Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

Meccanica Applicata alle Macchine

Ciclo XX

Settore scientifico disciplinare di afferenza: ING-IND/13

TOPOLOGICAL ANALYSIS OF SINGULARITY LOCI FOR SERIAL AND PARALLEL MANIPULATORS

Presentata da: Davide Paganelli

Coordinatore Dottorato

Chiar.mo Prof. Vincenzo Parenti Castelli

Relatore

Chiar.mo Prof. Carlo Innocenti

Esame finale anno 2008

Topological Analysis of Singularity Loci for Serial and Parallel Manipulators

Davide Paganelli

Contents

Acknowledgment

I am grateful to Professor Carlo Innocenti for his invaluable help during the period of my Phd.

My wife Chiara and my son Damiano deserve special thanks, for their patience and their help while writing this thesis.

Introduction

Singularities of robot manipulators have been intensely studied in the last decades by researchers of many fields.

Until the end of the eighties, it was commonly believed that serial singularities had to be met whenever a serial manipulator performs a posture-changing path. Many examples disproved this belief, and gave rise to an increasing interest of many researchers, who investigated the conditions that made a singularity-free posture change possible. This led to the study of the partition induced by the singularity locus in the jointspace of serial manipulators, in order to understand into how many pieces the jointspace is cut by the singularity-locus, and how the different postures are distributed into such pieces. However, the motivation to the study of serial singularities is mainly theoretical, because they cause no damage to the manipulator, unless possibly some local loss of dexterity.

Parallel machines became more and more attractive for many applications in the nineties, because of their higher stiffness, speed, and pay-load. Unfortunately, one of their main drawbacks are parallel singularities. Unlike serial manipulators, when a parallel manipulator crosses a parallel singularity, the control of the platform is locally lost, which can jeopardize the manipulator itself, or the environment wherein it is working. Therefore, all possible care must be taken for avoiding parallel singularities, which is again strictly related to the shape of the singularity locus, and to how it splits the workspace of the parallel machine.

Both problems of serial and parallel singularities lead to the study of the shape of the singularity locus, with particular attention to connectedness. The mathematical tools that study the shape of objects have been developed in the last two century by topologists. These tools will be recalled in Chapter 1, and will be used in Chapter 2 to develop a numerical procedure capable of analyzing the singularity locus of a given manipulator.

This procedure, which is the core of this work, is able to answer the ensuing two questions, for a given manipulator:

- How mamy singularity-free regions are partitioned by the singularity

in the space containing all possible positions of a manipulator?

- Given any two configuration of a manipulator, is it possible to connect them through singularity-free paths?

Chapter 3 and Chapter 4 present some examples of application of the algorithm described in Chapter 2 to serial and parallel manipulators respectively.

Chapter 5 recalls some of the methods used in this work for solving polynomial equations, which are usually the most difficult step of the proposed algorithm, and Chapter 6 concludes the paper.

This work was developed during a three-year Phd. at the department of Mechanical Engineering of the University of Bologna (DIEM), and was financed by the Italian Ministry for Research.

Chapter 1

Differential Topology

This Chapter recalls the mathematical tools that will be used in the entire work. The material for Section 1.1 and Section 1.2 has been mainly retrieved from textbooks such as [1], [2], [3], and [4], whereas the material about Morse Theory, reported in Section 1.3, stems from [5]. The main difficulty in understanding the meaning of some definitions and theorems, especially in Topology, is the very abstract nature of many concepts, which hardly admits a physical interpretation. To this extent, the on-line free encyclopedia www.wikipedia.org has been of great help for grasping the core ideas of many results.

1.1 Historical Background

Differential topology and differential geometry are two branches of mathematics strictly related to each other, with the common aim of studying the shape of smooth objects.

The idea of differential geometry is as old as Euclidean geometry. Indeed, since Euclid included in its *Elements* the famous "fifth postulate", many geometers throughout two thousand years tried to prove it by means of the others. The reason why such an effort has been made to prove this postulate, is that anyone feels it is true, but at the same time not as self-evident as the others. Nevertheless, any attempt to prove the fifth postulate led only to realize that removing the fifth postulates from the body of assumptions only produced conclusions which feel absurd, but neither contradict the first four postulates, nor appear more self-evident than the fifth one.

Only in the early nineteenth century the mathematicians realized that the problem was not the fifth postulate itself, but in a possible misunderstanding of the definitions, and of the "primitive concepts" that had been

implicitly taken for granted by Euclid. For example, the concepts of plane and straight line never seemed to need any definition, for they appeared to be naturally nested in the human mind. However, a casual reader of Euclid's *Elements* might fancy a "plane" as spherical surface, and a "straight line" as a maximum circle on it. Such reader would soon discover that all first four postulates make sense, but the fifth one is definitely wrong. In this way, new geometries can be constructed on objects like the spherical surface, where the fifth postulate does not hold, but many results that seem absurd on a plane become reasonable. For example, the sum of internal angles in a triangle is greater than π on a sphere.

The basis of such new noneuclidean geometries is the concept of *manifold*, first introduced by Bernard Riemann, while formally developing for the first time noneuclidean geometries. The German word used by Riemann was *Mannigfaltigkeit* (see [6] page 33), which literally means something multifarious, which can vary in many ways. Two main features are essential to understand the idea of manifold. First, a manifold can be considered as an object on its own, i.e. all its properties can be defined intrinsically, with no need of a Euclidean space containing it. For example, there is no need to think of a spherical surface embedded in a three-dimensional Euclidean space containing it: all its essential properties can be defined just using points *inside* the sphere. When the spherical surface is considered as a manifold, it becomes the entire universe, and what lays *outside* the sphere looses any meaning, exactly in the same way the outside of the Euclidean plane becomes meaningless while studying Euclidean geometry. Second, the manifold *locally resembles* a Euclidean space. For example, in order to realize that the earth surface has the geometry of a sphere, it is necessary to undertake a long travel, or to look *outside* the earth, for all the geometric properties of the neighborhood of each point cannot be told apart from those of a plane.

Differential geometry mainly developed the study of the *local* properties (such as curvature) of smooth manifolds, i.e. manifolds which are similar enough to a Euclidean space to enable the use of differential calculus. On the other hand, another branch of mathematics, topology, developed more or less in the same period the study of *global* properties of objects. Indeed, there are some properties, such as connectedness, or the number of holes, that do not depend on the local features of the object at hand, and do not change even after heavy deformations of an object.

The first to analyze a problem under this global point of view was probably Euler. A common pastime in the city of Königsberg where Euler lived was to try to find a path crossing all seven bridges of the city once, and only once (see Figure 1.1). In 1736, Euler published a paper proving that this was impossible. The title of the paper was (translated in English) "*The solution*

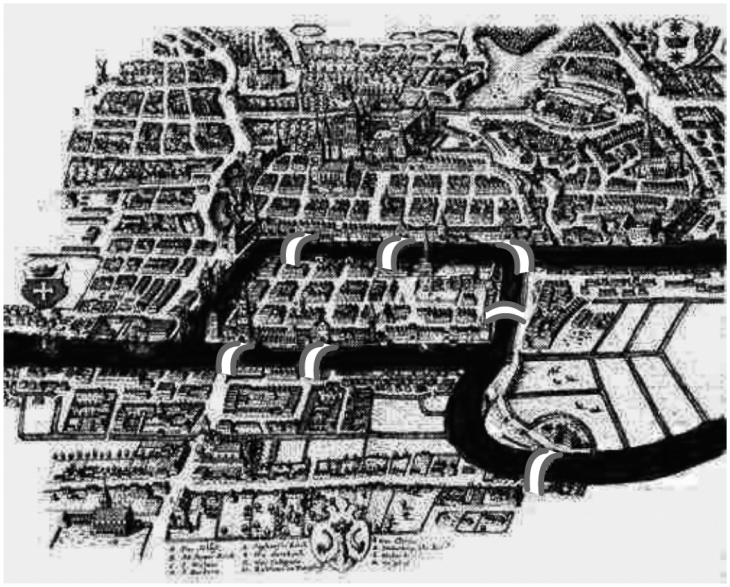


Figure 1.1: The seven bridges of Königsberg

of a problem relating to the geometry of position", implicitly meaning that the problem tackled in the paper dealt with a new kind of geometry, that did not care about the exact measures of distances between points, but was mainly concerned with the global shape of objects. Indeed, the possibility of finding a path crossing all seven bridges did not depend on the distance between the bridges, or on their length, but on how the river cut the city, and which ones of the different islands cut by the river were joined by the bridges, which is preeminently a topological problem.

Differential topology studies the topological properties of smooth manifolds. One of the first and more important tools for this purpose is Morse theory, whose development was started by Marston Morse in the twenties. This theory establishes a strong relationship between the number and the type of critical points of a smooth real-valued function defined on a smooth manifold and the shape of the manifold. Therefore, by studying the critical points of a suitable function defined on a manifold it is possible to obtain information about its shape, and, conversely, by knowing the shape of a manifold it is possible to foresee that any function defined on it will have a least amount of critical points with certain characteristics.

In the next sections, the main topological concepts will be recalled first, and the formal definition of manifold will be given. Then, the main results of Morse theory will be stated and discussed.

1.2 Basic Concepts

1.2.1 Topological space

The first thing to be defined is the object studied by topology, i.e. the *topological space*:

Definition 1.1 *A* **topological space** *is a set $\mathscr{X}$ together with a collection $\mathscr{T}$ of subsets of $\mathscr{X}$ such that*

- *the empty set and the whole set $\mathscr{X}$ are contained in $\mathscr{T}$*
- *the union of any collection of elements of $\mathscr{T}$ is also contained in $\mathscr{T}$*
- *the intersection of any finite collection of elements of $\mathscr{T}$ is also contained in $\mathscr{T}$*

The collection of subsets $\mathscr{T}$ is said to be a *topology* on $\mathscr{X}$. The elements of $\mathscr{X}$ are called points, whereas the elements of $\mathscr{T}$ are called *open sets*. The *closed sets* are defined as the complements of the open sets.

Note that, according to this definition, the idea of open set depends on the topology defined on the set $\mathscr{X}$, and that different topologies can be generated by changing the definition of what an open set is. A distance d[1] always produces a *metric* topology on a set $\mathscr{X}$. When a distance is defined, it is immediately defined the concept of open ball centered on a point:

Definition 1.2 *An* **open ball** *of radius r centered at a point P of $\mathscr{X}$ is a subset of $\mathscr{X}$ containing all points $Q \in \mathscr{X}$ such that $d(P,Q) < r$.*

Then, a subset of $\mathscr{X}$ can be defined as open if it contains at least one open ball centered at each of its points. The collection of open sets just defined is the metric topology on $\mathscr{X}$.

Consider for example the sets $\mathscr{X}_1$ and $\mathscr{X}_2$ represented in Figure 1.2, containing all the points X of the Euclidean plane such that $d\,(X,(0,0)) \leq 1$, and $d\,(X,(3,0)) \leq 1$, respectively, where d is the usual Euclidean distance on the Euclidean plane. Then let the set $\mathscr{X}$ be the union of $\mathscr{X}_1$ and $\mathscr{X}_2$.

The Euclidean distance restricted to the set $\mathscr{X}$ defines a distance on $\mathscr{X}$, and therefore a *metric* topology. Note that, according to the definition of topology, the whole set $\mathscr{X}$ and the empty set are both closed and open, or, as it is often said, *clopen*. Moreover, each of the two circles $\mathscr{X}_1$ and $\mathscr{X}_2$ is clopen. Indeed, the circle $\mathscr{X}_1$ is open, because, for each of its point, one

[1] A distance is a real valued function $d(X,Y)$ of two points X and Y of $\mathscr{X}$, such that $d(x,y) \geq 0$, $d(x,y) = d(y,x)$, and $d(x,z) \leq d(x,y) + d(y,z)$.

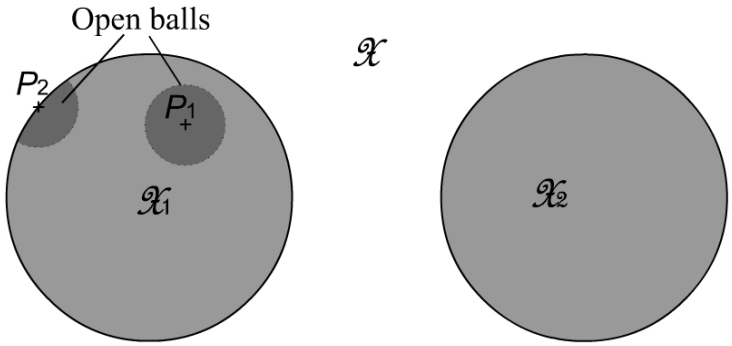


Figure 1.2: Example of a topological space

can find an open ball centered at this point and completely contained in $\mathscr{X}_1$, as shown in Figure 1.2. For the same reason also $\mathscr{X}_2$ is open. However, since the union of $\mathscr{X}_1$ and $\mathscr{X}_2$ yields the whole set $\mathscr{X}$, such two sets are the complement of each other, and therefore also closed.

1.2.2 Relationships between topological spaces

With the definition of topological space,, it is now possible to define how to establish whether two spaces have the same shape or not. The idea is that if a topological space can be deformed into another by stretching or bending it, but without cutting it, or gluing pieces to it, then its basic shape features, such as number of holes or disconnected pieces remain the same. A common joke about the topologist, is that he cannot tell the donut he is eating from the coffee cup he is drinking from. Indeed, both of them are made of one piece and have one hole, and one could easily make a clay donut out of a clay coffee-cup by simply stretching and bending it, with no need of cutting or adding pieces.

The concept of such a continuous deformation is formalized in many different ways, all based on continuous functions. The first one is homeomorphism.

Definition 1.3 *A function f from a topological space $\mathscr{A}$ to a topological space $\mathscr{B}$ is a* **homeomorphism** *if*

- *f is a bijection*
- *f is continuous*
- *the inverse function $f^{-1} : \mathscr{B} \to \mathscr{A}$ is continuous.*

$\mathscr{A}$ and $\mathscr{B}$ are said ***homeomorphic*** *if there exists a homeomorphism between them. If the function f is also differentiable, then it is named a* ***diffeomorphism*** *and $\mathscr{A}$ and $\mathscr{B}$ are said to be* ***diffeomorphic****.*

For example, a ball can be stretched to a cube through homeomorphism, but not through diffeomorphism. Diffeomorphisms are a little bit more demanding, for they require differentiability, therefore they cannot produce the edges of the cube out of the smooth boundary of the ball. Being homeomorphic is an equivalence relation, and partitions the set of topological spaces into equivalence classes. The same can be said for being diffeomorphic, which, adding more requirements subdivides each of the previous classes into subclasses of diffeomorphic spaces.

The classification induced by homeomorphisms is very strong, so strong that in topology it is said that a topological space *is* a sphere if the topological space is homeomorphic to the sphere. Yet, there are many topological features which are the same for non-homeomorphic topological spaces. For instance, a three-dimensional ball and a point possess a similar shape in many ways, for they are both composed of a single connected piece and have no holes. Nevertheless, a ball is obviously not homeomorphic to a point. In order to introduce a more flexible classification, homotopy equivalence has been introduced. To understand this relation, it is first necessary to define an equivalence relation among functions:

Definition 1.4 *Two continuous functions f and g from a topological space $\mathscr{A}$ to a topological space $\mathscr{B}$ are* **homotopic** *if there exists a continuous function $h : [0,1] \times \mathscr{A} \to \mathscr{B}$ such that $h(0,P) = f(P)$ and $h(1,P) = g(P)$ for any point P in $\mathscr{A}$.*

The homotopy function h is used to smoothly deform the function f into the function g. This smooth deformation is conveyed on the topological spaces through the ensuing definition:

Definition 1.5 *Two topological spaces $\mathscr{A}$ and $\mathscr{B}$ are* **homotopy equivalent** *if there exist two functions $f : \mathscr{A} \to \mathscr{B}$ and $g : \mathscr{B} \to \mathscr{A}$ such that $g \circ f$ is homotopic to the identity on $\mathscr{A}$ and $f \circ g$ is homotopic to the identity on $\mathscr{B}$.*

Homotopy is an equivalence relation in the space of functions, and homotopy equivalence is an equivalence relation in the set of topological spaces.

It is useful, in order to prove the theorems of Section 1.3, to give a restricted definition of homotopy, for specific application to subsets.

Definition 1.6 *A function $r : \mathscr{A} \to \mathscr{S}$ from a topological space $\mathscr{A}$ to a subset $\mathscr{S} \subset \mathscr{A}$ is a* **retract** *if the restriction of r to $\mathscr{S}$ is the identity function on $\mathscr{S}$.*

A retract is thus a function that brings the whole space into a subset, in such a way that the points of the subset remain fixed. Conversely, the inclusion function can be defined as follows:

Definition 1.7 *A function $i : \mathscr{S} \to \mathscr{A}$ from a subset $\mathscr{S} \subset \mathscr{A}$ to a topological space $\mathscr{A}$ is an* **inclusion** *if i delivers each element of S to the same element in A.*

Inclusion and retract can be used to define this special case of deformation retract:

Definition 1.8 *A subset $\mathscr{S}$ of a topological space $\mathscr{A}$ is a* **deformation retract** *of $\mathscr{A}$ if there exists a retract r from $\mathscr{A}$ to $\mathscr{S}$ such that, given the inclusion i, $i \circ r$ is homotopic to the identity map on $\mathscr{A}$.*

The deformation retract is just a special case of homotopy equivalence to subsets. If a subset is a deformation retract of the whole space, than the subset and the whole space are always homotopy equivalent.

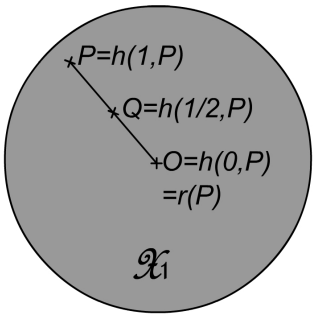


Figure 1.3: Example of deformation retract.

Consider for example Figure 1.3. The topological space $\mathscr{X}_1$ is defined as in Figure 1.2, i.e. it is the closed ball of radius 1 centered at the origin of the Euclidean plane. Let $\mathscr{S}$ be the subset of $\mathscr{X}_1$ containing the center of the ball O. A retract $r : \mathscr{X}_1 \to \mathscr{S}$ can be defined, such that $r(P) = O$ for any $P \in \mathscr{X}_1$. The inclusion composed to this function is homotopic to the identity on $\mathscr{X}_1$. To see this, it is easy to construct the ensuing homotopy function $h : [0, 1] \times \mathscr{X}_1 \to \mathscr{X}_1$:

$$h(t, P) = O + (P - O)t \tag{1.1}$$

Since h is continuous, $h(1, P)$ is the identity on $\mathscr{X}_1$, and $h(0, P)$ is $i \circ r(P)$, then $\mathscr{S}$ is a deformation retract of $\mathscr{X}_1$. It is easily proved that $\mathscr{X}_1$ and $\mathscr{S}$ are also homotopy equivalent. Consider the retract r and the inclusion i. By definition of retract and inclusion, $r \circ i$ is the identity on $\mathscr{S}$, and therefore trivially homotopic to the identity on $\mathscr{S}$. The function $i \circ r$ is homotopic to the identity, because $\mathscr{S}$ is a deformation retract of $\mathscr{X}_1$, therefore $\mathscr{S}$ and $\mathscr{X}_1$ are homotopy equivalent, by Definition 1.5. The continuous shrinkage of the space $\mathscr{X}_1$ into the subset $\mathscr{S}$ can be visualized by considering, for any P, the paths $h(t, P)$, while t evolves from 0 to 1. It can be therewith realized that each point is continuously conveyed to the center upon radial continuous paths, and that $h(t, \mathscr{X}_1)$ is a ball centered at O with radius t.

1.2.3 Properties of Topological spaces

Many topological properties can be defined to characterize the shape of a topological space. This section will recall some of them, that will be used in the following chapters.

The first important topological property is compactness, which is the base to prove many cornerstone theorems. The easiest definition for compactness applies to subsets of Euclidean spaces: a subset of $\mathscr{R}^n$ is compact if it is closed and bounded. Unfortunately, this definition is not intrinsic to the subset, because to decide whether a space is compact or not it is always unavoidable to embed it in a Euclidean space of proper dimension. Consider for example the open interval $\mathscr{I} = (0, 1)$. If $\mathscr{I}$ is included in the Euclidean space $\mathscr{R}$, with the usual metric topology, then $\mathscr{I}$ is not compact, because it is open. On the other hand, if the same definition were applied to $\mathscr{I}$ considered as topological space on its own, with the same metric topology, then it would be compact, because it is closed (see Section 1.2.1) and bounded. Therefore, in order to formalize the idea of compactness with no need of a broader Euclidean space, the ensuing definition is given:

Definition 1.9 *A topological space $\mathscr{X}$ is* **compact** *if any open cover of $\mathscr{X}$ has a finite subcover.*

To understand this definition, it is necessary to recall what an open cover is. Let an *indexed family* be a function $\mathscr{U}(A) : \mathscr{A} \to \mathscr{T}$, that associates to each point A of a given set $\mathscr{A}$ an open set contained in the topology $\mathscr{T}$ on $\mathscr{X}$. The collection of all the open sets $\mathscr{U}(A)$, contained in the image of the indexed family, is an open cover of $\mathscr{X}$ if their union contains the whole set $\mathscr{X}$. A subcover is a subset of the cover that is still a cover of $\mathscr{X}$. If, for every open cover it is possible to extract a subcover composed of a finite number of elements, then the space is compact.

Consider again the interval $\mathscr{I} = (0, 1)$. Then, define the indexed family $\mathscr{U}(r) : \mathscr{R}^+ \to \mathscr{T}$ that associates to any positive real number r the open interval $(1/r, 1)$. The sets of all $\mathscr{U}(r)$ is an open cover of $(0, 1)$, but there exists no finite subcover, hence $\mathscr{I}$ is not compact, regardless of the topological space that might contain it.

Another crucial property for this work is *connectedness*, which formalizes the idea that a topological space is composed of one single piece. The first, and simplest definition of connectedness is the following:

Definition 1.10 *A topological space is* **disconnected** *if it can be obtained as the union of two disjoint open sets,* **connected** *otherwise.*

In Definition 1.10, *disjoint* means that the intersection is the empty set. For example, the topological space $\mathscr{X}$ of Figure 1.2 can be obtained through the union of the two disjoint open sets $\mathscr{X}_1$ and $\mathscr{X}_2$, therefore $\mathscr{X}$ is disconnected.

It might seem obvious that a finite path between any points of a connected topological space always exists. Surprisingly, this is not true: there are some examples[2] of connected manifolds, whose points are not always mutually reachable through finite paths. Therefore, a stronger definition is required, to include this property. First, let a *path* be defined as follows:

Definition 1.11 *A* **path** *between two points A and B of a topological space $\mathscr{X}$ is a continuous function $p : [0, 1] \to \mathscr{X}$ such that $p(0) = A$ and $p(1) = B$. If $A \equiv B$, then the path is* **closed**.

It is now possible to define path-connectedness:

Definition 1.12 *A topological space $\mathscr{X}$ is path-connected if there exists a path between any two points in $\mathscr{X}$.*

For the purpose of this work, path-connectedness will be the essential and more important notion, therefore, from now on, the word connected will be used to mean path-connected.

If a set is connected, we know that it is composed of just one single piece. On the other hand, an important property of a non-connected topological space is the number of connected pieces it is composed of. Let first the following equivalence relation on the points of the topological space $\mathscr{X}$:

Definition 1.13 *The point B is* **reachable** *from the point A if there exists a path between A and B.*

[2]For example the topologist's sine curve is connected but not path-connected. For more details see http://en.wikipedia.org/wiki/Topologist%27s_sine_curve.

It is now possible to define the disjoint regions composing $\mathscr{X}$:

Definition 1.14 *The equivalence classes containing the reachable points are named* **disjoint regions**.

The number of disjoint regions is preserved through homeomorphism and homotopy equivalence. This important property will be very useful in the following work, therefore a proof will be hereafter provided.

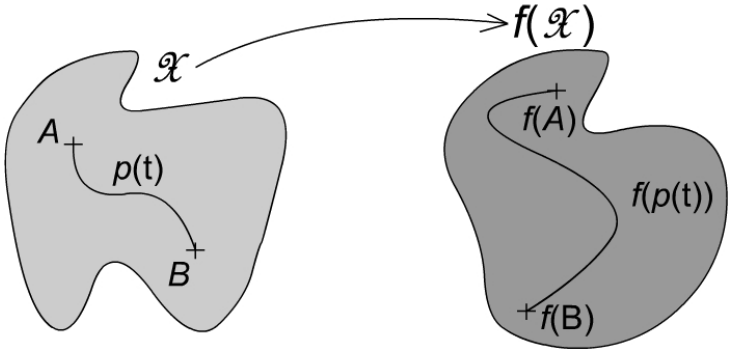


Figure 1.4: The image of a connected space through a continuous function is connected.

Let us first prove the ensuing lemma:

Lemma 1.1 *The image of a connected space through a continuous function is connected.*

With reference to Figure 1.4, consider a connected topological space $\mathscr{X}$, and its image $f(\mathscr{X})$ through a continuous function f. By definition of image, for any two points A' and B' in $f(\mathscr{X})$ there exist two points A and B in $\mathscr{X}$ such that $f(A) = A'$ and $f(B) = B'$. Since $\mathscr{X}$ is connected, there exists a path $p(t)$ in $\mathscr{X}$ such that $p(0) = A$ and $p(1) = B$. The function $q = f \circ p$ is a path, for f is continuous, and $q(0) = A'$, $q(1) = B'$. Such a path q can be found for any A' and B' in $f(\mathscr{X})$, therefore $f(\mathscr{X})$ is connected.

The ensuing theorem will be hereafter proved:

Theorem 1.2 *Two homeomorphic topological spaces are composed of the same number of disjoint regions.*

With reference to Figure 1.5, consider two homeomorphic spaces $\mathscr{X}$ and $\mathscr{Y}$. By definition, there exists a homeomorphism $f : \mathscr{X} \to \mathscr{Y}$. Let m be the number of disjoint regions of $\mathscr{X}$, and n the number of disjoint regions of

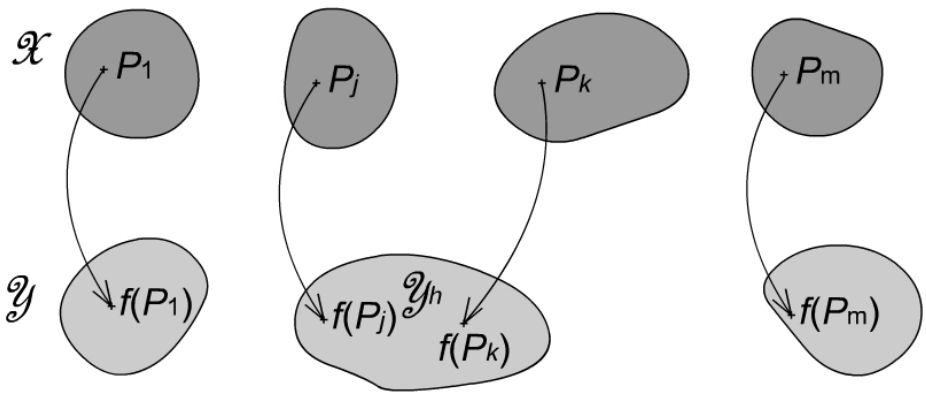


Figure 1.5: Two homeomorphic spaces are composed of the same number of disjoint regions.

$\mathscr{Y}$. m and n will be supposed finite, which is enough for the applications in this work. Suppose now, by absurd, that $m > n$. It is possible to find a point P_i, $i = 1, \ldots, m$, contained in each of the m disjoint regions of $\mathscr{X}$, therefore none of the points P_i is reachable from any other point P_j, if $i \neq j$. There must exist at least two images $f(P_j)$ and $f(P_k)$, with $j \neq k$, contained in the same disjoint region $\mathscr{Y}_h$ of $\mathscr{Y}$, because $m > n$ (see Figure 1.5). The image $f^{-1}(\mathscr{Y}_h)$ therefore contains both points $f(P_j)$ and $f(P_k)$, and must be disconnected. This conclusion is absurd, because $\mathscr{Y}_h$ is connected, and its image through the continuous function f^{-1} must be connected, which proves that $m \leq n$. Symmetrically, it can be proved that $n \leq m$, therefore m and n must be equal. Obviously, Theorem 1.2 is valid also for diffeomorphic spaces, because if two spaces are diffeomorphic they are always homeomorphic.

The same result can be stated also for homotopy equivalent spaces:

Theorem 1.3 *Two homotopy equivalent topological spaces are composed of the same number of disjoint regions.*

With reference to Figure 1.6, consider again two homotopy equvalent spaces $\mathscr{X}$ and $\mathscr{Y}$. By definition of homotopy equivalece, there exist two continuous functions $f : \mathscr{X} \to \mathscr{Y}$ and $g : \mathscr{Y} \to \mathscr{X}$, such that their compositions are homotopic to the identities on the two topological spaces. As for the previous case, m and n are supposed to be the finite number of disjoint regions composing $\mathscr{X}$ and $\mathscr{Y}$, respectively, and, by absurd, m is supposed greater than n. Consider again m points P_i, $i = 1, \ldots, m$, contained in different disjoint regions of $\mathscr{X}$. Since $m > n$, there must be at least two images $f(P_j)$ and $f(P_k)$, with $j \neq k$, contained in the same disjoint region $\mathscr{Y}_h$ of $\mathscr{Y}$. If these

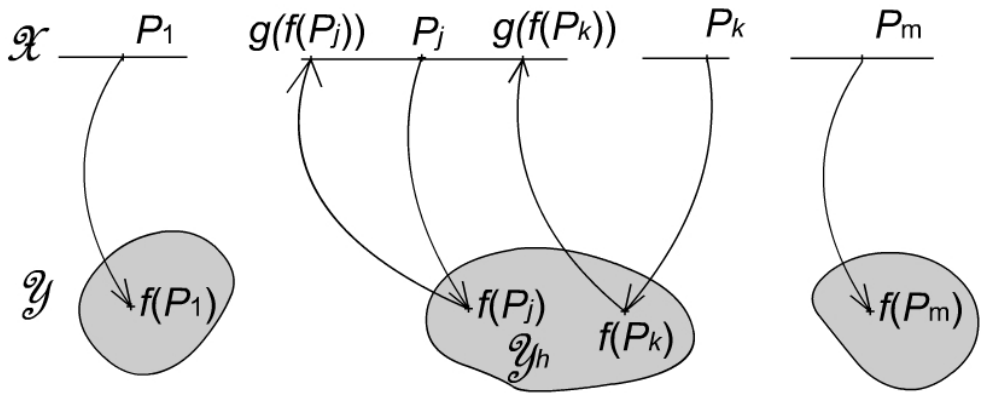


Figure 1.6: Two homotopy equivalent spaces are composed of the same number of disjoint regions.

two images are sent back to $\mathscr{X}$ through g, they must be delivered to the same disjoint region of $\mathscr{X}$, for g is continuous. Therefore, either $g(f(P_j))$ does not belong to the same disjoint region as P_j, or $g(f(P_k))$ does not belong to the same disjoint region as P_k. Without loss of generality, it will be supposed (see Figure 1.6) that $g(f(P_k))$ does not belong to the same disjoint region as P_k, therefore there exists no path from P_k to $g(f(P_k))$. However, $g \circ f$ must be homotopic to the identity, therefore there must be a continuous function $h : [0,1] \times \mathscr{X} \to \mathscr{X}$ such that $h(0,P) = P$ and $h(1,P) = g(f(P))$ for any $P \in \mathscr{X}$. The univariate function $p = h(t, P_k)$, for $t \in [0,1]$, is a path between P_k and $g(f(P_k)$, which contradicts the hypothesis $m > n$, thus $m \leq n$. It can be symmetrically proved that $n \geq m$, therefore $m = n$.

Another important property, which characterizes the shape of a topological space, is the number of holes. Unfortunately, unlike the number of disjoint regions, the number of holes is not easily reduced to a formal definition, because there exist different types of holes with different topological properties. For example, it is evident that both a sphere and a circle have a *hole*, but the topology of such a hole is very different: every closed path on a sphere can be continuously shrunk to a point (i.e. the sphere is simply connected), wheres on a circle not.

The number of holes of different shapes is formally characterized by *Betti number sequence*. The 0^{th} Betti number of a topological space is the number of disjoint regions composing it. The first Betti number can be intuitively defined as the number of cuts which can be made on a topological space, before it becomes disconnected. For instance, if the operation of *cutting* a torus is defined as the removal of a closed path from the surface, then the

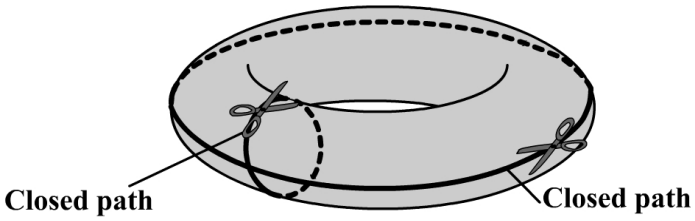


Figure 1.7: The two-dimensional torus can be cut twice, before falling apart.

torus can be cut at most twice, in order to preserve connectedness (see Figure 1.7), therefore the first Betti number of the torus is two.

The rigorous definition of the n^{th} Betti number for $n > 1$ is rather complex, and beyond the scope of this work. The k^{th} Betti number B_k is often referred to as the number of k^{th}-dimensional holes of a topological space, even though this is definition is useless, for it is hard to understand what a k^{th}-dimensional hole is, other than through vague examples, such as that the hole inside a circle is one-dimensional, and the hole inside a sphere is two-dimensional. For a more rigorous definition, refer to [3], suffice it to say here that the sequence of Betti numbers can be defined for most topological spaces that will be considered in this work, as reported in Table 1.1.

Topological Space	k^{th} Betti number B_k
n-dimensional Euclidean Space	1 if $k = 0$, 0 otherwise
n-dimensional sphere	1 if $k = 0$ or $k = n$, 0 otherwise
n-dimensional torus	$\binom{n}{k}$
n-dimensional real projective space	1 if $k \leq n$, 0 otherwise

Table 1.1: Betti numbers of some topologicla spaces

As for connectedness, Betti number sequence is a topological invariant, and it is preserved by homeomorphism and homotopy equivalence.

1.2.4 Manifolds

The objects that will be dealt with in this work, are a special kind of topological space, with more specific properties, i.e. manifolds, that are defined as follows[3]:

[3]Usually, this definition is stronger, for it includes also that the topological space is Hausdorff and second-countable. However, such details are beyond the aim of this work, for more information, see [2].

Definition 1.15 *An n^{th}-dimensional* **manifold** *is a topological space that is locally homeomorphic to an open ball in the n^{th}-dimensional Euclidean space.*

Locally homeomorphic means that there exists a neighborhood of any point of a manifold, which is homeomorphic to an open ball of the n^{th}-dimensional Euclidean space. This definition formalizes the idea that a manifold locally resembles the Euclidean space, as discussed in Section 1.1. The homeomorphisms map each neighborhood into the Euclidean space, and vice versa, exactly like topographic charts are used to represent on a plane small parcels of the earth globe. Therefore, such homeomorphisms are called *charts*, or *local coordinate systems*, because they associate to each point of $\mathscr{R}^n$ one and only one point of a portion of the manifold.

This definition is useful, but still does not include an important class of topological spaces. Consider for example the topological space $\mathscr{X}$ of Figure 1.2. The topological space $\mathscr{X}$ is no manifold, in the sense of Definition 1.15, because none of the points of the two circles $d(X,(0,0)) = 1$ and $d(X,(3,0)) = 1$ has a neighborhood homeomorphic to an open ball in the Euclidean plane. To include such cases, the concept of ***manifold with boundary*** has been introduced:

Definition 1.16 *An n^{th}-dimensional* **manifold with boundary** *is a topological space that is locally homeomorphic to an open ball in the n^{th}-dimensional Euclidean space, or to a half ball in the n^{th}-dimensional Euclidean space.*

The half ball is the set containing all the n-tuples $(x_1, x_2, \ldots x_n)$ of the Euclidean space $\mathscr{R}^n$ such that $x_1^2 + x_2^2 + \ldots + x_n^2 < r^2$ and $x_1 \geq 0$.

Also the concept of *interior* and *boundary* can now be defined for manifolds with boundary:

Definition 1.17 *A point of a manifold with boundary is said to be an* **internal point** *if there exists a neighborhood of this point homeomorphic to an open ball in $\mathscr{R}^n$, a* **boundary point** *otherwise.*

Therefore, points such as P_1 in Figure 1.2 are internal points of $\mathscr{X}$, whereas points such as P_2 are boundary points. Note that the concepts of *interior* and *boundary* have been therewith defined intrinsically, without involving points *outside* the manifold, for *outside* has no meaning, for a manifold.

The definition of boundary usually given for subsets of $\mathscr{R}^n$ coincides only partially with Definition 1.17. According to such definition, any neighborhood of a boundary point of a subset of $\mathscr{R}^n$ contains at least one point of the set itself, and one point not contained in the set. The main drawback of this definition is that being or not a boundary point of a set depends on what there is *outside* the set, and might thus be ambiguous. For example,

if the set $\mathscr{X}$ of Figure 1.2, composed of two disks is considered embedded in the plane, then the origin $(0, 0)$ is an interior point. On the other hand, if the whole plane containing $\mathscr{X}$ were considered embedded in the three-dimensional space, any point of the plane would be a boundary point, and therefore any point of $\mathscr{X}$ would be a boundary point. Therefore, the term *boundary* will be henceforth used to designate the boundary of a set considered as a manifold, not the boundary of a subset of $\mathscr{R}^n$.

1.3 Morse Theory

Morse theory is an important branch of differential topology, which tries to gain information about the topological properties (such as connectedness, or Betti numbers, as recalled in Section 1.2.3) of a manifold, from the number and the type of critical points of a smooth function defined on it. The main ideas of Morse Theory will be exposed through an example in the next section, whereas in the following sections rigorous results will be stated and proved.

1.3.1 Visualization

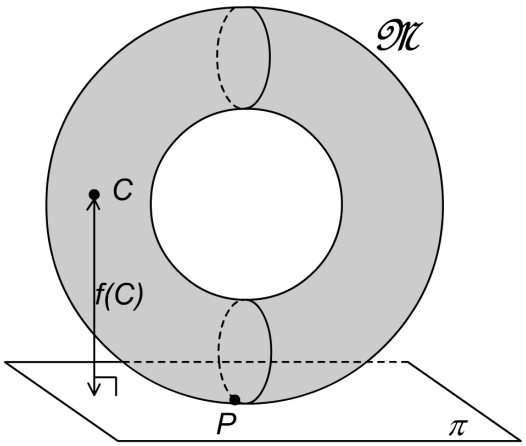


Figure 1.8: Example on the torus $\mathscr{M}$.

Consider a torus $\mathscr{M}$, touching a plane π at the point P, as in Figure 1.8. Then, define the function $f : \mathscr{M} \to \mathscr{R}$ as the height of any point of the torus above the plane π. Consider the set $\mathscr{M}^a$

$$\mathscr{M}^a = f^{-1}(-\infty, a] \tag{1.2}$$

containing all the points of $\mathscr{M}$ below a certain height a. We now observe the changes in the topological properties of $\mathscr{M}^a$ while the height a grows from $-\infty$ to $+\infty$. Such changes are described through the changes in the homotopy type.

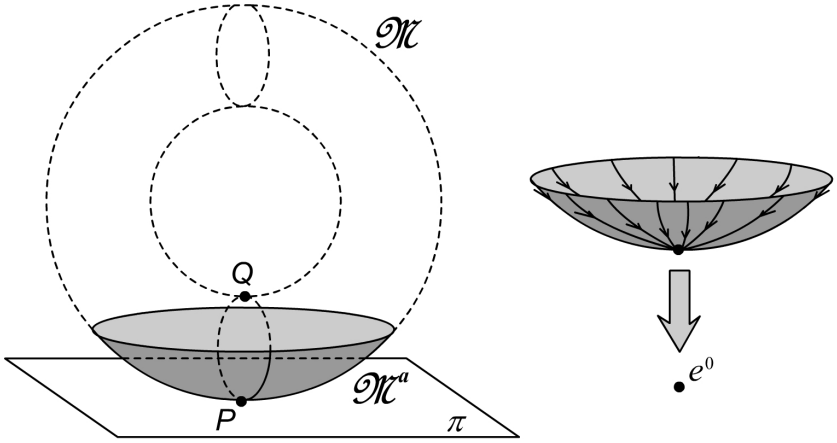


Figure 1.9: If $f(P) < a < f(Q)$, $\mathscr{M}^a$ is homotopy equivalent to a point.

As long as a remains lower than 0, $\mathscr{M}^a$ is the empty set. As soon as a grows above 0-level, the set $\mathscr{M}^a$ changes its topological properties, and becomes homotopy equivalent to a point (see Figure 1.9). Indeed, the whole set $\mathscr{M}^a$ could be continuously shrunk to the point P.

Let a *k-cell* be rigorously defined as follows:

Definition 1.18 *A* **k-cell** e^k *is the k-dimensional ball with boundary of radius 1, i.e. the set containing all points* $\mathbf{x} \in \mathscr{R}^n$ *such that* $\|\mathbf{x}\| \leq 1$.

Therefore the k-cell is a manifold with boundary, and the boundary is the $(k-1)$-dimensional sphere $\|\mathbf{x}\| = 1$. The change in the topology of $\mathscr{M}^a$, when a changes its sign, can be described as adding a 0-cell e^0 to the empty set.

The point Q is the closest point to P where the tangent plane is parallel to π. It easy to see that, as long as $0 < a < f(Q)$, the set $\mathscr{M}^a$ stretches, but is still contractible to a point, therefore its topological properties are preserved.

As soon as a reaches the height $f(Q)$, the topological properties of the set $\mathscr{M}^a$ undergo again a sudden change. A "handle" is glued to the set $\mathscr{M}^a$, which is now not contractible to a point anymore, but is homotopy equivalent to a circle. The operation of gluing a handle to $\mathscr{M}^a$, can be formally defined as *attaching a k-cell*:

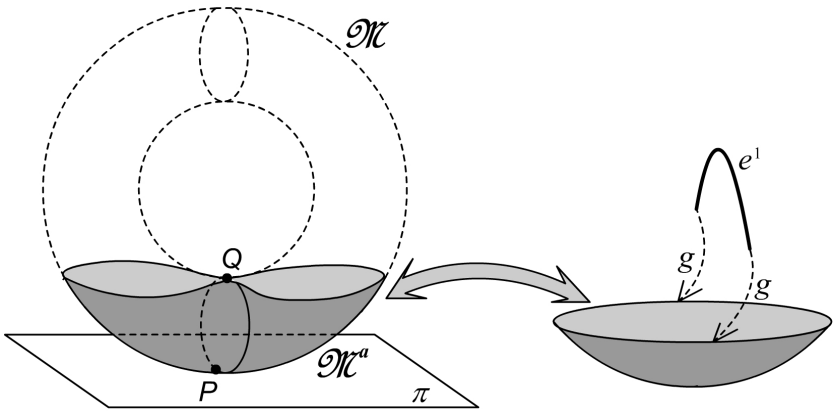


Figure 1.10: As soon as $a = f(Q)$, a 1-cell is attached to $\mathscr{M}^a$.

Definition 1.19 *By* **attaching a k-cell** *to a topological space $\mathscr{X}$, a topological space $\mathscr{Y}$ is obtained through the ensuing operations:*

- *$\mathscr{Y}$ is initially the empty set;*
- *the disjoint union of the k-cell and $\mathscr{X}$ is included in $\mathscr{Y}$;*
- *a continuous function g from the boundary of the k-cell to $\mathscr{X}$ is defined;*
- *each point of the boundary of the k-cell is identified with its image in $\mathscr{X}$, i.e. each point and its image are henceforth considered as the same point of $\mathscr{Y}$.*

Thus, as soon as $a = f(Q)$, a 1-cell is attached to $\mathscr{M}^a$ by means of a continuous function g, and becomes the "handle" we need to describe the variation of the homotopy type of $\mathscr{M}^a$ (see Figure 1.10).

If at the point R the tangent plane of the torus is again parallel to π (see Figure 1.11), then we see that so far as $f(Q) < a < f(R)$ the set $\mathscr{M}^a$ grows while a grows, stretches and bends, but still remains homotopy equivalent to a 1-cell attached to a 0-cell. Therefore, the number of pieces and the number of holes in $\mathscr{M}^a$ remain the same as long as $f(Q) < a < f(R)$.

As soon as a reaches the height of R, the topology of $\mathscr{M}^a$ changes abruptly once more: another 1-cell is attached to it (see Figure 1.12). Again, between the heights of the two points R and S, where the tangent plane is parallel to π, the topological properties remain unchanged.

Eventually, the maximum height $f(S)$ is reached by a, which produces the last change in topological properties of $\mathscr{M}^a$: a 2-cell is attached (see Figure 1.13). From $f(S)$ upward, the set $\mathscr{M}^a$ coincides with the whole torus $\mathscr{M}$,

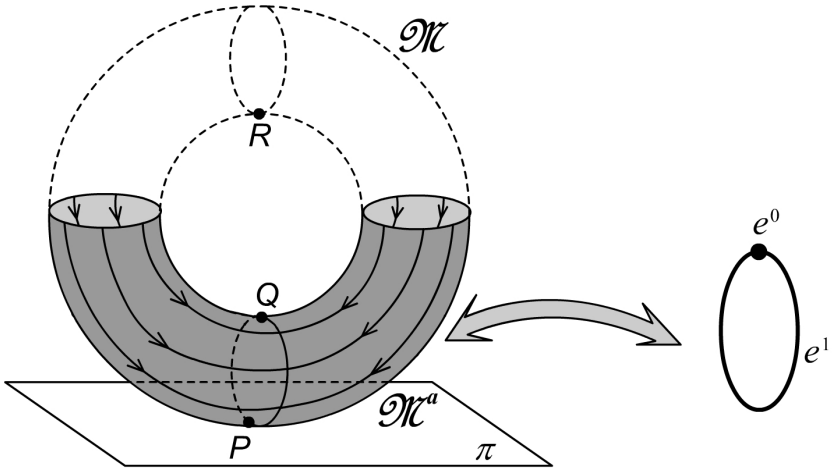


Figure 1.11: While $f(Q) < a < f(R)$, the topological properties of $\mathscr{M}^a$ do not change.

