

INDUCING CONSTRAINT ACTIVITY IN INNOVATIVE DESIGN

JONATHAN CAGAN¹ AND ALICE M. AGOGINO²

¹Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 and ²Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94720, U.S.A.

In this paper, a methodology for inducing trends in a first principle reasoning system for design innovation is presented. Dimensional Variable Expansion is used in 1stPRINCE (FIRST PRINciple Computational Evaluator) to create additional design variables and introduce new prototypes. Trends are observed at each generation of the prototype and induction is used to predict optimal constraint activity at the limit of the iterative procedure. The inductive mechanism is applied to a constant-radius beam under flexural load and a tapered beam of varying radius and superior performance is derived. A circular wheel is created from a primitive-prototype consisting of a rectangular, spinning block that is optimized for minimum resistance to spinning. Although presented as a technique to perform innovative design, the inductive methodology can also be utilized as an AI approach to shape optimization.

1. Introduction

In an earlier paper (Cagan and Agogino, 1987), the 1stPRINCE (FIRST PRINciple Computational Evaluator) methodology was introduced as a means to innovate new design concepts and expand the variable space in optimal design. Because each iteration of 1stPRINCE has the potential to create new design variables and constraints, the constraint activity is re-evaluated for each design generation. In this paper we present a method to induce constraint activity that potentially leads to the limit of the converging design of the innovating 1stPRINCE procedure. Application of this inductive procedure to a beam under flexural load innovates a tapered beam, optimized for minimum weight. The induced optimally directed shape of a spinning block that minimizes resistance to spinning leads to the discovery of a circular wheel.

The 1stPRINCE methodology innovates new prototypes based on an original primitive-prototype. Here we define a *primitive-prototype* as the model of a design problem specified by an objective function and a set of inequality and equality constraints within a design space. A *prototype* is a potentially optimal set of active constraints from a primitive-prototype that can be instantiated to at least one feasible design artifact. We then classify designs based on the prototype as *routine* or *non-routine*. We define a *routine design* as a prototype with the same set of variables or features as a previous prototype; the

structure of the prototype does not change. We define a *non-routine design* as a prototype with an expanded set of variables or features as compared to a previous prototype; new variables or features are introduced in the structure of the prototype.

Non-routine designs are further classified as *innovative* or *creative*. *Innovative designs* demonstrate new design variables or features in a prototype based on existing variables or features from a previous prototype. *Creative designs* introduce new design variables or features in a prototype demonstrating no obvious similarity to variables or features in a previous prototype. As the design features produced by 1stPRINCE are derived from existing prototypes, the method is classified as an innovative design process.

2. 1stPRINCE

The 1stPRINCE design methodology expands the design space by dividing a body along a critical variable where a critical variable is one that affects the objective function and when expanded will create new variables which also influence the objective. The expansion is performed by a formal technique called *Dimensional Variable Expansion* (DVE) presented by Cagan (1990). A less formal presentation but more intuitive description of DVE was presented by Cagan and Agogino (1987) as *integral division* which will be briefly summarized below.

A continuous integral of a function of variables χ

Received 30 January 1991 and in revised form 21 May 1991.

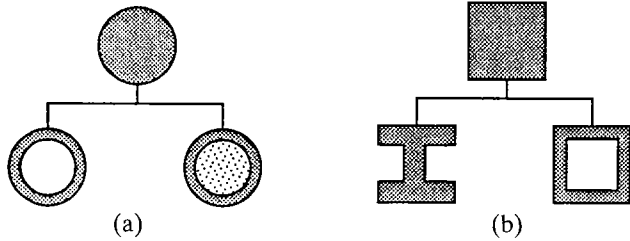


FIGURE 1. 1stPRINCE innovation on (a) solid rod and (b) solid bar primitives-prototypes

and w , with integration limits z_i to z_f , can be divided into a series of continuous integrals over m smaller ranges as:

$$\int_{z_i}^{z_f} f_i(\chi, w) dw = \int_{z_0}^{z_1} f_i(\chi^1, w) dw + \dots + \int_{z_{m-1}}^{z_m} f_i(\chi^m, w) dw, \quad (1)$$

where limit of integration z_0 replaces initial limit z_i , and z_m replaces initial limit z_f . If the body remains homogeneous after division, the equality in equation (1) remains consistent; analysis of either side of the equality should produce the same results and the same design. 1stPRINCE uses the right hand side of equation (1) as a starting point for its procedure. The integral is discretized over a critical variable and then discontinuities in properties within each subregion are permitted. Thus after application of 1stPRINCE, the equality in equation (1) no longer applies; rather, a completely different prototype than the one implied in the left hand side of the equation may result.

By division of an integral over a critical variable and by permitting discontinuities across the geometric axes, 1stPRINCE expands the design space through the introduction of new variables and constraints. In all the examples in this paper, integrals are divided into two smaller-ranged integrals ($m=2$) at each iteration. By minimizing weight, 1stPRINCE has innovated hollow tubes and composite rods from a solid cylindrical rod under torsion load [Figure 1(a)] where the solutions are presented in closed form and are optimally directed (Cagan and Agogino, 1987). From a solid rectangular cross-section rod under flexural load [Figure 1(b)], a hollow tube and I-beam are innovated¹ (Cagan, 1990), Figure 1(b). This paper will examine the application of 1stPRINCE to a beam

¹ Note, this points out an error in the analysis of 1stPRINCE in an earlier publication by Howard *et al.* (1989, p. 118). The authors in that publication claim that 1stPRINCE cannot innovate an I-beam. Quite the contrary, it is exactly this kind of first principle reasoning about stress distributions and material properties that 1stPRINCE excels at.

under flexural load by consideration of length as a critical variable while minimizing weight and to a spinning block while minimizing resistance to spinning.

3. Monotonicity analysis and symbolic computation

The prototypes innovated by 1stPRINCE are *optimally directed*. Optimally directed design is an approach to design which attempts to determine optimal regions of the design space by directing the search toward improving the objectives and eliminating suboptimal or dominated regions. The result is to reduce the size of the search space and gain insight as to the desirable directions for improving the design variables. "Optimally directed design" is distinguishable from "optimal design" in that the goal is not to find a single optimum for a specific set of numerical parameters, but to identify, in symbolic form where possible, insight as to how the design might be improved relative to the objectives.

Given a primitive-prototype, 1stPRINCE utilizes monotonicity analysis to search the design space in an optimally directed manner and identify critical variables for Dimensional Variable Expansion. Monotonicity analysis is a symbolic approach to non-linear optimization problems based on a qualitative form of the Karush-Kuhn-Tucker optimality conditions. Originally developed by Papalambros and Wilde (1979, 1982) and Wilde (1986) and automated symbolically with AI technology by Choy and Agogino (1986), the analysis utilizes qualitative first derivative information (i.e., the algebraic sign of the direction of change of a variable) to help determine which constraints will be active or inactive for a possible solution to the optimization problem.

Three rules of monotonicity analysis define well-constrained optimization problems (Papalambros and Wilde, 1979, 1982) without overconstrained cases (Wilde, 1986).

Rule One: If the objective function is monotonic with respect to (w.r.t.) a variable, then there exists at least one active constraint that bounds the variable in the direction opposite of the objective. A constraint is active if it acts at its lower or upper bound.

Rule Two: If a variable is not contained in the objective function then it must be either bounded from both above and below by active constraints or not actively bounded at all (i.e. any constraints monotonic w.r.t. that variable must be inactive or irrelevant).

Rule Three (The Maximum Activity Principle): The

number of non-redundant active constraints cannot exceed the total number of variables.

