EMERGENCE OF IMPERFECT SYMMETRY IN ENGINEERING OPTIMISATION AND EVOLUTION

DOCTORAL (PH. D.) DISSERTATION

PÉTER VÁRKONYI

2006

EMERGENCE OF IMPERFECT SYMMETRY IN ENGINEERING OPTIMISATION AND EVOLUTION

A Dissertation
Submitted to the Budapest University of Technology and Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Péter Várkonyi

Supervisor:
Gábor Domokos

2006

TABLE OF CONTENTS

TABLE OF CONTENTS **1**

ABSTRACT **3**

ABSTRACT IN HUNGARIAN **4**

1.1 A DISSZERTÁCIÓ TÉMÁJA, ALAPKÉRDÉSEK 4
1.2 SZERKEZETOPTIMALIZÁLÁS 5
1.3 ASZIMMETRIA AZ EVOLÚCIÓBAN 6

CHAPTER 1 INTRODUCTION **8**

CHAPTER 2 STRUCTURES **13**

2.1 INTRODUCTION TO STRUCTURAL OPTIMISATION 13
2.1.1 Problem statement *13*
2.1.2 An illustrative example: bifurcation of the symmetric optimum *14*
2.1.3 Principal results and structure of Chapter2 ..... *16*
2.2 GENERAL DEFINITION OF THE PROBLEM 18
2.3 REFLECTION SYMMETRY 20
2.3.1 Optimisation with one variable *20*
2.3.2 Exceptional cases *21*
2.3.3 Several variables *22*
2.4 GENERAL SYMMETRY 23
2.4.1 Introduction *23*
2.4.2 Exact condition of robust optimum *23*
2.4.3 Application of representation theory to optimisation problems *25*
2.4.3.1 Orbits in the optimisation problem 25
2.4.3.2 Analysis of the induced representation 26
2.5 NUMERICAL OPTIMISATION EXAMPLES 28
2.5.1 Main steps of the analysis *29*
2.5.2 Example 1 ..... *31*
2.5.3 Example 2 ..... *32*
2.5.4 Example 3 ..... *34*
2.5.5 Example 4 ..... *37*
2.6 BIFURCATION ANALYSIS OF OPTIMUM DIAGRAMS 38
2.6.1 Typical bifurcations in case of D_1 symmetry *39*
2.6.1.1 Reflected potentials 39
2.6.1.2 Multiple reflected potentials 41
2.6.2 Examples in engineering: optimisation of structures *42*
2.6.2.1 Unstable X bifurcation. 43
2.6.2.2 Point-like bifurcation 44
2.6.2.3 Five-branch pitchfork bifurcation 45
2.6.2.4 Three-branch pitchfork bifurcation 45
2.6.2.5 Wedge bifurcation 46
2.6.3 Bifurcation analysis at different symmetries *46*

2.7 EXCEPTIONAL STRUCTURES 48
2.8 SUMMARY 49

CHAPTER 3 EVOLUTION 51

3.1. INTRODUCTION TO EVOLUTION 51
3.1.1 Problem statement 51
3.1.1. Principal results and the structure of Chapter 3 53
3.2. ADAPTIVE DYNAMICS IN CONSTANT ENVIRONMENT 54
3.2.1. Fitness concept 55
3.2.2. Directional evolution 56
3.2.3. Properties of singular strategies 57
3.3. BASIC ASSUMPTIONS 59
3.4. TWO TYPES OF SYMMETRY-BREAKING 60
3.5. EMERGENCE OF ASYMMETRY 62
3.5.1. Frequency independent models 62
3.5.2. Weak symmetry 62
3.5.3. Strong symmetry in frequency dependent models 63
3.6. A NOVEL WAY OF EVOLUTIONARY BRANCHING 64
3.7. A MODEL EXAMPLE 66
3.7.1. Description of the model 66
3.7.2. Singular strategies and coalitions in the model 68
3.7.3. Branching in the model 69
3.8. BIOLOGICAL EXAMPLES OF SYMMETRICAL STRATEGIES 70
3.9. DISCUSSION 72

CHAPTER 4 SUMMARY AND PRINCIPAL RESULTS 73

ACKNOWLEDGEMENTS 77

REFERENCES 78

APPENDIX I REPRESENTATION THEORY 81

I.1 GROUPS 81
I.2 SYMMETRIES OF REAL ENGINEERING STRUCTURES 83
I.3 REPRESENTATIONS 84
I.3.1 Definition, Basic properties 84
I.3.2 Decomposition of representations 86
I.3.3 The regular representation 87
I.4 ORBITS 87
I.4.1 Orbits and invariant points 88
I.4.2 Dimensionality of orbits 89
I.5 REAL-VALUED REPRESENTATIONS 91

ABSTRACT

This thesis is devoted to the relation between imperfect symmetry (i.e. slight asymmetry) and optimisation in two distinct fields of science. The first part is motivated by the observation that engineering structures with imperfect symmetries are extremely rare, however both perfectly symmetrical and strongly asymmetrical structural solutions are frequent. It is demonstrated that perfectly symmetrical configurations are typically local optima in some class of structural optimisation problems, supporting the above observation. However, improving such configurations is often possible by introducing an adequate set of small perturbations. It is shown how to choose such a set, without performing detailed structural analysis of the structures. This result helps to improve symmetrical structures by minor perturbation of their symmetries. The emergence of local optima with imperfect symmetry is also investigated in structural optimisation. The second part of the thesis is based on the fact that imperfectly symmetrical body plans are common in the flora and fauna, in contrast to the world of structures. Such biological organisms are results of evolutionary development, which suggests their optimality. The aim of this part is the theoretical modelling of the loss of perfect symmetry in evolutionary development within the framework of Adaptive Dynamics: the ecological types of symmetry-breaking as well as the generic evolutionary patterns of the emergence of asymmetry are presented.

ABSTRACT IN HUNGARIAN
MAGYAR NYELVŰ ÖSSZEFOGLALÓ

TÖKÉLETLEN SZIMMETRIA MEGJELENÉSE A SZERKEZETOPTIMALIZÁLÁSBAN ÉS AZ EVOLÚCIÓBAN

1.1 A DISSZERTÁCIÓ TÉMÁJA, ALAPKÉRDÉSEK

A szimmetria kitüntetett szerepet játszik az emberi kultúra és megismerés minden területén. Az ember ősidők óta a szimmetria számos megjelenési formáját figyelte meg az őt körülvevő természetben a hópihék formájától kezdve egy virág felépítéséig, és a szimmetriára régtől fogva, mint a harmónia, tökéletesség szimbólumára tekintenek. Az ókori görög művészetben csakúgy, mint a tudományos világképükben, a szimmetria fő szervezőelemként jelentkezik. Számos példa mutatja ugyanakkor, hogy *tökéletlen szimmetriájú* (vagyis a szimmetrikustól csak kismértékben különböző) kompozíciókkal is kísérleteztek, például a szobrászatban vagy a templomépítészetben. A későbbi művészeti stílusok, azon belül is talán legjobban az építészeti stílusok, előnyben részesítették a tökéletes, illetve tökéletlen szimmetriájú kompozíciókat. A modern stílus volt az első, amely elvetette a szimmetria kitüntetett voltát, sőt tudatosan kerülte azt. A művészetekhez hasonlóan központi szerepet kapott a szimmetria a tudományok fejlődésében, elég, ha a kvantumfizika csoportreprezentációkon alapuló elméletére gondolunk.

A mérnöki szerkezetek között is gyakoriak a szimmetrikus formák, de az építészettel ellentétben *tökéletlen szimmetriát* csak elvétve látunk ezek között. A mérnöki alkotások tervezésekor általában célszerűségi szempontok az elsődlegesek, ezért a fenti megfigyelés azt az intuíciót sugallja, hogy egy tökéletlen szimmetriájú szerkezet nem lehet optimális, sőt, rosszabb, mint a tökéletesen szimmetrikus forma. A disszertáció egyik fele ennek az okát kutatja. Bemutat egy egyszerű optimalizálási feladat-típust ahol ez az intuíció helyesnek bizonyul. Ezen túlmenően két kérdést vizsgál: meg lehet-e a szimmetriát mégis zavarni úgy, hogy az a szerkezeten *javítson*, illetve egy tökéletlen szimmetriájú szerkezetalak *lehet-e optimális*.

Az élő természetben az evolúcióra gyakran úgy tekintenek, mint egy önszabályozó optimalizálási folyamatra, amely során a legéletképesebb életformák kifejlődnek és kiszorítják a kevésbé tökéleteseket. Ilyen értelemben az evolúció analógnak tűnik a mérnöki optimalizálással. Az evolúció során az élőlények testfelépítésének szimmetriája is változik (pl. a gerincesek általában kétoldali, míg a csalánzók sugaras szimmetriával rendelkeznek). A szimmetria általános az állatvilágban, de gyakran *tökéletlen*. A véletlenszerűen kialakuló hibákon kívül számos élőlény szimmetriájában genetikusan öröklött tökéletlenségek találhatók, ilyen például az emberi jobb/balkezesség és a hozzá kapcsolódó agyi aszimmetria, amely egyértelműen előnyösebb, mint a tökéletesen szimmetrikus felépítés. A dolgozat

második része az aszimmetria kialakulásának lehetőségeit vizsgálja az Adaptív Dinamika eszköztárának felhasználásával: fő célja annak felderítése, milyen típusai vannak az aszimmetria megjelenésének és ezekhez milyen evolúciós mintázatok kapcsolódnak.

A dolgozat tehát két jelentősen eltérő tudományterületen vizsgálja a szimmetria és optimum viszonyát. A mérnöki részben a tökéletlen szimmetriájú szerkezetek hiánya motiválta a kutatást, míg az evolúciós fejezet alapfelismerése a tökéletlen szimmetria gyakori volta az élővilágban. Mindkét terület szorosan kapcsolódik az optimalizálás témaköréhez, bár az evolúció számos aspektusa nem érthető meg, ha pusztán optimalizálási folyamatnak tekintjük. Valójában a két téma matematikai szempontból a sima függvénycsaládok (potenciálok) szinguláris pontjaival foglalkozó elemi katasztrófaelmélet közismert eredményeinek két eltérő jellegű általánosítása és alkalmazása. Míg a szerkezetoptimalizálási részben a potenciál nem sima volta jelenti a többletet, az evolúciós rész az „általánosított potenciálként" is felfogható fittnesz-függvények szinguláris pontjait vizsgálja. Mindkét esetben olyan eredményeket láthatunk, amelyek az elemi katasztrófaelméleti ismeretek alapján szokatlanok: a tartószerkezeti részben olyan bifurkációk jelentkeznek, ahol a szimmetrikus megoldás optimalitása a bifurkációs pontban nem változik, míg az evolúciós kutatás eredményei azt mutatják, hogy szimmetrikus élőlények között egy új, aszimmetrikus forma el tud terjedni anélkül, hogy kipusztítaná a szimmetrikus formát.

A két témakörhöz kapcsolódó vizsgálatokat részletesebben a következő két pontban foglalom össze.

1.2 SZERKEZETOPTIMALIZÁLÁS

Szerkezetoptimalizálási feladatokban gyakori, hogy egyes szerkezeti elemek jósága külön-külön van definiálva, és a legkedvezőtlenebb elem határozza meg a teljes szerkezet jóságát, azaz potenciálját. A fejezet gondolatmenete abból a felismerésből indul ki, hogy a fenti feladattípusnál egy tengelyes vagy egyéb szimmetriájú szerkezet szimmetriáját megzavarva a tökéletes szerkezet általában lokális, "robusztus" optimum, azaz ilyen feladatokban a szimmetria kis mértékű megzavarása ront a szerkezeten, mégpedig a romlás mértéke a zavarással lineárisan nő, ellentétben egy sima optimummal, ahol a minőségcsökkenés csak a zavarás négyzetével lenne arányos. A lokális optimum természetesen nem zárja ki más, erősen aszimmetrikus lokális optimumok létét. Ez a tény egyrészt hozzájárulhat annak az alapvető felismerésnek a magyarázatához, hogy miért olyan ritka a tökéletlen szimmetria mérnöki szerkezetek geometriájában, másrészt egyenesen következik belőle a kérdés, hogyan lehetne mégis javítani egy szimmetrikus szerkezeten, a szimmetria kismértékű zavarásával.

Ha egy helyett több szimmetria-sértő változót vezetünk be, akkor a szimmetrikus konfiguráció javíthatósága szempontjából az alábbi esetek fordulhatnak elő:

- A: a szimmetrikus konfiguráció robusztus optimum
- B1: a szimmetrikus konfiguráció optimum, de nem robusztus, azaz van a változóknak olyan kombinációja, amellyel a szerkezet minősége nem lineárisan romlik.
- B2: a szimmetrikus konfiguráció nem optimum, azaz a változóknak van olyan kombinációja, ami javít a szerkezeten.

Javíthatóság szempontjából a B2 típus van kitüntetett helyzetben. Ugyanakkor az, hogy egy példa B1 vagy B2 típusba tartozik-e, csak részletes erőtani számítások alapján dönthető el, míg az A és a B típusok között pusztán a szerkezet szimmetriaviszonyai és a változók

ismeretében is különbséget tehetünk. Definiáltam ezért a *potenciális javíthatóság* fogalmát, ami annyit jelent, hogy a tökéletes szerkezet B típusú, és a klasszikus reprezentációelmélet eredményeinek felhasználásával **adott változóhalmazra és szimmetriatípusra a potenciális javíthatóság egyszerűen ellenőrizhető feltételét határoztam meg** (I. tézis). Ennek segítségével könnyen kiválaszthatók olyan változók, amelyekkel egy szerkezetet kis zavarással potenciálisan javítható, ezen belül a tényleges javíthatóság kérdése részletes erőtani számításokkal dönthető el. **A potenciális javíthatóságra olyan szükséges, illetve elégséges feltételeket is meghatároztam, amelyek kizárólag a változók számára vonatkoznak** (II. tézis).

A javíthatóság mellett további kérdésként merült fel, lehetséges-e, hogy a szimmetriát kis mértékben megzavarva (lokális) optimumot kapunk, azaz a szimmetrikus szerkezet *optimálisan javítható*. Majdnem szimmetrikus optimumokat egy p paraméter bevezetésével előállított *feladatcsaládban* találhatunk a paraméter azon p_0 értéke környezetében, ahol a szimmetriatörő változók $\mathbf{x}$ vektorának optimális értékeit p függvényében ábrázolva az $(\mathbf{x},p)=(0,p_0)$ pontból aszimmetrikus ($\mathbf{x}\neq 0$) optimumok ágaznak el. Ezért az optimális javíthatóság kérdése a potenciálfüggvények bifurkációanalízisére vezet. Nem sima függvényekről lévén szó, az elemi katasztrófaelmélet eredményei nem alkalmazhatóak, a sima, szimmetrikus potenciálok szokásos villa-elágazása helyett más bifurkációs mintázatok jelentkeznek tipikusan. Tengelyes szimmetriájú szerkezetekre a vizsgálatot elvégezve megállapítottam, hogy **egy szimmetriatörő változó esetén speciális kivételektől eltekintve a szerkezetcsaládban nincsenek optimum-elágazások** (III.1 tézis), ez az eredmény azt mutatja, hogy a majdnem szimmetrikus szerkezeti optimumok rendkívül ritkák. **Ugyanakkor példát mutattam olyan, más szimmetriával rendelkező szerkezetek családjára, ahol a szimmetrikus szerkezetek potenciálisan nem javíthatók, a szerkezetcsaládban mégis vannak optimálisan javítható elemek** (III.2 tézis). Ezekben az esetekben, meglepő módon, az optimális javíthatósághoz kevesebb szimmetria-sértő változó bevezetése szükséges, mint a potenciális javíthatósághoz, annak ellenére, hogy a természetes intuíció szerint az előbbi tulajdonság tűnik speciálisabbnak.

1.3 ASZIMMETRIA AZ EVOLÚCIÓBAN

A dolgozat második része azzal az evolúciós jelenséggel foglalkozik, amikor kétoldali szimmetriával rendelkező élőlények testfelépítésében valamilyen öröklött aszimmetria jelenik meg, azaz szimmetriájuk tökéletlenné válik az evolúció során. Az evolúció sok szempontból tekinthető optimalizálódásnak, de valójában több annál, hiszen nem egy „optimális” faj egyeduralkodóvá válásához vezet, hanem a természetben tapasztalt sokféleséghez. Ennek megfelelően az evolúciós folyamatok modellezése mutat ugyan matematikai hasonlóságokat a szerkezetoptimalizálással, de az analógia csak részleges: az optimalizálási feladatok egy adott megoldás jóságát kifejező potenciálon alapulnak, ezzel szemben a biológiai modellekben ennek megfelelő fittnesz-függvény egy élőlény életképességét *adott környezetben* adja meg. A környezeti viszonyokra az adott életközösségben együtt élő összes élőlény hatással van, tehát, egy élőlény „jóságát” önmaga mellett a vele együtt élő versenytársak gyakorisága és típusa is befolyásolja. Ezt a tulajdonságot *gyakoriságfüggésnek* nevezik.

Az aszimmetria kialakulásának vizsgálatára az adaptív dinamika eszköztárát használtam, amely az evolúció fenti aspektusát figyelembe veszi, de számos, matematikailag nehezebben kezelhető tényezőt (pl. az evolúció részletes genetikai hátterét, összetett populációdinamikai

jelenségeket, nagy változással járó evolúciós lépések lehetőségét) elhanyagol, így aránylag könnyen alkalmazható számos evolúciós jelenség modellezésére.

Az aszimmetria megjelenésének típusait két szempont szerint osztályoztam. A megjelenést kiváltó ok szempontjából:

- Egyszerűbb élőlények genetikai rendszerei gyakran nem teszik lehetővé az aszimmetria kódolását. Az evolúciós fejlődés során a genetikai rendszer komplexebbé válik, és így lehetséges lesz az aszimmetria kialakulása. Ha az adott környezeti feltételek mellett előnyös, ki is alakulnak aszimmetrikus élőlények. Ezzel leegyszerűsített módon egy olyan tényezőt veszünk figyelembe (a genetikai korlátokat), amelyekkel az adaptív dinamika általában nem foglalkozik.
- Az aszimmetria kialakulása genetikailag lehetséges, de ökológiailag nem előnyös, majd a külső környezet megváltozása miatt előnyössé válik. Ennek modellezéséhez időfüggő adaptív dinamikai modellt kell vizsgálni.

A szimmetriatörés ökológiai jellegét tekintve pedig az alábbi típusokat különítettem el:

(A) A modell gyakoriságfüggő. **Ezen belül két, az irodalomban nem tárgyalt altípust vezettem be** (IV. tézis):
  (A1) ha két aszimmetrikus élőlény egymás tükörképe, akkor egymással felcserélhetőek, a modell viselkedésének megváltozása nélkül. Ezt *erős szimmetriának* neveztem el.
  (A2) ha két aszimmetrikus élőlény egymás tükörképe, akkor sem azonos a szerepük. Ezt *gyenge szimmetriának* neveztem el.

(B) A modell nem gyakoriságfüggő (tehát a vizsgált evolúciós folyamat optimalizálási feladatra vezethető vissza). Ekkor a tükörkép élőlények szükségképpen felcserélhetőek.

A dolgozatban mind a hat esetre **felsoroltam az aszimmetria kialakulásának tipikus mintázatait, összesen három különbözőt** (V.1 tézis). Időfüggő modellben, az (A1) esetben **kimutattam egy szokatlan elágazás-típus lehetőségét is: egy szimmetrikus populációban megjelenhet és elterjedhet egy új, aszimmetrikus típus, amely együtt él a szimmetrikus ősökkel** (V.2 tézis). Ez jellegében különbözik az adaptív dinamikában szokványos evolúciós elágazásoktól, amelyek során az elágazás előtti őstípus mindig eltűnik. Az új mintázat annak köszönhetően alakul ki, hogy erős szimmetria esetén az időfüggő modell fittnesz-függvénye diszkrét időpillanatokban degenerálttá válhat, ugyanakkor a mutációs lépések véges (nem infinitezimálisan kicsiny) mérete miatt a degenerált modellre jellemző viselkedés hosszabb ideig is fennáll, és azalatt néha számottevő evolúciós fejlődés is lezajlik. Ez az eset azt is mutatja, hogy **az aszimmetria megjelenésében a külső környezeti változások fontos szerepet játszanak** (V.3. tézis).

Az erős és a gyenge szimmetria közti különbséget valódi példákon, míg az új típusú evolúciós mintázatot egy klasszikus modellen szemléltettem.

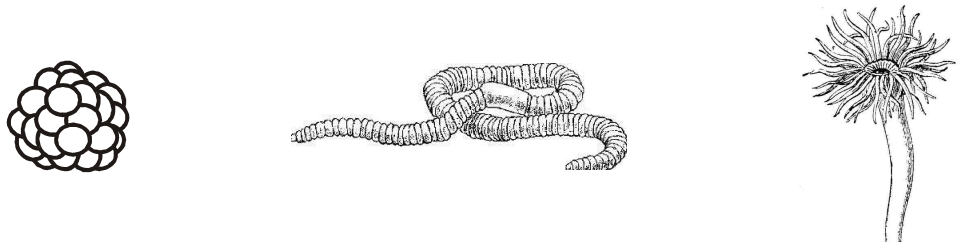
CHAPTER 1 INTRODUCTION

"Yet each in itself- this was the uncanny, the anti-organic, the life-denying character of them all-each of them was absolutely symmetrical, icily regular in form. They were too regular, as substance adapted to life never was to this degree - the living principle shuddered at this perfect precision, found it deathly, the very marrow of death - Hans Castorp felt he understood now the reason why the builders of antiquity purposely and secretly introduced minute variation from absolute symmetry in their columnar structures."

(Thomas Mann: The magic mountain, 1928)

The identification of 'symmetry' and 'perfectness' is probably as old as aesthetics itself and it is fundamental part of human art, science, and philosophy. This idea probably originates from multiple empirical observations of symmetry in the physical world.

The symmetries emerging in natural patterns, such as crystal structures, snowflakes, or water waves (Weyl, 1989) are widely considered as representative examples of beauty and harmony of Nature. This idea was already present in the ancient Greek culture where the word "symmetry" originates from. In fact, symmetry was considered as a main organising principle of the world. One of the first structured cosmic models of Anaximander describes the world as a system with spherical layers (Couprie et al, 2003), while the widely known theory of Platon identified four of the platonic solids with the four basic elements of the world (Cooper et al, 1997). Though models of the material and the universe have changed radically during the past millenia, symmetry still seems to play central role in understanding the physical world. In the 20th century many fields of science were built on group theory. The corresponding literature is huge, we only mention some general books on symmetry, Rosen, 1995, Hargittai et al., 1994) and a book on its application in quantum physics (Jones, 1998).



A **B** **C**

Figure 1.1 Examples of body plans with different symmetries: quasi-spherical body of mezosoans (A), a cylindrical earthworm (B), radial symmetry of anemones (C)

Symmetry seems to be a strong feature of the living world, as well (Purves et al, 2003). The majority of animals and also many plants have some kind of symmetry (Figure 1.1). The quasi-spherical symmetry of the most ancient multi-cellular creatures (mesozoans), the cylindrical symmetry of filamentous algae or various worms (e.g. earthworm), the radial

symmetry of anemones or the bilateral symmetry of the most vertebrates all show the presence of symmetry in common body plans. Evolution seems to reduce symmetry in most cases, i.e. more complex body plans usually have lower-order symmetries, even though there are examples of secondary symmetries as well (e.g. the secondary radial symmetry of sea urchins and starfish, which evolved from a bilateral body structure). At the same time, we can find a few completely asymmetrical animals (e.g sponges) as well as many creatures with genetically inherited *imperfectness of symmetry*. The handedness or the position of the heart in the human body, are nice examples of the latter category, where the symmetrical basic body plan is preserved despite the imperfectness.

A B C D

Figure 1.2 Ancient Greek sculptures of different ages A: 'Cycladian goddess', Amorgos, 2000 BC; B: 'Kyros', Anavissos, 520 BC.; C: 'Youth of Anitikitera', Anitikitera, 340 BC.; D: 'Aphrodite', Myrina, 2nd century AD. Photos are taken from Petrakos, 1993 (A,B,C) and Kunze et al, 1992 (D).

Beyond Nature and Science, symmetry gains main role in all fields of human creativity. In the ancient Greek culture symmetry was an aesthetical category rather than a pure mathematical definition, which emerged as an unavoidable ingredient of art and philosophy, as well. As an example, ancient Greek sculpture shows perfect symmetry, while gradual emergence of asymmetry can be observed in later representations of the human body, cf. Figure 1.2 (Marótzy, 2005). In accordance with the hint of Thomas Mann, architecture also preferred symmetry in the ancient times and later. All historical styles of architecture considered symmetrical forms as perfect. Accordingly, early works of a style usually show rigorous symmetry, while *imperfectness* (i.e. slight violation) *of symmetry* is a frequent indicator of the claim for renewal during the disintegration of architectural styles (Figure 1.3). The modern movement of the early XXth century was the first one to reject the aesthetical superiority of symmetry (Preziosi, 1998).

While architectural forms are strongly determined by aesthetical considerations, engineering structures, such as bridges, towers, or shells of major size are developed primarily by virtue of *practical optimality criteria*. Still, engineering structures tend to be symmetric as well, see Figure 1.4 for examples of several types of dihedral symmetries, all of which are common among tower-like structures. At the same time, we can find many asymmetrical engineering structures as well: Figure 1.5 shows two bridges of Seville, one of which is a classical form with reflection symmetry, while the other one is a popular asymmetrical structural solution.

Figure 1.3 Example of the imperfect symmetry in architecture: the late-renaissance Castle of Chambord, France (Domenico da Cortona, 1537). Notice the difference in the number of windows of the left and the right wing and several other 'imperfect' details.

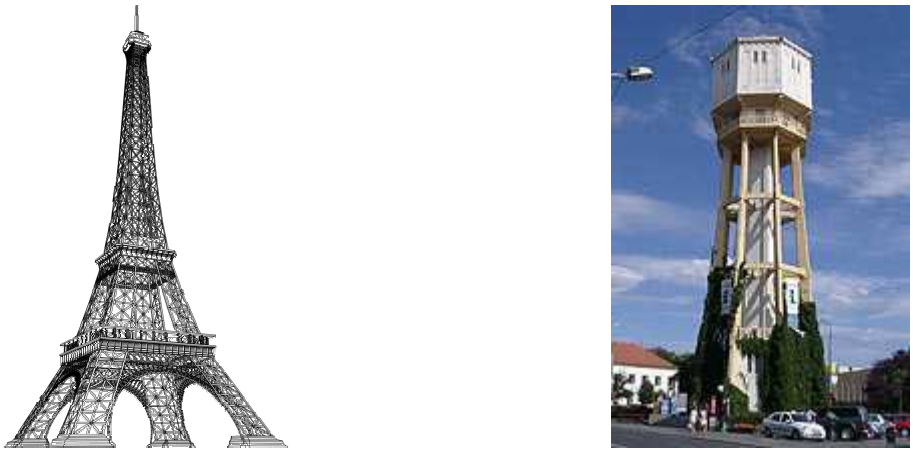


Figure 1.4 Left panel: Eiffel tower, Paris, France with D_4 symmetry. Right panel: Water tower in Siófok, Hungary with D_8 symmetry.

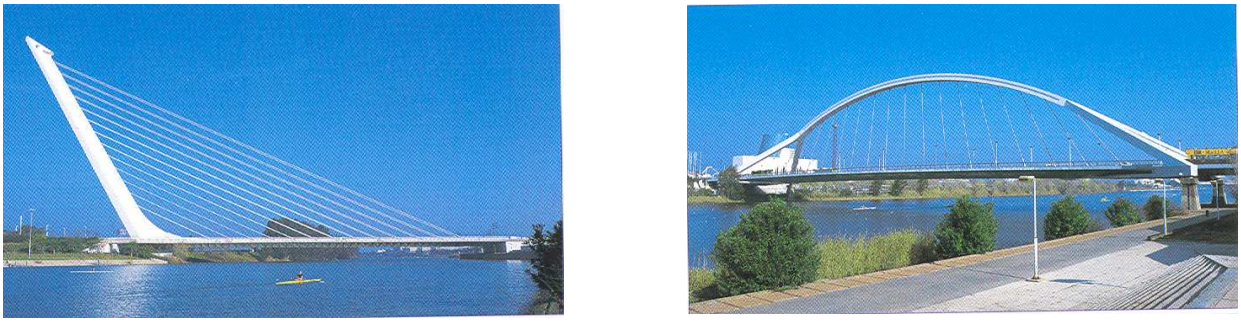


Figure 1.5 Left panel: El Alamillo bridge, Seville, Spain (designed by S. Calatrava, 1992). Right panel: La Barqueta bridge, Seville, Spain (designed by Juan J. Arenas and Marcos J. Pantaleón, 1992).

As already shown, *imperfect symmetry* seems to be a distinct category in evolution as well as in arts and aesthetics. It is plausible to pose the question, whether this category may emerge as an *optimal design* among engineering structures. Needless to say, among existing structures such constructions are extremely rare. This observation is in tune with the engineer's general intuition, which suggests that small perturbation of the symmetry yields an imperfect, i.e. worsened structural configuration.

One of our goals is to find arguments *supporting* the engineer's intuition, i.e. we would like to understand why slight perturbation of the symmetry of an engineering structure *often* weakens the quality of the structure. On the other hand, it is also our goal to pinpoint cases where the engineer's intuition *fails,* i.e. to find cases where *imperfect symmetry* proves to be optimal. This idea leads naturally to the construction of unusual, though *optimal structural shapes with imperfect symmetry*, which are, not in the geometric but in the structural sense, more perfect than their symmetric counterparts. Chapter 2 of my work is devoted to these questions.

The heart of structural optimisation is a 'goodness measure', which allows to distinguish between better/worse configurations. Similar potentials emerge in various fields of scientific research where some kind of optimisation is dealt with. According to the Darwinian theory, evolution is ruled by natural selection, which can be considered as a self-optimisation process of biological systems. Thus, optimal structural shapes carry a close analogy to the form of biological organisms created in evolutionary processes, such as the human body. In particular, evolutionary development in changing environment can be considered as downhill motion of an evolving variable x (scalar or vector) on a $U(x,t)$ time-dependent 'potential-landscape'. Motivated by the similarity between evolution and engineering optimisation, Chapter 3 of my work deals with the evolutionary modelling of *the emergence of bilateral asymmetry,* which is a common evolutionary phenomenon and usually results in a body plan with *imperfect symmetry*. The main goal of this part is to determine the generic temporal patterns of the emergence of new, asymmetric branches of the evolutionary tree within the framework of Adaptive Dynamics.

Despite the obvious analogy between engineering optimisation and evolution, two basic differences need to be outlined. First, simplifying evolution as optimisation is misleading: this point of view cannot explain the diversity in Nature. To solve this contradiction, evolutionary biology adopted the basic idea of game theory, namely that one's fitness (potential) depends on its own strategy *and also on the competitors' strategy* (Hofbauer et al, 1998). With other words, coexisting populations modify the environmental conditions (e.g. the abundance of food sources), and this way they influence each other's fitness, which may stabilise their coexistence. Thus, the 'fitness functions' of many biological models are generalised 'potentials' of the form $s([c],x,t)$, where the meaning of x and t are the same as in case of optimisation potentials, while the new argument $[c]$ symbolises data on the number and types of competitors.

Furthermore, the mathematical analogy of optimisation and evolution hides a major physical difference between the two tasks. Evolution is a dynamical process, which develops in time, while structural optimisation is a static procedure. In the latter case we can (and also we will) introduce a parameter (quasi-time), however this will result in families of separate optimisation examples, rather than one complex task of optimisation. By introducing a parameter, we can create optimum diagrams yielding bifurcation patterns of optima, which are mathematically somewhat analogous to evolutionary patterns (generated by the real time parameter). At the same time, the latter ones are spontaneous bifurcations, which can be

observed in Nature in temporal data-series, while the former ones do not emerge spontaneously, they just serve to give a broader view on structural optimisation.

The link between the two fields of my investigations is more than just the common basic concept of optimisation: the applied techniques and the aims of the two parts also show remarkable similarities. In the structural part, *generic results* on the optimality/improvability of a symmetric structural configuration against a given number or a given set of perturbing variables is investigated, based on the *truncated Taylor expansion* of its potential function. The aim of the investigations is to derive conditions of improvability, which do not call for detailed analysis of the specific potential functions. Analogously, the evolutionary part operates with the truncated Taylor-expansions of fitness functions to find generic evolutionary patterns without working out detailed ecological models and exact fitness functions.

CHAPTER 2 STRUCTURES

2.1 INTRODUCTION TO STRUCTURAL OPTIMISATION

2.1.1 Problem statement

This chapter of the thesis deals with the role of symmetry in structural optimisation. As discussed in Chapter 1, engineering structures are often symmetric, which may have objective cause, i.e. the symmetrical form may often be the optimal solution of a design problem. In the engineering praxis symmetrical forms are considered to be better than their perturbed, slightly asymmetrical variants. A small asymmetry is usually regarded as an imperfection of the structure; *almost* symmetrical design is extremely rare. In this section we will determine exact criteria for the local optimality of symmetric shapes. In particular, we will be interested in general conditions and an algorithm to determine *whether a symmetric structure may be improved via small geometric perturbations.*

Optimisation is often based on a scalar 'goodness measure', which is determined for all possible solutions of a problem and the biggest/smallest value corresponds to the best solution. We will follow this tradition and associate optima with minima, 'pessima' with maxima of a scalar potential. The literature for shape optimisation is extremely rich, for reference we mention Hemp (1973), Rozvány (1989), Banichuk (1990) and Sokolowski et al (1992). Optimising structural topology is also a popular field of research (see e.g. Bendsoe, 1995, Allaire, 2002). It is remarkable that almost all examples discussed in the literature exhibit some degree of symmetry. We do not intend to challenge the validity of these results; on the contrary, they illustrate one side of the landscape we are interested in. Our goal is to draw attention to the existence of *the other side, i.e. optimal structural solutions with slight asymmetry.*

In this work we will study one-parameter (p) families of structures, depending on a vector of "symmetry-breaking" scalar variables $\mathbf{x}=[x_1\ x_2...x_d]$. we will investigate the *optimal value* of the variables according to an arbitrarily chosen scalar measure of quality $U(p,\mathbf{x})$, associated with the *weakest point* of the structure. This implies that U is an *upper envelope* of the individual, smooth potentials, associated with the *weak points* of the structure. (Alternatively, one may look for an optimum based on a single, *global criterion*, e.g. minimisation of the total mass of a structure. Such problems typically lead to a *single, smooth* potential. Although, mathematically the latter one is undoubtedly a much simpler scenario, it does not fit to many real-life engineering problems: optimality of a structure consisting of several, identical elements, leads to the above-discussed concept of weak points. The next subsection will illuminate in detail the difference between the two approaches, based on a simple example.)

