

# Technical Briefs

## Notes on Monotonicity Principles

Yeh-Liang Hsu<sup>1</sup> and D. J. Wilde<sup>2</sup>

### Introduction

The branch of optimization theory known as “Monotonicity Analysis” (Wilde, 1975; Papalambros, 1979; Papalambros and Wilde, 1988) seeks to identify rigorously, in advance of any extensive numerical computation, which combinations of inequality constraints can be active, that is, satisfied with strict equality at the optimum. There are two Monotonicity Principles: the First Monotonicity Principle (MP1) deals with variables which are in the objective function, and the Second Monotonicity Principle (MP2) deals with variables which are not in the objective function. Although the Monotonicity Principles developed by Wilde and Papalambros deal only with strictly monotonic constraints, Hansen, Jaumard, and Lu have extended the principles to nonmonotonic ones (1989). This is an important advance since the extended principles allow us to derive conclusions and simplify problems more often than the previous version does.

The Monotonicity Principles are necessary conditions, applying only “if the problem is well constrained”. But before applying the Monotonicity Principles to an optimization problem, one rarely knows whether the problem is well constrained or not. Actually one of the important purposes of Monotonicity Analysis is to identify optimization models that are not well constrained. However, in the authors’ experience, when Extended Monotonicity Analysis is applied to not-well constrained problems, very often it reaches erroneous results instead of detecting that the optimization model is not-well constrained. This technical brief attempts to clarify some potential problems with applying Extended Monotonicity Analysis to not-well constrained problems, especially pointing out situations when the users should be alert that the conclusions from Extended Monotonicity Analysis may need to be checked by the Karush-Kuhn-Tucker Conditions.

For nonobjective variables, MP2 states that “Every monotonic nonobjective variable in a well-bounded problem is either irrelevant and can be deleted from the problem together with all constraints in which it occurs, or relevant and bounded by two active constraints, one from above and one from below” (P&W, 1988). This technical brief shows that the Second

Monotonicity Principle should be used only to find if a nonobjective variable is relevant to the problem, rather than to deduce activities of two constraints in one step. Moreover, if a nonobjective variable appears in only one non-monotonic constraint, it can still be relevant to the problem, as was indicated by the Extended Second Monotonicity Principle (Hansen, Jaumard, and Lu, 1989).

### The Optimality Conditions

We focus here on design optimization problems with only inequality constraints, expressed in the following form:

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ & && \mathbf{x} \in \mathbf{X} \subseteq \mathfrak{R}^n \end{aligned} \quad (1)$$

where  $\mathbf{X}$  is the set constraint. In engineering optimization problems,  $\mathbf{X}$  is usually the positive finite domain, that is,

$$\mathbf{X} = \varnothing = \{(x_i) \in \mathfrak{R}^n : 0 < x_i < \infty\} \quad (2)$$

Note that the set constraint cannot be satisfied with strict equality. Here the range constraints  $l_i \leq x_i \leq u_i$  are viewed as two sets of simple inequality constraints.

At the minimizer  $\mathbf{x}_*$ , the KKT conditions must be satisfied. The KKT conditions of Eq. (1) can be expressed as follows (P&W, 1988):

$$\begin{aligned} 1. & \mathbf{g}(\mathbf{x}_*) \leq \mathbf{0}, \\ 2. & \nabla f_* + \mu^T \nabla \mathbf{g}_* = \mathbf{0}^T \end{aligned} \quad (3)$$

where  $\mu \geq \mathbf{0}$ ,  $\mu^T \mathbf{g} = 0$ . The  $\mu$  are Lagrange multipliers.

Here the minimizer  $\mathbf{x}_*$  is assumed to be a regular point;  $f$  and  $\mathbf{g}$  are assumed differentiable. The complementary slackness condition  $\mu^T \mathbf{g} = 0$  implies that if  $\mu_k > 0$ , the corresponding inequality constraint  $g_k$  is active (tight), that is,  $g_k = 0$ .

In Eq. (1), the constraints define a constraint set  $\mathbf{K}$ . A feasible set is defined as  $\mathbf{F} = \mathbf{K} \cap \varnothing$ . Let  $f(\mathbf{x})$  be an objective function defined on  $\mathbf{F}$ . If a constrained minimum exists for  $\mathbf{x}_* \in \mathbf{F}$ ,  $f$  is said to be “well constrained” (P&W, 1988).