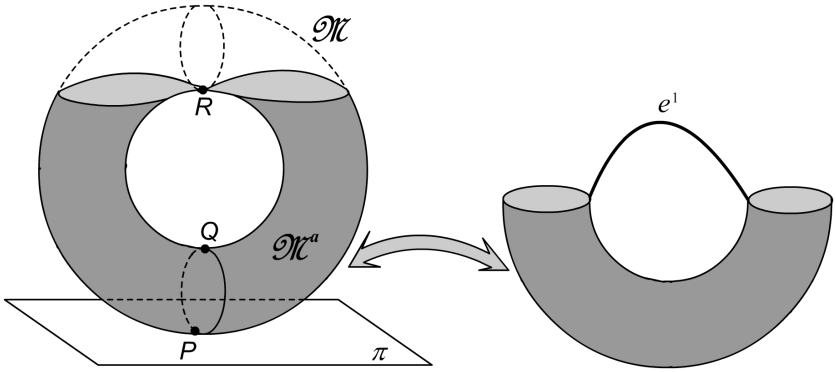


Figure 1.12: As soon as $a = f(R)$, another 1-cell is attached to $\mathscr{M}^a$.

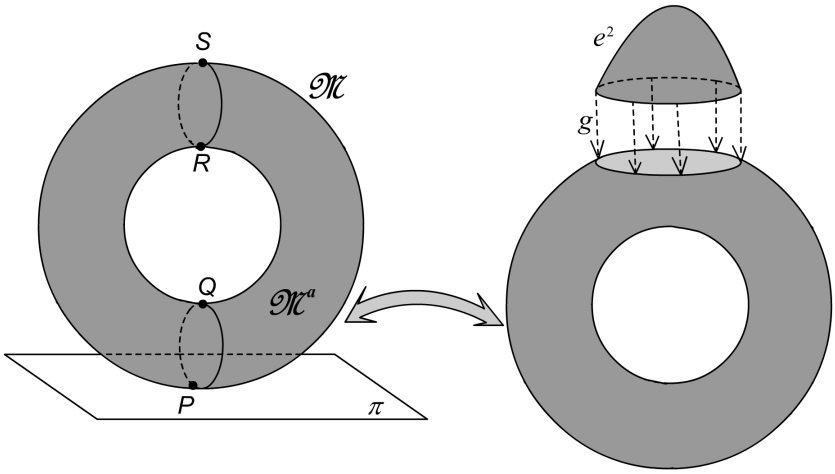


Figure 1.13: As soon as $a = f(S)$, a 2-cell is attached to $\mathscr{M}^a$.

thus the way $\mathscr{M}^a$ has been built adding k-cells provides information about how the whole manifold $\mathscr{M}$ can be built by successive attachment of k-cells.

Note that changes in the topological properties of $\mathscr{M}^a$ are discrete, and occur suddenly as soon as the points P, Q, R, and S are included into the manifold $\mathscr{M}^a$. It is easy to see that such points are critical points of the function f on $\mathscr{M}$, and that different types of critical points produce the attachment of different cells. Indeed, the minimum P attaches a 0-cell, the saddles Q and R attach 1-cells, and the maximum S attaches a 2-cell.

In the next section, the process just outlined will be formalized in a set of important theorems.

1.3.2 Definitions

The first concept to be defined is *smoothness*, both for manifolds and functions:

Definition 1.20 *A* **manifold** $\mathscr{M}$ *is* **smooth** *if all the homeomorphisms mapping neighborhoods of* $\mathscr{M}$ *into an open ball in the Euclidean space are of class* C^∞.

Definition 1.21 *A* **real-valued function** f *on a smooth manifold* $\mathscr{M}$ *is* **smooth** *if* $f \in C^\infty$ *on* $\mathscr{M}$.

We consider henceforth a smooth real-valued function f on a smooth n-dimensional manifold $\mathscr{M}$. In a neighborhood of every point, it is possible to

define a local coordinate system $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, which unambiguously identifies any point of the neighborhood. Thus, the *gradient* of the function f is an n-dimensional vector, defined as

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right) \tag{1.3}$$

The definition of *critical point* naturally stems from the definition of gradient:

Definition 1.22 *A point C of $\mathscr{M}$ is a* **critical point** *of f, if $\nabla f \mid_C = \mathbf{0}$.*

The second-order variations of f must be considered, to characterize the different effects of different types of critical points. Such second-order variations are controlled by the Hessian matrix of f at a critical point C, $H_f \mid_C$, which is defined as the $n \times n$ square matrix, whose elements h_{ij} are defined as:

$$h_{ij} = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_C \tag{1.4}$$

Definition 1.23 *A critical point of f is* **nondegenerate** *if the Hessian matrix $H_f \mid_C$ is nonsingular.*

If a critical point is nondegenerate, the Hessian matrix completely characterizes the shape of the function f in the neighborhood of a critical point (see Section 1.3.4). In particular, what is important to know is the *index*:

Definition 1.24 *The* **index** *of a nondegenerate critical point C is the number of negative eigenvalues of the Hessian matrix $H_f \mid_C$.*

It will be shown in the rest of this section that the index determines the dimension of the cell which is attached to the set $\mathscr{M}^a$, as soon as as it includes a new critical point of f.

The definitions of nondegenerate critical point and index have been here based upon a local system of coordinates on the manifold $\mathscr{M}^a$. Such definition can be given intrinsically (see [5], pp. 4-6), and it can be proven that neither being a nondegenerate critical point nor the index depend on the specific coordinate system used to calculate the gradient or the Hessian matrix.

1.3.3 Between critical points

Before analyzing in detail what happens when a critical point is included in the set $\mathscr{M}^a$, it will be hereafter proved that if a varies, but no critical points of f are met, the topological properties do not change, which is formalized through the ensuing theorem:

Theorem 1.4 *Let f be a smooth real-valued function on a smooth manifold $\mathscr{M}$. Let $a < b$, with $a, b \in \mathscr{R}$. Suppose that the set $f^{-1}[a, b]$ is compact and contains no critical points of f. Then $\mathscr{M}^a = f^{-1}(-\infty, a]$ is diffeomorphic to $\mathscr{M}^b = f^{-1}(-\infty, b]$. Furthermore, $\mathscr{M}^a$ is a deformation retract of $\mathscr{M}^b$*

An important tool for proving Theorem 1.4 is the *1-parameter group of diffeomorphisms*, defined as follows:

Definition 1.25 *A* **1-parameter group of diffeomorphisms** *of a smooth manifold $\mathscr{M}$ is a C^∞ function $\varphi : \mathscr{M} \times \mathscr{R} \to \mathscr{M}$ such that:*

- *for each $t \in \mathscr{R}$ the map $\varphi_t : \mathscr{M} \to \mathscr{M}$ defined by $\varphi_t(Q) = \varphi(t, Q)$ is a diffeomorphism of $\mathscr{M}$ onto itself.*
- *for all $t, s \in \mathscr{R}$, $\varphi_{t+s} = \varphi_t \circ \varphi_s$*

A 1-parameter group of diffeomorphisms can be *generated* by means of a vector field on $\mathscr{M}$. Let $\mathbf{x}(Q)$ be a vector field on $\mathscr{M}$, i.e. a smooth function which yields a vector in the tangent space of $\mathscr{M}$ at Q for every point $Q \in \mathscr{M}$. Suppose also that the vector field $\mathbf{x}$ vanishes outside of a compact subset $\mathscr{K} \subset \mathscr{M}$. Then, for any Q, define the function $\varphi(t, Q)$, as the solution curve $p : \mathscr{R} \to \mathscr{M}$, of the differential equation:

$$\mathrm{d}p(t)/\mathrm{d}t = \mathbf{x}(p(t)), \tag{1.5}$$

under the initial condition $p(0) = Q$.

It can be proved that, for every point of $\mathscr{M}$, there exists a neighborhood $\mathscr{U}$ and a real number ϵ such that the solution of Equation (1.5) exists and is unique for any $Q \in \mathscr{U}$ and $|t| < \epsilon$. Furthermore, the function $\varphi(t, Q)$ just defined is smooth both in t and in Q.[4]

Consider now the function $r(t) = p(s + t)$, where s is a real constant. It is easy to see that $r(t)$ is the solution of the ensuing Cauchy problem:

$$\begin{cases} \mathrm{d}r(t)/\mathrm{d}t = \mathrm{d}p(t+s)/\mathrm{d}t = \mathbf{x}(p(t+s)) = \mathbf{x}(r(t)) \\ r(0) = p(s) \end{cases} \tag{1.6}$$

[4]See [7], chapter IX.

Therefore, according to Equation (1.6) and Equation (1.5), $r(t) = \varphi(t, p(s))$, provided that $p(s) \in \mathscr{U}$ and $|t + s| < \epsilon$. This proves that $\varphi(t, \varphi(s, Q)) = \varphi(t, p(s)) = r(t) = p(t + s) = \varphi(t + s, Q)$, which satisfies the second requirement of Definition 1.25.

The only thing that remains to do, is to extend the function $\varphi(t, Q)$ outside of a local neighborhood $\mathscr{U}$. The set containing all the neighborhoods $\mathscr{U}_P$, with $P \in \mathscr{K}$, where the solution exists and is unique, is a cover of $\mathscr{K}$. Since $\mathscr{K}$ is compact, there exists a finite number of neighborhoods covering the whole $\mathscr{K}$, and each of such neighborhoods has its own value ϵ, which limits the variation range of t. Denote with ϵ_0 the smallest of such numbers. For any value of t there exist a natural number m and a real remainder $w < \epsilon_0/2$, such that $t = m\epsilon_0/2 + w$. Therefore, the function $\varphi(t, Q)$ can be defined as:

$$\varphi_t(Q) = \underbrace{\varphi_{\epsilon_0/2} \circ \varphi_{\epsilon_0/2} \circ \ldots \circ \varphi_{\epsilon_0/2}}_{m \text{ times}} \circ \varphi_w(Q) \tag{1.7}$$

Equation (1.7) is thus the needed extension of $\varphi(t, Q)$ for any real number, and for any point of the compact set $\mathscr{K}$. The function $\varphi(t, Q)$ generated by the vector fiel $\mathbf{x}$ is unique. Outside the compact set $\mathscr{K}$, the function $\varphi(t, Q)$ always yields Q for any t.

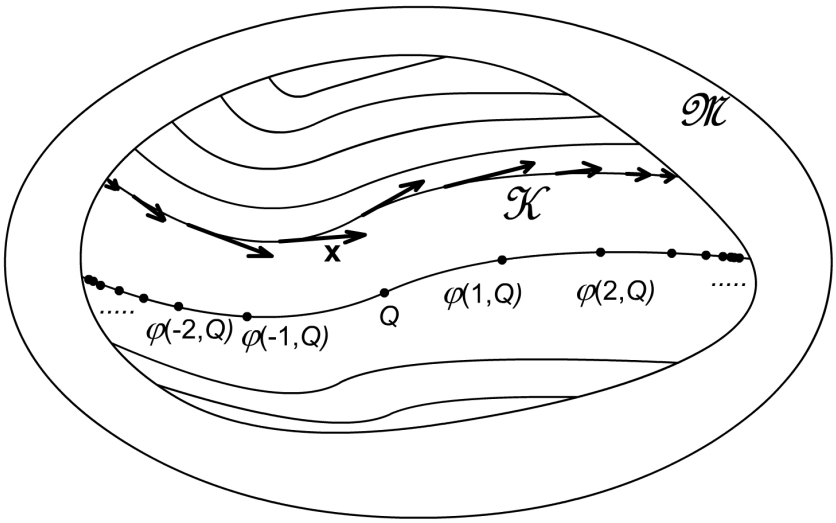


Figure 1.14: Generation of a 1-parameter group of diffeomorphisms.

The function $\varphi(t, Q)$ generated in this way from the vector field $\mathbf{x}$ is therefore a 1-parameter group of diffeomorphisms. The group of diffeomorphisms

$\varphi(t, Q)$ can be viewed as a bundle of curves on $\mathscr{M}$ (see Figure 1.14). All the curves of the bundle are smooth, and tangent to the direction of the vector field at each of their points. Each point in $\mathscr{M}$ is the initial point for one of the curves, therefore the curves never intersect, because at each point there must exist only one solution to Equation (1.5). The t^{th} diffeomorphisms φ_t "pushes" each point of $\mathscr{M}$ along the afore mentioned curves, following the vector field $\mathbf{x}$, and the larger is t, the further along the curve the point is delivered. The points where the vector field vanishes remain fixed under any diffeomorphism of the group, therefore, the closer the points are to the region where $\mathbf{x} = 0$, the shorter their displacements are.

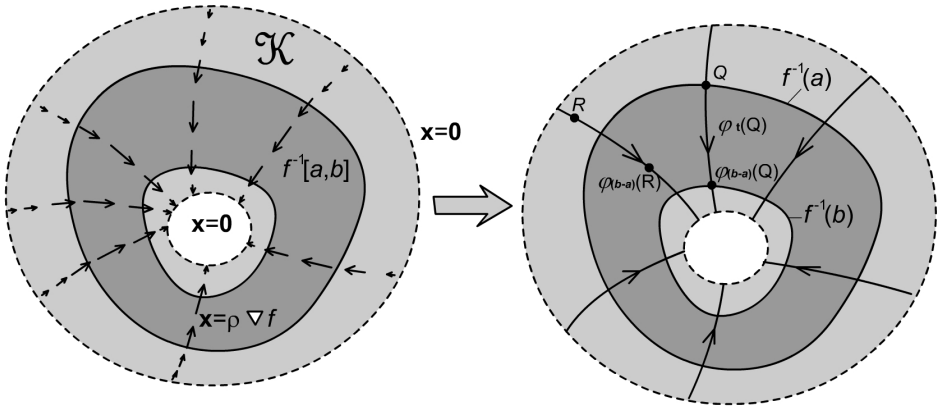


Figure 1.15: Generation of a 1-parameter group of diffeomorphisms through the gradient of f.

In order to prove Theorem 1.4, a 1-parameter group of diffeomorphisms will be generated, by means of the vector field defined by the gradient of f. Let the function $\rho : \mathscr{M} \to \mathscr{R}$ be a smooth function, which is equal to the scalar product of the gradient of f with itself: $\rho = \nabla f \cdot \nabla f$ inside the set $f^{-1}[a, b]$, and which vanishes outside of a compact set $\mathscr{K} \supset f^{-1}[a, b]$. Then let the vector field $\mathbf{x}$ be defined as

$$\mathbf{x}(Q) = \rho(Q)\, \nabla f|_Q \tag{1.8}$$

The vector field $\mathbf{x}$ generates a unique 1-parameter group of diffeomorphisms φ_t, as shown in Figure 1.15. Consider now, for a given point Q, the curve composed of the points $\varphi_t(Q)$, while t varies. By applying the derivation chain rule,the derivative of f with respect to t along this curve can be calculated:

$$\frac{\mathrm{d}f(\varphi_t(Q))}{\mathrm{d}t} = \frac{\mathrm{d}\varphi_t(Q)}{\mathrm{d}t} \cdot \nabla f|_Q = \mathbf{x}(Q) \cdot \nabla f|_Q = 1 \tag{1.9}$$

Thus the value of f increases linearly along with the variable t, upon any of such curves, as long as the points of the curves remain within the set $f^{-1}[a,b]$. For example, if $f(Q)=a$, then $f(\varphi_t(Q))=a+t$. Thus, as shown in Figure 1.15, the diffeomorphism of the group φ_{b-a} carries any point of the boundary $f^{-1}(a)$ of $\mathscr{M}^a$ to a point of the boundary $f^{-1}(b)$ of $\mathscr{M}^b$. Furthermore, any point inside $\mathscr{M}^a$ is carried inside $\mathscr{M}^b$, therefore φ_{b-a} is a diffeomorphism between $\mathscr{M}^a$ and $\mathscr{M}^b$, which proves the first half of Theorem 1.4.

For the second half, consider the retract $r:\mathscr{M}^b\to\mathscr{M}^a$:

$$r(Q)=\begin{cases} Q & \text{if } Q\in\mathscr{M}^a \\ \varphi_{(a-f(Q))}(Q) & \text{if } Q\in f^{-1}[a,b] \end{cases} \tag{1.10}$$

The retract r brings all the points on the curves $\varphi_t(Q)$ in $f^{-1}[a,b]$ to the point of intersection between the curve and the level set $f^{-1}(a)$. Then consider the homotopy function $h:[0,1]\times\mathscr{M}^b\to\mathscr{M}^b$:

$$h(t,Q)=\begin{cases} Q & \text{if } Q\in\mathscr{M}^a \\ \varphi_{t(a-f(Q))}(Q) & \text{if } Q\in f^{-1}[a,b] \end{cases} \tag{1.11}$$

It is easily verified that $h(0,Q)$ is the identity on $\mathscr{M}^b$ and $h(1,Q)$ is the retract r composed with the inclusion, which proves that $\mathscr{M}^a$ is a deformation retract of $\mathscr{M}^b$.

Theorem 1.4 proves that if the set $\mathscr{M}^a$ does not include new critical points, while a varies, then it always remains of the same diffeomorphism type, therefore its topological properties do not change.

1.3.4 Effect of critical points

Before analyzing in detail the effect of critical points on the topological properties of $\mathscr{M}^a$, it is necessary to prove two lemmas, which outline the shape of the function f in the neighborhood of a critical point.

Lemma 1.5 *Let f be a smooth real-valued function in a convex neighborhood $\mathscr{U}$ of $\mathbf{0}$ in $\mathscr{R}^n$, with $f(\mathbf{0})=0$. Then*

$$f(x_1,\ldots,x_n)=\sum_{i=1}^{n}x_i g_i(x_1,\ldots,x_n)$$

for some suitable smooth functions g_i defined in $\mathscr{U}$, with $g_i(\mathbf{0})=\partial f/\partial x_i|_{\mathbf{0}}$.

Lemma 1.5 can be easily proved, for

$$f(x_1,\ldots,x_n)=\int_0^1\frac{\mathrm{d}f(tx_1,\ldots,tx_n)}{\mathrm{d}t}\mathrm{d}t=\sum_{i=1}^{n}x_i\int_0^1\left.\frac{\partial f}{\partial x_i}\right|_{(tx_1,\ldots,tx_n)}\mathrm{d}t \tag{1.12}$$

Thus, the unknown functions g_i are defined as:

$$g_i = \int_0^1 \left.\frac{\partial f}{\partial x_i}\right|_{(tx_1,\ldots,tx_n)} \mathrm{d}t \tag{1.13}$$

Lemma 1.6 (lemma of Morse) *Let f be a smooth real-valued function on a smooth manifold. Let P be a nondegenerate critical point for of f. Then there is a local coordinate system $(y_1, \ldots, y_n)$ in a neighborhood $\mathscr{U}$ of P with $y_i(P) = 0$ for any i and such that the identity:*

$$f = f(P) - y_1^2 - \ldots - y_k^2 + y_{k+1}^2 + \ldots + y_n^2$$

holds throughout $\mathscr{U}$, where k is the index of the critical point P.

It is immediately clear that if such an expression for f exists, then k is the index of P. Indeed, the Hessian matrix written in the coordinate system $(y_1, \ldots, y_n)$ is a diagonal matrix, with k negative elements on the diagonal.

It will be hereafter shown that such a suitable coordinate system always exists. First of all we choose a coordinate system $(x_1, \ldots, x_n)$ such that P is mapped into the origin of $\mathscr{R}^n$. By applying Lemma 1.5 to the function $f(x_1, \ldots, x_n) - f(\mathbf{0})$, one obtains:

$$f(x_1, \ldots, x_n) - f(\mathbf{0}) = \sum_{i=1}^{n} x_i g_i(x_1, \ldots, x_n) \tag{1.14}$$

which holds in some neighborhood of P. Since P is a critical point,

$$g_i(\mathbf{0}) = \left.\frac{\partial f}{\partial x_i}\right|_P = 0 \qquad \forall i \tag{1.15}$$

then Lemma 1.5 can be applied to the functions g_i, too:

$$g_i(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j a_{ij}(x_1, \ldots, x_n) \tag{1.16}$$

Thus, by inserting Equation (1.16) into Equation (1.14), one obtains:

$$f(x_1, \ldots, x_n) - f(\mathbf{0}) = \sum_{i,j=1}^{n} x_i x_j a_{ij}(x_1, \ldots, x_n) \tag{1.17}$$

By defining $b_{ij} = (1/2)(a_{ij} + a_{ji})$, the previous equation assumes the ensuing symmetric form:

$$f(x_1, \ldots, x_n) - f(\mathbf{0}) = \sum_{i,j=1}^{n} x_i x_j b_{ij}(x_1, \ldots, x_n) \tag{1.18}$$

where $b_{ij} = b_{ji}$. Let the matrix $\mathbf{B}$ be the matrix containing the elements b_{ij}. Then $\mathbf{B}(\mathbf{0})$ is equal to half the Hessian matrix of f at the point P, and, being P a generic critical point by hypothesis, $\mathbf{B}(\mathbf{0})$ must be nonsingular.

Equation (1.18) can be diagonalized in the same way quadratic forms are usually diagonalized. It is first supposed by induction that there exists a local coordinate system $(u_1, \ldots, u_n)$ in a neighborhood $\mathscr{U}_1$ of $\mathbf{0}$ such that:

$$f(u_1, \ldots, u_n) - f(\mathbf{0}) = \pm u_1^2 \pm \ldots \pm u_{r-1}^2 + \sum_{i,j \geq r} u_i u_j c_{ij}(u_1, \ldots, u_n) \quad (1.19)$$

The matrix $\mathbf{C}$, containing the elements c_{ij} of Equation (1.19) for $i, j \geq r$, and whose other elements are equal to 0 outside the diaglonal and to ± 1 upon the diagonal, is also supposed by induction to be symmetric. $\mathbf{C}(\mathbf{0})$ is again half the Hessian matrix of f at P, in the coordinate system $(u_1, \ldots, u_n)$, therefore $\mathbf{C}(\mathbf{0})$ is nonsingular. Furthermore, it can be assumed that $\mathbf{C}(\mathbf{0})$ is diagonal[5], thus element $c_{rr}(\mathbf{0})$ cannot vanish. This immediately implies that there exists a neighborhood $\mathscr{U}_2$ of $\mathbf{0}$, possibly smaller than $\mathscr{U}_1$, where the function c_{rr} does not vanish.

Equation (1.19) can be rewritten as:

$$\begin{aligned} f(u_1, \ldots, u_n) - f(\mathbf{0}) &= \pm u_1^2 \pm \ldots \pm u_{r-1}^2 + 2 \sum_{i \geq r} u_i u_r c_{ir}(u_1, \ldots, u_n) + \\ &\quad + \sum_{i,j > r} u_i u_j c_{ij}(u_1, \ldots, u_n) \end{aligned} \quad (1.20)$$

Consider now the function $\mathbf{v} : \mathscr{R}^n \to \mathscr{R}^n$ such that:

$$\begin{cases} v_i = u_i & \text{if } i \neq r \\ v_r(u_1, \ldots, u_n) = \sqrt{2\,|c_{rr}(u_1, \ldots, u_n)|}\left[u_r + \sum_{i>r} u_i \frac{c_{ir}(u_1, \ldots, u_n)}{2c_{rr}(u_1, \ldots, u_n)}\right] \end{cases} \quad (1.21)$$

It can be easily verified that the Jacobian matrix of $\mathbf{v}$ is nonsingular at $\mathbf{0}$, therefore, in a neighborhood $\mathscr{U}_3$ of $\mathbf{0}$, possibly smaller than $\mathscr{U}_2$, the function $\mathbf{v}$ can be inverted, and the inverse function is continuous, by virtue of the inverse function theorem. Thus $\mathbf{v}$ is a local coordinate system on $\mathscr{U}_3$. Moreover, it is easily verified that:

$$2 \sum_{i \geq r} u_i u_r c_{ir}(u_1, \ldots, u_n) = v_r^2 - \sum_{i,j > r} v_i v_j \frac{c_{ir}(\mathbf{u}(\mathbf{v}))c_{jr}(\mathbf{u}(\mathbf{v}))}{2c_{rr}(\mathbf{u}(\mathbf{v}))} \quad (1.22)$$

[5]If $\mathbf{C}(\mathbf{0})$ were not diagonal, a simple linear change of coordinates would easily diagonalize it.

By inserting Equation (1.22) and Equation (1.21) into Equation (1.20) one obtains:

$$f(u_1,\ldots,u_n) - f(\mathbf{0}) = \pm v_1^2 \pm \ldots \pm v_r^2 + \sum_{i,j\geq r+1} v_i v_j d_{ij}(v_1,\ldots,v_n) \quad (1.23)$$

Equation (1.18) can be considered as the starting point of the induction, which, after n steps, leads to the proof of Lemma 1.6.

A direct consequence of Lemma 1.6 is the following:

Theorem 1.7 *Nondegenerate critical points are always isolated.*

Lemma 1.6 is very important, because it not only states that any function can be approximated by a quadratic form, in the neighborhood of a critical point for a given coordinate system, but also that f *is* a quadratic form, within a sufficiently small neighborhood and in a proper coordinate system. This is essential for proving the ensuing fundamental result:

Theorem 1.8 *Let $f : \mathscr{M} \to \mathscr{R}$ be a smooth function, and let P be a nondegenerate critical point of index k. Setting $f(P) = c$, suppose that $f^{-1}[c-\epsilon, c+\epsilon]$ is compact, and contains no critical points of f other than P, for some $\epsilon > 0$. Then, for all sufficiently small ϵ, the set $\mathscr{M}^{c+\epsilon}$ has the homotopy type of $\mathscr{M}^{c-\epsilon}$ with a k-cell attached.*

Indeed, according to Lemma 1.6, it is possible to choose a coordinate system $(u_1,\ldots,u_n)$ such that the critical point P coincides with the origin, and in a neighborhood $\mathscr{U}$ of P the ensuing identity holds:

$$f = c - u_1^2 - \ldots - u_k^2 + u_{k+1}^2 + \ldots + u_n^2 \quad (1.24)$$

A real ϵ is chosen, so that:

- the region $f^{-1}[c-\epsilon, c+\epsilon]$ is compact and contains no critical points other than P;
- the closed ball $u_1^2 + \ldots + u_n^2 \leq 2\epsilon$ is contained in the neighborhood $\mathscr{U}$.

We consider also the set e^k, defined as follows:

$$e^k = \left\{ P \in \mathscr{U} : u_1^2 + \ldots + u_k^2 \leq \epsilon \text{ and } u_{k+1} = \ldots = u_n = 0 \right\} \quad (1.25)$$

Clearly, e^k is homeomorphic to the k-cell, therefore we will say henceforth that e^k *is* the k-cell.

The resulting situation is represented in two dimensions in Figure 1.16:

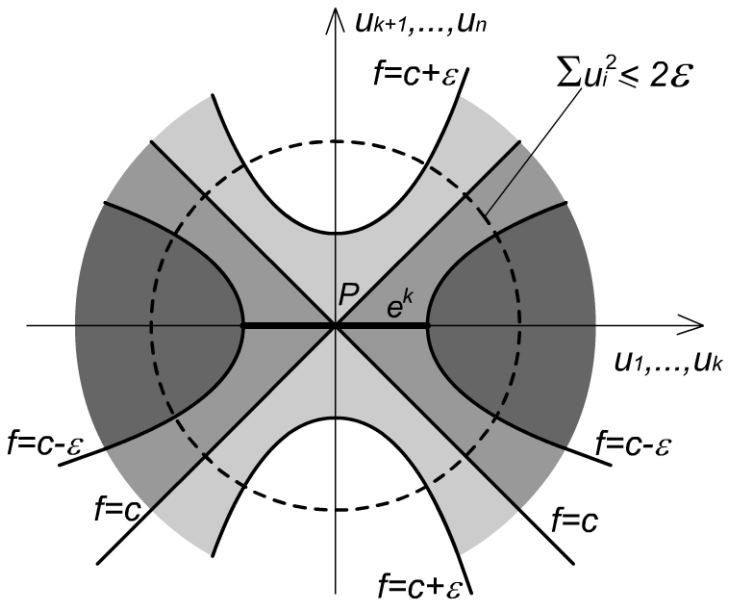


Figure 1.16: Proof of Theorem 1.8.

- the coordinate lines represent the planes $u^{k+1} = 0, \ldots, u^n = 0$ and $u^1 = 0, \ldots, u^k = 0$;
- the dashed circle is the boundary of the ball of radius $\sqrt{2\epsilon}$;
- the hyperbolas represent the hypersurfaces $f^{-1}(c-\epsilon)$ and $f^{-1}(c+\epsilon)$;
- the region $\mathscr{M}^{c-\epsilon}$ is painted with dark gray;
- the region $f^{-1}[c-\epsilon, c]$ is painted with medium gray;
- the region $f^{-1}[c, c+\epsilon]$ is painted with light gray;
- the horizontal dark line through P is the k-cell.

Note that the boundary of the k-cell is on the set $\mathscr{M}^{c-\epsilon}$, therefore the k-cell is indeed attached to the set $\mathscr{M}^{c-\epsilon}$.

It will be hereafter proved that $\mathscr{M}^{c-\epsilon} \cup e^k$ is a deformation retract of $\mathscr{M}^{c+\epsilon}$. To do so, a new function F will be constructed, which is equal to f outside the ball of radius $\sqrt{2\epsilon}$, but lesser than f inside the ball. Let $\mu : \mathscr{R} \to \mathscr{R}$ be a smooth function, satisfying the ensuing conditions:

- $\mu(0) > \epsilon$,

- $\mu(r) = 0$ for $r \geq 2\epsilon$,
- $-1 < \mu'(r) \leq 0$ for all r,

where $\mu' = \mathrm{d}\mu/\mathrm{d}r$. Now let the function $F : \mathscr{M} \to \mathscr{R}$ coincide with f outside the neighborhood $\mathscr{U}$, and be defined inside $\mathscr{U}$ by the ensuing equation:

$$F = f - \mu(\xi(\mathbf{u}) + 2\eta(\mathbf{u})) \tag{1.26}$$

where ξ and η are the following two functions:

$$\begin{aligned} \xi(\mathbf{u}) &= u_1^2 + \ldots + u_k^2 \\ \eta(\mathbf{u}) &= u_{k+1}^2 + \ldots + u_n^2 \end{aligned} \tag{1.27}$$

Thus f and F can be written as

$$\begin{aligned} f(\mathbf{u}) &= c - \xi(\mathbf{u}) + \eta(\mathbf{u}) \\ F((\mathbf{u})) &= -\xi(\mathbf{u}) + \eta(\mathbf{u}) - \mu(\xi(\mathbf{u}) + 2\eta(\mathbf{u}) \end{aligned} \tag{1.28}$$

It is easy to verify that $F = f$ outside the ellipsoid $\xi + 2\nu \leq 2\epsilon$, whereas inside such ellipsoid the ensuing inequality holds:

$$F \leq f = c - \xi + \eta \leq c + (1/2)\xi + \eta \leq c + \epsilon \tag{1.29}$$

which proves that the region $F^{-1}(-\infty, c+\epsilon]$ coincides with $f^{-1}(-\infty, c+\epsilon]$, which is $\mathscr{M}^{c+\epsilon}$.

Furthermore, the critical points of F are the same as those of f. Indeed, the critical points of F inside $\mathscr{U}$ do satisfy the ensuing condition:

$$\frac{\partial F}{\partial u_i} = \frac{\partial F}{\partial \xi}\frac{\partial \xi}{\partial u_i} + \frac{\partial F}{\partial \eta}\frac{\partial \eta}{\partial u_i} = 0 \qquad i = 1, \ldots, n \tag{1.30}$$

However the partial derivatives of F with respect to ξ and η never vanish, because:

$$\begin{aligned} \partial F/\partial \xi &= -1 - \mu'(\xi + 2\eta) < 0 \\ \partial F/\partial \eta &= 1 - 2\mu'(\xi + 2\eta) \geq 1 \end{aligned} \tag{1.31}$$

Since $\partial\xi/\partial u_i$ and $\partial\eta/\partial u_i$ vanish only at the origin, the only critical point inside $\mathscr{U}$ is P. Outside $\mathscr{U}$ F and f do coincide, therefore the two functions have the same critical points over the whole $\mathscr{M}$.

Consider now the region $F^{-1}[c-\epsilon, c+\epsilon]$. Since it has just been shown that $F^{-1}(-\infty, c+\epsilon]$ coincides with $f^{-1}(-\infty, c+\epsilon]$, and since $F \leq f$, it follows that

$$F^{-1}[c-\epsilon, c+\epsilon] \subset f^{-1}[c-\epsilon, c+\epsilon] \tag{1.32}$$

Therefore $F^{-1}[c-\epsilon, c+\epsilon]$ is compact, and it cannot contain any critical points, except possibly P. However, since $F(P) = c - \mu(\mathbf{0}) < c - \epsilon$, P is not contained in $F^{-1}[c-\epsilon, c+\epsilon]$, thus such region is free of critical points.

Therefore, according to Theorem 1.4, the region $F^{-1}(-\infty, c-\epsilon]$ is a deformation retract of $F^{-1}(-\infty, c+\epsilon]$, which coincides with $\mathscr{M}^{c+\epsilon}$.

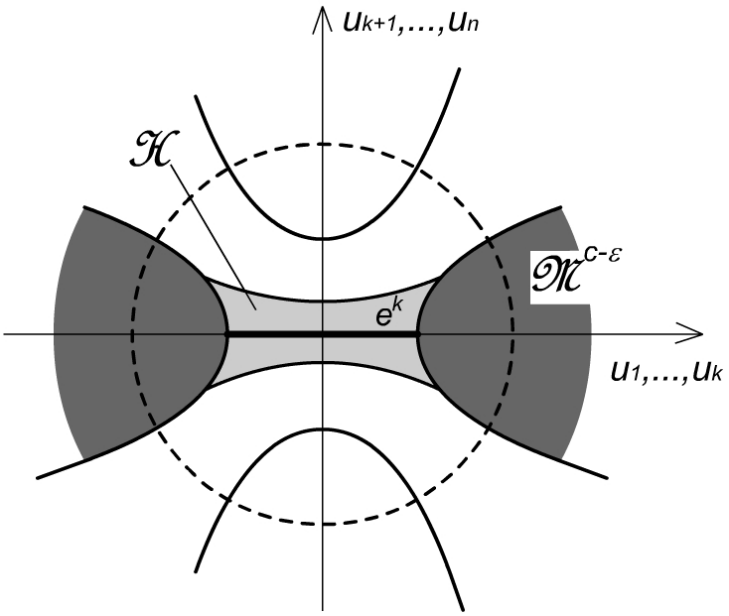


Figure 1.17: The handle $\mathscr{H}$.

The only thing that remains to be proved is that $\mathscr{M}^{c-\epsilon} \cup e^k$ is a deformation retract of $F^{-1}(-\infty, c-\epsilon]$. Consider Figure 1.17. The set $F^{-1}(-\infty, c-\epsilon]$ can be viewed as the union between $\mathscr{M}^{c-\epsilon}$ and a "handle" $\mathscr{H}$, where $\mathscr{H}$ denotes the closure of $F^{-1}(-\infty, c-\epsilon] - \mathscr{M}^{c-\epsilon}$. The handle $\mathscr{H}$ represents the additional region, in the neighborhood of P, which is included into $F^{-1}(-\infty, c-\epsilon]$, for $F \leq f$ in the neighborhood $\mathscr{U}$.

Let now the retract $\mathbf{r}$ from $\mathscr{M}^{c-\epsilon} \cup \mathscr{H}$ to $\mathscr{M}^{c-\epsilon} \cup e^k$ be defined as depicted in Figure 1.18, according to three different cases.

- CASE 1: within the region $\xi \leq \epsilon$, $\mathbf{r}$ is defined as:

$$\mathbf{r}(u_1, \ldots, u_n) = (u_1, \ldots, u_k, 0, \ldots, 0) \tag{1.33}$$

  therefore all points are delivered into the k-cell.

- CASE 2: within the region $\epsilon \leq \xi \leq \eta + \epsilon$, $\mathbf{r}$ is defined as:

$$\mathbf{r}(u_1, \ldots, u_n) = (u_1, \ldots, u_k, su_{k+1}, \ldots, su_n) \tag{1.34}$$

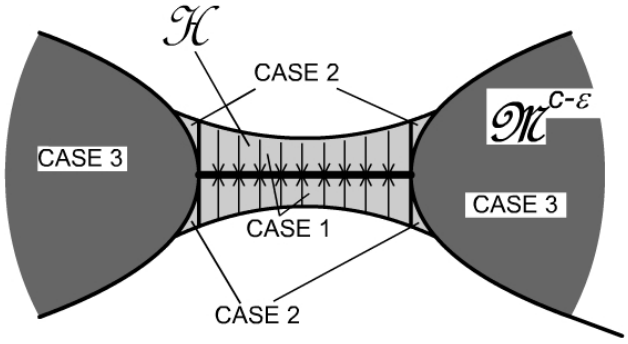


Figure 1.18: Retract from $\mathscr{M}^{c-\epsilon} \cup \mathscr{H}$ to $\mathscr{M}^{c-\epsilon} \cup e^k$.

where

$$s = [(\xi - \epsilon)/\eta]^{1/2} \tag{1.35}$$

therefore all points are delivered onto the hypersurface $f^{-1}(c - \epsilon)$.

- CASE 3 within the region $\eta + \epsilon \leq \xi$, i.e., within $\mathscr{M}^{c-\epsilon}$, $\mathbf{r}$ is defined as the identity.

It can be easily verified that the afore defined retract is continuous.

Consider now the ensuing homotopy function $\mathbf{h} : [0, 1] \times \mathscr{M}^{c-\epsilon} \cup \mathscr{H} \to \mathscr{M}^{c-\epsilon} \cup \mathscr{H}$:

- CASE 1: within the region $\xi \leq \epsilon$, $\mathbf{r}$ is defined as:

$$\mathbf{h}(t, u_1, \ldots, u_n) = (u_1, \ldots, u_k, tu_{k+1}, \ldots, tu_n) \tag{1.36}$$

which, for $t = 1$ is the identity, whereas for $t = 0$ coincides with the retract composed with the inclusion.

- CASE 2: within the region $\epsilon \leq \xi \leq \eta + \epsilon$, $\mathbf{r}$ is defined as:

$$\mathbf{h}(t, u_1, \ldots, u_n) = ((u_1, \ldots, u_k, s_t u_{k+1}, \ldots, s_t u_n) \tag{1.37}$$

where

$$s_t = t + (1 - t)\,[(\xi - \epsilon)/\eta]^{1/2} \tag{1.38}$$

which, for $t = 1$ is the identity, whereas for $t = 0$ coincides with the retract composed with the inclusion.

- CASE 3 within the region $\eta + \epsilon \leq \xi$, i.e., within $\mathscr{M}^{c-\epsilon}$, $\mathbf{h}$ is defined as the identity for any t

It can be easily verified that this homotopy function is continuous, therefore the retract composed the inclusion is homotopic to the identity, and $\mathscr{M}^{c-\epsilon} \cup e^k$ is a deformation retract of $\mathscr{M}^{c-\epsilon} \cup \mathscr{H}$, which proves Theorem 1.8.

By remembering that nondegenerate critical points are always isolated (Theorem 1.7), it is immediately possible to generalize Theorem 1.8 to the case $f^{-1}(c)$ contains many critical points: a k-cell is attached for each critical point of index k (see [5]).

Eventually, if the set $\mathscr{M}_+^a$ is defined as $f^{-1}[a, +\infty)$, then the ensuing two analogous theorems can be easily proved, by considering the function $-f$ in place of f:

Theorem 1.9 *Let f be a smooth real-valued function on a smooth manifold $\mathscr{M}$. Let $a < b$, with $a, b \in \mathscr{R}$. Suppose that the set $f^{-1}[a, b]$ is compact and contains no critical points of f. Then $\mathscr{M}_+^a$ is diffeomorphic to $\mathscr{M}_+^b$. Furthermore, $\mathscr{M}_+^b$ is a deformation retract of $\mathscr{M}_+^a$*

Theorem 1.10 *Let $f : \mathscr{M} \to \mathscr{R}$ be a smooth function, and let P be a nondegenerate critical point of index k. Setting $f(P) = c$, suppose that $f^{-1}[c - \epsilon, c + \epsilon]$ is compact, and contains no critical points of f other than P, for some $\epsilon > 0$. Then, for all sufficiently small ϵ, the set $\mathscr{M}_+^{c-\epsilon}$ has the homotopy type of $\mathscr{M}_+^{c+\epsilon}$ with a $(n-k)$-cell attached, where n is the dimension of $\mathscr{M}$.*

1.3.5 Morse inequalities

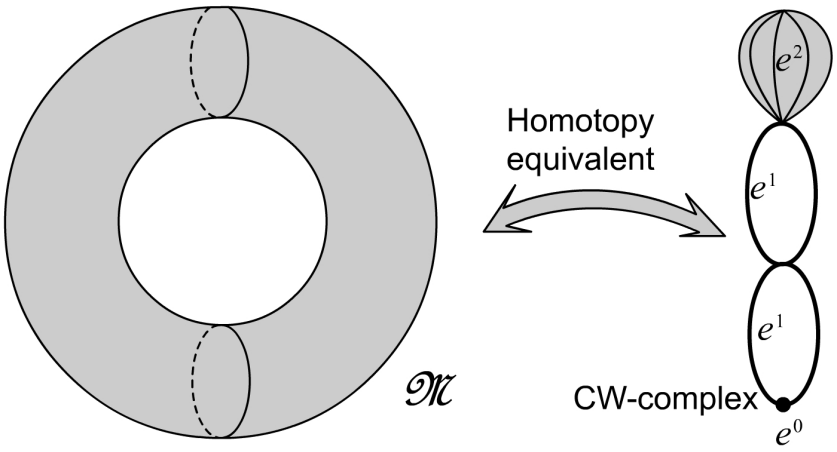


Figure 1.19: The torus is homotopy equivalent to a CW-complex.

The results of the previous sections make it possible to build a "skeleton", by attaching cells upon each other, which is homotopy equivalent to any compact manifold $\mathscr{M}$[6]. Such a skeleton is usually named *CW-complex*. For example, the height function defined on the two dimensional torus in Section 1.3.1 proves that the torus is homotopy equivalent to a CW-complex composed by one 0-cell, attached by the minimum P, two 1-cells, attached by the saddles Q and R, and one 2-cell, attached by the maximum S, as shown in Figure 1.19.

Clearly, there is a minimum of k-cells that must be used to build a CW-complex homotopy equivalent to a given manifold. For example, the torus of Figure 1.19 cannot be build using only 0-cells, otherwise it would be composed by a certain number of disjoint regions, and each of such regions would be contractible to a point. It can be proved that, if B_k is the k^{th} Betti number of a manifold, and c_k is the number of k-cells used to build a CW-complex homotopic to the same manifold, then the ensuing inequality holds:

$$B_k \leq c_k \tag{1.39}$$

which immediately leads to an analogous constraint on the number of critical points of index k of any smooth real-valued function defined on such manifold[7] .

For example, if a manifold is composed of a n disjoint regions, i.e. the 0^{th} Betti number is equal to n, then any homotopy equivalent CW-complex must be built by using at least n 0-cells. Furthermore, any function defined on such manifold must have at least n minima and n maxima, which could be also easily proved by means of Weierstraß theorem applied to each disjoint region. The torus depicted in Figure 1.19 can be reduced to a CW-complex containing at least one 0-cell, two 1-cells, and one 2-cell (see Table 1.1 for the Betti numbers of the torus), therefore the height function defined in Section 1.3.1 has the least possible number of critical point for each index.

This result will be used in the following sections, but the proof will be omitted. For further details, refer to [5], chapter 5.

[6]See Theorem 3.5 at page 20 of [5].

[7]see Theorem 5.2 at page 29 of [5].

Chapter 2

Analysis of Singularity Loci

In this chapter, the concepts of serial and parallel singularities will be recalled. By means of the results of Chapter 1, a numerical method will be proposed, capable of finding out a singularity-free path between any two configurations of a manipulators, if there exist any, and to count the number of maximal singularity-free regions.

2.1 Singularities

2.1.1 Serial Singularities

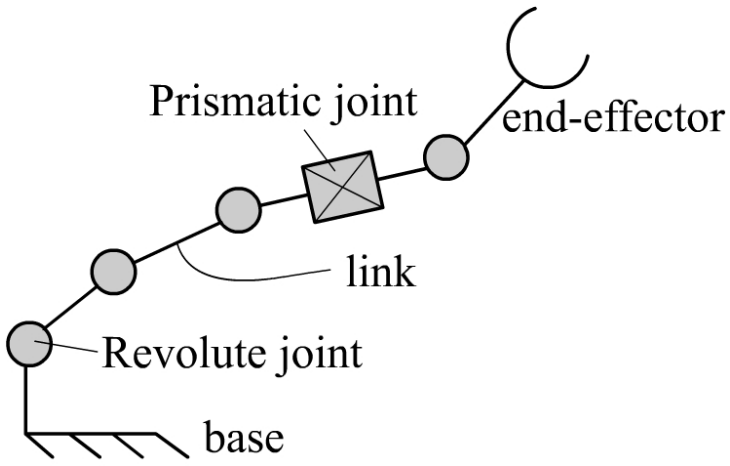


Figure 2.1: Serial manipulators.

Serial manipulators (see Figure 2.1) are open serial kinematic chains,

where each joint is actuated. The *links* are the rigid bodies connecting the joints. Each link of a serial manipulator connects at most two joints, and each joint connects at most two links. Usually, the joints are revolute or prismatic kinematic pairs. The first link of the chain is fixed in most applications, and therefore it will be named *base*.

Serial manipulators are the most widely used in industry, for many different tasks, ranging from pick-an-place applications to machining and welding. Their main features are their flexibility, and their easily controllable kinematics.

