3$, $v_0 = -3$, $t_0 = 1$, and $t_f = 11$.

Now, consider the following choice of the scaling constants:

$$\alpha_t = \sqrt{\frac{m}{k}}, \quad \beta_t = t_0, \quad \alpha_y = y_0 \tag{12}$$

Then the scaled system of DAEs is:

$$\frac{d^2\bar{y}(\bar{t})}{d\bar{t}^2} = -\bar{y}(\bar{t}) \tag{13a}$$

$$\bar{y}(0) = 1, \quad \frac{d\bar{y}}{d\bar{t}}(0) = \frac{v_0}{y_0}\sqrt{\frac{m}{k}} := \bar{\rho}_1 \tag{13b}$$

We see that this choice of constants results in a differential equation with unity coefficients. The original system had five parameters, but this scaled system now has only one, denoted $\bar{\rho}_1$, and is dimensionless. Furthermore, the initial position does not depend on any parameters. The scaling constant $\alpha_t = \sqrt{m/k}$ is typically referred to as the characteristic time constant for this system (also known as the reciprocal of the natural frequency).

Both the scaled and original solutions to this initial value problem are shown in Fig. 1. We see the stretching, squeezing, and shifting of the trajectory based on the scaling rules. The scaled trajectory might be more favorable to numerical approximation methods since the average magnitude of its derivative is closer to unity and consistent across the range of parameter values.

2.2 Scaling Dynamic Optimization Problems

We first denote the original problem formulation as P with optimality conditions O and optimal solution $\boldsymbol{x}^*$. Now the scaled formulation is denoted $\bar{P}$ with optimality conditions $\bar{O}$ and optimal solution $\bar{\boldsymbol{x}}^*$. For each problem, there may be problem parameters each denoted $\boldsymbol{\rho}$ and $\bar{\boldsymbol{\rho}}$, respectively. The optimization variables are related with the scaling function S: $\boldsymbol{x} = S(\bar{\boldsymbol{x}}, \boldsymbol{\rho})$ where for each optimization variable, a linear scaling law defined in Eqn. (4) exists that may depend on the problem parameters. Now for a given value of $\boldsymbol{\rho}$, the optimal solution to the two problems are clearly equivalent: if $\boldsymbol{x}^*$ solves P, then $\bar{\boldsymbol{x}}^* = S^{-1}(\boldsymbol{x}^*, \boldsymbol{\rho})$

solves $\bar{P}$; if $\bar{\boldsymbol{x}}^*$ solves $\bar{P}$, then $\boldsymbol{x}^* = S(\bar{\boldsymbol{x}}^*, \boldsymbol{\rho})$ solves P [24].

In some design studies, finding the optimal solution with respect to a single set of parameter values is sufficient to complete the design task. However, many require solutions that vary with respect to the parameters, i.e., $\boldsymbol{x}^*(\boldsymbol{\rho})$ for various values of $\boldsymbol{\rho}$. We can utilize the scaled problem in this task. If we are given two sets of parameters, $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$, such that $\bar{\boldsymbol{\rho}}_1 = \bar{\boldsymbol{\rho}}_2$, then we can determine the optimal solution with respect to $\boldsymbol{\rho}_2$ utilizing the solution found with $\boldsymbol{\rho}_1$:

$$\boldsymbol{x}^*(\boldsymbol{\rho}_2) = S(\bar{\boldsymbol{x}}^*(\bar{\boldsymbol{\rho}}_1), \boldsymbol{\rho}_2) \tag{14}$$

The condition that $\bar{\boldsymbol{\rho}}_1 = \bar{\boldsymbol{\rho}}_2$ is potentially restrictive [6], but it also can be quite useful. This implies that we only need to generate solutions for all relevant values of $\bar{\boldsymbol{\rho}}$ rather than $\boldsymbol{\rho}$, which is favorable since typically $\bar{\boldsymbol{\rho}}$ contains fewer elements that the original $\boldsymbol{\rho}$.

Additional scaling might be possible depending on what can be discerned from the activity of the inequality constraints [8] but will be problem dependent. If certain constraints are found to be inactive for particular ranges of the problem parameters and the reduced form of the original optimization problem eliminates any dependence on a particular parameter, then solutions may be scaled irrespective of that parameter (as long as the inactivity holds in the original problem). In DO, there may be infinite-dimensional path constraints where determining the activity can be challenging (one of the primary motivations for direct methods [17]). This may lead to solutions that are widely different for minor changes in the problem parameters, so scaling must be done carefully (this is shown in the example in Sec. 3.1).

To determine how the optimality conditions are related between the original and scaled problems, additional scaling constants for the multipliers need to be introduced. Then the correct values for these scaling constants must be determined such that a map is known between O and $\bar{O}$.

The relationships between the optimization formulations, optimality conditions, and optimal solutions for both the original and scaled problems are shown in Fig. 2. These relations will be discussed in the context of different paths that may be taken to obtain solutions to a particular DO problem. Consider first the standard case where the original problem is proposed and then an optimal solution is found (e.g., using DT and an NLP solver):

$$P \rightarrow \boldsymbol{x}^*$$

An alternative is to utilize the optimality conditions in Eqn. (3) to obtain an analytical solution or using a numeric indirect method:

$$P \rightarrow O \rightarrow \boldsymbol{x}^*$$

Now using scaling, we could instead transform to P to $\bar{P}$, and solve the scaled problem. Then the scaled solution can be mapped back to obtain the original problem's solution:

$$P \rightarrow \bar{P} \rightarrow \bar{\boldsymbol{x}}^* \rightarrow \boldsymbol{x}^*$$

