Full Analytical Sensitivities in NURBS based Isogeometric Shape Optimization

Xiaoping Qian *
Illinois Institute of Technology
Chicago, IL 60062

February 27, 2010

Abstract

Non-uniform rational B-spline (NURBS) has been widely used as an effective shape parameterization technique for structural optimization due to its compact and powerful shape representation capability and its popularity among CAD systems. The advent of NURBS based isogeometric analysis has made it even more advantageous to use NURBS in shape optimization since it can potentially avoid the inaccuracy and labor-tediousness in geometric model conversion from the design model to the analysis model.

Although both positions and weights of NURBS control points affect the shape, until very recently, usually only control point positions are used as design variables in shape optimization, thus restricting the design space and limiting the shape representation flexibility.

This paper presents an approach for analytically computing the full sensitivities of both the positions and weights of NURBS control points in structural shape optimization. Such analytical formulation allows accurate calculation of sensitivity and has been successfully used in gradient-based shape optimization.

The analytical sensitivity for both positions and weights of NURBS control points is especially beneficial for recovering optimal shapes that are conical e.g. ellipses and circles in 2D, cylinders, ellipsoids and spheres in 3D that are otherwise not possible without the weights as design variables.

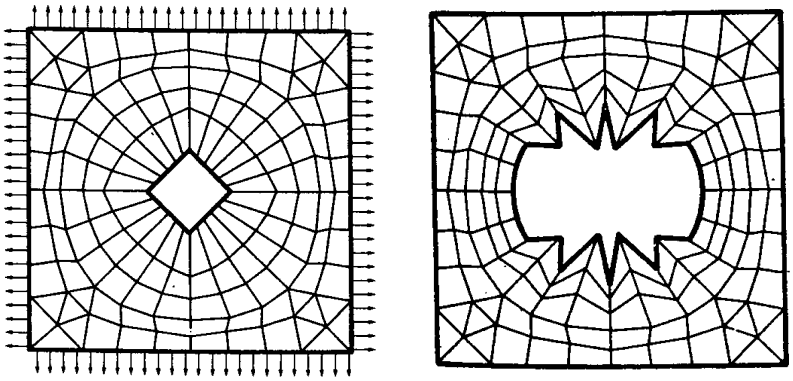
Keywords: Shape optimal design, NURBS, Isogeometric analysis

1 Introduction

This paper presents an approach for analytically computing the full sensitivities of both the positions and weights of NURBS (non-uniform rational B-spline) control points in structural shape optimization.

* *Email address:* qian@iit.edu.

Even since the work of Braibant and Fleury [1], B-spline and its generalized representation, NURBS, have been widely used in shape parameterization in structural optimization. It has become the method of choice for parameterizing freeform shape in structural optimization [2] for two important reasons: 1) With a few control points, NURBS can represent complex freeform shape. The alternative representation, the use of finite element nodes as design variables, would often lead to wiggly, irregular shape. Figure 1 gives one such example. 2) The output of NURBS-based shape optimization can be directly linked to a computer-aided design (CAD) system since NURBS is the standard shape representation underlying all major CAD software.



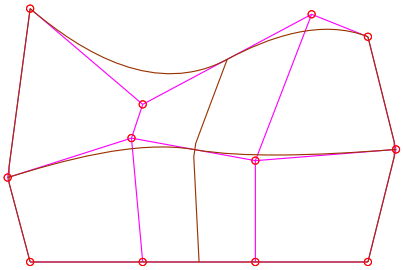
(a) Initial design (b) Optimized design

Figure 1: Optimized shape of a hole in a plate where elements nodes are used as design variables. Figures are taken from [1].

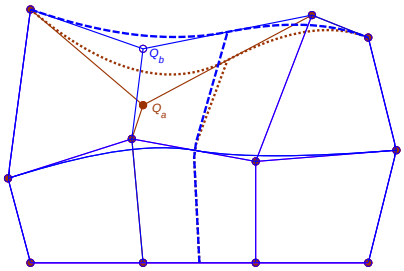
The recent advent of NURBS based isogeometric analysis [3] has made it even more advantageous to use NURBS in shape parameterization for design optimization since NURBS can not only be used to represent the geometry, it can also be used as a basis for approximating the physical fields. The use of the NURBS basis in finite element analysis has exhibited superior numerical properties, e.g. in terms of per-degree-of-freedom accuracy, over traditional finite element analysis [3]. Further, the tri-variate B-spline representation has been extended to represent both geometry and material composition in functionally gradient materials (FGM) parts [4] and used in the B-spline basis based graded finite element analysis of FGM objects [5]. It thus allows closer integration with CAD since the exact geometry and even material composition can be used in both design and analysis through the NURBS representation.

NURBS represented shape is affected by both the positions and weights of its control points. Figure 2 presents a 4×3 control net for a NURBS surface consisting of 2×2 knot spans with degree 2 in ξ_1 direction and degree 1 in ξ_2 direction. Figure 2.b shows when a control point changes its position from Q_a to Q_b, the underlying surface and knot spans change. Figure 2.c shows the surface change and the knot span change when the weight of control point Q

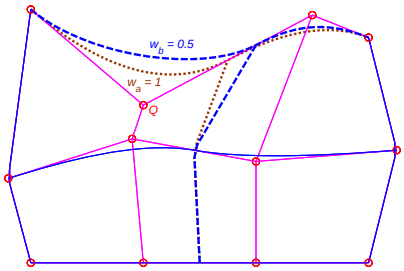
changes from 1 to 0.5.



(a) Initial NURBS surface



(b) Modified NURBS surface



(c) Modified NURBS surface

Figure 2: Both position and weight of a control point can change the NURBS surface. a) The initial NURBS surface; b) Modified NURBS surface after the position of control point Q is changed from Q_a to Q_b, c) Modified NURBS surface after the weight of the control point Q is changed from $w_a = 1$ and $w_b = 0.5$.

Although both positions and weights of control points affect the NURBS geometry, as demonstrated in Fig. 2, until very recently, usually only positions of control points are used as design variables [1, 6, 7, 8, 9, 10]. In these work, analytical sensitivities when used are only given for positions of control points, thus they are referred to as *partial sensitivity* in this paper. Isogeometric analysis has recently been successfully applied in structural shape optimization [11, 12] where, again, only analytical sensitivities for positions of control points are given.

Thus far, the use of both weights and positions of control points as design variables has only occasionally been explored, e.g. in [13]. In particular, there has been a lack of analytical formula for sensitivity calculation for shape optimization with both positions and weights of control points as design variables. A notable exception is very recent work in [14] where both control points and weights are used to optimize one-dimensional beam structures with sensitivities analytically evaluated.

Analytical formula for computing the sensitivity of physical quantities over both positions and weights is important for the following reasons:

- Analytical formulas leads to more accurate and efficient calculation of derivative information required in gradient-based optimization. The finite difference based method for gradient calculation suffers from the "*step-size dilemma*" due to the potential truncation error and round-off error [15]. It is also inefficient. If we need to find the derivatives of the structural response with respect to n design variables, the forward-difference approximation requires n analyses, while the central-difference approximation would require $2n$ analyses.

- The use of NURBS weights as design variables in structural optimization, in addition to positions of control points, lends more flexibility in shape representation and enlarges the design space, which can lead to better design. In particular, it makes it possible to recover a class of optimal shapes such as conic curves, e.g. ellipses and circles, and surfaces, e.g. cylinders, spheres and ellipsoids, which are otherwise not possible. Note, without weights, NURBS shape degenerates into B-spline shape and B-spline representation cannot exactly represent these conic curves and surfaces.

The absence of analytical sensitivities of physical quantities such as compliance, displacement and stress over shape parameters, i.e. both positions and weights, is due perhaps to the seeming complexity of such a derivation. A naïve way of obtaining these sensitivities would be to expand the integrand involved in calculating these physical quantities into explicit expressions of design variables, which would be of daunting complexity. In the context of isogeometric shape optimization, the derivation could be more involving since the NURBS basis function used in analysis also becomes affected by the design variables (weights). For example, although the usefulness of weights as design variables was recognized in recent work in [11], the derivation of analytical sensitivities for weights was not available and was deemed "*more complex*".

In this paper, we give a set of compact formulas for computing analytical sensitivities for both control point positions and weights, thus referred to as *full sensitivity*. Our approach follows that of [16, 17, 18], an isoparametric based technique for differentiating stiffness matrix and force vectors with respect to discrete design variables. We use the chain rule of differentiation and Jacobi's formula for the derivative of a determination to derive these compact formulas. These formulas are applicable to both traditional finite element based NURBS shape optimization and isogeometric shape optimization.

The calculation of these sensitivities involve two terms: *analysis terms* that are encountered during the usual finite element analysis and isogeometric analysis and *geometric sensitivities* that are represented as the derivatives of positions and weights over design variables. Mesh refinement is often required for accurate finite element analysis. This is especially true in isogeometric shape optimization since the weights are now design variables which can induce distorted distribution of isoparametric curves (element boundaries). Thus, an analytical method is also given for propagating geometric sensitivities of the control points in the design model to those of the control points in the refined analysis model.

Our numerical implementation is based on the isogeometric analysis due to its numerical advantages. Numerical examples demonstrate the availability of such analytical formulas has both theoretical implication and practical significance. Theoretically, they can be used in interpreting the optimality conditions and understanding behaviors of physical systems which may not been seen directly from the problem. For example, they can be used in determining whether an exact circle is the optimal shape for a hole in a plate under bi-axial load. In practice, they enlarge the design space, allow flexibility in shape representation and lead to better designs.

In the remainder of this paper, Section 2 gives a brief introduction on the NURBS basis and NURBS geometry and introduces some key notations used in sensitivity derivation. Section 3 gives a general formulation of shape optimization and the role of sensitivity in shape optimization. Section 4 gives the analytical formulas for sensitivities over positions and weights of NURBS control points. Section 5 discusses how the geometric sensitivity of a design model can be analytically propagated to that in the refined analysis model. Section 6 discusses the result of our numerical implementation on some common shape optimization problems. This paper concludes in Section 7. The derivation of the analytical formulas is given in the Appendix.

2 Introduction to NURBS

This section gives a brief introduction on NURBS basis functions and NURBS geometry. It introduces some notations that will be used in deriving the analytical sensitivity in following sections. For details on NURBS, refer to [19].