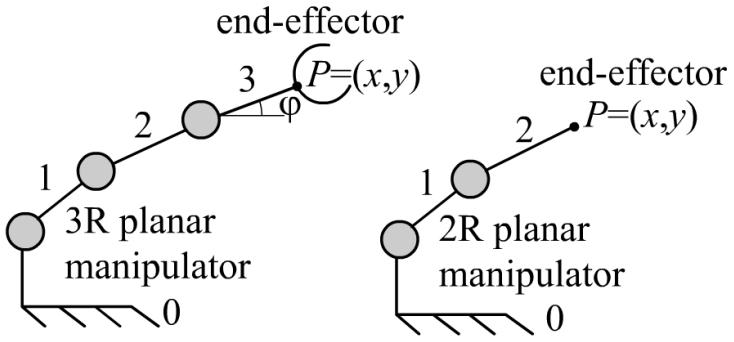


Figure 2.2: End effectors of different planar manipulators.

The *end-effector* of a serial manipulator is defined according to the task that must be fulfilled by the manipulator itself. For example, consider the 3R planar manipulator depicted in Figure 2.2. The task of this manipulator is to position link 3 in the plane, thus link 3 is the end-effector and its position can be identified through the coordinates (x, y) of a point P of the end-effector, and an angle φ representing the orientation of the end-effector in the plane. On the other hand, the task of the 2R planar manipulator shown in Figure 2.2 is to place the point P in the plane, therefore the point P is the end-effector, and its position can be identified through the coordinates (x, y) only. Thus, different end-effectors might be defined for the same manipulator, according to the different tasks that can be accomplished through that manipulator.

If the number of degrees of freedom of the end-effector is equal to the number of joints, then the serial manipulator is *nonredundant*. Only nonredundant manipulators will be considered in this work

The *jointspace* $\mathscr{J}$ of a serial manipulator is a manifold whose points are defined by the positions of all actuators of the manipulator, whereas the *workspace* $\mathscr{W}$ is the manifold containing all possible positions of the end-effector.

The kinematics of serial manipulators can be described through a function f from the jointspace to the workspace, which associates the corresponding position of the end-effector to a set of positions of the actuated joints. If Q is a point of the jointspace, and X is a point of the workspace, this function can be written as:

$$X = f(Q) \tag{2.1}$$

If, at a given point Q of the jointspace, the virtual displacement of the end-effector, $\partial X = \partial f(Q)$, vanishes for a certain nonzero virtual displacement of the point in the jointspace, $\partial(Q)$, then the manipulator is at a *serial singularity*. In other words, the manipulator is crossing a serial singularity if it is possible to impress a nonzero velocity to some actuators producing a corresponding zero velocity of the end-effector.

Singularities can also be defined by means of a given coordinate system. If the variable q_i locally describes the position of the i^{th} actuated joint, then the vector $\mathbf{q} = (q_1, \ldots, q_n)$ is a local coordinate system for the jointspace. Analogously, a vector $\mathbf{x} = (x_i, ..., x_n)$ is supposed to provide a local coordinate system for the workspace. Then, the function f can be locally written as follows:

$$\mathbf{x} = f(\mathbf{q}) \tag{2.2}$$

The relationship between the first order variations of $\mathbf{x}$ and $\mathbf{q}$ then becomes:

$$\partial\mathbf{x} = \mathbf{A}(\mathbf{q})\partial\mathbf{q} \tag{2.3}$$

where the matrix $\mathbf{A}(\mathbf{q})$ is the Jacobian matrix of the function f in the local coordinate system $(q_1, \ldots, q_n)$. If, according to the given definition, $\partial\mathbf{x}$ must vanish for a certain nonzero $\partial\mathbf{q}$, then the matrix $\mathbf{A}(\mathbf{q})$ must be singular, at a serial singularity. Thus, a position of a serial manipulator is a singular position if $J(\mathbf{q}) = \det \mathbf{A}(\mathbf{q}) = 0$. $J(\mathbf{q})$ is the Jacobian determinant in the chosen local coordinate system. If we assume that there exists a set of local coordinate systems such that J is defined as a smooth function on the whole jointspace, serial singularities can be described as the points Q of the jointspace where the real valued function $J(Q)$ vanishes. Therefore serial singularities are organized in a *singularity locus*, or *surface*, of dimension $n - 1$, where n is the dimension of the jointspace, and such locus cuts the jointspace into two or more disconnected pieces. In one or more of such pieces the Jacobian determinant is positive, and in one or more pieces it is negative.

A simple example of serial singularity is shown in Figure 2.3. Serial singularities are harmless for serial manipulators, however, crossing a serial singularity means loosing the capability of moving the end-effector along

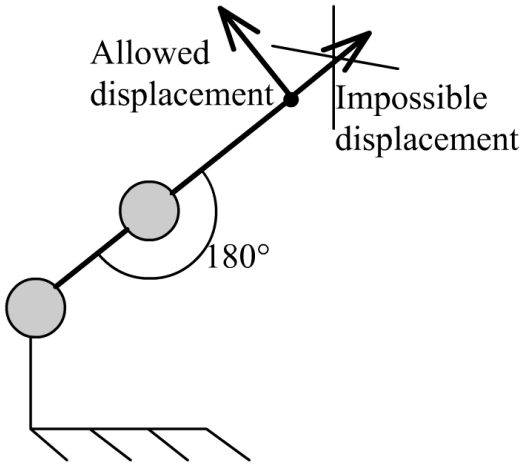


Figure 2.3: End effector of different planar manipulators.

a particular direction, therefore it might be undesirable for some applications. Contrarily to some commonly given definitions, some examples (see [8]) showed that serial singularities do not always occur for positions of the end-effector on the boundary of the workspace. In fact, the image of a serial singularity may be inside the workspace, too, and work as a wall, precluding the possibility of following a given trajectory of the end-effector. Therefore, it could be interesting to know whether a singularity-free path between any two points in the jointspace exists or not, and in the first case to find it out. Clearly, this question is strictly related to the number of disjoint regions into which the jointspace is cut by the singularity locus.

Furthermore, it has long been believed (see [9]) that a serial singularity must be crossed during any change of posture of a manipulator. Two different postures of a manipulators are two points of the jointspace with the same image through f in the workspace, and a change of posture is a path in the jointspace connecting two different postures of a manipulator. Many researches pointed out that singularity-free posture changes of a serial manipulators are possible, and provided significant examples (see [10], [11], [12]). However, no general methods have been given to determine whether, given any two postures, there exists a singularity-free posture change connecting them or not.

2.1.2 Parallel Singularities

Parallel manipulators are often complex machines, composed of many links, each of which can be connected to any number of other links through a

proper number of joints. The field of application of parallel machines i still much narrower than that of serial manipulators. The main advantages of parallel manipulators is due to their closed-loop structure, which increases stiffness and structural strength. Therefore they seem promising for applications such as light-weight space robots, high-precision machining tools or high-speed pick-and-place machines, for they provide higher stiffness and pay-load despite their lower weight. Unfortunately, the closed-loop structure is also the main drawback of parallel robots, because it increases the complexity of the kinematic model, the possibility of collisions between the links, and the danger of loosing control of the machine.

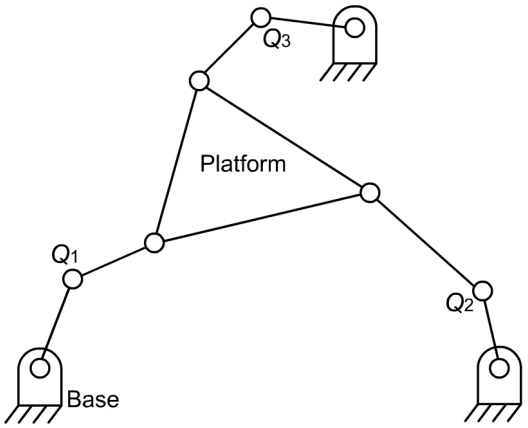


Figure 2.4: A fully-parallel planar manipulator.

Analogously to serial manipulators, the end-effector can be defined also for parallel manipulators, according to the specific task to be performed. The rigid body upon which the end-effector is placed is the *platform*.

The huge variety of parallel manipulators demands a more restricted definition, in order to bound the subject of this work:

Definition 2.1 *A* **fully-parallel manipulator** *is a parallel machine such that the platform is connected to the base through a number of serial kinematic chains (***legs***) equal to the number of degrees of freedom of the manipulator. Furthermore, each serial leg contains one actuated joint only.*

Similar definitions can be found in [13], [14]. The *jointspace* $\mathscr{J}$ and the *workspace* $\mathscr{W}$ of a parallel manipulator can be defined, analogously to serial manipulators, as the manifold containing the positions of all actuators and the manifold containing all possible positions of the end-effector. We define also the *configuration* of the manipulator as an ordered set containing the

positions of all the rigid bodies of the manipulator. Then, the *configuration space* can be defined as the manifold containing all the configurations of the manipulator allowed by the constraints.

An example of fully-parallel planar manipulator is shown in Figure 2.4. The task of the manipulator depicted in Figure 2.4 is to position the platform in the plane, therefore the end-effector and the platform do coincide. The workspace is the manifold containing all possible positions of a rigid body in the plane. For every leg there is an actuated joint Q_i, and the ordered sets containing the positions of the three actuators are all points of the jointspace. A configuration of the manipulator is a snapshot of the positions of all the links: for example, Figure 2.4 is a configuration. The manifold containing all possible snapshots like Figure 2.4, is the configuration space of the 3RRR planar manipulator.

The configuration space of the fully-parallel manipulators that will be analyzed can be described as the level set $g^{-1}(\mathbf{0})$ of a smooth function $g : \mathscr{J} \times \mathscr{W} \to \mathscr{R}$. According to this definition, the configuration space is defined as the points (Q, X) for which the ensuing equation holds:

$$g(Q, X) = \mathbf{0} \tag{2.4}$$

where Q is a point of the jointspace and X is a point of the workspace. Clearly, the definition of the configuration space could be easily extended to serial manipulators, too: in this case, the configuration space is easily described by the jointspace itself.

The points (Q, X) of the configuration space such that there exists a nonzero virtual displacement of the actuators, ∂Q, and a zero virtual displacement of the end-effector, ∂X, which satisfy the constraints, i.e. produce a zero variation of g, are named serial singularities. Indeed, the effect of serial singularities is the same on parallel and serial manipulators, for they cause in both cases a local loss of mobility of the end-effector.

The points (Q, X) of the configuration space such that there exists a nonzero virtual displacement of the end-effector, ∂X, and a zero virtual displacement of the actuators, ∂Q, which satisfy the constraints, i.e. produce a zero variation of g, are named parallel singularities. At a parallel singularity, small displacements of the end-effector along a certain direction are allowed, even though all actuators are kept locked.

The effect of parallel singularities is much more dangerous than that of serial singularities, for they might jeopardize the structural integrity of the machine, or produce the loss of control of the platform. This hazard can be understood by means of the principle of virtual work. If a proper set of forces or torques $\mathbf{w}$ is applied to the end-effector, and a set of forces or toques $\mathbf{t}$ is

applied to the actuated joints, then the manipulator is at equilibrium if and only if the sum of the virtual works produced by $\mathbf{w}$ and $\mathbf{t}$ vanishes for any virtual displacement in the configuration space, i.e.:

$$\mathbf{w} \cdot \partial X + \mathbf{t} \cdot \partial Q = \mathbf{0} \tag{2.5}$$

By the aforementioned definition, at a parallel singularity there exist two virtual displacements ∂X_s and ∂Q_s, of the end effector and of the joints respectively, which are allowed by the constraints and such that $\partial X_s \neq \mathbf{0}$ and $\partial Q_s = \mathbf{0}$. Along this particular displacements, Equation (2.5) becomes

$$\mathbf{w} \cdot \partial \mathbf{x}_s = \mathbf{0} \tag{2.6}$$

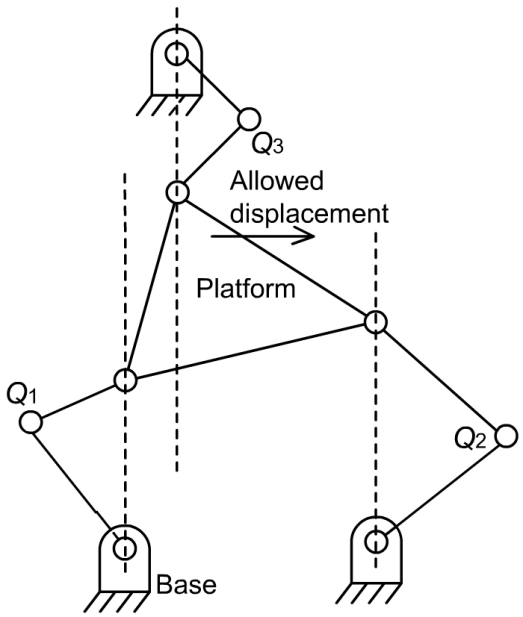


Figure 2.5: A fully-parallel planar manipulator.

If a set of actions $\mathbf{w}$ with a nonzero component along ∂X_s is applied to the end-effector, the equilibrium is impossible and either the control of the platform is lost, or in the worst case some of the components of the machine are damaged. The set of actions $\mathbf{w}$ needs not to be an external load applied to the platform: it might also be, by virtue of D'Alembert's principle, the set of inertial loads acting on the platform while operating along a certain trajectory. It is therefore utterly important to avoid crossing parallel singularities when moving a parallel manipulator from a position to another, lest inertial or external loads might not be withstood by the actuators. An

example of parallel singularity is depicted in Figure 2.5: the platform can undergo infinitesimal horizontal translations even though all three actuators are locked. An external horizontal force applied to the platform cannot be balanced by any set of torques applied by the actuators.

If local coordinate systems $(q_1, \ldots, q_n)$ and $(x_1, \ldots, x_n)$ are defined in the jointspace and in the workspace respectively, a virtual variation of Equation (2.4) can be expressed as follows:

$$\partial g(\mathbf{q}, \mathbf{x}) = \mathbf{A}(\mathbf{q}, \mathbf{x})\partial\mathbf{q} + \mathbf{B}(\mathbf{q}, \mathbf{x})\partial\mathbf{x} \tag{2.7}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the Jacobian matrices of g with respect to the coordinates $(q_1, \ldots, q_n)$ and $(x_1, \ldots, x_n)$ respectively. Then, according to the previous definitions, serial singularities occur for all points $(\mathbf{q}, \mathbf{x})$ of the configuration space such that $\det \mathbf{A} = 0$, and parallel singularities occur for all points $(\mathbf{q}, \mathbf{x})$ of the configuration space such that $J(\mathbf{q}, \mathbf{x}) = \det \mathbf{B} = 0$. This classification is similar to that proposed in [22]. Analogous to serial singularities, if we suppose that the Jacobian determinant J can be define as a smooth real-valued function on the whole configuration space, also parallel singularities are organized in a singularity locus, which is the zero level set J.

Many authors coped with the problem of avoiding parallel singularities. The easiest way to tackle this problem is to reduce the configuration space of the parallel manipulator, so that no parallel singularity lies inside the configuration space. However, this solution is also the most burdensome one, because it decreases the reachable space of the machine. An alternative way to eliminate singularities from the configuration space is to add redundant actuators to the manipulator (see, for example [15] and [16]), but this means to increase the complexity of the machine, which might be undesired.

If none of the two previous solutions is viable and the configuration space contains parallel singularities, there are only two possibilities left. The former is to try to control the manipulator as it meets a singularity, by taking into account the dynamics of the manipulator (as proposed in [17], [18], [19]). The latter is to plan a singularity-free path in the configuration space ([20] and [21]), given the starting and ending point of the path. Unfortunately, to the author knowledge, the path-planning strategies available so far in the literature are all local, which means that if they fail to find a singularity-free path it is not sure that a singularity free path does not exist at all.

Obviously, the problem of singularity-free path-planning is strictly related to the problem of identifying and characterizing the different disjoint regions into which the configuration space is split by the singularity locus. As it has just been pointed out in this Section 2.1, it will be supposed that the singularity locus of a manipulator, either serial or parallel, can be defined

through the zero level set of a smooth real-valued function J. In Section 2.3 a method capable of identifying the disjoint regions cut by the set $J = 0$ in the configuration space of a parallel or serial manipulator will be presented. The following section will first specialize to this extent the results of Chapter 1 to the number of disjoint regions composing a manifold.

2.2 Results about connectedness

Given a smooth real-valued function f on a smooth manifold $\mathscr{M}$, we define, analogous to the previous chapter, the set $\mathscr{M}_-^a = f^{-1}(-\infty, a]$, and $\mathscr{M}_+^a = f^{-1}[a, +\infty)$. Theorem 1.4 and Theorem 1.9 can be specialized as follows:

Theorem 2.1 *If $a < b$ and $f^{-1}[a, b]$ is compact and contains no critical points of f, then $\mathscr{M}_-^a$ is composed of the same number of disjoint regions as $\mathscr{M}_-^b$, and $\mathscr{M}_+^a$ is composed of the same number of disjoint regions as $\mathscr{M}_+^b$*

The proof immediately stems from Theorem 1.4, Theorem 1.9 and Theorem 1.2.

Analogously, by considering Theorem 1.3, Theorem 1.8 and Theorem 1.10 can be focused on the change of the number of disjoint regions produced by critical points:

Theorem 2.2 *Let P be a nondegenerate critical point of index k. Setting $f(P) = c$, suppose that $f^{-1}[c - \epsilon, c + \epsilon]$ is compact, and contains no critical points of f other than P, for some $\epsilon > 0$. Then, for all sufficiently small ϵ, the set $\mathscr{M}_-^{c+\epsilon}$ has the same number of regions as $\mathscr{M}_-^{c-\epsilon}$ with a k-cell attached, and the set $\mathscr{M}_+^{c-\epsilon}$ has the same number of disjoint regions as $\mathscr{M}_+^{c+\epsilon}$ with a $(n - k)$-cell attached, where n is the dimension of $\mathscr{M}$.*

It is now necessary to understand the effect of attaching a k-cell on the number of regions of a topological space. The first result is the following:

Theorem 2.3 *If a k-cell is attached to a topological space $\mathscr{X}$, then*

- *if $k = 0$, the number of disjoint regions of $\mathscr{X}$ is always increased by one;*
- *if $k \neq 0$, the number of disjoint regions of $\mathscr{X}$ does not increase.*

Theorem 2.3 can be proved as follows. The 0-cell has no boundary, therefore the image of the function g, used to glue the 0-cell, is empty. Recalling Definition 1.19, the result of attaching the 0-cell to $\mathscr{X}$ is just the disjoint

union of the two sets, which means that the 0-cell has generated a new disjoint region. Any other k-cell, with k greater than zero, has a boundary, therefore, when the k-cell is attached to $\mathscr{X}$ at least one point C of the cell is identified with one point X of $\mathscr{X}$. Since the cell is connected, C is reachable from any point of the cell, and, since C after the attachment coincides with X, all points of the cell are contained in the same disjoint region of X, which existed before the attachment of the cell.

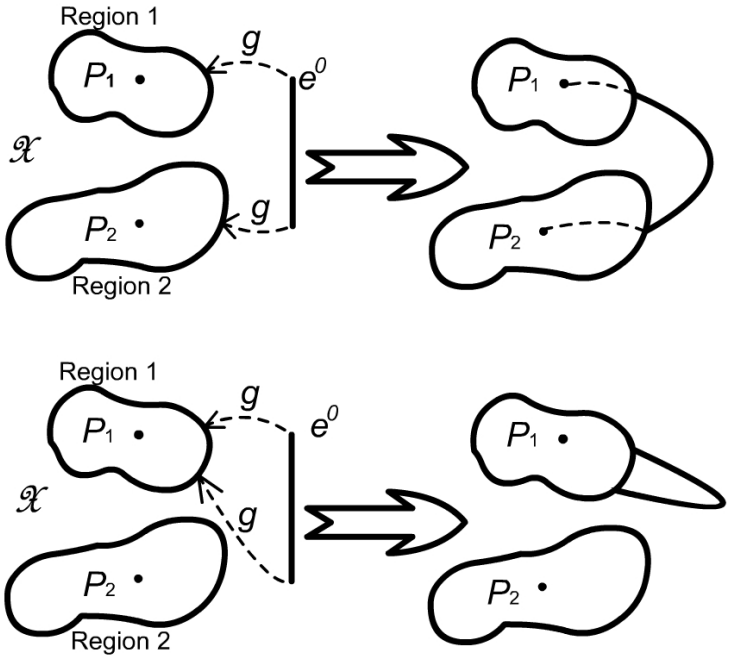


Figure 2.6: Attaching a 1-cell to a topological space, may or may not decrease by 1 the number of disjoint regions.

The ensuing result defines when the number of disjoint regions can decrease:

Theorem 2.4 *If a k-cell is attached to a topological space $\mathscr{X}$, then*

- *if the number of disjoint regions of $\mathscr{X}$ decreases, then it decreases by 1, and $k = 1$;*
- *if $k \neq 1$, the number of disjoint regions of $\mathscr{X}$ does not decrease.*

Indeed, suppose the number of regions decreases after attaching the k-cell. If so, after the attachment of the k-cell, there exists a continuous path going

from a point of a region of $\mathscr{X}$ to a point belonging to another. This path should start from the first point, enter the cell somewhere inside the first region through the image of g, exit the cell inside the second region through the image of g and reach the second point (see Figure 2.6). In order to allow this path, the image of g must contain at least two points belonging to two different disjoint regions, i.e. be disconnected. But the gluing function g is required to be continuous, thus, if its domain is continuous, its image must also be continuous, by virtue of Lemma 1.1. Any k-cell with $k > 1$ has a continuous boundary, therefore the image of g must be continuous and the number of cells cannot decrease.

If $k = 1$, then the number of disjoint regions may or may not decrease. With reference to Figure 2.6, the 1-cell can be attached as a bridge between two disjoint regions, connecting them, or as a handle to one disjoint region, producing a hole. At any rate, the 1-cell can connect at most two disjoint regions, because its boundary contains two points only.

A direct consequence of Theorem 2.3 and Theorem 2.4 is that the only k-cells which can affect the number of disjoint regions are the 0-cell and the 1-cell. Any other k-cell is unable to change the number of disjoint regions of a topological space, through the attachment.

2.3 Numerical Method

As pointed out in Section 2.1, two questions frequently arise when coping with singularities, most of all with parallel ones:

- Given any two configurations of the manipulator, is it possible to connect them through a singularity-free path?
- How many singularity-free disjoint regions do exist in the configuration space?

In this section, a numerical procedure will be developed, capable of answering these two questions.

It will be hereafter supposed that the configuration space $\mathscr{C}$ of a manipulator is a smooth compact n-dimensional manifold. The singularity locus will be supposed to be defined as the zero level set, $J^{-1}(0)$, of a smooth real valued function J defined on $\mathscr{C}$, which might be the Jacobian determinant, as discussed in Section 2.1.

In order to visualize the proposed method, imagine that the graph of the function J is plotted as a landscape upon the configuration space $\mathscr{C}$. The landscape can be imagined as partially covered with water and the level of

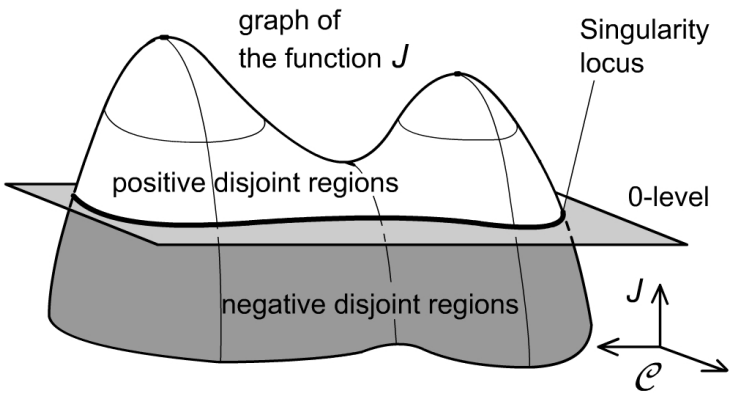


Figure 2.7: Graph of the function J on $\mathscr{C}$.

water is equal to zero. Figure 2.7 provides a lower-dimensional visualization of this representation, where the manifold $\mathscr{C}$ is supposed to be two-dimensional, and the graph of J is supposed to be a three-dimensional graph on $\mathscr{C}$. According to this representation, the shore-line is the singularity locus $J^{-1}(0)$. The set $\mathscr{C}^0_+ = J^{-1}[0, +\infty)$ is the set containing the *positive disjoint regions*, i.e. the disjoint regions composing $\mathscr{C} - J^{-1}(0)$ where the function J is positive, and is represented as the islands above the shore-line. The set $\mathscr{C}^0_- = J^{-1}(-\infty, 0]$ is the set containing the *negative disjoint regions*, i.e. the disjoint regions composing $\mathscr{C} - J^{-1}(0)$ where the function J is negative, and is represented as the region below the shore-line.

Since the function J is smooth, and is defined upon the compact manifold $\mathscr{C}$, there are an absolute maximum and an absolute minimum of J on $\mathscr{C}$. We can then imagine that the level of water is raised above the level of the absolute maximum, until the whole landscape is completely flooded with water. From this point, we imagine that the level of water a slowly decreases, down to zero level, and we observe the evolution of the set $\mathscr{C}^a_+ = J^{-1}[a, +\infty)$, composed of the islands above the current water level.

Suppose now that the graph of J is similar to that depicted in Figure 2.8, with two nondegenerate maxima M_1 and M_2 and one nondegenerate saddle point S. S is supposed to be a $(n-1)$-saddle point, where n is the dimension of the manifold $\mathscr{C}$, and a k-saddle point is defined as a saddle point with index k. If the water level a is higher than the level of the absolute maximum M_1, then the set $\mathscr{C}^a_+$ is empty. As soon as the water level reaches the height of the highest peak, M_1, an island crops out from the water. As soon as a critical point of J is met, the number of disjoint regions composing $\mathscr{C}^a_+$ changes. Before meeting the absolute maximum M_1, $\mathscr{C}^a_+$ was empty: it

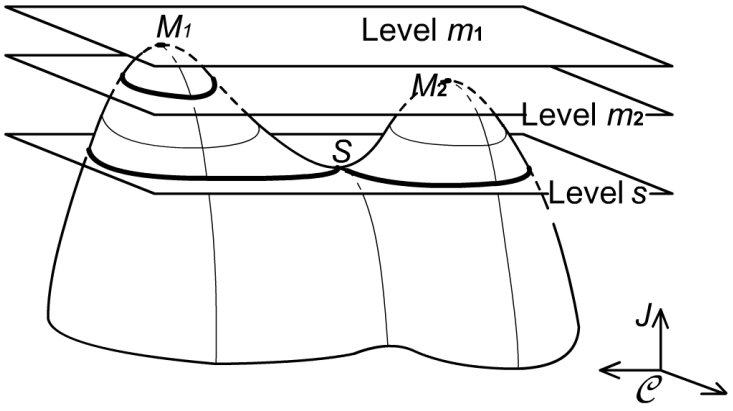


Figure 2.8: The number of islands above the water is observed, while the level of water decreases.

contained zero disjoint regions. After meeting the absolute maximum, the number of disjoint regions composing it changes as if a k-cell were attached to it . The maximum is a critical point of index n and the dimension of the manifold $\mathscr{C}_+^a$ is also n, thus, k equals 0 (Theorem 2.2). By virtue of Theorem 2.3, if a 0-cell is attached to the set $\mathscr{C}_+^a$ one disjoint region is added, thus, after passing the maximum the number of regions is 1.

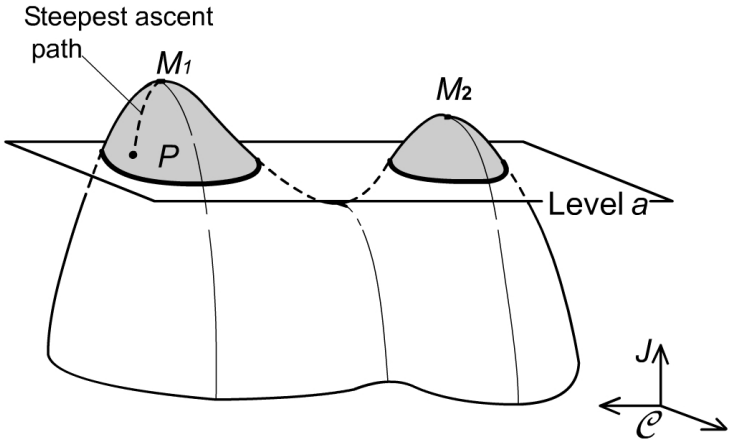
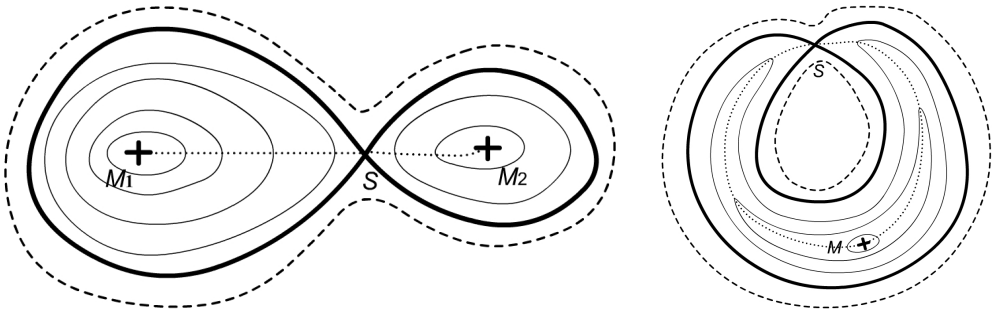


Figure 2.9: The steepest ascent path starting from P determines the disjoint region to which P belongs.

The level of water a keeps on decreasing: as long as it remains between m_1 and m_2, the heights of the two maxima of Figure 2.8, the number of

disjoint regions remains equal to 1, by virtue of Theorem 2.1. There exists only one island above the water. As soon as the maximum M_2 is reached, another island appears and a new disjoint region of $\mathscr{C}^a_+$ is generated. Passing the level m_2 the number of disjoint regions changes as if another 0-cell were attached to $\mathscr{C}^a_+$ for Theorem 2.2, thus the number of disjoint regions now equals 2 for Theorem 2.3. The number of disjoint regions remains equal to 2 until the saddle point S is reached by virtue of Theorem 2.1. Consider a point P of $\mathscr{C}^a_+$, with a contained in the open interval (s, m_2). It is possible to assess whether P belongs to the disjoint region generated by M_1 or to the disjoint region generated by M_2. The *steepest ascent* path starting from P must reach one of the two maxima M_1 or M_2: P belongs to the disjoint region generated by the maximum point reached (see Figure 2.9). Thus, the maxima work as labels for the disjoint regions: each disjoint region can be identified by means of the maximum contained in it. The steepest ascent path is a path which is tangent to the direction of the gradient at each of its points, and which is oriented as the gradient. Such paths can also be viewed as the curve $\varphi_t(P)$ of the 1-parameter group of diffeomorphisms generated in Section 1.3.3: every point can be delivered through the diffeomorphism φ_t arbitrarily close to the maximum which generated the disjoint region containing the point.

As soon as the level a reaches the height of the saddle point S, another variation of the number of disjoint regions of $\mathscr{C}^a_+$ is expected. Since the index of S is equal to $n-1$, the number of regions composing $\mathscr{C}^a_+$ changes as if a 1-cell were attached to it. By virtue of Theorem 2.4, the number of regions may be diminished by one or remain constant.



a) ***b)***

Figure 2.10: A $(n-1)$-saddle point may join two disjoint regions (a), or produce a new hole (b).

In Figure 2.10, the two possibilities are shown. The boundary of $\mathscr{C}^a_+$ is drawn with thin solid line before meeting the saddle, with bold line at the level of the saddle and with dashed line after having met the saddle.

The maxima are represented as crosses, while the saddles are the "touching points" of the boundary of $\mathscr{C}_+^a$. Two disjoint regions may touch and join at the saddle, as occurs in Figure 2.8, or a disjoint region may join with itself. In the former case, the number of disjoint regions is diminished by 1. In the latter, the topology of Figure 2.8 changes, because a hole appears after meeting the saddle, but the number of disjoint regions remains constant. To decide whether or not the number of disjoint regions has decreased, it is necessary to find out to which one of the existing regions the saddle point belongs. The method to reach this goal is identical to that proposed for a noncritical point: the steepest ascent path is followed starting from the saddle, until a maximum is reached. There are two different steepest ascent path starting from a $(n-1)$-saddle point (dotted lines in Figure 2.10): they leave the saddle point along the direction of the eigenvector associated to the positive eigenvalue of the Hessian matrix. If the steepest ascent paths reach the same maximum (cross in Figure 2.10), then a region joins with itself, and the number of regions remains constant (Figure 2.10b). If the steepest ascent paths reach two different maxima, the regions generated by the two maxima join together (Figure 2.10a).

To identify the new region generated by the joining, the maxima inside it (in case of Figure 2.10a, the two maxima) can be used: the steepest ascent path starting from any point inside the new disjoint region will lead to one of its critical points. In some rather imprbable cases, such paths may lead to the saddle point S: in such cases it is possible to follow any of the two steepest ascent paths starting from the saddle point, in order to reach one of the maxima contained in the new region.

From now on it is possible to proceed iteratively. Each maximum generates a new region, and each $(n-1)$-saddle may connect two existing regions. Following the two steepest ascent paths as for the first saddle, two maxima are reached: if they belong to two different regions, such two regions have joined. If the reached critical points belong to the same region, the number of regions remains constant.

As the level a reaches the value 0, the number of regions that compose $\mathscr{C}_+^a$ is determined. Each of these regions is provided with a set of maxima that completely characterizes it. Furthermore, all the maxima are connected to each other through a network of singularity-free paths.

Note that the singular critical points must not be considered: if the singular critical point is a maximum, an isolated singular point appears in the jointspace, if it is a $(n-1)$-saddle point, two regions "touch" on their boundary, but no connection is established between them.

The critical points with indexes different from n or $(n-1)$ are ignored. When such a critical point is met, a k-cell with $k > 1$ is attached to $\mathscr{C}_+^a$,

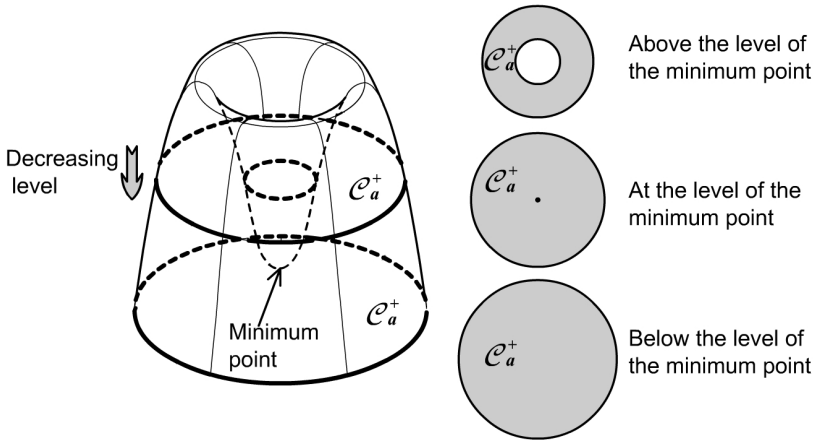


Figure 2.11: The critical points with indexes different from n or $(n-1)$ are ignored.

therefore, for Theorem 2.3 and Theorem 2.4, the number of regions remains the same. Figure 2.11 shows an example for a two-dimensional manifold $\mathscr{C}$: in the neighborhood of a non-degenerate positive minimum point, the graph of the function J looks like a bowl, because the value of J increases for any small enough displacement starting from the minimum point. While approaching the level of the minimum, a hole appears sooner or later in $\mathscr{C}_+^a$. Passing the minimum, the effect on $\mathscr{C}_+^a$ is the same as if a rubber patch, a 2-cell, were attached to it. The hole is closed by the patch, but the number of regions composing $\mathscr{C}_+^a$ does not vary, because the patch must be glued along its boundary, which is connected, thus the patch cannot work as a bridge between two disjoint regions. Recalling the analogy with landscape and water, there is a lake on an island. As the level of water decreases, the lake disappears, but the number of islands remains obviously the same.

If a degenerate critical point is met, it is not possible to know whether the number of regions composing $\mathscr{C}_+^a$ is changing by means of the Hessian matrix only. Higher derivatives have to be considered: the point might be a maximum, thus a new region is born. Or it might be neither a maximum nor a minimum and two or more regions could join together. Figure 2.12 shows an example of a *monkey saddle*[1], which joins three disjoint regions.

[1]The monkey saddle is for example the point $(0,0)$ for the function $f(x,y) = -x^3+3xy^2$. The monkey saddle is usually depicted as in Figure 2.12, but with opposite sign, and the name is due to the fact that its graph might be a good saddle for monkeys. Indeed, since the monkey has the tail, it needs three sides of the saddle pointing downward, and only the front side pointing upward.

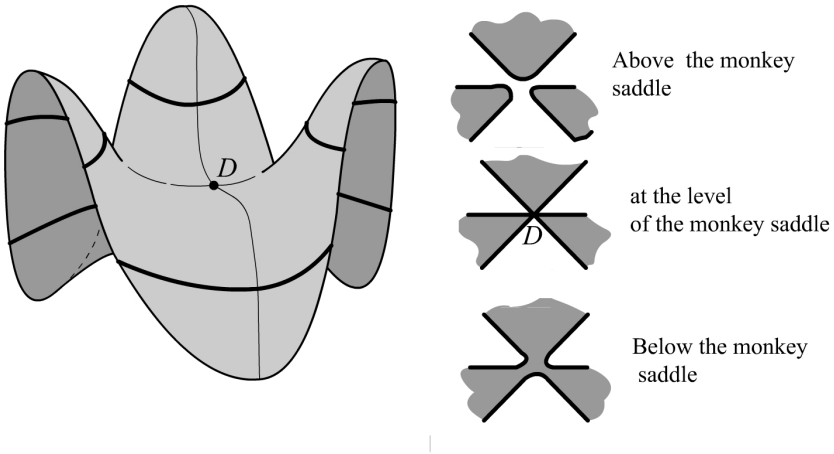


Figure 2.12: A monkey saddle joins three disjoint regions.

Of course, an analogous method can be used to count and identify the number of regions composing $\mathscr{C}_-^a$, thus, at the end of this procedure, it is possible to establish to which region any non-singular point belongs. Summarizing the procedure, the following operations have to be performed:

1. Determine all the critical points of J. This is the toughest part of the procedure, and will be developed according to the specific case at hand.

2. By studying the eigenvalues of the Hessian matrix of J, each critical point is identified as a degenerate or nondegenerate critical point, and the indexes of nondegenerate critical points are calculated.

3. The value of J is calculated at each critical point, in order to recognize those belonging to $\mathscr{C}_+^a$, those belonging to $\mathscr{C}_-^a$ and the singular ones. If any nonsingular degenerate critical point is found, it is impossible to proceed, for Morse theory does not hold anymore.

4. The maxima of $\mathscr{C}_+^a$ are sorted from the highest to the lowest, according to the value of J. The same is done with the $(n-1)$-saddles of $\mathscr{C}_+^a$, while other critical points in $\mathscr{C}_+^a$ are ignored.

5. Starting from each positive $(n-1)$-saddle, the two steepest ascent paths are followed, until they reach any of the maxima higher than the saddle. All maxima belonging to the same region of the two reached maxima are assigned to the same region. This operation is performed from the highest to the lowest saddle point. The singularity-free steepest ascent paths connecting the positive maxima are stored.

6. The minima of $\mathscr{C}_-^a$ are sorted from the lowest to the highest, according to the value of J. The same is done with the 1-saddles of $\mathscr{C}_-^a$, while other critical points in $\mathscr{C}_-^a$ are ignored.

7. Starting from each negative 1-saddle, the two steepest descent paths are followed, until they reach any of the minima lower than the saddle. All the maxima belonging to the same region of the two reached maxima are assigned to the same region. This operation is performed from the lowest to the highest saddle point. The singularity-free steepest descent paths connecting the negative minima are stored.

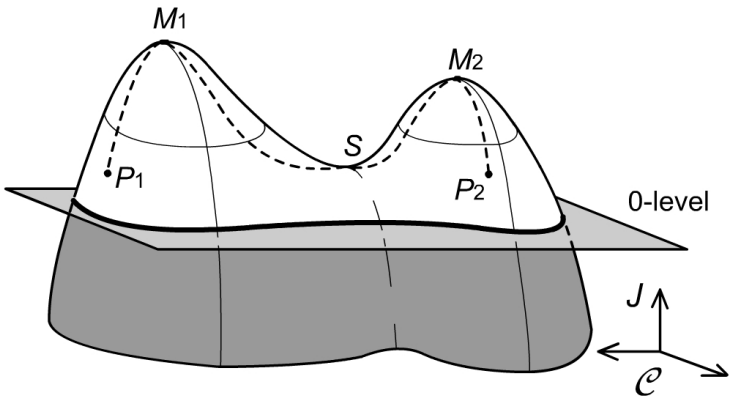


Figure 2.13: A singularity-free path between any two points in the same disjoint region is always found.

At this point, given any two configurations P_1 and P_2 it is always possible to find a singularity-free path connecting them, if it exists. The values of $J(P_1)$ and $J(P_2)$ are calculated. If the sign of $J(P_1)$ is different from the sign of $J(P_2)$, then there exists no singularity-free path between P_1 and P_2 because on any path J will vanish as its sign changes.

If $J(P_1)$ and $J(P_2)$ are both positive, then the steepest ascent paths starting from P_1 and P_2 are generated. If the two paths reach maxima belonging to two different positive disjoint regions, then there exists no singularity-free path between P_1 and P_2. If the two reached maxima belong to the same disjoint region, then a singularity-free path between P_1 and P_2 can be obtained by joining the two steepest ascent paths from P_1 and P_2 to the maxima, and any path in the singularity-free network connecting the maxima (see Figure 2.13). An analogous method can be used if $J(P_1)$ and $J(P_2)$ are both negative.

Chapter 3

Serial Manipulators

In this Chapter, the serial singularity loci of two types of serial manipulators will be defined and analyzed by means of the numerical method proposed in Chapter 2.

3.1 Spatial 3R regional serial Manipulators

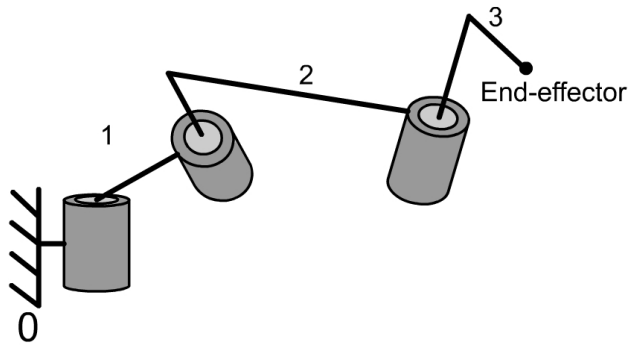


Figure 3.1: A spatial 3R regional serial manipulator.

A spatial 3R regional serial manipulator (3R manipulator henceforth) is a serial manipulator, composed of four rigid links, connected by three actuated revolute pairs. The task of such 3R manipulators is to place a point, the end-effector, in the three-dimensional Euclidean space (see Figure 3.1).

The serial singularity locus of 3R manipulators has been actively studied in the last two decades, probably because such singularity locus can be written as a relatively simple equation and can be easily visualized through a two-dimensional representation. Nevertheless, the shape of such singularity

locus is not trivial at all, which justifies a deeper search of a better geometrical insight.

3.1.1 Jointspace and Workspace

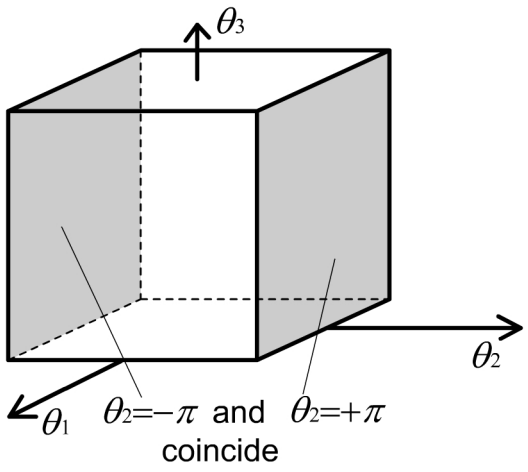


Figure 3.2: Jointspace of a spatial 3R regional serial manipulator.

The points in the jointspace of a 3R manipulator are the ordered sets containing the positions of the three actuated revolute joints. If the position of the i^{th} revolute joint is defined through the angle θ_i between the two links connected by the revolute joint, then each triplet $(\theta_1, \theta_2, \theta_3)$ identifies a set of positions of the actuators. Therefore, the jointspace of the 3R manipulator is homeomorphic to the set containing all possible triplets of real numbers modulo 2π, for the angles θ_i and $\theta_i + 2\pi$ do coincide. The jointspace is homeomorphic to the Cartesian product $\mathscr{S}_1 \times \mathscr{S}_1 \times \mathscr{S}_1$, where $\mathscr{S}_1$ is the one-dimensional circle. This Cartesian product is a 3-torus, (see [9]), and is homeomorphic to a three-dimensional cube, whose opposite faces are identified. This cube is depicted in Figure 3.2: the triplets $(\theta_1, \theta_2, \theta_3)$ are plotted in the three-dimensional Euclidean space, and the cube represents the region where all three variables are greater than $-\pi$ and lesser that $+\pi$. The opposite faces $\theta_i = -\pi$ and $\theta_i = +\pi$ identify the same points in the jointspace, and therefore are "glued" together, similar to the attachment of k-cells in the previous chapters.

The workspace of 3R manipulators contains all the points of the three-dimensional Euclidean space that can be reached by the end-effector, and can be defined as the image of the aforementioned 3-torus through a proper

function f. A global coordinate system (x, y, z) can be introduced in the workspace of 3R manipulators, because it is a subset of the three-dimensional Euclidean space. The function f which links each position of the end-effector (x, y, z) to each point in the jointspace $(\theta_1, \theta_2, \theta_3)$, can be found by means of 4×4 homogeneous transformation matrices, first introduced by Denavit and Hartenberg (see [23] and [24], Chapter 2).