The key here is that $\bar{\boldsymbol{x}}^*$ can be used to generate multiple $\boldsymbol{x}^*$ using Eqn. (14). One final approach that may yield the most in-

$$\begin{tikzcd}
P \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] \arrow[dd, bend right] & \bar{P} \arrow[d, leftrightarrow] \arrow[dd, bend left] \\
O \arrow[r, leftrightarrow] \arrow[d] & \bar{O} \arrow[d] \\
\boldsymbol{x}^* \arrow[r, leftrightarrow] & \bar{\boldsymbol{x}}^*
\end{tikzcd}$$

FIGURE 2: Relationships between the optimization formulations, optimality conditions, and optimal solutions for both the original and scaled problems.

formation from the scaling procedure is utilizing the optimality conditions:

$$P \rightarrow O \rightarrow \bar{O} \rightarrow \bar{P} \rightarrow \bar{\boldsymbol{x}}^* \rightarrow \boldsymbol{x}^*$$

With this approach, all of the suggestions in this section can be explored such as constraint activity.

The following section will use the theory in this section to illustrate scaling in DO with some motivating examples.

3 Motivating Examples

In this section, a number of examples of scaling in DO are presented. Some examples have direct application to existing design problems while others provide a more conceptual illustration. The examples in this section and the many uses described in Sec. 1.1 provide a strong foundation for using scaling in novel DO design problems.

3.1 Example 1: Simple SASA Problem

This first example was developed to better understand the observed trends from an existing design study. The co-design study focused on developing a novel strain-actuated solar array (SASA) system for spacecraft pointing control and jitter reduction [15]. In this system, distributed piezoelectric actuators were used to strain the solar arrays, causing reactive forces that could be used to control the bus (spacecraft body) with higher precision, higher bandwidth, and reduced vibrations.

The observed trend was an optimal constant ratio between the natural period of the first mode, T_1, and time allotted for performing the maneuver, t_f (which is visualized in Fig. 10 of Ref. [15]). There was much discussion on why this ratio seemed to be constant and if there was any significance to the value of this optimal ratio (approximately $T_1/t_f = 4.41$). To help provide some insight into these questions, a much simpler, but still representative, design problem was proposed. In Fig. 3, both the original and simplified SASA systems are visualized. The simplified system modeled many of the fundamental phenomena present in the original high-fidelity system using a small number of lumped parameters.

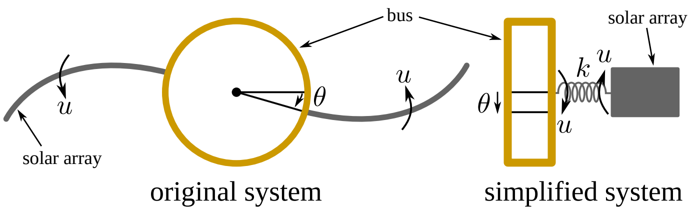


FIGURE 3: Illustrations of original and simplified strain-actuated solar array systems in Sec. 3.1.

The problem formulation for the simplified system is:

$$\min_{k,u(t)} \quad -\theta(t_f) \tag{15a}$$
$$\text{subject to:} \quad J\ddot{\theta}(t) + k\theta(t) = u(t) \tag{15b}$$
$$\theta(0) = \dot{\theta}(0) = 0 \tag{15c}$$
$$\dot{\theta}(t_f) = 0 \tag{15d}$$
$$|u(t)| \le u_{\max} \tag{15e}$$

where θ is the relative displacement of the bus, J is related to the inertia ratio between the solar arrays and bus, u is an open-loop control moment applied to the solar array and is bounded by $u_{\max}$, and k is the stiffness in the solar array. The boundary constraints enforce the system to start at rest with zero energy and end at rest. The optimization variables are both k and $u(t)$. Scaling can be applied to help analyze this DO problem.

3.1.1 Standard Form

First, we write the original DO problem in Eqn. (15) the standard form:

$$\min_{k,u(t)} \quad -\xi_1(t_f) \tag{16a}$$
$$\text{subject to:} \quad \dot{\boldsymbol{\xi}} = \begin{bmatrix} 0 & 1 \\ -k/J & 0 \end{bmatrix}\boldsymbol{\xi} + \begin{bmatrix} 0 \\ 1/J \end{bmatrix} u \tag{16b}$$
$$\phi_1 := \xi_1(0) = 0, \quad \phi_2 := \xi_2(0) = 0 \tag{16c}$$
$$\phi_3 := \xi_2(t_f) = 0 \tag{16d}$$
$$C_1 := u - u_{\max} \le 0, \quad C_2 := -u - u_{\max} \le 0 \tag{16e}$$
$$\text{where:} \quad \xi_1 = \theta, \quad \xi_2 = \dot{\theta} \tag{16f}$$

The Hamiltonian is then:

$$H = \lambda_1\xi_2 + \lambda_2\left(-\frac{k}{J}\xi_1 + \frac{u}{J}\right) + \mu_1(u - u_{\max}) + \mu_2(-u - u_{\max}) \tag{17}$$

Now using the conditions in Eqn. (3), the additional conditions for optimality are:

$$\dot{\lambda}_1 = \frac{k}{J}\lambda_2, \quad \dot{\lambda}_2 = -\lambda_1 \tag{18a}$$
$$0 = \frac{\lambda_2}{J} + \mu_1 - \mu_2 \tag{18b}$$
$$0 = \mu_1(u - u_{\max}),\ \mu_1 \ge 0, \quad 0 = \mu_2(-u - u_{\max}),\ \mu_2 \ge 0 \tag{18c}$$
$$0 = \nu_1\xi_1(0),\ \nu_1 \ne 0, \quad 0 = \nu_2\xi_2(0),\ \nu_2 \ne 0 \tag{18d}$$
$$0 = \nu_3\xi_2(t_f),\ \nu_3 \ne 0 \tag{18e}$$
$$0 = \lambda_1(0) - 1 + \nu_1, \quad 0 = \lambda_2(0) + \nu_2 \tag{18f}$$
$$0 = \lambda_1(t_f), \quad 0 = \lambda_2(t_f) - \nu_3 \tag{18g}$$
$$0 = \int_0^{t_f} \lambda_2 \frac{\xi_1}{J} dt \tag{18h}$$