### First Monotonicity Principle

The First Monotonicity Principle deals with objective function variables that are monotonic.

**First Monotonicity Principle (MP1) (P&W, 1988).** In a well constrained objective function, every (strictly) increasing (decreasing) variable is bounded below (above) by at least one active constraint.

Assuming differentiability of the functions, we can interpret MP1 using the KKT conditions. For a well constrained minimization problem, there exists at least one  $\mathbf{x}$  that satisfies the

<sup>1</sup> Associate Professor, Department of Mechanical Engineering, Yuan-Ze Institute of Technology, 135 Yuan-Tung Road, Taoyuan, Taiwan 320, ROC

<sup>2</sup> Professor Emeritus, Design Division, Department of Mechanical Engineering, Stanford University, Stanford, CA 94305

Contributed by the Design Automation Committee for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received May 1996; revised Dec. 1996. Associate Technical Editor: A. R. Parkinson.

optimality conditions Eq. (3). For a strictly increasing variable  $x_i$ , the second condition in Eq. (3) can be written as

$$\left(\frac{\partial f}{\partial x_i}\right)_* + \sum_j \mu_j \left(\frac{\partial g_j}{\partial x_i}\right)_* = 0, \quad (4)$$

where  $\mu_j \geq 0$ . But since  $f$  is strictly increasing in  $x_i$ ,  $(\partial f/\partial x_i)_* > 0$ . So if Eq. (4) is to be satisfied, there exists at least one constraint  $g_k$  where  $(\partial g_k/\partial x_i)_* < 0$  and  $\mu_k > 0$ . This also applies to the Extended First Monotonicity Principle proposed by Hansen, Jaumard and Lu (HJL, 1989).

**Extended First Monotonicity Principle (EMP1) (HJL, 1989).** Let  $f(x_1, x_2, \dots, x_n)$  be monotonous in  $x_i$  over  $\mathbf{X}$ , and let all  $g_k(x_1, x_2, \dots, x_n)$  involving  $x_i$  be continuous over  $\mathbf{X}$ . Then either  $\mathbf{X} = \emptyset$  or there exists a globally optimal solution such that at least one (global or range) constraint with different monotonicity in  $x_i$  from that of  $f$  is tight.

In the extended principle, when  $f$  is strictly increasing in  $x_i$ , the constraint with different monotonicity in  $x_i$  can be not only strictly decreasing, but also nonmonotonic in  $x_i$ . The local condition  $(\partial g_j/\partial x_i)_* < 0$  includes both cases.

Note that differentiability is not assumed by any of the Monotonicity Principles, and the proofs of the Monotonicity Principles do not involve the KKT conditions at all. Monotonicity Principles should not be seen as another form of the KKT conditions. The reason for interpreting the Monotonicity Principles using the KKT conditions is to clarify the potential problems with Extended Monotonicity Analysis, as illustrated below.

A hypothesis not stated in EMP1 is that the problem be well constrained. A simple one-variable example below illustrates how EMP1 may lead to a maximum if the problem is not well constrained.

**Example 1: Convex Constraint With One Positive Root**

$$\begin{aligned} &\text{minimize } x \\ &\text{subject to } g(x) = x^2 - x - 2 \leq 0, \\ &x \in \emptyset \subset \mathbb{R}^1 \end{aligned}$$

Since  $x$  is strictly increasing in the objective function, EMP1 concludes immediately that the nonmonotonic constraint  $g(x)$  is tight. The constraint function has two roots,  $x = -1$  and  $x = 2$ . Since the negative root lies outside the positive finite domain, one might be tempted to assume that the minimum is at  $x = 2$ . But it is easy to verify that this point is the maximizer of the problem instead of the minimizer!

The reason is that though  $(\partial g/\partial x)$  can be negative since  $g(x)$  is nonmonotonic, at  $x = 2$  where  $g(x)$  is tight,  $(\partial g/\partial x)_{x=2} > 0$ . Example 1 is not well constrained since  $x$  can go all the way to 0, violating the set constraint  $x \in \emptyset \subset \mathbb{R}^1$ . The hypothesis of EMP1 is not satisfied. As indicated in the interpretation using the KKT conditions above, when  $g_j$  is nonmonotonic, it is tight only if  $(\partial g_j/\partial x_i)$  has the correct sign at the solution point. Similar situation occurs when the constraint functions are not differentiable. To avoid errors of this sort, pay attention to the local monotonicity at any root considered a candidate for the minimum.