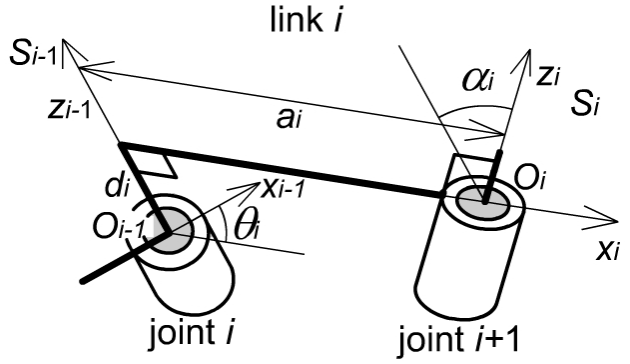


Figure 3.3: DH-parameters of a rigid link.

First of all, a reference frame is attached to each rigid link of the manipulator with two revolute joints. (see Figure 3.3). The axis z_i of the reference frame S_i, attached to the i^{th} link, is directed along the revolute joint closest to the end-effector, and the axis x_i is orthogonal to the axes of both revolute joints of the link.

The reference frame S_{i-1} can be obtained from the reference frame S_i by means of the ensuing subsequent operations:

1. the reference frame S_i is rotated about x_i axis an angle α_i, in order for axes z_i and z_{i-1} to become parallel;

2. the reference frame obtained at point 1 is translated along x_i axis a distance a_i, in order for axes z_i and z_{i-1} to coincide;

3. the reference frame obtained at point 2 is translated along z_{i-1} axis a distance d_i, in order for the origins of the two frames to coincide;

4. the reference frame obtained at point 3 is rotated about z_{i-1} axis an angle θ_i, the i^{th} joint angle, which eventually superimposes the two frames;

Therefore the three link parameters a_i,d_i, and α_i and the joint angle θ_i completely define the relative position of two reference frames fixed upon two adjoining links. The three aforementioned link parameters are usually named Denavit-Hartemberg parameters (DH-parameters henceforth).

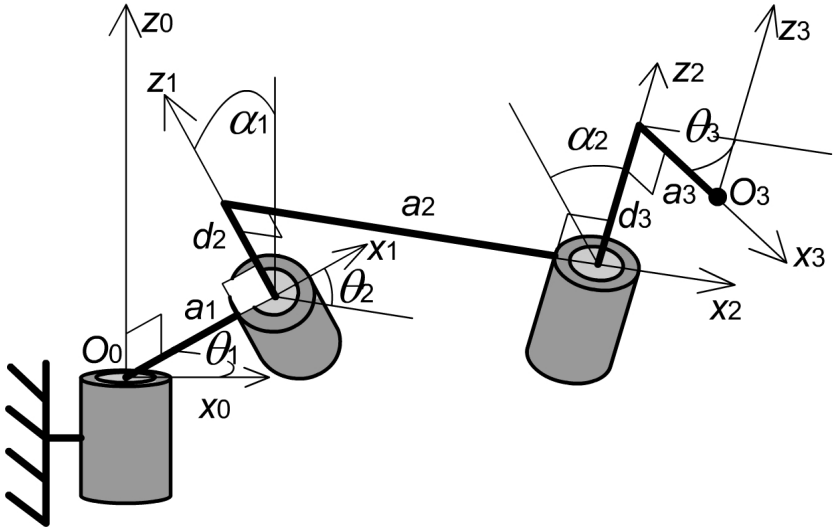


Figure 3.4: DH-parameters for 3R manipulators.

For 3R manipulators, two reference frames can be attached upon links with two revolute joints, i.e. links 1 and 2, as discussed above. A reference frame S_0 is defined upon base link 0, so that the axis z_0 is the axis of the first revolute joint, the origin O_0 lies upon axis x_1. Whatever fixed axis normal to z_0 at O_0 can be used to define x_0.

A reference frame S_3 is defined upon the last link 3, so that the origin O_3 coincides with the end-effector, the z_3 axis is parallel to the axis of the last revolute joint, and the x_3 axis is the normal line from O_3 to z_2.

Eventually, seven DH-parameters and three joint angles can be used to characterize the relative position of the four reference frames S_0, S_1, S_2, and S_3 , as shown in Figure 3.4 and summarized in Table 3.1.

The seven DH-parameters completely define the geometry of a 3R manipulator, whereas the three joint angles define its configuration.

The coordinate change from the frame S_i to the frame S_{i-1} can be defined by means of a 4×4 homogeneous transformation matrix. If $(x_i, y_i, z_i, 1)^{\mathrm{T}}$ is a vector containing the homogeneous coordinates of a point with respect to frame S_i, then the vector $(x_{i-1}, y_{i-1}, z_{i-1}, 1)^{\mathrm{T}}$, containing the coordinates of the same point with respect to S_{i-1} can be expressed as:

Reference frames	DH-parameters	Joint angle
$S_1 \rightarrow S_0$	a_1, α_1	θ_1
$S_2 \rightarrow S_1$	a_2, d_2, α_2	θ_2
$S_3 \rightarrow S_2$	a_3, d_3	θ_3

Table 3.1: DH-parameters defining the geometry of a 3R manipulator.

$$\begin{pmatrix} x_{i-1} \\ y_{i-1} \\ z_{i-1} \\ 1 \end{pmatrix} = \mathbf{A}_{(i-1),i} \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} \tag{3.1}$$

The matrix $\mathbf{A}_{(i-1),i}$ is a 4×4 matrix, composed as follows:

$$\mathbf{A}_{(i-1),i} = \begin{pmatrix} \mathbf{R}_{(i-1),i} & \mathbf{u}_{i-1}(O_i) \\ 0 \quad 0 \quad 0 & 1 \end{pmatrix} \tag{3.2}$$

where

- $\mathbf{R}_{(i-1),i}$ is the 3×3 rotation matrix from frame S_i, to frame S_{i-1}, whose columns are the directions cosines of the axes of S_i with respect to the axes of S_{i-1};
- $\mathbf{u}_{i-1}(O_i)$ is the vector containing the coordinates of the origin O_i of S_i with respect to S_{i-1}.

In terms of DH-parameters a_i, d_i, α_i, and of the joint angle θ_i, the matrix $\mathbf{A}_{(i-1),i}$ is can be written as follows:

$$\mathbf{A}_{(i-1),i} = \begin{pmatrix} \mathrm{c}\theta_i & -\mathrm{c}\alpha_i\mathrm{s}\theta_i & \mathrm{s}\alpha_i\mathrm{s}\theta_i & a_i\mathrm{c}\theta_i \\ \mathrm{s}\theta_i & \mathrm{c}\alpha_i\mathrm{c}\theta_i & -\mathrm{s}\alpha_i\mathrm{c}\theta_i & a_i\mathrm{s}\theta_i \\ 0 & \mathrm{s}\alpha_i & \mathrm{c}\alpha i & d_i \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.3}$$

where s and c are acronyms for sine and cosine.

The transformation matrix from end-effector frame S_3 to base frame S_0 is obtained as follows:

$$\mathbf{A}_{0,3}(\theta_1, \theta_2, \theta_3) = \mathbf{A}_{0,1}(\theta_1)\mathbf{A}_{1,2}(\theta_2)\mathbf{A}_{2,3}(\theta_3) \tag{3.4}$$

Finally, the position of the end-effector related to three generic joint angles is determined through matrix $\mathbf{A}_{0,3}$ as the coordinates $(x, y, z, 1)^T$ in

frame S_0 of the origin of frame S_3, namely:

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \mathbf{A}_{0,3}(\theta_1, \theta_2, \theta_3) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.5}$$

From Equation (3.5) the function f such that

$$(x, y, z) = f(\theta_1, \theta_2, \theta_3) \tag{3.6}$$

can be easily inferred. The image of the torus through such function also defines the workspace of the 3R manipulator.

Equation (3.6) yields the unique solution to the direct kinematics problem. The inverse kinematics problem, i.e. the problem of finding all points in the jointspace with the same given image in the workspace, can be reduced to finding out all the roots of a fourth-order polynomial in one joint angle. This solution will be omitted here, for more details see [25], annex 1.

3.1.2 Serial singularity locus

Serial singularities of 3R manipulators can be found by differentiating Equation (3.6):

$$\begin{pmatrix} \partial x \\ \partial y \\ \partial z \end{pmatrix} = \mathbf{Q}_f(\theta_1, \theta_2, \theta_3) \begin{pmatrix} \partial\theta_1 \\ \partial\theta_2 \\ \partial\theta_3 \end{pmatrix} \tag{3.7}$$

where $\mathbf{Q}_f$ is the Jacobian matrix of f. Serial singularities are the configurations of the manipulator where there exists a virtual displacement in the jointspace which produces no virtual displacement of the end-effector. Therefore, serial singularities are the points of the jointspace $(\theta_1, \theta_2, \theta_3)$ such that:

$$J(\theta_1, \theta_2, \theta_3) = \det \mathbf{Q}_f = 0 \tag{3.8}$$

The function J can be written as follows (see also [26] and [9]):

$$J(\theta_2, \theta_3) = a_3 \left(V_1(\theta_3)\mathrm{c}\theta_1 + V_2(\theta_3)\mathrm{s}\theta_2 + V_3(\theta_3)\right) \tag{3.9}$$

where:

$$\begin{aligned} V_1(\theta_3) &= m_1\mathrm{c}^2\theta_3 + m_4\mathrm{s}\theta_3\mathrm{c}\theta_3 + m_6\mathrm{c}\theta_3 + m_7\mathrm{s}\theta_3 \\ V_2(\theta_3) &= m_2\mathrm{c}^2\theta_3 + m_3\mathrm{s}^2\theta_3 + m_5\mathrm{s}\theta_3\mathrm{c}\theta_3 + m_8\mathrm{c}\theta_3 + m_9\mathrm{s}\theta_3 \\ V_3(\theta_3) &= m_{10}\mathrm{s}\theta_3\mathrm{c}\theta_3 + m_{11}\mathrm{c}\theta_3 + m_{12}\mathrm{s}\theta_3 \end{aligned} \tag{3.10}$$

and:

$$\begin{aligned}
m_1 &= a_3 d_2 \mathrm{s}\alpha_1 \mathrm{s}\alpha_2 \\
m_2 &= a_2 a_3 \mathrm{c}\alpha_1 \mathrm{s}\alpha_2 \\
m_3 &= a_2 a_3 \mathrm{s}\alpha_1 \mathrm{c}\alpha_2 \\
m_4 &= a_1 a_3 \mathrm{c}\alpha_1 \mathrm{c}\alpha_2 \mathrm{s}\alpha_2 - a_2 a_3 \mathrm{s}\alpha_1 \\
m_5 &= -a_3 d_2 \mathrm{s}\alpha_1 \mathrm{c}\alpha_2 \mathrm{s}\alpha_2 \\
m_6 &= a_2 d_2 \mathrm{s}\alpha_1 \mathrm{s}\alpha_2 - a_1 d_3 \mathrm{c}\alpha_1 \mathrm{s}^2\alpha_2 \\
m_7 &= -a_2^2 \mathrm{s}\alpha_1 \\
m_8 &= d_2 d_3 \mathrm{s}\alpha_1 \mathrm{s}^2\alpha_2 + a_1 a_2 \mathrm{c}\alpha_1 \mathrm{s}\alpha_2 \\
m_9 &= -a_2 d_3 \mathrm{s}\alpha_1 \mathrm{s}\alpha_2 \\
m_{10} &= -a_1 a_3 \mathrm{s}\alpha_1 \mathrm{s}^2\alpha_2 \\
m_{11} &= -a_1 d_3 \mathrm{s}\alpha_1 \mathrm{c}\alpha_2 \mathrm{s}\alpha_2 \\
m_{12} &= -a_1 a_2 \mathrm{s}\alpha_1
\end{aligned} \tag{3.11}$$

It is evident from Equation (3.9) that if $a_3 = 0$ any position of the manipulator is trivially singular. It will be hereafter assumed that a_3 is equal to the unit length, and that a_1, a_2, d_2, and d_3 are measured accordingly.

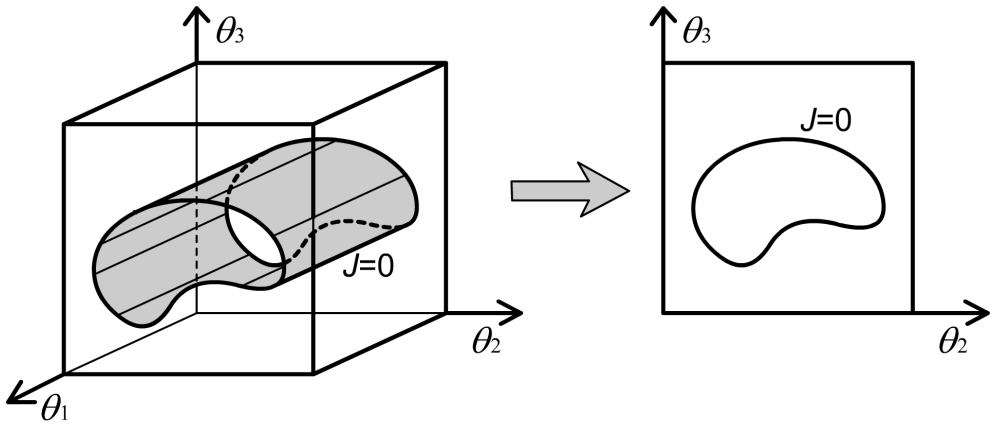


Figure 3.5: The singularity locus is an extruded surface along the θ_1 generator of the torus.

Equation (3.9) shows that the singularity locus can be expressed as the zero level set of a smooth function J, defined over all the jointspace. Furthermore, the function J does not depend on the first joint angle θ_1. The singularity surface can be obtained by extruding the section on plane $\theta_2\theta_3$

along θ_1 generator, as shown in Figure 3.5. Thus, the geometry of the singularity locus can be better studied and represented by considering only its section through the plane $\theta_2\theta_3$, which completely characterizes it (see [25]).

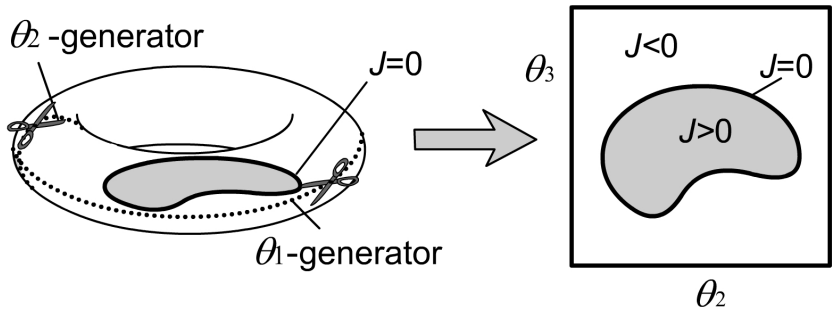


Figure 3.6: The section of the jointspace through $\theta_2\theta_3$ plane is a two-dimensional torus.

The section through $\theta_2\theta_3$ plane of the jointspace is a square with the opposite sides identified. This is a representation of a two-dimensional torus, cut along its two generators and laid flat on a plane, as shown in Figure 3.6. Since this work is focused on the topology of the singularity locus, the 2-torus described by the two angles θ_2 and θ_3 will be henceforth referred to as the whole jointspace of the 3R manipulator. The singularity locus is represented as a set of curves upon this 2-torus, which will be henceforth named singularity curves.

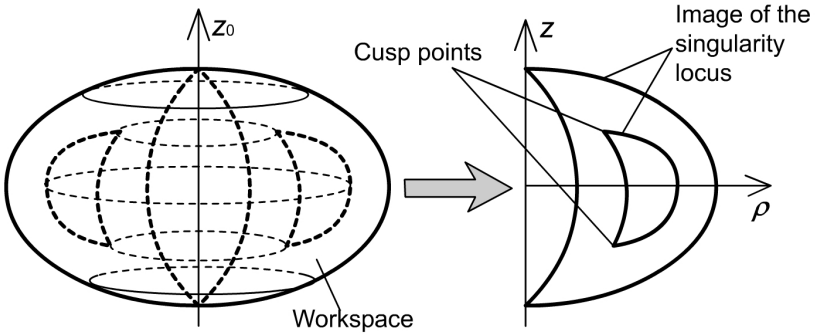


Figure 3.7: Only a radial section of the workspace is considered.

The workspace of 3R manipulators is clearly a solid obtained by revolution of a certain section about the axis of the first revolute joint. Since the image of the singularity locus in the workspace is a revolute surface about the axis

of the first revolute joint too, only a radial section of the workspace will be henceforth considered. With reference to Figure 3.7, the radial section of the workspace will be represented on a Cartesian plane, where the two axis z and ρ represent the coordinate z and the distance from z_0 axis of the end-effector, respectively. Therefore $\rho = \sqrt{x^2 + y^2}$.

The singularity locus of 3R manipulators has been intensely studied, with particular attention to the capability of changing posture along a singularity-free path. In [27], [28], and [29] it was proved that a 3R manipulator can change posture without crossing a singularity if and only if there are cusp points in its workspace. Such manipulators were also named *cuspidal* manipulators. A point of the workspace is a cusp if the inverse kinematics problem admits three coincident solutions for such point. Cusps usually appear as sharp corners of the image of the singularity curves in the workspace, as depicted in Figure 3.7. The dynamics of the singularity-free posture change has been analyzed and discussed in [30], [31], and [32].

However, no method was given to find out a singularity-free path between any two postures of a manipulator, or to count and identify the singularity-free regions in the jointspace. The general method derived in Chapter 2 has been applied in [33] to 3R manipulators, and will be discussed in the next Section.

3.1.3 Analysis of serial singularity locus for 3R manipulators

With the same notation as in Chapter 2, the serial singularity locus of a 3R manipulator is defined as the zero level set of the smooth function J on the configuration space, which coincides with the jointspace for serial manipulators.

The first thing needed for the application of the method of Chapter 2 is all critical points of J on the 2-torus representing the jointspace, which can be obtained by imposing that the gradient of J vanishes. The system of equations $\nabla J = \mathbf{0}$ can be written as follows:

$$\begin{pmatrix} \partial J/\partial\theta_3 \\ \partial J/\partial\theta_2 \end{pmatrix} = \begin{pmatrix} V_1'(\theta_3) & V_2'(\theta_3) \\ V_2(\theta_3) & -V_1(\theta_3) \end{pmatrix} \begin{pmatrix} \mathrm{c}\theta_2 \\ \mathrm{s}\theta_2 \end{pmatrix} = \begin{pmatrix} -V_3'(\theta_3) \\ 0 \end{pmatrix} \tag{3.12}$$

where $V_1'(\theta_3)$, $V_2'(\theta_3)$ and $V_3'(\theta_3)$ are the first derivatives of $V_1(\theta_3)$, $V_2(\theta_3)$, and $V_3(\theta_3)$ with respect to θ_3.

The sine and cosine of θ_2 are immediately obtained from Equation (3.12),

if $V_1'V_1 + V_2'V_2 \neq 0$:

$$\begin{pmatrix} \mathrm{c}\theta_2 \\ \mathrm{s}\theta_2 \end{pmatrix} = \frac{1}{V_1'V_1 + V_2'V_2} \begin{pmatrix} -V_3'V_1 \\ -V_3'V_2 \end{pmatrix} \tag{3.13}$$

By summing up the squares of the sine and cosine of θ_2 and subtracting 1, a polynomial equation in the only joint angle θ_3 is obtained:

$$(V_1'V_1 + V_2'V_2)^2 = (V_3')^2 \left(V_1^2 + V_2^2\right) \tag{3.14}$$

It can be easily checked that Equation (3.14) stems from Equation (3.12) also under the hypothesis $V_1'V_1 + V_2'V_2 = 0$. Therefore, all values of θ_3 at critical points are always solutions to Equation (3.14). If the sine and cosine of θ_3 are replaced in Equation (3.14) by their expressions in function of the tangent of $\theta_3/2$, Equation (3.14) becomes a polynomial of degree sixteen in the tangent of $\theta_3/2$. At most sixteen different values of θ_3 are expected at critical points, which can be easily found numerically at any desired precision.

The unique value of θ_2 corresponding to each θ_3 is found through Equation (3.13), if $V_1'V_1 + V_2'V_2 \neq 0$. If $V_1'V_1 + V_2'V_2 = 0$ there are three possible cases:

1. $V_1 = V_2 = V_3' = 0$, with the following sub-cases:
   - $V_1' = 0$ and $V_2' = 0$. In this particular case any value of θ_2 satisfies Equation (3.12): a horizontal line of critical points is found. Since Theorem 1.7 states that nondegenerate critical points are always isolated, these points must be all degenerate. If $J \neq 0$ at one of these points the algorithm must stop.
   - $V_1' \neq 0$ and $V_2' = 0$: $\theta_2 = \pm\pi/2$, two critical points are found.
   - $V_2' \neq 0$: $\tan\theta_2 = -V_1'/V_2'$, two critical points are found.
2. $V_1 = V_2 = 0$ and $V_3' \neq 0$. Equation (3.12) reduces to the first equation only, that can be easily solved leading to zero, one or two critical points.
3. $V_1^2 + V_2^2 \neq 0$ and $V_3' = 0$. Equation (3.12) reduces to two equivalent equations, and at least one of the coefficients of the second never vanishes. The second equation can be solved in a similar way as the last two sub-cases of case 1, bringing two critical points.

Therefore, there are at most sixteen real solutions θ_3 of Equation (3.14) where $V_1'V_1 + V_2'V_2 \neq 0$, for which there is only one value of θ_2 such that (θ_2, θ_3) is a critical point. For each real solution θ_3 of Equation (3.14) where $V_1'V_1 + V_2'V_2 = 0$ there can be zero, one, two or infinite critical points. If there are infinite critical points, they are all degenerate. If $V_1'V_1 + V_2'V_2 = 0$

at a solution θ_3 of Equation (3.14), then its multiplicity must be greater than one, for the first derivative of Equation (3.14) vanishes. Thus there can be at most eight solutions θ_3 of Equation (3.14) which identify two different nondegenerate critical points. This proves that, at any rate, there can be at most sixteen nondegenerate critical points of J upon the 2-torus.

By differentiating the gradient of J, the Hessian matrix at critical points can be easily calculated, therefore all critical points are classified according to their index, and subdivided into positive or negative according to the value of J.

Positive maxima and negative minima generate positive and negative regions respectively. On the other hand, negative maxima and positive minima shall be ignored, because they produce the attachment of 2-cells, unable to change the number of regions.

In this particular case, the dimension n of the jointspace is two, therefore any saddle point is both a 1-saddle and and a $(n-1)$-saddle point at the same time. Indeed, the positive saddle points glue the boundaries of the positive regions, whereas the negative saddles glue the boundaries of the negative regions.

The steepest ascent paths can be constructed by subsequent displacements along the direction of the gradient. Starting from the 1-saddles, two first small displacements are performed along the direction of the eigenvector corresponding to the positive eigenvalue of the Hessian matrix. The next point of each path is found by adding a small displacement along the direction of the gradient to the previous point. Before continuing with another displacement, two failsafe checkups are performed: it is ascertained that the value of J always increases along the path, and it is ascertained that the direction of the gradient does not undergo angular variations greater than a given threshold, which would mean that a critical point has already been passed. If the two failsafe checkups are both positive, then the second point is accepted, and the length of the next displacement is increased. If one of the two failsafe checkups yields negative response, then the next point is rejected, and the displacement is repeated with a shorter length. The presented algorithm is rather stable and it never occurred that it failed to track steepest ascent paths.

After each small displacement, it is checked that any of the maxima of J is closer than a given threshold. If this control is positive for one maximum, then such maximum is reached. The same can be done for steepest descent paths.

3.1.4 Examples

Some numerical examples of application of the proposed method will be hereafter presented.

Consider the manipulator M_1, whose DH-parameters are reported in Table 3.2.

a_1	a_2	a_3	d_2	d_3	α_1	α_2
4	2	1	0	3	$-80°$	$-80°$

Table 3.2: DH-parameters of manipulator M_1

The critical points and the steepest ascent paths are found as discussed above. The positive critical points are one maximum M_1, two saddles S_1 and S_2 and a minimum N_1. In Figure 3.8, the positive maximum is depicted as a cross, the positive saddles as stars, the positive minimum as a circle, and the steepest ascent paths as dash-dotted lines.

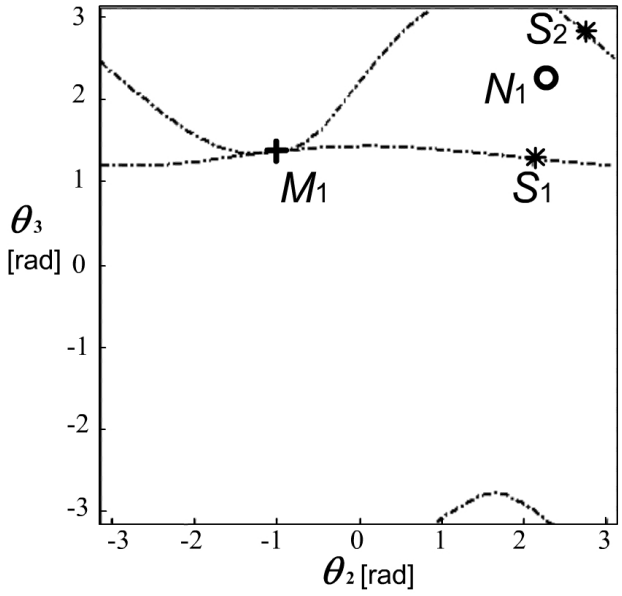


Figure 3.8: Positive critical points of manipulator M_1.

Since the jointspace is two-dimensional in this case, the level sets of the Jacobian determinant can be easily plotted, in order to verify the procedure proposed in Chapter 2. Figures 3.9, 3.10, and 3.11 show the evolution of the level sets while the level decreases. A new region is generated at the

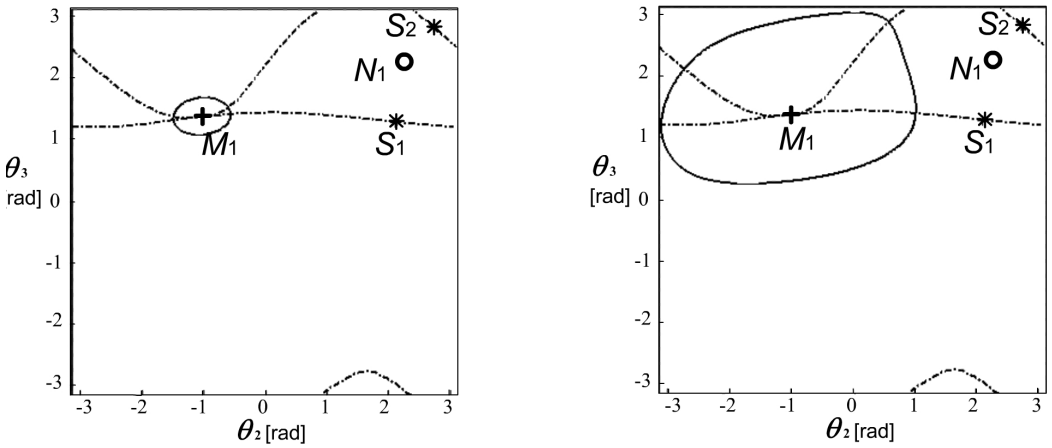


Figure 3.9: The maximum generates the positive region.

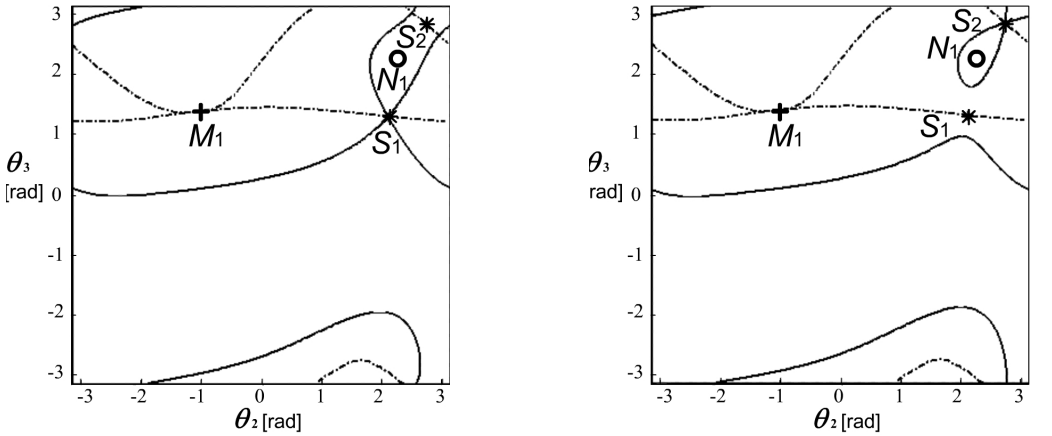


Figure 3.10: The two saddles glue the boundary of the positive region, producing two holes.

maximum M_1 (Figure 3.9), the two saddles glue the boundary of the level set producing two holes (Figure 3.10), and the minimum N_1 closes one of such gaps.

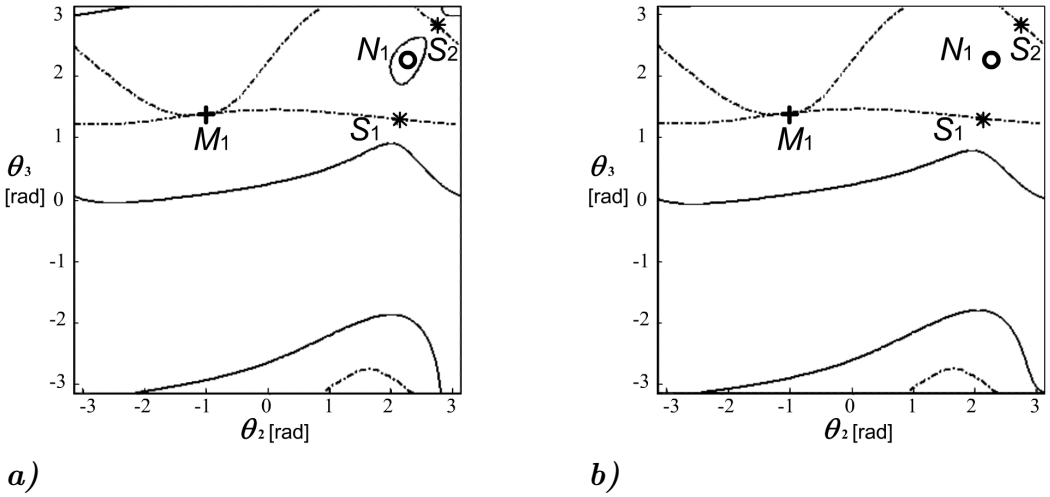


Figure 3.11: The minimum closes one hole.

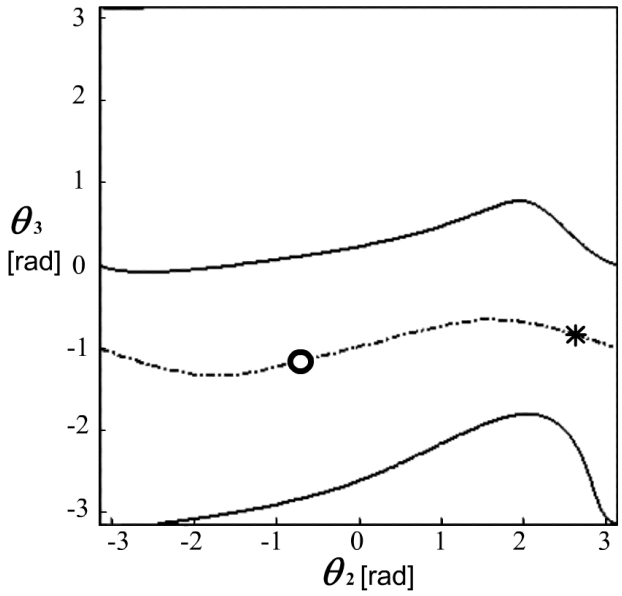


Figure 3.12: Negative critical points of manipulator M_1.

The negative critical points are reported in Figure 3.12, there are only one saddle and one minimum. Thus, since there is only one maximum and

one minimum of J for manipulator M_1, there are one positive region and one negative region in its jointspace. All points with the same sign of the Jacobian determinat can be connected by singularity free paths. In order to find such paths, it is sufficient to follow the steepest ascent paths for positive points, and the steepest descent paths for negative points, which always lead to the absolute maximum or to the absolute minimum respectively.

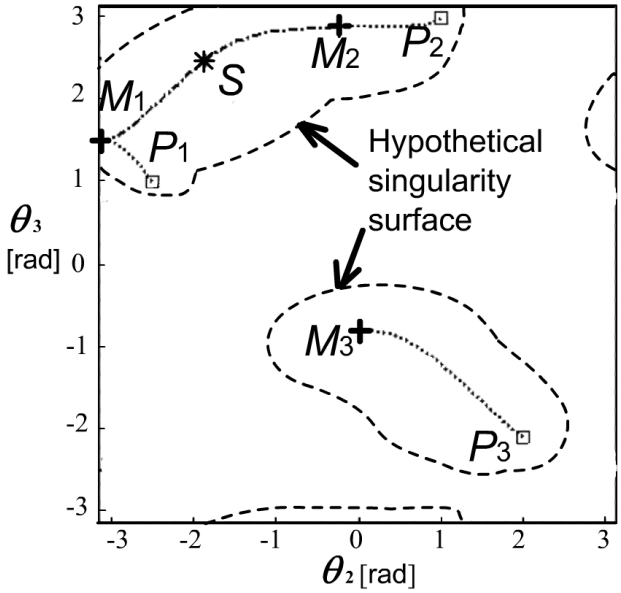


Figure 3.13: Positive critical points of manipulator M_2.

The second example is manipulator M_2, defined by DH-parameters reported in Table 3.3.

a_1	a_2	a_3	d_2	d_3	α_1	α_2
0.4	0.8	1	0.4	0.04	90°	90°

Table 3.3: DH-parameters of manipulator M_2

The positive critical points and the steepest ascent paths connecting them are shown in Figure 3.13. There are three maxima M_1, M_2, and M_3, and one saddle point S. The steepest ascent paths starting from S reach the two maxima M_1 and M_2, which are joined, whereas maximum M_3 generates a separate positive region. Thus we know that there are two positive regions, with no need of plotting them: Figure 3.13 shows a hypothetical singular boundary enclosing the two positive regions.

Consider now the three points $P_1 = (-2.5\text{rad}, 1\text{rad})$, $P_2 = (1\text{rad}, 3\text{rad})$ and $P_3 = (2\text{rad}, -2.1\text{rad})$, where J is positive. The steepest ascent paths starting from P_1, P_2, and P_3 lead to the maxima M_1, M_2, and M_3 respectively. The steepest ascent paths and the three points are reported in Figure 3.13. Therefore P_1 and P_2 belong to the same disjoint region, and can be connected by a singularity-free path. A singularity-free path from P_1 to P_2 is obtained by joining the steepest ascent paths from P_1 to M_1, from the saddle S to M_1 and M_2, and from P_2 to M_2.[1] On the other hand, since P_3 belongs to another positive disjoint region, it is not possible to reach P_1 or P_2 starting from P_3 without crossing a singularity.

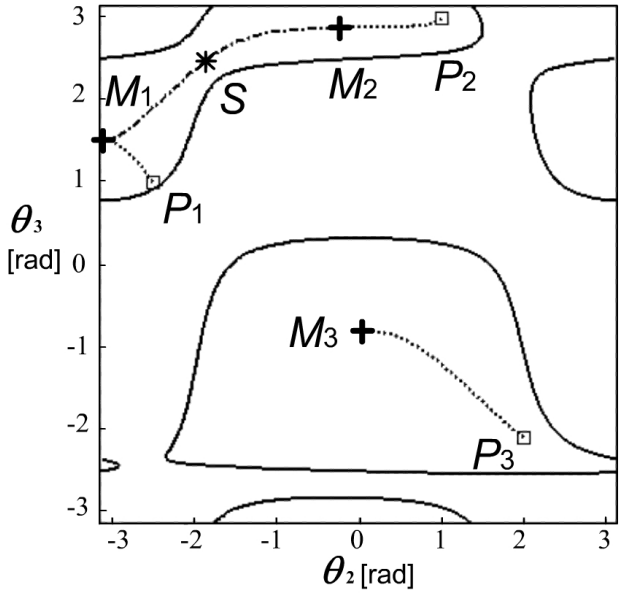


Figure 3.14: Singularity curves of manipulator M_2.

The results just obtained can be easily verified by plotting the singularity curves, as shown in Figure 3.14.

The analysis of the negative critical points shows that there are four minima, all connected through a network of steepest descent paths starting from six negative saddles (Figure 3.15). Therefore there is only one negative disjoint region, and all points with negative Jacobian determinant can

[1]The first joint angle θ_1 is not considered, because, as already discussed, it does not affect the singularity locus. If the values of θ_1 were different at points P_1 and P_2, it would be possible to find a singularity-free path by using the proposed path for the last two angles, and a simple linear interpolation between the two values of θ_1 at P_1 and P_2 for the first.

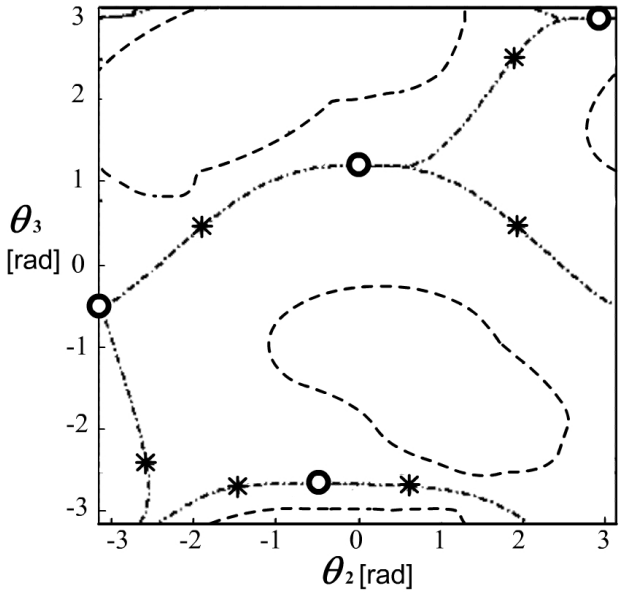


Figure 3.15: Negative critical points of manipulator M_2.

be connected through singularity-free paths. Consider for example the two negative points $P_4 = (-1, 2)$ and $P_5 = (0, -2.8)$. The steepest descent path connecting P_4 and P_5 to two minima can be calculated (dotted lines in Figure 3.15). It is now necessary to find a path connecting the two minima, but in this case there are many possibilities, since the steepest descent paths connecting the saddles to the minima compose a rather complicated network. If paths with no loops are searched for, the following procedure can be used. Starting from one of the two minima, all the "nearby" saddles are found, i.e. the saddles that are directly connected by a steepest descent path to the minima, and the paths connecting the saddles to the minimum. From this saddles, another steepest descent path starts, leading to other minima. If one of these minima is the one to be reached, the path followed is a possible connecting path, if not, the same procedure is repeated for it. If, during these procedure, one of the minima has already been reached before, the path is discarded, because a loop has been completed. In the case of these two points this procedure finds out six different paths without loops. Two of them are shown in Figure 3.16.

The two manipulators M_3 and M_4, whose DH-parameters are reported in Tables 3.4 and 3.5, show the effect of singular critical points on the singularity curves.

The positive critical points and the singularity curves of manipulators

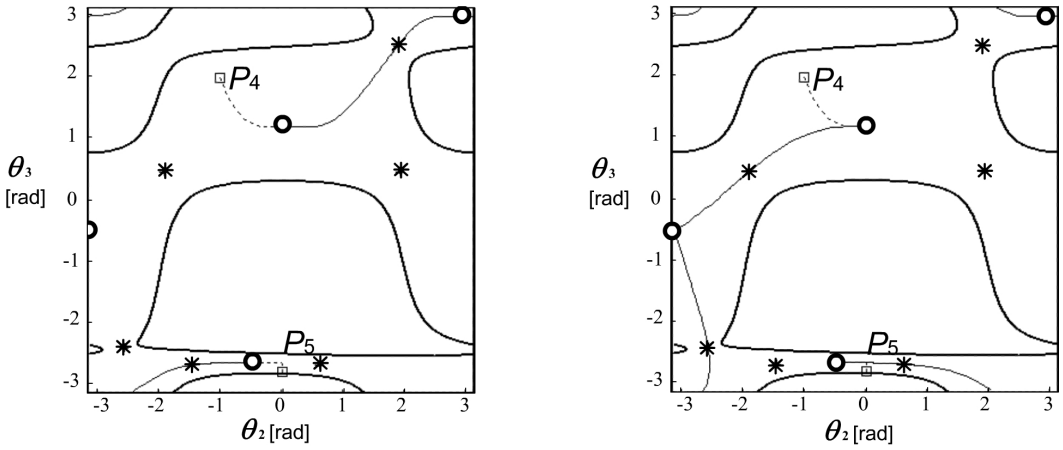


Figure 3.16: Two possible singularity-free paths connecting two negative points in the jointspace.

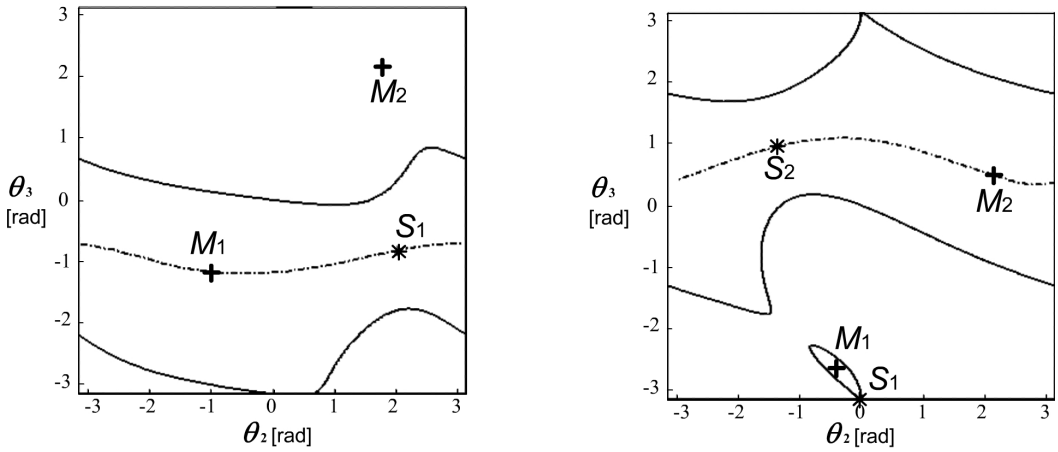
a_1	a_2	a_3	d_2	d_3	α_1	α_2
3.886172281172819	2	1	0	3	80°	−80°

Table 3.4: DH-parameters of manipulator M_3

a_1	a_2	a_3	d_2	d_3	α_1	α_2
3	1	1	0	3	−60°	60°

Table 3.5: DH-parameters of manipulator M_4

M_3 and M_4 are shown in Figure 3.17. In the jointspace of M_3 there is an isolated critical point[2], the maximum M_2, which obviously does not produce the generation of a new positive region. Manipulator M_4 features a singular saddle point S_1, where the boundaries of two positive regions "touch", without joining.



a) *b)*

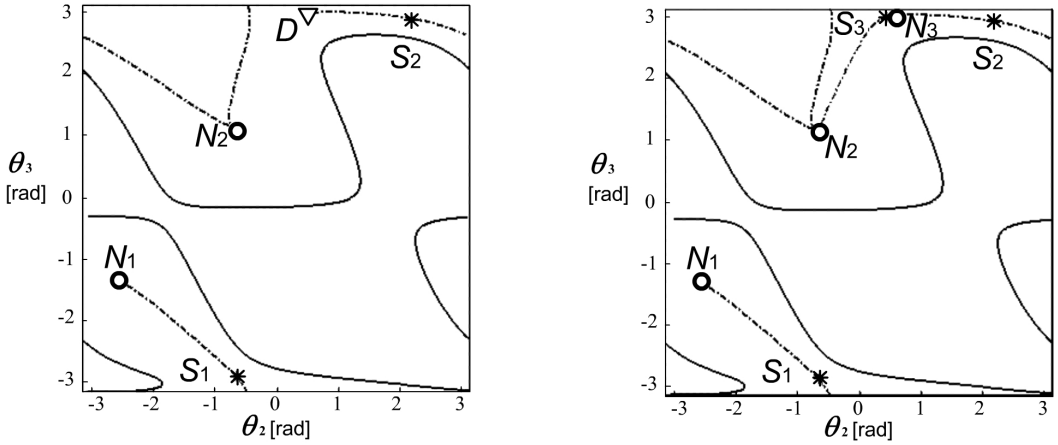
Figure 3.17: Two manipulators M_3 (a) and M_4 (b) with singular critical points.

The proposed method is unable to analyze a manipulator, if J has a degenerate non-singular critical point. Such manipulators are very rare, but not impossible: Figure 3.18a, shows manipulator M_5 with a degenerate critical point. The negative degenerate critical point D is reached by the steepest descent path starting from the negative saddle S_2, but at the same time a steepest descent path should connect it to the minimum N_2. It can be proved[3] that any function can be approximated by a function with no degenerate critical points. In most of cases an approximating function can be searched by means of a slight perturbation of the DH-parameters. This can be done if the manipulator has no singular critical points: in this case it is possible to slightly perturb its geometry, preserving all topological properties of its singularity locus[4]. In Figure 3.18b the manipulator M_5 has been

[2]This example proves that isolated singular points, or one-dimensional singularity loci if angle θ_1 is considered, do exist. By adding the continuity of J for manipulator M_3, this disproves what stated in [26] at page 83.