A NURBS curve of degree p is defined as follows

$$\boldsymbol{x}(\xi) = \frac{\sum_{i=0}^{n} B_{i,p}(\xi) w_i P_i}{\sum_{j=0}^{n} B_{j,p}(\xi) w_j}, \quad 0 \le \xi \le 1, \tag{1}$$

where $\{P_i\} = (x_{i_1}, x_{i_2})$ represents the coordinate positions of a set of $i = 0, \ldots, n$ control points, $\{w_i\}$ is the corresponding weight, and $\{B_{i,p}\}$ is the degree p B-spline basis function defined on the knot vector

$$\Xi = \{\xi_0, \xi_1, \ldots, \xi_{n+p+1}\}.$$

The i-th ($i = 0, \ldots, n$) B-spline basis function can be defined recursively as

$$\begin{aligned} B_{i,p}(\xi) &= \frac{(\xi - \xi_i) B_{i,p-1}(\xi)}{\xi_{i+p} - \xi_i} + \frac{(\xi_{i+p+1} - \xi) B_{i+1,p-1}(\xi)}{\xi_{i+p+1} - \xi_{i+1}} \\ B_{i,0}(\xi) &= \begin{cases} 1 & \xi_i \le \xi \le \xi_{i+1} \\ 0 & \text{Otherwise} \end{cases} \end{aligned}.$$

The derivative of the i-th B-spline basis function can be computed as follows

$$\frac{dB_{i,p}(\xi)}{d\xi} = \frac{p}{\xi_{i+p} - \xi_i} B_{i,p-1}(\xi) - \frac{p}{\xi_{i+p+1} - \xi_{i+1}} B_{i+1,p-1}(\xi).$$

A NURBS surface of degree p in ξ_1 direction and degree q in ξ_2 direction is a bivariate vector-valued piecewise rational function of the form

$$\boldsymbol{x}(\xi_1,\xi_2)=\frac{\sum_{k=0}^{n}\sum_{l=0}^{m}B_{k,p}(\xi_1)B_{l,q}(\xi_2)w_{k,l}P_{k,l}}{\sum_{s=0}^{n}\sum_{t=0}^{m}B_{s,p}(\xi_1)B_{t,q}(\xi_2)w_{s,t}},\quad 0\leq\xi_1,\xi_2\leq 1 \tag{2}$$

The $\{P_{k,l}\}$ form a $(n+1)\times(m+1)$ bidirectional control net, $\{w_{k,l}\}$ are the weights, and the $\{B_{k,p}\}$ and $\{B_{l,q}\}$ are the B-spline basis functions defined on the knot vectors Ξ_1 and Ξ_2.

Without loss of generality, we here consider a NURBS surface on a knot-span basis, defined by an array of $n_{en}=(p+1)\times(q+1)$ control points. Note, the NURBS basis function has local influence property, i.e. within a given knot span, only $(p+1)\times(q+1)$ number of non-zero basis functions. So the total number of nodes per element (knot span) is $n_{en}=(p+1)\times(q+1)$.

The NURBS basis function $N_{k,l}$ for the control point $P_{k,l}$ ($k=0,\ldots,n$ and $l=0,\ldots,m$) can be written as

$$N_{k,l}(\xi_1,\xi_2)=\frac{B_{k,p}(\xi_1)B_{l,q}(\xi_2)w_{k,l}}{\sum_{s=0}^{n}\sum_{t=0}^{m}B_{s,p}(\xi_1)B_{t,q}(\xi_2)w_{s,t}} \tag{3}$$

where p and q are degrees of the non-rational B-spline basis functions $B_{k,p}$ and $B_{l,q}$.

For notational convenience, we change the matrix form of the NURBS basis function into the column form by converting the matrix index (k,l) into a column index $i=k*(q+1)+l$. We further note the NURBS basis in the following form

$$\begin{aligned}R_i(\xi_1,\xi_2)&=B_{k,p}(\xi_1)B_{l,q}(\xi_2),\\ N_i(\xi_1,\xi_2)&=\frac{R_iw_i}{\mathbf{R}^T\mathbf{W}},\end{aligned}$$

where $\mathbf{R}$ and $\mathbf{W}$ represents the column collection of R_i and w_i for $i=1$ to n_{en}.

We thus can rewrite Eq. (2) as

$$\boldsymbol{x}(\xi_1,\xi_2)=\mathbf{N}^T\mathbf{P}=\sum_{i=1}^{n_{en}}N_iP_i. \tag{4}$$

The derivative of R_i over its parametric coordinates can then be computed as

$$\frac{\partial R_i}{\partial\xi_1}=\frac{dB_{k,p}(\xi_1)}{d\xi_1}B_{l,q}(\xi_2),\qquad \frac{\partial R_i}{\partial\xi_2}=B_{k,p}(\xi_1)\frac{dB_{l,q}(\xi_2)}{d\xi_2} \tag{5}$$

3 Shape optimization

We use a 2D elasticity problem as an example for shape optimization and give its weak form equilibirum equations. The terms in the weak form such as stiffness matrix and force vectors will be used in composing objective functions and constraints for structural shape optimization.

3.1 Linear elasticity analysis

We here consider a 2D linear elasticity problem. The strong form for linear elasticity [20] is as follows

$$\begin{cases} \nabla \cdot \sigma_x + b_x = 0 \text{ and } \nabla \cdot \sigma_y + b_y = 0 \text{ on } \Omega \\ \sigma = \mathbf{D}\nabla_s \mathbf{u} \\ \sigma_x \cdot n = t_x \text{ and } \sigma_y \cdot n = t_y \text{ on } \Gamma_t, \\ u = \bar{u} \text{ on } \Gamma_u. \end{cases}$$

where Γ_t is portion of the boundary where traction is specified and Γ_u is portion of the boundary where displacement is specified.

The underlying discrete equilibrium equation is

$$\mathbf{Ku} = \mathbf{f}.$$

where $\mathbf{K}$ is the stiffness matrix, $\mathbf{u}$ is the displacement vector and $\mathbf{f}$ is the external force vector. The stiffness matrix $\mathbf{K}$ can be assembled from the element stiffness matrix $\mathbf{K}_e$. Likewise, the force vector $\mathbf{f}$ can be assembled from the element force vector $\mathbf{f}_e$.

The element stiffness matrix is computed as follows.

$$\mathbf{K}_e = t_e \int_{\widehat{\Omega}_e} \mathbf{B}^T \mathbf{D} \mathbf{B} |\mathbf{J}| d\widehat{\Omega} \tag{6}$$

where t_e is the plate element thickness and $\widehat{\Omega}$ is the parametric domain of the structure in the $\xi_1\xi_2$ space.

The integral is integrated numerically by determining the value of the integrand at Gauss points in the element. The strain-displacement matrix $\mathbf{B}$ is

$$\mathbf{B} = \begin{bmatrix} \frac{\partial N_1}{\partial x_1} & 0 & ... & \frac{\partial N_{n_{en}}}{\partial x_1} & 0 \\ 0 & \frac{\partial N_1}{\partial x_2} & ... & 0 & \frac{\partial N_{n_{en}}}{\partial x_2} \\ \frac{\partial N_1}{\partial x_2} & \frac{\partial N_1}{\partial x_1} & ... & \frac{\partial N_{n_{en}}}{\partial x_2} & \frac{\partial N_{n_{en}}}{\partial x_1} \end{bmatrix} \tag{7}$$

where N_i is the basis function for finite element analysis and is the NURBS basis function in isogeometric analysis.

For plane stress, the stress strain matrix $\mathbf{D}$ is written as

$$\mathbf{D} = \frac{E}{1-v^2} \begin{bmatrix} 1 & v & 0 \\ v & 1 & 0 \\ 0 & 0 & \frac{1-v}{2} \end{bmatrix}.$$

where E is Young's modulus and v is Poisson's ratio.

The Jacobian matrix is given by

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial x_1}{\partial \xi_1} & \dfrac{\partial x_2}{\partial \xi_1} \\ \dfrac{\partial x_1}{\partial \xi_2} & \dfrac{\partial x_2}{\partial \xi_2} \end{bmatrix} \tag{8}$$

which maps the points from the parametric coordinates to the world coordinates.

The force vector on element e may be written

$$\mathbf{f}_e = \int_{\widehat{\Omega}_e} \mathbf{N}^T \mathbf{b} |\mathbf{J}| t_e d\widehat{\Omega} + \int_{\widehat{\Gamma_{t_e}}} \mathbf{N}^T \mathbf{t} |\mathbf{J}| d\widehat{\Gamma} \tag{9}$$

where $\mathbf{b}$ is the body force (force per unit area), $\mathbf{t}$ is the traction on the boundary, and $\widehat{\Gamma}_t$ is the parametric domain of the traction boundary in the $\xi_1\xi_2$ space.

3.2 Structural shape optimization

The general mathematical formulation of a structural optimization problem can be stated as follows

$$\begin{cases} \min\limits_{\alpha_s} & f(\mathbf{u}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) & \\ \text{s.t.} & h_i(\mathbf{u}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) = 0, & \text{i} = 1 \text{ to } n_h \\ & g_j(\mathbf{u}(\boldsymbol{\alpha}), \boldsymbol{\alpha}) \leq 0, & \text{j} = 1 \text{ to } n_g \\ & \alpha_{\min_s} \leq \alpha_s \leq \alpha_{\max_s}, & s = 1 \text{ to } n_{eq} \end{cases},$$

where the objective function f is a function of the state variable, e.g. displacement $\mathbf{u}$ and the design variables $\boldsymbol{\alpha}$, n_h is the number of equality constraints, n_g is the number of inequality constraints, and n_{eq} is the number of design variables. The behavior constraints are represented by equality and inequality constraints h_i and g_j.

To solve the generally nonlinear optimization problem, both gradient based and gradient-less methods can be applied. In this paper, we focus on a gradient based approach where both the structural response and its sensitivity over design change is required. The specific optimization algorithm used in this paper is the gradient-based method of moving asymptotes (MMA) [21].

For example, a commonly used design formulation for structural shape optimization is to minimize the mean compliance of a structure under a fixed amount material through a volume fraction constraint $V \leq V^*$. Its discrete form reads

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}} & f = \mathbf{f}^T \mathbf{u} \\ \text{s.t.} & V \leq V^* \\ & \mathbf{K}\mathbf{u} = \mathbf{f} \end{cases}.$$

Another objective function used in this paper is to minimize the displacement at the point of load under a volume fraction constraint, i.e.

$$\begin{cases} \min\limits_{\boldsymbol{\alpha}} & f = \mathbf{l}^T \mathbf{u} \\ \text{s.t.} & V \leq V^* \\ & \mathbf{K}\mathbf{u} = \mathbf{f} \end{cases}$$

where $\mathbf{l}$ is a zero vector with a 1 corresponding to the point of load.

Using the gradient-based optimization approach such as MMA to solve the above optimization problems requires the sensitivities of objective functions and constraints over the design variables, i.e. $\partial f/\partial \boldsymbol{\alpha}$, $\partial h_i/\partial \boldsymbol{\alpha}$ and $\partial g_j/\partial \boldsymbol{\alpha}$. Evaluating these sensitivities requires sensitivities of physical quantities such as the stiffness $\mathbf{K}$ and force vector $\mathbf{f}$ over the design variables.

4 Sensitivity analysis

Sensitivity is useful in evaluating the robustness of a particular design and in determining search directions during structural optimization. During the optimization process, the geometric domain Ω_e will change due to the change of design variables $\boldsymbol{\alpha}$, however the corresponding parametric domain $\widehat{\Omega}_e$ does not under certain constraints such that the mesh remains in a good quality. This can be ensured by checking the Jacobian of the mapping.