We will assume that for all values of the parameter p, the structure associated with $\mathbf{x}=0$ possesses a finite, non-trivial symmetry group Γ and furthermore we will assume that none of the other, individual ($\mathbf{x}\neq 0$) structures is Γ-invariant, however, the total, d-dimensional *set of structures* is Γ-invariant. (For more detail, see conditions (i) and (ii) in Section 2.2). Since U

is generated as an *upper envelope* of smooth functions, the symmetrical, $\mathbf{x}=0$ configuration falls typically into one of the following categories:

(A) Non-smooth (robust) optimum (minimum)
(B) Partially or completely smooth optimum, pessimum (maximum) or saddle.

An optimum is called robust if U is growing linearly in *every* direction. Partial smoothness means that a smooth submanifold is passing through $\mathbf{x}=0$. A rigorous definition of these categories will be given in Section 2.2, see *Definition 2.1*.

A symmetrical structure at $\mathbf{x}=0$ will be called *potentially locally improvabe* if it falls into category (B) (cf. *Definition 2.2*). In particular, pessimum or saddle, correspond to *actual* improvability. Whether a *potentially* locally improvable structure is *actually* locally improvable, one has to perform structural analysis of internal forces and stresses. We will show examples of such computations in subsections 2.5, 2.6.2, and 2.7.

A one-parameter (p) symmetrical structure family will be called *optimally improvable* at $p=p_0$ if the symmetrical optimum at $\mathbf{x}=0$ bifurcates at $p=p_0$ as the parameter p is varied (cf. *Definition 2.3*, Section 2.2). Structures, which are either actually improvable or optimally improvable, fall into the category of 'imperfect symmetry' discussed in Chapter 1.

Before stating the principal claims, some of the above key concepts (e.g. optimum based on global criterion vs. weak points, smooth optimum vs. robust optimum, potentially locally improvable vs. actually locally improvable) are illustrated on a simple example.

2.1.2 An illustrative example: bifurcation of the symmetric optimum

The symmetry of structures corresponds to the symmetry of the optimisation potentials. Smooth potentials (studied extensively in Golubitsky et. al., 1992) are adequate to model many optimisation problems in engineering, however, the classical pitchfork bifurcation of smooth, reflection-symmetric potentials predicts that the optimal symmetric solution will become pessimal, beyond a critical parameter value of a one-parameter family of optimisation problem.

This prediction may be correct in some cases, but apparently not in each one: the diagram of the optimal/pessimal values of x versus p match the general predictions if the total mass is minimised assuming constant safety against buckling in a one-parameter family of three-hinged structures (Figure 2.1/A). This is an optimisation problem *based on a global criterion* (cf. the optimisation diagram of Figure 2.1/C). However, if we assume that the two bars have given, *equal cross-sections* and the safety against buckling is investigated, the symmetrical ($x=0$) configuration proves to be (locally) optimal for all values of parameter p, despite the bifurcation in the optimum diagram (cf. Figure 2.1/B). The latter optimisation problem is based on the *concept of weak points.*

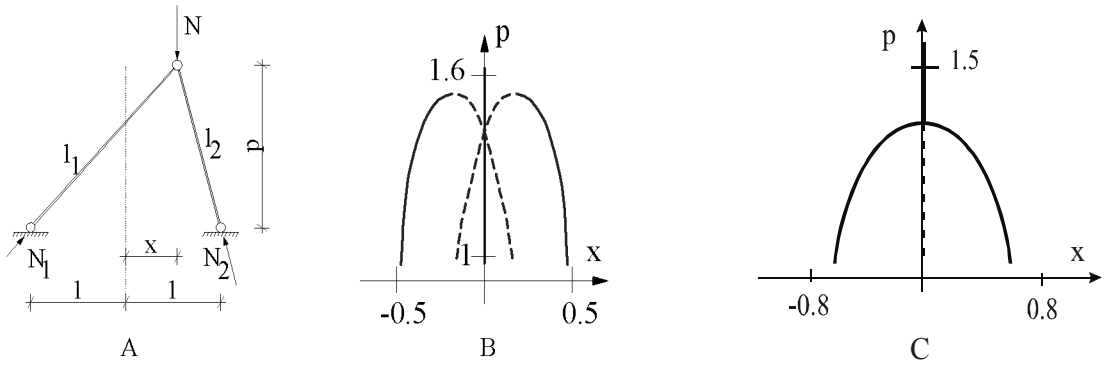


Figure 2.1 A: A simple three-hinged model loaded by the concentrated force *N*. N_i, and l_i stand for the internal forces, and the lengths of the bars, respectively. B: Optimisation diagram of safety against buckling if the cross sections are equal and independent of *x* C: Optimisation diagram of total mass if the cross-sections of the bars are circles (each of them with necessary diameter) and constant safety against buckling is demanded. (continuous line: optimum, dashed line: pessimum).

The discrepancy between the classical model's prediction and the actual behaviour of the second example can be explained if we try to define a suitable 'potential' for the optimisation problem. The safety of both bars (defined as the compressive force in the bar divided by its critical force) behaves smoothly but the envelope of these 'local potentials' will be, in general, non-smooth. In our case, the two local potentials: $f_1(p,x)$ and $f_2(p,x)$ are

$$f_1(p,x)=\frac{N_1(x,p,N)}{N_1^{cr}(x,p)}=\frac{l_1^2(x,p)}{\pi^2\cdot EI}\cdot N_1(x,p,N)=\frac{N\cdot(1-x)}{2p\pi^2\cdot EI}\cdot\left[p^2+(1+x)^2\right]^{3/2} \tag{2.1}$$

$$f_2(p,x)=\frac{N_2(x,p,N)}{N_2^{cr}(x,p)}=\frac{l_2^2(x,p)}{\pi^2\cdot EI}\cdot N_2(x,p,N)=\frac{N\cdot(1+x)}{2p\pi^2\cdot EI}\cdot\left[p^2+(1-x)^2\right]^{3/2} \tag{2.2}$$

where N_1, N_2 denote the compressive forces in the members, N_1^{cr}, N_2^{cr} stand for the corresponding Euler buckling loads, l_1, l_2 are the lengths, and EI is the flexural rigidity of the bars. Observe that $f_2(p,x)=f_1(p,-x)$. If p is constant, the 'optimisation potential' $U(x)$ of the structure can be generated from $f(x)$ as

$$U(x)=\max\big(f(x),f(-x)\big) \tag{2.3}$$

(see also Figure 2.2). It is easy to see that $U(x)$ has a (non-smooth) local minimum at $x=0$ for almost all values of p, thus this simple example *suggests* that the symmetric configurations of engineering structures are *robust (non-smooth) optima*, whenever the global optimum is determined by the worst of a discrete assembly of 'weak points'.

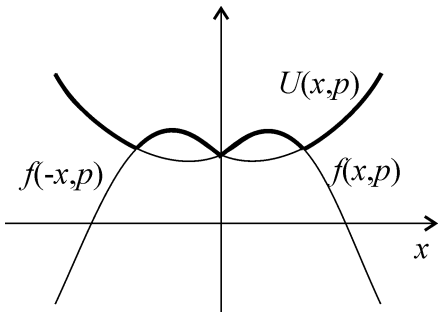


Figure 2.2: The reflected potential *U(x,p)* at *p*=0.5, generated via (2.3) from (2.1).

This type of potential is rather common in engineering practice: load-bearing structures are the most often designed based on strength conditions of the form $s \leq s_u$, where s and s_u are, respectively, the design and ultimate value of an internal force or stress. Since all parts of a structure have to meet these conditions, a straightforward choice of potential is $\max(s/s_u)$ for the whole object, which is often of type (2.3).

We can also introduce two or more perturbing variables. We will discuss later the example of Figure 2.4, which is the same as the previous one (Figure 2.1/A,C) with one additional symmetry-breaking variable. If only one of the variables is considered, $x_1=0$ or $x_2=0$ proves to be local optimum (see the $x_2=0$ and $x_1=0$ planes in Figure 2.4/B,C). However, for *simultaneous* optimisation of both variables, *robust optimality vanishes*, because the potential is typically either a partially smooth optimum (Figure 2.4/B) or a saddle point (Figure 2.4/C). In neither of the cases is it a robust optimum. According to our previous definition, it falls into category (B), i.e. it is *potentially, locally improvable,* however only the saddle point (Figure 2.4/C) corresponds to *actual* improvability.

This simple example illustrates that some symmetrical structures are *potentially,* or even *actually* improvable, other structures are robustly optimal in the symmetric configuration and cannot be improved with the given set of variables.

Our aim is to investigate this type of optimisation problems with principal focus on the question, under which conditions proves the symmetrical configuration to be a non-smooth, robust optimum (similar to the first introductory example). The answer is used to throw light on some questions of practical interest: the optimality of $x=0$ means that any minor perturbation of the symmetrical structure makes it worse. However, if one can find adequate variables where the optimality vanishes, the structure *can be improved via a small perturbation*. This way we can create unusual structural configurations with 'imperfect' symmetry, which are better than the usual, symmetric ones. However, if a small perturbation results in improvement, a bigger perturbation is likely to make the structure even better, i.e. the optimal configuration is usually strongly asymmetric. Thus, a further question is whether *an optimal structure* might or might not have 'imperfect' symmetry, i.e. slight asymmetry.

2.1.3 Principal results and structure of Chapter2

My main goal is to give simple algorithms determining the potential or optimal improvability of structures *without actually performing structural analysis,* i.e. without computing internal forces, stresses, etc. Below I give the list of my principal results, with references to the exact sources. Standard concepts of group and representation theory (see the Appendix for a summary or Jones, 1998) are used in the formulations.

First I state the most general criterion, which provides an easy-to-handle algorithm to decide whether a given structure can be locally improved in a given set of variables:

I. **I have proved that the sufficient and necessary condition of *potential local improvability* is that the representation of the Γ symmetry group of the structure in the space of variables is not sub-representation of the regular representation of Γ.** This statement is based on *Theorem 2.5*, *Theorem 2.3*, and *Definition 2.4*. The representation in the space of the variables is rigorously defined by eq. (2.4). I illustrated the application of this algorithm on many structural examples (Subsections 2.5, 2.6.2 and

2.7). My numerical computations show that *potential improvability* very often implies *actual improvability.*

My next goal was to determine weaker criteria of potential improvability.

II. **I have determined both the sufficient and the necessary criteria for potential local improvability, based solely on the number *d* of variables.** In particular,

II.1 **I proved that the typically sufficient condition of potential, local improvability is $d \geq O(\Gamma)$ where $O(\Gamma)$ denotes the order of Γ.** This condition yields for planar reflection symmetry $d \geq 2$, in case of C_m (cyclic groups) and D_m (dihedral groups) symmetry it yields $d \geq m$ and $d \geq 2m$, respectively. This condition is based on *Theorem 2.6* in Subsection 2.4.3.2

II.2 **I have proved that the typically necessary condition of potential local improvability is $d \geq 2\dim(\Gamma)$, where $\dim(\Gamma)$ denotes the dimension of the smallest real-valued representation of Γ, which has no trivial component** (cf. *Theorem 2.7* in Subsection 2.4.3.2, and *Definition I.19*). The necessary condition yields $d \geq 2$ for D_m symmetry and C_{2k} symmetry, in case of C_{2k+1} symmetry it yields $d \geq 4$. In case of C_2 and D_1, this condition agrees with both the necessary condition in Principal Result II.1 and the sufficient and necessary condition in Principal Result I. For C_3 symmetry, this result seems to contradict II.1 and I if the number of variables is 3. In fact this is not a contradiction, since an adequate set of variables cannot consist of 3 variables in this case.

II.3 **I have also proved that in case of D_1 symmetry** (e.g. planar reflection symmetry) **there exist *special, atypical* structures which can be locally improved by using only $d=1$ variable.** This statement is based on *Theorem 2.1*, Subsection 2.3.2. I determined the exact criteria for these special cases. Based on an example with D_2 symmetry (Example 3/D in Section 2.5) I demonstrated that there exist special, atypical cases (contradicting the general criteria) in other symmetry groups, as well (cf. Section 2.7).

III. **I proved the following statements regarding *optimal improvability*:**

III.1 **I have proved that in case of D_1 or C_2 symmetry** (e.g. planar reflection symmetry) and $d=1$ variable a typical, one-parameter family of structures *cannot be optimally improved*, i.e. **the typical, necessary condition of optimal improvability is $d \geq 2$**(cf. *Theorem 2.8*). In the proof I listed the possible optimum/pessimum bifurcations and provided structural examples for each listed case (cf. Section 2.6).

III.2 **I have provided an example of a structural family and a set of variables, which cannot be improved locally, however, it can be improved optimally.** Thus, I have showed, that in case of some symmetry groups, optimal improvability can be achieved with a smaller number of variables than local improvability (as opposed to D_1 symmetry). This statement is based on Subsections 2.5.5, 2.6.3.

Principal Results II.3, III.1, and a special case (reflection symmetry) of Principal Results I, II.1-2 have been published in Várkonyi et al.(in press). The publication of the rest of the results is in preparation.

Section 2.2 specifies the problem and the questions more precisely. The analysis of local optimality of the perfect configuration is investigated in parts 2.3 and 2.4. Section 2.5 illustrates the results via a number of numerical examples. In 2.6, the question of optimal

improvability is examined, while 2.7 shows and classifies exceptional cases. Finally the results are summarised in 2.8. Appendix I presents some elements of representation theory, which serves as mathematical background of Section 2.4.

2.2 GENERAL DEFINITION OF THE PROBLEM

The example in the Introduction (cf. Section 2.1.2) is generalised in two ways. First, the symmetries of structures are allowed to be different from reflection-symmetry; second, the number of symmetry-breaking variables can be arbitrary. Thus, in general, the object of our investigation is a structure, which is invariant under the elements of a finite symmetry group Γ. Such objects will be referred to as 'Γ-symmetrical'. In practical engineering problems, Γ is usually a cyclic or dihedral group or the trivial group, which corresponds to the lack of symmetry (see the arguments in Section I.2 in the Appendix). It is required that the loads, the internal forces, and, in fact, any external condition which has an effect on the optimisation process, support the 'Γ-symmetry' of the structure.

The optimisation is based on a scalar potential, which corresponds to some structural property, e.g. risk of buckling, total mass, maximum of bending moment, etc. Structures with lower potentials are considered as better.

The symmetry of the structure should be disturbed by some geometrical variables $\mathbf{x}=[x_1\ x_2 \ldots x_n]$, $\mathbf{x}\in\mathrm{R}^n$ (R stands for the set of real numbers). These variables represent the set of structures, which are considered as possible solutions of the optimisation problem. The structure corresponding to $\mathbf{x}=\mathbf{x}_0$ is referred to as $\boldsymbol{\mathcal{S}}(\mathbf{x}_0)$. The variables have to fulfill two restrictions:

(i) $\boldsymbol{\mathcal{S}}(\mathbf{x})$ is Γ-invariant if and only if $\mathbf{x}=0$.
(ii) the set $\{\boldsymbol{\mathcal{S}}(\mathbf{x}), \mathbf{x}\in\mathrm{R}^n\}$ is Γ-invariant.

Condition (i) is a consequence of the fact that our investigation relies to the optimality of the Γ-symmetrical configuration $\mathbf{x}=0$ *compared to non-symmetrical ones*, i.e. disturbed configurations should not be Γ-symmetrical. Condition (ii) means on the level of the engineering problem that if an asymmetrical configuration $\boldsymbol{\mathcal{S}}(\mathbf{x})$ is a potential solution then the transformed configuration $\gamma_i(\boldsymbol{\mathcal{S}}(\mathbf{x}))$, $\gamma_i\in\Gamma$ is also potential solution. Since the set of possible solutions is limited primarily by external conditions, which should not break the Γ-symmetry, (ii) is a *natural symmetry condition*. Finally, we require that all group elements $\gamma_i\in\Gamma$ (i=1,2,…r) should be represented by real-valued matrices $\mathbf{D}_i$ (i=1,2,…r) *in the space of the variables* $\boldsymbol{x}$, i.e. the symmetry transformations correspond to simple matrix multiplication:

$$\gamma_i(\boldsymbol{\mathcal{S}}(\mathbf{x}))=\boldsymbol{\mathcal{S}}(\mathbf{D}_i\mathbf{x}), \tag{2.4}$$

In fact, the set D of matrices $\mathbf{D}_i$ is a representation of Γ in the mathematical sense (see I.3); it will be referred to as the 'induced representation of Γ'. The above requirement is purely technical, which makes the application of representation theory on our problem easier. According to our experience, this condition does not preclude any optimisation problem of practical interest.

It is worth remarking that Γ also has a representation in the *physical space* of the structure, since the elements of Γ correspond to matrix transformations in an adequate physical co-

ordinate system. The latter representation of Γ will not gain importance during the following investigations.

As shown in the introductory part, the structural optimisation example based on a local criterion (Figure 1.1/A,B) yielded unexpected results, namely local optimality of the symmetrical configuration, which did not vanish at bifurcation points. We want to extend this result to other optimisation problems, where local goodness measures f_i correspond to elements/points of the structure (called weak points) and the worst of the weak points determines the global goodness (i.e. the potential $U(\mathbf{x})$) of the structure:

$$U(\mathbf{x}) = \max_i f_i(\mathbf{x}) \qquad i = 1,2,...,k \tag{2.5}$$

The functions $f_i(\mathbf{x})$ are supposed to be analytic, which allows approximating them by truncated Taylor-expansions if $|\mathbf{x}|<<1$.

The symmetrical configuration $x=0$ in the first example was not only a local optimum, but it was a "robust" one, i.e. for $x<<1$, we had $U(x)-U(0)\approx c|x|$ (where $c>0$ is a constant). (At smooth optima we have $U(x)-U(0)\approx c|x|^2$, for $x<<1$.) As we will see, this kind of non-smooth optimum is a characteristic property of similar examples. Before going into details, we give an exact definition of robust optima, which applies for problems with arbitrary number of variables:

Definition 2.1: The point $\mathbf{x}=0$ *is a robust local optimum (or minimum) of the scalar function* $U(\mathbf{x})$, $\mathbf{x}\in\mathbb{R}^n$ *if there exist scalars* $\delta,\varepsilon>0$ *such that* $|\mathbf{x}|<\delta$ *yields* $U(\mathbf{x})-U(0)\geq\varepsilon\cdot|\mathbf{x}|$. *(*$|\mathbf{x}|$ *denotes the* l_2*-norm of the vector* $\mathbf{x}$.

Based on *Definition 2.1*, symmetrical structures can be classified according to the following, simple scheme:

(A) $\mathbf{x}=0$ corresponds to a robust optimum. In this case, $\boldsymbol{S}(0)$ *cannot* be improved via small perturbations.
(B) $\mathbf{x}=0$ does not correspond to a robust optimum. Then we have two possibilities:
  (B1) $\boldsymbol{S}(0)$ *cannot* be improved via small perturbations.
  (B2) $\boldsymbol{S}(0)$ *can* be improved via small perturbations

As we will show, one can decide whether $\boldsymbol{S}(0)$ belongs to (A) or (B) *without* computing structural behaviour, solely based on the symmetry group Γ and the variables x_i. Here, the main goal of the thesis is to describe this algorithm and also, to formulate sufficient and necessary criteria for $\boldsymbol{S}(0)$ belonging either to (A) or to (B).

If $\boldsymbol{S}(0)$ belongs to (B), one can decide only after performing structural computations on the individual structure whether it belongs to (B1) or (B2). Although we will provide specific examples of such computations, we do not give any general method to distinguish between structures in (B1) and (B2), hence structures in category (B) are called *"potentially improvable"*. This concept is formalised in

Definition 2.2: A symmetrical structure is called 'potentially locally improvable' or simply 'potentially improvable' in a given set of variables $\mathbf{x}$*, if* $\mathbf{x}=0$ *is not robust, local optimum.*

In case of a one-parameter *family* of structures $\boldsymbol{S}(p,\mathbf{x})$ (where for each value of p criteria (i) and (ii) are satisfied) one can look for bifurcations of optima resulting in "slightly

asymmetrical optima". This concept will be also defined via the notion of "optimal improvability", in the following manner:

Definition 2.3: A one-parameter family of symmetrical structures is '(locally) optimally improvable' in a given set of variables $\mathbf{x}$ *at* $p=p_0$ *if for any* $\delta,\varepsilon>0$ *there exist* $\mathbf{x}_1$, p_1 *so that* $|\mathbf{x}_1|\le\delta$, $|p_1-p_0|\le\varepsilon$ *and* $\mathbf{S}(p_1,\mathbf{x}_1)$ *is locally optimal in* $\mathbf{x}$.

Obviously, *Definition 2.3* does not mean that $\mathbf{S}(p_0,0)$ can be optimised with a small perturbation of the symmetry, but it means that there are $\mathbf{S}(p_1,0)$ members of the family close to $\mathbf{S}(p_0,0)$, which own this property. In fact, *Definition 2.3* implies the existence of asymmetrical optima bifurcating from the $\mathbf{x}=0$ line (which is always a critical point, i.e. local optimum, pessimum, or saddle due to the symmetry of the potential) at $p=p_0$. For smooth potentials, such branches emerge in classical pitchfork bifurcations (e.g. Figure 2.1/B). On the other hand, bifurcations from non-smooth, robust optima seem to emerge in unusual bifurcation patterns (cf. Figure 2.1/C) where the bifurcating branches do not carry optimal solutions. Thus, "optimally improvable" structures are likely to be even rarer as locally improvable structures. This property will be investigated later via bifurcation analysis of optimum diagrams.

In the next unit, we start the examination of potential, local improvability in the simplest case of reflection-symmetrical structures.

2.3 REFLECTION SYMMETRY

In case of reflection symmetry (D_1 group), the elements of Γ are $\gamma_0\equiv$identity, and $\gamma_1\equiv$reflection, which will be represented by the real-valued matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ in the space of the variables: $\mathbf{D}_0=\mathbf{I}_k$ ($k\times k$ identity matrix) and $\gamma_1\gamma_1=\gamma_0$ yields $\mathbf{D}_1^2=\mathbf{I}_k$. Thus, if there is one perturbing variable, then $\mathbf{D}_0=1$, $\mathbf{D}_1=\pm1$. Since $\mathbf{D}_1=1$ contradicts condition (i), the only possibility is: $\mathbf{D}_0=-\mathbf{D}_1=1$. Similarly, one can derive $\mathbf{D}_0=-\mathbf{D}_1=\mathbf{I}_k$ from condition (i) in case of several variables. Thus, if D_1-symmetrical structures are optimised with the given conditions (i),(ii), reflection of the structures corresponds to changing the variables from $\mathbf{x}$ to $-\mathbf{x}$. Notice that the example of Figure 2.1 also had this property.

We remark that the group C_2 is isomorphic to D_1, i.e. the results of this section are valid for C_2-symmetry, as well.

2.3.1 Optimisation with one variable

The optimisation potential of a D_1-symmetrical example is either of the form of (2.3), or if the structures have more than two local weak points, it is of the form

$$\hat{U}(x)=\max\left(U_i(x)\right),\qquad i=1,2,\ldots,n, \tag{2.6}$$

where the functions $U_i(x)$ are defined by (2.3). Examples of the latter type are presented later, in subsection 2.6.2.

As demonstrated in the Introduction, potentials of type (2.3) (and, in fact, also (2.6)) have robust optima at $x=0$, unless $df/dx|_{x=0}=0$. Thus, reflection-symmetrical structures are typically

locally non-improvable with one perturbing variable. This property is generalised later to arbitrary symmetry and stated in Principal Result II.

2.3.2 Exceptional cases

There are special types of structures where x=0 is not a robust optimum, because $df/dx|_{x=0}\equiv 0$. We show two examples in Figure 2.3. In both cases, U is defined as the maximum of the bending moment along the beams.

In case A, the global moment maximum (i.e. the weakest point) of the symmetrical structure is at the middle of the beam if p is adequately big. The weakest point is invariant to the reflection of the structure, thus, the corresponding potential is symmetrical ($f_2(x)= f_2(-x)$), its first derivative is always 0 and $U(x)=\max(f_2(x), f_2(-x))$ is smooth at x=0. $\boldsymbol{S}(0)$ is potentially improvable in the sense of *Definition 2.2* (and, in this case, it actually *is* improvable).

In case B, the weakest point is not in the middle, i.e. the above 'symmetry argument' does not apply. Still the structure proves to be locally improvable, due to a specific property of its moment diagram: if small displacements are assumed (i.e. the secondary effect of the deformations on the loads is neglected), the emerging moment diagram happens to be symmetric, even if $x\neq 0$. Thus, $f_1(x)=f_1(-x)$ again holds. Here again, potential improvability implies actual improvability. We conclude with

Theorem 2.1: There exist exceptional reflection-symmetric structures, which are potentially improvable with one variable. Examples where the weakest point is invariant to reflection as well as some structures with atypical potential functions belong to this category.

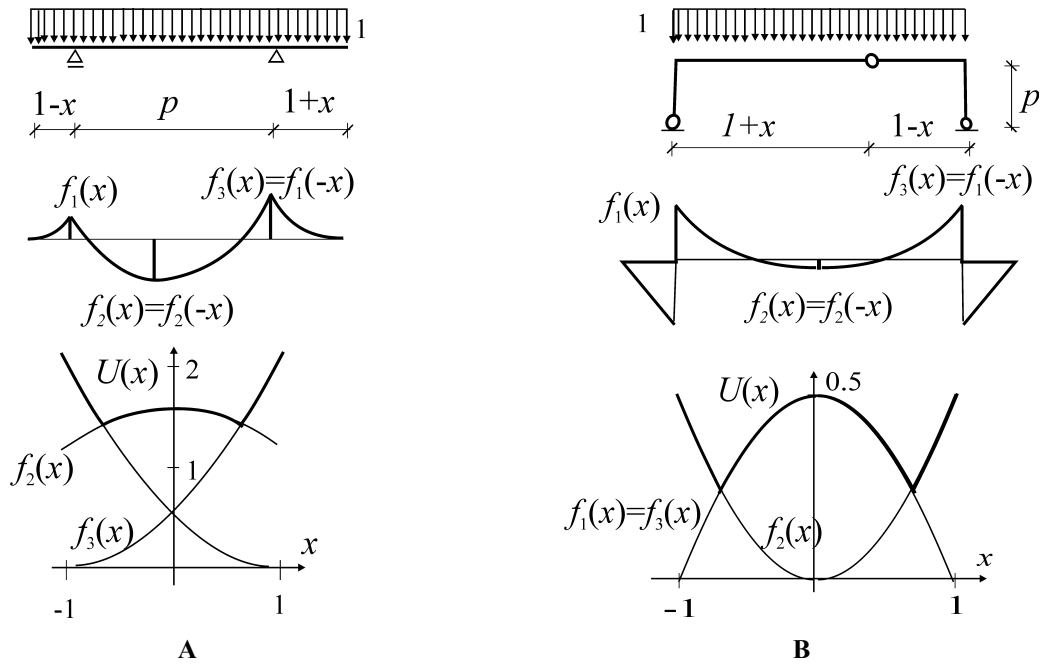


Figure 2.3 A: A beam, its moment diagram and the corresponding potential $U=\max(f_i)$ if p=4. The symmetrical configuration is potentially, locally improvable, because the weakest point of the beam is invariant to reflection. B: another example, where the symmetrical structure is locally improvable, because $f_1(x)\equiv f_3(x)$ holds for this specific example

Notice that the behaviour of the example "A" follows from simple symmetry-considerations. Identifying such a case needs only partial analysis (i.e. finding the weakest point of the perfect structure). However case "B" is completely example-specific, it could not be recovered but via determining the inner forces of the structure for all values of x. More details about types of exceptions are presented in part 2.7. Other exceptional examples are discussed in subsection 2.6.2 (D_1-symmetry) and 2.5 (other symmetries).

2.3.3 Several variables

One also can introduce two or more perturbing variables. The example of Figure 2.4 is the same as our first example (Figure 2.1/A,B) with one additional symmetry-breaking variable. The variables fulfil conditions (i)-(ii). If only one of them is considered, x_1=0 or x_2=0 proves to be local optimum (see the x_2=0 and x_1=0 planes in Figure 2.4/B,C). For simultaneous optimisation of the two variables, robust optimality vanishes, because the potential of the form $U(\mathbf{x})=\max[f(\mathbf{x}),f(-\mathbf{x})]$ has a smooth subspace across the point $\mathbf{x}$=0 (the 'wedge', where $f(\mathbf{x})=f(-\mathbf{x})$), consequently $\mathbf{x}$=0 is typically either optimum (Figure 2.4/B) or a special type of saddle point (Figure 2.4/C). In neither of the cases is it a robust optimum (because of the smooth subspace). According to *Definition 2.2*, it is *potentially, improvable,* however only the saddle point (Figure 2.4/C) corresponds to *actual* improvability. Detailed numerical analysis of the two-variable example is discussed in part 2.5.2.

We would like to remark here that this kind of saddle is 'almost' an optimum in the sense that an infinitely small *random perturbation* of the symmetry typically makes the structure worse (Figure 2.4/D). This property is true for arbitrary symmetry group and number of variables provided that $\mathrm{grad}f(\mathbf{x})|_{\mathbf{x}=0}\neq 0$ ('grad' denotes the gradient vector of scalar functions).

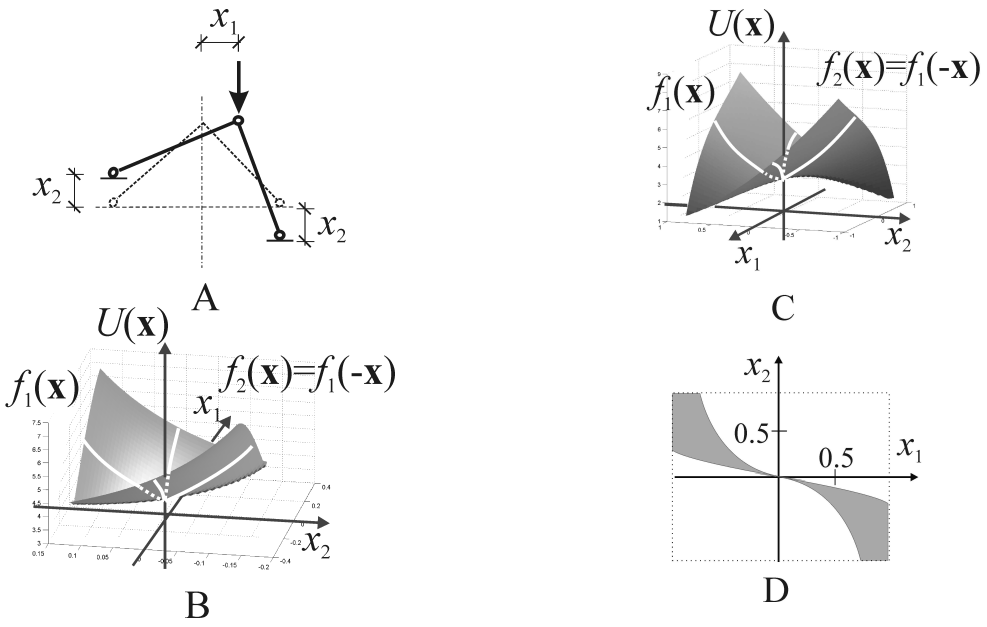


Figure 2.4 A: A simple three-hinged model with two perturbing variables. B: optimisation potential of the structure if p=0.15. The symmetrical configuration is wedge-like (i.e. not robust) optimum. Notice that optimisation with only one of the variables (white sections of the surface) would result in robust optimum. C: optimisation potential if p=2. The symmetrical configuration is a wedge-like "saddle" i.e. not optimum. D: The grey domain indicates values of x, for which $U(p,\mathbf{x})<U(p,0)$ if p=2. Notice that a randomly chosen small ($|\mathbf{x}|<<1$) value of x is typically out of this range. Thus, a small, random perturbation of the symmetry typically spoils the structure.

2.4 GENERAL SYMMETRY

2.4.1 Introduction

The previous subsection investigated optimality of reflection-symmetrical structures in problems where we had a set of local optimality criteria and the worst one determined the global 'goodness' of the structure. We have also demonstrated in the introduction that this kind of optimisation problems is rather general in the engineering practice.

Planar structural models with reflection symmetry are often called simply 'symmetric'. On the other hand, spatial structural models have many different types of symmetries. The goal of the forthcoming part is to generalise the assessments of Section 2.3 for structures with arbitrary symmetry. Several concepts and results of group and representation theory are used in this part. These are summarised in Appendix I.

As primary result for reflection symmetry, it has been demonstrated that in case of optimisation with one variable x, the symmetrical configuration x=0 was a local, 'robust' optimum. Conversely, when two (or more) variables were optimised simultaneously, robust optimality of the $\mathbf{x}$=0 configuration vanished. The order of the group D_1, associated with reflection symmetry, is 2, which yields the primary intuition that structures with r-order symmetries are potentially locally improvable if and only if the number d of variables is equal or more than r. The next subsections will show that only the 'if' part of the previous conjecture is true, however we can still determine *sufficient and necessary* conditions of potential local improvability, based on symmetry considerations.

First, in subsection 2.4.2 we derive an exact, however *example-specific* sufficient and necessary condition of robust optimality. This is further developed in subsection in 2.4.3 to find *typical* conditions, which are based *solely on symmetry considerations*. The latter conditions are the primary results of this part of the thesis.

2.4.2 Exact condition of robust optimum

In this part we derive a necessary and sufficient condition of the robust optimality of $\mathbf{S}(0)$, which is based on the knowledge of the inner forces of the specific example. Later we will show that in non-degenerate, typical cases this result can be generalised to structures with *unknown inner forces*, yielding the principal results of the thesis.

The potential of the whole structure is of the form of (2.5). Local properties of $U(\mathbf{x})$ at $\mathbf{x}$=0 are only influenced by the weakest points of the perfect configuration $\mathbf{S}(0)$, i.e. those $f_i(\mathbf{x})$ functions for which $f_i(0)= U(0)$. At the same time, the perfect structure $\mathbf{S}(0)$ has usually more than one 'weakest' points due to its symmetry.

