Integrating topology optimization in precision motion system design for optimal closed-loop control performance

van der Veen, Gijs; Langelaar, Matthijs; van der Meulen, Stan; Laro, Dick; Aangenent, Wouter; van Keulen, Fred

DOI
10.1016/j.mechatronics.2017.06.003

Publication date
2017

Document Version
Accepted author manuscript

Published in
Mechatronics

Citation (APA)
van der Veen, G., Langelaar, M., van der Meulen, S., Laro, D., Aangenent, W., & van Keulen, F. (2017). Integrating topology optimization in precision motion system design for optimal closed-loop control performance. *Mechatronics*, *47*, 1-13. https://doi.org/10.1016/j.mechatronics.2017.06.003

Available online at www.sciencedirect.com
ScienceDirect
Mechatronics 00 (2017) 1–21
Mechatronics

Integrating topology optimization in precision motion system design for optimal closed-loop control performance

Gijs van der Veen[a,c], Matthijs Langelaar[a], Stan van der Meulen[b], Dick Laro[c], Wouter Aangenent[b], Fred van Keulen[a]

[a]*Delft University of Technology, Mekelweg 2, 2628 CD, Delft, The Netherlands*
[b]*ASML Research Mechatronics & Control, P.O. Box 324, 5500 AH Veldhoven, The Netherlands*
[c]*MI-Partners BV, Dillenburgstraat 9N, 5652 AM, Eindhoven, The Netherlands*

Abstract

In pursuit of better accuracy, higher speed and larger scale, manufacturers of high-performance devices increasingly rely on components which have been designed with a multidisciplinary approach from the outset. In the context of motion systems, this means that for instance structural mechanics, control engineering and thermal analysis are considered early in the design. In addition, the prospect of producing freeform device components using additive manufacturing at full scale allows designers to even further refine components to a specific purpose, or even integrate multiple functions into a single component. The design freedom offered by additive manufacturing is far greater than that offered by traditional techniques. To exploit this freedom a topology optimization framework is proposed that allows to determine the optimal material quantity and distribution within a design volume. In particular, this article focuses on the closed-loop control performance of a motion system component, while simultaneously ensuring that mechanical requirements are met. Based on an example, it is demonstrated that this leads to nontrivial and non-intuitive designs which provide improved performance at lower structural mass compared to eigenfrequency designs. The framework allows rapid development of prototype designs, which may eliminate some of the costly design iterations which are currently made in industrial practice.

Keywords: Topology optimization, motion systems, mechatronics, integrated design, closed-loop performance, design sensitivity analysis
2010 MSC: 74S05, 74P05, 74P15, 49Q12, 70Q05

1. Introduction

The next generation of high-performance devices relies on components which have been designed with a multidisciplinary approach from the outset [1, 2]. In the context of motion systems, this means that for instance structural mechanics, control engineering and thermal analysis are already considered in the initial design. Tight demands on precision, scale and speed have made this timely multidisciplinary approach a necessity, since later on in the design process unfavorable characteristics of the design may have become hard to rectify. Furthermore, recent developments in additive manufacturing allow refining components even further by placing material only where it is functional. The design freedom offered by additive manufacturing is far greater than that offered by traditional techniques. To exploit this freedom a topology optimization framework is proposed that allows to determine the optimal material distribution within a design volume. Topology optimization [3] is already established as a design tool in, e.g., the aerospace and automotive industries, where it is mostly used to optimize components for minimal mass subject to strength and

stiffness requirements. Nevertheless, significant developments are required to bring these techniques to a level where industry can apply them to complex multidisciplinary design problems with more sophisticated performance metrics.

For controlled flexible structures it has long been recognized that it makes sense to design the structure and controller simultaneously, rather than designing the controller as a second step [1, 2]. Hence, simultaneous design of structures and controllers has received attention throughout the last three decades. Control design must necessarily account for internal flexibilities in the cases of large-scale or high-precision systems. This was first explored for large-scale spacecraft where the focus was often on analytical multivariable control techniques in combination with discretely parameterized structures. The latter refers to structures with only a few design variables (sizes, cross-sections) and truss structures. These optimal control frameworks optimize criteria in the time domain which relate to actuator activity, kinetic energy and/or strain energy. Examples are found where these problems are solved using gradient-based optimization [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] and randomized algorithms [14, 15]. The latter do not lend themselves to cases with many design parameters as they scale poorly [16].

A limitation of many of these approaches is that only a few structural parameters are considered, allowing limited design freedom and certainly not topology design, which typically involves many thousands of design variables. A further practical limitation of these efforts is that the considered control frameworks (LQ, LQR, etc...) have seen very little application in industrial practice due to complexity and robustness issues. They are rarely used in the frequency-domain control design practice for motion systems [1].

A few examples exist in literature in which topology optimization is coupled with control design [17, 18, 19, 20]. Silveira and Fonseca [20] propose the use of control performance a secondary objective. Zhu et al. [18, 19] propose a nested optimization approach and also consider actuator placement is also considered. The controller is optimized in a nested loop using standard tools, however no details are revealed as to how design gradient information was obtained and whether this was done in a consistent manner. Accurate design gradient information is essential for efficient and stable convergence in topology optimization [3].

To develop a feasible approach which simultaneously incorporates controller and topology design, this article focuses on a rather simple controller framework. Use is made of static (rigid-body) decoupling of the system in combination with independent, fixed-structure PID controllers [21]. In fact, this is a fairly standard approach for motion systems. Instead, complexity in terms of design freedom is offered by coupling the framework to topology optimization, which provides great freedom with regards to the material distribution [3], making it a natural counterpart to additive manufacturing. A frequency-domain approach to the integrated design of control system and structure was previously developed in [22] and to the best of our knowledge that was the first attempt in literature. Recent years have seen an increased interest in integrated optimization in the field of mechatronics [23, 24], but no systematic approaches are available for joint topology and frequency-domain controller optimization.

A heuristic that could be used for motion systems is to design structures for maximum natural frequencies [25, 26]. From the point of view of control design these should usually be as high as possible. This criterion by itself, however, does not convey any information about mode *shapes* and their effects on the frequency response functions seen by the controller. Therefore, this approach may lead to conservative or poorly balanced designs because determining the relevant modes a priori is very hard.

It is unlikely that optimizing for closed-loop performance alone will completely define a mechanical design, since it only involves the dynamical behavior between actuators and sensors. This situation may be improved by distinguishing between point of control (e.g., a sensor) and point of interest (e.g., location on a substrate), but it is likely the case that other mechanical requirements must be taken into account to fully define the design. Thus, it makes sense to include most other mechanical requirements from the start, since the design needs to satisfy them anyway.

This article considers the topology design of a high-precision motion stage with three degrees of freedom. There is a strong link between the dynamics of the stage and the achievable closed-loop performance (such as bandwidth) and the stage cannot be modelled as a rigid body. A topology optimization formulation is developed which directly designs for closed-loop performance (based on first ideas discussed in [22]) subject to closed-loop stability requirements. The stage is simultaneously required to satisfy stiffness requirements with regards to the deformations that occur during motions and accelerations. A final important consideration is that the dependence of the dynamics on the operating point is taken into account: depending on the position of the stage, the sensor locations change and thereby the observed dynamics.

Compared to our previous work in [22], which was restricted to a single-input-single-output system with only one control loop and 2 control parameters, now multiple-input-multiple-output (MIMO) systems with rigid-body decou-

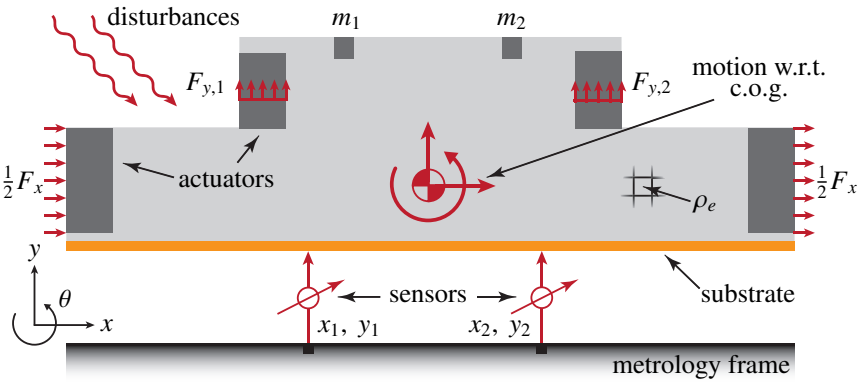


Figure 1: Simplified representation of a free-floating motion stage which can be controlled in three planar degrees of freedom (DoF). Each of the two sensors measures x and y displacements. Two actuators are available in each of the x and y directions to provide 3-DoF controllability. The light gray area may be modified by topology optimization.

pling and several control loops are considered. Also, the design process is set-up so that the control system functions when the sensor location changes. Finally, mechanical requirements on the design are added in this framework.

Next to presenting an integrated topology optimization formulation, the main contributions of this article comprise systematic and consistent design sensitivity analysis, which is crucial to the success of gradient-based optimization. This involves:

- Design sensitivity analysis of reduced order modal models of detailed finite element models;
- Design sensitivity analysis of metrics related to closed-loop MIMO transfer functions;
- Design sensitivity analysis of the response to static and self-equilibriated loads;
- Actuator masses which scale with the mass of the overall design;
- Control design which is robust against varying sensor locations;
- Design-dependent static decoupling of rigid-body motions.

In this article all these aspects are incorporated into a single optimization framework to address realistic problems.

1.1. Outline and guide to readers

Sections 2 and 3 describe the modelling and optimization procedure and are sufficient to comprehend the essential ideas behind the methodology and the case study in Sec. 5. The interested reader can (selectively) consult Sec. 4, which presents in more detail the design sensitivity analysis which is at the heart of the optimization framework.

2. Setup and modelling

A simplified 2D model of a positioning stage is considered, as shown in Fig. 1. This model represents a component in a precision device and is treated as a flexible body, suspended and positioned by actuator forces F_x and F_y. The forces F_x always act simultaneously, whereas $F_{y,1}$ and $F_{y,2}$ can be applied separately, leading to three independent force inputs. Furthermore, there is a contactless measurement setup. It is assumed that the coordinate x is measured as the average between the two sensors, whereas the coordinates y_1 and y_2 are measured independently. Thus, there are also three independent measurements. Together, the inputs and outputs allow positioning of the stage in any of the three rigid-body degrees of freedom (defined as x, y and θ, cf. Fig. 1) and allow unique measurement of this position. Note that depending on the position of the stage with respect to the metrology frame, the sensors see different points on the substrate, giving rise to *position-dependent dynamics* (see Secs. 2.1 and 3). Within a particular control task, however, the stage can be considered to be at a fixed x-position as far as the position-dependent dynamics

are concerned (small motion ranges about a given position are assumed), so that the local dynamics are linear and time-invariant.