[3]In [5], corollary 6.8, it is proved that any function on a manifold that can be embedded in a Euclidean space can be uniformly approximated through functions with non-generic critical points

[4]In other words, the manipulator is generic, as defined in [34], and recalled in Section



a) *b)*

Figure 3.18: (a) A manipulator M_5 with a degenerate nonsingular critical point D; (b) a manipulator M_6 without degenerate critical points obtained by perturbing M_5

slightly perturbed, obtaining manipulator M_6: the degenerate critical point disappears, splitting into the minimum N_3 and the saddle S_3, nevertheless the topology of the singularity locus remains almost the same. The DH-parameters of manipulators M_5 and M_6 are reported in Table 3.6 and Table 3.7.

a_1	a_2	a_3	d_2	d_3	α_1	α_2
1/3	5/6	1	0	0.5	100°	−29.81375632659814°

Table 3.6: DH-parameters of manipulator M_5

a_1	a_2	a_3	d_2	d_3	α_1	α_2
2/3	5/6	1	0	0.5	100°	−30°

Table 3.7: DH-parameters of manipulator M_6

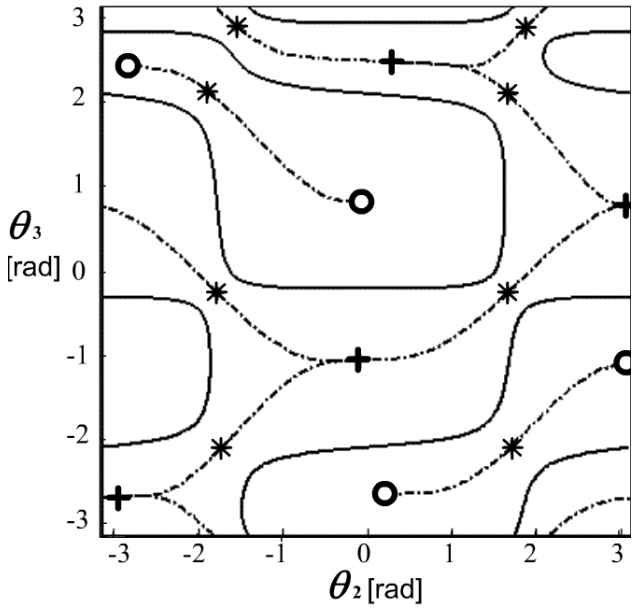


Figure 3.19: A manipulator with 16 critical points of J, DH-parameters $a_1 = 1/12$, $a_2 = 1/2$, $a_3 = 1$, $d_2 = 2/15$, $d_3 = 1/12$, $\alpha_1 = 90°$, and $\alpha_2 = -90°$.

3.1.5 Upper-bound for number of singularity-free regions

An upper-bound can be found for the number of singularity-free disjoint regions of 3R manipulators, by exploiting the maximum number of critical points of the function J upon the jointspace. This upper-bound will be searched for manipulators with nondegenerate critical points only. Section 3.1.3 has proved that there can be at most 16 nondegenrate critical points of J. Figure 3.19 shows the singularity curves of a 3R manipulators with 16 critical points of J.

First, the disjoint regions where J is positive are considered, together with its singular boundary. Each of such positive regions is a compact set, and J is continuous, therefore each positive region must contain at least a maximum and a minimum of J, by virtue of Weierstraß theorem. The minimum is located on the singular boundary, where $J = 0$, and in general is no critical point of J on the torus. On the other hand, the maximum must be located inside the positive region, and must be one of the 16 critical points of J on the torus. The torus is a compact set, thus there must be also a minimum

3.1.6.

of J on the torus, which must be a critical point of J. Thus there can be at most 15 maxima of J on the torus, i.e. at most 15 positive regions.

Suppose that there were 15 maxima and one minimum of J upon the torus. According to the hints of Morse theory recalled in Chapter 1, the two torus would be homotopy equivalent to a CW-complex composed of 15 0-cells, and one 2-cell. This is impossible, because the torus would be composed of 15 disjoint regions, according to Theorem 2.3, whereas the torus is a connected manifold. By virtue of Theorem 2.4, a 1-cell is always needed to attach two disjoint regions, therefore, if there existed 15 maxima, there should also be 14 positive or negative saddles on the torus, in order to connect the 15 regions generated by the maxima, which is still impossible, because it would exceed the maximum number of 16 nondegenerate critical points. This leads to the conclusion that there can be at most 8 maxima on the jointspace, because there must also be 7 saddles, in order for the disjoint regions generated by the maxima to be joined.

When a 1-cell is attached to two 0-cells in such a way that the two 0-cells are joined, the two 0-cells are identified with the boundary of the 1-cell, therefore the result is again a 1-cell. All cell are contractible to a single point, therefore, when two 0-cells are attached through a 1-cell, the result is one single 0-cell. If there where 8 positive maxima and 7 saddles, the 8 maxima would generate 8 0-cells, that would be progressively attached in pairs by the seven saddles, leading to one single 0-cell. This would imply that the torus would be homotopy equivalent to a CW-complex composed of one 0-cell and one 2-cell, i.e. a sphere, which is not true.

The Betti numbers of the torus are $1, 2, 1, 0, 0, ...$, as reported in Table 1.1, therefore, according to Morse inequalities, recalled in Section 1.3.5, at least two 1-cells are needed to build a two-dimensional torus. There can be at most 7 maxima, 6 saddles to join them into one single connected piece, 2 saddles to generate the two holes of the torus, and 1 minimum to close the surface.

This implies that there can be no more than seven singularity-free disjoint regions with the same sign of the Jacobian determinant, if there are only non-degenerate critical points. If there are seven positive singularity-free regions, there can be at most one negative singularity-free region, because there can be at most one minimum. For each negative singularity-free region that is added, one minimum must replace a maximum, and the saddle which is not needed to join the removed positive region is needed to join the new negative one. Therefore there cannot be more than 8 singularity-free regions at all, if there are only nondegenerate critical points.

3.1.6 Homotopy classes

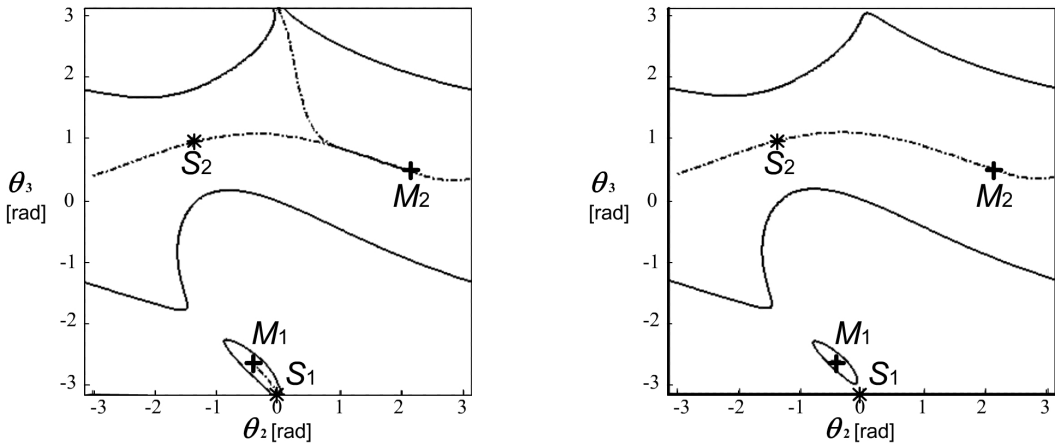


Figure 3.20: Two manipulators obtained through small perturbations of manipulator M_4 with a singular saddle point

Paï and Leu introduced in [34] the important notion of generic manipulator. Generic manipulators have the property that the topological features of their singularity surfaces are preserved under small perturbations of the DH-parameters.[5] They proved that a 3R manipulator is generic if and only if there exist no points in the jointspace where

$$\begin{aligned} J &= 0 \\ \nabla J &= \mathbf{0} \end{aligned} \tag{3.15}$$

In other words, a manipulator is generic if all its critical points are nonsingular. A 3R manipulator is said nongeneric otherwise. The gradient of J is normal to any level-set of the function J. If a 3R manipulator is generic, the normal vector to the singularity curve, which is a level-set of the function J, is always defined. Therefore also the tangent to the singularity curve is defined, i.e. the singularity curve is smooth. Neither corners nor self-intersections are allowed for the singularity curve of a generic 3R manipulator. An example of nongeneric manipulator is manipulator M_4, whose DH-parameters are reported in Table 3.4 and whose singularity curves are depicted in Figure 3.17. M_4 features a self-intersection of the singularity curves, which are not

[5]The definition of genericity is based upon the concept of transversal manifolds, whose intersections are stable under small perturbations of the manifolds themselves. For further details see [36] and [34].

smooth. If a small perturbation (± 0.1 degrees) is applied to the value of DH-parameter α_1, M_4 turns into a generic manipulator, but the topological properties of its singularity locus change. With reference to Figure 3.20, if the singular saddle becomes negative, a bubble detaches from the singularity curve, generating a positive disjoint region, whereas if the singular saddle becomes positive, the bubble is glued to the rest of the singularity curve, and the positive disjoint region disappears.

In [26], Burdick proposed a classification method for generic 3R manipulators, based on the ensuing concepts. A closed loop on the 2-torus $\mathscr{J}$ representing the jointspace can be defined as a univariate function $f : [a, b] \rightarrow \mathscr{J}$ such that $f(a)$ equals $f(b)$. The set of closed loops can the be subdivided in equivalent classes of homotopic loops, according to Definition 1.4.

The singularity curve of a generic 3R manipulator can be subdivided into disjoint regions, according to Definition 1.14. The disjoint regions composing the singularity locus were named *branches* in [26]. Since the branches composing the singularity curve of generic 3R manipulators are always closed, they can be considered as the image of closed loops on the jointspace. Two generic 3R manipulators are then said homotopic if their singularity curves are homotopic. This definition is the base of the classification proposed in [26]: homotopy is an equivalence relation and subdivides the set of generic 3R manipulators into different classes of homotopic manipulators. Non-generic manipulators lie on the boundary of such classes in the space of all 3R manipulators, as the example of Figure 3.20 shows.

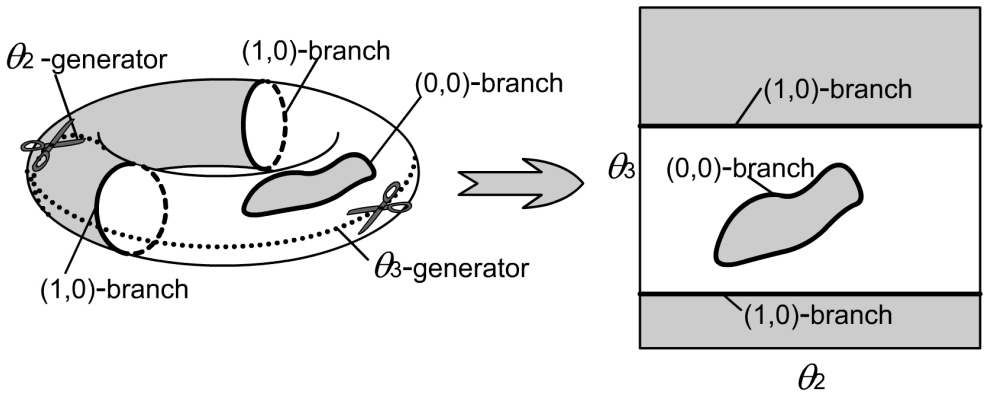


Figure 3.21: A 1(0,0)+2(1,0) manipulator.

It can be proved that two closed loops on a torus, and therefore two branches on the jointspace, are homotopic if and only if they wrap the same number of times around the two generators of the torus. Each branch can

be identified through a couple of integers (n_2, n_3): n_2 (respectively n_3) represents the number of times the branch encircles the torus along the θ_2- (respectively θ_3-) generator. In Figure 3.21 an example is shown: on the left the singularity curve is represented on the torus, in order to show how the branches encircle it, while on the right the singularity curve is represented on the square obtained by cutting the torus along its generators. In Figure 3.21 the singularity curve is composed of two branches of homotopy class (1,0), that encircle the torus once along θ_2-generator, and one (0,0) branch, that encircles no generators at all. Thus the manipulator of Figure 3.21 belongs to the homotopy class 1(0,0)+2(1,0).

Only quaternary manipulators, which have four inverse kinematic solutions at some positions of the end-effector in the workspace, were considered in [35], where it was stated that only eight homotopy classes of quaternary generic 3R manipulators can exist, namely 1(0,0), 2(0,0), 1(0,0)+2(1,0), 2(1,0), 4(1,0), 2(0,1), 2(1,1), and 2(2,1). In [37] a manipulator has been presented that belongs to a new homotopy class. Such a manipulator will be hereafter presented and discussed.

In Figure 3.22 the three-dimensional kinematic sketch of the manipulator M_7, defined by the set of DH-parameters reported in Table 3.8, is shown. The center of the sphere represents the end-effector, i.e. the point to be positioned in the workspace.

a_1	a_2	a_3	d_2	d_3	α_1	α_2
1.178	0.339	1	0.32	0.67	1.55rad	-1.124rad

Table 3.8: DH-parameters of manipulator M_7

M_7 is a generic 3R manipulator, and Figure 3.23a shows the singularity curve in the jointspace of M_7. It can be easily verified that M_7 belongs to the homotopy class 2(0,0)+2(1,0) which is not included among the eight classes enumerated in [35].

Figure 3.23b shows the workspace, with the same representation discussed in Section 3.1.1. The cross in Figure 3.23b indicates the position $P = (0.26, 0, 0)$ of the end-effector in the workspace. The four crosses in Figure 3.23a represent the projections on a $\theta_2\theta_3$-plane of the inverse kinematic solutions pertaining to P. The numbers in Figure 3.23b indicate the number of inverse kinematic solutions inside each region separated by the image of the singularity surface. The region with four inverse kinematic solutions that does not contain P is very narrow, yet it has non-zero volume. The four branches in the jointspace have been identified trough letters b'_1, b'_2, b'_3,

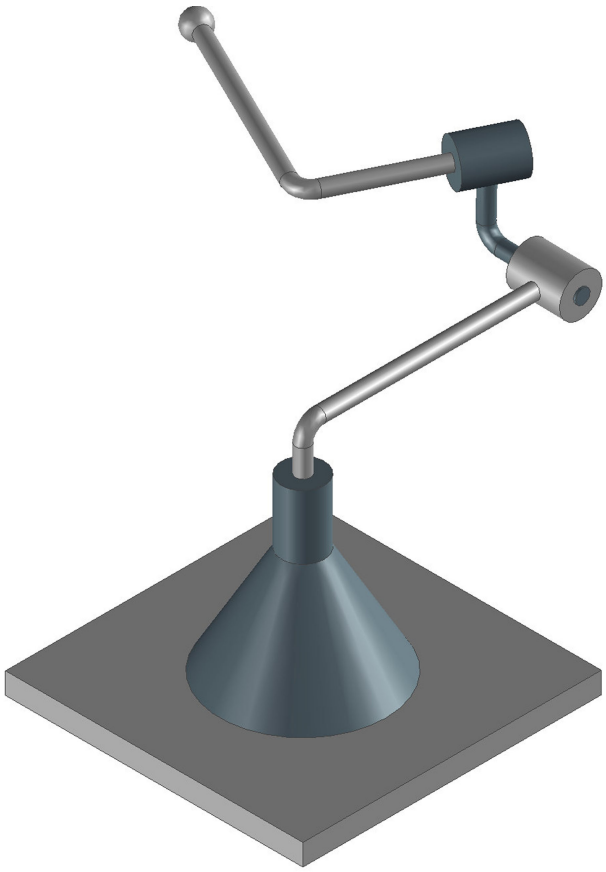


Figure 3.22: Manipulator M_7.

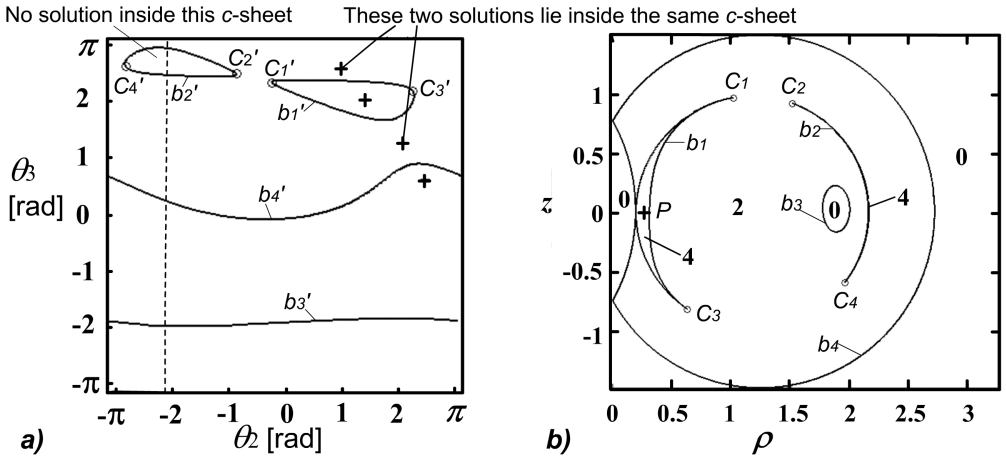


Figure 3.23: M_7 is a 2(0,0)+2(1,0) manipulator.

b'_4 in Figure 3.23a, while the respective images of the four branches in the workspace have been named b_1, b_2, b_3, b_4 in Figure 3.23b.

The manipulator M_7 features two properties that no manipulator belonging to the eight classes already found in [35] exhibits. First, although M_7 has four singularity-free disjoint regions, there is a different number of inverse kinematic solutions in each region, also for those points of the workspace reachable through four postures. On the contrary, the generic 3R manipulators with four disjoint regions found in [35] have one solution per singularity-free disjoint region.

In particular, one singularity-free disjoint region contains two inverse kinematic solution, while another contains none. This leads to the second property: the manipulator is capable of non-singular posture changes, because two postures lie in the same disjoint region, and is therefore cuspidal, as proved in [27]. Indeed, in the four points of the jointspace C'_1, C'_2, C'_3, C'_4, marked with circles in Figure 3.23a, there are three coincident inverse kinematic solutions, therefore their images in the workspace (C_1, C_2, C_3, and C_4 respectively) are cusp points (see [27]). Thus, M_7 disproves Theorem 9 of [35], which states that a generic quaternary 3R manipulator is non-cuspidal if and only if it has four disjoint regions.

The manipulator [35] has been found through a random automatic search. The manipulator [35] has three singularity-free disjoint regions in its jointspace, where the Jacobian determinant J has the same sign (the two regions enclosed by the two (0,0)-branches and the band-shaped region enclosed by the two (1,0)-branches that does not contain the (0,0)-branches), while all the already known generic manipulators have at most two. The procedure developed in this work has been used to analyze a large number of randomly generated manipulators, until a manipulator with more than two singularity-free disjoint regions where the sign of the Jacobian determinant is the same has been found.

Theorem 7 of [35] states that a (0,0)-branch may either appear alone, together with an additional (0,0)-branch, or with two (1,0)-branches. By virtue of Theorem 7, the existence of manipulators like M_7 was excluded in [35]. To prove Theorem 7, the fact that there can be at most four intersections between a θ_3-generator and the singularity curve was exploited, for Equation (3.9) can be written as a fourth order polynomial in the tangent of $\theta_3/2$. Analogously, it can be proved that there can be at most two intersections between a θ_2-generator and the singularity curve. But Figure 3.23a shows that 2(0,0)-branches and 2(1,0)-branches can coexist without exceeding four intersections with any θ_3-generator and two intersections with any θ_2-generator.

Theoretically, any number of (0,0)-branches could coexist, if they were

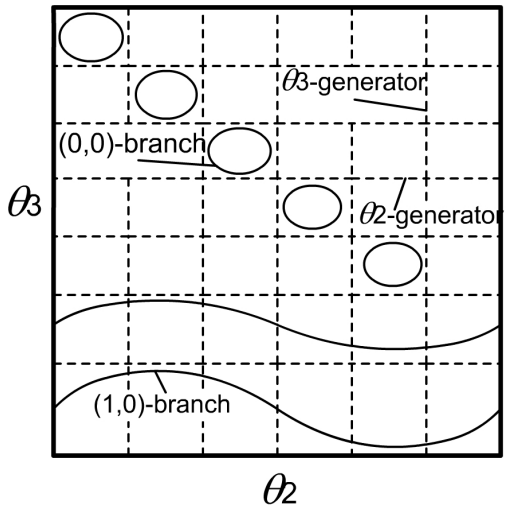


Figure 3.24: Any number of (0,0)-branches can coexist with two (1,0) branches.

only limited by the number of intersections with the generators of the torus, as shown in Figure 3.24. It will be hereafter proved that the singularity curve of a generic 3R manipulator can be composed of at most four branches of any kind, thus Figure 3.24 cannot represent the singularity curve of a generic 3R manipulator.

Wenger [35] proved that only five homotopy types of branches are allowed for generic 3R manipulators, namely (0,0), (1,0), (0,1), (1,1) and (2,1). He proved also that if there is a (0,1)-, a (1,1)- or a (2,1)-branch then there must be only two branches of the same homotopy type. The problem is therefore to establish how many (0,0)- and (1,0)-branches can coexist.

Both (0,0)- and (1,0)-branches are compact subsets of the torus. It is easy to see that there always exists a θ_2-generator that does not intersect any (0,0)- or (1,0)-branches, for any generic manipulator. This stems from the fact that any θ_2-generator can intersect the singularity curve at most twice, and that the singularity curves of generic manipulators are always smooth. It will be assumed that the origin of the angle θ_3 is such that the equation $\theta_3 = \pi$ defines a θ_2-generator that does not intersect any (0,0)- or (1,0)-branches. It is now possible to define a function $f : \mathscr{J} \rightarrow \mathscr{R}$ that associates to each point of the torus its θ_3-coordinate. The function f is not continuous upon the torus, but, according to the just defined origin for θ_3, the restriction of f upon any (0,0)- or (1,0)-branch is always continuous. By virtue of Weierstraß theorem, there must be at least a maximum and a

minimum of f upon any (0,0)- or (1,0)-branch. The problem is to establish how many critical points of the function $f(\theta_2, \theta_3) = \theta_3$ can exist upon the constraint $J(\theta_2, \theta_3) = 0$.

f and its constraint are smooth, because the manipulator is generic, thus the critical points of f can be found by means of Lagrange's multipliers. The critical points satisfy the condition:

$$\begin{cases} J = 0 \\ \lambda \partial J/\partial \theta_2 = 0 \\ \lambda \partial J/\partial \theta_3 = 1 \end{cases} \tag{3.16}$$

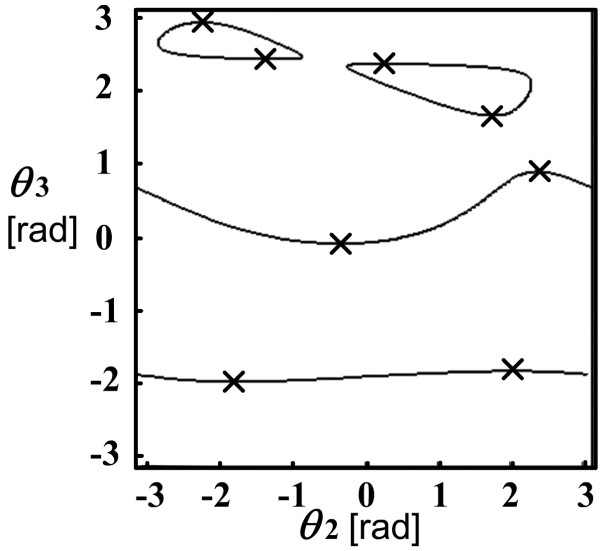


Figure 3.25: The tangent of the singularity curve is parallel to θ_2-generator at solutions of Equation (3.16).

A geometric interpretation can be given to Equation (3.16): its solutions are points of the singularity curve where the tangent is parallel to the θ_2-generator. In Figure 3.25 the solutions to Equation (3.16) are marked: it can be checked that each (0,0)- and (1,0)-branch must contain at least two solutions of Equation (3.16).

If Lagrange's multiplier λ vanished, the third equation of Equation (3.16) could not be satisfied, thus λ can be simplified in the second equation. θ_2 can be eliminated from the first two equations, yielding the ensuing equation in θ_3:

$$V_1(\theta_3)^2 + V_2(\theta_3)^2 - V_3(\theta_3)^2 = 0 \tag{3.17}$$

Equation (3.17) can be written as an eighth order polynomial in the tangent of $\theta_3/2$, thus there can be 2,4,6,8 or infinite real solutions to Equation (3.17), counting each solution with its multiplicity. Suppose there exist infinite real solutions to Equation (3.17). If V_3 vanished at each of the infinite real solutions, also V_1 and V_2 would have to, and the manipulator would be nongeneric, for J would be constantly equal to zero. Thus there must exist one of the infinite solutions to Equation (3.17), θ_3^*, such that $V_3(\theta_3^*)$ does not vanish. The first two equations of Equation (3.17) yield the ensuing equations:

$$\begin{cases} -V_3(\theta_3^*)\cos\theta_2^* &= V_1(\theta_3^*) \\ -V_3(\theta_3^*)\sin\theta_2^* &= V_2(\theta_3^*) \end{cases} \tag{3.18}$$

where θ_2^* is the value of θ_2 that satisfies Equation (3.17) along with θ_3^*. If there are infinite solutions to Equation (3.17), the left hand side of Equation (3.17) must be identically equal to zero, and its derivative must be identically equal to zero too. By substituting Equation (3.18) into the derivative of Equation (3.17), one can obtain:

$$-V_3(\theta_3^*)\left(\left.\frac{\partial V_1}{\partial\theta_3}\right|_{\theta_3^*}\cos\theta_2^* + \left.\frac{\partial V_2}{\partial\theta_3}\right|_{\theta_3^*}\sin\theta_2^* + \left.\frac{\partial V_3}{\partial\theta_3}\right|_{\theta_3^*}\right) =$$

$$= -V_3(\theta_3^*)\left.\frac{\partial J}{\partial\theta_3}\right|_{(\theta_2^*,\theta_3^*)} = 0 \tag{3.19}$$

where $\partial/\partial\theta_3|_P$ is the derivative with respect to θ_3, computed at the point P.

Equations 3.19 and 3.16 imply that both J and its gradient vanish at the point (θ_2^*, θ_3^*), i.e. the manipulator is nongeneric. There can be at most eight points satisfying Equation (3.16), because there cannot exist infinite real solutions to Equation (3.17) without contradicting the hypothesis of genericity of the manipulator. Therefore, at most four of such branches can coexist, because each (0,0)- and (1,0)-branch must contain at least two points satisfying Equation (3.16). This also proves that generic manipulators cannot have more than four singularity-free disjoint regions with the same sign of the Jacobian determinant, and, by analyzing all classes that have not yet been excluded, not more than five singularity-free disjoint regions at all.

Wenger proved that (1,0)-branches must appear in pair or sets of four (theorem 6 of [35]), thus all the classes containing (1,0)-branches have already been found. Two other possibilities remain, namely 3(0,0) and 4(0,0), but the author has been so far unable to prove that they are to be excluded.

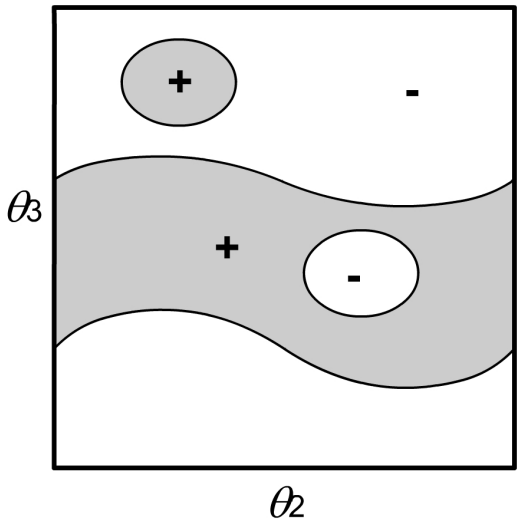


Figure 3.26: A manipulator of class 2(0,0)+2(1,0) can have two positive and two negative disjoint regions.

Manipulators of class 2(0,0)+2(1,0) may also have two positive and two negative singularity-free disjoint regions, unlike M_7 which possesses three disjoint regions with the same sign of the Jacobian determinant. None of such manipulators has been found so far, but Figure 3.26 shows a possible example.

On the other hand, 3(0,0) and 4(0,0) manipulators, if any existed, should possess three and four disjoint regions with the same sign of the Jacobian determinant. Indeed, each (0,0)-branch cuts two regions on the torus: one is homotopically contractible to a point, the other not. Inside the region contractible to a point there can be no other singular points, lest the two allowed intersections with a θ_2-generator are exceeded. Therefore, any other (0,0)-branch must be inside the region which is not contractible to a point, and such region is cut by any other (0,0)-branch into two more regions, one contractible, and one not. If there are only n (0,0)-branches, they cut the torus into n contractible regions, all surrounded by one region, not contractible. Therefore, all n contractible regions cut by (0,0)-branches must have the opposite sign with respect to the noncontractible region, as shown in Figure 3.27.

Several hundred thousands randomly generated 3R manipulators have been analyzed, but the only manipulators with more than two singularity-free disjoint regions with the same sign of the Jacobian determinant that have been found are like M_7.

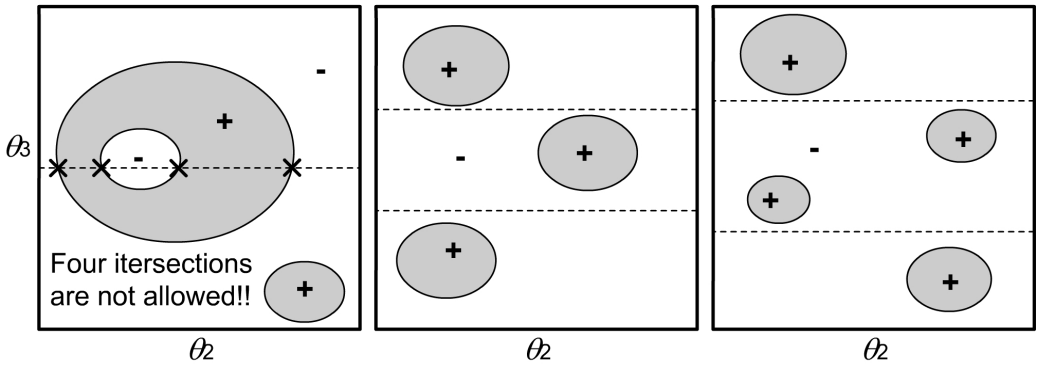


Figure 3.27: Manipulators of class 3(0,0) or 4(0,0) should have three or four singolarity-free disjoint regions with the same sign.

3.2 6R Manipulators

3.2.1 Singularity locus

6R manipulators are analogous to 3R manipulators, but they have six revolute joints, and their aim is to position a rigid body, the end effector, in the three-dimensional space. Figure 3.28 shows an example of a general 6R manipulator.

Analogously to 3R manipulators, the jointspace is homeomorphic to a 6-torus, obtained by the Cartesian product of the six one-dimensional circles $\mathscr{S}$, representing the positions of the six revolute joints, and parameterized through six joint angles θ_i. However, unlike 3R manipulators, the workspace is not in general homeomorphic to a subset of a Euclidean space, therefore it is impossible to find a global coordinate system for the workspace. Indeed, the workspace is a subset of the manifold containing all possible positions of a rigid body in the three-dimensional Euclidean space. This manifold can be represented as the Cartesian product of the three-dimensional Euclidean space, containing the position of one point of the end-effector, and the three-dimensional projective space, containing all possible orientations of the end-effector.

In order to define the singularity-locus as the level set of a smooth function on the jointspace, it is not convenient to differentiate the function f, that links each point in the jointspace with the corresponding position of the end-effector in the workspace. In fact, in order to write such function, a parameterization of the position of the end-effector would be needed, but any possible parameterization of the orientation of a rigid body either con-

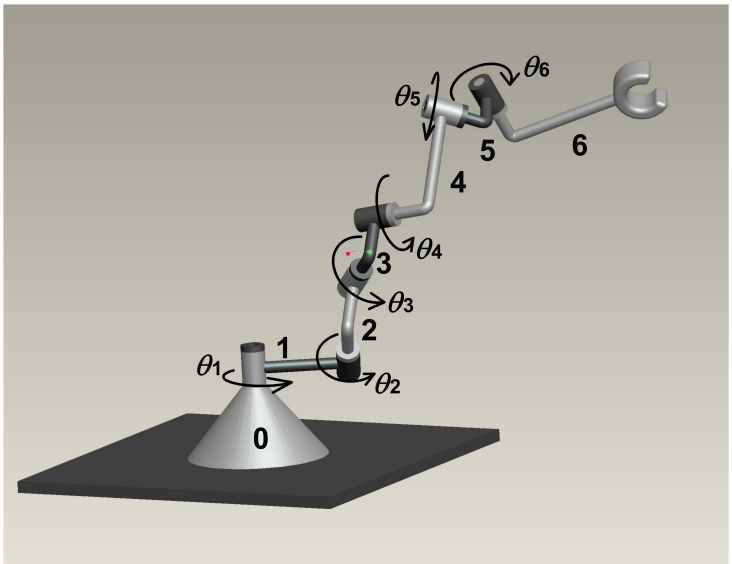


Figure 3.28: A 6R manipulator. The rigid links are numbered from 0 to 6.

tains itself singularities, which are not singular positions of the manipulator, or needs redundant parameters tied by means of proper equations, which increases the complexity of the problem. This is a direct consequence of the fact that the manifold containing all orientations of a rigid body and the three-dimensional Euclidean space are not homeomorphic, and therefore no global system of coordinates can be found.

Probably, the smartest way for computing the Jacobian matrix of a 6R manipulator is the screw-based Jacobian, proposed by Waldron et al.(see [38], [39], [40] and [24], pp.193-195). This geometrical method will be hereafter recalled,without mentioning explicitly the idea of screws. For more details about screws, see for example [41].

If a three-dimensional Cartesian reference frame (x, y, z) is chosen, then a local coordinate system for the position of a rigid body can be defined, in order to identify all possible virtual displacements of the rigid body from a given position. We can consider the point of the rigid body that in the given position coincides with the origin of the reference frame, and identify the displaced position of such point through a vector $(\partial x, \partial y, \partial z)$. Any virtual variation of the orientation of the rigid body can be identified through the vector $(\partial\xi, \partial\eta, \partial\zeta)$, containing the virtual rotations around the axes x, y, and z of the chosen reference frame. Therefore, once a proper reference frame is chosen, the vector $(\partial x, \partial y, \partial z, \partial\xi, \partial\eta, \partial\zeta)$ completely defines a virtual displacement of the end-effector of a 6R manipulator.

In this local coordinate frame, it is possible to write the virtual displacement of the end-effector in function of the six virtual variations of the six joint angles θ_i:

$$\begin{pmatrix} \partial x \\ \partial y \\ \partial z \\ \partial \xi \\ \partial \eta \\ \partial \zeta \end{pmatrix} = \mathbf{D} \begin{pmatrix} \partial \theta_1 \\ \partial \theta_2 \\ \partial \theta_3 \\ \partial \theta_4 \\ \partial \theta_5 \\ \partial \theta_6 \end{pmatrix} \tag{3.20}$$

where $\mathbf{D}$ is the Jacobian matrix of the manipulator. The virtual displacements of the end-effector can also be viewed as the sum of the displacements produced by the variation of each joint angle alone, with all other actuators locked. This is equivalent to splitting the vector-matrix product on the right-hand side of Equation (3.20) as a linear combination of the columns of $\mathbf{D}$ through the coefficients $\partial\theta_i$, i.e.:

$$\begin{pmatrix} \partial x \\ \partial y \\ \partial z \\ \partial \xi \\ \partial \eta \\ \partial \zeta \end{pmatrix} = \sum_{i=1}^{6} \mathbf{S}_i \partial \theta_i \tag{3.21}$$

where $\mathbf{S}_i$ are the columns of the Jacobian matrix.

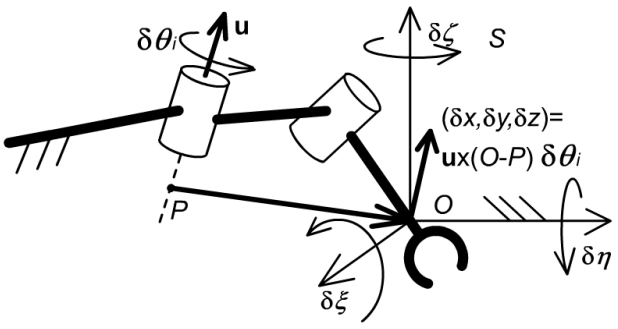


Figure 3.29: Determination of the i^{th}-column of the Jacobian matrix.

In order to find $\mathbf{S}_i$, we suppose that $\partial\theta_j = 0$, for $j \neq i$, i.e. all actuators but the i^{th} are locked. With reference to Figure 3.29, by choosing a generic

fixed reference frame S, it is easy to see that:

$$\begin{pmatrix} \partial x \\ \partial y \\ \partial z \\ \partial \xi \\ \partial \eta \\ \partial \zeta \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ \mathbf{u} \times (O - P) \end{pmatrix} \partial\theta_i = \mathbf{S}_i \partial\theta_i \tag{3.22}$$

where $\mathbf{u}$ is a three-dimensional vector containing the components of the unit vector directed along the axis of the i^{th} revolute joint, P is a generic point on the axis of the i^{th} revolute joint, and O is the origin of the frame S.

The optimal choice for the reference frame S, is a fixed frame which coincides in the actual configuration of the manipulator with Denavit-Hartenberg frame (see Figure 3.3) attached to one of the middle links of the manipulator, such as link number 4 of Figure 3.28. This frame will be named S_4

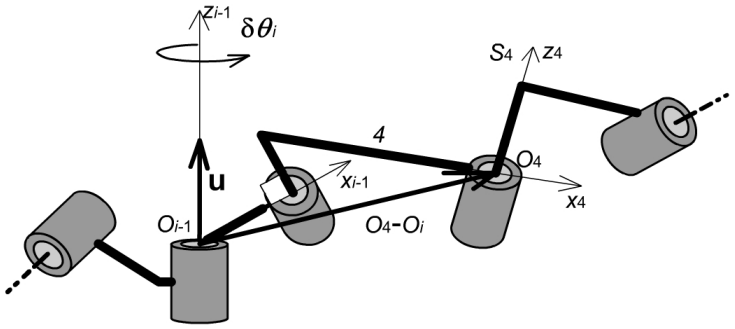


Figure 3.30: The DH-frame attached to link 4 is the optimal choice.

Referring to Figure 3.30, the vector $\mathbf{u}$, if expressed in the DH reference frame attached to the $(i-1)^{\text{th}}$ link, is equal to $(0,0,1)^{\text{T}}$. Therefore, the components $\mathbf{u}_{i-1}$ of $\mathbf{u}$ in S_4 can be obtained as:

$$\begin{pmatrix} \mathbf{u}_{i-1} \\ 0 \end{pmatrix} = \mathbf{A}_{4,(i-1)} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \tag{3.23}$$

where $\mathbf{A}_{4,(i-1)}$ is the 4×4 homogeneous transformation matrix between the DH-frames attached to links $(i-1)$ and 4. $\mathbf{A}_{4,(i-1)}$ can be obtained as:

$$\mathbf{A}_{4,(i-1)} \quad = \quad \mathbf{A}_{4,5} \ldots \mathbf{A}_{(i-3),(i-2)} \mathbf{A}_{(i-2),(i-1)} \quad \text{if } (i-1) > 4$$

$$\begin{aligned} \mathbf{A}_{4,i} &= \mathbf{A}_{3,4}^{-1} \dots \mathbf{A}_{i,(i+1)} \mathbf{A}_{(i-1),i}^{-1} \qquad \text{if } i < 4 \\ \mathbf{A}_{4,4} &= \mathbf{I} \end{aligned} \tag{3.24}$$

where matrices $\mathbf{A}_{(i-1),i}$ have already been defined in Equation (3.3). Analogously, the vector $(O_4 - O_{i-1})$ in frame S_4 can be obtained as

$$\begin{pmatrix} (O_4 - O_{i-1}) \\ 1 \end{pmatrix} = \mathbf{A}_{4,i-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.25}$$

with these two vectors, the i^{th} column of the Jacobian matrix $\mathbf{S}_i$ is expressed in the reference frame S_4 as

$$\mathbf{S}_i = \begin{pmatrix} \mathbf{u}_{i-1} \\ \mathbf{u}_{i-1} \times (O_4 - O_{i-1}) \end{pmatrix} \tag{3.26}$$

By composing the six column obtained in this way, the Jacobian matrix of a generic 6R manipulator is obtained, and the Jacobian determinant J can be easily calculated.

It can be easily seen (see [10], [11], [12]) that J does not depend on the first and the last joint angles θ_1 and θ_6. Furthermore, the sines and cosines of θ_2 and θ_5 appear with degree one in J, and the sines and cosines of θ_3 and θ_4 appear with degree two. Therefore, the singularity locus depends only upon four angles, and can be viewed as a locus on a 4-torus.

The critical points of J on the 4-torus can be found by imposing that the gradient of J vanishes. Unfortunately, unlike 3R manipulators, no general elimination method has been found to solve this problem. In some particular simple cases, the critical points can be found by means of ad hoc elimination methods, but in the most general cases only homotopy continuation method is able to tackle the problem. In Section 3.2.2 a couple of examples thereof are reported.

3.2.2 Examples

Without loss of generality, we can suppose that the reference frame S_0, attached to the base has the origin on the axis x_1 of the reference frame attached to the first link, and that the reference frame on the end-effector S_6 has the same origin and z-axis as the reference frame attached to link 5. According to these simplifications, fourteen DH-parameters are necessary to identify the geometry of a general 6R manipulator. Indeed, the six coordinate transformations from S_6 to S_0 can be completely defined through the fourteen DH-parameters and the six joint angles summarized in Table 3.9

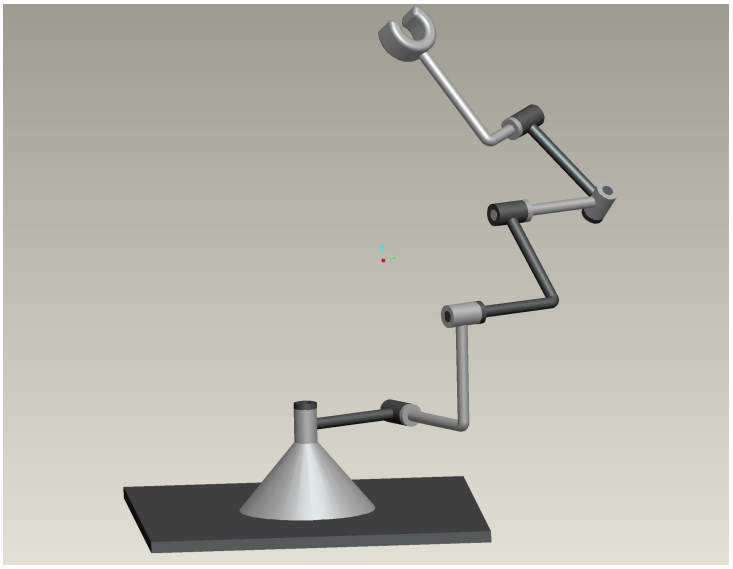


Figure 3.31: 6R manipulator M_8.

Reference frames	DH-parameters	Joint angle
$S_1 \to S_0$	a_1, α_1	θ_1
$S_2 \to S_1$	a_2, d_2, α_2	θ_2
$S_3 \to S_2$	a_3, d_3, α_3	θ_3
$S_4 \to S_3$	a_4, d_4, α_4	θ_4
$S_5 \to S_4$	a_5, d_5, α_5	θ_5
$S_6 \to S_5$		θ_6

Table 3.9: DH-parameters defining the geometry of a 6R manipulator.

The first example is manipulator M_8, defined by DH-parameters reported in Table 3.10 and shown in Figure 3.31.

a_1	a_2	a_3	a_4	a_5	d_2	d_3	d_4	d_5	α_1	α_2	α_3	α_4	α_5
1	1	1	0	1	1	1	1	0	90°	90°	0	90°	90°

Table 3.10: DH-parameters of manipulator M_8

The Jacobian determinant of manipulator M_8 is equal to:

$$\begin{gathered} J = -c_3\,c_4\,s_2\,s_4 - {c_3}^2\,c_4\,s_2\,s_4 - c_4\,s_2\,s_3\,s_4 - c_4\,s_2\,{s_3}^2\,s_4 - c_3\,s_2\,{s_4}^2 + \\ + s_2\,s_3\,{s_4}^2 - c_2\,c_3\,{c_4}^2\,s_5 - c_2\,{c_3}^2\,{c_4}^2\,s_5 - 2\,c_3\,{c_4}^2\,s_2\,s_5 + {c_4}^2\,s_3\,s_5 + \\ + c_2\,{c_4}^2\,s_3\,s_5 + c_3\,{c_4}^2\,s_3\,s_5 + c_2\,c_3\,{c_4}^2\,s_3\,s_5 + 2\,{c_4}^2\,s_2\,s_3\,s_5 - \\ - c_2\,c_3\,{s_4}^2\,s_5 - c_2\,{c_3}^2\,{s_4}^2\,s_5 - 2\,c_3\,s_2\,{s_4}^2\,s_5 + s_3\,{s_4}^2\,s_5 + c_2\,s_3\,{s_4}^2\,s_5 + \\ + c_3\,s_3\,{s_4}^2\,s_5 + c_2\,c_3\,s_3\,{s_4}^2\,s_5 + 2\,s_2\,s_3\,{s_4}^2\,s_5 \end{gathered} \tag{3.27}$$

where c_i and s_i stand for the cosine and sine of θ_i respectively.