1stPRINCE performs a qualitative analysis of an optimal design problem through monotonicity analysis. Certain valid, unique solution domains occur via unconditionally active and conditionally active constraints. Important information can be found from the relationships of the various parameters within those active constraints. A mathematical analysis can then be performed on each of the cases by back-substituting the known, active information into the objective function using symbolic algebra routines, and symbolically applying the Karush–Kuhn–Tucker first order conditions of optimality (as detailed in Agogino and Almgren, 1987; Almgren and Agogino, 1989). When an inequality constraint is deemed active, that constraint is at its bound and the values of the variables are known to equal the bounded values. The constraint can then be assumed to be a relevant equality constraint within the given solution. All active constraints must be used during back-substitution since each constraint contains important information about the optimal solution. Back-substitution of the prototype constraints leads to partial solutions in symbolic form that require less numerical computation than that of the original problem (e.g., see Jain and Agogino, 1990). In many cases, the complete solution can be obtained in closed form and no further numerical analysis is required. Detailed application of monotonicity analysis within the 1stPRINCE method is given in Cagan and Agogino (1987). In this paper only the results from the analysis will be presented.

4. Inducing constraint activity

In this paper, inductive learning is presented as a descriptive generalization for learning from examples to discover patterns in observational data (Michalski, 1983). The algorithm described in this section utilizes a heuristic concept of induction whereby if some set of facts Π is true for a sequence of n steps, then Π is induced to be true for all steps greater than n . As n approaches infinity, Π is continuously true. As a program generates data, patterns can be observed. If the patterns are valid for n consecutive iterations then they can be induced to be continuously valid.

1stPRINCE performs inductive inference by observing patterns on constraint activity. Monotonicity analysis is used to derive sets of active constraints. Each iteration of 1stPRINCE applies a mathematical manipulation, requiring a new monotonicity analysis

to be performed. Constraints of each generation are mapped back to the constraints from which they were derived in regions of previous generations. A prototype is considered the next *generation* of a different prototype if the former prototype was derived from the latter by a single iteration of 1st-PRINCE. Patterns are observed in constraint activity for each generation. In the implementation of 1st-PRINCE, the number of steps (n) observed by the process can be chosen by the user but is defaulted to be three. We define the following:

Inductively active: If a constraint is active for n consecutive generations, then it is induced to be continuously active. As with unconditionally active constraints, variable back-substitution can be performed.

Inductively inactive: If a constraint is inactive for n consecutive generations, then it is induced to be continuously inactive. As with unconditionally inactive constraints, the constraint can be removed from the constraint set.

Inductively active and inactive constraints are defined over any set of generations, not just for DVE. However, when the design space expansions are caused by expansion of a set of dimensional variables, inductively active constraints contain powerful information about the limit solution. The extended 1st-PRINCE algorithm including the induction technique is presented in Figure 2.

From multiple DVE over a set of co-ordinates, x , a series of subregions along x is generated. As the number of expansions approaches infinity, so does the number of regions; since the body is of bounded size,¹ the size of the regions along x approaches zero. If a constraint is inductively active over the regions being divided, then in the limit, it is assumed active over an infinite number of infinitely small regions. In that limit, the set of dimensional variables along x approaches a continuum described by x . Thus in the inductively active constraints, the dimensional variables associated with x can be replaced by the co-ordinate variables x . If other variables in those constraints are a function of the dimensional variables associated with x , they become functions of x . Since the constraints are of the same form in each region, the substitution for x creates an infinite number of equivalent constraints in an infinite number of regions. Thus the substitution need be made only for any single region with the resulting set of equations

¹ Each iteration must have a feasible solution to continue induction in the algorithm, and thus the body has to be bounded and of finite size.

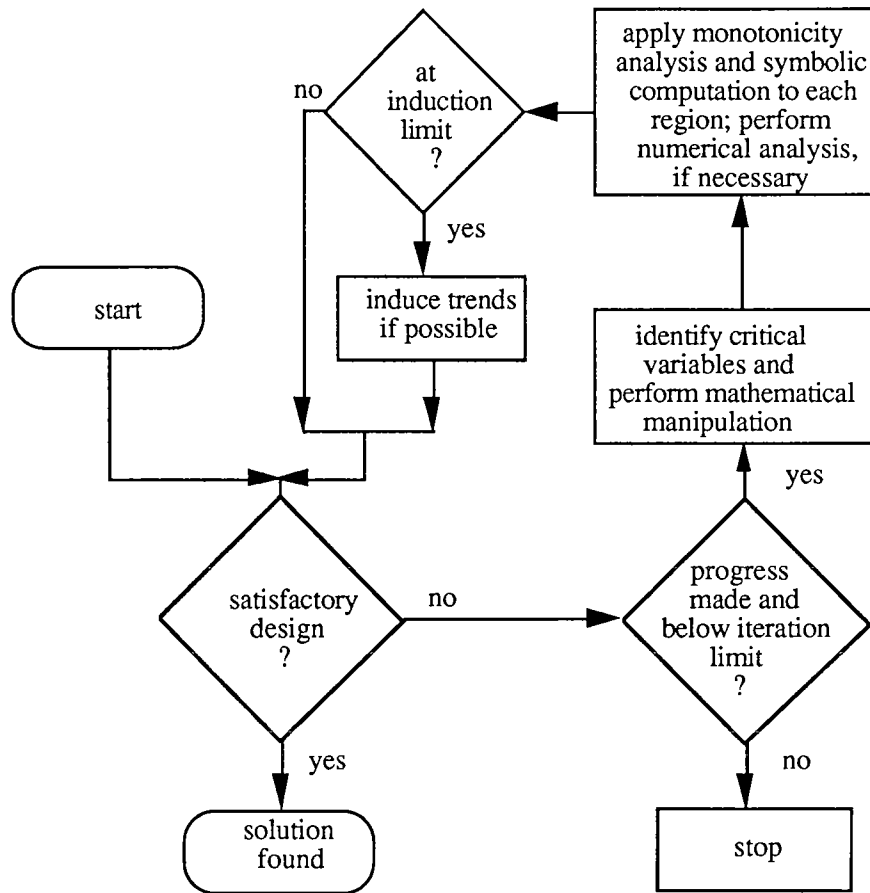


FIGURE 2. 1stPRINCE algorithm extended to include the induction of constraint activity

representative of the complete set over the entire body. The resulting set of constraints with the substitution made in the single region models a continuum solution over a newly formed single region, rather than a body of discrete regions, and the constraints are said to be *active across the continuum*.

As demonstrated in the example in the next section, the induction process described above is good for inducing trends, but even though 1stPRINCE can give closed form equations of the continuum solution, the solution may violate some constraint. This is a fundamental problem with heuristic-based inductive techniques in general. Some constraint may become violated in the limit of the trend induced which may not have been active for a smaller number of regions. Thus if one constraint is induced to be active across the continuum, then all constraints must also be checked across the continuum to guarantee that no constraint is violated by the inductive leap.

A constraint will be continuously active if the constraint is dominant in all generations of prototypes. A prototype which maintains the same activity as a prototype of a previous generation is

called a *generation-dominant prototype*; 1stPRINCE induces trends on sequences of generation-dominant prototypes. If different constraints are dominant then different paths may be generated and pursued by 1stPRINCE. If each constraint remains dominant within the path, then different prototypes will be induced. When 1stPRINCE induces a set of constraints to be continuously active or inactive over a sequence of generation-dominant prototypes, it forms a *dominant prototype* of the inductively active and inactive constraints. As the inductive procedure is heuristic, we can not guarantee that the limit induced into a dominant prototype is optimal nor feasible. The goal is to discover new prototype concepts that may not have been considered before.