3.1.2 Scaled Form

Consider the following change of variables based on linear scaling described in Sec. 2.1:

$$t = \alpha_t\bar{t}, \quad u(t) = \alpha_u\bar{u}(t), \quad \theta(t) = \alpha_\theta\bar{\theta}(t) \tag{19}$$

Next, the differential equation in Eqn. (15b) is parameterized as:

$$J\frac{\alpha_\theta}{\alpha_t^2}\frac{d^2}{d\bar{t}^2}\bar{\theta} + k\alpha_\theta\bar{\theta} = \alpha_u\bar{u} \Rightarrow \frac{d^2}{d\bar{t}^2}\bar{\theta} := \bar{\theta}'' = -k\frac{\alpha_t^2}{J}\bar{\theta} + \frac{\alpha_u\alpha_t^2}{J\alpha_\theta}\bar{u} \tag{20}$$

The three parameters in the problem are $\{J, u_{\max}, t_f\}$. There are three scaling constants to choose. The following points are considered when deciding how to scale the problem based on some previous intuition:

- Since all initial and final state conditions are zero, we should not start with α_θ
- The time horizon should be fixed, i.e., independent of t_f
- The magnitude of the control force should be unity to remove dependence on $u_{\max}$
- It is fine to have k multiplied by some constants since k is an optimization variable and is not directly constrained
- The natural period of this second-order system is $T = 2\pi\sqrt{J/k}$

With these considerations, we select the values of the scaling parameters as:

$$\alpha_t = \frac{t_f}{2\pi}, \quad \alpha_u = u_{\max}, \quad \alpha_\theta = \frac{u_{\max}t_f^2}{4\pi^2 J} \tag{21}$$

Note that the horizon is now fixed between 0 and 2π. Therefore, the scaled optimization problem is:

$$\min_{k,\bar{u}(\bar{t})} \quad -\frac{u_{\max}t_f^2}{4\pi^2 J}\bar{\theta}(2\pi) \tag{22a}$$
$$\text{subject to:} \quad \bar{\theta}'' = -\frac{kt_f^2}{4\pi^2 J}\bar{\theta} + \bar{u} \tag{22b}$$
$$\bar{\theta}(0) = \bar{\theta}'(0) = 0 \tag{22c}$$
$$\bar{\theta}'(2\pi) = 0 \tag{22d}$$
$$|\bar{u}(\bar{t})| \le 1 \tag{22e}$$

The scaled problem has two dimensionless quantities:

$$\bar{\rho}_1 = \frac{kt_f^2}{4\pi^2 J} \equiv \frac{t_f^2}{T^2}, \quad \bar{\rho}_2 = \frac{u_{\max}t_f^2}{4\pi^2 J} \tag{23}$$

First, we have $\sqrt{\bar{\rho}_1} = t_f/T$, the exact ratio we are investigating. We also have $\bar{\rho}_1$ as the only part of the formulation that depends on k and since there are no constraints on k, we are effectively designing the quantity $\bar{\rho}_1$ directly. Second, $\bar{\rho}_2$ is a positive constant linear factor in the objective function, so it will not affect the nature of the solutions [24], and it can be removed temporarily from the scaled formulation. Therefore, the formulation is equivalent to the following DO problem in the standard form:

$$\min_{\bar{\rho}_1, \bar{u}(\bar{t})} \quad -\bar{\xi}_1(2\pi) \tag{24a}$$

$$\text{subject to:} \quad \dot{\bar{\boldsymbol{\xi}}} = \begin{bmatrix} 0 & 1 \\ -\bar{\rho}_1 & 0 \end{bmatrix} \bar{\boldsymbol{\xi}} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \bar{u} \tag{24b}$$

$$\phi_1 := \bar{\xi}_1(0) = 0, \quad \phi_2 := \bar{\xi}_2(0) = 0 \tag{24c}$$

$$\phi_3 := \bar{\xi}_2(2\pi) = 0 \tag{24d}$$

$$C_1 := \bar{u} - 1 \leq 0, \quad C_2 := -\bar{u} - 1 \leq 0 \tag{24e}$$

The dependence on the parameters $\boldsymbol{\rho} = \{J, u_{\max}, t_f\}$ has been removed completely; therefore, finding the single solution, $\bar{\boldsymbol{x}}^*$, to this scaled formulation will give all solutions for any valid $\boldsymbol{\rho}$! We have also derived relationships to indicate how different values for these parameters directly affect the solution, i.e., the function S.