Each of the actuators is defined by a mass associated to the dark gray areas in the diagram. The mass of each actuator depends upon the total mass of the stage (see Sec. 2.7), as a result of actuator scaling to achieve the desired accelerations. Finally, there is a substrate layer which has different material properties.

2.1. Mechanical modelling

The system is modelled using a finite element (FE) discretization. It is assumed that the actuator forces are uniformly distributed throughout the actuator bodies. Mathematically, the dynamics can be written as:

$$M(\theta)\ddot{x} + K(\theta)x = B\tilde{u}, \tag{1a}$$

$$\tilde{y} = Cx, \tag{1b}$$

in which $M(\theta)$ and $K(\theta)$ are the mass and stiffness matrices which depend on the design θ, x is the vector of nodal displacements, B is a collection of 3 normalized actuation vectors corresponding to 1 N of actuator force, $\tilde{u} \in \mathbb{R}^{n_u}$ contains the inputs (forces), C is a matrix of 3 normalized sensing vectors and $\tilde{y} \in \mathbb{R}^{n_y}$ is a vector of outputs. Note that C is itself a function of the stage x-position, due to the relative motion between the stage and the metrology frame.

The dynamic behavior of the stage is approximated by a reduced-order model which is based on the first n modes of the structure (including the rigid-body modes), using a mode truncation approach. These dynamics can be written as:

$$\ddot{\eta} + 2Z\Omega\dot{\eta} + \Omega^2\eta = \Phi^T B\tilde{u}, \tag{2a}$$

$$\tilde{y} = C\Phi\eta. \tag{2b}$$

The matrices have the following structure:

$$\Omega = \mathrm{diag}\,(\omega_1,\ \omega_2,\ \ldots,\ \omega_n),$$
$$\Phi = \begin{bmatrix}\phi_1 & \phi_2 & \cdots & \phi_n\end{bmatrix},$$
$$Z = \mathrm{diag}\,(\zeta_1,\ \zeta_2,\ \ldots,\ \zeta_n),$$

in which (ω_i, ϕ_i) is the i-th eigenpair and ζ_i is the corresponding modal damping. Note that modal damping has been introduced, which is common in the modelling of high-performance mechatronic systems [1]. Damping ratios of 0.5% are assumed in this work. The transfer function related to this model (the Laplace transform of (2)) is denoted by $P(s)$:

$$P(s) = C\Phi\left(s^2 I + 2Z\Omega s + \Omega^2\right)^{-1}\Phi^T B. \tag{3}$$

2.2. Rigid-body decoupling

Static decoupling of rigid-body motions is obtained by computing pre and post compensation matrices $W_\mathrm{A} \in \mathbb{R}^{n_u \times n_u}$ and $W_\mathrm{S} \in \mathbb{R}^{n_y \times n_y}$ [21] to result in a statically decoupled plant $G(s)$:

$$G(s) = W_\mathrm{S} P(s) W_\mathrm{A}. \tag{4}$$

This means that the compensated system has a new set of input and output signals $u = \begin{bmatrix}F_u & F_y & F_\theta\end{bmatrix}$ and $y = \begin{bmatrix}x & y & \theta\end{bmatrix}$[1], for which it holds that a constant input to each of the inputs gives a constant rigid-body acceleration, which affects only the corresponding output. The advantage is that this makes it possible to design independent single-input-single-output (SISO) controllers thanks to limited interaction between the control loops.

[1] For simplicity it is assumed that the system is neither under nor over-actuated regarding rigid-body motions, such that W_A and W_S are square and unique.

Considering only rigid-body modes Φ_r and using (2), equation (4) can be rewritten as:

$$\ddot{x} = \Phi_\mathrm{r}\Phi_\mathrm{r}^T B W_\mathrm{A} u,$$
$$\ddot{y} = W_\mathrm{S} C \Phi_\mathrm{r}\Phi_\mathrm{r}^T B W_\mathrm{A} u$$

Then, two conditions are imposed. First, the independent inputs should only activate independent user-defined motions Φ_d[2] as seen from the outputs, expressed by $C\ddot{x} \equiv C\Phi_\mathrm{d} u$. This leads to:

$$W_\mathrm{A} = (C\Phi_\mathrm{r}\Phi_\mathrm{r}^T B)^{-1} C\Phi_\mathrm{d}. \tag{5}$$

Second, a normalization condition is imposed such that the compensated rigid-body system behaves as a unit mass with an identity inertia tensor, independent of the design. This simplifies controller tuning and is expressed as $\lim_{s\to 0} s^2 G(s) = I$, which ultimately leads to:

$$W_\mathrm{S} = (C\Phi_\mathrm{d})^{-1}.$$

2.3. Controller

In this article the focus is not on exploring state-of-the-art controller synthesis procedures. Rather, the focus is on combining topology optimization and controller tuning into a single framework. Hence, a simple but industry-standard control configuration is chosen, in which each rigid-body motion $G_{ii}(s)$ is controlled by a PID controller of the following type (Fig. 2):

$$C_i(s;k,\omega_\mathrm{b}) = k\frac{s+\frac{1}{5}\omega_\mathrm{b}}{s}\frac{3s+\omega_\mathrm{b}}{s+3\omega_\mathrm{b}}\frac{5\omega_\mathrm{b}}{s+5\omega_\mathrm{b}}. \tag{6}$$

This controller structure is often applied in the field of motion systems [27, 1] (e.g., systems with a double integrator) and provides integral action up to $\frac{1}{5}\omega_\mathrm{b}$, phase lead between $\frac{1}{3}\omega_\mathrm{b}$ and $3\omega_\mathrm{b}$ and first order roll-off beyond $5\omega_\mathrm{b}$ (Fig. 3). This leads to sufficient phase lead in the crossover region, low frequency disturbance rejection and limited high-frequency control action.

Note that the methodology described in this article is not limited to the PID control scheme described here, but in principle can treat any parameterized and differentiable controller which has a frequency-domain representation.

For purposes of optimization the controller parameters k and ω_b are scaled such that their expected ranges lie approximately between 0 and 1.

2.4. Closed-loop performance and constraints

The closed-loop disturbance sensitivity function is defined as [28, 21]:

$$S(s;\theta) = [I + L(s;\theta)]^{-1}, \tag{7}$$

[2] Φ_d is further defined in Sec. 4.1.1, but usually contains intuitive unit rigid-body motions such as uniform x and y translation and rotation about the z-axis.

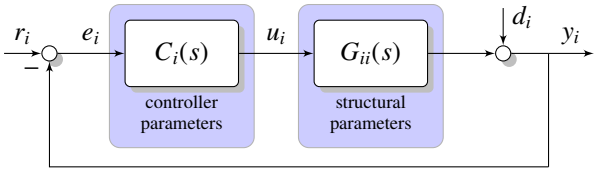


Figure 2: Closed-loop configuration of one of the statically decoupled control loops. The output y_i is regulated to track a reference signal r_i. The controller uses the error signal e_i and compensates tracking errors due to disturbance signal d_i by generating a control signal u_i.

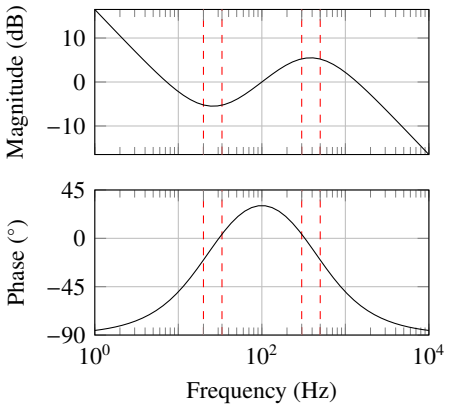


Figure 3: Frequency response of the PID controller (6) for $k = 1$ and $\omega_b = 100\,\text{Hz}$. The dashed lines indicate the frequencies $\frac{1}{5}\omega_b$, $\frac{1}{3}\omega_b$, $3\omega_b$ and $5\omega_b$.

in which L is the loop transfer function $L = GC$. Also, the dependence upon controller parameters and structural parameters, gathered in the vector θ, has been made explicit. The sensitivity function is an n_y-by-n_y transfer function matrix, and in this article the focus is on the diagonal elements. As described in Sec. 2.2, the plant P has been converted to a statically decoupled plant G. In that case the diagonal elements of $S(s;\theta)$ describe the sensitivity of each of the rigid-body motions to a perturbation of the same motion. Cross-sensitivities have not been considered, but an important consideration when examining cross-sensitivities is that these are not necessarily dimensionless and normalised, a fact which holds by definition for the diagonal components.

Given the simple controller structure employed (see Sec. 2.3), the diagonal elements of the sensitivity function will have a magnitude response which is similar to Fig. 4. The low-frequency asymptotic behavior, with a $60\,{}^{\text{dB}}/_{\text{decade}}$ slope, motivates the use of a single frequency as performance indicator. That is, the magnitude at a single frequency, say 10 Hz, will be indicative of performance (i.e., rejection of output disturbances) across the low-frequency range.

In SISO controller design, a sensitivity function which does not exceed an amplification of 6 dB implies that the loop gain remains outside the hatched circle in the Nyquist plot of the loop gain (see Fig. 5) [28] and implies a gain margin $GM \geq 6\,\text{dB}$ and a phase margin $PM \geq 29^\circ$ [21][3]. This still holds if each controller is tuned with the other loops closed.

A baseline configuration is created using a sequential loop closure approach [21], in which controllers are sequentially designed for the three rigid-body motions, based on the initial dynamics of the stage. It is important to realise that subsequently, during optimization, the controllers are simultaneously updated and the performance in all three loops is evaluated simultaneously. In that sense the procedure is more favourable than a sequential redesign, since each of the loops is tuned in the presence of the other control loops (all loops remain closed). This also means that the influence of each controller's parameters on the performance and constraints in all the other loops is accounted for.