The constraint function in Example 1 is convex, since its second derivative is the strictly positive number 2. An interesting situation occurs when the constraint is concave, in this continuous case leading to a negative second derivative. Consider the following example:

**Example 2: Concave Constraint With One Positive and One Negative Root.**

$$\begin{aligned} &\text{minimize } x \\ &\text{subject to } g(x) = -x^2 + x + 2 \leq 0, \\ &x \in \emptyset \subset \mathbb{R}^1 \end{aligned}$$

This constraint is the negative of that in Example 1, making it concave rather than convex. The constraint roots are still  $-1$ , and  $2$ , but this time it is the larger one  $2$  that has the correct local monotonicity to bound the variable  $x$  from below. Figure 1 shows it is the minimizing solution. But to prove that  $x = 2$  is the globally optimal solution as stated in EMP1, one really must verify that the smaller root  $-1$  is not in the positive finite domain  $\emptyset$ , for if it were, the problem would not have a globally optimal solution. Example 3 illustrates this latter situation.

**Example 3: Concave Constraint With Two Positive Roots**

$$\begin{aligned} &\text{minimize } x \\ &\text{subject to } g(x) = -x^2 + 5x - 4 \leq 0, \\ &x \in \emptyset \subset \mathbb{R}^1 \end{aligned}$$

This constraint has two roots  $1$  and  $4$ , and only the larger root  $4$  has the correct local monotonicity to bound the variable  $x$  from below. Though at this solution point the KKT condition Eq. (4) can be satisfied,  $x = 4$  is only a local minimum. As shown in Figure 2, the feasible domain splits into two disjoint regions. The smaller root  $x = 1$  is in  $\emptyset$ , which generates an unbounded region within  $\emptyset$ . No globally optimal positive solution exists.

**The Second Monotonicity Principle**

The Second Monotonicity Principle deals with variables not in the objective function, which are called nonobjective variables.

**Second Monotonicity Principle (MP2) (P&W, 1988).** Every monotonic nonobjective variable in a well-bounded problem is either

- (a) irrelevant and can be deleted from the problem together with all constraints in which it occurs, or
- (b) relevant and bounded by two ‘‘active’’ constraints, one from above and one from below.

The word ‘‘active’’ is in fact an overstatement, so it is here placed in quotation marks. Assuming differentiability of the functions, we can interpret MP2 using the Karush-Kuhn-Tucker (KKT) conditions. In a well-bounded problem, there exists at least one  $\mathbf{x}$  that satisfies the optimality conditions. For a nonob-

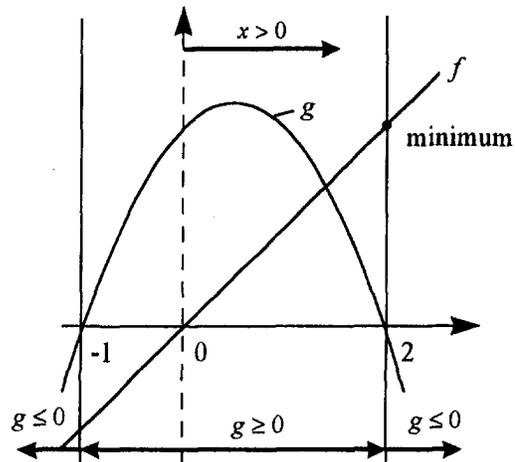


Fig. 1 Concave constraint, the smaller root  $-1$  is not in the positive finite domain

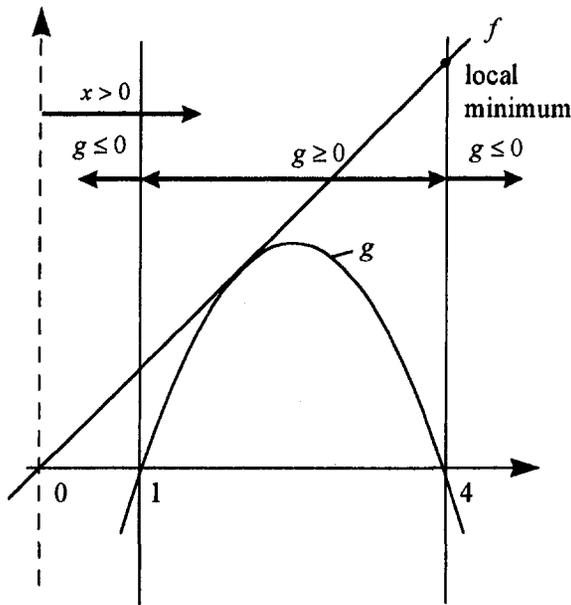


Fig. 2 Concave constraint, disjoint feasible regions

jective monotonic variable  $x_i$ ,  $(\partial f / \partial x_i)_* = 0$ , the KKT conditions become