3.1.3 Equivalence of the Optimality Conditions

It is still illustrative to show that there is a linear mapping between the O and $\bar{O}$ for Probs. (16) and (24). The Hamiltonian for the scaled problem is:

$$\bar{H} = \bar{\lambda}_1 \bar{\xi}_2 + \bar{\lambda}_2 \left(-\bar{\rho}_1 \bar{\xi}_1 + \bar{u} \right) + \bar{\mu}_1 (\bar{u} - 1) + \bar{\mu}_2 (-\bar{u} - 1) \tag{25}$$

We expect some scaling of the multipliers between the two problems so we will define a scaling rule for each:

$$\begin{gathered} \lambda_1 = \alpha_{\lambda_1} \bar{\lambda}_1, \quad \lambda_2 = \alpha_{\lambda_2} \bar{\lambda}_2, \quad \mu_1 = \alpha_{\mu_1} \bar{\mu}_1, \quad \mu_2 = \alpha_{\mu_2} \bar{\mu}_2 \\ \nu_1 = \alpha_{\nu_1} \bar{\nu}_1, \quad \nu_2 = \alpha_{\nu_2} \bar{\nu}_2, \quad \nu_3 = \alpha_{\nu_3} \bar{\nu}_3 \end{gathered} \tag{26}$$

To determine the proper values of the scaling constants, we need to substitute the change of variables into the optimality conditions in Eqn. (18):

$$\frac{\alpha_{\lambda_1}}{\alpha_t} \dot{\bar{\lambda}}_1 = \frac{k}{J} \alpha_{\lambda_2} \bar{\lambda}_2, \quad \frac{\alpha_{\lambda_2}}{\alpha_t} \dot{\bar{\lambda}}_2 = -\alpha_{\lambda_1} \bar{\lambda}_1 \tag{27a}$$

$$0 = \alpha_{\lambda_2} \frac{\bar{\lambda}_2}{J} + \alpha_{\mu_1} \bar{\mu}_1 - \alpha_{\mu_2} \bar{\mu}_2 \tag{27b}$$

$$0 = \alpha_{\mu_1} \bar{\mu}_1 (\alpha_u \bar{u} - u_{\max}), \; \alpha_{\mu_1} \bar{\mu}_1 \geq 0 \tag{27c}$$

$$0 = \alpha_{\mu_2} \bar{\mu}_2 (-\alpha_u \bar{u} - u_{\max}), \; \alpha_{\mu_2} \bar{\mu}_2 \geq 0 \tag{27d}$$

$$0 = \alpha_{\nu_1} \bar{\nu}_1 \alpha_\theta \bar{\xi}_1(0), \; \alpha_{\nu_1} \bar{\nu}_1 \neq 0 \tag{27e}$$

$$0 = \alpha_{\nu_2} \bar{\nu}_2 \frac{\alpha_\theta}{\alpha_t} \bar{\xi}_2(0), \; \alpha_{\nu_2} \bar{\nu}_2 \neq 0 \tag{27f}$$

$$0 = \alpha_{\nu_3} \bar{\nu}_3 \frac{\alpha_\theta}{\alpha_t} \bar{\xi}_2(2\pi), \; \alpha_{\nu_3} \bar{\nu}_3 \neq 0 \tag{27g}$$

$$0 = \alpha_{\lambda_1} \bar{\lambda}_1(0) - 1 + \alpha_{\nu_1} \bar{\nu}_1, \quad 0 = \alpha_{\lambda_2} \bar{\lambda}_2(0) + \alpha_{\nu_2} \bar{\nu}_2 \tag{27h}$$

$$0 = \alpha_{\lambda_1} \bar{\lambda}_1(2\pi), \quad 0 = \alpha_{\lambda_2} \bar{\lambda}_2(2\pi) - \alpha_{\nu_3} \bar{\nu}_3 \tag{27i}$$

$$0 = \alpha_t \int_0^{2\pi} \alpha_{\lambda_2} \bar{\lambda}_2 \frac{\alpha_\theta \bar{\xi}_1}{J} d\tau \tag{27j}$$

With the substitution applied, Eqn. (27) above should match the optimality conditions for Prob. (24) exactly. This is accomplished with the following relationships:

$$\alpha_{\lambda_1} = \alpha_{\nu_1} = 1, \quad \alpha_{\lambda_2} = \alpha_{\nu_2} = \alpha_{\nu_3} = J\alpha_{\mu_2} = J\alpha_{\mu_2} = \alpha_t \tag{28}$$

Thus, a simple linear map exists between the optimality conditions between the original and scaled forms. The two Hamiltonians are related by[1]:

$$H = \alpha_\theta \alpha_t \bar{H} \tag{29}$$

3.1.4 Bounded Period

An additional inequality was necessary to explain further the observed results from the original study. This constraint was in the form of a bound on the period:

$$T = \frac{t_f}{\sqrt{\bar{\rho}_1}} \leq T_{\max} \tag{30}$$

which can be written as:

$$\frac{t_f^2}{T_{\max}^2} - \bar{\rho}_1 := \bar{\rho}_3 - \bar{\rho}_1 \leq 0 \tag{31}$$

where $\bar{\rho}_3$ is an additional dimensionless parameter. We can denote $\bar{\rho}_1^\dagger$ as the optimal value for $\bar{\rho}_1$ without the additional constraint in Eqn. (31). If $\bar{\rho}_3 < \bar{\rho}_1^\dagger$, then the constraint is inactive, and the previous solution is valid. However, if $\bar{\rho}_3 \geq \bar{\rho}_1^\dagger$, then the constraint is active and we would need to find solutions for every required value of $\bar{\rho}_3$ since the form of the optimal control will vary for each value of $\bar{\rho}_3$. In this problem, the control remains bang-bang in nature but the number of switches increases and the switching locations vary (see Fig. 4b).