2.4.1. Performance

The scalar performance function is a weighted combination of the magnitudes on the diagonal of S, i.e.:

$$f_0(\theta) = \sum_{i=1}^{n_y} w_i \, |S_{ii}(j\omega^\star;\theta)|_{\text{dB}}, \tag{8}$$

in which w_i is a weight attached to the i-th control loop and $\omega^\star$ denotes the frequency at which the performance is evaluated. The notation $|\cdot|_{\text{dB}} = 20\log_{10}|\cdot|$ refers to the complex modulus converted to decibels, usually a negative value. Decibels are used to accomodate the potentially large dynamic range of $|S_{ii}(j\omega^\star;\theta)|$. Each component $|S_{ii}(s^\star;\theta)|_{\text{dB}}$ may have a rather different initial value, so that a reasonable initial choice is to take w_i equal to the reciprocal of

[3] Except in special cases, see [29].

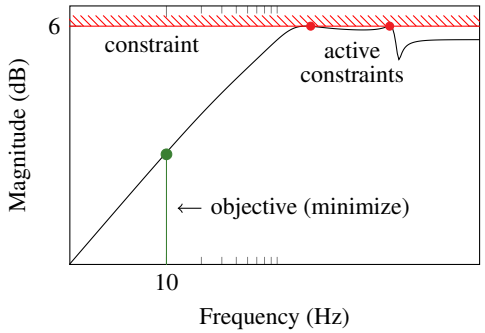


Figure 4: Typical shape of the disturbance sensitivity for one control loop. Attenuation is optimized at 10 Hz and amplification must remain below 6 dB everywhere.

$||S_{ii}(s^\star;\theta_0)|_{\text{dB}}|$ for the initial design θ_0. In that way, $f_0(\theta_0)$ sums to n_y for the initial parameter set and each sensitivity contributes equally to the objective in a relative sense.

While the disturbance sensitivity has been used here as a specific example of a performance measure, any open or closed-loop transfer function, or a combination of those, could be used. This means that a wide range of problems can be addressed using the generic methodology presented here, as long as a limited number of frequencies (in this case only $\omega^\star$) needs to be considered.

2.4.2. Constraints

For each diagonal element of $S(s)$, the constraint $|S_{ii}(j\omega)| \leq S_{\max}$ is enforced (often $S^{\text{dB}}_{\max} = 6\,\text{dB}$). To achieve this, the frequency response is scanned for all peaks (using the method described in [30]) and each of these peaks is monitored. The largest possible number of peaks in a sensitivity frequency response is related to the number of modes. Each of these constraints can be formulated as:

$$f_c(\theta) = |S_{ii}(j\omega_k;\theta)|_{\text{dB}} \leq S^{\text{dB}}_{\max}, \quad \begin{matrix} i = 1,\ldots,n_y \\ k = 1,\ldots,n_f^i, \end{matrix} \tag{9}$$

in which n_f^i is the number of peaks detected in the i-th sensitivity function and ω_k is the frequency at which a peak occurs.

Note that as the design (i.e., θ) changes and the optimization proceeds, the number of peaks in each transfer function may change. Hence, the set of constraints (9) can be a set which changes in size. This also means that these constraints are not globally continuously differentiable. Still, the local gradient gives information on feasible design changes.

A further constraint is that all closed-loop eigenvalues should be in the open left-half plane. Assume that state-space realizations of the reduced-order decoupled system (4) as well as the combined controllers (6) are given as:

$$G(s): \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \qquad C(s): \begin{cases} \dot{x}_c = A_c x_c + B_c y \\ u = C_c x_c \end{cases}$$

It can be shown that the closed-loop system matrix for a unity negative feedback loop between $G(s)$ and $C(s)$ satisfies:

$$\Psi = \begin{bmatrix} A & BC_c \\ -B_c C & A_c \end{bmatrix}. \tag{10}$$

All eigenvalues of this matrix must have negative real parts, which can be expressed as a further constraint:

$$f_\lambda(\theta) = \psi^p_{\max}\left(\mathcal{R}e\left\{\lambda_i(\Psi)\right\}\right) < 0, \quad i = 1,\ldots,n_\lambda. \tag{11}$$

In this equation $\psi^p_{\max}(\cdot)$ denotes a smooth approximation to the maximum function (to be defined in Sec. 4.6) [31, 32], which is used in order to aggregate individual eigenvalue constraints into a single constraint imposed upon the (approximate) largest eigenvalue.

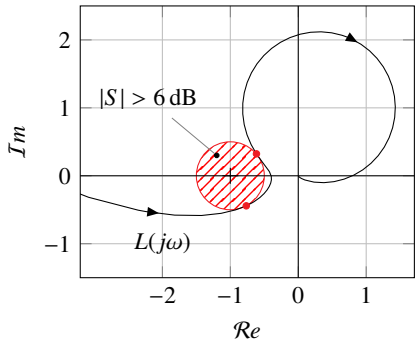


Figure 5: A typical Nyquist plot of a control system showing $L(j\omega)$, the critical point $(-1, 0)$ and the region $|S(j\omega)| > S_{\max}$.

2.5. Mechanical requirements

Since the stage is used for very precise positioning, it must meet several mechanical stiffness requirements in addition to providing optimized closed-loop positioning performance at the points of control (i.e., the measurement locations).

To achieve peak uniform acceleration $\ddot{x}_{\mathrm{u}}$ of the stage, certain (design-dependent) actuator forces $f_{\mathrm{accel}}(\theta)$ are applied which are in equilibrium with the inertia forces of the accelerating stage according to d'Alembert's principle. This causes a static deformation of the stage. Under these conditions the shear stress between the stage and the substrate may not exceed a certain level to avoid damage. This constraint can be formulated as:

$$f_\sigma(\theta) = \psi^p_{\max}\left(\sigma_i(\theta)\right) \leq \sigma_{\max},\ i \in \mathcal{I} \tag{12a}$$

$$\sigma(\theta) = Dx(\theta), \tag{12b}$$

$$K(\theta)x(\theta) = f_{\mathrm{accel}}(\theta) - M(\theta)\ddot{x}_{\mathrm{u}}, \tag{12c}$$

in which D is a matrix which computes the shear stress values at all the nodes and $\mathcal{I}$ is the set of nodes along the interface between the stage and the substrate. Note that the right hand side of (12c) must be self-equilibriated such that no rigid-body motions are induced. Further, $\psi^p_{\max}(\cdot)$ is the same aggregation function used for the eigenvalue constraints (11) in order to combine all stress constraints into a single constraint.

Similarly, certain external disturbance forces f_{dist} act on the stage. These disturbances are counteracted with corrective actuator forces $f_{\mathrm{corr}}(\theta)$ by the control system. Under the influence of these combined forces, the stage deforms. These deformations may not allow any point on the substrate to displace by more than a set amount. Such constraints can be expressed as:

$$f_x(\theta) = \psi^p_{\max}\left(\delta x_i(\theta)\right) \leq \delta x_{\max},\ i \in \mathcal{K} \tag{13a}$$

$$K(\theta)x(\theta) = f_{\mathrm{corr}}(\theta) + f_{\mathrm{dist}}, \tag{13b}$$

in which $\mathcal{K}$ is a set of degree-of-freedom indices. Here too, the right hand side of (13b) must be self-equilibriated.

2.6. Topology parameterization

The structure is parameterized using the density approach in topology optimization [3]. This implies that each finite element is assigned a design value ρ_e between $\rho_e^{\min}$ and 1 (see Fig. 1). Note that $\rho_e^{\min} > 0$ in order to avoid a singular stiffness matrix (cf. (14a)). Then, the properties of that element are scaled with the design value according to an interpolation function. A modified form of the SIMP (Solid Isotropic Material with Penalization) approach is used, which is widely used in topology optimization [3]. Specifically, for an element e, this results in:

$$K_e = \rho_e^3 K_e^{(0)}, \tag{14a}$$

$$M_e = \begin{cases} \rho_e M_e^{(0)}, & \rho_e > 0.1, \\ \rho_e^5 M_e^{(0)}, & \rho_e \leq 0.1, \end{cases} \tag{14b}$$

in which $K_e^{(0)}$ and $M_e^{(0)}$ are the element stiffness and mass matrices for solid material. These interpolations ensure that intermediate-density elements tend to underperform (i.e., for most intermediate density values the stiffness is far lower than the mass), encouraging a black-and-white design. The special mass interpolation (14b) ensures that for low design values the element mass tends to zero faster than its stiffness. This is necessary to suppress localized vibration modes in low-density areas [25, 33, 34, 22].

The element matrices are assembled into the global finite element matrices K and M according to [35]:

$$K = \mathop{\mathsf{A}}_{e=1}^{n_e} K_e, \quad M = \mathop{\mathsf{A}}_{e=1}^{n_e} M_e, \tag{15}$$

with the assembly operator performing the scattering of element matrices to the global degrees of freedom.

A spatial filter is applied to the design values to prevent mesh-dependent solutions, to avoid so-called "checkerboarding", and to introduce a minimum length scale [36, 37]. The filter is defined as

$$\tilde{\rho}_e = \frac{\sum_{k=1}^{n_e} w_{ek}\rho_k}{\sum_{k=1}^{n_e} w_{ek}}, \quad w_{ek} = \max\left(0,\ r_{\min} - \mathrm{dist}(k,e)\right).$$

Hence, the design values $\tilde{\rho}_e$ becomes a distance-weighted average of itself and its nearest neighbors. The *filtered design values* $\tilde{\rho}_e$ replace the *design values* ρ_e in (14a–14b). It is useful to note that the global filtering operation can be written and implemented as:

$$\tilde{\rho} = F\rho, \tag{16}$$

in which F is a sparse matrix.

Referring to the diagram (Fig. 1), only the light gray areas have been parameterized, which implies that only the topology in these areas may be modified.

2.7. Design-dependent actuator masses

The actuators shown in Fig. 1 are assumed to be of the electromagnetic type. It is assumed that the mass of each of these actuators can be expressed as a fraction of the overall structural mass, since accelerating a heavier stage will require larger actuators. Specifically, the mass of an actuator a can be expressed as:

$$m_{\mathrm{a}}(\theta) = \mu_{\mathrm{a}} m_c(\theta),$$

in which $m_c(\theta)$ is the total mass of the component excluding all actuators and $\mu_{\mathrm{a}}$is the mass fraction[4] of the actuator. Design-dependent actuator masses are expected to have a significant influence on the dynamic behavior of the stage.