Let $f_i(\mathbf{x})$, i=1,2,...,k denote the potentials associated with the set of weakest points of $\mathbf{S}(0)$:
The linear approximation of U is

$$U(\mathbf{x}) = U(\mathbf{0}) + \max_i \left(\mathbf{g}_i^T \mathbf{x}\right) + o\left(|\mathbf{x}|^2\right) \qquad i = 1,2,...,k \quad \text{if } |\mathbf{x}| << 1 \tag{2.7}$$

where $\mathbf{g}_i$=grad$f_i(\mathbf{x})|_{\mathbf{x}=0}$. With these notations,

Theorem 2.2: The configuration $\mathbf{S}(0)$ *is a robust optimum if and only if* $\mathbf{x}=0$ *is an internal point of the convex hull of the endpoints of vectors* $\mathbf{g}_i$*, i=1,2,...,k (points at the border are not considered as internal points).*

Proof of the "if" part of Theorem 2.2: Assume that $\mathbf{x}=0$ is inside the convex hull. In this case it can be written as a convex combination of the nodes of the convex hull:

$$0=\sum_{i=1}^{k} c_i \mathbf{g}_i \qquad \text{where} \qquad c_i>0, \qquad \sum_{i=1}^{k} c_i = 1 \tag{2.8}$$

Transposing both sides of eq. (2.8) and multiplying them by a unit vector ($\mathbf{v}\in\mathrm{R}^d$, $|\mathbf{v}|=1$) yields

$$0=\sum_{i=1}^{k} c_i \mathbf{g}_i^T \mathbf{v} \tag{2.9}$$

In the above sum, either all components are 0 or some of the components are positive. If there exists a $\mathbf{v}=\mathbf{v}_0$, for which all terms are zero, then $\mathbf{v}_0$ is orthogonal to the vectors $\mathbf{g}_i$ i.e. the vectors do not span the d-dimensional space of the variables R^d, which means that their convex hull has no internal point at all in R^d. This is in contradiction with the initial assumption. Thus, there must be a positive component in (2.9) for arbitrary $\mathbf{v}$:

$$\max_{1\le j\le k}\left(\mathbf{g}_j{}^T \mathbf{v}\right)>0 \quad \text{for arbitrary} \quad |\mathbf{v}|=1, \tag{2.10}$$

The function on the left side of eq. (2.10) is continuous in $\mathbf{v}$ and the set $\{\mathbf{v}\in\mathrm{R}^d, |\mathbf{v}|=1\}$ is compact. According to the Extreme Value Theorem (see e.g. Malik, 1992), such functions have a global minimum, which is positive, due to eq. (2.10)

$$m=\min_{|v|=1}\left(\max_{1\le j\le r}\left(\left(\mathbf{D}_j\mathbf{g}\right)^T\mathbf{v}\right)\right)>0, \tag{2.11}$$

Eq. (2.7) can be rearranged as

$$U(\mathbf{x})=U(\mathbf{0})+\max_{1\le i\le r}\left(\left(\mathbf{g}_i{}^T\bar{\mathbf{x}}\right)\cdot|\mathbf{x}|+o\left(|\mathbf{x}|^2\right)\right), \tag{2.12}$$

where $\bar{\mathbf{x}}=\mathbf{x}/|\mathbf{x}|$ and $|\bar{\mathbf{x}}|=1$. From (2.11) and (2.12) we have

$$U(\mathbf{x})\ge U(\mathbf{0})+m\cdot|\mathbf{x}|+o\left(|\mathbf{x}|^2\right), \tag{2.13}$$

which means that $\mathbf{x}=0$ is robust optimum of $U(\mathbf{x})$ (cf. *Definition 2.1*). Q.e.d.

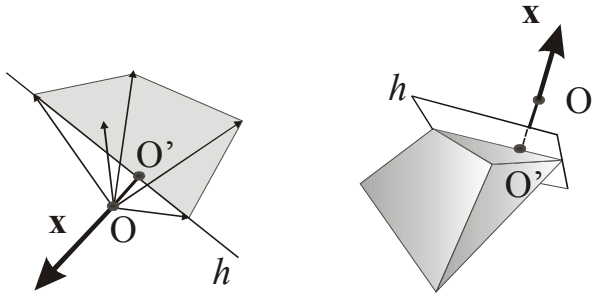


Figure 2.5: Illustration to the 'only if' part of *Theorem 2.2* if *d*=2 (left panel) and *d*=3 (right panel). Grey colour denotes the convex hull of the gradient vectors. The vectors themselves are hidden in the 3D case.

Proof of the "only if" part of Theorem 2.2: Instead of a rigorous proof, we show an illustrative sketch of proof, which is easy to imagine if $d\le3$ and applies for arbitrary d. (See also Figure 2.5) Assume that $\mathbf{x}=0$ (denoted by "O") is not in the convex hull! Consider the nearest point O' (in sense of l_2-norm) of the border of the convex hull to O (This might be O itself). For every border point P of an arbitrary d-dimensional convex object, we can find a 'tangent' d-1 dimensional hyperplane h (h is a plane if d=3,it is a line if d=2), which contains

P and has the whole convex object on one side. Consider a tangent hyperplane of the convex hull at O'. Let **x** be the normal vector of *h*, which is on the opposite side of *h* than the convex hull. In this case, $\mathbf{g}_i^T\mathbf{x}\le 0$ for arbitrary $1\le i\le r$. According to (2.12), $U(\mathbf{x})\le U(0)+o(|\mathbf{x}|^2)$, i.e $\mathbf{x}=0$ cannot be robust optimum: the symmetric structure is potentially improvable. Q.e.d.

2.4.3 Application of representation theory to optimisation problems

In the following part, the *typical* conditions for the robust optimality / potential improvability of the symmetric configuration $\mathbf{S}(0)$ are derived. These conditions refer to the number and type of the perturbing variables. As we will see, the results are *typical* but not exact in the sense that they do not hold for degenerate types of potential functions (e.g. in the reflection-symmetrical part, $U(x)$ potentials were exceptional if generated from $f(x)$ with zero derivative at $x=0$). We use many results of representation theory, which are summarised in Appendix I.

In this subsection we will use the following notations (in accordance with previous ones) : vector **x** for the perturbing variables and *d* for the number of variables, $\mathbf{S}(\mathbf{x})$ for the corresponding structures, $\Gamma\equiv\{\gamma_1,\gamma_2,...,\gamma_r\}$ for the symmetry group of $\mathbf{S}(0)$, $D\equiv\{\mathbf{D}_1, \mathbf{D}_2,..., \mathbf{D}_r\}$ for the elements of the induced representation in the space of the variables, $f_i(\mathbf{x})$ ($i=1,2,...,k$) for the potentials associated with the weakest points of $\mathbf{S}(0)$, *P* for one of the weakest points, $f_P(\mathbf{x})$ for the corresponding potential and finally $\mathbf{g}_i=\mathrm{grad}f_i(\mathbf{x})|_{\mathbf{x}=0}$ and $\mathbf{g}=\mathrm{grad}f_P(\mathbf{x})|_{\mathbf{x}=0}$.

In subsection 2.4.3.1, we outline that the set of weakest points can typically be generated from one such point *P* as the orbit of *P* with respect to Γ (i.e. $\{\gamma_i(P)$, $i=1,2,...,r$, see more about orbits in part I.4). This way we find a simple connection between the functions $f_i(\mathbf{x})$, which yield a characterisation of the convex hull of the endpoints ofvectors $\mathbf{g}_i$ (Subsection 2.4.3.2). Thus, we can apply *Theorem 2.2* without detailed computation of a specific example.

2.4.3.1 Orbits in the optimisation problem

Let *P* denote one of the weakest points of $\mathbf{S}(0)$ (such as the left bar of the introductory example of . Figure 2.1/A). Due to the Γ-symmetry of $\mathbf{S}(0)$, each weak point $\gamma_i(P)$ is identical to *P*, thus their potentials are equal if $\mathbf{x}=0$ (in the introductory example $\gamma_0(P)$ and $\gamma_1(P)$ are the left and the right bar, respectively). It is possible, however atypical, that some additional points have the same potential.

If the potential associated with *P* is $f_P(\mathbf{x})$, then the potentials of other weakest points are $f_i(\mathbf{x}) = f_{\gamma_i(P)}(\mathbf{x}) = f_P(\mathbf{D}_i\mathbf{x})$. Thus, the resultant optimisation potential $U(\mathbf{x})$ of the structure is typically of he form

$$U(\mathbf{x}) = \max_{i=1}^{r} f_P(\mathbf{D}_i\mathbf{x}) \qquad if \quad |\mathbf{x}| << 1, \tag{2.14}$$

and '>' emerges in eq. (2.14) instead of '=' atypically. (This is the case at the example of Figure 2.3 if *p* is chosen so that $f_1(0)=f_2(0)$).

Let $\mathbf{g}$ denote $\mathrm{grad}(f_P(\mathbf{x}))|_{\mathbf{x}=0}$. With this notation, $\mathbf{g}_i=\mathrm{grad}(f_P(\mathbf{D}_i\mathbf{x}))|_{\mathbf{x}=0}=\mathbf{D}_i^T\mathbf{g}$. Notice that the vectors $\mathbf{g}_i=\mathbf{D}_i^T\mathbf{g}$ are the orbit of $\mathbf{g}$ with respect to the representation $D^T\equiv\{\mathbf{D}_i^T\ j=1,2,...,r\}$. Thus from eq. (2.14) and *Theorem 2.2*, we have

Lemma 2.1: Let $\mathbf{g}$ *denote the gradient of the potential associated with one of the weakest points of* $\mathbf{S}(0)$. *If* $\mathbf{x}=0$ *is an inner point of the convex hull of the orbit of* $\mathbf{g}$ *with respect to*

$D^T \equiv \{\mathbf{D}_j^T\ j=1,2,...,r\}$, $\mathbf{x}=0$ is a robust optimum. Vice versa, if $\mathbf{x}=0$ is a robust optimum, $\mathbf{x}=0$ is typically inside the convex hull.

The application of *Lemma 2.1* still assumes the knowledge of **g**, which is example-specific. In the next subsection, an improved version of *Lemma 2.1* is derived, which is based solely on the type of the induced representation. We will use the fact that matrix transposition is a unitary transformation, i.e. D^T is equivalent to the induced representation D (cf. *Definition I.10*).

2.4.3.2 Analysis of the induced representation

Now we can formulate the most general theoretical results. The goal of this part is to give *typical* predictions on the optimality/improvability of a symmetrical structure ($\mathbf{x}=0$), based solely on the symmetry group Γ and the induced representation D. We show first that the analysis of Γ and D indicates if an example violates condition (i). Second, we derive the typical, necessary *and* sufficient conditions of potential improvability of $\mathbf{S}(0)$ based on Γ and D, and finally we give *separate* necessary and *separate* sufficient conditions for the potential improvability based only on the number d of perturbing variables.

The forthcoming statements are based on classical results of representation theory, which are summarised in Appendix I. One of the basic elements of the theory is the decomposition of representations to the direct sum of irreducible components, which is unique and easy to construct for representations of finite groups. The technique of creating such a decomposition is also discussed in the Appendix. The only remarkable difference between the classical theory and the structural applications is that the former one applies for complex-valued representations, however the representations emerging in structural optimisation are necessarily real-valued. Thus, some results are modified to apply for real-valued representations (Subsection I.5). Among others, we define the notion of 'half-irreducibility', which means irreducibility among real-valued matrices (*Definition I.19*). The half irreducible decomposition of a representation is also unique and simple to create.

We remark that two special types of representations play an important role in structural optimisation: one is the trivial representation (*Definition I.8*), in which every group element is represented by the scalar 1, and the other one is the regular representation of Γ (*Definition I.13*), which consists of $r \times r$ matrices and which has among others a trivial component in its irreducible decomposition.

Effect of condition (i)

Condition (i) excludes the Γ-symmetry of $\mathbf{S}(\mathbf{x})$ if $\mathbf{x} \neq 0$. The structure $\mathbf{S}(\mathbf{x})$ would be Γ-symmetric if and only if we had $\mathbf{S}(\mathbf{x})=\gamma_i(\mathbf{S}(\mathbf{x}))=\mathbf{S}(\mathbf{D}_i\mathbf{x})$ for every $1 \leq i \leq r$. This would imply that $\mathbf{D}_i\mathbf{x}=\mathbf{x}$ for every $1 \leq i \leq r$, i.e. the induced representation would have a nontrivial invariant point. Such points occur if and only if D has a trivial component in its irreducible decomposition (see a summary on the decomposition of representations in I.3.2 and also *Lemma I.1* in I.4.1). Hence,

Theorem 2.3: Condition (i) is satisfied iff the induced representation has no trivial component.

Although problems violating condition (i) are not subject of this research, we remark that there is no technical difficulty of analysing such optimisation problems. In that case one can

prove (see *Lemma I.3*), that **x**=0 is usually not inside the convex hull of the endpoints of vectors $\mathbf{D}_i^T\mathbf{g}$, i=1,2,...,r**,** thus *Lemma 2.1* yields

Lemma 2.2: If the induced representation has a trivial component (i.e. condition (i) is not satisfied), the unperturbed configuration **x**=*0 is typically not a robust optimum.*

Example 2/B in Subsection 2.5.3 is a numerical illustration of this case.

Typical conditions of potential improvability

If condition (i) is satisfied, the induced representation has no trivial component (cf. *Theorem 2.3*) and **x**=0 is a convex combination of the orbit of **g** (the coefficients in eq. (2.8) are c_i=1/k according to *Lemma I.2*). Hence, **x**=0 is inside the convex hull of the orbit of **g** unless the convex hull is degenerate (Degeneracy means that the elements of orbit do not span $\mathbb{R}^d$, in which case their convex hull has no "inside", only border points). Thus, *Lemma 2.1* yields:

Lemma 2.3: If the orbit of **g** *with respect to the representation* D^T *spans* $\mathbb{R}^d$ *and* D^T *(or, equivalently, D) has no trivial component, the symmetrical configuration* **x**=*0 is a robust local optimum. Otherwise,* **x**=*0 is typically not a robust local optimum.*

The 'otherwise' part of *Lemma 2.3* also follows from *Lemma 2.1* if the induced representation has no trivial component (i.e. condition (i) is not violated); it is a consequence of *Lemma 2.2* if *D* has a trivial component.

Lemma 2.3 reduces the question of robust optimality to deciding whether the orbit of **g** spans $\mathbb{R}^d$ or not, which depends primarily on the type of the induced representation *D*. If *D* is half-irreducible, the condition of *Lemma 2.3* is true for arbitrary **g**≠0 (see Lemma I.9). Thus we have:

Theorem 2.4: If the induced representation is half-irreducible but it is not the trivial representation and the gradient **g** *is non-zero, then* **x**=*0 is a robust local optimum of potential U(***x***), i.e.* **$**(**x**)** *is not potentially improvable.*

It is also shown in the Appendix (Lemma I.8) that if the induced representation is not half-reducible but it is a sub-representation of the regular representation of Γ (these basic concepts are defined in the Appendix), then the orbit of a *typical* vector **g** spans $\mathbb{R}^d$. Since the regular representation has one "forbidden" trivial component, it is worthwhile to introduce

Definition 2.4: A representation is called cyclic if it is a sub-representation of the regular representation of Γ*, and it has no trivial component.*

Based on *Lemma 2.3* and the above results, we can now formulate

Theorem 2.5: **$**(0) *is typically robustly optimal iff the induced representation is cyclic*.

"Typically" means on the one hand that an *adequately chosen* gradient **g** may make **$**(0) potentially improvable if *D* is cyclic, but with randomly chosen **g**, the chance of getting robust optimum at **x**=0 is 1. On the other hand, robust optimality of **$**(0) is possible although atypical if *D* is not cyclic. The latter option originates from the possibility of inequality in eq. (2.14).

Weakened results based on the number of variables

The conditions of the previous unit (*Theorem 2.4*, *Theorem 2.5*) need only the decomposition of the induced representation as input, which is much more simple to perform than detailed structural analysis. At the same time, we can often predict optimality/improvability of $\mathbf{S}(0)$ if only the number of variables is known.

The regular representation of the group Γ of order r is r-dimensional (i.e. it consists of $r\times r$ matrices). At the same time, it has one trivial component. Thus, cyclic representations in the sense of *Definition 2.4* are at most r-1 dimensional, i.e. *Theorem 2.5* yields

Theorem 2.6: If $d\geq r$, then $\mathbf{S}(0)$ *is typically potentially improvable.*

At the same time, half-irreducible representations are always sub-representations of the regular representation of Γ, i.e. they are cyclic (except if trivial, which is excluded by condition (i)). A one-dimensional representation is obviously irreducible (i.e. also half-irreducible), thus the perturbation of a structure's symmetry with one variable does not yield potential improvability due to *Theorem 2.5*. At the same time, some groups have only k or more dimensional half-irreducible (non-trivial) representations, where $k\geq 2$. In these cases a non-cyclic representation is at least $2k$-dimensional, i.e.

Theorem 2.7: If $d<2\dim(\Gamma)$ (where $\dim(\Gamma)$ means the dimension of the smallest non-trivial, half-irreducible representation of Γ *and d is the number of perturbing variables) and condition (i) is satisfied,* $\mathbf{S}(0)$ *is typically robustly optimal.*

In practical optimisation problems, Γ is a cyclic (C_n) or dihedral (D_n) group (see the meaning of the notations and the origin of this fact in Section I.2). Among these groups, C_n has only two-dimensional half-irreducible components if n is odd (see in part I.5), thus in this case $\mathbf{S}(0)$ is typically a robust optimum if it is perturbed by less than 4 arbitrary variables. (It follows from the same fact that the number of variables satisfying (i) and (ii) can be 2 but cannot be 1 or 3) In case of dihedral symmetries and cyclic ones of even order, only optimisation with one *arbitrary* variable yields robust optimality.

Notice that *Theorem 2.6* (necessary condition of potential improvability), *Theorem 2.7* (sufficient condition) and *Theorem 2.5* (necessary and sufficient condition) are equivalent in case of C_2 and D_1-symmetries. The latter one imeans reflection-symmetry. Hence it is not surprising that predictions about the potential improvability of reflection-symmetrical structures were based only on the number of variables.

2.5 NUMERICAL OPTIMISATION EXAMPLES

This part contains the detailed analysis of several structural optimisation problems. The main steps of the investigations are described below (part 2.5.1). After the general description, the examples are presented in 2.5.2-2.5.5. Example 1 is a reflection-symmetrical structure, the rest are D_2 or D_3 symmetrical. The two versions of Example 2 illustrate the analytical results. Example 3 has four variants, one of which proves to be exceptional according to the numerical results. Finally, Example 4 illustrates another exceptional case.

2.5.1 Main steps of the analysis

The analysis starts with prediction of the potential improvability of $\mathcal{S}(0)$ based on the results of Section 2.4. After that, detailed numerical analysis follows in each case. Details of the steps are described below.

Analytical part

It is decided analytically whether $\mathbf{x}=0$ is typically a robust optimum or not. Steps of the analysis are the following:

Step I. Determination of the induced representation of the problem. This is an intuitive but not difficult step. Notice that this step requires the fulfilment of condition (ii) and also that the elements of the group Γ are represented by linear matrix transformations in the space of the variables.

Step II. Construction of the irreducible and half-irreducible decompositions of the induced representation. (The technique is summarised in Section I.3.2).

Step III. ‚Categorisation of the optimisation problems according to *Theorem 2.3*, *Theorem 2.4*, and *Theorem 2.5* (1: condition (i) does not hold; 2: robust optimum is atypical; 3: robust optimum is typical; 4: robust optimum is sure unless $\mathbf{g}=0$).

For exact results concerning the robust optimality of $\mathbf{x}=0$, further steps could follow:

Step IV. Selection of one of the weakest points (denoted by P so far). Some exceptional structures (similar to the example of Figure 2.3/A) can be detected this way.

Step V. Determination of the gradient $\mathbf{g}$ of the potential $f_P(\mathbf{x})$ at $\mathbf{x}=0$, to check robust optimality directly. Atypical but possible results as well as further exceptional structures (such as that of Figure 2.3/B) can be detected this way.

The benefit of the last two steps is the recognition of atypical results (e.g. where $\mathbf{g}$ happens to be 0) and exceptional cases. Further information on types and handling of exceptions can be found in part 2.7. In case of simple examples, steps IV, and V. can be performed analytically, however, in case of more complicated structures they may include numerical work as well.

Numerical part

Beyond the local, analytical investigation at $\mathbf{x}=0$, global numerical optimality analysis yielding all local minima, maxima, and saddle points (i.e. all critical points) of the potential function U was performed in case of some of the examples. These calculations have been done on a PC in MATLAB environment using the Simplex Method (Allgover et al, 1990). Most of the examples have two symmetry-breaking variables ($d=2$), and one symmetry-preserving parameter, thus the results are plotted in 2+1 dimensional optimum diagrams. The parameter was included to get an overview on a family of structures and the number of variables was restricted because increasing the number of variables results in exponential growth of computational time. According to my experience, computing structural families with three or more variables would need more sophisticated tools (e.g. path following algorithms, see Domokos et al, 1995 or parallel processing in clusters, see Domokos et al, 2001) and developing an ‘optimising software’ was not a primary goal of this work.

The bifurcation diagrams show the location and types of the *critical points* of the potential function. Since non-smooth potentials of type (2.5) are not everywhere differentiable (e.g. usually at $\mathbf{x}=0$), we need a generalised interpretation of critical points. Critical points of a

smooth $h(\mathbf{x})$ function are the solutions of the grad($h(\mathbf{x})$)=*0* equation. The first derivatives $U_i(\mathbf{x})=dU(\mathbf{x})/dx_i$ of $U(\mathbf{x})$ suffer discontinuities of the first kind at some points, i.e. the left-hand and right-hand derivatives exist but are not identical. To overcome this problem, we apply the concept of *interval derivative*, (see e.g. Korn et al, 1968) which results in the interval $[U_{i\ \mathrm{left}}(\mathbf{x}),\ U_{i\ \mathrm{right}}(\mathbf{x})]$. (For example, the interval derivative of $f(x)=|x|$ at $x=0$ is the interval [-1,1]. At smooth points the interval derivative gives a scalar, identical to the classical derivative. With this concept, the generalised gradient is a 'vector', the entries of which are intervals and

*Definition 2.5: A point of U(***x***) is called critical if each entry of the generalised gradient contains zero as an element.*

In case of optimisation with 2 variables (d=2), there are three distinct kinds of critical points (see also Figure 2.6):

- smooth critical points of U are also critical points of some of the generating functions $f_i(\mathbf{x})$. where $\mathbf{x}=[x_1\ x_2]$. These can be determined from the equation
$$\mathrm{grad}\left(f_i(\mathbf{x})\right)=[0\quad 0]^{\mathrm{T}}, \tag{2.15}$$
which should be solved for every $1\le i\le n$ (n denotes the number of weak points of the structure). As supplementary condition, one should also check that the solutions $\mathbf{x}=\mathbf{x}_0$ really emerge in U, i.e. that $f_i(\mathbf{x}_0)=\max[f_h(\mathbf{x}_0)\ h=1,2,...,n]$
- wedge-like (i.e smooth in one direction, non-smooth in the other) critical points emerge where two intersecting functions f_i and f_j form a 'wedge' and the bottom of the wedge (which is a smooth line) has a critical point. These are determined from the
$$\begin{cases} f_i(\mathbf{x})=f_j(\mathbf{x}) \\ \mathrm{grad}\left(f_i(\mathbf{x})\right)\times\mathrm{grad}\left(f_j(\mathbf{x})\right)=0 \end{cases} \tag{2.16}$$
equations, which should be examined for every pair $1\le i,j\le n$. In the second equation, '×' denotes vector product and this equation corresponds to the condition that the line of intersection of the two functions (the bottom of the wedge) has zero slope. The supplementary condition $\mathrm{grad}[f_i(\mathbf{x})]^{\mathrm{T}}\cdot\mathrm{grad}[f_j(\mathbf{x})]<0$ ensures that the line of intersection of the functions is really a 'wedge'; $f_i(\mathbf{x})=\max[f_h(\mathbf{x})\ h=1,2,...,n]$ should also be checked.
- robust (i.e. sharp in all directions) optima may emerge when three (or more) of the generating functions coincide. The corresponding equations are
$$f_i(\mathbf{x})=f_j(\mathbf{x})=f_k(\mathbf{x}), \tag{2.17}$$
which should be solved for every triple $1\le i,j,k\le n$. The functions enclose a robust optimum, if and only if the point $\mathbf{x}=0$ is inside the convex hull of their gradient vectors (cf. *Theorem 2.2*); $f_i(\mathbf{x}_0)=\max[f_h(\mathbf{x}_0)\ h=1,2,...,n]$ should also be checked.

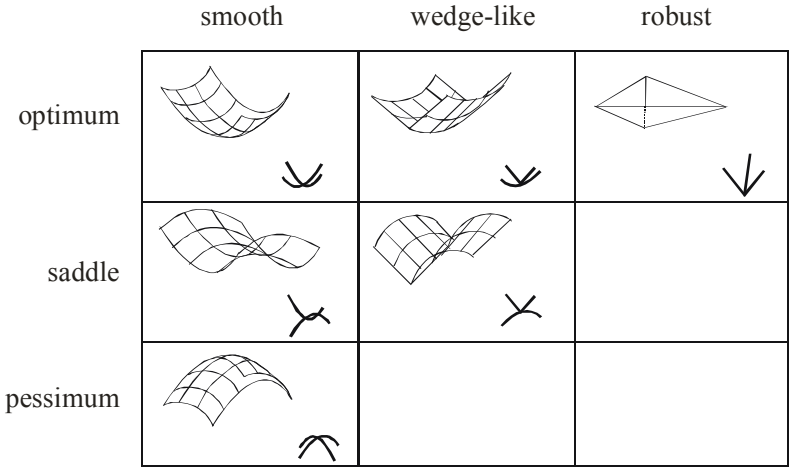


Figure 2.6: Generic types of critical points at functions of the form of (2.5). Symbolic pictograms of each type are also presented. We use these later to identify the critical points in the bifurcation diagrams.

2.5.2 Example 1

As first example, the structure of Figure 2.4 is analysed. It has been pointed out in Section 2.3 that there is typically no robust optimum at reflection-symmetric examples with two or more variables. We can regain these results with the general tools based of representation theory (Section 2.4): the symmetric structure has two invariant transformations: $\gamma_0 \equiv$identity and $\gamma_1 \equiv$reflection to the y axis; it is D_1-symmetrical. The corresponding elements of the induced representation are $\mathbf{D}_0$=<1 1> and $\mathbf{D}_1$=<-1 -1>, where <*> denotes a diagonal matrix with elements *. The irreducible decomposition of this representation consists of two I_1 components (see the notations of Table I.4). Since the regular representation of D_1 contains only one example of I_1, the induced representation is not cyclic, $\mathbf{x}$=0 is potentially locally improvable.

Numerical analysis was also performed for arbitrary positive value of p (Figure 2.7). According to the results, $\mathbf{x}$=0 is typically wedge-like optimum or saddle, i.e. either potentially or even actually improvable.

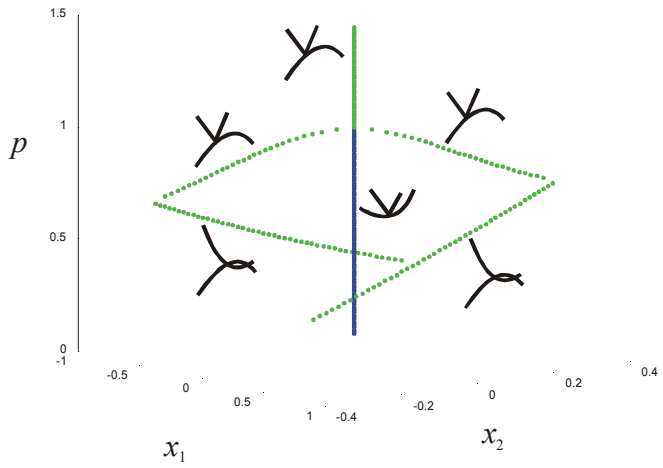


Figure 2.7: Optimum diagram of example 1. Blue and green points denote local minima and saddle points of the $U(x_1,x_2)$ functions. Small pictograms indicate the local shape of U (wedge-like or smooth) at these points using the notations of Figure 2.6.

2.5.3 Example 2

The second example (Figure 2.8/A,B) is shown to illustrate the role of the *regular representation* in structural optimisation problems. We concentrate on the analytical investigation and we will see that the numerical results are trivial in this case.

The bars of the D_3-symmetrical perfect structure have equal cross-sections of unit area, their ultimate normal forces are also assumed to be unit. The structure is optimised for maximal safety in the strength of the bars i.e. the potentials associated with the bars are the quotients of their internal forces and the ultimate force. The cross-sections of five bars are perturbed (so that the areas are $1+x_1,1+x_2,...,1+x_5$ in Figure 2.8/C), while the total mass of the structure is constant (i.e. the cross-section of one bar is $1- x_1- x_2-...- x_5$). The steps of the analysis follow section 2.5.1:

Step I. The elements of the induced representation are collected in Table 2.1. Notice that the symmetry transformations simply permute the six perturbed legs. Accordingly, the elements of the induced representation are almost permutation matrices.

Step II. We follow the technique of creating the irreducible decomposition discussed in I.3.2: the character of the induced representation is χ =[-1 -1 5 -1 -1 -1]. Group D_3 has 3 irreducible representations (Table I.4), their characters are χ_0=[1 1 1 1 1 1], χ_1=[1 1 1 -1 -1 -1], χ_2=[-1 -1 2 0 0 0]. The solution of eq. (I.4) is n_0=0, n_1=1, n_2=2, i.e. there is no trivial (I_0) component and there is one example of I_1 and two examples of I_2 irreducible component in the decomposition of the induced representation. The regular representation of D_3 has 1, 1, and 2 examples of the three components, respectively. The half-irreducible decomposition of D is the same as the irreducible one.

Step III. According to the previous results of the decompositions, condition (i) is satisfied; the induced representation is not half-irreducible, but it is cyclic, i.e. robust optimum is typical although not true for arbitrary $\mathbf{g}\neq 0$.

Step IV. At the given geometry, the weakest points are the six perturbed legs.(Calculations are omitted here, we just mention that the structure is statically determinate, i.e. the internal forces can be determined from equilibrium equations of the bars and hinges.) The bar with area $1+x_1$ is one of the weakest points of the perfect structure.

Step V. The perturbation has no effect on the equilibrium equations, i.e. it does not modify the internal force N in the weakest point. The ultimate force of this bar is proportional to the cross-sectional area, i.e. $N_{u1}(\mathbf{x})=1+x_1$. Consequently the gradient of the corresponding potential $f_1(\mathbf{x})=N/N_{u1}(\mathbf{x})$ is $\mathbf{g}=[N\ 0\ 0\ 0\ 0]^T$. The orbit of $\mathbf{g}$ is $[N\ 0\ 0\ 0\ 0]^T$, $[0\ N\ 0\ 0\ 0]^T$, $[0\ 0\ N\ 0\ 0]^T$, $[0\ 0\ 0\ N\ 0]^T$, $[0\ 0\ 0\ 0\ N]^T$ $[-N\ -N\ -N\ -N\ -N]^T$, which span $\mathbb{R}^5$. Thus we conclude that there is robust optimum at $\mathbf{x}$=0 (cf. *Lemma 2.3*).

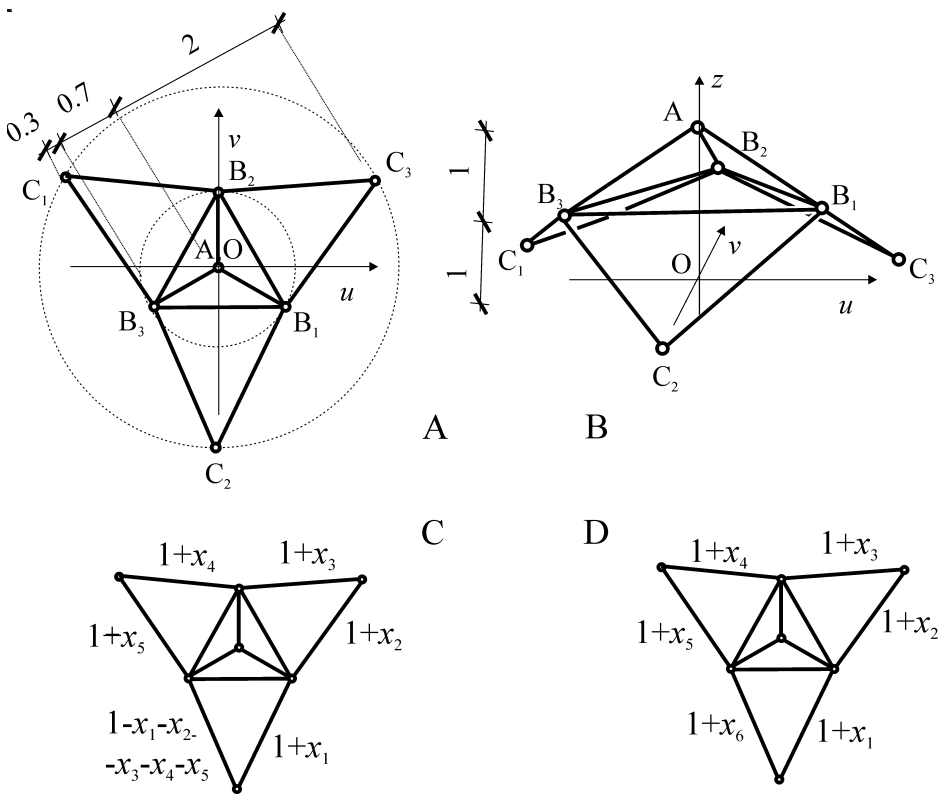


Figure 2.8: Upper (A) and 3D (B) view of Example 2, with two different perturbations (C, D). Each of the perturbations refer to the cross-sectional areas of the C_iB_j bars.

a_1	a_2	a_3
$\begin{bmatrix} & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ -1 & -1 & -1 & -1 & -1 \\ 1 & & & & \end{bmatrix}$	$\begin{bmatrix} & & & & 1 \\ -1 & -1 & -1 & -1 & -1 \\ 1 & & & & \\ & 1 & & & \\ & & 1 & & \end{bmatrix}$	$\begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$

b_1	b_2	b_3
$\begin{bmatrix} & 1 & & & \\ 1 & & & & \\ -1 & -1 & -1 & -1 & -1 \\ & & & & 1 \\ & & & 1 & \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ & & & & 1 \\ & & & 1 & \\ & & 1 & & \\ & 1 & & & \end{bmatrix}$	$\begin{bmatrix} & & & 1 & \\ & & 1 & & \\ & 1 & & & \\ 1 & & & & \\ -1 & -1 & -1 & -1 & -1 \end{bmatrix}$

Table 2.1: The induced representation (second row) corresponding to symmetry transformations (first row) of Example 2/A. The elements of the D_3 symmetry group are the following: a_i is rotation by $2i\pi/n$ around the line OA (see Figure 2.8). b_i is reflection to the plane OAB_i.

The last result is not surprising, since an arbitrary $\mathbf{x}\in\mathbb{R}^5$ perturbation weakens at least one of the six bars due to the total mass constraint, but does not modify the internal forces. Thus, in this case $\mathbf{x}=0$ is a unique and global optimum. Notice that D_3 is of order 6, while the induced representation is cyclic and 5-dimensional. According to *Theorem 2.6*, this is the maximal number of variables, which might yield robust optimum. In this case, the induced

representation is a 'truncated' version of the regular representation of D_3, from which only the trivial component has been removed.

We could analyse the same problem without the total mass constraint (Figure 2.8/D). In this case the areas of the six perturbed legs are $1+x_i$, i=1,2,...,6. Since the symmetry transformations of the perfect structure permute the six legs, the induced representation consists of 6×6 permutation matrices, it is the regular representation of D_3. According to *Theorem 2.3*, this example does not satisfy condition (i), which is again not surprising: if $x_1=x_2=...=x_6$, the D_3 symmetry is preserved. Despite the violation of (i), local optimality of the perfect configuration could be investigated: $\mathbf{x}=0$ is typically not a robust optimum according to *Theorem 2.5*. In our specific case, this is again trivial, no deeper analysis is needed, since if all variables are positive, all legs are strengthened, i.e. the structure is improved.