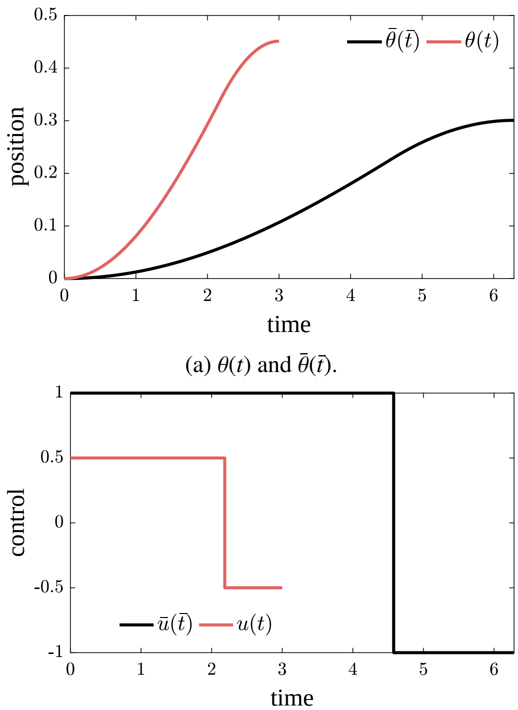
3.1.5 Solution

With the optimization problem, optimality conditions, and optimal solutions thoroughly characterized, we can compare this simple SASA problem to the original design study in Ref. [15]. Both the scaled and unscaled solutions for the simple SASA problem without the additional bound on the period being active are shown in Fig. 4. There are a number of similarities between the trajectories in Ref. [15] and in Fig. 4, including the bang-bang nature of the control when the period constraint is not active, and the general shape of the bus angle trajectories.

The primary comparison is Fig. 5, which includes the results from Fig. 10 of Ref. [15]. The figure plots the natural period of the first mode vs. t_f. There is an observed linear trend until a period limit seems to be reached. This is present for both design representations used: piecewise linear segments (PLS) and variable length (VL). Both of these representations can change the structural properties of the solar array. Although it is tough to see in the figure, there are coinciding data points at $t_f = 0.12$ s for both cases indicating that there is a similar optimal value for T_1 (and ratio).

Observing Fig. 5, there is a direct parallel between the data points and the simple SASA problem. For the scaled problem, $\bar{\rho}_1^\dagger = 0.0866$ while the results from Ref. [15] indicate $\bar{\rho}_1$ should be approximately 0.0513. The difference may be attributed to

[1]Using the objective $-\bar{\rho}_2 \bar{\theta}(2\pi)$, the relationship is $H = \alpha_\theta^2 \alpha_t \bar{H}$ with all multipliers containing an additional α_θ term.



(a) $\theta(t)$ and $\bar{\theta}(\bar{t})$.

(b) $u(t)$ and $\bar{u}(\bar{t})$.

FIGURE 4: Scaled and unscaled solutions for the simple SASA problem with $u_{\max} = 0.5$, $t_f = 3$, $J = 3$.

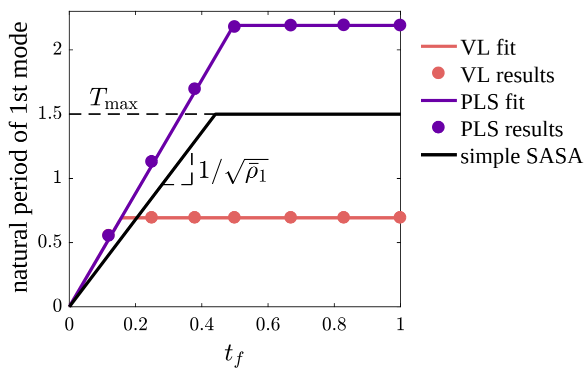


FIGURE 5: Natural period vs. t_f results for the simple SASA problem and original design study in Ref. [15].

the representation and constraints in the original design study. Initial discussions of the results tried to pin the ratio t_f/T_1 to $1/4$ due to the periodic resonance of a simple beam. However, the simple SASA problem suggests that this ratio is simply an arbitrary constant dependent on the interaction between the bang-bang controller and dynamics. Continuing with the comparisons, the $T_{\max}$ bound seems to equal about 2.19 s for the PLS case and 0.69 s for the VL case. If the physical-system constraints could be lifted, we have a reasonable prediction for how T_1 should vary.

This example illustrated how scaling can be utilized effectively to obtain insights in a design study.

3.2 Example 2: Co-Design Transfer Problem

In this example, we will use scaling to transform a co-design problem into a form with a known solution. Consider the following co-design problem:

$$\min_{k,u(t)} \quad \int_0^{t_f} u^2 dt \tag{32a}$$
$$\text{subject to:} \quad \ddot{y} = -ky + u \tag{32b}$$
$$y(0) = y_0, \quad \dot{y}(0) = v_0 \tag{32c}$$
$$y(t_f) = 0, \quad \dot{y}(t_f) = 0 \tag{32d}$$

where k is the time-independent, physical-system design variable and $u(t)$ is the open-loop control design variable. Through scaling, we can transform Prob. (32) above into the problem in Ref. [23, pp. 166–167]. To accomplish this, consider the following change of variables:

$$t = \alpha_t \bar{t}, \quad u = \alpha_u \bar{u} \tag{33}$$

We choose the following values for the scaling constants:

$$\alpha_t = \frac{1}{\sqrt{k}}, \quad \alpha_u = k \tag{34}$$

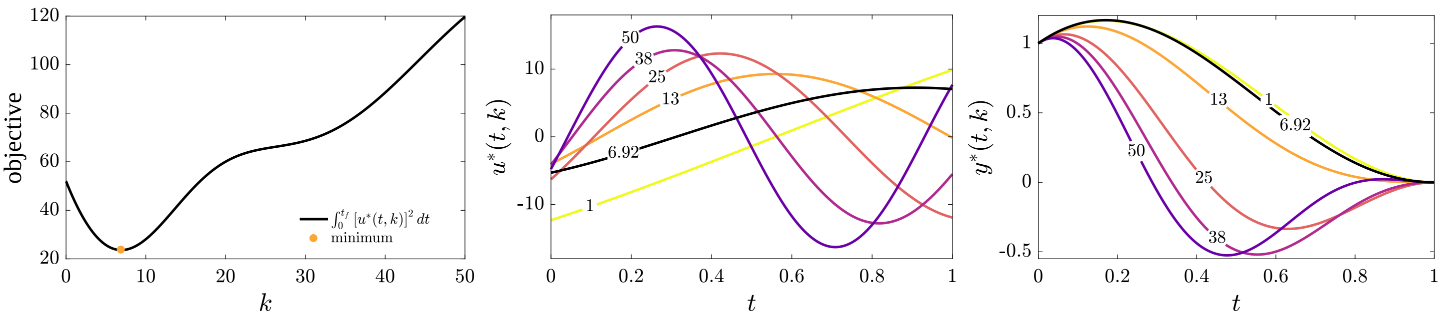
where $k > 0$. The case for $k = 0$ needs to be handled separately. Substituting in these scaling laws results in the following problem formulation:

$$\min_{k,\bar{u}(\bar{t})} \quad k^{3/2} \int_0^{\bar{t}_f} \bar{u}^2 d\bar{t} \tag{35a}$$
$$\text{subject to:} \quad y'' = -y + \bar{u} \tag{35b}$$
$$y(0) = y_0, \quad y'(0) = \bar{v}_0 \tag{35c}$$
$$y(\bar{t}_f) = 0, \quad y'(\bar{t}_f) = 0 \tag{35d}$$
$$\text{where:} \quad \bar{t}_f = \sqrt{k} t_f, \bar{v}_0 = v_0 / \sqrt{k} \tag{35e}$$

The solution to Prob. (35) above is readily available in Ref. [23]. Using this result and applying the scaling rules in Eqn. (33), we obtain the following solution for $u(t)$ as a function of k:

$$\begin{aligned} u^*(t,k) &= -\frac{2k}{k t_f^2 - \sin^2\left(\sqrt{k} t_f\right)} \left(c_1(t,k) x_0 + c_2(t,k) \frac{v_0}{\sqrt{k}} \right) \\ c_1(t,k) &= \sin\left(\sqrt{k}(t_f - t)\right) \sin\left(\sqrt{k} t_f\right) - \sqrt{k} t_f \sin\left(\sqrt{k} t\right) \\ c_2(t,k) &= -\cos\left(\sqrt{k}(t_f - t)\right) \sin\left(\sqrt{k} t_f\right) + \sqrt{k} t_f \cos\left(\sqrt{k} t\right) \end{aligned} \tag{36}$$

For certain problem parameter values, the solution for various values of k is shown in Fig. 6. The original objective function can now be computed analytically with Eqn. (36). With this closed-form expression for the objective function, an optimality condition for k can be derived. This property makes this problem quite suitable for investigations between nested and simultaneous co-design methods and can be used to illustrate a number of



(a) Objective function values.

(b) $u^*(t,k)$ for various values of k.

(c) $y^*(t,k)$ for various values of k.

FIGURE 6: Co-design transfer problem with $t_f = 1$, $x_0 = 1$, and $v_0 = 2$ (with $k^* = 6.924$).

concepts found in many co-design problems (such as open-loop control, multiple optima, degenerate plant designs, and a system-level objective function) [17].

For many DO problems, obtaining a closed-form solution for a portion of the design variables may not be possible. Observing different (but equivalent) forms of the same design optimization formulation can facilitate a better understanding of the design problem. In this example, we see that for larger values of k, the scaled problem has a larger time horizon but smaller initial velocity.

In addition, these different forms may be more suitable computation depending on the parameter values. In this example using the original Prob. (32) formulation, large values of k would result in highly oscillatory solutions. With the time-horizon fixed, large discretization errors would start appearing in the solutions with a fixed number of time points if a DT solution method was utilized. This helps motivate the final example.

3.3 Example 3: Direct Transcription Discretization

DT is a popular solution method for solving DO problems [1, 12, 19, 21]. In this approach, the dynamics are approximated with a large number of equality constraints, termed defect constraints. Equality constraints in a NLP are typically of the form $h(\boldsymbol{x}) = 0$. However, most numerical algorithms implement this constraint type as two inequality constraints as:

$$-\epsilon \le h(x) \le \epsilon \tag{37}$$

where ϵ is a small tolerance number (e.g., the default tolerance in MATLAB's `fmincon` function is $\epsilon = 10^{-6}$ [26]). In this example, we will demonstrate how the absolute magnitudes of the state variables can greatly degrade direct transcription approximations.

Consider the simple scalar system:

$$\dot{y} = ay, \quad y(0) = y_0 \tag{38}$$

where the solution to this set of DAEs is $\hat{y}(t) = y_0 e^{at}$. We can introduce a change of variables as:

$$t = \alpha_t \bar{t}, \quad y = \alpha_y \bar{y}$$

where the scaling constants are $\alpha_t = |a|$ and $\alpha_y = y_0$, resulting in the following scaled system:

$$\bar{y}' = \bar{y}, \quad \bar{y}(0) = 1 \tag{39}$$

We can now see how the relative error is influenced by the different systems.

A basic single-step DT method is the trapezoidal rule and the defect constraint that ensures accurate dynamics has the following form for a scalar ODE:

$$0 = y_k - y_{k-1} - \frac{h_k}{2}\Big(f(t_{k-1}, y_{k-1}) + f(t_k, y_k)\Big) \tag{40}$$

where y_k is the approximate value of $y(t)$ at t_k, $h_k = t_k - t_{k-1}$ is the step-size parameter, and $f(t_k, y_k)$ is the value of the derivative function at t_k.