3. Optimization problem formulation

At this point it is useful to recall the goal of this article, which is to develop a procedure to optimize the closed-loop performance (defined in Sec. 2.4.1) of a motion system, subject to constraints and mechanical requirements on the structure (defined in Secs. 2.4.2 and 2.5). To this end, a gradient-based optimization algorithm is used, which makes use of the objective and constraint values and their gradients. The computation of these design gradients is discussed in the next section, Sec. 4. Fig. 6 shows the outline of the initialization and optimization process.

Use is made of sequential quadratic programming (SQP), following the ideas in [38, 39]. A fairly conservative optimization strategy was chosen, in order to reduce the likelihood of ending up in an infeasible design (meaning that the combination of the structure and the controller is unstable) from which it is hard to recover. The chosen SQP algorithm makes use of the objective and constraint function values and their gradients with respect to the parameters. Second order gradient information is not computed, but estimated by means of a reciprocal intervening variable approach [38, 39].

[4] These mass fractions can usually be estimated from actuator force density data or empirical data.

All design variables are collected in a vector θ, which then consists of the topology design variables $\rho_e,\ e = 1,\ldots,n_d$ and the controller variables $k_i, \omega_{\mathrm{b},i},\ i = 1,\ldots,n_c$. Obviously, the topology parameters ρ_e affect the closed-loop performance and stability as well as the mechanical constraints via the mass and stiffness matrices. The controller parameters on the other hand only affect closed-loop performance and stability.

The optimization problem is formulated as:

$$\begin{aligned} \underset{\theta}{\text{minimize}} \quad & \max_{x\in\mathcal{X}} f_0(\theta, x) \\ \text{subject to} \quad & f_i(\theta, x) \le f_i^{\max} \ \ \forall \ \begin{matrix} i = 1,\ldots,n_c, \\ x \in \mathcal{X}. \end{matrix} \end{aligned} \tag{17}$$

In this problem $f_0(\theta, x)$ is the scalar objective function (8) and $f_i(\theta, x)$ are the, possibly vector-valued, constraints.

Note that in the objective and constraints a parameter x has been introduced. The dynamics seen by the control system depend on the sensor positions as the stage moves with respect to the metrology frame. To account for this fact, the design is performed simultaneously for multiple positions $x \in \mathcal{X}$, where $\mathcal{X}$ is a set of locations.

In the examples in this article three x-locations are considered: the center position and the extreme left/right positions of the stage. In each design iteration the worst case performance out of three sensor locations is optimized, while ensuring that the constraints are simultaneously satisfied for each of the three locations, effectively leading to a robust design.

3.1. Optimization problem convergence and uniqueness

Due to the non-linear and non-convex nature of the objective function and the constraint functions and the use of a gradient-based optimization algorithm, the existence and uniqueness of a global solution is not guaranteed. This is commonly accepted in topology optimization [3, 16]. A requirement of the proposed procedure is that one starts with a feasible design. This means that the starting combination of the topology (the structural lay-out) and the controller should satisfy the constraints and have a reasonable closed-loop performance, given the structural dynamics. This can be achieved by manually tuning the controller to the initial structural design. From this starting point onwards, the gradient-based search will attempt to improve the performance subject to the constraints. The optimization process may be terminated when the relative change in the objective value becomes small, or a finite number of iterations has been performed.

4. Design sensitivity analysis

The optimization framework presented in this article relies heavily upon the ability to efficiently compute design sensitivity information. First, for a given design, say the initial topology with initial controller parameters defined by θ_0, the performance and all constraint functions are computed. Then, the gradients of these functions with respect to all design variables θ are computed.

Although finite-differencing is a simple technique to obtain sensitivity information, the number of design variables (in particular the structural design variables) is very significant and can range into the millions for detailed finite element meshes, making finite-differencing infeasible. Since this number is large in comparison with the number of objective and constraint functions, use is made of adjoint sensitivity analysis where possible.

The following subsections treat the sensitivity analysis of the various analysis components which constitute the closed-loop and mechanical analysis of the design problem.

4.1. Mode sensitivities

The reduced-order system model (3) is based on the natural frequencies and eigenmodes of the system. Hence, the sensitivity of the reduced order model depends on the sensitivities of the modes themselves. This section describes how eigenvector and eigenvalue sensitivities are computed using an adjoint method according to [40, 41]. The eigenvalue problem related to the dynamics (1) is:

$$(K - \omega_i^2 M)\phi_i = 0, \qquad \phi_i^T M \phi_i = 1.$$

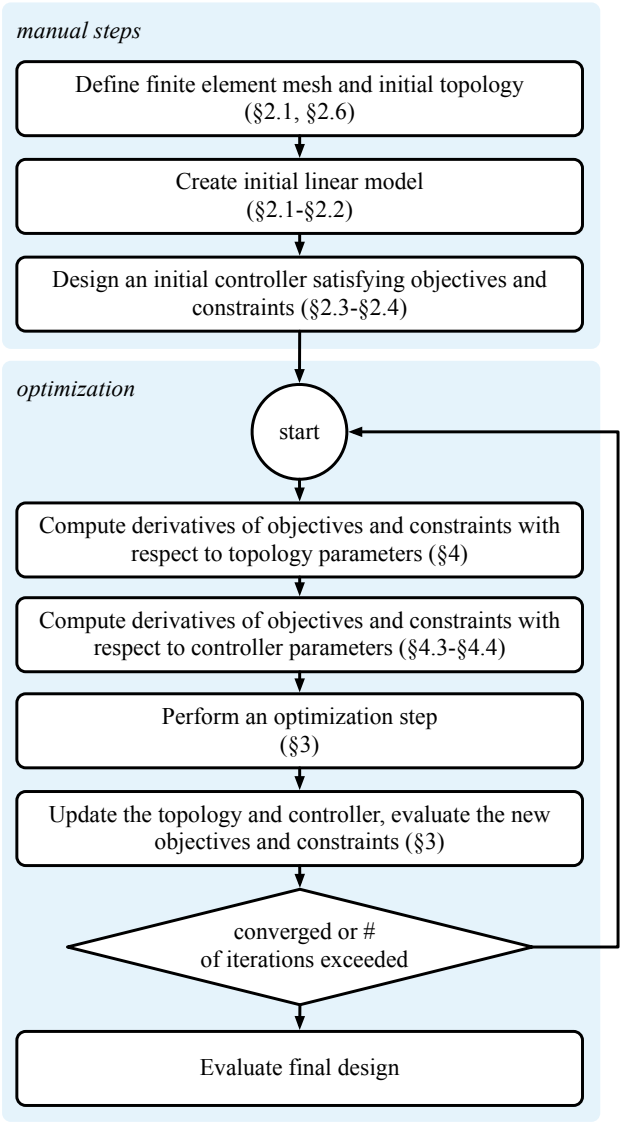


Figure 6: Flowchart outlining the steps in the algorithm.

Only the sensitivities of certain linear combinations of the modes are computed. This can be expressed as:

$$\tilde{\phi}_i = P^T \phi_i, \tag{18}$$

in which P computes certain linear combinations of the DoFs of interest (for instance based on the locations where a load is applied or where a displacement is measured). In the simplest case, P could be a vector which selects one element from ϕ_i.

Next, (18) is augmented with Lagrange multipliers to perform adjoint sensitivity analysis:

$$\tilde{\phi}_i^\star = P^T \phi_i + \mathcal{M}_i^T (K - \omega_i^2 M)\phi_i + \nu_i \left(\frac{1}{2} - \frac{1}{2} \phi_i^T M \phi_i \right).$$

$\mathcal{M}_i$ and ν_i are arbitrary but non-zero matrix-valued and vector-valued Lagrange multipliers, respectively, both related

to the i-th mode. Computing the gradient leads to:

$$\frac{\mathrm{d}\tilde{\phi}_i^\star}{\mathrm{d}\theta_j} = P^T \frac{\mathrm{d}\phi_i}{\mathrm{d}\theta_j} + \mathcal{M}_i^T \left(\frac{\mathrm{d}K}{\mathrm{d}\theta_j} - \omega_i^2 \frac{\mathrm{d}M}{\mathrm{d}\theta_j} - \frac{\mathrm{d}\omega_i^2}{\mathrm{d}\theta_j} M \right) \phi_i + \mathcal{M}_i^T \left(K - \omega_i^2 M \right) \frac{\mathrm{d}\phi_i}{\mathrm{d}\theta_j} + \nu_i \left(-\frac{1}{2} \phi_i^T \frac{\mathrm{d}M}{\mathrm{d}\theta_j} \phi_i - \phi_i^T M \frac{\mathrm{d}\phi_i}{\mathrm{d}\theta_j} \right). \quad (19)$$

Gathering all terms involving $\frac{\mathrm{d}\phi_i}{\mathrm{d}\theta_j}$ and $\frac{\mathrm{d}\omega_i^2}{\mathrm{d}\theta_j}$ the following adjoint equation can be defined:

$$\begin{bmatrix} K - \omega_i^2 M & -M\phi_i \\ -\phi_i^T M & 0 \end{bmatrix} \begin{bmatrix} \mathcal{M}_i \\ \nu_i^T \end{bmatrix} = \begin{bmatrix} -P \\ 0 \end{bmatrix}. \quad (20)$$

Given this choice, what remains from (19) is:

$$\frac{\mathrm{d}\tilde{\phi}_i}{\mathrm{d}\theta_j} = \mathcal{M}_i^T \left(\frac{\mathrm{d}K}{\mathrm{d}\theta_j} - \omega_i^2 \frac{\mathrm{d}M}{\mathrm{d}\theta_j} \right) \phi_i - \nu_i \frac{1}{2} \phi_i^T \frac{\mathrm{d}M}{\mathrm{d}\theta_j} \phi_i. \quad (21)$$

Hence, for each mode i, (20) must be solved once, after which (21) can be used to compute sensitivities of $\tilde{\phi}_i$ with respect to each design variable θ_j.