5. Design example: flexural beam

5.1 THE FIRST ITERATION

In this section 1stPRINCE is applied to a solid cylindrical cross-section beam under a flexural load

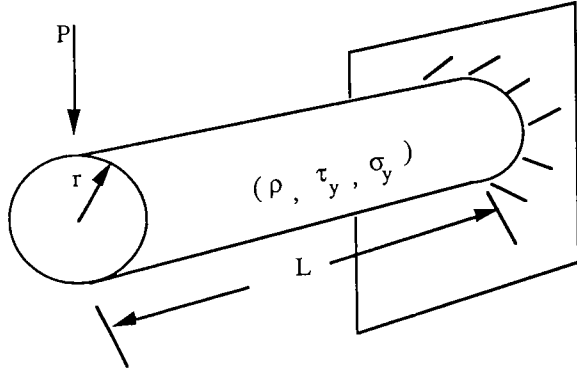


FIGURE 3. Cylindrical, solid beam which is clamped at one end and subject to a transverse load at the free end

(Figure 3). A beam of minimum weight is to be designed to withstand a transverse load (P) such that the maximum bending stress in the beam remains below the yield stress (σ_y) and the maximum shear stress remains below its yield stress (τ_y); both yield stresses may be divided by some factor of safety (ρ , σ_y , and τ_y are material properties). During a heuristic search, a design system may select the solid cylindrical beam (of radius r) as an initial primitive-prototype.

A primitive-prototype formed from elementary equations of the weight and stress of a circular beam under transverse load, based on elementary beam theory, is given as:

$$\begin{aligned} & \min: W \\ & \text{s.t.} \\ (a) \quad & W = \int_0^L \pi \rho_1 r^2 dx = \pi \rho_1 L r^2, \\ (b) \quad & \sigma = \frac{Mc}{I} = \frac{4Px}{\pi r^3}, \\ (c) \quad & \tau = \frac{P}{A} = \frac{P}{\pi r^2}, \\ (d) \quad & \sigma \leq \sigma_y, \\ (e) \quad & \tau \leq \tau_y, \\ (f) \quad & \sigma_y = K_1(\rho), \\ (g) \quad & \tau_y = K_2(\rho), \end{aligned} \quad (2)$$

where ρ is the mass density (related to other material properties, such as yield strength as demonstrated in constraints 2f and 2g¹), x is the distance along the beam of length L , P is the transverse load, M is the bending moment (equal to Px), c is the maximum

distance from the neutral axis (equal to r), I is the moment of inertia (equal to $\pi r^4/4$), and A is the area (equal to πr^2). Note that the weight which is to be minimized is in integral form, integrated over the critical co-ordinate variable x . To simplify the presentation, this example only considers integration over the length of the beam.²

There are two paths of dominant solutions. In the first, bending stress (constraints 2b and d) is dominant and shear stress (constraints 2c and e) is neglected. This path will be pursued in sections 5.1, 5.2, 5.3, and 5.4. Note that in the numerical examples, shear stress is verified to be inactive during the first three iterations. The second path requires that shear stress be dominant. This case is presented in section 5.5. Shear dominance is not an interesting case for this problem and is rare for most real design problems; however, it is a valid prototype which presents the state of shear in the beam and will be referred to in section 5.6. After new variables are introduced through Dimensional Variable Expansion, additional cases show mixed dominance; shear and bending stresses can each be active in different regions. This discussion will be presented in section 5.6.

The first iteration analysis for dominant bending stress [details of the analysis have been omitted but constraints (2b and d) are active] produces an optimally directed prototype with the stress at its yield limit at the maximum bending moment and with a radial dimension of

$$r = \left(\frac{4PL}{\pi \sigma_y} \right)^{1/3}. \quad (3a)$$

The beam weight is then found as

$$W = \rho \pi \left(\frac{4PL^{5/2}}{\pi \sigma_y} \right)^{2/3}. \quad (3b)$$

The weight and radius are now functions of yield stress, beam length, load, and mass density. Finding materials with low mass density and high yield stress will lead to better designs. Equations (3) can be derived via computer by using symbolic algebra routines to back-substitute active constraints into the objective function. Note that the prototype of equations (3) has 1 degree-of-freedom (DOF) (selection of a single material fixes density and yield stress).

If the constraints have indeed been met and the weight is satisfactory then the design is complete. If, however, a material cannot be found which meets the

¹The functions K_1 and K_2 designate some monotonic function of material density ρ .

²Integration over multiple axes can be performed sequentially on each axis. Cagan (1990) discusses flexure of a rectangular beam integrating over the 2-axis cross-section in which an I-beam is derived.

constraints, or if a designer wishes to improve on the weight, then modifications are desired. 1stPRINCE's second iteration innovates a new beam design.

5.4 THE SECOND ITERATION

On second iteration 1stPRINCE expands the primitive-prototype by DVE into two regions over length. As previously mentioned, weight equation (2a) is in integral form. 1stPRINCE divides this integral into two regions, the first from 0 to l_1 and the second from l_1 to L , and permits each region to have independent properties:

$$\int_0^L \pi \rho r^2 dx = \int_0^{l_1} \pi \rho_1 r_1^2 dx + \int_{l_1}^L \pi \rho_2 r_2^2 dx. \quad (4)$$

This division is represented in Figure 4. The objective function becomes the sum of the weights of the two regions, new variables are created, constraints are reformulated and new constraints are introduced, and infeasible geometries are eliminated. Note that 1stPRINCE creates new variables in the process: l_1 , ρ_1 , ρ_2 , r_1 , and r_2 that are not in the original solid beam primitive-prototype, replacing the original variables r and ρ .

The optimally directed innovative prototype has each region at its yield stress at the maximum bending moment and can be solved in symbolic form as:

$$W = \pi \rho_1 r_1^2 l_1 + \pi \rho_2 r_2^2 (L - l_1), \quad (5)$$

and

$$r_1 = \left(\frac{4Pl_1}{\pi\sigma_{y1}} \right)^{1/3} \quad \text{and} \quad r_2 = \left(\frac{4PL}{\pi\sigma_{y2}} \right)^{1/3}.$$

The prototype is of 3 DOF where the two materials and the intermediate length, l_1 , must be determined by further optimization techniques. The equation for l_1 can be found by setting the first partial derivative of the weight with respect to l_1 to zero ($\partial W / \partial l_1 = 0$), the first-order necessary conditions for optimality:

$$l_1 = \left(\frac{3}{5} \right)^{3/2} \left(\frac{\rho_2}{\rho_1} \right)^{3/2} \left(\frac{\sigma_{y1}}{\sigma_{y2}} \right) L. \quad (6)$$

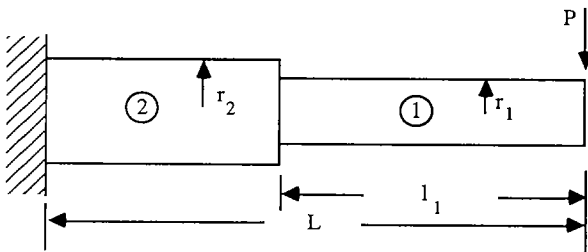


FIGURE 4. Expansion of beam by the second iteration of 1stPRINCE

As the problem is convex, the solution in equation (6) is a global optimum. If both materials are the same, the optimal geometry has

$$l_1 = \left(\frac{3}{5} \right)^{3/2} L, \quad r_2 = \left(\frac{4PL}{\pi\sigma_{y1}} \right)^{1/3}, \quad \text{and} \quad r_1 = \left(\frac{3}{5} \right)^{1/2} r_2. \quad (7)$$

Note that if the materials are not the same then r_2 could be greater than r_1 . For a constant cross-sectional beam, $r_1 = r_2$, giving

$$l_1 = \left(\frac{\sigma_{y1}}{\sigma_{y2}} \right) L, \quad \text{and} \quad \frac{\rho_2}{\rho_1} = \frac{5}{3}. \quad (8)$$

Once again, if a maximum weight constraint is not met or a better design is desired, then as long as progress is being made 1stPRINCE goes through a third iteration. Progress implies that a new design is superior in its design objective as compared to its predecessor designs.