We will consider the initial step, i.e., $k = 1$. Then the DT constraint for Eqn. (38) is:

$$0 = y_1 - y_0 - \frac{h_1}{2}(ay_0 + ay_1) \tag{41}$$

Using the tolerances in Eqn. (37), we can derive the minimum and maximum values for y_1 that still satisfy the constraint:

$$\frac{(1 + h_1 a/2)y_0 - \epsilon}{(1 - h_1 a/2)} \le y_1 \le \frac{(1 + h_1 a/2)y_0 + \epsilon}{(1 - h_1 a/2)} \tag{42}$$

Alternatively, the bounds for $\bar{y}_1$ are:

$$\frac{(1 + \mathrm{sign}(a)h_1/2) - \epsilon}{(1 - \mathrm{sign}(a)h_1/2)} \le \bar{y}_1 \le \frac{(1 + \mathrm{sign}(a)h_1/2) + \epsilon}{(1 - \mathrm{sign}(a)h_1/2)} \tag{43}$$

The relative error is defined as:

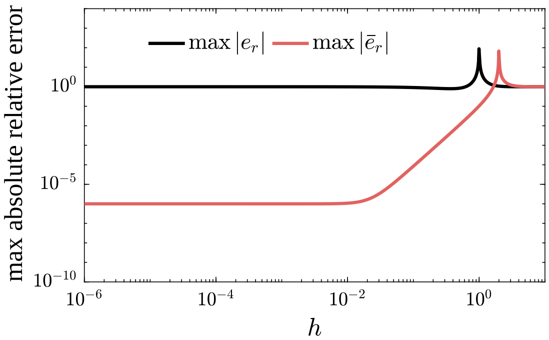
$$e_r = \frac{y_1 - \hat{y}_1}{\hat{y}_1} = \frac{y_1 - y_0 e^{at_1}}{y_0 e^{at_1}}, \quad \bar{e}_r = \frac{\bar{y}_1 - e^{\mathrm{sign}(a)t_1}}{e^{\mathrm{sign}(a)t_1}} \tag{44}$$

Now, consider the following values for the problem and DT parameters: $y_0 = 10^{-6}$, $a = 2$, $\epsilon = 10^{-6}$, $h = 10^{-3}$. For these values, we have the following relative error bounds:

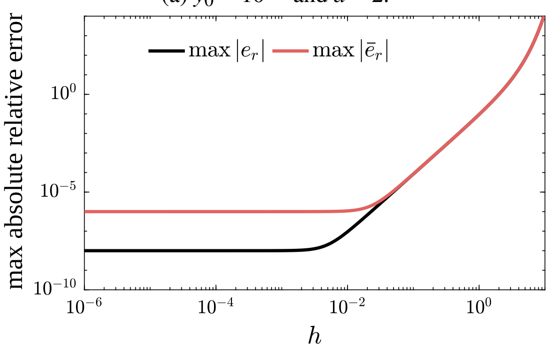
$$-9.99 \times 10^{-1} \le e_r \le 9.99 \times 10^{-1} \tag{45a}$$

$$-9.99 \times 10^{-7} \le \bar{e}_r \le 9.99 \times 10^{-7} \tag{45b}$$

We see that the e_r is quite poor using the original system while $\bar{e}_r$ is right around the expected ϵ tolerance. Consider one more



(a) $y_0 = 10^{-6}$ and $a = 2$.



(b) $y_0 = 10^2$ and $a = -1$.

FIGURE 7: Maximum absolute relative error vs. step size for original and scaled defect constraints.

choice of parameters: $y_0 = 10^3$, $a = -1$, $\epsilon = 10^{-6}$, $h = 10^{-3}$ and error bounds:

$$-1.08 \times 10^{-9} \le e_r \le 9.13 \times 10^{-10} \tag{46a}$$

$$-1.00 \times 10^{-6} \le \bar{e}_r \le 1.00 \times 10^{-6} \tag{46b}$$

Now we see that the relative error bounds are tighter than ϵ which could lead to numerical issues. The scaled form, however, is still right around the ϵ tolerance. The maximum absolute relative error vs. step size is shown in Fig. 7. From these plots, we see that for a sufficiently small step size, the scaled system has error bounds near ϵ.

Both of these parameter sets show the importance of scaling the dynamics to be near unity when using a DT implementation. Since DT implementations use a large number of linked defect constraints, the errors can compound, potentially producing highly inaccurate results. Additionally, scaling the time horizon provides a more uniform method for selecting the h to achieve suitable accuracy without too many optimization variables (note that $h = 10^{-3}$ for $\bar{e}_r$ is quite good in both plots). Finally, scaling the states to be near unity is also produces computationally favorable matrices (Hessian/Jacobians) and finite differencing [9, 12].

4 Conclusion

In this article, we explored scaling in DO with a particular focus on how to leverage scaling in design optimization. A review of the large amount of work around scaling was provided. The necessary theory for scaling DO formulations was presented and a number of novel motivating examples were provided. Scaling was shown to help facilitate finding accurate, generalizable, and intuitive information.

At a basic level, scaling is simply the stretching, squeezing, and shifting of the problem elements such as the time continuum, design variables, constraints, and objective function. The unique structure of DO suggests that scaling can be utilized in novel ways to provide better analysis and formulations more favorable for different solution methods. The mechanics of scaling are fairly straightforward but proper utilization of scaling is heavily reliant on the creativity and intuition of the designer. This nebulous qualification is the primary limitation on the use of scaling. However, scaling is also limited by the scaling law chosen, the structure of the problem, and the number of free parameters. Furthermore, scaling does not remove optimization; we still need to solve some form of the optimization problem. Improper use of scaling can lead to incorrect solutions/insights or amplification of numerical issues rather than mitigation.