2.5.4 Example 3

In this part, the structure of Figure 2.9 is optimised with four different perturbations The perfect configuration is D_2-symmetrical. This structure is statically indeterminate of degree one, i.e. its inner forces can be determined from equilibrium conditions and one additional compatibility equation on the deformation of the bars. This time, safety against buckling is optimised and the cross-sections of the bars are assumed to be equal (similar to the example of Figure 2.1/A,C). The potential associated with one of the bars is

$$f(\mathbf{x}) = N(\mathbf{x}) / N_{cr}(\mathbf{x}) = \frac{N(\mathbf{x}) l^2(\mathbf{x})}{EI\pi^2}, \tag{2.18}$$

where N and N_{cr} are the internal force and the critical force in the bar, E, I and l stand for the modulus of elasticity, minimal inertia, and length of the bar, respectively.

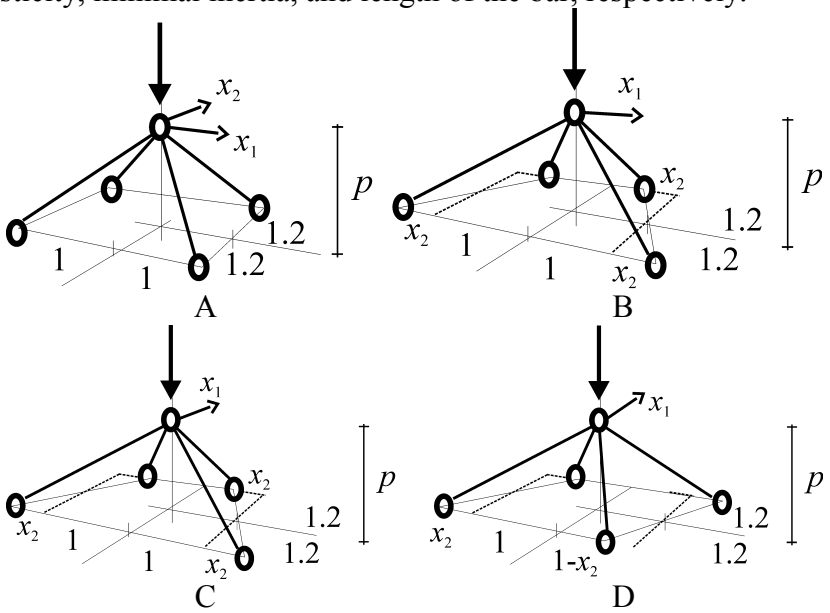


Figure 2.9: Example 3 with four different perturbations.

The analytical results are the following:

Step I.& II: The decompositions of the induced representations (Table 2.2) show that all examples are reducible; A, B, D are cyclic but C is not (it contains too many examples of the I_3 component).

Step III. According to *Theorem 2.5*, $\mathbf{x}=0$ is typically robust optimum in examples A, B, and D but it is not a robust optimum in example C.

Step IV. is trivial (all bars are weakest points).

Step V. has not been performed.

Numerical analysis of the four examples was also executed (Figure 2.10). The results match the expectations at examples 3/A,B,C:

- At C, $\mathbf{S}(0)$ is potentially improvable (it is a 'wedge-like' optimum if $p<0.5$ or $p>3$, or saddle otherwise).
- At A and B, $\mathbf{S}(0)$ is typically not potentially improvable, though there are exceptional points (e.g. the bifurcation point $[p,x_1,x_2]\approx[0.9\ 0\ 0]$ at example B), where the gradient vector $\mathbf{g}$ assumes an atypical value and the robust optimum vanishes.

The numerical results of example 3/D do not match the predictions: as the optimum diagram shows, $\mathbf{x}=0$ is *not a robust optimum*. Instead, it is typically a wedge-like optimum (if $p>1.8$ approximately) or saddle point (if $p<1.8$). The unexpected result indicates a special property of the optimisation potential, which cannot be derived from simple symmetry conditions (similar to the example of Figure 2.3/B). We can simply verify this hypothesis: if $x_1=0$, the structure is invariant under rotation by π around the z-axis. The displacement of the middle-hinge E due to the load is also invariant under this transformation, i.e. it is vertical. Consider one of the bars (Figure 2.11) and let the displacement of its upper end be denoted by Δz ($\Delta z<<1$). The potential associated with this linear, elastic bar is (cf. (2.18))

$$f(x_2)=N(x_2)\frac{l^2(x_2)}{EI\pi^2}=\Delta l\ \frac{EA}{l\ (x_2)}\cdot\frac{l^2(x_2)}{EI\pi^2}\approx\frac{\Delta zp}{l(x_2)}\cdot\frac{EA}{l(x_2)}\cdot\frac{l^2(x_2)}{EI\pi^2}=\frac{\Delta zpA}{I\pi^2}, \tag{2.19}$$

which is independent of x_2. (N, Δl, A, E denote compressive force, shortening, area of cross-section and modulus of elasticity of the bar, respectively.) Thus, $\partial f/\partial x_2=0$ at $\mathbf{x}=0$. The gradient of f is of the form $\mathbf{g}=[*\ 0]^T$. Vectors of this form happen to be an invariant subspace of the transposed induced representation (i.e. $\mathbf{D}_i^T\mathbf{g}=[*\ 0]^T$ for $i=1,2,3,4$), thus the convex hull of the orbit is degenerate and there is no robust optimum at $\mathbf{x}=0$ according to *Lemma 2.3*. However, if the same structure was optimised with a different kind of potential (e.g. safety of compressive strength) we would find robust optimum in accordance with the initial prediction.

irreducible/half-ireducible components of D_2		I_0	I_1	I_2	I_3
number of components in the regular representation of D_2		1	1	1	1
number of components in the induced representation of	Example 2/A	0	0	1	1
	Example 2/B	0	0	1	1
	Example 2/C	0	0	0	2
	Example 2/D	0	1	0	1

Table 2.2: Irreducible/half-irreducible decompositions corresponding to examples 2/A-D. The technique of creating the decompositions is discussed in part I.3.2. Names of the representations are taken from Table I.4. Neither of the induced representations has trivial components (I_0); C is not cyclic (because it has two I_3 components).

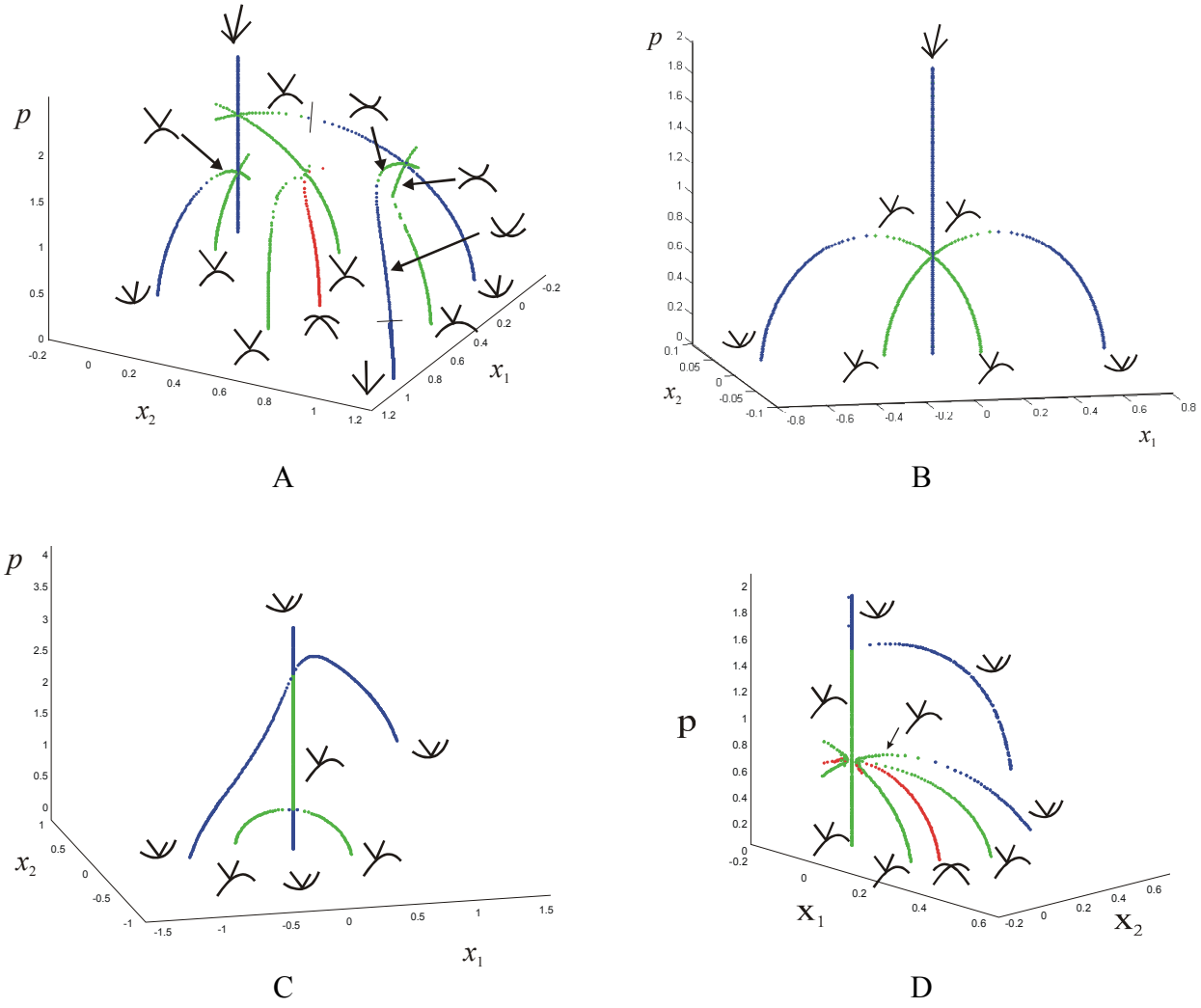


Figure 2.10: Optimum diagrams of examples 2/A-D (see Figure 2.9). Blue, green and red points denote local minima, saddle points and maxima of the $U(x_1,x_2)$ functions. Small pictograms indicate the local shape of U (sharp, wedge-like or smooth) at these points. Notice that sharp critical points are always robust optima and wedge-like points are optima or saddles. For better clarity, only the domain $x_1, x_2 \geq 0$ has been plotted in A and D.

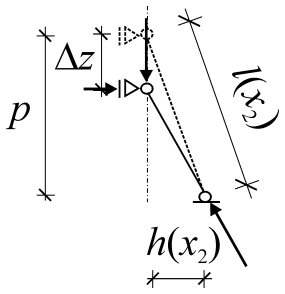


Figure 2.11: Deformation of a bar of example 3/D if x_1=0. As it has been shown, the displacement of the upper end (Δz) is vertical.

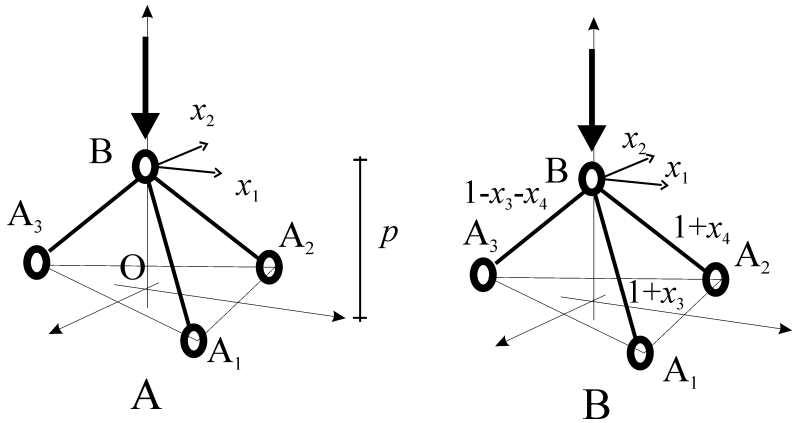


Figure 2.12: Example 4 with two different perturbations. $A_1A_2A_3$ is a regular triangle. There are two variables at A, while at B, the number of perturbations is 4; x_3 and x_4 refer to the cross-sectional areas of the bars.

2.5.5 Example 4

The structure is similar to example 3, but now it has only three legs and $\mathcal{S}(0)$ is D_3-symmetric. The optimisation potential is the same as before, see (2.18). Two variables perturb the structure at example 4/A (similar to example 3/A), while the number of variables is four at 4/B: x_1 and x_2 are displacements of the middle hinge, while x_3 and x_4 perturb the cross-sections of the bars assuming constant total mass and constant shape of the cross-section. Conditions (i) and (ii) are satisfied in both cases; the decompositions of the induced representations are shown in Table 2.3. Example 4/A has a half-irreducible induced representation, while the other one is reducible but cyclic. In both cases, robust optimality of $\mathbf{x}=0$ is predicted.

The structure has only three bars, all of them are weakest points of $\mathcal{S}(0)$. However, notice that the D_3 group is of order 6, i.e. the perfect structure should have 6 weakest points! This contradiction emerges because each of the weak points are invariant under one of the symmetry transformations of the perfect structure (bar A_iB is invariant to reflection to the plane $0BA_i$). The situation is similar to the case of Figure 2.3/A, where the weakest point was invariant under reflection and the resulting potential did not match the general predictions. Analogously, the predictions might prove incorrect at this example: the invariance of a weak point under some of the symmetry-transformations (e.g. γ_j) means that the gradient of corresponding potential satisfies $\mathbf{g}=\mathbf{D}_j^{\mathrm{T}}\mathbf{g}$, which is an unexpected degeneracy and might change the typical results. As the most remarkable difference, the orbit of $\mathbf{g}$ consists of 6 vectors, but only 3 non-identical ones. Better predictions can be obtained if the structure is considered as only C_3- instead of D_3-symmetric (i.e. the symmetry transformations, for which some of the weakest points are invariant are neglected). In this case the number of weakest points (and the corresponding gradient vectors) is predicted correctly by equation (2.14). This means a new decomposition of the induced representation (Table 2.4). The result predicts that 4/B is not cyclic, i.e. $\mathbf{x}=0$ is typically not a robust optimum. On the other hand, 4/A is half-irreducible, the prediction is unchanged in this case.

The numerical results support the improved predictions. Example 4/B has too many variables to determine its optimum diagram with the existing software, but we can determine the diagram of restricted versions of the problem. In fact, 4/A is a restricted version of 4/B (with $x_3=x_4=0$, see Figure 2.13/A) another possibility is the restriction $x_2=0$ and $x_3=x_4$ (Figure 2.13/B). At Example 4/A, $\mathbf{x}=0$ is typically robust optimum, while at the other one, it is not. It follows from the latter result that there is no robust optimum in the original problem 4/B, either.

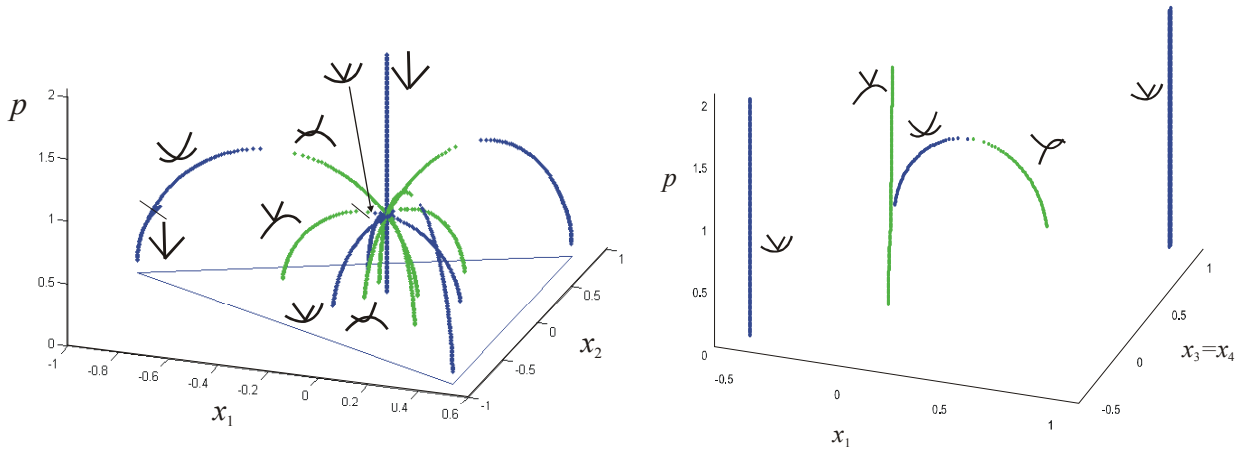
Irreducible/half-irreducible components of D_3	I_0	I_1	I_2
number of components in the regular representation of D_3	1	2	1
number of components in the induced representation of	Example 3/A	0	1	0
	Example 3/B	0	2	0

Table 2.3: irreducible/half-irreducible decompositions corresponding to Examples 3/A-B. (The two kinds of decompositions are identical.) The technique of creating the decompositions is discussed in part I.3.2. Names of the representations are taken from Table I.4. Neither of the induced representations has trivial components; both are cyclic.

irreducible components of C_3		I_0	I_1	I_2
number of components in the regular representation of C_3		1	1	1
number of components in the induced representation of	Example 3/A	0	1	1
	Example 3/B	0	2	2

half-irreducible components of C_3		I_0	S_1
number of components in the regular representation of C_3		1	2
number of components in the induced representation of	Example 3/A	0	1
	Example 3/B	0	2

Table 2.4: irreducible (left panel) and half-irreducible (right panel) decompositions corresponding to Examples 3/A-B, considered as only C_3-symmetric. The technique of creating the decompositions is discussed in part I.3.2. Names of the representations are taken from Table I.4, table 6.



A B

Figure 2.13: Optimum diagrams of example 4/A and a restricted version of 4/B. (Blue, green and red points denote local minima, saddle points and maxima of the $U(x_1,x_2)$ functions. Small pictograms indicate the local shape of U (robust, wedge-like or smooth) at these points.

2.6 BIFURCATION ANALYSIS OF OPTIMUM DIAGRAMS

Until now, individual structures have been investigated and predictions concerning their potential improvability have been derived. In this part, the object of our research is extended to one-parameter families of structures, which allows investigation of *optimal improvability* with small perturbations of the symmetry.

As defined in the introduction (*Definition 2.3*) optimal improvability corresponds to slightly asymmetrical optima bifurcating from the **x**=0 line. Such lines often appear at bifurcations of smooth functions with one variable (e.g. Figure 2.1/B), where the symmetrical optimum switches to pessimum. Similarly, bifurcating asymmetrical optima often emerge at examples, where the optimality of **x**=0 is not typical, i.e. the optimum may change to another type (saddle or pessimum), see Figure 2.7, Figure 2.10/C,D, Figure 2.13/B. On the contrary, if **x**=0

is typically optimal, there are usually no bifurcating optima (Figure 1.1/A, Figure 2.10/C,D), although we have an exceptional diagram (Figure 2.13/A).

It seems rather hopeless to decide in general whether a one-parameter family of structures with a given kind of symmetry and a given set of perturbing variables may or may not yield bifurcating optima. Such a result would need either a general statement about this property or the bifurcation analysis for all finite groups (infinite number!) and each of their representations. Due to the probable difficulties, we confine ourselves to the case of D_1 (reflection) symmetry: in the forthcoming part, we take a systematic approach to the bifurcations associated with potentials of the form (2.3) and (2.6), which we refer to as "reflected potentials" and "multiple reflected potentials", respectively. Section 2.6.2 provides structural engineering examples for each bifurcation, and finally the bifurcation analysis connected to the exceptional diagram Figure 2.13/A is presented in 2.6.3.

Elementary catastrophe theory (Poston et al, 1978) determines the bifurcations of smooth potentials. Bifurcations associated with special *non-smooth* potentials can also be found in Poston et al (1978), Section 16, where a generalisation of Thom's theorem is introduced in case of the so-called *conditional* catastrophes, however symmetrical potentials are not investigated. The forthcoming part is analogous to these ones, although its significance is much more modest. Elementary catastrophe theory has many other applications to engineering problems, e.g. Thompson et al (1973, 1984). Similarly to former works, we are looking for the *Taylor series expansion* of the generating, smooth potentials $f(x)$ at $x=0$. In our case, this provides a classification of bifurcation points for the non-smooth optimisation potential $U(p,x)$, containing both 'classical' cases as well as new ones.

2.6.1 Typical bifurcations in case of D_1 symmetry

Our goal is to give a local classification of one-parameter (p) classes of reflected (part 2.6.1.1) or multiple reflected (2.6.1.2) potentials $U(x,p)$ at $x=0$; this is an analogue to Thom's theorem for smooth functions. The local classification of U is reduced to the local classification of the smooth f generating potentials.

2.6.1.1 Reflected potentials

Thom's theorem shows that the local classification of a smooth function is usually determined by the lowest order non-vanishing term(s) of the function's Taylor expansion. Let $T_f^{(n)}$ denote the truncated Taylor series of the function $f(x,p)$ up to the n^{th}-order term.

At a general point on the p axis ($x=0$) $T_f^{(1)}$ does not vanish typically. At the same time, there exist typically a finite number of isolated points along the $x=0$ line, where $T_f^{(1)}$ vanishes, and there is typically no point where $T_f^{(2)}$ vanishes.

If $T_f^{(1)}$ does not vanish, $f(x,p)$ is, according to Thom's theorem, locally equivalent of the (0,0) point of the $f^{(1)}(x,p)$ function:

$$f^{(1)}(x,p) = x. \tag{2.20}$$

Consequently, the (0,0) point of the reflected $U(x,p)$ function generated from f via (2.3) is locally equivalent to the (0,0) point of $U^{(1)}(x,p)$ generated from the $f^{(1)}$ potential in (2.20).

This type of point is analogous to non-degenerate critical points of smooth functions, but it is non-smooth (Figure 2.14), more specifically it is *robust minimum* (cf. *Definition 2.1*).

If $T_f^{(1)}$ vanishes but $T_f^{(2)}$ does not, $U(0,p)$ is locally equivalent of $U^{(2)}(0,0)$ generated from one of the following two $f^{(2)}$ functions:

$$f^{(2)}(x,p) = p \cdot x \pm x^2 \tag{2.21}$$

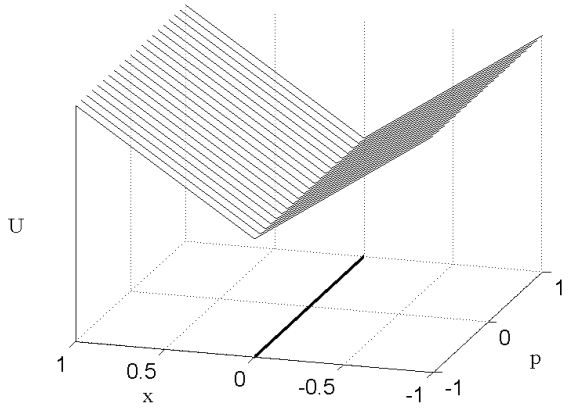


Figure 2.14: Generic point of reflected functions

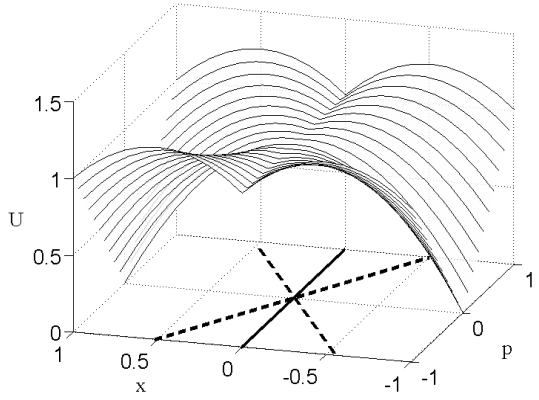


Figure 2.15: Unstable-X catastrophe

The (0,0) point of $U^{(2)}$ is analogous to a fold catastrophe point of a smooth function. It has two dual forms, the corresponding bifurcations are the unstable-X (Figure 2.15) and the point-like "bifurcation" (Figure 2.16). They appear different because the different role of maxima and minima in case of reflected functions. Figure 2.1/B, associated with the three-hinged example, also shows an unstable-X type bifurcation. As already introduced, higher degeneracy of $f(x,p)$ is atypical and there are no more typical catastrophes of one-parameter families of reflected functions.

Notice that neither the X- nor the point-like bifurcation contains bifurcating asymmetrical optima. Thus,

Theorem 2.8: If a one-parameter family of D_1 symmetrical structures is optimised with one variable, there is typically no value of the parameter where the structure is optimally improvable.

At the same time, there are applications, where $f(x,p)$ is, for some reason, odd or even (the even or the odd terms of the Taylor expansion vanish) and some other catastrophes are typical. We have already seen two examples of the latter case (cf. Figure 2.3).

In the case when $f(x,p)$ is odd (i.e. $f(x,p)+c=-(f(-x,p)+c)$), the typical robust optimality of $x=0$ is unchanged, but there exist isolated points, where U is locally equivalent of $U^{(3)}(0,0)$ generated of $f^{(3)}$:

$$f^{(3)}(x,p) = p \cdot x + x^3 \tag{2.22}$$

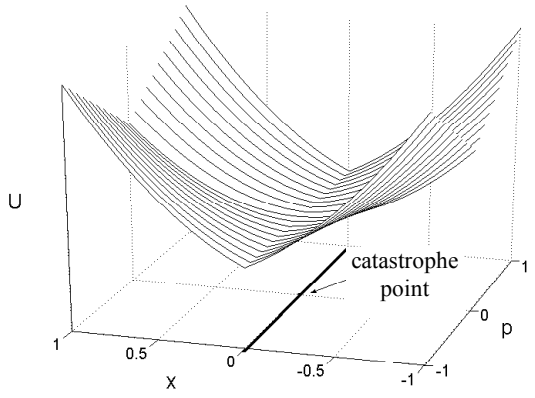


Figure 2.16: Point-like catastrophe

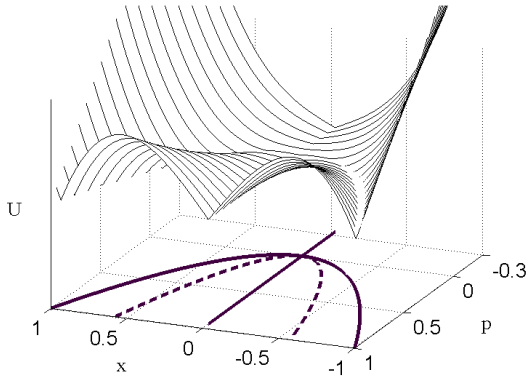


Figure 2.17: Five-branch pitchfork

This corresponds to a 'five-branch pitchfork' (Figure 2.17), which has no dual form. This pattern contains bifurcating optima, i.e. optimal improvability is not atypical in this special case, as opposed to *Theorem 2.8*.

In the case when $f(x,p)$ is even (i.e. $f(x,p) = f(-x,p)$), U is a smooth, symmetric function. The two emerging classes are well-known: the first (typical) one is equivalent of $U^{(4)}(0,0)$ generated of $f^{(4)}$

$$U^{(4)}(x,p) = f^{(4)}(x,p) = \pm x^2 \tag{2.23}$$

which is a one dimensional Morse-saddle, i.e. a smooth, non-degenerate critical point. Beyond this, there are typically isolated points where U is locally equivalent of $U^{(5)}(0,0)$ generated of $f^{(5)}$:

$$U^{(5)}(x,p) = f^{(5)}(x,p) = p \cdot x^2 \pm x^4 \tag{2.24}$$

These are the well-known standard and dual cusp catastrophe points (Figure 2.18), producing the 'stable' and 'unstable' symmetric bifurcation. This is the typical bifurcation occurring in a one-parameter family of *symmetric, smooth functions*. In the following, this bifurcation will be called 'three-branch pitchfork'. Here again, bifurcating optima emerge, but this is not surprising, since x=0 changes from optimum to pessimum at bifurcation points.

2.6.1.2 Multiple reflected potentials

Among this kind of functions (see (2.6)), the typical bifurcations are the same as those of reflected potentials: the bifurcations of $\hat{U}$ are a subset of the bifurcations of the individual U_i functions. A bifurcation of U_k at $(x,p)=(0,p_0)$ appears in $\hat{U}$, if

$$U_k(0,p_0) = \max_i\left(U_i(0,p_0)\right). \tag{2.25}$$

At the same time there are isolated points V at the $x=0$ axis where

$$U_i(0,p) = U_j(0,p), \qquad i \neq j \tag{2.26}$$

At these points, no bifurcation emerges if U_i and U_j are not exceptional, i.e. both have robust minima at V. However, a special 'wedge-bifurcation' may appear (there is an example in Figure 2.19), if U_i has a local minimum, and U_j has a local maximum at V. The latter can occur if U_j is generated from even f function (i.e. it is of type $U^{(4)}$ or $U^{(5)}$)

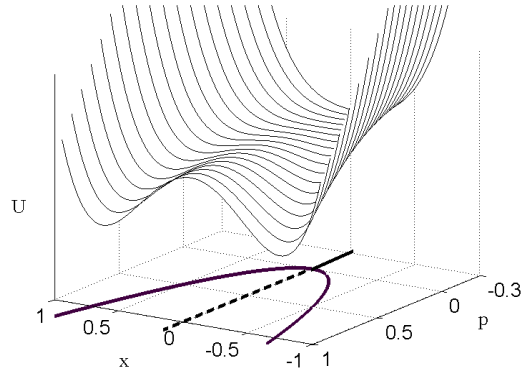


Figure 2.18: Standard cusp catastrophe, or stable three-branch pitchfork

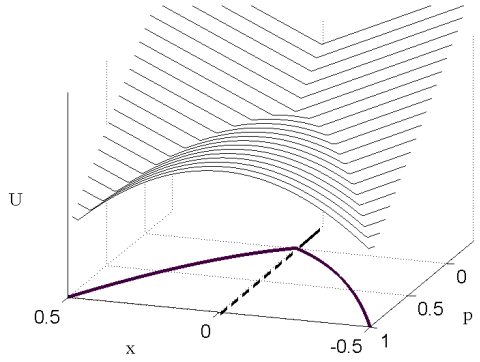


Figure 2.19: An example of the wedge-bifurcation

2.6.2 Examples in engineering: optimisation of structures

In this part we provide a list of examples (Figure 2.20), illustrating all the bifurcations described in part 2.6.1. Our goal was to make this illustration homogeneous and easy to follow in the sense that each bifurcation type is demonstrated on the *same type* of structure (continuous beam with four supports). As a result, some illustrations are somewhat artificial. Including a larger variety of structures yields other illustrations, however, their description would be more lengthy. Similar examples have been studied in Buella (2002), Alkér (2001).

Structures similar to our beams are usually designed based on *strength conditions* of the form $f \leq f_u$, where f and f_u are, respectively, the design and ultimate value of the bending moment. Such conditions have to be met by all parts or points of the structure. If the design variable x is optimised for this kind of condition, it is plausible to define an 'optimisation potential' $\hat{U}(x,p)$ as the maximum of $f(x,p)$ for all points of the structure, a 'better' structure corresponding to smaller values of $\hat{U}(x,p)$. As already shown previously, if

- a one-parameter (p) family of structures is examined,
- $f(x,p)$ is a smooth function at all points or parts of the structure,
- x satisfies condition (i)-(ii), which means in our case that $\gamma_1(\mathbf{S}(\mathbf{x}))= \mathbf{S}(-\mathbf{x})$ (with $\gamma_1 \equiv$reflection)

$\hat{U}(x,p)$ is typically a multiple reflected potential, thus the examples are likely to produce the bifurcations in question.

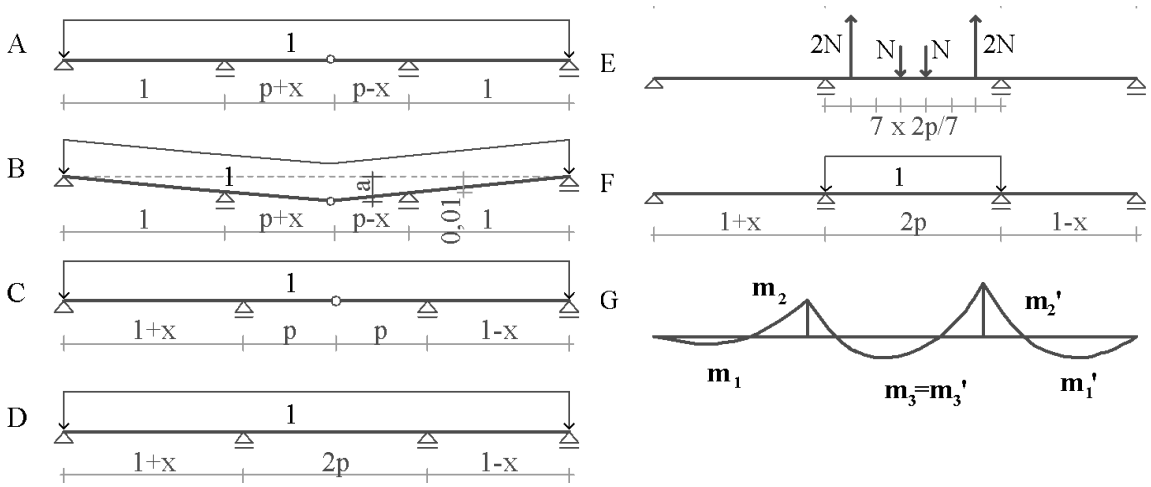


Figure 2.20 A-F: various parameterised beams and loads; G: qualitative moment diagram of the structures A-D

2.6.2.1 Unstable X bifurcation.

Consider the uniform, linearly elastic beam in Figure 2.20/A with four supports, subjected to uniform vertical load. Our goal is to optimise the position x of the hinge, making the maximum $\hat{U}$ of the bending moment as small as possible.

Calculating the support and hinge reactions under the assumption of small deformations (linear theory) is a common structural engineering problem. Solution techniques are available in advanced undergraduate textbooks; most easily it can be solved by the force method (cf. Gere et al., 1990), yielding finally the internal bending moment acting at an arbitrary point of the beam. The qualitative moment-diagram is illustrated in Figure 2.20/G. There are three pairs of local maxima in the moment diagram denoted by f_i and $f_i^{'}$; i=1,2,3. So $\hat{U}(x,p)$ is now the maximum of three pairs of local maxima:

$$\hat{U} = \max(U_1, U_2, U_3), \tag{2.27}$$

where the U_i's are reflected functions as defined in (2.3):

$$U_i = \max(f_i, f_i^{'}), \qquad i=1,2,3 \tag{2.28}$$

In our example the f_i functions can be determined analytically (cf. Gere et al, 1990) as:

$$f_2(x,p) = \frac{(p+x)\cdot\left[x+4p\left(xp^2+x^3+(p-x)^3+(p-x)^2\right)\right]}{8\cdot\left(p^3+3px^2+p^2+x^2\right)} \tag{2.29}$$

$$f_1(x,p) = \begin{cases} \tfrac{1}{2}\left(\tfrac{1}{2}-f_2(x,p)\right)^2 & \text{if} \quad f_2(x,p) \le \tfrac{1}{2} \\ 0 & \text{if} \quad f_2(x,p) > \tfrac{1}{2} \end{cases} \tag{2.30}$$

$$f_3(x,p) = \tfrac{1}{2}\left[x+\frac{f_2(-x,p)-f_2(x,p)}{2p}\right]^2 \tag{2.31}$$

We computed the bifurcation points of the U_i functions as the solutions of the $df_i(x,p)/dx=0$ equation analytically and found that there is an unstable-X bifurcation in $U_2(x,p)$ at point

$$P = \left(0; \frac{1}{6}\left(19+3\sqrt{33}\right)^{\frac{1}{3}} + \frac{2}{3}\left(19+3\sqrt{33}\right)^{-\frac{1}{3}} - \frac{1}{3}\right) \approx (0, 0.420) \tag{2.32}$$

Since $U_2>U_1$ at P, this X-bifurcation of U_2 occurs in the $\hat{U}$ function as well (cf. (2.25) and the corresponding remarks). A representative domain of the bifurcation diagram is plotted in Figure 2.21. At point $V=(0,\sqrt{2}-1)$ on axis p, we have $U_1=U_2$, so we could expect a wedge-bifurcation based on equation (2.26). However, since both U_1 and U_2 have local minima at V, no bifurcation occurs (cf. the comments after equation (2.26)).