5.3 THE THIRD ITERATION

In the third iteration, each region of the primitive-prototype is further expanded into two regions. The objective is now to minimize the total weight of the four regions as shown in Figure 5. From the expansion, 1stPRINCE has created new variables including: l_1 , l_2 , l_3 , r_1 , r_2 , r_3 , and r_4 . As the radius of each region closer to the clamped end is not specified to be greater than those nearer the free end (i.e., $r_i \geq r_{i-1}$, $i = 2, 3, 4$, not specified), the optimal design could actually appear as shown in the figure given different materials in each region.

1stPRINCE demonstrates that the prototype of dominant constraint activity again has each region at yield at its maximum bending moment. The radii and weight are given as:

$$r_i = \left(\frac{4Pl_i}{\pi\sigma_{y_i}} \right)^{1/3}, \quad i = 1, 2, 3, 4, \quad (9)$$

and

$$W = \sum_{i=1}^4 \pi \rho_i r_i^2 (l_i - l_{i-1}),$$

where $l_0 = 0$, and $l_4 = L$.

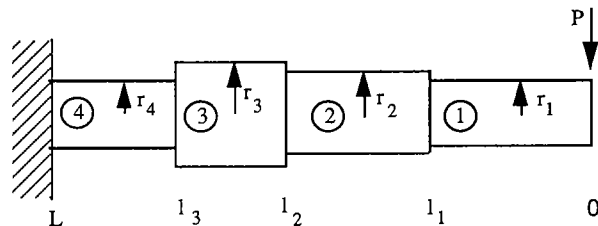


FIGURE 5. Expansion of beam by third iteration of 1stPRINCE

The prototype of equations (9) is 7 DOF, where the four materials and three intermediate lengths need to be determined to specify an interior solution. Those intermediate lengths can again be found by setting the partial derivatives of weight with respect to each length l_i , $i = 1, 2, 3$, to zero ($\partial W/\partial l_i = 0$):

$$\begin{aligned} \frac{5}{3}\pi\rho_1\left(\frac{4P}{\pi\sigma_{y_1}}\right)^{2/3}l_1^{2/3} - \pi\rho_2\left(\frac{4P}{\pi\sigma_{y_2}}\right)^{2/3}l_2^{2/3} &= 0, \\ \frac{5}{3}\pi\rho_2\left(\frac{4P}{\pi\sigma_{y_2}}\right)^{2/3}\left[l_2^{2/3} - \frac{2}{5}l_1l_2^{-1/3}\right] - \pi\rho_3\left(\frac{4P}{\pi\sigma_{y_3}}\right)^{2/3}l_3^{2/3} &= 0, \quad (10) \\ \frac{5}{3}\pi\rho_3\left(\frac{4P}{\pi\sigma_{y_3}}\right)^{2/3}\left[l_3^{2/3} - \frac{2}{5}l_2l_3^{-1/3}\right] - \pi\rho_4\left(\frac{4P}{\pi\sigma_{y_4}}\right)^{2/3}L^{2/3} &= 0. \end{aligned}$$

By letting

$$a_i = \pi\rho_i\left(\frac{4P}{\pi\sigma_{y_i}}\right)^{2/3},$$

equations (10) can be solved for the lengths as:

$$\begin{aligned} l_1 &= \left(\frac{3a_2}{5a_1}\right)^{3/2}l_2, \\ l_2 &= \left(\frac{3a_3}{5a_2}\right)^{3/2}\left[\frac{1}{1 - \frac{2}{5}\left(\frac{3a_2}{5a_1}\right)^{3/2}}\right]^{3/2}l_3, \\ l_3 &= \left(\frac{3a_4}{5a_3}\right)^{3/2}\left[\frac{1}{1 - \frac{2}{5}\left(\frac{3a_3}{5a_2}\right)^{3/2}\left[\frac{1}{1 - \frac{2}{5}\left(\frac{3a_2}{5a_1}\right)^{3/2}}\right]^{3/2}}\right]^{3/2}L. \end{aligned} \quad (11)$$

Equations (11) simplify for the same material used in each region. Note that the prototype DOF for any iteration, n , is $2^n - 1$. As the following numerical comparisons demonstrate, the design of each iteration is superior in weight to its predecessor.

5.4 NUMERICAL EXAMPLE

The numerical comparison of design artifacts instantiated from the three symbolic prototypes for a structural steel beam (ASTM-A36: $\rho = 7860 \text{ kg/m}^3$, $\sigma_y = 250 \text{ MPa}$) with a load of 10 kN and beam length of 1 m can be found in Table 1. Each iteration beyond the first has an innovated design derived from first principle knowledge which is superior in weight to its predecessor. Note that geometrically each improved design approaches a tapered beam which is known to be a weight-efficient beam for transverse loading. The lengths of the sub-regions are constant for all homogeneous material artifacts, as can be seen by equations (7) and by equations (11), letting $a_1 = a_2 = a_3 = a_4$. As mentioned previously, these designs are valid only for elementary beam theory. The authors have verified that the numerical solutions satisfy the theory; an additional constraint could have been included to require the ratio of radius to length be 1-to-20, generating alternative cases of active constraints. However, one limitation of our theory is that without such a constraint, the designs generated by 1stPRINCE may be infeasible due to the violation

TABLE 1. Innovated designs for solid steel beam of 1 m length with 10 kN transverse load

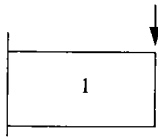
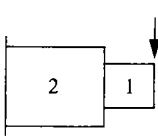
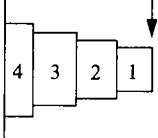
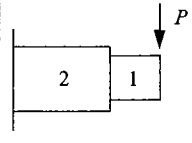
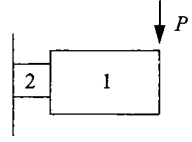
Iteration	Beam	Sub-lengths	Radii	Weight
1			$r_1 = 0.0371 \text{ m}$	33.92 kg
2		$l_1 = 0.465 \text{ m}$	$r_1 = 0.0287 \text{ m}$ $r_2 = 0.0371 \text{ m}$	27.59 kg
3		$l_1 = 0.212 \text{ m}$ $l_2 = 0.456 \text{ m}$ $l_3 = 0.720 \text{ m}$	$r_1 = 0.0220 \text{ m}$ $r_2 = 0.0285 \text{ m}$ $r_3 = 0.0333 \text{ m}$ $r_4 = 0.0371 \text{ m}$	24.13 kg

TABLE 2. Innovative designs for the second iteration of 1stPRINCE for combinations of steel and aluminum; the load is 10 kN and length is 1 m

Combination	Beam	Construction	Radii and sub-length	Weight
1		Steel in region 1 Aluminum in region 2	$r_1 = 0.0232$ m $r_2 = 0.0512$ m $l_1 = 0.248$ m	20.05 kg
2		Steel in region 2 Aluminum in region 1	$r_1 = 0.0489$ m $r_2 = 0.0371$ m $l_1 = 0.872$ m	22.08 kg

of unstated constraints; the method will generate new concepts but the user must further examine each one to guarantee their validity.

Since material is considered a variable in 1stPRINCE, each region can contain a different material. Table 2 compares the two-region beam of equations (5) for composite artifacts of the steel with aluminum (1100-H14: $\rho = 2710$ kg/m³, $\sigma_y = 95$ MPa). The prototype assumes that these beams can be manufactured with rigid bonds between regions; although not realistic with current manufacturing technology, the concept shows the potential benefits for designer materials and advanced manufacturing practices in composite design. First, for steel in region 1 and aluminum in region 2, the beam weighs 20.05 kg. If the steel is in region 2 and the aluminum in region 1, then the beam weighs 22.08 kg. The first combination is the lighter of the two, but both combinations are lighter than even the third iteration of the homogeneous steel beam. Not only is the second combination heavier than the first, it is also oddly shaped. However, although the second combination would likely not be employed in *most* design situations, it does not violate any constraints and is a feasible design which may have interesting applications. This example illustrates an important feature of 1stPRINCE; the methodology allows a designer to conceptualize different feasible designs which would likely have not been considered otherwise.