Aiding the designer in constructing the appropriate scaled formulations needs compelling examples. In the simple SASA problem, scaling was used to understand observed results from more complete, higher-fidelity design study. The simpler scaled optimization problem and dimensionless variables provided a number of insights. In the second example, the solution to a co-design problem was found by leveraging a scaled formulation with the known solution. Observing different (but equivalent) forms of the same design optimization formulation can facilitate a better understanding of the design problem. In the final example, the discretization error of one of the popular solution methods (namely direct transcription) for DO was explored. The scaled system showed much more favorable properties than the original. Additional work is needed to develop general guidelines to aid in balancing the many uses of scaling.

References

[1] Allison, J. T., and Herber, D. R., 2014. "Multidisciplinary Design Optimization of Dynamic Engineering Systems". *AIAA Journal,* ***52***(4), Apr., pp. 691–710. doi: 10.2514/1.J052182

[2] Holmes, M. H., 2009. *Introduction to the Foundations of Applied Mathematics*. Springer, ch. Dimensional Analysis, pp. 1–42. doi: 10.1007/978-0-387-87765-5

[3] van Groesen, E., and Molenaar, J., 2007. *Continuum Modeling in the Physical Sciences*. SIAM, ch. Dimensional Analysis and Scaling, pp. 1–29.

[4] Çengel, Y., and Cimbala, J., 2006. *Fluid Mechanics Fundamentals and Applications*, 1st ed. McGraw Hill.

[5] Buckingham, E., 1914. "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations". *Physical Review,* ***4***(4), Oct., pp. 345–376. doi: 10.1103/PhysRev.4.345

[6] Kittirungsi, B., 2008. "A Scaling Methodology for Dynamic Systems: Quantification of Approximate Similitude and Use in Multiobjective Design". Ph.D. Dissertation, The University of Michigan.
[7] Khalil, H. K., 2002. *Nonlinear Systems*, 3rd ed. Prentice Hall.
[8] Papalambros, P. Y., and Wilde, D. J., 2017. *Principles of Optimal Design*, 3rd ed. Cambridge University Press.
[9] Rao, A. V., 2009. "A Survey of Numerical Methods for Optimal Control". In AAS/AIAA Astrodynamics Specialist Conference, no. AAS 09-334.
[10] Bergamaschi, L., Gondzio, J., and Zilli, G., 2004. "Preconditioning Indefinite Systems in Interior Point Methods for Optimization". *Computational Optimization and Applications,* ***28***(2), July, pp. 149–171. doi: 10.1023/B:COAP.0000026882.34332.1b
[11] Benzi, M., 2002. "Preconditioning Techniques for Large Linear Systems: A Survey". *Journal of Computational Physics,* ***182***(2), Nov., pp. 418–477. doi: 10.1006/jcph.2002.7176
[12] Betts, J. T., 2010. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. SIAM. doi: 10.1137/1.9780898718577
[13] Ghanekar, M., Wang, D., and Heppler, G., 1997. "Scaling Laws for Linear Controllers of Flexible Link Manipulators Characterized by Nondimensional Groups". *IEEE Transactions on Robotics and Automation,* ***13***(1), Feb., pp. 117–127. doi: 10.1109/70.554352
[14] Brennan, S., and Alleyne, A., 2001. "Robust Scalable Vehicle Control via Non-Dimensional Vehicle Dynamics". *Vehicle System Dynamics,* ***36***(4-5), Nov., pp. 255–277. doi: 10.1076/vesd.36.4.255.3551
[15] Chilan, C. M., Herber, D. R., Allison, J. T., Nakka, Y. K., Chung, S.-J., Aldrich, J. B., and Alvarez-Salazar, O. S., 2017. "Co-Design of Strain-Actuated Solar Arrays for Spacecraft Precision Pointing and Jitter Reduction". *AIAA Journal (to appear)*.
[16] Herber, D. R., 2014. "Dynamic System Design Optimization of Wave Energy Converters Utilizing Direct Transcription". M.S. Thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA, May.
[17] Herber, D. R., and Allison, J. T., 2017. "Nested and Simultaneous Solution Strategies for General Combined Plant and Controller Design Problems". In ASME International Design Engineering Technical Conferences, no. DETC2017-67668.
[18] Wang, Q., and Arora, J. S., 2009. "Several Simultaneous Formulations for Transient Dynamic Response Optimization: An Evaluation". *International Journal for Numerical Methods in Engineering,* ***80***(5), Oct., pp. 631–650. doi: 10.1002/nme.2655
[19] Biegler, L. T., 2010. *Nonlinear Programming*. SIAM. doi: 10.1137/1.9780898719383
[20] Fathy, H. K., Papalambros, P. Y., Ulsoy, A. G., and Hrovat, D., 2003. "Nested Plant/Controller Optimization with Application to Combined Passive/Active Automotive Suspensions". In American Control Conference, pp. 3375–3380. doi: 10.1109/ACC.2003.1244053
[21] Allison, J. T., Guo, T., and Han, Z., 2014. "Co-Design of an Active Suspension Using Simultaneous Dynamic Optimization". *ASME Journal of Mechanical Design,* ***136***(8), June, p. 081003. doi: 10.1115/1.4027335
[22] Deshmukh, A. P., and Allison, J. T., 2016. "Multidisciplinary Dynamic Optimization of Horizontal Axis Wind Turbine Design". *Structural and Multidisciplinary Optimization,* ***53***(1), Jan., pp. 15–27. doi: 10.1007/s00158-015-1308-y
[23] Bryson Jr., A. E., and Ho, Y.-C., 1975. *Applied Optimal Control*, revised ed. CRC Press.
[24] Boyd, S., and Vandenberghe, L., 2009. *Convex Optimization*, 7th ed. Cambridge University Press.
[25] Strang, G., 1991. *Calculus*, 1st ed. Wellesley-Cambridge Press.
[26] The MathWorks. fmincon. Online. Accessed on Feb. 6, 2017. https://www.mathworks.com/help/optim/ug/fmincon.html