Since the plate thickness t_e and the stress-strain matrix $\mathbf{D}$ are constant, differentiation of Eq. (6) gives

$$\frac{\partial \mathbf{K}_e}{\partial \alpha_s} = \int_{\widehat{\Omega}_e} \left(\frac{\partial \mathbf{B}^T}{\partial \alpha_s} \mathbf{D}\mathbf{B}|\mathbf{J}| + \mathbf{B}^T\mathbf{D}\frac{\partial \mathbf{B}}{\partial \alpha_s}|\mathbf{J}| + \mathbf{B}^T\mathbf{D}\mathbf{B}\frac{\partial |\mathbf{J}|}{\partial \alpha_s} \right) t_e \, d\widehat{\Omega}. \tag{10}$$

Differentiating Eq. (9) gives

$$\begin{aligned} \frac{\partial \mathbf{f}_e}{\partial \alpha_s} = &\int_{\widehat{\Omega}_e} \left(\frac{\partial \mathbf{N}^T}{\partial \alpha_s}\mathbf{b}|\mathbf{J}| + \mathbf{N}^T \left(\frac{\partial \mathbf{b}}{\partial x_1}\frac{\partial x_1}{\partial \alpha_s} + \frac{\partial \mathbf{b}}{\partial x_2}\frac{\partial x_2}{\partial \alpha_s} \right) |\mathbf{J}| + \mathbf{N}^T\mathbf{b}\frac{\partial |\mathbf{J}|}{\partial \alpha_s} \right) t_e d\widehat{\Omega} \\ &+ \int_{\widehat{\Gamma_{t_e}}} \left(\frac{\partial \mathbf{N}^T}{\partial \alpha_s}\mathbf{t}|\mathbf{J}| + \mathbf{N}^T \left(\frac{\partial \mathbf{t}}{\partial x_1}\frac{\partial x_1}{\partial \alpha_s} + \frac{\partial \mathbf{t}}{\partial x_2}\frac{\partial x_2}{\partial \alpha_s} \right) |\mathbf{J}| + \mathbf{N}^T\mathbf{t}\frac{\partial |\mathbf{J}|}{\partial \alpha_s} \right) d\widehat{\Gamma}_t \end{aligned}. \tag{11}$$

Our goal is now to find analytical formulas for $\partial \mathbf{B}^T/\partial \alpha_s$, $\partial |\mathbf{J}|/\partial \alpha_s$, $\partial \mathbf{N}/\partial \alpha_s$ and $\partial \boldsymbol{x}/\partial \alpha_s$.

4.1 Full Analytical Sensitivity in NURBS Isogeometric Shape Optimization

First, we define two additional matrices:

$$\mathbf{G} = \begin{bmatrix} \frac{\partial N_1}{\partial x_1} & \frac{\partial N_2}{\partial x_1} & \cdots & \frac{\partial N_{n_{en}}}{\partial x_1} \\ \frac{\partial N_1}{\partial x_2} & \frac{\partial N_2}{\partial x_2} & \cdots & \frac{\partial N_{n_{en}}}{\partial x_2} \end{bmatrix},$$

$$\widehat{\mathbf{G}} = \begin{bmatrix} \frac{\partial N_1}{\partial \xi_1} & \frac{\partial N_2}{\partial \xi_1} & \cdots & \frac{\partial N_{n_{en}}}{\partial \xi_1} \\ \frac{\partial N_1}{\partial \xi_2} & \frac{\partial N_2}{\partial \xi_2} & \cdots & \frac{\partial N_{n_{en}}}{\partial \xi_2} \end{bmatrix}.$$

In the following, for notational convenience, let us denote $\partial/\partial \alpha_s$ by a prime ($'$) and the derivative over ξ_j, $\partial()/\partial \xi_j$, as as $()_{,\xi_j}$. Note, if we know $\mathbf{G}'$, $\mathbf{B}'$ can be drawn from it.

Thus for the calculation of the sensitivities of stiffness matrix and force vector over design variables α_s, it suffices to find expressions for $\mathbf{N}'$, $\mathbf{G}'$, $\boldsymbol{x}'$, and $|\mathbf{J}|'$.

We present below our results concerning analytical sensitivities for shape optimization based on NURBS isogeometric analysis. The detailed proof is provided in the Appendix.

Theorem 1 *(Full sensitivity for* $\mathbf{P}$ *and* $\mathbf{W}$ *in isogeometric optimization)*

$$|\mathbf{J}|' = |\mathbf{J}| tr\left(\mathbf{G}\mathbf{P}' + \mathbf{J}^{-1}\widehat{\mathbf{G}}'\mathbf{P}\right), \tag{12}$$

$$\mathbf{G}' = \mathbf{J}^{-1}\widehat{\mathbf{G}}'(\mathbf{I} - \mathbf{P}\mathbf{G}) - \mathbf{G}\mathbf{P}'\mathbf{G}, \tag{13}$$

$$\boldsymbol{x}' = \mathbf{N}^T\mathbf{P}' + (\mathbf{N}')^T\mathbf{P}, \tag{14}$$

$$N_i' = \frac{R_i w_i'}{\mathbf{R}^T\mathbf{W}} - \frac{R_i w_i \mathbf{R}^T\mathbf{W}'}{(\mathbf{R}^T\mathbf{W})^2}, \tag{15}$$

$$\begin{aligned}\left(N_{i,\xi_j}\right)' = &\frac{R_{i,\xi_j} w_i'}{\mathbf{R}^T\mathbf{W}} - \frac{R_{i,\xi_j} w_i \mathbf{R}^T\mathbf{W}' + R_i w_i'(\mathbf{R}_{,\xi_j})^T\mathbf{W} + R_i w_i(\mathbf{R}_{,\xi_j})^T\mathbf{W}'}{(\mathbf{R}^T\mathbf{W})^2} \\ &+ 2\frac{R_i w_i(\mathbf{R}_{,\xi_j})^T\mathbf{W}\mathbf{R}^T\mathbf{W}'}{(\mathbf{R}^T\mathbf{W})^3}.\end{aligned} \tag{16}$$

Note, $\mathbf{I}$ is an identity matrix. $\left(N_{i,\xi_j}\right)'$ in Eq. (16) is used in describing $\widehat{\mathbf{G}}'$. Inserting the above equations into Eqs. (10) and (11) gives the complete analytical sensitivities of stiffness matrix and force vector over the design variables. Since both the effect of positions $\mathbf{P}$ and weights $\mathbf{W}$ of control points are considered, we thus refer to the resulting sensitivity as total sensitivity.

Such sensitivity information reflects the effect of change from both control point positions and weights. It should be pointed out, although we present these formulas in the context of a 2D elasticity problem, they are exactly applicable to 3D problems. To the author's best knowledge, this is the first reported analytical sensitivity for NURBS based shape optimization, taking into account the effect of the NURBS weights and control points.

If the weights do not change with respect to design variables, i.e. $w_i' = 0$, this leads to $N_i' = 0$ and $\left(N_{i,\xi_j}\right)' = 0$, and thus $\widehat{\mathbf{G}}' = 0$. The above sensitivity equations would then become a partial sensitivity for $\mathbf{P}$ as follows.

Corollary 2 *(Partial sensitivity for* $\mathbf{P}$ *in isogeometric optimization)*

$$\begin{aligned}|\mathbf{J}|' &= |\mathbf{J}| tr(\mathbf{G}\mathbf{P}'), \\ |\mathbf{G}|' &= -\mathbf{G}\mathbf{P}'\mathbf{G}, \\ \boldsymbol{x}' &= \mathbf{N}^T\mathbf{P}', \\ \mathbf{N}' &= 0, \\ \left(N_{i,\xi_j}\right)' &= 0,\end{aligned}$$

which are identical to the forms presented in [17, 18], except that instead of using nodal coordinates, we use control points' coordinates $\mathbf{P}$ and instead of

using the Lagrange basis function in FEA, we use the NURBS basis function in isogeometric analysis. In this sense, the full sensitivity presented in Theorem 1 generalizes the sensitivity in structural optimization from the classical Lagrange shape function based isoparametric finite element analysis *where the basis functions do not change with respect to design variables* to NURBS based isogeometric analysis *where the basis functions could change*.

The total analytical sensitivities presented in Theorem 1 can also be extended to traditional finite element based shape optimization. In traditional FEA, the Lagrange basis function does not change w.r.t to the design variables, i.e. $N'_{i,\xi_j} = 0$, we thus have the following corollary for total analytical sensitivities for both $\mathbf{P}$ and $\mathbf{W}$ in FEA based shape optimization:

Corollary 3 *(Full sensitivity for* $\mathbf{P}$ *and* $\mathbf{W}$ *in FEA based optimization)*

$$
\begin{aligned}
|\mathbf{J}|' &= |\mathbf{J}| tr(\mathbf{G}\mathbf{P}'), \\
|\mathbf{G}|' &= -\mathbf{G}\mathbf{P}'\mathbf{G}, \\
\boldsymbol{x}' &= \mathbf{N}^T\mathbf{P}', \\
N'_i &= \frac{R_i w'_i}{\mathbf{R}^T\mathbf{W}} - \frac{R_i w_i \mathbf{R}^T\mathbf{W}'}{(\mathbf{R}^T\mathbf{W})^2}, \\
\left(N_{i,\xi_j}\right)' &= 0,
\end{aligned}
$$

In Corollary 3, the sensitivities for both positions and weights of control points are given. The term $\boldsymbol{x}'$ and consequently the term N'_i are needed to calculate the physical coordinates' derivatives over the design variables, e.g. in the body force term in Eq. 11 and in structural grid generation where element nodes are generated from a NURBS representation.

The significance of Theorem 1 on total sensitivities for both $\mathbf{P}$ and $\mathbf{W}$ is obvious for the following reasons.

- *Low computational overhead*: the terms used in Eqs. (12) - (16) can be divided into two parts: 1) *Geometric sensitivity* as represented by $\mathbf{P}'$ and $\mathbf{W}'$, which measures the sensitivities of positions and weights of NURBS control points over design variables. These terms are thus referred to as *geometric sensitivity* in NURBS geometric optimization since the NURBS shape can be completely characterized by $\mathbf{P}$ and $\mathbf{W}$ for given degrees and knot vectors in the NURBS geometry. 2) *analysis terms*. All the remaining terms are calculations already incurred during the usual isogeometric or finite element analysis. Together, geometric sensitivity in conjunction with the calculations incurred during the usual analysis leads to physical sensitivity, thus low computational overhead is required in the analytical sensitivity analysis. That is, to compute sensitivities of physical quantities such as mass, stiffness and force, all the extra terms we need are the derivatives of control points $\mathbf{P}$ and weights $\mathbf{W}$ with respect to design variables. The computational implementation of these analytical sensitivities only takes a few extra lines of code.

- *Flexible design parameterization and re-parameterization*: For a given NURBS geometry (surface or solid), different design parameterization, i.e. controlling the NURBS geometry via different design variables α_s, can be easily supported via the calculation of geometric sensitivity $\mathbf{P}'$ and $\mathbf{W}'$. The corresponding physical sensitivities can then be easily calculated through Eqs. (12) - (16). Any design re-parameterization for a given NURBS solid would not require any modification of analysis related terms.

- *Enlarged design space*: Since the weights are represented as variables in the above equations, the design space has been enlarged from control point positions only to control point positions and weights. This can lead to better optimal design. In particular, due to the use of rational form of Basis functions and the use of weights as variables in the optimization, the solution space thus admits conic curves and surfaces.

- *Checking Karush-Kuhn-Tucker (KKT) optimality conditions*: The analytical sensitivity allows the check of KKT conditions, i.e. gradients of objective function and constraints over design variables to see if a given shape is optimal.

5 Sensitivity propagation from the design model to the analysis model

In this section, we illustrate how analytical sensitivities in Theorem 1 are computed in isogeometric shape optimization, through sensitivity propagation from the design model to the analysis model.