The sensitivity of the eigenvalue ω_i^2 can be computed using [3]:

$$\frac{\mathrm{d}\omega_i^2}{\mathrm{d}\theta_j} = \phi_i^T \left(\frac{\mathrm{d}K}{\mathrm{d}\theta_j} - \omega_i^2 \frac{\mathrm{d}M}{\mathrm{d}\theta_j} \right) \phi_i. \quad (22)$$

Note that the sensitivity expressions for mode shapes and natural frequencies in this subsection are only valid for modes which have multiplicity one. For modes with multiplicity greater than one[5], one must resort to specialized techniques [42, 43].

4.1.1. Sensitivity analysis of rigid-body modes

Rigid-body modes can usually be defined by physical reasoning. I.e., for a free 3D structure one may define 3 translations along the coordinate axes and 3 rotations about these axes, for instance about the center of mass. However, unless the body has three axes of symmetry and a diagonal inertia tensor, these modes will *not* be mass-orthogonal. Assume that such "intuitive" modes have been computed and collected in the matrix Φ_d, then a non-unique transformation matrix T can be computed which makes these modes mass-orthonormal, by requiring:

$$\Phi_\mathrm{d,m} T = \Phi_\mathrm{d}, \quad (23)$$

such that:

$$\Phi_\mathrm{d,m}^T M \Phi_\mathrm{d,m} = T^{-T} \Phi_\mathrm{d}^T M \Phi_\mathrm{d} T^{-1} = I,$$

or $\Phi_\mathrm{d}^T M \Phi_\mathrm{d} = T^T T$. One admissible T can be found by requiring T to be symmetric, such that:

$$\Phi_\mathrm{d}^T M \Phi_\mathrm{d} = T^2, \quad (24)$$

in which T is the principal square root of $\Phi_\mathrm{d}^T M \Phi_\mathrm{d}$ [44]. The sensitivity of $\Phi_\mathrm{d,m}$ follows by differentiating (23):

$$\frac{\mathrm{d}\Phi_\mathrm{d,m}}{\mathrm{d}\theta_j} = -\Phi_\mathrm{d} T^{-1} \frac{\mathrm{d}T}{\mathrm{d}\theta_j} T^{-1},$$

The sensitivity of the transformation matrix T follows by differentiating (24):

$$\Phi_\mathrm{d}^T \frac{\mathrm{d}M}{\mathrm{d}\theta_j} \Phi_\mathrm{d} = \frac{\mathrm{d}T}{\mathrm{d}\theta_j} T + T \frac{\mathrm{d}T}{\mathrm{d}\theta_j}.$$

which is a *continuous Lyapunov equation* that can be solved for $\frac{\mathrm{d}T}{\mathrm{d}\theta_j}$ using a range of techniques, for instance making use of the Kronecker product [45, 46].

[5] This also occurs if there are multiple rigid-body modes, since these all have zero natural frequency, cf. Sec. 4.1.1.

4.2. Sensitivity of the decoupled plant G

Making use of (3) and defining $\Sigma = s^2 I + 2Z\Omega s + \Omega^2$, the gradient of $P(s)$ can be computed:

$$\frac{\mathrm{d}P(s)}{\mathrm{d}\theta_j} = C\left(\frac{\mathrm{d}\Phi}{d\theta_j}\Sigma^{-1}\Phi^T - \Phi\Sigma^{-1}\frac{\mathrm{d}\Sigma}{\mathrm{d}\theta_j}\Sigma^{-1}\Phi^T + \Phi\Sigma^{-1}\frac{\mathrm{d}\Phi^T}{d\theta_j}\right)B,$$

with:

$$\frac{\mathrm{d}\Sigma}{\mathrm{d}\theta_j} = 2Z\frac{\mathrm{d}\Omega}{\mathrm{d}\theta_j}s + \frac{\mathrm{d}\Omega^2}{\mathrm{d}\theta_j}.$$

The sensitivity of the decoupled plant $G(s)$ then follows from the chain rule:

$$\frac{\mathrm{d}G(s)}{\mathrm{d}\theta_j} = W_\mathrm{S}\frac{\mathrm{d}P(s)}{\mathrm{d}\theta_j}W_\mathrm{A} + W_\mathrm{S}P(s)\frac{\mathrm{d}W_\mathrm{A}}{\mathrm{d}\theta_j}.$$

Note that W_S is design-independent. For W_A one finds the following design sensitivity from (5):

$$\frac{\mathrm{d}W_\mathrm{A}}{\mathrm{d}\theta_j} = -(C\Phi\Phi^T B)^{-1}C\left(\frac{\mathrm{d}\Phi}{\mathrm{d}\theta_j}\Phi^T + \Phi\frac{\mathrm{d}\Phi^T}{\mathrm{d}\theta_j}\right)BW_\mathrm{A}.$$

The sensitivities of the controllers with respect to the gain k and the bandwidth ω_b are readily obtained by differentiating (6), but could equally well be obtained by finite differencing due to the small number of parameters involved.

4.3. Design sensitivity of closed-loop transfer functions

It can be seen by inspection that in terms of sensitivity analysis both (8) and (9) require computation of the gradient of $|S_{ii}(j\omega)|_\mathrm{dB}$ for a given frequency ω. These sensitivities are readily derived based on the results of the previous sections.

Since this article focuses on the sensitivity functions $|S_{ii}(j\omega)|_\mathrm{dB}$, originally defined in (7), its design sensitivity will be derived as an example. However, the same principles can be applied to any other closed-loop function. For brevity the explicit dependencies of $S(s;\theta)$ on s and θ are dropped. Then, from (7) it follows that:

$$\frac{\mathrm{d}S}{\mathrm{d}\theta_j} = -S\left(\frac{\mathrm{d}P}{\mathrm{d}\theta_j}C + P\frac{\mathrm{d}C}{\mathrm{d}\theta_j}\right)S. \tag{25}$$

However, in most cases one is interested only in the complex modulus $|S_{kl}|$ of one element (k,l) of the matrix-valued sensitivity function, for which it is straightforward to show that:

$$\begin{aligned}\frac{\mathrm{d}|S_{kl}|}{\mathrm{d}\theta_j} &\equiv \frac{\mathrm{d}}{\mathrm{d}\theta_j}\sqrt{\mathcal{R}e\{S_{kl}\}^2 + \mathcal{I}m\{S_{kl}\}^2} \\ &= \frac{1}{|S_{kl}|}\mathcal{R}e\left\{S_{kl}^\star\frac{\mathrm{d}S_{kl}}{\mathrm{d}\theta_j}\right\},\end{aligned}$$

in which the * denotes complex conjugation. Note that the latter equation can be evaluated based on the result of (25).

Finally, frequent use is made of values in decibels (cf. Sec. 2.4.1), for which the following relation is useful:

$$\frac{\mathrm{d}|S_{kl}|_\mathrm{dB}}{\mathrm{d}\theta_j} = \frac{20}{\ln 10}\frac{1}{|S_{kl}|}\frac{\mathrm{d}|S_{kl}|}{\mathrm{d}\theta_j}.$$

4.4. Sensitivity of closed-loop eigenvalues

The eigenvalues λ_i and left and right eigenvectors $\phi_{i,\mathrm{l}},\ \phi_{i,\mathrm{r}}$ of the closed-loop system matrix (10) are described by the equations:

$$\Psi\phi_{i,\mathrm{r}} = \lambda_i\phi_{i,\mathrm{r}}$$
$$\phi_{i,\mathrm{l}}^T\Psi = \lambda_i\phi_{i,\mathrm{l}}^T$$

Sensitivities are found by differentiating the right-eigenvector problem according to the chain rule:

$$\frac{\mathrm{d}\Psi}{\mathrm{d}\theta_j}\phi_{i,\mathrm{r}} + (\Psi - \lambda_i I)\frac{\mathrm{d}\phi_{i,\mathrm{r}}}{\mathrm{d}\theta_j} - \frac{\mathrm{d}\lambda_i}{\mathrm{d}\theta_j}\phi_{i,\mathrm{r}} = 0.$$

Pre-multiplying with the corresponding left-eigenvector and rearranging results in:

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}\theta_j} = \frac{\phi_{i,\mathrm{l}}^T\frac{\mathrm{d}\Psi}{\mathrm{d}\theta_j}\phi_{i,\mathrm{r}}}{\phi_{i,\mathrm{l}}^T\phi_{i,\mathrm{r}}}.$$

This equation allows one to compute the derivatives $\frac{\mathrm{d}\lambda_i}{\mathrm{d}\theta_j}$ of the closed-loop eigenvalues.

4.5. Sensitivity analysis for mechanical constraints

All mechanical constraints are of the form $f_i(x) - f_i^{\max} \leq 0$, in which x is the displacement field following from a static finite element computation. For instance, $f_i(x)$ could describe the relative displacement between two points on the structure, the absolute displacement of a set of points or the stresses at certain points. In each case, it is assumed that $f_i(x)$ is a user-defined and continuously differentiable (vector) function of the deformation field x.

To efficiently compute the sensitivity of the constraint functions the adjoint method is used [3, 47]. This involves converting the function $f_i(x)$ to a Lagrangian which incorporates the constraints. In this case, the "constraints" are represented by the equilibrium equation $Kx = f$, with f the applied load corresponding to the case under consideration. The resulting expression is:

$$f_i^\star(x) = f_i(x) + \lambda^T(Kx - f), \tag{26}$$

which implies that $f_i^\star(x) = f_i(x)$ for any nonzero vector λ. Differentiating this expression with respect to a parameter θ_j yields:

$$\frac{\mathrm{d}f_i^\star(x)}{\mathrm{d}\theta_j} = \frac{\partial f_i(x)}{\partial\theta_j} + \frac{\partial f_i(x)}{\partial x^T}\frac{\mathrm{d}x}{\mathrm{d}\theta_j} + \lambda^T\left(\frac{\mathrm{d}K}{\partial\theta_j}x + K\frac{\mathrm{d}x}{\partial\theta_j} - \frac{\mathrm{d}f}{\mathrm{d}\theta_j}\right). \tag{27}$$

The essence of the adjoint approach is to note that the term $\frac{\mathrm{d}x}{\mathrm{d}\theta_j}$ is expensive to evaluate. Hence, the freedom in choosing λ will be exploited to eliminate it from (27) by choosing:

$$K\lambda = \frac{\partial f_i(x)}{\partial x}. \tag{28}$$

Eq. (28) is the *adjoint equation* related to (26). With this choice of λ, (27) reduces to:

$$\frac{\mathrm{d}f_i^\star(x)}{\mathrm{d}\theta_j} = \frac{\partial f_i(x)}{\partial\theta_j} + \lambda^T\left(\frac{\mathrm{d}K}{\partial\theta_j}x - \frac{\mathrm{d}f}{\mathrm{d}\theta_j}\right). \tag{29}$$

Note that (29) is computationally relatively cheap to evaluate for each element θ_j of θ, mainly because (28) needs to be solved only once.