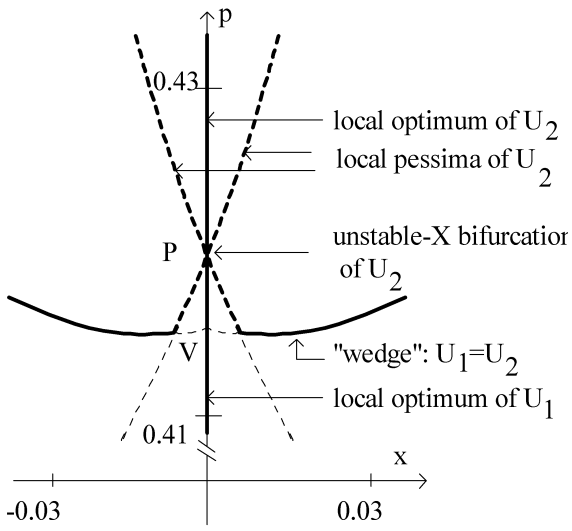


Figure 2.21: An example of the X-bifurcation

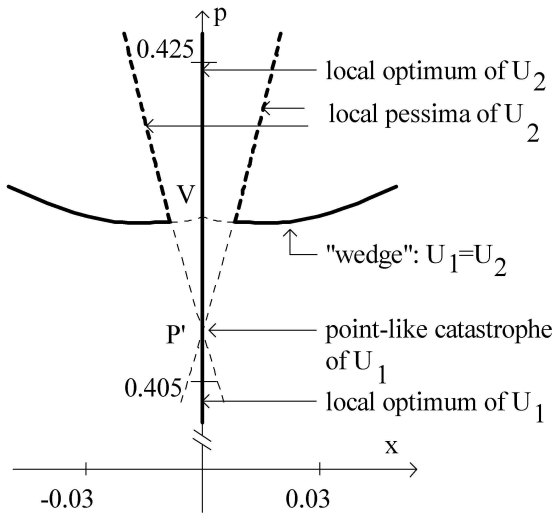


Figure 2.22: An example of the point-like bifurcation

2.6.2.2 Point-like bifurcation

In the previous example, equation (2.30) provides a simple relationship between f_1 and f_2, which shows that the critical points of U_1 and U_2 typically coincide. Furthermore, the following form of (2.30) (where f_2^2 is approximated by its truncated Taylor expansion)

$$f_1(x,p)=\frac{¼-f_2(x,p)+f_2^2(x,p)}{2}\approx\frac{¼+f_2^2(0,p)-f_2(x,p)\cdot\left(1-2\cdot f_2\,(0,p)\right)}{2} \qquad (2.33)$$

shows that an unstable-X bifurcation point of U_2, corresponds to a dual, point-like bifurcation point of U_1 if $f_2(0,p)<1/2$ (which is true for $p<1$). So U_1 has a point-like bifurcation at $P\approx(0,0.420)$ (cf. Figure 2.21), however, it is hidden, because $\hat{U}\neq U_1$ at point P. In order to make the point-like bifurcation at P appear in $\hat{U}$, we change the geometry of the structure slightly.

The new geometry is illustrated in Figure 2.20/B: the two inner supports are both symmetrically moved down by the distance a (this could be the result of soil settlement). This modification causes, according to our computations, the following effects:

- moves the critical point P downward in the bifurcation diagram ,
- does not change the position of point V because the moment diagrams are unchanged if $x=0$.

If, e.g. a=0.003, the new bifurcation point P' is under V, and the point-like bifurcation of U_1 appears in $\hat{U}$ (Figure 2.22).

2.6.2.3 Five-branch pitchfork bifurcation

The example of Figure 2.20/C is similar to the previous ones, but the position of the middle supports is optimised instead of the position of the hinge. Analysis is done in the same way as at the first example.

The bending moment f_{mid} at the middle of the structure is zero at arbitrary (x,p) values, since there is a hinge. On the other hand f_{mid} can be expressed from f_2 and f_2' as:

$$f_{\text{mid}} = (f_2 + f_2')/2 - p^2/2 \tag{2.34}$$

Combining (2.34) with $f_{mid}=0$ yields

$$p^2 = f_2(x,p) + f_2'(x,p) = f_2(x,p) + f_2(-x,p) \tag{2.35}$$

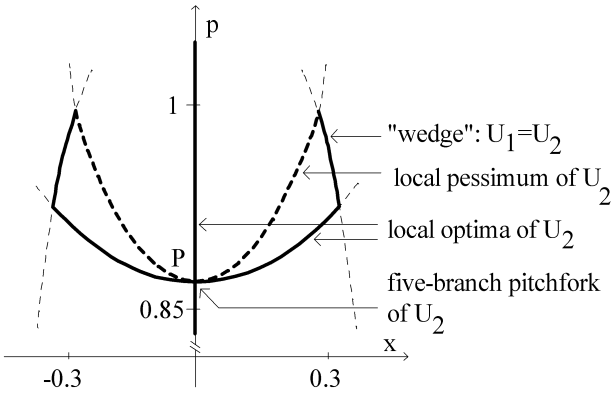


Figure 2.23: An example of the five-branch pitchfork.

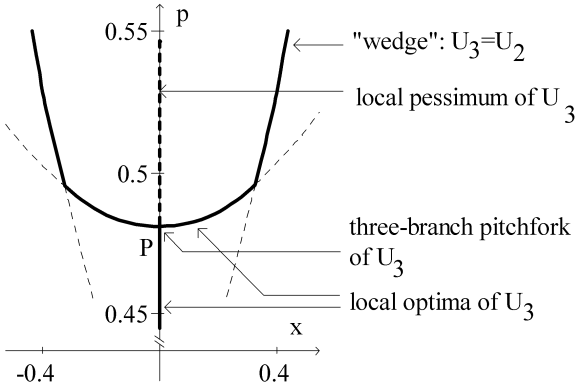


Figure 2.24: An example of the 'stable' three-branch pitchfork.

According to (2.35), f_2 is an odd function of x (the non-vanishing constant term does not influence the critical points), so at the bifurcation point $P=\left(0,\sqrt{3}/2\right)$ it is locally equivalent to $f^{(3)}$ (defined in (2.22)), thus the bifurcation of U_2 is a five-branch pitchfork. In the neighbourhood of P, $U_2>U_1$, so, based on (2.25), this bifurcation occurs in $\hat{U}$ as well (cf. Figure 2.23.)

2.6.2.4 Three-branch pitchfork bifurcation

The beam of Figure 2.20/D is again slightly different from the previous ones: the hinge is missing. This structure is statically indeterminate of the second degree, so two compatibility equations are needed beyond the equilibrium equations. The solution is constructed in the same way as at the other examples.

As f_3 occurs at the symmetry axis of the structure, we have:

$$f_3(x,p) = f_3(-x,p) \tag{2.36}$$

and

$$f_3(x,p) = f_3'(x,p) = U_3(x,p) \tag{2.37}$$

Since U_3 is always a smooth, symmetric function of x, the typical bifurcation of U_3 is the (stable or unstable) three-branch pitchfork. In our example, U_3 has a stable pitchfork at point $P\approx(0,0.4805)$. (The second co-ordinate of P has been computed numerically as a root of $f_3''=0$, leading to a sixth-order polynomial equation.) Since U_3 is not the global maximum

of the bending moment at P, the structure has to be modified in order to have the pitchfork in $\hat{U}$ as well. One example of such a modification is adding the loads of Figure 2.20/E to the structure. This load has the following properties:

- it leaves the moment diagram in the outer spans invariant and only changes the moment diagram in the middle span: it increases U_3 and does not influence U_2 and U_1. If N is chosen appropriately, U_3 becomes global maximum.
- the effect of the load is independent of x, so the character of the bifurcation remains unchanged.

Figure 2.24 shows the bifurcation diagram for N=1. At this value of N we can observe the 'stable' symmetric bifurcation in $\hat{U}$.

2.6.2.5 Wedge bifurcation

Let us regard Figure 2.20/F. The structure is the same as the one in Figure 2.20/D, however, the load on the outer spans is now zero. At point V=(0,1) we have U_2=U_3. At this point, U_2 has a local minimum and U_3 (which is a $U^{(4)}$ type potential, cf. (2.23)) has a local maximum. The two functions form a wedge-bifurcation, which appears in $\hat{U}$. The corresponding bifurcation diagram is illustrated in Figure 2.25.

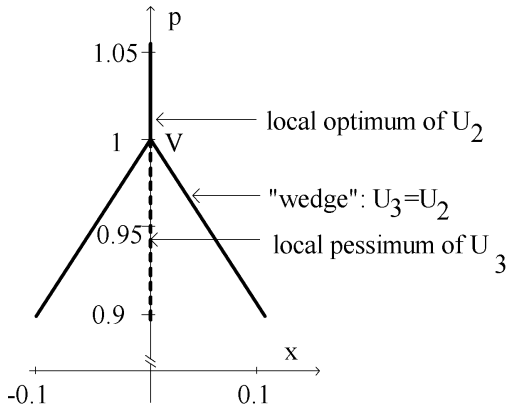


Figure 2.25: An example of the wedge-bifurcation.

2.6.3 Bifurcation analysis at different symmetries

The analysis of part 2.6.1 showed that there are typically no bifurcating, asymmetrical optima in a one-parameter family of D_1-symmetrical examples, at which x=0 is robust optimum (cf. *Theorem 2.8*). The only exception was a special case where the generating function f had the odd property as an example-specific degeneracy.

This fact suggests that bifurcating optima might be atypical in one-parameter families of structures with arbitrary symmetry, in which $\mathbf{x}$=0 is robust optimum. (In other words: the lack of potential improvability implies the lack of optimal improvability.) However, we have found a D_3-symmetrical counter-example (example 4/A, Figure 2.13/A), which indicates either that the counter-example has a specific degeneracy or that the above generalisation fails.

To decide this question, we performed a partial bifurcation analysis based only on the symmetry and the induced representation of example 4/A, regardless to the inner forces of the

specific structure to find some generic bifurcation patterns. The technique of the examinations is analogous to that of part 2.6.1.

We consider a one-parameter family of potential functions $U(x_1,x_2,p)$. According to eq. (2.14), U is generated from a smooth function $f(x_1,x_2,p)$ (i.e. the potential of bar A_1B) via

$$U(\mathbf{x}, p) = \max_{i=0}^{2} \left[f\left(\mathbf{D}_i \mathbf{x}, p\right) \right], \tag{2.38}$$

where $\mathbf{x}=[x_1\ x_2]^T$ and the matrices $\mathbf{D}_i$ are elements of the induced representation of example 4/A., i.e.

$$\mathbf{D}_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \quad \mathbf{D}_1 = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad \mathbf{D}_2 = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \tag{2.39}$$

The local classification of U is derived from the classification of $f(\mathbf{x},p)$. The truncated Taylor expansion of $f(x_1,x_2)$ up to the second-order term is the following:

$$T_f^{(2)} = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{20}x_1^2 + a_{11}x_1x_2 + a_{02}x_2^2 \tag{2.40}$$

As already shown in part 2.5.5, bar 1 is invariant under reflection to the OBA_1 plane (see Figure 2.13/A). Due to this symmetry, the potential of the bar satisfies $f(x_1,x_2)= f(x_1,-x_2)$, i.e. $a_{01}=a_{11}=0$ in (2.40). Beyond that, we can assume $a_{00}=0$ because the constant term is indifferent from the point of view of bifurcation analysis.

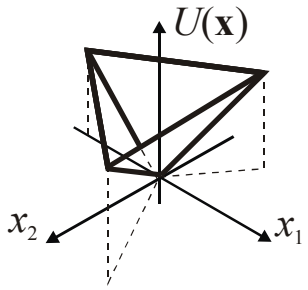


Figure 2.26: typical local configuration of the $U(\mathbf{x})$ function at x=0

At *typical points*, $T_f^{(1)}$ does not vanish, *f(x,p)* is locally equivalent of the (0,0,0) point of $f^{(1)}(x_1,x_2,p)=x_1$. The U function, generated from $f^{(1)}(\mathbf{x},p)$ via (2.38) has a robust optimum at $\mathbf{x}=0$ (Figure 2.26). At generic *bifurcation points*, $T_f^{(1)} = a_{10}x_1$ vanishes and *f(x,p)* is locally equivalent to the (0,0,0) point of

$$f^{(2)}\left(x_1, x_2, p\right) = px_1 + a_{20}x_1^2 + a_{02}x_2^2 \tag{2.41}$$

The type of the emerging bifurcation depends on a_{20} and a_{02}. The local shape of $U(x_1,x_2,p)$ cannot be plotted but in a four dimensional diagram, however the bifurcation diagrams are 3 dimensional, i.e. we can plot the latter ones. There are numerous typical patterns, some of them are presented in Figure 2.27. As the results suggest, there is a range of the parameters, where the *bifurcation pattern contains asymmetrical optima*, i.e. such optima are not atypical, in contrast with the case of D_1-symmetry.

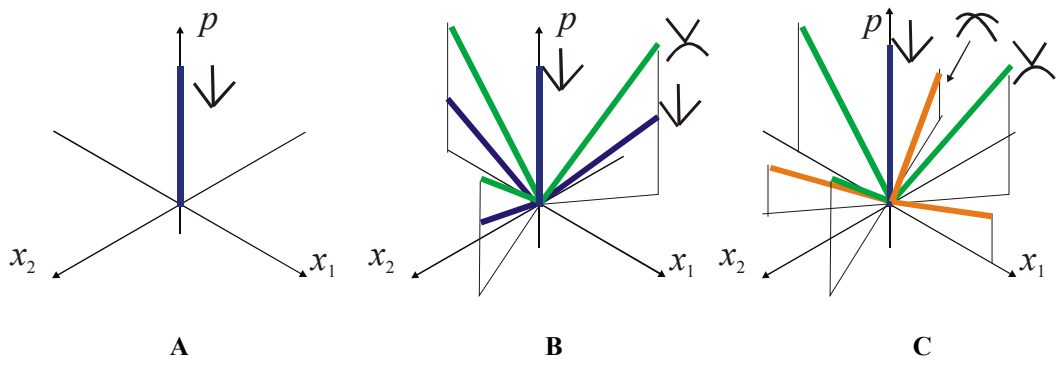


Figure 2.27: Numerically determined bifurcation patterns at some values of the coefficients a_{20} and a_{02}. A: both coefficients are positive B: $0<a_{20}\approx -a_{02}$ C: $a_{02}\approx 2a_{20}<0$. In all diagrams, red, green and blue points denote pessima, saddle points and optima, respectively. The figures show only half ($p>0$) of the bifurcations. Notice that B is of the same type as the bifurcation in Figure 2.13/A.

2.7 EXCEPTIONAL STRUCTURES

The results of the previous sections predict *typical* properties of optimisation problems. The exact behaviour of an example depends on the inner forces of the specific structure, e.g. the typical predictions on robust optimality of $\mathbf{x}=0$ in Section 2.4.3 were in some cases modified by the exact value of the gradient $\mathbf{g}$. At some of the numerical examples, the lack of checking the value of $\mathbf{g}$ yielded misleading results. Such difficulties have already emerged at the D_1-examples of Section 2.3. The exceptional cases can be classified into two distinct classes:

1: The weakest points of $\mathcal{S}(0)$ are invariant to some (Example 4) or all (example of Figure 2.3/A) elements of Γ. As shown in part 2.5.5, the typical predictions can be applied for these examples, provided that the structure is considered as only Γ^* symmetric, where $\Gamma^* \equiv \Gamma \setminus \{i_i\}$ and i_i denotes elements of Γ, to which some of the weakest points are invariant. If $\Gamma \cong \{i_i\}$, then Γ^* is the trivial group, i.e. the structure has typically no robust optimum at $x=0$ even if optimised by only one variable. To recognise these exceptions, one has to find (one of) the weakest points of the perfect structure.

2: The potential associated with the weakest point has a special property, which does not follow from the symmetry of the structure (such examples are that of Figure 2.3/B, where $df/dx \equiv 0$ was a surprising identity or example 2/D, where we had $\partial f/\partial x_2 \equiv 0$). The general predictions often fail in these cases. Such exceptions are difficult to recognise: one has to determine the exact gradient of the potential function associated with one of the weakest points.

As most outstanding property of exceptional examples, the local optimality of $\mathbf{x}=0$ differed from the predictions. At the same time, there are examples, where the optimality of $\mathbf{x}=0$ is unchanged, however atypical bifurcations emerge in the optimum diagrams. We mention the beam of Figure 2.20/C, which is a type 2 exceptional example: the potential of the weakest point satisfied $d^2f_2/dx^2 \equiv 0$. This identity did not change the robust optimality of $x=0$, but called forth a novel bifurcation pattern (the five-branch pitchfork). Thus, in this case the general prediction concerning *potential improvability* of the structure was correct, but the *optimal improvability* of the structural family changed.

2.8 SUMMARY

In this section the improvability of symmetrical structures via small perturbations of their symmetry has been examined in cases where the quality of the structures was determined by the worst one of several smooth 'local' potentials associated with a set of weak points.

It has been shown that the symmetric configuration of such examples is often a 'robust' local optimum. Exact (based on the structure's inner forces) and typical (based on the structure's symmetry and the choice of variables) conditions of robust optimality have been derived. The latter ones are considered as the main results of the research. Since robust optima cannot be improved by small perturbations, these results are applied to formulate conditions of the 'potential improvability' of symmetrical structures.

We have showed necessary as well as sufficient conditions for the *number of symmetry breaking variables*, which make a symmetrical structure potentially improvable, these results were based on the type of symmetry of the structure. We have also determined if *a given set of variables* typically yielded local, potential improvability or not. Both conditions are *much easier to handle* than an explicit verification of optimality at specific examples. The latter one means practically the calculation of the exact potential of a wide family of structures (the members of which are determined by arbitrary $\mathbf{x} \in \mathbf{R}^d$).

These results help to *improve symmetrical structures with small perturbations of the symmetry*. Using the typical condition, one can determine an adequate (and small) set of variables. After that, numerical analysis of the structures serves to decide if the given variables yield actual improvability and which combination of the variables should be applied to get improvement. Without preliminary verification of the variables, one would either have to choose many variables or one would risk choosing 'hopeless' variables. The former one is disadvantageous, because numerical computational time grows exponentially with the increased number of variables.

'Optimal improvability' of a symmetrical structure was also examined. We demonstrated that slightly asymmetrical local optima are extremely rare among reflection-symmetrical structures: there are typically no such optima in a one-parameter family of examples if perturbed by one variable. (We did not show, although one can easily verify that there are typically finite number of such examples in a two parameter family.) This result puts further light on the widely-known observation that structures with imperfect reflection-symmetry are very rare. We also demonstrated that structures with higher symmetry groups may behave rather differently in this respect. Bifurcation analysis concerning another specific symmetry group as well as a numerical example showed that slightly asymmetrical optima may emerge in a typical manner in a one-parameter family of structures.

All the shown results are 'typical' but exceptions are not excluded. Accordingly, we demonstrated the existence of exceptional examples in connection with all results. Most of the general statements were based on the first- or second-order terms of the Taylor-expansions of potentials associated with weak points of a structure. In cases where some of these terms had special values (e.g. they vanished), but their speciality did not follow from basic symmetry of

the structures/variables, the ‘typical’ predictions failed. However, such exceptional cases are not too frequent, i.e. our results are applicable in most cases.

We also showed several types of structures with detailed numerical analysis as illustration. We believe that the perspective offered by bifurcation and representation theory may be helpful in the understanding of optimisation problems related to symmetrical engineering structures.

CHAPTER 3 EVOLUTION

3.1. INTRODUCTION TO EVOLUTION

3.1.1 Problem statement

Symmetry and asymmetry are central concepts in understanding both phylogeny and ontogeny of animals (Moore, 2001). Except for sponges, all animal taxa can be characterised either by 'bilateral' or by 'radial' symmetry of their basic body plan as already demonstrated in Chapter 1. This distinction is based on having one or several planes of reflection symmetry passing through the oral-aboral axis of the animal. The actual body structure is often less symmetric than the basic body plan, due to secondary loss of symmetry. In particular, the left-right symmetry of the bilateral animals is rarely perfect. Different kinds of asymmetries emerge in different time scales of evolution. On the one hand, asymmetric locations of some organs, as the heart or the liver, are as old as the *Vertebrates* themselves. On the other hand, functional asymmetry of the human brain is probably very recent. While Chapter 2 was devoted to the role of structures with imperfect symmetry in optimisation problems (which was motivated by the lack of such solutions in the engineering praxis), now we are interested in understanding the bifurcation structure of evolutionary transitions from perfect to imperfect symmetry (which seem to occur often and to be the result of adaptation). For the sake of simplicity, we will replace the world 'imperfect symmetry' by simply 'asymmetry' in this chapter.

Evolution is inherently related to optimisation. 'Fitness functions' of the first (which will be defined later) can be regarded as the analogues of the potential functions of the second. Some models (called 'frequency independent' in the biological literature) show complete analogy to engineering optimisation, however the mechanism of evolution is in general more than just optimisation: a pre-defined global fitness function would predict a single winner of selection; optimisation itself is unable to explain the origin of *biological diversity*. To account for the coexistence of parallel branches of the evolutionary tree, one should take into account 'frequency dependence', i.e. the fact that the fitness function depends on the relative sizes of competing populations. In case of frequency dependence, evolution itself modifies the fitness function all the way. Consequently, one cannot rely on a global optimality criterion for predicting the outcome of evolution. According to the theory of adaptive dynamics or AD (Dieckmann et al., 1996; Metz et al., 1996; Geritz et al. 1997, 1998; Meszéna et al., 2005), directional evolution via small mutational steps still proceeds in the direction of the *current* fitness gradient. However, the 'uphill' evolution on the 'fitness landscape' is no longer guaranteed to end up at a local optimum, a local pessimum can be equally reached (Eshel, 1983; Taylor, 1989; Christiansen, 1991; Abrams, 1993). In the latter case, the theory predicts branching in the evolutionary process (Geritz et al. 1997, 1998).

Due to the possibility of frequency-dependence in biological systems, the mathematical background of evolution is more general than that of engineering optimisation. At the same time, this chapter will be more specific from another point of view: while engineering

structures with arbitrary finite symmetry have been objects of the investigations of Chapter 2, the evolutionary research is confined to the case of bilateral (i.e. D_1-) symmetry, since the real species, which motivated this study (see e.g. 3.8) are all examples of imperfect *bilateral* symmetry.

The primary goal of AD theory is to demonstrate the possibility of evolutionary branching, which can be initiated in two different ways (Metz et al., 1996; Geritz et al, 2004):

I. In a constant environment evolution converges to a branching point and branches there immediately.
II. The population evolves to an evolutionary stable strategy and waits there until an environmental change bifurcates this strategy to a branching point. Evolutionary branching occurs as a response to the modified conditions.

Most AD models concentrate on Scenario I, however it is a general perception (cf. punctuated equilibrium, Eldredge et al., 1972) that the bulk of evolutionary change is restricted to short transitional periods, i.e. most of the time evolution stops, and is waiting for an environmental change which will trigger the new phase of rapid evolution. This implies that the seemingly more complicated Scenario II. is more relevant for the real process (cf. Geritz et al, 2004).

Our present aim is to apply the adaptive dynamics framework to modelling the emergence of asymmetry instead of evolutionary branching. As we will see, asymmetry can emerge via, but also without branching. We want to get a deeper insight to the ecological types and background of symmetry as well as to decide whether environmental change is an important ingredient of this phenomenon or not (cf. the comments on evolutionary branching above). The final goal is to determine the generic evolutionary patterns of the emergence of asymmetry.

In principle, there are three possible scenarios for the evolutionary loss of symmetry (Figure 3.1). In the simplest case (a), an initially symmetrical population evolves to be asymmetric. This scenario does not contain branching and it can be fully described within the confines of the optimisation picture of evolution.

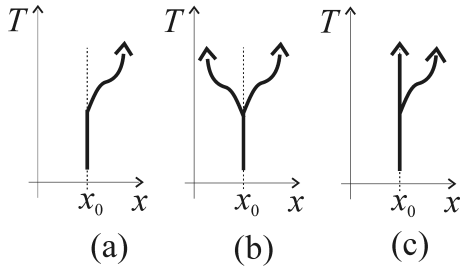


Figure 3.1. Three fundamental patterns for the emergence of asymmetry (referred to as type (a), (b) and (c)). In each case, *T* means time (in evolutionary time-scale) and *x* is a scalar variable corresponding to an evolving phenotypic value of individuals. $x=x_0$ corresponds to perfect bilateral symmetry, while $x\neq0$ means an asymmetrical phenotype.

In the second scenario (b), *two* asymmetric populations (which are symmetric mirror images of each other), emerge. If we assume a fixed potential (fitness function) then the slightest violation of the reflection symmetry between the two asymmetric populations would result in a temporary advantage of one of the populations and competitive loss of the other one. This scenario becomes robust only by assuming a frequency dependent fitness function.

The third scenario (c) assumes a different kind of evolutionary branching. A new, asymmetric form speciates away from he original, symmetric one, however, the symmetric one survives, as well. This scenario is inconceivable under the assumption of a fixed potential. On the one hand, the asymmetric form cannot appear while the symmetric form is optimal, while on the other hand, the symmetric form cannot survive when it becomes a local pessimum. Nevertheless, the scenario makes sense from the biological point of view: a new species acquires a new way of life and does not disturb its ancestor. Hence, frequency dependence is a necessary, however not sufficient ingredient of such situations. Any typical branching pattern is locally symmetric according to the conventional AD theory. Branching into an unchanged and an evolving branch is beyond the confines of the existing approach. As we will see, this type of evolution is possible due to the higher-order terms of fitness functions, which are usually neglected, but become relevant here, as a consequence of symmetry conditions. Thus, this work will conclude that all the three scenarios are likely to emerge in Nature.

3.1.1. Principal results and the structure of Chapter 3

Below I give the list of the principal results of this chapter. Standard concepts of Adaptive Dynamics are used in the formulations. I remark that a short overview on Adaptive Dynamics can be found in part 3.2 to enhance readability of the main text.

I started with the identification of the types of symmetry-breaking, which are likely to produce significantly different behaviour. The role of frequency dependence in evolution and, in particular, in adaptive dynamics is widely known. Considering the evolution of symmetry, I found two sub-classes (called strong vs. weak symmetry) in the frequency-dependent case, which are, according to my knowledge, not present in the current literature. The corresponding research is summarised in the following principal result:

IV: **I introduced a novel classification of symmetry in frequency-dependent ecological models, which I called strong/weak symmetry** (cf. Section 3.4). I determined the corresponding symmetry constraints in the fitness functions (‘strong’ symmetry yielded a more specific constraint than ‘weak’ symmetry), and showed that the two cases produce different evolutionary behaviour (see further details in Principal Result V). I also demonstrated the difference between the two classes on several real examples (Section 3.8)

After separating the qualitatively different cases (frequency-dependent vs. independent selection, strong vs. weak symmetry, changing environment vs. constant environment) I performed a systematic description of the patterns of the emergence of asymmetry in each case. My approach focused on the truncated Taylor expansions of the fitness functions in the light of the emerging symmetry-constraints.

V: **Studying the evolutionary patterns of the emergence of asymmetry**,

- V.1 **I listed the generic patterns in adaptive dynamics models** (all scenarios of Figure 3.1 occurred in some of the cases, see Section 3.5), I also determined the exact conditions of the emergence of each one.
- V.2 **I demonstrated the possibility of a novel evolutionary bifurcation pattern**, in which an asymmetrical evolutionary branch develops in a population with bilateral symmetry and the new branch coexists with the symmetrical ancestors (cf. Section 3.6). I also simulated this pattern numerically on a classical model of Levene (Section 3.7)

V.3 **I showed that the novel pattern occurs only in case of changing environment** Hence, environmental change is even more important ingredient than in the classical AD theory, where evolutionary branching can be demonstrated on autonomous models.

Principal Results IV and V have been published in Várkonyi et al (accepted for publication).

The necessary elements of AD for constant environment are summarised in Section 3.2, whereas Section 3.3 specifies the problem and the basic assumptions more precisely. Section 3.4 introduces a distinction between 'weak' and 'strong' symmetry. Section 3.5 analyses the types of the emergence of asymmetry. Section 3.6 summarises the patterns of evolutionary branching, Section 3.7 provides a model example. In the last unit, a few real examples are reviewed.

3.2. ADAPTIVE DYNAMICS IN CONSTANT ENVIRONMENT

Here we summarise the essentials of AD theory in constant environment, following Geritz et al. (1997, 1998).

We consider evolution of a continuous inherited trait x, referred to as phenotype or strategy. (Later we will identify this trait as the symmetry breaking parameter.) We assume that the investigated population is large, well-mixed, and it may consist of several sub-populations with different strategies $x_1,x_2,...,x_L$. It is assumed that an underlying model specifies the joint dynamics of these strategies. We further assume that this dynamics reaches a unique, global, and 'simple' (i.e. fix point, periodic or quasi-periodic but not chaotic) attractor on the fast time scale, except in degenerate cases (such as the coexistence of identical strategies).

From time to time, the dynamical system is perturbed by the emergence of a new, random 'mutant' strategy y with a small initial number of individuals. The mutant strategy y is always similar to an already existing one, which is considered as the ancestor of the mutant. The mutants appear on a slower time scale, i.e. when the already existing strategies have already reached the global fixed point.

The goal of AD is to understand the generic properties of the emerging evolutionary process, independently from the specific dynamical system governing the fast time scale changes of the populations.

It is an ongoing debate in evolutionary biology whether AD is a proper description of the evolutionary process. (See, for instance, the target review by Waxmann et al. (2005) and the related commentaries.) This debate is about the relative importance of *ecological* and *genetic* factors in evolution (cf. Schluter, 2001). Adaptive Dynamics concentrates on the former aspect and strongly simplifies the latter one (through the assumptions of clonal reproduction and small mutational steps in x). Using this theory enables us to find similarities between structural optimisation and evolution, since the ecological process of adaptation carries a close analogy to engineering optimisation.

As already mentioned, AD considers any evolutionary phenomenon in an asexual model, however AD-based models with complete sexual genetics (e.g. Dieckmann et al., 1999) seem

to support the possibility that speciation of sexual organisms is based on the phenomenon of AD-style evolutionary branching (Metz et al., 1996; Geritz et al., 2004). Analogously, a complete analysis of the emergence of asymmetry should include the consequences of sexual reproduction, but such an extension is beyond the confines of my work.

The following three points introduce three main elements of AD: the concept of fitness functions (3.2.1), evolutionary behaviour at non-singular points of models (3.2.2), and behaviour at singularities (3.2.3)

3.2.1. Fitness concept

A standard definition of the fitness of a population is its logarithmic per-capita growth rate, i.e. the difference between the birth and the death rates under specific environmental conditions. A population grows when its fitness is positive, i.e. when its rate of births is higher than its rate of deaths. In particular, one can asses the fitness of a newly emerged, and still rare, mutant strategy y when the 'resident' strategies $x_1,x_2,...,x_L$ are in equilibrium. This fitness is the so-called 'invasion fitness' $s_{x_1,x_2,...x_L}(y)$. There are three possible scenarios with respect to the fate of strategy y:

- It spreads and the new equilibrium will contain this new strategy. (The transition may, or may not, involve extinction of some of the residents.) This case corresponds to positive invasion fitness, i.e. $s_{x_1,x_2,...x_L}(y) > 0$.
- It becomes extinct ($s_{x_1,x_2,...x_L}(y) < 0$).
- Finding the consequences of the case $s_{x_1,x_2,...x_L}(y) = 0$ needs more detailed analysis. The mutant may spread, disappear or stay sparse according to higher-order effects in density. This situation appears generically if y is identical to x_i or in case of a linear fitness function (e.g. evolutionary game theory, cf. Maynard-Smith, 1982, Meszéna et al., 2001 or resource competition with substitutable resources, see e.g. Schreiber et al., 2003). The latter case is not relevant for us.

Henceforth we will mainly concentrate on the invasion against a single resident, for which the invasion fitness $s_x(y)$ trivially satisfies

$$s_x(x) = 0. \tag{3.1}$$

As a consequence of Eq. (3.1), the Taylor expansion of $s_x(y)$ at $(x,y)=(x_1,x_1)$, in the variables $\Delta x=(x-x_1)$ and $\Delta y=(y-x_1)$,) can be written as

$$s_x(y)\Big|_{x,y\approx x_1} = (\Delta y - \Delta x)\cdot(a_{00} + a_{10}\Delta x + a_{01}\Delta y + a_{20}\Delta x^2 + a_{11}\Delta x\Delta y + a_{02}\Delta y^2 + a_{30}\Delta x^3 + ...) \tag{3.2}$$

The sign of the function $s_x(y)$ can be conveniently plotted in a pairwise invasibility plot (PIP). (See Figure 3.2/A for an example.) In this plot, horizontal and vertical axes correspond to the resident (x) and the rare mutant (y) strategies, respectively. The dark region represents the strategy combinations for which the mutant can spread against the resident, i.e. $s_x(y)>0$. Observe that the main diagonal is always a borderline between the black and white regions, due to Eq. (3.1)

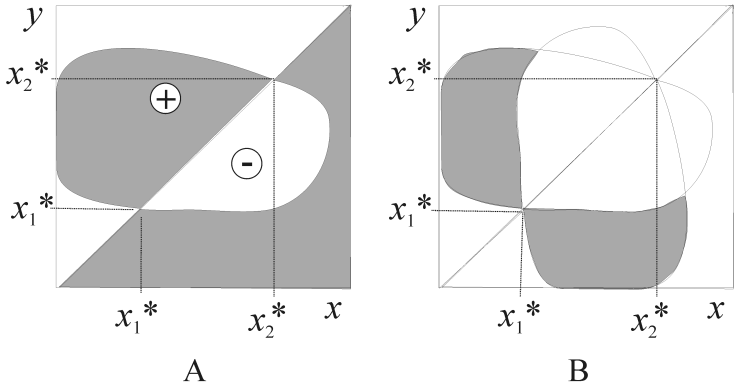


Figure 3.2. A: An example of the pairwise invasibility plot (PIP) with two singular strategies. x_1* is neither convergence stable (i.e. local attractor) nor ESS (i.e. evolutionary stable singular point), nor invasion stable. x_2* is convergence stable, invasion stable, and ESS. (The stability properties are defined in 3.2.3) x decreases in evolution (i.e. the local fitness gradient is negative, see part 3.2.2) if $x>x_2$* or $x<x_1$* and it increases (i.e. the local fitness gradient is positive) if x_2*>x>x_1* B: The grey area indicates the area of mutual invasibility.