5.1 Design and analysis models

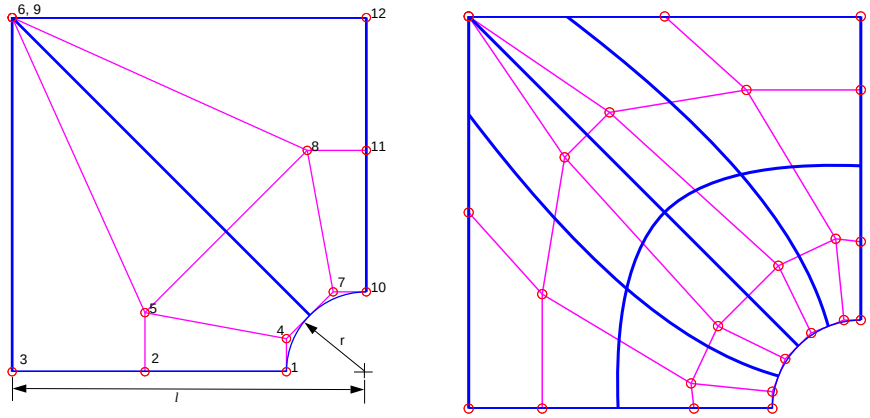
In our isogeometric optimization approach, we make distinction between the geometric model for design parameterization (referred to as a design model) and that for analysis (referred to as an analysis model).

In design parameterization, the structure shape is controlled by a set of design variables. The design control net is only as dense as necessary for defining the boundary shape. Excessive number of design variables could lead to wriggly optimal shapes. Figure 3.a shows two parameterizations of a hole profile within a square plate (only a quarter of the plate is shown).

- If a freeform hole profile is desired, the positions and weights for control points 1, 4, 7, and 10 can be chosen as independent design variables. This is indeed the shape parameterization used in the design of the plate with a hole in Section 6.1.

- Alternatively, if the hole profile is constrained to be circular, the design variable is the circle radius r.

In each case, the geometric sensitivities of 12 control points $\mathbf{P}$ and the corresponding weights $\mathbf{W}$ in Figure 3.a can be derived with respect to the respective design variables. Once $\mathbf{P}'$ and $\mathbf{W}'$ are known, they can be used in conjunction with the the design model based analysis terms in equations (12) to (16) to compute physical sensitivities such as $\mathbf{K}'$ and $\mathbf{f}'$.



(a) Design parameterization (b) Analysis refinement

Figure 3: Design parameterization of a hole profile in a plate and the refined analysis model. Only a quarter of the plate is shown. a) Design model used in shape parameterization; b) Analysis model through the refinement from the design model. Bold blues lines mark the element boundaries and red circles are control points.

However, the mesh at such density is likely not sufficient for accurate analysis required in optimization. This is especially true, given the fact that the weights are now design variables which can distort the isoparametric curve (element boundaries). Numerical results in Section 6 will further attest to this. Thus, the analysis model often requires much finer mesh. Mesh refinement techniques such as h-refinement, p-refinement, and k-refinement [3] can be used for this purpose. Figure 3.b shows an analysis model that is resulted from the original design model through knot insertion in both ξ_1 and ξ_2. The design model consists of 2×1 elements. The analysis model now consists of 4×2 elements and has 24 control points and corresponding weights. With such mesh refinement, sensitivity propagation is then needed that can automatically calculate the geometric sensitivities of control point positions and weights in the analysis model (i.e. the 24 control points in Fig. 3.b) based on the information on the geometric sensitivities of control point positions and weights in the design model (i.e. the 12 control points in Fig. 3.a). The subsection below describe how such sensitivity propagation can be conducted.

5.2 Sensitivity propagation from the design model to the analysis model

Here we focus on the sensitivity propagation during h-refinement. Sensitivity propagation formulation for other refinement can be derived similarly.

The basic procedure for h-refinement is through knot insertion. We describe below first the knot insertion for a B-spline curve, then its extension to a NUBRS curve.

Knot insertion refers to adding a new knot into the existing knot vector without changing the shape of the curve. Because the fundamental equality for B-spline curve $m = n + p + 1$ where $m + 1$ is the total number of knots for a degree p B-spline curve with $n + 1$ control points, inserting a new knot leads to a new control point to be added. More precisely, some existing control points are removed and new ones are added.

Given a set of $n+1$ control points $P_0, P_1, ..., P_n$, a knot vector $\{\xi_0, \xi_1, ..., \xi_m\}$ and a degree p, we can insert a new knot $\bar{\xi}$ into the knot vector without changing the shape of the B-spline curve $\boldsymbol{x}(\xi)$ as follows. Assuming we need to insert a knot $\bar{\xi}$ into the knot span $[\xi_l, \xi_{l+1}]$, we have the following basic knot insertion procedure for a B-spline curve:

- Find l such that $\bar{\xi}$ lies in the knot span $[\xi_l, \xi_{l+1}]$.
- Find $p + 1$ control points $P_{l-p}, P_{l-p+1}, ..., P_l$.
- Compute p new control points Q_i from the above $p + 1$ control points by using the formula
$$Q_i = (1 - \beta_i)P_{i-1} + \beta_i P_i,$$
where the ratio β_i is computed as below:
$$\beta_i = \frac{\bar{\xi} - \xi_i}{\xi_{i+p} - \xi_i} \qquad \text{for } l - p + 1 \leq i \leq l.$$

Thus the new knot vector becomes $\{\xi_0, \xi_1, ..., \xi_l, \bar{\xi}, \xi_{l+1}, ..., \xi_m\}$. The new control points are $\{P_0, P_1, ..., P_{l-p}, Q_{l-p+1}, Q_{l-p+2}, ..., Q_l, P_l, P_{l+1}, ..., P_n\}$.

The knot insertion for a NURBS curve is typically done by converting the given NURBS curve in three-dimensional (3D) to a B-spline curve in 4D, performing knot insertion in this four-dimensional (4D) B-spline curve, and then projecting the new set of control points back to 3D to form the new set of control points for the given NURBS curve. In this paper, we present the control points in 2D for convenience. Let $P_i = (x_i, y_i)$, then the control points in 3D is $P_i^w = (w_i x_i, w_i y_i, w_i)$. Then the new control point Q_i^w is calculated as follows

$$Q_i^w = (1 - \beta_i)P_{i-1}^w + \beta_i P_i^w.$$

The position of its projection in 2D becomes

$$Q_i = \frac{(1 - \beta_i)P_{i-1}^w + \beta_i P_i^w}{(1 - \beta_i)w_{i-1} + \beta_i w_i}, \tag{17}$$

and the weight is

$$w_{Q_i} = (1-\beta_i)w_{i-1} + \beta_i w_i. \tag{18}$$

Equations (17) and (18) form the basis for sensitivity propagation, i.e. propagating the sensitivity of control point position P_i and weight w_i over design variable α_s, i.e. (P_i', w_i'), to the new control point Q_i and w_{Q_i} after the knot insertion, i.e. (Q_i', w_{Q_i}').

From Eq. (18), we know

$$w_{Q_i}' = (1-\beta_i)w_{i-1}' + \beta_i w_i'. \tag{19}$$

From Eq. (17) and the new sensitivity for the weight w_{Q_i}, we have

$$\begin{aligned} Q_i' = \quad & \frac{(1-\beta_i)\left(P_{i-1}' w_{i-1} + P_{i-1} w_{i-1}'\right) + \beta_i \left(P_i' w_i + P_i w_i'\right)}{w_{Q_i}} \\ & - \frac{(1-\beta_i)P_{i-1}w_{i-1} + \beta_i P_i w_i}{w_{Q_i}^2} w_{Q_i}' \end{aligned} . \tag{20}$$

Therefore equations (18), (19), and (20) give all necessary equations for sensitivity propagation for h-refinement of a NURBS curve. A NURBS surface is defined with an array of control points and knot vectors. By repeating the above knot insertion procedure and the sensitivity propagation to all rows and all columns of control points, sensitivity propagation can thus be applied to 2D surface, and similarly to a tri-variate NURBS volume in 3D.

6 Computational examples

In this section, we present three numerical examples that use our full analytical sensitives for shape optimization. These examples, drawn primarily from [11], are commonly used examples in the shape optimization literature [22, 23, 24].

All problems are under plane stress conditions and the plate thickness $t_e = 1.0$. Unless otherwise specified, the convergence criteria used is the change of objective function values, i.e.

$$\varepsilon = \left| \frac{f^{(k)} - f^{(k-1)}}{f^{(0)}} \right|$$

where $f^{(k)}$ is the objective function value at the $k-$th iteration.

For the sake of simplicity in implementation, when mesh refinement is used, knots are recursively inserted at the parametric middle-point of every knot span.

6.1 Design of a plate with a hole

We begin our numerical examples with a classic shape optimization problem: optimizing the hole profile in a large plate under a biaxial stress field. The objective is to minimize the plate compliance under the constraint of material

volume. For an infinitely large plate, this problem has an analytical solution: circle under symmetric load and ellipse under asymmetric load. This makes it especially suitable to examine the role of weights in shape optimization since, without varying weights, NURBS degenerates into non-rational B-spline and cannot represent conic sections exactly.

The initial design, including the structural dimensions and loads, is shown in Fig. 4. It is modeled as a bi-quadratic NURBS surface with 4×3 control points. The knot vectors are {0,0,0,0.5,1,1,1} and {0,0,0,1,1,1}, thus leading to 2×1 knot spans. The Young's modulus is 210 and Poisson's ratio is 0.3.

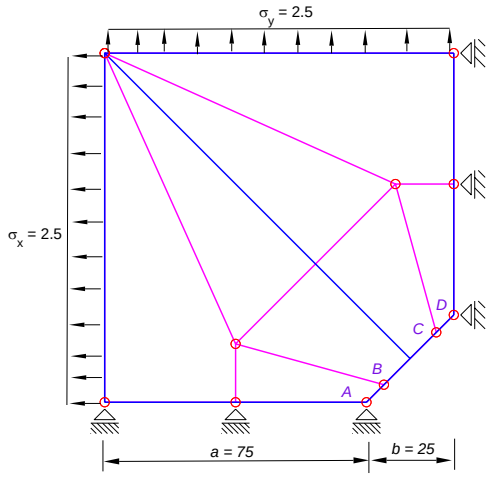
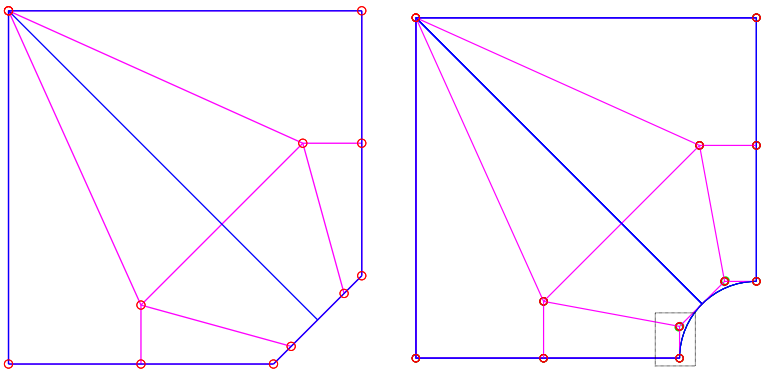


Figure 4: Initial design for a hole in a plate.