5.5 SHEAR DOMINANCE

As mentioned in section 5.1, the second solution path requires that shear stress (constraints (2c, e)) shows dominance. When shear stress is active, the

radius is at a constant value¹ of

$$r = \left(\frac{P}{\pi \tau_y} \right)^{1/2}. \quad (12)$$

DVE across the length leads to no improvement in the prototype and thus a beam of constant radius as defined by equation (12) is the optimally directed prototype for active constraints (2c, e) with 1 DOF. In reality, a shear beam as described in equation (12) would be short and stubby and not satisfy the assumptions of elementary beam theory; however, equation (12) does give the state of shear stress in the beam and is a valid prototype resulting from a monotonicity analysis.

5.6 INDUCING CONSTRAINT ACTIVITY

This discussion has followed two solutions cases. Figures 6(a–c) demonstrate the three iterations of bending stress dominant for a single material. For the purpose of induction, we will only consider the designs with a single material. In the implementation of 1stPRINCE, the user decides the number of generation-dominant prototypes necessary for observations toward inducing trends; in this example that number is chosen to be three. Thus, at this point, 1stPRINCE recognizes that constraints (2b and d) have been active for three iterations; the inductive mechanism then determines that these constraints should be *inductively active*. Similarly, the shear constraints (2c and e) were never active and can be assumed to be *inductively inactive*. Because the generations were created by DVE on dimensional variables associated with co-ordinate x , 1stPRINCE

¹ If τ_y were a function of x , then r would vary over x .

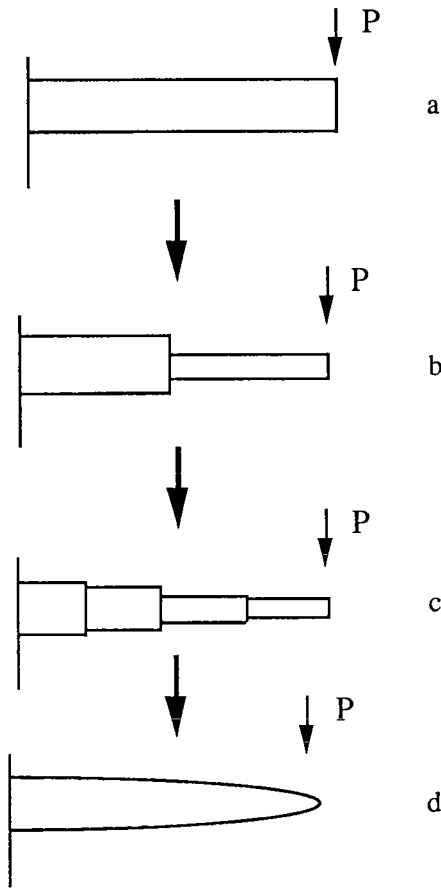


FIGURE 6. Induction over three consecutive generation-dominant prototypes on a clamped beam under flexural load for bending stress dominant

suggests that the bending stress constraints should be active across the entire continuum of x . Solving the constraint set for radius (r) as a function of continuous length (x) gives the equation:

$$r = \left(\frac{4Px}{\pi \sigma_y} \right)^{1/3} \quad (13)$$

Figure 6(d) shows that this dominant prototype, for a constant material, is a tapered beam which demonstrates the optimal trend given by constraint activity. Calculation of the weight of the tapered beam instantiated for the steel material in section 5.4 gives a value of 20.36 kg which is lighter than the stepped beams for the same load and beam length. Thus the induced design further improves the objective function.

In the second solution case shear stress [constraints (2c and e)] shows dominance. When shear stress is active, the radius is at a constant value as given in equation (12). Because DVE across the length leads to no improvement in the solution with constant

material, the constant-radius beam is the optimally directed solution for active constraints (2c and e). As discussed in section 5.5, this solution gives the state of shear in the beam. However, the actual state of stress in the beam is a combination of shear and bending.

A third solution case exists for multiple region, constant material beams where the bending stress is active for the regions closer to the clamped end and the shear stress is active for the regions closer to the point of load application. This is the most realistic case presented. However, the individual dominant prototypes presented do not recognize this case: in the bending dominant prototype it is not recognized that shear becomes dominant and the shear dominant prototype ignores the bending stress. In the bending dominant prototype, the beam tapers down to a point at load contact. The shear stress equality constraint (2c) indicates that as radius goes to zero, shear stress tends to infinity. The shear stress inequality constraint (2e) bounds the shear stress by its yield stress such that at

$$x = \frac{\pi \sigma_y}{4P} \left(\frac{P}{\pi \tau_y} \right)^{3/2}, \quad (14)$$

approaching the end of the beam, the shear stress constraints become active. As mentioned in section 4, this is a fundamental flaw in most heuristic-based induction techniques in that these constraints weren't active for reasonable numerical cases considered up to three iterations, and so were induced to be never necessary. However, as the radius gets small enough the constraints are dominant. Thus even though constraints are deemed to be *inductively inactive*, all constraints must be considered across the continuum of the induced trend to check for constraint violations. The important value of this inductive mechanism is to show optimal *trends* to the designer. However, by considering all constraints across the continuum, Figure 7 demonstrates that the optimally directed, feasible beam tapers [equation (13)] until the radius

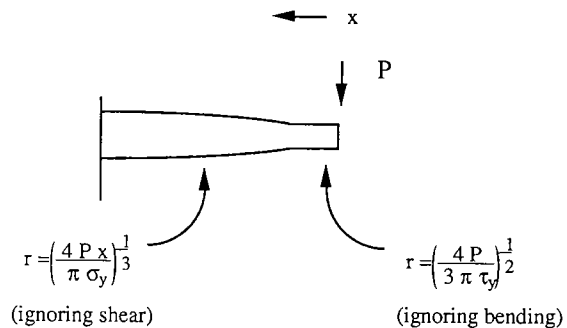


FIGURE 7. The continuum prototype for the clamped beam under flexural load with trade-off of dominance

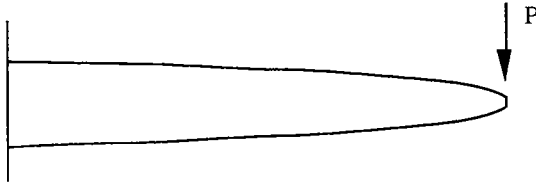


FIGURE 8. Combined loading solution to flexural beam problem

meets the shear stress [equation (12)] where the radius maintains that constant value.

These prototypes are valid only for elementary beam assumptions ignoring combined loading. Around the point of crossover of dominance between dominant prototypes, the stresses are at the same order of magnitude and thus combined stresses need to be considered. The actual solution around the point of cross-over of dominance is slightly different from that given in equations (12), (13) and (14), and a more detailed analysis is required. In addition, at the point of cross-over of dominance, Figure 7 shows a discontinuity in slope which would lead to stress concentrations; the preferred prototype has a continuous change of slope as would be dictated by consideration of combined loading. However, the trends associated with these solutions are valid and indicate the form of the optimally directed solution.

The actual combined loading situation can be found by a Mohr's circle analysis using the maximum-shear-stress criteria for failure; analysis shows the equality stress constraint inductively active as well as the inequality yield stress limit:

$$\frac{\sigma_y}{2} = \sqrt{\left(\frac{2Px}{\pi r^3}\right)^2 + \left(\frac{P}{\pi r^2}\right)^2}. \quad (15)$$

A sixth order polynomial in radius results:

$$\pi^2 \sigma_y^2 r^6 - 4P^2 r^2 - 16P^2 x^2 = 0. \quad (16)$$

This solution has no discontinuities in slope and appears in Figure 8.

6. Inventing the wheel

1stPRINCE has been applied to structural beam problems thus far. In this section the 1stPRINCE methodology is applied to a dynamics problem which we call "inventing the wheel". This example is only briefly discussed to demonstrate the concepts introduced in this paper and to motivate further research in inductive techniques for design; details of this section can be found in Cagan and Agogino (1989).