4.6. Constraint aggregation

For some of the introduced constraints, like local displacement and stress constraints, it is not feasible to add each local constraint individually. In these cases, constraint aggregation is applied [31, 32]. The reasoning can be explained for the example of displacements: if each local displacement shall be less than an allowed value, the maximum of all local displacements must necessarily also be less than that value. Hence, the following implication holds:

$$\delta x_i < \delta x_{\max} \; \forall \; i \implies \max_i \{\delta x_i\} < \delta x_{\max},$$

effectively allowing N constraints to be replaced by a single one. Since the maximum operator is non-differentiable, an approximation is introduced in the form of the Kreisselmeier-Steinhauser function [32]:

$$\psi^p_{\max} = \frac{1}{p} \ln\left(\sum_{i=1}^{N} e^{p x_i}\right).$$

This approximation is also continuously differentiable, leading to the approximate sensitivity:

$$\frac{\mathrm{d}\psi^p_{\max}}{\mathrm{d}\theta_j} = \frac{1}{\sum_{i=1}^{N} e^{p x_i}} \sum_{i=1}^{N} e^{p x_i} \frac{\mathrm{d}x_i}{\mathrm{d}\theta_j}.$$

See [32] for important aspects regarding the numerical implementation of this function.

4.7. Density filtering

Gradients of the mass and stiffness matrices can be computed with respect to the filtered design density field $\tilde{\rho}_e$ by making use of their definitions in (14) and (15), however, the relation to the actual design values ρ_e must be accounted for. Suppose a gradient $\mathrm{d}f_i/\mathrm{d}\tilde{\rho}$ has been computed with respect to the entire design value vector ρ. Then, the following relation can be derived from (16):

$$\frac{\mathrm{d}f_i}{\mathrm{d}\rho} = F^T \frac{\mathrm{d}f_i}{\mathrm{d}\tilde{\rho}}.$$

4.8. Design-dependent actuator mass

Design-dependent actuator masses were discussed in Sec. 2.7. This section will show how such dependencies can be incorporated into the sensitivity analysis.

It is assumed that the actuators are of a given geometry (i.e., the internal actuator topology is not parameterized and fixed) and that the element stiffness matrices in the parts therefore remain unaffected as the design proceeds. It is further assumed that the initial mass of an actuator equals m^0_{a}. The mass fraction μ_{a} specifies that the part has a mass of μ_{a} times the mass of the component excluding all the actuators m_{c}. Using this information, each scaled element mass matrix within the part is computed according to:

$$M_e = \frac{\mu_{\mathrm{a}}}{m^0_{\mathrm{a}}} m_{\mathrm{c}} M^0_e, \tag{30}$$

in which M^0_e is the original element mass matrix. Further, note that m_{c} is defined as:

$$m_{\mathrm{c}} = \sum_{e \in \mathcal{E}} \tilde{\rho}^p_e m^0_e, \tag{31}$$

in which p is the penalization coefficient according to (14b) and m_e is the unscaled mass of element e. Note that e in the summation runs only over the component elements excluding actuators $\mathcal{E}$ (i.e., does not include the elements inside the actuators).

The sensitivity of (30) is then readily obtained as:

$$\frac{\mathrm{d}M_e}{\mathrm{d}\theta_j} = \left(\sum_{e \in \mathcal{E}} \frac{\mathrm{d}\tilde{\rho}^p_e}{\mathrm{d}\theta_j} m^0_e\right) \frac{\mu_{\mathrm{a}}}{m^0_{\mathrm{a}}} M^0_e,$$

showing that the sensitivities of the element mass matrices within the actuators are related to the densities $\tilde{\rho}_e$ in a straightforward way.

5. Case study

As an example design problem the 2D stage shown earlier in Fig. 1 is considered.

5.1. Problem definition and results

Optimization is performed according to (17). The objective function is composed of the disturbance attenuation at $\omega^\star = 0.1$ Hz in each control loop, as given in (8). The total mass of the design is also added to the objective function with a small penalty factor (about 5% of the total objective value). The rationale behind this is that lighter designs are preferred, but the stimulus is small enough not to compromise the main objective of disturbance attenuation.

Furthermore, several constraints are imposed. For the control loops there are constraints on sensitivity function peaks (9) and on closed-loop stability (11). Mechanical constraints comprise constraints on shear stresses between the stage and the substrate (12) and deformations of the substrate (13).

The starting point for the optimization runs is a design in which the domain is filled with 75% density material (which corresponds to $E = 0.42E_0$ and $\rho = 0.95\rho_0$, following (14a-14b)) with $E_0 = 70\,\text{GPa}$ and $\rho_0 = 2700\,\text{kg/m}^3$ for the design area (aluminium) and $E_0 = 210\,\text{GPa}$ and $\rho_0 = 7000\,\text{kg/m}^3$ for the solid actuator areas (steel). To compare the integrated approach, an eigenfrequency optimization has also been performed. High natural frequencies are usually favourable for controller design. Hence, one can maximize certain eigenfrequencies of the stage, while still imposing the same set of mechanical design constraints and adding the same mass penalty. A certain spacing between eigenfrequencies is also desirable from a control perspective. Several problem formulations for eigenfrequency design are possible, for instance:

1. Maximize f_1 subject to $f_2 \geq 1.1 f_1$;
2. Maximize f_1 subject to $f_2 \geq 1.1 f_1$ and $f_3 \geq 1.1 f_2$;
3. Maximize $1/f_1 + 1/f_2$ subject to $f_2 \geq 1.1 f_1$ [48].

Here, f_i is the frequency of the i-th flexible mode. Only the results for option 2 are shown, as it was the best-performing option of the three. Option 3 in particular resulted in a poorly defined geometry with large gray areas. The result of option 2 is shown in Fig. 7 (frq). It is important to note that these eigenfrequency optimizations have been performed under the requirement that all mechanical constraints are satisfied. Furthermore, in order to determine the performance of the eigenfrequency designs, the resulting structural design was used in closed loop, and the controller was optimized in the same way as for the integrated design. This means that the only difference between designs 7 (frq) and 7 (int) is that the former is optimized for maximum first eigenfrequency and subsequenctly maximum control performance, whereas the latter is optimized directly for closed-loop performance: all other constraints are identical and a similar mass penalty is included.

The topology optimization results for the integrated approach in which controller and structure are optimized simultaneously, is shown in Fig. 7 (int). The fact that gray areas are present in both designs presents a challenge for practical purposes. Several options exist to convert these designs to pure solid–void designs, for instance based on image-processing or thresholding techniques [49]. See [50] for examples in the present context.

With regards to computational time, both the integrated and eigenfrequency designs usually converge within about 30-40 iterations, requiring about 5 minutes for eigenfrequency optimization and about 15 minutes for the integrated approach, where the bulk of the time is spent computing mode sensitivities (cf. Sec. 4.1). This can possibly be improved in future work by exploiting parallel computing capabilities or fast re-analysis techniques. For both optimizations convergence is reached when the objective does not change more than a set tolerance between iterations in a relative sense.

The performance of both designs is compared in Figures 8 to 11. Note that all frequencies have been scaled with a constant factor for intellectual property reasons. The attenuation results in Fig. 8 are based upon the disturbance attenuation in the three control loops at 0.1 normalized frequency units on a decibel scale. First, it appears from Fig. 8 that the integrated approach (int) yields a far better improvement in disturbance attenuation compared to the eigenfrequency design (frq) in each of the motion control loops. Second, from Fig. 9 (int), it is clear that this performance is achieved while the first three natural frequencies are all *lower* than for the eigenfrequency design. So, in this case, it is not at all beneficial to raise the first three natural frequencies. Clearly, eigenfrequency design can only be performed

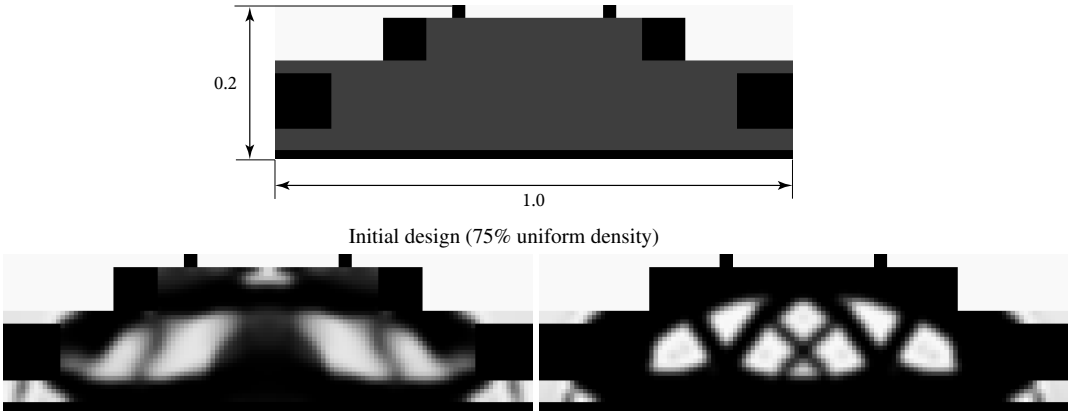


Figure 7: Example topologies comparing integrated and eigenfrequency designs. The performance figures are compared in Figures 8 to 11. Note that the initial design does not meet the mechanical requirements (Sec. 2.5).

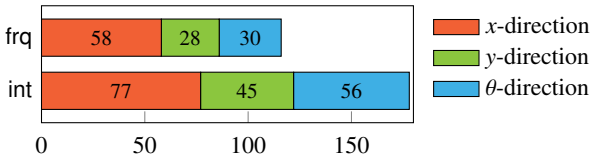


Figure 8: Disturbance attenuation figures for the two designs (larger values imply better attenuation). The abbreviations frq and int refer to the designs shown in Fig. 7.