Design with both positions and weights

In this example, to ensure the optimality condition, we use the KKT norm used in the MMA algorithm as the termination criteria and it should be less than 6.0e-6. We optimize the hole shape under two different volume fraction constraints: $V \leq V^*$, 1) $V^* = 96\%$ and 2) V* = 99%. Here '%' refers to the percentage of the total volume $(10,000)$ of the square plate without the hole.

For the first type of design with the volume constraint $V^* = 96\%$, the initial hole shape is a straight line (Fig. 5). The model is refined 3×4 times, leading to $(2 \times 2^3) \times (1 \times 2^4)$ elements (knot spans). The optimization is conducted with 8 design variables, including x-coordinates for control points A, B, C and y-coordinates for control points B, C, D and weights for control points B, C. The optimized design is shown in Fig. 5. In addition, following the work in [11], we also optimized the hole profile by pre-setting the control points for node B and C as $w_B = w_C = 1$ (with these weights, the NURBS curve degenerates into a B-spline curve) and $w_B = w_C = (1+\sqrt{2}/2)/2$ (with these weights, the NURBS curve can exactly represent the circular profile). We plot the three optimized



(a) Initial design

(b) Optimized design

(c) Local view of optimized design

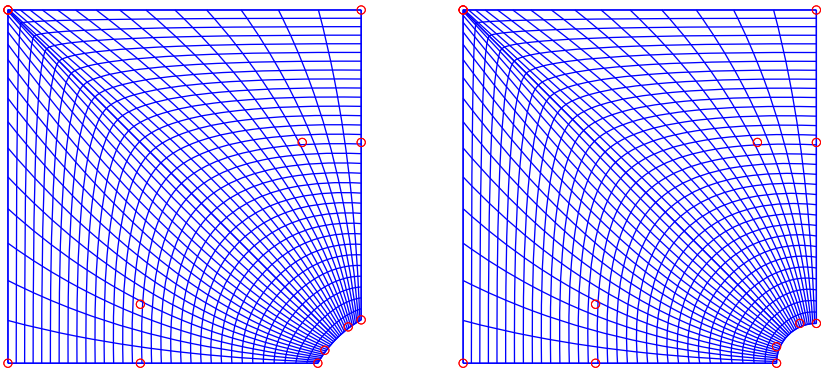
Figure 5: Hole profile optimization under $V^* = 96\%$ with both control points and weights as design variables.

Table 1: Optimal designs under the constraint $V^* = 96\%$

Type	Iteration	Compliance	Volume
Theoretical	N/A	466.5701	9600.0000
Initial	N/A	462.5555	9687.5000
8 variables,	23	466.5699	9599.9997
6 variables, $w_B = w_C = 1$	31	466.5708	9599.9996
6 variables, $w_B = w_C = 0.8536$	31	466.5701	9599.9996

profiles and the exact circule arc in Fig. 5.b. Visually speaking, the differences among the four profiles are indiscernable. In the view (Fig. 5.c) magnified at a scale comparable to that in [11], even though the control points are different for the three sets of optimized profiles, due to the influence of the weights, the hole profiles remain inseparable . This higher accuracy than that in [11] can be ascribed to the use of refined analysis model in our optimization.

The detailed comparison of compliance for the initial design, the optimized designs, and iteration times are shown in Table 1. This table demonstrates that optimization with both positions and weights of control points as design variables leads to smaller objective function value, i.e. compliance, than those without.



(a) Initial design (b) Optimized design

Figure 6: Hole profile optimization under $V^* = 99\%$ with both control points and weights as design variables. Red circles are control points for the design model. The blue curves are knot curves of the refined analysis model.

For the second type of design with the volume constraint $V^* = 99\%$, the initial hole shape and the final optimized shape after mesh refinement are shown in Fig. 6. Note, the initial 2×1 knot spans undergo 4×5 subdivisions, leading to $(2 \times 2^4) \times (1 \times 2^5)$ elements. Again, this optimized design with 8 design variables is compared with designs optimized with only control point positions as design variables and weights are set at 1 and $(1 + \sqrt{2}/2)/2$ for control point B, C. The

Table 2: Optimal designs under the constraint $V^* = 99\%$

Type	Iteration	Compliance	Volume
Theoretical	N/A	428.7086	9900.0000
Initial	N/A	428.1381	9910.5851
8 variables	6	428.7087	9899.9997
6 variables $w_B = w_C = 1$	5	428.7092	9899.9995
6 variables $w_B = w_C = 0.8536$	5	428.7087	9899.9995

detailed comparison of the optimization results and iteration times for these designs are shown in Table 2. Again, optimization with both positions and weights of control points as design variables leads to smaller objective function value, i.e. compliance, than those without.

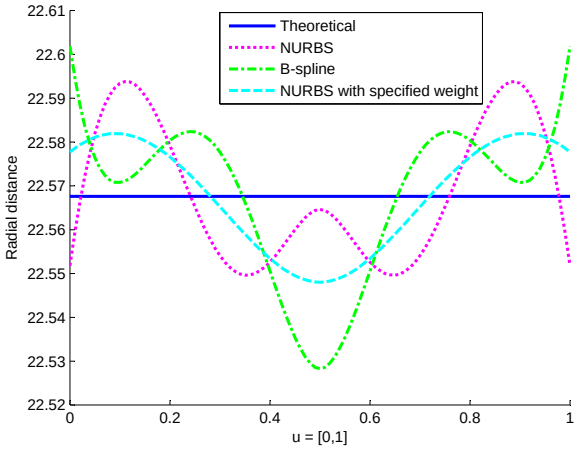


Figure 7: Radial distances from optimized holes to the circle center with both control points and weights as design variables. These holes are derived from different shape representations under $V^* = 96\%$.

Since the four profiles (theoretical circular arc, the NURBS profile optimized with 8 design variables, the B-spline profile optimized with 6 design variables and weights set to 1, and the NURBS profile optimized with 6 design variables but $w_B = w_C = (1 + \sqrt{2}/2)/2)$ are so close under each volume constraint, we plot in Fig. 7 and Fig. 8 the radial distances from points on these four curves to the theoretical circle center. Figure 7 shows, with $V^* = 96\%$, the deviation from the theoretical circle is about 0.33% percent for the B-spline representation (6 design variables) derived design and 0.18% for the NUBRS representation (8 design variables) derived design.

As V^* approaches 1, the finite plate approximates the infinite plate better and we thus expect the optimal profile closer to an exact circle. Figure 8 shows,

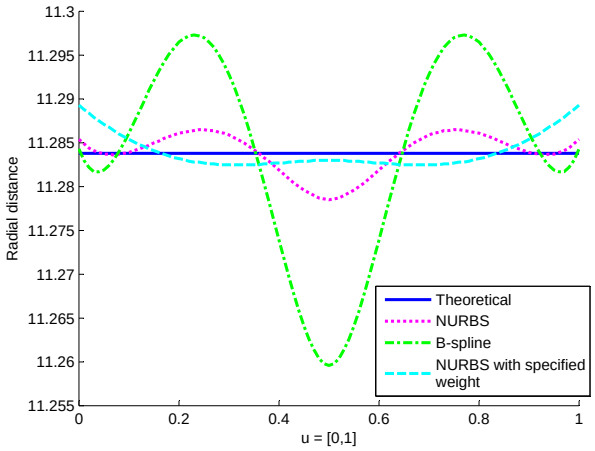


Figure 8: Radial distances from optimized holes to the circle center with both control points and weights as design variables. These holes are derived from different shape representations under $V^* = 99\%$.

with $V^* = 99\%$, the deviation is approximately 0.33% for the B-spline representation (6 design variables) derived design while less than 0.08% for the NUBRS representation (8 design variables) derived design. That is, for NURBS representation derived optimal designs, the deviations have become smaller from $V^* = 96$ to $V^* = 99$, yet this is not true for B-spline representation derived optimal designs. This is perhaps due to the fact B-spline cannot represent accurately the circular profile. For the NURBS representation with pre-specified weights ($w_B = w_C = (1 + \sqrt{2}/2)/2$), the deviation on the resulting profile is about the same as that on the NURBS repersentation derived profile. This further illustrates the significance of the use of weights as design variables in shape optimization.

Design with weights

Due to the use of both weights and positions of control points as design variables, the optimized profiles in the previous subsection are closer to the exact circle than those with positions only as design variables. In order to better understand and examine if exact circles can be recovered, we use the weights of Point B and Point C as design variables, and positions of control points are set for representing the exact circle and satisfying the volume constraint (Fig. 9). This figure shows the initial design with volume constraint of 96% and all weights are set to 1. The termination criteria is KKTnorm from the MMA code to be smaller than 1.0e-6. In the optimization, the initial model is subdivided 2×3 times for analysis.

Figure 10 shows, at three different initial conditions, 1) $w_B = w_C = 0.1$, 2) $w_B = w_C = 10$, and 3) $w_B = 0.1$, and $w_c = 10$, the optimized profiles recover

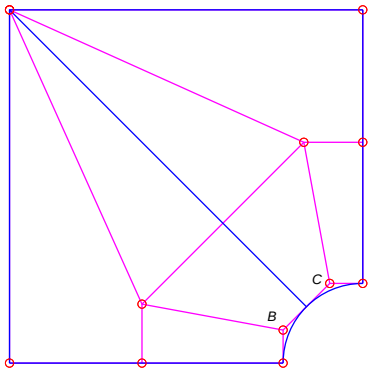


Figure 9: Control point positions for optimization with w_B and w_C as design variables.

well the exact circle (Fig. 9.d). Table 3 shows the recovered weights under these initial conditions. They deviate from the nominal weights $(1 + \sqrt(2)/2)/2 = 0.853553391$ for an exact circle less than 0.0001. Since the optimized profiles in the previous experiment have been shown to be very close to the exact circle, it is of no surprise that exact weights can be very well recovered.

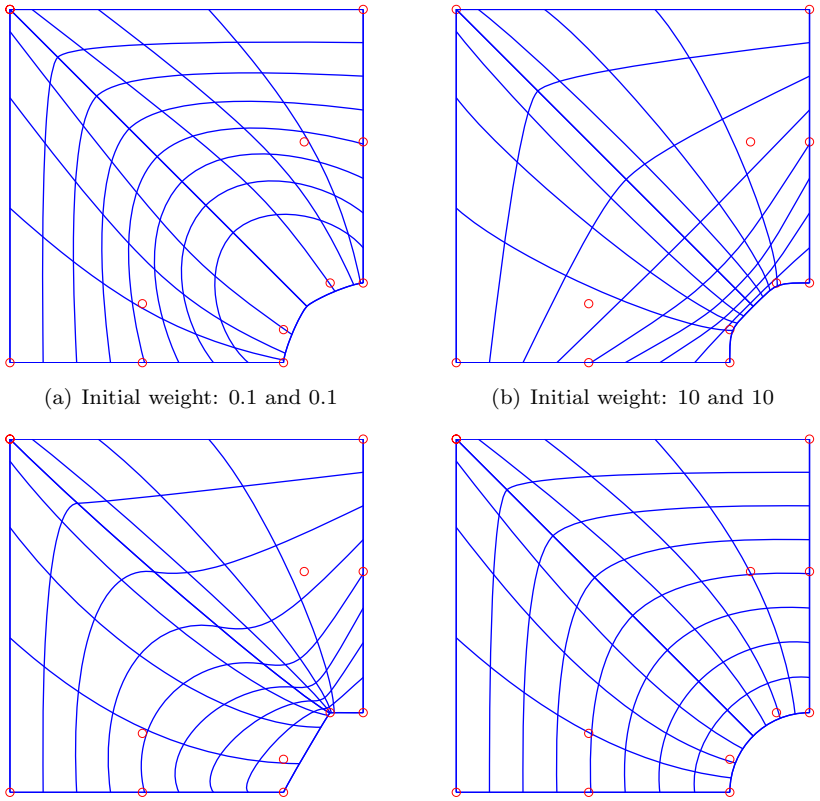
Table 3: Recovered weights under different initial conditions

$w_B^{(0)}$	$w_C^{(0)}$	Iteration (k)	$w_B^{(k)}$	$w_C^{(k)}$
0.1	0.1	5	8.535909e-001	8.535909e-001
10	10	8	8.535914e-001	8.535914e-001
0.1	10	12	8.535274e-001	8.536550e-001

Design with weights under asymmetric loads

We extend the above design with weights as design variables to design under asymmetric loads, i.e. $\sigma_x = -5$ and $\sigma_y = 2.5$. The theoretical prediction for an infinite plate is that the optimal hole is an ellipse with the ratio of major/minor axes equals to that of the load.