By imposing that the gradient of J vanishes, the ensuing four equations are obtained, whose solutions are the critical points of J:

$$\begin{aligned} &- c_2\,c_3 + c_2\,c_3\,{c_4}^2 + c_2\,s_3 - c_2\,{c_4}^2\,s_3 - c_2\,c_4\,s_4 - \\ &- c_2\,c_3\,c_4\,s_4 - c_2\,c_4\,s_3\,s_4 - 2\,c_2\,c_3\,s_5 + c_3\,s_2\,s_5 + \\ &{c_3}^2\,s_2\,s_5 + 2\,c_2\,s_3\,s_5 - s_2\,s_3\,s_5 - c_3\,s_2\,s_3\,s_5 \;=\; 0 \end{aligned} \tag{3.28}$$

$$\begin{aligned} &- c_3\,c_4\,s_2\,s_4 + c_4\,s_2\,s_3\,s_4 + c_3\,s_2\,{s_4}^2 + s_2\,s_3\,{s_4}^2 - \\ &- s_5 - c_2\,s_5 + c_3\,s_5 + c_2\,c_3\,s_5 + 2\,{c_3}^2\,s_5 + 2\,c_2\,{c_3}^2\,s_5 + \\ &+ 2\,c_3\,s_2\,s_5 + c_2\,s_3\,s_5 + 2\,c_2\,c_3\,s_3\,s_5 + 2\,s_2\,s_3\,s_5 \;=\; 0 \end{aligned} \tag{3.29}$$

$$\begin{aligned} &- s_2\,\big(-1 - c_3 + 2\,{c_4}^2 + 2\,c_3\,{c_4}^2 - s_3 + 2\,{c_4}^2\,s_3 + \\ &+ 2\,c_3\,c_4\,s_4 - 2\,c_4\,s_3\,s_4\big) \;=\; 0 \end{aligned} \tag{3.30}$$

$$\begin{aligned} &- c_5\,\big(c_2\,c_3 + c_2\,{c_3}^2 + 2\,c_3\,s_2 - s_3 - c_2\,s_3 - \\ &- c_3\,s_3 - c_2\,c_3\,s_3 - 2\,s_2\,s_3\big) \;=\; 0 \end{aligned} \tag{3.31}$$

The critical points can be found exploiting the fact that the last two equations can be factored. In particular, the ensuing cases have to be considered:

- $s_2 = c_5 = 0$, in this case the last two equations are satisfied, and the values for θ_3 and θ_4 can be found from the first two equations by

replacing the tangents of half angles and by means of Sylvester dialytic elimination method (see Section 5.1);

- $c_5 = 0$, $s_2 \neq 0$, in this case it can be easily seen that J vanishes, therefore none of these points has to be considered;

- $c_5 \neq 0$, $s_2 = 0$, in this case Equation (3.28) and Equation (3.29) are linear in c_2 and s_2, that can be easily eliminated; Equation (3.30) is already free of c_2 and s_2, therefore a set of two equations in the tangents of half θ_3 and θ_4 can be obtained, and solved through Sylvester dialytic elimination method.

- $c_5 \neq 0$, $s_2 \neq 0$, in this case the first equation yields immediately s_5 as a function of the other variables, c_2 and s_2 can be eliminated from the second and the fourth equation, whereas the third equation contains only variables θ_3 and θ_4; in this way two equations in the tangents of half θ_3 and θ_4 can be obtained, and solved as in the previous point.

In this way 128 nonsingular critical points of J on the 4-torus are obtained for M_8. Of these critical points, 8 are positive maxima, 10 are negative minima, 44 are positive 3-saddles and 46 are negative 1-saddles. The remaining 20 critical points shall not be considered, because their index does not produce any generation or attachment of regions.

By analyzing the steepest ascent paths, all maxima turn out to be connected through steepest ascent paths starting from the positive 3-saddles. Analogously, all minima are connected through steepest descent paths starting from the negative 1-saddles. Therefore there is only one positive singularity-free disjoint region and one negative singularity-free disjoint region in the jointspace of M_8

The second example is manipulator M_9, with DH-parameters summarized in Table 3.11, and depicted in Figure 3.32.

a_1	a_2	a_3	a_4	a_5	d_2	d_3	d_4	d_5	α_1	α_2	α_3	α_4	α_5
2	1	3	3	2	2	3	1	4	1.1rad	1.3rad	2.1rad	0.8rad	2.9rad

Table 3.11: DH-parameters of manipulator M_9

For manipulator M_9 no elimination method has been found to determine all critical points. Therefore it is necessary to resort to homotopy continuation method, which is the only numerical method able to find in any case all solutions to a system of polynomial equations. Homotopy continuation method and its application to this problem are better detailed in Section

5.2. In case of manipulator M_9, 1120 finite solutions have been found, 140 of which are real solutions, i.e. critical points of the function J.

Among these critical points there are 6 positive maxima, 6 negative minima, 37 positive 3-saddles, 33 negative 1-saddles, and 58 critical points unable to produce generation or attachment of new regions. Also in the case of M_9, all maxima are connected by a network of steepest ascent paths, and all minima are connected by a network of steepest descent paths, therefore there is only one positive singularity-free region and one negative singularity free region in the jointspace of M_9.

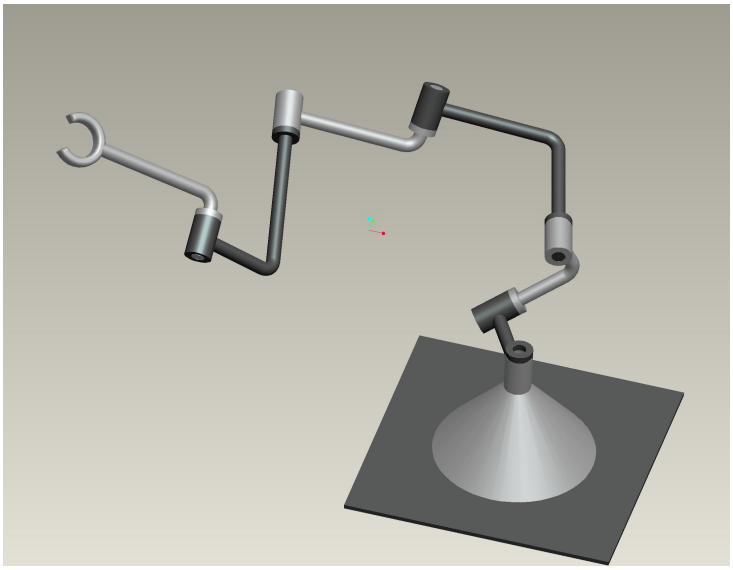


Figure 3.32: 6R manipulator M_9.

Chapter 4

Parallel Manipulators

In this Chapter, some examples of application of the method proposed in Chapter 2 to parallel manipulators will be presented. As discussed in Chapter 2, the serial singularities are harmless for parallel manipulators, and will therefore be ignored. On the other hand, parallel singularities might produce the loss of control of the platform, and therefore are to be avoided. Therefore, the proposed method will be used to find out safe paths free of parallel singularities, and to count and identify the parallel-singularity-free regions in the configuration space of parallel manipulators.

4.1 Spherical Wrists

The analysis of the partition induced by the parallel singularity locus on the workspace of some classes of spherical wrists was presented in [42], and will be discussed in this section.

4.1.1 3UPS Spherical wrists

Spherical wrists are manipulators whose task is to position a rigid body with a fixed point. Thus, by moving the actuators, the orientation of the platform is varied. 3UPS spherical wrists devise a simple parallel architecture to reach this target, which is depicted in Figure 4.1.

A 3UPS spherical wrist is composed of a platform, connected to the base by a spherical joint, and three legs, composed of two rigid bodies connected through a prismatic joint. The three legs are connected to the base and to the platform by means of a spherical joint and a universal joint. The universal joint could possibly be replaced by a spherical joint, but the legs would gain a passive rotational degree of freedom which might be undesired.

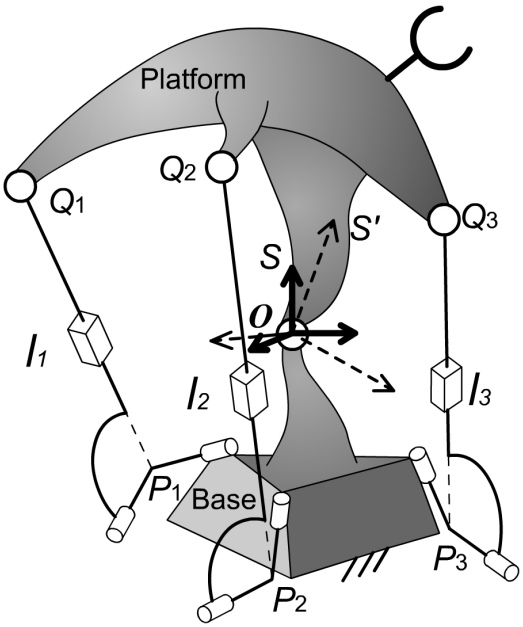


Figure 4.1: A 3UPS spherical wrist.

Let S and S' be two reference frames, attached to the base and to the platform respectively, and with the origin in the center of the spherical joint between the platform and the base. Let the three points P_1, P_2, and P_3, be the centers of the joints between the base and the legs, and the three points Q_1, Q_2, and Q_3 be the centers of the joints between the platform and the legs. The kinematic architecture of any 3UPS wrist is identified by the three vectors $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$, containing the coordinates of points P_1, P_2, and P_3 relative to frame S, along with the three vectors $\mathbf{q}_1$, $\mathbf{q}_2$, and $\mathbf{q}_3$, containing the coordinates of points Q_1, Q_2, and Q_3 relative to frame S'.

3UPS spherical wrists were first studied in [43], where the direct kinematics problem was solved. In [44], an alternative elimination method for the solution of the direct kinematics was proposed. In [45] the singularity locus of 3UPS spherical wrists was studied, and a representation method was proposed. However in the following sections a completely different parameterization and visualization will be adopted.

4.1.2 Configuration space

The workspace of a 3UPS spherical wrists contains all possible orientations of the platform. According to Euler theorem, any possible orientation of a

rigid body with a fixed point can be obtained from a given reference position of the body by performing a rotation of an angle θ about an axis directed as a unit vector $\mathbf{u}$ and containing the fixed point. Therefore, adopted a reference position where frames S' and S coincide, any orientation of the platform can be defined by way of a unit vector $\mathbf{u}$ and an angle θ. The orientations of the platform associated to $\mathbf{u}$ and θ, and to $-\mathbf{u}$ and $-\theta$ always coincide, thus the variation range of the angle θ can be restricted to the interval $[0, \pi]$.

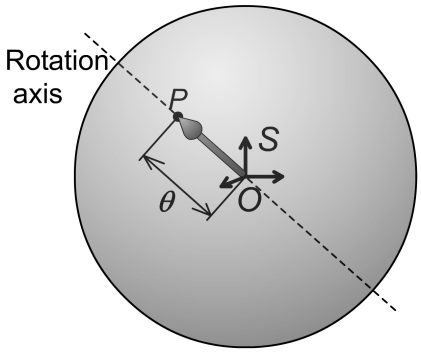


Figure 4.2: Visualization of the workspace of a spherical wrist.

A possible visualization of the workspace can be obtained by considering a ball of radius π in the three-dimensional Euclidean space. With reference to Figure 4.2, every point P inside the ball represents the orientation of the platform identified by the unit vector directed as the position vector of P, and by the angle θ equal to the length of the position vector of P. Thus, every orientation of the platform with an angle θ lesser than π corresponds to only one point inside the ball, whereas any orientation with $\theta = \pi$ is identified by two diametrically opposite points on the boundary sphere of the ball.

Analogous to the torus for the jointspace of serial manipulators, the workspace of a wrist is thus homeomorphic to a ball in the three-dimensional space, where any two diametrically opposite points on the boundary sphere are indeed the same point. For instance, with reference to Figure 4.3, a path starting from a point A to a point B of the ball can exit the ball on a point of the boundary sphere and enter back again through a diametrically opposite point.

It is useful to introduce a more homogeneous parameterization of the orientation of a rigid body, i.e. Euler parameters, which will enable an easier determination of the critical points of the Jacobian determinant. We consider the vector $\mathbf{e}$, containing the four parameters (e_0, e_1, e_2, e_3) defined as follows:

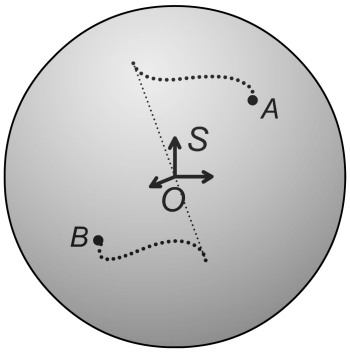


Figure 4.3: Diametrically opposite points on the boundary sphere are identified.

$$\begin{aligned} e_0 &= \cos\frac{\theta}{2} \\ e_1 &= u_1 \sin\frac{\theta}{2} \\ e_2 &= u_2 \sin\frac{\theta}{2} \\ e_3 &= u_3 \sin\frac{\theta}{2} \end{aligned} \tag{4.1}$$

where $\mathbf{u} = (u_1, u_2, u_3)$. It is easy to see that, since $\mathbf{u}$ is a unit vector:

$$e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1 \tag{4.2}$$

Moreover, e_0 is always not lesser that zero, because $0 \leq \theta \leq \pi$, and, if e_0 is equal to zero, the vectors $\mathbf{e}$ and $-\mathbf{e}$ represent the same orientation of the platform. Therefore the workspace is homeomorphic to a half three-dimensional sphere in the four-dimensional Euclidean space, where the diametrically opposite points on the cutting plane are identified.

Eventually, it is evident from Euler parameterization that any any orientation of the platform can be bijectively associated to a direction in $\mathscr{R}^4$, which means that the workspace of a wrist is homeomorphic to the three-dimensional projective space $\mathscr{P}^3$, whose points are the equivalence classes of proportional vectors in $\mathscr{R}^4$.

Each point of the jointspace can be identified via the three length of the legs, i.e. the three distances l_i between points P_i and Q_i of Figure 4.1. Therefore, the jointspace is always a subset of the three-dimensional Euclidean space, whose points can be identified by the vector $\mathbf{l} = (l_1, l_2, l_3)$.

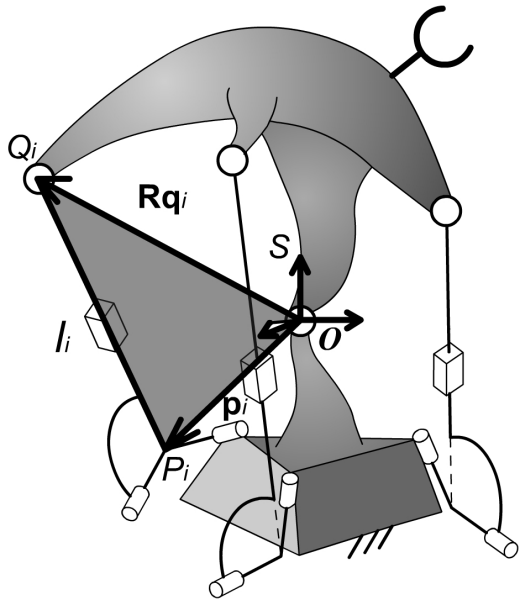


Figure 4.4: Constraints of the spherical wrist through Carnot theorem.

The vector $(l_1, l_2, l_3, e_0, e_1, e_2, e_3)$ can be used to identify a configuration of the parallel wrist, yet, not any such vector determines an allowed configuration of the wrist, for three constraints must be satisfied. The equations expressing these constraints are derived by means of Carnot theorem applied to the three triangles P_iOQ_i, as shown in Figure 4.4. For each of such triangles one can write the constraint:

$$l_i^2 = \mathbf{p}_i^T \mathbf{R}\mathbf{q}_i \qquad\qquad i = 1, 2, 3 \tag{4.3}$$

where $\mathbf{R}$ is the rotation matrix ruling the coordinate change from S' to S, whereas $\mathbf{p}_i$ and $\mathbf{R}\mathbf{q}_i$ are the column vectors containing the coordinates of points P_i and Q_i in the fixed frame S.

The rotation matrix can be written as a quadratic function of the four Euler parameters, according to the following equation:

$$\mathbf{R} = \begin{pmatrix} 1 - 2e_2^2 - 2e_3^2 & 2(e_1e_2 - e_3e_0) & 2(e_1e_3 + e_2e_0) \\ 2(e_1e_2 + e_3e_0) & 1 - 2e_1^2 - 2e_3^2 & 2(e_2e_3 - e_1e_0) \\ 2(e_1e_3 - e_2e_0) & 2(e_1e_0 + e_2e_3) & 1 - 2e_1^2 - 2e_2^2 \end{pmatrix} \tag{4.4}$$

Equation (4.3) represents a set of three quadratic equations in Euler parameters and leg lengths. If only positive leg lengths are accepted, which indeed does not exclude any configuration of the wrist, there is only one set of leg lengths for any orientation of the platform. Thus, the workspace alone can be used to represent the whole configuration space of the wrist.

4.1.3 Singularity Locus

In order to determine the singularity locus of the 3UPS spherical wrist, Equation (4.3) is differentiated:

$$\mathbf{A}\begin{pmatrix} \partial l_1 \\ \partial l_2 \\ \partial l_3 \end{pmatrix} = \mathbf{B}\begin{pmatrix} \partial e_0 \\ \partial e_1 \\ \partial e_2 \\ \partial e_3 \end{pmatrix} \tag{4.5}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the Jacobian matrices of Equation (4.3) with respect to leg lengths and Euler parameters respectively. Note that $\mathbf{B}$ is a 3×4 rectangular matrix.

Moreover, not any virtual variation of Euler parameters is allowed, for Equation (4.2) must hold for first order variations too. Thus, differentiation of Equation (4.2) yields the ensuing constraint upon the virtual variations of Euler parameters:

$$e_0\partial e_0 + e_1\partial e_1 + e_2\partial e_2 + e_3\partial e_3 = 0 \tag{4.6}$$

If Equation (4.6) and Equation (4.5) are put together, the ensuing relation is obtained:

$$\begin{pmatrix} & \mathbf{A} & \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} \partial l_1 \\ \partial l_2 \\ \partial l_3 \end{pmatrix} = \begin{pmatrix} & \mathbf{B} & \\ e_0 & e_1 & e_2 & e_3 \end{pmatrix}\begin{pmatrix} \partial e_0 \\ \partial e_1 \\ \partial e_2 \\ \partial e_3 \end{pmatrix} \tag{4.7}$$

Parallel singularities occur whenever a nonzero virtual displacement of the platform is allowed by the constraints, although the actuators undergo no virtual displacements. This implies that the determinant of the matrix at the right-hand side of Equation (4.7) vanishes. Thus the parallel singularity locus is defined as the zero level set of a function J on the configuration space, which contains all possible orientations of the platform. The function J can be obtained as:

$$J = \det\begin{pmatrix} & \mathbf{B} & \\ e_0 & e_1 & e_2 & e_3 \end{pmatrix} \tag{4.8}$$

Each element of matrix $\mathbf{B}$ is linear and homogeneous in the four Euler parameters, therefore J is a fourth-order homogeneous polynomial in the four Euler parameters. The singularity locus $J = 0$ can be represented as a two-dimensional surface cutting the workspace, as schematically depicted in

Figure 4.5. In the next section, the method developed in Chapter 2 will be used to determine how many parallel-singularity-free regions are partitioned by the singularity locus, and to find out whether it is possible or not to reach a desired position in the workspace without crossing a parallel singularity.

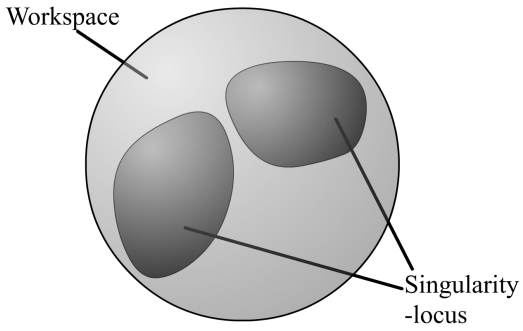


Figure 4.5: The singularity locus can be represented as a surface in the workspace.

4.1.4 Analysis of Singularity Locus

The toughest and most important task is the determination of all critical points of the function J on the configuration space, which coincides in this case with the workspace. For the case at hand, a redundant parameterization is used, because the four Euler parameters, tied by Equation (4.2), identify a point of the three dimensional manifold containing all possible orientation of a rigid body.

The most straightforward way to tackle the problem is to resort to Lagrange's multipliers. At the critical points of J the gradient of J is parallel to the gradient of the constraint, formalized by Equation (4.2), i.e.:

$$\begin{aligned} \partial J/\partial e_0 &= \lambda e_0 \\ \partial J/\partial e_1 &= \lambda e_1 \\ \partial J/\partial e_2 &= \lambda e_2 \\ \partial J/\partial e_3 &= \lambda e_3 \end{aligned} \tag{4.9}$$

where λ is Lagrange multiplier. λ can be easily eliminated considering the ensuing equation-set:

$$(\partial J/\partial e_1)\, e_0 - (\partial J/\partial e_0)\, e_1 = 0$$

$$\begin{aligned} (\partial J/\partial e_2)\, e_0 - (\partial J/\partial e_0)\, e_2 &= 0 \\ (\partial J/\partial e_3)\, e_1 - (\partial J/\partial e_1)\, e_3 &= 0 \end{aligned} \tag{4.10}$$

Equation (4.10) has been obtained by Equation (4.9) by multiplying the i^{th} equation by e_j and by subtracting the result from the product of the j^{th} equation by e_i, with a proper choice of i and j.

Equation (4.10) is set of three homogeneous fourth-order polynomial equations, in the four Euler parameters. Each solution in the projective space of such equation, when properly normalized, is set of Euler parameters defining a critical point of J on the workspace, except some extraneous solutions introduced while passing from Equation (4.9) to Equation (4.10). Such extraneous solutions are obtained when e_0 or e_1 are posed equal to zero. If $e_0 = 0$ Equation (4.10) becomes:

$$\begin{aligned} (\partial J/\partial e_0) &= 0 \\ (\partial J/\partial e_0) &= 0 \\ (\partial J/\partial e_3)\, e_1 - (\partial J/\partial e_1)\, e_3 &= 0 \end{aligned} \tag{4.11}$$

where the first two equations degenerate into the same one. Therefore, Equation (4.11) is a set of two homogeneous equations, the first of degree three and the second of degree four, in three unknowns. By virtue of Bezout theorem, Equation (4.11) admits 12 solutions, which are extraneous solutions to Equation (4.9), that does not admit, in general, solutions with $e_0 = 0$.

Analogously, if $e_1 = 0$, Equation (4.10) degenerates into the two ensuing equations

$$\begin{aligned} (\partial J/\partial e_1) &= 0 \\ (\partial J/\partial e_2)\, e_0 - (\partial J/\partial e_0)\, e_2 &= 0 \end{aligned} \tag{4.12}$$

that yield twelve extraneous solutions, too.

Equation (4.10) is a set of three homogeneous equations of degree 4, therefore, by virtue of Bezout theorem, it admits $4^3 = 64$ solutions in the complex projective space. Since 24 solutions are extraneous for Equation (4.9), there are 40 solutions to Equation (4.9), and the real ones are critical points of J.

Such forty solutions can be obtained by partial homogenization. First of all, Equation (4.10) is transformed into a non homogeneous system of equations, by posing $e_0 = 1$. In this way, any homogeneous solution with $e_0 = 0$ becomes a solution at infinity, included the twelve extraneous solutions. Then, Equation (4.10) is partially homogenized, by posing $e_2 = x_1/x_0$ and $e_3 = x_2/x_0$, and by simplifying the denominators. In this way, Equation

(4.10) becomes a system of three homogeneous equations of degree four in the three variables x_1, x_2, and x_0, where variable e_1, that has been left out of partial homogenization, is hidden in the coefficients. Variables x_1, x_2, and x_0 can be got rid of by means of classical elimination methods, as recalled in Section 5.1, obtaining a polynomial in the hidden variable e_1.

Coming from a homogeneous equation set that should have 64 solutions, the resultant polynomial should be of degree 64. However, since the homogeneous equation set always possesses 12 solutions with $e_0 = 0$, the resultant polynomial must have at least 12 solutions at infinity, and its degree will be at most 52.

Furthermore, since there are always twelve extraneous solutions with $e_1 = 0$, the resultant polynomial will be divisible by the monomial e_1^{12}. By dividing the resultant by e_1^{12}, a final equation of degree 40 is obtained, that is completely purged from extraneous solutions.

The polynomial of degree 40 is solved numerically, and the values of e_2 and e_3 corresponding to each solution in e_1 are found as recalled in Section 5.1. The values obtained are homogeneous solutions with $e_0 = 1$. In order to obtain the four Euler parameters identifying the orientation of the rigid body, the four values just obtained must be normalized, so that Equation (4.2) is satisfied. Should there be any critical point with $e_0 = 0$, this would be another solution at infinity to the resultant polynomial, whose degree would be lesser than 40. In this case, the loss of a solution is easily detected by the loss of one degree of the final polynomial, and the lost solution can easily be found by substituting $e_0 = 0$ into Equation (4.9).

In this way, all 40 complex solutions to Equation (4.9) are found, and the real ones are the critical points of the function J. These critical points must be classified into maxima, minima, 1-saddles and 2-saddles. In order to perform the classification, a local coordinate system could be chosen, and the Hessian matrix could be calculated and analyzed. However, this is not the most straightforward way to proceed, for the parameterization used henceforth is redundant, and represents no local coordinate system. A method to determine the type of critical points with a redundant parameterization is the Hessian matrix of Lagrangian equation, also called sometimes bordered Hessian, which will be hereafter briefly recalled. For further details see for example [46].

Lagrange multipliers made it possible to determine the critical points of J under the constraint defined by Equation (4.2). Lagrange multipliers are based upon the intrinsic property that, at a critical point, the first order increment of the function J along any direction contained in the tangent space of the constraint must vanish. In other words the gradient of J must be parallel to the gradient of the constraint.

At the same way, it is possible to analyze the second order variations of J in the neighborhood of a critical point. The maximum increase and maximum decrease directions of the second order approximation of the function J are searched for, in the tangent space of the constraint. A maximum will then be a critical point where there are no steepest increase directions in the tangent space of the constraint, a 1-saddle is a point where there is one steepest decrease direction and two steepest increase directions, and so forth.

Let $\tilde{\mathbf{e}}$ be a set of Euler parameters defining a critical point of J. Let c be defined as:

$$c(\mathbf{e}) = e_0^2 + e_1^2 + e_2^2 + e_3^2 - 1 \tag{4.13}$$

Thus $c = 0$ defines the constraint under which the critical points have been found.

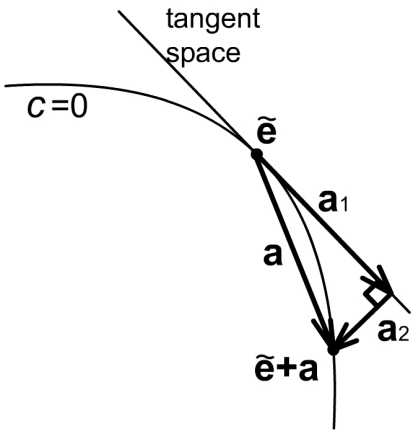


Figure 4.6: Small displacement from the critical point.

Suppose now that a small displacement $\mathbf{a}$ is performed on the constraint, starting from the critical point $\tilde{\mathbf{e}}$. The displaced point $\tilde{\mathbf{e}} + \mathbf{a}$ must belong to the constraint, too, thus the ensuing equation holds:

$$c(\tilde{\mathbf{e}} + \mathbf{a}) = 0 \tag{4.14}$$

The displacement $\mathbf{a}$ can be decomposed into a first order displacement, $\mathbf{a}_1$, and a higher order displacement $\mathbf{a}_2$, as shown in Figure 4.6. The first order displacement $\mathbf{a}_1$ is the projection of $\mathbf{a}$ onto the tangent space of the constraint, therefore the ensuing condition holds:

$$\mathbf{a}_1^T \nabla c = 0 \tag{4.15}$$

Furthermore, only displacements with the same arbitrary small length l will be considered, therefore all displacements are constrained by the ensuing condition:

$$\mathbf{a}^T\mathbf{a} - l^2 = 0 \tag{4.16}$$

If only infinitesimals up to the second order are considered, Equation (4.16) becomes:

$$\mathbf{a_1}^T\mathbf{a_1} - l^2 = 0 \tag{4.17}$$

The increment of the function J along the displacement $\mathbf{a}$ is approximated up to the second order, as follows:

$$\Delta J(\mathbf{a}) = J(\tilde{\mathbf{e}} + \mathbf{a}) - J(\tilde{\mathbf{e}}) = \mathbf{a}_2^T \nabla f|_{\tilde{\mathbf{e}}} + \frac{1}{2}\mathbf{a}_1^T \mathbf{H}|_{\tilde{\mathbf{e}}}\, \mathbf{a}_1 \tag{4.18}$$

where $\nabla f|_{\tilde{\mathbf{e}}}$ and $\mathbf{H}|_{\tilde{\mathbf{e}}}$ are the gradient and the 4×4 Hessian matrix of J at the critical point $\tilde{\mathbf{e}}$, obtained by differentiating J with respect to all four Euler parameters.

Equation (4.14) can be approximated up to the second order, obtaining:

$$c(\tilde{\mathbf{e}} + \mathbf{a}_1 + \mathbf{a}_2) = \mathbf{a}_2^T \nabla c|_{\tilde{\mathbf{e}}} + \frac{1}{2}\mathbf{a}_1^T \mathbf{H}_c|_{\tilde{\mathbf{e}}}\, \mathbf{a}_1 = 0 \tag{4.19}$$

where $\nabla c|_{\tilde{\mathbf{e}}}$ and $\mathbf{H}_c|_{\tilde{\mathbf{e}}}$ are the gradient and the Hessian matrix of the constraint with respect to the four Euler parameters.

Since $\tilde{\mathbf{e}}$ is a critical point, the ensuing relation holds:

$$\nabla f|_{\tilde{\mathbf{e}}} = \lambda\, \nabla c|_{\tilde{\mathbf{e}}} \tag{4.20}$$

By composing Equations 4.19 and 4.20, one obtains:

$$\mathbf{a}_2^T \nabla f|_{\tilde{\mathbf{e}}} = -\lambda \frac{1}{2}\mathbf{a}_1^T \mathbf{H}_c|_{\tilde{\mathbf{e}}}\, \mathbf{a}_1 \tag{4.21}$$

By inserting Equation (4.21) into Equation (4.18), the ensuing expression for the increment of J is derived:

$$\Delta J(\mathbf{a}_1) = \frac{1}{2}\mathbf{a}_1^T (\mathbf{H}|_{\tilde{\mathbf{e}}} - \lambda\, \mathbf{H}_c|_{\tilde{\mathbf{e}}})\, \mathbf{a}_1 \tag{4.22}$$

The steepest increase and decrease directions, in the neighborhood of the critical point $\tilde{\mathbf{e}}$, can be found as the critical points of the function $\Delta J(\mathbf{a}_1)$, under the constraints defined by Equations 4.15 and 4.17, which impose that the first order components lie in the tangent space, and that the displacements have equal lengths. Lagrange multipliers method is again the most suited tool to tackle the problem, yielding the ensuing equations:

$$\nabla_{\mathbf{a}_1}(\Delta J) = (\mathbf{H}|_{\tilde{\mathbf{e}}} - \lambda\, \mathbf{H}_c|_{\tilde{\mathbf{e}}})\, \mathbf{a}_1 = \alpha \mathbf{a}_1 + \beta\, \nabla c|_{\tilde{\mathbf{e}}} \tag{4.23}$$

where $\nabla_{\mathbf{a}_1}$ is the gradient calculated with respect to the first order components of the displacement vector $\mathbf{a}$.

Equation (4.23) can be rewritten as follows:

$$(\mathbf{H}|^*_{\tilde{\mathbf{e}}} - \alpha \mathbf{I}^*)\, \mathbf{a}^* = \mathbf{0} \tag{4.24}$$

where:

- $\mathbf{H}|^*_{\tilde{\mathbf{e}}}$ is the bordered Hessian, built as follows:

$$\mathbf{H}|^*_{\tilde{\mathbf{e}}} = \begin{pmatrix} (\mathbf{H}|_{\tilde{\mathbf{e}}} - \lambda\, \mathbf{H}_c|_{\tilde{\mathbf{e}}}) & -\nabla c|_{\tilde{\mathbf{e}}} \\ -\nabla c|^T_{\tilde{\mathbf{e}}} & 0 \end{pmatrix} \tag{4.25}$$

  If the Lagrangian function $L(\mathbf{e}, \lambda) = J(\mathbf{e}) - \lambda c(\mathbf{e})$ is introduced to build Equation (4.9), then the bordered Hessian matrix is the 5×5 Hessian matrix of L, calculated with respect to all its five variables. The value of λ necessary to calculate $\mathbf{H}|^*_{\tilde{\mathbf{e}}}$ at each critical point is easily determined through Equation (4.9);

- $\mathbf{I}^*$ is equal to the 5×5 identity matrix, save the fifth element of the fifth row, which is equal to zero;

- $\mathbf{a}^*$ is a vector obtained by appending to vector $\mathbf{a}$ the multiplier β, which is an unimportant dummy variable.

Therefore Equation (4.24) is a common eigenvalue problem, and the steepest increase or decrease directions are the directions for which the ensuing condition is satisfied:

$$\det\,(\mathbf{H}|^*_{\tilde{\mathbf{e}}} - \alpha \mathbf{I}^*) = 0 \tag{4.26}$$

Equation (4.26) is always a third order polynomial in the eigenvalue α, therefore three solutions are expected. For each positive solution, Equation (4.24) yields the corresponding steepest increase direction, whereas for each negative solution a steepest decrease direction is found. The index of the critical point is the number of negative solutions to Equation (4.26), which enables the classification of any possible critical point of J.

Also the generation of the steepest ascent or descent paths does not require the use of a local coordinate system. A method analogous to that presented in Section 3.1.3 can be used, and the steepest ascent direction is easily obtained as the component of the gradient of J normal to the gradient of the constraint c. Whenever, while following a steepest ascent or descent

path, a set of Euler parameters $\mathbf{e}$ with $e_0 < 0$ is reached, it is possible to immediately replace it with $-\mathbf{e}$, which is the same position of the platform, in order that e_0 is always greater than zero, as discussed in Section 4.1.2.

Analogous to Section 3.1.5, it is possible to exploit the maximum number of critical points of J to give an upper bound to the number of singularity-free regions composing the workspace of a 3UPU spherical wrist. By considering that there can be at most 40 real solutions to Equation (4.9), and that the Betti number sequence of the workspace is 1,1,1,0,0,0,. . . , as recalled in Table 1.1, there can be no more than 20 parallel-singularity-free regions with the same sign of the Jacobian determinant, and not more than 21 regions at all.

4.1.5 Example

The 3UPS spherical wrist W_1 will be used to show the application of the discussed method. The parameters defining manipulator W_1 are reported in Table 4.1, according to the convention discussed in Section 4.1.1.

$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}3$
(1,0,0)	(0,1,0)	(0,1,1)	(-9,2,6)/11	(6,6,7)/11	(-1,1,0)

Table 4.1: Parameters defining manipulator W_1.

Through the elimination method discussed in the previous section, 32 critical points are determined, among which there are 4 positive maxima, 2 positive 2-saddles, 4 negative minima, 12 negative 1-saddles, and 10 singular 2-saddles.

Figure 4.7 shows the workspace of the spherical wrist, through the representation discussed in Section 4.1.2, according to which the workspace is a three-dimensional ball with diametrically opposite boundary points identified. The four positive maxima M_1, M_2, M_3, and M_4, are depicted as cones, and the two positive 2-saddles S_1 and S_2, depicted as spheres. In this representation, the coordinate of these critical points are reprted in Table 4.2.

The steepest ascent paths starting from the two 2-saddles S_1 and S_2 join M_1 to M_2 and M_3 to M_4 respectively, thus there are two positive regions free of parallel singularities. The steepest ascent paths are represented in Figure 4.7 as black lines.

Given the three points $P_1 = (-1.85, 0.78, -1.55)$, $P_2 = (-0.50, -0.75, -1.90)$, and $P_3 = (0.30, 1.14, -2.47)$, where J is positive, it can be assessed to which one of the two positive regions they do belong by following the steepest ascent paths (black lines in Figure 4.7). The steepest ascent paths starting from

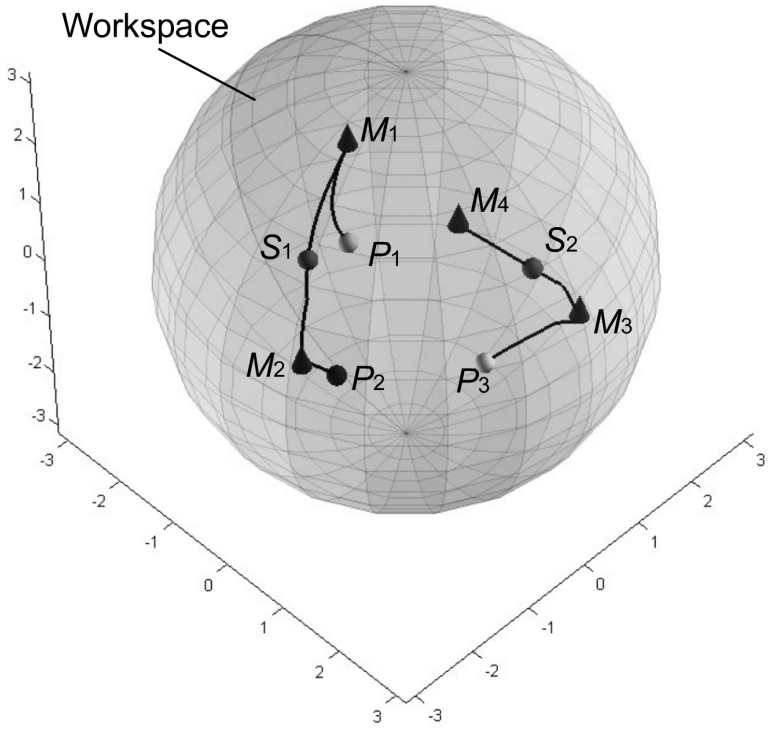


Figure 4.7: Positive critical points in the workspace of W_1.

Critical point	Coordinates
M_1	(-2.19,1.13,-0.27)
M_2	(-0.69,-1.17,-1.57)
M_3	(1.36,1.74,-1.30)
M_4	(0.69,0.22,0.91)
S_1	(-1.69,-0.08,-1.14)
S_2	(1.16,1.06,-0.08)

Table 4.2: Coordinates of the positive critical points of W_1.

P_1, P_2, and P_3 reach the maxima M_1, M_2, and M_3, respectively. Therefore $P_1$and P_2 belong to the same aspect, and the path P_1-M_1-S_1-M_2-P_2, connecting P_1 to P_2 is singularity-free. The steepest ascent path starting from P_3 reaches M_3, which belongs to a different region, therefore there exists no singularity-free path at all to reach P_1 or P_2 starting from P_3.

Figure 4.8 shows the evolution of the level sets of the function J while the level decreases. Four regions are generated by the four maxima, and are joined in pairs by the two 2-saddles.

The four minima are all connected through a network of steepest descent paths starting from the saddles, therefore there is only one negative singularity-free region.

4.1.6 3UPU Spherical wrist

The 3UPU architecture for a spherical wrist was first proposed in [47], where it was proved that it is able to perform local spherical motions, whereas in [48] and [49] its capability of generating a finite set of spherical motions was proved.

The 3UPU wrist, depicted in Figure 4.9, is very similar to the 3UPS spherical wrist, for it consists of a platform connected to the base through three legs composed of two universal joints and one actuated prismatic joint. The two universal joints are such that the axes of the revolute joints attached to the platform and to the base converge in the same point O, and the axes of the two remaining revolute joints of each leg are parallel. Analogous to the 3UPS wrist the architecture of the 3UPU wrist can be described through the coordinate vectors $\mathbf{p}_i$ and $\mathbf{q}_i$ of the centers P_i and Q_i of the universal joints of the base an the platform respectively. Furthermore, the equations describing the kinematics of the wrist, and therefore the parallel singularity locus are identicl to those of the 3UPS wrist.

The main difference is that the center of spherical motion is determined by the particular geometry and assemblage of the legs of the machine in the 3UPU wrist, and not through a physical joint as in the 3UPS wrist. With reference to Figure 4.10, if the platform is considered as the end-effector of the serial chain given by the i^{th} leg, it is easy to see that the point of the platform coincident with point O of the base can only undergo virtual displacements ∂O contained in the plane determined by points P_i, O, and Q_i. Therefore, whenever the normals to the three such planes determined by the three legs are linearly independent, the only allowed virtual displacement of point O is zero, i.e. the platform is constrained to spherical motion with center O.

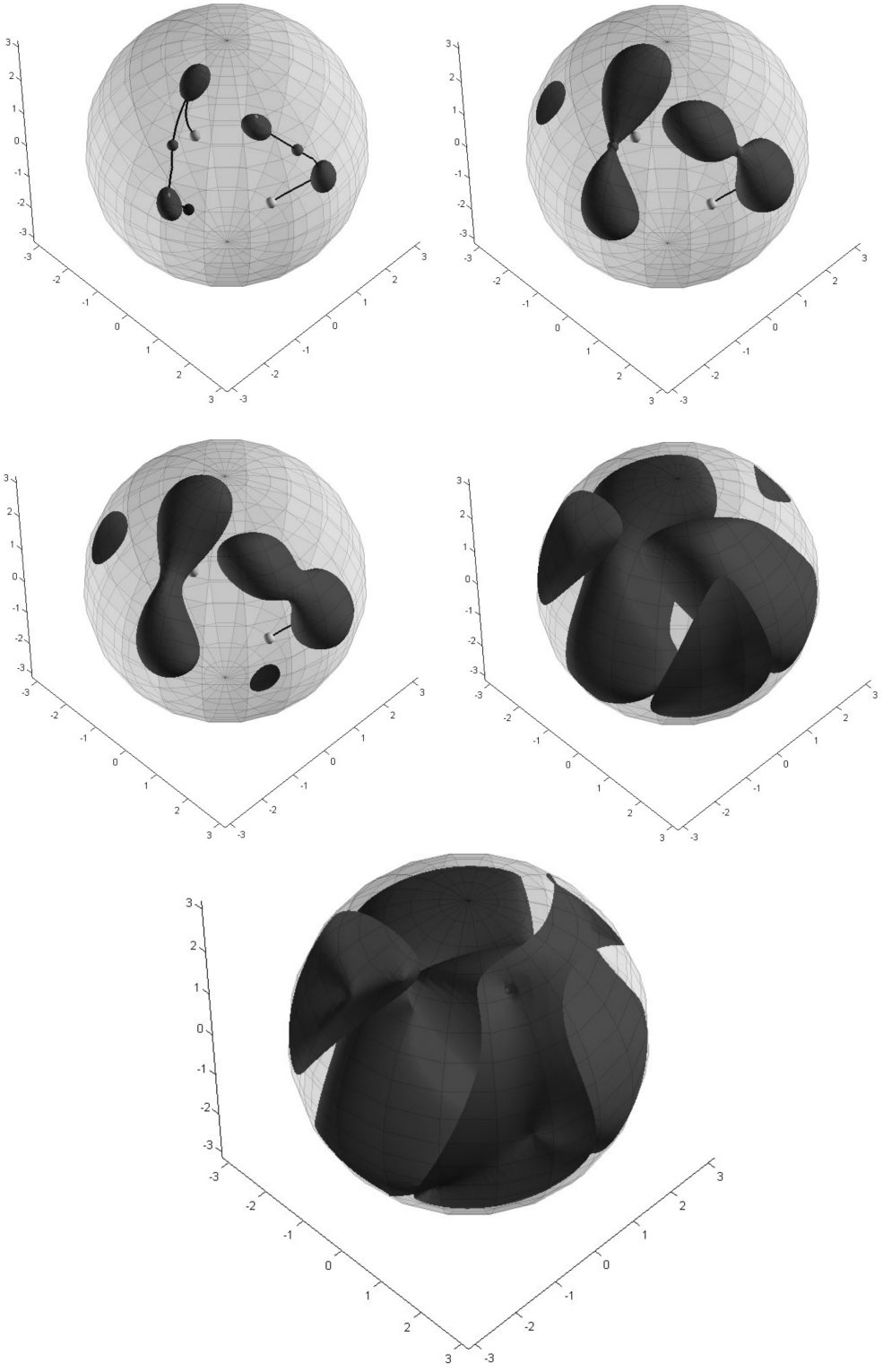


Figure 4.8: Evolution of the level sets of J as the level decreases.

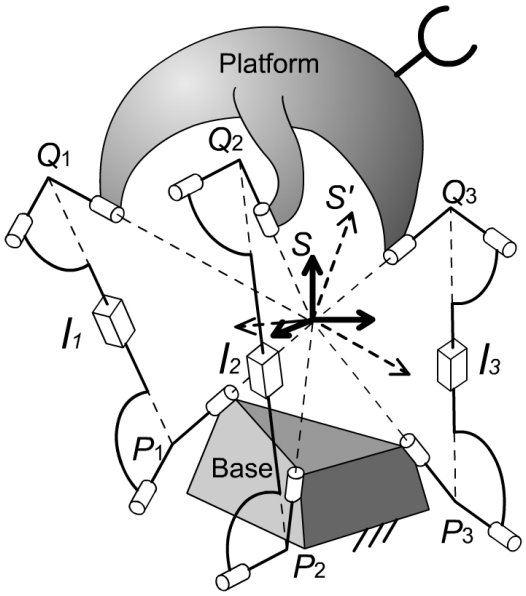


Figure 4.9: A 3UPU spherical wrist.

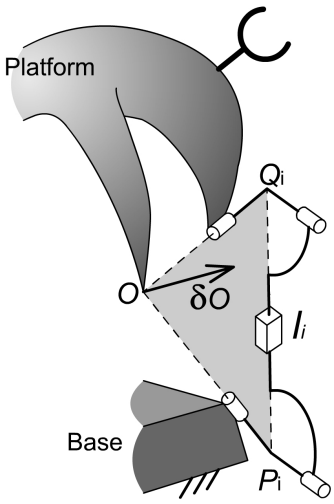


Figure 4.10: The center of the spherical motion of 3UPU wrists is determined by the architecture of the legs.