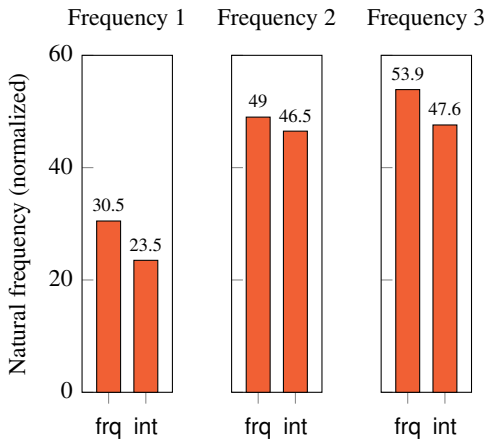


Figure 9: Natural frequencies (scaled) of the two designs. The abbreviations frq and int refer to the designs shown in Fig. 7. Mode shapes are shown in Figs. 13 and 14.

with a good understanding of which subset of modes is most relevant to overall performance, an understanding which is very hard to develop in general. In fact, the eigenfrequency design (frq) results in a lower average performance than the solid design while its lowest natural frequency is 30% higher than for the integrated design. Finally, the integrated approach also yields a slightly lower mass (Fig. 11), which is 5% below the eigenfrequency design.

The results obtained for this example demonstrate that non-trivial and fairly well-defined designs can be obtained, which could form a starting point for subsequent design. The integrated design (int) has on average about 20 dB better

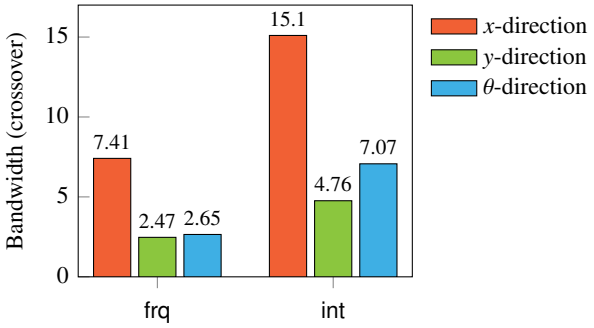


Figure 10: Bandwidth values for the two designs. The abbreviations frq and int refer to the designs shown in Fig. 7.

disturbance attenuation at 0.1 normalized frequency units on each of the motion control axes.

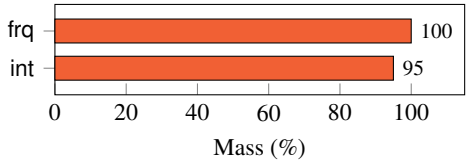


Figure 11: Mass comparison (smaller is lighter). The abbreviations frq and int refer to the designs shown in Fig. 7.

5.2. Analysis of limitations

A careful analysis of the transfer functions shown in Fig. 12 can help to understand the performance differences between the eigenfrequency design and the integrated design. These figures show the effective loop gains (i.e., each rigid-body loop gain transfer function is generated with the other control loops closed). Two sensor locations are considered: a center location, Fig. 12a and the leftmost location, Fig. 12b (the results for the rightmost location are identical to the leftmost). It is important to realize that for any of the control loops, only one of these sensor locations will usually dictate the performance limitations encountered.

Analysing the three control loops (x, y and θ) for both designs reveals the following observations:

- Eigenfrequency design: For each of the rigid-body motion control loops, performance is limited by non-collocated flexible modes, which is observed from the 180° phase loss associated with these modes. For the x and θ control loops this is the 3$^{\text{rd}}$ mode shown in Fig. 13c, which is what appears to be a second asymmetric beam-like bending mode of the stage. For the y control loop this is the 2$^{\text{nd}}$ mode shown in Fig. 13b, which is a symmetric second bending mode.

- Integrated design: The x control loop is limited by mode 6 shown in Fig. 14f, which is a complicated mode shape resembling a second asymmetric bending mode. The θ control loop is limited by mode 3 shown in Fig. 14c which also looks like a second asymmetric bending mode. For the y control loop it is clear that for the leftmost sensor position, a non-minimum phase zero is introduced when the other control loops are closed, which limits the bandwidth. This zero appears to be related to the first mode of the structure, shown in Fig. 14a.

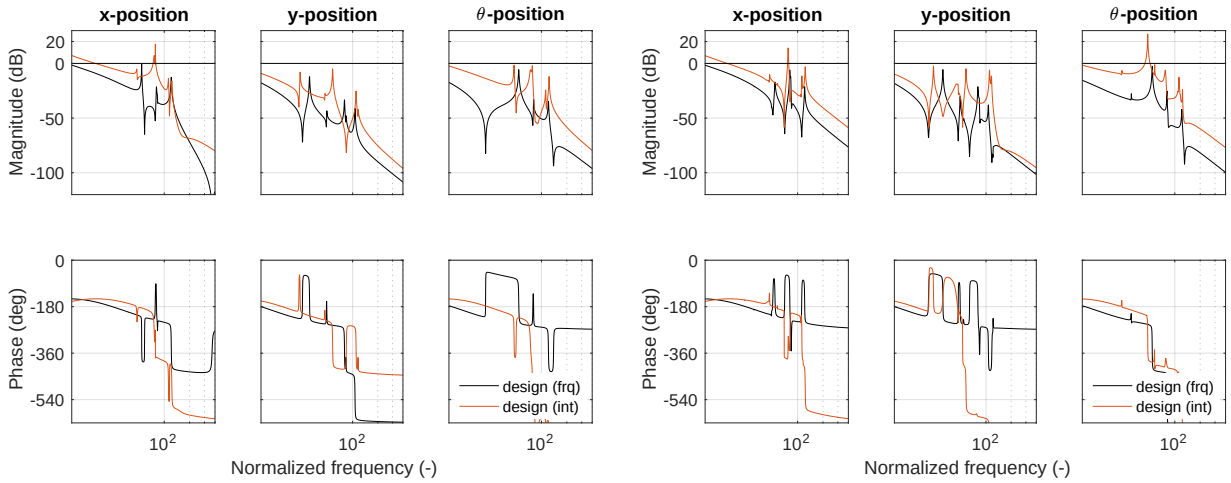
Comparing the red and black lines, it is also clear that in all control loops the integrated design leads to higher loop gains (i.e., better disturbance rejection) subject to the same robustness and stability constraints.

The previous analysis, along with the mode shapes shown in Figs. 13 and 14, shows that there may be a clear advantage to the integrated design approach. It is very hard to predict a priori, for a complicated geometry, which modes are bandwidth-limiting. This results in a limited scope for eigenfrequency optimization procedures, unless one has very detailed knowledge on how each mode affects the controlled dynamics of the structure. Still, when the structural topology changes, this knowledge may become obsolete, since mode shapes and their spatial character may

change drastically. In contrast, the integrated approach directly takes into account the controlled dynamics, and is therefore able to modify all structural modes based on their relevance to achievable performance in a coordinated way.

While it is hard to draw general conclusions from this case study, a good illustration of the advantages of the integrated design approach is the following: the integrated approach results in the fact that only the 6^{th} mode limits the x position control, whereas all other modes do not hinder the performance since they have been made barely observable or collocated (see the red lines in the x position dynamics in Figs. 12a and 12b).

Another observation is that the intricate topology of the integrated design leads to a denser eigenfrequency spectrum at low frequencies. Figs. 13 and 14 show that the first 4 modes of design (frq) span the same frequency range as the first 6 modes of design (int).



(a) *Center* sensor location. (b) *Leftmost* sensor location.

Figure 12: Effective loop gain transfer functions for two different sensor locations, comparing the eigenfrequency design (frq, black lines) with the integrated design (int, red lines).

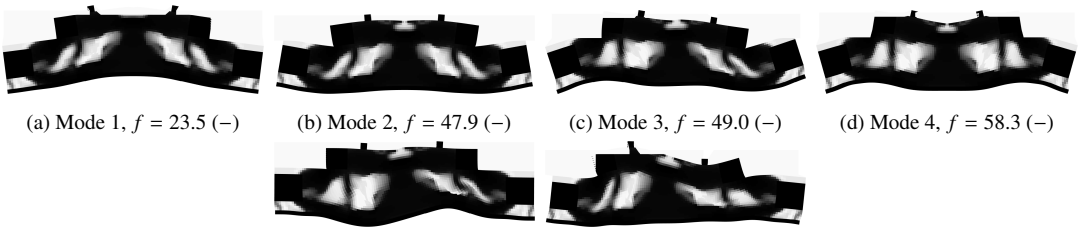
A fundamental assumption underlying the gradient-based optimization approach is that it relies on small changes in the plant dynamics between iterations, meaning that the nature of consecutive poles and zeros in the transfer functions does not change. The process has difficulty when flexible modes start to cross each other. In those cases, the controlled system dynamics change drastically, implying that the controller parameters need to be changed abruptly. This may be less of an issue if the algorithm is applied to modify existing designs so that only small changes in the dynamics occur, but it needs attention to allow application to a wider set of problems. The issue could also potentially be overcome by performing separate controller optimization iterations after each structural design modification.

6. Discussion

Based on the example of optimizing a motion stage for closed-loop performance, the potential of the integrated gradient-based design methodology as a useful tool for the precision industry has been demonstrated. It can deal with sophisticated objective functions related to closed-loop performance, while including requirements and specifications on mechanical behavior. The case study demonstrated that superior designs are generated using an integrated approach, compared to the conventional eigenfrequency maximization approach. The main focus of subsequent research will be to incorporate more constraints originating from actual practice and to move towards 3D design problems. While in this article the focus was on mechanical and control system specifications, ultimately, incorporating thermal specifications will also be a necessity for precision applications.

(a) Mode 1, $f = 30.3$ (–) (b) Mode 2, $f = 49.7$ (–) (c) Mode 3, $f = 54.7$ (–) (d) Mode 4, $f = 82.3$ (–)

Figure 13: The first four flexible mode shapes of the eigenfrequency design (frq).



(a) Mode 1, $f = 23.5$ (–) (b) Mode 2, $f = 47.9$ (–) (c) Mode 3, $f = 49.0$ (–) (d) Mode 4, $f = 58.3$ (–)

(e) Mode 5, $f = 79.7$ (–) (f) Mode 6, $f = 83.8$ (–)

Figure 14: The first six flexible mode shapes of the integrated design (int).