Figure 11 shows the initial position of the hole with weights $w_B = 0.1$, $w_C = 10$ and the resulting design under the volume constraint $V^* = 96\%$. Similar optimizations have been done for different volume constraints. The results are described in Table 4 where $w_B^0 = w_C^0 = (1+\sqrt{2}/2)/2$. The KKT norm 1.0e-6 is used for convergence termination. The analysis meshes are obtained with 4×5 subdivisions from the initial NURBS model, except for the 99.9% volume constraint with 4×6 subdivisions. The optimizations all converge after 9 to 12 iterations.



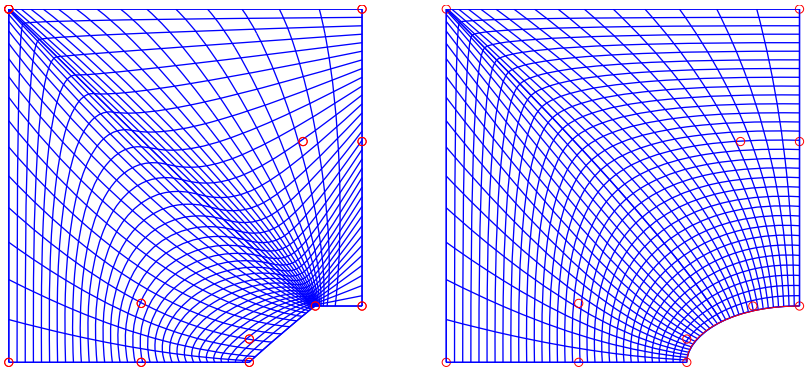
(a) Initial weight: 0.1 and 0.1

(b) Initial weight: 10 and 10

(c) Initial weight: 0.1 and 10

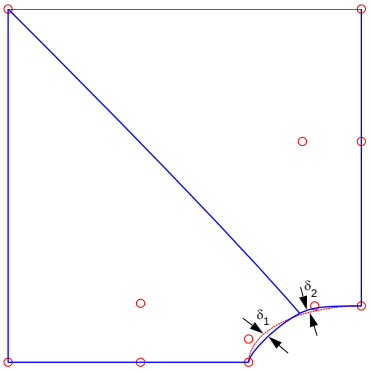
(d) Optimized design

Figure 10: Design optimization with weights as design variables under different initial conditions. Red circles are control points for the design model. The blue curves are knot curves in the refined analysis model.

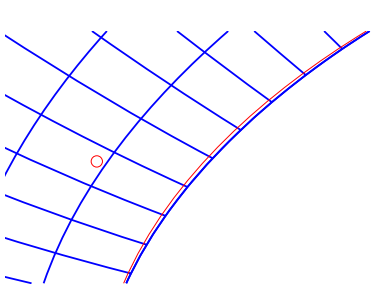


(a) Initial design (b) Optimized design

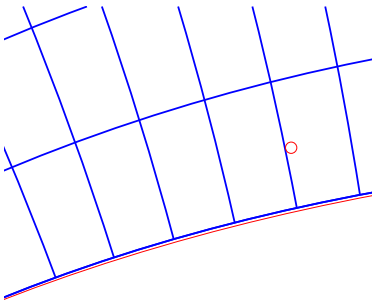
Figure 11: Optimized design under asymmetric load with $V^* = 96\%$ with weights as design variables. The red profile is an exact ellipse satisfying the volume constraint.



(a) general trend



(b) Magnified view of Fig. 11.b near control point B



(c) Magnified view of Fig. 11.b near control point C

Figure 12: Optimized hole shape under asymmetric load with weights as design variables. Dotted red curve representing the exact ellipse.

Table 4: Optimal designs under asymmetric load

Volume fraction	w_B	w_C	$(w_B - w_B^0)/w_B^0$	$(w_C - w_C^0)/w_C^0$
85%	0.78752	0.99992	-7.74%	17.15%
90%	0.78772	0.99880	-7.71%	17.02%
96%	0.81705	0.90446	-4.28%	5.96%
99%	0.84072	0.86822	-1.50%	1.72%
99.5%	0.84148	0.86756	-1.41%	1.64%
99.9%	0.84373	0.86748	-1.15%	1.63%

Table 4 shows the computational results for the finite-dimensioned plate nearly agree with the optimal result of the infinite plate. Note, the optimal designs for the finite and infinite plates are distinct [24, 25]. As the hole becomes smaller (the volume fraction of the solid approaches unity), the difference between the two sets of optimal design also become smaller as seen in Table 4. At 99.9%, the errors for the weights of the point B and point C are about a little over one percent.

An interesting trend can also be observed in Table 4: w_B is consistently below the weight for the exact ellipse and w_C consistently above it. As the volume fraction for the solid increases, the deviations become smaller. This suggests the optimal shape for a finite plate under asymmetric load would look as exaggerated in Fig. 12.a. As the volume fraction increases, δ_1 and δ_2 representing the deviation to the exact ellipse as shown in Fig. 12.a become smaller. A magnified view of the optimized profile and the exact ellipse in Figure 11.b is shown in Fig. 12.b and it demonstrates just this.

Optimal or not

One major advantage of having analytical sensitivities is that they can be used in understanding whether a given shape is optimal or not by checking the KKT condition.

For an objective function $f(\mathbf{x})$ subject to the constraint $g(\mathbf{x}) \leq 0$, when point $\mathbf{x}^*$ is a regular point, then if $\mathbf{x}^*$ is an optimal point, we have $\nabla f + \lambda \nabla g = 0$ where λ is the Lagrange multiplier, and ∇ is the gradient w.r.t. the design variables α_s. Here we assume the constraint is active.

We thus propose the use of KKT norm as defined below to determine whether a given shape is optimal or not. For each design variable α_s, we define a number μ_s as follows:

$$\mu_s = -\frac{\partial f}{\partial \alpha_s} \Big/ \frac{\partial g}{\partial \alpha_s}.$$

From μ_s, we can define a pseudo Lagrange multiplier $\widetilde{\mu}$

$$\widetilde{\mu} = \frac{\sum_{s=1}^{n_{eq}} \mu_i}{n_{eq}}$$

where n_{eq} is the number of design variables.

The KKT norm with the pseudo Lagrange multiplier $\widetilde{\mu}$ is then defined as follows

$$\mathbf{V} = \left[\frac{\partial f}{\partial \alpha_i} + \widetilde{\mu}\frac{\partial g}{\partial \alpha_s}\right], \quad \text{where } s = 1 \text{ to } n_{eq},$$

$$\text{KKTNorm} = \sqrt{\mathbf{V}^T\mathbf{V}}.$$

When a given shape is optimal, we would have the pseudo Lagrange multiplier $\widetilde{\mu}$ equal the real multiplier λ, i.e. $\mu_s = \widetilde{\mu} = \lambda$, and the KKT norm equal zero.

We compute the above defined KKT norm for a circular hole profile under symmetric load $\sigma_x = -\sigma_y$ and elliptical hole profiles under asymmetric load $\sigma_x = -2 * \sigma_y$. The results are shown in Table 5. The analytical hole (circular and elliptical) profiles are generated to satisfy the volume constraints listed in the table. There are 8 design variables, including positions of four control points and weights of the two middle control points. The mesh density is also described in the table.

Table 5: KKT norm of circles and ellipses under symmetric and asymmetric loads

Load	Mesh	Volume fraction	KKT norm
symmetric	32×32	96%	1.4611e-5
asymmetric	32×32	85%	0.0045
asymmetric	32×32	90%	0.0019
asymmetric	32×32	96%	3.0563e-4
asymmetric	32×32	99%	2.8213e-5
asymmetric	32×32	99.5%	1.9374e-5
asymmetric	32×64	99.9%	1.1935e-5

Table 5 demonstrates the following:

- Circular hole under symmetric load at 96% volume constraint is very close to the optimal shape since its KKT norm is small. This explains why the weights computed earlier are very close to the analytical weights for an exact circle.
- Under asymmetric load, below 96% volume constraints, the elliptical profiles are clearly not optimal since their corresponding KKT norm is relatively large.
- As the hole becomes even smaller, i.e. V^* becomes larger, the KKT norm becomes smaller. This suggests the elliptical profile approximates the true optimal profile better. This agrees well with the theoretical result that the optimal hole profile under asymmetric load is elliptical for an infinitely large plate.

6.2 Design of a cantilever beam

The goal in the design of a cantilever beam is to minimize the vertical displacement at the point of load under a volume constraint.

The initial design of the beam is shown in Fig. 13. The beam has a predetermined length $l = 30$ and an allowable height of $1.5 \leq h \leq 10$. The beam is modeled using a 6×2 control net, quadratic in the ξ_1 and linear in ξ_2 directions. The knot vector is $\{0, 0, 0, 0.25, 0.5, 0.75, 1, 1, 1\}$ in ξ_1 and $\{0, 0, 1, 1\}$ in ξ_2. leading to 4×1 knot spans. The initial design is a rectangle with $h = 6$. The initial weights are set to 1 and the weights are bound to be between 0.1 and 1 during the iteration process. The Young's modulus is 200×10^3 and Poisson's ratio is 0.3. The convergence criteria is $\varepsilon = 1.0e - 7$.

We consider two design cases. In the first one, there are total 6 design variables and they are vertical positions of control points in the upper row. In the second one, there are 12 design variables and they include both positions and weights of control points in the upper row. In each case, the volume is constrained to be 70% of the maximum area of 300. The load is 10.

To ensure reasonably accurate analysis during the optimization, the design model is subdivided 4×5 times, leading to total $(4 \times 2^4) \times (1 \times 2^5) = 64 \times 32 = 2048$ elements for analysis.