Figure 3.2/B represents *mutual* invasibility: gray region corresponds to strategy pairs (x,y) for which *both* $s_x(y) > 0$ *and* $s_y(x) > 0$. The joint dynamics of such strategies should have an internal stable fixed point corresponding to positive number of individuals for both strategies, i.e. such a strategy pair (x,y) is able to *coexist.* Conversely (since we assumed that an internal attractor is globally attracting) coexistence implies the non-negativeness of the two growths rates. If the degenerate cases ($s_x(y)$=0 or $s_y(x)$=0) are not considered (cf. the comments at the beginning of this subsection), coexistence implies mutual invasibility.

In many evolutionary models there exists a potential function $U(y)$ (also referred to as fitness in the biological literature), with the property that the strategy with the larger potential outcompetes any strategy with a lower potential. This potential-optimisation picture (which emerges also in structural optimisation) can be connected to the concept of invasion fitness via the identification

$$s_x(y) = U(y) - U(x), \tag{3.3}$$

i.e., the invasion fitness of a mutant corresponds to its *advantage* in potential-fitness. No mutual invasibility, i.e. no coexistence is possible in such models.

Evolutionary problems, which are characterised by an invasion fitness of type (3.3), are considered as frequency-independent, because fitness advantages/disadvantages do not depend on the relative frequencies (abundances) of the competing strategies. In this case,

$$\frac{\partial s_x(y)}{\partial x \partial y} = a_{10} - a_{01} = 0 \tag{3.4}$$

follows from (3.3).

3.2.2. Directional evolution

The direction of evolution via small mutational steps is determined by the 'local fitness gradient'

$$D(x) = \left[\frac{\partial s_x(y)}{\partial y} \right]_{y=x} = a_{00} \tag{3.5}$$

provided that it is non-zero. If $D(x)>0$, a mutant with strategy $y>x$ invades the resident population with strategy x, whereas if $D(x)<0$, mutants with $y<x$ can spread. Here we assume that $|y-x|$ is small enough to guarantee that the linear term dominates the fitness advantage/disadvantage of the mutant. Moreover, $s_x(y) \approx D(x) \cdot (y-x) > 0$ implies $s_y(x) \approx D(y) \cdot (x-y) \approx D(x) \cdot (x-y) < 0$ in this context, i.e. the initial advantage of the mutant ensures that it ousts and replaces the resident, provided that $D(x) \neq 0$.

As newer and newer mutants arrive and replace their ancestors, this 'trait substitution process' constitutes a more-or-less continuous evolution in the direction determined by the local fitness gradient. See Dieckmann et al. (1996) for the deterministic approximation of this stochastic evolutionary process. This 'directional' evolution proceeds until a 'singular' strategy x^* is reached, for which $D(x^*)=0$.

In a PIP, evolution to the positive direction is represented by having a black region immediately above the main diagonal (strategies between x_1^* and x_2^* in Figure 3.2/A; see also Figure 3.3/B). Conversely, a black region immediately below the main diagonal represents evolution to the negative direction (strategies smaller than x_1^* or larger than x_2^* in Figure 3.2; see also Figure 3.3/A). Consequently, singular strategies are characterised by intersection points of the main diagonal and another borderline (Figure 3.3/C-J).

3.2.3. Properties of singular strategies

Three distinct kinds of stability can be associated with singular strategies. A singular strategy x^* is a local attractor (or *convergence stable*) if and only if $D(x)$, which determines the direction of evolution, is positive for $x< x^*$ and negative for $x> x^*$ in the vicinity of the singular point. In the generic case, this yields the condition

$$\left. \frac{dD(x)}{dx} \right|_{x=x^*} = \left. \frac{\partial^2 s_x(y)}{\partial y^2} \right|_{y=x=x^*} + \left. \frac{\partial^2 s_x(y)}{\partial x \partial y} \right|_{y=x=x^*} = a_{10} + a_{01} < 0 \tag{3.6}$$

Note that a convergence stable singular strategy need not be a local fitness maximum. Strategy x^* is a local fitness maximum (or *evolutionary stable strategy,* ESS) in the generic case, if

$$\left. \frac{\partial^2 s_x(y)}{\partial y^2} \right|_{y=x=x^*} = a_{01} < 0 \tag{3.7}$$

Finally, a rare x^* strategist mutant can invade a population with slightly different strategy x (x^* is *invasion stable*), if $s_x(x^*) > 0$, which yields generically the condition

$$\left. \frac{\partial^2 s_x(y)}{\partial x^2} \right|_{y=x=x^*} = a_{10} < 0 \tag{3.8}$$

The three conditions coincide for frequency independent fitness by Eq. (3.4) (that is why there are only 'stable/optimal' and 'unstable/pessimal' points in a potential $U(y)$), but not in

general. For example, there are singular strategies, which are convergence stable, but evolutionary unstable (Eshel, 1983, Taylor, 1989; Christiansen, 1991; Abrams, 1993).

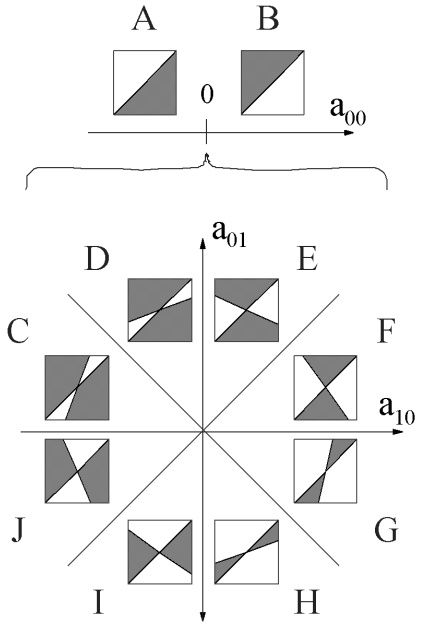


Figure 3.3: Local PIP-s around the point (x,x) at non-singular (A, B) and typical singular (C-J) x strategies. In all cases, grey/white colour corresponds to positive/negative fitness value.

At a *typical* singular strategy, the fitness function is dominated by the a_{10} and a_{01} coefficients, thus the local PIP contains two intersecting lines (one of these is the main diagonal), which divide the plot into four regions. (Later we will encounter cases when the first nonzero term is of higher order.) Figure 3.3/C-J represent the possible local configurations of the PIP around a singular strategy. The singular strategy is an ESS, if the vertical line through the intersection point lies in white regions (cases G-J) and it is invasion stable if the horizontal line lies in the black part (cases C,D,I,J). Convergence stability is indicated by a black region above the main diagonal on the left side and below the main diagonal on the right (cases C,H-J).

The really important singular points are the convergence stable ones, because an evolving population does not come close to a convergence-unstable strategy. At the same time, if a population's strategy is already x^*, the two other stability criteria determine its fate.

- If x^* is an ESS (cases G-J), it cannot be invaded by any similar mutant, i.e. it is a final rest point of the evolutionary process.
- If it is neither ESS nor invasion stable (cases E,F), similar mutants spread in a population of x^* strategists and the latter ones get extinct. (The overall result is generically divergence from x^* because the E and F type singularities are not convergence stable.)
- Finally, at an evolutionary unstable but invasion stable strategy (cases C,D), both the resident and the mutant are preserved and evolutionary branching occurs in such a way that both sub-populations diverge from the singularity. This branching process is discussed in Section 3.6. We remark that case D is usually not considered as a branching strategy, because its convergence instability prevents populations from converging to it, i.e. branching practically cannot occur.

3.3. BASIC ASSUMPTIONS

My goal is to describe symmetry-breaking via the evolution of an inherited continuous scalar strategy *x,* in accordance with the introduced framework of AD. Similarly to the symmetry-breaking optimisation variables of Chapter 2, the variable x should fulfil two restrictions, which are analogues of points (i) and (ii) in Section 2.2:

(i) There is one and only one strategy $x=x_0$, which corresponds to bilaterally symmetric body structure. The evolutionary development is assumed to start with a population with exact symmetry (i.e. with strategy x_0). This condition enables us to simulate symmetry-breaking by the model.

(ii) The set of strategies $x\in\mathbb{R}$ is invariant to reflection, i.e. for arbitrary $x_1\in\mathbb{R}$ there exists $x_2\in\mathbb{R}$ so that the mirror image of an $x=x_1$ strategist is an $x=x_2$ strategist. The lack of this property would mean that our simplified model violates the symmetry of the biological system. Furthermore- as a purely technical condition-we assume that the mirror image of an $x_0+\Delta x$ strategist is an $x_0-\Delta x$ strategist (cf. the case of reflection symmetry and one variable in Section 2.3).

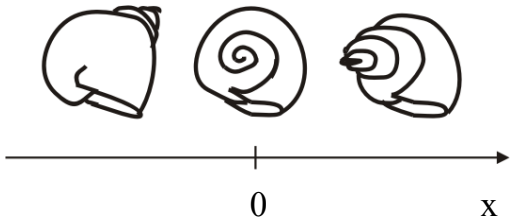


Figure 3.4: An example of symmetrical strategies: x is the slope of the spiral axis of the shell. $x>0$ means dextral while $x<0$ means sinistral shell.

If both conditions are fulfilled, we call x_0 'symmetrical strategy'. As an illustration, consider a geometrical model of snail shell forms (Raup, 1962) with three parameters, one of which is the slope x of the spiral (Figure 3.4). If $x=0$, we have a curve in a plane, generating a flat shell, reflection-symmetric with respect to this plane. On the other hand, if $x>0$, the shell is peaked and asymmetrical (dextral). With a negative value of x, the result is a reflected (sinistral) shell. In such a situation $x=0$ is a 'symmetrical strategy'.

We assume in line with AD methodology (see more details in Section 3.2) that strategy x can be modified only by small (though not infinitely small) mutational steps. In particular, we do not allow such "macro" mutations, via which 'left-handed' offspring of a 'right-handed' parent appear. See the Discussion for the consequences of some different assumptions.

We have to take into account that ontogeny of a symmetric body plan is simpler (and more ancient) than that of an asymmetric one. That is why we assume *exact* body symmetry, as a starting point. Then, emergence of asymmetry can be initiated in two ways (analogous to the two categories of evolutionary branching in the Introduction):

I. A *change in the developmental program* allows body asymmetry and asymmetry proves to be advantageous. This scenario can happen in constant environment.

II. The possibility for asymmetry is already present and an *environmental change* makes asymmetry advantageous.

Beyond the above two categories, we will distinguish between the following types of symmetry-breaking from the ecological point of view:

- symmetry-breaking in frequency dependent models. Within this class, we introduce the following two subclasses (see Section 3.4)
  - case of *strong symmetry*: the model behaviour depends on the relative sizes of competing populations in general, but not on the relative sizes of "lefty" and "righty" populations, which are mirror images of each other. With other words, asymmetrical individuals and their mirror images are ecologically identical.
  - case of *weak symmetry*: asymmetrical individuals and their mirror images are not identical.
- symmetry-breaking in frequency independent models.

Due to the two types of classifications we will investigate 3×2=6 separate cases in Section 3.5.

3.4. TWO TYPES OF SYMMETRY-BREAKING

The question of evolutionary advantage/disadvantage of symmetry breaking is relevant only if the *environment* itself possesses the symmetry in question, that is, if replacing *all individuals* of the model by the reflected ones does not affect the model behaviour. In the frequency-independent models, this condition yields

$$U(x_0+\Delta x)= U(x_0-\Delta x), \tag{3.9}$$

where Δx denotes x-x_0. According to eq. (3.9), the fitness of an individual is independent of its left/right handedness. This means that 'left-' and 'right-handed' individuals are completely equivalent from the point of view of ecological interactions: if only *a part of the individuals* are replaced by reflected ones in a population, this change does not affect the model behaviour either.

In the frequency-dependent case, the analogue of eq. (3.9) is

$$s_{x_0+\Delta x}(x_0+\Delta y) = s_{x_0-\Delta x}(x_0-\Delta y)\,, \tag{3.10}$$

which does not necessarily mean the equivalence of left- and right-handed individuals. Thus, two levels of symmetry can be distinguished in case of frequency dependence, i.e. when the interactions between the individuals affect the fitness function. We call a symmetrical strategy *strongly symmetrical* if all of the interactions are independent of left/right handedness. In this case, one can replace some (but not necessarily all) individuals by reflected ones and find the same model behaviour. In contrast, if the interactions depend on the handedness of the affected individuals, only the simultaneous reflection of all individuals is an invariant transformation of the model. The latter situation will be referred to as *weak symmetry*.

The symmetry condition (3.10) implies that all terms of odd order vanish in the Taylor expansion of the invasion fitness function at point (x_0, x_0). Thus, in case of weak symmetry, the general form of the expansion is more specific than Eq. (3.2):

$$s_x(y)\Big|_{x,y\approx x_0} = (\Delta y-\Delta x)\cdot(a_{10}\Delta x + a_{01}\Delta y + a_{30}\Delta x^3 + a_{03}\Delta y^3 + a_{21}\Delta x^2\Delta y + a_{12}\Delta x\Delta y^2 + ...)\,. \tag{3.11}$$

Since a_{00} vanishes, the symmetrical strategy x_0 is always singular. Since the coefficients a_{10} and a_{01} remain generically non-zero, the classification of the possible PIP-s for a weak symmetry remains the same as in Figure 3.3/C-J.

In contrast, the strong symmetry is characterised by a more restrictive condition:

$$s_{x_0+\Delta x}(x_0+\Delta y) = s_{x_0+\Delta x}(x_0-\Delta y) = s_{x_0-\Delta x}(x_0+\Delta y) = s_{x_0-\Delta x}(x_0-\Delta y. \tag{3.12}$$

which corresponds to the fact that reflection of only the residents or only the mutants are invariant model transformations. In this case the general form of the invasion fitness function is:

$$s_x(y)\Big|_{x,y\approx x_0} = (\Delta y^2 - \Delta x^2)\cdot(b_{00} + b_{10}\Delta x^2 + b_{01}\Delta y^2 + b_{20}\Delta x^4 + b_{11}\Delta x^2\Delta y^2 + b_{02}\Delta y^4 + ...)\cdot \tag{3.13}$$

The expansion contains only the terms, which are even *in both variables*, due to the more restrictive symmetry condition. Comparison with (3.11) yields $a_{01}=a_{10}=b_{00}$ and similar relations for the higher-order coefficients.

For strong symmetry, $b_{00}<0$ (Figure 3.5/A) implies convergence, evolutionary and invasion stability, because

$$\left.\frac{\partial^2 s_x(y)}{\partial x^2}\right|_{y=x=x_0} + \left.\frac{\partial^2 s_x(y)}{\partial x \partial y}\right|_{y=x=x_0} = 2b_{00} + 0 \tag{3.14}$$

$$\left.\frac{\partial^2 s_x(y)}{\partial y^2}\right|_{y=x=x_0} = 2b_{00}, \tag{3.15}$$

$$\left.\frac{\partial^2 s_x(y)}{\partial x^2}\right|_{y=x=x_0} = -2b_{00}, \tag{3.16}$$

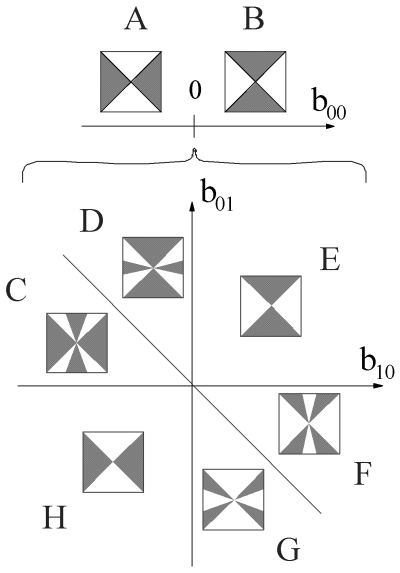


Figure 3.5: Local PIP-s for strong symmetry. A, B: x_0 is a generic strongly symmetrical strategy ($b_{00}\neq 0$); C-H: x_0 is a degenerate strongly symmetrical strategy ($b_{00}=0$). Grey/white colour corresponds to positive/negative fitness values.

Such a strategy is an attractive endpoint of evolution. Conversely, $b_{00}>0$ leads to a singularity, which is unstable in all senses (Figure 3.5/B), i.e. it is a repellor.

Later, we will also be interested in the case of vanishing b_{00}. If b_{00}=0, the character of the singular point is typically determined by b_{10} and b_{01} (Figure 3.5/C-H). The six emerging configurations are partly invasion stable (C,D,H), partly ESS (F-H), and the two properties are not equivalent. In particular, C and D are *branching strategies*.

Note the geometrical interpretation of the two kinds of symmetry. Weak symmetry is equivalent to the fact that the PIP is invariant under a rotation of 180° around the point (x_0,x_0). For strong symmetry, the PIP has a vertical and a horizontal symmetry axis at the point (x_0,x_0).

As already noted, weak and the strong symmetry are equivalent in frequency independent models, because frequency-independence means that the strategy of the competitors (including the handedness) does not affect the fitness of a strategy. Moreover, we have in this case

$$\frac{1}{4}\frac{\partial^4 s_x(y)}{\partial x^2 \partial y^2} = b_{10} - b_{01} = 0 \tag{3.17}$$

3.5. EMERGENCE OF ASYMMETRY

In this Section we study the evolutionary loss of bilateral symmetry. We mentioned in Section 3.3 that it can occur in constant environment (case I) or it can be induced by environmental change (case II). In the latter case, we suppose that, initially, the symmetrical strategy is evolutionary stable and the population assumes this strategy. The phenomenon will be discussed separately for models without frequency dependence (Section 3.5.1), as well as for weak symmetry (Section 3.5.2), and strong symmetry (Section 3.5.3). with frequency dependence

3.5.1. Frequency independent models

It has been demonstrated in Section 3.2 that evolutionary and invasion stability are equivalent in frequency independent models and branching cannot occur. Thus asymmetry can only emerge via type (a) divergence from the symmetrical strategy in constant as well as in changing environment. (Divergence can be realised at Figure 3.5/B type strategies.). This phenomenon is closely related to structural optimisation with a *global criterion* (yielding a smooth potential, see Sections 2.1.1-2). At the example of Figure 2.1/C, the symmetrical optimum bifurcates into a pessimum if a model parameter (p) is varied. The same phenomenon initiates the evolutionary emergence of asymmetry in case of changing environment (i.e. time-dependent model parameters).

3.5.2. Weak symmetry

It has been shown in Section 3.4 that the classification of generic weakly symmetrical strategies is the same as that of singular strategies without symmetry (see Figure 3.3). In constant environment (case I), asymmetry can emerge via type (a) divergence (cf. Figure 3.1) if the possibility of asymmetry develops when the population is at a repellor strategy (such as

Figure 3.3/E,F). Alternatively, type (b) branching may occur if asymmetry becomes reachable at a branching strategy (Figure 3.3/C,D). Notice that evolution starts exactly from the singular point, so convergence stability is irrelevant and Figure 3.3/D is also a branching point. The steps of this kind of branching process are summarised in Section 3.6 parallel with a different branching pattern.

Changing environment (case II) can be described by a moving point in the a_{10}-a_{01} plane (Figure 3.3), which is originally located in the ESS region. There are two generic possibilities for loosing evolutionary stability: reaching the border in a non-invasion stable or in an invasion stable state (Figure 3.6, case 1 and 2). In case 1, type (a) divergence from the symmetrical strategy occurs, while in case 2, an ordinary (type (b)) evolutionary branching is initiated.

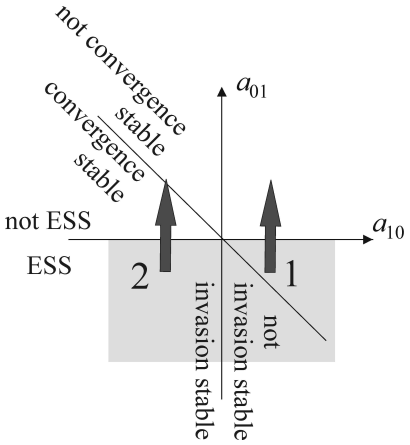


Figure 3.6: The parameter plane for weak symmetry (cf. Figure 3.3, eq. (3.11)). (3.13)). Our models are assumed to start from the ESS (grey) region. Arrows indicate the two generic ways (1, 2) of losing the ESS property in a time-dependent model.

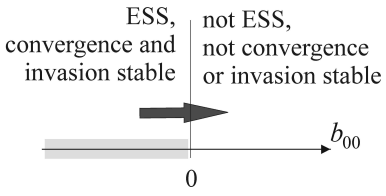


Figure 3.7: The parameter line for strong symmetry (cf. Figure 3.5, eq. (3.13)). Our models are assumed to startfrom the ESS (grey) region. The arrow indicates the way of losing the ESS property in a time-dependent model.

3.5.3. Strong symmetry in frequency dependent models

Despite frequency-dependence, the ESS and the invasion stability conditions are *generically* equivalent at strongly symmetrical strategies. Thus, the common way of the emergence of asymmetry is of type (a), analogously to part 3.5.1. However, as the degenerate cases of Figure 3.5/C-H break the equivalence, a different and surprising scenario may be realised in case II (changing environment) in presence of slow variation of b_{00}. The bifurcation process (Figure 3.7) has the following main steps:

- Initially, $b_{00}<0$ and the symmetrical strategy is Figure 3.5/A type (a stable evolutionary endpoint)
- The coefficient b_{00} approaches zero and one of the configurations of Figure 3.5/C-H emerges temporarily
- After some time, b_{00} gets far from zero on the positive side. The degenerate configuration disappears and the strategy becomes Figure 3.5/B type (repellor).

If the mutation step was infinitesimally small, the higher-order terms would dominate the Taylor expansion (3.13) only for infinitesimally short period, not long enough to have any effect on the evolution of the population. However, we consider *small, but finite* steps in the strategy space. In this case the 4^{th}-order terms dominate the quadratic ones in a finite interval of b_{00}, which may correspond to a long time interval, if the environmental change is

sufficiently slow. Here evolutionary development of the population depends on the properties of the temporarily emerging, degenerate configuration:

- If the degenerate state is neither invasion stable nor an ESS (Figure 3.5/E), type (a) divergence occurs as soon as the close-to-degenerate state is reached.
- If the degenerate state is an ESS (Figure 3.5/F-H), the population stays symmetric, but later, as the degenerate state is replaced by a Figure 3.5/B type repelling strategy, divergence occurs again.
- If the degenerate state is invasion stable but not evolutionary stable (Figure 3.5/C,D) an evolutionary branching occurs in the close-to-degenerate state. In Section 3.6, we describe this branching process in detail and show that it is of type (c).

3.6. A NOVEL WAY OF EVOLUTIONARY BRANCHING

The main goal of this section is to describe the details of the novel branching process of a population with *strongly symmetric* strategy, which was recognised in Section 3.5.3. (This is the situation $b_{10}<b_{00}\approx 0<b_{01}$, see Figure 3.5). This process differs significantly from the generic pattern of branching *without symmetry* (Geritz et al., 1998). We describe the two ways of branching simultaneously to highlight the similarities and differences. Notice that the generic branching pattern in case of *weak symmetry* (Section 3.5.2) is the same as the latter one.

The steps of the two processes are collected in the left (standard case) and right (strongly symmetric case) column of Table 3.1. In both cases, row 1 presents the fitness functions before branching, row 2 shows why two evolving branches coexist, and row 3 presents the corresponding fitness functions. It is demonstrated in row 4, that the number of coexisting branches cannot be more than two. Finally, the directions of evolution are determined in row 5.

In the standard case, the branching type evolution starts with the arrival of a mutant, which is located on the opposite side of the singularity x^* than the ancestor (row 2, left column). The consecutive mutation events always end up with extinction of the middle strategy (row 5, left column), i.e. two sub-populations evolve away from each other, resulting in a type (b) branching. In the strongly symmetric case, branching starts with the coexistence of a new, asymmetric mutant and its symmetric ancestors (row 2). The sequence of mutation-extinction steps results in a branching, in which one of the strategies stays symmetric while the other one evolves away; that is, a symmetric-asymmetric pair emerges in a type (c) branching (row 5, right column).

Evolution follows the introduced patterns as long as both sub-populations are close to the singular strategy. Later, the asymmetrical branch (at type (c) branching) or both branches (at type (b) branching) continue to evolve directionally according to their respective local fitness gradient, as demonstrated in Section 3.2.2 for a lone strategy.

–	**standard case (Figure 3.3/C)**	**case of strong symmetry (Figure 3.5/C,D)**
1	The fitness function $s_{x^*}(y)$, as a function of y, has a minimum at $y=x^*$. Locally, it can be approximated as (cf. Eq. (3.2), Figure 3.8/A). $$s_{x^*}(y) \approx a_{01}(y-x^*)^2$$	The fitness function $s_{x_0}(y)$, as a function of y, has a minimum at $y=x_0$. Locally, it can be approximated as (cf. Eq. (3.13), Figure 3.9/A) $$s_{x_0}(y) \approx b_{01}(y-x_0)^4 \quad (3.18)$$
2	Two strategies near to, but at the opposite sides of the singularity (i.e. $x_1 \le x^* \le x_2$) mutually invade each other and, consequently, are able to coexist.	If x_1 is near to x_0, x_1 and x_0 mutually invade each other, i.e. they are able to coexist
3	If x_1 and x_2 are coexisting (cf. row 2) and both of them are near to x^*, the invasion fitness is $$s_{x_1x_2}(y) \approx a_{01}(y-x_1)(y-x_2)$$ (see Figure 3.8/B, eq. (3.2)), because $x_1, x_2 \approx x^*$ implies $s_{x_1x_2}(y) \approx s_{x^*}(y)$ and $s_{x_1x_2}(x_1) = s_{x_1x_2}(x_2) = 0$ by definition .	If the $x_1=x_0+\Delta x_1$ and x_0 strategies are coexisting (cf. row 2) and x_1 is near to x_0, the invasion fitness has a double root at x_0 and two roots arranged symmetrically around x_0: $$s_{x_0x_1}(y) \approx b_{01}(y-x_0)^2(y-x_0-\Delta x_1)(y-x_0+\Delta x_1) \quad (3.19)$$ (see Figure 3.9/B, eq. (3.13)), because $x_1 \approx x_0$ implies $s_{x_0x_1}(y) \approx s_{x_0}(y)$ and $s_{x_0x_1}(x_0) = s_{x_0x_1}(x_1) = 0$ by definition, and finally $s_{x_0x_1}(x_0-\Delta x) = s_{x_0x_1}(x_0+\Delta x)$ for any Δx, due to eq. (3.12).
4	If more than two strategies coexisted, the corresponding fitness function would be 0 at each of them. The locally second-order invasion fitness function cannot have more than two zeros, i.e. coexistence of more than two strategies is impossible in the vicinity of x^*.	Generically only one strategy can coexist with x_0, because the arrival of two strategies with exactly the same distance from x_0 ($x_1=x_0+\Delta x$, $x_2=x_0-\Delta x$) is improbable and otherwise ($x_1=x_0+\Delta x_1$, $x_2=x_0+\Delta x_2$) the fitness function should have zeros at $x_0 \pm \Delta x_1$ and $x_0 \pm \Delta x_2$ and a double root in x_0. This is impossible, because it has only four roots in the vicinity of x_0.
5	If a new mutant emerges in presence of a coexisting pair, one of the three should become extinct by row 4. The strategy becoming extinct should have a negative growth rate when it has become rare already. As $a_{01}>0$, this condition holds only for the middle strategy, i.e. if $x_1<x_2<x_3$ are the three strategies, x_2 will become extinct independently of which of them was the mutant. (Figure 3.8/C.)	When a new mutant appears at the equilibrium of x_0 and another strategy, one of the three strategies (ie. x_0, $x_1=x_0+\Delta x_1$ and $x_2=x_0+\Delta x_2$) should become extinct by row 4. Assume that $\lvert\Delta x_1\rvert < \lvert\Delta x_2\rvert$. Then x_1 will become extinct, because the strategy becoming extinct should have a negative growth rate when it has become rare already (Figure 3.9/C).

Table 3.1: The course of the branching process at a generic branching strategy x^* (standard case, left column), and the branching process emerging at a degenerate, strongly symmetric strategy x_0 (right column).

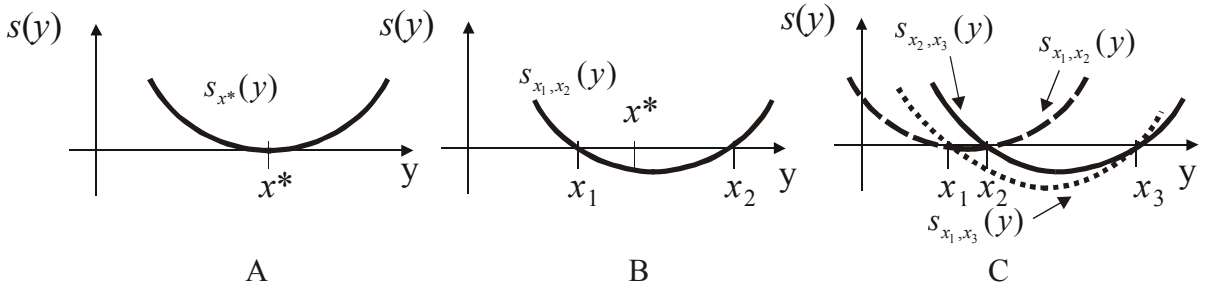


Figure 3.8: Fitness of possible mutants at a standard branching strategy without symmetry. A: before branching B: after branching C : fitness functions related to the coexistence of all pairs of strategies from x_1, x_2, and x_3.

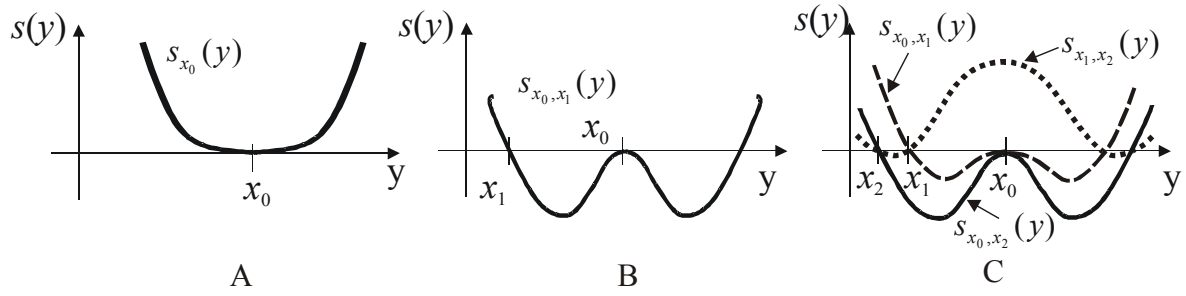


Figure 3.9: Fitness of possible mutants at a degenerate, strongly symmetric branching strategy. A: before branching B: if x_0 and another strategy coexist C: fitness functions related to the coexistence of all pairs of strategies from x_0, x_1, and x_2.

3.7. A MODEL EXAMPLE

In this Section we present a specific model to illustrate the type (c) branching. It is based on the examples of Levene (1953), Geritz et al. (1998). There are two parameters in the model, b and T, the latter representing the *time-dependence* of the model.

3.7.1. Description of the model

Consider a population of $x_1,x_2,\ldots,x_n$ strategists, the number of individuals is $N_1,N_2,\ldots,N_n$, respectively. The model assumes non-overlapping generations, which live in a spatially heterogeneous environment consisting of *two different patches*. A limited number of individuals, denoted by K_1 and K_2, live in each of the patches. The total number of individuals is constant:

$$N_1 + N_2 + \ldots + N_n = K_1 + K_2. \tag{3.20}$$

The lifecycle of each generation consists of three parts.

- During dispersal, the offspring is distributed randomly in both patches; the frequency of a strategy x_k among the offspring is proportional to the frequency of the parents with the same strategy, i.e. to N_k.
- In the second phase, the offspring is subjected to frequency-independent selection, which changes the relative frequencies of the strategies in both patches

independently. The chance of an x_k strategist in the i^{th} patch of surviving this phase is proportional to a given function $f_i(x_k)$.

– In the third phase, the survivors spread in both patches until their numbers reach the capacities (K_1 and K_2) of the patches. The relative frequencies of the strategies in each of the patches are constant in this phase.

In this model, the chance of surviving the second phase is

$$f_1(x) = e^{-2b_1^2 \cdot x^2 - x^4}. \tag{3.21}$$

$$f_2(x) = e^{2b_2^2 \cdot x^2 - x^4}. \tag{3.22}$$

with b_1 and b_2 positive parameters (see also Figure 3.10). Since both functions are symmetrical, x=0 is a symmetrical strategy. Observe that this is an example of *strong symmetry*, hence there is no difference between x and $-x$ strategists.

The optimal strategy is $\pm b_2$ in the second patch, i.e. there is an *asymmetrical optimum*. In the first patch, there is a symmetrical optimum the 'strength' of which is determined by b_1

Figure 3.10: The functions $f_1(x)$ and $f_2(x)$.

Consider a rare mutant with strategy y in an equilibrium population of $x_1,x_2,\ldots,x_n$ strategists with equilibrium numbers $\tilde{N}_1$, $\tilde{N}_2,\ldots$, $\tilde{N}_n$. If N_y is the (small) number of mutants in a generation, the N_y' number of mutants in the next generation can be approximated as

$$N_y' = K_1 \cdot \frac{f_1(y) \cdot N_y}{\sum_{k=1}^{n} f_1(x_k)\tilde{N}_k} + K_2 \cdot \frac{f_2(y) \cdot N_y}{\sum_{k=1}^{n} f_2(x_k)\tilde{N}_k}. \tag{3.23}$$

Consequently, the logarithmic per-capita growth rate of the rare mutants is

$$s_{x_1,x_2,\ldots,x_n}(y) = \log\left(\frac{N_y'}{N_y}\right) = \log\left(K_1 \cdot \frac{f_1(y)}{\sum_{k=1}^{n} f_1(x_k)\tilde{N}_k} + K_2 \cdot \frac{f_2(y)}{\sum_{k=1}^{n} f_2(x_k)\tilde{N}_k}\right). \tag{3.24}$$

To reduce the number of model parameters, assume that b_1=b_2=b and let the parameter T be defined as

$$T = \frac{K_2}{K_1 + K_2}. \tag{3.25}$$

We can determine the fitness function of rare mutants in this model in case of a monomorphic resident population (with strategy x):

$$s_x(y) = \log\left[(1-T) \cdot e^{-2b^2 \cdot (y^2 - x^2) + x^4 - y^4} + T \cdot e^{2b^2 \cdot (y^2 - x^2) + x^4 - y^4}\right]. \tag{3.26}$$

As it is expected, the fitness function satisfies the condition of strong symmetry (3.12).