7. Acknowledgements

The authors gratefully acknowledge Jan van Eijk (Mice), Jeroen van de Wijdeven and Ton de Groot (ASML) and Pascal Etman (Eindhoven University of Technology) for their valuable contributions to this research. Funding: This work was funded by a Point-One university-industry interaction grant from the Dutch government in the context of the sector “Hightech Systems and Materials”.

References

[1] R. Munnig-Schmidt, G. Schitter, A. Rankers, J. van Eijk, The Design of High Performance Mechatronics: High-tech Functionality by Multidisciplinary System Integration, 2nd Edition, Delft University Press, 2014.
[2] G. E. van Baars, J. Smeltink, J. van der Werff, M. Limpens, M. Barink, J. de Vreugd, O. Galaktionov, G. Witvoet, Additive manufacturing for high-tech systems performance, Mikroniek (6) (2014) 5–11.
[3] M. Bendsøe, O. Sigmund, Topology Optimization: Theory, Methods and Applications, Engineering online library, Springer, 2003.
[4] A. L. Hale, W. E. Dahl, R. J. Lisowski, Optimal simultaneous structural and control design of maneuvering flexible spacecraft, Journal of Guidance, Control, and Dynamics 8 (1) (1985) 86–93.
[5] D. F. Miller, J. Shim, Gradient-based combined structural and control optimization, Journal of Guidance, Control, and Dynamics 10 (3) (1987) 291–298.
[6] M. Salama, J. Garba, L. Demsetz, F. Udwadia, Simultaneous optimization of controlled structures, Computational Mechanics 3 (4) (1988) 275–282.
[7] M. Milman, M. Salama, R. Scheid, R. Bruno, J. Gibson, Combined control-structural optimization, Computational Mechanics 8 (1) (1991) 1–18.
[8] T. Zeiler, M. Gilbert, Integrated control/structure optimization by multilevel decomposition, Structural optimization 6 (2) (1993) 99–107.
[9] R. A. Canfield, L. Meirovitch, Integrated structural design and vibration suppression using independent modal space control, AIAA Journal 32 (10) (1994) 2053–2060.
[10] F. Y. Cheng, D. Li, Multiobjective optimization of structures with and without control, Journal of Guidance, Control, and Dynamics 19 (2) (1996) 392–397.
[11] A. Messac, Control-structure integrated design with closed-form design metrics using physical programming, AIAA journal 36 (5) (1998) 855–864.
[12] X. Liu, D. W. Begg, On simultaneous optimisation of smart structures – part i: Theory, Computer Methods in Applied Mechanics and Engineering 184 (1) (2000) 15 – 24.
[13] G. Zhao, B. Chen, Y. Gu, Control–structural design optimization for vibration of piezoelectric intelligent truss structures, Structural and Multidisciplinary Optimization 37 (5) (2009) 509–519.
[14] X. Liu, D. W. Begg, R. J. Fishwick, Genetic approach to optimal topology/controller design of adaptive structures, International Journal for Numerical Methods in Engineering 41 (5) (1998) 815–830.
[15] B. Xu, J. Jiang, J. Ou, Integrated optimization of structural topology and control for piezoelectric smart trusses using genetic algorithm, Journal of Sound and Vibration 307 (3–5) (2007) 393 – 427.
[16] O. Sigmund, On the usefulness of non-gradient approaches in topology optimization, Structural and Multidisciplinary Optimization 43 (5) (2011) 589–596.
[17] J. Ou, N. Kikuchi, Integrated optimal structural and vibration control design, Structural optimization 12 (4) (1996) 209–216.

[18] Y. Zhu, J. Qiu, H. Du, J. Tani, Simultaneous optimal design of structural topology, actuator locations and control parameters for a plate structure, Computational Mechanics 29 (2) (2002) 89–97.

[19] Y. Zhu, J. Qiu, H. Du, J. Tani, Simultaneous structural-control optimization of a coupled structural-acoustic enclosure, Journal of Intelligent Material Systems and Structures 14 (4-5) (2003) 287–296.

[20] O. A. A. da Silveira, J. S. O. Fonseca, Simultaneous design of structural topology and control for vibration reduction using piezoelectric material, in: E. Dvorkin, M. Goldschmit, M. Storti (Eds.), Mecánica Computacional, Vol. XXIX, 2010, pp. 8375 – 8389.

[21] S. Skogestad, I. Postlethwaite, Multivariable feedback control, analysis and design, John Wiley and Sons, 1996.

[22] G. J. van der Veen, M. Langelaar, F. van Keulen, Integrated topology and controller optimization of motion systems in the frequency domain, Structural and Multidisciplinary Optimization 51 (3) (2015) 673–685.

[23] A. Albers, J. Ottnad, J. Minx, P. Häußler, Topology and controller parameter optimization in dynamic mechatronic systems, in: 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, American Institute of Aeronautics and Astronautics, 2008, pp. 1–12. doi:doi:10.2514/6.2008-5898.

[24] M. M. da Silva, Computer-aided integrated design of mechatronic systems, Ph.D. thesis, Katholieke Universiteit Leuven (2009).

[25] N. Pedersen, Maximization of eigenvalues using topology optimization, Structural and Multidisciplinary Optimization 20 (1) (2000) 2–11.

[26] K. Vandyshev, M. Langelaar, F. van Keulen, J. van Eijk, Combined topology and shape optimization of controlled structures, in: 3rd International Conference on Engineering Optimization, Rio de Janeiro, Brazil, 2012, pp. 1–9.

[27] M. Steinbuch, M. Norg, Advanced motion control: An industrial perspective, European Journal of Control 4 (4) (1998) 278 – 293. doi:http://dx.doi.org/10.1016/S0947-3580(98)70121-9.
URL http://www.sciencedirect.com/science/article/pii/S0947358098701219

[28] K. Åström, R. Murray, Feedback Systems: An Introduction for Scientists and Engineers, Princeton University Press, 2010.

[29] M. Šebek, Z. Hurák, An often missed detail: Formula relating peek sensitivity with gain margin less than one, in: M. Fikar, M. Kvasnica (Eds.), Proceedings of the 17th International Conference on Process Control '09, Slovak University of Technology in Bratislava, Štrbské Pleso, Slovakia, 2009, pp. 65–72.

[30] N. Bruinsma, M. Steinbuch, A fast algorithm to compute the $\mathcal{H}_\infty$-norm of a transfer function matrix, Systems & Control Letters 14 (4) (1990) 287 – 293. doi:http://dx.doi.org/10.1016/0167-6911(90)90049-Z.
URL http://www.sciencedirect.com/science/article/pii/016769119090049Z

[31] P. Duysinx, O. Sigmund, New developments in handling optimal stress constraints in optimal material distributions, in: 7th Symposium in Multidisciplinary Analysis and Optimization, AIAA/USAF/NASA/ISSMO, 1998, pp. 1501–1509.

[32] J. R. R. A. Martins, N. M. K. Poon, On structural optimization using constraint aggregation, in: 6th World Congress on Structural and Multidisciplinary Optimization, 2005, pp. 1–10.

[33] D. Tcherniak, Topology optimization of resonating structures using simp method, International Journal for Numerical Methods in Engineering 54 (11) (2002) 1605–1622.

[34] J. Du, N. Olhoff, Topological design of freely vibrating continuum structures for maximum values of simple and multiple eigenfrequencies and frequency gaps, Structural and Multidisciplinary Optimization 34 (2) (2007) 91–110.

[35] T. J. R. Hughes, The Finite Element Method: Linear Static and Dynamic Finite Element Analysis, Dover Civil and Mechanical Engineering, Dover Publications, 2012.

[36] T. E. Bruns, D. A. Tortorelli, Topology optimization of non-linear elastic structures and compliant mechanisms, Computer Methods in Applied Mechanics and Engineering 190 (26–27) (2001) 3443 – 3459.

[37] O. Sigmund, J. Petersson, Numerical instabilities in topology optimization: A survey on procedures dealing with checkerboards, mesh-dependencies and local minima, Structural optimization 16 (1) (1998) 68–75.

[38] A. A. Groenwold, L. F. P. Etman, D. W. Wood, Approximated approximations for SAO, Structural and Multidisciplinary Optimization 41 (1) (2010) 39–56.

[39] L. F. P. Etman, A. A. Groenwold, J. E. Rooda, First-order sequential convex programming using approximate diagonal QP subproblems, Structural and Multidisciplinary Optimization 45 (4) (2012) 479–488.

[40] T. H. Lee, An adjoint variable method for structural design sensitivity analysis of a distinct eigenvalue problem, KSME International Journal 13 (6) (1999) 470–476.

[41] T. H. Lee, Adjoint method for design sensitivity analysis of multiple eigenvalues and associated eigenvectors, AIAA Journal 45 (8) (2007) 1998–2004.

[42] A. P. Seyranian, E. Lund, N. Olhoff, Multiple eigenvalues in structural optimization problems, Structural Optimization 227 (1994) 207–227.

[43] M. I. Friswell, The derivatives of repeated eigenvalues and their associated eigenvectors, Journal of Vibration and Acoustics 118 (3) (1996) 390–397.

[44] N. Higham, Functions of Matrices, Society for Industrial and Applied Mathematics, 2008. arXiv:http://epubs.siam.org/doi/pdf/10.1137/1.9780898717778.

[45] R. H. Bartels, G. W. Stewart, Solution of the matrix equation AX + XB = C [F4], Commun. ACM 15 (9) (1972) 820–826.

[46] J. W. Brewer, Kronecker products and matrix calculus in system theory, IEEE Transactions on circuits and systems 25 (9) (1978) 772–781.

[47] K. Choi, N. Kim, Structural Sensitivity Analysis and Optimization 1: Linear Systems, Mechanical Engineering Series, Springer, 2005.

[48] Z.-D. Ma, N. Kikuchi, H.-C. Cheng, Topological design for vibrating structures, Computer Methods in Applied Mechanics and Engineering 121 (1–4) (1995) 259 – 280.

[49] O. Sigmund, Morphology-based black and white filters for topology optimization, Structural and Multidisciplinary Optimization 33 (4-5) (2007) 401–424.

[50] G. J. van der Veen, M. Langelaar, S. H. van der Meulen, W. H. T. M. Aangenent, D. A. H. Laro, F. van Keulen, Topology optimization for the conceptual design of precision mechatronic systems, in: Proceedings of the 30th ASPE Annual Meeting, American Society for Precision Engineering, 2015, pp. 136–141.