$$\sum_j \mu_j \left( \frac{\partial g_j}{\partial x_i} \right)_* = 0, \quad (3)$$

since  $(\partial f / \partial x_i)_* = 0$ . But the Lagrange multipliers  $\mu_j \geq 0$ , so there are two situations in which Eq. (3) can be satisfied:

- $\mu_j = 0$  for all  $(\partial g_j / \partial x_i)_* \neq 0$ . That is, all constraints containing  $x_i$  are inactive,  $x_i$  is irrelevant to the problem and can be deleted.
- If  $x_i$  is strictly increasing in  $g_k$ , and  $\mu_k > 0$ , then  $\mu_k (\partial g_k / \partial x_i)_* > 0$ . To satisfy Eq. (3), there must exist at least one constraint  $g_l$  in which  $x_i$  is strictly decreasing and  $\mu_l > 0$ , so that  $\mu_l (\partial g_l / \partial x_i)_* < 0$ .

However, the basic hypothesis of MP2 is that the problem is well constrained. The following example shows that MP2 can lead to a maximum solution if incorrectly applied to a problem that is not well constrained.

#### Example 4

$$\begin{aligned} \text{minimize } & x_1 \\ \text{subject to } & g_1 = -x_1 + 2x_2 - 4 \leq 0, \\ & g_2 = x_1 - x_2 - 6 \leq 0. \end{aligned}$$

In the objective function, variable  $x_1$  is monotonically increasing, and needs to be bounded from below. By the First Monotonicity Principle,  $g_1$  is active for  $x_1$ , so  $x_2$  is relevant to the problem. By MP2, both  $g_1$  and  $g_2$  are alleged to be "active" for  $x_2$ . Solving  $g_1 = 0$  and  $g_2 = 0$ , we get  $(x_1, x_2) = (16, 10)$ . It is easy to verify that this point is a maximizer instead of a minimizer. Example 4 is not well constrained since  $x_1$  can go all the way to zero, violating the set constraint, while still satisfying the model constraints.

This case can be easily explained using the KKT conditions. For this problem, the optimality conditions for variables  $x_1$  and  $x_2$  can be written as  $1 - \mu_1 + \mu_2 = 0$ ,  $2\mu_1 - \mu_2 = 0$ . Solving this set of equations, we get  $\mu_1 = -1$ ,  $\mu_2 = -2$ , which contradict the sign requirement for the multipliers in the KKT conditions. When both  $g_1$  and  $g_2$  have the correct monotonicities required by the Monotonicity Principles, Eq. (3) can also be satisfied

when both multipliers are negative. That is, at the solution point of the "active" constraints, there is a possibility that the corresponding multipliers are negative.

Instead of checking the signs of the multipliers at the solution point, we can prevent this case from happening. One way to avoid it is not to use MP2 to determine activities of two constraints in one step. Any correct conclusion found by MP2 can be obtained by applying MP1 twice. Unboundedness, as in Example 4, can also be detected by MP1.

In Example 4, after  $g_1$  is proved active by MP1, we should use it to eliminate  $x_1$  and get the following subproblem:

$$\begin{aligned} \text{minimize } & 2x_2 - 4 \\ \text{subject to } & g_2 = x_2 - 10 \leq 0. \end{aligned}$$

Note that the monotonicity of  $x_2$  in  $g_2$  changes from decreasing to increasing. Applying MP1 again, we will detect the fact that the problem is not well constrained. Therefore, it is suggested that MP2 be used only to find if a nonobjective variable is relevant to the problem, not to deduce constraints activities.

In deducing relevance of variables, there is also a case which has not been explicitly addressed in the literature. Consider first the Extended Second Monotonicity Principle proposed by Hansen et al. (1989):

**Extended Second Monotonicity Principle (EMP2).** Let  $f(x_1, x_2, \dots, x_n)$  be independent of  $x_k$  and all  $g_i(x_1, x_2, \dots, x_n)$  involving  $x_k$  be continuous over  $\mathbf{X}$ . Then either  $\mathbf{X} = \emptyset$  or there exists a globally optimal solution  $\mathbf{x}_*$  such that at least one constraint not monotonically increasing in  $x_k$  is tight and a globally optimal solution  $\bar{\mathbf{x}}$  (with possibly  $\mathbf{x}_* = \bar{\mathbf{x}}$ ) such that at least one constraint not monotonically decreasing in  $x_k$  is tight.