Optimized designs for the two cases are shown in Fig. 14 and Fig. 15. These designs are reached after 12 and 43 iterations with the convergence criteria of $\varepsilon = 1.0e - 7$. The convergence history of the vertical displacement and volume constraints for both cases are shown in Fig. 16 and 17. The second design with both positions and weights as design variables leads to smaller displacement, -1.037738e-2 versus -1.055000e-2 in the first design. The weights for the first, third and sixth control points become 0.1 while others remaining nearly at 1.0, thus leading to internal isoparametric curves moving away from these control points, as shown in Fig. 15

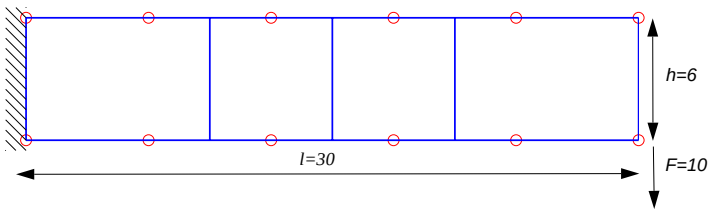


Figure 13: Initial design of a cantilever beam, control net (consisting of a 6×2 control points in red circles) and the loading condition.

The weight-induced uneven distribution of isoparametric curves in the optimized result (Fig. 15) is very apparent. Without mesh refinement, (in this case 4×5 subdivisions), the analysis on the distorted and sparse mesh would

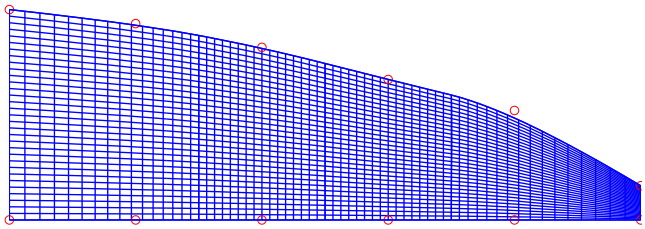


Figure 14: The optimized design and the knot curves used in analysis where only positions of control points at the top row are used as design variables.

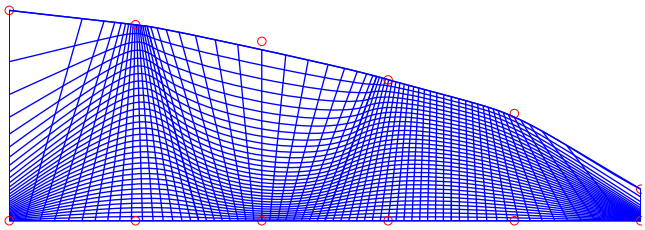


Figure 15: The optimized design and the knot curves used in analysis where both positions and weights of control points at the top row are used as design variables.

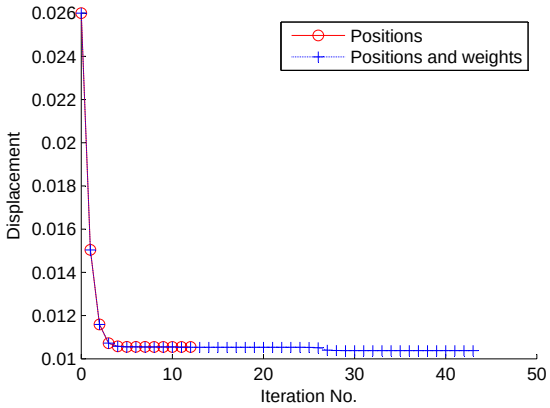


Figure 16: Convergence history for vertical displacement at the point of load for two design cases: positions of control points as design variables and both positions and weights of control points as design variables .

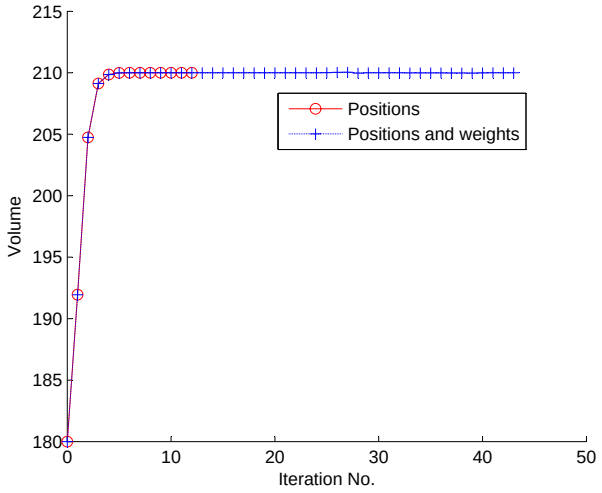


Figure 17: Convergence history for the volume constraint for two design cases: positions of control points as design variables and both positions and weights of control points as design variables.

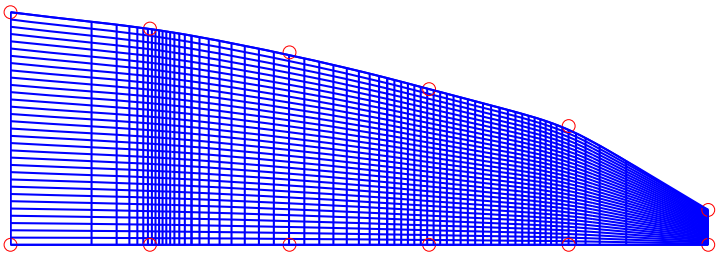


Figure 18: The optimized design and the knot curves used in analysis where both control point positions and weights at the top row are used as design variables and the weights of the control points in the bottom row are set the same as their counterparts in the top row.

not be accurate. Since the knots are inserted in the parametric middle point of each knot span, the distribution of mesh elements is still not uniform. More sophisticated mesh smoothing methods such as those used for addressing the problem of "floating meshes" [26] can perhaps be developed to insert the knots in an efficient manner that would result in a more uniformly distributed elements with fewer knot insertions. A simpler alternative, but not a complete remedy, for this example is implemented. It set the weights of control points in the bottom row equal to their counterparts in the top row. The result is shown in Fig. 18. The displacement after optimization is -1.030668e-2 and it took 28 iterations.

6.3 Design of an open spanner

The goal is to design the outer shape of a full open end spanner. The design objective is to minimize the displacements in two loading cases $u_{F_A} - u_{F_B}$ for a given material volume, with u_{F_A} and u_{F_B} as vertical displacements at the point of loads F_A and F_B respectively. The boundary condition for load F_A and the initial design is shown in Fig. 19. The boundary condition for load F_B is symmetric to that of F_A. The Young's modulus is 210×10^3 and Poisson's ratio is 0.3. The convergence criteria is $\varepsilon = 1.0e - 6$.

The structure has a pre-defined length $l = 25$, maximum height $h = 10$ and minimum handle thickness 2. The bolt shape and size ($b = 2$) is fixed. We use a bi-quadratic control net to represent the structure. In total we have 10×9 control points. The initial weights are all set to 1. The volume for the optimal design is constrained to be 35% of the initial volume.

We consider the following two design scenarios:

- *Control points alone as design variables*: The vertical positions of the control points at the top and bottom sides are design variables. In order to ensure the handle is straight, the three control points at the top and bottom right ends are set to equal to each other. Thus, there are total 12 design variables.

- *Control points and weights as design variables*: In addition to the 12 control points as in the above case, weights for five control points at the top and bottom left end are set as design variables. The outer limits for the weights are 0.1 and 1. Thus, there are total 22 design variables in this case.

To ensure the analysis accuracy, each knot span in ξ_1 and ξ_2 is subdivided into two. The original internal nodes are unchanged during the optimization and the newly created internal nodes from subdivision are updated automatically according to the coordinates and weights of outer control points.

The optimized solutions are reached after 18 iterations in the first case and 17 in the second one. The results are shown in Fig. 20 and 21. Both designs recover well the known shape of a spanner and compare well with the results obtained in the literature [22, 23, 11]. The resulting displacement at the point

of loads in the first design (0.2886) is slightly larger than that (0.2411) in the second design. Further, (nearly) sharp transitions occur in the upper and lower side of the second design (*i.e.* optimized with both control points and weights as design variables shown in Fig. 21). In this optimized design, the weight for the protruded control point (the third from the left end) near the sharp transition is 0.4365 while the weights for the neighboring control points are 0.1, thus the appearance of the somewhat sharp transition. However, since there is no knot repetition at this point, the contour is still C^1 smooth.

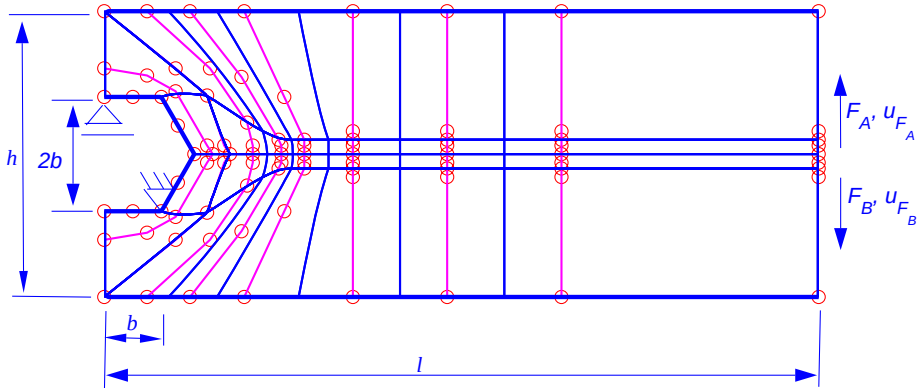


Figure 19: Initial design and two loading conditions of the open spanner problem. Red, bold outer shape represents the initial design. The bi-quadratic control net (consisting of 10×9 control points) is shown in magenta and knot curves in blue.

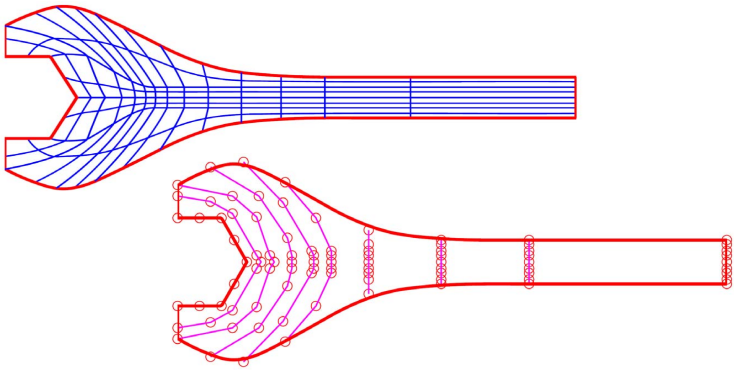


Figure 20: Optimal design and the knot curves of the open spanner with control points at the top and bottom rows as design variables.

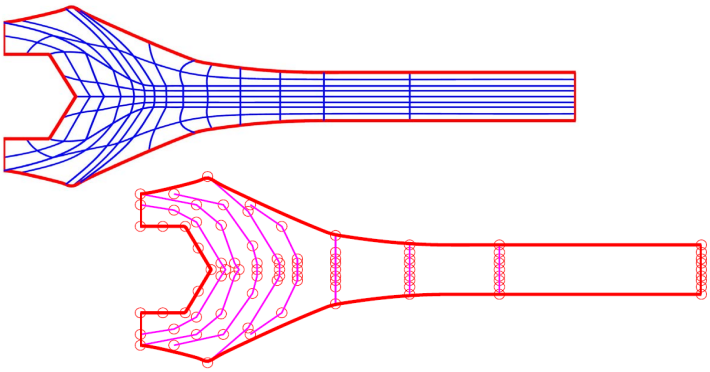


Figure 21: Optimal design and the knot curves of the open spanner with both control point positions and weights for five control points at the top and bottom left end as design variables.