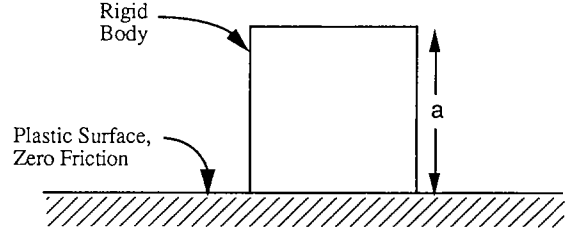


FIGURE 9. A rigid square body on a plastic surface of zero friction

Consider a square, rigid body (of side length a) in rectangular co-ordinates resting on a plastic surface with zero friction (Figure 9). The problem is to find the shape which will minimize the resistance to spinning, i.e., minimize the change in energy over the distance travelled. The initial potential energy of the system is set equal to that required to just provide enough rotational (kinetic) energy to flip the block in Figure 9 onto its side. Since the body's surface contact is not point-wise continuous (the primitive-prototype is square in cross-section), it will fall onto its side and energy will be lost. Figure 10 shows the potential energy of the block in its initial position [Figure 10(a)] and at its maximum potential on its corner at 45° [Figure 10(b)]. The minimum required energy to rotate the block is the difference between the energies of the two configurations. Since energy (E) is the weight (W) times the height (h), the energy (E^*) imparted to the body is given as

$$E^* = Wh = W\left(\frac{\sqrt{2}}{2} - \frac{1}{2}\right)a. \quad (17)$$

The weight is given as the material density (ρ) times the area of the body,

$$W = \int \rho dA. \quad (18)$$

Thus the total energy imparted to the system is $E^* = 0.2071\rho ta^3$, where t is the thickness. In addition, a minimum area constraint is dictated.

The objective is to minimize the resistance to spinning over the perimeter distance travelled given the input of energy, E^* . As the body in its present configuration flips onto its side (travelling a perimeter distance a) it will lose all of its E^* energy and stop. If the configuration was modified so that it lost less energy as it flipped onto its side, then it could continue to flip until all of the energy was lost. Thus in order to minimize the resistance to spinning, the loss in energy as the body flips around must be reduced.

At present, we assume symmetry of the body around the x - and y -axes through the centre so that

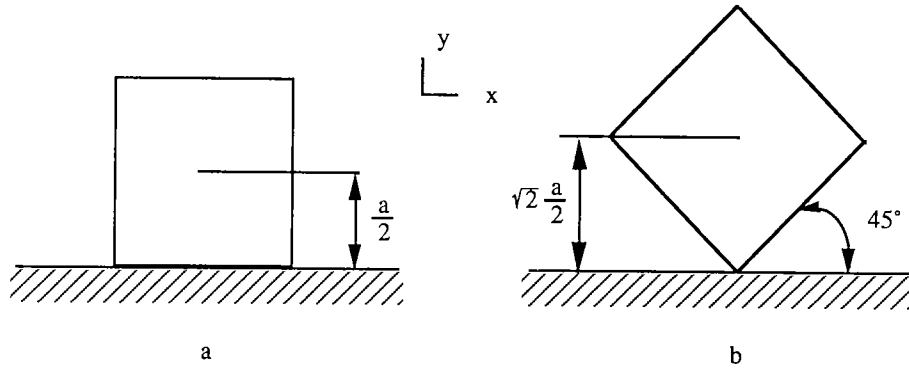


FIGURE 10. Potential energy at the minimum (10a) and maximum positions (10b)

the body can appropriate any stable point as its starting point, and symmetry at 45° so that energy can be imparted to the body in both positive (clockwise) and negative (counter-clockwise) directions. If symmetry is maintained, then any loss of total energy is due to impact with the plastic surface. Therefore we wish to minimize the total change in energy over perimeter distance travelled (p), where p is called the *envelope perimeter*:

$$\min: \int \frac{|dE|}{p} = \int W \frac{|dh|}{p} = \sum W \frac{|\Delta h|}{p}. \quad (19)$$

The minimum area constraint is defined as:

$$A \geq A_{\min}. \quad (20)$$

The change in height (Δh) is specified by the following constraints:

- (a) $\Delta h = \sum |d_i - h_i|$ (for each region i that is an edge region),
- (b) $d_i = [x_i^2 + y_i^2]^{1/2}$ (for each region i that is an edge region),
- (c) $h_i =$ perpendicular distance from surface to parallel line to surface through centre of mass of body of each edge surface (found from intersection of edge and perpendicular line through origin for each surface i that is an edge surface).

Calculation of h_i comes from intersecting a line through two edge vertices which form a surface edge and a line of negative, inverted slope passing through the origin; this gives the distance from the surface to the centre of mass height. Variable d_i is the distance from the centre of mass to the edge vertex. Variable lengths h_i and d_i are shown in Figure 11(b).

The perimeter distance travelled is defined by:

$$p = \sum [(x_{e1} - x_{e2})^2 + (y_{e1} - y_{e2})^2]^{1/2} \text{ (for edge vertex } e1 \text{ next to edge vertex } e2 \text{ along the perimeter envelope of the body, summed over all edge vertices).} \quad (22)$$

Also material density can either equal ρ or 0 (ignoring composite designs.)

Maintaining symmetry and following the solution path which leads to the minimum objective function, the body is divided in the x - and y -directions, material in the outer-most corners is removed, and then the body is optimized numerically, as sequentially demonstrated in Figure 11. In each iteration, the objective function decreases from value $0.0518\rho ta^2$ to $0.0064\rho ta^2$, as demonstrated in Table 3 which gives the dimensions for Figure 11 and the objective function for each iteration. In the limit we induce that the objective function approaches zero.¹

Observing the constraint activity on each iteration, minimum area inequality constraint (20) is active and height variation and perimeter distance equality constraints (21) and (22) are relevant. We now induce that these constraints should be *inductively active* and thus active across the continuum. These equations can now be solved by considering symmetry and observing limits. In the limit the objective function is observed to approach zero and thus Δh approaches zero, assuming finite envelope perimeter. Setting $\Delta h = 0$ in the inductively active constraint (21a) implies:

$$\sum |d_i - h_i| = 0, \quad \forall i. \quad (23)$$

Thus,

$$|d_i - h_i| = 0, \quad \forall i. \quad (24)$$

For equation (24) to be valid,

$$d_i = h_i, \quad \forall i. \quad (25)$$

Due to symmetry, we know that $d_i = h_i, \forall i \geq 1$ along the edge. Thus it can be inferred that this can only be valid if $d_i = h_i = h_{i-1} = \text{constant}$. Let us call that constant, h .

¹The automated induction method can only induce trends in constraint activity.