EMP2 implies an important fact: if a nonobjective variable appears in only one nonmonotonic constraint, it can still be relevant to the problem, and the optimization problem is still well constrained. Example 5 illustrates this case.

#### Example 5

$$\begin{aligned} \text{minimize } & x_1 \\ \text{subject to } & -x_1 + (x_2 - 1)^2 + 1 \leq 0 \end{aligned}$$

The minimizer of this problem is (1, 1), and the constraint is active. Though the nonobjective variable  $x_2$  appears in only one constraint, it is still relevant to the problem. At the minimizer,  $\mu = 1$ ,  $(\partial g / \partial x_2)_* = 0$ , and the nonobjective variable  $x_2$  is a stationary point in the active constraint, so Eq. (3) is satisfied.

MP2 implies that in a well constrained problem, if a nonobjective monotonic variable does not appear in two constraints with opposite monotonicities, it has to be irrelevant to the problem. MP2 excludes nonmonotonic constraints like that in Example 5. Example 5 shows that if the nonobjective variable appears in only one nonmonotonic constraint, it can satisfy both conditions of EMP2 and still be relevant to the problem.

#### Conclusions

The Extended Monotonicity Analysis applies to non-monotonic functions, which greatly expands the number of problems amenable to Monotonicity Analysis. However, there is a complication introduced which can cause error if not taken into account. This springs from the fact that whereas monotonic functions can have no more than one root, non-monotonic functions can have several.

The decision of which root to use can be a difficult one. The examples in this technical brief suggest that if a non-monotonic constraint is concluded to be tight by EMP1, pay special attention to the local monotonicity at any root considered a candidate for the minimum. If none of the roots in  $\emptyset$  has the correct local monotonicity, i.e., satisfies the KKT conditions, this non-

monotonic constraint is not active and the optimization model may still not be well constrained.

It is also suggested that MP2 be used only to find if a nonobjective variable is relevant to the problem, rather than to deduce activities of two constraints in one step. If a nonobjective variable appears in only one non-monotonic constraint, it can still be relevant to the problem.

## References

- Hansen, P., Jaumard, B., and Lu, S. H., "Some Further Results on Monotonicity in Globally Optimal Design," *ASME Journal of Mechanisms, Transmissions, and Automation in Design*, Vol. 111, No. 3, 1989, pp. 353–361.
- Papalambros, P., and Wilde, D. J., *Principles of Optimal Design—Modeling and Computation*, Cambridge University Press, New York, 1988.
- Papalambros, P., "Monotonicity Analysis in Engineering Design Optimization," Ph.D. Thesis, Department of Mechanical Engineering, Stanford University, California, 1979.
- Wilde, D. J., "The Monotonicity and Dominance in Optimal Hydraulic Cylinder Design," *ASME Journal of Engineering for Industry*, Vol. 97, 1975, pp. 1390–1394.

# Operational Envelope of a Spatial Stewart Platform

F. A. Adkins<sup>1</sup> and E. J. Haug<sup>2</sup>

*This technical brief presents the operational envelope for the spatial Stewart platform and dome of a six degree of freedom driving simulator, extending prior work that has been limited to planar mechanisms and manipulators. The set of all points in space that can be occupied by any point in the dome of the simulator is defined as its operational envelope. The geometry of the driving simulator's dome and unilateral constraints on actuator lengths are incorporated and details of the defining equations for the operational envelope are given. Numerical methods are used to calculate the boundary of the operational envelope and results are presented in graphical form.*

## 1 Introduction

Workspace analysis for mechanisms and manipulators has conventionally focused on determining the set of all points in space that can be reached by a specific point on a working body, called the accessible output set [2, 7, 8, 11]. While knowledge of the accessible output set is important in defining the working capability of a working body, an extension of the formulation is required to determine the set of all points in space that can be occupied by any point on the working body to avoid interference with its surroundings. Such an extension to the operational envelope has been presented by Haug, Luh, and Adkins [4]. Analytical criteria and numerical methods for mapping the boundary of the accessible output set are also given by Haug, Luh, Adkins and Wang [5]. These methods apply directly for numerical calculation of the boundary of the operational envelope of a six degree of freedom Stewart platform with unilateral actuator constraints and the large projection dome of a driving simulator.