This condition is almost always verified, except for some singular positions, where the constraint of spherical motion is lost, and the wrist gains a translational degree of freedom. These singular positions were named by Zlatanov et al. in [50] *constraint singularities*, because at such positions some constraints of the parallel architecture are locally lost. Constraint singularities are typical of lower mobility parallel manipulators where the platform possesses less than six degrees of freedom. In such manipulators, some of the six degrees of freedom of the platform are controlled through the actuators of the manipulator, whereas some other (the translational ones, in the case of 3UPU spherical wrist) are passively constrained through the geometry of the legs. Parallel singularities are always detected by differentiating the equations connecting the input variables of the jointspace to the output variables of the workspace, but constraint singularities may not. In order to detect constraint singularities is always necessary to consider all six degrees of freedom of the platform, and to investigate under which conditions the constraints upon the degrees of freedom that are not controlled by the actuators might fail. For more detail about constraint singularities refer for example to [50] or [51].

In case of 3UPU wrists, as briefly shown, constraint singularities occur when the normals to the three planes containing the points P_i, O, and Q_i are linearly dependent. This geometrical condition coincides, in the case of 3UPU wrists, with the condition for the occurrence of parallel singularities $J = 0$, where J is defined as in Equation (4.8). Thus the singular positions $J = 0$ for 3UPU wrists are twice as dangerous as the same positions of 3UPS wrists, because, in addition to a local rotation, a local translation is gained by the manipulator, and the spherical constraint might be lost by the manipulator.

However, any other position where $J \neq 0$ is safe, and the machine is well constrained, thus an identical method to the one proposed for 3UPS manipulators can be used to find out safe paths, free of both parallel and constraint singularities.

A more rigorous treatment of parallel and constraint singularities of the 3UPU wrist can be found in [49], were Di Gregorio rigorously proves for the first time that parallel and constraint singularities of the 3UPU wrist do coincide. In [49] the explicit equation of the singularity locus of 3UPU wrists is given in terms of Rodrigues parameters, which is equivalent to the condition $J = 0$ determined in Section 4.1.3 in terms of Euler parameters.

4.2 3-UPU Translational Manipulator

4.2.1 Architecture

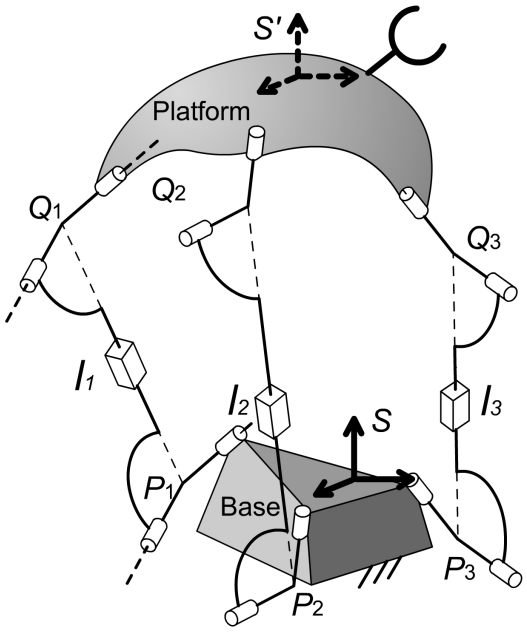


Figure 4.11: Kinematic architecture of a 3UPU translational manipulator.

3UPU translational manipulators have been proposed and analyzed in [52], [53], and [54].

The architecture that will be hereafter considered is that proposed in [53], and sketched in Figure 4.11. The platform is connected to the base by means of three legs, consisting of two links connected to each other by a prismatic joint and to the base and the platform through universal joints. The universal joints satisfy the ensuing two geometrical requirements:

- in each leg, the axes of the two revolute joints connected to the base and to the platform are parallel;
- in each leg, the axes of the two middle revolute joints, not connected to the base nor to the platform, are parallel.

It can be proved that this architecture constrains the platform to pure translational motions (see [53]).

With reference to Figure 4.12, the geometry of 3UPU translational manipulators, can be parametrized in the ensuing way:

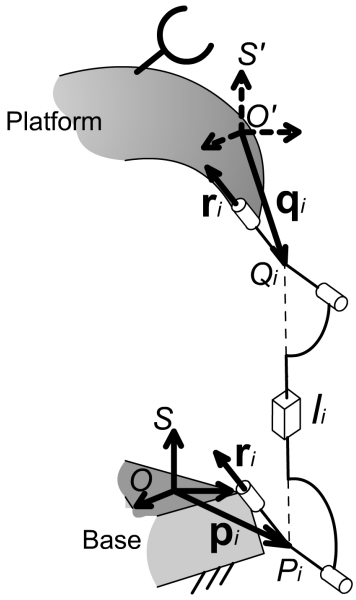


Figure 4.12: Parametrization of the geometry of a 3UPU translational manipulator.

- two reference frames S and S' with parallel axes, attached to the base and to the platform respectively, are defined;
- on the i^{th} leg, the center P_i of the universal joint attached to the base is identified through its coordinate vector $\mathbf{p}_i$ in frame S;
- on the i^{th} leg, the center Q_i of the universal joint attached to the platform is identified through its coordinate vector $\mathbf{q}_i$ in frame S';
- on the i^{th} leg, the common directions of the axes of the revolute joints attached to the base or to the platform is identified by way of a unit vector $\mathbf{r}_i$

Therefore, the nine vectors $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ completely define the kinematic architecture of 3UPU translational manipulators.

4.2.2 Configuration space

The workspace of a 3UPU translational manipulator is the manifold containing all possible positions of the platform. Each point of the workspace can be identified by means of the coordinate vector $\mathbf{x} = (x, y, z)$ of a point, for

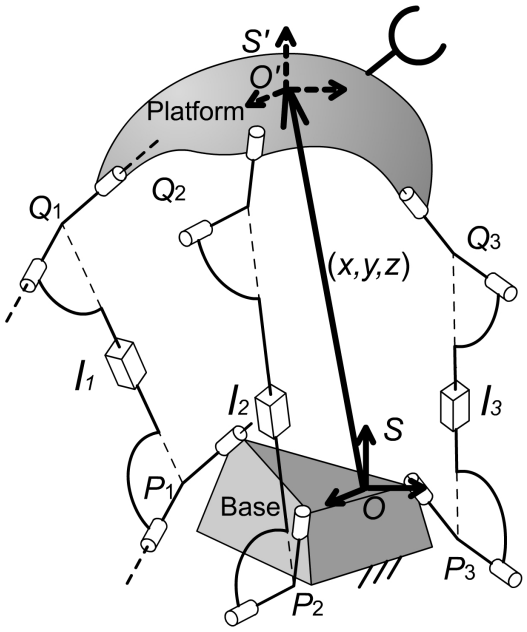


Figure 4.13: The position of the platform is determined by the position vector of O'.

example the origin O' of S', with respect to the fixed frame S'. Therefore the workspace is the whole three dimensional Euclidean space.

Any point of the jointspace is defined by the lengths l_1, l_2, and l_3 of the three actuated legs, thus the jointspace is a subset of the three dimensional Euclidean space, too. More specifically, the length l_i is equal to the distance between points P_i and Q_i.

Therefore, the vector (l_1, l_2, l_3, x, y, z) identifies a configuration of the manipulator only if the ensuing constraints are satisfied:

$$l_i^2 = (Q_i - P_i)^2 = (\mathbf{q}_i + \mathbf{x} - \mathbf{p}_i)^2 \tag{4.27}$$

Like the spherical wrists, if only positive lengths are accepted to describe the length of the legs, there exists only one point of the jointspace that defines a configuration along with a given point in the workspace, which means that the workspace and the configuration space are homeomorphic, and can be considered as the same manifold. In other words, the vector (x, y, z) identifies both a position of the platform and a configuration of the manipulator.

4.2.3 Singularity locus

Equation (4.27) can be differentiated, obtaining the ensuing relation:

$$\mathbf{A} \begin{pmatrix} \partial l_1 \\ \partial l_2 \\ \partial l_3 \end{pmatrix} = \mathbf{B} \begin{pmatrix} \partial x \\ \partial y \\ \partial z \end{pmatrix} \tag{4.28}$$

where the i^{th} row of matrix $\mathbf{B}$ is the vector $2(Q_i - P_i)^T$, which can be written as $2(\mathbf{x} + \mathbf{c_i})^T$, where $\mathbf{c_i}$ is a constant vector for the i^{th} row. This means that the determinant of $\mathbf{B}$ is linear in the variables x, y, and z.

The parallel singularity locus is therefore a plane in the three-dimensional Euclidean space, because it is determined by the equation:

$$J_p = \det \mathbf{B} = 0 \tag{4.29}$$

which is linear in the variables x, y, and z.

Analogous to 3UPU spherical wrists, 3UPU translational manipulators possess constraint singularities, because the constraint to translational motion is only ensured by the particular geometry of the legs. However, unlike the 3UPU wrist, constraint singularities do not coincide with parallel ones, for 3UPU translational manipulators. Constraint singularities of translational 3UPU manipulators were studied and rigorously determined by Parenti Castelli and Di Gregorio in [51], where the equation defining the singularity locus of a 3UPU translational manipulator is explicitly reported.

Constraint singularities can be determined in an similar manner to 3UPU spherical wrists. By observing the architecture of each leg of the 3UPU translational manipulator (Figure 4.13), it can be easily inferred that each leg allows only two rotational degrees of freedom of the platform. If the directions of the axes of the two revolute pairs composing each universal joint are denoted through the unit vectors $\mathbf{r}_i$ and $\mathbf{s}_i$, each leg allows only virtual rotations about axes directed as linear combinations of $\mathbf{r}_i$ and $\mathbf{s}_i$. Therefore, any virtual rotation about a vector with nonzero component along a vector $\mathbf{t}_i$, normal to both $\mathbf{r}_i$ and $\mathbf{s}_i$, is hindered by the i^{th} leg. Thus, if the three vectors $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ of the three legs are linearly independent, the platform can perform no virtual rotation. On the other hand, if $\mathbf{t}_1$, $\mathbf{t}_2$, and $\mathbf{t}_3$ are linearly dependent, the platform can undergo an uncontrolled virtual rotation, i.e. the manipulator is at a constraint singularity. Such constraint singularities are organized in a locus, which is defined by the ensuing equation:

$$J_c = \det \mathbf{C} = 0 \tag{4.30}$$

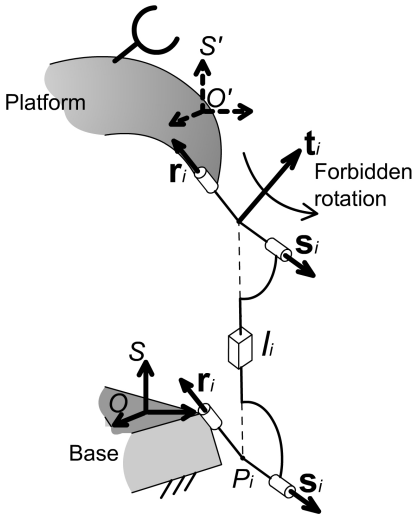


Figure 4.14: Each leg hinders a virtual rotation of the platform.

where $\mathbf{C}$ is a matrix whose columns are the components of vectors $\mathbf{t}_i$ with respect to frame S'. Such components can be expressed as follows:

$$\mathbf{t}_i = (Q_i - P_i) - [(Q_i - P_i)\mathbf{r}_i]\,\mathbf{r}_i \tag{4.31}$$

Equation (4.30) is a third order polynomial in the coordinates x, y, and z, identifying the configurations of the manipulator.

4.2.4 Analysis of singularity locus

Both constraints and parallel singularities are equally dangerous for the 3UPU translational manipulator, and must be avoided while moving from a configuration to another. In this section, the method developed in Chapter 2 will be applied to 3UPU translational manipulators, in order to find out, if it exists, a path free of constraint and parallel singularities connecting any two configurations of the manipulator.

The surface to be avoided is the zero level set of the function $J_p J_c$ on the workspace of the 3UPU translational manipulator. Unfortunately, the workspace of a translational manipulator is the three-dimensional Euclidean space, which is not compact. Thus, the method developed in Chapter 2 cannot be straightforwardly applied, because it works on compact manifolds only.

Yet, it is possible to transform the three-dimensional Euclidean space into a compact manifold. First of all, consider the three-dimensional real

projective space associated to the three-dimensional Euclidean space, i.e. each vector (x_0, x_1, x_2, x_3) of the projective space is such that the coordinates (x, y, z) in the Euclidean space satisfy the ensuing condition:

$$\begin{aligned} x &= x_1/x_0 \\ y &= x_2/x_0 \\ z &= x_3/x_0 \end{aligned} \tag{4.32}$$

Each point of the projective space corresponds to one point of the Euclidean space, except the points with $x_0 = 0$, i.e. points at infinity, that do not exist in the Euclidean space.

We can imagine the workspace of the 3UPU manipulator as the projective space, where the points with $x_0 = 0$ must never be crossed, exactly like singularities. Thus, the locus of "forbidden" points, the points that must not be crossed, is defined by the ensuing equation in the real projective space:

$$J = x_0 J_p J_c = 0 \tag{4.33}$$

where J_p and J_c are properly converted to homogeneous coordinates.

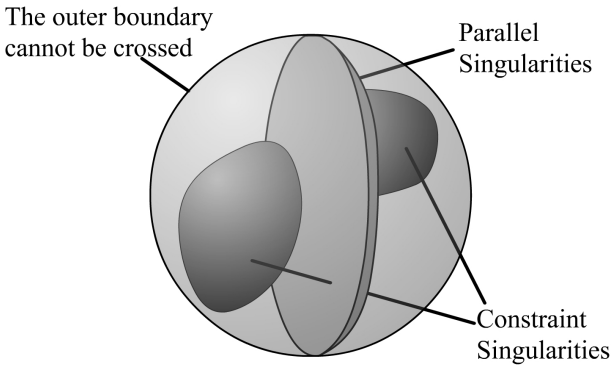


Figure 4.15: Singularity locus of the 3UPU translational manipulator.

The real projective space can be represented as a ball, analogous to the manifold containing all orientations of a rigid body. With reference to Figure 4.15, a three-dimensional ball with radius 1 is considered. For any point inside the ball, the coordinates of the point are the homogeneous coordinates x_1, x_2, and x_3 of the corresponding point in the projective space. The projective coordinate x_0 is defined as $x_0 = \sqrt{1 - x_1^2 - x_2^2 - x_3^2}$. In this way, all the points of the ball correspond to one point of the projective space. Furthermore, all points on the spherical boundary of the ball are points with

$x_0 = 0$, i.e. points at infinity. Like Euler parameters, we represent a point of the projective space as a four-dimensional vector (x_0, x_1, x_2, x_3), with the constraint:

$$x_0^2 + x_1^2 + x_2^2 + x_3^2 = 1 \tag{4.34}$$

and with $x_0 >= 0$.

The set of points of the projective space that must not be crossed, expressed formally by Equation (4.33), is the union of the ensuing sets:

- the locus of parallel singularities, linear in the four projective coordinates;
- the locus of constraint singularities, which is a third order polynomial in the four projective coordinates;
- the set of points at infinity with $x_0 = 0$, which is the spherical boundary of the ball in Figure 4.15

As usual, the critical points of J on the configuration space must be found. The configuration space is the three-dimensional real projective space, and the critical points can be found through Lagrange multipliers method:

$$\begin{aligned} \partial J/\partial x_0 &= \lambda x_0 \\ \partial J/\partial x_1 &= \lambda x_1 \\ \partial J/\partial x_2 &= \lambda x_2 \\ \partial J/\partial x_3 &= \lambda x_3 \end{aligned} \tag{4.35}$$

where λ is Lagrange multiplier. λ can be easily eliminated considering the ensuing equation-set:

$$\begin{aligned} T_1 = (\partial J/\partial x_1)\, x_0 - (\partial J/\partial x_0)\, x_1 &= 0 \\ T_2 = (\partial J/\partial x_2)\, x_0 - (\partial J/\partial x_0)\, x_2 &= 0 \\ T_3 = (\partial J/\partial x_3)\, x_0 - (\partial J/\partial x_0)\, x_3 &= 0 \end{aligned} \tag{4.36}$$

obtained by extracting λ from the first equation and then substituting it into the remaining three. Equation (4.36) is a set of three homogeneous equations of degree five. Any critical point with $x_0 = 0$ must not be considered, because it lies on the locus $J = 0$. Therefore it is possible to substitute $x_0 = 1$ into Equation (4.36), and to search for the solutions in terms of x_1, x_2, and x_3.

Since J is divisible by J_c and J_p, Equation (4.36) can be written in the ensuing form:

$$\mathbf{M} \begin{pmatrix} J_c \\ J_p \\ J_c J_p \end{pmatrix} = \mathbf{0} \tag{4.37}$$

where the elements $\mathbf{M}$ are polynomials in variables x_1, x_2, and x_3. Therefore all points where $J_c = 0$ and $J_p = 0$ are critical points of J. This points form in general a curve in the workspace, and, not being isolated, are always degenerate critical points. Fortunately, the critical points where $J_c = J_p = 0$ are all singular, and must be ruled out of the analysis. Thus, only critical points where either J_c or J_p do not vanish must be considered, which, along with Equation (4.37) yields the ensuing additional equation:

$$T_4 = \det \mathbf{M} = 0 \tag{4.38}$$

Equation (4.38) is a third order equation in x_1, x_2, and x_3 and can be used to reduce the degree of Equation (4.36). Each of the three polynomials T_1, T_2, and T_3 can be written as follows:

$$T_i = T_4 Q_i + R_i \tag{4.39}$$

where Q_i and R_i are the quotient and the remainder of a polynomial division performed on T_i through the divisor T_4, with respect to a given variable x_j. At every points where all T_i vanish along with T_4, all remainders R_i must vanish too. Therefore the equation set

$$\begin{aligned} R_1 &= 0 \\ R_2 &= 0 \\ R_3 &= 0 \\ T_4 &= 0 \end{aligned} \tag{4.40}$$

is always equivalent to Equation (4.36), along with the condition $T_4 = 0$. If R_1, R_2, and R_3 are remainders of polynomial divisions performed with respect to x_2, x_1, and x_1 respectively, R_1, R_2, and R_3 are polynomials of degree four in the two variables x_2 and x_3. Therefore the equation set:

$$\begin{aligned} R_1 &= 0 \\ R_2 &= 0 \\ R_3 &= 0 \end{aligned} \tag{4.41}$$

can be solved with a method similar to Section 4.1.4. Variable x_1 can be hidden in the coefficients, and a partial homogenization with respect to x_2 and x_3 yields a set of three homogeneous equations in three unknowns of degree four. A resultant polynomial in x_1 can then be found through the elimination method reported in Section 5.1.

Unfortunately, in this way the condition $T_4 = 0$ has not been directly imposed: Equation (4.41) is not completely equivalent to Equation (4.40),

which introduces extraneous solutions. The author has found no way to factor out such extraneous solutions from the resultant polynomial, however they can be easily detected, for they do not satisfy the condition $T_4 = 0$.

Once all real solutions have been found by solving numerically the resultant polynomial, and all extraneous solutions have been canceled, all critical points of J are known. The classification of critical points, and the determination of steepest ascent paths is then analogous to the one proposed in Section 4.1.4 for spherical wrists.

4.2.5 Example

According to the conventional parameterization proposed in Section 4.2.1, manipulator T_1, defined by vectors of Table 4.3, is considered.

$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	$\mathbf{q}_1$	$\mathbf{q}_2$	$\mathbf{q}_3$	$\mathbf{r}_1$	$\mathbf{r}_2$	$\mathbf{r}_3$
(1,0,0)	(0,1,0)	(0,1,-1)	(1,1,1)	(0,1,-1)	(1,1,1)	$\frac{(1,0,2)}{\sqrt{3}}$	$\frac{(-1,0,1}{\sqrt{2}}$	$\frac{(-1,-1,0)}{\sqrt{2}}$

Table 4.3: Parameters defining manipulator T_1.

In the workspace of T_1 there are three positive maxima and two positive 2-saddles, whose homogeneous coordinates are listed in Table 4.4, and visualized in Figure 4.16 through the conventional representation proposed in Section 4.2.2.

Critical point	Homogeneous coordinates
M_1	(0.722, -0.505, -0.425, 0.209)
M_2	(0.575, 0.326, 0.491, 0.568)
M_3	(0.606, 0.596, -0.313, -0.423)
S_1	(0.657, 0.681, -0.156, 0.282)
S_2	(0.079, 0.417, 0.902, -0.076)

Table 4.4: Coordinates of the positive critical points of T_1.

Figure 4.16 also shows the steepest ascent paths (black lines), departing from the positive 2-saddles and reaching the maxima. It can be seen that maxima M_2 and M_3 are joined, while no paths reach maximum M_3. Therefore there are two positive regions, free of parallel and constraint singularities.

Figure 4.17 shows the evolution of the level sets of the function J, while the level decreases from the height of the maximum to that of the lowest

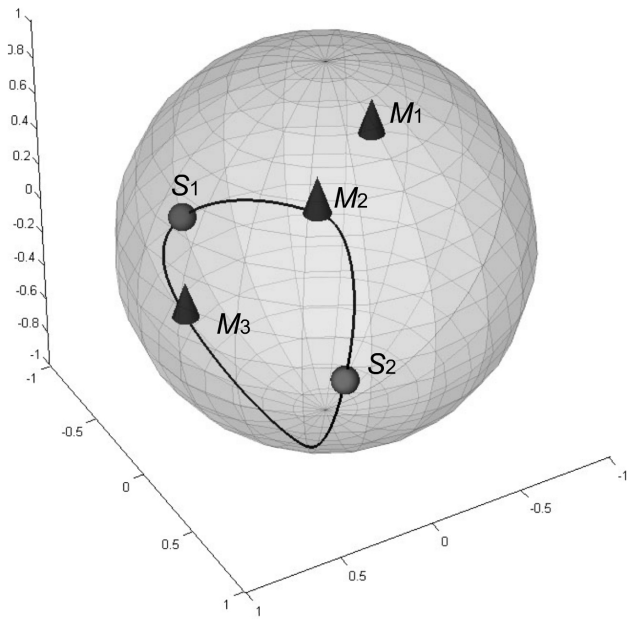


Figure 4.16: Positive critical points of manipulator T_1.

saddle. Three regions are generated by the three maxima, and two of them are joined twice by the two 2-saddles.

Figure 4.18 shows the level set $J = 0$. The outer spherical boundary belongs to the locus, but it has not been plotted, in order for the inside of the ball to be visible. The darker surface inside the ball represents the locus of parallel singularities, whereas the brighter surface the locus of constraint singularities. The intersection curve of the two surfaces is a set of singular degenerate critical points, that have been ruled out from the determination of critical points by means of the polynomial division.

There are five negative minima and four negative 1-saddles, and the network of steepest descent paths is such that there are also two negative regions.

Figure 4.19 shows all relevant critical points: the positive maxima are depicted as upward bound cones, the negative minima as downward bound cones, and the saddle points as spheres. The network of singularity-free steepest ascent and descent paths is represnted as black lines. Figure 4.20 shows two rotated views of the locus $J = 0$, where it is possible to verify that the steepest ascent and descent paths never cross the spherical boundary, nor the parallel and constraint singularity loci.

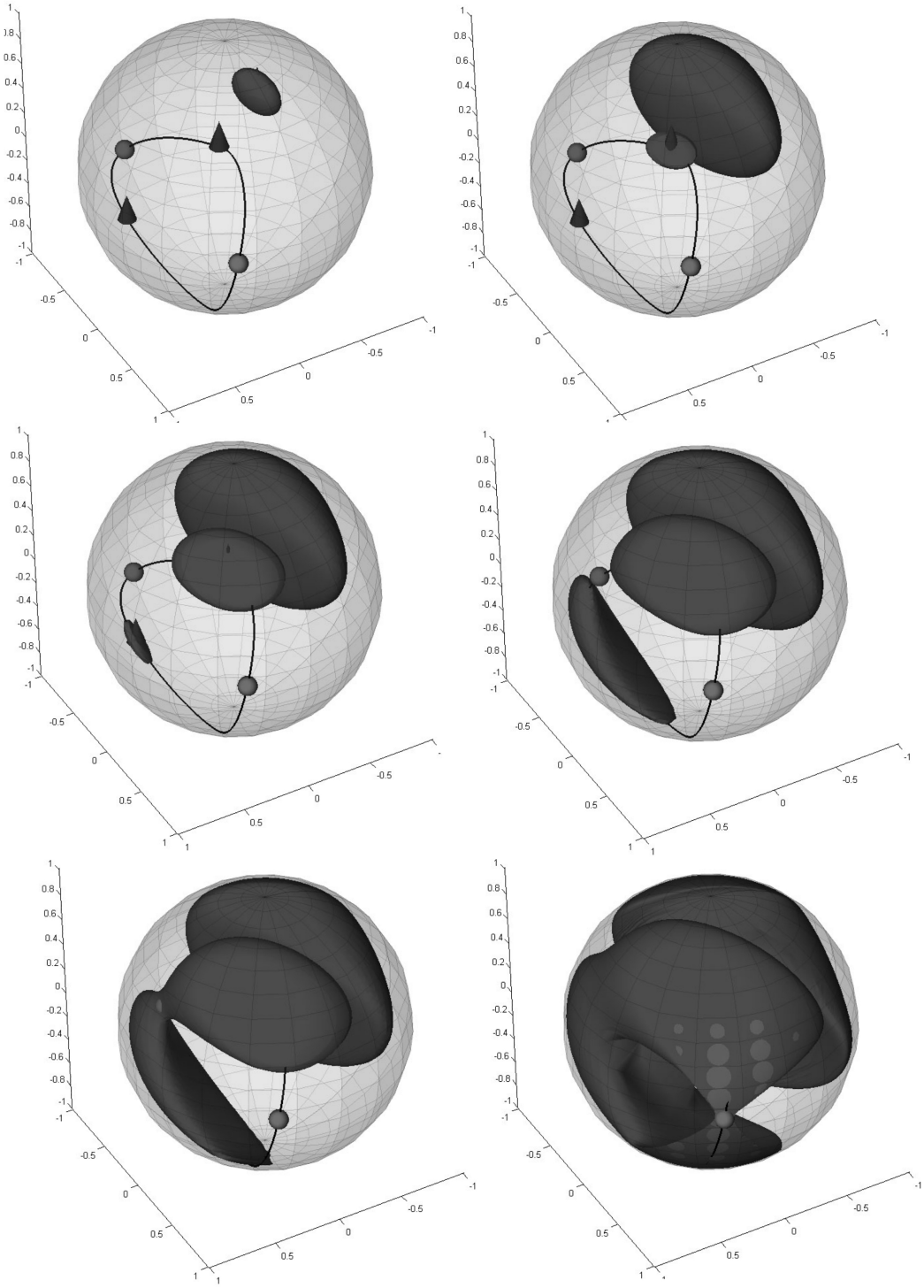


Figure 4.17: Evolution of the level sets of J as long as the level decreases.

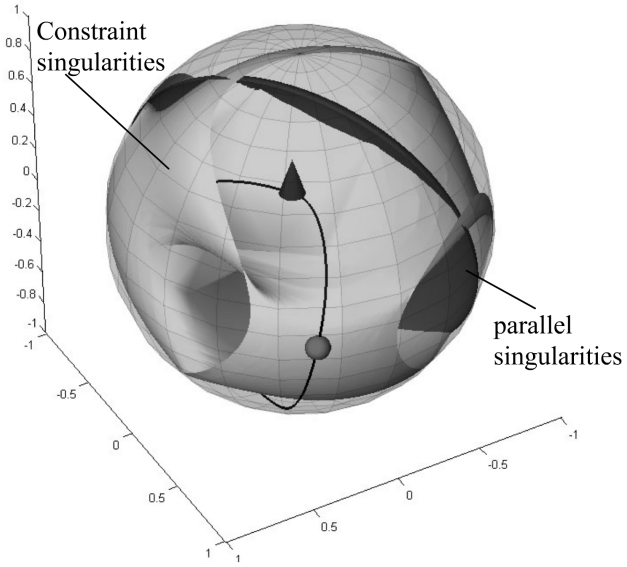


Figure 4.18: Locus defined by $J = 0$.

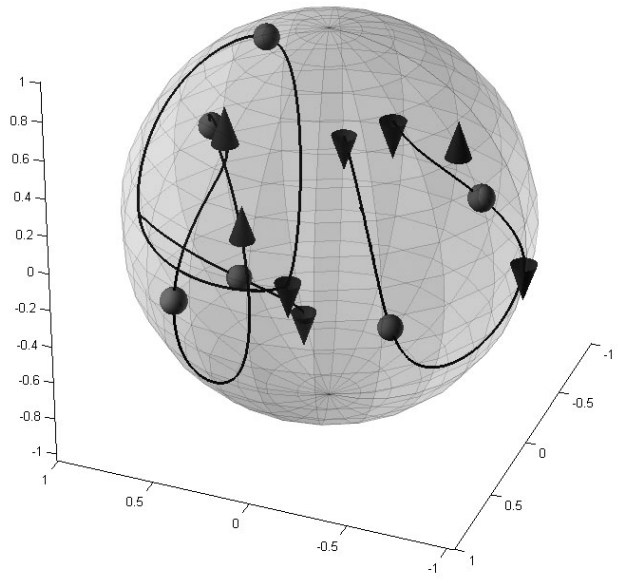


Figure 4.19: All relevant critical points of T_1.

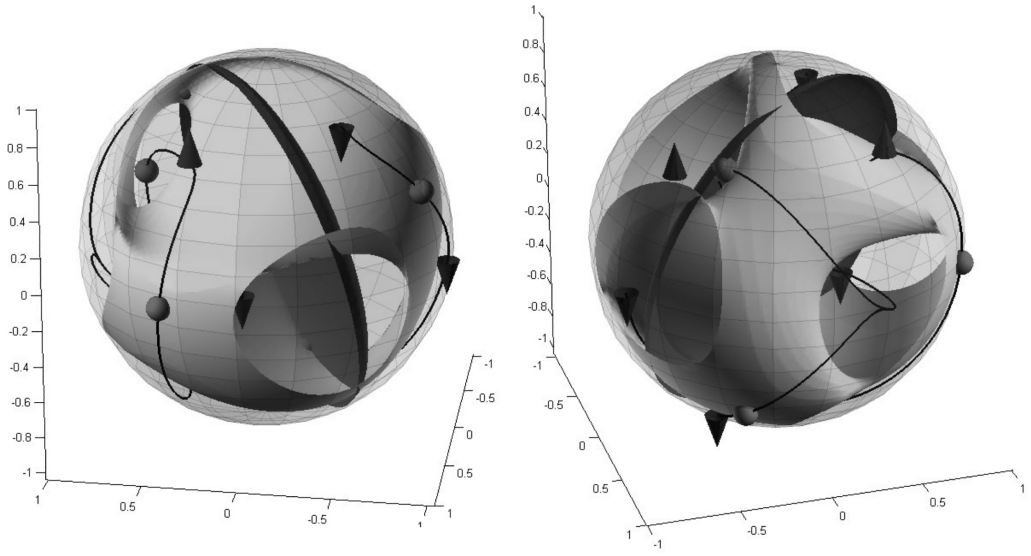


Figure 4.20: The steepest ascent and descent paths are all singularity-free.

4.3 3RRR Planar Manipulators

This section collects the results of the application of the analysis method developed in Chapter 2 to 3RRR planar manipulator, published in [66].

4.3.1 Architecture

A 3RRR planar manipulator with general structure is depicted in Figure 4.21. The platform is connected to the rigid frame through three legs, composed of two connecting rods and three revolute joints, with the middle one actuated. The center of the i^{th} leg revolute joint on the fixed frame is indicated by P_i, whereas the center of the i^{th} leg revolute joint on the platform is indicated by Q_i. The center of the actuated revolute joint of the i^{th} leg is denoted by R_i.

The kinematical structure of the platform can be determined through the three parameters u_2, u_3, and v_3, defining the coordinates of Q_1, Q_2, and Q_3 in the reference frame uQ_1v attached to the platform, as shown in Figure 4.21. Analogously, the kinematic structure of the fixed frame is given by the three parameters a_2, a_3, and b_3, defining the coordinates of P_1, P_2, and P_3 in the fixed reference frame xP_1y. The i^{th} leg can be defined through the lengths of the two connecting rods: l_i and m_i (see Figure 4.21). Thus twelve parameters are used to define a 3RRR manipulator.

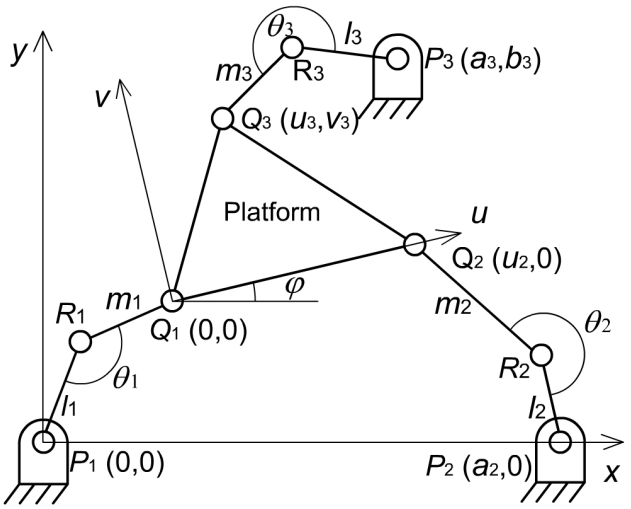


Figure 4.21: A 3RRR manipulator.

This class of planar manipulators have been widely studied, and often used as an example, due to its simple kinematic architecture. Workspace analysis methods for similar manipulators were proposed in [55] and [56], and the singularity locus of analogous manipulators was defined and studied in [57] and [58].

4.3.2 Configuration space: the assembly configurations

The workspace of a 3RRR planar manipulator is a subset of the manifold containing all possible positions of the platform in the plane. Each points of the workspace will be identified by the coordinates x and y of point Q_1 in the fixed reference frame xP_1y, and by the angle φ between x- and u-axes.

The position of the i^{th} actuator is given by the angle θ_i, between the two rods composing each leg. Any point in the jointspace is therefore identified by the three angles $(\theta_1, \theta_2, \theta_3)$.

Any configuration of the manipulator can be represented through the six parameters $(x, y, \varphi, \theta_1, \theta_2, \theta_3)$. However, not any combination of these six parameters identifies a configuration of the manipulator, for the ensuing constraints imposed by the three legs must be satisfied:

$$\mathbf{f} = \mathbf{0} \tag{4.42}$$

where $\mathbf{f} = (f_1, f_2, f_3)$, and

$$f_i = (P_i - Q_i(x, y, \varphi))^2 - l_i^2 - m_i^2 - 2l_i m_i \cos\theta_i \tag{4.43}$$

Equation (4.43) can be easily derived by expressing the coordinates of each Q_i in the fixed reference frame xP_1y, and by applying Carnot theorem to the three triangles $P_iQ_iR_i$.

The configuration space is thus represented as the three dimensional manifold $\mathscr{C}$ described by Equation (4.42) and embedded in the six dimensional manifold containing all the possible vectors $(x, y, \varphi, \theta_1, \theta_2, \theta_3)$.

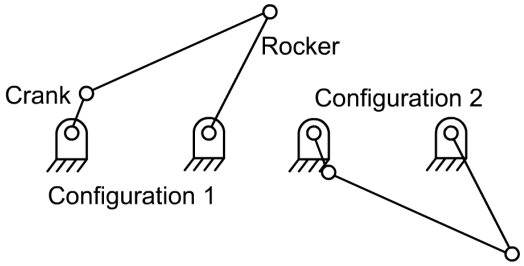


Figure 4.22: A Grashof four-bar linkage.

Unlike the manipulators presented so far, the configuration space of planar 3RRR manipulators coincides neither with the workspace nor with the jointspace. Moreover, $\mathscr{C}$ is not always connected, and might be composed of one or more disjoint regions. If two configurations belong to two different of such disjoint regions, then there are no feasible paths connecting them, which physically means that to bring the manipulator from one configuration to the other the kinematic chain must be dismantled. This problem is typical of parallel architectures. A classical single-degree-of-freedom example is Grashof four-bar linkage (see [59]), such as the one depicted in Figure 4.22. If the linkage of Figure 4.22 is at configuration 1, it is impossible to reach configuration 2 without disassembling the kinematic chain.

The disjoint regions composing the configuration space were named *assembly circuits* for single-degree-of-freedom mechanism in [60], where an interesting discussion is provided to discriminate circuits from branches. The name *circuit* can be understood by considering the four-bar linkage. If a configuration of the four-bar linkage is identified by means of the two angles θ_1 and θ_2, between the crank and the frame, and between the rocker and the frame, than the configuration space is described by an equation like $g(\theta_1, \theta_2) = 0$. If the set $g(\theta_1, \theta_2) = 0$ is plotted on the torus spanned by the two angles θ_1 and θ_2, the disjoint regions of the configuration space appear as closed curves, or, in other words, circuits. Many authors tackled the problem of counting the different assembly circuits in single-dof mechanisms. For example, refer to [60], [61], [62], and [63].

The denomination assembly configuration was introduced in [60], to generalize the notion of assembly circuit to multi-degree-of-freedom mechanisms. A criterion was given in [63] to determine the number of ACs composing the configuration space of any single-loop planar kinematic chain, which was proved to be at most two. A counting method for the ACs of two-dof multi-loop mechanisms was given in [64] and [65]. In Section 4.3.5 the method proposed in Chapter 2 will be extended, in order to be able to count and identify the assembly configurations of 3RRR planar manipulators, and to find a feasible path between any two configuratons, whenever it exists.

Furthermore, the Configuration space $\mathscr{C}$ might not be a smooth manifold, which hinders the application of the proposed method. This happens whenever the gradients of f_1, f_2, and f_3 are linearly dependent. These cases are very rare, however, should any such manipulator be analyzed, it will be detected while finding the critical points, and excluded from the analysis, as discussed in Section 4.3.4.

4.3.3 Singularity locus

In order to derive the equation of the singularity locus, the relationship between the virtual displacements of the platform and the actuators is needed. Such relationship is obtained by differentiating Equation (4.42):

$$\mathbf{A} \begin{pmatrix} \partial\theta_1 \\ \partial\theta_2 \\ \partial\theta_3 \end{pmatrix} + \mathbf{B} \begin{pmatrix} \partial x \\ \partial y \\ \partial\varphi \end{pmatrix} = 0 \tag{4.44}$$

where $\mathbf{A}$ and $\mathbf{B}$ are the Jacobian matrices of the constraints $\mathbf{f}$, with respect to jointspace and workspace variables, respectively. Parallel singularities occur when the platform can perform virtual displacements, even though all actuators are locked. Thus all parallel singularities must satisfy the following condition:

$$J(x, y, \varphi) = \det \mathbf{B} = 0 \tag{4.45}$$

Serial singularities occur when $\det \mathbf{A} = 0$, but they will be ignored, because they are not dangerous for the manipulator. The singularity locus is a two-dimensional manifold defined by the zero level-set of the function J, on the three-dimensional configuration space $\mathscr{C}$. In the next section, the method developed in Chapter 2 will be applied to the analysis of the parallel-singularity locus of 3RRR planar manipulators.

4.3.4 Analysis of singularity locus

According to Lagrange's optimization method, the Lagrangian function L can be defined as:

$$L(x, y, \varphi, \theta_1, \theta_2, \theta_3, \lambda_1, \lambda_2, \lambda_3) = J - \lambda_1 f_1 - \lambda_2 f_2 - \lambda_3 f_3 \tag{4.46}$$

where f_1, f_2, and f_3, are defined by Equation (4.43). The critical points of J constrained on $\mathscr{C}$ are the points where the gradient of L with respect to all its nine variables vanishes. By equating to zero the derivatives of L with respect to the i^{th} actuator angle θ_i, the ensuing equations are obtained:

$$\lambda_i \sin\theta_i = 0 \tag{4.47}$$

Therefore, the following four cases are obtained.

Case a: All Lagrange's multipliers are not equal to zero.

In this case the sine of the three angles θ_i must vanish (Equation (4.47)), thus all three legs are completely outstretched or folded-up, for i is equal to 0 or π. Such positions can be obtained by substituting all possible combinations of 0 and π into each θ_i of Equation (4.42), which is reobtained as derivatives of L with respect to Lagrange's multipliers.

By subtracting the first equation of Equation (4.42) from the last two, two linear equations in x and y are obtained. From these linear equations, x and y can be determined as functions of the sine and cosine of φ , and back substituted into the first of Equation (4.42), yielding a trigonometric equation in φ. This last equation is easily solved by expressing the sine and cosine of φ in function of $\tan\varphi/2$, and then solving the resulting single-variable polynomial equation. Lagrange's multipliers, which will be useful for the classification of critical points, can be determined through the remaining derivatives of L.

Case b: i^{th} Lagrange's multiplier is equal to zero.

In this case, only the sines of θ_j and θ_k vanish, with j and k different from i. Analogous to the previous case, two equations for x, y, and φ are obtained by substituting all possible combinations of 0 and π into θ_j and θ_k, in the j^{th} and the k^{th} equations of Equation (4.42). By subtracting one of such equations from the other, a linear equation in x is obtained:

$$g(y, \varphi)x - h(y, \varphi) = 0 \tag{4.48}$$

which yields x as a function of y and φ .

By equating to zero the derivatives of L with respect to x, y, and φ, the ensuing equation is obtained:

$$\mathbf{A}(x, y, \varphi) \begin{pmatrix} 1 \\ \lambda_i \\ \lambda_j \end{pmatrix} = \mathbf{0} \tag{4.49}$$

where $\mathbf{A}$ is a 3×3 matrix, whose columns contain the gradients of J, f_j, and f_k with respect to variables x, y, and φ . Equation (4.49) implies that the determinant of $\mathbf{A}$ must vanish, which yields the third condition in x, y, and φ . By substituting the expression of x obtained from Equation (4.48) into this equation and into the j^{th} equation of Equation (4.42), the variable x is eliminated, and two polynomial equations in y and the tangent of $\varphi/2$ are obtained, which can be easily solved through Sylvester dialytic elimination method (see Section 5.1).

Among the solutions just obtained, there are some extraneous solutions, which can be easily detected and got rid of, for at such solutions the two coefficients g and h of Equation (4.48) vanish. The angles θ_j and θ_k are equal to 0 or π , whilst the angle θ_i can be derived from the i^{th} equation of Equation (4.42). The j^{th} and k^{th} Lagrange's multipliers are obtained from Equation (4.49), and the i^{th} is obviously zero.

Case c: i^{th} and j^{th} Lagrange's multipliers vanish, but the k^{th} not.

By equating to zero the derivatives of L with respect to x, y, and φ, the ensuing equation is obtained:

$$\mathbf{B}(x, y, \varphi) \begin{pmatrix} 1 \\ \lambda_k \end{pmatrix} = \mathbf{0} \tag{4.50}$$

where $\mathbf{B}$ is a 3×2 matrix, whose columns contain the gradients of J and f_k with respect to variables x, y, and φ. Equation (4.50) implies that all the three 2×2 minors of $\mathbf{B}$ are singular, which yields three equations. By considering two of the three conditions just derived, along with the equation obtained by substituting 0 or π into θ_k in the k^{th} equation of Equation (4.42), three equations in the variables x, y, and φ are obtained, which can be solved analogously to case b.

It is possible to prove that, by equating to zero only two of the three 2×2 minor determinants, some extraneous solutions are introduced, which do not make the determinant of the third minor vanish. By imposing this last condition, it is possible to detect and get rid of such extraneous solutions.

Case d: All Lagrange's multipliers are equal to zero.

In this case the gradient of J with respect to x, y, and φ must vanish, which yields two linear equations in x and y, and a quadratic equation in x and y. This equation-set can be solved by techniques analogous to case a.

In this way, all possible critical points are determined. Note that this solution method finds all the points where the gradient of J and the gradients of the constraints with respect to the six configuration variables are linearly dependent. Thus, if at any configuration the gradients of the three constraints are linearly dependent, i.e. the configuration space is not smooth, this configuration is included among the critical points. In order to assess whether $\mathscr{C}$ is smooth, and this method is applicable, as pointed out in Section 4.3.2, it is therefore sufficient to check that, at all critical points, the gradients of the three constraints are not linearly dependent.

Once the critical points are determined, they are all classified by methods analogous to Section 4.1.4. The maximum increase and decrease directions, in the neighborhood of a critical point, are the six-dimensional vectors $\mathbf{a}_1$ satisfying the ensuing eigenvalue problem:

$$(\mathbf{H}^* - \alpha \mathbf{I}^*)\mathbf{a}^* = \mathbf{0} \tag{4.51}$$

where:

- $\mathbf{H}^*$ is the Hessian matrix of the Lagrangian L, calculated with respect to all its nine variables, at the critical point to be classified.
- $\mathbf{I}^*$ is equal to the 9×9 identity matrix, with the last three diagonal elements set equal to zero.
- $\mathbf{a}^*$ is a nine-dimensional vector obtained by appending three dummy variables to $\mathbf{a}_1$.

The equation

$$\det(\mathbf{H}^* - \alpha \mathbf{I}^*) = 0 \tag{4.52}$$

is a third order polynomial in α. Positive and negative solutions to Equation (4.52) represent maximum increase and maximum decrease directions, therefore the number of negative solutions is the index of the critical point. The maximum increase or decrease directions corresponding to each solution can be determined through Equation (4.51).

The steepest ascent or descent paths can be generated starting form 2- or 1-saddles along the maximum increase or decrease eigenvector. Between the critical points, the steepest ascent direction is obtained by projecting the gradient upon the tangent space of the constraints, as in Section 4.1.4. With

the steepest ascent direction at any point, steepest ascent or descent paths are easily generated.

4.3.5 Analysis of assembly configurations

The method proposed in Chapter 2 can be extended, in order to count the assembly configurations, too. In Chapter 2 the evolution of the set $\mathscr{C}_+^a$ has been considered, with level a decreasing form the absolute maximum of J down to zero.

Suppose now that the level a keeps on decreasing, below zero level. The process of generation and joining of disjoint regions continues just the same as above zero level: the negative maxima generate new disjoint regions, whereas negative 2-saddles may join existing disjoint regions, but now if the steepest ascent paths starting from negative 2-saddles reach positive maxima, they are not singularity-free anymore. However, they are still feasible paths, even though dynamic-control techniques shall be devised in order to keep the platform under control while crossing parallel singularities.