7 Concluding remarks

This paper presents analytical formulas for computing full sensitivities of both positions and weights of NURBS control points in shape optimization. Such analytical formulation allows accurate calculation of sensitivity and is useful in gradient-based shape optimization. The analytical sensitivity is also useful in determining whether a given shape is optimal.

The full analytical sensitivity for both positions and weights of NURBS control points is especially beneficial for recovering optimal shapes that are conical e.g. ellipses and circles in 2D, cylinders, ellipsoids and spheres in 3D that are otherwise not possible without the weights as design variables.

NURBS weights have lent extra flexibility for shape parameterization. The use of combined control points and weights has demonstrated consistently better design in terms of lowering the objective function.

Our implementation is based on isogeometric analysis since its computation is more accurate on a per-node basis. However, the analytical formulas of full sensitivities are not limited to isogeometric analysis, these formulas are also applicable to standard FEM based shape optimization.

Due to the use of weights as design variables, isogeometric shape optimization may lead to uneven NURBS weights, thus leading to distorted distribution of knot curves and element boundaries. Mesh refinement through knot insertion has shown to be effective in resolving such distortion. Future work will further investigate the use of mesh smoothing techniques to guarantee high-quality discretization.

Acknowledgments

The author is thankful for the financial support from the US National Science Foundation, including grant #0900597, grant #0900170 and grant #0800912.

The author would like to thank Krister Svanberg from Royal Institute of Technology, Sweden, for sharing his MMA code. The author also wants to thank Pinghai Yang at Illinois Institute of Technology for his contribution in the earlier coding of isogeometric analysis.

The communication with W. A. Wall and M. A. Frenzel at Technical University of Munich, Germany, has been helpful. Reviewers' constructive remarks have also improved this paper.

Appendix: Proof on analytical sensitivities in NURBS based isogeometric shape optimization

We show below the derivation process of obtaining the expressions for $\mathbf{G}'$, $|\mathbf{J}|'$, $\left(N_{i,\xi_j}\right)'$, N_i', and $\boldsymbol{x}'$ in Theorem 1. They are obtained through the chain rule of differentiation and Jacobi's formula. The approach we are going to follow below was initially reported in [16] and neatly organized in [17, 18].

For the two matrices ($\mathbf{G}$ and $\widehat{\mathbf{G}}'$) defined in Section 4.1 , we have $\widehat{\mathbf{G}} = \mathbf{JG}$. Taking the derivative with respect to the design variable α_s, we have $\widehat{\mathbf{G}}' = \mathbf{J}'\mathbf{G} + \mathbf{J}\mathbf{G}'$. Thus, we obtain the expression for $\mathbf{G}'$ as follows

$$\mathbf{G}' = \mathbf{J}^{-1}\left(\widehat{\mathbf{G}}' - \mathbf{J}'\mathbf{G}\right). \tag{21}$$

Meanwhile, from the NURBS geometry definition $\boldsymbol{x} = \mathbf{N}^T\mathbf{P}$, we have $\mathbf{J} = \widehat{\mathbf{G}}\mathbf{P}$. Taking the derivative on both sides with respect to the design variable α_s, we have

$$\mathbf{J}' = \widehat{\mathbf{G}}\mathbf{P}' + \widehat{\mathbf{G}}'\mathbf{P}. \tag{22}$$

Combinging Eq. (21) and Eq. (22), we have

$$\begin{aligned}
\mathbf{G}' &= \mathbf{J}^{-1}\widehat{\mathbf{G}}' - \mathbf{J}^{-1}\left(\widehat{\mathbf{G}}\mathbf{P}' + \widehat{\mathbf{G}}'\mathbf{P}\right)\mathbf{G} \\
&= \mathbf{J}^{-1}\widehat{\mathbf{G}}' - \mathbf{J}^{-1}(\widehat{\mathbf{G}}\mathbf{P}')\mathbf{G} - \mathbf{J}^{-1}\widehat{\mathbf{G}}'\mathbf{P}\mathbf{G} \\
&= \mathbf{J}^{-1}\widehat{\mathbf{G}}'(\mathbf{I} - \mathbf{P}\mathbf{G}) - \mathbf{G}\mathbf{P}'\mathbf{G}.
\end{aligned}$$

This corresponds to Eq. (13) in Theorem 1.

In order to compute $|\mathbf{J}|'$, we utilize the Jacobi's formula, $|\mathbf{A}|' = |\mathbf{A}|tr(\mathbf{A}^{-1}\mathbf{A}')$, where A is a nonsingular matrix function.

Therefore, we have the following expression for $|\mathbf{J}|'$

$$\begin{aligned}
|\mathbf{J}|' &= |\mathbf{J}|tr\left(\mathbf{J}^{-1}\mathbf{J}'\right) \\
&= |\mathbf{J}|tr\left(\mathbf{J}^{-1}(\widehat{\mathbf{G}}\mathbf{P}' + \widehat{\mathbf{G}}'\mathbf{P})\right) \\
&= |\mathbf{J}|tr\left(\mathbf{J}^{-1}\widehat{\mathbf{G}}\mathbf{P}' + \mathbf{J}^{-1}\widehat{\mathbf{G}}'\mathbf{P}\right) \\
&= |\mathbf{J}|tr\left(\mathbf{G}\mathbf{P}' + \mathbf{J}^{-1}\widehat{\mathbf{G}}'\mathbf{P}\right)
\end{aligned}, \tag{23}$$

which corresponds to Eq. (12) in Theorem 1.

The terms in $\widehat{\mathbf{G}}'$ include $\partial^2 N_i/\partial\xi_j\partial\alpha_s$. Since we need to calculate $\partial N_i/\partial\xi_j$ anyway in finite element analysis, the additional task would be to calculate their derivative over α_s.

The NURBS basis function's derivative over the parametric coordinates is

$$N_{i,\xi_j} = \frac{R_{i,\xi_j} w_i}{\mathbf{R}^T\mathbf{W}} - \frac{R_i w_i (\mathbf{R}_{,\xi_j})^T \mathbf{W}}{(\mathbf{R}^T\mathbf{W})^2}.$$

Taking another derivative over the design variable leads to the expression in Theorem 1.

References

[1] Braibant, V., and Fleury, C., 1984. "Shape optimal design using b-splines". *Computer Methods in Applied Mechanics and Engineering,* **44**, pp. 247 – 267.

[2] Samareh, J. A., 1999. *A Survey Of Shape Parameterization Techniques.* NASA Langley Technical Report.

[3] Hughes, T. J. R., Cottrell, J. A., and Bazilevs, Y., 2005. "Isogeometric analysis: Cad, finite elements, nurbs, exact geometry and mesh refinement". *Computer Methods in Applied Mechanics and Engineering,* **194**, pp. 4135–4195.

[4] Qian, X., and Dutta, D., 2003. "Physics-based modeling for heterogeneous objects". *ASME Transactions Journal of Mechanical Design,* **125**, pp. 416–427.

[5] Yang, P., and Qian, X., 2007. "A b-spline based approach to heterogeneous object design and analysis". *Computer-Aided Design,* **34**(2), pp. 95–111.

[6] Schramm, U., and Pilkey, W. W., 1993. "The coupling of geometric descriptions and finite elements using nurbs - a study in shape optimization". *Finite Elements in Analysis and Design,* **15**, pp. 11 – 34.

[7] Choi, J. H., 2002. "Shape design sensitivity analysis and optimization of general plane arch structures". *Finite Elements in Analysis and Design,* **32**, pp. 119 – 136.

[8] Nadir, W., Kim, I. Y., and de Weck, O. L., 2004. "Structural shape optimization considering both performance and manfuacturing cost". In 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference.

[9] Zhang, X., Rayasam, M., and Subbarayan, G., 2007. "A meshless, compositional approach to shape optimal design". *Computer Methods in Applied Mechanics and Engineering,* **196**, pp. 2130 – 2146.

[10] Silva, C. A. C., and Bittencourt, M. L., 2007. "Velocity fields using nurbs with distortion control for structural shape optimization". *Struct. Multidiscip. Optim.,* **33**, pp. 147 – 159.

[11] Wall, W. A., Frenzel, M. A., and Cyron, C., 2008. "Isogeometric structural shape optimization". *Comput. Methods Appl. Mech. Engrg.,* **197**, pp. 2976 – 2988.

[12] Cho, S., and Ha, S. H., 2009. "Isogeometric shape design optimization: exact geometry and enhanced sensitivity". *Struct. Multidiscip. Optim.,* **38**, pp. 53 – 70.

[13] Poueymirou, D., Tribes, C., and Trepanier, J. Y., 2004. "A nurbs-based shape optimization method for hydraulic turbine stay vane". In Proceedings of the Third International Conference on Computational Fluid Dynamics, ICCFD3, Toronto, 12-16 July 2004, pp. 415 – 421.

[14] Nagy, A. P., Abdalla, M. M., and Gurdal, Z., 2010. "Isogeometric sizing and shape optimization of beam structures". *Computer Methods in Applied Mechanics and Engineering,* **199**, pp. 1216–1230.

[15] Haftka, R. T., and Gurdal, Z., 1992. *Elements of Structural Optimization.* Kluwer Academic Publishers.

[16] Brockman, R. A., 1987. "Geometric sensitivity analysis with isoparametric finite elements". *Com. Appl. Numer. Methods,* **3**, pp. 495–499.

[17] Haslinger, J., and Makinen, R. A. E., 2003. *Introduction to Shape Optimization: Theory, Approximation, and Computation.* SIAM, Philadelphia.

[18] Christensen, P. W., and Klarbring, A., 2009. *An Introduction to Structural Optimization.* Springer.

[19] Piegl, L., and Tiller, W., 1997. *The NURBS Book.* Springer-Verlag, New York.

[20] Fish, J., and Belytschko, T., 2007. *A First Course in Finite Elements.* John Wilsey and Sons, Ltd.

[21] Svanberg, K., 1987. "The method of moving asymptotes: A new method for structural optimization". *International Journal of Numerical Methods in Engineering,* **24**, pp. 359 – 373.

[22] Herskovits, J., Dias, G., Santos, G., and Soares, C. M., 2000. "Shape structural optimization with an interior point nonlinear programming algorithm". *Struct. Multidiscip. Optim.,* **20**, pp. 107–115.

[23] Wilke, D., Kok, S., and Groenwold, A., 2006. "A quadratically convergent unstructured remeshing strategy for shape optimization". *Int. J. Numer. Meth. Engrg.,* **65**, pp. 1–17.

[24] Norato, J., Haber, R., Tortorelli, D., and Bendsoe, M. P., 2004. "A geometry projection method for shape optimization". *International Journal for Numerical Methods in Engineering,* **60**, pp. 2289–2312.

[25] Pedersen, P., 2000. "On optimal shapes in materials and structures". *Structural and Multidisciplinary Optimization,* **19**, pp. 169 – 182.

[26] Bletzinger, K. U., Firl, M., Linhard, J., and Wüchner, R., 2010. "Optimal shapes of mechanically motivated surfaces". *Computer Methods in Applied Mechanics and Engineering,* **199**, pp. 324 – 333.