<sup>1</sup> Department of Applied Mathematical and Computational Sciences, University of Iowa, Iowa City, IA 52242.

<sup>2</sup> Department of Mechanical Engineering, University of Iowa, Iowa City, IA 52242.

Contributed by the Mechanisms Committee for publication in the *JOURNAL OF MECHANICAL DESIGN*. Manuscript received June 1996; revised Nov. 1996. Associate Technical Editor: B. Ravani.

## 2 Analytical Formulation

The Stewart platform [10] shown in Fig. 1 consists of a moving upper platform attached to ground by six extendable actuators. The lengths of the six actuators serve as inputs, with upper and lower bounds imposed on the lengths of each, to control the motion of the six-degree-of-freedom platform. The Iowa Driving Simulator (IDS) shown in Fig. 2 is a high fidelity ground vehicle simulator that incorporates such a Stewart platform as the motion base for six-degree-of-freedom motion control. Details of the kinematic formulation are given in Ref. 6 and a complete analysis of the accessible output set calculation is given in Ref. 9.

Attached to the motion base is a projection dome for wide-field visual display. This projection dome is modelled as a body of revolution about the vertical axis of the cross section shape presented in Fig. 3.

A brief discussion of the kinematics formulation and the associated accessible output set is given here. Limitations on the lengths of the six actuators are specified as unilateral constraints of the form

$$0 < l_i^{\min} \leq l_i \leq l_i^{\max}, \quad i = 1, \dots, 6 \quad (1)$$

They can be accounted for by introducing auxiliary input variables  $v_i$  through the relations

$$l_i = l_i^0 + l_i' \sin v_i \quad (2)$$

where  $l_i^0 = (l_i^{\max} + l_i^{\min})/2$ ,  $l_i' = (l_i^{\max} - l_i^{\min})/2$ ,  $i = 1, \dots, 6$ . As  $v_i$  varies from  $-\pi/2$  to  $\pi/2$ , the length  $l_i$  varies from  $l_i^{\min}$  to  $l_i^{\max}$ , ensuring satisfaction of the actuator length unilateral constraints. Kinematic constraint equations for the Stewart platform manipulator are composed of six distance constraints associated with the six actuators and one normalization constraint on the Euler rotation vector  $\beta$  [6]. The distance constraint equation for the  $i$ -th actuator is

$$\Phi_i = (\mathbf{r} + \mathbf{A}(\alpha, \beta)\mathbf{a}_i' - \mathbf{t}_i)^T(\mathbf{r} + \mathbf{A}(\alpha, \beta)\mathbf{a}_i' - \mathbf{t}_i) - (l_i^0 + l_i' \sin v_i)^2 = 0 \quad (3)$$

where  $\mathbf{r}$  is the vector from the origin of the  $x - y - z$  frame to the origin of the  $x' - y' - z'$  frame in Fig. 1 and  $\mathbf{a}_i'$  and  $\mathbf{t}_i$  are the constant local coordinate vectors to the upper and lower actuator attachment locations. Actual data from the Iowa Driving Simulator's Stewart Platform motion base [9] are used in the computations described below. The matrix [6]

$$\mathbf{A}(\alpha, \beta) = \cos \alpha \mathbf{I} + (1 - \cos \alpha)\beta\beta^T + \sin \alpha \tilde{\beta} \quad (4)$$

is the orientation transformation matrix for the upper platform, where  $\alpha$  is the angle of rotation of the upper platform about a

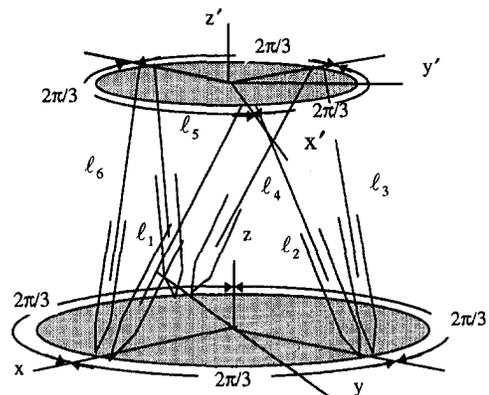


Fig. 1 A spatial Stewart platform