There must exist an absolute minimum of the function J on $\mathscr{C}$, for $\mathscr{C}$ is compact and J is continuous. As soon as level a reaches the absolute minimum level, the manifold $\mathscr{C}_+^a$ coincides with the whole configuration space $\mathscr{C}$. Therefore, the disjoint regions composing $\mathscr{C}_+^a$ are indeed the assembly configurations composing the whole configuration space $\mathscr{C}_+^a$. As for the singularity-free regions, each assembly configuration is endowed with a set of maxima of the function J, which completely defines it. All the maxima contained in the same assembly configuration are connected through a network of steepest ascent feasible paths.

In order to assess whether two points belong to the same assembly configuration, the steepest ascent paths starting from such points can be followed, until any of the maxima is reached. If the two maxima belong to the same assembly configuration, then there exists at least one feasible path connecting them, which can be obtained by joining the steepest ascent paths from the two points to the reached maxima, and any of the feasible paths among the network connecting the two maxima. This path is singularity-free only if the two points also belong to the same singularity-free region which can be assessed through the method described in Chapter 2.

This process can be analogously repeated for the manifold $\mathscr{C}_-^a$, letting the level a increase from the absolute minimum to the absolute maximum. This second procedure is redundant for the purpose of determining the assembly configurations, but it might be useful to find out which negative singularity-free regions belong to which assembly configuration, and to cross-check the results hitherto obtained.

The procedure just described can be summarized in the following steps:

1. All critical points of the Jacobian determinant J on the configuration space $\mathscr{C}$ are determined.
2. The critical points are classified into positive and negative maxima and into positive and negative 1- and 2-saddles.
3. The two steepest ascent paths are followed, starting from each positive 2-sadlle up to two positive maxima. The two positive maxima, and any maxima belonging to their singularity-free region are assigned to the same singularity-free region. After all the positive 2-saddles have been processed, the positive maxima belonging to each positive singularity-free region are stored.
4. The two steepest ascent paths are followed, starting from each negative 2-saddle up to two maxima. The two maxima, and any maxima belonging to their assembly configuration are assigned to the same assembly configuration. After having processed all negative 2-saddles, the maxima of each assembly configuration are stored.
5. Step 3) is repeated, suitably modified, for the negative singularity-free regions, to find the negative minima contained in each negative singularity-free regions.
6. Step 4) is repeated, suitably modified, for the positive 1-saddles, to find the minima contained in each assembly configuration.

At the end of this procedure, it is always possible to assess whether a feasible or a singularity-free path exists between any two configurations of the manipulator, by simply following the steepest ascent or descent paths starting from the two points and considering the maxima they reach.

4.3.6 Examples

a_2	a_3	b_3	u_2	u_3	v_3	l_1	m_1	l_2	m_2	l_3	m_3
10	3	10	10	3	3	1	2	10	2	6	7

Table 4.5: DH-parameters of manipulator P_1

Consider manipulator P_1, whose kinematical structure is summarized in Table 4.5, according to the parameterization defined in Section 4.3.1.

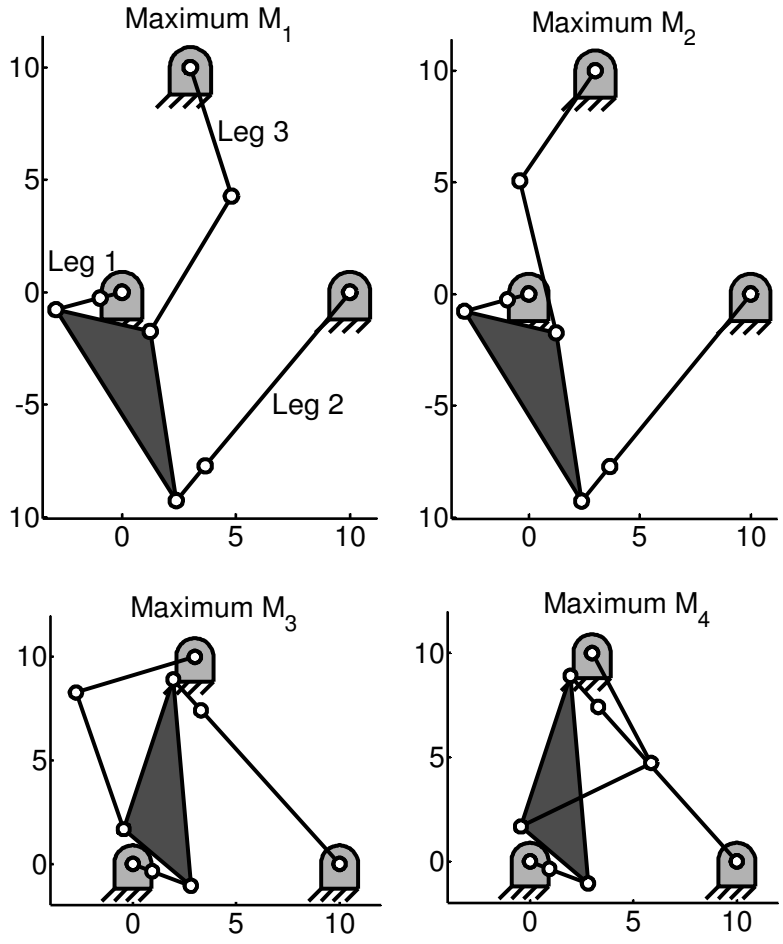


Figure 4.23: The four positive maxima of manipulator P_1.

For P_1, there are four positive maxima, nine positive 2-saddles, four negative 2-saddles, and no negative maxima. The four positive maxima are shown in Figure 4.23. Maxima M_1 and M_2 are joined by steepest ascent paths starting from some of the positive 2-saddles, while M_3 and M_4 are not connected to any other maximum by any steepest ascent path starting from any positive or negative 2-saddle. Therefore, there are three assembly configurations: one containing M_1 and M_2, and the other two containing M_3 and M_4. P_1 was generated by searching a manipulator such that the loop composed by leg 1, leg 2, the platform and the frame has two assembly configurations, according the condition derived in [63], and that leg 3 be able to completely outstretch in one of such assembly configuration, but not in the other. Therefore one of the two assembly configurations of the loop is split into two assembly configurations by the fact that leg 3 can never outstretch, nor fold back.

The analysis of the negative critical points shows that each of the tree assembly configurations is split into two singularity-free regions, one positive, and one negative. Therefore, if the sign of the Jacobian determinant is the same at two configurations belonging to the same assembly configuration, a singularity-free path connecting them always exists.

a_2	a_3	b_3	u_2	u_3	v_3	l_1	m_1	l_2	m_2	l_3	m_3
10	3	10	10	3	3	1	2	4	2	5	6

Table 4.6: DH-parameters of manipulator P_2

However, the assembly configurations are not always split into two singularity-free regions only, as manipulator P_2 shows. In P_2 there is only one assembly configuration, therefore any configuration of the manipulator is reachable, but this assembly configuration is split into four singularity-free regions, three positive and one negative. Figure 4.24 shows three positive maxima belonging to the three positive singularity-free regions: many feasible paths connect these three configurations where the Jacobian determinant is positive, but none of them is free of parallel singularities.

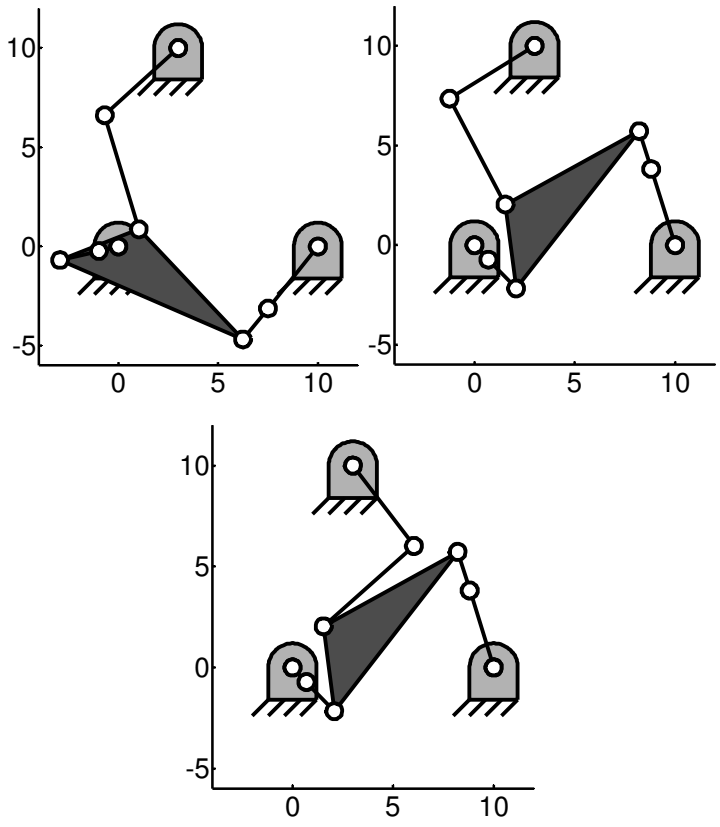


Figure 4.24: Three positive maxima of manipulator P_2, belonging to three different singularity-free regions.

Chapter 5

Solution Methods for polynomial Equations

As shown in the previous Chapters, the application of the proposed method always needs the determination of all critical points of a function on a manifold. In case of the robot manipulators presented above, this problem always reduces to finding all roots of a polynomial equation-set. In this Chapter, some of the standard methods that have been used to solve polynomial equations are reported and discussed.

5.1 Algebraic elimination methods

The most classical solution methods for polynomial equations make use of algebraic elimination, whose application to some standard problems will be recalled in this section. For further details, refer for example to [67].

We suppose to deal with a system on n polynomial equations. The standard method for algebraic elimination is then to proceed with partial homogenization, with respect to $n-1$ variables, and to hide the remaining one in the coefficients. After partial homogenization, a system of n homogeneous polynomial equations in n variables is available, whose coefficients depend on the hidden variable. A condition on the coefficients of the homogeneous system is now required, in order for the n homogeneous equations to have a common solution in the n variables.

The target of algebraic elimination is to find out enough different equivalent equations to the original homogeneous system to write it as:

$$\mathbf{M}\mathbf{y} = \mathbf{0} \tag{5.1}$$

where $\mathbf{M}$ is a square matrix whose elements depend on the hidden variable only, and $\mathbf{y}$ is in general a vector containing monomials obtained as products

of some homogeneous variables. The homogeneous variables cannot be all equal to zero simultaneously, thus, if $\mathbf{y}$ is properly built, $\mathbf{y}$ never vanishes, too.

Eventually, if it is possible to write the original system in this desired form, Equation (5.1) admits a set of common homogeneous solutions if

$$\det \mathbf{M} = \mathbf{0} \tag{5.2}$$

which is the desired condition in the coefficients, and therefore in the hidden variable. $\det \mathbf{M}$ is also named the *resultant* of the polynomial equation set.

In the rest of this section, some particular cases will be considered. It will be supposed that partial homogenization has already been performed, and it will be shown how the homogeneous system can be written in the form of Equation (5.1).

5.1.1 Sylvester's dialytic elimination method

This method applies to two homogeneous polynomial equations in two variables. Consider, for example, the ensuing two polynomial equations:

$$\begin{aligned} p_1(x_0, x_1) = a_{m0}x_1^m + a_{(m-1)1}x_1^{(m-1)}x_0 + \ldots + a_{0m}x_0^m &= 0 \\ p_2(x_0, x_1) = b_{n0}x_1^n + b_{(n-1)1}x_1^{(n-1)}x_0 + \ldots + b_{0n}x_0^n &= 0 \end{aligned} \tag{5.3}$$

where x_0 and x_1 are the homogeneous variables and a_{ij} and b_{ij} are the coefficients of the monomial $x_1^i x_0^j$ in the first and the second equations respectively, which contain the hidden variable. The first and the second equations are of degree m and n respectively.

Consider now the n equations obtained by multiplying p_1 by the n monomials of degree $(n-1)$, $x_1^{(n-1)}, x_1^{(n-2)}x_0, \ldots, x_1x_0^{(n-2)}, x_0^{(n-1)}$, and the m equations obtained by multiplying p_2 by the m monomials of degree $(m-1)$, $x_1^{(m-1)}, x_1^{(m-2)}x_0, \ldots, x_1x_0^{(m-2)}, x_0^{(m-1)}$. In this way, $m+n$ equations of degree $m+n-1$ have been obtained, which are satisfied by all solutions to the original Equation (5.3). The $m+n$ equations just obtained can be written in the form:

$$\mathbf{M} \begin{pmatrix} x_1^{(m+n-1)} \\ x_1^{(m+n-2)}x_0 \\ \vdots \\ x_1x_0^{(m+n-2)} \\ x_0^{(m+n-1)} \end{pmatrix} = \mathbf{0} \tag{5.4}$$

Since all $m+n$ monomials $x_1^i x_0^j$ of degree $m+n-1$ cannot vanish simultaneously, Equation (5.4) leads to the desired condition in the coefficients a_{ij}

and b_{ij} of the two equations:

$$\det \mathbf{M} = 0 \tag{5.5}$$

If the coefficients hide a nonhomogenized variable, then Equation (5.5) is the resultant polynomial equation in that variable, that can be solved numerically.

Whenever Equation (5.5) is verified, then Equation (5.3) is verified for a point (x_0, x_1) in the projective space. This point can be found by posing $x_0 = 1$ in Equation (5.4), which yields the ensuing linear system:

$$\mathbf{M}^* \begin{pmatrix} x_1^(m+n-1) \\ x_1^{(m+n-2)} \\ \vdots \\ x_1 \end{pmatrix} = -\mathbf{m}^* \tag{5.6}$$

where $\mathbf{M}^*$ is a matrix obtained by $\mathbf{M}$ through removal of the last column and of one of its rows, and $\mathbf{m}^*$ is the last column of $\mathbf{M}$ with the same row removed. The solution of this linear system yields the powers of the unknown variable x_1, and the last unknown is the desired value of x_1 that is associated to $x_0 = 1$. Should Equation (5.6) possess no solution, then the imposition $x_0 = 1$ is wrong, which implies that $x_0 = 0, x_1 = 1$ is a solution to Equation (5.3).

5.1.2 Sylvester's elimination method for three equations

Consider the ensuing three homogeneous polynomial equations in three variables, with the same degree n:

$$\begin{aligned} p_1(x_0, x_1, x_2) = a_{n00}x_2^n + a_{(n-1)10}x_2^{(n-1)}x_1 + \ldots + a_{00n}x_0^n &= 0 \\ p_2(x_0, x_1, x_2) = b_{n00}x_2^n + b_{(n-1)10}x_2^{(n-1)}x_1 + \ldots + b_{00n}x_0^n &= 0 \\ p_3(x_0, x_1, x_2) = c_{n00}x_2^n + c_{(n-1)10}x_2^{(n-1)}x_1 + \ldots + c_{00n}x_0^n &= 0 \end{aligned} \tag{5.7}$$

where x_0, x_1 and x_2 are the homogeneous variables and a_{ijk}, b_{ijk} and c_{ijk} are the coefficients of the monomial $x_2^i x_1^j x_0^k$ in the first, the second and the third equation respectively.

Let the degree n of the two equations be equal to two. In this case, the ensuing result can be exploited to generate more equivalent homogeneous equations of degree two:

Theorem 5.1 *Let* $\mathbf{p}(\mathbf{x})$ *be a homogeneous n-dimensional function in* $\mathbf{x}$*, with all components of the same degree. Let* $J(\mathbf{x})$ *be the determinant of the Jacobian matrix of* $\mathbf{p}$*, calculated at point* $\mathbf{x}$*, and* $\nabla J|_{\mathbf{x}}$ *be the gradient of* J *calculated at point* $\mathbf{x}$*. If the point* $\tilde{\mathbf{x}}$ *is a solution to the set of homogeneous equations* $\mathbf{p}(\mathbf{x}) = \mathbf{0}$ *of the same degree, then* $\tilde{\mathbf{x}}$ *is also a solution to the homogeneous equations:*

$$J(\tilde{\mathbf{x}}) = 0 \tag{5.8}$$
$$\nabla J|_{\tilde{\mathbf{x}}} = \mathbf{0} \tag{5.9}$$

For a proof of Theorem 5.1 refer to [67] p.84.

If the degree of $\mathbf{p} = (p_1, p_2, p_3)^T$ is equal to two, as the case at hand, then the elements of its Jacobian matrix are linear in the homogeneous variables, and the degree of the Jacobian determinant is three. By virtue of Theorem 5.1, the ensuing second-order equations must hold at each solution of Equation (5.7):

$$\nabla J|_{(x_0,x_1,x_2)} = \mathbf{0} \tag{5.10}$$

By adding to Equation (5.10) the original equation set, the ensuing set of six second-order equations is obtained:

$$p_1(x_0, x_1, x_2) = 0$$
$$p_2(x_0, x_1, x_2) = 0$$
$$p_3(x_0, x_1, x_2) = 0 \tag{5.11}$$
$$\nabla J|_{(x_0,x_1,x_2)} = \mathbf{0} \tag{5.12}$$

Since there are exactly six homogeneous second-order monomials of degree 2 in three variables, Equation (5.7) can be rewritten as follows:

$$\mathbf{M} \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_0x_1 \\ x_1x_2 \\ x_0x_2 \end{pmatrix} = \mathbf{0} \tag{5.13}$$

which obviously yields the desired resultant equation $\det \mathbf{M} = 0$. Whenever the resultant equation is satisfied, the set of values for x_0, x_1, and x_2 that satisfy Equation (5.13) can be found in analogous manner as Section 5.1.1: $x_0 = 1$ is substituted into Equation (5.13) and the resulting linear equation is solved.

If the system $\mathbf{p} = \mathbf{0}$ is composed of three equations of degree three, Theorem 5.1 can be again used to find the resultant. In this case, the elements of the Jacobian matrix are of degree two, and the Jacobian determinant is a sixth-order polynomial. By imposing that the gradient of the Jacobian determinant vanishes, three additional equations of degree five are obtained.

If each third-order equation of the original set is multiplied by the six second-order monomials in three variables, 18 equations of degree five are obtained. By adding the gradient of the Jacobian determinant, a set of 21 equations is obtained. Since there are exactly 21 fifth-order monomials in three variables, the set of 21 homogeneous equations can be written as:

$$\mathbf{M}\mathbf{y} = 0 \tag{5.14}$$

where $\mathbf{M}$ is 21×21 matrix containing the coefficients of the 21 fifth-order equations, and $\mathbf{y}$ is a vector containing all 21 monomials of degree five, that can never vanish simultaneously. The determinant of $\mathbf{M}$ is again the desired resultant polynomial.

If the degree of the three equations is greater than three, it is not possible to exploit the gradient of the Jacobian determinant anymore. However, a general algebraic elimination method, also due to Sylvester, is still available.

Suppose for instance that the equation-set is composed of three homogeneous equations of degree four, as occurred in Sections 4.1.4 and 4.2.4. 18 equations of degree six can be obtained by multiplying each of the three equations by the six second-order monomials $x_1^2, x_2^2, x_3^2, x_0x_1, x_1x_2, x_0x_2$. Since the number of monomials of degree six is equal to 28, ten more equations are needed, in order to write an equation analogous to Equation (5.14).

The three fourth-order equations can be decomposed in the ensuing form:

$$\begin{aligned} u_1x_0^4 + v_1x_1 + w_1x_2 &= 0 \\ u_2x_0^4 + v_2x_1 + w_2x_2 &= 0 \\ u_3x_0^4 + v_3x_1 + w_3x_2 &= 0 \end{aligned} \tag{5.15}$$

where u_i, v_i, and w_i are polynomials in the homogeneous variables.

Equation (5.15) takes the ensuing matrix form:

$$\begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix} \begin{pmatrix} x_0^4 \\ x_1 \\ x_2 \end{pmatrix} \tag{5.16}$$

Since the homogeneous monomials x_0^4, x_1, x_2 cannot vanish simultaneously, the ensuing condition must hold:

$$\det \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix} = 0 \tag{5.17}$$

Equation (5.17) is a sixth-order polynomial equation, that can be added to 18 already found. Through a similar procedure, by decomposing the three original equations into a form like the following:

$$\begin{aligned} u_1 x_0^\alpha + v_1 x_1^\beta + w_1 x_2^\gamma &= 0 \\ u_2 x_0^\alpha + v_2 x_1^\beta + w_2 x_2^\gamma &= 0 \\ u_3 x_0^\alpha + v_3 x_1^\beta + w_3 x_2^\gamma &= 0 \end{aligned} \tag{5.18}$$

(5.19)

where $\alpha + \beta + \gamma = 6$, for all possible combinations, the remaining nine equations are found.. In this way, 28 sixth-order equations in the 28 sixth-order monomials are available, and straightforwardly lead to the elimination of the homogeneous variables.

This method can be directly generalized to the case of n equations in n unknowns of the same degree. For more details see [67], page 86.

5.2 Homotopy Continuation Method

Homotopy continuation method is a method for solving polynomial equations that has been widely applied in many technological and scientific fields. In particular, in robotics it has been used to solve direct and inverse kinematic problems for many parallel and serial manipulators, whenever a specific elimination method was not available. Homotopy continuation method will be hereafter outlined, for more details refer to [68], [69], and [70].

The core idea of homotopy continuation method is to start form an easily solvable system of polynomial equations, and then smoothly and continuously deform it into a desired, hardly-solvable system of polynomial equations. The solutions of the easily solvable system will then be smoothly and continuously deformed into the desired solutions of the complicated one.

Consider, for instance, the ensuing second order equation in one variable:

$$p(x) = x^2 - 5x + 6 = 0 \tag{5.20}$$

and the factored equation:

$$p_s(x) = x(x-1) = 0 \tag{5.21}$$

whose solutions are obviously $x = 0$ and $x = 1$.

Consider now the ensuing linear combination of the previous two equations:

$$h(x, \lambda) = \lambda p(x) + (1 - \lambda) p_s(x) = 0 \tag{5.22}$$

Clearly, h is a homotopy function between p_s and p: h is continuous ,$h(x,0) = p_s(x)$, and $h(x,1) = p(x)$. While λ varies smoothly from 0 to 1, the function h turns smoothly from p_s into p. Also the solutions to the polynomial equation $h(x,\lambda) = 0$ will move along smooth paths from the solutions to $p_s = 0$, $x = 0$ and $x = 1$, to the solutions of $p = 0$, $x = 2$ and $x = 3$, while λ continuously changes from zero to one. Such smooth paths can be followed by imposing small increments $d\lambda$ on λ, and then by searching for solutions to $h(x, \lambda + d\lambda) = 0$ close to the solutions of $h(x,\lambda) = 0$, that are known from the previous step. In this way, step by step, the solutions of $p_s(x) = 0$ are turned into solutions of $p(x) = 0$, with no actual need of solving this last equation.

If p_s is chosen as a third-order equation, instead of a second-order one, then the starting equation will have three solutions instead of two. Equation $h(x,\lambda) = 0$ is at the beginning a third-order equation, but, as long as λ approaches one, the coefficient of x^3 approaches zero, therefore one solution diverges to infinity, whereas the remaining two approach to the desired two solutions of $p = 0$.

On the other hand, if p_s is chosen as a linear function, then the starting equation will have only one solution instead of two. Equation $h(x,\lambda) = 0$ is initially linear, but, as soon as λ becomes greater than zero, the coefficient of x^2 will not vanish, and a second solution will appear, and evolve from infinity down to one of the solutions of $p(x) = 0$ while λ approaches one. Therefore, if only the single solution of the initial linear equation is tracked, only one solution to $p(x) = 0$ is found, while the other one is lost.

It is therefore crucial to start from an initial equation p_s with the same degree as the equation to be solved. If the degree of the initial equation is greater than that of the equation to be solved, then some of the paths starting from its solutions lead to infinity, which is a useless loss of computational time, because no actual solutions are found. On the other hand, if the degree of the initial equation is lesser, some solutions are lost, which is unacceptable for the purposes of this work, because all critical points must always be found.

Analogous problems arise when more equations are considered, but, unfortunately, the notion of "degree" of an equation-set is not straightforwardly generalized from the one-dimensional case. Like the one-dimensional example, the starting equation-set for a polynomial continuation must have a number of finite solutions not lesser than the equation-set to be solved, lest some solutions are lost. However, the number of solutions must be as close as possible to that of the final equation-set, or lots of computational effort is lost while tracking paths leading to infinity.

We consider as an example the equation-set of Section 3.2.2, that is used to find all critical points of the Jacobian determinant on the Jointspace

of a 6R serial manipulator. The gradient of the Jacobian determinant of a 6R manipulator can be written as polynomial equation in the variables $s_1, c_1, s_2, c_2, s_3, c_3, s_4$, and c_4, where s_i and c_i are the sine and the cosine of the i^{th} joint angle θ_i. Imposing that the gradient vanishes leads to the equation-set $\mathbf{p} = \mathbf{0}$, i.e.:

$$\begin{aligned} p_1 = g_1(s_1, c_1, s_2, c_2, s_3, c_3, s_4, c_4) &= 0 \\ p_2 = g_2(s_1, c_1, s_2, c_2, s_3, c_3, s_4, c_4) &= 0 \\ p_3 = g_3(s_1, c_1, s_2, c_2, s_3, c_3, s_4, c_4) &= 0 \\ p_4 = g_4(s_1, c_1, s_2, c_2, s_3, c_3, s_4, c_4) &= 0 \end{aligned} \tag{5.23}$$

$$\begin{aligned} p_5(s_1, c_1) = s_1^2 + c_1^2 - 1 &= 0 \\ p_6(s_2, c_2) = s_2^2 + c_2^2 - 1 &= 0 \\ p_7(s_3, c_3) = s_3^2 + c_3^2 - 1 &= 0 \\ p_8(s_4, c_4) = s_4^2 + c_4^2 - 1 &= 0 \end{aligned} \tag{5.24}$$

The first four components of $\mathbf{p}$ are equal to g_1, g_2,g_3, and g_4, i.e. the components of the gradient of the Jacobian determinant. Due to the structure of the Jacobian determinant, these first four equations are of degree 1 with respect to variables s_1 and c_1, of degree 2 with respect to variables s_2 and c_2, of degree 2 with respect to variables s_3 and c_3, and of degree 1 with respect to variables s_1 and c_1.

The Bezout degree of Equation (5.23) is equal to 20736, because it is the product of the degrees of all eight equations, and the first four are of degree six, whereas the last four are of degree two. By virtue of Bezout theorem, the number of finite solutions to Equation (5.23) will be at most 20736, therefore any equation-set with the first four equations of degree six and the last four equations of degree two will be a possible starting system for a homotopy continuation method.

However, most solutions would lead to infinity, because the structure of the first four equations of Equation (5.23) is particular. For example, no variable appears with degree greater than two, whereas an equation of degree 6 contains in general the sixth powers of all variables. The structure of Equation (5.23) must be better characterized, in order to find a starting system with a lower number of solutions.

This characterization can be found via multi-homogenization methods (see [71], [72], and [73]). Suppose that a partition in the set of variables is introduced, for example we consider the following sets of variables: $\{s_1, c_1\}$, $\{s_2, c_2\}$, $\{s_3, c_3\}$, and $\{s_4, c_4\}$. Then we can multi-homogenize the equation-set: we replace s_i and c_i with x_i/z_i and y_i/z_i respectively, and clear all the

denominators. A multi-homogeneous equation-set is obtained, i.e. a set of equations homogeneous with respect to the triplets of variables x_i, y_i, and z_i.

It can be proved that any multi-homogeneous equation-set, homogeneous with respect to the same triplets of variables, has the same number of solutions in the multi-homogeneous projective space as Equation (5.23), if it has the same degree with respect to all multi-homogeneous triplets of variables.

For example,consider the equation-set

$$\mathbf{p}_s = (q_1, q_2, q_3, q_4, r_1, r_2, r_3, r_4)^T = \mathbf{0} \tag{5.25}$$

The elements q_i are defined as:

$$\begin{aligned} q_i = \quad & a_{1i}(x_1, y_1, z_1)a_{2i}(x_2, y_2, z_2)a_{3i}(x_2, y_2, z_2) \\ & a_{4i}(x_3, y_3, z_3)a_{5i}(x_3, y_3, z_3)a_{6i}(x_4, y_4, z_4) \end{aligned} \tag{5.26}$$

where the a_{ji} are homogeneous linear functions of their variables, with randomly generated coefficients. Each q_i has degree 1 with respect to the group of variables $\{x_1, y_1, z_1\}$, degree 2 with respect to the group of variables $\{x_2, y_2, z_2\}$, degree 2 with respect to the group of variables $\{x_3, y_3, z_3\}$, and degree 1 with respect to the group of variables $\{x_4, y_4, z_4\}$, exactly as the first four equations of Equation (5.23).

The elements r_i are defined as:

$$r_i = b_{1i}(x_i, y_i, z_i)b_{2i}(x_i, y_i, z_i) \tag{5.27}$$

where the b_{ji} are homogeneous linear functions of their variables, with randomly generated coefficients. Each r_i has degree 2 with respect to the variable group $\{x_i, y_i, z_i\}$, exactly as the last four equations of Equation (5.23).

Therefore Equation (5.25) and Equation (5.23) have the same structure in the multi-homogeneous projective space defined by the considered group of variables, where they have the same number of solutions. Furthermore, Equation (5.25) is easily solvable, for its equations are all products of linear factors, and the number of such solutions is 1536, which is far lesser than the number of solutions estimated through Bezout theorem.

Then the ensuing homotopy function can be constructed

$$\mathbf{h}(\mathbf{x}, \lambda) = \lambda \mathbf{p}(\mathbf{x}, \lambda) + (1 - \lambda)\mathbf{p}_s(\mathbf{x}, \lambda) \tag{5.28}$$

where $\mathbf{x}$ is a vector containing all multi-homogenized coordinates. The solutions of the starting equation-set (5.26) can all be determined, and by means of small increments of λ and of Newton method such solutions can be smoothly brought into the solutions of Equation (5.23).

Notwithstanding the closest guess of the number of finite solutions of the original system, Equation (5.23) admits less than 1536 solutions, therefore some paths might diverge to infinity. It is therefore more convenient to track paths using the homogenized variables: in this way paths diverging to infinity simply converge to solutions where one or more variables z_i vanish, but the norm of the homogeneous triplets can always be kept equal to 1. Thanks to this trick, the distance between the solutions at two nearby steps never approaches infinity, which might cause numerical problems. In the case of the example reported in Section 3.2.2, 1120 finite solutions have been found, while 416 solutions approached infinity. However, although 416 paths have been uselessly followed, all solutions to Equation (5.23) have been found.

Chapter 6

Conclusion

This work presented a numerical method able to count and identify the singularity-free regions carved by the singularity locus in the configuration space of a manipulator, and its application to some serial and parallel manipulators.

In principle, this method works for any manipulator, but some very particular cases, where there are degenerate critical points of the Jacobian determinant, or the configuration space is not a smooth manifold. The application is rather simple, except the determination of all critical points of the Jacobian determinant on the configuration space. This part of the procedure reduces in most cases to the determination of all solutions to a polynomial equation-set, that might be a very hard task in practice, although it is always theoretically possible.

However, if the determination of the critical points of the Jacobian determinant is viable, like the presented examples, the proposed method represents a stable and powerful tool for analyzing the topology of the singularity locus and for planning singularity-free paths of the manipulator.

The proposed method does not take into account the possible reduction of configuration space of a manipulator due to the mechanical interference between the links. The analysis of the singularity locus under the additional constraint that no collision between the links takes place is a possible future development of the proposed method, as well as its application to more parallel manipulators with six degrees of freedom.

Bibliography

[1] Dugundji, J.; 1966, "Geometry, Topology, and Physics,"
Institute of Physics Publishing, London.

[2] Nakahara, M.; 1990, "Algebraic Topology,"
Cambridge University Press.

[3] Hatcher, A.; 2002, "Algebraic Topology,"
Cambridge University Press.

[4] Whitehead, G.,W.; 1978, "Elements of Homotopy Theory,"
Graduate texts in Mathematics, Vol.61, Springer-Verlag.

[5] Milnor, J.; 1963, "Morse Theory,"
Annals of Mathematics Studies, Study 51, Princeton University press.

[6] Riemann, B.; 1851, "Grundlagen für eine allgemeine Theorie der Functionen einer veränderlichen complexen Grösse,"
Inauguraldissertation, Göttingen, available on the web at
http://www.emis.de/classics/Riemann/Grund.pdf.

[7] Graves, L.M.; 1956, "The Theory of functions of Real Variables,"
McGraw-Hill Book Company, Inc.

[8] Wenger, P.; 2004, "Uniqueness Domains and Regions of Feasible Paths for Cuspidal Manipulators,"
IEEE Transactions on Robotics, Vol.20, No.4, pp.-745-750

[9] Burdick, J.W., 1988, "Kinematic Analysis and Design of Redundant Robot Manipulators,"
Ph.D. Thesis, Department of Mechanical Engineering, Stanford University.

[10] Parenti Castelli, V., Innocenti, C.; 1988, "Spatial Open Kinematic Chains: Regions and Subregions,"

Proceedings of the 7th CISM IFToMM Symposium on Theory and Practice of Robots and Manipulators: RoManSy 8, Udine, Italy.

[11] Parenti Castelli, V., Innocenti, C.; 1988, "Position Analysis of Robot Manipulators: Regions and Subregions,"
Advances in Robot Kinematics, 1988, Ljubliana.

[12] Parenti Castelli, V., Innocenti, C.; 1992, "Singularity-Free Evolution from One Configuration to Another in Serial and Fully-Parallel Manipulators,"
Robotics, Spatial Mechanisms, and Mechanical Systems, Vol. 45, pp.553-560.

[13] Chablat, D. and Wenger, P., 1998, "Working modes and aspects in fully parallel manipulators,"
proceedings of IEEE int. conf. on Robotics and Automation, Leuven, Belgium, pp. 1964-1969.

[14] Merlet, J.P.; 1997, "Les robots parallèles,"
seconde édition, Paris, 1997

[15] Collins, C.L. and Long, G.L., 1995, "The Singularity Analysis of an In-Parallel Hand Controller for Force-Relected Teleoperation,"
IEEE Transactions On Robotics and Automation, vol. 11, No.5, pp. 661-669.

[16] Dasgupta, B. and Mruthyunjaya, T., 1998, "Force redundancy in parallel manipulators: theoretical and practical issues,"
Mechanism and Machines Theory, vol. 33, No.6, pp. 727-742.

[17] Nenchev, D.N. and Uchiyama, M., 1996, "Dynamic analysis of parallel-link manipulators under the singularity-consistent formulation,"
Proceedings of IROS 96, vol. 33, No.6, pp. 1227-1233.

[18] Kevin Jui, C.K. and Sun, Q., 2003, "Path trackability and verification for parallel manipulators,"
Proceedings of IEEE international conference on Robotics and Automation, pp. 4336-4341.

[19] Kevin Jui, C.K. and Sun, Q., 2005, "Path tracking of parallel manipulators in the presence of force singularity,"
Journal of Dynamic Systems, Measurement and Control, Vol. 127, pp. 550-563.

[20] Dasgupta, B. and Mruthyunjaya, T., 1998, "Singularity-free path planning for the Stewart platform manipulator,"
Mechanism and Machines Theory, vol. 33, No.6, pp. 711-725.

[21] Bhattacharya, S., Hatwal, H., and Ghosh, A., 1998, "Comparison of an exact and an approximate method of singularity avoidance in platform type parallel manipulators,"
Mechanism and Machines Theory, vol. 33, No.7, pp. 965-974.

[22] Gosselin, C., and Angeles, J., 1990, "Singularity analysis of closed-loop kinematic chains,"
IEEE Transactions On Robotics and Automation, Vol. 6, No.3, pp. 281-290.

[23] Denavit, J. and Hartenberg, R.S., 1955, "A kinematic Notation for Lower Pair Mechanisms Based on Matrices,"
ASME Journal of Applied Mechanics, vol. 77, pp. 215-221.

[24] Tsai, L.W.; 1999, "Robot Analysis,"
John Wiley and sons, inc.

[25] El Omri, J.; 1996, "Analyse Géométrique et Cinématique des Mécanismes de Type Manipulateur,"
Thése de Doctorat, Ecole Centrale de Nantes.

[26] Burdick, J.,W., 1995, "A Classification of 3R Regional Manipulator Singularities and Geometries,"
Mechanisms and Machines Theory, vol. 30, No.1, pp. 71-89.

[27] Wenger,P., El Omri, J., 1994, "On the kinematics of Nonsingular and Singular Posture Changing manipulators,"
Advances in Robot Kinematics and Computational Geometry, J.-P. A.J. Lenarcic and B.B. Ravani (eds.), Kluwer Academic Publisher, pp. 29-38.

[28] El Omri, J., Wenger,P., 1995, "A General criterion fro the Identification of Nonsingular Posture Changing 3-DOF Manipulators,"
Computational Kinematics, J.-P. Merlet and B.Ravani (eds.), Kluwer Academic Publisher, pp. 153-162.

[29] El Omri, J., Wenger,P., 1995, "How to Recognize Simply a Nonsingular Posture Changing 3-DOF Manipulator?,"
Proceedings of IEEE International Conference on Advanced Robotics, pp. 215-222.

[30] Wenger,P., 1992, "A New Formalism for the kinematics Analysis of All Nonredundant Manipulators,"
Proceedings of IEEE International Conference on Robotics and Automation, pp. 442-447.

[31] Wenger,P., El Omri, J., 1996, "Changing Posture for Cuspidal Robot Manipulators,"
Proceedings of IEEE International Conference on Robotics and Automation, pp. 3173-3178.

[32] El Omri, J., Wenger,P.,1996, "Kinematic Analysis of Cuspidal Positioning Manipulators in the Presence of Obstacles,"
Proceedings of ISRAM'96.

[33] Paganelli, D., 2005, "Regional 3R Spatial Manipulators: a New Method to Characterise the Jointspace Partition Induced by Singularities,"
proceedings of the XVII AIMeTA national congress.

[34] Paï, D.K., and Leu, M.C., 1992, "Genericity and Singularities of Robot Manipulators,"
IEEE Transactions on Robotics and Automation, Vol.8, pp.545-559.

[35] Wenger, P., 1998, "Classification of 3R Positioning Manipulators,"
Journal of Mechanical Design, Vol. 120, pp. 327-332.

[36] Golubitsky, M., and Guillemin, V. 1973, "Stable mappings and their singularities,"
Springer.

[37] Paganelli, D., 2007, "The Ninth Homotopy Class of Spatial 3R Serial Regional Manipulators,"
Journal of Mechanical Design, Vol. 129, pp.445-448.

[38] Waldorn, K.,J., 1982, "Geometrically Based Manipulator Rate Control Algorithms,"
Mechanisms and Machines Theory, Vol. 17, No.6, pp.379-385.

[39] Waldorn, K.J., Wang, S.L., and Bolin, S.J. 1985, "A Study of the Jacobian Matrix of Serial Manipulators,"
Journal of Mechanisms, Transmission, and Automation in Design, Vol. 107, No.2, pp.230-238.

[40] Wang, S.L., and Waldorn, K.J., 1987, "A Study of Singular Configurations of Serial Manipulators,"

Journal of Mechanisms, Transmission, and Automation in Design, Vol. 109, No.1, pp.14-20.

[41] Hunt, K.H., 1978, "Kinematic Geometry of Mechanisms," *Claredon Press, Oxford.*

[42] Paganelli, D., 2007, "Avoiding Parallel Singularities of 3UPU and 3UPS Spherical Wrists," *Proceedings of ICRA 2007 IEEE International Congress of Robotics and Automation.*

[43] Innocenti, C., Parenti Castelli, V., 1993, "Echelon Form Solution of Direct Kinematics for the General Fully-Parallel Spherical Wrist," *Mechanism and Machines Theory*, Vol. 28, No.4, pp. 553-561.

[44] Paganelli, D., Innocenti, C., 2006, "Determining the 3x3 Rotation Matrices That Satisfy Three Linear Equations in the Direction Cosines," *Advances in Robot Kinematics, Springer.*

[45] Sefrioui, J., Gosselin, C., 1994, "Etude et Representation des Lieux de Singularite des Manipulateurs Paralleles Spheriques a Trois Degres de Liberte avec Actionneurs Prismatiques," *Mechanisms and Machines Theory*, Vol.29, No.4, pp.559-579.

[46] Fletcher, R., 1987, "Practical Methods of Optimization," *John Wiley and Sons*

[47] Karouia, K., and Hervé, J.M., 2000, "A three-dof tripod for generating spherical rotation," *Advances in robot kinematics, Kluver Academic Publishers*, pp. 395-402.

[48] Di Gregorio, R., 2003, "Kinematics of the 3-UPU Wrist," *Mechanism and Machine Theory*, Vol. 38, pp. 253-263.

[49] Di Gregorio, R., 2004, "Statics and Singularity Loci of the 3-UPU Wrist," *IEEE Transactions On Robotics*, Vol. 20, No.4, pp. 630-635.

[50] Zlatanov, D., Bonev I.A., and Gosselin, C.N., 2002, "Constraint Singularities of Parallel Mechanisms," *Proceedings of the 2002 IEEE International Conference on Robotics 8 Automation*, pp. 496-502.

[51] Di Gregorio, R., and Parenti Castelli V., 2002, "Mobility Analysis of the 3-UPU Parallel Mechanism Assembledfor a Pure Translational Motion," *Journal of mechanical Design*,Vol. 124 pp. 259-264.

[52] Tsai, L. W., 1996, "Kinematics of a Three-dof Platform With Three Extensible Limbs,"
Recent Advances in Robot Kinematics, Lenarcic, J., and Parenti-Castelli, V., Eds., Kluwer Academic Publishers,pp. 401410.

[53] Di Gregorio, R., and Parenti-Castelli, V., 1998, "A Translational 3-dof Parallel Manipulator,"
Advances in Robot Kinematics: Analysis and Control Lenarcic, J., and Husty, M. L., Eds., Kluwer Academic Publishers, pp. 4958.

[54] Parenti-Castelli, V., Di Gregorio, R., and Lenarcic, J., 1998, "Sensitivity to Geometric Parameter Variation of a 3-dof Fully-Parallel Manipulator,"
Proceedings of the 3rd International Conference on Advanced Mechatronics, pp. 364369.

[55] Pennock, G. R., and Kassner, D.J., 1993, "The Workspace of a General Geometry Planar Three-Degree-Of-Freedom Platform-Type Manipulator,"
Journal of Mechanical Design, Vol. 115, No. 1, pp. 269-276.

[56] Merlet, J.P., Gosselin, C.M., and Mouly, N., 1998, "Workspaces of Planar Parallel Manipulators,"
Mechanism and Machine Theory, Vol. 33, No. 1-2 pp.7-20.

[57] Sefrioui, J. and Gosselin, C.M., 1995, "On the Quadratic Nature of the Singularity Curves of Planar Three-Degree-of Freedom Parallel Manipulators,"
Mechanism and Machine Theory, Vol. 30, No. 4, pp.533-551.

[58] Wang, J. and Gosselin, C.M., 1997, "Singularity Loci of Planar Parallel Manipulators with Revolute Actuators,"
Robotics and Autonomous Systems, Vol. 21, pp.377-398.

[59] Paul, B., 1979, "A Reassessment of Grashof's Criterion,"
Journal of Mechanical Design, Vol. 101, pp. 515-518.

[60] Chase, T.R. and Mirth, J.A., 1993, "Circuits and Branches of Single-Degree-Of-Freedom Planar Linkages,"
Journal of Mechanical Design, Vol. 115, No.2 pp. 223-230.

[61] Mirth, J.A. and Chase, T.R., 1993, "Circuit Analysis of Watt Chain Six-Bar Mechanisms,"
Journal of Mechanical Design,, Vol. 115, No.2 pp. pp.214-222.

[62] Midha, A., Zhao, Z.L., and Her, I., 1985, "Mobility Conditions for Planar Linkages Using Triangle Inequality and Graphical Interpretation,"
Journal of Mechanical Design,, Vol. 107, No.3, pp. 394-400.

[63] Foster, D.E. and Cipra, R.J., 1998, "Assembly Configurations and Branches of Single-Loop Mechanism with Pin Joints and Sliding Joints,"
Journal of Mechanical Design,, Vol. 120, No.3, pp. 387-391.

[64] Foster, D.E. and Cipra, R.J., 2002, "An Automatic Method for Finding the Assembly Configurations of Planar Non-Single-Input-Dyadic Mechanisms,"
Journal of Mechanical Design,, Vol. 124, No.1, pp. 58-67.

[65] Dou, X. and Ting, K.L., 1998, "Identification of Singularity Free Joint Rotation Space of Two-DOF Parallel Manipulators,"
proceedings of 1998 ASME DETC.

[66] Paganelli, D., 2007, "A New Singularity-Free Path-Planning Method for a Class of Planar Parallel Manipulators,"
Proceedings of 2007 ASME IDETC/CIE.

[67] Salmon, G. 1885, "Modern Higher Algebra,"
Hodges, Figgis, and Co..

[68] Allgower, E.L., and Georg, K. 1990, "Numerical Continuation Methods: an Introduction,"
Series in Computational Kinematics, Springer, Vol 13.

[69] Allgower, E.L., and Georg, K. 1992, "Continuation and Path Following,"
Acta Numerica, pp.1-64.

[70] Wise, S.M., 1998, "POLSYS PLP: A Partitioned Linear Product Homotopy Code for Solving Polynomial Systems of Equations,"
Virginia Polytechnic Institute, Master Thesis.

[71] Morgan, A.P., Sommese, A.J., and Wampler, C.W., 1995., "A Product Decomposition bound for Bezout Numbers,"
SIAM Journal on Numerical Analysis, Vol. 32,pp. 1308-1325.

[72] Wampler, C. W., 1994, "An Efficient Start System for Multi-Homogeneous Polynomial Continuation,"
Journal of Numerical Mathematics, Vol. 66, pp.517-523.

[73] Ting, L., Zhenjiang, L., and Fengshan B., 2003, "An Efficient Start System for Multi-Homogeneous Polynomial Continuation,"
Journal of Applied Mathematics and Computation, Vol. 146, pp.237256.