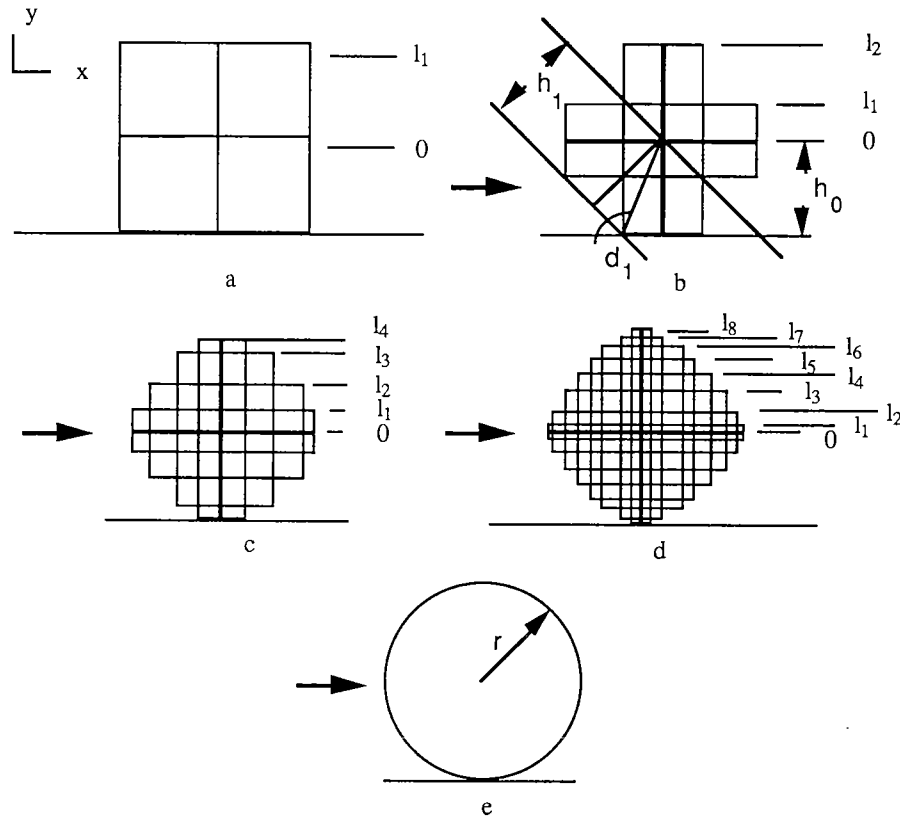


FIGURE 11. The body during its first four iterations and induction step by 1stPRINCE

Thus, substituting $d_i = h$ into inductively active constraint (21b) gives:

$$[x_i^2 + y_i^2]^{1/2} = h$$

(continuously across the body perimeter). (26)

Equation (26) designates the equation for a circle of radius h . Thus, in the limit, the optimal shape to minimize the change in energy over distance travelled is a circle; $A = 0.25a^2$, $h = 0.564a$. Therefore, the

optimal shape of the edge of the body is

$$[x_i^2 + y_i^2]^{1/2} = 0.564a$$

(continuously across the body perimeter). (27)

Figure 11(e) shows the optimally directed shape of the body, a circle of radius $0.564a$. Starting with a square block in rectangular coordinates, and aided by the 1st-PRINCE methodology, we have derived a circular body as the optimally directed shape, and thus have "invented the wheel."

TABLE 3. Values of dimensions and objective for each iteration shown in Figure 11

Iteration	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8	Objective
1	0.5a								0.0518 ρta^2
2	0.255a	0.617a							0.0249 ρta^2
3	0.120a	0.341a	0.505a	0.597a					0.0123 ρta^2
4	0.054a	0.173a	0.272a	0.375a	0.461a	0.529a	0.568a	0.592a	0.0064 ρta^2
Induction	$r = 0.564a$								0 ρta^2

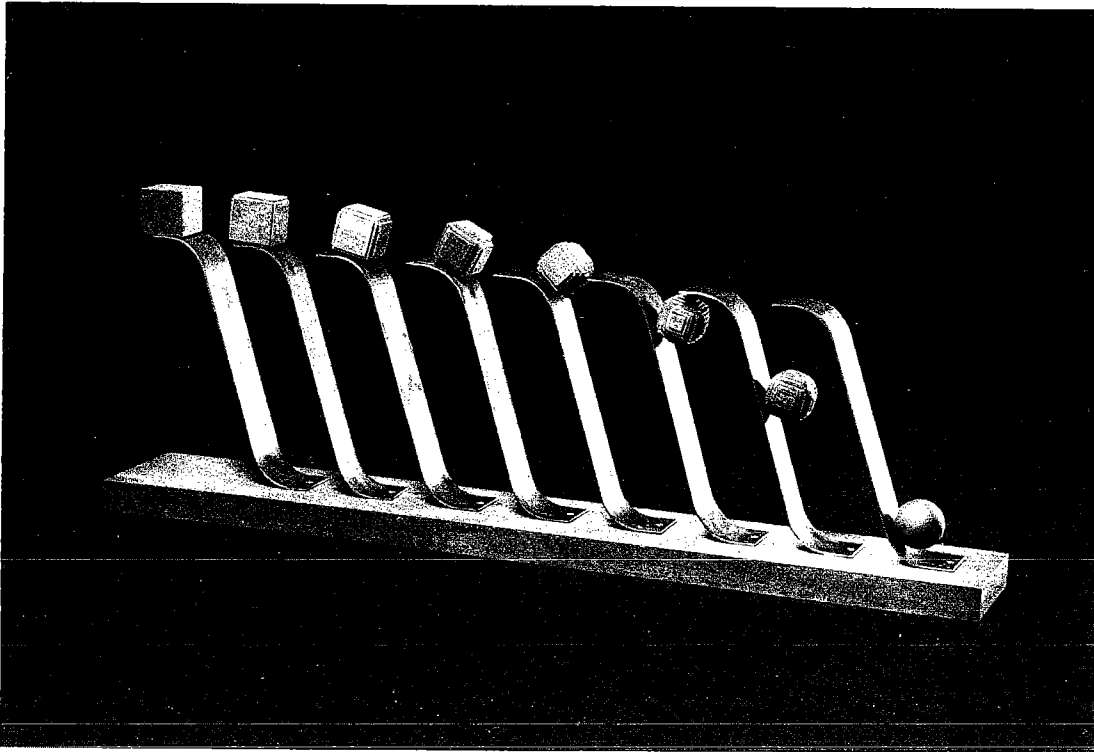


FIGURE 12. *Integrate to Decrease Rolling Resistance* by Steve Slominski, media: wood and steel, 1991

The problem formulation assumed a frictionless surface. With such a surface the wheel has no traction and thus will produce no forward motion but rather spin in place, since only rotational energy is imparted to the wheel. The problem could be formulated with the more realistic frictional surface, but the set of the resulting equations would be more complicated. Note that the wheel shape which results from the frictionless analysis can also be utilized on a frictional surface.

It is interesting to note that if we did not require symmetry about 45° other shapes could have potentially been invented, because there is no loss of energy due to friction. These shapes would need to be continuous convex surfaces, such as an ellipse, because energy would be lost due to impact with the surface caused by any non-convexity. As a final point of interest, the sculpture of Figure 12 was inspired by a lecture on 1stPRINCE and innovation in engineering design.

7. Implementation and discussion

Dimensional Variable Expansion methods in 1st-PRINCE are implemented in Allegro CommonLISP and Flavors on a MACII. The inductive program is

implemented in FranzLISP on Vax series computers running under Unix. Monotonicity analysis, back-substitution using symbolic algebra, and symbolic algebra algorithms for solving the Karush-Kuhn-Tucker conditions were implemented by Choy and Agogino (1986) and Agogino and Almgren (1987) in the SYMON/SYMFUNE systems written in VAXIMA, a symbolic algebra environment implemented in FranzLISP on Vax series computers. 1st-PRINCE builds on previous work embedded in the SYMON/SYMFUNE program and each part of the process is run independently by the user.

The flexural beam problem, as described in this paper, was derived entirely on the computer by 1st-PRINCE. Dominant prototypes were derived in closed form as limit solutions for cases with dominant constraints. However, as demonstrated in the beam problem, it is possible that one prototype is optimal over one part of a region and another prototype over a different region. This suggests using the union of the prototypes in proposing innovative variants of the original primitive-prototype. In the beam problem, two cases of dominance were derived leading to the tapered beam (active bending stress constraint) and the constant radius beam (active shear constraint) dominant prototypes. Figure 7 illustrates their union; the resulting prototype exhibiting the trend of the

tapered beam with a shear end. This combined prototype is more practical than either of the two individual prototypes as no cantilever beam under point load can taper to an endpoint without shear taken under consideration; however, shear stress is so small that away from the point of load application bending is the larger stress and shear is insignificant. This illustrates the need for a procedure for joining prototypes that is not arbitrary, but determined by crossovers in constraint dominance. Such procedures are not yet to be automated and are the subject of current research.