The results for the dimorphic case (two resident populations) are more involved. First, the equilibrium densities $\tilde{N}_1$, $\tilde{N}_2$ of the two resident populations have to be determined from the following two equations:

$$\widetilde{N}_i = K_1 \cdot \frac{f_1(x_i)\cdot \widetilde{N}_i}{f_1(x_1)\widetilde{N}_1 + f_1(x_2)\widetilde{N}_2} + K_2 \cdot \frac{f_2(x_i)\cdot \widetilde{N}_i}{f_2(x_1)\widetilde{N}_1 + f_2(x_2)\widetilde{N}_2} \qquad \text{for} \qquad i = 1,2 \tag{3.27}$$

Second, the results for $\tilde{N}_1$ and $\tilde{N}_2$ and Eq. (3.21), (3.22) and (3.25) are substituted into (3.24) to obtain the fitness function. The results are quite complicated and they have to be analysed numerically.

The fitness function for three or more coexisting strategies is uninteresting, since our analysis (in Section 3.7.2) shows that the maximal number of strategies in stable coexistence is two.

3.7.2. Singular strategies and coalitions in the model

We investigated the behaviour of the model at different values of b and T. Analysis of the fitness function (3.26) yielded the following results:

- x_0=0 is singular strategy, since it is a symmetrical strategy.
- We determined the fitness gradient (see Eq. (3.5)) by deriving Eq. (3.26) with respect to y. Solving $D(x^*)$=0, we found another pair of singular strategies:

$$x^*(b,T) = \pm b \cdot \sqrt{2\cdot T - 1} \qquad \text{if} \qquad T > \tfrac{1}{2}. \tag{3.28}$$

- We analysed the stability properties of the singular strategies by substituting Eq. (3.26) into the conditions (3.6) and (3.7). The x_0=0 strategy is ESS, invasion and convergence stable if $T < \frac{1}{2}$, it is degenerate if $T = \frac{1}{2}$ and it is unstable in all senses if $T > \frac{1}{2}$.
- The asymmetrical singular strategy is ESS invasion and convergence stable if $b<2^{-1/4}$, or if $b>2^{-1/4}$ and $T>T^*$ with

$$T^*(b) = \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{8b^4}}\,. \tag{3.29}$$

  Otherwise it is a branching strategy (convergence and invasion stable but not ESS).
- In the degenerate state ($T = \frac{1}{2}$) state, the stability of the symmetrical strategy can be determined from fourth derivatives of (3.26) with respect to x and y, which determine the b_{10} and b_{01} coefficients (see Eq. (3.13) and Figure 3.5). The symmetrical strategy is ESS (Figure 3.5/H type) if $b<2^{-1/4}$ and it is branching strategy (of type Figure 3.5/C) if $b>2^{-1/4}$.

Further, numerical computations showed that:

- There exists a convergence stable and ESS coalition of symmetrical x_1=0 and asymmetrical $x_2(b)$ strategists at appropriate parameter values.
- The value of $x_2(b)$ is independent of T.
- The coalition exists if $T_{min}(b)<T<T_{max}(b)$. If $T<T_{min}$(b), the asymmetrical strategy vanishes, while if $T>T_{max}(b)$, the symmetrical strategy gets extinct.

The PIP associated to the fitness function (3.26) is illustrated in Figure 3.11 for some values of b and T. We can also construct an evolutionary bifurcation diagram of the model, which

shows the singular strategies and coalitions at specific values of b (Figure 3.12). We also plotted $x_2(1)$, $T_{min}(1)$ and $T_{max}(1)$ in Figure 3.12. For other values of b, the functions $x_2(b)$ and $T_{min}(b)$ can be determined numerically and $T_{max}(b)$ is the solution of $x_2(b)=x^*(b,T_{max}(b))$ (cf. Eq. (3.28)).

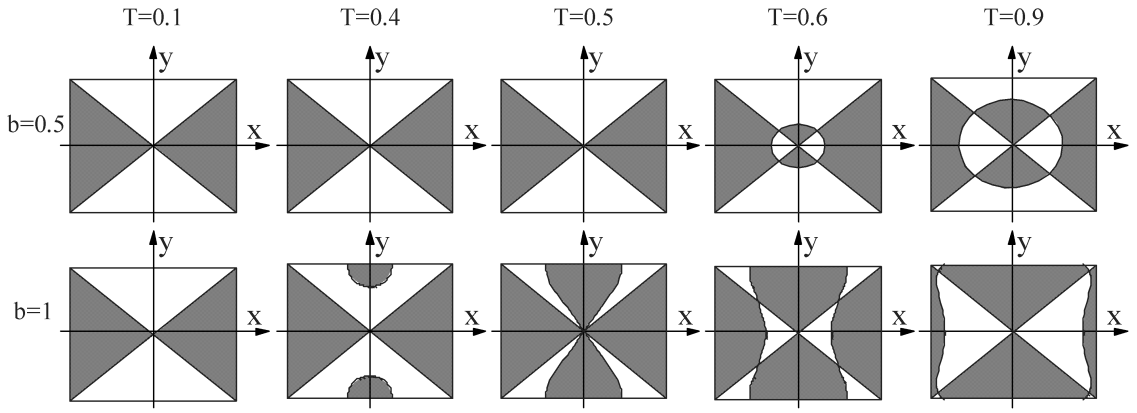


Figure 3.11: PIP of the model at specific parameter values (the grey region means positive fitness and the white means negative)

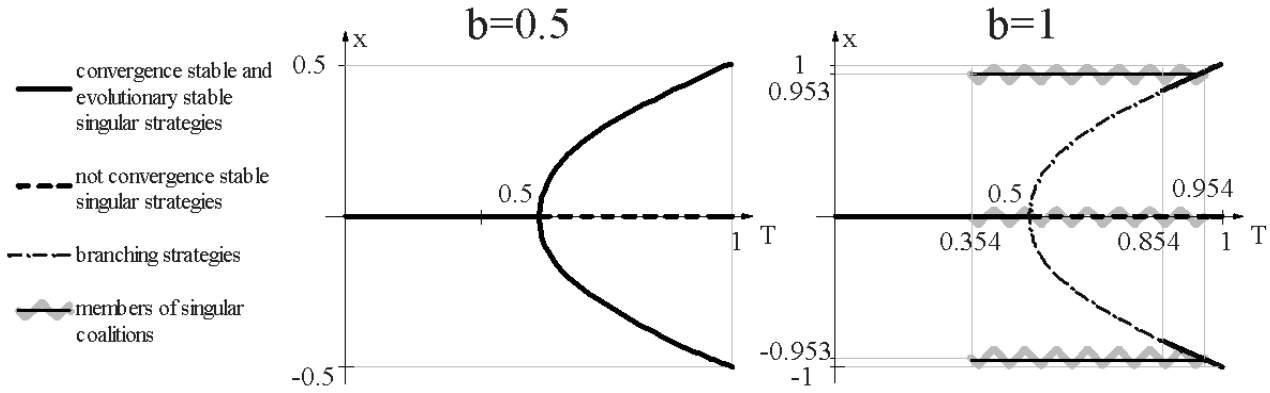


Figure 3.12: Bifurcation diagram of the model at b=0.5 and b=1.

3.7.3. Branching in the model

As we already showed, the model has a degenerate, symmetrical branching strategy at $T = \frac{1}{2}$ and $b>2^{-1/4}$. This means that a type (c) branching occurs at appropriate values of b (e.g. b=1), if the parameter T (representing the capacity of the second patch relative to the first one) *slowly increases* on evolutionary time scale and it reaches 1/2. Figure 3.13/A, illustrates this branching in numerical simulations.

If the increase of T is *faster*, the model behaviour is different: type (c) branching is replaced by type (a) divergence from the symmetrical strategy, followed by an 'ordinary' branching (Figure 3.13/B).

If $b<2^{-1/4}$, no branching occurs. If T is increased and it reaches 1/2, the population diverges from the symmetrical strategy (type (a)) and converges to the asymmetrical singular strategy (Figure 3.13/C), which itself slowly moves with the increase of T.

At our example the increased speed of environmental change modified the pattern of the emergence of asymmetry (the type (c) branching of Figure 3.13/A was replaced by type (a) divergence and a standard branching in an asymmetrical state, as seen in Figure 3.13/B), but not the final outcome. There are other models where the higher speed of environmental change prevents branching, and modifies the evolutionary outcome as well.

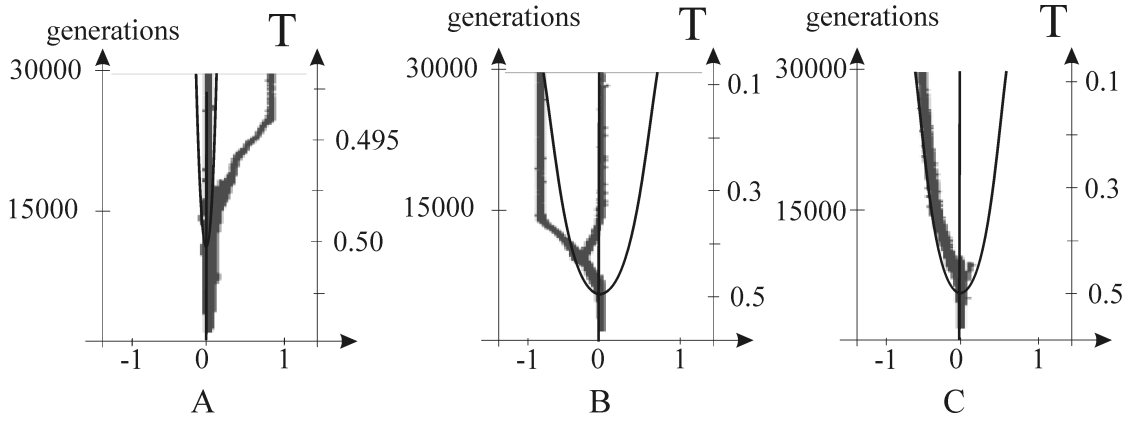


Figure 3.13: Numerical simulations of the model. Thin lines indicate the singular strategies in the model as functions of T. A-B: with b=1 and different speeds of environmental changes (T). In both cases, the two coexisting branches converge to the stable coalition (x_1;x_2)≈(0;±0.953), cf. Figure 3.12/B. C: with b=0.5. Branching does not occur, evolution converges to the stable singular strategy.

3.8. BIOLOGICAL EXAMPLES OF SYMMETRICAL STRATEGIES

Some illustrative examples of *strongly* and *weakly symmetrical* strategies based on real populations are summarised in this section.

A widely known example of the secondary loss of bilateral symmetry is the beak of crossbills, which we introduce based on Benkman (1996), see also other works of the same author. The asymmetry of the beak is measured by the angle x of the lower mandible of crossbills: x=0, x<0 and x>0 correspond to straight, leftward curved and rightward curved lower mandibles, respectively. Needless to say, x=0 is a symmetrical strategy.

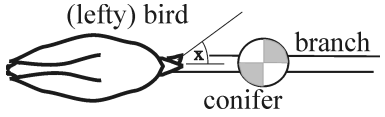


Figure 3.14: Schematic upper view of a crossbill (head to the right) with lefty beak standing on a branch next to a conifer. The bird can pick seeds from the lower left quarter of the conifer (in grey colour). The bird could stand on the other side of the conifer as well, in that case it could reach the seeds in the upper right quarter (also grey). Seeds in the white quarters of the conifer are reachable only by a righty beak.

Crossbills use their special beaks to pick out seeds from pinecones. Many of them, such as the White winged crossbill subspecies *Loxia leucoptera megaplaca* forage on pinecones, which cannot be twisted or removed from the trees (Figure 3.14). Individuals can pick out seeds from only a part of the conifers depending on the direction of their beak. Thus, 'lefties' and 'righties' are ecologically *different*: the rarer one has ecological advantage in comparison

with the more common one. The difference between lefties and righties is also indicated by the stable 1:1 ratio of the two morphs. This is an example of a *weak symmetry*. In contrast, the subspecies *Loxia leucoptera leucoptera* and *bifasciata* forage on different conifers, which are easily removed or twisted. In this case no ecological difference seems to exist between the two types of beaks. Accordingly, significant variance in the ratio of the two morphs was observed in different populations. This is an example for strong symmetry.

Different species of *Cichlid* fishes in Lake Tanganyika provide another pair of examples. The scale-eating *Perissodus microlepis* attack other species from behind and try to byte scales from the left or the right side of the victim **(**Takahashi et al, 1994). They have two asymmetrical forms in correspondence with the hunting strategy: Some of them open their mouth to the left, while the other ones have right-sided mouths. If x is the angle of mouth opening (x=0 for symmetrical mouth, x<0 for left-sided and x>0 for right-sided mouth), x=0 is again a *symmetrical* strategy. It is *weakly* symmetrical, because a small group of $-x$ in a big population of x strategists would have higher fitness than the frequent phenotype, because of the unexpected way of attacking the victims and the inequality $s_x(-x)>0$, contradicting Eq. (3.12).

The herbivorous species *Telmatochromis temporalis* has similar, asymmetrical mouth, used to bite weed from the side of rocks while swimming along them (Mboko et al., 1998). As the weed does not adapt itself to the 'hunting strategy' of the fish, the x and $-x$ strategists are ecologically equivalent in this case. Thus, $x_0=0$ can be considered as a *strongly symmetrical* strategy.

More recent studies of Lake Tanganyika populations show, that the slightly asymmetrical body structure of many *Cichlid* species might have a different reason: it is an adaptive result of cross-predation in food chains. (Lefty predators tend to prefer righty victims and vice versa, see Nakajima et al., 2004). According to these results, all these species are examples of weak symmetry.

Finally, the shell chirality of snails, introduced in Section 3.4, becomes important at mating (Asami et al., 1998). The mating strategy of some *pulmonate* land snail species, which have relatively flat shells, prevents mating with individuals of opposite chirality, while a different mating behaviour of other, tall-shelled species permits it. The different chirality has in the latter case only minor disadvantage according to experiments of Asami et al (1998). The first situation is a typical example of weak symmetry and the second is close to strong symmetry (which would be perfect if there was no disadvantage of cross-mating at all).

	weak	strong
non frequency-dependent		(a)
frequency dependent	(a) (b)	(a), +(c) only in time-dependent models

Table 3.2: Types of emergence of asymmetry. Note that symmetry is always strong in absence of frequency dependence. (a), (b) and (c) refer to the scenarios of Figure 3.1.

3.9. DISCUSSION

In this chapter, we examined the evolutionary patterns of the emergence of secondary (partial) asymmetry in species with bilateral symmetry in their basic body plan. Three distinct scenarios have been described, as illustrated in Figure 3.1. Two levels of bilateral symmetry ('strong' and 'weak') have been defined and the difference has been illustrated on biological examples. We determined the typical evolutionary patterns in different classes of models concerning symmetry.

The results are summarised in Table 3.2: the type (a) emergence of asymmetry (when the superior asymmetrical form outcompetes the inferior symmetrical one) is possible in all three cases. Type (b) (when two asymmetrical variants emerge, avoiding competitive exclusion) requires weakly symmetrical frequency dependence. Finally, type (c) (when an asymmetrical form branches away from the unchanged and surviving symmetrical form) is restricted to the case of frequency-dependent strong symmetry.

Type (c) is a novel way of evolutionary branching. It differs from the usual pattern since the initial speed of divergence is not equal for the two branches. It relies on the transient dominance of the higher-order terms in the evolutionary models, i.e. on a sufficiently slow change of the environmental parameters. We simulated this type of branching on a symmetrical version of Levene's classical multi-patch model.

Our study assumes that the degree and the direction of asymmetry (both determined by the phenotypic value x) are inherited from the parents and mutations cause small deviation in x. In some cases, the direction of asymmetry develops randomly at some stage of the individual development (see Brown et al, 1990, Govind, 1989). This different inheritance mechanism would leave type (b) unchanged, and modify type (a) or (c) in such a way that a second asymmetrical branch with opposite handedness also appears.

It is also possible that, while handedness is inherited from parents, a special 'reflected' mutation (i.e. an offspring with opposite handedness) may occur with some probability. This is the case e.g. if the handedness is determined by a simple two-allele locus. If this type of mutation is frequent enough, again, the asymmetric variants will populate both asymmetric branches in types (a) and (c). However, if the reflected mutations are exceedingly rare, the relative frequencies of the lefties and the righties will change randomly.

An interesting way to continue our research would be to detect the patterns of the emergence of asymmetry in Nature. Empirical study of speciation is very difficult: while it is too slow for direct observation, simultaneously it is too fast to leave a fossil record (Eldredge & Gould, 1972). As a consequence, theoretical insight always played an important role in this field (Turelli et al. 2001). We contributed to this endeavor by investigating the bifurcation patterns of emergence of body asymmetry. We are intrigued to learn whether the novel way of evolutionary branching we uncovered is a part of the natural process of evolution.

CHAPTER 4 SUMMARY AND PRINCIPAL RESULTS

In this work I studied the relation between symmetry and optima in two distinct fields of science. My main goal was to study the existence and emergence of objects with "imperfect" symmetry, i.e. objects with slight asymmetry. The structural topic was motivated by the apparent lack of engineering structures with imperfect symmetry, while the research concerning evolution was based on the numerous observations of imperfect symmetry in biology. Both tasks are connected to optimisation, although, as it has been shown, evolution is more than a simple optimisation process. In fact, both topics can be considered mathematically as generalisations of elementary catastrophe theory, classifying the singularities of families of smooth potentials. In structural optimisation, the non-smoothness of the potentials lead to the generalisation of the classical theory, while in case of evolutionary models the fitness functions can be regarded as a generalisation of potentials.

Not surprisingly, generalisation of elementary catastrophe theory led to results, which are not predicted by the classical theory. In case of engineering structures I identified symmetrical optima surviving beyond bifurcation points while in case of evolutionary models I could identify scenarios where the symmetrical strategy survived after the emergence of a new, asymmetrical branch. These new phenomena not only proved to be physically relevant, the two mentioned examples also indicate a strong analogy between the two studied fields.

In Chapter 2, I examined the local improvability of structures supporting a finite symmetry group Γ with respect to a number of symmetry-breaking scalar variables. If the global optimisation potential of the structures is determined by the upper envelope of several smooth local potentials (associated with points or elements of the structures) the symmetrical configuration tends to be local optimum, i.e. the perfect configuration cannot be improved by small perturbations of the symmetry in the majority of the cases. I introduced the concept of 'potential improvability' (which often implied actual improvability, see *Definition* 2.2 and the related comments) and determined the following typical condition, which is one of the Principal Results of my thesis:

P.R. I: **The sufficient and necessary condition of *potential improvability* is that the representation of the symmetry group Γ of the structure in the space of variables is not sub-representation of the regular representation of Γ.** (cf. *Theorem 2.5*, *Theorem 2.3*, and *Definition 2.4*).

I illustrated the application of this algorithm on many structural examples (Subsections 2.5, 2.6.2 and 2.7). My numerical computations show that *potential improvability* very often implies *actual improvability.*

This condition yields an easy-to-handle algorithm to decide whether a given structure can be locally improved in a given set of variables, without performing detailed calculations of the structure. The application of this algorithm has been illustrated on many structural examples

My numerical computations confirmed that *potential improvability* often yields *actual improvability.*

The above condition needs a short analysis of the representation of Γ emerging in the space of the variables (which was called induced representation). In some cases, the number of variables itself determines whether the structure is potentially improvable or not. In particular I proved the following two statements:

P.R. II.1 **The typically sufficient condition of potential improvability is $d \geq O(\Gamma)$ where $O(\Gamma)$ denotes the order of Γ** (*Theorem 2.6*)**.** This condition yields for planar reflection symmetry $d \geq 2$, in case of C_m and D_m symmetry it yields $d \geq m$ and $d \geq 2m$, respectively.

P.R. II.2 **The typically necessary condition of potential improvability is $d \geq 2\dim(\Gamma)$, where $\dim(\Gamma)$ denotes the dimension of the smallest real-valued representation of Γ, which has no trivial component** (cf. *Theorem 2.7* in Subsection 2.4.3.2 and *Definition I.19*). The necessary condition yields $d \geq 2$ for D_m symmetry and C_{2k} symmetry, in case of C_{2k+1} symmetry it yields $d \geq 4$. In case of C_2 and D_1, this condition agrees with both the necessary condition in Principal Result II.1 and the sufficient and necessary condition in Principal Result I. For C_3 symmetry, this result seems to contradict II.1 and I if the number of variables is 3. In fact this is not a contradiction, since an adequate set of variables cannot consist of 3 variables in this case.

However, I also demonstrated the existence of exceptional, atypical structures where the above conditions fail:

P.R. II.3 **In case of D_1 symmetry** (e.g. planar reflection symmetry) **there exist *special, atypical* structures which can be locally improved by using only d=1 variable (***Theorem 2.1***).** I determined the exact criteria for these special cases. Based on a special example with D_2 symmetry I demonstrated that there exist special, atypical cases (contradicting the general criteria) in other symmetry groups, as well.

Chapter 2 investigated a modified version of the basic question, as well: if the symmetric configuration is improvable by a small perturbation, the perturbed configuration is still usually not locally optimal (i.e. the bigger the perturbation is, the better the structure becomes). However one can find slightly asymmetrical optima (cf. *Definition 2.3*) in a one-parameter family of structures if asymmetrical optima bifurcate from the symmetrical configuration ($\mathbf{x}$=0). In connection with such bifurcations, I proved

P.R. III.1 **In case of C_2, D_1 symmetries** (e.g. planar reflection symmetry) and $d=1$ variable a typical, one-parameter family of structures *cannot be optimally improved*, i.e. **the typical, necessary condition of optimal improvability is $d \geq 2$.**(cf. *Theorem 2.8*). In the proof I listed the possible optimum/pessimum bifurcations and provided structural examples for each listed case.

Since the proof relies on the symmetry- and variable-specific bifurcation analysis of the optimisation diagrams, I could not extend this result to arbitrary symmetry. In addition, I found a case, which seems to contradict the potential generalisation of the above claim:

P.R. III.2 **I provided an example for a structural family and a set of variables, which cannot be improved locally, however, it can be improved optimally.** Thus, I showed that in case of some symmetry groups, optimal improvability can be achieved with a smaller number of variables than local improvability. (as opposed to D_1 symmetry).

The elements of Principal Result III put more light on the observation that imperfectly symmetrical structures are rare. At the same time, Principal Results I and II. can be applied in the engineering practice to improve a given symmetrical structure with a small number of variables. The results help to choose an adequate set of variables; this should be followed by numerical analysis, which decides if the potentially improvable structure is actually improvable or not. Although it is beyond the confines of the present work, it would be interesting to apply the method to real, large-scale engineering structures. This opens a challenging avenue of future research.

Evolution is more complex than optimisation of engineering structures and it is a spontaneous, dynamical process. Despite these basic differences, temporal evolutionary patterns are analogous to structural optimisation diagrams. The second part of my Thesis deals with patterns of the emergence of imperfect symmetry in the course of evolution, which is motivated by the fact that animals with imperfect symmetry are (unlike structures with imperfect symmetry) common in Nature.

The complex genetic background of evolution is strongly simplified by the method of Adaptive Dynamics. I applied this framework to the study of emerging asymmetry. While classifying the types of symmetry-breaking,

P.R. IV: **I introduced a novel classification of symmetry in frequency-dependent ecological models, which I called strong/weak symmetry.** I determined the corresponding symmetry constraints in the fitness functions (‘strong’ symmetry yielded a more specific constraint than ‘weak’ symmetry), and showed that the two cases produce different evolutionary behaviour (see further details in Principal Result V). I also demonstrated the difference between the two classes on several real examples (see Section 3.8).

Using the above classification, I performed a systematic approach to the evolutionary patterns of the emergence of asymmetry, and

P.R. V.1 **I listed the generic evolutionary patterns of the emergence of asymmetry in adaptive dynamics models** (see Figure 3.1), I also determined the exact conditions of the emergence of each one.

One of the emerging patterns proved to be especially interesting. More specifically,

P.R. V.2 **I demonstrated the possibility of a novel evolutionary bifurcation pattern**, in which an asymmetrical evolutionary branch develops in a population with bilateral symmetry and the new branch coexists with the symmetrical ancestors. I also simulated this pattern numerically on a classical model of Levene in Section 3.7.

The novel pattern occurs in case of strong symmetry, when a time-dependent evolutionary system moves through a degenerate state. While the model is degenerate only in a specific moment in time, the effect of the degeneracy extends over a finite time-interval, during which the model is close to be degenerate. This fact follows from the discreteness of adaptive dynamics: it consists of small but discrete evolutionary steps. Thus, degenerate evolutionary patterns may develop in the model if the environmental variation is slow, i.e. if the close-to-degenerate state lasts adequately long. This phenomenon raises the significance of environmental change:

P.R. V.3 I showed that the novel pattern occurs only in case of changing environment Hence, environmental change is an even more important ingredient in this context than in the classical AD theory, where evolutionary branching exists in autonomous models.

The unfolded list of evolutionary patterns is a set of theoretical possibilities. We have no evidence of the physical existence of these patterns (in particular the novel pattern): such an evidence could be based only on fossil data, but according to many results, fast transitional phenomena of evolution (such as branching or the sudden emergence of asymmetry) seem to be too fast to leave a remarkable fossil record. Nevertheless, theoretical modelling plays an important role in this field, repeatedly predicting phenomena, which are later verified either by experiments or by collected data.

Principal Results I, II, and III have been partially published in Várkonyi et. al. (in press). Publication of the rest of these results is in preparation (see more details in part 2.1.3). Principal Results IV, V have been published in Várkonyi et. al. (accepted for publication).

ACKNOWLEDGEMENTS

First of all, I would like to express my heartfelt thanks to my supervisor, Gábor Domokos for his tireless work, support, and encouragement. His never-ending stream of new ideas not only gave our joint work a push whenever we faced unexpected difficulties, but also made me the past three years unforgettable. His example showed me that research is more than a job yielding interesting challenge: it is also an open and fascinating approach to the world.

I am also grateful to my wife Dóri, who saw mostly the shady side of the task, namely the lack of time and me, but still, stood by me throughout. Also our son, Gáspár was highlight of many of our evenings and holidays over the past one and a half year.

Géza Meszéna was our partner and co-author in the evolutionary part. Working with him was a great pleasure, and I learnt a lot from him. I also enjoyed working with Kristóf Juhász, who has joined the structural optimisation task recently; he transmitted the cited thoughts of Thomas Mann on symmetry to me.

I thank both Dr. Tibor Tarnai and Dr. István Scheuring for their review during the 'in-house' discussion of this work. I also thank my mother Annamária Várkonyi-Kóczy for her hard work of carefully reading and commenting the manuscript. Their suggestions helped to improve the language and the readability of the text a lot.

Many other people helped me with their answers, comments and, advice. Among others, I thank Hans Metz, Éva Kisdi, Claus Rüffler, Tom van Dooren, Lajos Rónyai, Zsolt Gáspár, Kata Marótzy and Dávid Czabán for their assistance.

Completing this work would not have been possible without the support of my close colleagues, first of all of Péter Árva, who have undertaken much of my duty of education during the past months. I am thankful for all of them.

Parts of this work have been supported by OTKA funds T046646 and TS049885.

REFERENCES

Abrams, P. A., Matsuda, H., Harada, Y., 1993. Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits, *Evol. Ecol.* 7, 465-487.

Alkér, K., 2001. *Aszimmetria és optimum a szerkezettervezésben (Asymmetry and optimum in structural design, in Hungarian)*, Scientific paper presented at the annual Students' Competition, Technical University of Budapest.

Allaire, G., 2002. *Shape optimisation by the homogenization method*, Springer, New York.

Allgower, E. L., Georg, K., 1990. *Numerical continuation methods: an introduction*, Springer, Berlin,

Asami, T., Cowie, R. H., Ohbayashi, K., 1998. Evolution of mirror images by sexually asymmetric mating behaviour in hermaphroditic snails, *Am. Nat.* 152, 225-236.

Banichuk, N. V.,1990. *Introduction to optimisation of structures*, Springer, New York.

Bendsoe, M., 1995. *Methods for optimisation structural topology, shape and material*, Springer, Berlin.

Benkman, C. W., 1996. Are the ratios of bill crossing morphs in crossbills the result of frequency-dependent selction? *Evol. Ecol.* 10, 119-126.

Brown, N. A., Wolpert, L., 1990. The development of handedness in left/right asymmetry. *Development* 109(1), 1-9.

Buella, Cs., 2002. *Szerkezettervezés: aszimmetriák, analógiák és katasztrófák* (*Structural design: asymmetry analogies and catastrophes, in Hungarian)*, Scientific paper presented at the annual Students' Competition, Technical University of Budapest.

Christiansen, F. B., 1991. On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* 138, 37-50.

Cooper, J. M., Hutchinson, D. S., (Eds.), 1997. *Plato: Complete Works*, Hackett Publishing Co. Inc.

Couprie, D. L., Hahn, R., Naddaf, G., 2003. *Anaximander in Context: New Studies in the Origins of Greek Philosophy*, State University of New York Press, N.Y.

Coxeter, H. S. M., 1973. *Introduction to geometry*, 2nd Ed. Wiley & Sons Inc., New York, Hungarian Translation: Műszaki Könyvkiadó, Budapest.

Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes, *J. Math. Biol.* 34, 579-612.

Dieckmann, U., Doebeli, M. 1999. On the origin of species by sympatric speciation. *Nature* 400, 354-357.

Dieckmann, U., Doebeli, M., Metz, J. A. J., Tautz, D. (Eds.), 2004. *Adaptive Speciation*, Cambridge University Press, Cambridge.

Domokos, G., Gáspár, Zs., 1995. A global, direct algorithm for path-following and active static control of elastic bar structures, *International J. of Structures and Machines*, 23(4), 549-571.

Domokos, G., Szeberényi, I., 2001. A hybrid parallel approach to one-parameter nonlinear boundary value problems, *Computer Assisted Mechanics and Engineering Sciences*, 11, 1-20.

Eldredge, N., Gould, S.J., 1972. Punctuated equilibria: an alternative to phyletic gradualism. In *Models In Paleobiology* (Ed. by T. J. M. Schopf). Freeman, Cooper and Co., San Francisco.

Eshel, I., 1983. Evolutionary and continuous stability, *J. Theor. Biol.* 103, 99-111.

Gavrilets, S., 2005. 'Adaptive speciation': it is not that simple, *Evolution* 59, 696-699 .

Gere, J. M. and Timoshenko, S. P., 1990. *Mechanics of Materials*, 3rd edition, PWS-Kent, Boston.

Geritz, S. A. H., Kisdi, É., Meszéna, G., Metz, J. A. J., 1997. The dynamics of adaptation and evolutionary branching, *Phys. Rev. Lett.* 78, 2024-2027. URL http://angel.elte.hu/~geza/GeritzPRL.pdf

Geritz, S. A. H., Kisdi, É., Meszéna, G., Metz, J. A. J., 1998. Evolutionary singular strategies and the evolutionary growth and branching of the evolutionary tree, *Evol. Ecol.* 12, 35-57.

Geritz, S. A. H., Kisdi, É., Meszéna, G., Metz, J. A. J., 2004. Adaptive Dynamics of Speciation. In: *Adaptive Speciation*, Dieckmann, U., Doebeli, M., Metz, J. A. J., Tautz, D. (Eds.) Cambridge University Press, Cambridge.

Golubitsky, M., Stewart I., Schaeffer, D. G., 1982. *Singularities and Groups in Bifurcation Theory*, Vol. 1., Springer Verlag, New York.

Govind, C.K., 1989. Asymmetry in lobster claws, *Am. Sci.* 77, 468-474.

Hargittai, I., Hargittai, M., 1994, *Symmetry: a unifying concept*, Shelter Publishing Co, Bolinas.

Hemp, W., 1973. *Optimum structures*, Clarendon Press, Oxford.

Hofbauer, J., Sigmund, K.,1998. *Evolutionary games and population dynamics*, Cambridge Univ. Press, Cambridge.

Jones, H. F., 1998. *Groups, representations and physics*, 2nd ed. IOP Publishing Ltd, London.

Korn, G. A., Korn, T. M., 1968. *Mathematical handbook for engineers*, 2nd edition, McGraw-Hill, New York.

Kunze, M., Jakob-Rost, L., 1992. *Pergamon Museum*, Verlag Philipp von Zabern, Berlin.

Levene, H., 1953. Genetic equilibrium when more than one niche is available, *Am. Nat.* 87, 331-333.

Malik, S. C., 1992. *Mathematical analysis*, 2nd ed., Wiley, New York.

Mann, T., 1928. *The magic mountain*, (Translated by H. T. Lowe-Porter.) Pinguin Edition, Secher&Warburg, London.

Marótzy, K., 2005. *Personal communication*.

Maynard-Smith, J., 1982. *Evolution and the theory of games*, Cambridge University Press, Cambridge.

Meszéna, G., Czibula, I., Geritz, S. A. H., 1997. Adaptive dynamics in a 2-patch environment: a toy model for allopatric and parapatric speciation. *J. Biol. Syst.* 5, 265-284. URL http://angel.elte.hu/~geza/MeszenaEtal1997.pdf

Meszéna, G., Kisdi, É., Dieckmann, U., Geritz, S. A. H., Metz, J. A. J., 2001. Evolutionary optimisation models and matrix games in the unified perspective of adaptive dynamics. *Selection* 2, 193-210.

Meszéna, G., Gyllenberg, M., Jacobs, F. J., Metz, J. A. J., 2005. Link between population dynamics and dyanamics of Darwinian evolution. *Physical Review Letters* 95(7), 78-105. URL http://angel.elte.hu/~geza/PhysRevLett_95_078105.pdf

Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A., van Heerwaarden, J. S., 1996. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction I, In: S.J. van Strien, S.M. Verduyn Lunel (Eds.) *Stochastic and spatial structures of dynamical systems*, North Holland, Elsevier, pp. 183-231.

Mboko, S. K., Kohda, M., Hori, M., 1998. Asymmetry of mouth-opening of a small herbivorous Cichlid fish Telmatochromis Temporalis in lake Tanganyika, *Zool. Sci.* 15, 405-408.

Moore, J., 2001. *An introduction to invertebrates.* Cambridge University Press, Cambridge.

Nakajima, M., Matsuda, H., Hori, M., 2004. Persistence and fluctuation of lateral dimorphism in fishes. *Am. Nat.* 163, 692-698.

Petrakos, B*.,* 1993. *National Museum*, Clio, Athens.

Poston, T., Stewart, J., 1978. *Catastrophe theory and its applications*, Pitman, London.

Preziosi, D., 1998. *The art of art history: a critical anthology*, Oxford University Press, Oxford.

Purves, W. K., Sadava, D., Orians G. H., Heller, H. C., 2003. *Life: the science of biology*, Sinauer Associates, Stanford.

Raup, D. M., 1962. Computer as aid in describing form in gastropod shells, *Science* 138, 150-152.

Rosen, J., 1995. *Symmetry in Science: an introduction to the general theory*, Springer-Verlag, New York.

Rozvány, G., 1989. *Structural design via optimality criteria*, Kluwer Academic Publishers, Dordrecht.

Schluter, D., 2001. Ecology and the origin of species, *Trends Ecol. Evol.* 16, 372-380

Schreiber, S., J., Tobiason, G. A., 2003. The evolution of resource use, *J. Math. Biol.* 47, 56-78

Sokolowski, J., Zolesio, J.P., 1992. *Introduction to shape optimisation. Shape sensitivity analysis*, Springer Series in Computational Mathematics, Springer, Berlin.

Takahashi, S., Hori, M., 1994. Unstable evolutionary stable strategy and oscillation: a model of lateral asymmetry in scale-eating cichlids, *Am. Nat.* 144, 1001-1020.

Taylor, P. D., 1989. Evolutionary stability in one-parameter models under weak selection. *Theor. Pop. Biol.* 36, 125-143.

Thompson, J. M. T. and Hunt, G. W., 1973. *A General Theory of Elastic Stability*, Wiley, London.

Thompson, J. M. T., Hunt, G. W., 1984. *Elastic Instability Phenomena*, Wiley, Chichester.

Turelli, M., Barton, N. H., Coyne, J. A., 2001. Theory of speciation, *Trends Ecol. Evol*. 16, 330-343

Várkonyi P.,L., Domokos G., in press. Symmetry, optima and bifurcations in structural design, *Nonlinear Dynamics*.

Várkonyi, P. L., Meszéna, G., Domokos G., accepted for publication. Emergence of asymmetry in evolution. *Theor. Pop. Biol.*

~~Via, S., 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* 16, 381-390.~~

Waxman, D., Gavrilets, S., 2005. 20 questions on Adaptive Dynamics: a target review. *J. Evolution. Biol.* 18, 1139-1154.

Weyl, H., 1989. *Symmetry.* Princeton Univ. Press, Princeton.

APPENDIX I REPRESENTATION THEORY

This section contains a brief summary of some basic results of group representation theory. My aim is not a systematic description of the theory, rather the introduction of only the elements, which are necessary to understand the results of this work. More general descriptions can be found in text books such as Coxeter (1973), Jones (1998). The latter one served as source of sections I.1-5, with minor modifications in the notations and formulations. On the other hand, Sections I.4-7 contain more specific results, which are according to my knowledge not of primary interest in representation theory, but play key role in this work. In these subsections sketchy proofs are also attached.

Part I.1 is an introduction to groups, while the specific groups emerging in engineering problems are listed in I.2. Part I.3.1 defines representations. I.3.2 is devoted to the most basic results of representation theory, which allow to create a unique decomposition of representations to the direct sum of a few simple ones. In I.3.3 a special kind of representation is analysed, which plays important role in structural optimisation. Part I.4 deals with the orbit of vectors with respect to a representation. Orbits emerge explicitly in the conditions of improvability of structures (see *Lemma 2.3*). Finally, differences between complex- (for which classical results apply for) and real-valued representations (which emerge in the engineering problems) are collected in I.5.

I.1 GROUPS

*Definition I.1: a group is a finite or infinite set Γ of elements and a binary operation * ('group operation'), with the following four properties:*

- *closure of group operation: if $\gamma,\eta \in \Gamma$, $\gamma*\eta \in \Gamma$*
- *associativity of group operation: if $\varphi,\gamma,\eta \in \Gamma$, $(\varphi*\gamma)*\eta=\varphi*(\gamma*\eta)$*
- *existence of identity element: there is an $i \in \Gamma$, so that for any $\gamma \in \Gamma$, $\gamma*i=i*\gamma=\gamma$*
- *existence of inverse: for any $\gamma \in \Gamma$, there is a group element γ^{-1} satisfying $\gamma*\gamma^{-1}=i$.*

The simplest ('trivial') group has one single identity element i and the corresponding group operation acts as $i*i=i$. There is one group with two elements $\{i,\gamma\}$, for which $i*i=\gamma*\gamma=i$ and $i*\gamma=\gamma*i=\gamma$. Though a group might be infinite as well, we mainly study finite groups. A finite group can be conveniently characterised by a table, which contains the effect of the group operation on the group elements. This form of the two-element group and another one with four elements are shown in Table I.1. Notice that each row and column is a permutation of the group elements; the identity element of the group corresponds to an unperturbed row as well as an unperturbed column.

Two seemingly different groups may have the same structure according to:

Definition I.2: two groups Γ and Λ are isomorphic ($\Gamma \equiv \Lambda$) if there is a one-to-one correspondence between their elements $\gamma_i \Leftrightarrow \lambda_i$ so that $\gamma_j*\gamma_k=\gamma_l$ if and only if $\lambda_j*\lambda_k=\lambda_l$

Isomorphic groups are often considered as identical. These classes of groups are called 'abstract groups' and are fully defined by a group table. On the other hand, two isomorphic groups may consist of rather different elements and operators. A number of isomorphic groups (all belonging to the abstract group of Table I.1/B) are collected in Table I.2.

Two further basic definitions of group theory are

Definition I.3: The order r of a group Γ is the number of elements in the group.

Definition I.4: An abstract group $\Lambda=\{\lambda_i\}$ is subgroup of the abstract group $\Gamma=\{\gamma_i\}$ if there is a homomorphism $H: \Gamma\rightarrow\Lambda$, which satisfies $H(\gamma_j)*H(\gamma_k)=H(\gamma_l)$ if and only if $\gamma_j*\gamma_k=\gamma_l$. A real group Λ is subgroup of the real group Γ if its elements are a subset of the elements of Γ and the two group operations have the same effect on the elements of Λ.

.

	a_0	a_1
a_0	a_0	a_1
a_1	a_1	a_0

A

	a_0	a_1	b_0	b_1
a_0	a_0	a_1	b_0	b_1
a_1	a_1	a_0	b_1	b_0
b_0	b_0	b_1	a_0	a_1
b_1	b_1	b_0	a_1	a_0

B

Table I.1.: Group tables of the two-element abstract group (A) and a four-element abstract group (B).

elements	group operation
$\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}, \begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}, \begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix}, \begin{bmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix}$	multiplication
0,1,2,3	addition modulo 3
1, -1, i, -i	multiplication
the invariant euclidean 2D transformations of a rectangle (identity, rotation by π, reflection with respect to two lines)	product of transformations
the invariant euclidean transformations of an oriented rectangle embedded in 3D space. (identity, rotation by π, reflection to the plane of the rectangle and reflection to the centre of the rectangle)	product of transfomations

Table I.2. Different groups, which belong to the four-element abstract group of Table I.1/B. The fourth one is a symmetry group called D_2 (see Table I.3), the fifth one is also symmetry group. The first one is the regular representation of D_2 (see Section I.3.3)

In this work, we concentrate on *symmetry groups*, i.e. groups with euclidean transformations as elements and the product of these transformations as group operation. There are symmetry groups in many abstract groups, e.g. rotation by 180° around a point in 2D space and identity

transformation (called C_2) belong to the two element abstract group of Table I.1/A. The last two examples in Table I.2 are also symmetry groups.

Some geometric objects (such as structures) are called 'symmetric' (e.g. reflection symmetric). With group theoretical terms, this means that these objects are invariant to a set of euclidean transformations (e.g. reflection and identity in case of reflection symmetry). Such a set of invariant transformations always forms a symmetry group Γ. This yields a more precise definition of symmetry:

Definition I.5: A geometrical object is 'Γ-symmetrical' if there is a symmetry group Γ, the elements of which are invariant transformations of the object Γ will also be referred to as the symmetry of the object in question.

I.2 SYMMETRIES OF REAL ENGINEERING STRUCTURES

In this part, we introduce the specific groups, which occur as symmetries of real engineering structures. Since our goal is to optimise structures with respect to their *inner forces* (generated by external loads), we modify the *Definition I.5* as:

Definition I.6: A load-bearing structure is 'Γ-symmetrical' if its geometry, loads and inner forces are invariant to the elements of a symmetry group Γ.

Notice that real examples on ground are primarily subjected to *gravitation*, which determines the special 'down' direction. The symmetry transformations of such structures must preserve this direction (i.e. the image of a vertical vector pointing down is also pointing down). Among the euclidean transformations of the 3D space, this property is owned by transformations, which map a point (x,y,z) to $(f_x(x,y),f_y(x,y),z+c)$ (z stands for the vertical co-ordinate, c is constant). In case of *finite* symmetry groups, c must be 0, otherwise repeating the transformation would never end up in identity transformation, i.e. it would generate infinitely many group elements. Thus, the symmetry transformation do not modify z, they are '2-dimensional'. Hence, only the symmetry groups observed in *two-dimensional space* can emerge as symmetries of real structures. If we had e.g. a regular tetrahedron shaped framework (this kind of symmetry cannot be observed in 2D space), its internal forces (generated mostly by gravity) would break the symmetry of the tetrahedron. The resultant symmetry of the structure (in the sense of *Definition I.6*) would be reduced to a subgroup of the original one. Conversely, cosmic structures (such as satellites) are likely to have '3-dimensional' symmetries, like those of the Platonian solids.

In the following, we introduce the finite symmetry groups of the two-dimensional space. The two dimensional euclidean transformations are

- identity
- shifting by an arbitrary vector
- rotation by an arbitrary angle
- reflection to an arbitrary line

These transformations form two distinct types of finite groups (and many infinite ones which are not discussed here). Basic properties of the *cyclic groups* (denoted by C_n, $n\geq1$) and the *dihedral groups* (D_n, $n\geq1$) are collected in Table I.3. Two of these symmetry groups (C_1 and

D_2) are isomorphic, i.e. they belong to the same abstract group. From the point of view of structural optimisation, C_2 and D_1-symmetrical examples will behave in the same way.

name of the group	C_n ($n\geq1$)	D_n ($n\geq1$)
order	n	$2n$
notation of elements	$a_0,a_1,...,a_{n-1}$	$a_0,a_1,...,a_{n-1}$ $b_0,b_1,...,b_{n-1}$
meaning of elements in 2D space	a_i is rotation by $2i\pi/n$ around the point (0,0)	a_i is rotation by $2i\pi/n$ around a point (0,0); b_i is reflection to the line $y\cos(\pi i/n) = x\sin(\pi i/n)$
identity element	a_0	a_0
effect of group operation	$a_i*a_j=a_{i+j \bmod(n)}$	$a_i*a_j=a_{i+j \bmod(n)}$ $a_i*b_j=b_{j-i \bmod(n)}$ $b_i*a_j=b_{i+j \bmod(n)}$ $b_i*b_j=a_{j-i \bmod(n)}$
planar objects with this symmetry	for n=1: asymmetrical objects for n=2: paralelogramma; for n>2: oriented regular n-gon	for n=1: deltoid for n=2: rectangle; for n>2: non-oriented regular n-gon

Table I.3: The finite symmetry groups of the 2D space. The two columns correspond to two classes of groups. Row 3 determines the group elements (the exact transformations may appear different in different coordinate systems, the transformations in row 3 provide an example) Row 5 defines the group table of these groups in a condensed form. C_1 is the trivial group, which corresponds to the lack of symmetry. We remark that C_2 (180° rotation symmetry) and D_1 (reflection symmetry) are isomorphic, they belong to the same abstract group

I.3 REPRESENTATIONS

I.3.1 Definition, Basic properties

In structural optimisation, the type of the symmetry-breaking variables is characterised by the representation of the symmetry transformations of the perfect structure in the space of the variables. A representation of a group Γ is defined as

Definition I.7: a representation of group d *is a homomorphism of the elements* γ_i *of* Γ *to complex square matrices* $\gamma_i\rightarrow\mathbf{D}_i$*, for which* $\gamma_j*\gamma_k=\gamma_l$ *implies* $\mathbf{D}_j\mathbf{D}_k=\mathbf{D}_l$.

The dimension of a representation is straightforward: the order of the matrices in the representation.

It can happen that the same matrix belongs to two different group elements in a representation. In particular, every group has

Definition I.8: The trivial representation is the homomorphism $\gamma_i\rightarrow1$,

in which every group element is mapped to the *to* 1×1 identity matrix. On the other hand,

Definition I.9: a representation is called faithful if different group elements are mapped to different transformations.

Notice that a faithful representation of Γ itself forms a group, which is isomorphic to Γ. The representations emerging in engineering optimisation problems are always faithful as a consequence of condition (i).

Representations can be classified according to

Definition I.10: Two representations are equivalent if there is a (complex-valued) unitary transformation, which transforms elements of one to the elements of the other one. We denote equivalence of two representations as $D_1 \equiv D_2$.

Equivalent representations can be considered as the same representation in different coordinate systems. Equivalence classes of representations can conveniently be characterised by

Definition I.11: The character of a representation $\mathbf{D}_1$, $\mathbf{D}_2$,..., $\mathbf{D}_r$ *is a vector of length r, namely* $[\, tr(\mathbf{D}_1)\ tr(\mathbf{D}_2)...tr(\mathbf{D}_r)]$, *where* $tr(\mathbf{X})$ *denotes the trace of the matrix* $\mathbf{X}$.

The characters of equivalent representations are equal, because the trace of a matrix is equal to the sum of its eigenvalues, which are invariant to unitary transformations. On the other hand, it can be proven that two representations with equal characters are necessarily equivalent. Thus, the character itself determines the equivalence class of a representation.

If we restrict ourselves to the representations of *finite* groups, we can define the most basic property of representations as

Definition I.12: A representation D is reducible if there exists a representation D' for which $D \equiv D'$ *and D' consists of block-diagonal matrices (each of which has the same block-structure). Another equivalent definition of reducibility is presented later. Irreducible representations are often called simply irreps.*

According to *Definition I.10*, an appropriate unitary transformation decomposes a reducible transformation to the direct sum of irreps. What is more, the irreducible components in two different decompositions of a representation are equivalent, i.e. *the decomposition is unique* up to the level of equivalence classes. This situation is somewhat similar to the prime factorisation of integers. As an example of the decomposition, consider a unitary transformation of three rotating matrices (the middle ones on the left side of the equations), which are a representation of C_3:

$$\begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}\mathrm{i} \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}\mathrm{i} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}\mathrm{i} \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}\mathrm{i} \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{I.1}$$

$$\begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{1}{2}+\frac{\sqrt{3}}{2}i & 0 \\ 0 & -\frac{1}{2}-\frac{\sqrt{3}}{2}i \end{bmatrix} \quad \text{(I.2)}$$

$$\begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \\ \frac{\sqrt{2}}{4} & -\frac{\sqrt{2}}{4}i \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{1}{2}-\frac{\sqrt{3}}{2}i & 0 \\ 0 & -\frac{1}{2}+\frac{\sqrt{3}}{2}i \end{bmatrix} \quad \text{(I.3)}$$

The decomposition shows that this representation is reducible, namely the direct sum of two one-dimensional irreps. The technique of creating this kind of decomposition is introduced in the next part.

Any group has infinite number of representations, but a restricted number of irreducible ones, according to

Theorem I.1 (Dimensionality theorem): If a group of order r has n_k equivalence classes of k-dimensional irreducible representations, $n_1^2 + n_2^2 + \ldots = r$.

The list of irreducible representations can be constructed for simple groups by techniques, which are not discussed here. For us, the symmetry groups of the 2D space are of special interest (see Section I.2). The lists of their irreps are collected in Table I.4.

symmetry group		irreducible representations		
name	order	name	dimension	character
trivial	1	I_0	1	{1} (trivial representation)
C_n	n	I_l	1	$\{ a_k: (-1)^{2lk/n} \}$ with $0 \le l \le n-1$ (trivial representation if $l=0$)
D_n if n is even	$2n$	I_0	1	$\{ a_k: 1, b_k: 1\}$ (trivial representation)
		I_l	2	$\{ a_k: 2\cos(2lk\pi/n), b_k: 0\}$, with $1 \le l \le n/2-1$
		$I_{n/2}$	1	$\{ a_k: 1, b_k: -1\}$
		$I_{n/2+1}$	1	$\{ a_k: -1^k, b_k: -1^k \}$
		$I_{n/2+2}$	1	$\{ a_k: -1^k, b_k: -1^{k+1}\}$
D_n if n is odd	$2n$	I_0	1	$\{ a_k: 1, b_k: 1\}$ (trivial representation)
		I_l	2	$\{ a_k: 2\cos(2lk\pi/n), b_k: 0\}$, $1 \le l \le (n-1)/2$
		$I_{(n+1)/2}$	1	$\{ a_k: 1, b_k: -1\}$

Table I.4: Irreducible representations of the 2-dimensional symmetry groups. The meaning of the notations a_k, b_k can be found in Table I.3. Notice that the characters might be complex numbers.

I.3.2 Decomposition of representations

The decomposition of reducible representation plays a primary role in our work. The process is introduced through the example of equations (I.1)-(I.3). As already mentioned, these three matrices are a representation of the group C_3. The characters of the irreps of C_3 are (cf. Table I.4):

group elements	a_0	a_1	a_2

characters of the irreps	I_0: χ_0=	[1 1 1]
	I_1: χ_1=	[1 $(-1)^{2/3}$ $(-1)^{4/3}$]
	I_2: χ_2=	[1 $(-1)^{4/3}$ $(-1)^{2/3}$]

The character of the representation is the trace of the three matrices, i.e. χ= [2 –1 –1]. Hence the trace of a block diagonal matrix is the sum of the traces of its blocks, we only have to produce χ as a linear combination of χ_0, χ_1, χ_2:

$$\chi = \sum_{k=0}^{2} n_k \chi_k \tag{I.4}$$

where the resulting coefficients $n_k \in \mathbb{N}$ indicate the number of the three irreps in the decomposition of the representation ($\mathbb{N}$ stands for the set of natural numbers). The solution exists and it is unique. In our case $[n_0\ n_1\ n_2]$ =[0 1 1]. This is the awaited result, hence equations (I.1)-(I.3) already presented the decomposition of the representation to the direct sum of these two irreps.

I.3.3 The regular representation

All groups, and in particular symmetry groups, have a special representation. Consider the group table of a group Γ of order r (see for example Table I.1). The i^{th} row contains a permutation of the group elements. The corresponding $r \times r$ size permutation matrices $\mathbf{R}_i$ form a representation of Γ.

Definition I.13: The representation R, which consists of the above defined matrices $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_r$ *is called the regular representation of* Γ.

As an example see the regular representation of D_2 (Table I.1/B) in the top row of Table I.2. One can show that each k-dimensional irrep of a group Γ appears k times in the decomposition of the regular representation of Γ. (This is in accordance with the *Dimensionality theorem* (*Theorem I.1*), since the regular representation is r-dimensional.)

I.4 ORBITS

In this part the orbit of a vector with respect to a representation is introduced. Notice that the typical condition of potential improvability (*Lemma 2.3*) applies to orbits, which are defined as:

Definition I.14 : The vectors $\{\mathbf{D}_1\mathbf{v}, \mathbf{D}_2\mathbf{v}, \ldots, \mathbf{D}_r\mathbf{v}\}$ *are called the orbit of* $\mathbf{v}$ *with respect to the representation* $D=\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_r\}$.

Before going into details, some new definitions are needed:

Definition I.15: S is an invariant subspace of a representation $D=\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_r\}$ *if* $\mathbf{v} \in S$ *implies,* $\mathbf{D}_i\mathbf{v} \in S$ *for every* $1 \le i \le r$.

Definition I.16: Invariant points of a representation $D=\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_r\}$ *are the vectors* $\mathbf{v}$*, for which* $\mathbf{D}_i\mathbf{v}=\mathbf{v}$ *for every* $1 \le i \le r$.

These concepts are used to give another definition of reducibility.

Definition I.17: A representation is reducible iff it has a non-trivial invariant subspace.

The new definition is equivalent of *Definition I.12*: consider the block-diagonal form of a reducible representation $\{\mathbf{D}_i=\mathrm{diag}(\mathbf{D}_i^{(1)},\mathbf{D}_i^{(2)}),\ i=1,2,...,r\}$, where the sizes of $\mathbf{D}_i^{(1)}$and $\mathbf{D}_i^{(2)}$ are $k\times k$ and $(r-k)\times(r-k)$. If a vector is of the form $\mathbf{v}=[v_1\ v_2...\ v_k\ 0\ 0...0]^T$, then $\mathbf{D}_i\mathbf{v}$ inherits this form, thus we have found an invariant subspace. Conversely, the existence of invariant subspaces yields the block-diagonality of the representation in an appropriate co-ordinate system. The proof of the latter statement is not discussed here.

The following two subsections focus on two properties of orbits.

I.4.1 Orbits and invariant points

The point $\mathbf{x}=0$ is the trivial invariant point of any representation, hence $\mathbf{M}\cdot 0=0$ for arbitrary matrix $\mathbf{M}$. However, some representations have additional invariant points, as well.

Assume that we have a $\mathbf{v}\neq 0$ invariant point of $D=\{\mathbf{D}_i,\ i=1,2,...,r\}$. Consider a transformation matrix $\mathbf{T}$, which moves $\mathbf{v}$ into $\mathbf{Tv}=[1\ 0\ 0...0]$. This vector is an invariant point of $\mathbf{T}D\mathbf{T}^{-1}=\{\mathbf{T}\mathbf{D}_i\mathbf{T}^{-1},\ i=1,2,...,r\}$. Consequently the matrices $\mathbf{T}\mathbf{D}_i\mathbf{T}^{-1}$ are all of the form

$$\mathbf{T}\mathbf{D}_i\mathbf{T}^{-1}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & * & * & *\\ \vdots & * & * & *\\ 0 & * & * & *\end{bmatrix}, \tag{I.5}$$

i.e. there is a trivial representation among the irreducible components of the representation $\mathbf{T}D\mathbf{T}^{-1}$, which is equivalent of D. Conversely, a trivial component implies the existence of invariant points. Thus we can formulate

Lemma I.1: A representation has non-trivial invariant points iff it has a trivial component.

Now we can continue with some properties of orbits. The sum of the elements of an orbit is an invariant point, hence

$$\mathbf{D}_i\mathbf{D}_1\mathbf{v}+\mathbf{D}_i\mathbf{D}_2\mathbf{v}+...+\mathbf{D}_i\mathbf{D}_r\mathbf{v}=\mathbf{D}_{i_1}\mathbf{v}+\mathbf{D}_{i_2}\mathbf{v}+...+\mathbf{D}_{i_r}\mathbf{v}, \tag{I.6}$$

where $i_1,i_2,...,i_r$ are a permutation of $1,2,...,r$. Thus,

Lemma I.2: If a representation $D=\{\mathbf{D_1},\ \mathbf{D_2},...,\mathbf{D}_r\}$ has no trivial component, the sum of the elements of the orbit of any vector $\mathbf{v}$ *is*

$$\mathbf{D}_1\mathbf{v}+\mathbf{D}_2\mathbf{v}+...+\mathbf{D}_r\mathbf{v}=0, \tag{I.7}$$

As a consequence of *Lemma I.2*, 0 is convex combination (linear combination with positive coefficients) of the vectors $\mathbf{D}_i\mathbf{v}$, which gains importance in structural optimisation. However, the same property is usually not true for representations with trivial components. To demonstrate this, consider a representation with a trivial component, which has the form of the right side of eq. (I.5) and a vector $\mathbf{v}=[v_1\ v_2...v_r]$, with $v_1\neq 0$. Each element of the orbit of $\mathbf{v}$ is of the form $\mathbf{D}_i\mathbf{v}=[v_1\ *...*]$. Thus,

Lemma I.3: If $\mathbf{v}$ *is a vector, representation* $D=\{\mathbf{D}_1, \mathbf{D}_2,...,\mathbf{D}_r\}$ *has a trivial component, and the constants* $c_1,c_2,...,c_r$ *satisfy* $c_1+c_2+...+c_r \neq 0$*, then*

$$c_1\mathbf{D}_1\mathbf{v} + c_2\mathbf{D}_2\mathbf{v} + ... + c_r\mathbf{D}_r\mathbf{v} \neq 0\,, \tag{I.8}$$

typically.

According to *Lemma I.3*, zero typically cannot be generated as a convex combination of the vectors $\mathbf{D}_i\mathbf{v}$ in this case.

I.4.2 Dimensionality of orbits

In this part, we focus on the question, under which conditions the orbit of a d-dimensional vector $\mathbf{v}$ spans the d-dimensional complex space $\mathbb{C}^d$. The first, trivial fact is that the orbit consists of r vectors, thus $r \geq d$ is a necessary condition. Another trivial requirement is that $\mathbf{v} \neq 0$. We find more precise conditions in the following.

The orbit of a vector either spans $\mathbb{C}^d$ or an S subspace of it. Assume the latter one. Then, any vector $\mathbf{w} \in S$ can be generated as the linear combination of the elements of the orbit, i.e.

$$\mathbf{w} = \sum_{k=1}^{r} c_k \mathbf{D}_k \mathbf{v} \qquad c_k \in \mathbb{C} \tag{I.9}$$

If $\mathbf{w}$ is multiplied by any element $\mathbf{D}_i$ of the representation, the result

$$\mathbf{D}_i\mathbf{w} = \sum_{k=1}^{r} c_k \mathbf{D}_i\mathbf{D}_k \mathbf{v} \qquad c_k \in \mathbb{C} \tag{I.10}$$

is again a linear combination of the orbit elements, hence $\{\mathbf{D}_i\mathbf{D}_k\ k=1,2,...,r\}$ is a permutation of $\{\mathbf{D}_k\ k=1,2,...,r\}$. Thus, $\mathbf{D}_i\mathbf{w} \in S$, which means that *$S$ is an invariant subspace of the representation*. Hence irreducible representations have no non-trivial invariant subspaces in $\mathbb{C}^d$ (cf. *Definition I.17*)

Lemma I.4: If D is an irreducible d-dimensional representation and $\mathbf{v} \neq 0$ *is a d-dimensional vector, the orbit of* $\mathbf{v}$ *with respect to D spans* $\mathbb{C}^d$.

In the following we study reducible representations. Consider the irreducible decomposition of D. We prove

Lemma I.5: If there exists a d_1-dimensional irreducible component $D^{(1)}$, which emerges at least d_1+1times in the d-dimensional representation D, then the orbit of an arbitrary real vector $\mathbf{v}$ *with respect to D does not span* $\mathbb{C}^d$.

Proof of lemma: let the elements of D and $D^{(1)}$ be denoted by $\mathbf{D}_j$ and $\mathbf{D}_j^{(1)}$ $j=1,2,..,r$, respectively. Without loss of generality, we can assume that $\mathbf{D}_j$ is in block-diagonal form, i.e.

$$\mathbf{D}_j = \begin{bmatrix} \mathbf{D}_j^{(1)} & & & & \\ & \ddots & & & \\ & & \mathbf{D}_j^{(1)} & & \\ & & & * & * \\ & & & * & * \end{bmatrix} \Bigg\} d_1 + 1 \tag{I.11}$$

Consider an arbitrary *d*-dimensional vector in the form

$$\mathbf{v} = \left[\mathbf{v}_1{}^T \quad \mathbf{v}_2{}^T \quad \cdots \quad \mathbf{v}_{d_1+1}{}^T \quad * \quad *\right]^T, \tag{I.12}$$

where the $\mathbf{v}_j$-s are d_1-dimensional vectors. Hence d_1+1 d_1-dimensional vectors are linearly dependent, they have a non-trivial zero linear combination:

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ...c_{d_1+1}\mathbf{v}_{d_1+1} = 0 \qquad c_i \in C, \tag{I.13}$$

This equation yields

$$c_1\mathbf{D}_j{}^{(1)}\mathbf{v}_1 + c_2\mathbf{D}_j{}^{(1)}\mathbf{v}_2 + ...c_{d_1+1}\mathbf{D}_j{}^{(1)}\mathbf{v}_{d_1+1} = 0 \qquad \text{for any arbitrary } j \quad . \tag{I.14}$$

The elements of the orbit of **v** are

$$\mathbf{D}_j\mathbf{v} = \begin{bmatrix} \mathbf{D}_j^{(1)} & & & & \\ & \ddots & & & \\ & & \mathbf{D}_j^{(1)} & & \\ & & & * & * \\ & & & * & * \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_{d_1+1} \\ * \\ * \end{bmatrix} = \begin{bmatrix} \mathbf{D}_j^{(1)}\mathbf{v}_1 \\ \vdots \\ \mathbf{D}_j^{(1)}\mathbf{v}_{d_1+1} \\ * \\ * \end{bmatrix} \tag{I.15}$$

Due to equation (I.14), every element of the orbit is in the subspace

$$c_1\left[x_1 \quad x_2\right] + c_2\left[x_3 \quad x_4\right] + ...c_{d_1+1}\left[x_{2d_1+1} \quad x_{2d_1+2}\right] = 0 \quad , \tag{I.16}$$

where $x_i \in \mathbb{C}$ denotes the i^{th} coordinate of a point in the *d*-dimensional complex space. Q.e.d.

Notice that *Lemma I.5* does not apply to the regular representation of a group, because it contains each irreducible d_1-dimensional representation exactly d_1 times (see before). Now we show

Lemma I.6: Let R denote the regular representation of a group Γ of order r. There exists a vector **v***, the orbit of which with respect to R spans* $\mathbb{C}^r$.

Proof of *Lemma I.6*: Hence the regular representation consists of permutation matrices, each element of the orbit of $\mathbf{v}=[1\ 0\ 0...0]^T$ is a vector with one '1' and r-1 '0'-s. Each element contains the '1' in a different positions, due to the fact that every column of a group table contains each group element only ones (cf. the definition of the regular representation). Thus, the orbit of **v** is a permutation of the following vectors: [1 0 0...0],[0 1 0....0],...,[0 0...0 1]. These vectors span $\mathbb{C}^r$. Q.e.d.

We remark without exact proof, that *Lemma I.6* states practically more than the existence of *one adequate vector* **v**: it also yields

Lemma I.6/A: Let R denote the regular representation of a group Γ of order r. The orbit of a typical vector **v** *with respect to R spans* $\mathbb{C}^r$.

The statement of *Lemma I.6/A* is somewhat obvious if we consider the orbit of **v** as rows of an $r \times r$ square matrix **O**(**v**). The orbit does not span $\mathbb{C}^d$ if and only if peig[**O**(**v**)]=0, where peig() means product of eigenvalues. Such an equation is either identity (i.e. true for arbitrary **v**) or it is typically not true. Because of *Lemma I.6,* only the latter one is possible.

Using

Definition I.18: If D and E are representations of a group Γ, we call E sub-representation of D, if their irreducible decompositions contains $d_1,d_2,...,d_k$ and $e_1,e_2,...,e_k$ examples of the irreducible representations of Γ and for every $1 \le i \le k$, $e_i \le d_i$. This an analogue of the divisors of integers.

we can generalise *Lemma I.6*/A as

Lemma I.7: Let R denote the regular representation of a group Γ of order r and let R' be a d-dimensional sub-representation of R. The orbit of a typical vector $\mathbf{v}'$ *with respect to R' spans* $\mathbb{C}^d$.

Proof of Lemma I.7: Let $\mathbf{R}_i$ and $\mathbf{R}_i$' denote the elements of *R* and *R*'. Then, $\mathbf{R}_i$ can be transformed by a unitary transformation to the following block-diagonal form:

$$\mathbf{T}\mathbf{R}_i\mathbf{T}^{-1} = \begin{bmatrix} \mathbf{R}_i' & \\ & * \end{bmatrix}, \tag{I.17}$$

An arbitrary vector $\mathbf{v} \in \mathbb{C}^r$ can be decomposed as

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}' \\ * \end{bmatrix}, \tag{I.18}$$

where $\mathbf{v}$' is *d*-dimensional. The orbit of $\mathbf{Tv}$ with respect to $\mathbf{TRT}^{-1}$ typically spans $\mathbb{C}^r$ (*Lemma I.6*/A). Since

$$\mathbf{T}\mathbf{R}_i\mathbf{T}^{-1} \cdot \mathbf{T}\mathbf{v} = \mathbf{T}\begin{bmatrix} \mathbf{R}_i' & \\ & * \end{bmatrix}\begin{bmatrix} \mathbf{v}' \\ * \end{bmatrix} = \mathbf{T}\begin{bmatrix} \mathbf{R}_i'\mathbf{v}' \\ * \end{bmatrix}, \tag{I.19}$$

the vectors $\mathbf{R}_i$'$\mathbf{v}$' also span $\mathbb{C}^d$ typically. Q.e.d.

Lemma I.5 and *Lemma I.7* can be united in

Lemma I.8: Let D and R denote a d-dimensional representation and the regular representation of a group Γ, respectively. The orbit of a typical vector $\mathbf{v}$ *with respect to D spans* $\mathbb{C}^d$*, iff D is a sub-representation of R.*

I.5 REAL-VALUED REPRESENTATIONS

All results of representation theory introduced so far, hold for *complex-valued representations*, however, in structural optimisation problems, the emerging representations are necessarily real-valued. Some of the results can be applied in a different form for real-valued representations. As main difference, some representations are reducible, however any of their decompositions contains complex entries. This is the case obviously at the example of equations (I.1)-(I.3), because the characters of its two components are complex-valued, however the traces of real-valued matrices are always real. Such a representation is irreducible among real-valued representations.

Definition I.19: A (real-valued) representation is called half-irreducible if it is either irreducible or any of its decompositions contains complex elements.

This definition is again equivalent of having no non-trivial invariant subspace, but this time in $\mathbb{R}^d$. The list of half-irreducible representations of a finite group is again finite, although the *Dimensionality theorem* does not hold for the number and dimensions of half-irreducible representations. These representations of the symmetry groups of the 2D space are collected in Table I.5.

If we are restricted to real representations, *Lemma I.4* can be improved in a straightforward manner:

Lemma I.9: If D is a half-irreducible, d-dimensional, nontrivial real representation and $\mathbf{v}\neq 0$ *is a d-dimensional real vector, the orbit of* $\mathbf{v}$ *with respect to D spans* $\mathbb{R}^d$.

Similarly, one could replace the word 'irreducible' to 'half-irreducible' in *Lemma I.5*, however this would weaken its statement. Thus, we apply *Lemma I.5* and the consequent *Lemma I.8* in their original form for real-valued representations.

symmetry group		half-irreducible representations		
name	order	name	dimension	character
trivial	1	I_0	1	{1} (trivial representation)
C_n if n is even	n	I_0 S_l $I_{n/2}$	1 2 1	$\{a_k: 1\}$ (trivial representation) $\{a_k: 2\cos(2kl\pi/n)\}$ with $1\le l\le n/2-1$ $\{a_k: (-1)^k\}$
C_n if n is odd		I_0 S_l	1 2	$\{a_k: 1\}$ (trivial representation) $\{a_k: 2\cos(2kl\pi/n)\}$ with $1\le l\le (n-1)/2$
D_n if n is even	$2n$	I_0 I_l $I_{n/2}$ $I_{n/2+1}$ $I_{n/2+2}$	1 2 1 1 1	$\{a_k: 1, b_k: 1\}$ (trivial representation) $\{a_k: 2\cos(2lk\pi/n), b_k: 0\}$, with $1\le l\le n/2-1$ $\{a_k: 1, b_k: -1\}$ $\{a_k: -1^k, b_k: 1\}$ $\{a_k: -1^k, b_k: -1\}$
D_n if n is odd	$2n$	I_0 I_l $I_{(n+1)/2}$	1 2 1	$\{a_k: 1, b_k: 1\}$ (trivial representation) $\{a_k: 2\cos(2lk\pi/n), b_k: 0\}$, $1\le l\le (n-1)/2$ $\{a_k: 1, b_k: -1\}$

Table I.5: Half-irreducible representations of the symmetry groups of Table I.3. The meaning of notations a_k, b_k can be found in Table I.3. Notice that the only difference compared to Table I.4 is that some pairs of one-dimensional irreps of the cyclic groups are replaced by two-dimensional half-irreps. The names of the representations is *I* if they are irreducible (cf. Table I.4) and *S* if they are reducible but half-irreducible.