In the spinning block problem, much of the geometric derivations were done manually utilizing the 1stPRINCE design methodology. The complexity of the equations involved went beyond the state of the art in computer algebra algorithms and the number of possible prototype generations made a general implementation unwieldy. Also, the induction technique presented in this paper is based on constraint activity; trends not based on this activity, such as recognizing that the objective function approaches zero value, are currently not automated.

Another area deficient in the methodology and pursued as current research is the use of a language of features. In the wheel problem, we specified that only the edge regions would be considered for manipulation. However, there is currently no mechanism to specify such criteria in the design problem formulation. Such a language would aid in problem specification and in the automated conversion of a problem description to a computational description.

8. Conclusions

1stPRINCE is an interactive methodology which innovates new design prototypes using first principle reasoning. By determining which variables and constraints are pertinent for optimization and applying operators such as Dimensional Variable Expansion and induction processes, 1stPRINCE discovers optimally directed novel designs. New design variables are obtained by transformations of the design space and not from new, exterior knowledge.

Induction is an important technique which aids in concept discovery. This paper demonstrates that inductive techniques can be applied to optimally directed mechanical/structural design. Monotonicity analysis performs a qualitative optimization to determine candidate sets of active constraints. After a number of mutations by 1stPRINCE, inductive techniques observe trends from the constraint activity which may lead to new prototypes in the limit.

In the case of the flexural beam, the process induces a tapered beam from a solid rod, an innovative design under our definition of innovation. In the case of the spinning square, the 1stPRINCE method induces a circular wheel; this design has little resemblance to the original artifact and could be considered a creative design. Either way, these designs derived from 1stPRINCE are non-routine in that new variables are created. Inductive techniques play an important role in the derivation of these final artifact configurations; as the process derives better designs to satisfy the objective, the inductive algorithm in 1stPRINCE learns from the examples and recognizes what optimal trends the artifact is demonstrating. Research into inductive techniques and applications have a promising future in aiding non-routine design.

Although presented as a technique to perform innovative design, the methodology can also be utilized as an approach to shape optimization. Unlike traditional shape optimization techniques like the calculus of variations (Wylie and Barrett, 1982; Courant and Hilbert, 1937) or finite element applications (Vanderplaats, 1984; Azarm *et al.*, 1988; Yang and Botkin, 1987; Haftka and Grandhi, 1986), the form of the solution, the relations between variables, and the potential discontinuities in the variables do not need to be known *a priori*. Rather the system identifies critical variables for expansion and reasons about trends to suggest optimal shape concepts. In addition, the methodology takes advantage of non-numerical symbolic techniques for optimal design that can lead to complete or partial solutions presented in closed form.

Acknowledgements

This research is supported by The National Science Foundation under grant DMC-8451622 and an equipment grant from the Digital Equipment Corporation. The authors would like to thank Stuart Russell for his suggestion to "reinvent the wheel" and Steve Slominski for creating the sculpture in Figure 12.

References

- Agogino, A. M. and Almgren, A. S. 1987. Techniques for integrating qualitative reasoning and symbolic computation in engineering optimization. *Engineering Optimization* **12**(2), 117-135.
- Almgren, A. and Agogino, A. M. 1989. A generalization and correction of the welded beam optimal design problem using symbolic computation. *ASME Journal of Mechanisms, Transmissions, and Automation in Design* **111**(1), 137-140.

- Azarm, S., Bhandarkar, S. M. and Durelli, A. J. 1988. On the experimental vs. numerical shape optimization of a hole in a tall beam. In *Proceedings of ASME Design Automation Conference, Kissimmee, FL, September 25-28*, pp. 257-264.
- Cagan, J. 1990. Innovative Design of Mechanical Structures from First Principles. Ph.D. Dissertation, University of California, Berkeley, CA, April.
- Cagan, J., and Agogino, A. M. 1987. Innovative design of mechanical structures from first principles. (*AI EDAM*) **1**, 169-189.
- Cagan, J. and Agogino, A. M. 1989. Inducing optimally directed non-routine designs. *Preprints of International Round-Table Conference: Modeling Creativity and Knowledge-Based Creative Design, Heron Island, Queensland, Australia, December 11-14*, pp. 95-117. A portion to be reprinted in *Modeling Creativity and Knowledge-Based Creative Design*, Gero, J. S. and Maher, M. L., (Eds), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Choy, J. K. and Agogino, A. M. 1986. SYMON: Automated SYmbolic MONotonicity Analysis System for qualitative design optimization. In *Proceedings of ASME 1986 International Computers in Engineering Conference, Chicago, July 24-26*, pp. 305-310.
- Courant, R. and Hilbert, D. 1937. *Methods of Mathematical Physics*, Vol. 1. New York, Interscience.
- Haftka, R. T. and Grandhi, R. V. 1986. Structural shape optimization—a survey. *Computer Methods in Applied Mechanics and Engineering*, **57**, 91-106.
- Howard, H. C., Wang, J., Daube, F. and Rafiq, T. 1989. Applying design-dependent knowledge in structural engineering design. (*AI EDAM*) **3**, 111-123.
- Jain, P. and Agogino, A. M. 1990. Theory of design: an optimization perspective. *Journal of Mechanism and Machine Theory* **25**(3), 287-303.
- Michalski, R. S. 1983. A theory and methodology of inductive learning. In *Machine Learning: An Artificial Intelligence Approach* (Michalski, R. S., Carbonell, J. G. and Mitchell, T. M., eds), pp. 83-134. Los Altos, CA: Morgan Kaufman.
- Papalambros, P. 1982. Monotonicity in goal and geometric programming. *Journal of Mechanical Design* **104**, 108-113.
- Papalambros, P. and Wilde, D. J. 1979. Global non-iterative design optimization using monotonicity analysis. *Journal of Mechanical Design* **101**, 645-649.
- Vanderplaats, G. N. 1984. *Numerical Optimization Techniques for Engineering Design: With Applications*. New York: McGraw-Hill.
- Wilde, D. J. 1986. A maximal activity principle for eliminating overconstrained optimization cases. *Transactions of the ASME, Journal of Mechanisms, Transmissions, and Automation in Design* **108**, 312-314.
- Wylie, C. R. and Barrett, L. C. 1982. *Advanced Engineering Mathematics*. New York: McGraw-Hill.
- Yang, R. J. and Botkin, M. E. 1987. A modular approach for three-dimensional shape optimization of structures. *AIAA Journal* **25**(3), 492-497.

Jonathan Cagan is an Assistant Professor of Mechanical Engineering at Carnegie Mellon University. Dr Cagan's research interests focus around conceptual design methodologies, emphasizing optimization techniques, artificial intelligence techniques, computational representations, manufacturing processes, and techniques to generate creative design configurations. Professor Cagan, a member of Tau Beta Pi, Phi Beta Kappa and Sigma Xi honor societies, received his B.S. and M.S. degrees from the University of Rochester in 1983 and 1985, and his PhD from the University of California at Berkeley in 1990. He has spent significant time in industry and is a licensed professional engineer in the states of Pennsylvania and California.

Alice M. Agogino is an Associate Professor of Mechanical Engineering at the University of California at Berkeley where she heads the Berkeley Expert Systems Technology (BEST) and the Concurrent, Collaborative, Computer-Aided Design (C³AD) Laboratories. Dr Agogino received a PhD from Stanford University in 1984. She has six years of industrial experience and is a registered Professional Mechanical Engineer in California. Dr Agogino received an NSF Presidential Young Investigator Award in 1985; IBM Faculty Development Award, 1985/86; Pi Tau Sigma Award for Excellence in Teaching, 1986; Ralph R. Teeter Educator Award, 1987; and SME Young Manufacturing Engineer of the Year Award, 1987/88. Her research interests include concurrent engineering, intelligent CAD, non-linear and multiobjective optimization, probabilistic modeling, intelligent control and manufacturing, and decision and expert systems.

